Comparing 3D Vector Field Visualization Methods: A User Study

Andrew S. Forsberg, *Member, IEEE*, Jian Chen, *Member, IEEE*, and David H. Laidlaw, *Senior Member, IEEE*

Abstract—In a user study comparing four visualization methods for three-dimensional vector data, participants used visualizations from each method to perform five simple but representative tasks: 1) determining whether a given point was a critical point, 2) determining the type of a critical point, 3) determining whether an integral curve would advect through two points, 4) determining whether swirling movement is present at a point, and 5) determining whether the vector field is moving faster at one point than another. The visualization methods were line and tube representations of integral curves with both monoscopic and stereoscopic viewing. While participants reported a preference for stereo lines, quantitative results showed performance among the tasks varied by method. Users performed all tasks better with methods that: 1) gave a clear representation with no perceived occlusion, 2) clearly visualized curve speed and direction information, and 3) provided fewer rich 3D cues (e.g., shading, polygonal arrows, overlap cues, and surface textures). These results provide quantitative support for anecdotal evidence on visualization methods. The tasks and testing framework also give a basis for comparing other visualization methods, for creating more effective methods, and for defining additional tasks to explore further the tradeoffs among the methods.

Index Terms—3D vector fields, visualization, user study, tubes, lines, stereoscopic and monoscopic viewing.

---

1 INTRODUCTION

Little data exists about the relative merits of 3D vector visualization methods evaluated in the context of real-world tasks. This lack is recognized as a top visualization challenge [20] since knowledge from evaluations would be extremely useful for working scientists and visualization researchers. Our aim here was to conduct a controlled study that contributed specific results and was extensible.

We present an instance in a planned series of studies that investigates the problem of formally evaluating visualization methods. We intended to simulate an exploration scenario in which the scientist did not know in advance the answer or ideal visualization parameters. Given a correct answer, visualization methods and parameters can often be selected that lead to a very effective visualization. Our interest here was evaluating visualization methods that could help scientists in the discovery process.

As in any study of this type, a number of our experimental design decisions might have made differently knowing the results. Nonetheless, we feel that the results have the potential to inform the development of effective visualizations and their evaluation.

The two main challenges in designing the study were defining "realistic scenarios" consisting of data and tasks, and choosing which visualization methods to investigate.

1.1 Choice of scenarios

Our approach to selecting tasks was to interview scientists and base our tasks on the kinds of visual search tasks they perform using their 3D vector fields. Sometimes we could extend the discussion of tasks by proposing 3D versions of the tasks in Laidlaw et al.'s earlier study of 2D visualization methods [13]; scientists would either confirm their relevance (e.g., tracing the path of a streamline) or explain other tasks more relevant to their work (e.g., a task of identifying swirling movements was not used in Laidlaw's work). For this controlled experiment requiring over a hundred unique datasets we could not use a scientist's actual datasets, but instead attempted to ensure that our synthesized datasets included features of research interest. We interviewed scientists studying arterial blood flow and bat flight, as well as more general flow problems. The tasks we defined test the ability either to deliver direction information, to convey magnitude information, or

both. Some tasks depend on local information around a particular 3D point, while others require more global information. Two tasks involve critical points (CPs). While CPs are not generally what 3D flow scientists study, they are reasonable choices for points of interest because, as commonly noted in our discussions with scientists, *the behavior of the neighboring vector field is interesting to understand and visualize.* Note that no single task represented all the vector-related scientific areas we considered- for example, gauging the speed of movement at a point is very different from characterizing the patterns of movement that might reveal the type of CP at a specific point. Participants completed several instances of five simple but representative tasks involving CPs, advection, swirling movement, and comparative speed. While these are not a complete set of vector visualization tasks (e.g., they do not test vorticity), they do cover an important range of vector field analysis tasks, as confirmed by other expert flow scientists who have participated in our study. Pilot studies have revealed that these tasks are challenging for both novice and flow experts.

1.2 Choice of visualization methods

We used prior work to help identify promising methods as well as methods motivated by those in production tools like TecPlot [1] and ParaView [2]. Specifically, we varied viewing conditions (stereoscopic and monoscopic) and integral curve rendering (lines and tubes) across five tasks. Many other visualization methods and parameterizations of visualizations (including user interaction styles) could have entered into the design; however, we had to make a careful design choice that limited the duration of the study for participants' comfort. For example, each additional visualization method added over an hour to the study, and our design already took 2.5 hours on average. We thus view this study as the first in a series, believing that evaluating first very frequently used visualization methods would provide a useful baseline. In addition, stereo displays are rare in visualization facilities, even though stereo viewing is an inherent human ability and helps understand 3D geometries [25][21]. The results of the work presented here and similar work may motivate the development of improved visualization methods based on stereo viewing and help facilities decide whether to make stereo displays available to scientists.

1.3 Hypotheses

Dependent variables were completion time, accuracy, confidence, and subjective responses. Our hypotheses were:

- The tube method would outperform the line method,
- Stereo viewing would outperform mono viewing, and
- The combination of tube and stereo representation would be best.

- *Authors with Computer Science Department, Brown University, E-mail: {asf,jchen,dhl}@cs.brown.edu .*

Fig. 1. A participant views 3D vector field visualizations at a stereo monitor, wearing active stereo shutter glasses for both stereoscopic and monoscopic visualizations. The keyboard is used both to rotate the dataset and to specify answers and confidence levels.

Our results indicate the methods vary in performance across tasks. The mean time was always faster for the line-stereo (LS) method. Some monoscopic methods were more accurate for "Type of CP?" and "Is swirling?".

Our main contribution is the results of a formal study comparing four visualization methods for five tasks.

2 RELATED WORK

2.1 Visualization methods

Many methods have been developed and combined to visualize 3D vector fields. All production-use visualization systems such as Tec-Plot [1] or ParaView [2] offer methods to visualize 3D curves using glyphs, lines, or surfaces. Other methods visualize 3D curves with textures and animation. Derived data like λ_2 (the second eigenvalue of the pressure gradient) [11] may be visualized to help find features of interest like swirl. Some workers discuss visualization needs with artists (i.e., expert visual communicators) who then help design an effective visualization [3]. Our work is related in that we aim to evaluate previously developed visualization methods.

