What Do People See in a Twenty-Second Glimpse of Bivariate Vector Field Visualizations?

Henan Zhao, Student Member, IEEE, Garnett W. Bryant, Wesley Griffin, Judith E. Terrill, and Jian Chen, Member, IEEE

Abstract—Little is known about how people learn from a brief glimpse of three-dimensional (3D) bivariate vector field visualizations and about how well visual features can guide behavior. Here we report empirical study results on the use of color, texture, and length to guide viewing of bivariate glyphs: these three visual features are mapped to the first integer variable \( v_1 \) and length to the second quantitative variable \( v_2 \). Participants performed two tasks within 20 seconds: (1) MAX: find the largest \( v_2 \) when \( v_1 \) is fixed; (2) SEARCH: find a specific bivariate variable shown on the screen in a vector field. Our first study with eighteen participants performing these tasks showed that the randomized vector positions, although they lessened viewers’ ability to group vectors, did not reduce task accuracy compared to structured vector fields. This result may support that these color, texture, and length can provide to a certain degree, guide viewers’ attention to task-relevant regions. The second study measured eye movement to quantify viewers’ behaviors with three-errors (scanning, recognition, and decision errors) and one-behavior (refixation) metrics. Our results showed two dominant search strategies: drilling and scanning. Coloring tended to restrict eye movement to the task-relevant regions of interest, enabling drilling. Length tended to support scanners who quickly wandered around at different \( v_1 \) levels. Drillers had significantly less errors than scanners and the error rates for color and texture were also lowest. And length had limited discrimination power than color and texture as a 3D visual guidance. Our experiment results may suggest that using categorical visual feature could help obtain the global structure of a vector field visualization. We provide the first benchmark of the attention cost of seeing a bivariate vector on average about 5 items per second.

Index Terms—Visual guidance, viewing behavior, bivariate glyph, evaluation metric, eye-tracking

1 INTRODUCTION

Seminal concepts in visualization have portrayed visual understanding as a reconstruction of the input from local features (e.g., edges, orientations, colors, and shapes) integrated into increasingly complex global objects and structures. In contrast, however, some recent visualization and vision science studies have suggested that recognition of real-world scenes may be initiated from the high-level global configuration [4,21]. In the former, humans are viewed as perceiving geometry forms (e.g., edges) before seeing objects and finally details in a scene, so that visual recognition is akin to a stimuli-driven bottom-up process [19]. In the latter, visual recognition uses the statistical attributes of the scene information to derive structures [36] and then guides attention to the task-related function through detailed scrutiny. In other words, viewers see forests before trees.

Anecdotal evidence from studies of scientific vector field exploration and visual studies of visualization [7,8,13] suggests that the second mechanism is likely to govern complex scene understanding, or at least that global and high-level scene features may significantly impact on guide users viewing behaviors. Research in vision and information visualization has also shown that during the initial glance of a visual stimulus, people perceive the higher-level spatial and functional components of the natural scene [36,42,51].

Take the example in Figure 1, a three-dimensional (3D) simulation result in quantum physics, a motivating application domain in this work. The vector magnitude at each location shows two variables: an ordinal visual stimulus, people perceive the higher-level spatial and functional

Fig. 1: An example showing human’s visual capabilities to quickly and extract statistics information of the global scene without identifying the individual datum. Viewers subsequent visual search can be guided by this structural division. Scientists frequently encounter complex simulation results as vectors shown on multiple cutting planes like this one, from a complex quantum physics simulation. The charge densities (scalars) are encoded by 8 different \( v_1 \) ranges as shown in the legend with a segmented extended blackbody colormap. The length is mapped to a continuous variable \( v_2 \).

vector length at each sampling site. Here we immediately see the chunk of colors and associated global structural distribution of \( v_1 \), without knowing the details of \( v_1 \) or its composite relationships of \( v_1 \) and \( v_2 \) at each sampling site.

This anecdotal evidence and recent experimental results may support an aspect of human visual intelligence: a cascade of human visual learning may exist that uses rapid global representations prior to slow presumably serial and attention-demanding visual scrutiny. Categorical visual features can guide the human attention to significantly reduce the search space and thus speed up visual queries [51]. If these mechanisms are supported, we may need to design not only for individual
locations but also for global viewing and a new process, i.e., for a viewing hierarchy from global guidance to scan of complex details, both bestowed by visual features. Though an extensive literature is devoted to this dual-viewing hierarchy and global scene guidance from understanding natural scenes and critical medical diagnosis from imaging techniques [50, 51], little is known about the design, evaluation, and viewers’ strategies in 3D vector field visualizations, when viewers can be benefited from seeing from complex data.

Here we study the possibility of visual guidance by features of color, texture, and length (Figure 2) and explore the following questions: Do different categorical features have different scene-guidance power and lead to more effective behaviors? What are the types and characteristics of users behaviors? Are certain viewing strategies more efficient than others? We perform two experiments to answer these questions and demonstrate such a human visual intelligence in two tasks: SEARCH (find a specific bivariate drawn on the screen) and MAX (find a local or global maximum). While one might argue that it is the structural content of the data (i.e., groups of similar variables as in the example in Figure 1) that constrains the possible location of task-relevant structures rather than feature guidance, global view guidance by visual features may not be as strong as we have anticipated. We thus in the first experiment randomize the vector positions and show that participants still achieve comparable accuracy. Since spatial uncertainty is inherent to feature uses and human errors, the goal of the eye-tracking experiment is to quantify how these features work through new human errors measurement metrics.

Our contributions include the following:

- A first look into the broader issues of perceptual speed of global scene feature guidance by extending to visualization ideas in vision science to study complex bivariate vector field visualizations.
- First benchmark of the attention cost of “seeing” a vector (on average about 5 items/second) and the tripartite eye-movement-based analysis of errors in scanning, recognition, and decision.
- New knowledge about categorical-cue-guidance of color, texture, and length and their rank order for degree of guidance; And explanations of observed performance patterns and implications for design from behavior measures.

2 Related Work

This section presents research that influences our work in visual guidance in both visualization and vision science and 3D glyph design for vector field visualizations.

