

Dartmouth College

Dartmouth Digital Commons

Dartmouth Scholarship

Faculty Work

1-31-2008

A Scale Invariant Measure of Clutter

Mary J. Bravo
Rutgers University

Hany Farid
Dartmouth College

Follow this and additional works at: <https://digitalcommons.dartmouth.edu/facoa>



Part of the [Physiology Commons](#)

Dartmouth Digital Commons Citation

Bravo, Mary J. and Farid, Hany, "A Scale Invariant Measure of Clutter" (2008). *Dartmouth Scholarship*. 3899.

<https://digitalcommons.dartmouth.edu/facoa/3899>

This Article is brought to you for free and open access by the Faculty Work at Dartmouth Digital Commons. It has been accepted for inclusion in Dartmouth Scholarship by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.

A scale invariant measure of clutter

Mary J. Bravo

Psychology Department, Rutgers,
Camden NJ, USA



Hany Farid

Department of Computer Science, Dartmouth College,
Hanover NH, USA



We propose a measure of clutter for real images that can be used to predict search times. This measure uses an efficient segmentation algorithm (P. Felzenszwalb & D. Huttenlocher, 2004) to count the number of regions in an image. This number is not uniquely defined, however, because it varies with the scale of segmentation. The relationship between the number of regions and the scale of segmentation follows a power law, and the exponent of the power law is similar across images. We fit power law functions to the multiple scale segmentations of 160 images. The power law exponent was set to the average value for the set of images, and the constant of proportionality was used as a measure of image clutter. The same 160 images were also used as stimuli in a visual search experiment. This scale-invariant measure of clutter accounted for about 40% of the variance in the visual search times.

Keywords: natural images, clutter, visual search, image segmentation

Citation: Bravo, M. J., & Farid, H. (2008). A scale invariant measure of clutter. *Journal of Vision*, 8(1):23, 1–9, <http://journalofvision.org/8/1/23/>, doi:10.1167/8.1.23.

Introduction

Visual search is a popular experimental paradigm that has been employed in thousands of studies. Visual search is also an everyday task familiar to anyone with car keys or reading glasses. Despite the vast amount of research on this topic, we are far from making quantitative predictions about everyday search. The gap between research and real-life exists in part because the stimuli used in research bear little resemblance to real images. Visual search research has generally used simple shapes arranged in a regular array on a blank background, and it is unclear how such stimuli relate to the complex and continuous images encountered in real life. As a result, current search models cannot account for such commonplace observations, as “I can’t find my keys because of all this clutter on my desk.” The models cannot address this observation because they cannot yet quantify the clutter found in real images.

The goal of this study was to find a measure of clutter that could be used to predict search times with real images. To develop this measure, we sought a search task that would produce a robust and a reliable effect of clutter. Not all search tasks show such an effect: Even on a cluttered desk, it is easy to find keys that are attached to a bright red key fob or that are in a remembered location. Previous research suggests that there are three conditions that are important for observing an effect of clutter. First, observers must be uncertain of the target’s color, orientation, and other simple features. This uncertainty prevents them from selectively attending to a particular feature to guide their search (Wolfe, 1994). Second, the images must be unstructured, so that observers cannot

predict where the target is likely to appear (Torralba, Oliva, Castelhano, & Henderson, 2006). And third, the targets must be drawn from the same pool of objects as the distractors, so that they are not especially salient (Itti & Koch, 2000).

With these requirements in mind, we selected a task that involved looking for common objects in the contents of handbags. Our stimuli were photographs downloaded from the “What’s in your bag?” group on the photo-sharing Web site, <http://flickr.com> (Figure 1). Although the images had diverse arrangements, backgrounds, and lighting conditions, their content was typically limited to a small number of object categories. The task we gave our observers was to search the set of images several times, each time looking for an object from a different category. Because the observers knew only the category to which the target belonged, they were uncertain of its simple features. Thus, the observers could not use selective feature attention to exclude much of the clutter. And similarly, because the images had an unpredictable arrangement, observers were uncertain of the target’s location. This prevented the observers from using selective spatial attention to exclude much of the clutter. And finally, because the observers searched each image several times for different targets, the targets were, on average, no more salient than the distractors. This meant that the targets did not reliably attract attention, and so observers had to search through the clutter to find them.

