An Intuitive Model of Perceptual Grouping for HCI Design

Ruth Rosenholtz
MIT
Cambridge, MA
rruth@mit.edu

Nathaniel R. Twarog
MIT
Cambridge, MA
ntwarog@mit.edu

Nadja Schinkel-Bielefeld
MIT
Cambridge, MA
schinkel@mit.edu

Martin Wattenberg
IBM Watson Research
Cambridge, MA
mwatten@us.ibm.com

ABSTRACT

Understanding and exploiting the abilities of the human visual system is an important part of the design of usable user interfaces and information visualizations. Good design enables quick, easy and veridical perception of key components of that design. An important facet of human vision is its ability to seemingly effortlessly perform "perceptual organization"; it transforms individual feature estimates into perception of coherent regions, structures, and objects. We perceive regions grouped by proximity and feature similarity, grouping of curves by good continuation, and grouping of regions of coherent texture. In this paper, we discuss a simple model for a broad range of perceptual grouping phenomena. It takes as input an arbitrary image, and returns a structure describing the predicted visual organization of the image. We demonstrate that this model can capture aspects of traditional design rules, and predicts visual percepts in classic perceptual grouping displays.

Author Keywords

Perceptual organization, grouping, good continuation, proximity, similarity, Gestalt, contour integration.

ACM Classification Keywords

H5.2. User interfaces, theory & methods.

INTRODUCTION

Design of user interfaces and information graphics is poorly understood, and somewhat hit-or-miss in terms of effectiveness. A number of issues influence the success of a design, and these run the gamut of the underlying human behavior. A design must be good cognitively (can the user easily understand the semantic structure of the design?), perceptually (can they effortlessly interpret the visual information present in the design?), and socially (does the design fit into the user's workflow? will they want to use it?). Here we focus on perceptual aspects of design.

Perhaps the most important aspect of human vision for design is perceptual organization. Perceptual organization refers to phenomena in which the visual system quickly and seemingly effortlessly transforms individual feature esti-

mates into perception of coherent regions, structures, and objects. These phenomena were first studied in detail by
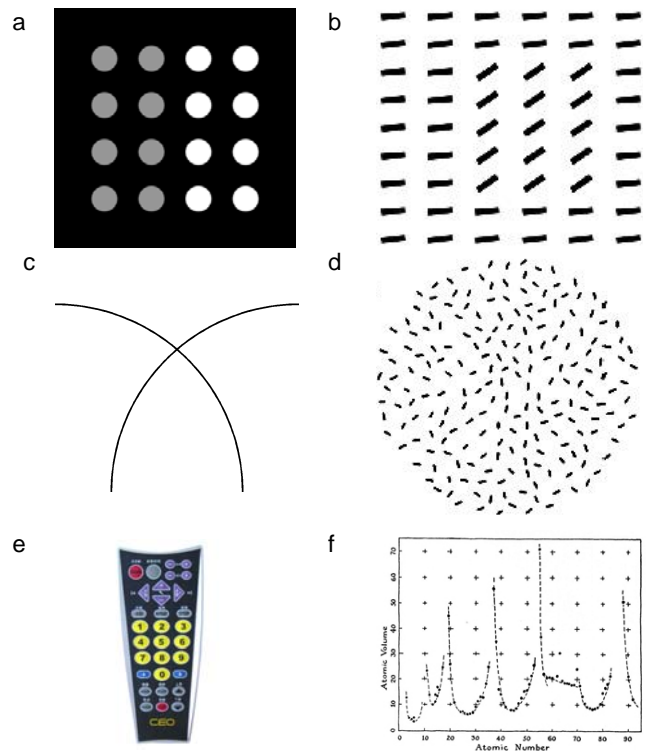

Figure 1: Perceptual grouping examples, including grouping by proximity & similarity (a, b), and grouping by good continuation (c, d). (e) A user interface; what is the percept? (f) A graph, from [5]. Will a user perceive the trend of the data?

the Gestalt psychologists, who produced a set of qualitative Gestalt principles that govern pattern perception [1, 2], including but not limited to: the tendency of things to group if they are nearby (the Gestalt law of proximity); if they share similar features (the law of similarity), or are smooth and continuous (the law of good continuation). The duals of perceptual grouping are important phenomena in their own right: we quickly and effortlessly perceive boundaries between certain visual textures, perceive edges between coherent regions in an image, and quickly detect unusual items that seem to "pop out" from the background. Examples of perceptual grouping phenomena are given in Figure 1.

Following the visual system's "rules" of visual organization makes interpretation of visual aspects of designs effortless: a user easily sees which labels refer to which parts of a dia-

gram, notices a trend in data, and makes connections between the "overview" and the "detail" in a map. Good designs use the natural perceptual processing power of the brain, and interpretations of such designs are fast, robust to instruction, and cross-cultural [3]. With poor visual design, the grouping structure may not match the structure of the information, leading to confusing displays [for examples, see 4, 5]. A user might see columns of information where in fact there were intended to be rows; incorrectly group two regions of a graphic that have no relation to each other, and so on. In Figure 1f, do the +'s at the grid points interfere with perceiving the data curves?

However, models for how users extract meaning from visual displays are incomplete. Designers often "eyeball" it–i.e. try to judge perceptual groupings using their own visual systems, perhaps using tricks such as "squinting" at the design or viewing it from far away to bring out coarser scale groupings. Various researchers have suggested general guidelines for design [3, 5, 6, 7]. Many existing models in use in the HCI and information graphics field are specific to particular types of displays e.g. alphanumeric [8], text documents [9, for a review], visual basic dialog boxes [10], and sketch editing with vectorized input [11]. In addition, some work has been done to translate very basic rules of thumb about "preattentive" (i.e. fast and effortless) visual processing to the design of visualizations [12].

Rules of thumb, based on simple behavioral experiments, are useful in understanding and guiding design. However, designers may have difficulty applying them to more complex displays. Ideally, one would prefer a model that could predict the likely perceptual groups for an arbitrary design. Such a model would be most useful if its mechanisms and output were easy to understand, as this transparency would aid a designer in making changes to a poor design.

