Visual Encodings that Support Physical Navigation on Large Displays

Alex Endert, Christopher Andrews, Yueh Hua Lee, Chris North

Virginia Polytechnic Institute and State University

ABSTRACT

Visual encodings are the medium through which information is displayed, perceived, interpreted, and finally transferred from a visualization to the user. Traditionally, such encodings display information as representations of length, color, size, slope, position, and other glyphs. Guidelines for such encodings have been proposed, but they generally assume a small display, small datasets, and a relatively static user. Large, high-resolution visualizations are able to display far more information simultaneously, allowing users to leverage physical navigation (movement) as an effective interaction through which to explore the data space. In this paper, we analyze if and how the choice of visual encodings for large, high-resolution visualizations affects physical navigation, and ultimately task performance for a spatial information visualization task.

KEYWORDS: large, high-resolution display, information visualization , aggregation, perceptual scalability

INDEX TERMS: H.5.m [Information Interfaces and Presentation]: Miscellaneous

1 INTRODUCTION

Large, high-resolution displays are compelling tools for information visualization. The number of available pixels allows for large quantities of information to be simultaneously displayed. These displays typically exceed the limits of visual acuity, providing the user with the opportunity to employ *physical navigation* to explore the data space. In other words, rather than using tools such as pan and zoom, the user moves his or her eyes, head, and body to navigate the space. For example, to perform the common visualization task of moving between overview and detail, the large, high-resolution display user can simply move back or step closer to change the level of detail. Physical navigation allows the user to leverage physical and cognitive tools like proprioception, optical flow, and spatial memory to enhance his or her comprehension of the visualization. The environmental cues provide context, coherency, and a reduced need for internal memory, which can improve user performance for basic visualization tasks [1, 22].

In order to gain insight from visualizations, users rely heavily on perception as a means for interpreting and decoding the embedded information they see. The properties of human perception provide both an opportunity as well as a challenge for visualization designers. By leveraging the properties of human perception, visualization designers can improve both the quality and the quantity of information displayed [19]. Conversely, human perception can also be "fooled" by misleading or unintended patterns in visualizations, leading to misapprehensions and misunderstandings [14]. Thus, it is critical to understand the properties of different visualizations and how they will be perceived.

The introduction of physical navigation creates an environment where visualizations are potentially observed from many distances, viewing angles, and scales. This has the potential to create additional perceptual issues due to: (1) the increased role of peripheral vision, (2) potential distortions of certain encodings caused by extreme viewing distances and angles, and (3) different encodings' abilities to display meaningful information when observed beyond the limits of basic visual acuities. We also need to consider how encodings interact as more of them become visibly available – do they combine into meaningful patterns, or do they interfere with one another and create false patterns or no patterns. We refer to the overall effectiveness of a visualization as it scales up as its *perceptual scalability*.

Of particular interest is the perceptual scalability of visual encodings with respect to *physical zooming*, that is, moving forwards and backwards to traverse the various levels of detail from full overview to the details of a single glyph. Ideally, as the user moves back away from the display, encodings should "adapt" and continue to provide the user with meaningful information. While this may be provided actively by tracking the user and updating the visualization, we are primarily concerned with how the actual attributes of the encodings are perceived as the user moves around.

As the user moves back away from the display, it creates an effect we refer to as *visual aggregation*. Visual aggregation occurs through a combination of visual acuity (details of individual glyphs start to be lost with distance) and perceptual effects caused by the introduction of increased numbers of glyphs into the visual field. The painting technique of pointillism makes use of this phenomenon to turn what appears to be a field of individual colored dots into recognizable scenes as the viewer steps back from the painting. Similarly, the goal of visual aggregation in

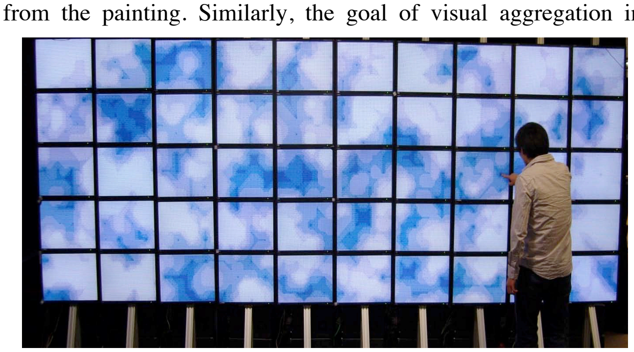


Figure 1. A user standing while analyzing a 100 megapixel large display visualization encoded using color.

{aendert, cpa}@vt.edu, alvalee99@gmail.com, north@vt.edu

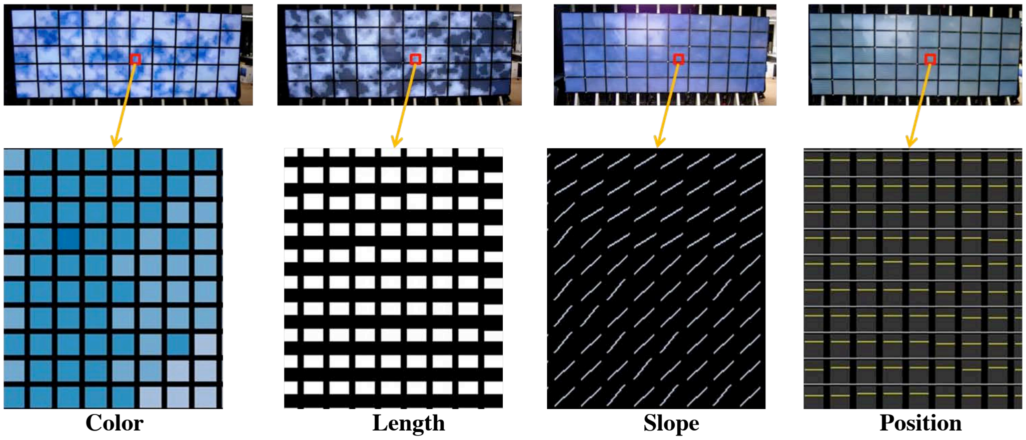


Figure 2. Overview with corresponding enhanced views of a small portion of the visualization for each encoding.

