An Evaluation of Methods for Encoding Multiple, 2D Spatial Data

Mark A. Livingston, Jonathan Decker, and Zhuming Ai

Naval Research Laboratory, 4555 Overlook Ave, Washington, DC, USA 20375

ABSTRACT

Datasets over a spatial domain are common in a number of fields, often with multiple layers (or variables) within data that must be understood together via spatial locality. Thus one area of long-standing interest is increasing the number of variables encoded by properties of the visualization. A number of properties have been demonstrated and/or proven successful with specific tasks or data, but there has been relatively little work comparing the utility of diverse techniques for multi-layer visualization. As part of our efforts to evaluate the applicability of such visualizations, we implemented five techniques which represent a broad range of existing research (Color Blending, Oriented Slivers, Data-Driven Spots, Brush Strokes, and Stick Figures). Then we conducted a user study wherein subjects were presented with composites of three, four, and five layers (variables) using one of these methods and asked to perform a task common to our intended end users (GIS analysts). We found that the Oriented Slivers and Data-Driven Spots performed the best, with Stick Figures yielding the lowest accuracy. Through analyzing our data, we hope to gain insight into which techniques merit further exploration and offer promise for visualization of data sets with ever-increasing size.

Keywords: Multi-variate visualization, scalar fields, spatial surfaces, glyphs, color encoding, user study

1. INTRODUCTION

Our ability to acquire data about our conceptual or physical environment continues to grow faster than our capacity for analysis and decision-making. Visualization of complex, multi-variate data offers the potential to apply the human intellect to problems for which automated analysis techniques have yet to be developed and verified. One area of long-standing interest in visualization is increasing the parameters of visual representations through which variables are mapped to properties of the visualization. While color and intensity, shape and orientation, and texture with attributes such as regularity or density have all been applied successfully for particular problems, there has been little comparison of diverse techniques for fundamental tasks in visualization. Our goals in this work were (first) to abstract a task fundamental to our end users' visualization needs and (second) to analyze, from both a logical/theoretical and quantitative basis, how well users could perform these tasks with a variety of multi-variate visualization techniques.

One characteristic of the GIS data with which our analysts work is the massive number of variables within the data (as many as 2000). Fundamental types of data include roads (vector data), land use (low-frequency scalar fields), event locations (point data), and demographic data (high-frequency scalar fields). Other variables of interest in certain applications might include pedestrian or vehicular traffic data (vector field). When we refer to data as complex, we mean that the variables of interest represent a broad selection of data types. It is not unusual for multiple variables of interest to be of the same type, nor is it unusual to be interested in variables of different types. But an analyst may wish to look at an arbitrary number of variables simultaneously. We began our investigation with visualization of three, four, or five scalar fields of data, though our ultimate goal (like many authors) is to push the limits of visualization techniques to allow more layers of diverse types to be simultaneously comprehended by the user.

One goal for our GIS analysts is to detect patterns among independent variables such as demographic information and urban development to try to predict locations which may fit a pattern of previous criminal activity. This can inform security forces which areas may need more patrols or for times of day in which attacks are

Corresponding author: mark.livingston@nrl.navy.mil; phone 202 767-0380; fax 202 404-1192

Visualization and Data Analysis 2011, edited by Pak Chung Wong, Jinah Park, Ming C. Hao, Chaomei Chen, Katy Börner, David L. Kao, Jonathan C. Roberts, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7868, 78680C · © 2011 SPIE-IS&T · CCC code: 0277-786X/11/$18 · doi: 10.1117/12.872576

more likely. In this application, being able to visualize the relationships between numerous variables can be advantageous. We previously applied coordinated multiple view strategies, but this alone did not aid in making comparisons across multiple layers of spatial data. We thus turned to techniques that explicitly represent multiple data layers with various graphical properties.

In applying these techniques to various data sets which our users analyze, we noticed several undesirable qualities inherent to these methods. We began to explore additional published techniques (Section 3), and then to consider on which aspects of the techniques we should form a basis for comparing the applicability to our visualization problems. This launched our logical (aspiring to theoretical) analysis (Section 4). Next, we wanted to experimentally verify whether our insights matched the users' performance on tasks that our domain analysis (assisted by subject matter experts) revealed as those on which we should focus our initial efforts. This culminated in pilot user studies (Section 5). The logical and experimental analysis of the multi-variate visualization techniques comprise the contribution of our ongoing work.

2. PREVIOUS WORK

Numerous authors have contributed to the body of anecdotal, theoretical, and quantitative evidence arguing for the design quality of a multi-variate visualization. We concentrate our review of this body of work on the last two contributions. One may begin by analyzing capabilities of human perception to derive design guidelines that may be applied to visualization tasks.[1] Among the type of tasks described were the "perception of emergent properties" made evident by visual presentation of data. With the advent of the conceptual framework of visual analytics[2] and its emphasis on analytical reasoning through visual interfaces, the importance of clarity of presentation for complex data was further stressed. These concepts are fundamental to our approach.

More directly informing our work are approaches that begin with understanding the data and then examining visual properties. Healey et al.[3] identified four pieces of information by which a user and visualization system may architect a visualization: the importance of each attribute, the spatial frequency, whether it is continuous or discrete, and the task (if any) the user wishes to perform on the attribute. The authors then discussed how this information may be used in combination with understanding of human perception, mixed-initiative interaction, and automated search strategies to create a mapping from data attributes to visual features. Features employed included luminance, hue, size (height of bars), density, orientation, and regularity to a grid. Earlier, Zhou and Feiner[4] characterized data in order that an automated method might craft visualizations. The dimensions in the taxonomy were type (divisible or atomic), domain (semantic, e.g. physical or abstract), attributes (e.g. shape), relations (connections between data), role (with respect to user goals), and sense (user visual preferences). These taxonomies sparked our thinking about what aspects of the data created difficulties for given visualization techniques.

