Subjective Quantification of Perceptual Interactions among some 2D Scientific Visualization Methods

Daniel Acevedo and David H. Laidlaw

Abstract— We present an evaluation of a parameterized set of 2D icon-based visualization methods where we quantified how perceptual interactions among visual elements affect effective data exploration. During the experiment, subjects quantified three different design factors for each method: the spatial resolution it could represent, the number of data values it could display at each point, and the degree to which it is visually linear. The class of visualization methods includes Poisson-disk distributed icons where icon size, icon spacing, and icon brightness can be set to a constant or coupled to data values from a 2D scalar field. By only coupling one of those visual components to data, we measured filtering interference for all three design factors. Filtering interference characterizes how different levels of the constant visual elements affect the evaluation of the data-coupled element. Our novel experimental methodology allowed us to generalize this perceptual information, gathered using ad-hoc artificial datasets, onto quantitative rules for visualizing real scientific datasets. This work also provides a framework for evaluating visualizations of multi-valued data that incorporate additional visual cues, such as icon orientation or color.

Index Terms—Perception models, 2D visualization methods, visualization evaluation, perceptual interactions, visual design.

1 Introduction

Modeling the space of possible visualization methods for a given scientific problem has challenged computer scientists, statisticians, geographers, and cognitive scientists for many years; it is still an open challenge. The goal of such models is to describe a searchable space where scientists can find visualization methods that optimally convey the information they require. Our approach to achieving this is to optimize the design of the images by studying how the components of a visualization method affect each other to facilitate, or complicate, data perception and comprehension.

The basic scientific visualization process involves symbolization, the translation of verbal and numerical information into graphic form [28], and comprehension, the analysis and understanding of the data presented. Our research is oriented towards developing exploratory data visualization methods, with the goal of prompting visual thinking and knowledge construction by presenting unknown and often raw data [26]. Understanding and insight are the main goals of scientific data visualization methods, but methods to represent known phenomena (e.g. turbulence in air flow or stress points in a structure) or geared towards performing specific tasks (e.g. finding extrema or identifying a type of turbulent flow) are qualitatively different from visualization methods designed for exploration of the data. Scientists usually utilize the latter during the early stages of their research, when they require visuals that provide a broad understanding of the data being presented. They begin posing hypotheses and asking questions about the data, which lead them towards task-oriented visualization methods for further analysis. Exploratory visualization methods allow them also, in a first approximation, to qualitatively assess the validity of their experimental and data gathering methods. At this stage, visualization is merely a tool to help scientists think about their problem [18].

We focus on visualization methods for multi-valued scientific datasets in 2D. These datasets are widely used in disciplines such as meteorology, geology, cartography, physics, and engineering. Even when scientists are studying three (or higher) dimensional phenomena, they often rely on 2D slices, such as cutting planes or isosurfaces, to explore and study the datasets.

• *Daniel Acevedo and David H. Laidlaw are with the Department of Computer Science at Brown University, E-mail: {daf,dhl}@cs.brown.edu.*

Manuscript received 1 March 2006; accepted 1 August 2006; posted online 6 November 2006.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

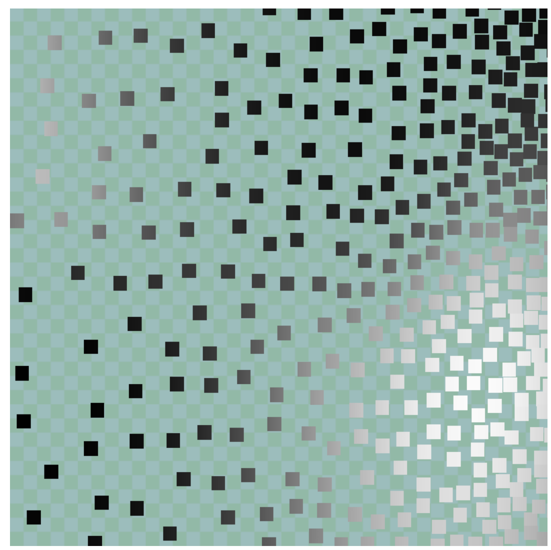

Fig. 1. An example of interference between icon spacing (representing a linear variable) and icon brightness (representing a more general scalar field). Areas of high brightness create false lower-spacing regions.

Common practice in scientific visualization is the mapping of scalar quantities to the visual qualities of surfaces containing the data, with color being the predominant example. Other visual qualities that can be used to represent a scalar field on a 2D surface belong not to the surface itself, but to glyphs or icons that can be placed on the surface. Color is again the initial choice for most applications, but size, distribution, and orientation of these icons can also be used to visually represent a scalar field. We call those visual characteristics our *visual elements*, since they will be the basic components of our study. Icons also have the advantage that they can be layered, increasing the number of variables being simultaneously shown.

A key issue here is how that layering property is used. Each of the visual elements can be mapped to represent a scalar variable but combining them in a single display will create visual artifacts. These perceptual conflicts can distract from correctly reading the information presented. See Figure 1 for an example of this.

The main novelty of our approach is the quantification and modeling of how the different visual elements interact with each other. This interaction can be explored at many levels [8] but the present study is limited to filtering interference among the various elements. This type of interaction is based on the visual elements being mapped to data one at a time, while the rest remain constant across the visualization. We chose to limit our experiment to three visual elements: icon brightness, icon size and icon spacing. This decision was made to decrease issues due to subject fatigue during the experiment. Size and density are elements that received a highly varied set of reactions during our previous experiments [1]. Also, very few studies have been published exploring these two elements together [36]. Icon brightness was chosen because it is an element that has been studied in depth, allowing us to compare our results with previous experiments. We realize that these choices greatly constrain our otherwise exponentially large exploration space but, with just three elements involved, we are able to generate an extensive set of examples for our experimental subjects to evaluate. We will explain in detail our experimental methodology in Section 4.

For our experiment we established a set of design factors that characterize the capabilities of a visualization method in displaying scientific data for exploration. These factors include constraints implied by the dataset, such as the number of levels of a data variable present and its minimum spatial feature size. A third factor indicates how visually linear the mapping between data and visual element is across the image. We will describe these in more detail in Section 3. To obtain numerical characterizations of our factors we created a framework for evaluating visualization methods through indirect perceptual tasks, making the experiment easier on the subjects but still powerful and generalizable from our perspective.