In developing a measure of clutter, we began with the assumption that when observers search an image for a target, they evaluate the largest image chunks that are likely to correspond to single objects. Ideally, these chunks would correspond to whole objects, but the



Figure 1. An example stimulus downloaded from the “What’s in your bag?” group on <http://flickr.com>.

selection of whole objects can require the application of top-down knowledge, and this is costly in terms of time and processing resources. For this reason, we think it is likely that the chunks for search are the regions defined by bottom-up segmentation (Neisser, 1967). These regions can be extracted without accessing object memory, but they are still likely to arise from single objects (Brunswick & Kamiya, 1953; Fine, MacLeod, & Boynton, 2003). Although it is only a conjecture that the amount of clutter is related to the number of image regions, we have reported results that are consistent with this notion (Bravo & Farid, 2004b). In this earlier experiment, we found that search times are not simply related to the number of objects in an image, they are also related to the number of distinct object parts. In this earlier paper, we viewed the number of object parts as a crude approximation of the number of image regions. Later in this paper, we re-examine these data using a direct measure of the number of image regions.

A great practical advantage of defining clutter in terms of regions rather than objects is that region extraction is a much more tractable problem than object extraction. There is a significant obstacle to counting either regions or objects, however, and that is the problem of scale. Scenes often have a hierarchical organization, and it is possible to define objects at many levels in this hierarchy. And so just as it is valid to label a bush as an object or its leaves as objects, it is also valid to segment the image of a bush into one big region or into many small regions. One

way to decide among the possible segmentations of an image is to consider the information needed to perform a particular task. Since observers engaged in visual search are looking for a specific object, one might assume that it is the size of this object that defines the appropriate scale for segmentation. But this presupposes that observers detect targets using information at the same scale as the target itself. This may not be true, especially in cluttered scenes. Clutter may camouflage the shape of a target and force observers to rely on smaller-scale features for target detection (Bravo & Farid, 2004a). So although we propose that the amount of clutter in an image is related to the number of regions produced by image segmentation, we also acknowledge the difficulty of uniquely determining this number. As we will describe in our [Methods](#) section, the difficulty of determining the appropriate scale for segmentation turns out to have a straightforward solution.

Methods

The segmentation algorithm

We used a segmentation algorithm developed by Pedro Felzenszwalb and Daniel Huttenlocher to segment our images. This algorithm is not a model of human image segmentation: It was not designed to be biologically

plausible, and it does not include all of the grouping rules used by humans. Nonetheless, like many segmentation algorithms, it produces results that are often consistent with the organization we perceive. And unlike other segmentation algorithms, it is extremely efficient, making it suitable for testing on hundreds of large images.

A full description of the algorithm can be found in Felzenszwalb and Huttenlocher (2004), and the source code is available on the authors' Web sites. In brief, the algorithm works by representing an image as an undirected graph with each vertex in the graph representing a pixel in the image. Neighboring vertices are connected by an edge, and the weight on the edge is proportional to the difference between the corresponding pixels. Segmenting the image into regions involves cutting edges in the graph to produce subcomponents, which are disjoint sets of interconnected vertices. The algorithm decides which edges to cut by comparing the minimum weight connecting two subcomponents with the maximum weight within the subcomponents. (This description of the algorithm is highly simplified; interested readers can find a more accurate description in Felzenszwalb & Huttenlocher, 2004.)

If left unchecked, this operation can cut all edges in a graph, producing regions that correspond to individual pixels. To avoid over-segmenting the image, the algorithm includes a penalty on cuts that produce small regions. The size of small is defined by a scale parameter that has the unit of pixels. When the scale parameter is set to 500, for example, subcomponents with an area smaller than 500 pixels are penalized by an amount that is inversely proportional to their area. This penalty does not impose a fixed size limit; instead, it imposes a greater burden of evidence for the segmentation of small regions. Figure 2A

shows the segmentations of one of the bag stimuli with six values of the scale parameter, and Figure 2C shows how the number of segmented regions changes with the scale parameter. Note that the relationship between the segmentation scale and the number of regions is well described by a power law.

We applied the segmentation algorithm to the bag images and found that they were all generally well fit by a power law with an exponent of -1.32 ($\sigma = 0.13$). The images differed, however, in the constant of proportionality of the power law, with highly cluttered images having larger constants. We used this constant as our measure of clutter. By defining clutter in this way, we avoid the issue of scale. The following sections describe a test of whether this measure of clutter can be used to predict observers' search times.

