Overview Use in Multiple Visual Information Resolution Interfaces

Heidi Lam, Tamara Munzner, *Member, IEEE*, and Robert Kincaid

Abstract—In interfaces that provide multiple visual information resolutions (VIR), low-VIR overviews typically sacrifice visual details for display capacity, with the assumption that users can select regions of interest to examine at higher VIRs. Designers can create low VIRs based on multi-level structure inherent in the data, but have little guidance with single-level data. To better guide design tradeoff between display capacity and visual target perceivability, we looked at overview use in two multiple-VIR interfaces with high-VIR displays either embedded within, or separate from, the overviews. We studied two visual requirements for effective overview and found that participants would reliably use the low-VIR overviews only when the visual targets were simple and had small visual spans. Otherwise, at least 20% chose to use the high-VIR view exclusively. Surprisingly, neither of the multiple-VIR interfaces provided performance benefits when compared to using the high-VIR view alone. However, we did observe benefits in providing side-by-side comparisons for target matching. We conjecture that the high cognitive load of multiple-VIR interface interactions, whether real or perceived, is a more considerable barrier to their effective use than was previously considered.

Index Terms—Multiple resolutions, overview use, user study.

✦

1 INTRODUCTION

Designers often face the screen-space challenge in which the amount of data to be displayed greatly exceeds output device capacity. One common information visualization approach is to provide multiple visual information resolutions (VIRs) in the interface, where an overview concentrates the data at a lower VIR to allow users to select a region of interest to explore in more detail at a higher VIR. When the data is structured at multiple levels that are relevant to the task, that structure can be used to create the lower-VIR views. When the dataset has only a single level of inherent structure, designers have little guidance on when multiple VIRs would be effective.

Generally, font readability characterizes the usefulness of displayed text. For graphical displays, the visual requirements are more difficult to define despite the rich history of perception research. In general, a visual object is *salient* when it attracts the user's attention more than its neighbours, and is therefore easily detected [8]. One way to achieve extreme visual salience is by visual pop-out, where visual objects with features that can be preattentively processed are spotted quickly and reliably on the display independent of the number of distractors and observer intent [17]. However, this extreme approach can be inappropriate when it is unclear *a priori* which of several aspects of the dataset should be emphasized. Instead, a more appropriate strategy would be to encode visual objects with sufficient salience to enable overview use without having one aspect overpower the others. The low-VIR view would contain a variety of items of similar salience, where the visual target does not draw more attention than the non-targets but can be serially searched. We investigated whether selective provision of high-VIR details, as in multiple-resolution interfaces, could relax perceptual requirements established for single low-VIR views.

We studied four interfaces: low VIR, high VIR, and two multiple-VIR interfaces where high and low VIRs were available in separate regions, or embedded together. Our study data were unordered collections of line graphs synthetically created for specific visual characteristics at low and high VIRs. At low VIRs, we used colour encoding for a heatmap effect; at high VIRs, we used height coding in conjunction with colour to show a more traditional plot. To better study interface preferences, our participants could use any combination of VIRs in the multiple-VIR interfaces.

Based on pilot study results, we selectively looked at two perceptual requirements: target visual complexity and visual span. We established the boundaries of these requirements by showing that our participants universally chose to use the low-VIR displays only when the visual targets were structurally simple and spanned a small visual angle. We then focused on situations where these visual requirements were not completely met. We found that in these cases, using our multiple-VIR interfaces did not result in better time and accuracy performance over the high-VIR interface, even when the multiple-VIR interfaces offered obvious benefits such as selectively embedding detailed plots in a complex-target matching task, or side-by-side display in a visual comparison task. In fact, we were intrigued to find that at least 20% of the participants chose to forego these benefits and devote the entire interface to the high-VIR display. We conjecture that our results reflect the high interaction costs of multiple-VIR interfaces, and the surprisingly stringent target visual requirements to enable effective overview use in multiple-VIR interfaces.

2 TERMINOLOGY

We use the term *visual information resolution* (VIR) as a measure of displayed visual information quality: displays with low VIR have comparatively less visual information than displays with high VIR. Much of the existing literature denotes these VIRs by their expected functions: for example, focus+context or overview+detail. In this paper, we name these VIRs based on their visual encodings; focus or detail can be thought of as a region of *high VIR*, while context or overview is of comparatively *low VIR*. In the case where the multiple VIRs are integrated and embedded spatially, we refer to those interfaces as *embedded* displays, for example, the focus+context visualizations. In cases where the different VIRs are displayed as separate views, we refer to these interfaces as *separate*, as in overview+detail displays. Since the different VIRs can occupy the entire window, or be integrated as part of a single window, we explicitly differentiate the two by using the term *view* to denote separate windows or panes, and the term *region* to denote an area within a view.

3 RELATED WORK

Human vision research has generally focused on visual search on a single VIR, while several usability studies have looked at the use of multiple VIRs using single-level data.

- *Heidi Lam is with University of British Columbia, E-mail: hllam@cs.ubc.ca.*
- *Tamara Munzner is with University of British Columbia, E-mail: tmm@cs.ubc.ca.*
- *Robert Kincaid is with Agilent Technologies, E-mail: robert_kincaid@agilent.com.*

Table 1. Instructions for the four study tasks

Domain Instruction	Visual Instruction
Max: Which location has the highest power surge for the time period shown on the screen?	Look for the brightest spot. You can mouse over and read the power off the tool-tip. Also notice the maximum power scale is shown above.
Most: Which location has the most number of power surges?	None needed.
Shape: A fault happened at location <x> at 6:00, causing a similar power surge in another location afterwards. Which one?	Look for a power surge of a similar shape as the one at location <x> at 6:00.
Compare: Find the power profile that is the same as that of location <x>.	All the profiles are exactly the same, except time-shifted by different amounts. The power surges of location <x> are in the middle of each column.

3.1 Human Vision Literature

A vast amount of human vision research has been done to measure and to understand visual salience. In terms of measurement, researchers have built predictive models to automatically evaluate interfaces. Using plain character displays, Tullis identified six display characteristics that correlated with visual search times [18]. Recent efforts are based on image processing and statistics models, for example, Rosenholtz *et al.*'s Feature Congestion model [13]. In terms of mechanisms, one well-studied area is preattentive vision (e.g., [11](p. 554-560), [17]). The information visualization community has incorporated much of this perceptual knowledge into its design guidelines for visual encoding [19](p. 151-156), as in our design of the low-VIR display. For displays that are visually cluttered, one proposed solution is attention filtering using colour and intensity coding to help users segregate their visual fields [21]. In our study we investigated the potential of using multiple VIRs as a different approach to address this problem.