Evaluating the effectiveness of visualization methods is difficult because tests to evaluate them meaningfully are hard to design and execute [22]. We have researched this issue previously in several user studies comparing 2D vector visualization methods. The first one [24] used scientists to evaluate 6 visualization methods, and the second one [19] studied the validity of subjective measures to evaluate the same methods using designers as subjects. Results indicated that the designers rated the visualization methods in a pattern similar to the results of the scientists.

Following these examples we conducted an initial experiment to characterize the effectiveness of 2D scientific visualization methods using visual design experts as our subjects [1]. In that study we utilized a superset of the design factors we used in the current experiment, trying to engage our expert subjects into the evaluation of scientific datasets. The difficulty of the tasks required, the high variance of the responses obtained, and the small subset of visual element combinations tested made our results difficult to generalize. The current experiment improves the tasks by making them more accessible to non-expert subjects, lowering the variance between subjects.

In the following sections we will first put our research effort in context with the state of the art in visualization synthesis. Section 3 will provide a detailed description of the different elements we have developed to generate and evaluate visualization methods. In that same section we will introduce the notation we will use in the remainder of the paper.

In Section 4 we will describe the components of the current study, followed by a description of its results in Section 5 and a discussion on how they compare to established visual design criteria. Finally we will conclude by introducing our ideas for follow-up studies and future research, along with the main lessons learned from the current study.

2 Previous Work

Visual designers and artists are trained on how to communicate messages visually. In our case the message is a scientific dataset. We have previously researched, and continue to pursue, the idea of using artistic techniques for scientific visualization [23], [21]. Our experience in this area, and our ongoing collaboration with the Rhode Island School of Design, helped us select the set of visual elements that form the *means* by which we communicate our message.

Wallschlaeger and Busic-Snyder [34] provide a very comprehensive classification of the different elements involved in the communication process. Although they provide a very clear description of each element (color, shape, texture, etc.), they fail to formalize the interaction among them and the issues arising from their simultaneous use, a key component in our research.

Outside of the academic literature for art and design, one of the first and most cited works in the classification and analysis of visual elements for data representation is Bertin's *Semiology of Graphics* [3]. Our approach is very similar to his in that we are trying to characterize the capabilities of each of our visual elements individually, and then build up a model of how they perform in combination. He acknowledges that any combination of visual elements is possible but he dedicates very few pages to formalizing the use of their combinations. Our studies are designed to gather knowledge and provide a basis for a formal model for the effective combination of visual elements. Our work also presents an opportunity to address a main criticism of Bertin's work, that he lacks experimental results for his factual presentation of visual properties, by providing quantifiable evidence of his theories.

Many researchers have followed and applied Bertin's work: Cleveland and McGill [11], Mackinlay [27], Casner [9], Robertson [31], Lange et al. [25], Card and Mackinlay [7], Nowell [30], Chi [10], Nagappan [29], Bokinsky [4] and Tory and Moller [33] are the most extensive works trying to create a model for visualization synthesis. Most of them, however, deal with what is known as information visualization methods. Hanrahan [14] recognizes the artificial nature of the separation between information and scientific visualization, but acknowledges that most of the research aimed at the definition and characterization of a space of visualization methods has been done in the information visualization field. Our work is very much inspired and guided by the classification models developed for information visualization.

For spatially referenced data, MacEachren [26] presents an excellent summary of previous research in cartographic visualization. He expanded Bertin's visual variables to include crispness, resolution, transparency and arrangement. He also divided Bertin's color into hue and saturation for a total of 12 visual variables. Although his classification is better supported by experimental references from map makers and perceptual scientists, we miss some discussion about the specific use of each variable, both individually and in combination (combinations of hue, lightness and saturation are briefly presented). He provides clues towards the generation of rules for map-making but does not go as far as presenting such rules.

We have also studied the visual perception literature to support our investigations, and the Gestalt laws of perception [13] are one of the earliest attempts to qualify how the human visual system recognizes relationships among visual elements.

Ware [35] provides an excellent reference towards the understanding of all perceptual processes involved in information comprehension. Color, texture, form, and motion are the main elements discussed in his work, beginning from the physiological elements involved in perceiving each of those, up to a series of recommendations for their use in displaying abstract information. Ware takes a broad approach at information visualization and, although continuous data is discussed in the book, it is not its main focus. He provides a very good introduction to the theory of integral and separable dimensions for visual attributes, but provides little quantifiable evidence for his classification. Our study provides such evidence for the displaying of continuous scalar data.

We have found little experimental evidence about the perception of visual element combinations. Callaghan studied how hue and lightness interact in a texture segregation task [5]. She also compared, in pairs, hue with form and line orientation [6]. Although she reached valid conclusions about which variables dominate and when they interfere, her stimulae were limited to two levels of the visual variable being analyzed (e.g. horizontal and vertical for the oriented lines), while the potentially interfering variable was randomized or kept constant.

Our experiment is very much inspired by Carswell and Wickens's work [8] in which they classify different graphical attributes into integral, separable or configurable dimensions depending upon how each attribute's reading is affected by the others, taken pairwise. They found that visual elements can help each other when displaying the same information (redundancy or performance facilitation), or inhibit each other when only one element is changing (filtering interference). They also describe a third type called condensation in which opposite variation of each variable occurs simultaneously. Their experimental displays are based on single icons, looked at in isolation. We are extending their experiments to more complex displays and, for now, limiting our analysis to filtering interference analysis.

Our goal is to find the visual characteristics of different visual elements when displaying quantitative information, where visual contrast is the property that makes one data value different from the next. The measurement of texture contrast thresholds is common in texture segregation studies [2]. Those studies utilize stimulae with regions where the particular visual element differs in some amount with respect to the surrounding region. Many stimulae are required to explore the full range of a visual element, and even more to include interference analysis with secondary elements. Our displays are designed to evaluate, with less iterations, a larger portion of the range for each element.

Most of the literature about perceptually effective data representation is based on experience. Authors define sets of guidelines that, in the absence of visual perception theories [32], follow common practice and established knowledge [12]. In general, these approaches rely on a clear definition of the task a visualization must fulfill, making them difficult to apply in our research. Our exploratory visualization methods are geared towards presenting the data as clearly and unbiased as possible for scientists to explore.

Healey has studied extensively the application of preattentive processing to visualization [16]. Preattentive processing allows detection of visual elements in a display without focusing attention on them. Initially, he focused on experiments comparing hue and orientation [17]. Subjects in his experiments were asked to perform numerical estimation tasks with varying hue and orientation differences, as well as varying display time. He also proposed ViA, a visualization system based on perceptual knowledge [15]. The goals of this system are very similar to the ones in our research. He builds, by hand, the perceptual knowledge-base used to suggest a visualization method, while we are gathering that knowledge through subjective evaluations. In general, Healey's experiments come the closest to our evaluation approach.

