Measuring visual clutter

Ruth Rosenholtz
Department of Brain & Cognitive Sciences, MIT,
Cambridge, MA, USA

Yuanzhen Li
Department of Brain & Cognitive Sciences, MIT,
Cambridge, MA, USA

Lisa Nakano
Department of Brain & Cognitive Sciences, MIT,
Cambridge, MA, USA

Visual clutter concerns designers of user interfaces and information visualizations. This should not surprise visual perception researchers because excess and/or disorganized display items can cause crowding, masking, decreased recognition performance due to occlusion, greater difficulty at both segmenting a scene and performing visual search, and so on. Given a reliable measure of the visual clutter in a display, designers could optimize display clutter. Furthermore, a measure of visual clutter could help generalize models like Guided Search (J. M. Wolfe, 1994) by providing a substitute for "set size" more easily computable on more complex and natural imagery. In this article, we present and test several measures of visual clutter, which operate on arbitrary images as input. The first is a new version of the Feature Congestion measure of visual clutter presented in R. Rosenholtz, Y. Li, S. Mansfield, and Z. Jin (2005). This Feature Congestion measure of visual clutter is based on the analogy that the more cluttered a display or scene is, the more difficult it would be to add a new item that would reliably draw attention. A second measure of visual clutter, Subband Entropy, is based on the notion that clutter is related to the visual information in the display. Finally, we test a third measure, Edge Density, used by M. L. Mack and A. Oliva (2004) as a measure of subjective visual complexity. We explore the use of these measures as stand-ins for set size in visual search models and demonstrate that they correlate well with search performance in complex imagery. This includes the search-in-clutter displays of J. M. Wolfe, A. Oliva, T. S. Horowitz, S. Butcher, and A. Bompas (2002) and Bravo and Farid (2004), as well as new search experiments. An additional experiment suggests that color variability, accounted for by Feature Congestion but not the Edge Density measure or the Subband Entropy measure, does matter for visual clutter.

Keywords: clutter, visual search, set size, Feature Congestion, Subband Entropy

Citation: Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision, 7*(2):17, 1–22, http://journalofvision.org/7/2/17/, doi:10.1167/7.2.17.

Introduction

Clutter is an important phenomenon in our lives and an important consideration in the design of user interfaces and information visualizations. It can interfere with searching for an important item, for example, a threat in a baggage X-ray, a document on our desktop, or a vehicle or pedestrian while driving. Clutter can interfere with quickly and veridically gathering visual information and making decisions. However, we lack a clear understanding of what clutter is; what features, attributes, and factors are relevant; why it presents a problem; and how to identify it. In practical applications, a computational measure of clutter could help either by allowing optimization of the level of clutter in displays over which we have control or by providing system alerts when clutter might impair task performance, for example, when road clutter might impair driving performance.

In addition, a measure of visual clutter would also be useful for basic research in visual search. Most visual search research has used simple displays like that shown in Figure 1A. Researchers interested in search in more complex, naturalistic displays, however, need to be able to perform experiments on images more like that in Figure 1B and to generalize models of visual search to such images. This introduces numerous difficulties. Researchers have begun to address some of these difficulties, for example, by generalizing models of visual saliency to more complex imagery (Itti, Koch, & Niebur, 1998; Rosenholtz & Jin, 2005) and by exploring the influence of top–down information in more natural tasks and images (Torralba, Oliva, Castelhano, & Henderson, 2006). One of the remaining difficulties in generalizing models to more complex and natural images concerns the notion of "set size," that is, the number of "items" in the display. Research on visual search has focused a great deal on reaction time (RT) versus set-size functions as a measure of search difficulty, and set size accounts for a significant proportion of the variance in simple search experiments. Furthermore, models such as Wolfe's (1994) Guided Search use the set size of the display to set a

doi: 10.1167/7.2.17 Received July 15, 2006; published August 6, 2007 ISSN 1534-7362 © ARVO

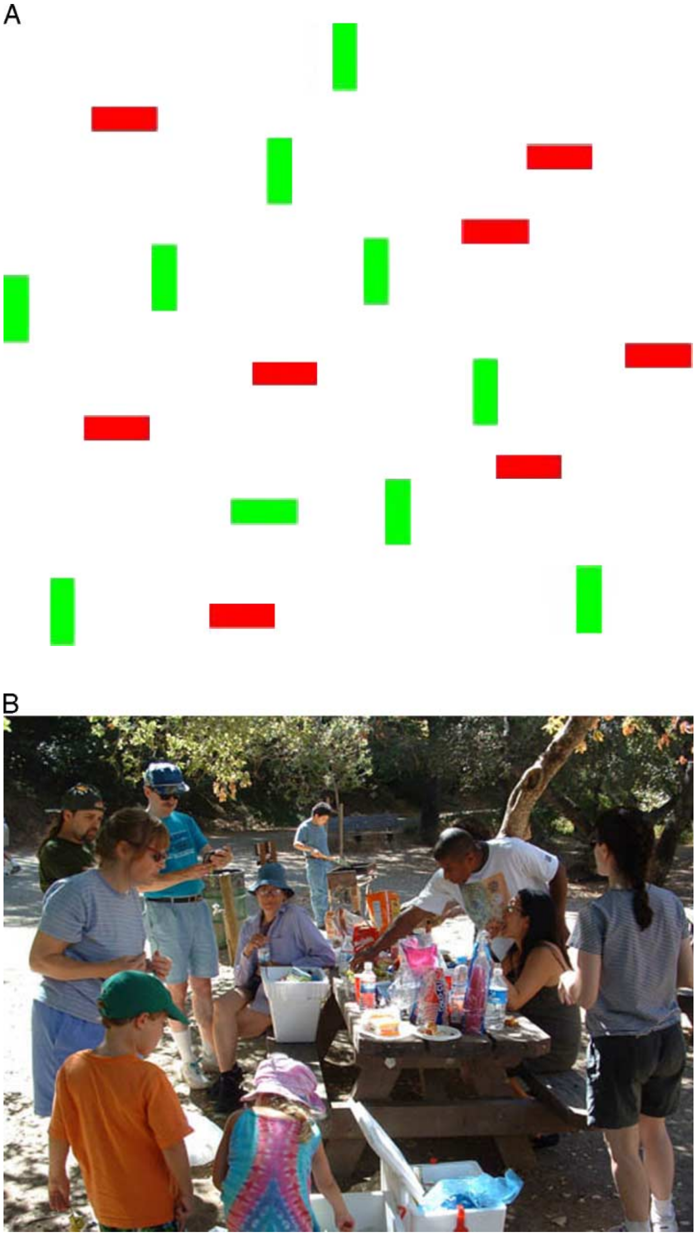


Figure 1. (A) Typical display for visual search experiments versus (B) a more complex, natural image.

criterion for when to stop looking when the observer has not found the target. However, even in complex displays generated by the experimenter, the set size of the display is often unclear. In the map shown in Figure 2, is a mountain range an item or multiple items? Does a single raindrop constitute an item? The outline of a state? In natural images, where the experimenter does not control the display, determining the number of items in the display becomes extremely difficult even given a reasonable definition of what constitutes an "item."

We propose that visual search research requires, for complex imagery, a concept of visual clutter as a stand-in for the standard concept of set size. In the following section, we discuss what one might want from a measure of clutter. Then, we suggest a broad operational definition, followed by more specific candidate definitions, which will allow us to derive several measures of the level of clutter in a display. One such measure, the Feature Congestion measure of clutter, makes use of extensive modeling of what makes items in a display visually salient. Another is based on the notion that visual clutter is related to the amount of visual information in a display. Finally, we test Edge Density measure of clutter, suggested by Mack and Oliva (2004), as a measure of subjective image complexity. In general, these clutter measures correlate well with the influence of a complex background on search performance both in previous search results from other researchers and in our own experiments.

Visual search is a common subtask in many real-world visual tasks. A user must find buttons or other components of a user interface. An alert system must draw attention to a relevant part of a display so that the user can find it easily. Comprehending information visualizations also has a significant visual search component. By having an understanding of how clutter plays a role in visual search, we take a significant step toward understanding the role of clutter in many real-world visual tasks.

Clutter as a stand-in for set size

In substituting a notion of clutter for one of set size, we do not merely want to try to count the number of items in more complex displays. In the first place, this is currently unrealistic, given the state of computer vision algorithms. Furthermore, as mentioned in the previous section, items are ill-defined. What is an item in a natural scene such as that in Figure 1B: Is it a person? A shirt? A cup in a stack of cups? A branch on a tree? Furthermore, Bravo and Farid (2004) have argued that items are not even the unit of relevance, suggesting instead that the number of "parts" is more relevant to search performance.

By analogy with the notion of set size, we should not expect a measure of clutter to predict, by itself, search performance. In standard RT versus set-size performance curves, as in Figure 3, set size interacts with something like target–distractor discriminability to determine search difficulty. (Here, we may not mean target–distractor discriminability in exactly the sense of threshold in a single-item discriminability experiment because we may need to take into account greater search inefficiency in spatial-configuration or conjunction search situations, which may not entirely be captured by traditional discriminability. We use the term "discriminability" here more loosely.) In a simple view of visual search, target–distractor discriminability determines which curve

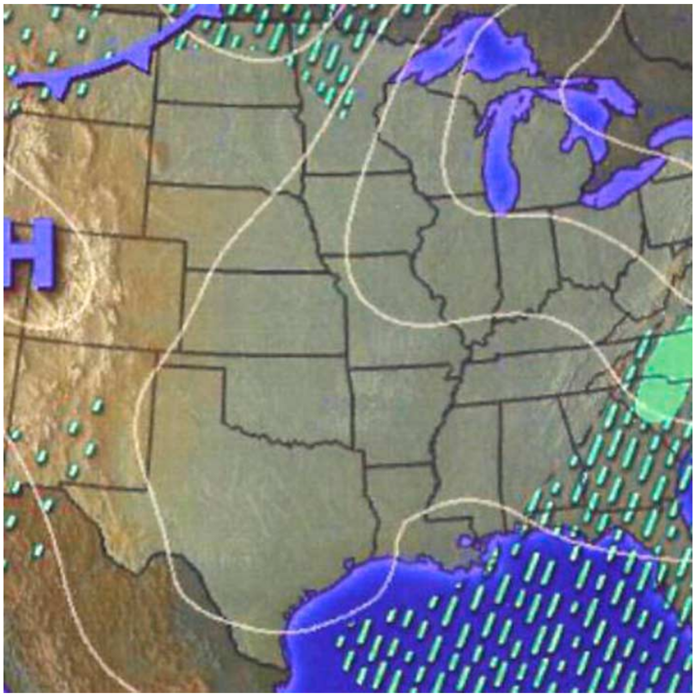


Figure 2. A portion of an information visualization: a map. What counts as an "item" in a display like this?

describes search performance. Set size determines essentially where a stimulus lies on a particular performance curve; that is, it determines the x-value. Together, the two (curve plus x-value) predict the performance, for example, RT. Similarly, we would expect clutter to interact with discriminability to predict search performance in complex imagery.

Additional factors also influence search performance. Of particular importance for the issues of clutter and set size is the effect of top–down information on visual search. When a cue indicates that the target will appear at a subset of possible locations in a display, RTs are a function of the relevant (cued) rather than the nominal set size (Palmer, 1994; Palmer, Ames, & Lindsey, 1993). Features of the target may also cue a subset of potential items as possible targets, again effectively reducing set size. Furthermore, researchers have shown that in complex natural scenes, expertise such as prior knowledge about regions likely to contain a target, such as a pedestrian, can limit eye movements during search to those regions (Torralba et al., 2006). In testing our measure of visual clutter, we propose to initially minimize these top–down effects by focusing on search for categorical targets in situations that minimize prior knowledge about likely target locations. (If one were trying to determine whether set size is relevant for search performance, one would not first perform experiments in which relevant set size is unknown and differed from nominal set size.) Once we have confidence in a measure of visual clutter, such a measure can help us better evaluate experimental results involving more complex stimuli and search tasks, including those with significant top–down components.

Our goals for this article are to derive, implement, and test several initial measures of visual clutter. These measures should be able to operate on arbitrary images, rather than requiring a list of the items in the display and their properties (e.g., "green tree at location (5, 2.6)"). The measures should behave sensibly on standard simple psychophysical displays. Furthermore, they should correlate well with performance in search experiments, at least when one (approximately) controls for target–distractor discriminability and when such experiments have a minimal top–down component to the visual search task.