2.2 Human-centered displays

Enhanced displays offering stereoscopic viewing, large fields of view, and high-resolution promise to make complex 3D data easier to understand by more closely impedance matching the human visual system [26][5][23][25]. Our work relates to this in that this study compares monoscopic and stereoscopic viewing.

2.3 Evaluation

Several kinds of evaluation of 3D vector visualization methods have been made. Some work has developed visualizations and tested them for specific applications [14]. Some evaluations seek to identify generally important tasks and evaluate multiple methods [13]; others compare visualization methods on different displays [21][6]. As a result of head-to-head comparisons with line representations, tubes have been recommended for representing 3D lines; for instance, Ware's earlier study [25] found that "even without stereo and motion depth cues, tubes allowed for surprisingly accurate judgments. Thus the strongest recommendation that comes from this study is that tubes should be used to render 3D pathlines or streamlines". Our work relates to different aspects of each of these examples in that we aim to evaluate methods for 3D vector field visualization and to verify prior results in a more realistic context and at a display representative of today's computer systems.
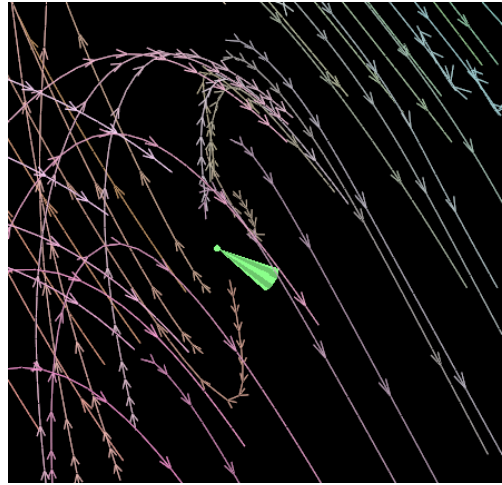

Fig. 2. A screengrab of the lines method: 1-pixel wide lines represent a set of integral curves seeded on a 4x4x4 regular grid for each dataset. Lines are colored so that similar curves have similar color. Arrow-like glyphs indicate direction and speed–the larger the gap between glyphs, the faster the vector field at that point.
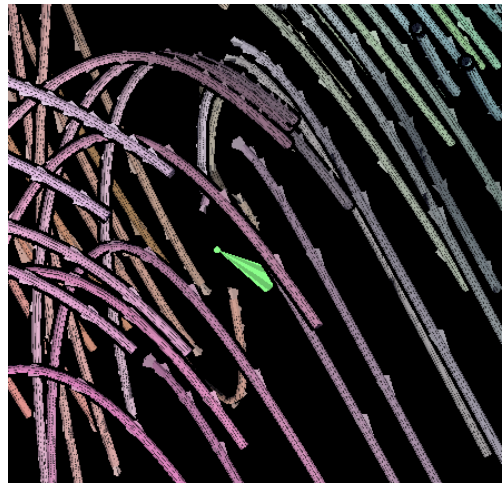

Fig. 3. Screengrab of the tubes method for the dataset in Figure 2. Surface shading, glyphs, surface texture, and "halos" around tubes provide spatial cues. Glyphs and arrows on the surface texture are spaced proportional to speed.

3 METHODS

3.1 Experimental design

We used a 2x2x5 within-participant design (i.e., each participant sees every experimental condition) with the following independent variables: method (line and tube), viewing condition (mono and stereo), and task (five instances).

3.2 Apparatus

The apparatus consisted of two stereo displays (Mitsubishi Diamond Pro 2070SB, 22" display, 1280x1024 resolution per display, approx. 83 dpi) arranged as a left and right monitor (see Figure 1). The frame rate was 15 fps. Each display has 1280x1024 pixels with a display area of 16 x 12 inches, giving an individual pixel size of 0.012 inches (0.03 cm). Head-tracking was not used in this study. Participants sat in a standard desk chair and viewing distance was 1.5 feet, although they were free to move their head closer or further away; we estimate that the viewing distance from the display could vary between 1 - 2 feet. Thus, the visual angle per pixel could vary between 3.44 minutes of

arc (1-foot viewing distance) and 1.72 minutes of arc (2-foot viewing distance).

Participants wore StereoGraphics active shutter glasses for both monoscopic and stereoscopic viewing conditions to keep display parameters like brightness constant. The left monitor displayed 3D visualizations and the right monitor was used for the "Type of CP?" trial to display the CPs shown in Figure 4 for reference.

Participants rotated the dataset about the X and Y axis relative to the screen using the cursor keys. Answers were given by typing on the keyboard. The room was darkened to provide a clearer view of the display, but indirect lighting helped illuminate the keyboard buttons when viewed through stereo glasses.

3.3 Conditions

Stereoscopic viewing: In the stereo viewing condition views were generated for each eye assuming an eye separation of 2.5 inches. For the monoscopic condition, the views presented to each eye were identical.

Integral curve renderings: All trials visualized a set of integral curves computed from a 4x4x4 array of seed points spaced regularly through the volume. Integral curves were computed using VTK's RungeKutta4 integrator [12] with propagation unit set to cell length, a maximum propagation of 100 units and a terminal speed of 0.1 units. Two rendering methods were used for the integral curves: a) 1-pixel wide lines and b) textured, lit tubes 8 pixels in diameter (see Figures 2 and 3, respectively). Both used anti-aliased rendering, and in both integral curves were colored so that curves with similar spatial position and shape had similar color [7]. In addition, each method draws geometric glyphs along the curve and spaces them proportionally to the magnitude or speed of the vector field at that location–the larger the gap between shapes, the faster the vector field along that segment of the line. The shapes themselves are arrow-like and indicate direction of movement along the line.