Another line of inquiry in vision research is to characterize separable visual features that can be used to encode multi-dimensional data such that the users can selectively focus on a single dimension without being affected by other visually encoded dimensions. Examples include texture and colours [3] and motion [6]. In our low-VIR display, we encoded one dimension with color, and the second with space.

While the vision science literature offers valuable advice to designers in their choice of visual encoding, to the best of our knowledge it does not consider the interplay between different visual resolutions in multiple-VIR displays. For that, we turn to usability studies in the fields of human-computer interaction and information visualization.

3.2 Usability Studies

Although multiple-VIR interface study results are sometimes characterized as mixed, the situation becomes clearer when we categorize the studies. In cases where the task involved multiple levels of the data, study results generally show that multiple-VIR interfaces outperformed their high-VIR counterparts. Examples include Schaffer's network repair task where the answers involved links at all levels of the network [15], and Hornbaek and Frokjaer's essay-writing task where the participants were required to summarize the main points of an electronic document [5].

In cases where the dataset structure had only a single intrinsic level, multiple-VIR interfaces were found to be beneficial when the low-VIR display provided perceivable details required by the task. For text, perceivability is simply readability. The situation is well illustrated by Baudisch *et al.*'s 2004 study on information searches, which shows performance benefits for their multiple-VIR interfaces, but only for selective tasks [2]. Their study interfaces displayed web documents with guaranteed legible keywords, but surrounding text could be too small to read. When the task only required reading the keywords, as in their *Outdated* task, their multiple-VIR interfaces outperformed their high-VIR browser. In contrast, when the task required reading text around these keywords, as in the *Analysis* task, having an extra overview did not offer any performance benefits. Similarly, in North and Shneiderman's 2000 study, their *separate* interface had a low-VIR view displaying the names of geographic states that acted as hyperlinks for the coordinated high-VIR view, and using that interface provided time performance benefits for participants [10].

For non-textual graphic displays, geographic maps demonstrate the delicate balance between the need for concise yet perceivable displays in low-VIR views. A 2002 study by Hornbaek *et al.* found that a low-VIR overview resulted in slower performance times and worse recall accuracy for their Washington map trials, and their Montana map trials had generally poor performance results [4]. Their results suggest that the failure of the overviews was partly due to insufficient details provided to support their study tasks: the Montana map itself was single-level and did not offer enough meaningful map contents at low VIRs to guide region selections, and the Washington map display did not show enough details at the overview level to support their tasks.

Given the delicate balance between the need for concise yet perceivable displays in low-VIR views, our study attempts to shed further light on the perceptual requirements for visual targets to be reliably detectable when showing non-textual data at multiple VIRs.

4 USER STUDY DESIGN

We studied four interfaces: two single-VIR (*LoVIR*, *HiVIR*) as comparison baselines, and two multiple-VIR (*Embedded*, *Separate*). We had four visual search and compare tasks, and collected three types of data: performance measurements as time and error rates; detailed observations of participant behaviours and strategies; and participant feedback from subjective questionnaires.

4.1 Study Tasks

We piloted a diverse range of operations constructed from published taxonomies [1][14][16][20]. We used a scenario of monitoring and managing electric power in a control room to develop concrete examples of these abstract operations. We selected four of the original twelve pilot tasks that addressed different aspects of the criteria, including the need for comparison. Table 1 presents the task code names, and the domain and visual instructions provided for each task to control for individual differences in visual analytical skills.

Based on pilot results, we identified two target characteristics that affected high- and low-VIR view use: complexity and visual span. Complexity referred to the number of peaks in the target, where simple targets had a single peak and complex ones had multiple peaks. Targets were considered local when they span a limited horizonal display width, or dispersed when targets span the entire display width. In summary, the task characteristics were:

- *Max*: simple, local, no comparison
- *Most*: complex, dispersed, no comparison
- *Shape*: complex, local, comparison
- *Compare*: simple, local, comparison

4.2 Study Data

We developed tight criteria for data generation to control the visual qualities of the data populations. In addition to the targets, we created two distractor and five background populations to avoid target pop-out by colour or position, and to control task difficulty. Each peak was created using a Gaussian function with a specified mean that translates to peak location, and variability that translates to peak width. The peak was scaled to the required height. In addition, we added a random noise of up to two pixels in absolute value to better mimic real-life data [7]. Fig 1 shows the targets and distractors for the four tasks. The parameters used were determined based on pilot results.

