Picturing Bivariate Separable-Features for Univariate Vector Magnitudes in Large-Magnitude-Range Quantum Physics Data

Henan Zhao, Student Member, IEEE, and Jian Chen, Member, IEEE

Abstract—We present study results from two experiments to empirically validate that separable bivariate pairs for univariate representations of large-magnitude-range vectors are more efficient than integral pairs. The first experiment with 20 participants compared: one integral pair, three separable pairs, and one redundant pair, which is a mix of the integral and separable features. Participants performed three local tasks requiring reading numerical values, estimating ratio, and comparing two points. The second 18-participant study compared three separable pairs using three global tasks when participants must look at the entire field to get an answer: find a specific target in 20 seconds, find the maximum magnitude in 20 seconds, and estimate the total number of vector exponents within 2 seconds. Our results also reveal the following: separable pairs led to the most accurate answers and the shortest task execution time, while integral dimensions were among the least accurate; it achieved high performance only when a pop-out separable feature (here color) was added. To reconcile this finding with the existing literature, our second experiment suggests that the higher the separability, the higher the accuracy; the reason is probably that the emergent global scene created by the separable pairs reduces the subsequent search space.

Index Terms—Separable and integral dimension pairs, bivariate glyph, 3D glyph, quantitative visualization, large-magnitude-range.

1 INTRODUCTION

Bivariate glyph visualization is a common form of visual design in which a dataset is depicted by two visual variables, often chosen from a set of perceptually independent graphical dimensions of shape, color, texture, size, orientation, curvature, and so on [1], [2]. A bivariate glyph design has been used to show univariates for quantum physicists at National Institute of Standards and Technology (NIST) to examine simulation results; thanks to their teams Nobel-prize-winning scalable simulations, quantum physicists world-wide can now simulate at any scale. A critical quantum-physics analysis task is to understand spin (often depicted as vector) magnitude variations because these magnitudes showing atom behaviors are large in range and are often not continuous where the magnitudes can vary greatly in local regions.

On the visualization side, the initial design and evaluation of large-magnitude-range vector visualizations use scientific notation to depict digit and power as two concentric cylinders [3]: inside and outside tube lengths ($length_y$, $length_x$) are mapped to digit and power accordingly (aka splitVectors, Figure 1). A three-dimensional (3D) bivariate glyph scene of this splitVectors design (Figure 1) achieved up to ten times greater accuracy than the traditional direct linear univariate mapping (linear) (Figure 2) for reading a vector ratio between two vector magnitudes. However, this bivariate splitVectors glyph also increases task completion time for an apparently simple comparison task between two vectors in 3D. Linear is a significantly more efficient approach than their new solution.

One may frame this large-magnitude-range issue as a visual design problem: how can we depict a univariate quantity using bivariate visual features or glyphs to help quantum physicists examine complex spatial data? Intuitively, the last empirical study result on vector magnitude comparisons agrees well with a design consensus: to obtain a single magnitude at each location, the human visual system integrates these two component parts (digit and exponent terms) into one gestalt. This integration is referred to as holistic processing by Ware [4] in visualization and as feature binding by Treisman [5] in vision science. Both study how our visual system combines separate object features such as shapes, color, motion trajectories, sizes, and distances into the whole object. Ware [4] (G5.14) further recommends that: “If it is important for people to respond holistically to a combination of two variables in a set of glyphs, map the variables to integral glyph properties.” Since comparison is a recognition task, to represent a univariate vector magnitude we should always use integral properties (visual properties perceived together as a unit) or linear visualizations, instead of the separable (features manipulated and perceived independently) length pair in splitVectors.

Essentially, the current bivariate glyph design treats visual design as a bottom-up stimulus-driven composition in which the visual properties of object features (e.g., orientations and colors) are combined into single objects (here vectors). In this work, we challenge this consensus and argue that for one, feature-binding needs not occur at the object (vector) level and, for the other, this bivariate splitVectors gives viewers a correspondence challenge that does not arise.

—We present study results from two experiments to empirically validate that separable bivariate pairs for univariate representations of large-magnitude-range vectors are more efficient than integral pairs. The first experiment with 20 participants compared: one integral pair, three separable pairs, and one redundant pair, which is a mix of the integral and separable features. Participants performed three local tasks requiring reading numerical values, estimating ratio, and comparing two points. The second 18-participant study compared three separable pairs using three global tasks when participants must look at the entire field to get an answer: find a specific target in 20 seconds, find the maximum magnitude in 20 seconds, and estimate the total number of vector exponents within 2 seconds. Our results also reveal the following: separable pairs led to the most accurate answers and the shortest task execution time, while integral dimensions were among the least accurate; it achieved high performance only when a pop-out separable feature (here color) was added. To reconcile this finding with the existing literature, our second experiment suggests that the higher the separability, the higher the accuracy; the reason is probably that the emergent global scene created by the separable pairs reduces the subsequent search space.

Index Terms—Separable and integral dimension pairs, bivariate glyph, 3D glyph, quantitative visualization, large-magnitude-range.

1 INTRODUCTION

Bivariate glyph visualization is a common form of visual design in which a dataset is depicted by two visual variables, often chosen from a set of perceptually independent graphical dimensions of shape, color, texture, size, orientation, curvature, and so on [1], [2]. A bivariate glyph design has been used to show univariates for quantum physicists at National Institute of Standards and Technology (NIST) to examine simulation results; thanks to their teams Nobel-prize-winning scalable simulations, quantum physicists world-wide can now simulate at any scale. A critical quantum-physics analysis task is to understand spin (often depicted as vector) magnitude variations because these magnitudes showing atom behaviors are large in range and are often not continuous where the magnitudes can vary greatly in local regions.

On the visualization side, the initial design and evaluation of large-magnitude-range vector visualizations use scientific notation to depict digit and power as two concentric cylinders [3]: inside and outside tube lengths ($length_y$, $length_x$) are mapped to digit and power accordingly (aka splitVectors, Figure 1). A three-dimensional (3D) bivariate glyph scene of this splitVectors design (Figure 1) achieved up to ten times greater accuracy than the traditional direct linear univariate mapping (linear) (Figure 2) for reading a vector ratio between two vector magnitudes. However, this bivariate splitVectors glyph also increases task completion time for an apparently simple comparison task between two vectors in 3D. Linear is a significantly more efficient approach than their new solution.

One may frame this large-magnitude-range issue as a visual design problem: how can we depict a univariate quantity using bivariate visual features or glyphs to help quantum physicists examine complex spatial data? Intuitively, the last empirical study result on vector magnitude comparisons agrees well with a design consensus: to obtain a single magnitude at each location, the human visual system integrates these two component parts (digit and exponent terms) into one gestalt. This integration is referred to as holistic processing by Ware [4] in visualization and as feature binding by Treisman [5] in vision science. Both study how our visual system combines separate object features such as shapes, color, motion trajectories, sizes, and distances into the whole object. Ware [4] (G5.14) further recommends that: “If it is important for people to respond holistically to a combination of two variables in a set of glyphs, map the variables to integral glyph properties.” Since comparison is a recognition task, to represent a univariate vector magnitude we should always use integral properties (visual properties perceived together as a unit) or linear visualizations, instead of the separable (features manipulated and perceived independently) length pair in splitVectors.

Essentially, the current bivariate glyph design treats visual design as a bottom-up stimulus-driven composition in which the visual properties of object features (e.g., orientations and colors) are combined into single objects (here vectors). In this work, we challenge this consensus and argue that for one, feature-binding needs not occur at the object (vector) level and, for the other, this bivariate splitVectors gives viewers a correspondence challenge that does not arise.

