The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements

John Henderson, Myriam Chanceaux, Tim Smith

The Influence of Clutter on Real-World Scene Search: Evidence from Search Efficiency and Eye Movements

John M. Henderson[1], Myriam Chanceaux[2], and Tim J. Smith[1]

[1]University of Edinburgh, [2]University of Grenoble

Running Head: Clutter and Scene Search

Address correspondence to:

John M. Henderson
Psychology Department
7 George Square
University of Edinburgh
Edinburgh EH8 9JZ
United Kingdom

Abstract

In the present study we investigated the relationship between visual clutter and visual search in real-world scenes. Specifically, we investigated whether visual clutter correlates with search performance in scenes as assessed both by traditional behavioral measures (response time and error rate) and by eye movements. Our results demonstrate that clutter is related to search performance in scenes. These results hold for both the traditional search measures and for eye movements. The results suggest that clutter may serve as an image-based proxy for search set size in real-world scenes.

A critical task in human vision is to locate a sought-after object in the visual environment. In traditional research on visual search, a key variable has been shown to be set size, or the number of items in the display. In search within well-controlled psychophysical displays, set size accounts for a large proportion of the variance in search time (Wolfe, 1998).

However, in pictures of natural scenes, it is far more difficult to identify *a priori* the number of "items" in the image. In fact, Neider and Zelinsky (2008) have argued that the concept of object-based set size has no clear meaning in real-world scenes (see also Bravo & Furid, 2004). Rosenholtz and colleagues have proposed that "clutter" may provide a stand-in or proxy for set size in natural scenes (Rosenholtz, Li, & Nakano, 2007; Rosenholtz, Li, Mansfield, & Jin, 2005). Rosenholtz et al. (2007) noted that in real-world scenes, set size would seem to be better captured by something like object parts than by whole items. Rosenholtz et al. operationalized clutter using three image-based properties: feature congestion, sub-band entropy, and edge density. Feature congestion can be thought of as local variability in specific image features such as color, orientation, and luminance. Sub-band entroy can be thought of as a measure of the efficiency with which an image can be encoded while maintaining perceived image quality (similar to the method used in JPEG image compression). Finally, edge density is a count of the number of edges in the image. An advantage of each of these measures is that they can easily be determined for complex scenes (compared to object-based set size) and can be precisely quantitatively specified.

Rosenholtz et al. (2007) conducted a series of visual search experiments in complex full-color images to compare the relationship of each of the three clutter

measures to search time. The experiments used images of geographic maps, and participants searched for either embedded Gabor patches or small arrows in each image. Though feature congestion was able to account for the additional effect of color variability on search performance that the other measures did not capture, all three measures predicted search times relatively well, supporting the idea that image-based clutter can provide a stand-in for set size in scene search.

In the present study, we sought to extend the research in this area in two important ways. First, we wanted to investigate the relationship between clutter and search in photographs of real-world scenes. Real-world scenes have particular properties that differentiate them from other complex images (Henderson, 2003; Henderson & Ferreira, 2004). Second, because search in complex scenes typically requires eye movements (Castelhano & Henderson, 2005; Henderson, Weeks, & Hollingworth, 1999), we wanted to examine the influence of clutter on both overall search time and on eye movement behavior. We specifically asked whether clutter, as assessed with the three image-based measures proposed by Rosenholtz et al. (2007), correlate with search performance in real-world scenes as assessed both by search performance and by eye movement.

To investigate these issues, we used a difficult visual search task in which viewers are asked to search for and discriminate small Ts and Ls embedded in photographs of real-world scenes (Brockmole & Henderson, 2006; Brockmole, Castelhano, & Henderson, 2006; see also Wolfe, Oliva, Horowitz, Butcher, & Bompas, 2002). If search becomes more difficult in more cluttered real-world scenes, as has been found in complex images of maps (Rosenholtz et al., 2007), then we should find see a negative relationship between degree of clutter and search efficiency.

Methods

Participants. Sixteen Edinburgh University first year Psychology students took part for £6 each.