For the *Max* task, the target peak was 10% higher, or 6% brighter on screen, than the distractor peaks, and at least 20% higher than the background peaks, as shown in Fig 1(b, c, d, e). For the *Most* task, the
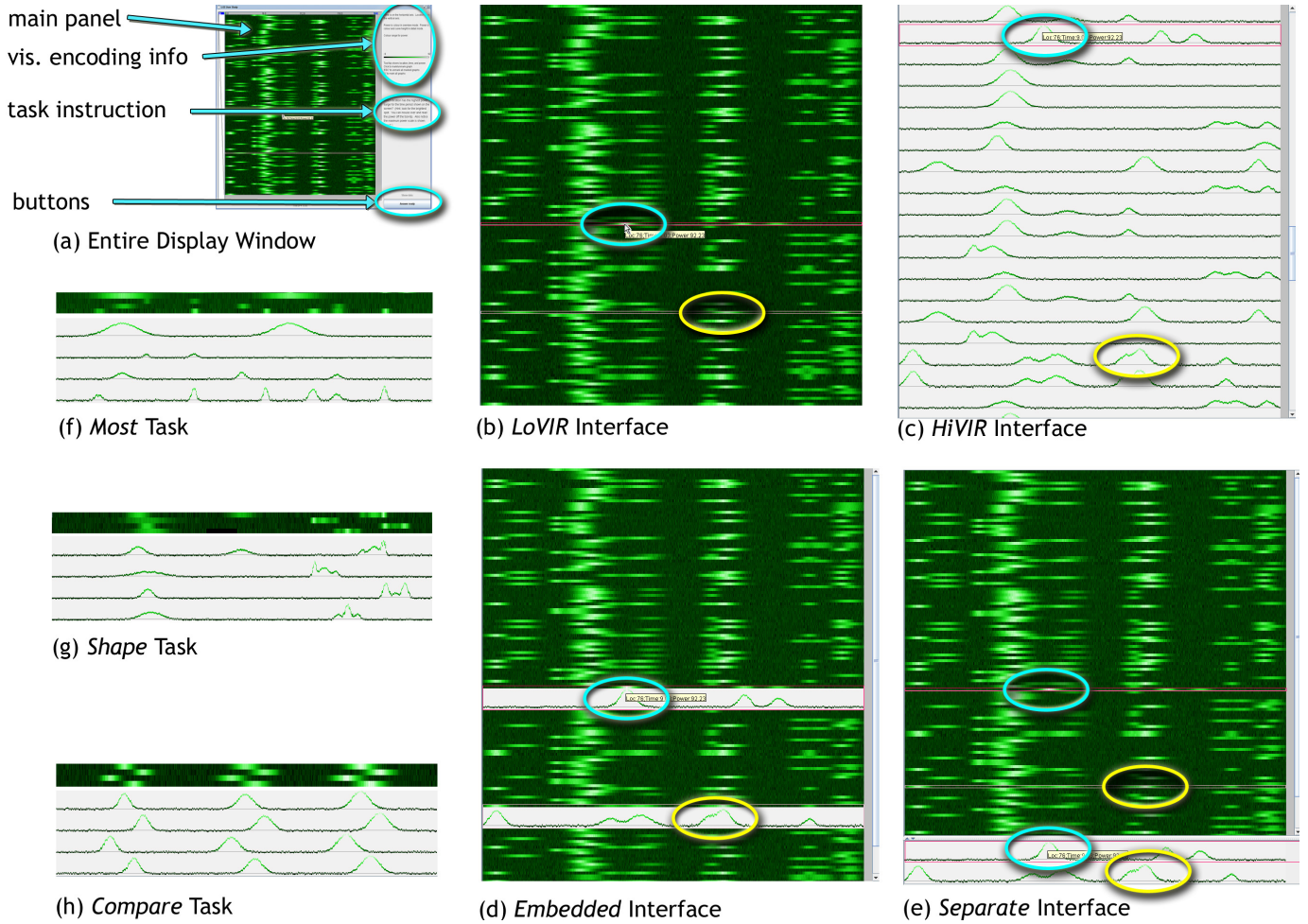
Fig. 1. Study interfaces and task data. **(a)** The full display window had a narrow region on the far left with strip/plot numbers, and then a main panel in the middle whose contents depended on the interface. The far right panel contained study instructions: on top, information on visual encoding and available interface interactions; beneath that, task instructions, as provided in Table 1; on the bottom, the `Show Data` and `Answer Ready` buttons. The main panel contents for each interface for the *Max* task: **(b)** *LoVIR*, **(c)** *HiVIR*, **(d)** *Embedded*, and **(e)** *Separate*. The target is circled in cyan, and one of the distractors is circled in yellow. We also show a closeup view of a few plots and strips for the other three tasks: **(f)** *Most*, **(g)** *Shape*, and **(h)** *Compare*.

target consisted of six random peaks of varying widths and heights, with the distractors having four peaks, and the background graphs having three peaks or less, as shown in Fig 1(f). For the *Shape* task, the target and distractors were peak clusters of three narrow peaks with similar widths and different heights out of four possible configurations, as shown in Fig 1(g). In the *Compare* task, the target was any of the peaks in a three-peak line graph. Both distractors and background contained the same peaks, but horizontally shifted by ±10, ±20, or ±30 pixels from the target, as shown in Fig 1(h).

For each task, we generated a collection of 140 line graphs, each with 800 data points, for a total of 112,000 data points. These numbers were determined by the horizontal and vertical resolution of the display area, so that the entire collection could be visible without scrolling in *LoVIR*.

4.3 Interfaces

We used two visual elements to show xy-data, inspired by the Line Graph Explorer system [7] that uses analogous but visibly different visual encodings for low- and high-VIR views. Both elements encoded the x-dimension in the same way, but their encodings of the y-data value differed: (1) Strip encoded the y-data with colour as a low-VIR strip of 6 pixels in height:

(2) Plot doubly encoded the y-data with both colour and vertical spatial position as a high-VIR plot of 45 pixels in height:

Colour encoding was achieved by mapping y-value to saturation and brightness in the HSB space. To maximize line-graph detail perceivability, we mapped the normalized y-value y to saturation s and brightness level b using a sigmoidal function:

$$s = \frac{2}{1 + e^{-4(1-y)}} - 1; \; b = \frac{2}{1 + e^{-4y}} - 1 \tag{1}$$

Using these two visual elements, we built the four interfaces shown in Fig 1: (b) *LoVIR*, (c) *HiVIR*, (d) *Embedded* and (e) *Separate*. The display area for all the interfaces was 872 x 880 pixels. *LoVIR* showed the data collection using only the strips, while the *HiVIR* interface displayed only the plots.

Both *Embedded* and *Separate* provided strips and plots, showing only strips initially. In *Embedded*, left clicking on a strip added or removed a corresponding plot directly below, with the pair bounded by a one-pixel perimeter box to visually reinforce the association.

In *Separate*, left clicking on a strip added or removed the corresponding plot in the bottom panel, and marked or unmarked both

the strip and the plot with separate perimeter boxes. The lower plot window automatically resized with newly added plots, up to half the screen height, after which a separate scrollbar provided navigation. Users could inactivate the automatic panel resizing by manually dragging the panel divider. Dragging the panel all the way to the top or the bottom of the screen allowed users to manually transform *Separate* to either *HiVIR* or *LoVIR*.

All interfaces had a panel on the far left to display the strip/plot numbers as text-strings for plots or as graphical bars for strips. Positions and states of the number displays were linked with those of the corresponding strips/plots.

4.3.1 Common interactions

For consistency, we standardized a number of interactions, adding only slight interface-specific adaptations.

- *Scrolling.* A scrollbar supported vertical scrolling when display height exceeded panel height. *LoVIR* never required scrolling while *HiVIR* always did. *Embedded* and *Separate* became scrollable once a plot was added. Both the top and the bottom panels were separately scrollable in *Separate*. None of the interfaces required horizontal scrolling.