—We present study results from two experiments to empirically validate that separable bivariate pairs for univariate representations of large-magnitude-range vectors are more efficient than integral pairs. The first experiment with 20 participants compared: one integral pair, three separable pairs, and one redundant pair, which is a mix of the integral and separable features. Participants performed three local tasks requiring reading numerical values, estimating ratio, and comparing two points. The second 18-participant study compared three separable pairs using three global tasks when participants must look at the entire field to get an answer: find a specific target in 20 seconds, find the maximum magnitude in 20 seconds, and estimate the total number of vector exponents within 2 seconds. Our results also reveal the following: separable pairs led to the most accurate answers and the shortest task execution time, while integral dimensions were among the least accurate; it achieved high performance only when a pop-out separable feature (here color) was added. To reconcile this finding with the existing literature, our second experiment suggests that the higher the separability, the higher the accuracy; the reason is probably that the emergent global scene created by the separable pairs reduces the subsequent search space.

Index Terms—Separable and integral dimension pairs, bivariate glyph, 3D glyph, quantitative visualization, large-magnitude-range.

1 INTRODUCTION

Bivariate glyph visualization is a common form of visual design in which a dataset is depicted by two visual variables, often chosen from a set of perceptually independent graphical dimensions of shape, color, texture, size, orientation, curvature, and so on [1], [2]. A bivariate glyph design has been used to show univariates for quantum physicists at National Institute of Standards and Technology (NIST) to examine simulation results; thanks to their teams Nobel-prize-winning scalable simulations, quantum physicists world-wide can now simulate at any scale. A critical quantum-physics analysis task is to understand spin (often depicted as vector) magnitude variations because these magnitudes showing atom behaviors are large in range and are often not continuous where the magnitudes can vary greatly in local regions.

On the visualization side, the initial design and evaluation of large-magnitude-range vector visualizations use scientific notation to depict digit and power as two concentric cylinders [3]: inside and outside tube lengths ($length_y$, $length_x$) are mapped to digit and power accordingly (aka splitVectors, Figure 1). A three-dimensional (3D) bivariate glyph scene of this splitVectors design (Figure 1) achieved up to ten times greater accuracy than the traditional direct linear univariate mapping (linear) (Figure 2) for reading a vector ratio between two vector magnitudes. However, this bivariate splitVectors glyph also increases task completion time for an apparently simple comparison task between two vectors in 3D. Linear is a significantly more efficient approach than their new solution.

One may frame this large-magnitude-range issue as a visual design problem: how can we depict a univariate quantity using bivariate visual features or glyphs to help quantum physicists examine complex spatial data? Intuitively, the last empirical study result on vector magnitude comparisons agrees well with a design consensus: to obtain a single magnitude at each location, the human visual system integrates these two component parts (digit and exponent terms) into one gestalt. This integration is referred to as holistic processing by Ware [4] in visualization and as feature binding by Treisman [5] in vision science. Both study how our visual system combines separate object features such as shapes, color, motion trajectories, sizes, and distances into the whole object. Ware [4] (G5.14) further recommends that: “If it is important for people to respond holistically to a combination of two variables in a set of glyphs, map the variables to integral glyph properties.” Since comparison is a recognition task, to represent a univariate vector magnitude we should always use integral properties (visual properties perceived together as a unit) or linear visualizations, instead of the separable (features manipulated and perceived independently) length pair in splitVectors.

Essentially, the current bivariate glyph design treats visual design as a bottom-up stimulus-driven composition in which the visual properties of object features (e.g., orientations and colors) are combined into single objects (here vectors). In this work, we challenge this consensus and argue that for one, feature-binding needs not occur at the object (vector) level and, for the other, this bivariate splitVectors gives viewers a correspondence challenge that does not arise.
Fig. 1: (a) Five bivariate configurations of univariate vector magnitude using the scientific notation. This example shows vector magnitude $440 \times 10^2$ with each depicted using two values: digit 4.4 and power 2. (b) Contours shown using the length-color (LC) pair. This work demonstrates that more separable pairs lead to efficient local comparisons of a couple of vectors. Global scene structures guided by more separable dimensions also led to more accurate and highly efficient strategies without the needs for synthesizing univariate magnitudes.

Fig. 2: Large-magnitude-range contours computed from a simulation result are shown using five bivariate feature-pairs and linear representation.
when integral dimensions or direct linear encoding is used - the need to relate these two quantitative variables to their visual features hampers its efficiency. There are two ways to describe human experiences. If the visual field has only one or two objects at a time and if splitVectors of length-pairs are used, a viewer would take longer to process information in order to determine which length is exponent and which is mantissa. We suggest that in this case, if correspondence errors account for the temporal costs with bivariate feature pairs, then techniques preventing this type of error can be as effective as a direct linear encoding without time-consuming correspondence search.

Now, for example, if we increase the feature separability by replacing the exponent-to-length mapping in Figures 1 and 2 to exponent-to-color mapping in Figures 1 and 2 for comparison tasks, it would be counterproductive for our attention first to visit each glyph to compute the magnitude (driven by bottom-up process). Instead, the global categorical color (hue) can guide our attention to first compare and categorize the colors, prior to a visual comparison when the colors are the same (within the same exponent) to compare vector lengths. In this case, no object-level binding is needed as long as the correspondence between the two visual features can be easily understood.

Further considering the viewers’ task relevant to multiple vector objects (e.g., find maximum), the same sequential viewing looking for the subregions followed by length-inspection works equally well. The reason is that feature binding need not occur at the object level, but can be done first at the scene level, and scene context benefits the reduction of search regions when there is no ambiguity in finding the correspondences. Coincidentally, this first impression of the data to drive statistical information is also called holistic or global pattern processing [6]. Wolfe called features guiding this top-down task-driven attention behaviors as scene features [7]. Here, we may refer to Ware’s holistic features [4] as object-level holistic and Biederman’s [6] and Wolfe’s [7] as scene-level holistic design thinking.

Reducing correspondence error is influenced by the choices of separable dimensions. According to Treisman [8] and Wolfe [9], the initial preattentive phase is the major step towards improved comprehension, more important than the attentive phase. We select the “most recognizable” features as color (Figures 1a, 1b, 2a, 2b) and texture (Figures 1c, 2c) and size (Figures 1d, 2d) dimensions. Size and color are preattentive and permit visual selection at a glance, at least in two-dimension (2D). We purposefully select texture patterns by varying the amount of dark on white, thus introducing luminance variations when many vectors are examined together (Figure 2d). Compared to the continuous random noise in Urness et al. [10], ours is for discrete quantities and thus uses regular scale variations. When coupled with integral and separable dimensions, we anticipated that preattentive pop-out features in separable dimension pairs might reduce correspondence errors compared to integral dimensions. Following this logic, we hypothesize that highly distinguishable separable dimension pairs might erase the costs associated with the correspondence errors to reduce task completion time and be more accurate.

We tested this hypothesis in two experiments with six tasks using four dimension pairs to compare against the length_x-length_y (separable) in Zhao et al. [3]; length_y-length_x (integral), length-color (separable), length-texture (separable), and length_y-color/length_x (redundant and separable). Since we predicate that separable dimensions with more preattentive features would reduce the task completion time, length-color and length_y-color/length_x might achieve more efficiency without hampering accuracy than other bivariate feature pairs.

This work makes the following contributions:

- Empirically validates that bivariate-glyphs encoded by highly separable dimensions would improve comparison task complete time (Exp 1).
- Is the first to explain the benefits of the global scene-guidance which expands the widely accepted object-level bivariate glyph design in visualization (Exp 2).
- Offers a rank order of separable variables for 3D glyph design and shows that the separable pairs length-color and length-texture are among the most effective and efficient feature pairs.

2 Theoretical Foundations in Perception and Vision Sciences

At least three perceptual and vision science theories have inspired our work: integral and separable dimensions [11], preattentive features ranking [12], [13], [14], [15], and monotonicity [2].

Integral and Separable Dimensions. Garner and Felfoldy’s seminal work on integral and separable dimensions [11] has inspired many visualization design guidelines. Ware [4] suggests a continuum from more integral to more separable pairs: (red-green)-(yellow-blue), size_x-size_y, color-shape/size/orientation, motion-shape/size/orientation, motion-color, and group position-color. His subsequent award-winning bivariate study [2] using hue-size, hue-luminance, and hue-texton (texture) supports the idea that more separable dimensions of hue-texton lead to higher accuracy. Our work follows the same ideas of applying integral and separable dimensions but differs from Ware’s texture selection in two important aspects. First, the Ware study focuses on finding relationships between two independent data variables, and thus its tasks are analytical; Second, our texture uses the amount of black and white to show luminance variations, in contrast to the discrete shape variation in texts. We anticipate that ours will be more suitable to continuous quantitative values [16]. No existing work we know of has studied whether or not the separable features can facilitate global comparisons and can be scaled to 3D vector field analysis.