None of the candidate clutter measures explicitly deal with objects, but they will be a function of the number of objects in the display, as well as of their appearance and organization. Furthermore, the measures may be applied to any static display because they take an image as input and do not require a list of items in the display.

What is clutter?

Clutter is the state in which excess items, or their representation or organization, lead to a degradation of performance at some task. Excess and/or disorganized display items can cause crowding (Stuart & Burian, 1962), masking (Legge & Foley, 1980), decreased object recognition performance due to occlusion, and impaired visual search performance (see Wolfe, 1998, for a review). More items can also stretch or exceed the limits of short-term memory (Miller, 1994). In the case of short-term memory, the relevant factor seems not to be merely the number of objects, but their features (color, orientation, etc.); the

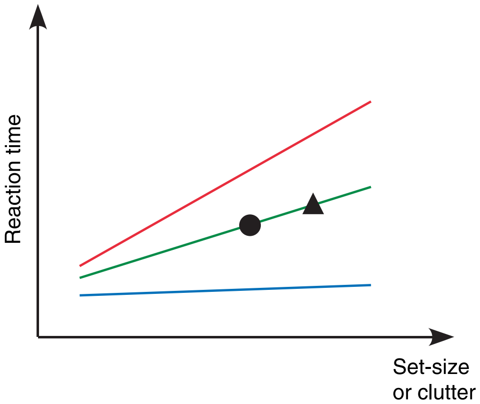


Figure 3. RT versus set-size (or clutter) curves. Set size (clutter) and target–distractor discriminability interact to predict performance. Target–distractor discriminability select which curve describes the RT data; for example, a difficult feature search might put performance on the green curve. Set size (clutter) selects location on this curve (circle or triangle), to predict RT.

capacity for certain simple features is higher than for more complex features (Alvarez & Cavanagh, 2004).

Sometimes, more items lead to performance benefits. This tends to occur when the items are low entropy, meaning that the appearance of one item is easily predicted from its neighbors. Under such conditions, it can be easier, for instance, to spot a trend in data when there are a larger number of points contributing to that trend or to notice a grouping of items with similar characteristics. Given a consistent trend or group of items, more items can aid detection of a deviation—an outlier with features or trend different from the group, or a boundary between two differing groups.

Because clutter is a state when excess items lead to performance degradation, ideally, a measure of visual clutter should be modulated by low-entropy conditions in which additional items may actually lead to performance benefits. It should also capture the various phenomena in which additional or disorganized items degrade performance. However, although all of these phenomena have been studied extensively, many lack adequate models. There exist a number of models of visual search, and in the next section, we first suggest a measure of visual clutter based on such a model.

A measure of visual clutter: Feature congestion

When is a desk cluttered?

In deriving a measure of visual clutter, consider a situation in which one wants to leave a note on a colleague's desk, with the hope that it will draw attention and the colleague will act upon it. If the colleague's desk is uncluttered, it seems easy to place a note such that one is confident that the colleague will notice it. If the colleague's desk is cluttered, one cannot be confident of drawing attention, and often, one leaves a note on an uncluttered chair instead of the desk.

This analogy suggests that the level of visual clutter in a display or scene is related to the ease of adding an attention-drawing target to that display or scene. However, predicting, given a background image, the difficulty of adding an attention-grabbing target is not a usual task for a visual search model. Most such models are instead built to predict the ease of searching for a particular target among particular distractors and would need to be run iteratively to predict the difficulty of adding an item that would draw attention. Luckily, one of the authors (Rosenholtz, 1999, 2001a, 2001b; Rosenholtz & Jin, 2005) has, for some time now, been developing the Statistical Saliency Model. Rosenholtz' model tries to capture human performance at a functional rather than biological level, utilizing the notion that the visual system is designed to characterize various statistical aspects of the visual display. This statistical framework leads to easier intuitions for why a target in a display is or is not salient and can suggest features for a salient target or predict the ease of selecting features to create a salient target. Furthermore, this model has recently been implemented so it can run on arbitrary images (Rosenholtz & Jin, 2005).

In the following subsection, we describe the Statistical Saliency Model for visual search and show how it can predict the ease with which one can add an attention-grabbing target to a display. Then, we will develop our measure of visual clutter and describe its implementation.

The Statistical Saliency Model

Rosenholtz begins with the premise that the visual system has an interest in detecting "unusual" items. She suggests that an item is unusual and, thus, salient if its features are outliers to the local distribution of features in the display. Following a long line of visual search researchers, these features are likely to include such things as contrast, color, orientation, and motion (see Wolfe, 1998, for a review). Rosenholtz suggests a measure like a *z*-score for the degree to which a feature vector, $\mathbf{T}$, is an outlier to the local distribution of feature vectors, represented by their mean, μ_D, and covariance, Σ_D. The saliency, Δ, is given by the following equation:

$$\Delta = \sqrt{(\mathbf{T} - \mu_\mathrm{D})' \Sigma_\mathrm{D}^{-1} (\mathbf{T} - \mu_\mathrm{D})} \tag{1}$$

where $(\mathbf{T} - \mu_\mathrm{D})'$ indicates a vector transpose. The higher the target saliency, the easier the predicted search. The saliency, Δ, can be thought of as a formalization of Duncan and Humphreys' (1989) notion of the different roles of target–distractor versus distractor–distractor similarity in search performance. Rosenholtz's model predicts the results of a wide range of search experiments involving basic features such as those mentioned above (Rosenholtz, 1999, 2001a, 2001b; Rosenholtz, Nagy, & Bell, 2004), including experiments that were previously thought to involve search asymmetries (Rosenholtz, 2001a). More recently, this model has been implemented to run on arbitrary images and shown to be predictive of eye movement data (Rosenholtz & Jin, 2005).

Consider the graphical interpretation in Figure 4. The Statistical Saliency Model essentially represents the local distribution of features by a set of covariance ellipsoids in the appropriate feature space, as shown. The local covariance, Σ_D, specifies the size, aspect ratio, and orientation of the covariance ellipsoids. The innermost, 1σ, ellipsoid indicates feature vectors 1 *SD* away from the mean feature vector, μ_D. The 2σ ellipsoid indicates feature vectors that are 2 *SD* away from the mean, and so on. A target with a feature vector on the $n\sigma$ ellipsoid will have saliency $\Delta = n$. The farther out the target feature

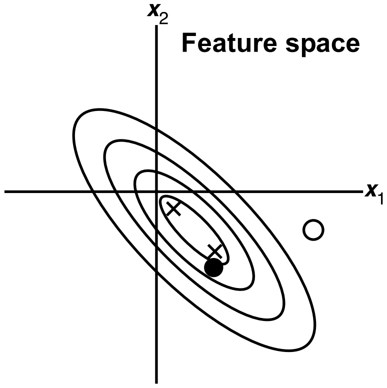


Figure 4. Graphical depiction of the Statistical Saliency Model. *X*s represent the actual local distribution of features. Ellipses represent the mean and covariance of this distribution. Ellipses correspond to points of equal saliency. Outer ellipses correspond to greater saliency and easier search; thus, the model predicts that a target with a feature vector represented by the open circle (○) is easier to search for than a target with a feature vector represented by the closed circle (•).

vector lies on these nested ellipsoids, the easier the predicted search.

One can also use this model to choose attention-drawing features for an item in the display. If a designer wants to add an item to a given portion of the display, such that the saliency of that item is at least *d,* then any features outside of the local $d\sigma$ covariance ellipse will suffice.

The Statistical Saliency Model can indicate the difficulty of adding a new, salient item to a local area of a display. Figure 5 demonstrates this. Figure 5A depicts a number of lines with identical orientation and variable luminance. Figure 5B shows a cartoon of the associated covariance ellipse, assuming an (orientation, luminance) feature space. The display in Figure 5A has high variance in the luminance direction and low variance in the orientation direction (its only variance is due to observation noise). Thus, the ellipsoid has low area, and it would be easy to draw attention to a target simply by giving it a significantly different orientation, as shown in Figure 5C. On the other hand, the display in Figure 5D has high variance in both orientation and luminance, as shown by the covariance ellipse in Figure 5E, which has a relatively high area. It would be difficult to draw attention to a target in this display using only orientation or luminance as features.

The volume of the local covariance ellipsoid represented by Σ_D therefore gives a measure of the local clutter in a display, that is, of the difficulty of adding a new, salient item to a local area of a display. Locally measuring the ellipsoid size and pooling over the relevant display area gives a measure of clutter for the whole display. We call this the Feature Congestion measure of visual clutter. Displays with high clutter, according to this measure, are cluttered because feature space is already "congested" (filled by the covariance ellipsoid), so that there is little room for a new feature to draw attention. Too many colors, sizes, shapes, and/or motions are already clamoring for attention.

The Feature Congestion measure of visual clutter

Our discussion leads us to a surprisingly simple measure of clutter—The clutter in a local part of a display is related to the local variability in certain key features. Implementation of the Feature Congestion clutter measure involves four stages: (1) compute local feature (co)variance at multiple scales and compute the volume of the local covariance ellipsoid, (2) combine clutter across scale, (3) combine clutter across feature types, and (4) pool over space to get a single measure of clutter for each input image.

In the current implementation, we use color, orientation, and luminance contrast as features. Color, luminance contrast, and orientation have been used to model a number of perceptual phenomena, for example, pattern discriminability (Watson, 2000) and preattentive texture segmentation (Malik & Perona, 1990). Color naturally seems important to our sense of clutter. Contrast-energy feature detectors are known to exist early in the visual system and can not only detect simple luminance contrast but also serve as a measure of size and shape (see Rosenholtz, 2000, for a review of evidence that such detectors may mediate preattentive processing of shape in the human visual system). Future implementations might include other features believed to be basic features in visual search and attention, for example, features suggested by the work of Treisman and Gelade (1980).

Below, we briefly describe our implementation. See Rosenholtz (2000) for more details on the feature covariance stage of processing because similar steps are used to segment an image into regions of different texture—a process thought to have much in common with visual search in the human brain.

Feature covariance

The Feature Congestion measure of visual clutter depends upon being able to estimate the perceptual distance between feature vectors. Therefore, we start by converting the input image into the perceptually based CIELab color space (C.I.E., 1978). We then process the image at multiple scales (currently, three) by creating a Gaussian pyramid by alternately smoothing and subsampling the image (Burt & Adelson, 1983).

Next, we find features at each scale. For luminance contrast, we compute a form of "contrast energy" by filtering the luminance band by a center-surround filter formed from the difference of two Gaussians and squaring the outputs. For

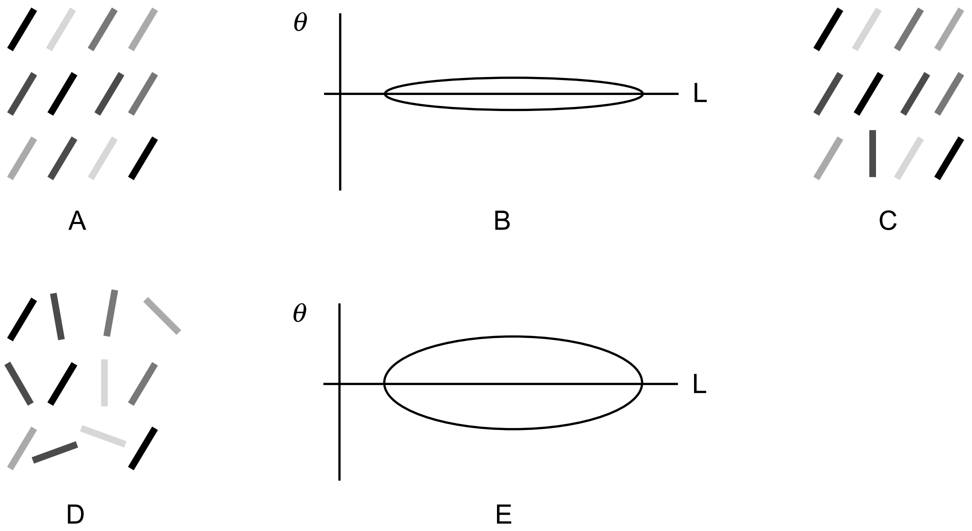


Figure 5. (A) High luminance variability, low orientation variability, as indicated in Panel (B). (C) One easily notices an item with an unusual orientation. (D) High variability in both luminance and orientation, as indicated in Panel (E). It would be difficult to draw attention to an item in this display using only the features of luminance and orientation.

color, we extract a local mean color at each scale by pooling with a Gaussian filter. For orientation, we compute oriented opponent energy, a la Bergen and Landy (1991), which gives us a two-vector, $(k \cos(2\theta), k \sin(2\theta))$, at each image location and scale, where θ is the local orientation and k is related to the extent to which there is a single strong orientation at the given scale and location.