Bag stimuli

Our stimuli were 160 photographs of the contents of handbags downloaded from the "What's in your bag?" group on <http://flickr.com>. The targets for our search experiment were eight objects that occurred frequently in the images: cell phones, ipods, keys, writing implements, eyeglasses, hair brushes, money, and blister packs. (A blister pack is a type of packaging used for pills; the blister refers to a clear plastic bubble backed with foil.) An image was downloaded only if it could be used in two target-present conditions and two target-absent conditions. To be used in a target-present condition, the image must have had only one example of the target object, or, in the case of keys, pens, or money, a single group of target objects.

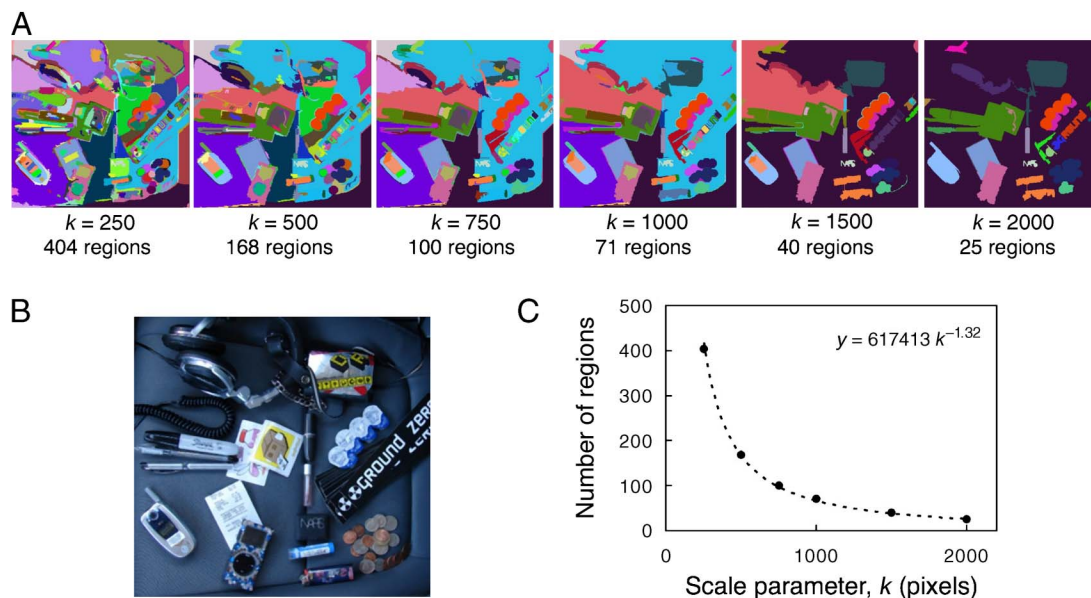


Figure 2. Along the top are six segmentations of image B. As the scale of segmentation increases, the number of segmented regions decreases. This function is well fit by a power law with an exponent of -1.32 (C).

Photographs were also selected based on image quality: Photographs that were blurry or that contained obvious quantization artifacts were not used. Also, a preference was given to photographs that depicted a random arrangement of objects. Many photographers laid out the contents of their handbags in an organized way. Although we included some of these images, our sample was biased toward images showing a haphazard arrangement. The rationale for this decision is explained later, but, in brief, we assumed that observers might ignore some regions if the image had an obvious orderly arrangement.

Many of the images were resized in Photoshop so that similar objects would have similar sizes throughout the experiment. This resizing involved matching the spatial extent of one of the target objects to a preset value. This resizing was crude in that it did not compensate for projection effects such as foreshortening. We resized the images because although our measure of clutter is invariant to the scale of segmentation, this measure is not invariant to the scale of the scene. Doubling the size of an image would increase the number of regions at all scales, and this would increase the measured clutter. But although changing image size will change measured clutter, we think it is unlikely that it will produce a commensurate change in search times. We assume that observers quickly apprehend the scale of a scene and adjust their search strategy accordingly. Since our measure of clutter does not adjust to the scale of the scene, we manually scaled the images. This rescaling could be automated or eliminated in applications having cameras at a known or fixed viewing distance.

Procedure

The observers were instructed to decide quickly but accurately whether a particular target was present in each image, and they were told to register their response using one of two keys. The observers were also instructed that when the target was present, it would be easily recognizable: An object that was too small, too occluded, or too dark to be identified was not the target.

