Vision Research 116 (2015) 142–151

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Clutter perception is invariant to image size

Gregory J. Zelinsky [a,b,*], Chen-Ping Yu [b]

[a] Department of Psychology, Stony Brook University, USA
[b] Department of Computer Science, Stony Brook University, USA

A R T I C L E I N F O

Article history:
Received 26 September 2014
Received in revised form 7 April 2015
Available online 14 May 2015

Keywords:
Visual clutter
Proto-objects
Image segmentation
Color clustering
Superpixels
Image size

A B S T R A C T

Two experiments evaluated the effect of retinal image size on the proto-object model of visual clutter perception. Experiment 1 had 20 participants order 90 small images of random-category real-world scenes from least to most cluttered. Aggregating these individual rankings into a single median clutter ranking and comparing it to a previously reported clutter ranking of larger versions of the identical scenes yielded a Spearman's $\rho = .953$ ($p < .001$), suggesting that relative clutter perception is largely invariant to image size. We then applied the proto-object model of clutter perception to these smaller images and obtained a clutter estimate for each. Correlating these estimates with the median behavioral ranking yielded a Spearman's $\rho = .852$ ($p < .001$), which we showed in a comparative analysis to be better than six other methods of estimating clutter. Experiment 2 intermixed large and small versions of the Experiment 1 scenes and had participants ($n = 18$) again rank them for clutter. We found that median clutter rankings of these size-intermixed images were essentially the same as the small and large median rankings from Experiment 1, suggesting size invariance in absolute clutter perception. Moreover, the proto-object model again successfully captured this result. We conclude that both relative and absolute clutter perception is invariant to retinal image size. We further speculate that clutter perception is mediated by proto-objects—a preattentive level of visual representation between features and objects—and that using the proto-object model we may be able to glimpse into this pre-attentive world.

1. Introduction

Everyone knows what clutter is; it is the typically negative percept resulting from the disordered organization of an excessive number of objects. Most previous work on clutter has focused on its consequences for task performance. The clearest example of this is the decrease in visual search efficiency accompanying an increase in clutter (Bravo & Farid, 2008; Henderson, Chanceaux, & Smith, 2009; Mack & Oliva, 2004; Neider & Zelinsky, 2011; Rosenholtz, Li, & Nakano, 2007). This effect of clutter on search efficiency led some researchers to suggest that clutter might be used as a surrogate measure of search set size (Neider & Zelinsky, 2011; Rosenholtz, Li, & Nakano, 2007), the number of objects appearing in a search display. This suggestion in turn led to the development of several computational methods for quantifying clutter (e.g., Bravo & Farid, 2008; Lohrenz et al., 2009; Rosenholtz, Li, & Nakano, 2007; van den Berg, Cornelissen, & Roerdink, 2009) so as to predict search efficiency in real-world scenes, stimuli in which a set of discrete objects cannot be defined objectively. A goal of our study is to further evaluate one of these methods for quantifying clutter—the proto-object model of clutter perception (Yu, Samaras, & Zelinsky, 2014).

There is a fundamental relationship between clutter and visual attention. Much attention research has been devoted to identifying those mental processes that can be performed simultaneously, without creating interference or performance costs, and those that result in performance costs when combined. The former have been termed *pre-attentive* and the latter *post-attentive* (or simply, *attentive*), referring to the fact that individual processes must be selected for serial execution so as to avoid incurring costs. This distinction largely shaped the massive literature on visual search (see Wolfe, 1998; for a review), but dates even farther back to the seminal attention studies using dichotic listening paradigms (see Pashler, 1998; for a review). In addition to a small set of basic visual features that can be extracted and used in parallel (Wolfe & Horowitz, 2004), our perception of clutter is likely pre-attentive; we seem able to effortlessly estimate how much "stuff" there is over a region of space (see Alvarez, 2011; for a review). This follows from the fact that clutter perception is likely derived from summary statistics computed over local pooling regions (Rosenholtz, Huang, & Ehinger, 2012; van den Berg, Cornelissen, & Roerdink, 2009), as has been proposed for crowding

* Corresponding author at: Psychology B-240, Stony Brook University, Stony Brook, NY 11794-2500, USA.
E-mail address: Gregory.Zelinsky@stonybrook.edu (G.J. Zelinsky).

http://dx.doi.org/10.1016/j.visres.2015.04.017

(Balas, Nakano, & Rosenholtz, 2009; Rosenholtz, Huang, Raj, et al., 2012). Setting aside for now the question of what units are actually summed, the assumption is that this summary is obtained pre-attentively and does not involve an actual count of discrete things; visual information is accumulated in parallel and summed to derive the clutter estimate. Indeed, our percept of clutter may be more than just the product of a pre-attentive process; it may be our perception of pre-attention.

Size also matters. Limiting this statement to vision, examples range from older work showing that larger shapes are given greater perceptual weight by the oculomotor system, causing saccades to land closer to bigger objects (Findlay, 1982), to recent work showing that common objects have a canonical size that affects their recognition speed and accuracy (Konkle & Oliva, 2011). Shifting focus from objects to scenes, another goal of our study asks how the retinal size of visual scenes impacts the perception of scene clutter.

