# Evaluating Multivariate Visualizations on Time-Varying Data

Mark A. Livingston, Jonathan W. Decker, and Zhuming Ai

Naval Research Laboratory, 4555 Overlook Ave SW, Washington, DC, USA

## ABSTRACT

Multivariate visualization techniques have been applied to a wide variety of visual analysis tasks and a broad range of data types and sources. Their utility has been evaluated in a modest range of simple analysis tasks. In this work, we extend our previous task to a case of time-varying data. We implemented five visualizations of our synthetic test data: three previously evaluated techniques (Data-driven Spots, Oriented Slivers, and Attribute Blocks), one hybrid of the first two that we call Oriented Data-driven Spots, and an implementation of Attribute Blocks that merges the temporal slices. We conducted a user study of these five techniques. Our previous finding (with static data) was that users performed best when the density of the target (as encoded in the visualization) was either highest or had the highest ratio to non-target features. The time-varying presentations gave us a wider range of density and density gains from which to draw conclusions; we now see evidence for the density gain as the perceptual measure, rather than the absolute density.

**Keywords:** Quantitative evaluation, multivariate visualization, visual task design, texture perception, relative texture density, user study

## 1. INTRODUCTION

Large, time-varying data sets present several challenges to visualization designers. Enabled by increasing power of graphics hardware, recent approaches have built on early work in glyph-based representations[1–3] to devise methods for presenting multiple variables simultaneously. Such techniques aspire to help the user to discern subtle patterns involving multiple variables, leading to analytical insights on the data.

Multivariate visualization (MVV) techniques thus may be considered attempts to take advantage of the perceptual capabilities of the human visual system to spot these subtle patterns and use them to interpret the variables and their relationships. These patterns often take the form of variation of simple properties of primitive shapes, such as length, width, size, orientation, hue, and intensity. These cues rely on preattentive visual processing to make different values stand out perceptually.[4] Building such elements into textures which are modulated by scalar field values is one common strategy for display of multiple scalar fields.[5] Textures are perceived primarily through their orientation, scale, and contrast, but dimensions such as density and regularity also may be used to convey field values.[6]

Since the parameter space of variations that may be introduced to the primitive elements or textures is quite large, we may find in the literature numerous MVV techniques, and as a small part of our work, we introduce a hybrid of two previous techniques. Our focus, however, is on the evaluation of these techniques, which is not as often a focus of MVV research. User-based evaluations are not easy to design for comparison of multiple, diverse MVV techniques. We have begun a line of research attempting to compare MVV techniques on a variety of fundamental visual analysis tasks appropriate for the visual analytics call to detect unexpected patterns in the data through the visual pattern analysis described above. Simple tasks of finding critical (maximal) values[7] and trend detection[8] led us to conclude that MVV techniques may not be an improvement over baseline techniques of presenting variables separately. However, a task of finding maximum overlap (equivalent to maximum sum) of six variables was demonstrated[9] to be significantly more difficult with the baseline case than three MVV techniques. Further, analysis indicated that user error rates correlated with the feature density of the resulting visualization in the target, or with the ratio of the feature density in the target to feature density in non-target regions.

Send correspondence to: Mark A. Livingston at mark.livingston@nrl.navy.mil, Telephone: 1 202 767 0380

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*



| 1. REPORT DATE | 2. REPORT TYPE | 3. DATES COVERED |
|---|---|---|
| **FEB 2013** | | **00-00-2013 to 00-00-2013** |

| 4. TITLE AND SUBTITLE | |
|---|---|
| **Evaluating Multivariate Visualizations on Time-Varying Data** | 5a. CONTRACT NUMBER |
| | 5b. GRANT NUMBER |
| | 5c. PROGRAM ELEMENT NUMBER |
| 6. AUTHOR(S) | 5d. PROJECT NUMBER |
| | 5e. TASK NUMBER |
| | 5f. WORK UNIT NUMBER |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>**Naval Research Laboratory,4555 Overlook Ave., SW,Washington,DC,20375** | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | 10. SPONSOR/MONITOR'S ACRONYM(S) |
| | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) |

12. DISTRIBUTION/AVAILABILITY STATEMENT

**Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES

**SPIE Visualization and Data Analysis, vol. 8654, Burlingame, CA, 4-6 Feb 2013.**

14. ABSTRACT

**Multivariate visualization techniques have been applied to a wide variety of visual analysis tasks and a broad range of data types and sources. Their utility has been evaluated in a modest range of simple analysis tasks. In this work, we extend our previous task to a case of time-varying data. We implemented  ve visualizations of our synthetic test data: three previously evaluated techniques (Data-driven Spots, Oriented Slivers, and Attribute Blocks), one hybrid of the  rst two that we call Oriented Data-driven Spots, and an implementation of Attribute Blocks that merges the temporal slices. We conducted a user study of these  ve techniques. Our previous  nding (with static data) was that users performed best when the density of the target (as encoded in the visualization) was either highest or had the highest ratio to non-target features. The time-varying presentations gave us a wider range of density and density gains from which to draw conclusions; we now see evidence for the density gain as the perceptual measure, rather than the absolute density.**

15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT<br>**unclassified** | b. ABSTRACT<br>**unclassified** | c. THIS PAGE<br>**unclassified** | **Same as Report (SAR)** | **14** | |

The identification of this task as one that clearly benefits from the application of MVV techniques raised questions that we attempt to answer in this work. First, we determine the effect of introducing time-varying data to the data; this multiplies the number of data values present. It also gives us the possibility of introducing target features that are not merely the most dense region of the visualization. Second, we introduce a new MVV that is a hybrid of two existing techniques and an implementation of an existing technique that enable us to smoothly extend it to the new data set.

## 2. RELATED WORK

Research on MVVs benefits from a user-centered approach, encompassing both perceptual and cognitive studies of human capabilities. Guidelines for perceptual discernment of subtle differences could improve performance on a wide variety of data-intensive tasks using MVV techniques. We briefly review techniques and evaluations, with comments on our implementations.

### 2.1 Multivariate Visualization Techniques

*Color Blending* is perhaps the oldest and conceptually simplest MVV. Each variable is assigned a particular color; the value of each pixel is computed to be the weighted sum of the colors, with the weights derived from the data values. Thus the dominant hue of a pixel or region in the visualization should indicate the greatest component value among the data values at that location. Each pixel is a visual sample, so the spatial resolution is equal to that of the display device, but expressing more than three independent variables through only three degrees of freedom requires a creative mapping. Even when displaying only three variables or with sufficient degrees of freedom in the display (printer or monitor, for example), perceptual limitations often interfere with the conveyed impression of data values.