Finally, Johnson [20], in his list of top scientific visualization problems, recognizes the quantification of the effectiveness of visualization methods as one of the major research areas in this field. He also included perceptual issues, multifield visualization and theory of visualization, all areas that we are addressing in our project.

3 Definitions

In this section, we will define the two main components of our study: our set of visualization methods and the design factors we defined to characterize them.

3.1 A Space of Visualization Methods $\mathbb{V}$

For the current study we will only work with icon brightness, icon size, and icon spacing, denoted as v_0, v_1, and v_2 respectively in the remainder of the paper. We utilize circular icons for all visualization methods shown in this experiment. Expert visual designers [1] form our first experiment suggested a circle as our test icon shape because if its neutrality. It avoids having preferred linear cues in a per-icon basis, leaving all cues to be obtained from the dataset and icon locations alone. A visualization method, $v \in \mathbb{V}$, takes a scientific dataset and produces a visualization display. The method corresponds to a layered combination of our visual elements where the different data variables being represented are mapped to one or more of the available elements. For the current experiment only one layer is created per visualization display and we are only looking for interactions among the three elements we chose. A visualization method is then expressed as follows:

$$v = \{(m_0, m_1, m_2), (r_0, r_1, r_2)\}$$

Each component of v refers to one of the three visual elements v_i:

		values			
		0.00	0.33	0.66	1.00
V_0	brightness	0.00	0.33	0.66	1.00
V_1	size (pixels)	2	5	7	10
V_2	spacing (pixels)	0	3	6	10

Table 1. Values used for each of our visual elements.

$$m_i = \begin{cases} 0 & v_i \text{ is not mapped} \\ d_i & v_i \text{ is mapped} \end{cases}$$

$$r_i = \begin{cases} c_i \in \mathbb{R} \in [0.0, 1.0] & m_i = 0 \\ (b_i, e_i) \in \mathbb{R}^2 \in ([0.0, 1.0], [0.0, 1.0]) & m_i \neq 0 \end{cases}$$

where d_i is the index of the data variable mapped into v_i. (b_i, e_i) indicates the range of linear mapping between v_i and d_i.

3.2 Design Factors

$\mathbb{W}(v) = (w_0(v), w_1(v), w_2(v))$ provides an evaluation of a visualization method, $v \in \mathbb{V}$. It produces a vector of values, each of which quantitatively characterizes the visualization method with respect to a specific design criterion.

The goal of our visualizations is exploratory: scientists need an accurate representation of their data but have no simple specific tasks in mind, other than exploring how the different variables interact. In this sense, the factors we define here provide information about the quality of the data presented and the capability of a visualization method to work in combination with other methods. Said factors are:

- *data resolution (w_0):* the number of different levels of a data variable that can be distinguished by a viewer;
- *spatial feature resolution (w_1):* the minimum spatial feature size that can be reliably represented with the method, expressed as a percentage of the image width;
- *visual linearity (w_2):* the degree to which subjects perceive the mapping from data value to visual property as linear; this factor is measured by asking subjects to indicate the locations where they see the values of 0, 0.25, 0.5, 0.75, and 1.0 along the image for a linear dataset visualization;

We derived these factors from our experience creating scientific visualizations for our collaborators in many disciplines and from our pilot study on designer-critiqued visualization methods [19]. Bertin [3] developed a similar classification of his *retinal properties* according to their level of organization (whether they could be used to represent quantitative, qualitative, or ordered information) and the number of steps they could take (our data resolution factor). Our factors introduce a new measure: visual linearity. Our data resolution and spatial feature resolution factors capture the fact that we are targeting quantitative datasets.

4 Methodology

Building on these visualization methods and design factors, we have developed an experimental approach for acquiring knowledge about our space of visualization methods.

4.1 Visualization Methods

Table 1 shows the values we chose for each of the three visual elements involved. Icon size and spacing are both measured in pixels. Size indicates the diameter of the circular icons, while spacing indicates the distance between two icons. As mentioned before, we utilize a Poisson-disk distribution to randomly place icons across the image. This distribution, as opposed to a regular grid, allows us to represent a

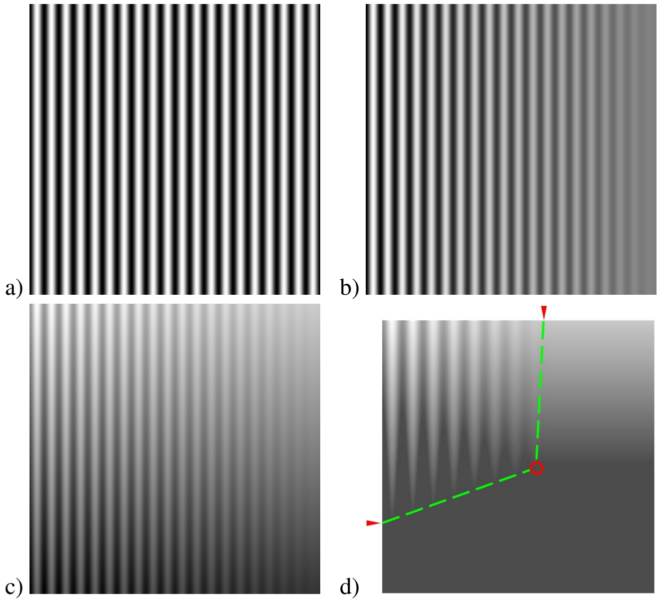

Fig. 2. Process for creating the stimuli for the data resolution identification task. (a) Shows a vertical sine-wave dataset. (b) Shows the same dataset with amplitude values a linearly decreasing from left to right. (c) Shows the final appearance of the datasets used for this task, where we also linearly move the zero value of the sine-wave from $a/2$ at the top of the image to $1-a/2$ at the bottom. (d) Shows how subjects would mark the area where they perceive the sine-wave pattern.

continuous scalar dataset with icon spacing by mapping the data variable to the distribution's disk size at each point sampled. We experimentally chose the upper limits for size and spacing so we could explore methods with reasonable spatial feature capabilities.

