Effects of Illumination, Texture, and Motion on Task Performance in 3D Tensor-Field Streamtube Visualizations

Devon Penney *
Brown University

Jian Chen†
University of Southern Mississippi

David H. Laidlaw‡
Brown University

Abstract

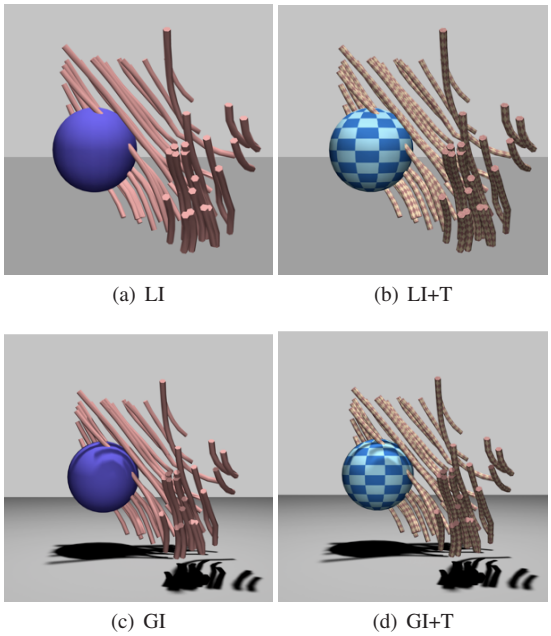
We present results from a user study of task performance on streamtube visualizations, such as those used in three-dimensional (3D) vector and tensor field visualizations. This study used a tensor field sampled from a full-brain diffusion tensor magnetic resonance imaging (DTI) dataset. The independent 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, and (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 reduced the errors in participants' answers over local OpenGL-style rendering for the visual-tracing task only when motion was present. Motion contributed 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 can help in the design of lighting model, such as flow or diffusion-tensor field visualizations and identify situations when the lighting is more efficient and accurate.

Index Terms: H.1.2 [Models and Principles]: User/Machine Systems-Human information processing—; H.5.2 [Information Interfaces and Presentation]: User Interfaces-Theory and methods—

1 Introduction

Three-dimensional (3D) streamtubes are popular for visualizing tensor and vector fields to provide valuable information about the underlying physical phenomenon [23]. However, dense streamtubes suffer from visual cluttering, which impedes identification of specific tracts of interest and slows user interaction. Accordingly, efforts have been made to improve structure illumination and rendering by adding specular reflections [12] or increasing visual realism [1]. While psychophysical studies have demonstrated that such illumination methods can enhance shape and depth perception and generate visually appealing interreflections (light reflecting diffusely from one surface onto another) and shadows in relatively simple scene settings [18], their effects have not been studied extensively from real-world task contexts. Thus, this work studies how illumination models can depict 3D scenes effectively for users to accomplish their tasks. 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.

We explore the effect of illumination models, motion, and texture on a set of tube visualization tasks. We compare two illumination models (Figure 1): global illumination (GI) and an OpenGL-based local illumination model (LI). GI simulates the complex behaviors of light around the model being viewed and can create realistic renderings [1]. LI is a baseline method for comparison. Here we also measure motion and texture, because these are believed to convey information effectively [24] and have been used in many platforms, such as ParaView for tube-based visualizations. By combining these factors, results from our study can give a picture of how real-world visualizations work under various display conditions. Our experimental approach was not to compare the entire system but to isolate those factors that would lead to better visualization task performance.



(a) LI (b) LI+T

(c) GI (d) GI+T

Figure 1: The four rendering styles studied, left to right: (a) OpenGL-style rendering without texture (LI), (b) LI with texture (LI+T), (c) global illumination rendering without texture (GI), and (d) global illumination rendering with texture (GI+T). Our study included conditions based on these four rendering styles shown both with and without motion. The sphere can be a proxy for a tumor in one of the tasks. The scene is also one of the small datasets used in the study.

Our experiment has three spatial analysis tasks: a depth-judgment task, a visual-tracing task, and a contact-judgment task. Each corresponds to tasks derived from dense tube visualizations. For example, the depth-judgment tasks are often used in psychophysical studies to measure depth perception given different visual cues; the visual-tracing tasks is analogous to an advection task in dense flow field [3]; the contact-judgment task can be used to evaluate a tumor in diffusion tensor magnetic resonance imaging (DTI) bundles. Independent variables include illumination model (LI and GI), motion (with and without), texture (with and without),

*e-mail: dmp@cs.brown.edu
†e-mail:jian.chen@usm.edu
‡e-mail:dhl@cs.brown.edu

and task type. Dependent variables include 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 aiding data understanding.