*Attribute Blocks* build on early visualizations that use a cluster of shapes or a divided shape to represent multiple values at sample points.[1–3] Each attribute may be visualized with a continuous variable, such as color or intensity; variables are separated by their location within the cluster or shape.[10] Dynamically changing the array's configuration and the size and origin of the individual components enables synthesizing higher resolution than the initial sampling of multivariate data. Issues arise in determining how to sample the underlying data fields, since (unlike Color Blending), the multi-valued representation requires more than a single pixel to represent one sample. Thus rich features may be observed, but at a cost of the spatial resolution. If a data value is not constant over the cell assigned to that data layer, then a spatial fusion technique must be applied in order to calculate the final color of the cell. We used an average of samples uniformly distributed over the area of the cell; in retrospect, a nearest-neighbor or maximum-value strategy may have led to better results. Several glyph-based techniques used similar properties to display data; *Stick Figures* used a torso-and-limb structure to encode data values with relative angles of the torso to the display and limbs to the torso.[11]

Several recent MVV methods have drawn design inspiration from artistic techniques. *Brush Strokes* compose a texture inspired by impressionist paintings; attributes of length, width, orientation, intensity, and hue enable five variables to be encoded in this MVV.[12] Strokes are placed randomly over the surface. One difficulty of this technique is that the parameters do not have the same resolution. Intensity and hue will have more output levels that the user may discern than width or length, due to properties of display hardware and human perception. We further observed that wide strokes can appear to be blurred.

*Oriented Slivers*[13] encodes each data layer with short, grayscale lines on a randomly jittered grid. The orientation differentiates the data layers; the intensity encodes the data values. Sliver density affects the frequency of the underlying data which may be reliably understood. Further, high sliver density, great width, or great length may prevent the user from distinguishing slivers. Still, the technique has the advantage of using few perceptually significant features, allowing the potential for many data layers to be visualized. We opted to restrict ourselves to the technique as defined[13] rather than invent extensions such as the use of color; however, our hybrid technique, introduced in Section 3, may be conceived as a color extension of this technique.

*Data-driven Spots*[14] (DDS) is similar in spirit to pointillist art techniques, using the fact that the human visual system naturally fills space between samples. DDS encode each data layer with Gaussian kernels on a randomly jittered grid. The layers are differentiated by the size and hue, while intensity encodes the data

value. Layers may also move over the surface to further perceptual distance between them and to synthesize resolution beyond that created by the size and spacing of the spots, albeit perhaps by raising a conflict with the jitter pattern. As with Oriented Slivers and Brush Strokes, spot density affects the perceptible frequency of the underlying data. *Color weaving*[15] similarly works on the same concept of overlaying color on a high-frequency texture pattern. This technique does not rely on features such as Gaussian kernels, but on a color field at the display resolution. The closest analogy for DDS is non-overlapping, space filling kernel sets, which is how we implement DDS. Color weaving has been shown to enable good performance in an evaluation, which is the topic of the next subsection.

### 2.2 Evaluating Multivariate Visualizations

A few authors have conducted evaluations of MVV techniques with quantitative and qualitative studies and a variety of tasks, resulting in an assortment of observations. Height and density of vertical bars over a 2D domain were easily identified, but certain combinations with background elements (such as salience or regularity of samples in a dense field) made it hard to understand the data.[6] Brush Strokes (using color, texture, and feature hierarchies among luminance, hue, and texture) enabled verification[16] that perceptual guidelines for visualization[4] apply to non-photorealistic visualizations as well. Oriented Slivers[13] enabled users to perceptually separate layers within a data set. To get the best performance on identifying the presence of a constant rectangular target in a constant background field required a minimum separation of 15° between layers; however, more complex fields may require 30° separation.[4]

A key type of evaluation is testing task performance with a MVV. DDS enabled users to discern boundaries amongst as many as nine layers of data.[14] Other art-inspired techniques such as pointillism, speed lines, opacity, silhouettes, and boundary enhancement enabled users to track a feature over time more accurately and with a subjective preference[17] compared to baseline visualization strategies of separate grayscale visual representations, whether separated spatially or temporally. Adding colors and altering texture properties such as line thickness or orientation in line-integral convolution created effective visualizations for multiple flow fields, as assessed by domain experts.[15] Ellipsoid glyphs were effective at showing tensor structure in diffusion tensor images, whereas layered Brush Strokes encoded field values and enabled users to understand relationships between layers, albeit with a potential for cluttered images.[18] This was not a serious problem in the task because the application displayed dependent variables (data layers).

Other studies have compared multiple, diverse visualization techniques; Table 1 summarizes these studies, the techniques compared, the tasks, and the findings. The most relevant study for our work compared Color Weaving and Color Blending.[19] Users were able to read combinations of 2, 3, 4, and 6 data values with error rates between 7% (two values) and 17% (six values) with color weaving, whereas error rates were between 11% (two values) and 28% (six values) with color blending. Data values were encoded via single-hued color scales that varied jointly in saturation and luminance; users (sequentially) moved six sliders to indicate their responses. When a visualization explicitly represented a feature sought[20] – e.g. showed the sign of vectors in the field, represented integral curves, and showed critical point locations – users performed better at finding the features. Experts and non-experts did not show significant differences. Line integral convolution was best for localizing critical points due to the density of streamlines. Grid-seeded streamlines were best overall across tasks and metrics. Image-guided streamline placement yielded mean error that was 1.5 standard deviations below the norm and mean response time that was 1.0 standard deviations below the norm for advection of a particle. Line integration convolution enabled low (1.0 standard deviations below the mean) errors in count, distance, and flow speed when localizing critical points as well as faster (1.0 standard deviations below the mean) response time; while brush strokes and grid-based streamlines achieved similar results on two of these metrics (but not all four). Grid-based streamlines were 1.0 standard deviations below the mean error and response time for identification of critical point types. Multi-layer texture synthesis enabled[21] users to perform with no significant difference from Brush Strokes for weather data visualization.