Tubes were shaded by two directional lights: light #1, a "headlight" always located at the viewer's eye position and shining in the direction of the virtual camera's viewing direction, and light #2, a light fixed in world-space at the location (-5, -5, -2) (at the start of each trial the light is "over the viewer's right shoulder"). (Note that the volume is centered at the world origin and has bounding volume (-1, -1, -1) to (1, 1, 1).) The tube surface was textured with an arrow texture that encoded the local speed and direction of the vector field. At a larger scale, "feathers" like those on an archer's arrow were drawn on the tubes to reflect speed and direction, just as the arrows do in the lines method. Finally, the tubes also had a black halo around them to emphasize their front-to-back ordering [22]. Pilot studies indicated that the halos on tubes were beneficial and no participants made negative comments about them. Tubes had a radius of 0.0115 (as explained in section 3.3.1) and halos were drawn by enabling OpenGL's front face culling and drawing each tube a second time with the radius of the tube doubled (i.e., a radius of 0.0230). Several radii were tested for the tube radius but pilot subjects noted that doubling the radius worked well.

3.3.1 Pilot studies

We ran a series of pilot studies to determine parameters for the seeding strategies as well as the visualization methods. In pilot studies we used the regularly spaced seeding method found to work best in [13]; we compared n x n x n arrays for n = 3, 4, 5, 6 and found n = 4 to be the best balance of clarity and information from streamlines. We did find that slightly different densities were best for different tasks so the density chosen was a compromise (practical constraints on study length required selecting a single density value). For the visualization methods, first one subject explored the parameter space and tuned variables like tube radius, halo radius, glyph spacing, glyph scale, texture shape, texture scale, texture spacing, etc. to his or her preferred settings. Then for each parameter we performed a "step wedge" by looking as a group of four at a range of parameters neighboring each setting selected by the subject. A step wedge is a controlled process for exploring the effect of some process on a range of inputs–for example, in image processing a step wedge could help see the effect of a

filter by applying the filter to a set of equally spaced gray values inputs and displaying the output values adjacent to each input value. Our step wedge explored not image processing filters but parameterizations of the visualization methods–our pilot subjects assigned a score to each parameter (e.g., seed spacing, tube radius, etc.) and we ultimately selected the average value for each parameter.

A difficult decision was what degree of interaction to allow, since while interactivity is a vital element of data visualization it can confound formal user study results. A highly interactive system allows many usage patterns, more training, and the development of different strategies by users. Because our main interest here was visual performance, we minimized user interaction. At first, participants could only rotate about the Y-axis, but since many requested rotation about the X-axis too it was added. Rotation was controlled through the cursor keys on the keyboard.

Tasks: Our five tasks all aimed at testing how well subjects understand "chunks of 3D flow". Informally, understanding chunks of 3D flow is the commonality we have noted from working on 3D flow tasks and developing visualization methods in collaboration with fluids researchers. While flow experts are often searching for different scientific features and often only indicate what they look for given a statement of the problem, our best general description of a flow scientist's task when visualizing 3D flow fields is that to varying degrees they all explore or study a localized point, often in considering the neighborhood around a region of interest. Finding and describing many common flow features fit this categorization including swirling flow, stagnation points, vortices, flow separation, flow reversal, and high-residence time. 3D flow scientists sometimes reduce their problems to 2D visualization or quantitative analysis, but in the context of this paper we consider that a different problem from 3D flow visualization.

Specifically, the five tasks were:

- Task 1: Is a given point a critical point (CP)?
- Task 2: Identify the type of a given CP
- Task 3: Does the field advect from point A to B?
- Task 4: Is there swirl at a given point?
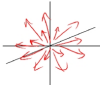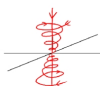- Task 5: Is the speed faster at point A or B?

Note that Task 5 tests whether the advection speed (i.e., magnitude of the vector field) is faster at point A or B.

All tasks were binary choices except for task #2 which involved picking one of eight CPs (see Figure 4). The total number of conditions was the product of the above, 2x2x5=20. Participants performed four instances of each condition (two for each possible binary answer) except for conditions involving task #2, for which participants performed two instances of the eight possible answers–pilot tests showed that more instances would have been too fatiguing. Thus the total number of trials was 2x2x4x4 + 2x2x8x2 = 128 trials.

3.4 Datasets

Each trial consisted of a dataset and visualization method pairing. In pilot studies we observed that task difficulty sometimes appeared to be a function of the dataset, and we thus wanted to ensure that all participants saw the same visual stimuli.

We required a controlled set of stimuli to perform this study. We generated 1000 3D vector fields and then selected a subset of 128 for use in the study (see details below). We use 3D vector fields generated from a Gaussian-based radial basis function [8]. Each field was generated by first selecting six random locations uniformly distributed on the volume [-1, 1] x [-1, 1] x [-1, 1]. (We tested using more and less random locations, but found six random locations generally gave a good level of vector field complexity.) At each random location, a vector was generated such that all three components of each random vector were chosen from a uniform random distribution between -1 and 1. Using the Gaussian-based radial basis function with a shape parameter of 1.2, we sampled the field to a 32 x 32 x 32 regular grid spanning the range [-1, 1] x [-1, 1] x [-1, 1]. An example of a full dataset presented to a participant during a training trial is shown in Figure 5.

Type 1	Type 2	Type 3	Type 4
Source	Sink	Saddle 1 in, 2 out	Saddle 2 In, 1 Out
Type 5	Type 6	Type 7	Type 8
Spiral Attract, In	Spiral Repel, Out	Spiral Repel, In	Spiral Attract, Out

Eight types of 3D critical points

Fig. 4. The eight types of critical points to be identified by participants (this figure was explained to participants in a training phase). "In" and "out" specifies whether the vector field moved in towards or out from the CP relative to each eigenvector axis. Attracting and Repelling describes whether spiral movement is moving towards or away from the CP.

We used the Newton-Raphson method and eigenanalysis to detect and classify critical points [27]. The eight types of first-order critical points in stable vector fields are represented in our datasets: two node types, two saddle types, and four spiral types (see Figure 4). We removed vector fields that did not have 1, 2, 3, or 4 critical points. Pilot studies [9] suggested that the fields were complex enough to measure the effectiveness of visualization methods.

From the pool of 1000 datasets we selected datasets for each task satisfying the conditions below. We required CPs to be located in the middle part of the dataset (i.e., the central third for the X, Y, and Z dimensions), so that its context was more likely to be useful. The following details the specific parameters used to select datasets used with each task. The trial generation phase produced 128 trials. All subjects experienced the same trial conditions, but in an order determined by Latin squares [19].