The 50-min experiment was conducted on an Apple PowerMac using MatLab and PsychToolbox routines (Brainard, 1997; Pelli, 1997). There were eight blocks of trials, one for each of the eight targets. The number of trials in a block ranged from 68 to 108, with an average of 83. (Ideally, each block would have had the same number of trials, but we had limited control over these stimuli, and we were more concerned that each image be used with four different targets.) Each block was preceded by 10 practice trials to allow the observers to adjust their perceptual set to the new category. Auditory feedback was given after incorrect trials. The observer initiated the first trial in each block; all subsequent trials began 1 s after a response.

Observers

Thirty observers participated in this experiment. The observers were recruited from the undergraduate subject pool at Rutgers-Camden. All observers reported normal color vision and normal or corrected-to-normal acuity.

Results and discussion

Figure 3A shows response times plotted against clutter for each of the eight target categories. Each data point shows the average time observers took to find a particular target in a particular image. We expected considerable scatter in these plots because the images differed in many ways that are known to influence search (e.g., the location and salience of the target). To isolate the effect of clutter, it is necessary to average-out these other effects. To this end, we had observers search each image for four different targets, twice with the target present and twice with it absent. Figure 3B shows, for each image, the averaged response times for two present trials and two absent trials. Figure 3C shows the normalized average response time for all four trials. (To normalize the present and absent data, we fit each data set with a line, subtracted off the intercepts, and then transformed the present data so that its line coincided with that of the absent data.) The clutter measure accounts for 38% of the variance of the averaged, normalized response times.

To understand why the clutter measure does not account for more of the search time variance, we examined the outliers in Figure 3C. Because a small change in the power law exponent can cause a large change in the constant of proportionality, we examined whether the outlier images had best-fitting exponents that differed consistently from the mean. We found no obvious relationship between the deviation of the best-fitting exponent and the dispersion in the search time data. We also visually inspected images that had similar amounts of clutter but different response times. For example, images A and B in Figure 4 are judged as similarly cluttered, but the average search time for A is much faster than that for B. A potentially relevant difference between these images is that A contains several patches of regular texture, and it is possible that observers treated the regions comprising these textures as a single chunk (Neider & Zelinsky, 2006). It is also instructive to compare images that have similar response times but different amounts of measured clutter. Image C is judged to have more clutter than image B, but observers searched these two images with similar ease. Again, we might understand this discrepancy by noting that although the textured background in C increases measured clutter, observers may recognize it as a texture and ignore it during search. A better measure of the clutter may require