Feature-Binding and Scene-Guidance Theories. Treisman and Gelades feature-integration theory of attention [8] showed that the extent of difference between target and distractors for a given feature affects search time. This theory may explain why splitVectors was time consuming; the similarity of the two lengths may make them interfere with each other in the comparison, thus introducing temporal cost. What we “see” depends on our goals and expectations. Wolfe et al. propose the theory of “guided search” [7], [17], a first attempt to incorporate users’ goals into viewing, suggesting that what users see is based on users’ goals.
Wolfe et al. further suggest that color, texture, size, and spatial frequency are among the most effective features in attracting the user’s attention.

**Preattentive and Attentive Feature Ranking.** Human visual processing can be faster when it is preattentive, i.e., perceived before it is given focused attention [8]. The idea of pop-out highlighting of an object is compelling because it captures the user’s attention against a background of other objects (e.g., in showing spatial highlights [18]). Visual features such as orientation and color (hue, saturation, lightness) can generate pop-out effects [8,19]. Healey and Enns [20] in their comprehensive review further remark that these visual features are also not popped-out at the same speed: hue has higher priority than shape and texture [21].

Visual features also can be responsible for different attention speeds, and color (hue) and size (length and spatial frequency) are among those that guide attention [16]. For visualizing quantitative data, MacKinlay [14] and Cleveland and McGill [15] leverage the ranking of visual dimensions and suggest that position and size are quantitative and can be compared in 2D. Casner [21] expands MacKinlay’s APT by incorporating user tasks to guide visualization generation. Demiralp et al. [22] evaluate a crowdsourcing method to study subjective perceptual distances of 2D bivariate pairs of shape-color, shape-size, and size-color. When adopted in 3D glyph design, these studies further suggest that the most important data attributes should be displayed with the most salient visual features, to avoid situations in which secondary data values mask the information the viewer wants to see.

**Monotonicity.** Quantitative data encoding must normally be monotonic, and various researchers have recommended a coloring sequence that increases monotonically in luminance [24]. In addition, the visual system mostly uses luminance variation to determine shape information [25]. There has been much debate about the proper design of a color sequence for displaying quantitative data, mostly in 2D [26] and in 3D shape volume variations [27]. Our primary requirement is that users be able to read large or small exponents at a glance. We chose four color steps in the first study and up to seven steps in the second study for showing areas of large and small exponents that are mapped to a hue-varying sequence. We claim not that these color sequences are optimal, only that they are reasonable solutions to the design problem [26].

### 3 Experiment I: Local Discrimination and Comparisons

The goal in this first experiment is to quantify the benefits of separable pairs for visual processing of a few items. This section discusses the experiment, the design knowledge we can gain from it, and the factors that influence our design.

#### 3.1 Methods

**3.1.1 Bivariate Feature-Pairs**

We choose five bivariate feature-pairs to examine the comparison task efficiency of separable-integral pairs.

- **Length$_x$-color/Length$_y$ (redundant and separable)** (Figure 1b). This pair compared to Length$_y$-Length$_x$ adds a redundant color (luminance and hue variations) dimension to the exponent and the four sequential colors are chosen from Colorbrewer [26].
- **Length-color (separable)** (Figure 1c). This pair maps exponents to color. Pilot testing shows that correspondence errors in this case would be the lowest among these five feature-pairs.
- **Length-texture (separable)** (Figure 1d). Texture represents exponents. The percentage of black color (Bertin [28]) is used to represent the exponential terms 0 (0%), 1 (30%), 2 (60%) and 3 (90%), wrapped around the cylinders in five segments to make them visible from any viewpoint.
- **Length$_x$-length$_y$ (splitVectors [3], separable)** (Figure 1e). This glyph uses splitVectors [3] as the baseline and maps both digit and exponent to lengths. The glyphs are semitransparent so that the inner cylinders showing the digit terms are legible.

Feather-like fishbone legends are added at each location when the visual variable length is used. The *tick-mark band* is depicted as subtle light-gray lines around each cylinder. Distances between neighboring lines show a unit length legible at certain distance (Figure 1 rows 2 and 3).

**3.1.2 Hypotheses**

Given the analysis above and recommendations in the literature, we arrived at the following working hypotheses:

- Exp I.H1. (Overall). The length-color feature-pair can lead to the most accurate answers.
  
  Several reasons lead to this conjecture. Color and length are separable dimensions. Colors can be detected quickly, so length and color are highly distinguishable. Compared to the redundant length$_y$-color/length$_x$, length-color reduces density since the feature-pairs are generally smaller than those in length$_y$-color/length$_x$.

- Exp I.H2. (Integral-separable). Among the three separable dimensions, length-color may lead to the greatest speed and accuracy and length-texture would be more effective than splitVectors.

  The hypothesis could be supported because color and length are highly separable.

- Exp I.H3. (Redundant hypothesis). The redundant pair length$_y$-color/length$_x$ will reduce time compared to splitVectors.

  This hypothesis could be supported because redundancy increases information processing capacity.

**3.1.3 Tasks**

Participants perform the following three task types as in Zhao et al. [3] so that results are comparable. They had unlimited time to perform these three tasks.

**Exp1.Task 1 (MAG): magnitude reading** (Figure 3a). What is the magnitude at point A? One vector is marked by a red triangle labeled “A”, and participants should report the magnitude of that vector. This task requires precise numerical input.

**Exp1.Task 2 (RATIO): ratio estimation** (Figure 3b). What is the ratio of magnitudes of points A and B? Two vectors are
to generate the data by randomly sampling some quantum
in Zhao et al. [3]: We replicate their data selection method
Because we are interested in comparing our results to those
3.1.4 Data Selection
Two vectors are marked
mark shapes or decipher each vector magnitude and com-
quantitative task [14]. Participants can either compare the
ratio judgment is the most challenging quantitative task [14]. Participants can either compare the
glyph shapes or decipher each vector magnitude and com-
Exp1.Task 3 (COMP): comparison (Figure 3c). Which
magnitude is larger, point A or B? Two vectors are marked
with red triangles and labeled “A” and “B”. Participants
select their answer by directly clicking the “A” or “B”
answer buttons. This task is a simple comparison between
two values and offers a binary choice of large or small.

(a) Exp1 MAG task: What is the magnitude of the vector at
point A? (answer: 636.30)

(b) Exp1 RATIO task: What is the ratio of the magnitude
between the vectors at points A and B? (answer: 3.60)

(c) Exp1 COMP task: Which magnitude is larger, point A or
point B? (answer: A on the right.)

Fig. 3: Experiment 1: Local discrimination and comparison
tasks.

marked with two red triangles labeled “A” and “B”, and
participants should estimate the ratio of magnitudes of these
two vectors. The ratio judgment is the most challenging
quantitative task [14]. Participants can either compare the
glyph shapes or decipher each vector magnitude and com-
pute the ratio mentally.

Exp1.Task 3 (COMP): comparison (Figure 3c). Which
magnitude is larger, point A or B? Two vectors are marked
with red triangles and labeled “A” and “B”. Participants
select their answer by directly clicking the “A” or “B”
answer buttons. This task is a simple comparison between
two values and offers a binary choice of large or small.

3.1.4 Data Selection
Because we are interested in comparing our results to those
in Zhao et al. [3]: We replicate their data selection method
to generate the data by randomly sampling some quantum
physics simulation results and produce samples within 3D
boxes of size 5×3×3. There are 445 to 455 sampling locations
in each selected data region.