Apparatus. Eye position was sampled with an SR Research Eyelink 1000 eyetracker sampled right at 1000Hz. The experiment was controlled with Experiment Builder software. Viewing was binocular but only the right eye was tracked. Saccades were detected using SR Research's saccade detection algorithm. This algorithm used a 17-sample model with a velocity criterion of 50°/s and a minimum amplitude of 0.5°. Fixations were defined as any time when the eyes were not in a saccade or blink. No minimum duration criterion was set for fixations. Images were presented on a 21" CRT with a refresh rate of 140 Hz.

Stimuli. Participants were presented wih 60 unique full-color 800 x 600 pixel 24 bit photographs of real-world scenes (subtending a visual angle of 25.7° x 19.4°) from a variety of sources (e.g., on-line and collections of lab members) depicting a variety of indoor and outdoor scene categories. Scenes did not include people or animals in order to minimise distractions from the search task. A gray letter T or L (Arial 9-point font) was superimposed onto the scenes (Brockmole & Henderson, 2006). Letters were used to ensure that the target location could not be predicated by local or global scene content. The letter subtended 0.3 deg horizontally and vertically. Placement of each search target was random although positions close to the screen centre were avoided. To ensure that the search target would be identifiable and would not be mistaken for an edge in the image, any search targets randomly positioned on the border between regions differing in color or texture were jittered a few pixels to a nearby constant region.

Procedure. Participants were instructed to search the scenes as quickly as possible for a small grey superimposed letter. Once found, participants indicated whether the target was a T or L by pressing the appropriate button on a joypad (Microsoft Sidewinder). Before beginning the main experiment, participants were shown two example scenes with the search targets highlighted and then given three practice trials. The main experiment only began once participants understood the search task and the appearance of the search targets. During the main experiment, in the event that the letter was not found, the trial timed out after about 6000ms.[1]

Results

To investigate search behavior as a function of clutter, we computed the three measures of clutter proposed by Rosenholtz and colleagues using the algorithms and code described by Rosenholtz et al. (2007). Figure 1 presents a visual illustration of feature congestion and edge density for two of the scenes used in the present study. (It is not possible to convert sub-band entropy into an equivalent graphical representation.)

[1] An additional manipulation was included in each trial, but this was irrelevant for the purposes of the current study. This manipulation involved presenting a brief luminance onset following the first 1000ms of scene exploration. After the onset, the scene remained in view for 5000ms. The data presented here did not include saccades and fixations immediately following the sudden onset.

Figure 1. Examples of two scenes used in the experiment (top panels), and a depiction of edge density (middle panels) and feature congestion (bottom panels) for those scenes. The edge density and feature congestion maps for the two scenes were each generated using the same scales.

To assess the relationship between each measure of clutter and search, we examined seven dependent measures derived from search behavior and analyzed them

as a function of the three clutter indexes: search efficiency assessed by search response time and missed targets; search initiation time measured by latency for first saccade following onset of the scene; mean fixation duration over the entire viewing episode and for the first 1000ms; and mean saccade amplitude over the entire viewing episode and for the first 1000ms. In a secondary analysis, we examined fixation duration as a function of the local clutter value around each fixation point.

Search Efficiency. We first examined two standard measures of search efficiency, response time and error rates, as a function of clutter. These results can be seen in Figures 2 and 3. Each point in the figures (and entered into the statistical analyses) represents the value of the dependent variable averaged over participants for one scene. For *response time*, all three measures were significantly related to response time (R^2=.27, .18, and .28 for feature congestion, sub-band entropy, and edge density respectively, *p*'s<.001), though numerically the correlations for feature congestion and edge density were larger. In the case of *search failure*, defined as trials in which the target was not located, all three clutter measures were significantly related to search performance (R^2= 0.16, 0.28, and 0.29, all p's < 0.005).