The study preparation involved a task parameter tuning phase whose objective was to cause participant accuracy to average about the midpoint between guessing and a perfect score. Through iterative testing we ultimately selected datasets and 3D points for the tasks (where applicable) using the criteria below. Below, "Near the center of the dataset" means that the x, y, and z coordinates of a point are in the range $[-\frac{1}{3}, \frac{1}{3}]$.

Task 1: Is a CP? CP near the center of the dataset, 4 instances per visualization method (4 x 4 = 16 trials), 2 instances are CPs, 2 instances not CPs.

Task 2: Type of CP? CP near the center of the dataset, 2 instances of each of 8 CPs per visualization method (16x4=64 trials).

Task 3: Advection task First point (P1) near the center of the dataset volume, second point (P2) on surface of sphere of radius $0.1 \times \sqrt{3}$ centered at P1, 4 instances per visualization method (4 x 4 = 16 trials), 2 instances on surface, 2 instances rotated by $20°$ about a random vector relative to P1.

Task 4: Is there swirl? Query point near the center of the dataset volume, 4 instances per visualization method (4 x 4 = 16 trials), instances per method in ranges [-4, -3], [-1, 0], [0, 1], and [3, 4]. The majority of λ_2 values for all datasets were in the range [-7, 7].

Task 5: Which point is faster? Query points near the center of the dataset volume, 4 instances per visualization method (4 x 4 = 16 trials), difference in speed between query points in range $speed_{max} \times [0.2, 0.5]$ (where $speed_{max}$ is the maximum speed for the specific dataset).

After the culling process, each dataset had between one and three CPs (66 datasets had 1, 53 had 2, and 9 had 3).
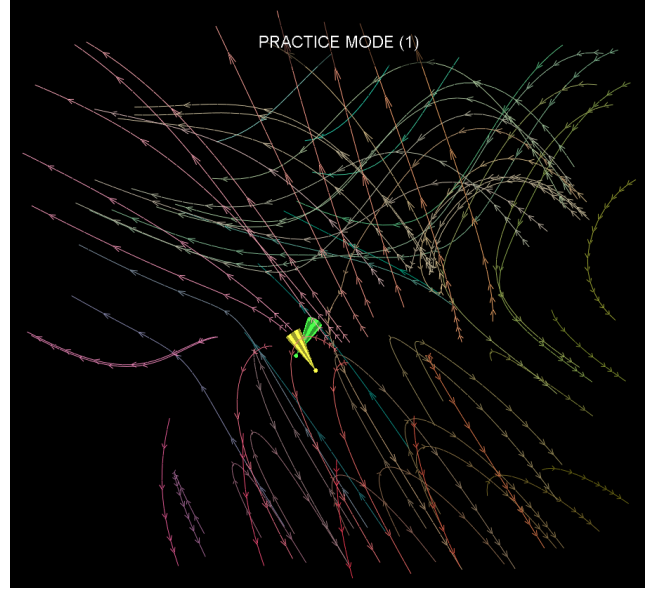

Fig. 5. A sample visualization of a full dataset using the participant-preferred lines visualization method. The task is the "Advection" task, which asks whether a line seeded at the tip of one cone will pass through the tip of the other cone.

3.5 User interaction

User interaction was minimized in this study–participants could use the cursor keys on the keyboard to rotate the dataset about its X and Y axes. This choice helped reduce variability in response time and is discussed in more detail in section 5.5.

3.6 Timing and training

Participants first completed an IRB consent form and pre-questionnaire. We then gave background information on 3D vector fields, integral curves, critical points, swirling, the tasks, and the visualization methods. Participants next completed the trials and events were logged to a data file. Participants then completed a post-questionnaire and there was a debriefing session. At the start of each group of trials for a particular task, we confirmed that participants were seeing both stereoscopic and monoscopic views of the visualization methods–all participants confirmed they could see the stereo methods properly and were not seeing double images.

All trials for a particular task were completed in series. Latin squares randomized the ordering of tasks and trials within each task across subjects. The average study ran 2.5 hours and participants took short breaks between tasks.

3.7 Participant pool

Six female and fourteen male subjects participated in the study. The mean age was 25. Thirteen subjects were undergraduates, four graduate students, one geoscience research staff member, one postdoc, and one faculty member. Participant areas of specialty were applied math, biomechanics, computer science, geoscience, statistics, anthropology, environmental studies, and English. Three participants were experts in that they had doctoral degrees and study vector fields regularly. By running twenty participants we collected data on all combinations of the conditions.

4 RESULTS

Discussion and details of the analysis follow, including thresholds and significance. *F* and *p* values [15] computed with SAS's General Linear Model (GLM) procedure are shown in Table 2. Tukey pairwise comparisons among dependent variables are detailed below. In the graphs and discussion we use the following abbreviations for the four
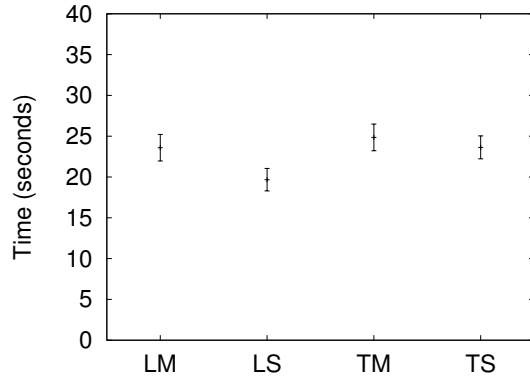
Fig. 6. Mean completion time across all tasks. LM = lines-mono; LS = lines-stereo; TM = tubes-mono; TS = tubes-stereo.

Fig. 8. Participant ratings of the difficulty of the five tasks; higher values indicate harder tasks.

Table 1. Differences as measured by time, accuracy, and confidence based on SAS's Tukey pairwise comparisons. Only statistically significant differences are listed – blank entries denote no differences. The notation $A > B$ indicates that method A was significantly ***more effective*** at the task than method B for the metric label at the top of the column.