Urness et al.[5] applied overlay and embossing to composite textures which encoded multiple 2D vector fields. By adding colors and altering texture properties such as line thickness or orientation in line-integral convolution, they created effective visualizations for multiple flow fields, as assessed by domain experts. Laidlaw et al.[6] visualized seven-layer diffusion tensor images using ellipsoid glyphs and brush strokes. They showed significant differences between healthy and unhealthy spinal cords in mice. The glyphs were effective at showing tensor structure everywhere within the images, whereas layered brush strokes encoded field values and enabled users to understand relationships between layers. The difficulty in this method was the potential for cluttered images. This was not a serious problem in their application because it involved a number of dependent variables.

Several user studies have examined the utility of individual techniques. Healey et al.[7] found that height and density of vertical bars over a 2D domain could be easily identified, but that certain combinations with background elements (such as salience of regularity of samples in a dense field) made it hard to understand the data. They validated their results with a user study on weather data. The introduction of Brush Strokes[8] (specifically, color, texture, and feature hierarchies among luminance, hue, and texture) enabled verification that guidelines for perception during visualization[1] applied to non-photorealistic visualizations as well. The authors also noted the aesthetic quality of such visualizations. Oriented Slivers[9] 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, they found a minimum separation of 15° between layers necessary.

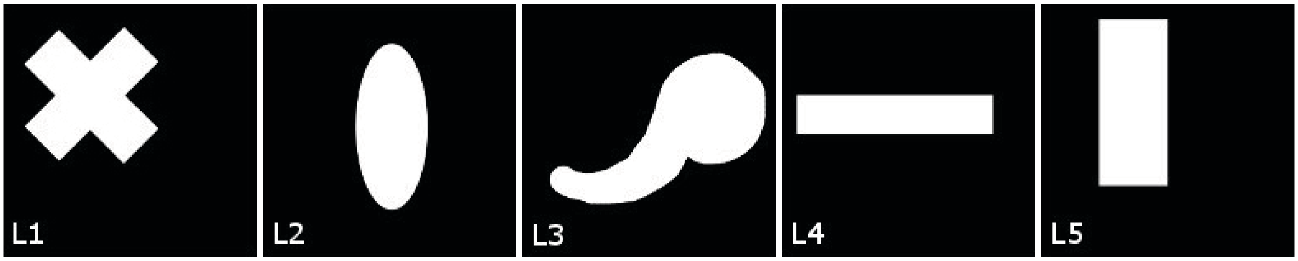

Figure 1. These five layers will be used to demonstrate how each of the multi-layer techniques encodes multiple layers into a single image.

Pointillism techniques for visualizing overlapping regions were employed by Jenks on a map of crops harvested.[10] One dot represented 10,000 acres of harvest crops, divided into 12 categories and identified by color. Color choice was aided by the distribution of the data; since crops of peanuts and sugar rarely intermingled, it was safe to give them a similar hue. But this mapping, although it was a more numeric approach to pattern mapping than employed later, demonstrated the power of sub-sampling spatial data in order to allow multiple layers of data to be clearly visible on the same surface.

Bokinsky[11] devised Data-Driven Spots (DDS), a collection of Gaussian spots and bumps with varying radii, color, opacity, and animation that enabled users to discern boundaries amongst as many as eight layers of data. Joshi[12] visualized time-varying fluid data using art-inspired techniques such as pointillism, speed lines, opacity, silhouettes, and boundary enhancement for weather and other data. Users were able to track a feature over time more accurately and expressed preferences for the illustration-inspired techniques.

Other studies have compared multiple, diverse visualization techniques. Laidlaw et al.[13] compared six techniques for 2D vector data, asking users to locate critical points, identify types of critical points, and advect a particle. Users performed better when the visualization explicitly represented the solution to the tasks – i.e. showed the sign of vectors in the field, represented integral curves, and showed critical point locations. Experts and non-experts did not show significant differences. Hagh-Shenas et al.[14] compared Color Blending and Color Weaving. Color Blending refers to interpolating colors specific to each layer to create continuous fields of combined color representations. Color Weaving refers to the pointillism techniques, such as DDS, discussed earlier. The name comes from the flow field color method,[15] which works on the same concept of separating colored elements so that multiple, overlapping features can be identified in the same spatial region. Maintaining the original colors as in Color Weaving outperformed Color Blending;[14] this difference increased with the number of components. Color selection for the various scales was a critical issue in the blending methods. Tang et al.[16] developed multi-layer texture synthesis for weather data visualization, varying scale, brightness, orientation, and regularity. Users in their study performed as well with this technique as with one using the Brush Strokes technique proposed by Healey et al.[8]

3. TECHNIQUES AND IMPLEMENTATIONS

This section describes the techniques used in our study. For demonstration purposes, we will use the set of five layers shown in Figure 1, each of which contains one shape with solid borders. Discrete features are more easily identifiable in all of these methods, since the contrast between borders is sharp, regardless of the visual mapping. This was chosen to aid the reader in understanding how to read these mappings. In our user study, we used Gaussian features as a better proxy for the real-world data in our intended application.