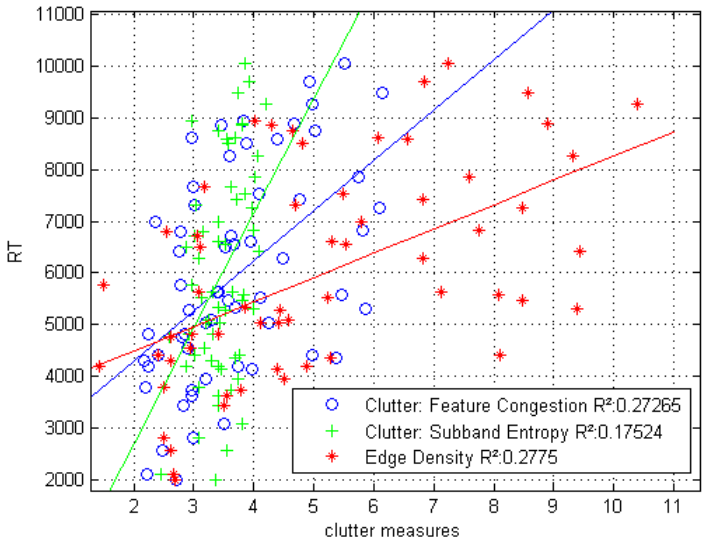

Figure 2. Search response time (ms) as a function of scene clutter.

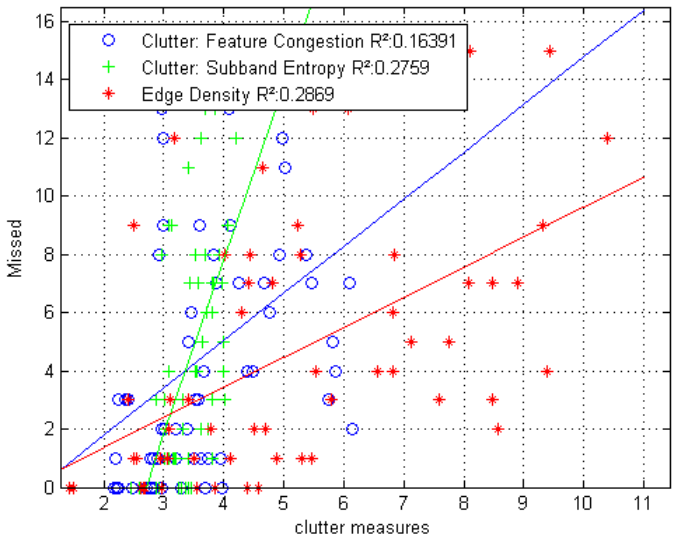

Figure 3. Search failure (missed targets) as a function of scene clutter.

Eye Movements. During search in real-world scenes, viewers typically make a series of eye movements as they search (Castelhano & Henderson, 2005). Therefore, in addition to response time and error rates, we also investigated the degree to which eye movements during search reflect visual scene clutter. Figure 4 presents a representative scan pattern of one participant in one search scene.