We select the data satisfying the same following condi-
tions: (1) the answers must be at locations where some
context information is available, i.e., not too close to the
boundary of the testing data. (2) no data sample is repeated
to the same participant; (3) Since data must include a broad
measurement, we select the task-relevant data from each
exponential term of 0 to 3.

3.1.5 Empirical Study Design

Design and Order of Trials. We use a within-subject design
with one independent variable of bivariate quantitative
feature-pair (five types). Dependent variables are relevant
task completion time. We also collect participants’ confi-
dence levels. The accuracy measure follows Zhao et al. [3] to
study how sensitive a method is to error uncertainty based
on the relative error (RE) or fractional uncertainty, calculated
as $RE = \frac{\text{correct answer} - \text{participant answer}}{\text{correct answer}}$.
This measure is used for MAG and RATIO tasks. The benefit
of this approach is that it takes into account the value of
the quantity being compared and thus provides an accurate
view of the errors.

Table 1 shows that participants are assigned into five
blocks in a Latin-square order, and within one block the
order of the five feature-pair types is the same. Participants
perform tasks with randomly selected datasets. Each par-
ticipant performed 60 trials (3 tasks × 4 random data × 5 feature-pairs). These four random data are from four
exponent ranges.

Participants. We diversify the participant pool as much
as possible, since all tasks can be carried out by those
with only some science background. Twenty participants
(15 male and 5 female, mean age = 23.3, and standard
deivation = 4.02) participated in the study, with ten in com-
puter science, three in engineering, two in chemistry, one in
physics, one in linguistics, one in business administration,
one double-major in computer science and math, and one
double-major in biology and psychology. The five females
are placed in each of the five blocks (Table 1). On average,
participants spent about 40 minutes on the computer-based
tasks.

Procedure, Environment, and Interaction. Participants
are greeted and complete an Institutional Review Board
(IRB) consent form. All participants had normal or

Table 1: Experiment I design: 20 participants are assigned
to one of the five blocks and use all five bivariate pairs. Here, $L_y L_z$: length$_y$-length$_z$ (splitVectors), $L_y L_x$: length$_y$-
length$_x$, LC: length-color, LT: length-texture, and LCL:
length$_y$-color/length$_z$.

| Block | Participant | Feature-pair |
|-------|-------------|--------------|
| 1     | P1, P6, P11, P16 | splitVectors, $L_y L_z$, LC, LT, LCL |
| 2     | P2, P7, P12, P17 | $L_y L_z$, LC, LCL, splitVectors, LT |
| 3     | P3, P8, P13, P18 | LC, LCL, LT, splitVectors, $L_y L_z$ |
| 4     | P4, P9, P14, P19 | LT, $L_y L_z$, splitVectors, LCL, LC |
| 5     | P5, P10, P15, P20 | LCL, LT, $L_y L_z$, LC, splitVectors |
TABLE 2: Summary statistics by tasks. The significant main effects and the high effect size (ES) are in **bold** (none in these observations) and the medium effect size is in *italic*. Effect size is eta-square labeled “small” (0.01 – 0.06), “medium” [0.06–0.14], and “large” > 0.14 effects following Cohen [29]. Post-hoc Tukey grouping results are reported "medium" corrected display with resolution 1920.

| Task | Variables | Significance | ES |
|------|-----------|--------------|----|
| **MAG** | time | $F_{(4,384)} = 6.8, \ p < 0.0001$ | 0.07 |
| relative error | $L_C, L_T, LCL, \text{splitVectors}$ $> L_x, L_x$ | 0.01 |
| **RATIO** | time | $F_{(4,395)} = 6.2, \ p < 0.0001$ | 0.06 |
| Three groups: A: LC, splitVectors, LT | 0.06 |
| B: splitVectors, LT, LCL | 0.39 |
| C: LT, LCL, $b_yL_x$ | 0.01 |
| **COMP** | time | $F_{(4,395)} = 10.4, \ p < 0.0001$ | 0.09 |
| Three groups: A: LCL, LT, LC | 0.01 |
| B: splitVectors | 0.04 |
| C: splitVectors, $L_yL_x$ | 0.03 |
| accuracy | $\chi^2 = 0.4, \ p = 0.98$ | 0.03 |

corrected-to-normal vision and passed the Ishihara color-blindness test. They filled in the informed consent form (which described the procedure, risks and benefits of the study) and the demographic survey. We showed feature-pair examples and trained the participants with one trial for every feature-pair per task. They were told to be as accurate and as quickly as possible, and that accuracy was more important than time. They could ask questions during the training but were told they could not do so during the formal study. Participants practiced until they fully understood the feature-pairs and tasks. After the formal study, participants filled in a post-questionnaire asking how these feature pairs supported their tasks and were interviewed for their comments.

Participants sat at a 27” BenQ GTG XL 2720Z, gamma-corrected display with resolution 1920 × 1080. The distance between the participants and the display was about 50cm. The minimum visual angle of task-associated glyphs was 0.2° in the default view where all data points were visible and filled the screen.

Participants could rotate the data and zoom in and out. Lighting placement and intensity were chosen to produce visualization with contrast and lighting properties appropriate for human assumptions and the spatial data. The screen background color was neutral stimulus-free gray background to minimize the discriminability and appearance of colors. Using black or white background colors will make the black and white texture stimuli disappear thus bias the results.

3.2 Experiment I: Results and Discussion

3.2.1 Analysis Approaches

We collected 400 data points for each task. In preparing the accuracy and task completion time for analysis, a trial was considered to have an answer of the first type of correspondence error if responses’ exponent value did not match the correct one for the MAG task. This correspondence errors occurred when participants had trouble differentiating the levels within a encoding.

We detected 11 instances of the first type of correspondence errors from MAG (these trials comprised 2.75% of the total: three splitVectors, five $length_y$-$length_x$, one $length-texture$, and two $length-y$-$color/length_x$). This correspondence error appeared to be influenced by the integral-separable dimension as well and the integral dimension $length_y$-$length_x$ had the highest (5) and $length-color$ had no instances. We used only the remaining correct ones in the statistical analysis because these errors would mask all other data by being at least one order of magnitude larger. For the remaining data in MAG and all data in RATIO and COMP tasks, we used standard outlier detection by first calculating the mean and standard derivation across all trials for each participant and pruning any trials that were +/- two standard derivations from that participant’s mean. With this approach, no outlier was detected in the MAG, RATIO, and COMP tasks.

Table 2 and Figure 4 show the $F$ and $p$ values computed with SAS one-way measures of variance for task completion time. A post-hoc analysis using Tukey’s Studentized Range test (HSD) was performed when we observed a significant main effect. When the dependent variable was binary (i.e., answer correct or wrong), we used a logistic regression and estimated parameters using the mean and standard derivation across all trials for each participant and pruning any trials that were +/- two standard derivations from that participant’s mean. With this approach, no outlier was detected in the MAG, RATIO, and COMP tasks.

Table 2 and Figure 4 show the $F$ and $p$ values computed with SAS one-way measures of variance for task completion time. A post-hoc analysis using Tukey’s Studentized Range test (HSD) was performed when we observed a significant main effect. When the dependent variable was binary (i.e., answer correct or wrong), we used a logistic regression and estimated parameters using the mean and standard derivation across all trials for each participant and pruning any trials that were +/- two standard derivations from that participant’s mean. With this approach, no outlier was detected in the MAG, RATIO, and COMP tasks.

### 3.2.2 Overview of Study Results

Our results clearly demonstrated the benefits in terms of task completion time of separable dimensions for comparison. We observed a significant main effect of feature-pair type on task completion time for all three tasks MAG, RATIO, and COMP, and the effect sizes were in the medium range (Table 2, Figure 4). Length-texture was the most efficient approach. For COMP, $length-color$, $length-texture$, and $length_y$-$color/length_x$ were most efficient for simple two-point comparison (Figure 4).

### 3.2.3 Separable Dimensions Are Better Than Integral Dimensions for Local Comparisons

Our separable-integral hypothesis (H1) was supported. In the MAG tasks, the integral $length_y$-$length_x$ was least efficient and all other separable-pairs were in a separate group, the most efficient one (Figure 4a). In the RATIO tasks, $length-color$, $length-texture$, and splitVectors were the most efficient group (Figure 4b); in the COMP tasks, the redundant $length_y$-$color/length_x$, $length-color$, and $length-texture$ were in the most efficient group (Figure 4c).