Then, for each feature, we compute the local (co) variance for each feature. This may be done efficiently, and in a biologically plausible way, through a combination of linear filtering (to average over a local area) and point-wise nonlinear operations (to compute the variance). From a covariance matrix, it is straightforward to compute the volume (area) of the covariance ellipsoid, our local measure of color (orientation) feature congestion, that is, clutter. The contrast feature congestion is simply the square root of the contrast variance.

Combine across scales

For each feature, we combine feature congestion across scale by taking the maximum at each pixel. We reason that a feature is locally congested if it is congested at any scale. Little has been done to examine the interaction of multiple scales in determining target saliency or the influence of clutter, and more basic research needs to be done to have a better understanding of how information combines across scale.

Combine across features

At this point, we have three clutter maps for the image, representing the “color congestion,” “texture congestion,” and “orientation congestion.”

Next, we combine color, contrast, and orientation clutter at each point. We first take the cube root of color clutter (a volume) and the square root of orientation clutter (an area) to make these more comparable to contrast energy clutter (a scalar). Even so, the three clutter measures are not scaled equivalently. The true measure of how congested a feature space is is how much of feature space is taken up by the covariance ellipsoid relative to how much feature space is available. Therefore, it is appropriate to scale the clutter value in each feature dimension by essentially the range of possible clutter values for that feature. Currently, we approximate that range by normalizing by the standard deviation of clutter values for a given feature over a wide range of input images. We then combine the scaled color, contrast, and orientation clutter at each point by taking their sum.

A final model of clutter will almost certainly involve a more complicated combination rule than that used here. The features might, for instance, be combined into a single large feature vector prior to computation of the covariance, as might be suggested by the search results and modelling of Eckstein, Thomas, Palmer, and Shimozaki (2000). Some features might have priority over other features; for example, Callaghan (1989) has suggested that color dominates over geometric form in texture segregation. Much basic research needs to be done to adequately model feature interaction. In the absence of such research, our aim was to see how far we could go with a simple measure. Early attempts to allow a general linear combination of clutter across features did not greatly improve performance of this clutter measure.

Combine across space

Finally, we pool over space to get a single measure of clutter for each display by taking the average clutter value over the entire image.

The Subband Entropy measure of visual clutter

For a given number of objects in a scene, the scene will appear less cluttered the more "organized" it is. Organization may involve grouping similar objects together; aligning them; and making many of the objects a similar hue, luminance, size, and so forth. This principle is well known in the literature on decluttering one's home, and research has shown that perceptual organization of this sort affects search performance (Treisman, 1982) and other visual tasks. The degree of organization of a scene can be thought of in terms of the extent to which each part of the scene is predictable from the rest of the scene, or in terms of the amount of redundancy in the scene. The Feature Congestion measure of clutter captures this concept of "organization" to some extent implicitly; by looking at feature covariance, it essentially captures some measure of the grouping by similarity + proximity in the display.

To the extent that an image contains redundancy, it can be represented, either in the brain or in a computer, with a more efficient code. If, for example, a region of the image forms a single homogeneous group, then that region can be encoded by noting its group characteristics and location instead of encoding each point within that region. Similarly, if a number of objects repeat regularly, they can be efficiently encoded by representing the object and its repetition pattern. Therefore, the less cluttered an image is, the more it is redundant and the more efficiently it can be encoded.

We propose a second measure of clutter based on attempts to measure the efficiency with which the image can be encoded while maintaining perceptual image quality. For this purpose, one would ideally like an encoder that detects the sorts of redundancies exploited by the visual system. Both low-level visual processing and mid-level visual processing seem aimed at capturing redundancies in natural images. Early filtering by V1 receptive fields can be thought of as capturing low-level redundancies in the input. Researchers have suggested that a way to achieve efficient coding of natural scenes is to choose early filtering operations that maximize sparse representation and that this principle leads to receptive fields like that in V1 (Olshausen & Field, 1996). Mid-level perceptual organization captures higher level redundancies in images by encoding groupings, symmetry, and so on.

However, although one would ideally like to know how efficiently one can encode an image by making use of the same sorts of redundancies as the visual system, this is difficult to do given the current state of understanding and modeling of perceptual organization and mid-level processing. Therefore, for our purposes, we initially suggest using encoding efficiency of current, highly successful subband image coding methods such as JPEG 2000. Our Subband Entropy clutter measure is based on the notion that clutter is related to the number of bits required for subband (wavelet) image coding. A wavelet coder first decomposes the image into a set of subbands with different orientations and spatial frequencies, which is analogous to the decomposition that occurs early in human vision. Thus, it captures some of the same redundancy in images as early vision. We use steerable pyramids as the basis for our Subband Entropy measure (Simoncelli & Freeman, 1995).

A number of effective image coders such as JPEG and JPEG 2000 follow a subband transform of the image with entropy encoding. Similarly, in our Subband Entropy measure of clutter, we next compute the Shannon entropy within each subband. Shannon entropy is defined as

$$\sum_i -p_i \log(p_i). \tag{2}$$

Here, p is the probability distribution of coefficients in each subband and is estimated by binning (i.e., quantizing) the subband coefficients into bins indexed by i and computing a histogram. This Shannon entropy essentially captures the bits required to encode the subband, for a given level of fidelity, as specified by the coarseness of the bins (quantization). Higher fidelity, that is, finer bins, requires more bits to encode. In our computations, the number of bins is equal to the square root of the number of coefficients, meaning that bands with fewer coefficients also have, on average, fewer coefficients per bin. This implicitly says that it is more important to faithfully reproduce lower frequencies than high frequencies; at lower frequencies, a wavelet transform has fewer coefficients, and thus, this strategy leads to finer bins, more bits required, and more faithful encoding.

The clutter measure is computed as a sum of these subband entropies.

The algorithm is as follows:

1. Convert the RGB image into CIELab.
2. Decompose the luminance (L) and the chrominance (a, b) into wavelet subbands using a steerable pyramid.
3. Bin the wavelet coefficients within each subband and compute the Shannon entropy within each subband according to Equation 2.

4. Sum the subband entropies for the luminance (L) and for the chrominance channels (a, b).
5. Compute a weighted sum of chrominance and luminance entropies. We used a weighting of 0.08 for each of the chrominance channels and 0.84 for the luminance channel. Image encoders typically use fewer bits for chrominance than luminance channels because chrominance need not be coded with as much fidelity, without compromising image quality. The Subband Entropy measure, however, is not very sensitive to changes in this weighting—a chrominance weighting of 0.22 to a luminance weighting of 0.56 gave nearly identical results on all examples in this article.

Previous measures of "clutter"

There have been several past attempts to quantify something like the "clutter" in a display. Researchers in the field of information visualization have attempted to measure what they call "information density." In information visualization, one often has decisions to make about how much information to present to the user in a given display or part of a display. The hope has been that with a measure of the information density, designers would be able to choose the optimal level of information to present, trading off giving the user more information against making it more difficult for the user to quickly and veridically extract that information from the display.

In this context, Woodruff, Landay, and Stonebraker (1998) experimented with several measures of information density: the number of visible objects and the number of vertices. Other researchers have suggested additional measures of information density or clutter. Dynamic Logic's (2001) study on the effect of clutter on the effectiveness of Web advertisements used the number elements on a Web page as a metric, where an element consisted of a word, graphic, or "interest area." Tufte (1983) suggested the number of entries in the source data matrix—that is, the table of data used to generate the visualization—per unit area. Nickerson (1994) enumerated a number of density measures, including the number of graphic tokens per unit area, the number of vectors needed to draw the visualization, and the length of program to generate the visualization. Frank and Timpf (1994) suggested the amount of "ink" per unit area as a metric for simple black and white maps.

These metrics have a number of difficulties, and few have actually been implemented. Certainly, the amount of clutter has some dependency upon the number of objects, graphic tokens, or entries in the source data matrix in the display. The number of objects (or, conversely, the amount of blank space) has been shown to influence task performance at both a directory assistance search task (Springer, 1987) and at a number of tasks with a map (Phillips & Noyes, 1982). However, column alignment helped in the directory assistance task, and Phillip and Noyes (1982) found that "like clutters like"; for example, additional lines on a map cause increased difficulty for tasks involving lines. Counting the number of objects does not take into account the appearance or organization of the objects. Merely counting the number of elements on a Web page did not prove to be a good measure of clutter in the sense of correlating with advertisement effectiveness (Dynamic Logic, 2001), although human clutter judgments did correlate with advertisement effectiveness. Furthermore, as mentioned above, for complex visualizations, the number of objects can be ill-defined. A vertex may be similarly ill-defined—How sharp a turn in a curve is required? The amount of ink is ill-defined for color displays. In addition to these difficulties in applying suggested clutter measures when we are given a list of drawing objects in a visualization, the situation becomes even bleaker if one wants to measure the clutter in an image, for example, in a natural scene.

Oliva, Mack, Shrestha, and Peeper (2004) have attempted to determine what factors influence the human representation of "complexity," a concept clearly related to that of clutter. They had users hierarchically sort photographs according to their complexity and indicate at each hierarchical level the basis for the sort. Users indicated that complexity depended on the quantity and variety of objects, detail, and color, as well as on higher level, more global concepts like the symmetry, organization, and "openness" of the depicted space. Our Feature Congestion and Subband Entropy measures of visual clutter, as we will see, correlate with quantity and variety of objects, although we do not explicitly find objects. Our Feature Congestion measure explicitly incorporates a notion of the variety of color in an image, and the Subband Entropy measure also captures this, because more color variability means less predictability, and thus, more information = entropy = clutter. Both measures will implicitly deal with certain aspects of perceptual organization, such as grouping by a combination of proximity and similarity. In a later version of the Feature Congestion measure, we are interested in incorporating some of the higher level components of perceptual organization that Oliva et al. suggest. Mack and Oliva (2004) have implemented early versions of several measures of complexity: quantity of contours, degree of symmetry, global color variability, and degree of openness. (We argue that local color variability is more appropriate because it implicitly responds to local groupings of color. Local variability has been shown to affect search performance; e.g., Nothdurft, 1993.) They compared these measures to the mean subjective rankings of complexity on indoor scenes and found a correlation of $r = .85$. Edge density—the percentage of pixels that are edge pixels—-alone led to a correlation with mean subjective rankings of $r = .83$. (Note, however, that the mean Spearman rankorder correlation between subjects was only $r = .61$, comparable with what we have found for subjective judgments of clutter in maps; Rosenholtz, Li, Mansfield, & Jin, 2005; $r = .70$. In that article, we found a correlation of $r = .77$ between median subjective judgments of clutter in maps and an earlier version of the Feature Congestion clutter measure. None of these differences in correlation coefficients is significant, $p > .05$.) This high correlation between subjective judgments of complexity in indoor scenes and such a simple measure as edge density suggests that this simple measure is worth examining further. In what follows, we also examine the performance of an Edge Density measure of visual clutter. To obtain the Edge Density measure for each image, we applied MATLAB's Canny edge detector to each image and measured the density of edge pixels. The Canny edge detector has several parameters: a low threshold, high threshold, and sigma. These parameters were set by hand to values that gave good results overall to the examples presented in this article. The low threshold and high threshold are used to find weak and strong edges, respectively, and the Canny edge detector keeps weak edges only if they are connected to strong edges. These thresholds were set to 0.11 and 0.27, respectively. The sigma parameter is the standard deviation of the Gaussian filter used in the computation of the gradient. It was set to the default, $\sigma = 1$, comparable with the finest scale in the Feature Congestion and Subband Entropy measures.