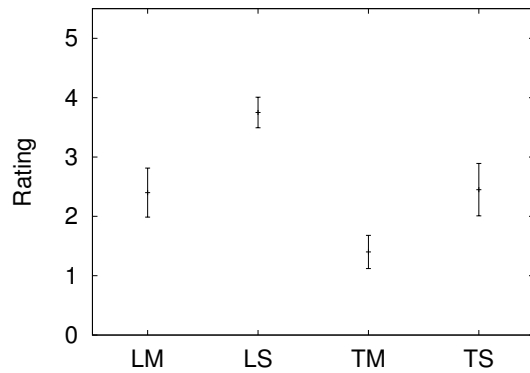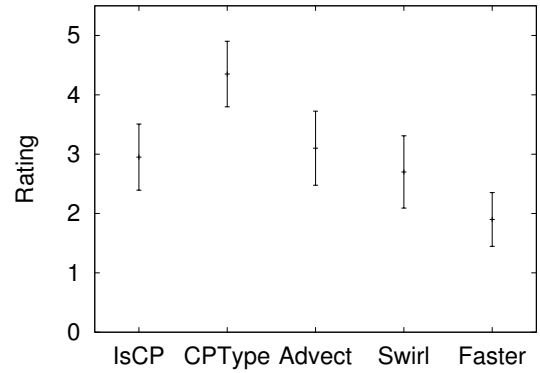| Task	Time	Accuracy	Confidence
Is a CP?		LM > LS	
		LM > TM	LS > LM
Type of CP?	LS > TS		
	LS > LM		LS > LM
	LS > TM		
Advection	LS > TM		
Swirling?	LS > TM	TM > LS	
Faster?	LS > LM		LS > TM
	LS > TM		LS > TS
	LS > TS		

Fig. 7. Participant ratings of the four methods for completing the tasks; higher values indicate greater preference for that method. LM = lines-mono; LS = lines-stereo; TM = tubes-mono; TS = tubes-stereo.

methods: TM = tubes mono; TS = tubes stereo; LM = lines mono; and LS = lines stereo. All graphs show mean values with 95% confidence intervals. Only graphs showing statistically significant differences are presented. (Hereafter, all uses of the word "significantly" refer to statistically significant differences.)

4.1 Quantitative and subjective summary

Across all tasks, participants finished trials significantly faster and with higher confidence using LS than all other methods (see Figure 6). There were no significant differences among the methods in terms of accuracy across all tasks. In the post-questionnaire, participants ranked LS most preferred, TS and LM second most preferred, and TM the least preferred for performing the tasks in the study (see Figure 7). Additionally, participants ranked CP-TYPE the most difficult task, ADVECT the second most difficult, and WHICH-FASTER the least difficult task; both IS-A-CP and SWIRL were as difficult as both ADVECT and WHICH-FASTER (see Figure 8).

The summary statistics in Table 2 show a high F value for the mean time by task. Below we report performance differences for each method across the tasks.

4.2 Quantitative summary by task

Table 1 summarizes differences among methods based on SAS's Tukey test. TM was the most accurate method for the "Is-a-CP?" and "Swirling?" tasks. For the three other tasks, all methods were equally accurate and LS was significantly faster. For the "Type of CP?" task, the mean accuracy was 46%. The mean accuracy of correctly identifying the category of CP type (i.e., node, saddle, and spiral) was 83%.

4.3 Stereo vs mono and rotation, tubes vs lines

To compare stereoscopic and monoscopic viewing, we grouped LS and TS into a single group, "Stereoscopic methods (SM)", and LM and TM into a single group, "Monoscopic methods (MM)".

SM was significantly faster than MM for the "Type-of-CP?" task (mean time 31.2 vs 28.0 seconds, respectively) and "Which-faster?" task (mean time 18.7 vs 15.4 seconds, respectively). For all tasks, the mean speed of SM was faster than MM. For the "Is-a-CP?" task, the mean score of MM (86%) was significantly better than SM (66%) (F(1, 22) = 17.53, $p < 0.0001$). Participants spent significantly less time per trial rotating datasets using SM than MM (mean 69.2% vs. 76.3%).

Line-based methods were faster than tube-based methods for each task except "Is-a-CP?". Stereo-based methods were faster than mono-based methods for "Type-of-CP?", "Advection", and "Which-faster?".

4.4 Novice vs expert

In terms of time, LS was significantly faster than LM for experts. For novices, LS was significantly faster than all other methods. In terms of accuracy, LS was more accurate than TS for novices. All other methods were comparable for both novices and experts. In terms of confidence, novices were more confident in answers given using LS than LM or TM. Experts were more confident with LS than TM.

4.5 Debriefing

In debriefing sessions participants generally said that the tubes occluded neighboring and more distant tubes and that made the tasks harder. Many participants commented on the "costs" associated with stereo viewing, such as wearing the glasses and perceiving a flickering in the display. Several participants said that during trials they were not
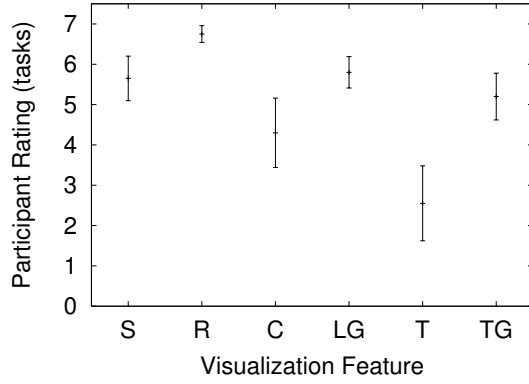
Fig. 9. Participant rankings of elements of the visualization methods. S = Stereo, R = Rotating dataset, C = Coloring similar lines a similar color, LG = glyphs on lines, T = Tube textures, and TG = tube glyphs (i.e., "feathers").

Table 2. Statistics for the various comparisons.