Our work makes several key contributions. We offer results from a comparative study on how illumination models, motion, and texture affect task completion time, error rates, and accuracy in 3D environments, excluding other factors such as interactivity. Our study also contributes to an understanding of visual cues, which could inform other kinds of dense-tube rendering such as general 3D vector field visualization and other tasks involving examining curves in 3D space.

2 Background and related work

2.1 Dense tube visualization

3D tensor and vector field visualizations often use tubes [23]. For example, DTI visualizations have 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 [27]. When applied to DTI, these streamlines or streamtubes or tensor glyphs 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, getting insights into dense datasets can be difficult.

There are at least two ways to improve visualization of these dense datasets: decluttering by clustering fiber bundles and conveying structures via perceptual principles. Clustering reduces the enormous quantity of individual fibers to a number that can still convey anatomical meaning and can be understood visually. By improving structure, a visualization system often constructs an illumination model to show shadows of the fibers, or allows users to query their data interactively [28]. While Moberts et al. [14] compare different clustering algorithms and Forsberg et al. [3] compare vector field visualization methods, we study structure from the perspective of illumination approaches.

2.2 Illumination model and relevant cues

Several illumination techniques are common in flow or tensor field rendering. Among them, lighting has profound effects on spatial understanding. For example, Zockler designed illuminated lines to enhance depth [28] and Interrante designed transparent surfaces, texture, and halo to convey shapes [5, 6]. Among illumination models, LI 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 simulates the behavior of light throughout the scene in order to increase visual realism [8]. For example, a photon-mapping algorithm can produce interreflections and soft shadows [1]. Lindemann and Ropinski compared seven volume illumination techniques on three depth and size perception tasks [11]. The results indicated that the more advanced lighting approximations, the better the task performance. Their scenes were static and the shapes did not have as many occlusions as we would encounter in dense tube visualizations.

Much evidence suggests that cues generated from GI by revealing spatial structure and orientation among surfaces, allow more accurate shape discrimination. For example, Weigle and Banks showed that GI was beneficial in visualizing highly dense tubes for detecting boundary shapes and reduced error rates in depth-judgment tasks, especially when a perspective view was used [25]. Our study advances the previous work by using experimental settings with more tasks requiring both local and global shape understanding. Also, the scenes were taken from DTI sub-volumes with three-level visual complexities, which produce tubes with more curvature than the tubes used in [25] and thus were more visually challenging.

2.2.1 Shadow and interreflection

Two types of shadows, extrinsic and intrinsic [9], can be generated via illumination models. Intrinsic shadows, also called shading, are the shadows an object makes on itself (Figure 1). They have long been used by artists and understood by psychologists to provide information on the convex or concave shapes of objects and the direction of illumination in a scene [9]. Such convex and concave shapes are also much more recognizable by human observers [15]. Extrinsic shadows, 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 [22]. For the purposes of our study, LI and GI can both produce intrinsic shadows on individual tubes, but only GI can produce extrinsic shadows between tubes.

It is generally agreed that shadows increase reported task accuracy. For example, Thompson et al. analyzed how shadow and interreflection affect depth perception [22]. They found 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 perception of spatial structure. However, Hubona et al. found that the effects of shadows were task-dependent [4]. Shadows enhanced the accuracy but not the speed for the object positioning tasks; But for the object resizing, the shadows effects is neither signficant on the accuracy nor the speed. Shadows are also subtle enough to distort the perception of 3D shapes [15].

Another difference between LI and GI is that GI implementations support interreflection. The light that ultimately reaches the eye and affects the image bounces more than once. In this way, surfaces not directly facing the light can be illuminated by other surfaces. Specifically to tensor field visualizations, Banks introduced the idea of maximizing the reflected light over the perimeter of an infinitesimally thin cylinder, treating diffuse and specular reflection separately. Hardware solutions to maximize reflection illumination modes have been applied to diffusion tensor imaging by Wenger et al. [26]. Obert et al. addressed the aesthetic drawbacks of the inflexibility of GI by creating a relighting tool for CG film making [16]. While many studies have focused on algorithm design, our goal here is to provide objective results of the impact of illumination method on task performance. We asked aesthetic effects in the post-questionnaire to collect some subjective comments on rendering approaches.

2.2.2 Texture and motion

Texture is also an important factor in the understanding of tube structures. It has been suggested that "lines that follow the form" convey shape effectively [6]. We use results of Jianu et al. to construct diamond textures that follow tube geometry [7], because that pilot study showed that texture could convey orientations of vector and tensor fields.

Motion is known to be a powerful cue to improve spatial understanding. Sollenberger and Milgram [20] demonstrated the utility of motion parallax in visualizing complex simulated blood vessel structures in the brain. Ware and Franck [24] found that motion produced by head-tracking had effects as powerful as stereo viewing on task completion times. We also wish to understand how the effects of motion and illumination model are combined.