3.1 Oriented Slivers

Oriented Slivers[9] places a pattern of short, white lines at randomly jittered grid positions on a black background. These slivers share an orientation within each pattern. A pattern is blended with a specific data layer, so that the source image can only be seen on the surface of the slivers. Therefore, a low density of slivers equates to a sparse sampling of the source layer. Furthermore, the pattern cannot be so dense that the orientation of the slivers are not distinct. For this reason, Oriented Slivers is not a good candidate for multi-dimensional data surfaces with high frequency spatial data. Figure 2 shows a composite of the five shape layers and a key which matches the

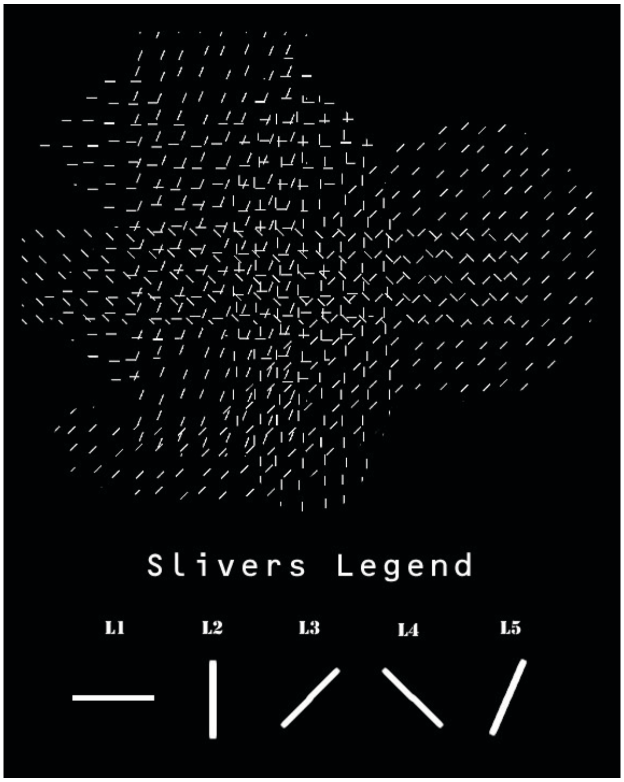

Figure 2. Our five layers (Figure 1) shown with Oriented Slivers. A legend explained the encoding to users.

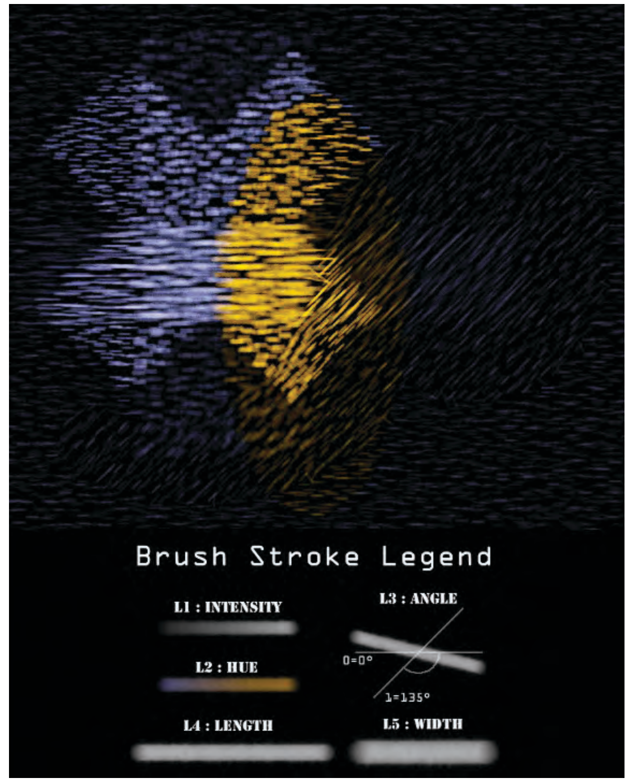

Figure 3. Brush Strokes image of our five layers, with its legend. Note that hue does not vary in grayscale.

orientation to each individual layer. This is the same encoding and legend as in our user study. The composite was scaled down so that it would fit in the paper, but the shapes can still be visually segmented. However, if the slivers were much smaller or denser, it would become hard to make out their individual orientations.

3.2 Brush Strokes

Brush Strokes are an example of a single pattern with several parameters that may encode data layers. This is as opposed to techniques like Oriented Slivers and Data-Driven Spots, which utilize a parameterized pattern repeated for multiple layers. The difference is that this single pattern's elements must have multiple attributes, and these elements will vary throughout the image based on the underlying data. The specific technique we employed in this space uses brush strokes to encode data. These strokes are randomly placed over the surface; they vary in the intensity and hue of their surface color, in orientation, and in their width and height. Figure 3 shows a Brush Stroke image of the five shape layers and the legend. Intensity and hue (not indicated in grayscale printing) are the clearest indicators, as the × (layer L1) and oval (layer L2) are distinctly defined. The third clearest attribute is the stroke orientation, which reveals layer L3. The length of the stroke encodes layer L4. This can be seen by examining the density of the strokes, as the longer strokes fill out the gaps in the image. Finally, the width of the stroke encodes layer L5. To our eyes, this manifests itself as blurring within the image.

3.3 Data-Driven Spots

Data-Driven Spots (DDS)[11] has roots in stippling techniques proposed by geographers to encode overlapping data.[10] It is similar to Oriented Slivers; instead of grayscale-encoded lines, DDS places small Gaussian kernels on a randomly jittered grid. We encode each layer with a different style of dot and offset the distributions, so that there is minimal overlap between spots from different layers (Figure 4, left). Specific layers can be identified by the spots' size and hue (latter not identifiable in grayscale). Layers can also be animated by slowly moving the spots across the surface; however, we leave this option for future work. Figure 4 (right) contains an image that