- *Mouse-click marking.* A left click toggle-marked a strip/plot. In *LoVIR*, *Embedded*, and *Separate*, the mark was a one-pixel box surrounding the entire strip in the low-VIR view. In *HiVIR*, we marked by coloring the plot background, because perimeter marking was not salient in the visually noisy plots.

- *Key-press global action.* For the single-VIR interfaces, participants could mark all strips/plots with the O key, and unmark them with the Esc key. For the multiple-VIR interfaces, pressing the O key added all plots to the high-VIR view in *Separate*, or opened all plots within *Embedded*. Pressing Esc restored the initial low-VIR view.

- *Mouseover highlighting.* For all the interfaces, a red one-pixel box appeared around the strip/plot perimeter on mouseover to provide visual feedback of the strip/plot in focus. In *Separate*, the strip-plot pair was highlighted for visual linking. Fig 1(b) shows mouse-over highlighting.

- *Mouseover tool-tips.* For all the interfaces, mouseover triggered a tool-tip to immediately appear, displaying the x- and the y-value of the data point under the cursor and the strip/plot number of that row.

4.4 Participants

24 participants, 15 of them female, were recruited using an online reservation system. The average age of the participants was 26 years and ranged between 19 to 40 years. Most were university students, with less than half from the Department of Computer Science.

4.5 Material

The study was conducted on a desktop machine with a 3.2GHz Intel P4 CPU, 1.5 GB of RAM, and Java 1.5.0_06, using a 19-inch LCD display with 1280 x 1024 pixels.

4.6 Study Design and Protocol

The study was a within-subject, two-factor design with interface and task being the two factors, each with four levels. All four interfaces were tested against the four tasks, with a different, but isomorphic, dataset for each trial. The order of presentation of the interfaces was counter-balanced between participants. Task ordering was randomized, and data ordering was fixed to avoid repeats in interface/data pairing between participants.

The experiment consisted of four interface sessions, with one training and one actual task for each of the four interface/task combinations. The experimenter began by explaining the compact visual encoding used in the low-VIR views. Participants were then told about

the structure of the study. They were encouraged to try out interface features and to explore new strategies for the different interfaces during training, as strategies developed for one interface might not be appropriate for another. The entire display window is shown in Fig 1(a). For each task, participants first read the instructions in the right-hand panel of the study interface. When ready, they would press the `Show Data` button to display the data using the session interface. Once an answer was found, the participants pressed the `Answer Ready` button to enter the answer in a dialogue box.

For each interface/task combination, we allotted at least 10 minutes for the participants to complete each training task. At the end of the 10 minutes, they had the option to end the training and be told the answer, or to continue the task. On average, the participants took (3 ± 2) minutes to finish the training tasks, with similar averaged time over the four tasks. In terms of interfaces, the *Separate* training trials took four minutes on average, which was one to two minutes longer than the rest. Actual tasks had 5-minute time limits, after which the participants had to proceed to the next task without being informed of the correct answer. Breaks were allowed in between tasks, and there was a mandatory five-minute break after two interface sessions.

For each task, the experimenter observed participant mouse actions, verbal comments, and non-verbal signals including large-scale eye movement and signs of frustration. These observations were translated into textual narrations. For example: "Look for target in low-res. Press O to switch to high-res. Scan and scroll from top. Found answer, visual check without using tool-tip". We used these observations to help us interpret our performance time and accuracy results. We also developed a coding scheme for two kinds of usage behaviours:

- *Interface mode used to locate final answer.* The three categories were `LoVIR` mode, `HiVIR` mode, and `both`. The observation, only recorded for the two multiple-VIR interfaces, was later corroborated by the electronically recorded log of user actions.

- *Answer confirmation method.* The two categories were `visual comparison`, and `tool-tip/numeric confirmation`, differentiated based on back-and-forth tool-tip activations of the target and the candidate line graphs.

- *Visual search mode.* This observation was only collected for the *LoVIR* interface. The two categories were `serial search`, where the participant systematically inspected one strip at a time and in sequence, and `visual spotting`, where they surveyed the entire display simultaneously. Due to the narrow strips in *LoVIR*, serial search required the visual guide provided by the mouseover framing box, as shown in 1(b). For visual spotting, participants simply gazed at the display without any mouse interactions.

After the four interface sessions, the participants filled out two questionnaires. The entire study took about two hours, and the participants were compensated with CDN $20.

4.6.1 Study Hypotheses

We developed three study hypotheses based on pilot observations and existing beliefs about multiple-VIR interface use. H1 aimed to establish boundaries of our two selected perceptual requirements:

H1 The targets should be simple and span a limited region for a single low-VIR display to be usable.

We believed that the *LoVIR* interface would be the most efficient for the *Max* task, where the visual target satisfied both criteria; insufficient but usable for *Shape* task, where the target was complex; and would be unusable for the *Most* task, where all three criteria were violated.

In cases where the visual requirements were not completely satisfied, we hypothesized that selective display of high-VIR plots would mitigate the adverse effects of the lost perceivability, especially when the interface obviously supported the task. More specifically, our hypotheses were:

H2 When the targets were visually complex and could not be easily detected in the strips, embedded display of high VIR plots alongside the low-VIR strips would prime the search by promoting the learning of the unfamiliar and abstract strip.

In other words, the *Embedded* interface would better support the *Shape* task than the *HiVIR* or the *Separate* interfaces.

H3 When the targets were visually simple but similar to the distractors, precise identification of these targets using the low-VIR view would be difficult. However, users should still be able to select rough matches from the low-VIR view. The interface that displayed these potential matches in high VIR that allowed side-by-side comparisons would better support the task.

In other words, the *Separate* interface would better support the *Compare* task than the *HiVIR* or the *Embedded* interfaces.

4.7 Study Design Choices

Our goal of filling specific gaps in our understanding of multiple-VIR interface use led to eight main design choices.

1. *Synthetic data.* To create multiple isomorphic datasets with tight control over the visual characteristics of target, distractor and background graphs, we chose to generate synthetic data with real-world data characteristics.

2. *Unordered data.* While we used the visual encoding of the Line Graph Explorer system to build our interfaces [7], we specifically avoided providing its sorting or clustering capabilities for two reasons. First, we wanted to focus on visual search and comparison based solely on visual qualities of individual targets, instead of the larger context. Pilot results showed that when the line graph collections as a whole showed larger trends, for instance clusters, the display was treated as a whole and participants did not selectively view individual line graphs in detail. Second, the power of reordering and clustering is already well understood [7, 12].

