Information Visualization (2004) 3, 123–133

www.palgrave-journals.com/ivs

Analyzing perceptual organization in information graphics

Martin Wattenberg[1]
Danyel Fisher[2]

[1]IBM Watson Research Cambridge, Cambridge, MA, U.S.A.; [2]University of California, Irvine, CA, U.S.A.

Correspondence:
Martin Wattenberg, IBM Watson Research, Cambridge, 1 Rogers Street, Cambridge, MA 02142, U.S.A.
E-mail: mwatten@us.ibm.com

Received: 24 November 2003
Revised: 30 January 2004
Accepted: 8 March 2004

Abstract

We propose a new method for assessing the perceptual organization of information graphics, based on the premise that the visual structure of an image should match the structure of the data it is intended to convey. The core of our method is a new formal model of one type of perceptual structure, based on classical machine vision techniques for analyzing an image at multiple resolutions. The model takes as input an arbitrary grayscale image and returns a lattice structure describing the visual organization of the image. We show how this model captures several aspects of traditional design aesthetics, and we describe a software tool that implements the model to help designers analyze and refine visual displays. Our emphasis here is on demonstrating the model's potential as a design aid rather than as a description of human perception, but given its initial promise we propose a variety of ways in which the model could be extended and validated.
Information Visualization (2004) **3,** 123–133. doi:10.1057/palgrave.ivs.9500070

Keywords: Screen design; software psychology; perceptual organization

Introduction

The design of information graphics remains a poorly understood, hit-or-miss process. Part of the difficulty is that models for how humans extract information from visual displays remain incomplete. Indeed, seemingly minor design variations can have dramatic effects on comprehensibility. As a result, creating effective displays often requires expensive user tests, time-consuming redesigns, and even a certain amount of guesswork.

Many researchers have recognized these problems and have investigated guidelines and models for the perception of information graphics.[1] Much work has been done on the efficacy of different visual encodings,[2,3] resulting in useful rules about the use of color, position, area, etc. to represent different types of variables. Others, for example Healy *et al*.[4] have investigated how models of preattentive processing can be used in designing visualizations.

But these lines of research do not address a key element in the efficacy of an information graphic: the degree to which its perceptual organization reflects the organization of the underlying data. Many authors have stressed that to design successful information graphics, one must take into account the effects of perceptual grouping. For instance, the work of Kosslyn[5] contains many examples in which unintentional grouping effects lead to confusing displays. It would therefore be useful to have a tool that helped designers assess the perceptual organization of their designs.

Some attempts have been made to model perceptual organization in information graphics. Tufte provides general guidelines, such as the 'Macro/Micro' principle.[6] But quantitative models suitable for software implementation are rare. Several authors have analyzed special classes of displays: Tullis[7] analyzes alphanumeric screens; Schneiderman *et al.*[8] investigate standard Visual Basic dialog boxes. The work of Saund[9] on deriving perceptual structure in the context of sketch editing is more ambitious, but still requires a vectorized version of a graphic as input. Because it is not amenable to the analysis of non-vector-based visualizations, it is problematic to apply this method to the output of existing programs. (Portions reprinted, with permission from Martin Wattenberg[10] (c) 2003 IEEE.)

In this paper, an extension of the work by Wattenberg and Fisher[10] we introduce a formal model of visual organization which can be applied to a broad class of information graphics. We present an algorithm that takes as input an arbitrary grayscale image, and returns as output an analysis of the image's organization that links perceived structures at different scales. We do not claim that this technique captures all or even most of the aspects of human visual perception — that would be far beyond any current system — but we do propose it as a potentially helpful new model of a particular aspect of perceptual organization.

We then describe a prototype software tool that applies this model to help designers see how an information graphic may be understood by viewers. We demonstrate the utility of the model by exhibiting a variety of examples in which it captures aspects of design aesthetics; we also show how it can be used in the redesign of a real-life visualization. Finally, we discuss directions for validating and extending the model.

Multi-scale model of visual organization

Motivation: importance of multiple scales

Most information graphics display structure at several different scales. That is, an image will contain large-scale organization as well as many smaller details. Our hypothesis is that at all these scales the visual structure should reflect the structure of the data being conveyed, with large-scale organization reflecting a broad overview or summary, and smaller details reflecting details of the data. As Bertin[11] puts it:

> A graphic should not show only the leaves; it should show the branches as well as the entire tree. The eye can then go from detail to totality and discover at once the general structure and any exceptions to it.

This intuition about multiple scales is shared by many visual designers. Typographers, for example, routinely speak of a visual hierarchy in text layouts. Figure 1 shows a hand-drawn example of such a hierarchy. (We have chosen a piece of text as an example for analysis in the next section since it has several natural, unambiguous scales: letter, word, line, and paragraph.)

Despite the general belief that multi-scale structure exists and is important, that structure can prove surprisingly elusive. Experienced designers use techniques such as looking at an image from across a room or holding it upside down to get a better sense of its organization. In many ways, it would be helpful to have a mathematical model that matched the standard designer's intuition. Such a model could be useful to designers, for instance, who could apply it to early designs to see if the structure matched what they wished to communicate. It could also be helpful in automating some aspects of design – for instance, a computer might try to use the model to optimize the correspondence between visual structure and data structure. All of these potential uses rely on a precise model that can implemented algorithmically.

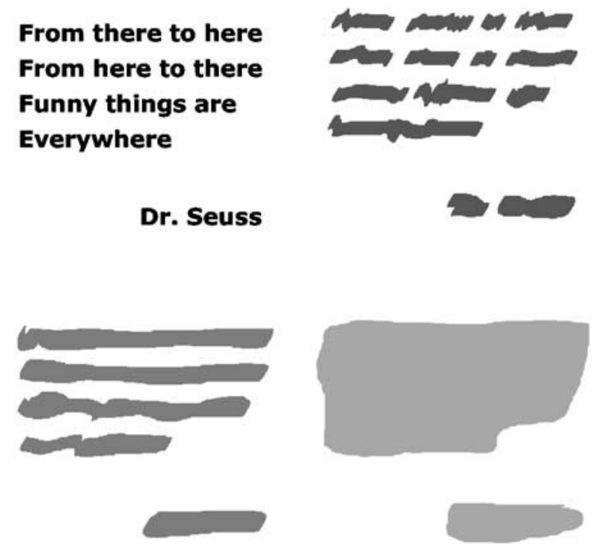

Figure 1 Visual hierarchy, hand-drawn, for a piece of text. (The 'Dr. Seuss' image.)