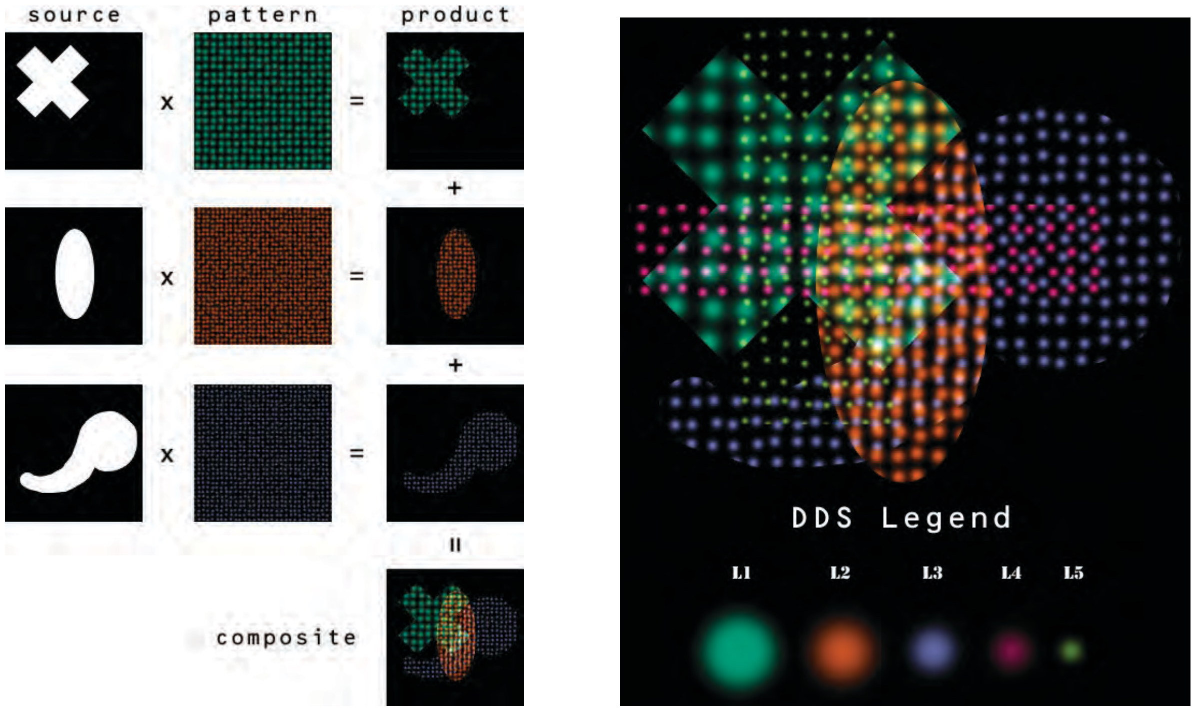

Figure 4. *Left*: DDS images are composited from source images blended with spot patterns which are parameterized so that they do not obscure each other. *Right*: a DDS image of our five layers with its legend.

encodes all five shape features, as well as the technique legend shown to subjects during the user study. This method works well with the high-frequency edges of these features; all the individual shapes are clearly visible.

3.4 Color Blending

Color Blending begins with a set of colors c_i, with each color chosen to represent an individual layer i. Each pixel of the composite image is created using the sum of the colors, weighted by the components of the normalized source vector v at that pixel: $\sum_{i=1}^{n} c_i v_i$. This sum is blended with the mean value at that pixel. This technique benefits from being one of the few that does not have to subsample the source. However, since it is blended with the mean value at each pixel, some light features can be hard to make out. Figure 5 demonstrates the technique when used on the shape layers from Figure 1, as well as a legend to aid users in decoding the blend. Even though the shapes are distinct, the blending of the colors produces some confusion in the overlapping regions. The pentagon shows how the colors blend, but this can be difficult to interpret for the user. Attempting to match the colors is betrayed since the colors are always scaled by the mean value. It completely relies on hue; grayscale versions look to have constant value, although the shapes are visible.

3.5 Stick Figures

One particularly abstract technique uses a Stick Figure to represent the value of each data vector.[17] The stick figure body is angled with respect to a "home" orientation and the angles of each limb with respect to the body. The space is divided into grid cells, each of which is represented by a unique instance of the Stick Figure. For our implementation (Figure 6), the body is vertical when layer L1 is zero, and it is oriented 135° degrees from vertical at the layer's maximum value (clockwise, as depicted in the legend). The limbs, when the underlying value is zero, are oriented at 10° from being parallel with the body. When positioned at the maximum value in their corresponding layer, they appear oriented 110° from the body. (The full range of a limb appears in the legend over the limb matching layer L4.) With this type of mapping, a grid of Stick Figures represents multiple layers of data at once.

We could not show the entire composite image of the shapes like we did for the other techniques, since the Stick Figures would appear too small to be readable. For demonstration, we provided a cropped portion of the

Figure 5. Color Blending image of shapes with its legend.

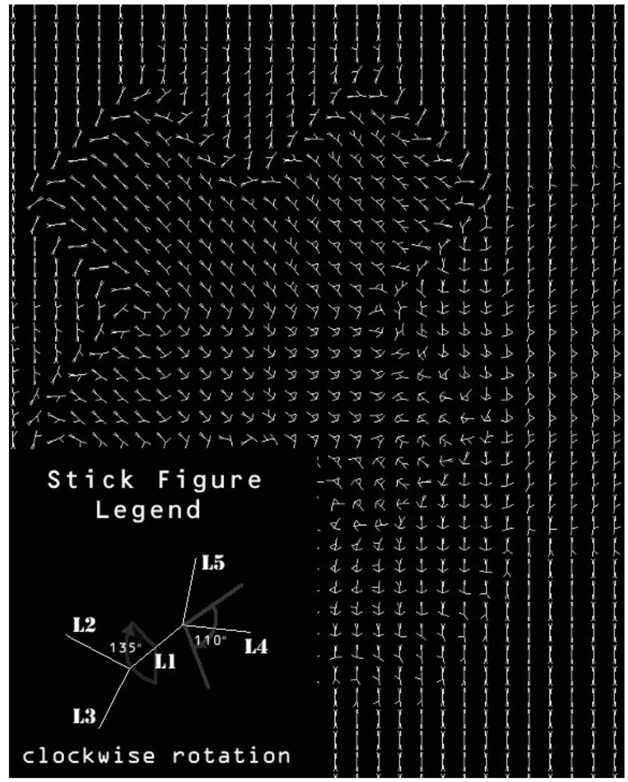

Figure 6. Stick Figures image of shapes with its legend.

composite in Figure 6. The × feature in layer L1 is the most distinct, because it is encoded with the body of the Stick Figures. We recommend that one find the other shapes by noticing that when all values are zero at a location, the Stick Figure there is completely vertical. When this begins to change, then the figures are transitioning into a feature. Stick Figures near the top, above the ×, show a lowered left arm, indicating that those Figures are within the tall box feature in layer L5. You can follow the Figures down and see that this arm remains lowered throughout the box, even with the body rotated in the region covered by the ×.