Understanding the relationship between retinal image size and clutter perception is important. We live in a very cluttered world, filled with busy city streets, messy desks, and computer screens packed with icons numbering in the dozens if not hundreds. The sizes of these screens, however, have taken two trends. One has been to make computer monitors bigger so as to fit even more things in our field of view. The other has been to decrease the size of these screens so that we can put them in our purses and pockets and carry them wherever we go. This latter trend creates an obvious problem: given that the screens through which we increasingly interact with the world have become smaller, but the number of apps and other icons that we put on these screens has increased or stayed about the same, our world is becoming perceptually compressed into increasingly smaller spaces.

What are the consequences of this compression for our perception of clutter? One possibility is that perceived clutter might increase with decreasing retinal size. Objects in smaller scenes are closer together. If the absolute distance between objects affects perceived clutter, small scenes should be perceived as being more cluttered than larger ones. Another possibility is that retinal size doesn't matter for clutter perception, and that what is important is the number of perceived objects (or proto-objects). If so, decreasing the size of an image should not affect perceived clutter so long as this manipulation is not so drastic as to change the number of objects that are perceived.

2. Experiment 1

To investigate the relationship between retinal image size and perceived clutter we adopt the joint behavioral and computational approach reported recently by Yu, Samaras, and Zelinsky (2014). These authors asked participants to rank order 90 images of random category scenes from least to most cluttered. Note that a time-unlimited clutter ranking task is perfectly suited to the broader goal of our study, to demarcate the boundary between pre-attention and attention by obtaining an explicit estimate of clutter (pre-attention) that is minimally confounded with more goal-directed (attentive) tasks such as search, scene memory, or even free viewing. Yu et al. then modeled this clutter ranking by computing proto-objects for each scene and ordering them based on the number of proto-objects in each. Model success was assessed by correlating these behavioral and computational rankings. In Experiment 1 we used the scenes from Yu et al. to obtain another behavioral ranking of clutter, only these scenes were one-quarter the size of those used in the earlier study. This consistency in both stimuli and task allows the new behavioral clutter ranking to be compared directly to the one from Yu and colleagues so as to determine the effect of retinal image size on clutter perception. Additionally, we test the proto-object model of clutter perception from Yu et al. to determine how well it predicts the effect of changing image size on these behavioral clutter rankings, and compare these predictions to those from other models of clutter perception.

2.1. The proto-object model of clutter perception

Central to our approach is the suggestion that clutter perception can be predicted by how much stuff appears in a scene, where "stuff" is quantified in terms of locally similar image features that become merged into perceptual fragments that we refer to as *proto-objects*. This definition of a proto-object loosely follows the original usage of the term as coined by Rensink and Enns (1995). These authors conceptualized proto-objects as being relatively small and highly volatile clusters of visual features, created by the pre-attentive application of local grouping processes, from which more extensive visual object representations are ultimately built (see also Rensink, 2000). We subscribe to all of these defining properties. Indeed, rather than substantively reconceptualizing what a proto-object is, we see our work as contributing to the further quantification of this construct and its application to real-world objects and scenes.

Several models of attention have appealed to proto-objects in the context of visually complex stimuli. These have taken two basic approaches. One has been to redefine proto-objects as regions of high image salience. For example, Walther and Koch (2006) used a saliency map (Itti & Koch, 2001; Itti, Koch, & Niebur, 1998) to identify salient points in an image, then spread the activation from each back through the intensity, orientation, and color feature maps to obtain a saliency-based segmentation that they referred to as proto-objects (see also Borji, Sihite, & Itti, 2013; Russell et al., 2014, for related approaches). Using this method, Nuthmann and Henderson (2010) compared these salient proto-objects to objects hand labeled from a scene to see which could better describe the *preferred viewing location* (PVL) of behavioral participants performing various scene inspection tasks. They found that proto-objects were less successful than actual objects in describing the PVL effect, at least for proto-objects defined by feature salience. Another approach has been to use color blob detectors (Forssén, 2004) applied directly to unprocessed images (Wischnewski et al., 2009) and video (Wischnewski et al., 2010) to define proto-objects. These proto-objects are then combined with the Theory of Visual Attention (TVA, Bundesen, 1990) to produce a priority map that is used to predict allocations of visual attention.

Our method for deriving proto-objects differs from previous methods in at least two key respects. First, saliency is not considered in our method. We quantify how much stuff there is in a scene, regardless of whether this stuff is salient or not. Second, rather than using blob detectors to segment proto-objects from an image, which at best restricts the shape of proto-objects to coarse elliptical regions, we use more sophisticated image segmentation techniques developed in the computer vision literature. Specifically, we combine superpixel image segmentation (Liu et al., 2011) with a clustering method (Comaniciu & Meer, 2002) to merge featurally-similar superpixels into proto-objects. Note that superpixels themselves are atomic regions of an image containing pixels that are similar in some feature space, but superpixel methods tend to over-segment images. For this reason we treat superpixel segmentation as a preprocessing stage, one that we follow with a merging stage in which neighboring superpixels that are similar in color are combined to create more spatially extended image fragments that we define as proto-objects. Our model then simply counts the number of proto-objects in an image to obtain an estimate of its clutter (Yu, Samaras, & Zelinsky, 2014).