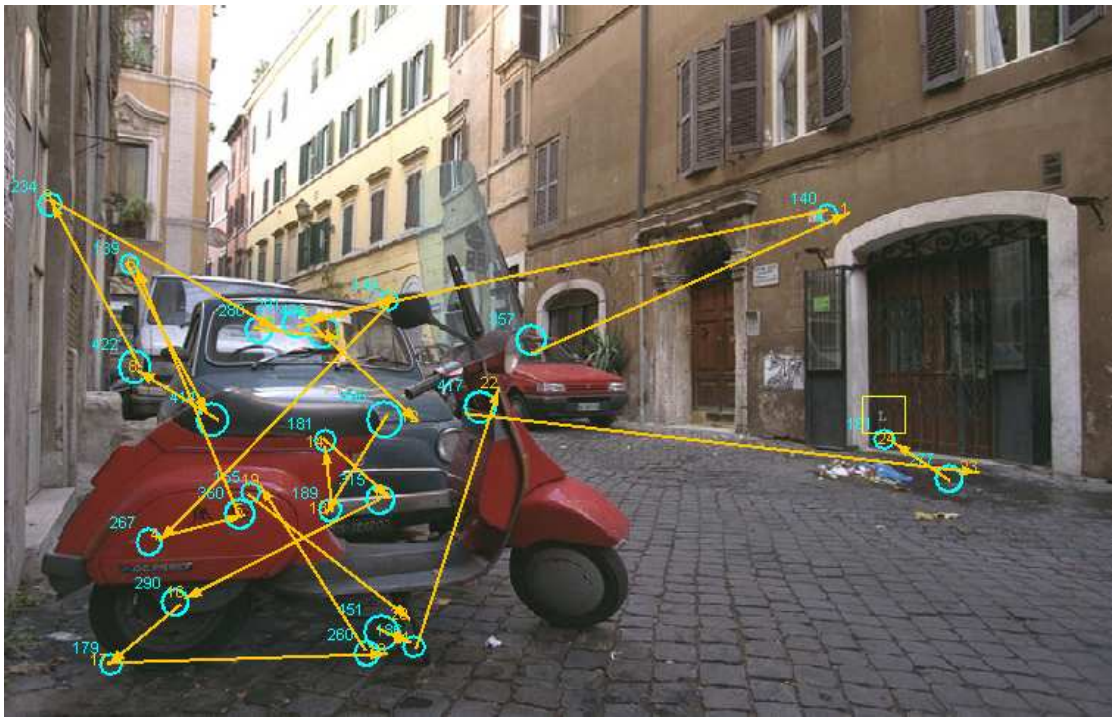

Figure 4. Example of a scan pattern during search. The blue circles depict fixations (circles scaled by fixation duration) and the yellow arrows depict saccades. The yellow box at the end of the scan pattern surrounds the target letter (the box was not presented to the subject).

Search Initiation Time. We first examined the influence of clutter on the latency of the initial saccade from the appearance of the scene. This measure serves as an index of the time taken by the viewer to begin the first scanning eye movement in the scene. Increasing set size has previously been shown to increase the time taken to search an array and the duration of the initial fixation (Zelinsky, 2001). We might expect that the time to begin the search would increase with clutter since it might take more time to choose a potential first saccade target. In agreement with this hypothesis, initial saccade latency showed a significant positive relationship with both sub-band entropy and edge density (R^2=.11 for both clutter measures, p's<.05); feature

congestion did not produce an effect (R^2=.01, ns). These results are shown in Figure 5.

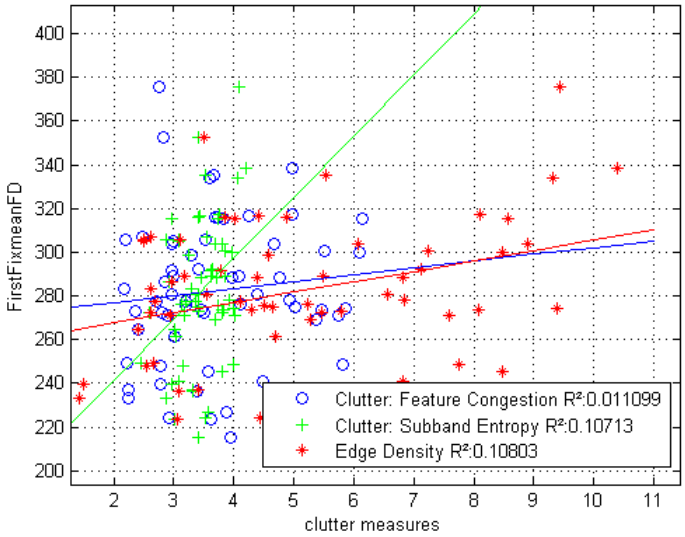

Figure 5. Search initiation time (ms) as a function of scene clutter.

Fixation Duration. To investigate the relationship between clutter and eye movements during search, we examined mean fixation durations. Given that fixation durations reflect moment-to-moment processing difficulty during both scene memorization Henderson & Pierce, 2008; Henderson & Smith, in press) and scene search (Henderson & Smith, in press), we might expect that a more cluttered scene would lead to longer average fixation durations. To test this prediction, we looked at the influence of clutter on mean fixation durations for the 5534 fixations within the first 1 s of search (Figure 6).[2] We observed a relationship between clutter and fixation duration for all three measures of clutter (R^2=.08, .17, and .14 for feature congestion, sub-band entroy, and edge density respectively, *p*'s<.05 for feature congestion and edge density, p<.001 for sub-band entropy).