3. *Task domain and visual instructions.* To control for individual differences in visual analytical skills between participants, we provided specific domain task instructions on control room monitoring and the visual operation on the encoded data. Our scenario provided a concrete unifying story, but did not require any specific expertise on the part of the participants.

4. *On-the-fly interface switching.* To observe our participants' interface choices as another indicator of interface effectiveness, we allowed our participants to switch to either VIR of the multiple-VIR interfaces at any point, even though we provided an automatic mechanism to allocate screen space between the two VIRs.

5. *Only two discrete VIRs.* Some previous multiple-VIR interface studies have found that distortion-based interaction across a continuous range of VIRs can decrease performance and satisfaction (e.g., [9]). In this study, we choose to focus on the issue of spatial arrangement of separating low-VIR regions from, versus embedding them within, high-VIR regions. We thus used only two discrete VIRs, as in systems like TableLens [12], to avoid conflating the question of spatial arrangement with that of distortion.

6. *Same platform and screen area across interfaces.* A common platform ensured consistent visual encoding, common interaction, and identical display areas.

7. *The full dataset is simultaneously visible from the low-VIR interface to be used as an overview.* Our dataset size was therefore limited to the display capability of the low-VIR view, which was 140 line graphs.

As a result the last three design choices, vertical scrolling was needed when users chose to display plots.

5 STUDY RESULTS

In this section, we present performance results for the actual tasks as time and error counts, coded observations, and subjective questionnaire results. We used the original interface grouping for all the results even when the participants switched to single-mode use in the multiple-VIR interface trials. In a separate analysis, we did not find significant differences between the single-mode use and the multiple-mode use populations in the multiple-VIR interface trials.

5.1 Performance time and error results

Performance time was defined as the period from which the participant pressed the `Show Data` button to the time when he pressed the `Answer Ready` button. We analyzed the time results using repeated measure two-factor Analysis of Variance (ANOVA) with interface and task as the two factors. When the sphericity assumption was violated, we used the Greenhouse-Geisser adjustment and marked the results as adjusted. Post-hoc analyses were performed with Bonferroni correction, and we report significant post-hoc results only.

Fig 2 shows the time results. A main effect was found in both interface ($F(3, 69) = 5.97$, $p = .001$), task ($F(3, 69) = 34.45$, $p < .0001$), and in interface-task interaction ($F(9, 207) = 11.20$, $p < .0001$, adjusted). For interface, post-hoc analysis indicated *LoVIR* trials were slower than *Embedded* or *Separate*. For task, all except the *Most* and the *Shape* results were different. For interface-task interaction, *HiVIR/Max* tasks were almost 3.5 times slower than the rest, *LoVIR/Most* almost 2 times slower, and *LoVIR/Shape* 1.7 times slower.

Error measures were binary for each task: 1 when the participant provided an incorrect answer, and 0 otherwise. We first analyzed the data using the Friedman test, and used the Mann-Whitney test with appropriate corrections for post-hoc analysis. We report significant results only. Fig 3 shows error results for each interface/task condition. Results showed that *LoVIR/Most* trials had 7 errors compared to the perfect scores of *HiVIR/Most* and *Embedded/Most*, and *LoVIR/Shape* trials had 6 errors compared to the perfect scores of *Embedded/Shape* and *Separate/Shape*. Along with the time results in Fig 2, we concluded that none of the interface/task results exhibited time-accuracy tradeoff: tasks that took longer also had more errors.

5.2 Observations

We quantified our observations by classifying each trial into one of the encoded categories. For multiple-VIR interfaces, we based our counts on the interface mode used at the time where the participants found the answers, and the count results are shown in Table 2. For all the interfaces, the methods for answer confirmation are summarized in Table 3. For the *LoVIR* interface, the visual search modes used to locate the visual targets are shown in Table 4.

5.3 Subjective preference and Questionnaire results

When asked to select the preferred interface overall, participants preferred both multiple-VIR interfaces over *LoVIR*, and *Separate* over *HiVIR* ($\chi^2(3, N = 14) = 15.00$, $p = .002$). None preferred *LoVIR*.

We also solicited two sets of subjective participant feedback with questionnaires. Results were first analyzed using the Friedman test, and the Mann-Whitney test was used for post-hoc analysis. The first questionnaire solicited subjective ratings of the four interfaces over the four tasks, as shown in Fig 4. To normalize the data, we divided the score for each interface by the sum of the scores for the task. Our results showed that *LoVIR* was preferred for the *Max* task, while *HiVIR* was thought to be most useful in the *Most* task. For the *Shape* and the *Compare* tasks, both *HiVIR* and *Separate* were preferred over *LoVIR*.

We also obtained feedback on the interfaces' ease of use with a 5-point rating scale. All except the `navigate` question produced significant results. As seen from Fig 5, *LoVIR* scored poorest in all the questions with significant findings, reflecting our participants' frustration with the interface. Only one question differentiated the other three interfaces: our participants found it easier to find data using *Separate* than *LoVIR* or *HiVIR*.
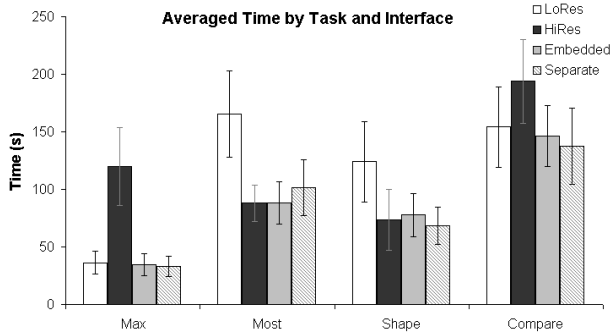
Fig. 2. Averaged time results by task and interface. The error bars show 95% confidence level.
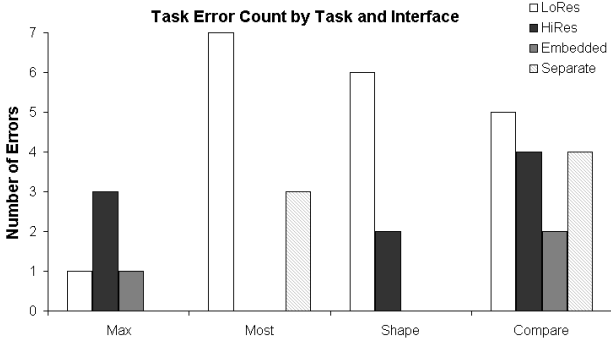