In this paper, we draw on tools from statistics as well as recent work in computer vision to propose a model of perceptual grouping that is simple to understand and implement, yet effective. This model can predict image segmentation, contour integration, segmentation of orientation-based textures, grouping by similarity and proximity in standard Gestalt displays, segmentation of natural images, and grouping in more complex diagrams.

PREVIOUS WORK
A great body of computer vision work exists on the topic of perceptual grouping. However, much of it is inadequate for user interface (UI) designs and information graphics.

After many years of human and computer vision research, results of image segmentation models are still often quite poor. For one thing, many of these models have inherent biases against extended groups. A classic result is the tendency to predict the perceptually invalid segmentation of a pure blue sky into 3-4 separate regions.

Furthermore, the better models often do not easily lend themselves to intuitions. For example, the normalized cut algorithm [13], which works fairly well, treats image segmentation as a graph partitioning problem, with each pixel a node. The similarity between two pixels determines the weight between the corresponding nodes. This weight can be thought of as the tightness of a spring. The algorithm partitions this spring-mass system into regions that will tend to move independently. This is an evocative description, but it does not lend itself to easy intuitions about the predicted segmentation, nor how one might change a display to obtain a different grouping.

In information graphics, often interesting groupings form between non-physically adjacent items. We want to know, for instance, whether a user will easily perceive the association between colored lines on a plot and the colors in a legend. Will it be obvious to a user that a set of buttons on a remote control perform related functions (Figure 1e)? The vast majority of computer vision algorithms group only contiguous regions. It is unclear how well these algorithms can extend to group over gaps between items. This is a serious problem for applying these algorithms to UI designs. Contour integration algorithms do group across gaps [14, 15, 16, 17], but have not been extended to grouping based upon other Gestalt principles. Techniques that cluster in luminance or color space, for example k-means and non-parametric equivalents [18, 19], will group across gaps in space. However, these techniques, as commonly used, take this loosening of the proximity constraint too far; they will tend to group independently of proximity.

Another difficulty is that typical perceptual organization models do not produce a hierarchical grouping, though see [20, 21, 22, 23, 9]. Something like a hierarchical percept clearly exists [24, 25]. In a text document, for instance, individual letters group to form lines of text, which form paragraphs, which form columns, and so on.

There has been little work comparing models of perceptual grouping with human perception, though for partial attempts see, for example, [26, 27, 28]. In part this is due to the lack of quantitative data on perceptual organization. More seriously, though, many of the algorithms perform too poorly to predict even what qualitative data exist. Predicting a wide range of simple qualitative perceptual organization behavior would be a significant step forward.

A more subtle difficulty with existing models of perceptual grouping is that nearly all handle only one kind of perceptual grouping, e.g. they only find boundaries between regions of natural images [20, 21, 22, 26, 29, 30, 31, 32, 33], perform grouping by proximity [23, 9], segment textures [27, 34, 35, 36, 28], or do contour integration [14, 15, 16, 17]. Certainly if separate models for the different grouping phenomena were required, we would use separate models. However, we will demonstrate that this is not necessary. Advantages of a single, unifying model include the fact that the outputs are compatible, and thus it will be easier to combine them to get an overall picture of the perceptual structure of a display. Furthermore, the difficulty of getting

intuitions about what will group is significantly reduced if there is essentially one model to understand instead of four. The brain itself contains the best perceptual organization system in existence, and it is generally believed to use similar mechanisms to perform related tasks.

More recent models have become more unified largely due to a unified vision of the purpose of perceptual organization. (Although see, for example [37], which captures a number of perceptual organization phenomena based on a unified view of the underlying neural operations.) This vision echoes that of Helmholtz [38], who argued that what we perceive is our mind's best guess of what is in the world, based on summary data (the input image) and prior experience. By this argument, the goal of perception is to create what the computer vision field refers to as a "generative model" for an image: what processes created the image, and where each of the processes operates. Perceptual grouping concerns itself with the latter. Based on this view of perceptual grouping, a number of computational models have been developed for image segmentation [39, 13], edge detection including texture segmentation [28, 26, 36, 40], contour integration [41, 42, 43, 44] and distinguishing figure from ground [45]. Our model similarly finds groups by attempting to infer the processes that generated the display.

THE MODEL
We motivate our model in detail, since the intuitions to be gained apply to design as well.

The Representation is Key
Why is it difficult to predict perceptual groupings? Many perceptual grouping problems appear difficult to explain via simple processing in the image domain. Consider the Gestalt disk array example in Figure 1a. The typical percept in this figure is of an array of disks with three groups: the gray disks on the left, the white disks on the right, and the background. It is this grouping we hope to mimic.

In determining what regions form groups, both proximity, i.e. difference in (x, y), and similarity, i.e. difference in feature (here, luminance) value, are relevant. Virtually any simple algorithm can group the pixels in each individual disk, since they are touching and contain pixels with exactly the same luminance. The difficulty comes from variations in luminance, and bridging the gaps between individual disks to group them in spite of the lack of spatial adjacency. However, spatial information is also important – you do not want to group together all pixels of the same luminance, regardless of how far apart they are.

One possible solution is to blur the image, then check whether this has merged the disks [20, 21, 22, 23, 29, 30, 31]. See Figure 2. This is a relatively old idea in image processing. The blur makes the algorithm somewhat robust to variation in luminance, and also allows grouping across small spatial gaps. This approach does reasonably well at capturing grouping by proximity when all foreground items share the same contrast with the background. However,

when items have different contrasts, as in Figure 1a, this approach performs poorly. Blurring the image not only merges neighboring regions, but also mixes the background luminance with that of the disks, as seen in Figure 2. This mixing makes such methods overly sensitive to the difference between foreground and background. Proximal high contrast items will tend to group with each other, whereas lower contrast items will tend to group with either neighboring high contrast items or with the background. Contrast does influence the spatial extent over which discontiguous items may group. However, algorithms that blur in
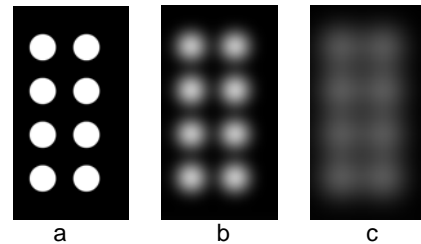

Figure 2: (a) Right side of Figure 1a, at successive levels of blur (b,c). Note that the disks "merge" in (c).

space overestimate this influence, making mid-gray disks on a black background far less likely to group with each other than white disks, and this is not the case.