Human and machine vision Psychologists have long studied perceptual organization and its multi-scale aspects. A full review of the psychological literature on this topic is beyond the scope of this paper, but we cite a few reference points. Gestalt psychologists, starting with Wertheimer,[12] have proposed a number of 'laws' for how the brain groups objects: by proximity, good continuation, and so on. Multi-scale aspects of grouping have also been addressed in several lines of research.[13,14] Many of these theories of grouping were qualitative, but investigators have worked on creating quantitative or algorithmic models as well. Typical examples from this large research area are Kubovy,[15] who treats grouping by proximity in dot lattices, and the work of Li[16] on neural network simulations of cortical processing. One interesting system is Logan's CODE theory of visual attention,[17] which has in fact been applied to information visualizations.[18] Like many other psychologically derived models, however, this system requires 'feature' constructs as input rather than direct pixel data, meaning that they cannot be directly applied to images of information graphics.

The field of machine vision, however, provides a different and more immediately fruitful perspective, and is a rich source of pixel-level models. Analyzing visual structure has long been recognized as an important component of computer vision (see Witkin and Tenenbaum[19]), and modern computer vision frameworks typically are designed to be applied to arbitrary images. In this paper, we highlight one particular framework, scale space theory, and through a series of examples suggest that it is particularly suitable for the analysis of information graphics. A natural future direction would be to reconnect this model with psychological work through experimental validation.

Limits and assumptions

Rather than attempting to model the full range of visual experience, we focus on non-interactive motionless grayscale images, and make no attempt to reconstruct a three-dimensional (3D) scene. By eliminating from consideration color, depth, motion, and interactivity, we simplify the domain considerably yet retain significant generality, for example encompassing a significant fraction of printed information graphics. Furthermore, even within the domain of static grayscale images, we do not attempt to create a complete model of visual grouping. Instead, as a first step, we focus on a single type of structure. Obviously, it would be desirable to have a model that eventually did account for the many other dimensions of visual perception, and in the final section we discuss potential generalizations.

Our model: mathematical definition

We now define our model. First we make precise the idea of 'scale.' Then we define a simple method of extracting structure at a given scale. Finally, we describe a technique for linking structures found at different scales. Throughout, we illustrate our progress through the algorithm with the data in the top-left of Figure 1.

Scale space We base our model on the classical machine vision concept of *scale space*. Scale space theory[19–22] is a formalism that describes the structure of a signal at many different scales at once. (Despite the similar name and notation, scale space in this sense is not directly related to the 'space-scale diagrams' of Furnas and Bederson,[23] an elegant application of Riemannian geometry to zooming user interface design.)

To define scale space precisely, we need some notation. First, we represent the input image as a function:

$$f : [0, L] \times [0, L] \rightarrow [0, 1].$$

That is, we take f to be a function on a square of side L, where a value of 0 corresponds to black, 1 to white, and values in between correspond to shades of gray.

Given the function f, we then extend its domain to a 3D 'scale space' by a special family of functions f_s, where $s \geq 0$. First, let G_s be a Gaussian kernel with 'width' s; more formally, let

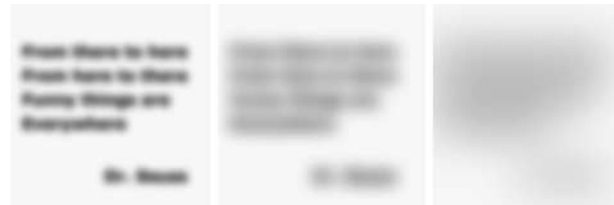

Figure 2 f_s for the Dr. Seuss image from Figure 1, where $s = 8$, 16, 44.

$$G_s(x, y) = \frac{1}{2\pi s^2} e^{-(x^2+y^2)/2s^2}.$$

We define then f_s by

$$f_s = f^* G_s,$$

where * represents convolution. Informally, the function f_s represents the original image having been blurred by a factor of s. Figure 2 shows f_s for three different values of s, using the same starting data as the top-left of Figure 1. The 3Ds pace formed by the spatial dimensions x,y and the new scale dimension s is known as scale space, and by analyzing the functions f_s on this 3D scale space we can uncover important structures in the original two-dimensional (2D) image. We refer to 's' as the 'scale' of the image, and so describe Figure 2 as being seen at three different scales.

Structure and segmentation Having defined scale space we now need a notion of structure or organization at a given scale. There are many possible ways to define a structure. We choose to define structure by creating a segmentation of the image at each scale. For a given scale s, we follow Marr and Hildreth[24] and consider the difference-of-Gaussians edge detection function

$$g_s = f_s - f_{3s/2}.$$

This function is one of the best-studied edge detectors, and has some correspondence to the responses of retinal neurons.[24] It is a close approximation of another classical edge detector, the Laplacian operator, but numerically more stable. Figure 3 shows the function g_s for the Dr. Seuss image at three different scales.

We can then naturally segment the square as follows. Consider the two (disjoint) regions where $g_s < 0$ and $g_s > 0$. We let the connected components of these regions form the elements of our segmentation. The sign of g_s is thus constant in a given segment and has some significance; it corresponds, very roughly, to whether the segment is brighter or darker than its neighbors.

Why use the difference-of-Gaussians edge detector? It has several advantages. First, simplicity: it is well-understood and efficient to calculate. Second, unlike several other popular edge detectors,[25,26] the difference-of-Gaussians method has the benefit of immediately producing closed contours, thus creating a segmentation without additional steps. Third, the sign of the function g_s is useful in creating an algorithmic version of the linking step below. Despite these advantages, it is important to note some well-known drawbacks to this technique: poor localization, rounded corners, and over-sensitivity.[27] A different edge detector would not, however, fundamentally alter the framework of our model.

Figure 4 shows the resulting segmentation at scales of 8, 16, and 44. In the top row the edges of segments are shown. In the bottom row, each segment has been filled with a single gray tone representing the average grayscale value of the pixels in the segment, a technique we call a *scale space cartoon*. The scale space cartoon itself is a small

Figure 3 Difference of Gaussians: $f_s - f_{3s/2}$, $s = 8$, 16, 44. 50% gray is zero; dark gray is negative; light gray is positive.