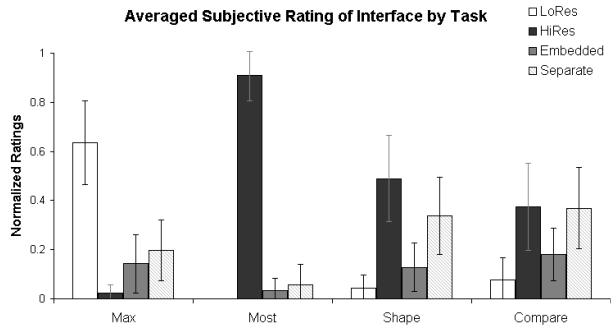Fig. 3. Total error results categorized by task and interface collected over 24 participants.

Fig. 4. Subjective ratings for the four interfaces for each task. Error bars show 95% confidence intervals.
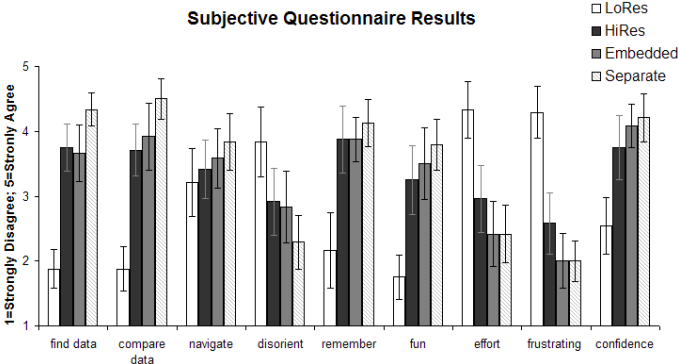

Fig. 5. Subjective questionnaire results for to solicit the ease of use for the four interfaces. Error bars are 95% confidence intervals.

Table 2. Coded behaviour: Interface mode use for the two multiple-VIR interfaces

Task	LoVIR	HiVIR	Both	LoVIR	HiVIR	Both
		Separate			*Embedded*	
Max	14	0	10	9	0	15
Most	0	23	1	0	21	3
Shape	0	13	11	0	14	10
Compare	1	4	19	0	7	17

Table 3. Coded behaviour: Answer confirmation mode: Vis = visual confirmation; Tip = Numeric read off tool-tips

Task	Vis	Tip	Vis	Tip	Vis	Tip	Vis	Tip
	LoVIR		*HiVIR*		*Separate*		*Embedded*	
Max	7	17	0	24	7	17	4	20
Most	24	0	24	0	24	0	24	0
Shape	18	6	19	5	22	2	19	5
Compare	3	21	2	22	13	11	2	22

Table 4. Coded behaviour: Visual search mode use for the *LoVIR* interface: Search = Serial Search; Spot = Visual Spotting

Task	Search	Spot	Task	Search	Spot
Max	2	22	*Most*	21	3
Shape	13	11	*Compare*	22	2

6 DISCUSSION

We investigated whether established perceptual requirements for low VIR could be relaxed in multiple-VIR interfaces when selective data are shown at high VIR. We first established the boundaries of the two perceptual requirements and showed that visual targets needed to be simple and span a limited visual angle to be reliably detected on the low-VIR overviews, thus confirming H1. Surprisingly, the merits of our multiple-VIR interfaces did not seem to relax these requirements based on participant interface choice and objective performances, thus we were unable to prove H2 or H3. We now discuss our three hypotheses in more detail, along with a more general discussion on multiple-VIR interface use.

H1: True. The low-VIR view alone is sufficient if the target is simple and spans a limited visual angle

For the visual complexity requirement, we compared the *Shape* to the *Max* task. The *Shape* task targets had three peaks, which were displayed as three bands with different colour intensities in the low-VIR view (Fig 1(g)). Since these tri-band targets were more visually complex than the single bands in the *Max* task, our participants could not easily find the targets in the low-VIR view. When forced to rely on the low-VIR view, as in the *LoVIR/Shape* tasks, we observed that 13 out of 24 participants resorted to serial search to locate the target. Even when the targets were found, some participants could not confirm their answers visually and needed to crosscheck the y-values using the tool-tips. Not surprisingly, our participants made more errors, took longer and assigned *LoVIR* the lowest subjective rating for the task.

In contrast, the *LoVIR* interface was effective for the *Max* task, where the majority of our participants (22 out of 24) could find the targets without resorting to serial search. Indeed, 63% of the participants considered the overview mode to be sufficient and preferred *LoVIR* for this task. On occasions where the plots were also available, participants only used them to confirm their answers. In short, the low-VIR view is extremely effective for the *Max* task.

The difference in results were large and surprising given the small difference in the two sets of visual targets. We believed that even though the three-peak targets in the *Shape* task were distinctive, the complex structure may be too difficult for the participants to process in the low-VIR view. Nonetheless, half of our participants still pre-

ferred to use both VIRs for target search in the multiple-VIR trials even though more than half (27 out of 48 times) bypassed the initial low-VIR view and switched to the high-VIR view.

The visual span boundary was established using the *Most* task and the *Max* task. Our *Most* task results show the extreme difficulty in using the low-VIR strip when the target spans a wide horizontal region. Unlike the case of the *Shape* trials, even serial searching in the low-VIR became difficult for participants in the *Most* trials, which required counting the number of peaks in each line graph. Participants almost always used the high-VIR plots for the *Most* task in the multiple-VIR tasks (45 out of 48 times), *LoVIR/Most* tasks were on average 89% slower and more erroneous than the *HiVIR/Most* tasks, and 91% preferred *HiVIR* for the task. This lopsided preference is worth noting since the participants were performing the same serial-search operation using either the strips or the plots. One reason may be that counting peaks is arguably a detailed visual task that requires foveal vision. In order to count the peaks, the participants would need to move their eyes along the strip to focus on individual peaks. Despite the visual aid provided by the framing highlight box, participants could not successfully fix their gaze on the desired strip, and frequently misread values from strips above or below the intended target. As a result, the low-VIR view was effectively unusable for the *Most* task due the dispersed nature of the targets. Our result is similar to Tullis' model indicating a positive correlation between target visual angle and search time [18], and we believe that the situation would be similar for long vertical targets.