Consider instead the example of Figure 1c. The typical percept is two smooth curves that cross in the middle. Here the pixels that form groups are contiguous; how do we separate those pixels into two separate curves, given that they touch and are the same color? Again, this is tricky in the image domain. We need to split the image into two contours in spite of the spatial proximity. Where the two curves cross, the local difference in orientation is sufficient to suggest a percept of two contours.

The key in both of these examples is to represent the original image in a higher dimensional space that incorporates both space and feature dimensions. For example, Figure 3 shows the representation of Figure 1a in (x, y, L) space, where L corresponds to a measure of luminance, e.g. the L component of CIEL*a*b [46]. This representation is an "indicator function," with 1's at all points (x, y, L) such that there exists a point in the original image with the given (x, y) location and the given luminance, L. The representation has 0's at all other points in the 3-D space. (Note: in Figure 3 we continue to show elements in their original luminance, to clarify the relationship to the original image.)

This representation in a higher dimensional feature space explicitly incorporates, in a way that the original image did not, the relevance of both spatial proximity and feature similarity. One can imagine that a simple grouping algorithm could easily find the desired perceptual groups. The gray disks lie near each other and not near anything else. One can make a similar observation about the white disks, and about the background. If one blurred *in this higher-dimensional space*, the gray disks would merge to form a
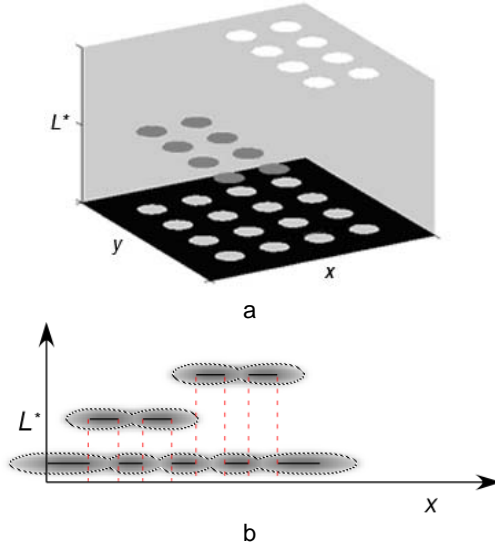
Figure 3: (a) Representation of Figure 1a in *(x, y, L)* space. (b) A slice *y=y₀* through this space (bold lines), superimposed over the same slice, blurred. Hashed black lines show "blob" boundaries, which define 3 groups.

single group, the white disks another, and the background a third. Note also that the predicted groupings will be quite robust to the amount of blur chosen. We have moved from a complicated 2D problem to a seemingly more tractable 3D problem. The right representation can make perceptual organization seem much easier.

A higher-dimensional representation like this makes image smoothing computationally simpler [33], and representation in both spatial coordinates and 3-dimensional color coordinates [e.g. 47, 39] forms the basis for an image segmentation technique known as "mean-shift".

In the case of contour integration (e.g. Figure 1cd), orientation is the key feature. The appropriate representation uses *(x, y, θ)* space, where θ is the local orientation estimate (Figure 4). In *(x, y, θ)* the two curves of Figure 1c do not even come near each other, making it almost impossible to group them in any way other than the desired percept. Again, the right representation can make a perceptual grouping problem much easier. We use steerable filtering [48] to extract the best local orientation at a given scale. The technique described by [49] returns, at each scale, two images, approximately corresponding to $k\cos2\theta$ and $k\sin2\theta$, where *k* indicates the strength or confidence in the orientation. Regions with strong, single orientations yield values of *k* near 1, whereas regions with poor contrast or multiple orientations, such as corners, yield values closer to 0.

We map each pixel to a point *(x, y, θ)*, where θ is the estimated orientation at pixel *(x, y)*. However, rather than placing simply a 1 at point *(x, y, θ)*, we instead put the weight, *k*. This causes our contour integration module to ignore unoriented or weakly oriented regions in the image.
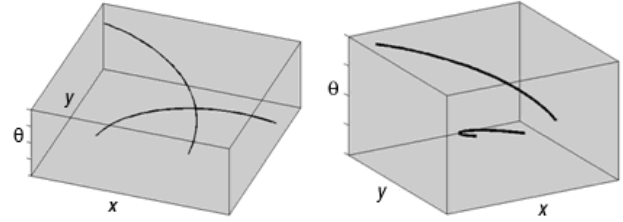

Figure 4: Representation of Figure 1c in *(x, y, θ)* space, from two viewing angles.

A Simple Blurring Merges Regions Into Groups

A good representation is key to making perceptual grouping tractable. Given this representation, how do we decide what groups? Whereas blurring in the image domain has undesirable effects, in the higher-dimensional space, simple blurring performs well at joining separate elements of a perceived group into a single "blob." It does so in a way that mimics human perception. Figure 3b shows a cartoon example of this blur through a slice of the disk array example of Figure 1a. Note that the blur correctly joins the light gray disks into a single "blob", and the same for the dark gray disks and for the background.

Additional blurring is not necessary for finding groups in the curve-crossing example of Figures 1c and 4, but the point of contour integration is often to connect separated contour elements, as in the more complicated contour integration example of Figure 1d. Here blurring in *(x, y, θ)* space helps to join the separated contour elements.

Note that in Figure 3b, we blur in both the feature dimension, *L*, and in the spatial dimension, *x*. One can think of blurring in the feature dimension as joining regions that are sufficiently similar in feature. Features, such as luminance, need to differ by a certain amount in order to be perceived as different by a human observer. If the blur we apply in the model is smaller than this "just noticeable difference," parts of the image whose features are for humans not distinguishable may be incorrectly segmented. This sets lower limits on the amount of feature blur.