In a previous study, we found[7] that the parameterized patterns of DDS and Oriented Slivers helped users perform critical point (maximum) detection more accurately and faster than glyph representations of Brush Strokes and Stick Figures and more accurately than Color Blending. We also found some techniques were sensitive to monitor settings (brightness and contrast) and room lighting conditions. On a trend detection

| Citation | Techniques | Task(s) | Findings |
|---|---|---|---|
| Laidlaw et al.,[20] 2005 | Gridded icons, Jittered grid icons, Brush strokes, Line integral convolution, Image-guided streamlines, Grid-seeded streamlines | Localize critical values, Identify critical point type, Advect particle | Image-guided streamlines 1.5 standard deviations (SD) below mean error and 1.0 SD below mean response time (RT) for advection; also 1.0 SD below mean error and RT for identifying critical point type, LIC 1.0 SD below mean errors and RT for critical point localization |
| Hagh-Shenas et al.,[19] 2006 | Color weaving, color blending | Read data values in sequence | Weaving better by 4-11 percentage points |
| Tang et al.,[21] 2006 | Multi-layer texture synthesis, Brush strokes | Localize critical values | Texture synthesis better by 7.5 percentage points |
| Livingston et al.,[7] 2011 | Data-driven spots, Oriented slivers, Brush strokes, Color blending, Stick Figures | Localize critical values | Data-driven spots best with 39-58% of mean error of other techniques; Oriented slivers next best with 44-65% of error; some techniques sensitive to monitor settings |
| Livingston & Decker,[8] 2011 | Data-driven spots, Oriented slivers, Brush strokes, Color blending, Dimensional stacking, baseline | Detect greatest trend | Baseline technique best with 48-57% of error; Data-driven spots next best with 61-75% of error; Users were fooled by extreme value and better if target was near a "distraction" |
| Livingston & Decker,[22] 2012 | Data-driven spots, Oriented slivers, Brush strokes, Color blending, Attribute blocks, baseline | Detect greatest trend | Attribute blocks better than Dimensional stacking in previous test; main effect of distance to distraction |
| Livingston et al.,[9] 2012 | Data-driven spots, Oriented slivers, Attribute blocks, baseline | Detect multi-way overlap | Multivariate techniques better than baseline; inconclusive evidence for density and for density gain as key aspect for success with MVVs |

Table 1. Table of previous evaluations of multiple multivariate visualization techniques, showing techniques used, task(s), and summary findings.

task,[8] DDS and a baseline case of separate grayscale visualizations outperformed Brush Strokes, Dimensional Stacking, Oriented Slivers, and Color Blending with respect to accuracy, but not with respect to response time. A follow-up study[22] found that the technique of Attribute Blocks improved greatly over Dimensional Stacking, but that adjustments to the DDS technique expected to improve performance (via improved contrast) worsened user performance. Previous exposure to techniques lowered response time and subjective workload, but not error on the trend localization task. As noted above, we found[9] that subjects appeared to use texture density or relative texture density (also called density gain) as a cue to determine the region of maximum overlap between six variables presented with DDS, Oriented Slivers, and Attribute Blocks. The baseline visualization technique of juxtaposed presentation of grayscale images representing a single variable led to poor performance.

## 3. STUDY DESIGN

We extended the task from our previous study,[9] since one of our strongest interests is to understand tasks that cannot be easily solved using baseline visual representations such as spatially distinct grayscale images of each variable. Our modified task (as well as previous suggestions offered numerous times by colleagues and reviewers)

Figure 1. In DDS, six time steps are displayed in six separate images; here, a close-up of an example stimulus shows a target in time step six (second from right) at the center of this cropped image set. In Oriented DDS (far right), all time steps are composited into a single image, with the target and all the distractions from each time step appearing. The orientation indicates the time step to which each feature belongs.

led us to create a hybrid visualization technique, which we term *oriented data-driven spots*. We also wanted to conduct a test that would clarify our previous, conflicting results for the perceptual cue (density versus density gain) that enabled users to solve the task. We describe the hybrid technique, the modified task, the independent variables in the task design, the dependent variables we measured, and our hypotheses about them. We then describe the subject population and the test environment. Finally, we present our results.

### 3.1 Oriented Data-driven Spots

A natural question that may be asked regarding two techniques that use independent perceptual cues to identify data layers (variables) is how one may combine them into a hybrid technique. With the success of both oriented slivers and DDS in our previous studies, we decided to combine these two techniques using the color (as in DDS) to separate data layers and orientation (as in oriented slivers) to separate time steps for each variable. To make the orientation salient, we extended the "spots" into ellipses rather than circles – i.e. anisotropic rather than isotropic kernels. This enabled us to merge* all 36 data values (six variables in six time steps) in a single image (Figure 1). The merging is performed in a straightforward manner: the relevant data slice is merely sampled by the same rules as in the original DDS and oriented slivers techniques (according to the texture pattern for the relevant data layer).

### 3.2 Experimental Task

In designing our task, we wanted to force users to attend to all variables presented in the visualization. Our previous studies on critical point (maximum) detection in a single layer[7] and trend detection[8,22] were not designed to meet this goal. As a result, we found in the trend detection task that the baseline technique (presenting variables in spatially separated grayscale visualizations) performed as well or better than the MVV techniques.

We built on the task used to study DDS presented in the original exposition of the technique,[14] estimating overlap between two target layers of binary data in the presence of zero to seven distraction layers. DDS enabled better performance than side-by-side presentation of the targets (which had no distractors); this was true for any number of distractors present in the DDS visualization. The original task was inspired by finding co-occurrence of elements in a chemical assay. We conceived of the task as having a user determine the region with the greatest number of variables (four to six) that were overlapping at their maximum value among the six variables shown. However, with the MVV representations we use, this task simplifies to finding the area of maximum texture density.

Our revised task extended the data configuration from six static variables to six time-varying variables, with six time steps in the data. Thus every point in the spatial extent of our synthetic data domain has 36 values. We synthesized the data set in order to be able to control variables that affect the task from a perceptual point of view. The data is quite sparse, so most values are zero. Target regions are squares in which four, five, or six variables achieve their maximum value. In the target, no other variables were non-zero other than the ones designated to be part of the target. Other regions were randomly constructed to distract the user. These were essentially targets involving fewer layers, but with two important differences. First, the target never overlapped

*Henceforth, when we use the term "merged" in this paper, we refer to showing data from separate time steps in the same image; we will often use the term "temporally merged" to emphasize this definition.

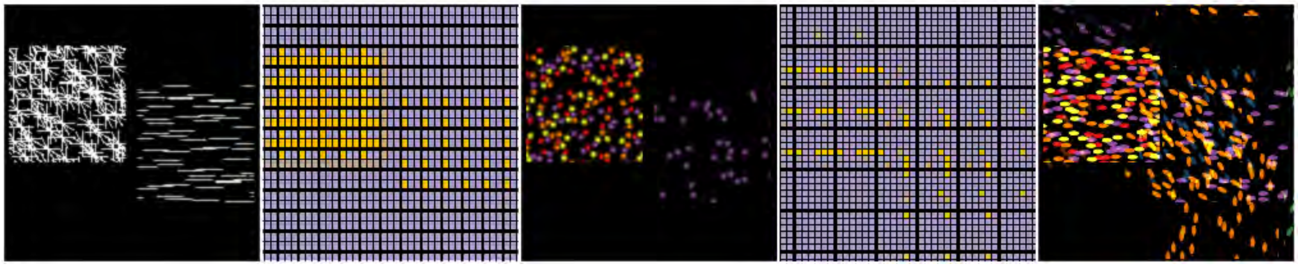

Figure 2. Images of the five MVV techniques in our study, with a four-variable target at the left side of the close-up view. The temporally separate techniques (leftmost three) are shown in the correct time step; the temporally merged techniques (rightmost two) show all time steps. From the left: Oriented Slivers, Attribute Blocks, Data-driven Spots, Temporal Attribute Blocks, Oriented Data-driven Spots.

with any distraction, but distractions could overlap each other. Second, the distractions had at least one more layer completely empty than the target had (e.g. if the target had five variables, no more than four layers were used for a distraction). Further, if only one more layer was empty in a distraction, then at least one other layer was limited to half the maximum value. This constraint is perhaps easiest to conceive as being placed on the sum of the variables in the target and distractions. If we think of each variable having a range of [0..1], then the target had a sum $s$ of four, five, or six, whereas the distractors had a sum of $[0..(s-1.5)]$. This also gives one example of a variable that we could control only in a synthetic data set; another example is the distance between the distractions and the target. The synthetic data set could still be considered a proxy for data from chemical assays; thus we feel this task is ecologically valid.