Mean time per method (across all tasks): $F(3,38) = 12.63$, $p < 0.0001$
Mean time by task (across all techs): $F(4,38) = 123.32$, $p < 0.0001$
Mean confidence by task (across all techs): $F(3,38) = 6.65$, $p = 0.0002$
Mean time per method (Type of CP?): $F(3,22) = 5.96$, $p = 0.0005$
Mean time per method (Advection): $F(3,22) = 6.01$, $p = 0.0005$
Mean time per method (Which faster?): $F(3,22) = 14.25$, $p < 0.0001$
Mean time per method (Swirling?): $F(3,22) = 3.65$, $p = 0.0131$
Accuracy per method (Is a CP?): $F(3,22) = 6.50$, $p = 0.0003$
Accuracy per method (Swirling?): $F(3,22) = 4.04$, $p = 0.0077$
Confidence per method (Which faster?): $F(3,22) = 7.88$, $p < 0.0001$
Confidence per method (Type of CP?): $F(3,22) = 3.34$, $p = 0.0186$
Participant rating of methods: $F(3,22) = 23.89$, $p < 0.0001$
Participant rating of tasks: $F(4,23) = 8.68$, $p < 0.0001$
Participant rating of factors: $F(5,24) = 23.22$, $p < 0.0001$
Mean Time, Stereo vs Mono (Type of CP?): $F(1,1) = 6.31$, $p = 0.0121$
Mean Time, Stereo vs Mono (Which faster?): $F(1,1) = 7.47$, $p = 0.0066$
Percent time rotating, Stereo vs Mono: $F(1,20) = 150.78$, $p < 0.0001$
Mean time, novice (all tasks): $F(3,35) = 11.28$, $p < 0.0001$
Mean time, expert (all tasks): $F(3,21) = 3.20$, $p = 0.0235$
Mean accuracy, novice (all tasks): $F(3,35) = 2.63$, $p = 0.0487$
Mean confidence, novice (all tasks): $F(3,21) = 4.77$, $p = 0.0026$
Mean confidence, expert (all tasks): $F(3,35) = 2.79$, $p = 0.0404$

consciously aware which methods were displayed stereoscopically vs. monoscopically. The majority said they would like to use LS if they did this kind of work professionally (i.e., every day for hours), although some noted that because of the ergonomics of stereo viewing such as the weight of the glasses they would generally use a monoscopic visualization, reserving stereoscopic visualization only for the hard tasks. Some participants said they liked the lines method with or without stereo–rotating was most important.

4.5.1 Feature ratings

In the post-questionnaire, participants rated the features of the visualization methods on a scale of 1 (did not make a difference) to 7 (made a significant difference); see Figure 9. The ability to rotate the dataset was rated highest. Stereo and the glyphs were rated next most important. Integral curve coloring was given the next highest ranking. The tube surface textures were rated lowest, but with the most variance.

4.5.2 Participants' suggestions for improvement

In the debriefing we asked participants for ideas to improve the visualization methods. Frequent or interesting suggestions were: reduce the tube radius, use animation in conjunction with the lines, vary tube radius so it indicates direction, use the mouse to make variable-speed rotation easier to control, let me zoom in for a better view, let me add a specific seed point, let me globally increase or decrease the number of lines, and get more comfortable stereo glasses.

5 DISCUSSION

Our expectations were that the tube methods would outperform the line methods, that stereo viewing would outperform monoscopic viewing, and that the TS method would be best. Our data did not agree with these expectations. No method performed best overall for all tasks and while LS generally finished the tasks faster and participants ranked the LS method highest, it was not more accurate. The following subsections discuss specific results.

5.1 Lines and tubes

Results. LS performed best for all metrics and participants ranked that method highest for this study. This differs from earlier work that found tubes performed best [25]. We might have expected our tubes to perform well since it had more 3D cues such as shading, surface texture (conveying both data information as well as helping the visual system identify the disparity between left and right images for the stereo conditions), and halos. It might be further surprising that LS performed best because we used the non-lit lines representation, which could be expected to perform worse than a line representation with more 3D cues like illuminated streamlines.

Participants' comments suggest the explanation may be primarily that the lines offer a clearer view of the data by not occluding each other, unlike the tubes. This may be especially true for the CP and swirl questions, which may be less affected by clutter or clear perception of line depth as long as participants can see into the space sufficiently.

Our results are not in agreement with Ware's study [25]. One obvious difference is the fidelity of the displays used. A higher resolution display might have improved perception of the tubes. This would not, however, necessarily have addressed the occlusion issue. We visualized 64 tubes per trial whereas Ware's study visualized one. Ware estimated that his study's 0.5 mm tube would be 2-3 pixels wide on a conventional screen. Our tubes were about 7-8 pixels wide.

Choice of lines and tubes. This paper compares only line-based techniques because they are important methods used by many scientists and the study design used was already 2.5 hours long. Future work could test other methods, including higher-dimensional primitives such as surfaces and volumes [27][16][17][18].

5.2 Accuracy

Results. Two particularly surprising results were that LM was so accurate for the "Is a CP?" task and that TM was more accurate than LS for the "Is swirling?" task. For the former, our working hypothesis is that "Is a CP?" may be a 2D pattern matching task–but if so, it is unclear why the other visualization methods were less accurate. Monoscopic methods (both TM and LM) had mean accuracy 85% and stereoscopic methods (LS and TS) mean accuracy 66%. We think that "Is swirling?" may be a "low-resolution" task – that is, the clarity participants reported LS offered was not critical in that task and the thick tubes helped emphasize swirling patterns and may have increased the accuracy.

Accuracy vs. other metrics. Fluid analysis experts said in discussing the results that in the end accuracy is the most important metric in their research, so in that sense for some tasks (like identifying swirl and determining if the vector at a point has zero magnitude), our data indicates that monoscopic visualizations will be more accurate under the conditions we tested than stereoscopic visualizations.

5.3 Seeding strategy

The 4x4x4 array of seed points spaced regularly through the volume is a simple seeding strategy that has been shown to be effective in general for similar tasks on 2D vector fields [13]. We believe it worked well

for a first view, but participants said that they sometimes wanted more lines or an additional specific line. Because we were simulating an exploration task we could not make seed placement a function of the correct answer. It would be interesting to compare seeding strategies to discover whether any are more effective for exploration. Similarly, as mentioned earlier, our specific study design did not provide user control for adding integral curve seed points because we were concerned with participant response to the visualization methods, not in seeding strategies of individuals and the efficacy of 3D interaction techniques. However, both of those topics are of interest for further research in the context of this study.

5.4 Swirling

Defining swirl. We instructed participants to look for patterns like water going down a drain and try to determine whether the point indicated by the marker was part of a rotating region that had a center point within itself. Most participants accepted this definition and became comfortable answering the question during the training, saying their strategy was to study the visualization and use their intuition to answer the question. However, a few participants had many questions about the definition and wanted a more specific definition of swirling. After some discussion, one participant finally asked for an example of a region that was *not* swirling. Further research in how to communicate verbally and visually the important feature "swirling" would be useful.