Edges of the Blobs Define Perceptual Groups

Blurring leads to blobs in the higher-dimensional representation. Neighboring regions in *(x, y, feature)* space will merge into a single blob, corresponding to a predicted perceptual group. Increasing the blur leads to bigger blobs, and larger-scale grouping of more disparate regions. Therefore, simply changing the blur from small to large produces a "hierarchy" of groupings (this may not produce a strict hierarchy, but as [23] points out, neither does perception).

The last step involves finding the boundaries of coherent blobs, and thus labeling groups. This is a challenge. Perceptual groups tend to correspond to broad, fairly flat regions in the higher-dimensional space. One wants to find the boundaries of those regions while ignoring smaller "bumps" in the blobs. Such smaller, noisy bumps can occur

due to feature variability within a group, and due to discontiguous groups.

Finding meaningful boundaries while ignoring noisy bumps is the classic problem of edge detection in computer vision. A number of existing edge detection algorithms generalize well to our higher-dimensional space. We use a robust version of the Marr & Hildreth [50] edge detection algorithm. Informal testing suggests that this method works well across a range of perceptual grouping examples, and is reasonably robust to choice of parameters.

At this point, it is worth mentioning the relationship of our model to two related suggestions for finding perceptual groups or image segmentations. Logan [51] provided a thumbnail sketch of an algorithm for grouping by similarity and proximity. Though his description differs greatly from ours, he effectively suggested estimating a surface much like our blurred, higher-dimensional representation, which he called the CODE surface, then looking at a slice $L = L_0$. Groups would derive from finding connected components above some threshold, T, within that slice. Logan suggested examining multiple values of L_0 and T to extract a grouping hierarchy. This differs from our proposed model both in looking at only one slice of the *(x, y, feature)* representation at a time, and in that it finds blobs by thresholding rather than edge detection. As a result, his suggestion is far more sensitive to choice of parameters than our model. Furthermore, a given choice of parameters L_0 and T will not tend to select meaningful groups throughout the image. Our method also requires a choice of parameters, but degrades more gracefully, as the necessary threshold for Logan will depend upon the unknown group size, and operating a slice at a time is an unnatural and less reliable way of understanding the full higher-dimensional blob structure.

Also closely related is a computer vision technique for segmenting color images, known as mean-shift [47]. Mean-shift typically represents the image in *(x, y, color)* space, with color typically in CIEL*a*b* color coordinates. Mean-shift then blurs with a separable Gaussian kernel (meaning one can blur first with a 1-D Gaussian in *x*, then, a 1-D blur in *y*, then *L*, and so on). It finds groups by finding peaks of the resulting function, and points associated with each peak. This technique has proven reasonably effective and efficient at segmenting natural scenes [39], but has not been applied to contour integration or perceptual grouping of disjoint regions. Its reliance on finding peaks in the higher dimensional representation makes it highly sensitive to the noisy "bumps" we wish to ignore. In practice, mean-shift has difficulty grouping broad, homogeneous areas into a single group. This is problematic enough in natural images, in which mean-shift often produces very robust over-segmentations that are perceptually invalid. In information visualizations or user interface designs, in which large flat regions abound, and groups may not be spatially contiguous, mean-shift will likely have difficulty deriving perceptually valid groups.

Further Discussion of the Model

Our perceptual grouping algorithm, then, represents the image in a higher dimensional *(x, y, feature)* space, blurs to merge regions into coherent "blobs," and performs edge detection to find meaningful groups. How can this algorithm perform grouping by similarity and proximity, and contour integration? Interpreting the algorithm in terms of the statistical goal of the visual system gives us a clue.

The visual system likely aims to infer a generative model of the image: what processes produced the image, and where they operate. Figure 1a, for instance, can be thought of as generated by one random process, active on the left side of the image, which produced dark gray disks, one random process that produced light gray disks on the right, and one process that produced the dark background. (Very little is "random" about these processes, but the visual system evolved to handle the randomness inherent in natural images, e.g. of trees, leopard skin, grass, etc.) To extract this interpretation of the image, one could gather "samples" from throughout the image, to estimate the distribution of features across space. This distribution is the probability density function (pdf) that "generated" the image. In this case, the pdf has 3 "modes," corresponding to the 3 groups.

Our representation of the image in *(x, y, feature)* space can be thought of as a histogram estimate of the underlying pdf, with each point in the image corresponding to a sample. Blur in the **feature** dimensions parallels a standard statistical technique for better estimating the pdf, known as Parzen windowing. In statistics, choice of the amount of blur is a classic unsolved problem. Too much blur overly smoothes the pdf estimate, which can lead to underestimates of the number of modes. Too little blur leads to a noisy pdf estimate, with too many modes. What the brain may do is try a range of blur, and construct a hierarchy of possible groups, from a fine segmentation based on small amounts of blur through a coarser segmentation based on larger amounts of blur. Perhaps we have a hierarchical grouping percept because the brain is not sure of the best parameters.

Blurring the pdf estimate in the *(x, y)* dimensions is justified based on the prior that processes in natural scenes tend to be localized, so that if one point in a scene came from a given process, that same process likely also generated neighboring points. Blurring in the *(x, y)* dimensions in effect allows us to improve our pdf estimate by collecting more samples, likely to be from the same process, from neighboring areas of the image.

The blur in *(x, y)*, then, should be related to the probability of finding pixels from the same process at a given distance from the current pixel. Our examination of hand-labeled groups in natural images (data from [26]) suggests that blur for grouping by similarity and proximity should be approximately isotropic. Groups in the world do not, by and large, tend to extend more in one direction than another, though there is a slight preference for horizontal and vertical directions. Furthermore, blur should be independent of the local

feature value, since e.g., brighter regions are not, in general, larger than dark regions. We approximate these constraints with a 2-D separable, isotropic Gaussian blur. A different blur may be more appropriate for information graphics, given how they differ from natural scenes. However, it is often assumed that perceptual organization processes in the brain are adapted to processing natural images.

For contour integration (Figures 1c & 4), other priors are more appropriate. Contour segments more likely come from the same process when they are roughly co-circular, with a preference for low curvature "circles", i.e. for collinear segments [41, 14]. Here we assume a simple prior, in which a local orientation estimate of α at point *(x, y)* indicates that we should look for additional samples from the same process in roughly direction α. This means in practice that for each slice $\theta=\alpha$ in *(x, y, θ)* space, we blur with an anisotropic Gaussian oriented at α degrees. We have found that an elongated Gaussian with an aspect ratio of at least 10 works quite well. Though this explicitly prefers collinear contours, in practise it does surprisingly well at joining roughly co-circular contour segments (an observation also recently made by [52]).