[2] The data pattern did not differ when fixations over the entire trial were included, but the onset at 1 s could potentially affect the next few fixations as shown by Brockmole & Henderson, 2006, so we based the reported analyses only on fixations prior to the onset event.

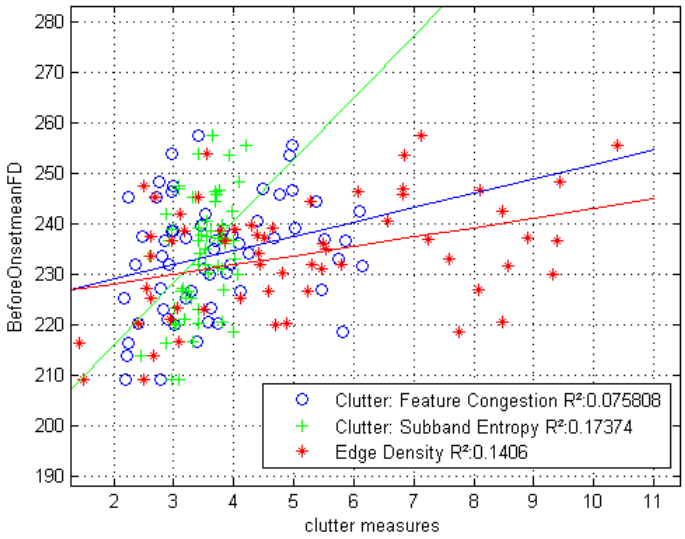

Figure 6. Mean fixation duration over first 1 s of viewing as a function of scene clutter.

Saccade Amplitude. We also examined the mean saccade amplitude as a function of the three clutter measures. As with fixation duration, we looked at initial saccade amplitude and saccade amplitude over the first 1 s of viewing. Unlike fixation duration, there was no relationship between clutter and saccade amplitude for any measure of clutter or any measure of amplitude (all $R^2 < .025$, ns).

Influence of Local Clutter on Fixation Duration. Given that we observed an influence of global scene clutter on fixation durations, we sought to examine the nature of this effect more carefully. Specifically, we examined the degree to which clutter around the current fixation point determined the duration of that fixation. To investigate this question, we defined square regions of two sizes (30 pixels and 100 pixels; approximately 1 and 3.3 deg of visual angle respectively) around each fixation point and determined the fixation duration for each of the 5534 fixations in the first 1 s of viewing. In contrast to global clutter, we observed no relationship between local clutter and fixation durations for regions of either size (all R^2=.00, ns; see Figure 7).

Together with the global clutter data, these results demonstrate that fixation durations are influenced by global scene clutter even when they are not influenced by local clutter around fixation.