information visualization is for the viewer to be able to perceive patterns in the data when the visualization is viewed from a distance. Thus, visual aggregation becomes important when considering physical navigation as the means for interaction, just as computational aggregation techniques [10] are important when users virtually navigate (e.g., panning, zooming, etc.) a visualization.

In this paper, we describe our study of four primary visual encodings (color, length, slope, position) on a large, high-resolution display. Our analysis examines the performance and the behavior of subjects performing a basic visualization task to provide insight in the perceptual scalability of the various encodings.

2 Related Work

Previous work details the properties and construction of visual encodings, and lists numerous encoding variants [3],[5],[18],[19],[13-14]. The design of encodings is more complex than may first seem, relying on a combination of the human visual and perceptual abilities [19]. A significant amount of work has been conducted to determine the perceptual effectiveness of many of these visual encodings. For example, Healey et al. examined which encodings support fast preattentive processing by the human visual system [11] Cleveland and McGill [8] comparatively tested user performance of a common set of visual encodings, determining a relative order of effectiveness for standard, paper-sized visualizations containing a small number of datapoints. Color, an inherently complex encoding, can be further analyzed in terms of its components (i.e. hue, saturation, luminance) [20].

However, these guidelines were developed assuming that the user is focused on a fairly small amount of data in visualizations on standard-sized displays. Furthermore, their studies assume the user is at a fixed viewing distance and angle from the visualization, where each encoding is perceived within the applicable visual acuity required to decode the glyph. This leads us to the question of how effective these encodings are when presented on a large high-resolution visualization, where users are constantly changing these view parameters, and greater amount of data is visualized.

Physical navigation is fundamental to the use of large, high-resolution displays, suggesting care be taken in the design of tools and visualization to ensure proper support. Thus, we are interested in studying how the visual encodings effect the physical navigation based on how well each encoding visually aggregates. Some initial work on the relevant issues may help to inform design:

In terms of (1) exploiting peripheral vision, color encoding should emphasize luminance over hue [19]. Motion and animation have shown to be effective for peripheral awareness [2, 16]. Chewar et al. found that position (given a common axis) and color were effective in busy dual task scenarios [7]. While these results say more about peripheral attention, peripheral vision likely played a role due to the visual separation of the dual tasks.

In terms of (2) distortions due to extreme viewing angles, Wigdor et al. examined how visual encodings are affected by the position of the viewer [21]. In particular, they looked at what happens when information must be displayed at oblique or right angles to the viewer. They found that the accuracy of length, position, and angle are significantly superior to area and slope under these conditions.

In terms of (3) visual aggregation, Yost et al. suggests the use of "filled" encodings such as colored bar-graph instead of line-graph encodings to better enable visual aggregation at a distance [22]. Eick et al. point out the importance of considering the scalability of the individual encodings (i.e. what is the minimum amount of pixels needed to produce an effective glyph given an encoding) [9]. To understand how these issues come together and develop guidelines for large display visualizations, we take a holistic approach to examine the effectiveness of various visual encodings in the context of large scale visualizations on large high-resolution displays.

3 Method

The purpose of this study is to analyze how the choice of visual encodings can directly impact physical navigation and task performance on large, high-resolution visualizations. In particular, we are interested in how the encodings impact physical navigation behaviors and strategies (especially whether they support or inhibit physical zooming) and how those behaviors translate into task performance. Ultimately this will help us to understand the issues that affect the perceptual scalability of various encodings. To accomplish these goals, this study examines user task performance and user behavior with four glyph-based encodings on small and large displays, where users are physically navigating in the presence of a large amount of uniformly spatially referenced single-dimensional data. We hypothesize that the design of the encoding directly impacts the users' ability to physically navigate, ultimately affecting task performance.

3.1 Equipment

The display used for this experiment is a tiled powerwall consisting of fifty twenty-inch LCD monitors arranged in a 10 x 5

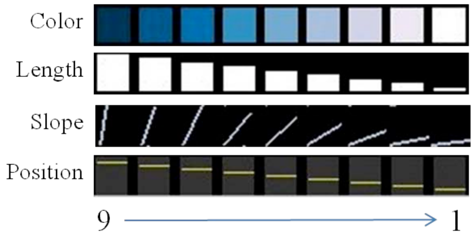


Figure 3. Legend of each encoding used, in order of their corresponding value, ranging from 9 (highest) to 1 (lowest).

matrix, providing a total resolution of 16,000 x 6,000 or 96 Megapixels (Figure 1). For the large display visualizations, the participants were shown a full-sized visualization covering the entire display, while the small display visualizations used a single monitor in the center of the display. The visualization was written in OpenGL, using Chromium to distribute it across the wall [12].