3 Experimental Design

The primary purpose of this study was to explore the effects of illumination model, motion, texture, and scene complexity on task per-

formance. 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 $2\times2\times2\times3\times3$ within-participant design with the following independent variables: illumination model (LI and GI), motion (presence and absence), texture (presence and absence), scene complexity (small, medium, and large), and task (depth-judgment, visual-tracing, and contact-judgment). We did not include more factors (e.g., interactivity) in this study to avoid lengthening the experiment and introducing fatigue that might confound the results.

Each participant performed 48 tasks, 16 instances of each of the three tasks. Task order was randomized for each participant. 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 and scene complexity

Our 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.

We selected a range of bounding volumes to make a set of scenes possible. The bounding volume of the full model was 217.6 mm $\times$ 217.6 mm $\times$ 153 mm. The box edge lengths were 8.8 mm, 14.08 mm, and 24.64 mm, each containing 30-50, 100-150, and 300-560 tubes respectively.

Most algorithms seed tubes using distance measures, such as that [27] used in our study. We did not control the number of bundles in each sub-volume explicitly but took random samples based on the three edge lengths. Our general observation was that the larger the volume, the greater the number of bundles. By doing this, we created three levels of scene complexity: small, medium, and large.

3.1.2 Illumination model

We used two rendering algorithms for this study: LI and GI. For LI 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 1600×1600, rendered in perspective. We carefully tuned the rendering to ensure important cues are present in GI (e.g., interreflection and projection) for depth judgment.

We used a traditional three-point lighting scheme plus several fill lights. Rim and key lights were placed in relation to a preset camera with 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. 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. Small changes in light placement and intensity can yield very different images. Our purpose here was to produce sufficiently good images for the purpose of testing.

We chose to study LI and GI because GI can produce a realistic scene [1] with most types of shadow present and LI was used as a baseline method for comparison. Although numerous studies have suggested that shadows cast on the ground improve depth judgment [4], the advantages of other types of shadows, such as interreflection in GI, were still unknown. Our goal was to study rendering in general, but not good and specific lighting approaches, such as local ray casting [17] or illuminated lines [28].

3.1.3 Motion

Motion was synthesized using a sequence of 23 images (110-degree viewing angle), each image sequence accounting for ±5 degrees of rotation in the local dataset coordinates about the vertical axis in the image plane. We did not allow free-form interaction because GI's high computational costs prohibited real-time viewpoint-based rendering. The choice of the ±5 degrees was made after a pilot study where participants reported that structure understanding was not jeopardized by the discrete views. The frame rate permitted relatively continuous motion between frames at a frame rate around 30 fps. The motion was automatic for the depth-judgment and contact-judgment tasks. The visual-tracing tasks used the arrow keys to traverse the frames.

3.2 Tasks and user interface

Three tasks were used in our study, all involving spatial understanding of the underlying geometrical tube structures from the data. All tasks were performed using a Dell 3007WFP monitor running at a resolution of 2560 $\times$ 1600 screen pixels. They are: (1) Depth-judgment (Figure 2(a)), e.g., which tube is closer, the blue or the green tube? (2) Visual-tracing (Figure 2(b)), e.g., where is the endpoint of the tube? and (3) Contact-judgment (Figure 2(c)), e.g., does the blue sphere intersect, touch, or have no contact with the tubes?

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

3.2.1 Depth-judgment task

Participants were shown a blue and a green tubes embedded in a dataset (Figure 2(a)), and were asked to report which tube was closer to their viewpoint. The tubes were selected at random with the constraint that they did not occlude each other from any sampled viewpoint. The distance between the two tubes was computed by sampling on a fixed interval across the dataset rotation range. We called this a global visualization task because making correct judgments may require visual scanning of neighboring tubes. Only datasets with a clear front and back when viewed in the 23 discrete frames were chosen for this task.