Evaluation of measures of visual clutter

Our goal in this article is to design and test several measures of visual clutter: Feature Congestion, Subband Entropy (refer to http://hdl.handle.net/1721.1/37593 where the two clutters were implemented), and Edge Density. Because a good part of our interest in such measures is to have a stand-in for set size in models of visual search, we need to test whether these measures correlate well with performance in visual search tasks. In particular, we should hope and expect them to perform well in tasks involving minimal top–down information about the appearance and location of the target because such information could lead to the equivalent of a "relevant set size or clutter" as opposed to straight set size or clutter. The influence of top–down information on visual clutter could be included in a later implementation of any of these measures but is not included at present. In Experiment 1, we test the three measures against RT in a visual search task involving searching for a category target in map images and show that the measures perform well. Reaction times, however, are governed by a number of complicated factors. In Experiment 2, we instead use limited display times and find contrast thresholds necessary to correctly identify the target at a given level of accuracy. We show that the clutter measures also correlate well with this more traditional psychophysical measure of performance. Next, we demonstrate that the clutter measures give sensible answers on standard simple search displays. Then, we test these measures against previous results of search in clutter, predominantly limiting ourselves again to those previous results with minimal top–down components. Experiment 3 is aimed at distinguishing between these measures through more careful selection of displays and, in particular, at examining whether color variance is an important aspect of visual clutter.

Experiment 1: Visual search in cluttered maps

In our first experiment, we examine the influence of visual clutter, as determined by our various measures, on RT in a visual search experiment.

In many visual search experiments, an observer searches for only one target per "scene," and only once. Although studied little in human vision, many real-world scenarios involve repeated search in the same scene for a number of different targets. We look for our keys, then our gloves, then our computer bag, and so on.

When an observer searches only once in each scene, the target saliency is a key factor determining performance. In a simple scene, a salient target can save the observer as much as a few hundred milliseconds. On the other hand, in repeated-search scenarios, observers cumulatively spend a lot of time interacting with each scene. As a result, properties of the scene that enable or impair search performance become important. This is why some people spend so much time trying to organize and declutter their offices, homes, and other scenes over which they have control and within which they do many different tasks. A cluttered scene can cost us, over time, far more than a few hundred milliseconds.

Due to our interest in more real-world search scenarios and in clutter, in particular, we use a repeated search task to study the extent to which the background scene within which the observer searches influences search performance. To focus on bottom–up influences such as clutter, as opposed to top–down influences such as priors on likely target location, the categorical targets can appear with equal probability in any of six isoeccentric locations. Rather than looking at performance at searching for a particular target in a particular location, which will tend to be strongly influenced by factors such as target saliency, we examine how, on average, across targets and locations, the background scene affects search performance.

Stimuli

Observers searched for a categorical target—a Gabor—in a number of geographic maps. Although Gabors do not typically appear in geographic maps, this choice provided several advantages. First, we could, in principle, have had a more natural task in which observers search for targets already existing in the maps. But first, symbols already present in a map can hardly be expected to appear all at the same eccentricity, thus adding another source of variance to the data. Furthermore, searching for symbols already present in the maps would have been a more difficult experiment to set up because all the maps contain different symbols, and we would need to both determine which symbols appeared in which maps and inform the observer of a different target on each trial. We wanted observers to search for a variety of targets, not already present in the maps, over which we had a fair degree of control over their appearance, while keeping instructions to observers simple by having them search for a categorical target. We chose targets consisting of 16 grayscale Gabors. (Earlier pilot experiments had observers search for a bull's-eye target, which one might think of as a "you are here" symbol, and the results in terms of correlation with various measures of visual clutter were virtually identical to those presented here.) The wavelength of the carrier sine wave was always twice the standard deviation of the envelope. The Gabors varied in scale (envelope σ = 0.55° or 0.83°), orientation (45°, 90°, 135°, or 180°), and phase (sine or cosine). Before the experiment, observers were shown a number of examples of Gabors, and each observer received a training block with feedback to ensure that he or she understood the task and nature of the targets. Each target appeared exactly once in each of the six general locations of the image. The exact location of the target was determined by the superposition of one of six locations in the image and a small random position jitter of up to 0.75° in both x and y directions. The prejitter target locations were isoeccentric at approximately 7.6° from the initial fixation.

Nineteen colored map images served as the background against which the targets were superimposed. The targets were added as a semitransparent layer on top of the maps. The map images were found using a random search for maps on the Web to test the three candidate clutter measures on typical maps. The maps were chosen so that 8 were at the scale of showing an entire country, 3 were at the scale that showed several cities, 4 showed a large portion of a city, and 3 showed only a few streets. Again, this was done to get a sampling of typical existing maps. The maps were cropped to be uniform in size, to avoid effects in which search might take longer simply because there was more area to search, and spanned about 24° of visual angle when viewed during the experiment from 15 in. The maps were corrected to have a mean luminance of mid-level gray and to have approximately the same mean color (also a mid-level gray) for all. In principle, this was done to minimize a possible source of variance in which search in one map might be easier or harder merely because the map was darker or closer to the target (gray) color than another image. In practice, a pilot experiment suggests that this manipulation had virtually no effect on the fits of the clutter measures to the experimental data. The mean luminance of the Gabor target was always set to the local mean luminance of the map, again to minimally control for local contrast of the target, which might affect search performance. Figure 6 gives an example of a map with superimposed Gabor target.

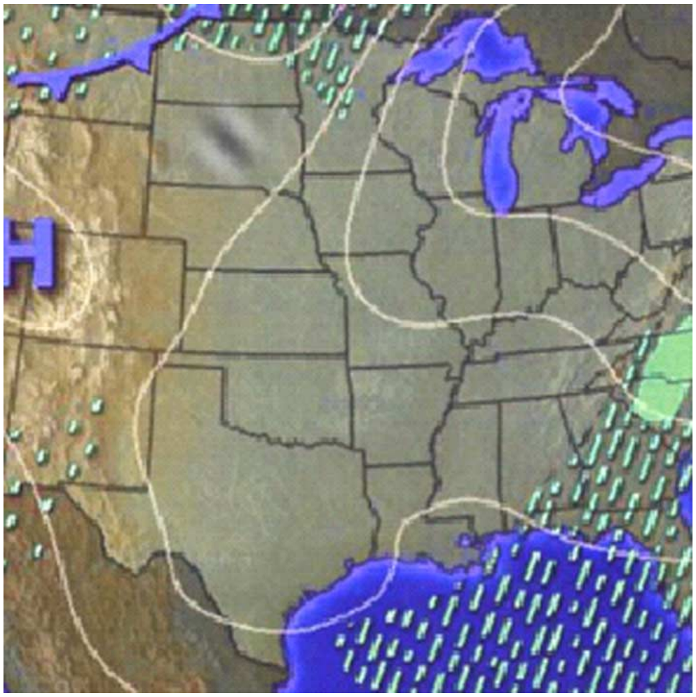


Figure 6. Example of background image with superimposed Gabor target.

Methods

Each observer was seated in front of a computer display. Observers were instructed to search for a Gabor target against each background image and to respond as quickly and accurately as possible as to whether a target was present or absent. The image was displayed until subjects indicated their response by pressing "f" for target absent or "j" for target present. After each trial, visual feedback was given to indicate whether the response was correct. After the feedback, a fixation-cross appeared in the center of the display for 500 ms before the next trial began. A total of 3,648 trials (16 targets × 19 images × 6 locations × target absent/present) were divided evenly among 12 blocks. We ran our experiments in MATLAB, using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Six subjects participated in Experiment 1. All observers had normal or corrected-to-normal vision and were compensated for their time.

Results

We are interested in how the clutter of an image correlates with search performance within that image. As a result, in what follows (both for this experiment and later experiments), we average performance across target locations within an image and average across subjects. Error bars (which show standard error) on the plots give an indication of the variability across subjects and target locations. We report correlation coefficients between clutter and mean performance, essentially asking what percentage of the variance in mean performance is accounted for by a given clutter measure.

All clutter measures were computed on the background images before adding the targets. Figure 7A shows mean log(RT) versus the Feature Congestion clutter measure. Figure 7B shows the results for the Subband Entropy measure. Figure 7C shows the results for the Edge Density measure. Each data point represents 1 of the 19 background images. We average RT from correct trials over target location, Gabor type, and subject to get a single measure of RT per image for both target-absent and target-present trials. We find a significant correlation between the average log(RT) and each of the clutter measures: Feature Congestion (r = .74, target present; r = .76, target absent; $p < .001$), Subband Entropy (r = .75, target present; r = .77, target absent; $p < .001$), and Edge Density (r = .83, target present and target absent; $p < .001$). All three measures of visual clutter do a good job of predicting the effect of the background image—that is, the effect of the display clutter—on search performance. None

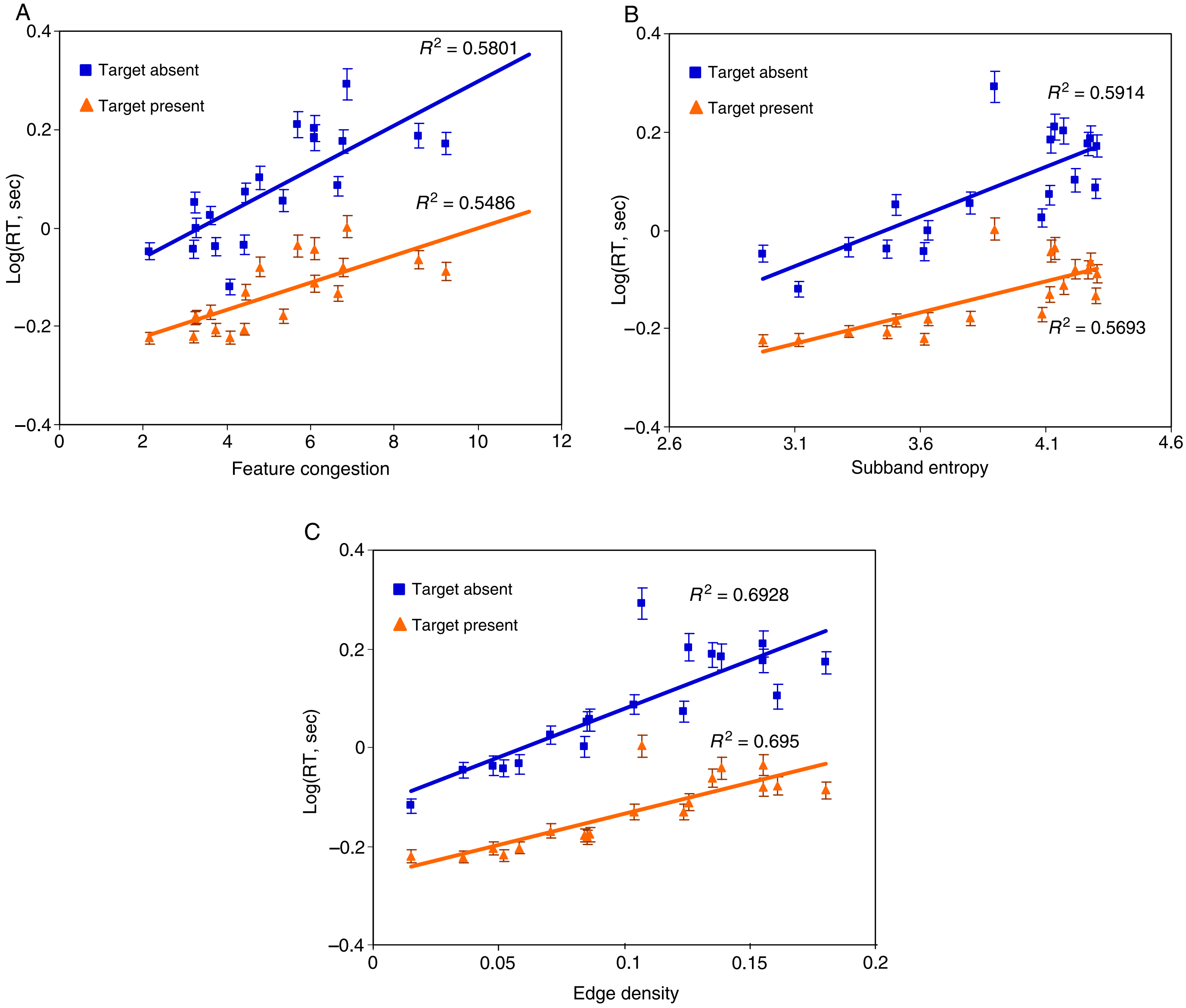


Figure 7. Mean log(RT) versus clutter measures, plotted with 95% confidence intervals. (A) Log(RT) versus Feature Congestion clutter, (B) log(RT) versus Subband Entropy clutter, and (C) log(RT) versus Edge Density clutter.