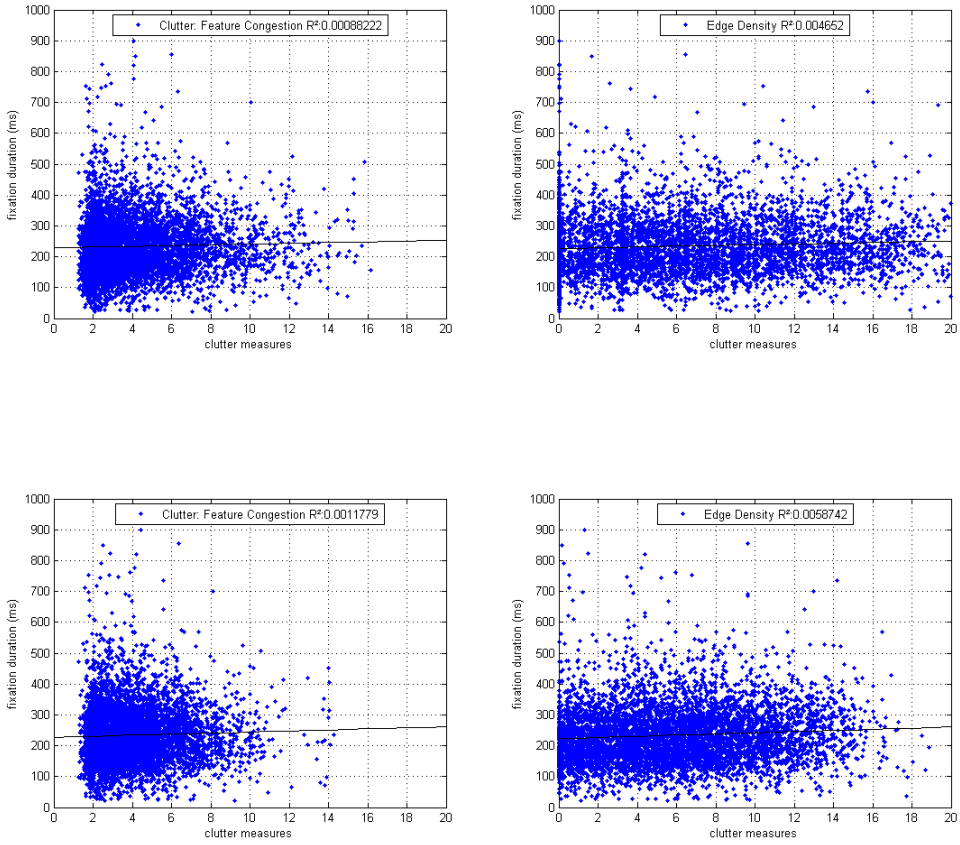

Figure 7. Fixation duration as a function of scene clutter surrounding the fixation point, for the first 1 s of viewing. Top panels show data for a region of 1 deg around fixation and bottom panels show a region of 3.3 deg. Left panels show feature congestion and right panels show edge density.

Influence of Local Clutter on Fixation Location. Finally, we examined the degree to which clutter around the current fixation point changed as a function of ordinal fixation number (1st, 2nd, 3rd, etc. fixation) for the first 6 fixations produced by participants following scene presentation. We again defined square regions of two sizes (30 pixels and 100 pixels; approximately 1 and 3.3 deg of visual angle

respectively) around each fixation point. We then determined the average degree of clutter for each fixation (blue lines in each panel of Figure 8) and for a random sample of locations within each scene (green lines). We found that the first viewer-determined fixation (Fixation 1 on the blue line in each panel of Figure 8) tended to be centered on a region of significantly higher clutter than would be predicted by chance (feature congestion, 1 deg,: $t(15)=9.774$, $p<.001$; 3.3 deg.: $t(15)=7.516$, $p<.001$; edge density, 1 deg.: $t(15)=7.085$, $p<.001$; 3.3 deg.: $t(15)=6.43$, $p<.001$). By the next fixation the local clutter is indistinguishable from the baseline. These results suggest that viewers start their search with more cluttered scene regions. This tendency could be a conscious strategy to initially exclude regions of the scene in which the search target would be harder to find. However, the tendency could also be an artifact of saliency as the clutter measures used, feature congestion and edge density are constructed from the same visual features as used in models of visual saliency (e.g. Itti & Koch, 2001). The local saliency around fixation has been shown to be significantly higher for the first viewer-determined fixation following scene presentation than subsequent fixations (Parkhurst, Law, & Niebur, 2002; Foulsham & Underwood, 2008). This relationship between clutter and saliency, both in terms of how the factors are modeled and how we, as viewers process and respond to them requires future investigation.