2.1 Visual Guidance

The concept of visual guidance has been explored in both the visualization and vision science communities. Both study how visual features can be manipulated to guide visual attention to important regions and surfaces. Most notable solutions in visualization are perhaps the mature and important volume-rendering techniques such as Kim and Varshney’s saliency-driven volume rendering [35], Rheingans and Ebert’s structural-revealing nonphotorealistic volume rendering [39], Kniss et al.’s and Kindlmann and Durkin’s opacity transfer functions (see recent excellent review by Ljung et al. [24]); Intarrante’s diligently designed texture alignments for emphasizing directional information [33]; and Bruckner et al.’s stylized rendering of direct distance encoding to reveal spatial structures [10]. Nearly all of these stylized methods require creative algorithmic solutions. Our work begins to connect design choices to visual behaviors, thus shedding light on what visual features may be more effective in guiding human attention and what viewer behaviors tend to be more perceptually accurate.

Quantifying visual saliency in choosing visual features has also been studied in both visualization and vision science. Designs implementing bottom-up control of the feature information have been exceptionally insightful and productive. For example, Healey et al. [20] used preattentive processing (hue and orientation) for high-speed target detection, boundary identification, and region detection. Their systematic investigations emphasize that effective glyphs can encode large amounts of data in limited screen space and that choosing preattentive cues permits effective information access. Vision science treats guidance as a signal-detection problem and Wolfe and Horowitz [50] reveal five factors that influence guidance: (1) bottom-up stimulus-driven guidance similar to that studied in visualization; (2) top-down, task-driven guidance, (3) scene guidance in which attributes of the scene drive attention to areas likely to contain targets (answers), (4) guidance based on the perceived value of features or items, and (5) prior search history. All these aspects of guidance act together to guide users behaviors in real-world tasks [50, 51].

Inspired by this pioneering work, our approach is largely to understand how this highly influential guided-search can be applied to 3D visualizations. Because the initial statistical understanding is likely to be governed by categorical information, our current work compares color (most effective categorical feature [33]), texture (luminance-varying), and length (effective in 2D but perhaps limited in 3D [35]) to study how viewers alter their behaviors when using these as scene guidance.

2.2 Understanding The Research Paradigm and a Holistic Account of Scene Guidance

Our validation of visual guidance is also strongly influenced by Hochstein and Ahissars reverse-hierarchy theory [22] in perceptual learning science. That work suggests that human visual processing follows a dual hierarchy: vision-at-a-glance captures a high-level, generalized, categorical scene interpretation and a later attentional scrutiny stage requires sequential scan of individual items.

Recent studies have demonstrated the power of vision-at-a-glance. A
mere glance can assist radiologists in detecting cancerous tissues [16],
can help a crowd escape from dangerous scenes or quickly detect scene
structures, even when these structures are spatially far apart [57], and
help drivers see pedestrians and other surroundings from a six-second
glimpse [17]. This first impression is often termed holistic or global
pattern processing and features guiding such processing are called
scene features [4]. Evidence from recent visualization results may also
suggest the existence of rapid holistic signals in visualizations. For
example, viewers can memorize infographics and spatial data in brief
200ms exposures [58, 31]. We can visually average a large collection
of spatial data even without precise assessment of each datum [13].

These pioneering results may have suggested that human use global
statistical information presented at least in natural scenes and in data
other than vector fields when search accuracy is crucial. In this work,
we expand the literature to understand vector field and study viewers’
behaviors when treating scene guidance as categorical information to
understand the power of glimpse of vector field for 20 seconds. As in
other bivariate visualization studies [48], 3D glyph studies [40], and
generic glyph design [34], we carefully select the features to measure
effectiveness and use eye tracking to quantify human behavior.

2.3 Empirical Studies of Spatial Glyphs

Our study is related to the visualization of spatially encoded vector
glyphs where the first variable is an ordered integer and the second
is quantitative and continuous. We choose to study glyphs because
they are a powerful communication tool for visualizing spatial data and
because of their impact on medical imaging and vector- and tensor-field
analysis in physical sciences [46, 53]. Many real-world glyph
scenes showing scientific data are also highly dense, and inaccurate
human judgment from visualization can lead to uncertainty and ineffi-
cient behavior. Solutions like sampling can declutter visualizations,
but in some applications where to sample is unknown and spatial glyph
placement is often crucial. Furthermore, no matter what visualization
techniques are used, human errors cannot be further reduced. Thus it
is crucial to learn and quantify human errors, e.g., using eye-tracking

technologies [5].

Our method of applying fundamental visual science in visualization
design is not new. For example, Ware [47] convincingly explained
Laidlaw et al.’s vector field study results [29] by pointing out that
the human visual cortex contains lines or edges of orientation and
location cells and how the human visual system gradually integrates
these features into the output to represent the 2D vector fields. Given
these successes, our work further distinguishes between these low-level
feature-driven bottom-up processes (in which the initial vision perceives
colors, orientations, and motions) and our view of ranking global scene-
feature guidance, i.e., perceiving conjoined forms of overall patterns
through we seem to break the feature rules of showing continuous edges
as in gestalt principles. To make the ensemble of object-features to
work, we further suggest the highly pop-out visual dimensions would
support the role of forming the similar object groups so participants’
inference of the visual structure will be drawn. As a result,
we use color and texture to compare against length, which might have
limited ability to generate the pop-out effect in 3D [35, 46]

The third type specific to bivariate data visualization is the corre-
spondence between the exponent and digit to its visual representations
-the fact that a viewer must be able to tell which part is exponent and
which digit [52].

3.3 Errors in Human Behavior

A third metric is human error. Examining human errors will help
explore means to overcome inefficient behaviors and can lend us knowl-
edge when to use automatic decision support to reduce them. Here we
borrow error terms from Kundel et al. [28] and Cain et al. [12] to define
new metrics, currently confined to medical imaging analysis.

1. Scanning error: the target (answer) was never fixated or appeared
in the visual scan path.

2. Recognition error: the missed target is briefly inspected but not
examined long enough to be properly evaluated.

3. Decision error: the missing target is searched for an appropriate
interval but is not correctly categorized.

4. Refixation: an object is visited repeatedly.

Often these error terms are weighted differently depending on the
application. In vector field analysis, scanning error is perhaps least
desirable since it represents misses and thus has a higher cost when
critical structures or data items must be recognized. Decision error is
likely to be relatively low in our case, we anticipate, because once fix-
ated, most viewers would be able to decide on their number, assuming
the tasks are intuitive and simple enough (this turns out not to be the
case from our study). Refixation is an interesting phenomenon and can
make an object “retrospectively visible” so items being missed at the
decision and recognition stages are found again.
The goal in this study is to determine how the random placement of variables to study the global viewing strategies in three vision attention [44]. As it may sound counter-intuitive, with salient representations, the experiment is visual feature type. For the bivariate representation, we collect subjective self-reported behaviors.