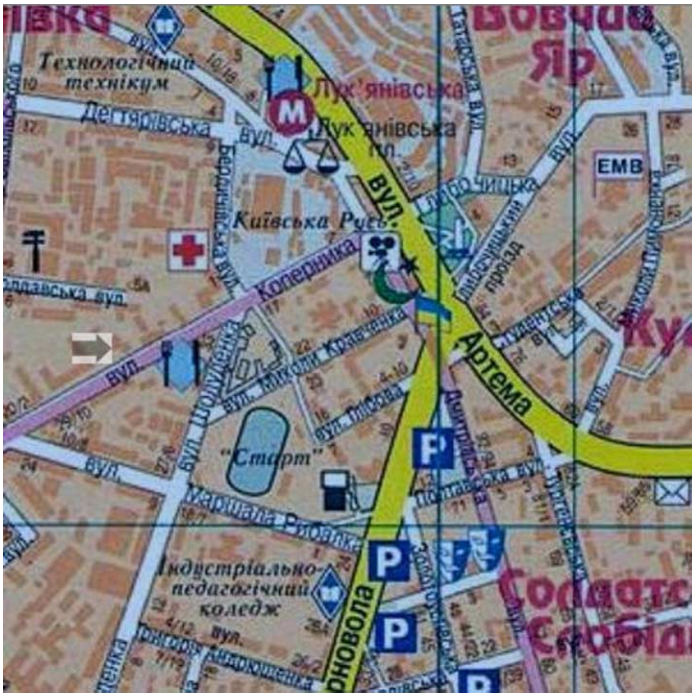


Figure 8. Example of map image with superimposed arrow target (center left).

of these correlation coefficients differ significantly from each other ($p > .05$).

Experiment 2: Contrast thresholds for visual search in maps

In this experiment, we compare the outputs of the three clutter measures to a more traditional psychophysical measure of performance: contrast thresholds necessary for a given level of performance at a visual search task. In this, as in later experiments, unless otherwise specified, the methods and stimuli are as in Experiment 1.

Stimuli

Observers searched for an arrow symbol in a number of geographic maps. The arrows were dark against a light rectangular background, so that their average color was mid-level gray. Each arrow pointed either to the left or to the right. Arrows were approximately 1.2° in length. Figure 8 shows an example of an arrow in a map.

Twenty colored map images served as the background images against which the targets were superimposed. Fifteen of these maps were the same as in Experiment 1. Four of those maps were removed from this experiment because they already contained arrows, and 5 maps of approximately the same clutter were added in their place. The arrows were again added transparently to the background images; this meant that at 0 contrast, the arrows were invisible.

Methods

An arrow appeared in each display at one of six possible target locations. Target locations were all at the same 7.6° of eccentricity. Search displays appeared for 1 s, followed by a random noise mask. Observers indicated via a key press ("d" or "k") whether the arrow pointed to the left or to the right, respectively. Target contrast was varied by a Robbins and Monro (1951) stochastic approximation staircase (see Treutwein, 1995, for a review), to arrive at a 75% correct threshold. We approximate target contrast as $(L_{\text{max}} - L_{\text{min}})/L_{\text{mean}}$, where L_{mean} is taken over a local neighborhood of the target. It is unclear what the best measure of target contrast is for a target placed in a complex and nonstationary environment. However, by the above measure of contrast, thresholds were relatively consistent from location to location within a given image. A different staircase was used for each target location in each image.

Four experienced psychophysical subjects participated in the experiment, including two of the authors. All observers had normal or corrected-to-normal vision.

Results

Contrast thresholds were fairly similar from location to location within an image, so we average thresholds within an image. Figure 9 shows the mean threshold target contrast for each image, averaged over target location and subject. Figure 9A shows the mean threshold target contrast versus the Feature Congestion clutter measure. Figure 9B shows the results for the Subband Entropy measure. Figure 9C shows the results for the Edge Density measure. Again, mean contrast threshold is significantly correlated with all clutter measures ($p < .001$): Feature Congestion, $r = .93$; Subband Entropy, $r = .68$; and Edge Density, $r = .83$. The Feature Congestion measure is significantly better than Subband Entropy ($p < .05$), but Edge Density is not significantly different from either Feature Congestion or Subband Entropy.

Performance of the Feature Congestion measure of visual clutter on standard simple visual search displays

If clutter is to act as a stand-in for set size, ideally, it should do something sensible for simple psychophysical displays. In particular, we would like a clutter measure to monotonically increase with nominal set size. We ran the three clutter measures on feature search displays (red disk target among green disks), *T* versus *L*, and conjunction

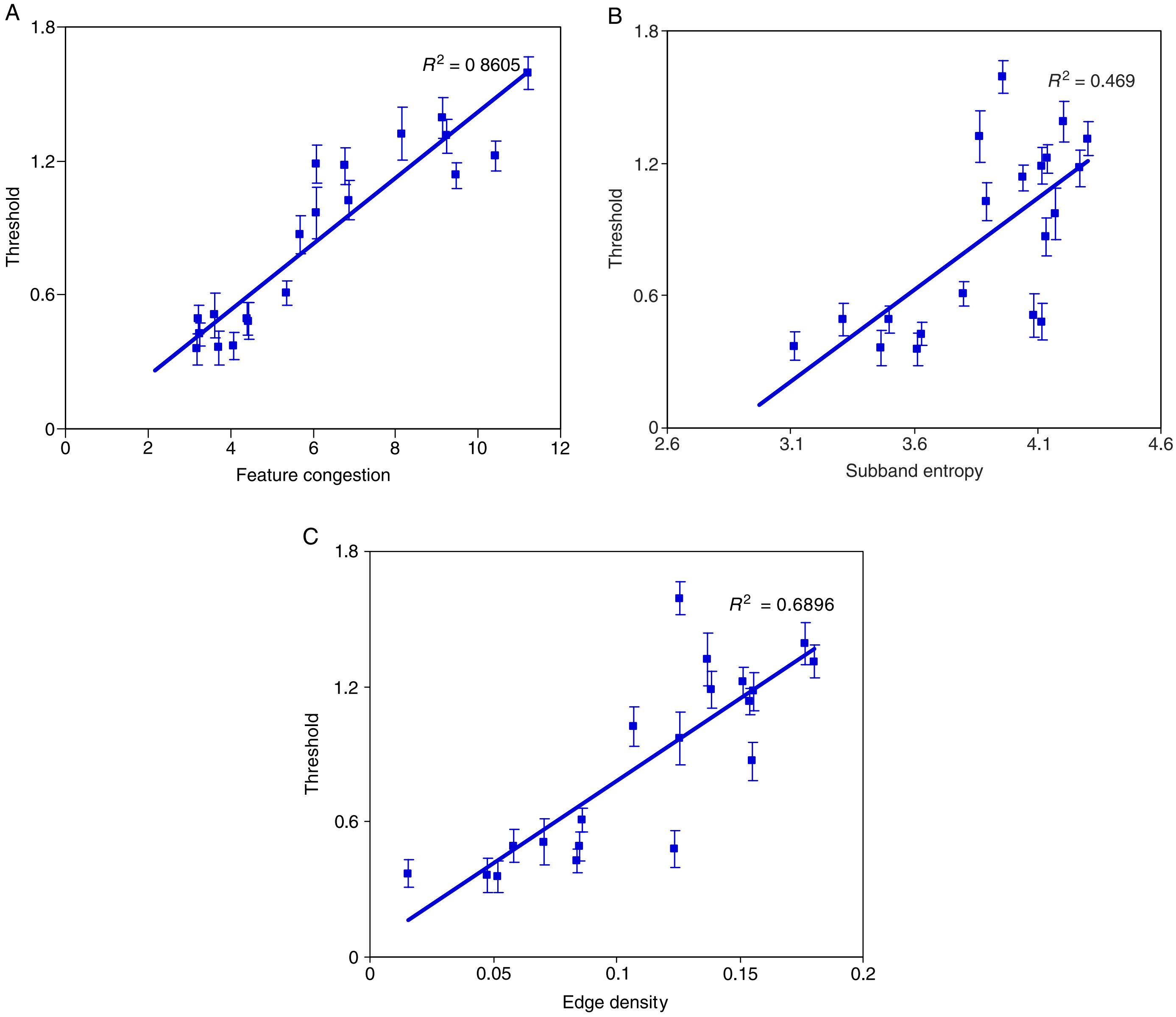


Figure 9. Mean threshold target contrast (threshold) versus clutter measures, plotted with 95% confidence intervals. (A) Threshold versus Feature Congestion clutter, (B) threshold versus Subband Entropy clutter, and (C) threshold versus Edge Density clutter.

search for a red horizontal bar among green horizontal and red vertical. Set sizes were 4, 8, 12, and 18.

In general, not surprisingly, all three measures of visual clutter monotonically increased with set size. Comparisons between types of search are trickier. The Edge Density measure, in particular, is quite sensitive to item size, making comparisons across search types uninformative. For both Subband Entropy and Feature Congestion, feature search displays appear more cluttered than conjunction search, which appears more cluttered than the search for a *T* among *L*s. Conjunction search is probably less cluttered than feature search for these examples because the red items in our conjunction search examples were actually much more similar to the background than the green items were and thus provided less clutter than the green items did. The ordering of clutter between feature and *T* versus *L* search makes some sense if we think of clutter as a more complicated stand-in for set size. Search performance is a matter of target–distractor discriminability and clutter or set size. The *T* versus *L* search is arguably difficult precisely because low target–distractor discriminability means that the display looks like a uniform (low clutter) texture. Target–distractor discriminability is high in a red among green feature search, so search is easy regardless of the level of "clutter."

Comparing the Feature Congestion measure of visual clutter with previous results on search in clutter

Comparison with Wolfe, Oliva, Horowitz, Butcher, and Bompas (2002): Search on cluttered desks

Wolfe et al. (2002) had observers search for a target *T* among distractor *L*s, where both target and distractors appear against one of three "desk" images: empty, clean, and messy. The *T* and *L*s appear in predictable locations, and in one condition, they appear on yellow "post-it" notes, so these experiments have a significant top–down component to the search, and as a result, we might expect there to be limits to the predictive value of any bottom–up clutter measure. Nonetheless, even with top–down information that could guide the observers to ignore the background, they find that more "messy" backgrounds lead to additive RT costs in their search task. We asked whether the candidate clutter measures could predict the increase in clutter from empty, to clean, to messy desk, and in fact, they can. All clutter measures were applied to the background images without a target *T* or distractor *L*s present. The Feature Congestion clutter measures for the three desk images shown in Figure 10 are 3.4, 4.3, and 6.1, respectively. The Edge Density and Subband Entropy measures give similar results, although the Edge Density measure is highly sensitive to parameter settings—with the wrong settings, the empty desk is actually more cluttered than the clean desk due to the wood grain. Many measures of clutter are likely to give this ordering on images with such different levels of clutter, but nonetheless, it is important to confirm that a measure of clutter gets reasonable results on one of the few existing search in clutter experiments in the human vision literature.