RESULTS
In this section, we show results of our algorithm, first on standard Gestalt displays, then on more complicated information graphics. In what follows, we present only the results by the grouping module(s) most applicable to the given display. We discuss future work on combining the results in the Discussion. Each predicted group is represented by a colored blob or blobs superimposed over the original image. These grouping images are best viewed in color.

Grouping By Proximity and Similarity
Figure 5 displays the grouping-by-luminance-similarity predictions on several simple Gestalt disk arrays. In Figure 5a the vertical spacing between disks is significantly less than the horizontal spacing. Because of this, the typical

hierarchical percept is individual disks, which group into 5 columns, and, on a larger scale, simply an array of disks. It is difficult to perceive rows of disks. Our algorithm correctly predicts this result. (The fine scale percept of individual disks is predicted, but not shown.)

Figure 5d shows a similar Gestalt array, in which rows of disks alternate in luminance. The similarity of disks within a row, and the difference across rows, overrides the column percept of grouping by proximity. We perceive rows of disks, and it is difficult to see columns, as predicted by the algorithm. (The coarse scale percept of an array of disks is predicted but not shown.)

Grouping Contours by Good Continuation
Figure 6 displays good continuation predictions on several stimuli. Figure 6a shows a Venn diagram, in which the obvious percept is of two circles overlapping, as predicted by our algorithm. Note that neither top-down knowledge of "circle" is required, nor a preference for closed contours.

Figure 6c shows a standard contour integration stimulus. Figure 6d shows the contours most salient to human observers, from [41]. Figure 6e shows the most salient contours predicted by our algorithm. Overall, the algorithm's performance is quite similar to the expected perceptual organization. In pilot experiments we have tested our algorithm more quantitatively against human performance at detecting, in less than 200 ms, whether the largest contour is on the left or right of fixation. Our model performs quite well at predicting the results of these experiments.

Text and Texture Examples
Textures, i.e. patterns of orientation, size, contrast, etc., are often used to visualize high dimensional data, because of the ease with which the visual system notices similarities and differences in such patterns. Figure 7 shows the results of grouping by similar orientation in a texture. Note that the algorithm correctly predicts the percept.
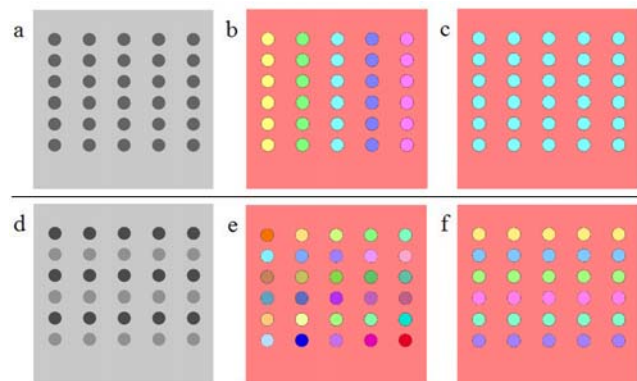

Figure 5: Grouping by luminance similarity. (a) & (d) Original images (500x500 pixels), with dominant column and row percept, respectively. (b) & (e) Results of the algorithm on (a) & (d), respectively. Spatial blur σ = 20 pixels, luminance blur σ_L = 4% of the range of luminance values. (c) & (f) show results of the algorithm for σ = 32 pixels, σ_L = 4%.
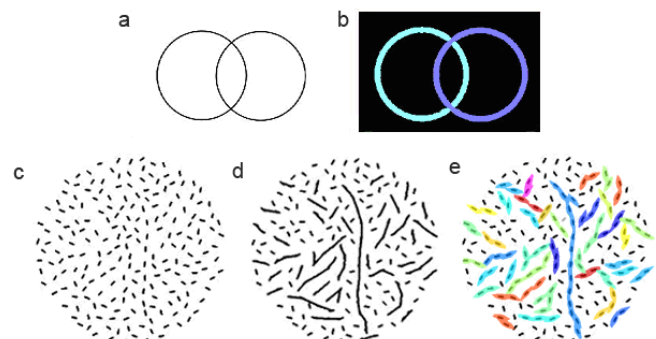

Figure 6: (a) Venn diagram. (b) Contour grouping results on (a). (c) Field of oriented elements demonstrating grouping by good continuation, from Geisler et al (2001). (d) Contours selected by Geisler et al. as being present in (c). (e) Most salient (longest) contours from our contour grouping algorithm.
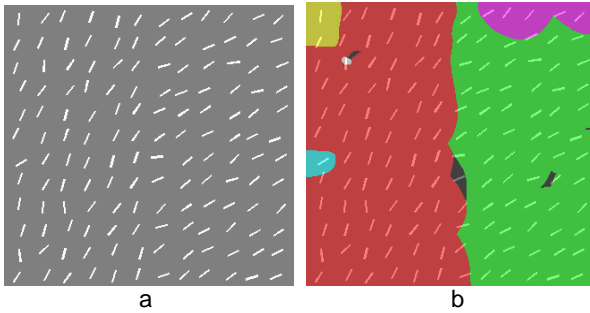
Figure 7: Segmentation results (b) on a texture defined by line orientation (a). Observers can rapidly localize the segmentation boundary in (a) [28].

A special kind of texture for UI design is text. Figure 8 shows our predicted hierarchical segmentation of a simple paragraph of text. As we progress up the hierarchy, we predict the percept of letters, then words, then lines, and finally the entire paragraph along with its header.