How the context affects the perceptual speed by randomization and sufficiently clear to guide human vision. Here, we attempt to understand guidance since the global semantic properties of the salient colors is color differences immediately attract attention, having colors next to absence of the detailed knowledge of scene layout [9, 11, 38, 43]. Such co-occurrence of location) and this process can occur even with the introduced by global statistics of object location (rather than mere organization. In the randomized condition, contextual influence is occurrence, with the recognition within an initial glance to see the scene structured conditions the objects that have been associated over space. For example, in the hand, context can be framed as the relationship between vector objects and then the context of a vector field is defined as a combination of objects that have been associated over space. For example, in the structured condition the scene context can be build by location co-occurrence, with the recognition within an initial glance to see the organization. In the randomized condition, contextual influence is introduced by global statistics of object location (rather than mere co-occurrence of location) and this process can occur even with the absence of the detailed knowledge of scene layout [9, 11, 38, 43]. Such salient features pop-out as the local or global scene feature to guide vision attention [44]. As it may sound counter-intuitive, with salient color differences immediately attract attention, having colors next to each other between adjacent locations would not improve global scene guidance since the global semantic properties of the salient colors is sufficiently clear to guide human vision. Here, we attempt to understand how the context affects the perceptual speed by randomization and collect subjective self-reported behaviors.

Visualization Methods. The second independent variable in this experiment is visual feature type. For the bivariate representation, we vary the first variable to study the global viewing strategies in three representations.

- **Color** (Figure 4a) is chosen because it is ranked top for showing categories [30] and because global viewing is believed to rely on categorical information. Here, colormaps were chosen from the sequential hue-varying maps in Colorbrewer [18].

- **Texture** (Figure 4b) uses the percentage of black on white, as inspired by a design by Bertin [3]. The visual effect is similar to luminance variation. For example, to represent the v1 term in five categories, we use the amounts of black 0 (0%), 1 (25%), 2 (50%), 3 (75%), and 4 (100%) all wrapped around the cylinders in five segments to be visible from any viewpoint.

- **Length** (Figure 4c) is chosen because of its best quantitative scale [33] and its ubiquitous use to represent vector magnitude. Here we use length as an ordered integer instead.

The second variable is quantitative and is always mapped to vector length.

4.1.2 Choosing Tasks

Fig. 5 shows two example tasks used in this study.

SEARCH (Figure 5a): An example task is: Find the vector shown on the lower-right part of the screen within 20 seconds.

This task involves a single well-defined target search from the entire vector field. The target vector was displayed on the lower-right corner of the screen and was varied in each trial.

MAX (Figure 5b): An example task is: Which point has maximum vector magnitude among all vectors when v1 = X? X in the study was an integer from 3 to 7.

This task is chosen because it involves searching for the extreme among many vectors. Also, participants must compare the current vector with those seen in the recent past in order to determine whether the current one is the answer, so that they must also hold their past search in their memory while going through the vector field. Thus, this task may require a greater cognitive load than the first SEARCH task.

To accomplish these tasks, viewers must examine the entire field to obtain an answer. We assume that if the global scene features (the three visual features) in fact guide attention, participants would obtain these questions from a sub-region by utilizing global statistical features, obtain an answer. We assume that if the global scene features (the three visual features) in fact guide attention, participants would obtain these global statistical structures and examine a relatively small sub-regions, without costly visual scrutiny of the entire field.

4.1.3 Hypotheses

Given our own observations and the literature, we have had several hypotheses when entering the experiment:

- **H1.** The accuracy of these three visual features follows the rank order of categorical performance. This is because the initial scene guidance is likely to be categorical and using categorical data is more effective in categorical tasks [49].

- **H2.** Randomization may not be a deterministic factor in task accuracy. This is because participants might be able to use pop-out features to obtain global data distributions.

- **H3.** Behavior. All participants would try to examine sub-regions only.

4.1.4 Choosing Vector Field Data

We selected the data carefully to avoid introducing a confounding factor of dataset. We generated the data by randomly sampling some very complex large-magnitude-range simulation results within 3D-box containing 445 to 455 vectors. That data are adjusted to a range between three (thus data range: [1, 100] to seven-levels (thus data range [1-10^6]) categories from random samples. The number of data points among the power categories are about the same, i.e., total number of points/ power level. Thus, if the participants answer questions from a sub-region by utilizing global statistical features, they would examine around 64 to 145 vector items. The categorical colormaps for the 3-7 levels were chosen from colorbrewer [18].

4 EXPERIMENT I

The goal in this study is to determine how the random placement of bivariate glyphs and visual feature mappings influence participants’ accuracy.

4.1 Methods

4.1.1 Independent Variables

**Randomness.** The first independent variable in this study is randomness. In the structured condition, vectors belonging to the same first-variable category are configured to be spatially proximate (Figure 3a). In the randomized condition, the corresponding vectors are positioned randomly in space (Figure 3b). Also note that the orientations are not continuous in this randomized condition, so that participants could not group vectors by orientation.

The effect of randomization is to alter the context. On the one hand, context can be framed as the relationship between vector objects and then the context of a vector field is defined as a combination of objects that have been associated over space. For example, in the structured conditions the scene context can be build by location co-occurrence, with the recognition within an initial glance to see the organization. In the randomized condition, contextual influence is introduced by global statistics of object location (rather than mere co-occurrence of location) and this process can occur even with the absence of the detailed knowledge of scene layout [9, 11, 38, 43]. Such salient features pop-out as the local or global scene feature to guide vision attention [44]. As it may sound counter-intuitive, with salient color differences immediately attract attention, having colors next to each other between adjacent locations would not improve global scene guidance since the global semantic properties of the salient colors is sufficiently clear to guide human vision. Here, we attempt to understand how the context affects the perceptual speed by randomization and collect subjective self-reported behaviors.

**Visualization Methods.** The second independent variable in this experiment is visual feature type. For the bivariate representation, we vary the first variable to study the global viewing strategies in three representations.

- **Color** (Figure 4a) is chosen because it is ranked top for showing categories [30] and because global viewing is believed to rely on categorical information. Here, colormaps were chosen from the sequential hue-varying maps in Colorbrewer [18].
were instructed to be as accurate as possible. They were told that all tasks had time constraints of 20s or until the participants practiced until they fully understood the visualization and could ask questions during training but not when testing begins. Participants finished the current test and pressed the 'next' button. Participants examined all three features under the randomized and structured conditions three times. Repetition was used so that participants would have enough time to form strategies. Half of them looked at the randomized data and other half looked at the structured first. All participants looked at data from all categorical ranges as well for a variety of experiences. As a result, each participant executed 90 trials for each task and 180 total for these two tasks.