H2: False. Embedding high-VIR plots in low-VIR strips did *not* enhance complex-target matching

We have established that the complex targets used in the *Shape* task could not be easily found by most of our participants, but they could still be serially searched. We now discuss if placing the more comprehensive and familiar plot alongside the unfamiliar and abstract tri-band strip would facilitate learning and prime visual search, enough to at least allow the participants to narrow their search space by selecting a set of potential candidate line graphs for detailed examination.

We concluded that H2 was not supported since we did not detect any performance differences between the *Embedded*, *Separate* and *HiVIR* trials for the *Shape* task. The close proximity of the strip and the plot provided by the *Embedded* interface did not seem to have been sufficient for our participants to learn the less familiar colour strips to prime the visual search. In fact, having the extra overviews did not seem to provide any performance benefits, regardless of the spatial arrangement of the two VIRs.

Our observations helped to interpret our performance results: half of our participants in the multiple-VIR interface trials switched to the high-VIR mode to complete the tasks, suggesting that selectively providing high-VIR plots did not seem to provide enough detail for the task. To us, the switch was perplexing, as the participants would have to memorize the visual target and scan six full screens to perform the task. This strategy turned out to be difficult for at least two of them, as they missed the targets in their first scan, and had to rescan the entire six screens to find the targets.

One possible explanation of the switching may be the visually different encodings of the VIRs. We attempted to minimize the effect of the difference by visually linking the two encodings with smooth animation in the *Embedded* interface, and also by instructing our participants with sample line graphs shown at the beginning of the training sessions. During the design of the visual encodings, we experimented with filling in the area beneath the line graphs, but that reduced the perceivability of the high-VIR encoding. Even though we cannot discount the different encodings as a factor that hindered the use of the *Embedded* interface without further investigation, we believed our participants' choice of switching to the high-VIR display was based more on the interaction costs of the multiple-VIR interfaces. We will further discuss the participants' choice at the end of the discussion, since it is not isolated to the *Shape* task.

H3: False. Providing side-by-side visual comparison with selective detailed plots did *not* enhance simple but similar target matching

Our last hypothesis studied whether providing obvious support for the task would relax the target perceptual requirements. We base our discussion on the results of the *Compare* task, where our participants were required to match the simple single-peak target that only differed from the distractors by a small horizontal shift.

Our results did not support H3, as we found that the participants were equally slow and error prone for all four interfaces in this task. In other words, we did not detect performance benefits provided by the extra overview of both multiple-VIR interfaces, or even by the side-by-side comparison capability of the *Separate* interface.

Our observations provided insights to the performance results: our participants derived a successful strategy to work with the single-VIR interfaces. In the *HiVIR* interface, for example, participants took advantage of the mouse wheel and scrolled vertically up and down with the cursor fixed horizontally at the horizontal point where the target peaked. As a result, all they needed to do was to find another peak at the same horizontal point numerically by reading off the tool-tips, thus avoiding the need to directly and visually compare the plots themselves. A similar strategy was used in the *LoVIR/Compare* trials. Instead of using the mouse wheel, which was not available as the interface was not scrollable, participants tried to keep the horizontal position of the mouse constant while mousing vertically up and down.

Due to the success of the strategy, a few participants voluntarily switched to the high-VIR plots for the *Compare* task. We saw this happen in 4 out of 24 cases for the *Separate* trials, and 7 out of 24 cases for the *Embedded* trials. One participant in the *Separate/Compare* trial even chose to use the low-VIR view exclusively for the task, a surprising choice given our H1 findings.

We did observe evidence to suggest the merits of using the *Separate* interface in its intended form. For the 20 participants that used both the strips and the plots for the task, 13 found the side-by-side comparison sufficient and did not crosscheck between the originals and the target peaks by reading off the numeric values from the tool-tips. In contrast, with the *Embedded* interface, only two participants relied on visual crosschecking without tool-tips. Perhaps that is why our participants preferred the *Separate* interface for the *Compare* task along with the *HiVIR* interface, even though we could not detect any performance benefits.

Interaction complexity and spatial arrangements

Our results suggest a surprisingly conservative set of visual requirements for the overviews to be usable in a multiple-VIR setting: our participants reliably used the low-VIR displays only when the target was simple and spanned a narrow visual angle. Any deviation from that composition, as in a three-peak target, severely reduced the usefulness of the low-VIR displays.

Given the fragility of the low-VIR interface, it is therefore of great interest to see if selective displays of high-VIR details could compensate for some of the lost perceivability, or offer enough benefits to tolerate the loss. In short, will the participants take advantage of the low-VIR interfaces to reduce the search space by first selecting a handful of potential candidates for further examination in detail? Even for tasks that seemed to be suitable for the multiple-VIR interfaces, at least 20% of our participants preferred the high-VIR displays, as indicated by their interface choice. Granted, using the high-VIR view alone may be impossible rather than simply difficult if a dramatically larger amount of data were used, for example, millions of points rather than the 112,000 used in our study. We did, however, observe considerable difficulties in using the high-VIR displays, and the need for seemingly elaborate strategies to enable their use.

We believe this interesting choice was due to the cost of interface interaction complexity, which may also explain the lack of performance benefits over the optimal single-VIR interface for each task. Although seemingly tedious and laborious, using the high-VIR plots has a low

cognitive load: the only navigation available is scrolling, a relatively passive exercise, and the answer will usually be apparent sooner or later. In contrast, navigation in a multiple-VIR interface is complex, as it involves active selection of potential target candidates, an action that requires mental and visual concentration. It also requires the physical effort in clicking on a relatively narrow strip and potentially the physical effort of scrolling. In our case, having only two VIRs instead of a zone of higher VIRs may have made the target more difficult to select in low VIR, but we believe the rest of the costs would still remain.

Switching from a multiple-VIR mode to a single VIR can thus provide an easily perceived short-term benefit of lower cognitive load, despite potentially increasing the total time required to complete the task. Our study training for the users required them to demonstrate proficiency in the use of all four interfaces, as is usual in single-session laboratory settings. We conjecture that users trained to demonstrate proficiency in a multiple-VIR interface may still not have internalized confidence in its use: that is, may not have adequately understood the longer-term cost of these short-term choices.