λ_2 **and divergent flow.** Our datasets are not guaranteed to be non-diverging flow. We recognized late in the project that Hussain's λ_2 value may not accurately identify swirling movement in the vector fields we used. Visually, however, we observed that swirling-like movement had negative λ_2 values and non-swirling movement had positive λ_2 values.

5.5 Choice of limiting user interaction

User-controlled interactive rendering is a critical part of visualization of complex data and has been reported since the first interactive systems became available [10]. A user's performance may depend on a variety of factors including task, data, visual design, interaction, and displays. For an evaluation to differentiate causes and effects, factors must be included and controlled explicitly [4]. Because our purpose was to study the visualization, we designed the interaction technique both to be reasonable for use in querying the data and answering questions and to be constant across task conditions, so that interaction was not a confounding factor in our experimental results.

It is hard to guess correctly how changing the user interaction would impact performance, and that is one value of empirical studies that help form and then test hypotheses. If zooming and panning were added, for example, performance might improve (because participants could zoom into the dataset for a clearer view) or might get worse (because participants have trouble navigating or lose orientation inside a dataset).

Another consequence of incorporating richer interaction is greater variance in user performance time, possibly making it harder to identify differences. Viewing this study as the first in a series we decided to provide the most basic user interaction in a very simple manner (cursor key controls) at this stage and to focus on the results for that context. We believe that, given the study design and the results, we have achieved a reasonable balance of the design parameters–the results confirm some expectations for the visualizations tested and disprove others, independent of the user interaction controls provided.

Future studies might augment our controls with zooming and panning and report the results. It would be useful to compare head-tracked view controls with device-driven controls.

5.6 Range of critical points

The datasets had between 1 and 3 critical points, resulting in a max-to-min ratio of 3, so that one might expect the dataset complexity to vary significantly (especially over a lower max-to-min ratio of 1.5, for example). However, this range was the result of our pilot study

process driven by performing the tasks on the various dataset parameterizations, and yielding datasets with comparable complexity for the tasks. No participant asked why some datasets were more complex than others. Also, every participant completed the same set of 128 (task, dataset) pairings, so all were exposed to the same max-to-min ratios. Additionally, we note that all datasets were produced using six random vectors and resulted in a variable number of critical points. Summing basis vector fields derived from CPs, as in van Wijk's work [24], may give greater control over the number and location of CPs.

5.7 Feedback on tasks by experts

In the debriefing we asked participants whether the tasks were important and if they would suggest other tasks. The purpose of this question was to help test whether the experts in our study believed the tasks (which were based on interactions with other scientists studying 3D datasets) were important. The experts all felt the tasks were relevant for their work. One suggested incorporating geometries interacting with the 3D vector fields. As part of this debriefing we discussed λ_2 further and our vector fields–see section 5.4.

5.8 Line representations

We used the "color similar curves similar colors" coloring scheme [7] because in pilot studies subjects said it made datasets "more approachable" than uniform or randomly colored tubes.

We also considered using illuminated streamlines [28] but reluctantly decided not to because of practical constraints on the study scope. We also were interested in baseline performance for a line representation with no lighting cues, which presumably would perform worse than illuminated streamlines for 3D tasks.

The tube texture is admittedly at the threshold of perception. Through pilot studies and a step wedge review evaluating a range of tube parameter settings, we optimized the texture by eventually reducing the tube radius to minimize occlusion while retaining a thick enough tube that the texture was just visible. As a result of this optimization process, at some distances the arrow in the tube texture cannot be seen, although one participant in the step wedge review pointed out that even if the arrow was not visible there was sufficient information in the texture for judging speed. While most participants said they did not use the surface texture, a few participants reported that they used it for some tasks. Furthermore, independent of helping judge speed and direction, the texture may have had subconscious benefits in stereo viewing by helping determine the disparity between features in the left and right images.

A thin streamline representation was selected because many production applications [1][2] provide it as a default. A tube representation was selected because it performed well in earlier studies [25]. Textured tubes were of particular interest due to Ware's recommendation: "Even without stereo and motion depth cues, tubes allowed for surprisingly accurate judgments. Thus the strongest recommendation that comes from this study is that tubes should be used to render 3D pathlines or streamlines." Furthermore, prior work has encoded data like speed and direction onto textures. We included monoscopic and stereoscopic viewing conditions to help test whether our natural ability to perceive the world stereoscopically made a difference [23].

5.9 Stereo

Stereo and fatigue. Can scientists use a stereo system over long periods to identify complex features? This paper's results suggest that stereo may reduce the time to complete some visual analysis tasks for 3D vector fields. Today's stereo displays can cause fatigue–for example, from cue conflicts. Over time, as technology and our understanding advance, higher quality displays well suited to the human stereo visual system may eliminate this problem.

Do tasks involve stereo perception? It is not clear whether participants utilized stereo in their strategies for completing our tasks. Some of our results indicate that tasks can be performed faster and subjects preferred doing them with stereo viewing, but unfortunately we did not collect data directly related to this question. In non-time-critical situations, accuracy is arguably the most important metric for scientists

and table 2 shows that participants were more accurate using mono-scopic than stereoscopic visualization methods for the "Is-a-CP?" and "Is swirling'?" tasks. Future studies should investigate whether par-ticipants utilized stereo in completing the tasks. Also, to control for the display brightness and contrast we required participants to wear stereo glasses even for monoscopic conditions. It would be interesting to have participants perform the same set of tasks without glasses, but time considerations precluded us from looking at this issue.

Stereo coupled with inteaction. One hypothesis future studies could explore is that stereo visualization alone may not provide a re-ally significant performance increase without being coupled with 3D interaction. For example, if this study was repeated but participants could use a 3D input device to specify additional seed points, then we might expect the stereo methods to outperform monoscopic methods because of the coupling of stereo viewing and 3D interaction.

5.10 Shortening length of study

This study can take up to 2.5 hours per person. A factor contribut-ing to this long duration is the number of datasets, but reducing the already small number of iterations per task might negatively impact the statistical analysis. The study might be shortened if fewer tasks or visualizations were tested.