Information Graphics

The previous grouping examples are highly relevant to designers, as similar groupings occur in UI designs and information visualizations. Even more relevant, we here test the algorithm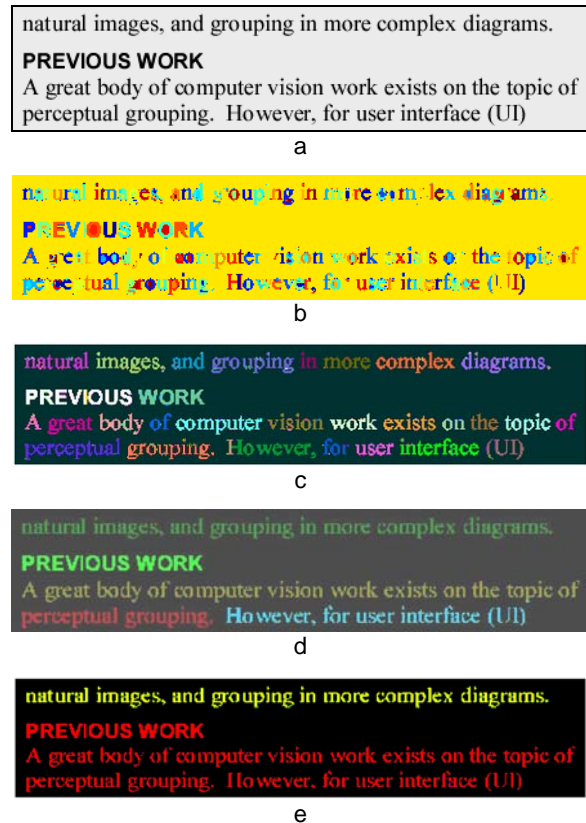 on several information graphics from Tufte [5]. These examples are particularly interesting, since we have Tufte's description of the "true" percept.

Figure 9 shows variations of a plot by Pauling. In Figure 9a one can clearly recognize the shape of the curve, as pointed out in [5] and shown by our algorithm (b). In Figure 9c, the dashed lines have been removed, and it is hard to see the peaks in the graph as the dots are quite scattered. In 9e, Tufte has inserted crosses in the background to facilitate reading of the actual values of the data, at the cost of a modest amount of additional clutter. The curves are still clear, as predicted by the perceptual organization algorithm.

Figure 10 shows a section of Marey's train schedule. The vertical axis represents location, i.e. train stations between Paris and Lyon. The horizontal axis denotes time. Diagonal lines indicate key information: trains traveling from one station to another over time. The version with the low-contrast grid is much easier to read [5]. This is also reflected in the model prediction; for the high-contrast grid, the model finds many grid lines but not all the diagonal lines representing the trains. For the low contrast grid all diagonal lines are found, but not the grid lines. Stops of trains result in offsets of the diagonal lines. However, if they are small, they do not alter grouping into a single "train."
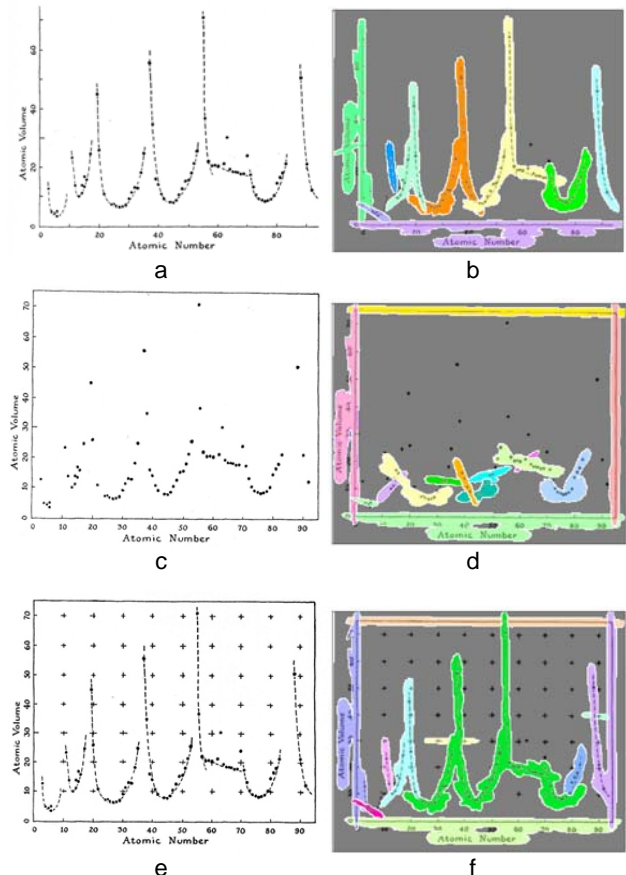

Figure 8: (a) Original text, along with the algorithm's predicted hierarchical grouping of letters (b), words (c), lines (d), and header grouped with lower paragraph (e).

Figure 9: Variations on a graph (a, c, e) from [5], along with the predicted contours (b, d, f, respectively).
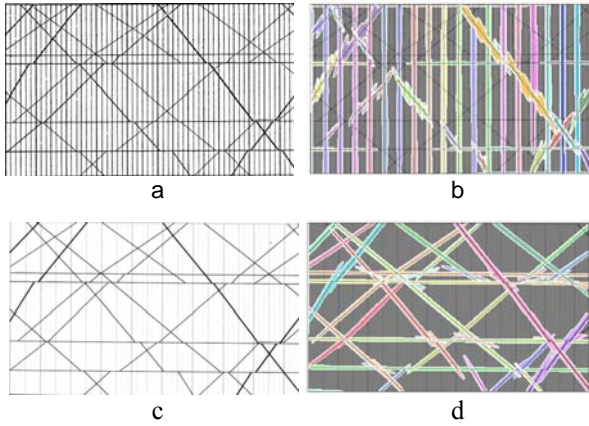
Figure 10: Sections from two variations of Marey's [5] train schedules (a,c) and the contour integration results (b,d).

Figure 11a shows cancer rates among white females, by U.S. county. Figure 11b shows the grouping-by-proximity predicted by our algorithm. The groups found agree well with Tufte's explication of the percept: a cluster of high cancer rates in the Northeast, with additional outlier clusters in southern California and northern Minnesota.

With our model in hand, one could have predicted a number of Tufte's observations on these designs. Clearly this should be useful for designers, as it enables generalization of design rules to arbitrary and complex displays.