Here another issue with this technique has presented itself; as the body rotates, the identity of each limb becomes hard to follow. Specifically in data sets with solid shapes like this one, in which the figure suddenly snaps to a new posture, it can be hard to keep track of the figure. Some implementations of this technique color the limbs, but this can be hard to read on a small glyph, leaving the alternative to further sub-sample the field. Another approach could have been to rotate the body of the figure half as much and offset the "home" orientation, so that the figure has a smaller range for its overall orientation.

4. ANALYSIS OF TECHNIQUES

In the above technique descriptions, we mentioned some of the limitations inherent to each of the methods. We will now compare the techniques to one another directly in several important areas.

4.1 Color Reliance

The requirement for color can limit a technique, specifically if it would be useful to overlay other data over the composite. As can be observed on a gray copy of this paper, several of these techniques rely on color. The colors we used in our study were taken from a qualitative set in ColorBrewer* and do not vary in intensity. Since only the hue is encoding the data, intensity is left to encode some other attribute. This can be seen in the legend for the Color Blending technique (Figure 5). In grayscale, this pentagon is a solid shade of gray, revealing that layers can no longer be distinguished. All that remains is the mean surface multiplied by the weighted colors. From this we can still see differences in intensity, which represent the number of layers overlapping at each pixel. The Brush Strokes method relies on color for one layer. In our study (Figure 3), layer L2 disappears in gray shades. Thus, this technique is still usable, but loses some scalability. Data-Driven Spots fairs much better; with the loss of color, it loses a redundant encoding, but layers may still be distinguished via the spot size. Oriented Slivers and Stick Figures do not use color, revealing that there may be benefits to combining these layers with color-enhanced techniques.

4.2 Scalability

Another major point of comparison is the maximum number of layers that can be represented with these techniques. The Brush Strokes are clearly limited to the number of distinct attributes for the texturing primitives. We are not quite at the maximum limit of this technique at five layers. For instance, we could include stereoscopic 3D imaging to make an additional layer pop out. Nevertheless, this technique spans the breadth of graphical mappings, which is relatively small compared to exploring the depth of one or two mappings. Another method with low scalability is Color Blending. While it is true that this method could combine any number of colors, it becomes increasingly difficult to provide a color set where any combination of selected colors is meaningful. We will discuss later how blending proved difficult for subjects to utilize when going from three to four and then to five layers (Figure 10). Stick Figures have similar issues with scalability, only in another direction. While one could add several more layers easily, this would immediately require the Stick Figures to be more spread out, decreasing the resolution of the encoding.

Data-Driven Spots does not have a demonstrated limit, but Bokinsky[11] was able to composite eight layers using Gaussian spots and rendered bumps. Given the density of regions in her composites, there doesn't seem to be much room for additional layers, and by including the lit bump texture, she is already using a hybrid of two globally parameterized patterns. Weigle et al. determined a minimum difference in sliver orientation of 15°, limiting the technique to 12 layers.[9] That would make it the most scalable method we tested; however, we will see in Section 5.2 that using this minimal difference in angle appeared to result in greater user error, which leads us to believe that we need see greater angles of separation for accuracy on our task.

4.3 Spatial Frequency

The final area of comparison is the amount of sub-sampling required by each method. With most of these techniques, the quality of the layout algorithm dictates the spatial frequency that can be represented. Much research has been done on algorithms to intelligently pack surface elements for dense, meaningful multi-layer visualizations.[8,18,19] We used a simple method where elements are randomly jittered from a uniform grid. This was enough for the frequency within our generated data sets in the study, but we wish to explore the limit of these techniques in this area in future work. For now, we can make some logical observations.

Stick Figures have the least potential for mapping high frequency data. In order for the Figures to be readable, they must not be so dense that they overlap. Thus each glyph should be given some space to itself. The DDS technique needs room for each of its dots, and the blurry edges of the dots make it hard to discern fine surface details on the elements. Slivers can be packed close together, but it becomes hard to read the orientations as more and more layers of dense slivers are composited together. Brush Strokes created a texture over the surface which can have some very fine detail. However, data values need to be averaged over the length and width of each stroke, making the technique limited in scalability to the maximum size of the strokes. Therefore, surface density may be limited by the desired scalability for this technique.

The only method which does not sub-sample the data surface is Color Blending, since every pixel is the weighted sum of layers' colors. This admits that Color Blending is perhaps best suited for high-frequency scalar fields with small pixel depths. Even in this case, the researcher must take care when choosing the color set. Also, the technique monopolizes the color attribute, making it difficult to overlay additional data, such as shape data.

*`http://www.colorbrewer2.org`

Figure 7. For the user study, we generated Gaussian kernels centered at random locations and varying slightly in their standard deviation. This data allows us to easily calculate the error based on the user's distance from the Gaussian center.

5. USER STUDIES

With the large parameter space of techniques and variations within each technique, we opted for testing the usability of the techniques with two small-scale, focused user studies. The emphasis here will be on qualitative results, although statistical tests will be used as supporting evidence for some of our conclusions.

5.1 Study Design

We selected our task in consultation with GIS domain experts; one fundamental task they identified was recognition of the maximum value of a scalar field in the presence of multiple fields. (This is akin to the task of finding critical points.[13]) We wanted to garner broad-based information about the utility of the five techniques described in Section 3. Thus our primary independent variable was Visualization, which took on the values Oriented Slivers, Data-Driven Spots, Brush Strokes, Color Blending, and Stick Figures. Since we were also concerned about the scalability of the techniques, we used a variable NumberOfLayers, which took on the value of 3, 4, or 5. We varied in which layer the subject searched for the maximum; TargetLayer took on the integer values 1-5. This variable was not crossed with Visualization and NumberOfLayers. It was also not distributed evenly because we limited the number of trials to nine for each Visualization, in order to keep each user's total time commitment to 30 minutes. TargetLayer took the value of 1 only once, whereas it took the remaining four values twice. Thus each subject completed nine trials with each Visualization.