### 3.3 Independent Variables

We evaluated five visualization techniques: DDS, Oriented DDS, Oriented Slivers, and the two forms of Attribute Blocks. The first form of Attribute Blocks (which shall henceforth be called by that name) used a $3 \times 2$ grid in each cell to represent the six variables; the time steps were shown in separate images. The second form, differing only in the arrangement of the cell, used a $6 \times 6$ grid in each cell to represent all six variables at all six time steps. We shall refer to this as Temporal Attribute Blocks, not to imply a variation on the technique, but in how we assigned the parameters of the single technique to create a different visual representation. The same color map was used in each version. Figure 2 shows examples of what data values look like with each technique; Figure 3 shows the legends for the techniques; these legends were present during the trials to assist users.

The visual representation was the independent variable of primary interest in our study. Since one of our long-term goals is to determine how many variables may be comprehended, a secondary independent variable was the number of layers (data variables) that were included in the target: four, five, or six. We also used three target sizes: 31, 61, and 91 pixels. We further varied the time step in which we placed the target in a controlled fashion (though this variable was not completely crossed with the others). In the analysis, we use trial count as an independent variable to look for fatigue effects.

### 3.4 Dependent Variables and Hypotheses

We measured the error with respect to target value. Specifically, the error was the value (number of overlapped layers) at the target minus the number of layers at the selected location. Since the maximum target value was six, this measure has (in theory) a range of [0,6]. We also measured response time and the number of times a user selected an answer. Response time was measured from the onset of the stimulus until the time of the selection of the final answer. We measured the number of answers selected; the users were informed that they could change their answer as many times as they wished. Finally, we measured the subjective workload associated with each technique through the NASA Task-load Index.[23] We formulated the following hypotheses based on previous results from our own work as well as the literature.

1. We expected the temporally merged techniques of Oriented DDS and Temporal Attribute Blocks to lead to the greatest error.

2. We further expected DDS to outperform Oriented Slivers and Attribute Blocks for error.
3. We expected users to be fastest with the temporally merged techniques.
4. We expected error to increase with increasing number of variables in the target.
5. We expected error to increase with smaller target size.

### 3.5 Subjects and Procedures

The control software was implemented as a set of web pages viewed with the Google Chrome browser (version 17.0.963.83m) under Windows XP (Service Pack 3). The user sat at a standard desktop environment and viewed the stimuli on a 30-inch Dell WFP3008 monitor running at 2560x1600 resolution. Factory default settings were maintained for brightness (75), contrast (50), sharpness (50), gamma ("PC"), color settings mode ("Graphics"), and Preset mode ("Desktop"). The room had standard fluorescent lights. We did not enforce a precise viewing distance; the desktop yielded a viewing distance of 67cm for a typical seated position (giving pixel pitch of 0.25mm). Figure 4 shows images of the entire data trial screen. This configuration is identical to the configuration in our previous studies,[7–9, 22] except for the browser version, though no new features were utilized in this study.

Fifteen subjects (10 male, 5 female) participated in the study; they averaged 35.3 years of age (range: 20-67). All self-reported having normal or corrected-to-normal visual acuity and normal color vision. All reported being heavy computer users; three had participated in previous MVV studies in our lab. The subject first read a set of instructions about the task, which included hints about the target (no overlap with distractors, binary values for the layers at the target location). The subject then proceeded through each technique. Each technique began with instructions specific to the technique. The subject then completed three practice questions, in which only one answer could be selected, but the correct answer was immediately shown. This was followed by the data trials; the order of trials within each technique was determined by random permutation. At the end of each technique, the user completed the NASA TLX. Each subject completed three repetitions of the combination of target size and number of variables in the target for each of the five visualization methods, for a total of $5 \times 3 \times 3 \times 3 = 135$ data points per subject (2025 total).

### 3.6 Study Results

We ran a series of repeated measures ANOVA calculations to determine statistically significant effects.

#### Error as a Function of Technique

The visualization technique had a significant main effect on the user error – $F(4, 56) = 7.747$, $p = 0.000$. As we predicted in our first hypothesis, the temporally merged techniques led to the greatest error. We can attribute a portion of this error to misinterpretation of which time step held the answer. Our error analysis looked up the value in the image for the time step specified in the response, so a correct location with an incorrect time step counted as an error in the analysis. We identified 71 errors of this type; 53 occurred with Oriented DDS (30.5% of errors made with Oriented DDS), 17 with Temporal Attribute Blocks (9.3%), and (perplexingly) one with (temporally separate) DDS (0.1%). If we disregard the time step selected and only look at the point selected to determine the error, however, there would still have been a significant difference between the visualization techniques. In either calculation of error, Oriented Slivers and Attribute Blocks outperformed the other techniques. So while we can support our first hypothesis (that the temporally merged techniques would yield the highest errors), we cannot support our second hypothesis (that DDS would perform best, as it did in our previous studies). We explore possible reasons for this in the Discussion (Section 4). Figure 5 shows a graph of the error with both calculations; for the two best techniques, the error did not occur and the boxes are identical.

### Response Time as a Function of Technique

We hypothesized that the temporally merged techniques would be faster than the temporally separated techniques. With all the data on one visualization, there was no need to page through six time steps or remember the value and location of a selected target. Indeed, a significant main effect of visualization technique was found on response time – $F(4, 56) = 3.384$, $p = 0.015$. Oriented Slivers (fastest temporally separated technique) was indeed slower than the temporally merged visualizations of Oriented DDS – $t(14) = 2.405$, $p = 0.031$ – and Temporal Attribute Blocks – $t(14) = 2.8454$, $p = 0.013$ (Figure 5). We note that Oriented Slivers led to faster times than DDS – $t(14) = 2.675$, $p = 0.018$, but not significantly faster than Attribute Blocks.