Independent and Dependent Variables. Independent variables include position distribution (structured and randomized) and feature type (three levels of color, texture, and length). The dependent variable in this controlled study is accuracy since all tasks were timed.

Participants. There were eighteen participants (12 male and 6 female, mean age = 23.8 and standard deviation = 4.9). They were graduate students from diverse backgrounds, the majority from computer science (7) and computer engineering (3). Others were two from information systems, two mechanical engineering, one chemical engineering, one physics, and one business school.

Experimental Procedure. Participants were greeted and completed an Institutional Review Board (IRB) consent form, which described the study’s procedure, risks, and benefits, and a demographic survey. All participants were tested using the Ishihara color-blindness test and could continue only if they had normal color vision and normal and corrected-to-normal vision. Participants were trained to understand these bivariate and were given practical trials using all conditions. They were tested using the Ishihara color-blindness survey. All participants were tested using the Ishihara color-blindness.

Full-factorial repetitive within-participant design was used: all participants examined all three features under the randomized and structured conditions three times. Repetition was used so that participants would have enough time to form strategies. Half of them looked at the randomized data and other half looked at the structured first. All participants looked at data from all categorical ranges as well for a variety of experiences. As a result, each participant executed 90 trials for each task and 180 total for these two tasks.

System log. The system automatically loaded the combination of visualizations and questions during the experiment without investigators’ intervention. Participants used a computer mouse to click on the answer (also called target) they found. Each answer or target selection was indicated by a yellow arrow superimposed on the display so that it could appear to be on top of any item. They could change their answer before the time ran out or before they clicked on the “next” button, when the next question was loaded and the user proceeded. The system recorded all user interactions (mouse movements and clicks) and participants answers.

4.2 Results and Discussion

We collected 1620 data points for each task. Data were analyzed by task. Table summarizes the statistical measures of the main effect of the independent variables (feature and randomness) on the relative error using the SAS MIX repeated measures. Here relative error is computed by measuring how far the answer was from the ground-truth or \( \frac{|\text{participant's answer} - \text{ground truth}|}{\text{ground truth}} \). A post-hoc analysis used Tukey’s Studentized Range test (HSD) to compare the independent variables levels when we observed a significant main effect. Figure shows the mean relative error (in gray) vs. randomness and visual feature. All error bars (maroon) represent 95% confidence intervals. Table also reports effect sizes using Cohen’s \( d \), labeled “small” \((0.1 − 0.21)\), “medium” \((0.21 − 0.35)\), and “large” \(≥ 0.35\) effects [14].
We computed the number of correspondence errors in each task. Here we present the results from 1620 questions. Table 1: Experiment I: the number of correspondence errors and subjective preference rating from 1 (worst) to 7 (best).

| Task | Variables | Significance | Effect Size |
|------|-----------|--------------|-------------|
| SEARCH | Total number of correspondence errors from 1620 questions | F(2,516) = 37.6, p < 0.0001 (color) > (texture) > (length) | 0.5 |
| | Average subjective rating | F(1,516) = 0.3, p = 0.60 | 0.02 |
| MAX | Total number of correspondence errors from 1620 questions | F(2,513) = 29.7, p < 0.0001 (color, texture) > (length) | 0.41 |
| | Average subjective rating | F(1,513) = 4.6, p = 0.03 | 0.1 |

4.2.1 Numbers of Correspondence Errors

We computed the number of correspondence errors in each task. Here an answer was considered having a correspondence error when participants selected an erroneous category from the first variable. Overall, participants were generally reasonably accurate in these tasks (Table 1). Color led to the fewest correspondence errors and texture and length the most. Participants also rated color the easiest to use and texture and length the most difficult. Among the correct answers, participants’ answers using color and texture were on average more accurate and were less than 8% away from the ground-truth for both tasks (Figures 6a and 6b, left). The least efficient length estimate was generally off by about 12%-14% on average (Figure 6c, right).

4.2.2 Main Effect of Visual Features and Randomness

H1 was supported. The results support that visual features guidance at different speed for bivariate data visualization. Participants had lowest accuracy with the length and this main effect of visual feature on relative error was significant for both SEARCH and MAX tasks. The rank order was supported by post-hoc analysis that color and texture in general lower relative errors than the use of length (Table 2). It is not surprising that participants should show a significant effect given that their areas of expertise and experience with vector fields varied widely. Participants self-reported that they had trouble drawing boundaries between the adjacent categories when the range of the visual feature was large. These results may suggest that the continuous quantitative encoding of texture and size made it harder to reject distractors.

We could speculate that our feature-ranking result was in line with vision science results in that global feature selection preferred categorical features of color/hue and luminance-varying texture and as a result incorporating categorical features or boundary-revealing into global visual guidance might be productive.

H2 was supported for SEARCH and only partially supported for MAX, as we can see that randomness was not a significant main effect on relative error for SEARCH but was for MAX (Table 2). The effect size of SEARCH was small and post-hoc analysis revealed that randomized and structured were in the same Tukey group. One may speculate that people could use features effectively as categorically distinct factors to reduce distractor heterogeneity via grouping, and that this grouping could be spatially distributed.

5 Experiment II: Eye Tracking Behavior Study

So far, Experiment I suggests that participants used scene guidance and categorical color as scene guidance would be most effective (lowest relevant error and lowest wrong answers). Texture, however, had the same relative error but the most wrong answers. Length perhaps was not effective. Though we could not say randomness was not a determining factor in forming global strategies because we did not see any significant differences. These results were used to guide this second study. Here our goals were to understand how viewers attend to objects and their behavior to achieve efficiency using color, texture, and length. Here we use eye-tracking data to assess not just the features but how they may lead to efficient behaviors.

5.1 Methods

5.1.1 Overview

The same three visualization methods were used. There were three main differences between this study and Experiment I. First, the only independent variable was feature, coupled with the structured condition for us to measure the visual behaviors. We did this for two reasons: (1) the effect size was small in Experiment I even when randomness was a significant main effect for MAX; (2) Eye-tracking with the chin-rest can introduce fatigue, and thus we wanted to limit the study time to 30 minutes. Second, we also removed interaction and participants only looked at two-dimensional (2D) static pictures from the same data used in Experiment I. The visual cues were large enough for participants to answer questions. Third, all participants in this study were eye-tracked.

5.1.2 Study Design

We used a full-factorial design and participants looked at 15 instances of the three visualization methods, leading to 45 trials per task. The study took about 30 minutes to complete (45 trials × 20 seconds × 2 tasks = 30 minutes).