With these parameters we can define six possible value ranges, pairs (b_i, e_i),for each visual element: $(0.00, 0.33)$, $(0.00, 0.66)$, $(0.00, 1.00)$, $(0.33, 0.66)$, $(0.33, 1.00)$, and $(0.66, 1.00)$. For icon brightness methods we combine these six ranges with all possible combinations for the other two elements, creating a total of 72 visualization methods that we will evaluate. For icon size and spacing methods we keep icon brightness constant at 1.00, so 24 combinations (6*x*4) can be defined for each of those two elements. Note that even constraining our experiment to a small number of elements, and only four possible values per element, the number of combinations is quite large: 120 different visualization methods.

The experiment consists of three different tasks, one per design factor. For each task subjects are shown a set of images using different visualization methods to represent simple scalar datasets.

4.2 Data Resolution Identification Task

For this task we are asking subjects to evaluate how many different levels of the data variable a method is able to represent. We asked this question directly during a previous pilot study [1], but subjects had a high variance in their responses, prompting us to design a new approach to ask this question. We created the new stimuli as follows.

Using a vertical sine-wave pattern with constant wavelength λ across the image (Figure 2 (a)), we linearly decrease the amplitude a from left to right (Figure 2 (b)). While the amplitude values remain constant vertically across the image, we linearly move the zero value of the sine-wave from $a/2$ at the top to $1-a/2$ at the bottom. Figure 2 (c) shows the final appearance of such a dataset using grayscale.

The question subjects must answer in this task is: in what region of the image do you see the sine-wave pattern? Since they are told the pattern will be more pronounced at the top left corner of the image, they just need to place 3 marks to approximately bound the region where they perceive the pattern. Figure 2 (d) shows an example of this.

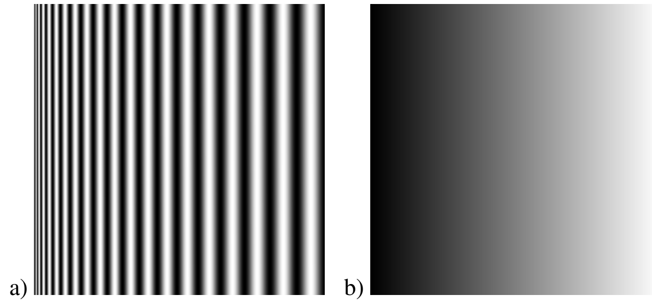

Fig. 3. (a) Shows an example stimulae for the spatial feature resolution identification task dataset, with wavelength λ linearly decreasing from right to left. (b) Shows the stimulae for the visual linearity perception task.

To create these datasets we have two extra variables to fix, the initial amplitude a and the wavelength λ. We tested several values for these variables and decided to evaluate eight different combinations using two amplitude values ($a = \{0.2, 0.6\}$) and four wavelength values ($\lambda = \{0.625, 1.25, 2.5, 5\}$ measured in percentage of the image width). To avoid multiplying by 8 the full set of 120 methods, we decided to use only combinations that utilize the full range of the data-mapped visual element, i.e. $b_i = 0.00$ and $e_i = 1.00$. During the analysis of the results we can still describe the data resolution capabilities for any subrange. Figure 4 (a), (b) and (c) show examples of images used for this task for each of the three visual elements.

Finally, to obtain actual data resolution values we developed the following process. The marks placed by a subject delimit a region on the image where the pattern is visible. The right and bottom boundaries indicate lines where the difference between the extremes values of the sine-wave are last perceived by the subject, i.e. the *just noticeable differences* (*jnd*) boundary. The basic idea to obtain data resolution values is to follow these boundary lines, starting from the top mark, jumping from one level to the next a distance equal to the amplitude at each point. With this process we will also obtain actual values, in the range $(0.00, 1.00)$, for each level identified.

Since there are two different initial amplitude values used, the results will overlap after a certain distance. The total data resolution of a visualization method will be given by the number of levels obtained for the $a = 0.2$ dataset, plus the number of levels for the $a = 0.6$ datasets with values greater than the maximum level obtained from the $a = 0.2$ dataset.

4.3 Spatial Feature Resolution Identification Task

For this task we are asking subjects to evaluate the size of the smallest spatial feature a method can represent. Again, during a previous pilot study [1] we asked this question directly. We used expert designers as subjects and we hoped their expertise would allow them to judge the capabilities of each method. Due to insufficient training for this task and inconsistent understanding of the concept of spatial features across subjects, our results suffered the high variance problems.

Our approach to fix this task was to indirectly ask the question by exploring the limits of each subject's visual perception. In this case our datasets are vertical sine-wave patterns that maintain constant amplitude a but linearly change their wavelength λ from left to right across the image. Figure 3 (a) shows an example of this dataset using brightness values from 0.0 to 1.0 ($a = 1.0$).

By asking subjects to place a mark when they stop perceiving the sine-wave pattern, we are obtaining our minimum feature size measurement. $\lambda/2$ at that point will be the minimum spatial feature a method can represent. For this task we use all 120 visualization methods mentioned before. The amplitude for each display is indicated by the range (b_i, e_i). Figure 4 (d), (e) and (f) show examples of images used for this task for each of the three visual elements.