The physical dimensions of the open floor in front of the display measures 17 ft. wide by 15 ft. deep. Subject behavior was recorded using two video cameras and a Vicon motion capture system that tracked the subject's head.

3.2 Visualization Design

For this study, the visualizations we developed are conceptually based on the embedded visualization recommended for large, high-resolution displays by Yost et al. [22]. In essence, rather than focusing on a single, simple visualization across the whole display, in which position would be the most obvious encoding, we are considering the case in which position on the display already has meaning – for example, population statistics or sensor data embedded within a geospatial visualization. Each visualization was generated using the Diamond Square algorithm [15], which produces realistic terrain visualizations (Figure 2).

The values we chose to represent ranged from one to nine. Due to the quantitative nature of our data, the encodings we chose to test were *color*, *length*, *slope*, and *position*. A sample of each glyph used in the study can be seen in Figure 3. Each glyph is 20 x 20 pixels, and a 5 pixel buffer separates the glyphs on each side. Using a constant glyph size allowed us to hold size, spacing, and data density constant independent of encoding or value. While a color glyph may occupy as little as one pixel, we found that a 20 x 20 glyph is a reasonable size for the other encodings, as their construction requires a higher number of pixels.

This glyph size allowed us to display 153,600 glyphs in the large visualization and 3072 glyphs on the small visualization. The size of the glyphs did not change from the large to the small condition, only the number of data points.

Color. Each glyph for the color encoding occupied the entire 400 pixels. The colors we used were originally developed by Brewer et al. [4], who found this particular color ramp effective for quantitative data such as ours.

Length. The length encoding consisted of a glyph including a 20 pixel wide, solid white bar varying in height depending on value. The tallest (i.e. highest value) bar was 18 pixels tall, while the shortest was only 2, meaning that each intervening value was represented by an additional two pixels of height. The reason that neither 0 nor 20 were used was because during testing we found that when the minimum and maximum values were represented with the two extremes (0 and 20 pixel tall bars, respectively), the glyphs were too definitive, removing the need for comparisons, making them unsuitable for this particular study.

Slope. The glyph utilizing the slope encoding was a line with varying slope, ranging from 5 degrees (representing the value 0) to 85 degrees (representing the value 9). The desire to avoid definitive encodings (0 and 90 degrees) again motivated the choice of range. Due to the total size of 400 pixels allocated to each encoding, aliasing is present with many of the slopes drawn.

Position. Position, in this study, is represented as a yellow, one pixel line drawn on a dark gray background. The position of this line from the bottom is equal to the height of the bar drawn in the length encoding glyphs. Each glyph has its own axis, as the spatial location occupies meaning in the visualizations used in this study, and thus a full display scatterplot with a common axis was not used.

These encodings were chosen based on their defining characteristics. For instance, position and length are different, as position can act as length if it is filled in (in this case, filling in the area under the yellow bar), but using these two allows us to explore this difference. The slope encoding, when used in this style of visualization, allows us to observe how visual aggregation creates a textural effect. We elaborate on these points in the discussion.

3.3 Experiment Design

The experiment had 12 participants (1 female and 11 male). All of them were undergraduate computer science students. The ages of the participants ranged from 20 to 23, with an average age of 21. No participant reported any known color blindness. Six participants reported corrected vision. None of the participants had prior experience with large display visualizations.

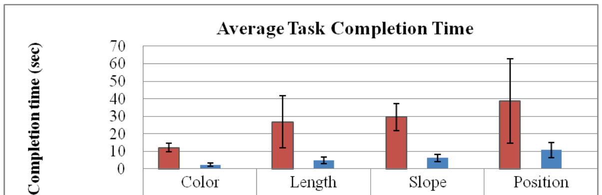
Each participant was shown 32 visualizations, broken down as follows: 16 large display visualizations (4 per encoding), and 16 small display visualizations (4 per encoding). For each trial, the ordering of the 32 visualizations was randomized to minimize any ordering effect. Prior to the study, each participant was shown a total of 8 practice visualizations (one of each encoding, both on the small and large display condition) to get comfortable with the task, display sizes, and encodings. The task for each visualization was to find the maximum value in the visualization, and walk up and point to it – an inherently multiscale task requiring both overview of the entire visualization as well as detail of specific regions. Two starting locations were marked on the floor: one for the small display visualizations placed 3 feet away, and one for the large display visualizations placed 9 feet away, both centered with respect to the display.

4 Results

We collected both quantitative and qualitative data during this study. Timing and accuracy data were obtained for each visualization in order to assess task performance. We asked participants to perform their task as "quickly and accurately as possible", informing them that they are only allowed one answer per visualization. VICON (motion tracking) data was recorded for analysis of their physical navigation and behavior. In addition, video and audio footage was captured for post-study analysis. Finally, following each trial, the participants were debriefed with a series of questions concerning their experience, strategies for solving the task, and comments regarding encoding preference.

4.1 Task Completion Time

The timing data corresponds to the time elapsed between the start of a single visualization (the participant was at their marked starting location and the visualization was shown) and the participant signifying their answer by pointing to a glyph on the



	Color	Length	Slope	Position
Large	12.085	26.813	29.502	38.699
SD	4.823	29.928	15.545	47.931
Small	2.346	4.857	6.284	10.656
SD	1.660	3.530	4.196	8.958

Figure 4. Average task completion time and standard deviations (SD).