**SplitVectors** was not as bad as we originally thought in handling correspondence errors, especially for the quantitative reading tasks of MAG and RATIO. SplitVectors belonged to the same efficient post-hoc group as $length-color$ and $length-texture$ for the RATIO tasks and these three were also most efficient for MAG.
We speculate that this result may indicate that when the comparison set size was small, participants did not need scene-level information to achieve accuracy. We anticipate that when the search space set-size increases, the search will become time-consuming and the lack of scene-level features would increase correspondence error and thus reduce effectiveness. We observe this in Experiment II.

3.2.4 Separable pairs of length-color and length_y-color/length_x achieved comparable efficiency to direct linear glyph

Critical for motivating this experiment was whether the separable pairs supported COMP and how the separable pairs compared in efficiency to the direct mapping. Since our study had the same numbers of sample data as Zhao et al. [3], we then performed a one-way t-test to compare the direct linear encoding in Zhao et al. [3]. Our separable-hypothesis (H2) was supported and our results indicated that COMP (judging large or small) from separable variables was no more time-consuming than direct linear glyphs. Our post-hoc analysis showed that length-color, length-color/length, and linear were in the same post-hoc group, i.e. that there were no significant differences between these features. We also observed that splitVectors dropped to the least efficient or most error-prone post-hoc groups (Fig. 4c). This result replicated the former study results in Zhao et al. [3] by showing that splitVectors impaired comparison efficiency.

This result may be explained by the idea that the highly separable pairs may turn the comparison into a single-dimension digit comparison tasks, since a viewer could quickly resolve the two exponents and thus reduce the correspondence error introduced by the splitVectors design.

3.2.5 Redundant Feature-Pairs Were Efficient

We also confirmed hypothesis H3. We were surprised by the large performance gain with the redundant encoding length_y-color/length_x of mapping color and length to the exponents in splitVectors. With redundant encoding, the relative error was significantly reduced and task completion time was much shorter (significantly shorter for MAG and COMP tasks). While Ware [4] confirmed that redundancy encoding was integrated into the encoded dimension, in our case, where color and size were separable, we suggested that the redundancy worked because participants could use either length or color in different task conditions. Since we could also consider that length_y-color/length_x was a redundant encoding with length_y-length_x and did better than length_y-color/length_x in some tasks (MAG and COMP), we may arrive at a design recommendation: when integral dimensions of length_y-length_x were less accurate, adding more separable color could compensate to aid participants in their tasks.

We had two considerations in setting up this experiment. The first was a statement about feature design relevant to the global holistic experience. If the vector field contains one

---

Fig. 4: Task completion time and relative error or accuracy by tasks. The horizontal axis represents the mean task completion time while the vertical axis showing the accuracy or relative error. Same letters represent the same post-hoc analysis group. Colors label the feature-pair types. All error bars represent 95% confidence interval.
object at a time, then the integral and separable dimensions and associated correspondence error may explain our object experience as we have shown in Experiment I. However, when the binding problem is raised by looking at multiple vectors, it is possible that object binding does not occur at the individual object-level but rather at the scene-level first, perhaps governed by global gestalt features.

The second consideration is relevant to the correspondence errors when managing holistic global viewing experiences that may not be significant when a few items are compared. Generally, subjective reports from the first study indicate that length-color and length-texture show similar perceptual speed. For a feature to actually guide attention, Wolfe suggests that a just-noticeable-difference for that feature is not sufficient and one must also look at feature distractors, whether or not they are heterogeneous, and that the efficiency of a scene guidance will decline as a function of the degree of distractor variation. Efficiency will be achieved if the target and distractors are “linearly separable” meaning that a line can be drawn separating the target from the distractors in the feature space. This is similar to the studies of Acevedo et al. for saliency measures and Urness et al. for “texture stitching”. Acevedo et al. attempted to show that features can be segmented and Urness et al. show boundaries from continuous flow fields using spatial frequency. Chung et al. showed the order and just noticeable differences of visual stimuli and suggested that size and hue had highest accuracy in 2D and texture value introduces greater ordering effects. In our study, we believe that colors (especially in multiple hues) are categorical and thus should produce better perceptual speed than texture and length. Performance of texture may decline faster than color as the data range increases because our vision is not as sensitive to luminance-variation as to hues. The efficiency of color in Experiment I could well arise because the range (of 4) was not large enough. The current study expands the data range from the single level in the first study to five ranges \( [3, 7] \) to understand feature-pair scalability. SplitVectors produces the second type of correspondence error between two lengths which can challenge human eyes to see the two component parts.

4.2 Method

4.2.1 Feature-Pairs

We used length-color, length-texture, and baseline splitVectors in Experiment II. These three visualizations were chosen because length-color and length-texture are among the best feature-pairs from Experiment I and because color and texture are among the most separable features according to Ware. To introduce a correspondence error or “distractor” experience, we vary the data range from the 4 levels in experiment I to 3-7 levels in Experiment II.

We had two reasons to use categorical hue instead of quantitative colormaps. The first was based on the subjective observation comparing a categorical colormap from Colorbrewer and a segmented continuous colormap by the number of exponents generated from the extended blackbody colormap. As we can see, the boundary detection with these colormaps might be associated with data density. We found that unless the data density was reasonably high, detecting the boundaries using continuous colormaps was harder than the Colorbrewer colormaps. The second reason is that the initial at-a-glance global statistical summary of the scene depends on categorical information - then categorical visual encoding may be more suitable. An informal observation on texture choices was that detecting maximum and minimum

![Fig. 5: Experiment II: An example data using a categorical and a segmented continuous colormaps with two data densities. The boundaries between the data categories are more recognizable when the data are dense in (a) and (c). The boundaries are more difficult to recognize in (b) than in (d). We use the categorical colormaps in Experiment II.](image)
regions was easier than any intermediate regions.

4.2.2 Hypotheses
We had the following hypotheses:

- **Exp I.H1. (Accuracy).** More separable pairs will be more effective. We thus anticipate a rank order of effectiveness from high to low: length-color, length-texture, and splitVectors.
  
  While we did not see a significant main effect in Experiment I and we believed that for tasks related to multiple objects (vectors), pop-out color features would reduce the correspondence error and facilitate scene-level feature binding.

- **Exp I.H2. (Correspondence error).** Less separable pairs would lead to more first type correspondence errors, when participants would choose the wrong exponent level.

- **Exp I.H3. (User behavior).** More separable feature-pairs would lead to optimal users’ behaviors: i.e., participants can quickly locate task-related regions for tasks that demand looking among many vectors.

4.2.3 Tasks
Participants performed three tasks in which they had to compare all vectors to obtain an answer.

**Exp II.Task 1 (SEARCH):** A vector search within 20 seconds (Figure 6a). Find the vector with magnitude X within 20 seconds. The target vector was shown at the bottom-right corner of the screen. Participants were asked to find this vector.

**Exp II.Task 2 (MAX):** An extreme value search within 20 seconds (Figure 6b). Within 20 seconds, locate the point has maximum magnitude when the exponent is X. X in the study was a number from 0 to the maximum exponent ($\in [2, 6]$). This was a global task requiring participants to find the extremum among many vectors.

**Exp II.Task 3 (NUMEROSITY):** within 2 seconds, estimate the total number of vector exponents (Figure 6c). Estimate the total number of vector exponents in the entire vector field within 2 seconds. Data are randomly chosen and modified to produce the 3 to 7 range. No data is used repeatedly in this experiment.

4.2.4 Data Choices
Data were first sampled using the same approach as Experiment I. We then modified the exponent range from 3 to 7 for the three tasks by normalizing the data to the desired new data range. Doing this let us preserve the critical domain-specific data attributes of their spatial structures and only altered the magnitude range to improve the applicability and reuse of our study results.