Independent and Dependent Variables. The independent variable was the visual feature. Dependent variables are errors and eye movement cost such as eye fixation and context switching.

Participants. Six participants of diverse engineering and science backgrounds participated in this study.

Experiment Procedure. An Eyelink1000 eyetracker (SR Research, Ontario, Canada) was used to sample the x- and y-position of the eye at 500 Hz (Figure 7). At the beginning of each session, we calibrated and validated the eye tracking using the manufacturer’s standard nine-point calibration procedure for each participant prior to each use.
We collected 540 data points for each task. Whether or not the eye fixation coincides with the ground-truth data for 40 milliseconds. The blue circles show the eye-fixation position and context-switch cost represent averages for each participant; these values were then used for statistical analyses and could be found in the supplementary material.

To investigate the participants strategies, we coregistered the x and y vector pixels at that location. The fixation time, fixation length, saccade and context-switch cost were used to characterize and compare these three visual features. Some participants were observed to be “drillers” who constrained their eye fixations within the target data regions of a single v1 area and quickly examined or drilled through vector magnitude in that subregion. Other participants exhibited a second fixation behavior: they were “scanners” who looked at much larger v1 regions in addition to those within the target.

We wanted to know if feature type would affect participants behaviors. Thus, we examine the distribution of drillers and scanners by feature type, with the results shown in Figure 7. The trials were grouped by the percentage of object fixation points in the v1 region, where 100% were that all fixations were within v1, 95% permits 5% noise swinging to regions other than v1, and so on. We used four thresholds - 100%, 95%, 90%, and 80% – and then summarized the percentage of the trials belonging to these categories. Figure 8 shows one example of each type of drillers and scanners of SEARCH and MAX tasks: the drillers showed focused exploration of the target region and the scanners showed multiple v1 regions’ fixations. Both drillers’ and scanners’ eye-movement behaviors seem from these samples to have good organizational structure.

We may observe that in general there were more color than texture and length drillers in all threshold conditions. For SEARCH, 34.4% of the color trials were full driller trials when participants’ fixations were in the target region only, compared to 16.7% texture and 6.7% length trials (Figure 7a, the first left-most group of 80%).

We observed that the scanning error for SEARCH using length (33%) almost doubled that of color (17%) (Figure 6c), and the feature type was not a significant main effect on scanning error. But for MAX, texture had the highest scanning error (24%) compared to color (13%) and

| Task    | Error | Significance | ES |
|---------|-------|--------------|----|
| SEARCH  | Scanning | $\chi^2=6.3$, $p=0.04$ | 0.16 |
|         | Recognition | $\chi^2=7.1$, $p=0.03$ | 0.24 |
|         | Decision | $\chi^2=2.6$, $p=0.28$ | 0.17 |
| MAX     | Scanning | $\chi^2=3.3$, $p=0.19$ | 0.11 |
|         | Recognition | $\chi^2=0.3$, $p=0.86$ | 0.04 |
|         | Decision | $\chi^2=1.0$, $p=0.60$ | 0.08 |

5.2 Results and Discussion

We collected 540 data points for each task.

5.2.1 Reconstruction of the Eye-Tracking Data

To investigate the participants strategies, we coregistered the x and y position of the eye fixation with a testing image to produce fixation points and the corresponding geometry values. Here an eye-fixation is each instance that the eye remained in an area (5 × 3 pixels) for at least 40 milliseconds. The blue circles show the eye-fixation position and the size of the circle shows the fixation time. We subsequently compute whether or not the eye fixation coincides with the ground-truth data for the guidance features (or v2 of the bivariate data).

We further measured object fixation when eye fixation was within a vector item. This object fixation contains a fixation cluster, defined as successive fixations that fall within a single vector area of interest. This metric undoubtedly depends on how one estimates the “useful” regions. Here we use the vector length and width to constrain the single area of interest provided that this is the unit on which participants need to fixate to get an answer.

Data presented in the next section state the fixation or object fixation in the analysis, where the fixations were either from the original blue circles from the device or the post-processed clustered fixations on the vector pixels at that location. The fixation time, fixation length, saccade and context-switch cost represent averages for each participant; these values were then used for statistical analyses and could be found in the supplementary material.

Fig. 7: Experiment II: percentage of drillers by varying the threshold of the percentage of fixation points within the ground-truth v1 region. Here, 100% means all fixations in that trial fell in the v1 region only.

Interaction. Participants in this experiment examined static images. No zooming, panning, or rotation is allowed. They selected the answers using the same approach as in the first experiment.

System Log. Eye movement data were recorded and categorized as either fixations or saccades depending on eye position velocity and acceleration using the commonly accepted standard from the manufacturer’ default settings. We also recorded participants’ answers.

Table 3: Experiment II: main effects of features on scanning, recognition and decision errors. ES represents effect size using Cramer’s V for binary choice. The significant main effects and the high effect size in bold (none was found) and the medium effect size is in italic.
length (18%) though no significant difference was observed (Table 4). We could speculate that the lack of significance was that our sample size was small.

**Recognition error** and **decision errors** relate to the missed targets. The recognition error concerns a target that was fixated but appeared on the scanning path during visual inspection but only briefly, not long enough for proper evaluation. In contrast, decision errors occurred when the target was fixated for a long enough time interval but was not chosen as the target. However, it was difficult to truly untangle these types of errors, because the fixation time needed to recognize an item has not been defined in vector field visualizations. Vision science has established a threshold of 20-50 items per second based on natural scenes for human eyes to recognize an object. The difficulty with applying these numbers to our study was that previous studies used visually distinct targets and distractor items (e.g., distinct numbers and letters or objects such as apple, orange or pear). Our object fixation was longer than that established in vision science: on average a single object fixation to make a decision about a target is about 200-250 ms.

This measure was stable for both tasks and thus may be used as a rough benchmark of about 5 items per second for one of a single fixation to recognize bivariates of a vector. In this work, we followed previous work in using the object fixation duration on distractor items to provide a reasonable basis for separating recognition and decision errors. In general, we used the third quartile from the distractors’ fixation time as the threshold, which led to 306 ms for SEARCH and 303 ms for MAX.

Figures 9a and 9b show the results. The figures at top with yellow bars showed the number of trials among the 90 trials when the recognition error occurred; those at bottom show the recognition error rate, i.e. the trials not chosen as the final answer within the corresponding yellow ones. Recognition errors nearly doubled and were high (47%) for length, compared to color (21%) and texture (17%) in SEARCH. Figure 9c showed that in general the decision errors were high, though we observed no significant differences among the three encoding methods. The high decision error and lack of differences may arise for two reasons: first, accurately perceiving a pattern or a numerical value in 3D can be challenging. Second, eventually, once a category is recognized, the tasks using all three approaches involved examining the length of the second variable $v_2$.