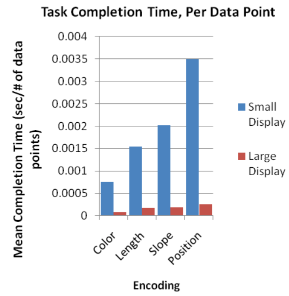


Figure 5. Average task completion time, normalized based on number of data points.

map (physically touching the glyph). We allowed only a single answer as we are interested in how quickly the participants found what they perceived as the correct answer (i.e. when they thought they were finished). They were informed of a two minute time limit for each visualization, but no participant ever reached the limit. However, a small number of "best guesses" were observed (especially for the position encoding on the large display), where participants essentially gave up and chose a nearby high-value glyph after exhaustive searching. The accuracy of their responses is analyzed further in section 4.2 in terms of "rate" (whether their answers were correct or not) and "response" (the value of their answer versus the correct answer).

There are two independent variables in this study, encoding type and display size. Table 1 shows the means and standard deviations for completion time for the two primary effects. The color encoding is the best among the four encodings in terms of completion time (7.216s). Position showed the poorest performance (24.822s). We performed a two-way within subject ANOVA to test the effects of our two factors (encoding type and display size). In order to meet ANOVA assumptions, square root values were used for analysis. We found a significant effect for both encoding type and display size: $F(3,33) = 11.10$, $p < 0.001$ and $F(1,11) = 92.419$, $p < 0.001$, respectively. An LSD post hoc test revealed that participants exhibited the shortest completion time for visualizations encoded using color ($p<0.05$). There is no significant difference in completion time between slope and length ($p>0.05$), and participants took the longest time to complete the task when position was the encoding used ($p<0.05$).

Table 1. Mean and standard deviations (SD) for task completion time in terms of the two primary factors (encoding and display size).

	Encoding				Display Size	
	Slope	Color	Length	Position	Large	Small
Mean	18.130	7.216	15.835	24.822	26.810	5.978
SD	16.331	6.070	17.326	26.208	21.989	4.635

As such, we found that there is no significant two-way interaction term. This means that in both display size conditions, color is the significantly better encoding than length and slope, and these two encodings are significant better than position.

Figure 4 shows the order of effectiveness in terms of completion time remains unchanged from the small to the large display trials and an increase in task completion times occurred between the large and small display condition. This is expected as the number of data points increased. Interestingly, while not the focus of our study, our results also duplicate a result reported by Yost et al. [22]. Upon analyzing the timing data, we see that completion times on the large display are larger, but they are not larger by the same factor by which the data was increased, a factor of fifty. A normalized comparison of the small and large display conditions can be seen in Figure 5, where we present a "time per data point" view of the results for each encoding, derived by dividing the mean completion time by the number of data points used for each display size.

4.2 Accuracy

We analyzed accuracy in two ways. First, we analyzed "rate", the percentage of "correct" (i.e. participant selected the value 9) versus "wrong" answers (i.e. participant selected any value other than 9). Second, we analyzed the actual responses, analyzing the participants' answers versus the correct (i.e. value 9) answer.

4.2.1 Rate

As mentioned before, participants were asked to find the maximum value and they were only allowed a single attempt per trial. There were 8 (4 encodings x 2 display size) combinations in the experiment, each repeated 4 times under each combination. The means and standard deviations are shown in Table 2. As shown, participants exhibited the best performance when using the color encoding (accuracy rate=0.979). In contrast, position performed the worst among these four encodings (accuracy rate=0.392). In terms of display size, participants had higher accuracy rate in the small display condition (accuracy rate=0.879) than in the large display condition (accuracy rate=0.331).

Table 2. Mean and standard deviations (SD) for accuracy rate in terms of the two primary factors (encoding and display size).

	Encoding				Display Size	
	Slope	Color	Length	Position	Large	Small
Mean	0.448	0.979	0.594	0.396	0.328	0.880
SD	0.410	0.071	0.422	0.382	0.340	0.186

A within subject ANOVA for encoding type and display size shows significant effects: $F(3,33) = 72.653$, $p < 0 .001$ and $F(1,11) = 3183546$, $p < 0.001$, respectively. An LSD post hoc test revealed that each encoding is statistically different from each other except for slope and position ($p>0.05$). This means that the accuracy rate is significantly higher for color than for length ($p<0.05$), and length is significantly higher than both slope and position ($p<0.01$). There is no significant difference between slope and position ($p>0.05$).

The two-way interaction term between encoding type and display size is significant, $F(3,33)=30.732$, $p<0.001$. Figure 6 shows the accuracy rate for each encoding given the display size. The effect of display size is not significant for the color encoding ($p>0.05$). This means that using color, participants' accuracy rates

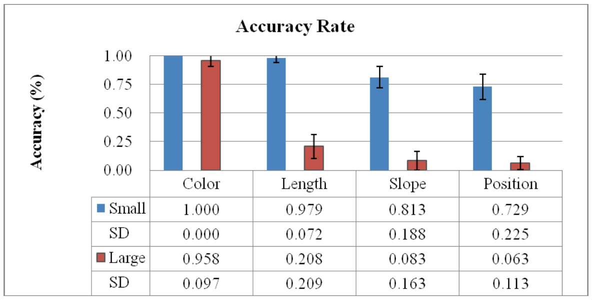


Figure 6. Accuracy rates and standard deviations (SD), where 1.0 denotes all responses for that condition were correct.