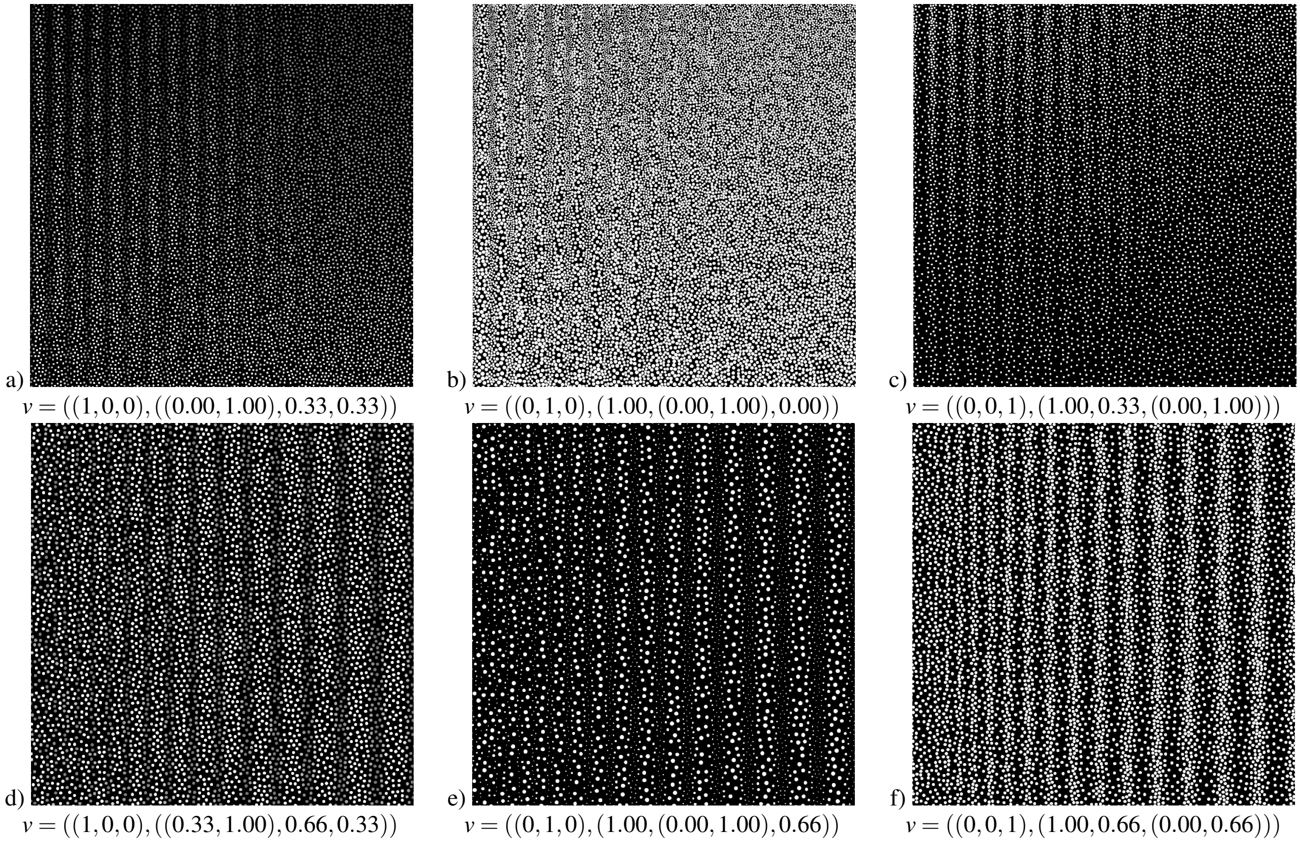

Fig. 4. Examples of various stimuli used for the experiment. (a), (b) and (c) show data resolution identification stimuli for icon brightness, size and spacing respectively. All of them with $\lambda = 5\%$ and $a = 0.6$. (d), (e) and (f) show spatial feature resolution identification stimuli for the same three visual elements. The actual method parameterizations are indicated below each image.

4.4 Visual Linearity Perception Task

In this task subjects are shown visualizations of a linear dataset that progresses from a value of 0 on the left of the image to a value of 1 on the right edge (see Figure 3 (b)). They are told that 0 and 1 are at the very edge of image and are asked to place five marks for the values $0.0, 0.25, 0.50, 0.75, 1.0$. The two extremes would indicate regions where they do not perceive a change in the visualization's border regions. A visually linear method would maintain a constant ratio between data changes and visualization changes.

4.5 Experimental Setup

We ran a full within-subjects pilot study where 6 computer science graduate students performed all three tasks. We randomized the order of the three tasks, the order of the visual elements being evaluated within each task and the order of the combinations for each element. The full study consisted of nine separate sections (3 tasks x 3 elements) with a training subsection and a real trial subsection within each one. Response time was recorded for the real trials. There was no time limit during any part of the study, although subjects were instructed to proceed as quickly and accurately as they could. Subjects took an average of an hour and forty minutes to complete the whole study and were paid for their participation. Subjects were given written instructions before each task. Stimuli for all tasks consisted of images of size 900x900 pixels displayed one by one on an LCD display.

During training, subjects were shown datasets where we artificially controlled the areas where patterns were present. For example, for the spatial feature resolution task, we reached $\lambda = 2$ pixels after only half the image width. Since our minimum icon size corresponds to 2 pixels, no visible pattern is possible for features of 1 pixel ($\lambda/2$). This allowed us to screen for subjects not understanding the task. All subjects performed correctly during training after only a few examples.

For all tasks subjects had the option of selecting a "No Pattern" button when they could not detect the sine-wave pattern at all or, in the case of the visual linearity task, when they could not see anything changing in the image.

After each section of real trials, subjects were asked to indicate their confidence level for the responses they gave and write down any comments they had about the task, the interface, or the visualization methods.

5 Results and Discussion

Data resolution results were plotted for each of the four λ values used for the data resolution task. Following the process indicated before, we calculated data resolution values for all 120 visualization methods. Spatial feature resolution and visual linearity results were also plotted for all visualization methods. These results successfully characterize the capabilities of each visual element using a variety of value ranges in combination with potentially interfering visual elements. We will discuss each design factor separately.

5.1 Spatial Feature Resolution

Figure 5 shows how spatial feature resolution values increase (the actual spatial feature size is measured in percentage of the image width) when icon size and icon spacing values grow larger for the same icon brightness range. The growth is faster with spacing than size. We also observed that, as expected, different ranges of brightness have different capabilities. All six brightness ranges show the same trend with respect to size and spacing as shown in Figure 5. The range $v_0 = (0.66, 1.0)$ showed the largest (poorest) feature size capabilities. In fact, most of our subjects declared not seeing a pattern at all in the displays corresponding to these particular methods.

The best results, with spatial feature resolutions of around 0.3% of the image width, were obtained for methods that used the smallest icon size (2 pixels) and the smallest spacing (0 pixels). This is reasonable

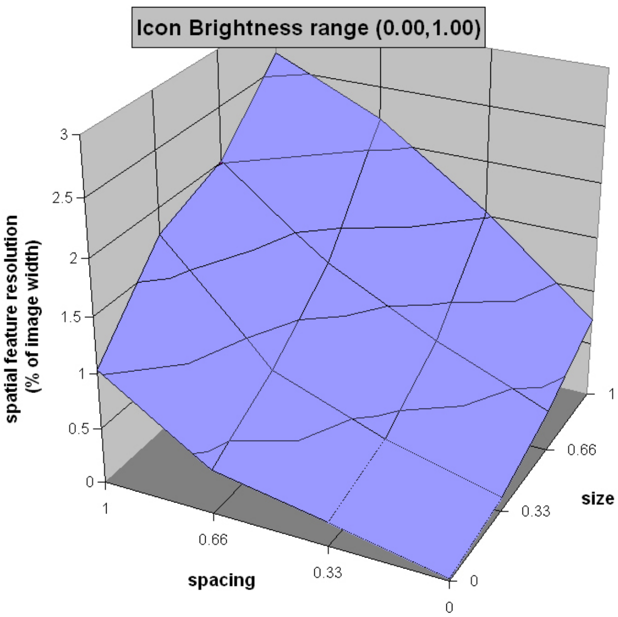

Fig. 5. Plot of the spatial feature resolution results for icon brightness methods ($v = ((1,0,0),((0.00,1.00),*,*))$). Note how spacing growth affects spatial feature resolution more than size growth.