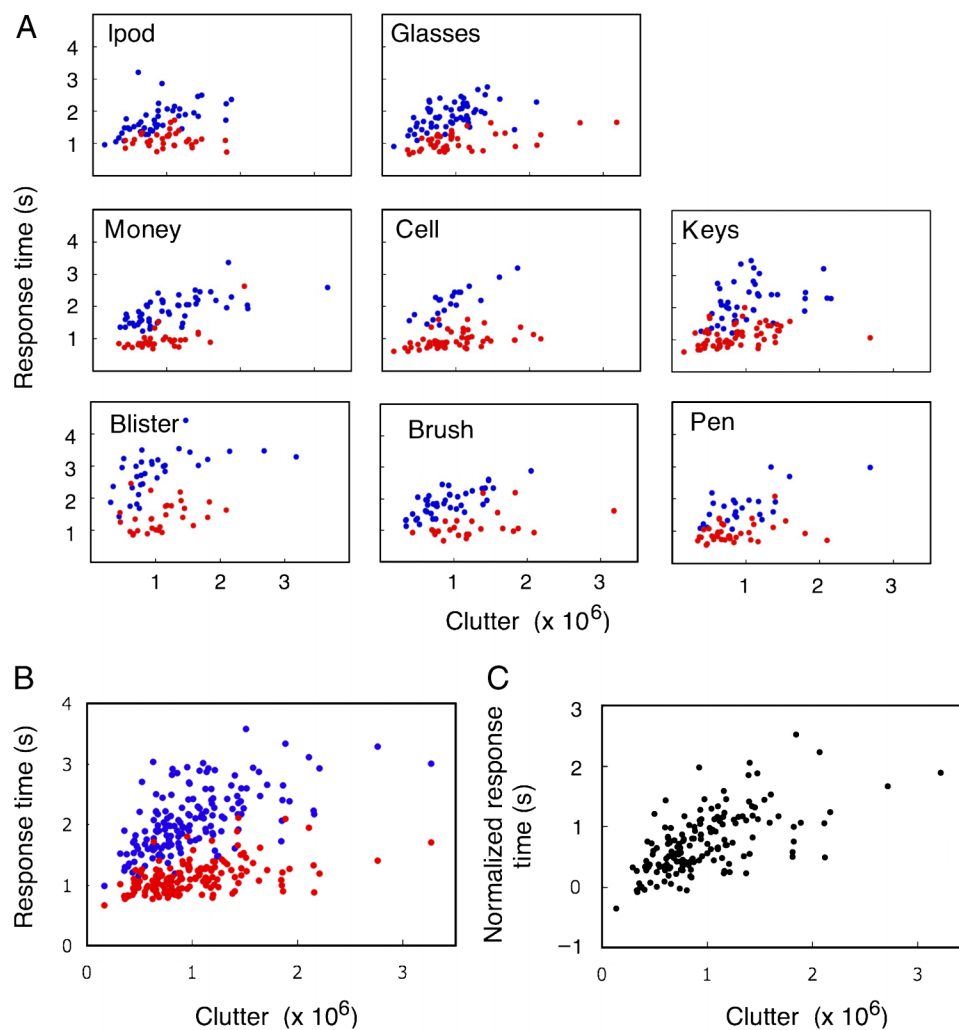


Figure 3. (A) Search times plotted against image clutter for the eight target categories. Each data point corresponds to the average response time for one target and one image, $N = 30$. Red dots correspond to target-present trials, and blue dots correspond to target-absent trials. (B) Each image was used with four target categories, twice with the target present and twice with the target absent. The graph shows for each image the average response times for two present trials and two absent trials, respectively. Present data: slope = 0.27, intercept = 0.81, $R^2 = .24$; absent data: slope = 0.66, intercept = 1.37, $R^2 = .36$. (C) The normalized, averaged response times for all four trials, $R^2 = .38$.

a segmentation algorithm that groups regular textures into a single region.

Although some of the scatter in the data is surely due to the limitations of our clutter measure, some of the scatter may be due to the limitations of our experimental design. In the introduction, we described three task requirements that are important for isolating the effect of clutter. One requirement is that observers be uncertain of the target's features. We hoped the wide range of exemplars comprising each target set would satisfy this requirement. But although simple features like color or orientation were not reliably associated with the targets, there may have been more complex, category-specific features that were. A second requirement is that the target be no more salient than the distractors. We attempted to satisfy this requirement by having observers search each

image for different targets and averaging the results. Because only a subset of the image objects served as targets, this requirement may also have been only partially satisfied. The third requirement for isolating the effect of clutter is that the stimuli be unstructured. Most images had a haphazard arrangement of objects, but some images had an orderly arrangement. If observers perceived this order, they may have assumed that objects were not piled on top of one another, and they may have ignored small regions that were clearly embedded in larger regions. So this requirement may also have been only partially satisfied.

These potential shortcomings in the experiment arose because we traded control over our stimuli for the diversity and the naturalism of images downloaded from the Internet. In a previously reported experiment, we made

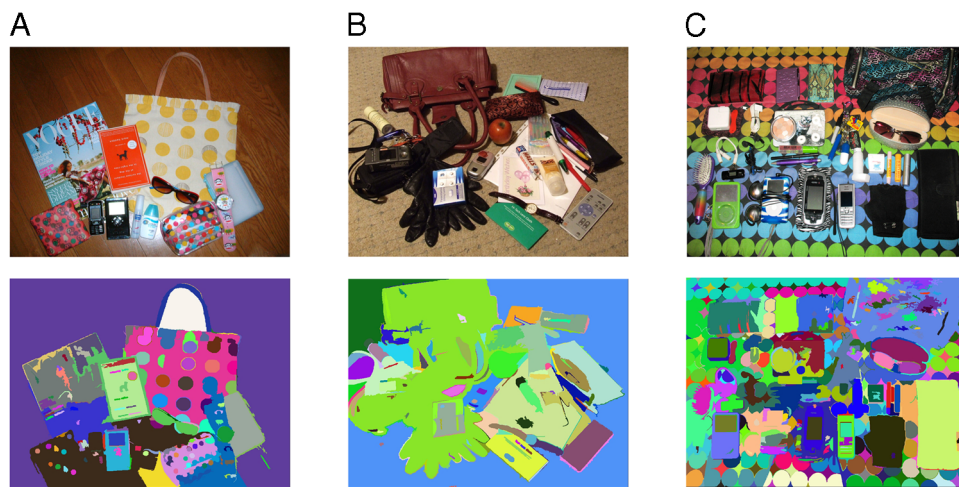


Figure 4. Three bag images and their segmentations ($k = 1000$). According to our measure, images A and B have similar amounts of clutter, but observers search image A faster than image B. Conversely, image C is judged to have more clutter than image B, but observers searched both images with similar speed.