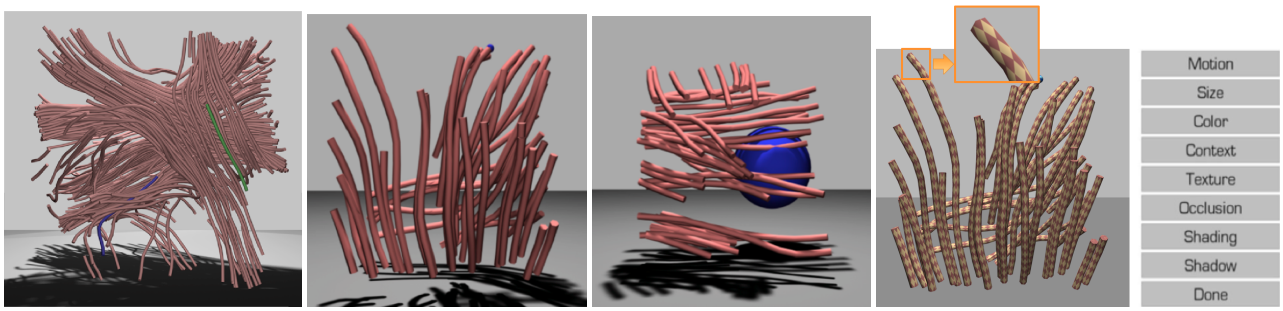
3.2.2 Visual-tracing task

This task required participants to trace a randomly selected tube with a blue sphere on one end and to mark the other endpoint (Figure 2(b)). The constraint for 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). Thus 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 in the motion condition. We did not use automated motion, because tracing a tube in a dynamic scene could be difficult and impractical. We called this a global visualization task because it also required visual scanning of neighboring tubes, as in the depth-judgment task.

We chose this task because DTI requires the understanding of bundles and flow requires the understanding of the direction of streams, for example an advection task in a flow field. This task is also popular in cognitive psychology to study human visual attention behaviors, which could be relevant to 3D vector and tensor field visualization.

3.2.3 Contact-judgment task

Participants were asked to judge whether and how the blue sphere intersected the tubes 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 (Figure 2(c)). Tangent means that the sphere grazed the tube(s) but did not fully intersect; full intersection means that the sphere intersected the tube(s). The sphere



(a) Depth-judgment task (b) Visual-tracing task (c) Contact-judgment task (d) User interface

Figure 2: The three tasks studied and the UI used. (a) Depth-judgement 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. (b) Visual-tracing task: participants were asked to trace a tube and mark its endpoint. When motion was enabled, participants could rotate the dataset using the arrow keys on the keyboard. (c) 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. (d) 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.

was randomly placed in the dataset with equal distribution among the three cases. We called this a local visualization task because participants needed to examine only the tubes around the sphere. GI is considered to show spatial relationship without rotation and this task places GI in a real-world context.

3.2.4 Cue choices and user interface

An example user interface is shown in (Figure 2(d)). The task dataset was displayed in the left window. After each task, participants identified the cues that were useful in performing that task, shown on the right side of the screen. The purpose of this step was to determine the perceived usefulness of cues and to find correlations between perceived usefulness and task completion time. We have included eight types of cues including motion, size, color, context (surrounding fibers), texture, occlusion, shading, and shadow. By context cue, we refer to the tube relationship with respect to its location in the environment and neighbors. These relationships should be appropriate and consistent with the vector field visualization experience: tubes tend to be seen grouping to bundles, are often exist among other bundles or tubes, do not intercept (in our cases), and the bundles are usually in proximity to each other and may have other properties in common. Participants were told what these cues were during the training session to avoid any confusions during the formal experiment.

3.3 Dependent variables

The dependent variables include task completion time and error rate. The criteria to measure error or correctness were: for depth-judgment and contact-judgment tasks in which the answer were binary choices and the correct answers were those that matched the true spatial relationship. For the visual-tracing task, the correct answers were measured by the distance between the marker the participant made and the true target location on the image plane. Distances within a threshold of 50 pixels in the screen coordinate (which was about 1.5 tube width) were marked as correct answers.

3.4 Participants

Twenty-six volunteers (11 males and 15 females) participated in this study. All were Brown University undergraduate and graduate students. The participants' areas of study were biomedicine (5), biology (5), neuroscience (3), geology (3), engineering (3), applied math(2), linguistics(2), history (2), and computer science (1). All participants reported correct vision and color vision. Two had extensive experience with human anatomy. Running 26 participants allowed us to collect data on all combinations of the conditions on all participants.

3.5 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 given 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 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 (8-task each), thus forming a total of 48 unique tasks. Participants were asked to balance the efficiency and accuracy. The cue question was answered after each task. Participants were told to finish the tasks as quickly and accurately as they could. Finally, participants' responses were collected in a post-questionnaire about the perceived usefulness of the proposed rendering techniques, task difficulty, and the aesthetics of the rendering schemes. Participants were also interviewed for additional comments.

4 Results

4.1 Statistical method and summary statistics

We performed a within-subject GLM (general linear model) procedure on illumination models (LI and GI), texture (presence and absence), motion (presence and absence), and task type (depth-judgment, visual-tracing, and contact-judgment). We measured F and p values [10]. When there was a significant main effect, we separated the levels to study the sources of the significance. Posthoc analysis of Tukey pairwise comparison among dependent variables and Tukey's Studentized Range (HSD) test were also used, which compared all possible pairs among the levels to measure. Those without differences were put in the same group. For those aggregated analysis (GI + motion), we used t-test to compare the mean task completion time or error rate. All analyses were conducted using SAS (statistical analysis software).