Prior literature used both synthetic data and real-world data to construct the data visualization as test scenarios, enabling tight control over the stimulus parameters. Most of the synthetic data in these studies were generated to replicate real-world data characteristics and others were explained in fictitious use scenarios. The goal was primarily to prevent preconceived user knowledge about the domain-specific attributes. As a result, the synthetic data strike the right balance between real-world uses and the data characteristics. In our cases, replicating characteristics in quantum physics data was challenging and indeed impossible, since atom behaviors in high-dimensional space were largely unknown and thus were not easily simulated. Our approach was therefore to randomly sample quantum physics simulation results to capture domain-specific attributes and then modify the data to suit evaluation purposes. We showed our data to our physicist collaborators to ensure their validity.

4.2.5 Empirical Study Design
**Dependent and Independent Variables.** We used a within-subject design with two independent variables of feature-pair (three levels: baseline splitVectors, length-color, and length-texture) and exponent range (five levels: 3-7). The dependent variable was relative error. We did not measure time since all tasks were time-constrained.
were used to give participants enough time to develop strategies. For NUMEROSITY tasks, the design runs 4 repetitions, resulting in 3 (feature-pair) strategies. For SEARCH and MAX, we measured relative error (which was the percentage the reported value was away from the ground truth) with SAS repeated measure for each task. Procedure, interaction, and environment were the same as those in the Experiment I.

Self-Reporting Strategies. Several human-computer interaction (HCI) approaches can help observe users’ behaviors. Answering questions can assist us to determine not just which technique is better but also the strategies humans adopt. For example, cognitive walkthrough (CTW) measures whether or not the users actions match the designers’ pre-designed steps. Here we predicted that participants would use the global scene-features as guidance to accomplish tasks. We interviewed participants and asked them to verbalize their visual observations in accomplishing tasks.

4.2.6 Participants

Eighteen new participants (12 male and 6 female, mean age = 23.8, and standard deviation = 4.94) of diverse backgrounds participated in the study (seven in computer science, four in computer engineering, two in information systems, three in engineering, one in business school, and one in physics). Procedure, interaction, and environment were the same as those in the Experiment I.

4.3 Experiment II: Results and Discussion

We collected 810 data points per task for the first two tasks of SEARCH and MAX and 1080 points for the third NUMEROSITY task.

4.3.1 Summary Statistics

For SEARCH and MAX tasks, we measured relative error (which was the percentage the reported value was away from the ground truth) with SAS repeated measure for significant main effects and the high effect size are in italic. Effect size is Cohen’s d for tasks SEARCH and MAX, and Cramer’s V for task NUMEROSITY (NUM). Post-hoc Tukey grouping results are reported for significant main effects, where > means statistically significantly better and enclosing parentheses mean they belong to the same Tukey group. Here, LC: length-color and LT: length-texture.

Fig. 7: Experiment II: Relative error in SEARCH and MAX and accuracy in NUM. Same letters represent the same post-hoc analysis group. All error bars represent 95% confidence intervals.

Literature suggests they simply look for the longest cylinder from the splitVectors. Length-color also led to the most accurate answers, and now splitVectors was better than length-texture for NUMEROSITY tasks. This result can be explained by participants’ behaviors - more than half the participants suggested they simply look for the longest cylinder from the
splitVectors since they know the numerical values in the test were continuous. This behavior deviated from our original purpose of testing the global estimate but did show two perspectives in favor of this work: (1) participants developed task-specific strategies during the experiment for efficiency; (2) 3D length still supported a reasonable pop-out feature, even though it was not as effective as color.

These subjective behaviors through self-report suggested that they adopted a sequential task-driven viewing strategy to first obtain gross regional distribution of task-relevant exponents. After this, a visual comparison within the same exponent region was performed especially when the features pop-out globally as scene features. With these two steps, judging large or small or perceiving quantities accurately from separable variables would not use object-binding. Participants in our study used a top-down control process to utilize these spatial constraints regardless of feature types (for NUMEROSITY) and modify how global structures are used to see the features between tasks.

4.3.3 The Cost of Correspondence Errors

Reducing correspondence error was influenced by the choices of separable dimensions. Our second hypothesis H2 (correspondence error) was also supported. We first tested the first type of correspondence error (those answers with different exponent values) in MAX and SEARCH in the same way as in Experiment I. We saw 36 instances from SEARCH (about 4.4% of all samples, or 1 length-color, 20 length-texture, and 15 splitVectors); 59 instances from MAX (about 7% of all samples or 6 length-color, 38 length-texture, and 15 splitVectors). These results when combined with those in Experiment I confirmed that length-texture had worse first type correspondence error. However when viewers in the correct data sub-categories, they could obtain as accurate answers as length-color.

All participants commented on how the number of powers in the data affected their effectiveness. For length-texture, 10 participants remarked that it was difficult to differentiate adjacent powers when the total power level is around 4-5 for length-texture. The white and black textures were very easy to perceive. All but two participants agreed that length-color could perhaps support up to 6. Chung et al. [32] studied ordering effects and it would be challenging to compare ours to their results because their visual stimuli were not shown as a scene-feature but a stimului alone. More than half of the participants felt that effectiveness of splitVectors was not affected by changing the number of powers, since they looked for the longest outer cylinder to help find the answer. These results may suggest that sub-region selection with length-texture can perhaps be better designed with interfaces when the users can interactively select a texture level.

5 General Discussion

We discuss the results from both experiments and suggest future directions.

5.1 Separable Dimensions for Univariate Data Visualization for Large-Range Quantum Physics Data

The results of Experiment I showed that separable dimensions could achieve the same efficiency as direct linear visualizations for the COMP task and was always more efficient than integral pairs. For these local-tasks, we didn’t observe significant error reduction. The results from Experiment II studied the rank order of the separable pairs and found that more separable pairs also improved accuracy for global tasks. length-texture and splitVectors in both experiments led to higher correspondence errors than length-color.

Visual variables that are separable (i.e. manipulated and perceived independently) would initially be considered problematic for encoding univariate data because of the known object-level feature-binding challenges involving in achieving integrated numerical readings by combining two visual features. Our experiment showed that binding does not have to be successful at the object-level. A viewer can adopt a sequential task-driven viewing strategy based on a view hierarchy: viewers first obtain global distributions of the scene. Then, a visual scrutiny is possible within a subregion. In other words, binding occurred at the scene level rather than the object level.

The separable-dimension pairs of length-color and length-texture worked because they supported the scene-centered structural perception in which the processing of global structure and the spatial relationships among components precede analysis of local details according to participants’ self-reports. Another possibility for texture to be effective is the ordering - participants could see large and small. From a practical perspective, our results may suggest that it was easiest for viewers to interpret a scene in which features are scene features for showing ordering and global structures. Scientific data are rarely unstructured. Using coloring to provide some initial regional division may be always better than not. Texture (luminance) could achieve similar accuracy and efficiency as long as the first-type of correspondence error was removed.

5.2 Feature Guidance vs. Scene Guidance

Taking into account both study results, we think an important part of the answer to correspondence error is guidance of attention. Attention in most task-driven fashion is not deployed randomly to objects. It is guided to some objects/locations over others by two broad methods: feature guidance and scene guidance.

Feature guidance refers to guidance by properties of the task-target as well as the distractors (leading to correspondence errors). These features are limited to a relatively small subset of visual dimensions: color, size, texture, orientation, shape, blur or shininess and so on. These features have been broadly studied in 3D glyph design (see reviews by Healey and Enns [29], Borgo et al. [34], Lie et al. [35], Ropinski et al. [36], and McNabb and Laramee [37]. Take one more example from quantum physics simulation results, but with a different task of searching for the structural distributions in the power of 3 in Figure 8 will guide attention to either the fat cylinders (Figure 8a) or the bright yellow color (Figure 8c) or the very dark texture (Figure 8c), depending on the feature-pair types.
Working with quantum physicists, we have noticed that the structure and content of the scene strongly constrain the possible location of meaningful structures, guided “scene guidance” constraints \[6\], \[7\]. Scientific data are not random and are typically structured. Contextual and global structural influences can arise from different sources of visual information. If we return to the MAX search task in Figure 8 again, we will note that the chunk of darker or lighter texture patterns and colors on these regular contour structures strongly influence our quick detection. This is a structural and physical constraint that can be utilized effectively by viewers. This observation coupled with the empirical study results may suggest an interesting future work and hypothesis: adding scene structure guidance would speed up quantitative discrimination, improve the accuracy of comparison tasks, and reduce the perceived data complexity.