We enrolled 15 subjects (11 male, 4 female; average age=35.4) in the experiment, yielding 675 data points. To discover the preliminary findings, we ran a 5 (Visualization) $\times$ 3 (NumberOfLayers) $\times$ 5 (TargetLayer) ANOVA using the Rweb1.03 server at `bayes.math.montana.edu`.[20] Dependent measures were distance-based error in normalized image space (range: 0-1) and response time (seconds). Users completed all tasks for a particular Visualization technique in a block; we recorded subjective workload evaluations (NASA Task-load Index,[21] range: 0-100) after each technique.

Users sat at a workstation computer, clicking on the critical point they identified with a mouse pointer. Each technique was given a single tutorial question, with the NumberOfLayers set to 5 and the TargetLayer selected from [1-5]. The order of the Visualizations was varied randomly for the first nine subjects, then counterbalanced in order to approximate a Latin Square design. For the practice questions, the correct maximum value was shown to the subject after they responded. After the tutorial pages, the subject was presented with a single task per screen, grouped by sequence. The task screens displayed the TargetLayer, the legend for the Visualization technique (shown in Figures 2, 4, 3, 5, and 6), the NumberOfLayers in the current Visualization, and finally the multi-layer Visualization image. At the bottom of every screen in the survey was a button labeled "submit and continue." User could change their responses until hitting this button. Response time was measured from initial display of a question until the clicking on their final location (not until the submit button was clicked).

5.2 Study Results

We found that Oriented Slivers and the Data-Driven Spots visualizations enabled users to find the maximum value of the requested field more accurately than the other visualizations – $F(4,56)=9.8364$, $p=0.000$ (Figure 8, left). Stick Figures yielded the greatest error, with Brush Strokes also yielding poorer than average performance. Users were fastest with Data-Driven Spots, Color Blending, and Oriented Slivers – $F(4,56)=34.0763$, $p=0.000$

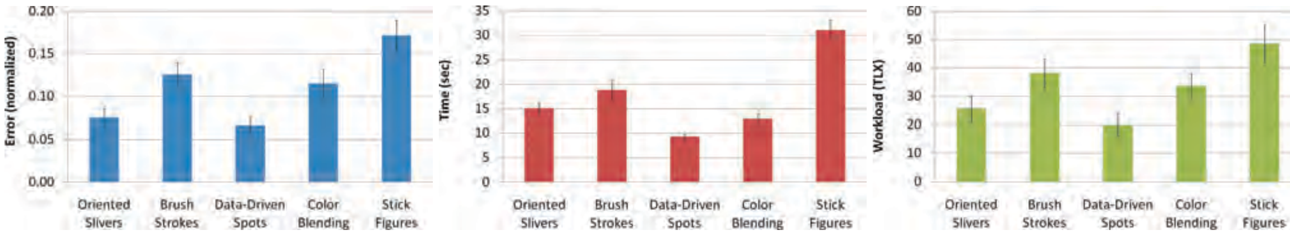

Figure 8. These three graphs summarize the effect of the Visualization on error (left), response time (center), and subjective workload (right). Subjects were most accurate with Data-Driven Spots and Oriented Slivers; they were fastest with Data-Driven Spots, Color Blending, and Oriented Slivers; they rated lower workload for Data-Driven Spots and Oriented Slivers. Note the similar pattern for all measures – two objective and one subjective.

(Figure 8, center). Stick Figures yielded the slowest performance. Subjective workload was lowest for Data-Driven Spots and Oriented Slivers – $F(4,56)=4.9599$, $p=0.002$ (Figure 8, right). This may be interpreted in several ways, but we prefer to think of the consistent pattern of results as an indication of the intuitiveness and general usability of the Data-Driven Spots and Oriented Slivers techniques. One may also reasonably infer that Stick Figures and Brush Strokes were less intuitive and the tutorial and legend less instructive than the instructional materials for the other visualizations. It may be that the proper interpretation of our results are that the consistency of the parameter representing field value increases the usability of a technique. Oriented Slivers and Data-Driven Spots always display field values with intensity, whereas Brush Strokes relies on a diverse array of parameters (length, width, angle, hue, and intensity of strokes). It may be that the angular measure of Stick Figures leads to confusion. Or perhaps presenting multiple visualizations with similar styles strengthened these techniques at the expense of the others. We noted that some visualizations led users to catastrophic errors (presumably searching the wrong layer). This again may signal a lack of clarity in the instructions for the interpretation of a particular visual representation. However, outlier analyses (using Studentized residual > 4.0 and a second analysis using > 3.0) indicated only seven and 24 outliers (1% and 3.6%, respectively) for error; removal of either set from the analysis served only to strengthen the results given above and below for error. (We report the weakest version of our results.)

The usability of various graphical parameters is also questioned by our finding of a significant difference in error depending on TargetLayer – $F(4,56)=7.9099$, $p=0.000$. The differences between techniques confound a detailed analysis, but this difference also was revealed by a significant interaction between the technique employed and the target layer – $F(16,224)=2.2134$, $p=0.006$ (Figure 9). We saw that poor performance with Stick Figures was caused not by layer L1 (represented with the torso), but by misinterpretation of the "limbs" of the figure (layers L2-L5). Similarly, it appears that the problems with Color Blending were caused almost entirely by the fifth layer, which had a green hue that – in retrospect and despite using a five-color scheme from the popular ColorBrewer palette – may not have had great enough separation from the first layer for the blending operation. Brush Strokes suffered most due to the hue key for the second layer, which ranged from blue to gold, but the intensity and the geometric cues (length, width, and angle of strokes) also did not fare as well as most other techniques. For Oriented Slivers, the 60° diagonal for layer L5 clearly yielded greater error than the other four. While Weigle et al.[9] proposed that 15° was the minimal separation necessary, our results lead us to wonder if greater separation would have been a more appropriate choice of orientations for the slivers. Finally, Data-Driven Spots used the same five colors as Color Blending, and it may have caused some catastrophic errors when considering layer L5, which saw the greatest error with DDS. There is not enough data to warrant statistical conclusions for these individual comparisons, but they certainly give us cues as to how to improve the study design and suggest hypotheses to make in future studies.