Figure 12 shows a final example of an early version of the algorithm that uses color as a feature for grouping by similarity. Buttons of a similar color are predicted to group if they are sufficiently close to each other.
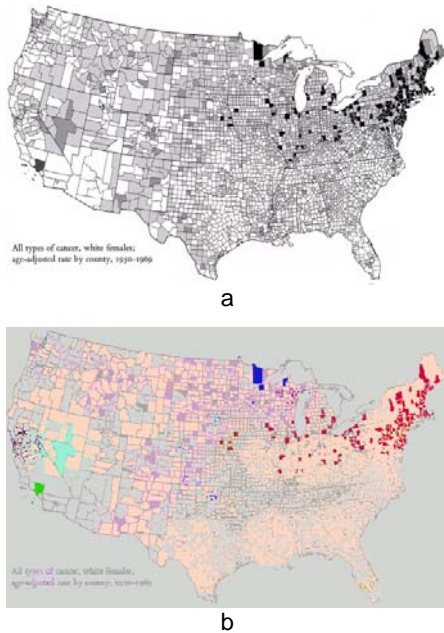

Figure 11: (a) Cancer rates for white females, from [5]. (b) Groupings found by the algorithm. We have adjusted the saturation in the results to emphasize interesting groups.
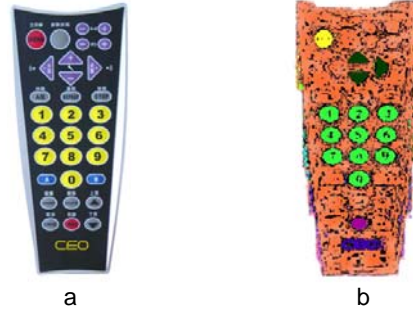

Figure 12: (a) Remote control. (b) Grouping by (color) similarity and proximity.

DISCUSSION & CONCLUSIONS

We have presented a conceptually simple model for perceptual grouping. The key idea is to translate a complicated two-dimensional image, in which segmentation is difficult, into a higher-dimensional representation where straightforward methods yield good results. Our particular technique uses a high-dimensional blur operation, which is simple to implement and understand. We may examine a hierarchy of groupings by varying the degree of blur.

In this work, we have extracted features, e.g. orientation, at a single scale. The visual system is known to extract features at multiple scales, and in future versions we will incorporate this. This is akin to standing farther away from a display, prior to extracting groups. Extracting features at multiple scales is not equivalent to changing the extent of the blur in $(x, y, feature)$ space, though in some cases they may have similar effects on the predicted groupings.

We have focused on a model based on luminance and orientation. A long line of research on the basic features influencing perceptual organization [53], however, suggests many other candidate features.

One issue in integrating multiple features is how to generalize the blur operation. For example, we would clearly like to model the effect of color, but it is not obvious how one should blur in the color dimensions. A unit step anywhere in CIELa*b* space has approximately the same perceptual discriminability, yet it seems unlikely that isotropic, separable filtering in L, a*, b* is optimal, as it implies that the likelihood of grouping two colors together depends only upon their discriminability, not on whether they differ in hue, saturation, or luminance. In the natural world, luminance varies within a given object due to shading, whereas hue changes occur less frequently. Such distinctions may well be present (whether learned or hardwired) in our perceptual system, Thus work on statistics of segmentation of natural images [e.g. 26, 54] may lend insight into how best to blur in color space.

Other likely features include some measure of size or shape. A simple stand-in may be contrast energy, as in [55, 27, 28], though ultimately a more complicated measure of shape may be required. An open question—a generaliza-

tion of the issue raised above with dimensions of color—concerns how these multiple features should be combined. One might, for instance, combine all the features into a feature vector, and find groups using a single high-dimensional representation (x, y, θ, L, a*, b*, …). However, this might depend upon whether such dimensions are "integrable" or "separable," as discussed in [3]. Groupings based on different features might require more complicated combination rules, since [56] shows that when color and geometric form lead to different texture segmentations, color dominates.

A related open question is how one should combine the results from, say, grouping by similarity, with grouping by good continuation. Contour integration may serve to "fix" under-segmentation when features of neighboring items are too similar, and we are exploring using our predicted contours to modulate the results of grouping by similarity.

Finally, some perceived groupings have greater strength than others, meaning they are more likely to be perceived. We are currently exploring, with some success, a measure of grouping strength based upon how stable a group is to changes in algorithm parameters.

Top-down influences are no doubt also important, though our current algorithm does quite a bit without them. We might incorporate top-down effects in our framework by rewarding groups that matched familiar shapes, either in the image or in the higher-dimensional representation.

Our simple algorithm works well at predicting grouping in Gestalt displays, as well as information visualizations like diagrams. This algorithm, and the intuitions associated with it, should be of use to designers wishing to ensure that the structure of the information presented agrees, as near as possible, with the likely perceptual structure of a display.

ACKNOWLEDGMENTS

REFERENCES

1. Wertheimer, M. *Laws of Organization in Perceptual Forms*. Harcourt Brace Jovanovich, London, 1938.

2. Koffka, K. *Principles of Gestalt Psychology*. Hartcourt, New York, 1935.

3. Ware, C. *Information Visualization: Perception for Design*. Elsevier, San Francisco, 2004.

4. Kosslyn, S. M. Understanding charts and graphs. *Applied Cognitive Psychology*, 3 (1989), 185-225.

5. Tufte, E. R. *The Visual Display of Quantitative Information*. Graphics Press, CT, 1983.

6. Cleveland, W. *The Elements of Graphing Data*. Wadsworth, Monterey, CA, 1980.

7. Mackinlay, J. D. Applying a theory of graphical presentation to the graphic design of user interfaces. In *Proc. SIGGRAPH (UIST)* (1988), ACM Press, 179-189.

8. Tullis, T. S. A computer-based tool for evaluating alphanumeric displays. *INTERACT 84* (1984), 719-723.

9. Vincent, L. Current topics in applied morphological image analysis. In Kendall, W. S. et al., eds., *Current Trends in Stochastic Geometry and Its Applications*. Chapman & Hall, 1997.

10. Shneiderman, B., Chimera, R., Jog, N., Stimart, R., & White, D. Evaluating spatial and texture style of displays. In MacDonald, L. W. and Lowe, A. C., eds., *Display Systems: Design and Applications*. John Wiley & Sons, Chichester, U.K., 1997.