Another structure acting as guidance is the size itself. It was used by participants seeking to resolve the NUMEROSTIY tasks to look for the longest outside cylinders. We have showed several examples like Figure 8, our collaborator suggested that the cylinder-bases of the same size with the redundant encoding (Figure 8b) also helped locate and group glyphs belonging to the same magnitude. This observation agrees with the most recent literature that guidance-by-size in 3D must take advantage of knowledge of the layout of the scene \[33\].

Fig. 8: Contours of simulation data. Size from this viewpoint can guide visual grouping and size in 3D must take advantage of knowledge of the layout of the scene \[33\].
of the layout of the scene [33].

Though feature guidance can be preattentive and features are detected within a fraction of a second, scene guidance is probably just about as fast (though precise experiments have not been done and our Experiment II only merely shows this effect). Scene ‘gist’ can be extracted from complex images after very brief exposures [6], [39]. This doesn’t mean that a viewer instantly knows, say, where the answer is located. However, with a fraction of a second’s exposure, a viewer will know enough about the spatial layout of the scene to guide his or her attention towards vector groups in the regions of interest.

A future direction, and also an approach to understanding the efficiency and the effectiveness of scene guidance, is to conduct an eye-tracking study to give viewers a flashview of our spatial structures and then let the viewer see the display only in a narrow range around the point of fixation: does this brief preview guide attention and the gaze effectively? Recently, work in the information visualization techniques domain has measured and correlated performance on the glance or global structure formation. Vision science discovered long ago that seeing global scene structures in medical imaging decision making guides experts attention (experts always know where to look) [42], [43].

5.2.1 Use Our Results in Visualization Tools and Limitations of Our Work

Visualization is used when the goal is to augment human capabilities in situations where the problems might not be sufficiently defined for a computer to handle algorithmically or to communicate certain information. One of these areas is quantum physics: simulation results are in high-dimensional space thus cannot be interpreted in computational solutions. As a result, quantum physicists count on visualization to detect patterns and trends. Our collaborators were amazed though not surprised by many design possibilities and the performance differences among them.

Our current study concerns bivariate data visualization in which the bivariate variables are component parts of a univariate variable. the first variable is always an integer and the second variable is bounded to a real number in the range [1, 10). Application domains carrying similar data attributes could reuse of work. The design principle of prompting scene-level guidance would be broadly applicable to 3D visualizations. Our design is somewhat limited to preliminary pop-out stimuli. Our design could have been improved by following advanced tensor glyph design methods especially those in tensor field visualizations. Both generic [44] and domain-specific requirements for glyph designs [27], [45], [46] have led to the summary of glyph proprieties (e.g., invariance, uniqueness, continuity) to guide design and to render 2D and 3D tensors. A logic step for us is to truly understand the quantum physics principles to combine data attributes and human perception to arrive domain-specific solutions.

One limitation of this work is that we measured only a subset of tasks crucial to showing structures and omitted all tasks relevant to orientation. However, one may argue that the vectors naturally encode orientation. When orientation is considered, we could address the multiple-channel mappings in two ways. The first solution is to use the length-texture to encode the quantitative glyphs and color to encode the orientations if we cluster the vectors by orientations. The second solution is to treat magnitude and orientation as two data facets and use multiple views to display them separately, with one view showing magnitude and the other for orientation (using Munzner’s multiform design recommendations [47]). The second limitation here was that our experiments were limited to a relatively small subset of visual dimensions: color, texture, and size. A future direction would be to try shapes and glyphs to produce novel and useful design.

6 Conclusion

This work shows that correspondence computation is necessary for retrieving information visually and that viewers strategies can play an important role. Our results showed that length-color with the separable pairs was most efficient and effective for both local and global tasks. Our findings in general suggest that, as we hypothesized, distinguishable separable dimensions perform better. Our empirical study results provide the following recommendations for designing 3D bivariate glyphs for representing univariate variables:

- Highly separable pairs can be used for quantitative comparisons as long as these glyphs are scene-structure forming. We recommend using length-color.
- Texture-based glyphs (length – texture) that introduces luminance variation can cause correspondence error and will only be recommended when task-relevant structures can be constrained.
- Integral and separable bivariate feature-pairs have similar accuracy when the tasks are local. They influence accuracy when the search space increases.
- 3D glyph scene would shorten task completion time when the glyph scene support structural feature guidances.
- The redundant encoding (lengthy-color/lengthy) greatly improved on task completion time of integral dimensions (splitVectors) by adding separable and preattentive color features.

Empirical study data and results can be found online at https://sites.google.com/site/interactivevisualcomputinglab/download/integral-and-separable-dimension-pairs.

Acknowledgments

The work is supported in part by NSF IIS-1302755, NSF CNS-1531491, and NIST-70NANB13H181. The user study was funded by NSF grants with the OSU IRB approval number 2018B0080. Non-User Study design work was supported by grant from NIST-70NANB13H181. 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...
A systematic review of experimental studies on data glyphs,” IEEE Transactions on Visualization and Computer Graphics, vol. 23, no. 7, pp. 1663–1679, 2017. [Online]. Available: https://doi.org/10.1109/tvcg.2017.2649018

C. Ware, “Quantitative texton sequences for legible bivariate maps,” IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 6, pp. 1523–1529, 2009. [Online]. Available: https://doi.org/10.1109/tvcg.2009.175

H. Zhao, G. W. Bryant, W. Griffin, J. E. Terrill, and J. Chen, “Validation of SplitVectors encoding for quantitative visualization of large-magnitude-range vector fields,” IEEE Transactions on Visualization and Computer Graphics, vol. 23, no. 6, pp. 1691–1705, 2017. [Online]. Available: https://doi.org/10.1109/tvcg.2016.2599949

C. Ware, Information Visualization: Perception for Design. Elsevier, 2012. [Online]. Available: https://www.elsevier.com/books/information-visualization/ware/978-0-12-381465-7

A. M. Treisman and G. Gelade, “A feature-integration theory of attention,” Cognitive Psychology, vol. 12, no. 1, pp. 97–136, 1980. [Online]. Available: http://doi.org/10.1016/0010-0285(80)90005-5

I. Biederman, “On processing information from a glance at a scene,” ACM SIGGRAPH Workshop on User-oriented Design of Interactive Graphics Systems, 1977. [Online]. Available: http://doi.org/10.1145/1223573.1223585

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, vol. 15, no. 12, p. 1106, 2015. [Online]. Available: https://doi.org/10.1167/15.12.1106

A. Treisman and S. Gormican, “Feature analysis in early vision: evidence from search asymmetries,” Psychological Review, vol. 95, no. 1, pp. 15–48, 1988. [Online]. Available: https://doi.org/10.1037/0033-295X.95.1.15

J. M. Wolfe and L. S. Utechkin, “What is a preattentive feature?” Current Opinion in Psychology, 2018. [Online]. Available: https://doi.org/10.1016/j.copsyc.2018.11.005

T. Urenss, V. Interrante, I. Marusic, E. Longmire, and B. Ganapathisubramani, “Effectively visualizing multi-valued flow data using color and texture,” IEEE Visualization, pp. 115–121, 2003. [Online]. Available: https://doi.org/10.1109/visual.2003.1250362

W. R. Garner and G. L. Felfoldy, “Integrity of stimulus dimensions in various types of information processing,” Cognitive Psychology, vol. 1, no. 3, pp. 225–241, 1970. [Online]. Available: https://doi.org/10.1016/0010-0285(70)90016-2