We analyzed each task separately because we were not concerned with task differences and because the task completion time

Table 1: Summary of differences as measured by task completion time and accuracy (or error rate) for the three tasks. Only statistically significant differences are listed. Notations are as follows: illumination model is denoted as GI (for global illumination model) or LI (for OpenGL-style illumination model); Motion is denoted as "motion" (M) when it is present; Texture was denoted as "texture" (T), otherwise, "no texture". The notation A>B indicates that method A was significantly more efficient or effective at the task than method B for the metric label at the top of the column. All graphs and tables in this paper use this same notation system.

Task	Completion Time	Accuracy
Depth-judgment	LI>GI NM>M LI>GI+M	M>NM
Visual-tracing	NM>M NT>T	M>NM
Contact-judgment	NM>M	M>NM

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

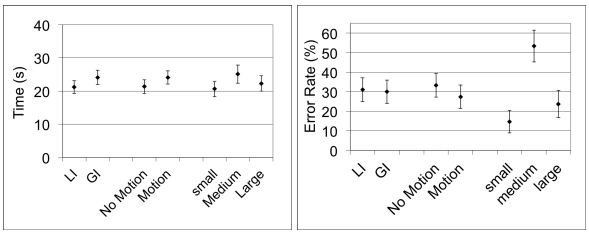
Task	Illumination	Motion	Texture	Complexity
Depth judgment	$p = 0.9$ $F = 0.001$	$p = 0.08$ $F = 3.0$	– –	p<0.0001* $F = 30.9$
Visual-tracing	$p = 0.04*$ $F = 4.3$	$p = 0.03*$ $F = 10.2$	$p = 0.07$ $F = 3.34$	p<0.0001* $F = 60.3$
Contact judgment	$p = 0.96$ $F = 0.01$	$p = 0.07$ $F = 3.36$	$p = 0.81$ $F = 0.05$	p<0.0001* $F = 59.5$

for different tasks was significant ($p = 0.0005$, $F(15,1184) = 12.07$). We conducted outlier detection using histograms on dependent measures. For those cases with skewed normal distributions, we removed outliers at the 99% percentile. We removed five outliers (leaving 446 observations), five outliers (leaving 446 observations), and six outliers (leaving 445 observations) from the three tasks accordingly. Texture was always present in the depth judgment 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. We compared the treatment within each main effect, as shown in Table 1. We measured the F and p values from the GLM procedures, shown in Tables 2 (on error rate) and 3 (on task completion time).

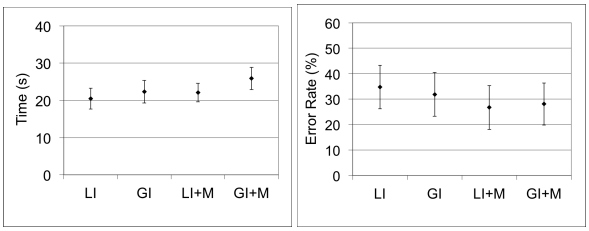
4.2 Depth-judgment task

Figure 3(a) illustrates a comparison of task execution time. The results indicates that LI outperformed GI, no motion outperformed motion, and small complexity outperformed large and medium complexity. Participants' task completion time was also significantly different. The HSD test for illumination model revealed that LI and GI were in different groups. By separating the presence and absence of the motion condition, we found the differences between LI 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 four conditions (Figure 3(c)), LI and GI+M were the only pairs in different groups.

The main effects of motion and illumination model on error rate are shown in Figure 3(b) and the summary statistics in Table 2. We found no significant difference in the main effects on error rate. No significant differences were found between LI, GI, LI+M, and GI+M (Figure 3(d)). We observed that the higher the complexity, the higher the error rate. The smallest box size led to a 14% error rate compared to 24.8% and 52% for middle- and large-size boxes.



(a) Time vs. independent variables (b) Error rate vs. independent variables



(c) Time vs. rendering style (d) Error rate vs. rendering style

Figure 3: Depth-judgment task: Effect of illumination model, motion, and scene complexity on task performance.

Table 3: Summary of the F and p values for the main effects on task completion time. "*" indicates the significant main effects.