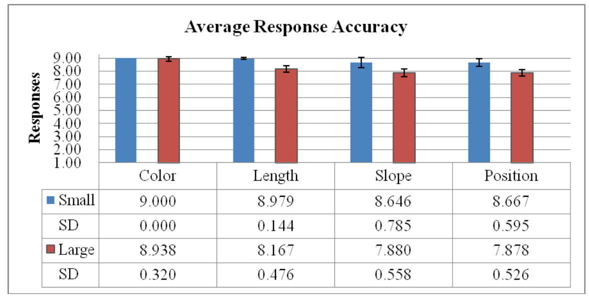


Figure 7. Average response accuracy (9 being correct).

do not show a significant difference between the small and large display conditions. In contrast, when using length, slope, or position, the effect of display size becomes significant. The small display condition performs significantly better than the large display condition for these three encodings ($p<0.001$).

There is no significant difference in terms of accuracy rate between color and length for the small display condition ($p>0.05$). However, the accuracy rate does significantly decrease from color to length in the large display condition ($p<0.001$). This suggests that for the large display condition, the color encoding maintains effective accuracy rates compared to the small display condition.

Further, a post-hoc Tukey test for the small display condition reveals the following groups based on similarity: (1) color and length, and (2) slope and position. Similar analysis of the large display condition reveals these groups: (1) color, (2) length, and (3) slope and position. This will be discussed in the next section.

4.2.2 Response

The second analysis of accuracy is in terms of the actual value of each user's response. We define the response as the average of each participant's value of their answer for the four times they were presented that encoding type and display size condition. The response analysis is meant to represents how close a participant's response to the correct answer. The larger the response value, the closer to it is the correct answer, 9.

Table 3 shows the means and standard deviations of each encoding type and display size for the participant responses. As shown, participants had the smallest amount of error from the correct answer using the color encoding. In contrast, position performed the worst, as the mean is the farthest from the accurate answer. Overall, the small display condition trials resulted in answers closer to the correct answer than large display condition trials. This result is expected, as the number of data points is significantly reduced (by a factor of 50), and the entire dataset was visible to the participants without any interaction or movement.

Table 3. Mean and standard deviations (SD) for user response accuracy in terms of the two primary factors (encoding and display size).

	Encoding				Display Size	
	Color	Length	Slope	Position	Large	Small
Mean	8.969	8.573	8.255	8.268	8.210	8.823
SD	0.111	0.428	0.493	0.493	0.481	0.289

A two-way within subject ANOVA was used to analyze the responses. This analysis reveals that there are two significant main effects, encoding type: $F(3,33)=44.642$, $p<0.001$ and display size: $F(1,11)=149.249$, $p<0.001$. An LSD post hoc test shows that color performs significantly better than length ($p<0.001$), and length is significantly better than both slope and position ($p<0.005$). There is no significant difference between slope and position ($p>0.05$).

The interaction between encoding type and display size is significant ($F(3,33)=3.184$, $p<0.001$) and shown in Figure 7. The effect of display size is not significant for the color encoding ($p>0.05$), which means that when using color, the participants' responses show no significant difference between the small and large display conditions. In contrast, when using length, slope, or position, the display size effect is obvious. These three encodings perform significantly better under the small display condition than the large ($p<0.001$). This further illustrates the point that although differences between encodings occur even on small display visualizations, the effects are amplified on large display visualizations.

There is no significant difference in response values between color and length for the small display condition ($p>0.05$). However, there is a significant decrease in performance from color to length on the large display condition ($p<0.001$). This suggests that when used for a large display visualization, the color encoding maintains a comparable response accuracy as the size of the visualization increases. On the small display, participants exhibit similar performance when using either the color or length encoding. These results again suggest that for both the small and large display conditions, color is the superior encoding in terms of response accuracy.

Further, a post-hoc Tukey test for the small display condition reveals the following groups based on similarity: (1) color and length, and (2) slope and position. The analysis of the large display condition shows these groups: (1) color, (2) length, and (3) slope and position.

A difference in the breakdown of the response accuracy can also be seen based on the encoding. For the small display condition, the majority of the answers were correct. However, there are significantly more errors for slope and position than for color and length – hence the two significantly different groups. For color on the small display condition, 100% of the reported answers were correct, while length only had one incorrect answer. In the large display condition, the groupings changed – color is significantly different from the remaining encoding types. This is reflected by color being the only encoding on the large display condition where the majority of the answers are correct (only 1 seven and 1 eight reported). Further, length was significantly different due the responses having a higher amount of correct answers compared to slope and position. The overall ordering of effectiveness in terms of accuracy remains unchanged between the small and large display.

We believe the difference of accuracy and task completion time between the large and the small display conditions was largely