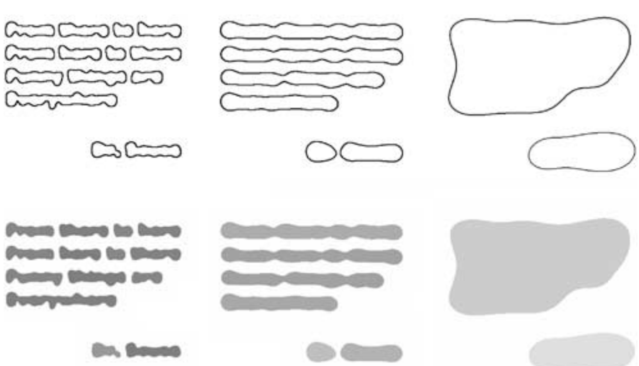

Figure 4 Algorithmically derived segmentation of the Dr. Seuss image for $s = 8$, 16, 44. Top: edges of segments. Bottom: filled segments, or *Gestalt cartoons*.

but interesting visualization issue: informal tests showed that for complex segmentations, users found these scale space cartoons easier to interpret than the outline view commonly seen in computer vision output. Note how closely the images in the bottom row match the hand-drawn diagrams of Figure 1.

The scale space cartoons do raise some new issues, however. One potential concern is that two adjacent segments with similar average values may be difficult to distinguish. In many cases, however, this difficulty simply reflects the fact that the visual difference between the two segments is relatively unimportant. In situations where drawing attention strongly to all segmentations is necessary, one might draw a faint outline around each segment.

Edge detection is not the only way to locate structure at a given scale. Probably the most common method – one used in many of the original scale space papers – is to analyze local maxima and minima of the function f_s.[21,28] Often this analysis is accompanied by some sort of watershed segmentation.[22,29] We tried several variants of this technique but found they produced poor results, possibly due to the non-generic nature of typical information graphics. Compared to images of natural scenes, diagrams and visualizations have an unusual number of areas of nearly uniform brightness. In many cases we found that f_s contained ridges, valleys, and plateaus that were almost but not quite level, leading to a proliferation of local extrema that did not correspond to useful features in the image. Figure 5 shows an example. On the left is a simple graph. On the right are the boundaries of regions found by watershed segmentation for local minima (as in Lindberg[22]) for scale $s = 16$. It is clear that the graph line itself has dissolved into many individual segments, because the smoothed function has many almost indistinguishable extrema in the area of the main 'graph line' in the image. This effect, which in no way reflects the visual experience of viewing the graph, is why we chose the edge-detection scheme described above.

Linking structures at different scales As described so far, the model finds structure only at a single scale. But the

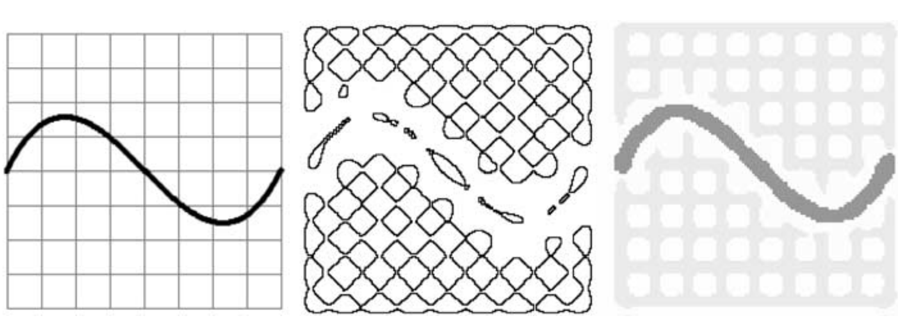

Figure 5 An image (left), and its boundaries as found by watershed segmentation (center). Contrast with the results of our edge-detection algorithm (right).

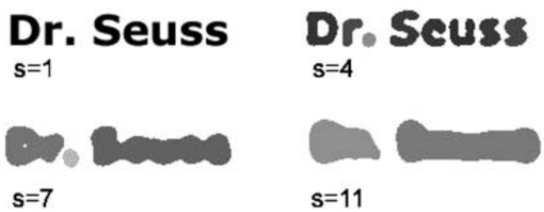

Figure 6 Four scales of Dr. Seuss.

perceptual structure of an image includes not just the structure at one scale, but the relationships between features at different scales. In the scale space literature, linking features between scales is often referred to as finding the *deep structure* of an image.[21] In this section we describe a novel method of finding this deep structure that is particularly useful for information graphics.

Consider the segmentations in Figure 6, shown as a series of scale space cartoons. It is visually clear that the two blobs in the $s=11$ view correspond to the individual letters of the words 'Dr.' and 'Seuss' respectively. The final part of our model is a method of making this intuition precise.

Let S_1 and S_2 be two image segments found at scales $s_1 \leq s_2$ respectively. We can naturally view S_1 and S_2 as embedded within the 3D scale space, that is, as the sets $\{s_1\} \times S_1$ and $\{s_2\} \times S_2$. We will say S_1 is *linked* to S_2, denoted by $S_1 \leq S_2$, if either $S_1 = S_2$ or there is a continuous curve C through scale space from a point on S_1 to a point on S_2, such that g_s maintains the same sign on C and s is monotonically increasing along C. It is easy to verify that the relation $\leq$ defines a partial order on the set of segments. It is also clear from the definition that this partially ordered set breaks into two disconnected components, one that corresponds to the subset of segments where $g_s < 0$, which we denote as L^- and one we call L^+ where $g_s > 0$. (It is possible for each of these two sets to have many maximal elements.) In some cases, L^- and L^+ turn out to correspond to foreground and background elements. For example, in the Dr. Seuss image, the segments corresponding to the text are represented in L^- while the whitespace is represented in L^+.

Figure 7 is a visualization of the results of connecting linked segments in L^- for the Dr. Seuss image.

The image shows a 3D view of scale space, with four separate planes highlighted (corresponding to $s=1$, 4, 7, 11). For each plane, we show the segmentation for the corresponding s value, and for each pair of linked segments in adjacent planes we have drawn a line between the segments' centroids. For simplicity, in this diagram we only show L^-, the segments with negative g_s, since they account for the main visual structure. The result is a tree structure on the words that corresponds to the intuitive hierarchical division of a phrase into words and words into letters.