Task	Illumination	Motion	Texture	Complexity
Depth judgment	$p = 0.75$ $F = 0.1$	p<0.0001* $F = 18.5$	– –	p=0.002* $F = 6.2$
Visual-tracing	p=0.13 $F = 3.3$	p=0.003* $F = 8.8$	p=0.005* $F = 7.9$	p<0.0001* $F = 90.1$
Contact judgment	$p = 0.5$ $F = 0.5$	p<0.0001* $F = 22.8$	$p = 0.51$ $F = 0.4$	$p = 0.34$ $F = 1.1$

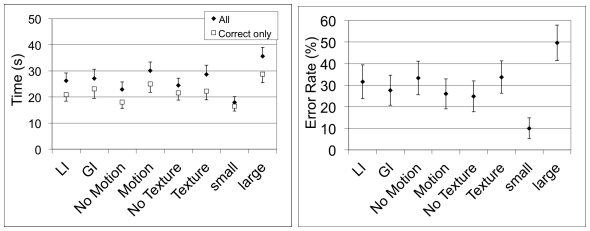
4.3 Visual-tracing task

The effects of illumination model, motion, texture, and scene complexity on task performance are shown in Figure 4. The F and p values are shown in Table 3. Motion and texture, all had a significant impact on task performance. With-motion had longer task-completion time than no-motion (mean=22.9s vs. 30.1s) (Figure 4(a)). With-texture also led to longer task completion time (mean=24.47s vs. 28.7s). The illumination model was not significant and LI showed slightly shorter task completion time (mean=26.3s vs. 27.1s). The main effects on error rate are shown in Figure 4(b) and the F and p values in Table 2. Illumination model, motion, and scene complexity had a significant impact on error rate.

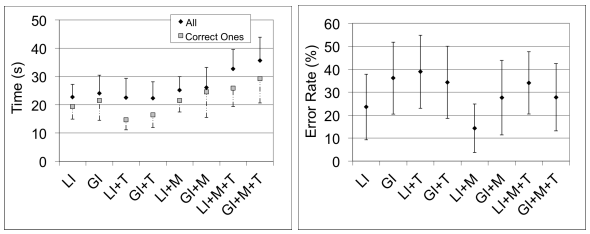
We measured the combined effect of rendering, motion, and texture (Figures 4(b) and 4(d)). The HSD test on task completion time suggested that GI+M+T was in a different group from LI, GI, LI+T, and GI+T accordingly, when all observations were used. LI+M+T and GI+M+T were in different groups from GI+T and LI+M when correct-only observations were used. LI+T led to the shortest task completion time, followed by GI+T (Figure 4(c)). All eight conditions were in the same group for error rate, though LI+M led to the lowest error rate (Figure 4(d)).

Accuracy was measured as the distance between the pointer and the true target location on the image plane. Distances within a threshold of 50 pixels in the screen coordinate (which was about 1.5 tube width) were marked as correct answers. We found scene complexity was significant ($p < 0.0001$). No other significant main effect (illumination model, motion, or texture) on accuracy was observed.

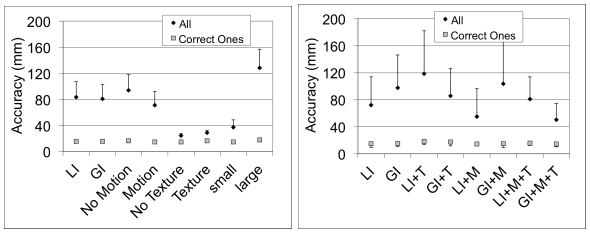
Also, motion had a significant impact on accuracy, leading to



(a) Time vs. independent variables (b) Error rate vs. independent variables



(c) Time vs. rendering style (d) Error rate vs. rendering style



(e) Accuracy vs. independent variables (f) Accuracy vs. rendering style

Figure 4: 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. (b) Illumination model, model and scene complexity had a significant impact on error rate. (c) Combined effect: GI+M+T was in different groups from LI, GI, LI+T, and GI+T accordingly. (d) Combined effect: LI+M led to the lowest error rate. (e) Only scene complexity was significant. (f) GI+M+T led to the most accurate answers.

higher accuracy when all observations were used (Figure 4(e)). The full-cue condition led to the longest task-execution time and the most accurate answers (Figure 4(f)).

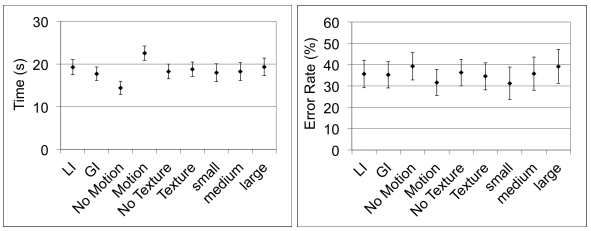
4.4 Contact-judgment task

One significant main effect was motion ($p < 0.0001$, $F(1,384) = 47.22$). Again, no motion led to shorter task completion times (mean=15.0s vs. 23.8s) (Figure 5(a)) and lower error rates (Figure 5(b)). None of the main effects was significant on error rate.