### Subjective Workload as a Function of Technique

We also note that visualization technique had a significant main effect on subjective workload – $F(4, 56) = 4.914$, $p = 0.002$ – which we believe reflects the above results. Users felt that the Oriented Slivers and Attribute Blocks were less work (average rating of 29.3 for each, on a scale of 1-100) than DDS – mean workload of 45.3, $t(14) = 2.219$, $p = 0.043$ – and Temporal Attribute Blocks – mean workload of 45.8, $t(14) = 2.253$, $p = 0.040$. Oriented DDS was not significantly different from any other technique in post-hoc t-tests (average rating of 41.2).

### Error as a Function of Number of Variables in Target

The number of variables present in the target had a significant main effect on error – $F(2, 28) = 8.321$, $p = 0.001$. Users accrued the lowest error with only four variables in the target. However, the pattern of error was not as we expected. Users were scored with the greatest error in the case of five variables in the target. Further, the effect was not consistent across visualization techniques, evidenced by a significant interaction between visualization and number of variables in target – $F(8, 112) = 8.789$, $p = 0.000$. Oriented Slivers had greater error with six variables present as opposed to four or five, whereas Temporal Attribute Blocks had the least error with four variables present rather than five or six. Attribute Blocks had the greatest error with five variables present, Oriented DDS had greater error with six variables, and DDS saw roughly equal errors for across the number of variables in the target. Thus we cannot support our fourth hypothesis.

### Response Time as a Function of Number of Variables in Target

The number of variables present in the target had a significant main effect on response time – $F(2, 28) = 4.282$, $p = 0.024$. Users were slightly slower with only four variables present (22.4 seconds) than with five variables (20.8 sec) or six variables (20.4 sec) present. One may be attribute this result to the fact that if a user saw a target with all six variables, there was no need to search further. This could be especially true for the temporally separated visualizations; if a six-variable target was found in (for example) the third time step, there was no need to search the remaining three time steps. However, this was a rarely-used strategy; in only 42 trials did the subject apply "short-circuit" evaluation – and in nine of these trials, it was done incorrectly (i.e. the subject committed an error, though we can't say if it were due to this strategy). Similar to the results for error, the response time was not consistent across techniques, evidenced by a significant interaction between visualization technique and number of variables in the target – $F(8, 112) = 2.151$, $p = 0.037$. For DDS and Oriented Slivers, subjects were slightly *faster* with only four variables present, in contrast to the remaining techniques.

### Error as a Function of Target Size

The target size had a significant main effect on error – $F(2, 28) = 7.349$, $p = 0.003$. In a somewhat perplexing result, the middle size (61 pixels) yielded the least error. Possible explanations are discussed in Section 4.

### Response Time as a Function of Target Size

The target size had a significant main effect on response time – $F(2, 28) = 7.770$, $p = 0.002$. In another somewhat surprising result, users were *slower* with the largest target size. Possible explanations also appear in Section 4.

| Technique | Pearson R | p-value | Slope | Intercept | |
|---|---|---|---|---|---|
| DDS | -0.4194 | 0.0297 | -0.7308 | 1.599 | |
| Slivers | -0.4472 | 0.01962 | -0.7757 | 1.266 | |
| *Attribute* | *-0.3067* | *0.1199* | *-0.8541* | *1.347* | *all data* |
| Attribute | -0.381 | 0.05513 | -0.7342 | 1.15 | outlier removed |
| Oriented DDS | -0.3517 | 0.07224 | -1.411 | 2.999 | |
| *Temporal AB* | *-0.2084* | *0.2971* | *-0.9138* | *2.233* | *all data* |
| Temporal AB | -0.4892 | 0.01339 | -1.774 | 2.858 | outliers removed |

Table 2. The Pearson correlation values and the statistical significance, along with slope and intercept, for each of the lines graphed in Figure 6. As noted, Attribute Blocks and Temporal Attribute Blocks required outlier removal to reach statistical significance.

### Other Observations

We did not see a significant main effect of the trial count on error – $F(26, 364) = 1.239$, $p = 0.197$. However, we did see a significant effect on response time – $F(26, 364) = 3.395$, $p = 0.000$. The first few trials saw a generally decreasing time with increasing trial number; by the seventh trial, users were generally at their fastest. This would indicate some learning was occurring for users. Given that we did not provide feedback after the practice questions, we consider this to be a typical learning effect; users got faster at the task, but not necessarily any better (since they were not being given information on their performance during the test). However, the effect is probably not meaningful (despite the statistical significance); users averaged 26.2 seconds on their first five trials (after the practice trials), and 20.1 seconds on the remaining 22 trials. This may indicate that subjects needed more than the three practice questions we offered.

## 4. DISCUSSION

With these statistical results in hand, we now turn to the interpretation and understanding of what caused the techniques to help or impede user performance. We previously identified two candidate reasons, based on the target and distraction feature density. In our previous study,[9] both the absolute target density and the "density gain" – i.e. ratio of the target's density to the most dense distraction showed some promise, but failed to explain the results for all techniques. We now demonstrate the success of the density gain – relative to the densest *half of the distraction set* – in explaining the core results in this study.

We define the density of the target and the distractions as the number of pixels that are at least 30% of the transition from indicating zero value to indicating full value. This does imply that some distractions had some pixels that were not counted as "on" because they were not of sufficient value; this did not occur in the target except in the case of anti-aliasing. Note that while this definition may be applied to all of the techniques, it has a slightly different meaning for oriented slivers, DDS, and Oriented DDS (pure intensity) and Attribute Blocks and Temporal Attribute Blocks (a combination of intensity and hue change). The figure of 30% was determined by a pilot test to determine smallest change detectable in Attribute Blocks. Further, for Attribute Blocks, the black grid lines should not be counted as basis (denominator) for the density, as they never change color value. The black background in DDS, Oriented DDS, and Oriented Slivers will be covered by data representations. Our DDS implementations use space-filling feature sets, so this does have the potential to cover all of the background (in theory). Oriented slivers could in theory be configured to do this, although our implementation, which centers slivers on grid points, cannot. However, this definition appears to suffice to explain our results.

We computed the correlation between the density gain from the distraction set to the target (as defined above) for each data trial and for each visualization technique. For the techniques of DDS, Oriented DDS, and Oriented Slivers, the correlation using the full set of data trials (27 questions) was statistically significant; for Attribute Blocks and Temporal Attribute Blocks, it was not. However, we identified one outlier that prevented Attribute Blocks from reaching statistical significance; we found two outliers that prevented Temporal Attribute Blocks from reaching statistical significance. Figure 6 shows the data plots and regression lines; Table 2 shows the correlation values.