**Refixations** occurred only rarely. The top graph in Figure 9d shows the total number of items refixated and the bottom shows the average number of refixations. Overall, refixation occurred no more than twice and for no more than three items in a scene in general.

### 5.2.4 Quantifying Drillers and Scanners’ Performance Results
So, does it matter if one adopts a driller or scanner strategy during the vector SEARCH and MAX task? And when is one strategy more effective than the other? We looked mainly at the relative error and used other measures for reasoning purposes, e.g., fixation time, total number of fixated objects, normalized context switching cost, and total context switching cost.

Table 4 and Figure 10 summarized the results. Being driller or not was a significant main effect for both SEARCH and MAX. Despite that the relative error doubled for scanners, we did not find significant differences between errors between drillers and scanners. We may say that the no-significant between comparison may have caused by our low number of participants in this eye-tracking study. However, drillers had significant longer object fixation time and scanned fewer objects
Table 4: Experiment II: Driller and scanners. Comparisons of the relative error, normalized object fixation time, total number of fixated object, and normalized and total context switching cost (ms) between drillers and scanners. ES represents effect size. The high effect size is in **bold** and the medium effect size in *italic*.

| Task      | Variables         | Significance | ES  |
|-----------|-------------------|--------------|-----|
| SEARCH   | Relative error    | F(1,38)=1.2, p = 0.28 | 0.39 |
|           | Fixation time (norm) | F(1,38)=30.1, p <0.0001 | 0.71 |
|           | # of fixated objects | F(1,38)=22.8, p =0.069 | 0.69 |
|           | Context switching (norm) | F(1,38)=0.2, p = 0.65 | 0.07 |
|           | Context switching (total) | F(1,38)=9.6, p =0.004 | 0.54 |
| MAX      | Relative error    | F(1,48)=1.0, p = 0.32 | 0.17 |
|           | Fixation time (norm) | F(1,48)=2.9, p = 0.09 | 0.23 |
|           | # of fixated objects | F(1,48)=0.04, p = 0.85 | 0.06 |
|           | Context switching (norm) | F(1,48)=2.9, p = 0.09 | 0.23 |
|           | Context switching (total) | F(1,48)=0.6, p = 0.43 | 0.09 |

Fig. 10: Experiment II: Quantifying Drillers and Scanners’ Eye-Movement Behaviors.

for SEARCH. But these factors were about even for MAX. The total context switching cost was significantly higher for drillers as well.

There were a number of reasons why drilling might be more effective. It may be that drilling was simply a better strategy. Certainly, this advantage depended on the task at hand. For MAX when multiple items were searched, drilling was only slightly better. At least one factor may make it unwise to argue too strongly that the drill method was a determining factor or inherently superior. This was because most color users were drillers. We believe that further research was necessary to determine if drilling was truly a better strategy for the task and whether or not encoding should be designed to encourage drilling behaviors. And if so, interface mechanisms could be provided to provide v1 only data to encourage drilling behaviors.

6 General Discussion

Like a dream come true, visualization tools are helping scientists achieve deep insights from increasingly rich and complex data inputs. This section summarizes the study outcomes and answer the questions we raised earlier.

6.1 Categorical Color was Most Effective 3D Scene-Guidance Feature Compared to Texture and Length

The very first at-a-glance view is not of such unattached free-floating visual properties (of edges etc.) but rather of coherent joined objects. Visual-features (here color) as well as those scene-features when pop-out can manifest attention. Later, explicit perception returns to those low-level visual cues for attention-demanded scrutiny and looking for details. In another words, the human vision system learns from easy to hard from visualization and the global holistic processing may happen before the learning from low-level visual properties directly.

Our first perspective in this work is to quantify human visual learning to see what features encourage visual guidance and the characteristics of these. We look at the outcome of the process and assume that, if such a process occurs, participants would use a drilling behavior to focus on task-relevant regions. Unlike the traditional supports a cascade of human visual learning that addresses the rapid, easy, and automatic global or holistic representation prior to slow presumably serial search of visual scrutiny. Our results seem to support that the classical studies of feature guidance might be reversed into scene guidance or the combination of two of feature+scene guidance in visualization, because participants tried to be driller as much as they could. Participants were also more effective when using categorical colors, followed by texture and length.

6.2 Measurement Metrics using Error

Our second perspective in this work is to quantify error measures and understand the visual guidance efficiency of a visualization. Most evaluation methods focus on understanding time and accuracy, central in ultimately determining which techniques are chosen for what goals and tasks [24]. Insight-based methods measure the numbers of insights, questions-based studies the numbers of research questions. In a nutshell, all these methods are driven by outcomes in the analytical stages; evaluation experts need to do some guesting work to understand “why” and “when” visualizations help or hinder insights or question formation.

In this work, our first experiment includes a larger participant pool so that we could establish the traditional statistical tests for effectiveness (accuracy), efficiency (respond speed), and subjective preferences. In the second study, we coupled eye tracking to empirical studies to reveal ongoing human visual processing. One obvious benefit of our method is for the future to determine when to introduce autonomous solutions to compensate inefficient human behaviors.

Another possible future direction to characterize these error attributes into layers and study how errors cascade, i.e., the inefficient behavior from a level above is transported to the level below and influences how humans see and act upon next. Differences and similarities in fixation and its dynamics are measured to understand the behaviors in short and long sequence analysis. Additionally, patterns revealed help establish whether there are distinct strategies that participants employee - what are those caused by visual marks and what else by users’ differences.

6.3 Eye-Tracking

In the long term, understanding how humans see permits us to learn how human expertise in many scientific fields reflects complex cognitive and perceptual processing. This knowledge can inform visualization design or even inform automatic approach to simulate human viewing process. Common eye tracking metrics such as fixation count, fixation duration, and time to first fixation were used as overall indicator of participants’ ease or difficulty using the studied methods. It has been used to measure multiview visualization efficiency to study whether or not visualization types in views influence efficiency and accuracy [15] [21]. Though eye tracking method is becoming trendy in visualization studies, analytical solution remains limited. Our own experience is that it is better to combine hypothesis and eye-tracking methods together to study when, why, and how visualization works.