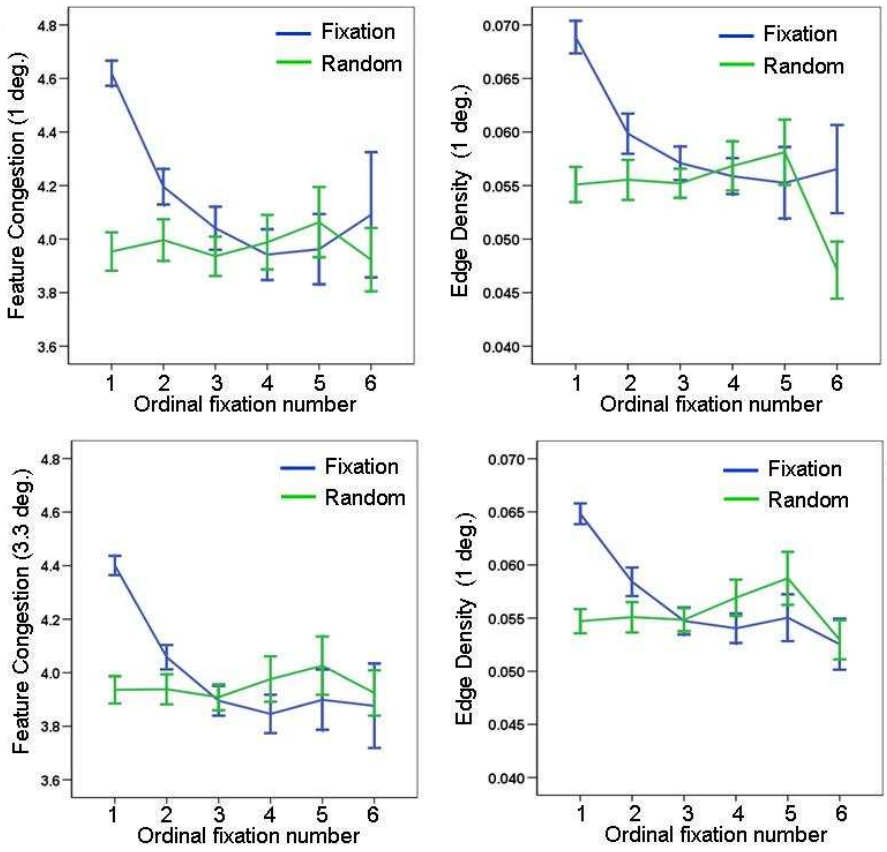

Figure 8. Average local clutter surrounding the fixation point as a function of ordinal fixation number, for the first 6 fixations. Left column = feature congestion, right column = edge density. Top row = clutter for 1 degree region around fixation, bottom row = clutter within a 3.3 degree region. Blue curves represent clutter around actual fixations. Green curves represent clutter around randomly chosen locations within the same scene. Error bars show 1 standard error of the mean.

General Discussion

The purpose of the present study was to investigate whether clutter, an image-based measure of visual complexity, can serve as a proxy for set size in real-world scene search. Rosenholtz et al. (2005; see also Bravo & Farid, 2008) observed a relationship between several measures of clutter and search in complex images of

maps. We sought to extend this work to search in photographs of real-world scenes, using both overall search time and eye movement behavior as dependent measures.

Overall, although there was some variability across clutter measures depending on the specific analysis, our results demonstrate that clutter does correlate with both global search efficiency (measured by search time and search failure) as well as with eye movement behavior during search. The latter result is novel and provides the first direct evidence that eye movements during scene search are influenced by the degree of clutter present in the scene.

It is interesting that edge density, despite failing to capture color variability, does not do significantly worse than feature congestion and sub-band entropy in predicting search efficiency. In fact, edge density was the only clutter measure to significantly correlate with all of the reported dependent measures. This effect is generally consistent with the observation that eye movements during scene viewing are very similar for color and gray-scale versions of the same pictures (Henderson & Hollingworth, 1998). The edge density result is also interesting in light of the finding that high spatial frequency contrast (i.e., edges) correlates with fixation location in real-world scenes better than luminance contrast (Baddeley & Tatler, 2006), and the finding that color is not a strong correlate of fixation location (Tatler, Baddeley, & Gilchrist, 2005). As in the case of feature congestion, visual saliency is a more complex measure than edge density because it takes additional image variability including color into account. The present results converge with those Baddeley and Tatler (2006) and Tatler et al. (2005) in suggesting that the most important image property for predicting search efficiency at both a macro (e.g., response time) and micro (e.g., eye movement) level of analysis may be edges.