The choice of a 3D display is a visualization exercise in its own right. We tried various alternatives, such as

Figure 7 Linked segments in L^- at different scales for part of the Dr. Seuss image.

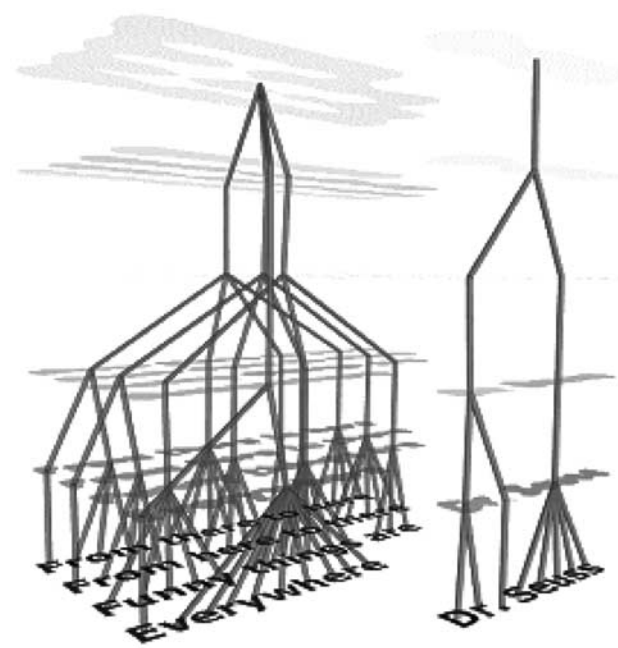

Figure 8 The linked structures in L^- for the entire Dr. Seuss image.

abstract graph-theoretic views of the lattice and a layout of 2D thumbnails with connections drawn between segments. In these cases, however, users were uniformly confused about the connection between the lattice structure and the image.

For completeness the L^- lattice for the entire Dr. Seuss image is shown in Figure 8. Again, the structure nicely corresponds to the intuitive hierarchy of paragraphs, lines, words, and letters.

Although linking structures at different scales by following zero-crossings of various operators is common in scale space theory,[22] the particular linking described here is unusual, and in fact is a key distinguishing feature of our model. Most scale space segmentation algorithms

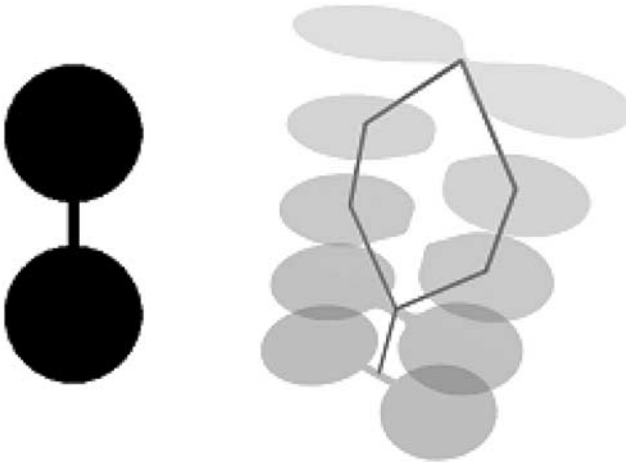

Figure 9 Image whose structure is not tree-like. Left: original image. Right: structure of L^{-}.

seek a hierarchical segmentation of an image, where the partial order is always a tree structure. The segmentation described above, however, can produce non-nested segments with non-tree lattices. In the context of scene segmentation and object recognition – the conventional applications of scale space theory – this is an undesirable property. But as several authors have pointed out,[9,29] a non-tree lattice seems to model well the visual experience of certain images. Indeed, given that the goal of many information graphics is to portray complex inter-relationships, any model that led to pure trees would be of limited applicability.

Figure 9 gives an example of an image whose visual structure is not tree-like. The barbell image, at a small scale, is one continuous object, at a slightly larger scale breaks into two main parts, and at a large scale merges into one object again.

Possible structures Since the analysis technique described here produces visual structures that have the form of a lattice, a natural question is whether all finite lattice structures can be represented, or *realized*, by some image. That is, how expressive is the system: Given a data set with an arbitrary lattice structure L, is there always an image whose visual structure is L? We do not know the answer, but conjecture that there are classes of lattices that are not realizable. Possibly this may be easier to prove when the segmentation is performed according to the zeros of the Laplacian operator, which is similar to the difference-of-Gaussians method used here but allows the introduction of mathematical machinery related to the heat equation.

At the same time, it seems likely that many structures are realizable. For example, all trees are realizable via a treemap-like diagram. (To see how this could work, consider only black and white images where the black region is the disjoint union of closed connected sets corresponding to 'leaves'. We can then show that any

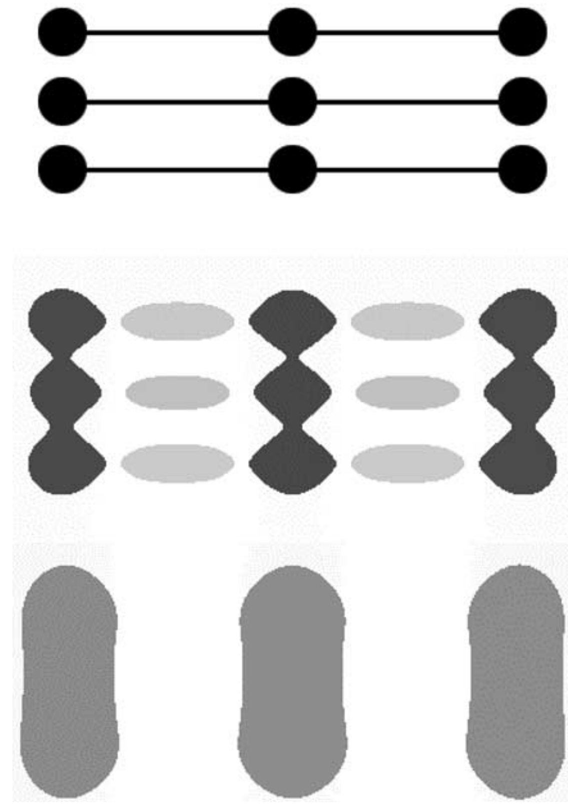

Figure 10 An image (top), with scale space cartoons at high (middle) and low (bottom) resolution.