6 CONCLUSIONS

Across trials, stereoscopic viewing and a thin-line representation helped participants complete trials significantly faster than the other visualization methods. In this study, participants liked the combina-tion of a clear visualization and stereoscopic viewing, although stereo viewing did not generally improve their accuracy.

Visualizations based on tubes should use a tube radius that does not lead to perceived occlusion among neighboring and more distant tubes and objects. Participants rated the ability to rotate the dataset interactively the most important feature in completing the tasks in the study. Furthermore, participants believed they would prefer variable-control rotation over fixed-speed rotation. Encoding direction and speed on a tube surface texture using our parameters was not useful for participants–the texture was too hard to see.

REFERENCES

[1] Tecplot, Inc., http://www.tecplot.com/, last accessed: June 2009.

[2] Kitware, Inc., http://www.paraview.org/, last accessed: June 2009.

[3] D. Acevedo, C. Jackson, D. H. Laidlaw, and F. Drury. Using visual design experts in critique-based evaluation of 2D vector visualization methods. *IEEE Transactions on Visualization and Computer Graphics*, 14(4):877–884, July 2008.

[4] D. A. Bowman, J. L. Gabbard, and D. Hix. A survey of usability evalu-ation in virtual environments: classification and comparison of methods. *Presence: Teleoperators and Virtual Environment*, 11(4):404–424, 2002.

[5] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the cave. In *Proceedings of ACM SIGGRAPH*, volume 27, pages 135–142. ACM, August 1993.

[6] C. Demiralp, C. D. Jackson, D. B. Karelitz, S. Zhang, and D. H. Laidlaw. Cave and fishtank virtual-reality displays: A qualitative and quantitative comparison. *IEEE Transactions on Visualization and Computer Graph-ics*, 12(3):323–330, 2006.

[7] C. Demiralp and D. H. Laidlaw. Similarity coloring of DTI fiber tracts. In *Proceedings of DMFC Workshop at MICCAI 2009*, 2009.

[8] H. Q. Dinh, G. Turk, and G. Slabaugh. Reconstructing surfaces by volu-metric regularization using radial basis functions. *IEEE Trans. on Pattern Analysis and Machine Intelligence archive*, 24(10):1358 – 1371, 2002.

[9] A. S. Forsberg, J. Chen, and D. H. Laidlaw. Towards comparing 3D flow visualization methods: A user study. IEEE Visualization Poster, 2008.

[10] W. Hibbard and D. Santek. Interactivity is the key. In *VVS '89: Proceed-ings of the 1989 Chapel Hill Workshop on Volume Visualization*, pages 39–43, New York, NY, USA, 1989. ACM.

[11] J. Jeong and F. Hussain. On the identification of a vortex. *Journal of Fluid Mechanics*, pages 69–94, 285 1995.

[12] Kitware, Inc. *The Visualization Toolkit User's Guide*, January 2003.

[13] D. H. Laidlaw, R. M. Kirby, C. D. Jackson, J. S. Davidson, T. S. Miller, M. da Silva, W. H. Warren, and M. J. Tarr. Comparing 2D vector field visualization methods: A user study. *IEEE Transactions on Visualization and Computer Graphics*, 11(1):59–70, 2005.

[14] R. S. Laramee, D. Weiskopf, J. Schneider, and H. Hauser. Investigating swirl and tumble flow with a comparison of visualization techniques. In *VIS '04: Proceedings of the Conference on Visualization '04*, pages 51–58, Washington, DC, USA, 2004. IEEE Computer Society.

[15] R. J. Larsen and M. L. Marx. *An Introduction to Mathematical Statistics and Its Applications*. Prentice Hall, third edition, 2000.

[16] L. Li and H.-W. Shen. Image-based streamline generation and rendering. *IEEE TVCG*, 13(3):630–640, 2007.

[17] O. Mallo, R. Peikert, C. Sigg, and F. Sadlo. Illuminated lines revisited. *IEEE Visualization Conference*, 2005.

[18] O. Mattausch, T. Theussl, H. Hauser, and E. Groller. Strategies for in-teractive exploration of 3D flow using evenly-spaced illuminated stream-lines. In *SCCG '03*, 2003.

[19] S. E. Maxwell and H. D. Delaney. *Designing Experiments and Analyzing Data: A Model Comparison Perspective*. Wadsworth, 1990.

[20] T. Munzner, C. Johnson, R. Moorhead, H. Pfister, P. Rheingans, and T. S. Yoo. NIH-NSF visualization research challenges report summary. *IEEE Computer Graphics and Applications*, 26(2):20–24, 2006.

[21] Prabhat, A. Forsberg, M. Katzourin, K. Wharton, and M. Slater. A com-parative study of desktop, fishtank, and cave systems for the exploration of volume rendered confocal data sets. *IEEE Transactions on Visualiza-tion and Computer Graphics*, 14(3):551–563, 2008.

[22] J. Sobel, A. Forsberg, D. H. Laidlaw, R. Zeleznik, D. Keefe, I. Pivkin, G. Karniadakis, P. Richardson, and S. Swartz. Particle flurries: Synoptic 3D pulsatile flow visualization. *IEEE Computer Graphics and Applica-tions*, 24(2):76–85, March/April 2004.

[23] A. van Dam, A. Forsberg, D. H. Laidlaw, J. LaViola, and R. M. Simpson. Immersive VR for scientific visualization: A progress report. *IEEE Com-puter Graphics and Applications*, 20(6):26–52, November/December 2000.

[24] J. J. van Wijk. Image based flow visualization. In *SIGGRAPH '02: Pro-ceedings of the 29th Annual Conference on Computer Graphics and In-teractive Techniques*, New York, NY, 2002.

[25] C. Ware. 3D contour perception for flow visualization. In *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization*, 2006.

[26] C. Ware, K. Arthur, and K. S. Booth. Fish tank virtual reality. In *Proceed-ings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 37–42, 1993.

[27] X. Ye, D. Kao, and A. Pang. Strategy for seeding 3D streamlines. In *IEEE Visualization 2005*, 2005.

[28] M. Zöckler, D. Stalling, and H.-C. Hege. Interactive visualization of 3D-vector fields using illuminated stream lines. In *IEEE Visualization*, 1996.