The influence of clutter in the present experiment, though statistically significant, generally accounted for a relatively small amount of the variance in each of the measures of search. Also, the influence of clutter was relatively small compared to the influence of set size typically observed in standard visual search tasks (Wolfe, 1998). Why was the influence of clutter not more pronounced? We suspect that there may be several reasons. The measures of clutter proposed by Rosenholtz and used here may be only an approximation of perceived clutter, which may also take into account other factors beyond those included in the three measures. For example, other image features such as contrast, crowding, masking, and so forth are very likely play a role. Higher-level image features related to scenes like those proposed by Torralba and Oliva (2003) in their spatial envelope theory (e.g., degree to which the scene is open-closed, natural-artificial, and near-far) could also be important for natural scenes. Finally, higher-level cognitive factors such as the semantic similarity of the objects in the scene, scene coherence, the degree to which the scene is familiar, and so on may all play a role in search efficiency (Henderson, 2007). These higher-level factors are clearly more difficult to capture in image-based measures. Both additional image features and higher-level factors would contribute error variance and reduce the correlations of search with clutter.

In summary, we found that clutter correlates with search performance in real-world scenes. Furthermore, we have provided the first evidence that clutter also predicts eye movement characteristics during real-world search. These data converge with those presented by Rosenholtz et al. in showing that an image-based proxy for search set size can be related to search performance in real-world scenes.

Acknowledgements

We thank Ruth Rosenholtz for generously providing her Matlab code for Feature Congestion, Sub-band Entropy, and Edge Density, and for her helpful discussion of this research. We also thank the other members of the Visual Cognition Lab for their critical comments. This project was supported by a grant from the Economic and Social Science Research Council of the UK (RES-062-23-1092) to JMH.

References

Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: a Bayesian system identification analysis. *Vision Research*, 46, 2824-2833.

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

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.

Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition, 13, 99-108*.

Brockmole, J. R., Castelhano, M. S., & Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 699-706.

Castelhano, M. S., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition, 12,* 1017-1040.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision, 8*, 1-17.

Henderson, J. M. (2003). Human gaze control in real-world scene perception. *Trends in Cognitive Sciences, 7*, 498-504.

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science, 16*, 219-222.

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson and F. Ferreira (Eds.), *The interface of language, vision, and*

action: Eye movements and the visual world (pp. 1-58). New York: Psychology Press.

Henderson, J. M., & Hollingworth, A. (1998). Eye Movements during Scene Viewing: An Overview. In G Underwood (Ed.), *Eye Guidance while Reading and While Watching Dynamic Scenes*. (pp. 269-293). Oxford: Elsevier.

Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review, 15*, 566-573.

Henderson, J. M., & Smith, T. J. (in press). How are eye fixation durations controlled during scene viewing? Evidence from a scene onset delay paradigm. *Visual Cognition*.

Henderson, J. M., Weeks, P. A. Jr., & Hollingworth, A. (1999). Effects of semantic consistency on eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 210-228.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews, Neuroscience*, 2, 194–203.

Neider, M. B., & Zelinsky, G. J. (2008). Exploring set-size effects in scenes: Identifying the objects of search. *Visual Cognition, 16*, 1-10.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42, 107–123.

Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005). Feature congestion, a measure of display clutter. *SIGCHI* (pp. 760-770). Portland, Oregon.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision, 7(2):17*, 1-22.

Tatler, B. W., Baddeley, R. J., & Gichrist, I. D. (2005) Visual correlates of eye movements: Effects of scale and time. *Vision Research,* 45 (5), 643-659.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems, 14,* 391-412.

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. J., & Bompas, A. (2002). Segmentation of objects from backgrounds in visual search tasks. *Vision Research, 42,* 2985-3004.

Zelinsky, G. J. (2001) Eye movements during change detection: Implications for search constraints, memory limitations, and scanning strategies. *Perception & Psychophysics, 63,* 209-225.