We noted a trend for an increase in the number of layers present from three to four and again to five to increase error – $F(2,28)=2.7236$, $p=0.083$ – and a significant effect on the response time – $F(2,28)=3.7298$, $p=0.037$. We expected this to be significant for both error and time, even in a small study. Data-Driven Spots and Oriented Slivers appeared to be affected very little by the change from three to five layers, whereas Color Blending changed performance by a statistically significant amount. This is somewhat encouraging that the visualizations may scale better than we had previously expected, but it clearly remains an effect to test in future studies.

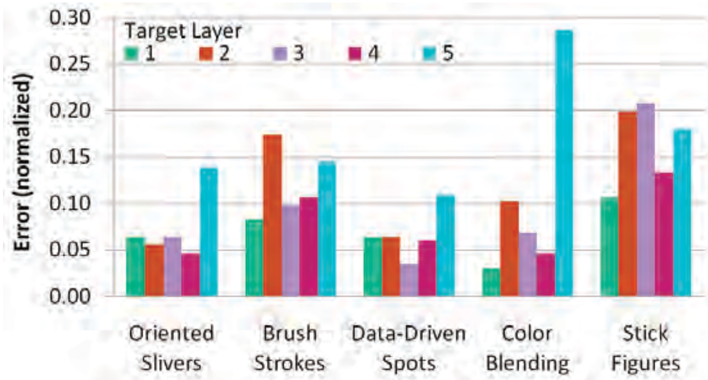

Figure 9. The error for some techniques varied widely with which layer was the target layer. Notably, Color Blending suffered perhaps some misidentification of the correct layer when targeting layer L5; this may have affected DDS for layer L5 as well. Stick Figures yielded better performance on the torso attribute than on the limbs. Brush Strokes fared more poorly when using hue as a key (for level L2) than intensity or the geometric cues. Oriented Slivers did not fare as well with the 60° orientation for layer L5. In color versions of this paper, the colors of the bars represent the colors of the layers in Color Blending and DDS.

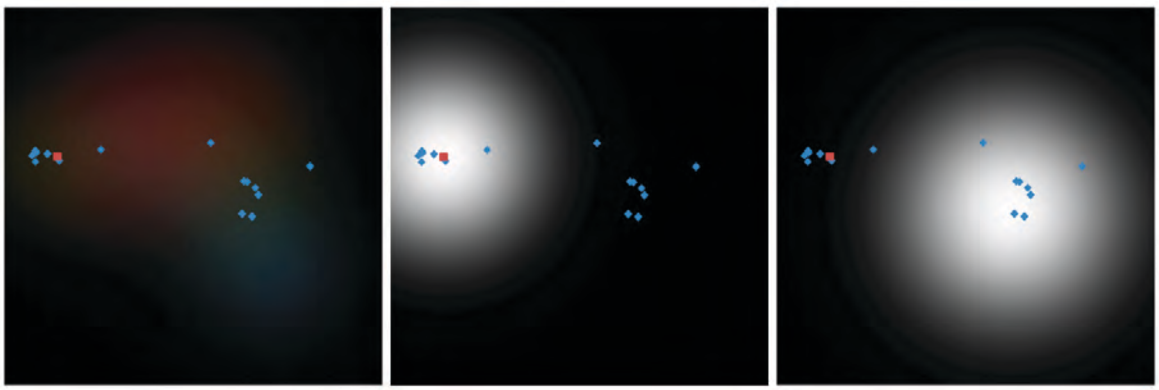

Figure 10. Subjects could confuse the shades of green in our color set. Here are user responses when asked to find the center of a Gaussian blob in one of five layers visualized with Color Blending. The left image is the actual composite, the center image is the target kernel (layer L5) and the right image is a kernel most subjects confused with the target kernel (layer L1). The square marks the center of the target kernel.

Several users commented that Stick Figures were confusing, which is clearly shown in the error measure. This was attributed, in one case, to the dependence of the limb angle on the body orientation to understand the data. Users had difficulty with Color Blending when there was large overlap between the target kernel and one or more of the distractors. Specifically, subjects could tell that they were mixing up the shades of green in our color set, and this can be easily seen when looking at their responses. Figure 10 shows the responses of the subjects to a question in the Color Blending session. The left image shows the Visualization with the responses overlaid. The Gaussian kernel from TargetLayer L5 is shown in the center image; the square marks its center, the correct solution. Most users gravitated toward the wrong shade of green (right image).

Some subjects raised concerns about the tutorials we used to explain each technique. One oversight in our Brush Stroke legend (Figure 3) was that we did not directly label which of the extremes (dark or light, thin or flat, blue or yellow, etc) corresponded to low values, and which corresponded to high values. In the tutorial, only the orientation was explained in detail, and users were given feedback on their understanding of the hue mapping in the tutorial example task. The mapping of intensity was straight-forward (low intensity proportional to a low value), but length and width were left completely to the user's personal visual reasoning. Given the data set, most users were able to see the region of long strokes or wide, blurry strokes, thus explaining the mapping. We should have clearly defined the mapping in every instance; only some users were able to fill in the missing information by reasoning from the provided image.