the opposite trade-off. The next section describes the application of the clutter measure to these less natural but more controlled stimuli.

Ancillary test

As mentioned in the introduction, we had previously tested the idea that search times are determined by the number of objects in the image. As Figure 5 shows, our stimuli were photo-collages of common objects. These objects were classified as “simple” if they had uniform color (Figure 5, top) and as “compound” if they had multiple parts with different colors (Figure 5, bottom). The target was a food item. Displays consisted of 6, 12, or 24 items, and half of the displays contained a target. All of the distractors in a display were of one type, either simple or compound; the target was of either type.

If search times depend only on the number of objects in the display, then one would expect similar search times for displays composed of simple and compound distractors. Instead, we found that search times were longer for displays composed of compound distractors (Figure 6). We interpreted this result as suggesting that search times depend on the number of regions defined by bottom-up segmentation. That is, we assumed that simple objects would often be segmented as a single region, whereas compound objects would often be over-segmented into multiple regions. Because displays composed of compound objects contain more regions, observers should take longer to search them.

This interpretation is only qualitative; we did not try to count the number of regions in the compound and simple displays. As a rough estimate, however, one could try to estimate this number from the number of object parts:



Figure 5. Photo-collage stimuli from Bravo and Farid (2004b). Both stimuli contain 24 objects, but the objects in the bottom stimulus have more parts than the objects in the top stimulus.

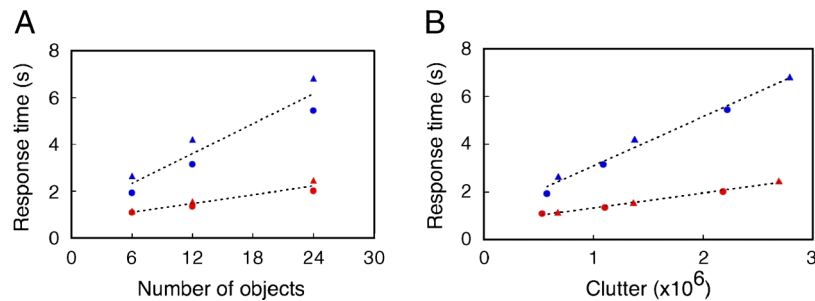


Figure 6. Search times plotted as a function of the number of objects (A) and as a function of the amount of image clutter (B). Red symbols correspond to the target-present condition, and blue symbols correspond to the target-absent condition. Circles correspond to simple distractors, and triangles correspond to compound distractors.

Since the compound objects had at least twice as many parts as the simple objects, compound displays should have twice as many regions as simple displays. By this reasoning, the slope of the search functions for compound displays should be twice that for simple displays. The actual slope ratio was closer to 1.4 (Figure 6A). This discrepancy might reflect a failure of our hypothesis, but it also might reflect problems with the region estimate. One potential problem is that because the objects were randomly arranged, occlusions may have altered the number of visible regions. That is, occlusions may have completely hidden some of the compound object parts, which were relatively small. At the same time, occlusions may have fragmented some of the simple objects. A better test of our hypothesis would involve a direct count of the regions in the displays rather than a count of the object parts used to make the displays.

To count the number of regions in the photo-collage stimuli, we segmented these images at the same six scales as the bag stimuli. We counted the regions produced by each segmentation and again found a power law relationship between this number and the segmentation scale. For these artificial stimuli, the average exponent of the power law was -1.13 . We fixed the exponent of the power law to this value and used the constant of proportionality as our measure of clutter. We then calculated the average clutter in each of the 12 types of photo-collage stimuli: simple vs. compound distractors; 6, 12, or 24 objects; and target present versus target absent. Finally, we replotted the response time data using average clutter as the independent variable. As Figure 6B shows, the clutter measure brings into register the results for the simple and the compound displays. Our clutter measure is clearly a better predictor of search times than is the number of objects.