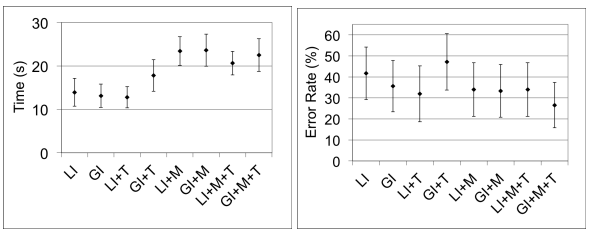
We also measured the combined effect of illumination model, motion, and texture on task completion time (Figure 5(c)) and error rate (Figure 5(d)). The HSD test suggested that each of LI+M, GI+M, LI+M+T, and GI+M+T was in different groups from LI, GI and LI+T. The full-cue condition (GI+M+T) had the lowest error rate, although task completion time was longer. LI+T had intermediate task completion time and intermediate error rate.

4.5 Subjective comments

Participants were asked in a post-questionnaire 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, with



(a) Time vs. independent variables (b) Error rate vs. independent variables



(c) Time vs. rendering style (d) Error rate vs. rendering style

Figure 5: Contact-judgment task: effect of illumination model, motion, texture, and scene complexity on task performance. (a) and (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.

1 being the worst and 7 the best. The perceived usefulness (subjective comments on the usefulness for the tasks studied in the experiment) of the image increased with the realism of the rendering, i.e., LI (3.27) < LI+T (3.78) < GI (4.29) < GI+T (5.27). The scores for aesthetic beauty were the reverse of the texture categories: LI + T (3.57) < LI (3.62) < GI+T (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 task 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 do not agree with the expectations that GI, motion, and texture would improve task completion time. Our data do support that motion is a strong cue to reduce error rate and improve the accuracy of visual-tracing. Interestingly, our results reveal that the accumulated cues conditions led to longer task completion times and often the lowest error rates.

5.1 Illumination model effects are task dependent

No effect of illumination model on task completion time was observed. LI and GI differed only when motion was present for the depth-judgment and 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. The significance again occurred when motion was presented, possibly indicating that using GI and motion together could led to better judgment. In general, GI slightly reduced error rates compared to LI, but at the cost of longer task completion time.

These results support only partially our hypothesis that GI would improve task performance. Because of the small reduction in 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 be more important in the contact-judgment task because interreflections could suggest orientations and distances between the sphere and the tube [1]. However, this effect was not supported in our study.

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 may have been difficult for participants to find which tube cast which shadow and where, especially when large numbers of tubes were presented. Our informal observation for images like Figure 1 suggests that the shadows cast on the ground in the GI conditions might help recognize large-scale structures such as bundles. However, such shadows might not support visualization at a fine level examining individual tubes. Finally, the GI effects depend on our parameter setup. For example, the ambient light can be optimized. The contributions of shadows to real-world scene understanding merit further study by altering parameters that would vary scene contrast and ambient lighting.

Another explanation is that LI shading used in our experiment produced salient depth effects that might be suffice to help participants perform tasks. Such an effect can be observed by comparing the image quality in our rendering and the ones studied by Weigle and Banks [25]. The image contrast in the visualizations produced using LI and GI was similar given GI no advantage. 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 was supported by our results at the cost of lengthened task completion time. We believe the time cost of motion is mostly due to the motor actions for changing views. The benefits of motion likely came 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 The importance of designing textures

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 the dataset used had many tubes, the entire scene was reduced to a single mass of diamond-textured shapes, that failed to convey the spatial structure. Participants had great difficulty in detecting a tube's direction when the context had similar frequency content, which causes spatial masking. In addition, the diamond shapes may have been poor at conveying orientation because they 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, so as to distinguish the two textures by at least 30 degrees [2]. Another way would be to orient the texture to near-horizontal or near-vertical lines [13] or make it follow the principal curvature directions. Studying texture design and use might be an interesting and challenging future direction.

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 depth-judgment task, a useful texture could be one that presents the relative component depth. When textures were presented in the visual-tracing and contact-judgment tasks, participants examined them in greater detail. Texturing methods such as dots [19], as proposed for vascular structure, are worth exploring within the context of different rendering methods, as they have been proven to improve task performance.

Shading was 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-tracing and contact-judgment (Figure 5(d)). We would expect shading and shadow to be used for the contact-judgment task because both convey relational properties. We did not observe this effect, probably because the large, round sphere did not require detailed visual examination. Another explanation may be that the scene complexity caused less clear inter-object relationships.