Thus it appears we have found decisive evidence that the density gain of the target region in the representation is the critical perceptual feature to complete the task of finding the maximum overlap. However, it should be noted that we did not in this analysis account for the time step in which the distractions are found. For the temporally separate visualizations, this is a fundamental difference, in that distractions in different time steps than the target must be compared via the user's memory, whereas for the temporally merged visualizations, the comparison can be perceptually accomplished by comparing regions of a single screen (assuming one may overcome the clutter). Thus one could argue that this evidence deserves further analysis. We leave that for future work.

We made some changes to the appearance of Attribute Blocks from our previous study. We made the individual cells within each $3 \times 2$ sample block smaller. This enabled a sample block to fit within the smallest target size. We expected the error for this case to drop. We found that the error on the smallest target size with the Attribute Blocks technique dropped to approximately one-quarter of its value (1.24 vs 0.32). However, the error on the largest target size increased notably, from 0.09 to 0.57. This is a result that we have yet to explain.

We also changed the angular distribution of the Oriented Slivers, using the full 180° range that may be used for six variables. We previously used 168° of the available range. We see that the overall error for Oriented Slivers dropped from 0.42 to 0.29. We note that the cases of five and six variables present in the target result in denser clusters and slivers, and it was in these two cases that the improvement occurred (four variables: from 0.18 to 0.22, five variables: from 0.46 to 0.23, six variables: from 0.63 to 0.42). Thus it appears that Oriented Slivers benefited from a small spreading out of the angular separation between layers. While this seems unlikely to have produced such a large effect, it was the only difference in the implementation of the technique. This also warrants further investigation.

Returning to the results with respect to target size noted above, we see in the low error with the middle target size and slower speed with the largest target size some possible evidence of the effort to understand both sparse and dense targets. The latter produces the condition of *operator overload*, in which too much information is presented for any of it to be easily used. The former may produce an analogous condition for the cognitive effort to process sparse data. This is another interesting avenue for future work.

## 5. CONCLUSION

We appear to have produced sufficient evidence to conclude that the perceptual factor determining whether users were able to solve the task of finding the greatest overlap is the ratio of density in the target region to density in non-target regions, which we refer to as *density gain*. We have evidence that a visualization method that creates density in a non-target region will impede users' ability to solve the task accurately. This is the type of perceptual feature that we believe it is critical to find when seeking to understand the usability of MVV techniques.

Our version of the task that introduced time-varying variables raises some interesting questions about how this critical perceptual feature interacts with the working memory of the user. We have also left analysis of other perceptual features, notably spatial distribution amongst target and distractions, for future work. We further noticed some parameters of the multivariate visualizations we studied that appeared to have a profound effect on the success of the techniques as we implemented them in this work versus previous work (in our lab and elsewhere). Thus another potentially fruitful avenue for future work is in-depth study of the individual techniques to study the effects of such parameters on common visual analysis tasks. We believe the success demonstrated by our users and the perceptual feature thus shown to lead to success can help guide this line of research as well.

## ACKNOWLEDGMENTS

The authors thank the anonymous subjects and reviewers. This work is funded under the NRL Base Program.

## REFERENCES


[1] Beddow, J., "Shape coding of multidimensional data on a microcomputer display," in [*Proceedings of IEEE Visualization*], 238–246 (Oct. 1990).

[2] LeBlanc, J., Ward, M. O., and Wittels, N., "Exploring N-dimensional databases," in [*Proc. of IEEE Visualization*], 230–237 (Oct. 1990).

[3] Levkowitz, H., "Color icons: Merging color and texture perception for integrated visualization of multiple parameters," in [*Proceedings of IEEE Visualization*], 164–170, 420 (Oct. 1991).

[4] Ware, C., [*Information Visualization: Perception for Design*], Morgan Kaufmann, 2nd ed. (2004).

[5] Kehrer, J. and Hauser, H., "Visualization and visual analysis of multi-faceted scientific data: a survey," *IEEE Transactions on Visualization and Computer Graphics* **18** (2012).

[6] Healey, C. G. and Enns, J. T., "Building perceptual textures to visualize multidimensional datasets," in [*IEEE Visualization*], 111–118 (1998).

[7] Livingston, M. A., Decker, J. W., and Ai, Z., "An evaluation of methods for encoding multiple, 2D spatial data," in [*SPIE Visualization and Data Analysis*], (Jan. 2011).

[8] Livingston, M. A. and Decker, J. W., "Evaluation of trend localization with multi-variate visualizations," *IEEE Transactions on Visualization and Computer Graphics* **17**, 2053–2062 (Dec. 2011).

[9] Livingston, M. A., Decker, J. W., and Ai, Z., "Evaluation of multivariate visualization on a multivariate task," *IEEE Transactions on Visualization and Computer Graphics* **18** (Dec. 2012).

[10] Miller, J. R., "Attribute blocks: Visualizing multiple continuously defined attributes," *IEEE Computer Graphics & Applications* **27**, 57–69 (May/June 2007).

[11] Pickett, R. M. and Grinstein, G. G., "Iconographic displays for visualizing multidimensional data," in [*Proc. of the 1988 IEEE Intl. Conf. on Systems, Man, and Cybernetics*], 514–519 (Aug. 1988).

[12] Kirby, R. M., Marmanis, H., and Laidlaw, D. H., "Visualizing multivalued data from 2d incompressible flows using concepts from painting," in [*Proceedings of IEEE Visualization '99*], 333–340 (1999).

[13] Weigle, C., Emigh, W., Liu, G., Taylor II, R. M., Enns, J. T., and Healey, C. G., "Effectively visualizing multi-valued flow data using color and texture," in [*Graphics Interface*], 153–162 (2000).

[14] Bokinsky, A. A., *Multivariate Data Visualization with Data-driven Spots*, PhD thesis, The University of North Carolina at Chapel Hill (2003).

[15] Urness, T., Interrante, V., Longmire, E., Marusic, I., O'Neill, S., and Jones, T. W., "Strategies for the visualization of multiple 2D vector fields," *IEEE Computer Graphics & Applications* **26**(4), 74–82 (2006).

[16] Healey, C. G., Tateosian, L., Enns, J. T., and Remple, M., "Perceptually based brush strokes for nonphotorealistic visualization," *ACM Transactions on Graphics* **23**(1), 64–96 (2004).

[17] Joshi, A., *Art-inspired techniques for visualizing time-varying data*, PhD thesis, The University of Maryland, Baltimore County (2007).

[18] Laidlaw, D. H., Ahrens, E. T., Kremers, D., Avalos, M. J., Jacobs, R. E., and Readhead, C., "Visualizing diffusion tensor images of the mouse spinal cord," in [*Proceedings of IEEE Visualization '98*], 127–134 (1998).

[19] Hagh-Shenas, H., Interrante, V., Healey, C., and Kim, S., "Weaving versus blending: a quantitative assessment of the information carrying capacities of two alternative methods for conveying multivariate data with color," in [*Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization*], 164 (2006).