Comparison with Bravo and Farid (2004): Categorical search in sparse versus cluttered arrays, with single versus multipart objects

Bravo and Farid (2004) had observers search for a categorical target (food). Because this is a categorical task in displays where target and distractors appear in random locations, we expect that a good bottom–up clutter measure should do a reasonable job of correlating with search performance. Their main results are as follows: Their observers search in both sparse and cluttered displays (see Figure 11), and Bravo and Farid find that search in cluttered displays is significantly worse than search in the sparse displays. They also find that search is significantly more difficult with complex (multipart) distractors than with simple distractors, but that the complexity of the target makes no difference. Search difficulty increases with the number of objects—the

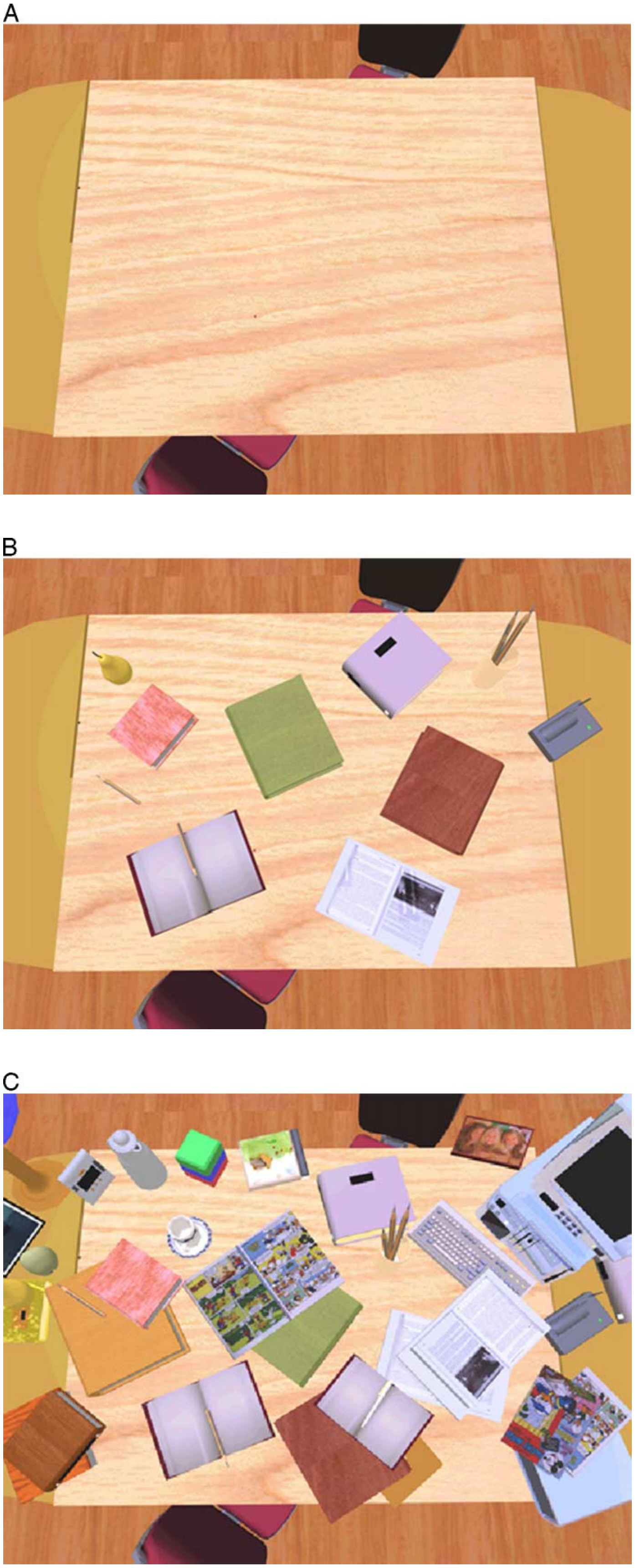


Figure 10. (A) Empty, (B) clean, and (C) messy desk images from Wolfe et al. (2002).

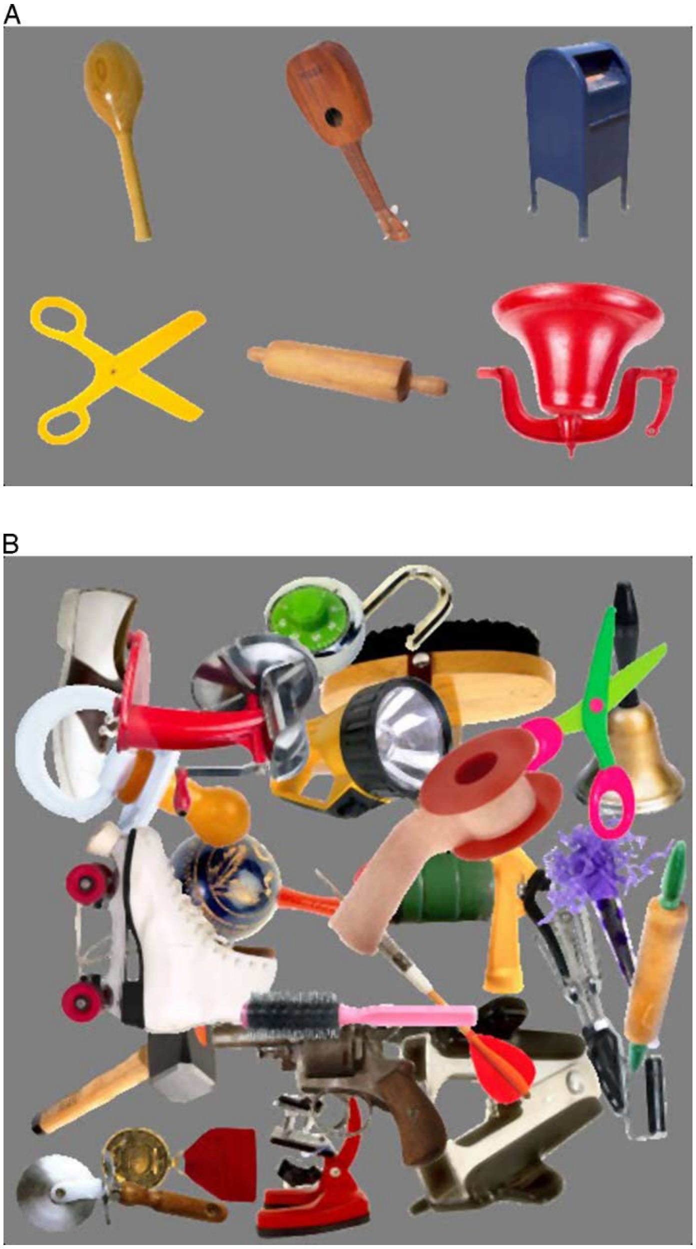


Figure 11. (A) Sparse arrangement and (B) cluttered arrangement, from Bravo and Farid (2004). In addition, Panel (A) shows simple objects, whereas Panel (B) shows complex (multipart) objects.

nominal set size—but they argue that, because of the difference between simple and multipart distractors, the number of parts is a more relevant measure of search difficulty.

We ran the three clutter measures on 960 of their images. The results for the Feature Congestion measure are shown in Figure 12; results for the other two measures were quite similar qualitatively, except that within sparse displays, there was no apparent difference between complex and simple distractors. Cluttered displays are significantly more cluttered than sparse displays. Images with complex distractors are significantly more cluttered than images with simple distractors, but there is no significant difference between simple and complex targets. Distractor type and display type interact, so there is a bigger effect of distractor type in complex displays than in sparse. Bravo and Farid (2004) find no significant difference of distractor type in simple displays, as found by Subband Entropy and Edge Density. But this difference from the predictions of the Feature Congestion measure could be due to added noise in the empirical data leading to a lack of significance, as they do show a consistent difference in their target absent trials. There is a significant effect of nominal set size, and this interacts with distractor type, so clutter versus set-size slopes are larger for complex than for simple distractors. Furthermore, clutter versus set-size slopes are larger in cluttered than in sparse display arrangements. There is strong agreement between these clutter results and the results of Bravo and Farid (2004), suggesting that our measures of clutter are predictive of many of their results. In some sense, this is disappointing because it does not allow us to distinguish between these measures of clutter. On the other hand, in a sense, it is an indication of the utility of a measure of clutter for understanding the results of search in complex displays. If any of our measures of clutter can predict the Bravo and Farid results, this suggests that the difference in performance between, say, simple and complex distractors may be due to a difference in clutter of the displays, measured in any of a number of ways, and not necessarily to a difference in the number of parts, which Bravo and Farid essentially give as their measure of clutter.

Experiment 3: Does color variability matter?

All three measures of visual clutter correlate well with the data from our Experiment 1 and with previous search experiments. A big difference between the three measures is in their handling of feature variability and, in particular, color variability. Experiments and modeling of visual search suggest that increased feature variability impairs search performance (Duncan & Humphreys, 1989; Rosenholtz, 1999, 2001b). Furthermore, Oliva et al. (2004) report an association between color variability and complexity judgments. These observations are often echoed in design guidelines. For example, the Windows MSDN Visual Design Guidelines (http://msdn2.microsoft.com/en-us/library/ms997613.aspx) warn against the use of too many colors in a user interface because this will add visual clutter.

The Feature Congestion clutter measure explicitly captures variability of color and other features. The Subband Entropy measure captures this more implicitly by looking essentially at the amount of high frequencies in the color bands, and the Edge Density measure ignores

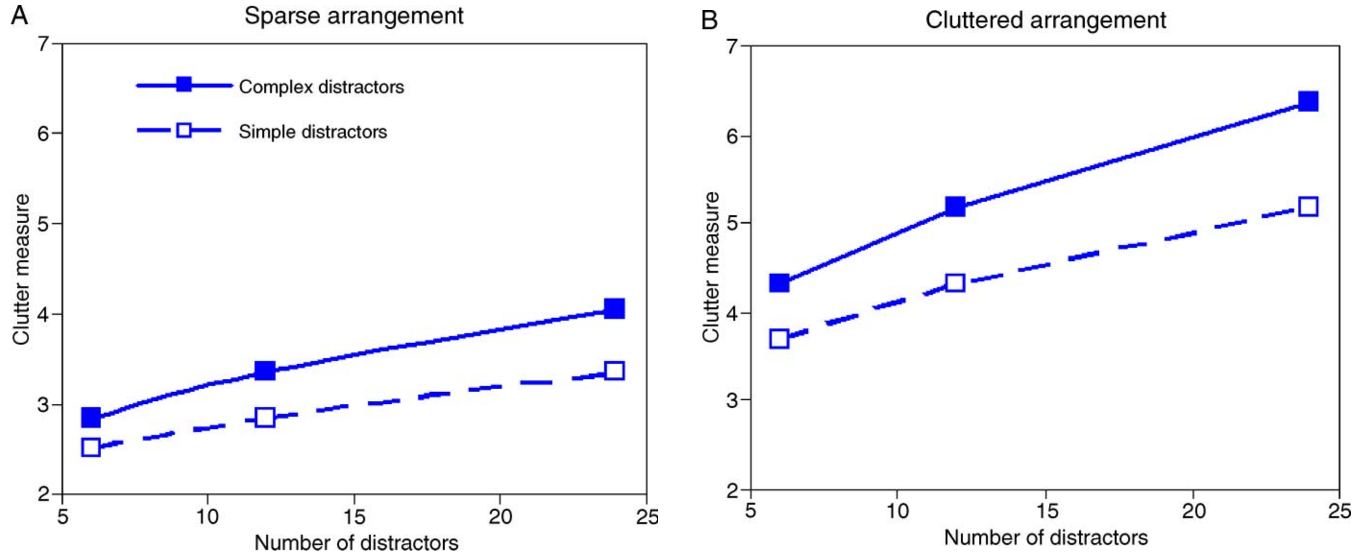


Figure 12. Results of the Feature Congestion clutter measure on images from Bravo and Farid (2004).

color variability. However, despite these differences, all three measures have performed well on the above examples.

How can we reconcile both research and common wisdom that suggest that color variability is an important factor for visual clutter and the fact that all three clutter measures have done well at predicting search results, even when they do not take color variability into account? Perhaps, for our particular choice of images, color variability covaries to a large extent with edge density—when there is a change in object, there is an edge, and there also tends to be a change in color. However, this is hardly true in general. Search experiments, for which we hope to have a useful measure to replace set size, have a wide variety of degrees of color variability. Designers of user interfaces and information visualizations often spend a great deal of time deciding how much color variability to use for their designs. As a result, many maps, Web pages, medical imaging displays, thermal imaging displays, and others, are monochrome or near monochrome, whereas others have a high degree of color variability (see Figure 13). In this experiment, we intentionally separate color variability from a change in objects to examine the importance of capturing color variability in a measure of visual clutter.