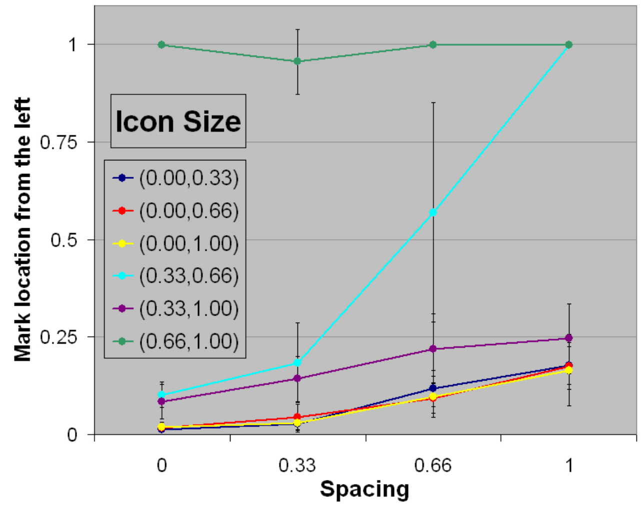

Fig. 6. Results for all 6 ranges of icon size with respect to the 4 different values of spacing for the spatial feature resolution task. The vertical axis shows the distance from the left border of the display where subjects placed the mark, the closer to zero the lower the spatial feature resolution is (see Figure 4(e)). All plots show 95% confidence intervals. Observe how, for the size range (0.66,1.00), almost all subjects reported a "No Pattern" result (plotted as 1 in the graph.) It can be observed that the larger the range the better resolution. For same size ranges, like (0.00,0.33), (0.33,0.66), and (0.66,1.00), the smaller the size the better results. Equivalent plots for spacing versus size show symmetric effects.

since these methods produce displays that are the closest to simple grayscale displays. Still, reasonable spatial feature resolution values, around 2.5% of the image width (around 22 pixels for the 900x900 images used in our experiment), were obtained for the extreme cases where maximum size and spacing were used. The 95% confidence intervals are very small across the board for all methods.

Interesting to note is the fact that ranges that included brightness values of zero showed, in general, smaller spatial feature resolution than the rest, despite the fact that the background for all images was also black, diminishing the contrast. The range of 0.3% to 2.5% is common for all brightness ranges that include black. After that, the more the lower end of the brightness range moves from black, the worse subjects performed.

More surprising are the spatial feature resolution values for size and spacing methods. Figure 6 shows the results for size. The overall trend is that they have symmetric interaction: larger icon size affects the reading of spacing values in the same way spacing affects the reading of icon size values. The unexpected result is that actual spatial feature resolution values are comparable to icon brightness methods, with the best results being around 0.3% of the image width for small spacing and small size of icons. The explanation for this unexpected good performance of size and spacing comes from the design of our experimental stimuli. Our sine-wave patterns for this task do not change vertically across the image. This produces very strong linear cues that induce subjects to continue perceiving the sine-wave pattern when, locally, there is no clear evidence of it. Real datasets do not usually exhibit this kind of linear structure, so our results would not be applicable in practice. A solution to this effect would be to run the experiment again using only narrow horizontal bands of our square displays.

5.2 Data Resolution

Figure 7 shows results for data resolution for the same full-range icon brightness visualization methods shown before. The trends are consistent with the results from spatial feature resolution. Maximum data resolution levels are around 21 levels in average, with 95% confidence intervals being ± 5. This result was obtained when size and spacing are at minimum values and $\lambda = 5\%$.

As mentioned before, for this task we presented subjects four different wavelength datasets. As expected, the number of recognizable levels of data grows larger as the wavelength increases. For example, for the same method mentioned before but with $\lambda = 2.5\%$, the data resolution is around 15 levels, whereas going all the way down to $\lambda = 0.625\%$ yields only around 9 recognizable levels. Note that in this particular case we are not yet at the spatial feature resolution limit for this particular method, which was around 0.3%.

The results for size and spacing are particularly interesting since, although they still follow the same trend as expected from the spatial feature resolution data, they are very different in absolute values (see Figure 8). Maximum data resolution values for icon size visualization methods are around 5 data levels, while maximum values for spacing methods get up to 9 levels. Given the problems with the stimuli for the spatial feature resolution task, we believe the better performance of spacing for this task reflects its real relationship with respect to size methods. Same as before, the resulting data resolution values clearly increase when the range of the visual element being used grows larger.x

5.3 Visual Linearity

All subjects reported difficulty completing this task. They easily placed the marks for the extreme values but they could not judge, in general, the 25% intermediate differences we were asking them to indicate, especially for icon brightness methods. Subjects also complained about possible inaccurate gamma calibration of the monitors used. We need to further explore this task and reimplement this portion of the experiment. It is still worth noting that practically all methods, for all three visual elements, exhibited clear constant-value areas for the extreme values, sometimes as large as 30% of the image width. This is consistent with our data resolution values where subjects indicate no *jnd*'s for those ranges.

Collecting accurate data about perceived visual linearity is important because our experimental tasks compound the effects of perceptual non-linearity with interference from the distractor visual parameters, which are constant in our case. To solve this, our visual linearity perception task is designed to provide us with the information to "fix" the parameterization. Given reliable perceived linearity solutions we could adjust the datasets for the other tasks to compensate for linearity, isolating the interference from the distractors.