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

[21] Tang, Y., Qu, H., Wu, Y., and Zhou, H., "Natural textures for weather data visualization," in [*Tenth International Conference on Information Visualization*], 741–750 (July 2006).

[22] Livingston, M. A. and Decker, J. W., "Evaluation of multi-variate visualizations: A case study of refinements and user experience," in [*SPIE Visualization and Data Analysis*], (Jan. 2012).

[23] Hart, S. G. and Staveland, L. E., "Development of NASA-TLX (task load index): Results of empirical and theoretical research," in [*Human Mental Workload*], Hancock, P. A. and Meshkati, N., eds., 239–250, Elsevier Science Publishers (1988).

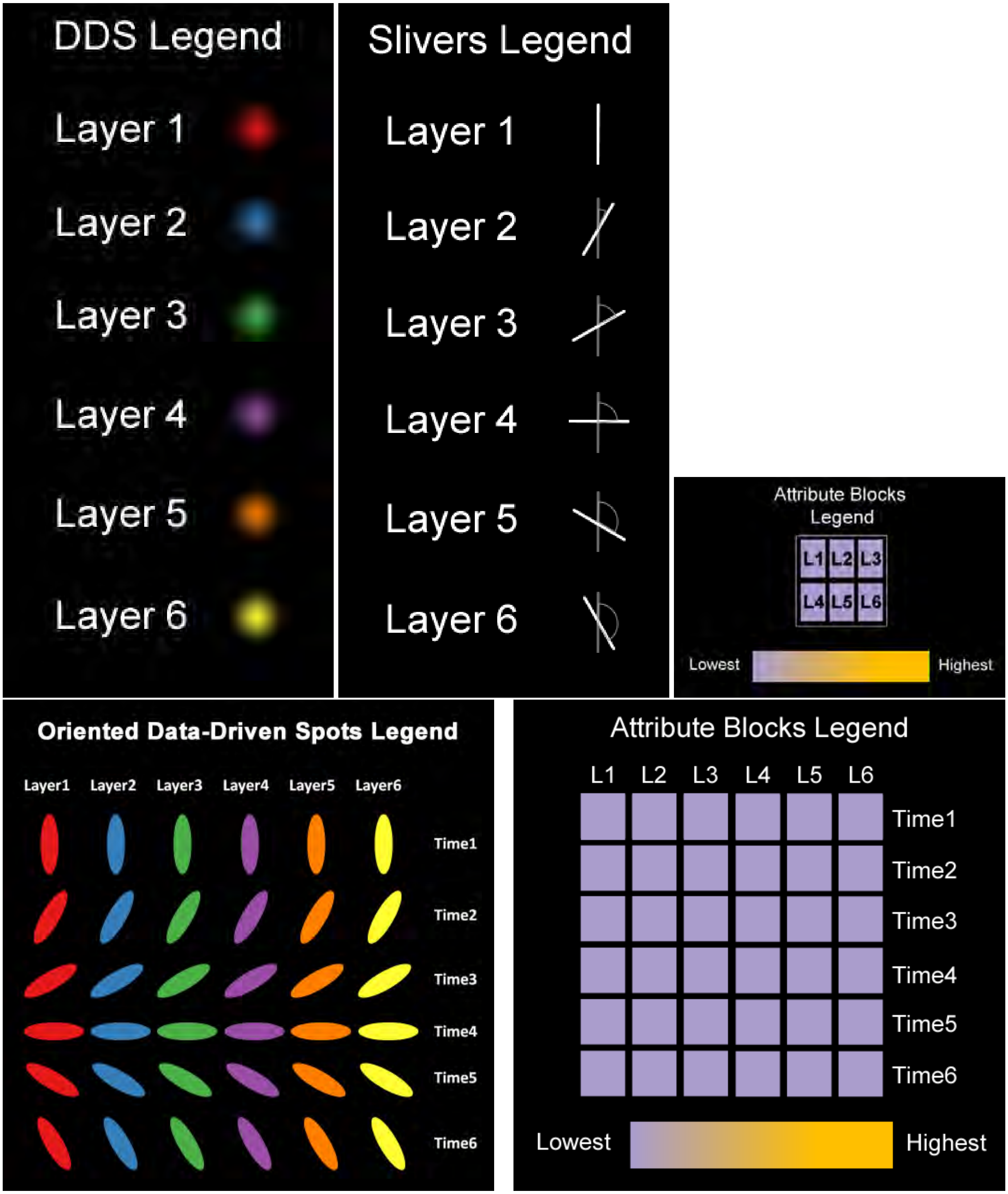


Figure 3. The legends for each technique were present for the data trials, respectively. Top row: temporally separate presentations of (from left) data-driven spots (DDS), oriented slivers, and attribute blocks. Bottom row: temporally merged techniques, requiring a single image for all 36 data values, which we call oriented data-driven spots (left) – a hybrid of DDS and oriented slivers – and temporal attribute blocks (right).

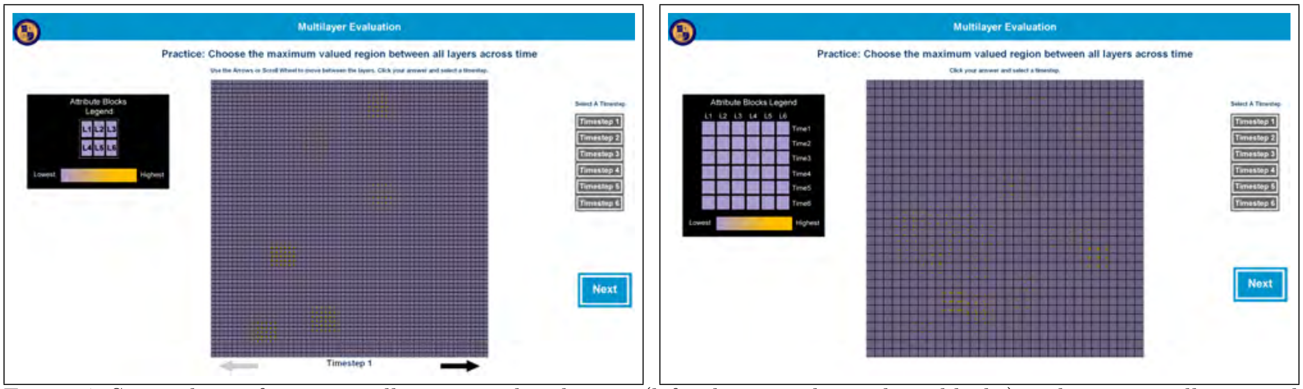


Figure 4. Screenshots of a temporally-separated technique (left, shown with attribute blocks) and a temporally merged technique (right, shown with temporal attribute blocks). The user selected the target's location in the image and, through the buttons at the right of the screen, the time step in which the target appeared. When the user was satisfied with the response, the "Next" button would move to the next trial.