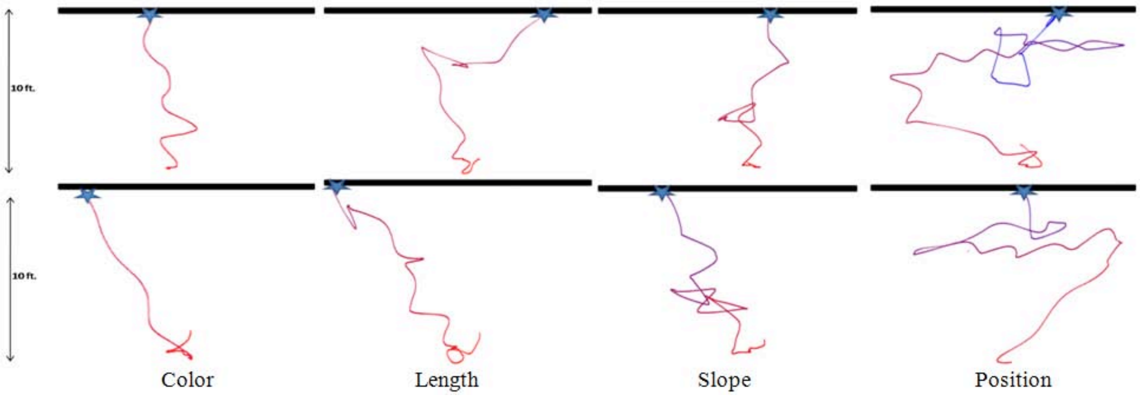


Figure 8. Motion traces showing a top-down view of a single user physically navigating while analyzing the large display visualization. Encodings shown below each column. The traces shift color from red to blue to indicate time (on a common scale). The distance from the starting point in the back of the space to the display at is 10 feet. The width of the display is over 14 feet.

due to the way each encoding visually aggregates. For the small display, the encodings were only observed from a detail level (i.e. the participants did not have to physically navigate to gain an overview). However, in the large display condition, it became critical for the encodings to function effectively in the aggregated (overview) state, allowing for effective physical navigation strategies. Through analyzing the participants' movements and behaviors, we gain further insight into how each encoding supported visually aggregation and physical navigation.

4.3 Movement and Behavior

We draw from both our observations during the trials as well as recorded motion capture data for our analysis of participants' physical movements and behaviors during their trials. This data provides valuable evidence as to the way each participant reacted to the visually aggregated form of each encoding.

From the motion capture data we extracted top-down traces of the movement of the participant through the space in front of the display over the course of a single task, as shown in Figure 8. The presented traces are representative samples of the behaviors exhibited across the participants in reaction to the different encodings. Each trace begins at a fixed starting point (center of the display, 9 feet back), and ends with the participant reporting their answer while touching the display (represented by the star). These views only show 2 dimensional movements and do not account for vertical head movement generated by participants leaning down or standing on their toes to view the low and high regions of the display. The movements associated with the small display visualizations were negligible, as the participants did not have to navigate physically to use the visualization. Each trace is on a common distance scale, as well as a common time scale (represented by the color shift from red to blue).

Color. As illustrated in the example motion traces, the actual identification of the target glyph or its close neighborhood could be done from the back of the space. Once the target region was identified the participants would move very directly up to the display to point at their answer. As we can see in the example traces, there was not a lot of re-examination of the visualization as they got closer, indicating that they were very sure of their response from a distance. There are a couple of explanations for this. First, the color glyphs make maximum use of the available pixels, making them easier to see from a distance. Second, color aggregates well into regions. As one participant noted, it was "easier to see what areas *not* to look at from back here". What he was referring to was the ability to stand back and spot areas where no potential targets existed, thus reducing the extent of the visualization that had to be considered and examined.

Length. The behaviors observed during the length encoding condition were similar to those displayed during the color condition. As with color, participants seemed to be able to quickly winnow down the regions of interest, but then they had to "zoom in" (walk closer) to look at the detail view for actual comparisons. This can be seen in the motion traces: as participants move towards the display, it is not to immediately identify the answer, but rather to compare a small number of possible targets, and only then making a decision. For example, the top length participant shown in Figure 8 moved towards the display to compare two possible targets. His first target was near the middle of the visualization. He approached the target, then backed up to his right, compared the two targets, and ultimately made a decision to select the target on his far right. Participants reported that the length encoding roughly aggregated to color regions that helped to reduce the potential regions of interest. As one participant pointed out while analyzing the visualization from the starting point, "I have to find areas with lots of white". However, individual differences between glyphs could not be seen from a distance, requiring participants to move closer to make comparisons between candidate targets.

Slope. For the slope encoding, we can see that the participants immediately moved closer to the display, stopping about halfway there, and then spending a large amount of time at that distance to analyze the visualization, before moving closer to confirm their answers. The movements indicate that they were unable to make out values (of both individual glyphs or clusters) from the starting point. After moving closer, they were able to see values of these clusters of interest, then move in yet closer to confirm their findings. One of the participants provided this explanation for the observed behavior: "With slope, I could see the boundaries of the clusters, but had to come closer to tell if they were high or low values".

Position. The movement traces show fairly clearly how ineffectual the position encoding was. The longer traces indicate a longer task completion time, with the movements and paths illustrating an inefficient method of physically navigation. The motion data shows both participants immediately moving right up the display (at a distance of less than 4 feet from the display) and performing "lawnmower" searches, considering each value individually. After scanning the entire display, they would move back to a point where they felt confident of their answer and report it. This strategy also made it difficult for potential candidate to be remembered and returned to since there was minimal context to help the participant relocate a particular glyph. This poor physical navigation strategy can be primarily attributed to the paucity of information available from the "overview" obtained from the back of the space (near the starting point, 9 feet from the display). Participants commented that they "cannot make out anything from [this distance]", and "it all looks the same". Instead of gaining an overview, one participant described the visualization from this distance as "distracting and not useful". This meant that no regions could be ruled out and every glyph had to be examined until the target was identified.