'leaf' region can be subdivided to create a new tree branch of child nodes without disturbing the rest of the structure.) Some complex non-tree structures are also realizable. For example, Figure 10 shows an image with 3 'leaf' nodes that split into many more items, and then merge back into 3 'parent' nodes, each of which is linked to one of the original leaves. This example indicates the variety of possible structures; classifying realizable structures completely is an interesting problem for future research. The example also shows a limitation of our method: the 'large-scale' structure is not always the most salient to a human observer: most people would likely describe the original diagram as three horizontal objects rather than 3 vertical ones,[30] although the vertical grouping is a notable feature of the graphic.

Related methods The general concept behind our construction, analyzing a signal at multiple resolutions, is found in many fields. One closely related technique of multi-scale analysis is the continuous wavelet transform. The difference-of-Gaussian operator used in our segmentation step is in fact a close approximation to the Mexican Hat wavelet.[31] Statisticians use convolution with Gaussian kernels of varying radii in *kernel density estimation*,[32] a non-parametric estimation technique; Leung *et al.*[29] have applied scale space theory to statistical clustering using a watershed-type segmentation technique. A third technique that is closely related is the multi-scale pyramid representation.[33] Originally used for image compression, it is interesting to note that this structure is now used in at least one sophisticated model of visual perception and attention.[34]

Results and applications of the model

To test our model, we built a software tool that applies the model to arbitrary input images. The tool was used to create all the images in this paper, with the exception of the hand-drawn Figures 1, 13, and 14. As a demonstration of our model, we apply it to three case studies, and show how it can be used in the redesign of a real-life visualization.

The software

The software tool contains the following numerical approximation of the model. We represented the image functions f_s as 2D arrays of floating-point values (one per pixel in the original image), and computed f_s for only a few discrete values of s. To perform linking, we looked at each pair of successive approximations to f_s, and connect any two segments that share a sign and which overlap. Our implementation is written in Java, and on a 700 MHz Pentium 3 PC requires up to a minute to perform a full structural analysis on a 800×600 pixel image at 15 scales. Once the analysis is performed, it is saved for viewing as both a series of grayscale images and as a 3D VRML file. We describe the interface of the application in more detail in Section 4.

Questions of scale An important consideration for computational algorithms is to ensure that they are able to handle large data sets. Because our system examines the output of a visualization algorithm — the pixels it produces — and not the underlying data set, it is limited by the size of the screen. The segmentation and linking steps each require time and memory proportional to the number of pixels, since each pixel is 'marked' once in an offscreen buffer. For a constant s, computing g_s is also linear in the number of pixels. Since the algorithm is not affected by the number of items in the underlying data set for a visualization, and indeed works by progressively simplification of an image, it is in practice computationally feasible. In addition, the examples below show that often only a small subset of scaling levels need to be examined in order to reveal useful information about perceptual grouping.

A simple example: graphs and grid lines

Our first example shows scale space cartoons of two versions of a simple graph (Figure 11). At top left is a graph with thin gridlines, at top right is a graph with overpoweringly thick ones. The segmented versions at scale $s=4$ are shown below. In the graph with thick gridlines the graph itself is not segmented from the background. This is an interesting indication of both the strength of our model and one of its limitations. A human can segment the graph in the second diagram by using orientation information, which our model ignores. Nonetheless, doing so places an additional cognitive burden on the viewer, and in fact it is a standard principle of information design that grid lines should be significantly lighter than lines representing 'foreground' data. Thus the model indicates, correctly, that there is a problem with the second graph. This situation – where a minor visual change has a large effect on comprehensibility – is exactly where it is useful to have a model.

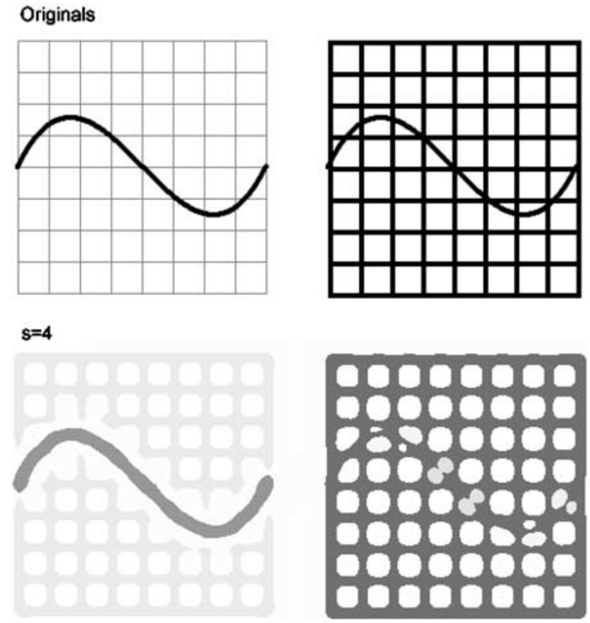

Figure 11 Gestalt cartoons showing differentiation of figure and ground in a graph. Left: thin grid lines. Right: thick grid lines.

A famous real-life example

How does the model fare on a real-life example? Figure 12 shows scale space cartoons for a complex scatterplot, the famous astronomical Hertzprung–Russell diagram. This scatterplot, which displays data on stars with temperature on the x-axis and absolute magnitude on the y-axis, plays a central role in scientists' conception of stellar evolution. The HR diagram at the top left of Figure 12 is reproduced directly from Spence and Garrison,[35] which contains a detailed discussion of this historically significant information graphic.

The segmentations in the scale space cartoons capture the intuitive experience of reading the diagram: the small-scale ($s=4$) view emphasizes the vertical structures, while at $s=8$ and 16 the large-scale clusters stand out. The areas highlighted for $s=16$ correspond nicely to the standard organization given by human experts: Figure 13 shows how an astronomer structures the diagram.