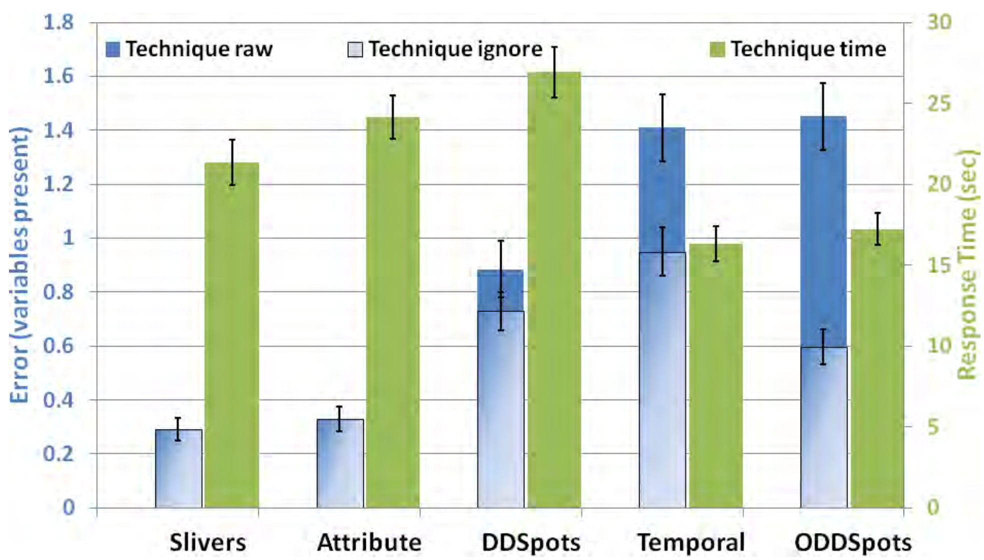


Figure 5. Graph showing the significant main effect of visualization technique on user error (blue); this effect occurs when error is measured using both location and time step selected, and when error is measured using only image location (and ignoring whether the user selected the correct time step). The change between the two error functions demonstrates the difficulty users had in selecting the correct time step when all time steps were presented in a single image (with Oriented DDS and Temporal Attribute Blocks, rightmost two sets of bars). This graph also shows the significant main effect of visualization technique on response time. Largely due to the need to page through six time slices, users were faster with the techniques that merged the time steps.

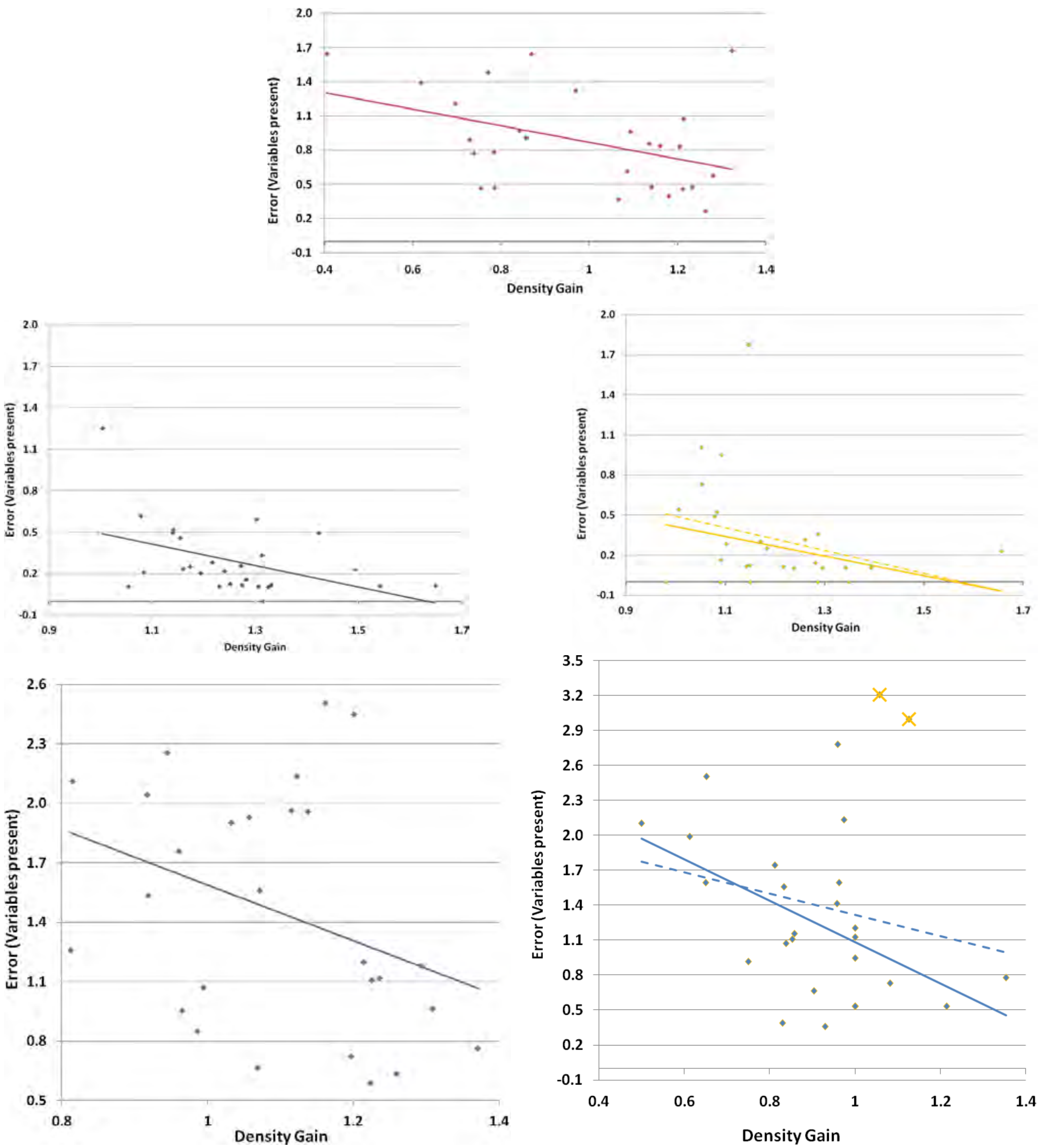


Figure 6. Graphs of the mean error for each trial question as a function of density gain. Top row: DDS. Middle row: Oriented Slivers and Attribute Blocks. Bottom row: Oriented DDS and Temporal Attribute Blocks. Note the outliers (points with × through them) in Attribute Blocks and Temporal Attribute Blocks. These points had to be eliminated in order to achieve statistical significance of the correlation.