11. Saund, E. Symbolic construction of a 2-D scale-space image. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 12, 8 (1990), 817-830.

12. Healey, C.G., Booth, K.S., & Enns, J.T. Harnessing preattentive processes for multivariate data visual-ization. *Proc. Graphics Interface '93* (1993), 107-117.

13. Shi, J. & Malik, J. Normalized cuts and image segmentations. *IEEE Conf. on Computer Vision and Pattern Recognition* (1997), 731-737.

14. Field, D. J., Hayes, A., & Hess, R. F. Contour inte-gration by the human visual system: evidence for a local association field. *Vision Research*, 33 (1993), 173-193.

15. Parent, P. & Zucker, S. W. Trace inference, curvature consistency, and curve detection. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 2, 8 (1989), 823-839.

16. Mumford, D. Elastica and computer vision. In *Algebraic Geometry and Its Applications*. Springer-Verlag, NY, 1994.

17. Grossberg, S. & Mingolla, E. Neural dynamics of per-ceptual grouping: textures, boundaries, and emergent segmentations. *Perception & Psychophysics*, 38 (1985), 141-171.

18. Duda, R. O. & Hart, P. E. *Pattern Classification*. Wiley, NY, 1973.

19. Ohlander, R., Price, K., & Reddy, R. Picture segmen-tation by a recursive region splitting method. *Computer Graphics and Image Processing*, 8 (1978), 313-333.

20. Witkin, A. P. Scale-space filtering. *Proc. 8th Int. Conf. on Artificial Intelligence* (1983), 1019-1022.

21. Koenderink, J. J. The structure of images. *Biological Cybernetics*, 50 (1984), 363-370.

22. Lindeberg, T. *Scale-Space Theory in Computer Vision*. Kluwer Academic, Dordrecht, Netherlands, 1994.

23. Wattenberg, M. & Fisher, D. A model of multi-scale

perceptual organization in information graphics. *Infovis 2003* (2003), 123-133.

24. Palmer, S. E. Hierarchical structure in perceptual representation. *Cognitive Psychology*, 9 (1977), 441-474.

25. Navon, D. Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9 (1977), 353-383.

26. Martin, D., Fowlkes, C., & Malik, J. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26, 5 (2004), 530-549.

27. Malik, J. & Perona, P. A computational model of texture segmentation. *Proc. Computer Vision and Pattern Recognition* (1989), 326-332.

28. Rosenholtz, R. Significantly different textures: A computational model of pre-attentive texture segmentation. *Proc. Europ. Conf. Computer Vision* (2000), 197-211.

29. Perona, P. & Malik, J. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 12 (1990), 629-639.

30. Geman, S. & Geman, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 6 (1984), 721-741.

31. Tomasi, C. & Manduchi, R. Bilateral filtering for gray and color images. *IEEE Int. Conf. on Computer Vision* (1998), 839-846.

32. Durand, F. & Dorsey, J. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans. on Graphics*, 21, 3 (2002), 257-266.

33. Paris, S. & Durand, F. A fast approximation of the bilateral filter using a signal processing approach. *Proc. European Conf. Computer Vision* (2006), 568-580.

34. Julesz, B. A theory of preattentive texture discrimination based on the first order statistics of textons. *Biological Cybernetics*, 41 (1981), 131-138.

35. Beck, J. Textural segmentation, second-order statistics, & textural elements. *Biol. Cybern.*, 48 (1983), 125-130.

36. Voorhees, H. & Poggio, T. Computing texture boundaries from images. *Nature*, 333 (1988), 364-367.

37. Li, Z. Pre-attentive segmentation in the primary visual cortex. *Spatial Vision*, 13 (2000), 25-50.

38. Helmholtz, H. *Handbook of Physiological Optics. Vol. 3, The Perceptions of Vision.* Optical Society of America, Rochester, 1925.

39. Paris, S. & Durand, F. A topological approach to hierarchical segmentation using mean shift. *IEEE Conf. Computer Vision & Pattern Recognition* (2007), 1-8.

40. Ruzon, M. A. & Tomasi, C. Edge, junction and corner detection using color distributions. *IEEE Trans. Pattern Analysis & Mach. Intell.*, 23, 11 (2001), 1281-1295.

41. Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. Edge co-occurrence in natural images predicts con-tour grouping performance. *Vision Research*, 41 (2001), 711-724.

42. Estrada, F. J. & Elder, J. H. Multi-scale contour extraction based on natural image statistics. *IEEE Conf. Computer Vision & Pattern Recognition Workshop* (2006), 183-190.

43. Schinkel-Bielefeld, N. *Contour integration models predicting human behavior*. University of Bremen, 2007. http://nbn-resolving.de/urn:nbn:de:gbv:46diss000-108845.

44. Mahamud, S., Williams, L. R., & Thornber, K. K. Segmentation of multiple salient contours from real images. *IEEE Trans. on Pattern Analysis & Machine Intelligence*, 25 (2003), 433-444.

45. Ren, X., Fowlkes, C., & Malik, J. Figure/ground assignment in natural images. *Proc. European Conf. on Computer Vision* (2006), 614-627.

46. C.I.E. Recommendations on uniform color spaces, color difference equations, psychometric color terms. *Supp. No. 2 to CIE publ. 15 (E.-1.3.1) 1971/(TC-1.3.)* (1978).

47. Comanciu, D. & Meer, P. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 24 (2002), 603-619.

48. Freeman, W. T. & Adelson, E. H. The design and use of steerable filters. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 13 (1991), 891-906.

49. Landy, M. S. & Bergen, J. R. Texture segregation and orientation gradient. *Vis. Research*, 31 (1991), 679-691.

50. Marr, D. and Hildreth, E. C. Theory of edge detection. *Proc. Royal Society, London B*, 207 (1980), 187-217.

51. Logan, G. The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review*, 103, 4 (1996), 603-649.

52. Watt, R., Ledgeway, T., and Dakin, S. C. Families of models for Gabor paths demonstrate the importance of spatial adjacency. *Journal of Vision*, 8, 7 (2008), 1-19. http://journalofvision.org/8/7/23/.