5.4 Scene complexity impacts performance for global tasks

Scene complexity was measured by the size of the bounding volume in the visualization. Three levels were used. Interestingly, the scene complexity had a significant impact on both task completion time and error rate in the depth-judgment and visual-tracing tasks, but not in the contact-judgment task. An intuitive explanation is that contact-judgment is a local task that requires only 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 the larger field of view introduced by the display hardware. For example, the visual-tracing task is simple when no occlusion occurs and when following the trace of the tube does not requires switching context. In these conditions, eyeball movement suffices to accomplish the task. However, head movement may be required when the display becomes larger and the tube crosses a large portion of the screen. The effort of changing gaze could introduce a greater mental and cognitive load and thus introduce time penalties. The error rates were especially high for the visual-tracing task. However, the large scene complexity had lower error rates than the medium sized ones for the depth-judgment task. This is partially because the distances between the tubes were greater, thus making size cues more salient in the large conditions.

5.5 Differences among tasks

The visual cues had a significant impact on task performance and error rate for global tasks. Our results suggested that special attention should be paid to supporting these kinds of tasks. An alternative 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 design control would be 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 affording such results are difficult to implement effectively.

5.6 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" [22]. The shadow in question was probably too dim to be a useful cue because of an overuse of fill lights and ambient illumination calculated from GI, which is a by-product of insufficient lighting design. An interesting future direction might be to use the non-photorealistic rendering (NPR) approach to explicitly visualize the occlusion effects [19]. This method could make the "dim" shadow to become clearer. Another interesting approach that might be worth comparison is to combine chromatic shadows [21] which could enhance depth and surface perception.

6 Conclusion

We have presented the results of a user study comparing task completion time and error rates for streamtube rendering tasks using illumination method, texture, and motion as the independent variables. The coding error was unfortunate, but does not reduce the utility of the data that was coded correctly, particularly for the visual-tracing and contact-judgment tasks. Nonetheless, our results do not agree with Weigle and Banks's study [25], broadening the basis for future work that explores this large design space. Specifically, (1) numerous psychophysical studies support that idea that GI can aid 3D visualization by providing shadow and interreflection for recovering 3D information from 2D screen images. Our study suggested that GI provided such benefits in a limited fashion with visually complex visualizations. Considering the computational cost of real-time rendering, GI could be detrimental in real-world tasks compared to other stronger 3D cues such as motion. Thus, 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. (2) 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 may be counterproductive because the scene could appear flattened and two-dimensional [5].

Our results suggest that, in general, the simpler stimuli (i.e., those with LI, 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. In conclusion, we did not find the benefits of using the GI model compared to the LI model, perhaps due to the high scene complexity from the real-world data and task settings. From a real-world application standpoint, the results from the experiment, most significantly those with motion, imply that interactivity can be a more effective means to design a visualization environment, where motion can be explicitly controlled by the user. There are also long-term possibility for studying the effects of shading cues to balance between LI and GI to find the optimal point to increase accuracy while maintaining interactivity.

Acknowledgements

The authors wish to thank the participants for their time and effort and the anonymous reviewers for their helpful remarks. This work was supported in part by NSF (CCF-1785542, IIS-1018769, and IIS-1016623), NIH (RO1-EB004155-01A1), and by a Brown University Center for Vision Research Fellowship. The authors would also like to thank David Banks and Chris Weigle for discussion of the tasks and data analysis and Katrina Avery for her excellent editorial support.

References

[1] D. C. Banks and C.-F. Westin. Global illumination of white matter fibers from DT-MRI data. *Visualization in Medicine and life sciene*, pages 178–234, 2007.

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

[3] A. Forsberg, J. Chen, and D. Laidlaw. Comparing 3D vector field visualization methods: a user study. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1219–1226, 2009.

[4] 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. *ACM Transactions on Human-Computer Interaction*, 6(3):214–242, 1999.

[5] 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.

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

[7] 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.

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

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

[10] R. Larsen and M. Marx. *An introduction to mathematical statistics and its applications*. Prentice Hall, 2000.

[11] F. Lindemann and T. Ropinski. About the influence of illumination models on image comprehension in direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):1922–1931, 2011.

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

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

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

[15] J. F. Norman, Y. Lee, F. Phillips, H. F. Norman, L. R. Jennings, and T. R. McBride. The perception of 3-D shape from shadows cast onto curved surfaces. *Acta Psychologica*, 131:1–11, 2009.

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

[17] T. Peeters, A. Vilanova, and B. Romeny. Interactive fibre structure visualization of the heart. In *Computer Graphics Forum*, volume 28, pages 2140–2150, 2009.

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

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

[20] 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.

[21] V. Šoltészová, D. Patel, and I. Viola. Chromatic shadows for improved perception. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*, pages 105–116, 2011.

[22] 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.

[23] 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.

[24] C. Ware and G. Franck. Evaluating stereo and motion cues for visualizing information nets in three dimensions. *ACM Transactions on Graphics*, 15(2):121–140, 1996.

[25] 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. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1723–1730, 2008.

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

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

[28] 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.