General discussion

Vision is such complex problem that it seems most amenable to a reductionist approach. The strength of this

approach is obvious: With simplified stimuli and tasks, it is possible to formulate precise questions and to obtain unambiguous answers. The weakness of this approach is also obvious: Fractionating and simplifying a complex problem can alter it in essential ways. Using isolated objects as stimuli has allowed us to learn a great deal about visual search, but it has also allowed us to neglect some of the fundamental challenges posed by real images. Only recently have researchers begun to consider how image clutter might alter the process of visual search (Rosenholtz, Li, & Nankano, 2007; Wolfe, Oliva, Horowitz, Butcher, & Bompas, 2002), as well as other visual processes, like object recognition (Rolls, Aggelopoulos, & Zheng, 2003; Sheinberg & Logothetis, 2000).

To examine how image clutter affects visual search, it is necessary to measure it. The goal of this study is to devise a measure of clutter that is both intuitive and feasible. Our measure is based on the idea that the chunks for visual search are the regions formed by perceptual organization (Neisser, 1967). Perceptual organization is assumed to involve fast, bottom-up processes that exploit the statistical regularities found within objects (Brunswick & Kamiya, 1953). So although these processes do not access object memory, they produce regions that likely correspond to single objects (Elder & Goldberg, 2002; Fine et al., 2003). The basic phenomena of human perceptual organization were described early last century (Wertheimer, 1938), and subsequent research has revealed much about the underlying processes. Still, there is currently no fully integrated model of human perceptual organization. To define the regions in our images, we borrowed an image segmentation algorithm from the computer vision community. As a model of human perceptual organization, the segmentation algorithm is too simplistic: It does not take into account symmetry, collinearity, parallelism, and other grouping cues that humans use. Also, the algorithm makes decisions based only on local information, and so it may not always produce the optimal global segmentation. But the simplicity of this segmentation algorithm is also its strength; the algorithm is extremely efficient, and this makes it feasible to use on large sets of big, color images.

The clutter measure we developed using this algorithm has the very useful property of scale invariance. This property is useful because it allows us to study how clutter affects search even when we do not know the features that are involved in this task. For example, we can study the effect of clutter on search for a hairbrush, although we do not know whether the observer searches for coarse-scale features associated with the object's shape, for fine-scale feature associated with its bristly texture, or for both features simultaneously. We can apply our clutter measure to this task because it characterizes the clutter over a range of scales.

In addition to being useful, the scale invariance we have observed may reveal something about the structure of natural images. Many image properties, including number of regions, vary with scale as a power law (Field, 1987; Martin, Fowlkes, Tai, & Malik, 2001; Ruderman, 1997). The scale invariance that this implies has been explained by the fractal-like structure of objects (Mumford & Gidas, 2001). An object may have several parts, and each of these parts may have several more parts, and these parts may have surface patterns due to the effects of lighting and texture. If images of objects have structure at arbitrarily small scales, then this could explain the power law relationship we observed between the number of image regions and the scale of segmentation. This could also explain why the bag images all had similar power law exponents. These stimuli differed primarily in the number but not the types of objects they contained; if one image contained more objects than another, it would likely contain more structure at all scales.

When we applied the clutter measure to the photo-collage stimuli from our 2004 experiment, we found that these stimuli were also well fit by a power law, but the average exponent differed from that of the bag images. This is not entirely surprising given the artificial nature of the photo-collage stimuli. But even natural images might have different power law exponents depending on their content. Images of landscapes, for example, often have large expanses of water, grass, sand, or rock. These large textured regions have much fine-scale structure but little coarse-scale structure. Thus, the number of regions in a landscape image might fall off very steeply with increasing scale. To test this possibility, we applied the segmentation algorithm to 60 images depicting man-made objects (tools, cars, room interiors, and buildings) and 60 images depicting nature (plants and landscapes). The average exponent for the images of artifacts (-1.31 , $\sigma = 0.14$) was comparable to that for the bag stimuli (-1.32 , $\sigma = 0.13$), but the average exponent for the images of nature was more negative (-1.51 , $\sigma = 0.19$). If different types of images have different exponents, then our clutter measure will work best for images with similar content.

A full model of visual search in natural images must consider several variables besides clutter. We intentionally minimized the role of these other variables in our task, but in many search tasks they dominate performance.