Several users complained about the format of the survey. Given the size of the images (1024 pixels square) and the vertically-oriented layout, some of the content (image, key, task instructions) may have been displaced off the screen, forcing the subjects to scroll the window to complete their task. The study should have been designed so that it would fit completely on a single screen, allowing the subjects to see the question, technique, and layer legend at a glance. This may have had a negative impact on workload.

5.3 Sensitivity to Monitor Settings

We conducted a small, separate study to determine whether the various techniques were likely to be sensitive to monitor settings, such as brightness, contrast, or gamma correction. We had three subjects (all of whom completed the first experiment as well) perform the study using a Dell 3008 WFP monitor under three conditions:

- factory default settings and lights on
- factory default settings and lights off
- altered contrast and gamma correction with lights on

The variables Visualization and Monitor Settings were crossed; each user completed nine trials for each pair in this cross product. NumberOfLayers and TargetLayer were randomly permuted in the same range as they (respectively) had in the first experiment. This yielded 3 (users) $\times$ 5 (Visualizations) $\times$ 3 (Monitor Settings) $\times$ 9 (trials) = 405 data points. We gathered summary statistics (mean and standard deviation) on sub-blocks for comparison with a series of Student t-tests. All tests were performed assuming unequal variances between groups using Welch's t-test and the Welch-Satterthwaite equation to compute degrees of freedom.

For Brush Strokes and Color Blending, we found no significant differences between the three environmental conditions listed above. We found this surprising for Color Blending; intuitively, it would be among the techniques most sensitive to color and intensity settings for the monitor. Subjects performed worst under the altered monitor settings with Oriented Slivers – $t(41)=3.227$, $p=0.003$ – and with Data-Driven Spots – $t(45)=4.110$, $p=0.000$. The subjects performed best with the altered monitor settings with Stick Figures – $t(28)=2.7665$, $p=0.010$. While this data involves only three subjects, the findings of significant differences are sufficient to raise concern that these techniques may well be sensitive to monitor settings. One practical implication of this is that we have dropped plans to implement a larger study using a web-based data collection instrument. We will instead have subjects come to our lab, where we can be sure that monitor settings will not confound the study design.

6. CONCLUSION

We have outlined the process and results of a usability study into several image encoding techniques for visualization of multi-layer data sets. Our logical analysis highlighted the potential for greater scalability with visualizations with repeated layers of few attributes. We also noted some trade-offs between scalability and spatial resolution. In our user studies, subjects performed significantly faster and more accurately when using techniques that employ a composite of parameterized patterns, like DDS and Oriented Slivers, over techniques that use complex surfaces with multiple local parameters, like Brush Strokes and Stick Figures. We also found a significant difference in the workload reported by users in order to complete a simple task of finding a critical point. We also saw that which layer was the target affected subjects' performance, further exemplifying that the number of parameters for surface elements can have an effect on subject comprehension. In defense of Stick Figures, other designs for Stick Figures may lead to different results. For all techniques, longer training, different tasks, and data types other than scalar fields all represent interesting avenues for future work. Further, a comparison of the best of these techniques against coordinated multiple views would be of great value. Animated extensions to these techniques (such as described for DDS[11]) add temporal sharing of the surface to the spatial sharing we employed thus far. We believe still images are limited in their usefulness in encoding multiple dimensions, and interaction is the key to seeing the full scope of the underlying data.

We also believe that multiple, globally-defined surface patterns (e.g. DDS or Oriented Slivers) may be combined to produce denser composites. Oriented Slivers and DDS composite layers separated by graphical attributes and using other attributes of sparse samples (glyphs) to show the data multiple patterns. Given the separation of shape and color between our DDS and Oriented Slivers implementations, a hybrid seems possible, where multiple globally parameterized patterns work together as a type of locally parameterized pattern with multiple surface elements. We wish to explore the usefulness of such hybrids in our ongoing work.

ACKNOWLEDGMENTS

The authors thank Denise Royle, the anonymous subjects, and the anonymous reviewers.

REFERENCES

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

[2] Thomas, J. J. and Cook, K. A., [*Illuminating the Path: The Research and Development Agenda for Visual Analytics*], IEEE Computer Society (2005).

[3] Healey, C., Kocherlakota, S., Rao, V., Mehta, R., and St. Amant, R., "Visual perception and mixed-initiative interaction for assisted visualization design," *IEEE Transactions on Visualization and Computer Graphics* **14**, 396–411 (March-April 2008).

[4] Zhou, M. X. and Feiner, S. K., "Data characterization for automatically visualizing heterogeneous information," in [*Proceedings of the 1996 IEEE Symposium on Information Visualization*], 13 (1996).

[5] 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).

[6] 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 [*IEEE Visualization*], 127–134 (1998).

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

[8] 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).

[9] 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).

[10] Jenks, G. F., "Pointilism as a cartographic technique," *The Professional Geographer* **5**, 4–6 (Sept. 1953).

[11] Bokinsky, A. A., *Multivariate data visualization with data-driven spots*, PhD thesis, The University of North Carolina at Chapel Hill (2003).

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

[13] 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).

[14] 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*], (2006).

[15] Urness, T., Interrante, V., Marusic, I., Longmire, E., and Ganapathisubramani, B., "Effectively visualizing multi-valued flow data using color and texture," in [*IEEE Visualization*], 115–121 (October 2003).

[16] 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).

[17] Pickett, R. and Grinstein, G., "Iconographic displays for visualizing multidimensional data," in [*Proceedings of the 1988 IEEE International Conference on Systems, Man, and Cybernetics*], 514–519 (Aug 1988).

[18] Kindlmann, G. and Westin, C.-F., "Diffusion tensor visualization with glyph packing," *IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization / Information Visualization 2006)* **12**, 1329–1335 (September-October 2006).

[19] 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*], 333–340 (1999).

[20] Banfield, J., "Rweb: Web-based statistical analysis," *Journal of Statistical Software* **4** (Mar. 1999). http://www.jstatsoft.org/v04/i04/paper.

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