C. G. Healey, K. S. Booth, and J. T. Enns, “Large datasets at a glance: Combining textures and colors in scientific visualization,” IEEE Transactions on Visualization and Computer Graphics, vol. 5, no. 2, pp. 145–167, 1999. [Online]. Available: https://doi.org/10.1109/2945.773807

C. G. Healey, K. S. Booth, and J. T. Enns, “Visualizing real-time multivariate data using preattentive processing,” ACM Transactions on Modeling and Computer Simulation, vol. 5, no. 3, pp. 190–221, 1995. [Online]. Available: http://doi.org/10.1145/217853.217855

J. Mackinlay, “Automating the design of graphical presentations of relational information,” ACM Transactions on Graphics, vol. 5, no. 2, pp. 110–141, 1986. [Online]. Available: https://doi.org/10.1145/22949.22950

W. S. Cleveland and R. McGill, “Graphical perception: Theory, experimentation, and application to the development of graphical methods,” Journal of the American Statistical Association, vol. 79, no. 387, pp. 531–554, 1984. [Online]. Available: https://doi.org/10.2307/2288400

J. M. Wolfe and T. S. Horowitz, “What attributes guide the deployment of visual attention and how do they do it?” Nature Reviews Neuroscience, vol. 5, no. 6, pp. 1–7, 2004. [Online]. Available: http://doi.org/10.1038/nrn1411

J. M. Wolfe, “Guided search 4.0,” Integrated Models of Cognitive Systems, pp. 119–149, 2007. [Online]. Available: http://doi.org/10.1117/1.33549

H. Strobelt, D. Ooiike, B. C. Kwon, T. Schreck, and H. Pfister, “Guidelines for effective usage of text highlighting techniques,” IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 1, pp. 489–498, 2016. [Online]. Available: https://doi.org/10.1109/tvcg.2015.2467799

C. Healey and J. Enns, “Attention and visual memory in visualization and computer graphics,” IEEE Transactions on Visualization and Computer Graphics, vol. 18, no. 7, pp. 1170–1188, 2012. [Online]. Available: https://doi.org/10.1109/tvcg.2011.129

T. C. Collagon, “Dominance in texture segregation: Hue, geometric form, and line orientation,” Perception, & Psychophysics, vol. 46, no. 4, pp. 299–311, 1989. [Online]. Available: https://doi.org/10.3758/bf03249894

S. M. Casner, “Task-analytic approach to the automated design of graphic presentations,” ACM Transactions on Graphics, vol. 11, no. 2, pp. 111–151, 1992. [Online]. Available: https://doi.org/10.1145/108360.108361

C. Demiralp, M. S. Bernstein, and J. Heer, “Learning perceptual kernels for visualization design,” IEEE Transactions on Visualization and Computer Graphics, vol. 20, no. 12, pp. 1933–1942, 2014. [Online]. Available: https://doi.org/10.1109/tvcg.2014.2346978

B. E. Rogowitz and A. D. Kalvin, “The Which Blair Project: A quick visual method for evaluating perceptual color maps,” IEEE Visualization, 2008. [Online]. Available: https://doi.org/10.1109/vis3d.2008.461501

J. P. O’Shea, M. Agrawala, and M. S. Banks, “The influence of shape cues on the perception of lighting direction,” Journal of Vision, vol. 10, no. 12, pp. 1–21, 2010. [Online]. Available: https://doi.org/10.1167/10.12.21

M. Harrower and C. A. Brewer, “Colorbrewer.org: An online tool for selecting colour schemes for maps,” The Cartographic Journal, vol. 40, no. 1, pp. 27–32, 2003. [Online]. Available: https://doi.org/10.1080/00087788.2003.998534

C. Zhang, T. Schultz, K. Lawonn, E. Eisenmann, and A. Vilanova, “Glyph-based comparative visualization for diffusion tensor fields,” IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 1, pp. 797–806, 2016. [Online]. Available: https://doi.org/10.1109/tvcg.2015.2467439

J. Berlin, Semiology of Graphics: Diagrams, Networks, Maps. University of Wisconsin Press, 1967.

J. Cohen, Statistical power analysis for the behavioral sciences. New York: Academic Press, 1988. [Online]. Available: https://doi.org/10.3234/9780203771587

C. D. Wickens and C. M. Carswell, “The proximity compatibility principle: Its psychological foundation and relevance to display design,” Human Factors, vol. 37, no. 3, pp. 473–494, 1995. [Online]. Available: http://dx.doi.org/10.1177/001872089503700303

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

D. H. Chung, D. Archambault, R. Borgo, D. J. Edwards, R. S. Laramee, and M. Chen, “How ordered is it? on the perceptual orderability of visual channels,” Computer Graphics Forum, vol. 35, no. 1, pp. 1–13, 2016. [Online]. Available: https://doi.org/10.1111/cgf.12889

M. P. Eckstein, K. Koehler, L. E. Welbourne, and E. Akbas, “Humans, but not deep neural networks, often miss giant targets in scenes,” Current Biology, vol. 27, 2017. [Online]. Available: https://doi.org/10.1016/j.cub.2017.07.068

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. [Online]. Available: http://diglib.eg.org/handle/10.2312/conf.EG2013.stars.039-063
[35] A. E. Lie, J. Kehrer, and H. Hauser, “Critical design and realization aspects of glyph-based 3D data visualization,” Proceedings of the Spring Conference on Computer Graphics, pp. 19–26, 2009. [Online]. Available: https://doi.org/10.1145/1980462.1980470

[36] T. Ropinski, S. Oeltze, and B. Preim, “Survey of glyph-based visualization techniques for spatial multivariate medical data,” Computers & Graphics, vol. 35, no. 2, pp. 392–401, 2011. [Online]. Available: https://doi.org/10.1016/j.cag.2011.01.011

[37] L. McNabb and R. S. Laramée, “Survey of surveys (SoS)-mapping the landscape of survey papers in information visualization,” Computer Graphics Forum, vol. 36, no. 3, pp. 589–617, 2017. [Online]. Available: https://doi.org/10.1111/cgf.13212

[38] A. Oliva, “Gist of the scene,” Neurobiology of Attention, vol. 696, no. 64, pp. 251–258, 2005. [Online]. Available: https://doi.org/10.1016/B978-012375731-9/50045-8

[39] 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, vol. 25, no. 1, pp. 872–881, 2019. [Online]. Available: https://doi.org/10.1109/TVCG.2018.2865264

[40] Z. Bylinskii, P. Isola, C. Bambridge, A. Torralba, and A. Oliva, “Intrinsic and extrinsic effects on image memorability,” Vision Research, vol. 116, pp. 165–178, 2015. [Online]. Available: https://doi.org/10.1016/j.visres.2015.03.005

[41] H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann, “A design space of visualization tasks,” IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 12, pp. 2366–2375, 2013. [Online]. Available: http://doi.org/10.1109/TVCG.2013.120

[42] G. Kindlmann and C.-F. Westin, “Diffusion tensor visualization with glyph packing,” IEEE Transactions on Visualization and Computer Graphics, vol. 12, no. 5, pp. 1329 – 1335, 2006. [Online]. Available: https://doi.org/10.1109/TVCG.2006.134

[43] T. Munzner, Visualization Analysis and Design. A K Peters Visualization Series. CRC Press, 2014. [Online]. Available: https://doi.org/10.1201/b17511

Henan Zhao received the B.E. degree in Computer Science and Information Security from Nankai University, China. She is a Ph.D. student in Computer Science and Electrical Engineering at University of Maryland, Baltimore County. Her research interests include design and evaluation of visualization techniques.

Jian Chen received her PhD degree in Computer Science from Virginia Polytechnic Institute and State University (Virginia Tech). She did her postdoctoral work in Computer Science and BioMed at Brown University. She is an Associate Professor in the Department of Computer Science and Engineering at The Ohio State University, where she directs the Interactive Visual Computing Lab. She is also affiliated with the Ohio Translational Data Analytics Institute. Her research interests include design and evaluation of visualization techniques, 3D interface, and immersive analytics. She is a member of the IEEE and the IEEE Computer Society.