For the spatial arrangement of the VIRs, we did not detect differences in participant performances between the embedded or separate ones. The differing costs of the multiple-VIR interfaces may explain the lack of demonstrated differences. For example, embedding plots within the stacked strips can potentially disrupt the overview effectiveness of the low-VIR view, as indicated some of our participants' quick successive opening and closing of the same plot in the *Embedded* trials: open to see plot details, and close to better see the overview for visual search. As for the *Separate* interface, we observed the well-known problem of associating between separate views, where the participants closed and reopened the plots in the high-VIR view, one at a time, to re-associate them to the strips.

Despite these costs, we did observe benefits in providing side-by-side comparisons in visual comparison tasks. When using the *Separate* interface, more participants relied on visual crosschecking without tool-tip activations to confirm their answers than in the *Embedded* trials. The merits of providing side-by-side visual comparison may explain the subjective preference results. The *Separate*, along with the *HiVIR* interface, were the preferred interfaces for the two visual comparison tasks, and overall, our participants found it easier to compare between line graphs when using the *Separate* interface than either single-mode interfaces.

7 CONCLUSION AND FUTURE WORK

Using a set of contrasting visual targets displayed on two VIRs, we started by establishing the boundaries of two visual qualities for an effective low-VIR view, namely that the target should be simple and span a limited visual angle. When either of these boundaries were crossed, multiple-VIR interfaces did not enhance visual search over using a single high-VIR view, even though the multiple-VIR interfaces provided obvious benefits for the tasks, for example, side-by-side comparison. We believe our results reflect the high cognitive costs of interaction with multiple-VIR interfaces.

This work is a first attempt to look at the interplay between high and low VIRs based on overview target perceivability. Obviously, a more systematic study of other established perceptual requirements, such as item density and grouping [18], is required to draw more precise conclusions. In addition, eye-tracking, instead of note-taking, would allow more precise and objective measurements of interface use. It would also be interesting to separately test participants' visual abilities, which might shed more light on interface choice.

Despite our efforts to ensure diversity and generalizability, our work was naturally limited by our visual encoding, interface, and study design choices. Nonetheless, as an initial step, our work indicates the delicate balance of visual perceivability and interaction requirements in a specific use of low-VIR overviews to facilitate visual search in a large information space. It would be interesting to investigate other potential uses of overviews, such as providing global context or affecting visual search, and how these effects interact.

REFERENCES

[1] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *Proc. IEEE Symposium on Information Visualization (InfoVis'05)*, pages 111–117, 2005.

[2] P. Baudisch, B. Lee, and L. Hanna. Fishnet, a fisheye web browser with search term popouts: a comparative evaluation with overview and linear view. In *Proc. ACM Advanced Visual Interface (AVI 2004)*, pages 133–140, 2004.

[3] C. G. Healey and J. T. Enns. Large datasets at a glance: Combining textures and colors in scientific visualization. *IEEE Trans. on Visualization and Computer Graphics (TVCG)*, 5(2):145–167, 1999.

[4] K. Hornbaek, B. Bederson, and C. Plaisant. Navigation patterns and usability of zoomable user interfaces with and without an overview. *ACM Trans. on Computer-Human Interaction (ToCHI)*, 9(4):362–389, 2002.

[5] K. Hornbaek, E. Frokjaer, and C. Plaisant. Reading patterns and usability in visualization of electronic documents. *ACM Trans. on Computer-Human Interaction (ToCHI)*, 10(2):119–149, 2003.

[6] D. E. Huber and C. G. Healey. Visualizing data with motion. In *Proc. IEEE Visualization 2005*, pages 527–534, 2005.

[7] R. Kincaid and H. Lam. Line Graph Explorer: Scalable display of line graphs using focus+context. In *Proc. ACM Advanced Visual Interface (AVI 2006)*, pages 404–411, 2006.

[8] F. Landragin, N. Bellalem, and L. Romary. Visual salience and perceptual grouping in multimodal interactivity. In *Proc. International Workshop on Information Presentation and Natural Mulimodal Dialogue (IPNMD)*, pages 151–155, 2001.

[9] D. Nekrasovski, D. Bodnar, J. McGrenere, T. Munzner, and F. Guimbretire. An evaluation of pan and zoom and rubber sheet navigation. In *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'06)*, pages 11–20, 2006.

[10] C. North and B. Shneiderman. Snap-together visualization: can users construct and operate coordinated visualizations. *Journal of Human-Computer Studies*, 53(5):715–739, 2000.

[11] S. E. Palmer. *Vision Science*. Bradford, Cambridge MA, 1999.

[12] R. Rao and S. Card. Table lens: Merging graphical and symbolic representations in an interactive focus plus context visualization for tabular information. In *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'94)*, pages 318–322, 1994.

[13] R. Rosenholtz, Y. Li, J. Mansfield, and Z. Jin. Feature congestion: A measure of display clutter. In *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'05)*, pages 761–770, 2005.

[14] S. F. Roth and J. Mattis. Data characterization for intelligent graphics presentation. In *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'90)*, pages 193–200, 1990.

[15] D. Schaffer, Z. Zuo, S. Greenberg, L. Bartram, J. Dill, S. Dubs, and M. Roseman. Navigating hierarchically clustered networks through fisheye and full-zoom methods. *ACM Trans. on Computer-Human Interaction (ToCHI)*, 3(2):162–188, Mar. 1996.

[16] M. Tory and T. Möller. Rethinking visualization: A high-level taxonomy. In *Proc. IEEE Symposium on Information Visualization (InfoVis'04)*, pages 151–158, 2004.

[17] A. Treisman. Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, 31(2):156–177, 1985.

[18] T. S. Tullis. A computer-based tool for evaluating alphanumeric displays. In B. Shackel, editor, *Human-Computer Interaction: INTERACT '84*, pages 719–723. Elsevier Science, 1985.

[19] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann, 2nd edition, 2004.

[20] W. Wehrend and C. Lewis. A problem-oriented classification of visualization techniques. In *Proc. IEEE Visualization (Vis'90)*, pages 139–143, 1990.

[21] M. Yeh and C. D. Wickens. Attentional filtering in the design of electronic map displays: A comparison of color-coding, intensity coding, and decluttering techniques. *The Journal of the Human Factors and Ergonomics Society*, 43(4):543–562, 2001.