6.4 Limitations and Future Work Visual Guidance

This work has limitations. We could have had more participants in the eye-tracking study so we may observe significance between driller and scanner behaviors. Visual search is one of most important spatial tasks since it can be relevant to many sensemaking tasks and since fundamental limits on visual processing make it impossible for humans to recognize everything at once. Future work might focus on the type of guidance and quantifying the factors that influence such guidance.

Compared to past empirical studies of using a few items, our work may suggest that for real-world uses, empirical study perhaps needs to
include more items to learn if global visual inspection would facilitate human sensemaking.

7 Conclusion

This study is the first to our knowledge to systematically compare and discuss human behaviors when participants were being guided by visual features. Results from the study lend the following insights to the visualization community for designing effective vector field visualizations: (1) Highly distinguishable categorical color map led to driller - viewers systematically explore the space. In contrast, length led to more scanners. Color-length pairs had the highest accuracy among the texture-length and length-length pairs. (2) Texture using black-and-white proportion might have been influenced by context and had the least power to reject distractors. When distractors were rejected properly (correct trials), texture had the same accuracy as color for the bivariate pairs. (3) Participants attempted to use the global statistical attributes to guide their attention and there may exist global visual processing prior to detailed visual inspection.

Acknowledgments

The work is supported in part by NSF IIS-1302755, NSF CNS-1531491, and NIST-70ANB13H181. The user study was funded by NSF grants with the OSU IRB approval number 2018B0080. Non-User Study The work is supported in part by NSF IIS-1302755, NSF CNS-1531491, and NIST-70ANB13H181. The last author would like to thank colleagues Prof. Jeremy Wolfe for suggesting the visual guidance work. Prof. Nuriit Gronau for introducing and discussing reverse-hierarchy theory, and Dr. Matthew S. Cain for error quantification. The authors also thank Prof. Andrew Leber for letting us use the EyeLink1000 facility to carry out the eye tracking study and Dr. Ayala Alon for training the eye tracker uses. Finally, the authors would like to thank Katrina Avery for her excellent editorial support and all participants for their time and contributions. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Certain commercial products are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the products identified are necessarily the best available for the purpose.

References

[1] D. Acevedo, J. Chen, and D. H. Laidlaw. Modeling perceptual dominance among visual cues in multilayered icon-based scientific visualizations. IEEE Visualization Posters, 2007.

[2] D. Acevedo and D. Laidlaw. Subjective quantification of perceptual interactions among some 2d scientific visualization methods. IEEE Transactions on Visualization and Computer Graphics, 12(5):1133–1140, 2006. doi: 10.1109/tvcg.2006.180

[3] J. Birtin. Semiology of graphics: diagrams, networks, maps. University of Wisconsin Press, 1967.

[4] J. Biederman. On processing information from a glance at a scene. ACM SIGGRAPH Workshop on User-oriented Design of Interactive Graphics Systems, 1977. doi: 10.1145/1024273.1024283

[5] T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. Visualization of eye tracking data: A taxonomy and survey. In Graphics Interfaces 2013, pp. 260–284. Wiley Online Library, 2017.

[6] R. Borgo, J. Kehrer, D. H. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen. Glyph-based visualization: Foundations, design guidelines, techniques and applications. Eurographics State of the Art Reports, pp. 39–63, 2013. doi: 10.2312/conf/EG2013/stars/039-063

[7] M. A. Borkin, Z. Bylinskii, N. W. Kim, C. M. Bainbridge, C. S. Yeh, D. Borkin, H. Pfister, and A. Oliva. Beyond memorability: Visualization recognition and recall. IEEE Transactions on Visualization and Computer Graphics, 22(1):519–528, 2016. doi: 10.1109/tvcg.2015.2467732

[8] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister. What makes a visualization memorable? IEEE Transactions on Visualization and Computer Graphics, 19(12):2306–2315, 2013. doi: 10.1109/tvcg.2013.234

[9] N. D. Bruce and J. K. Tsotsos. Salience, attention, and visual search: An information theoretic approach. Journal of Vision, 9(3):5–5, 2009. doi: 10.1167/9.3.5

[10] S. Brauckner, S. Grimm, A. Kanis, and M. E. Groller. Illustrative context-preserving exploration of volume data. IEEE Transactions on Visualization and Computer Graphics, 12(6):1559–1569, 2006. doi: 10.1109/tvcg.2006.96

[11] Z. Bylinskii, T. Judd, A. Oliva, A. Torralba, and F. Durand. What do different evaluation metrics tell us about saliency models? IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(3):740–757, 2019.

[12] M. S. Cain, S. H. Adamo, and S. R. Mitroff. A taxonomy of errors in interactive volume rendering. ACM Transactions on Visualization and Computer Graphics, 19(12):2818–2827, 2013. doi: 10.1109/tvcg.2013.205636

[13] J. Chen, G. Zhang, W. Chiou, D. H. Laidlaw, and A. P. Achour. Measuring the effects of scalar and spherical color maps on ensembles of DMBI tubes. IEEE transactions on visualization and computer graphics, 2019. doi: 10.1109/tvcg.2019.2989438

[14] J. Cohen. Statistical power analysis for the behavioral sciences. New York: Academic Press, 1988. doi: 10.4324/9780203771587

[15] G. Convertino, J. Chen, B. Yost, Y.-S. Ryu, and C. North. Exploring context switching and cognition in dual-view coordinated visualizations. Proceedings International Conference on Coordinated and Multiple Views in Exploratory Visualization, pp. 55–62, 2003. doi: 10.1109/cmrv.2003.1215003

[16] K. K. Evans, T. M. Haygood, J. Cooper, A.-M. Culpan, and J. M. Wolfe. A half-second glimpse often lets radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast. Proceedings of the National Academy of Sciences, 113(37):10292–10297, 2016. doi: 10.1073/pnas.1606187113

[17] L. Fidman, H. Toyoda, S. Seemann, B. Seppelt, L. Angell, J. Lee, B. Meller, and B. Reimer. What can be predicted from six distances of driver glances? Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 2805–2813, 2017. doi: 10.1145/3025453.3025929

[18] M. Harrower and C. A. Brewer. Colorbrewer.org: An online tool for selecting colour schemes for maps. The Cartographic Journal, 40(1):27–37, 2003. doi: 10.1002/9780470797587.ch134

[19] C. Healey and J. Enns. Attention and visual memory in visualization and computer graphics. IEEE Transactions on Visualization and Computer Graphics, 18(7):1170–1188, 2012. doi: 10.1109/tvcg.2011.127