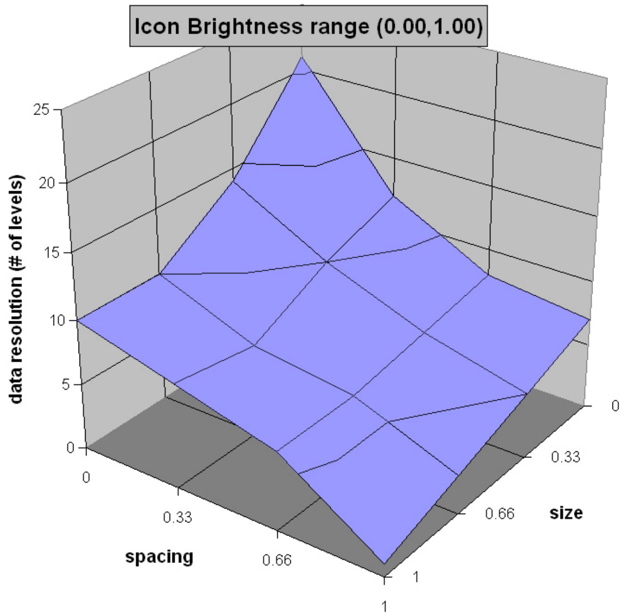

Fig. 7. Plot of the data resolution results for icon brightness methods ($v = ((1,0,0),((0.00,1.00),*,*))$). Consistent with the results for spatial feature resolution (the direction of the size and spacing axes is reversed with respect to Figure 5), spacing growth affects data resolution more than size growth. This plot corresponds to $\lambda = 5\%$, which gives the highest data resolution values.

5.4 Comparison with Previous Results

Bertin [3] provides one of the few examples of quantitative results for data resolution values. Although he does not explicitly measure them for icon brightness, he recognizes that the smaller the icons, the fewer levels our perception should be able to differentiate. Our results contradict that for all spacing levels (see Figure 7). For size, he proposes an average of 20 distinguishable levels when the ratio between the extremes of the range is 1 to 10. Our range is only 2 to 10 so our results should be expected to be smaller, but 5 levels is the maximum our subjects could differentiate. The more surprising result is that, for icon spacing, our subjects can differentiate a maximum of 9 levels, while Bertin does not expect more than 5.

Our spatial feature resolution results are hard to compare with existing psychophysical experiments, given our lack of control for values like distance from the screen or gamma correction. The values obtained, on the other hand, seem to follow our expectations. Our perceptual system contains specialized cells to detect brightness changes, whereas size and spacing changes seem to get processed differently. Our results validate this trend of better results for icon brightness. They also generate some surprising evidence for the perceptual ordering of icon size and spacing.

5.5 About the Experiment

Even though we dramatically reduced the number of combinations of visual elements we explore, the experiment posed a big design challenge. Subjects commented on its apparently extraordinary length due to the similarity of all the images. As we saw from the results, the actual values obtained establish clear differences. With these 6 subjects we were able to fully randomize the order of the three tasks to eliminate any possible learning effects. Nevertheless, fatigue was a big factor that, although it did not explicitly show up in the data collected, will require moving to shorter between-subjects design, or even multiple sessions.

With this experiment we solved the high variance issue we had during our pilot study with expert visual designers. The cost for this is two-fold. First, subjects evaluated methods one at a time, avoiding direct comparisons among displays that were possible in the first study.

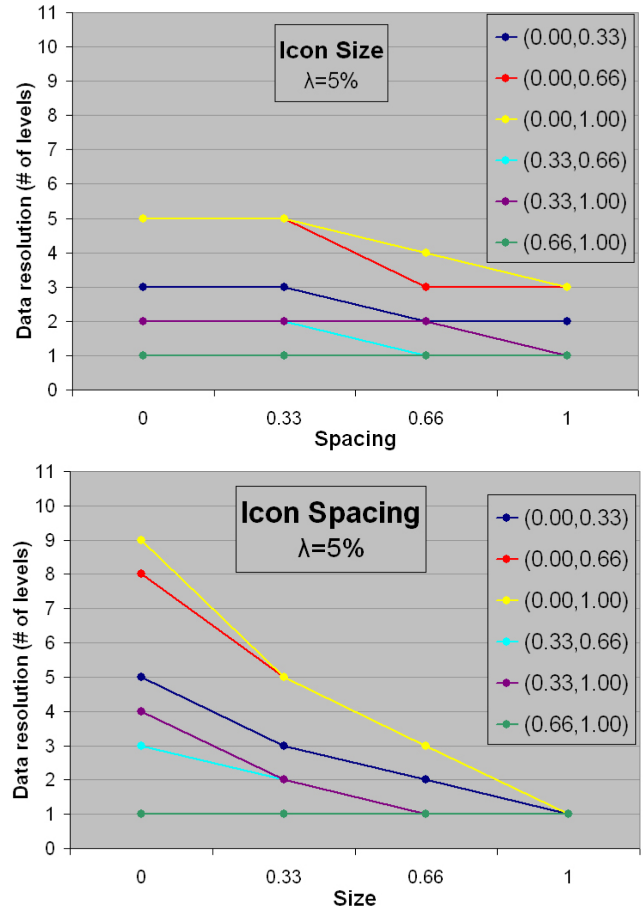

Fig. 8. Results of the data resolution task for size and spacing. Both plots correspond to $\lambda = 5\%$. 95% confidence intervals are around ± 2. Observe how increased values for the distractor variable decrease the data resolution results consistently across different ranges.

Comparative critique is a very useful tool design educators utilize constantly, but one that we had to sacrifice to improve the quality and quantity of data obtained. Secondly, we did not use expert visual designers as subjects, so we could not expect feedback on why a method performs as it does for a given task. Our tasks now are more perceptual than conceptual and the low variance of the data, along with consistent trends, validates our choice of non-expert subjects.

During the study, all wave-like patterns utilized a sine-wave function. Since we are looking for *jnd* values, a triangular function with more marked ridge and valley lines or even a step function could yield different results. A step function, for example, would allow us to generalize our results to discrete-valued datasets.

Our first priority after obtaining these encouraging results will be to increase the number of visual elements involved, including color and orientation. This will require moving to a between-subject design to avoid fatigue when running the experiment. Once we have more data about how the different elements interact, we will begin defining a model for higher order combinations that we would also need to evaluate on real datasets. We will use expert visual designers again at that point, since exhaustive exploration of such a high dimensional space would be impractical.

This experiment is an important first step in our very complex modeling project. The tasks chosen here measure characteristics that we will use to evaluate the effectiveness of a visualization method. This effectiveness will be measured by quantifying how well a method fulfills a set of given design goals such as how much data resolution is required or what minimum spatial feature resolution a visualization method should guarantee. These results begin to describe how our space of visualization methods is structured, so we can, ultimately, efficiently search within it for effective methods for exploratory scientific visualization.

6 Conclusion

In this experiment we characterized the capabilities of a total of 120 different visualization methods to represent 2D scalar fields effectively. Our results successfully reproduce what other perceptual experiments have obtained when describing the individual performance of visual elements: icon brightness easily outperforms icon spacing and size while being subject to their interferences. Icon spacing also outperforms icon size, contradicting some previous results. Our main contribution is the successful application of a new methodology capable of evaluating perceptual interactions among multiple visual elements, making numerically explicit how a change in one of the elements affects how a user reads the data in the visualization.