For example, some search targets, such as stop signs or exit signs, are designed to be easily found in background clutter. The conditions that produce a salient target have been well studied and extensively modeled (Itti & Koch, 2000; Nothdurft, 2002).

Recently, Rosenholtz et al. (2007) proposed a model of clutter that focuses on target saliency. The model measures the range of simple features in an image and uses this range to predict whether a target added to the image is likely to attract attention. If the image has a limited range of features, then it is likely that an added target will be salient; but if the image has a wide range of features, then it is likely that an added target will not be salient. Note that the model predicts fast search times for an image composed of many, very similar objects because a target object added to such an image would be easy to find. This prediction of fast search times is less likely to apply, however, when the observer searches for one of the objects already in the image. It is this type of search, search for a nonsalient target, that requires a measure based on image regions.

This paper proposes a measure of clutter that can predict search times when the target is not salient and when the target's simple features and location are not known. The measure is intuitive and feasible. And because the measure is scale invariant, it can be applied to tasks in which the nature of the relevant image information is unknown. This makes the measure especially useful for studying search tasks that we know little about, such as the search for real objects in real scenes.

Acknowledgments

This work was supported by a Guggenheim Fellowship awarded to HF.

Commercial relationships: none.

Corresponding author: Mary Bravo.

Email: mbravo@camden.rutgers.edu.

Address: 311 North Fifth St., Camden NJ 08102, USA.

References

- Brainard, D. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436. [PubMed]
- Bravo, M. J., & Farid, H. (2004a). Recognizing and segmenting objects in clutter. *Vision Research*, 44, 385–396. [PubMed]
- Bravo, M. J., & Farid, H. (2004b). Search for a category target in clutter. *Perception*, 33, 643–652. [PubMed]
- Brunswik, E., & Kamiya, J. (1953). Ecological cue-validity of proximity and of other Gestalt factors. *American Journal of Psychology*, 66, 20–32. [PubMed]

- Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision*, 2(4):5, 324–353, <http://journalofvision.org/2/4/5/>, doi:10.1167/2.4.5. [PubMed] [Article]
- Felzenszwalb, P., & Huttenlocher, D. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59, 167–181.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A, Optics and Image Science*, 4, 2379–2394. [PubMed]
- Fine, I., MacLeod, D. I., & Boynton, G. M. (2003). Surface segmentation based on the luminance and color statistics of natural scenes. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 20, 1283–1291. [PubMed]
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489–1506. [PubMed]
- Martin, D., Fowlkes, C., Tai, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *Proceedings of the IEEE International Conference on Computer Vision*, 2, 416–425.
- Mumford, D., & Gidas, B. (2001). Stochastic models of generic images. *Quarterly of Applied Mathematics*, 59, 85–111.
- Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, 46, 614–621. [PubMed]
- Neisser, U. (1967). *Cognitive psychology* (pp. 86–104). New York: Appleton-Century-Crofts.
- Nothdurft, H. C. (2002). Attention shifts to salient targets. *Vision Research*, 42, 1287–1306. [PubMed]
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442. [PubMed]
- Rolls, E. T., Aggelopoulos, N. C., & Zheng, F. (2003). The receptive fields of inferior temporal cortex neurons in natural scenes. *Journal of Neuroscience*, 23, 339–348. [PubMed] [Article]
- Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter [Abstract]. *Journal of Vision*, 7(2):17, 1–22, <http://journalofvision.org/7/2/17/>, doi:10.1167/7.2.17.
- Ruderman, D. L. (1997). Origins of scaling in natural images. *Vision Research*, 37, 3385–3398. [PubMed]
- Sheinberg, D. L., & Logothetis, N. K. (2000). Noticing familiar objects in real world scenes: The role of temporal cortical neurons in natural vision. *Journal of Neuroscience*, 21, 1340–1350. [PubMed] [Article]
- Torrallba, A., Oliva, A., Castelano, 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]
- Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. Ellis (Ed. & Trans.), *A source book of Gestalt psychology* (pp. 71–88). London: Routledge & Kegan Paul.
- Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1, 202–238.
- Wolfe, J. M., Oliva, A., Horowitz, T. S., Butcher, S. J., & Bompas, A. (2002). Segmentation of objects from backgrounds in visual search tasks. *Vision Research*, 42, 2985–3004. [PubMed]