[20] C. G. Healey, K. S. Booth, and J. T. Enns. High-speed visual estimation using preattentive processing. ACM Transactions on Computer-Human Interaction, 3(2):107–135, 1996. doi: 10.1145/230562.230563

[21] H.-Y. Ho, I.-C. Yeh, Y.-C. Lai, W.-C. Lin, and F.-Y. Cherng. Evaluating 2D flow visualization using eye tracking. Computer Graphics Forum, 34(3):501–510, 2015. doi: 10.1111/cgf.12662

[22] S. Hochstein and M. Ahissar. View from the top: Hierarchies and reverse hierarchies in the visual system. Neuron, 36(5):791–804, 2002. doi: 10.1016/s0896-6273(02)00578-x

[23] V. Interrante and C. Grosch. Strategies for effectively visualizing 3d flow with volume lic. IEEE Visualization, pp. 421–424, 1997. doi: 10.1109/visual.1997.663912

[24] T. Isenberg, P. Isenberg, J. Chen, M. Sedlmair, and T. Moller. A systematic review on the practice of evaluating visualization. IEEE Transactions on Visualization and Computer Graphics, 19(12):2818–2827, 2013. doi: 10.1109/tvcg.2013.126

[25] Y. Kim and A. Varshney. Saliency-guided enhancement for volume visualization. IEEE Transactions on Visualization and Computer Graphics, 12(5):925–932, 2006. doi: 10.1109/tvcg.2006.174

[26] W. K. Kundel, C. F. Nodine, and D. Carmody. Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. Investigative Radiology, 13(3):175–181, 1978. doi: 10.1097/00004424-19780500-00001

[27] D. H. Laidlaw, R. M. Kirby, C. D. Jackson, J. S. Davidson, T. S. Miller, M. Da Silva, W. H. Warren, and J. M. Tarr. Comparing 2D vector field visualization methods: A user study. IEEE Transactions on Visualization and Computer Graphics, 20(12):2394–2403, 2014. doi: 10.1109/tvcg.2014.236238
[30] R. S. Laramee, H. Hauser, H. Doleisch, B. Vrolijk, F. H. Post, and D. Weiskopf. The state of the art in flow visualization: Dense and texture-based techniques. Computer Graphics Forum, 23(2):203–221, 2004. doi: 10.1111/j.1467-8659.2004.00753.x

[31] R. Li and J. Chen. Toward a deep understanding of what makes a scientific visualization memorable. IEEE VIS and arXiv preprint, 2018.

[32] P. Ljung, J. Krüger, E. Groller, M. Hadwiger, C. D. Hansen, and A. Ynnerman. State of the art in transfer functions for direct volume rendering. Computer Graphics Forum, 35(3):669–691, 2016. doi: 10.1111/cgf.12934

[33] J. Mackinlay. Automating the design of graphical presentations of relational information. ACM Transactions on Graphics, 5(2):110–141, 1986. doi: 10.1145/22949.22950

[34] L. McNabb and R. S. Laramee. Survey of surveys (sos)-mapping the landscape of survey papers in information visualization. In Computer Graphics Forum, vol. 36, pp. 589–617. Wiley Online Library, 2017.

[35] T. Munzner. Visualization analysis and design. A K Peters Visualization Series. CRC Press, 2014.

[36] A. Oliva. Gist of the scene. Neurobiology of Attention, 696(64):251–258, 2005. doi: 10.1016/B978-012375731-9/50045-8

[37] A. Oliva and A. Torralba. Building the gist of a scene: The role of global image features in recognition. Progress in Brain Research, 155:23–36, 2006. doi: 10.1016/S0079-6123(06)55002-2

[38] K. Rayner. Eye movements and attention in reading, scene perception, and visual search. The Quarterly Journal of Experimental Psychology, 62(8):1457–1506, 2009.

[39] P. Rheingans and D. Ebert. Volume illustration: Nonphotorealistic rendering of volume models. IEEE Transactions on Visualization and Computer Graphics, 7(3):253–264, 2001. doi: 10.1109/2945.942693

[40] T. Ropinski, S. Oeltze, and B. Preim. Survey of glyph-based visualization techniques for spatial multivariate medical data. Computers & Graphics, 35(2):392–401, 2011. doi: 10.1016/j.cag.2011.01.011

[41] T. Ropinski, M. Specht, J. Meyer-Spradow, K. H. Hinrichs, and B. Preim. Surface glyphs for visualizing multimodal volume data. Vision, Modeling and Visualization, pp. 3–12, 2007.

[42] G. Ryan, A. Mosca, R. Chang, and E. Wu. At a glance: Pixel approximate entropy as a measure of line chart complexity. IEEE Transactions on Visualization and Computer Graphics, 25(1):872–881, 2019.

[43] A. Torralba, A. Oliva, M. S. Castelhano, and J. M. Henderson. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. Psychological Review, 113(4):766, 2006. doi: 10.1037/0033-295X.113.4.766

[44] A. M. Treisman and G. Gelade. A feature-integration theory of attention. Cognitive Psychology, 12(1):97–136, 1980. doi: 10.1016/0010-0285(80)90005-5

[45] M. O. Ward. Multivariate data glyphs: Principles and practice. In Handbook of Data Visualization, pp. 179–198. Springer Berlin Heidelberg, 2008. doi: 10.1007/978-3-540-33037-0

[46] C. Ware. Designing with a 2-D attitude. Information Design Journal, 10(3):258–265, 2001. doi: 10.1075/idj.10.3.07war

[47] C. Ware. Toward a perceptual theory of flow visualization. IEEE Computer Graphics and Applications, 28(2), 2008. doi: 10.1109/mcg.2008.39

[48] C. Ware and M. D. Plumlee. Designing a better weather display. Information Visualization, 12(3-4):221–239, 2013. doi: 10.1177/1473871612465214

[49] C. D. Wickens and C. M. Carswell. The proximity compatibility principle: Its psychological foundation and relevance to display design. Human Factors, 37(3):473–494, 1995. doi: 10.1518/001872095779049408

[50] J. Wolfe, M. Cain, K. Ehinger, and T. Drew. Guided search 5.0: Meeting the challenge of hybrid search and multiple-target foraging. Journal of Vision, 15(12):1106–1106, 2015. doi: 10.1167/15.12.1106

[51] J. M. Wolfe and I. S. Utochkin. What is a preattentive feature? Current Opinion in Psychology, 2018. doi: 10.1016/j.copsyc.2018.11.005

[52] H. Zhao and J. Chen. Bivariate separable-dimension glyphs can improve visual analysis of holistic features. arXiv preprint arXiv:1712.02333, 2017.