One subtlety is that at high scales some regions of the scale space cartoon become extremely faint, making them difficult to see. In Figure 12, for example, the area

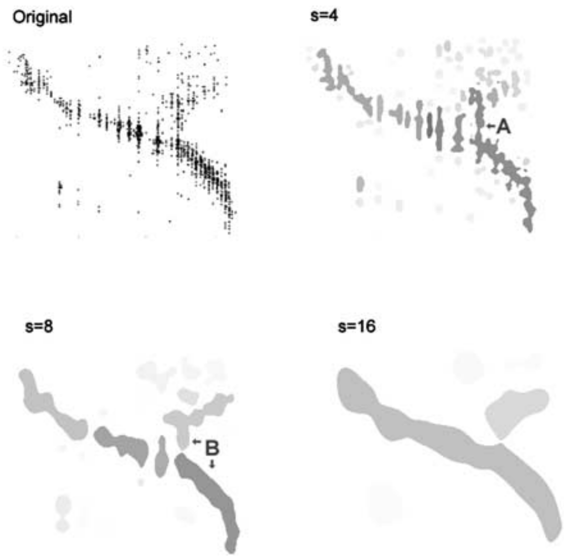

Figure 12 Original image and gestalt cartoons of the Hertzprung–Russell diagram. Reprinted with permission from The American Statistician. Copyright 1993 by the American Statistical Association. All rights reserved.

corresponding to the 'white dwarfs' section toward the bottom left has several subtle regions, barely visible at $s=8$. This problem could be fixed by outlining the regions in a dark color, but that tactic could bring its own problems, implying stronger salience for small regions than is actually the case.

The regions labeled A and B in Figure 10 show another example of how a non-tree structure can be an appropriate model. The region labeled A is a single large segment, reflecting the small-scale structure of a combined dense vertical and diagonal cluster. But a larger scale, $s=8$, that segment has broken into two parts, at B, corresponding to the giants and main sequence regions in Figure 13. Since at the highest scale these segments merge again, we see that the model produces a non-tree lattice structure. This contrasts with many clustering methods and with conventional scale-space segmentation techniques, which produce trees only.

A treemap redesign

Finally, we discuss how the model can inform the design of a visualization. We take as our example the SmartMoney Market Map[37] a treemap visualization[38] that displays data on several hundred publicly traded stocks. The first author of this paper, who led the design of the Market Map, has on many occasions heard the comment that the borders between regions are not strong enough. His intuition, however, was always that they were perfectly fine as is. Since this is exactly the kind of design issue where a perceptual model would be useful, we decided to apply our software tool. To make a comparison, we created a stylized version of the current Market Map and a

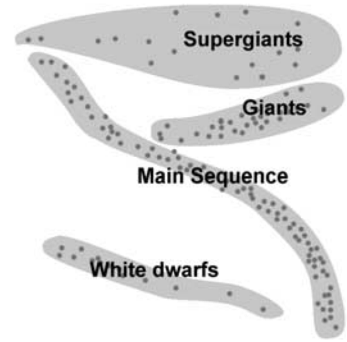

Figure 13 Human expert partitioning of HR diagram. After Fix.[36]

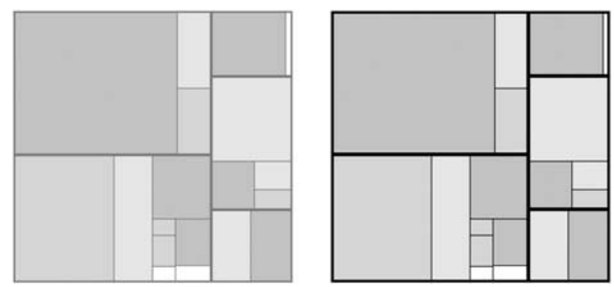

Figure 14 Left: sketch of portion of Market Map. Right: A redesign with stronger borders.

redesigned version with darker and thicker borders. (See Figure 14).

When we fed these images into our model, the results were clear. Figure 15 shows the structure derived for the current version. Note that the lattice structure is complex, confusing, and does not follow the underlying hierarchy of the data items. At point A in the diagram, for example, two items in different groups are spuriously joined. In Figure 16, the lattice structure is far simpler and close to a perfect tree. The key is that the borders demarcating regions are now significantly darker than the filled rectangles, and that widths of the borders at different levels of hierarchy are much different. This dramatic result has led to a reconsideration of the original design – exactly what we would want from a perceptual model.

The 'scale space cartoonist' application

In this section, we provide a more detailed description of the end-user application for applying our scale space model, currently named 'The Scale Space Cartoonist.' The application is written in Java and is platform-neutral. We will first describe the graphical pieces that make up the application. We believe that the general framework may be useful in for other applications aimed at automatically analyzing graphics.

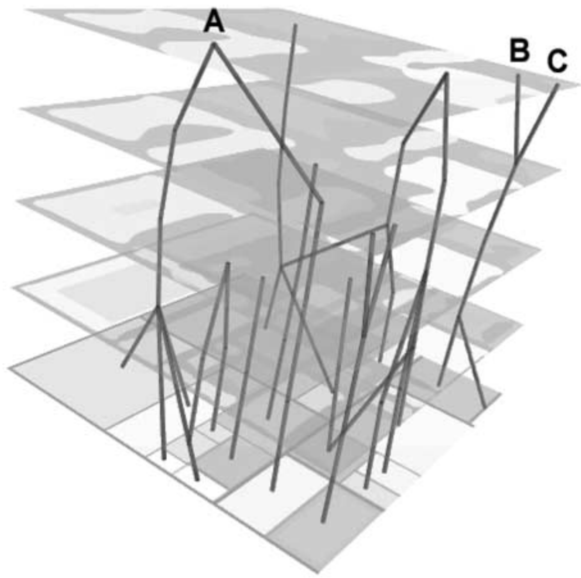

Figure 15 L^+ structure of original map design at scales up to $s=20$. Some flaws: A, two items in different groups are spuriously joined; B and C, a single group is spuriously separated.

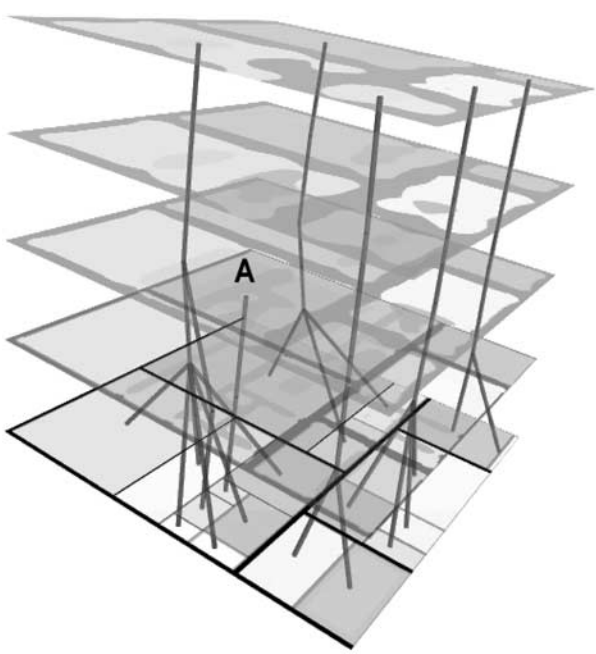

Figure 16 L^+ structure of redesigned map. Grouping is almost perfect; only flaw is an 'orphan' item at A.