Acknowledgements

This work was partially supported by the NSF ITR grant CNS-0427374. Thanks to the students that helped evaluate and run this experiment, and all the Visualization Research Group at Brown for their helpful comments. Special thanks to Cullen Jackson for his invaluable help with experimental design and data analysis.

References

[1] D. Acevedo, C. Jackson, D. H. Laidlaw, and F. Drury. Using visual design expertise to characterize the effectiveness of 2d scientific visualization methods. IEEE Visualization'05, Poster Compendium, October 2005.

[2] J. R. Bergen. Theories of visual texture perception. In *Spatial Vision*. CRC, Boca Raton, FL, 1991.

[3] J. Bertin. *Semiology of Graphics*. University of Wisconsin Press, 1983.

[4] A. A. Bokinsky. *Multivariate Data Visualization with Data-Driven Spots*. PhD thesis, University of North Carolina at Chapel Hill, 2003.

[5] T. C. Callaghan. Dimensional interaction of hue and brightness in preattentive field segregation. *Perception and Psychopysics*, 36(1):25–34, 1984.

[6] T. C. Callaghan. Interference and dominance in texture segregation: Hue, geometric form, and line orientation. *Perception and Psychopysics*, 46(4):299–311, 1989.

[7] S. K. Card and J. Mackinlay. The structure of the information visualization design space. In *Proceedings of IEEE Symposium on Information Visualization*, pages 92–99, 1997.

[8] C. M. Carswell and C. Wickens. The perceptual interaction of graphical attributes: Configurality, stimulus homogeneity, and object integration. *Perception and Psychopysics*, 47:157–168, 1990.

[9] S. M. Casner. A task-analytic approach to the automated design of graphic presentations. *ACM Transactions on Graphics*, 10(2):111–151, April 1991.

[10] E. H. Chi. A taxonomy of visualization techniques using the data state reference model. In *Proceedings of InfoVis 2000*, pages 69–76, 2000.

[11] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, September 1984.

[12] S. G. Eick. Engineering perceptually effective visualizations for abstract data. In *Scientific Visualization Overviews, Methodologies and Techniques*, pages 191–210. IEEE Computer Science Press, 1995.

[13] W. D. Ellis. *A source book of gestalt psychology*. Harcourt, Brace, and company, New York, NY, 1939.

[14] P. Hanrahan. Teaching visualization. *Computer Graphics*, 39(1):4–5, February 2005.

[15] C. G. Healey, R. S. Amant, and M. S. Elhaddad. Via: A perceptual visualization assistant. In *28th Workshop on Advanced Imagery Pattern Recognition*, 1999.

[16] C. G. Healey, K. S. Booth, and J. T. Enns. Harnessing preattentive processes for multivariate data visualization. In *Proceedings of Graphics Interface*, pages 107–117, Toronto, Canada, 1993.

[17] C. G. Healey, K. S. Booth, and J. T. Enns. High-speed visual estimation using preattentive processing. *ACM Transactions on Human-Computer Interaction*, 3(2):107–135, June 1996.

[18] B. Hibbard. The top five problems that motivated my work. *IEEE Computer Graphics and Applications*, 24(6):9–13, November/December 2004.

[19] C. Jackson, D. Acevedo, D. H. Laidlaw, F. Drury, E. Vote, and D. Keefe. Designer-critiqued comparison of 2D vector visualization methods: A pilot study. In *SIGGRAPH 2003 Sketches and Applications*. ACM SIGGRAPH, 2003.

[20] C. Johnson. Top scientific visualization research problems. *IEEE Computer Graphics and Applications*, 24(4):13–17, July/August 2004.

[21] D. Keefe, D. Karelitz, E. Vote, and D. H. Laidlaw. Artistic collaboration in designing VR visualizations. *IEEE Computer Graphics and Applications*, 25(2):18–23, March/April 2005.

[22] R. Kosara, C. G. Healey, V. Interrante, D. H. Laidlaw, and C. Ware. User studies: Why, how, and when. *Computer Graphics and Applications*, 23(4):20–25, July/August 2003.

[23] D. H. Laidlaw. Loose, artistic "textures" for visualization. *IEEE Computer Graphics and Applications*, 21(2):6–9, March/April 2001.

[24] D. H. Laidlaw, M. Kirby, C. Jackson, J. S. Davidson, T. Miller, M. DaSilva, W. Warren, and M. Tarr. Comparing 2D vector field visualization methods: A user study. *Transactions on Visualization and Computer Graphics*, 11(1):59–70, January-February 2005.

[25] S. Lange, H. Schumann, W. Muller, and D. Kromker. Problem-oriented visualisation of multi-dimensional data sets. In *Proceedings of the International Symposium and Scientific Visualization*, pages 1–15, 1995.

[26] A. MacEachren and M.-J. Kraak. Exploratory cartographic visualization: Advancing the agenda. *Computers and Geosciences*, 23(4):335–343, 1997.

[27] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, 5(2):110–141, April 1986.

[28] G. F. McCleary Jr. An effective graphic vocabulary. *IEEE Computer Graphics and Applications*, 3(2):46–53, March/April 1983.

[29] R. Nagappan. A compositional model for multidimensional data visualisation. In *In Proceedings of SPIE, Visual Data Exploration and Analysis VIII*, pages 156–167, January 2001.

[30] L. T. Nowell. *Graphical Encoding for Information Visualization: Using Icon Color, Shape, and Size to Convey Nominal and Quantitative Data*. PhD thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, November 1997.

[31] P. K. Robertson. A methodology for choosing data representations. *IEEE Computer Graphics and Applications*, 11(3):56–67, May/June 1991.

[32] H. Senay and E. Ignatius. A knowledge-based system for visualization design. *IEEE Computer Graphics and Applications*, 14(6):36–47, November 1994.

[33] M. Tory and T. Moller. Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics*, 10(1):72–84, January/February 2004.

[34] C. Wallschlaeger and C. Busic-Snyder. *Basic Visual Concepts and Principles for Artists, Architects and Designers*. McGraw Hill, 1992.

[35] C. Ware. *Information Visualization. Perception for Design.* Elsevier, 2nd edition, 2004.

[36] J. M. Wolfe. Visual search. In *Attention*. University College London Press, London, UK, 1998.