Stimuli

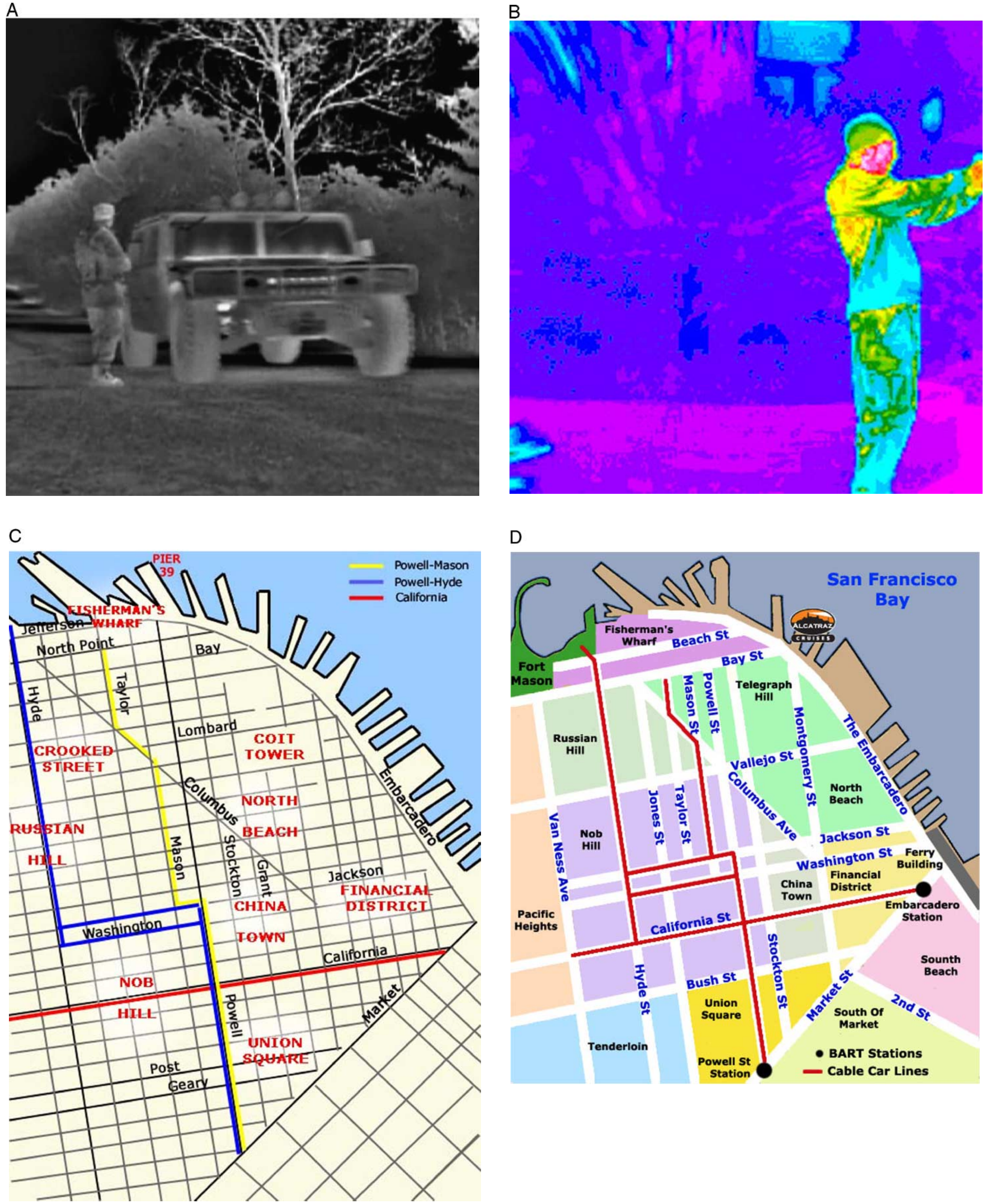
We chose six maps from Experiment 1 and modified them to create a set of 18 maps. For each map, we kept the original. In addition, we created from each map a gray map and a red map. The gray map was created by desaturating the original map using the “desaturate” command in Photoshop. The red map was created by converting the images to CIELab color space and then fixing the ratio of *a* to *b* to an approximately constant value specifying a reddish hue, while allowing the L channel to vary freely. The gray and red maps have considerably less color variability than the original images, while containing approximately the same edge density and subband entropy in the L channel as the original image.

The target was one of three Gabors. The Gabors varied only in color (green, yellow orange, and gray) and in orientation (45°, 135°, and 45°, respectively). The gray Gabor was equivalent to the Gabors used in Experiment 1. The sigma, wavelength, and phase of the Gabors remained constant: 0.55°, 1.1 deg/cycle, and 90°, respectively. We used fewer targets in this experiment than in Experiment 1 to reduce experiment time because comparison of Experiment 1 with a pilot experiment with fewer Gabors showed little difference in the search results or ability of the clutter measure to predict subjects’ performance.

Methods

The same protocol from Experiment 1 was followed. There were a total of 648 trials (3 targets × 18 maps × 6 locations × present vs. absent) split between two blocks.

Figure 13. Two pairs of similar imagery with similar numbers of edges and similar high frequency content but very different amounts of color variability. (A) Thermal imagery, monochrome, and (B) thermal imagery, with pseudocoloring. Panels (C) and (D) are cable-car maps of San Francisco, with similar content but very different use of color. The amount of color to use is a common design decision faced by designers.


A
B
C
PIER
39
Powell-Mason
Powell-Hyde
California
FISHERMAN'S
WHARF
Jefferson
North Point
Bay
Hyde
Taylor
Lombard
COIT
TOWER
CROOKED
STREET
Columbus
NORTH
BEACH
Embarcadero
RUSSIAN
HILL
Mason
Stockton
Grant
Jackson
FINANCIAL
DISTRICT
CHINA
TOWN
Washington
California
NOB
HILL
Powell
Market
UNION
SQUARE
Post
Geary
D
San Francisco
Bay
ALCATRAZ
CRUISES
Fisherman's
Wharf
Beach St
Fort
Mason
Bay St
Mason St
Powell St
Telegraph
Hill
Russian
Hill
Montgomery St
The Embarcadero
Vallejo St
Columbus Ave
North
Beach
Jones St
Taylor St
Van Ness Ave
Nob
Hill
Jackson St
Washington St
Ferry
Building
China
Town
Financial
District
Embarcadero
Station
Pacific
Heights
California St
Stockton St
South
Beach
Hyde St
Bush St
Union
Square
Market St
2nd St
South Of
Market
Tenderloin
Powell St
Station
BART Stations
Cable Car Lines

Four subjects participated in Experiment 3. All participants had normal or corrected-to-normal vision and were compensated for their time.

Predictions

The original images had significantly more clutter (M = 6.3) than either the red-map (M = 5.8, $p < .05$) or gray-map (M = 5.1, $p < .01$) images, as measured by Feature Congestion. The percentage of edge pixels was not significantly different between the red-map (M = 13.0%) and the original (M = 12.6%, $p > .05$) images and between the original and the gray-map images (M = 12.5%, $p > .05$). By the Subband Entropy measure, the original images (M = 4.1) were significantly *less* cluttered than the red-map images (M = 4.2, $p < .01$) and significantly more cluttered than the gray-map images (M = 3.4, $p < .01$).

From these differences, the Feature Congestion measure suggests that search should be more difficult in the original map images than in either the red maps or gray maps. The Edge Density measure suggests that search should be about the same in all three categories, possibly slightly more difficult in the red maps than in the original maps. The Subband Entropy measure suggests that search should be more difficult in the red-map images than in the original images and easier in the gray-map images than in the original images.

Results

We again looked at RTs for correct trials, separating target-present from target-absent trials. For each image, we averaged RT over Gabor type, target location, and subject. We found that mean RT was significantly slower in the original images than in the red-map images for both target-absent trials, M_{red} = 947 ms, M_{orig} = 1,179 ms, $t(23) = 6.1$ (paired t test), $p < .001$, and target-present trials, M_{red} = 552 ms, M_{orig} = 772 ms, $t(23) = 11.8$, $p < .001$. For target-present trials, the mean RT was significantly slower for original images than for gray-map images, M_{gray} = 619 ms, $t(23) = 8.3$, $p < .001$. For target-absent trials, there was no significant difference between RTs for the original images versus the gray-map images (M_{gray} = 1,172 ms, $p = .83$). The difficulty in target-absent trials for the gray-map images makes sense, given that one of the possible targets is also gray; in the absence of any target, observers take a long time to decide that there is no gray target against the gray background. This is also likely the explanation for the faster RTs for red-map images than gray-map images.

Ignoring the target-absent gray-map trials as being slow for a reason other than clutter, in all other cases, search in the original image is significantly slower than search in the red-map and gray-map images. This pattern of results is predicted by the Feature Congestion measure of visual clutter, and not by either the Edge Density measure or the Subband Entropy measure. Color variability is relevant to image clutter, and the important measure is the amount of color variability (the number of colors and how different they are) as opposed to the amount of high frequencies in the color channels.

Edge Density and the presence of high frequencies are clearly correlated with clutter, and one could perhaps argue that for certain classes of images, color variability will covary with things like high frequency content and edge density, and thus, measures such as Edge Density and Subband Entropy will do a reasonable job despite their lack of an explicit accounting for color variability.

Conclusions and future work

A measure of the visual clutter in an arbitrary image would be beneficial both for practical applications and for basic research into topics such as visual search. Clutter can degrade performance at a number of visual tasks, and having a measure of clutter could therefore be useful either to help design displays with an optimal level of visual clutter or to provide system alerts when the level of clutter might impair performance, for example, in searching for a threat in a baggage X-ray or in noticing pedestrians while driving. For basic research into visual search, we argue that a measure of clutter is a key component for moving visual search research into more natural, complex imagery because it could replace the notion of set size in simple psychophysical displays. A measure of clutter could allow us to better understand the inherent difficulty in searching through a complex display and thus allow us to better evaluate whether our performance with that display is due solely to more or less clutter or whether other factors such as top–down information play a role.

We have tested three measures of visual clutter, including two we suggested—Feature Congestion and Subband Entropy—and the Edge Density measure previously used to predict subjective judgments of image complexity (Mack & Oliva, 2004). The Feature Congestion measure was based on an analogy that the more cluttered a desk is, the more difficult it would be to add an attention-grabbing note to the desktop. This measure is based on Rosenholtz' Statistical Saliency Model. The Subband Entropy measure is based on the notion of clutter as related to the efficiency with which the image can be encoded and inversely related to the amount of redundancy and grouping in the image. The Edge Density measure attempts to capture the notion of clutter as number of objects by calculating the density of edges, as well as a likely correlation of clutter with high-frequency content.

There are, of course, a number of additional components or improvements that could be made to these measures. From the point of view of applications and, perhaps, also basic visual search research, it would be good to model other visual processes with degraded performance in the presence of additional items, such as object recognition under conditions of crowding. Although we have focused on modeling visual search, our Feature Congestion measure may actually already be a step toward capturing the likelihood of impaired object recognition due to crowding; if crowding occurs because features from flanking items "bleed" into the features of a central item, the more "congested" the feature space, perhaps, the more likely it is that this feature bleeding impairs recognition performance. A similar argument could be made for Subband Entropy: The higher the entropy, the less that the features of one object can be predicted from the features of surrounding objects, and thus, the more that "bleeding" of features could be expected to impair performance under crowding.

The amount of clutter or "visual information" in a display no doubt also has to do with the degree of organization of a display—to what extent features are grouped together by the visual system, due to similarity, alignment, symmetry, and so on. Such perceptual organization is known to affect performance at visual search (Treisman, 1982) and at other visual tasks. Our Feature Congestion measure captures a bit of perceptual organization implicitly—grouping by similarity + proximity—as does the Subband Entropy measure. Ideally, such a measure would more explicitly capture the effects of perceptual organization in human perception.

	FC	SE	ED
Run on arbitrary images as input	●	●	●
Mean RT in map search	●	●	●
Contrast threshold in map search	●	◉	●
Set size in standard search displays	●	●	●
Wolfe et al. (2002) messy desks	●	●	◉
Bravo and Farid (2004)	●	●	●
Color variability	●	○	○
Clutter due to different features	●	◉	○
Limited gamut displays	●	◉	○
Spatial distribution of clutter	●	○	●

Table 1. Summary chart comparing the three clutter measures: Feature Congestion (FC), Subband Entropy (SE), and Edge Density (ED). A filled black circle (●) indicates that the measure does well. A filled gray circle (◉) indicates that the measure does well, but less well than one of the other measures. An empty circle (○) indicates that the measure does not satisfy the given criterion. See text for further explanation.