Interface components

The application is composed from four main components along with an auxiliary module for 3D viewing (see Figure 17). The large panel at the right displays a cartoon of the current graphic. On the left is a pane with thumbnail views of the analysis of the image at different

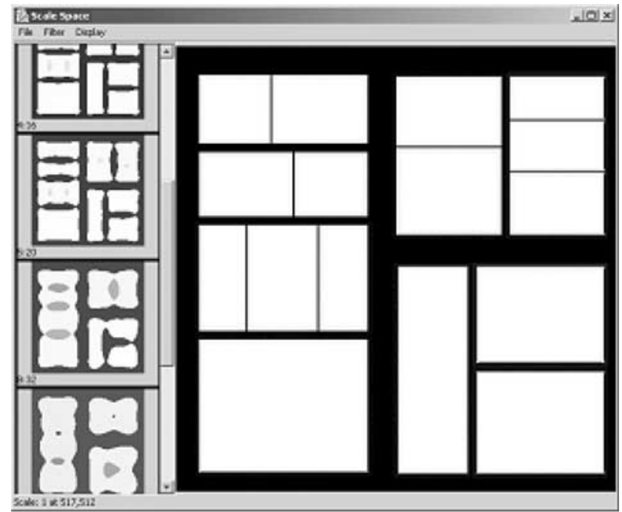

Figure 17 The Scale Space Cartoonist interface.

scales. In informal tests, the previews proved to be critical, allowing users to quickly view cartoons of the image at a variety of scales. Each thumbnail is clickable so that the user may easily select one for closer examination. As mentioned above, a full structural analysis may take up to a minute for a full-screen image, but the small size of the thumbnails, and the fact that they represent only a small subset of scale, means they can be generated in under a second.

The menus at top allow the user to customize the view and to perform simple input and output. Users may analyze an image either by loading a file or pasting an image from the clipboard. (The latter feature allows easy analysis of running programs: after capturing a screen dump, a user can simply paste the clipboard image into the Scale Space Cartoonist.) The 'Filter' menu allows the user to change between different visual filters. The default filter is the difference-of-gaussians segmentation described in the section on 'our model = mathematical definition', but the program also enables viewing of simple Gaussian blurs and a Laplacian convolution. The 'Display' menu allows the user a choice between different views of the processed image. The default is a scale space cartoon, but the user can also view raw image data as in Figures 2–4. One option that seemed initially promising but proved confusing for viewers was to use alpha blending to overlay a scale space cartoon (in a shade of orange) on top of the original black and white image. Although we had hoped that this would provide a good way for users to understand the connection between the cartoon and the image, it was hopelessly confusing in practice.

The application also allows two special viewing modes. In some cases, it is helpful to see the scale space cartoons for many scales simultaneously. A menu option lets the user choose to make a 'contact sheet' image, showing a packed array of cartoons at any set of scales desired. Such contact sheets are convenient for printing. A second menu option lets the user create a 3D VRML view of the scale space analysis, viewable in a web browser with a VRML 2.0 plug-in (e.g. Parallelgraphics' Cortona plug-in). The 3D view is good for detailed analysis and demonstrations, but is difficult to reproduce clearly in print.

Current limitations and future directions

We have a described a model that is by no means complete. Here we discuss various limitations and possibilities for addressing those limitations. Perhaps most important, the model is at its core a psychological hypothesis and therefore cries out for experimental validation. There are several natural directions to investigate. One tactic would be to compare the structures generated by our model with self-reports of users' perceptions. A more pragmatic validation would be to study whether, in using the software tool described here, creators of information graphics are able to modify their designs in a way that user studies show are beneficial.

Two obvious shortcomings of our model are that it applies only to grayscale images and that it addresses only one type of grouping mechanism. One of the reasons to choose scale space analysis as the basis for our method is that there is a rich body of research extending the basic idea to more general aspects of image structure. Theories that handle color or orientation have been proposed[14,34,40,41,43] and could be applied to our model. In addition, increasingly sophisticated visual perception models, based on psychology, are becoming increasingly available.[30,39] Some of those models fruitfully address notions of connectivity and suggest a broad set of ways that humans cluster information. Orientation-sensitive models have the potential to address the fact that our method often confers insufficient saliency on lines and curves, which can lead to unsatisfactory analyses for graphics such as node-and-link diagrams. It may also be advantageous to use a more sophisticated segmentation method than the difference-of-Gaussians edge detection employed here, since in some complicated images the simple segmentation algorithm described here can yield counterintuitive results. It would also be useful to investigate ways of optimizing the numerical algorithm to run in an interactive timeframe.

A deeper limitation is that the model relies on a directed graph to describe the perception of an image. Obviously, in many cases an information graphic conveys a completely different type of data. Consider a line graph. As we showed above, our scale space model can provide insight into whether the graph stands out from a background grid. The model does not, however, provide any insight into the numerical values conveyed by that graph – which is clearly the most important data it conveys. It is therefore unlikely that scale space analysis, no matter how it is extended, will become the sole tool for evaluating the perception of information graphics.

Finally, although the scale space method may help identify flaws in presentation, it currently does not give suggestions for fixing those flaws. A potentially fruitful area of long-term investigation would be using the model to automatically optimize information graphics. That is, given a known data structure one could attempt to find a method for displaying that structure in an optimal manner according to the model detailed here. This could ultimately involve either an algorithm that incrementally improved a representation, to find a local optimum, or even some method of mapping a structure directly to a globally optimal visual representation.

Conclusion

We proposed a new technique for modeling multi-scale perceptual organization in information graphics. The model is based on a classical machine vision technique, scale space, with a novel method of creating links between structures at different scales. We demonstrated how a software implementation of this model captures important aspects of design aesthetics for several information graphics, and gave an example of how it may be used to give input into questions of design. We believe there is sufficient evidence of promise that it is worth extending and validating the model.