5 Discussion

By combining the performance and behavioral data, we gain insight into the perceptual scalability of these encodings. The degree to which an encoding effectively aggregated is clearly shown in the patterns of behavior. There is then a clear link to the performance associated with each encoding, showing the importance of visual aggregation and perceptual scalability. In addition, we can point out that this also impacted the participants' encoding preference, as all 12 participants reported color as their preferred encoding in both conditions. Further, we can discuss factors to consider when choosing visual encodings intended for large, high-resolution visualizations.

The primacy of physical navigation in the use of large, high-resolution displays means that it must be considered during the development of any tool or visualization for these displays. As such, we must ask how the results of our study can inform us about the design of visual encodings (and their ability to cope with visual aggregation) to support physical navigation. Drawing from both the participants' behavior and their comments, we can further consider the properties of encodings that leads to useful visual aggregation. Each of these properties must be analyzed in terms of the visual and perceptual system of users, as both are inherently necessary for information visualization.

We can identify three levels of aggregation that are important. The first level is the ***glyph level*** – what happens to individual glyphs as the limits of visual acuity are reached. As the user moves away from the glyph, fine details will be lost, and we can roughly approximate the behavior of glyphs as the average of the pixel values within the region it is displayed (including any background used to differentiate it). This also means that pixels having a low contrast with respect to the background will also fade. We can take advantage of this effect to "remove" detailed information from the overview [22], but we also want the remaining encoding to provide at least an approximation of the original value. Here we see one of the real benefits of color as an encoding – the average across the glyph is the same as the glyph, so even as pixel-level acuity is lost, the glyphs retain their distinctness.

The second level of aggregation occurs when ***multiple glyphs merge***. Here, two or more adjacent glyphs appear to merge together, either through loss of visual acuity or through perceptual factors, such as the Gestalt principle of similarity, which causes objects with similar properties, such as color or shape, to be seen as being part of a group [6]. Color is particularly compelling because it is perceptually dominant and is robust in the face of feature removal (i.e., red remains red at a distance, while a pentagon may become indistinguishable from a circle). The factors for determining how well a glyph aggregates at the glyph level should be taken into account again here as well. There should be a smooth transition from one glyph to the next (in terms of pixel usage values).

Finally, the third level of aggregation is the ***aggregation of multiple glyphs into a field***. At this level, the combined contribution of a multitude of glyphs will result in fields of either color or texture depending on the underlying glyphs. As the goal is to detect patterns, there are two primary considerations for an encoding. The first consideration is for perceptual boundaries between regions containing different values. The second consideration is for the distinctiveness of those regions. In other words, can the viewer not only tell that there are two regions, but can he or she also tell what the dominant value is for the region.

It is instructive to now reconsider the encodings in light of this model. The color encoding is particularly effective because it scales well. The use of the entire glyph to represent the encoded value meant that individual values were still available in the overview. Perceptually, color groups well, so the delineation of regions was high. In addition, since the aggregated regions used the same encoding as the individual glyphs, it was easy for the participants to assign values to various regions.

The length encoding is weaker than color for glyph-level aggregation, functioning better on a detailed level where comparisons have to be made between values (e.g. what percent is X greater than Y). Additionally, there are clear differences from low to high values (as can be seen in Figure 3), but there is little to differentiate neighboring values. This is borne out by the behavior of the subjects. While they were able to use visual aggregation to rapidly select areas of interest, it was a rough "lighter" or "darker" differentiation, lacking the finer granularity provided by color.

The slope encoding is interesting because the individual glyphs aggregate into the same color. The only difference between regions was a textural one caused by the different orientation of the lines. The orientation contrast allowed the boundaries of regions to be perceived [17], but the lack of distinctiveness between regions made it impossible for the participants to say anything about the general value of a particular region. This explains the difficulty experienced by the subjects with this encoding. They could perceive patterns from a distance, but they could not actually know what value a region represented.

Considering the position encoding, we see it is lacking in both of these features. The value of the glyph is encoded entirely by the thin bar. As such, the color of the aggregated form of the encoding is the same for all values – a slightly yellowish gray that was only barely distinguishable from the background. The bar is also a feature that is rapidly lost due to limitations of visual acuity, so no remnant of the value remains in the full overview. Closer in, since the orientation of all of the bars is the same, there is also no

Table 4. Summary of visual aggregation

	Distant (~ 9 ft.)	**Intermediate (~ 5 ft.)**	**Close (~ 2 ft.)**
Color	Aggregates as color Distinguish individual and field values All tasks	(same)	(same)
Length	Aggregates as luminance Approximate field values Region Identification	Increased aggregate granularity Approximate individual values Candidate selection	 Distinguish individual values Comparison and selection
Slope	Aggregates as texture Recognize field boundaries Task not supported	Aggregates as flow fields Approximate field values Region identification, Candidate selection	 Distinguish individual values Comparison and selection
Position	No aggregation Task not supported	No aggregation Task not supported	"Lawnmower search"

variation in texture, and so no fields are perceivable. There is some merging between glyphs, but only horizontally to form "waves" along the rows, which no participant found useful for detecting the strictly two-dimensional patterns in the data.

Overall, the visual aggregation results from this study are summarized in Table 4. In general, for visual aggregation across uniformly placed representations, color should be a primary consideration. It is clear that the most control over the aggregated form is achieved by explicitly encoding the attributes containing potential patterns into color. This works best with ordinal data that can be encoded using some form of color ramp. This will allow similar values to be encoded into similar colors, making it easier to perceive regions that are not necessarily dominated by a single value, but containing a range of neighboring values. The length encoding worked without color, because the quantity of pixels varied between values, creating a perception of distinguishable regions of dark and light, and similar encodings that vary the quantity of pixels may also be similarly effective.