A comparison of the three measures of visual clutter is shown in Table 1. All three measures of visual clutter can run on arbitrary images as input. We have demonstrated that these measures of visual clutter perform well at predicting the results of our visual search experiments. They increase monotonically with nominal set size in standard psychophysical displays and correlate well with the results of previous search-in-clutter experiments from Wolfe et al. (2002) and Bravo and Farid (2004). They all correlate well with mean search time in our map search experiments and with contrast thresholds in Experiment 2. There were no significant differences between the three measures in terms of predicting mean RT in search within our maps. Feature Congestion does better at predicting mean contrast thresholds, and this is a significant improvement over Subband Entropy. The measures all perform about equally well on the results of previous search experiments.

In Experiment 3, we have shown that increased color variability does increase visual clutter in a way that Feature Congestion captures, but neither of the other two measures do. Given that choosing how much color to use in a display is a common decision for designers of information visualizations and user interfaces, this is an important advantage for Feature Congestion. Feature Congestion has other advantages as well. In this article, we have asked this measure to give only one number representing clutter for the entire image. The Feature Congestion clutter measure, unlike Subband Entropy and Edge Density, can let us know the difference between color, texture (contrast energy), and orientation clutter. Perhaps, this will allow Feature Congestion to make better predictions in situations in which the observer is known to be looking for a target defined, for instance, by a unique color.

Another issue is one of displays with limited gamut. As mentioned above, designers are often trying to decide whether to limit, for example, the colors allowed in a display. This limitation may apply to potential targets in a display and to the background or distractors. In this case, it may be misleading to label a display as "uncluttered" simply because it is monochrome; one might not be able to add a target that draws attention because of its color, because the display might be monochrome. The Feature Congestion measure of clutter can gracefully handle this situation by reporting essentially the fraction of the *available* feature space taken up by the covariance ellipsoid. The Subband Entropy measure may be able to handle particular gamut limitations, for example, monochrome displays, but cannot easily handle arbitrary limitations on the available feature space. We do not see a way for the Edge Density measure to handle a limited feature space.

Feature Congestion and Edge Density can also give an idea of the spatial distribution of clutter, which Subband Entropy cannot easily do. Figure 14 gives an example of a

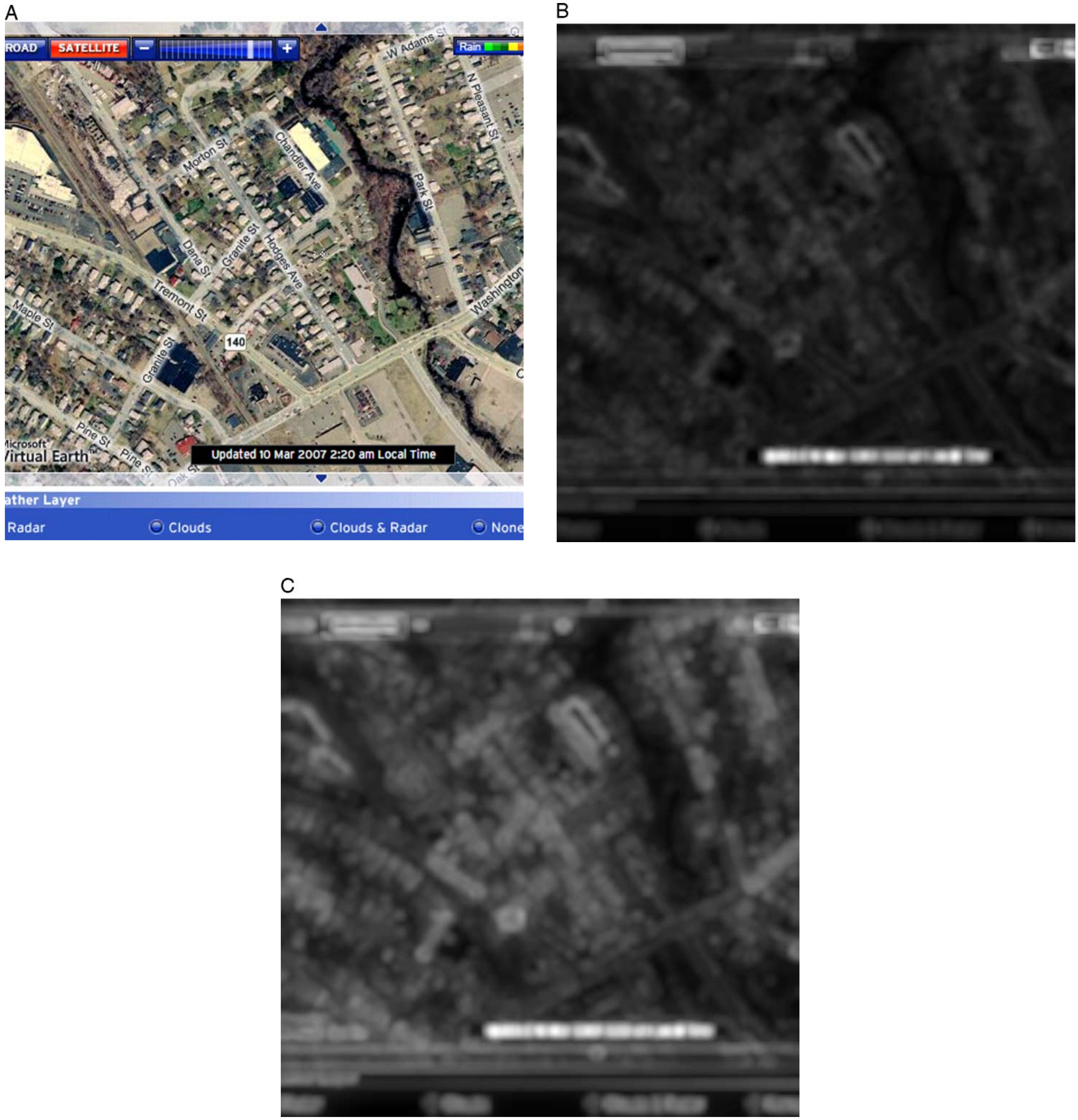


Figure 14. A sample map and its clutter maps output by the Feature Congestion measure. (A) Original map; (B) color variability clutter map; (C) full clutter map combining color, orientation, and contrast-energy clutter.

map and two clutter maps output by the Feature Congestion measure of visual clutter. The clutter maps clearly indicate that there is a good deal of clutter everywhere in the image, but there is color clutter only in a small portion; search for a color-defined target within this scene might be quite efficient.

Acknowledgments

This research was supported by ONR Grant N00014-01-1-0625 and NSF Grant BCS-0518157 to Ruth Rosenholtz. The authors thank Zhenlan Jin, Jonathan Mansfield, Naomi Kenner, Mary Bravo, and Jeremy Wolfe for their help on experiments and modeling.

Commercial relationships: none.
Corresponding author: Ruth Rosenholtz.
Email: rruth@mit.edu.
Address: 46-4115, MIT, Cambridge, MA 02139.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15,* 106–111. [PubMed]

Dynamic Logic. (2001). Beyond the click: Insights from marketing effectiveness research. [Link]

Bergen, J. R., & Landy, M. S. (1991). Computational modeling of visual texture segregation. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 253–271). Cambridge, MA: MIT Press.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. [PubMed]

Bravo, M. J., & Farid, H. (2004). Search for a category target in clutter. *Perception, 33,* 643–652. [PubMed]

Burt, P., & Adelson, E. H. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communication, COM-31,* 532–540.

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

C.I.E. (1978). *CIE Recommendations on Uniform Color Spaces, Colour-difference Equations, and Psychometric Colour Terms* (supp. no. 2). Paris: Bureau Central de la CIE.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458. [PubMed]

Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size in visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & Psychophysics, 62,* 425–451. [PubMed]

Frank, A. U., & Timpf, S. (1994). Multiple representations for cartographic objects in a multi-scale tree—An intelligent graphical zoom. *Computers & Graphics, 18,* 823–829.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 1254–1259.

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America, 70,* 1458–1471. [PubMed]

Mack, M. L., & Oliva, A. (2004). *Computational estimation of visual complexity*. Paper presented at the 12th Annual Object, Perception, Attention, and Memory Conference. Minneapolis, Minnesota.

Malik, J., & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A, Optics and image science, 7,* 923–932. [PubMed]

Miller, G. A. (1994). The magical number seven, plus or minus two: Some limits on our capacity for processing information. 1956. *Psychological Review, 101,* 343–352. [PubMed]

Nickerson, J. V. (1994). *Visual programming*. PhD dissertation, New York University, New York.

Nothdurft, H. C. (1993) The role of features in preattentive vision: Comparison of orientation, motion, and color cues. *Vision Research, 33,* 1937–1958. [PubMed]

Oliva, A., Mack, M. L., Shrestha, M., & Peeper, A. (2004). Identifying the perceptual dimensions of visual complexity of scenes. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society Meeting*. Chicago.

Olshausen, B. A., & Field, D. J. (1996). Natural image statistics and efficient coding. *Network, 7,* 333–339. [PubMed]

Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research, 34,* 1703–1721. [PubMed]

Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 108–130. [PubMed]

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. [PubMed]

Phillips, R. J., & Noyes, L. (1982). An investigation of visual clutter in the topographic base of a geological map. *Cartographic Journal, 19,* 122–132.

Robbins, H., & Monro, S. (1951). A stochastic approximation method. *Annals of Mathematical Statistics, 22,* 400–407.

Rosenholtz, R. (1999). A simple saliency model predicts a number of motion popout phenomena. *Vision Research, 39,* 3157–3163. [PubMed]

Rosenholtz, R. (2000). Significantly different textures: A computational model of pre-attentive texture segmentation. In D. Vernon (Ed.), *Proceedings of the European Conference on Computer Vision* (pp. 197–211). Berlin-Heidelberg: Springer Verlag.

Rosenholtz, R. (2001a). Search asymmetries? What search asymmetries? *Perception & Psychophysics, 63,* 476–489. [PubMed] [Article]

Rosenholtz, R. (2001b). Visual search for orientation among heterogeneous distractors: Experimental results and implications for signal-detection theory models of search. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 985–999. [PubMed]

Rosenholtz, R., & Jin, Z. (2005). A computational form of the Statistical Saliency Model for visual search [Abstract]. *Journal of Vision, 5*(8):777, 777a, http://journalofvision.org/5/8/777/, doi:10.1167/5.8.777.

Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005). Feature congestion, a measure of display clutter. *SIGCHI* (pp. 761–770). Portland, Oregon.

Rosenholtz, R., Nagy, A. L., & Bell, N. R. (2004). The effect of background color on asymmetries in color search. *Journal of Vision, 4*(3):9, 224–240, http://journalofvision.org/4/3/9/, doi:10.1167/4.3.9. [PubMed] [Article]

Simoncelli, E. P., & Freeman, W. T. (1995, October) *The steerable pyramid: A flexible architecture for multiscale derivative computation*. Paper presented at the 2nd Annual IEEE International Conference on Image Processing, Washington, DC.

Springer, C. J. (1987). Retrieval of information from complex alphanumeric displays: Screen formatting variables' effects on target identification time. In G. Salvendy (Ed.), *Proceedings of the 2nd International Conference on Human–Computer Interaction* (pp. 375–382). Honolulu, Hawaii: Elsevier Science.

Stuart, J. A., & Burian, H. M. (1962). A study of separation difficulty: Its relationship to visual acuity in normal and amblyopic eyes. *American Journal of Ophthalmology, 53,* 471–477. [PubMed]

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113,* 766–786. [PubMed]

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 194–214. [PubMed]

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136. [PubMed]

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research, 35,* 2503–2522. [PubMed]

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.

Van Wert, M. J., Horowitz, T. S., Place, S. S., & Wolfe, J. M. (2006). Errors in low prevalence visual search: Easy to produce, hard to cure [Abstract]. *Journal of Vision, 6*(6):444, 444a, http://journalofvision.org/6/6/444/, doi: 10.1167/6.6.444.

Watson, A. B. (2000). Visual detection of spatial contrast patterns: Evaluation of five simple models. *Optics Express, 6,* 12–33. [PubMed] [Article]

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202–238.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–73). London: University College London Press.

Wolfe, J. M., Oliva, A., Horowitz, T. S., Butcher, S., & Bompas, A. (2002). Segmentation of objects from backgrounds in visual search tasks. *Vision Research, 42,* 2985–3004. [PubMed]

Woodruff, A., Landay, J., & Stonebraker, M. (1998). Constant information density in zoomable interfaces. In *Proceedings of Advanced Visual Interfaces* (pp. 57–65). L'Aquila, Italy.