Acknowledgments

Thanks to Ben Shneiderman for helpful comments on the purpose of modeling perception, and to Dan Gruen for advice on cognitive psychology. Thanks to Kushal Dave, Steve Rohall and Bernard Kerr for comments on this manuscript, and to the IBM Research CUE group for support and encouragement. We also acknowledge the anonymous referees for several helpful suggestions and references. Figure 12 reprinted with permission from *The American Statistician*.

References

1 Ware C. *Information Visualization: Perception for Design*. Morgan-Kaufmann: San Francisco, 2000.
2 Cleveland W. *The Elements of Graphing Data*. Wadsworth: Monterey, CA, 1980.
3 Mackinlay JD. Applying a theory of graphical presentation to the graphic design of user interfaces. *Proceedings of the ACM SIGGRAPH Symposium on User Interface Software (UIST '88)*, 1988; 179–189.
4 Healey CG, Booth KS, Enns JT. Harnessing preattentive processes for multivariate data visualization. *Proceedings Graphics Interface '93*, 1993; 107–117.
5 Kosslyn SM. Understanding charts and graphs. *Applied Cognitive Psychology*, 1989; **3**: 185–226.
6 Tufte E. *Envisioning Information*. Graphics Press Cheshire, Connecticut, 1990.

7 Tullis TS. A computer-based tool for evaluating alphanumeric displays. *Proceedings of INTERACT'84*, 1984.
8 Shneiderman B, Chimera R, Jog N, Stimart R, White D. Evaluating spatial and textual style of displays. *Proceedings of Getting the Best from State-of-the-Art Display Systems '95*. London 1995.
9 Saund E. Symbolic construction of a 2-D scale-space image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990; **12**: 817–830.
10 Wattenberg M, Fisher D. A multi-scale model of perceptual organization in information graphics. *InfoVis 2003*, 2003.
11 Bertin J. *The Semiology of Graphics*. University of Wisconsin Press: Madison, WI, 1983.
12 Wertheimer M. Gestalt theory. In: Ellis WD (Ed). *A Sourcebook of Gestalt Psychology*. The Humanities Press: New York, 1924/1950; 1–110.
13 Navon D. Forest before trees: the precedence of global features in visual perception. *Cognitive Psychology*, 1977; **9**: 353–383.
14 Palmer SE. Hierarchical structure in perceptual representation. *Cognitive Psychology*, 1977; **9**: 441–474.
15 Kubovy M. The perceptual organization of dot lattices. *Psychonomic Bulletin & Review*, 1994; **1**: 182–190.
16 Li Z. A neural model of contour integration in the primary visual cortex. *Neural Computation*, 1998; **10**: 903–940.
17 Logan GD. The CODE theory of visual attention. *Psychological Review*, 1996; **103**: 603–649.
18 Pirolli P, Card SK, van der Wege MM. Visual information foraging in a focus+context visualization. *Proceedings, CHI01, Human Factors in Computing Systems*, 2001.
19 Witkin A, Tenenbaum M. On the role of structure in vision. In: Rosenfeld A (Ed). *Human and Machine Vision*. Academic Press: New York, 1983.
20 Iijima T. Basic theory of pattern observation. *Papers of Technical Group on Automata Control, IECE, Japan*, 1959.
21 Koenderink JJ. The structure of images. *Biological Cybernetics*, 1984; **50**: 363–370.
22 Lindeberg T. *Scale-Space Theory in Computer Vision*. Kluwer: Dordrecht, 1994.
23 Furnas GW, Bederson B. Space-scale diagrams: understanding multiscale interfaces. *Proceedings CHI 95. Human Factors in Computing Systems*, 1995.
24 Marr D, Hildreth E. Theory of edge detection. *Proceedings of the Royal Society of London, Series B*, 1980; **207**: 187–217.
25 Canny J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986; **8**: 679–698.
26 Shen J, Castan S. An optimal linear operator for step edge detection. *Computer Vision, Graphics, and Image Processing*, 1992; **34**: 321–343.
27 Parker JR. *Algorithms for Image Processing and Computer Vision*. John Wiley & Sons: New York, 1997.
28 Witkin A. 1983. *Scale-space filtering. Proceedings of the Eighth International Joint Conference on Artificial Intelligence (IJCAI '83)*. Karlsruhe 8–12 August, 1983..
29 Leung Y, Zhang JS, Xu ZB. Clustering by scale-space filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000; **22**: 12.
30 Palmer S, Rock I. Rethinking perceptual organization: the role of uniform connectedness. *Psychonomic Bulletin and Review*, 1994; **1**: 29–55.
31 Antoine JP, Carrette P, Murenzi R, Piette B. Image analysis with two-dimensional continuous wavelet transform. *Signal Processing*, 1993; **31**: 241–272.
32 Scott DW. *Multivariate Density Estimation*. Wiley-Interscience: New York, 1992.
33 Burt PJ, Adelson EH. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 1983; **31**: 532–540.
34 Itti L, Koch C. Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2001; **2**: 194–203.
35 Spence I, Garrison RF. A remarkable scatterplot. *The American Statistician*, 1993; **47**: 12–19.
36 Fix JD. *Astronomy: Journey to the Cosmic Frontier*. (Second edition, updated) McGraw-Hill: New York, 1999.
37 Wattenberg M. Visualizing the Stock Market. *CHI99 Extended Abstracts*, 1999.
38 Shneiderman B. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on*, 1992; **11**: 92–97.
39 Geisler WS, Super BJ. Perceptual organization of two-dimensional patterns. *Psychological Review*, 2000; **107**: 677–708.
40 ter Haar Romeny B, Geusebroek JM, Van Osta P, van den Boomgaard R, Koenderink JJ. Color differential structure. *Scale-Space 2001*, 2001; 353–361.
41 Kalitzin B, ter Haar Romeny B, Viergever M. Invertible orientation bundles on 2d scalar images. *Scale-Space Theories in Computer Vision*. Springer: Berlin, 1997; 77–88.
42 Mozer M, Zemel RS, Behrmann M, Williams CKI. Learning to segment images using dynamic feature binding. *Neural Computation*, 1992; **4**: 650–665.
43 Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990; **12**: 629–639.
44 Wickens CD, Carswell CM. The proximity compatibility principle: its psychological foundation and relevance to display design. *Human Factors*, 1995; **37**: 473–495.