6 Conclusion

In this study we analyze how visual encodings impact the effectiveness of large, high-resolution visualizations. Through analysis of the task performance metrics and the user movements and behaviors based on the encoding used, we found that the choice of visual encoding directly affects the physical navigation of participants, which can result in significantly impact task performance. Color visualizations on the large display produced 96% accuracy and were more than twice as fast in performance, while the remaining encodings were less than 25% accurate and significantly slower. This highlights color's ability to visually aggregate well, as evidenced by the observational data. The users were able to identify both clusters and targets from an overview level. In contrast, position did not visually aggregate well, forcing users to resort to naïve search while close to the display.

To support physical navigation, encodings need to have a balance between the expressiveness of glyphs and good visual aggregation properties. Visual aggregation needs to be considered both at the glyph level in terms of the distinctiveness of each aggregated glyph, and at the field level in terms of the coherency of clusters and regions.

Large, high-resolution visualizations have the ability to not only show more information, but also afford the opportunity for users to interact with their data via natural, physical movement. Such an advantage must be accompanied by encodings that support this movement. Therefore, as designers chose encodings for large display visualizations, we urge them to consider the key characteristic of each encoding – the way in which it visually aggregates.

Acknowledgments

This research is supported by the National Geospatial-Intelligence Agency (NGA), contract HM1582-05-1-2001.

References

[1] Ball, R., North, C. and A. Bowman, D. Move to improve: promoting physical navigation to increase user performance with large displays. In *CHI 2007* (San Jose, California, USA, 2007). ACM.

[2] Bartram, L., Ware, C. and Calvert, T. Moticons: detection, distraction and task. *Int. J. Hum.-Comput. Stud.*, 58, 5 2003), 515-545.

[3] Bertin, J. *Semiology of graphics*. University of Wisconsin Press, Madison, Wis., 1983.

[4] Brewer, C. A. *Guidelines for use of the perceptual dimensions of color for mapping and visualization*. SPIE, 1994.

[5] Card, S. K., Mackinlay, J. D. and Shneiderman, B. *Readings in information visualization: using vision to think*. Morgan Kaufmann Publishers Inc., 1999.

[6] Chang, D., Dooley, L. and E. Tuovinen, J. Gestalt theory in visual screen design: a new look at an old subject. In *Proceedings of the Seventh world conference on computers in education conference on Computers in education: Australian topics - Volume 8* (Copenhagen, Denmark, 2002). Australian Computer Society, Inc.

[7] Chewar, C. M., McCrickard, D. S., Ndiwalana, A., North, C., Pryor, J. and Tessendorf, D. Secondary task display attributes: optimizing visualizations for cognitive task suitability and interference avoidance. In *VisSym 2002* (Barcelona, Spain, 2002). Eurographics Association.

[8] Cleveland, W. S. and McGill, R. *Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods*. American Statistical Association, 1984.

[9] Eick, S. G. and Karr, A. F. Visual Scalability. *Journal of Computational and Graphical Statistics*, 11, 1 2002, 22-43.

[10] Elmqvist, N. and Fekete, J.-D. Rolling the Dice: Multidimensional Visual Exploration using Scatterplot Matrix Navigation. *IEEE Transactions on Visualization and Computer Graphics*, 142008, 1141-1148.

[11] Healey, C. G., Booth, K. S. and Enns, J. T. High-speed visual estimation using preattentive processing. *ACM Trans. Comput.-Hum. Interact.*, 3, 2 1996, 107-135.

[12] Humphreys, G., Houston, M., Ng, R., Frank, R., Ahern, S., Kirchner, P. D. and Klosowski, J. T. Chromium: a stream-processing framework for interactive rendering on clusters. *ACM Trans. Graph.*, 21, 3 2002, 693-702.

[13] MacEachren, A. M. *How Maps Work*. New York : Guilford Press, 1995.

[14] Mackinlay, J. Automating the design of graphical presentations of relational information. *ACM Trans. Graph.*, 5, 2 1986, 110-141.

[15] Martz, P. *Generating random fractal terrain, http://gameprogrammer.com/fractal.html*.

[16] McCrickard, D. S., Catrambone, R. and Stasko, J. T. Evaluating Animation in the Periphery as a Mechanism for Maintaining Awareness. *INTERACT*, 2001, 148-156.

[17] Nothdurft, H. C. Texture segmentation and pop-out from orientation contrast. *Vision Research*, 31, 6 1991, 1073-1078.

[18] Ward, M. O. A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization*, 1, 3/4 2002, 194-210.

[19] Ware, C. *Information visualization : perception for design*. Morgan Kaufman, San Francisco, CA, 2000.

[20] Ware, C. and Beatty, J. C. Using color dimensions to display data dimensions. *Hum. Factors*, 30, 2, 1988, 127-142.

[21] Wigdor, D., Shen, C., Forlines, C. and Balakrishnan, R. Perception of elementary graphical elements in tabletop and multi-surface environments. In *CHI 2007* (San Jose, California, USA, 2007). ACM.

[22] Yost, B., Haciahmetoglu, Y. and North, C. Beyond visual acuity: the perceptual scalability of information visualizations for large displays. In *CHI 2007* (San Jose, California, USA, 2007). ACM.