Informativeness of visual features guides search

Alex D. Hwang, Emily C. Higgins and Marc Pomplun*

Department of Computer Science, University of Massachusetts at Boston,

Boston, Massachusetts 02125, USA

*Corresponding author

Research Report

Revision submitted on December 5, 2008

Word count for abstract: 87

Word count for main text including acknowledgments: 1345

Number of references: 16

Abstract

While visual search is known to be heavily guided by visual features similar to those in the search target, little is known about how we select the feature dimensions – such as luminance or orientation - to guide search so efficiently. Here we introduce a novel measure of the informativeness of individual feature dimensions for a given search task. By analyzing human eye movements during search in real-world scenes, we show that guidance is heavily determined by the statistical differences in informativeness across feature dimensions for everyday search.

The ability to quickly locate objects in visual space is crucial for the tasks of everyday life. Besides contextual inferences as to where search 'targets' may be (e.g., "cars are more likely found on streets than on trees", see Neider & Zelinsky, 2006; Torralba, Oliva, Castelhano & Henderson, 2006), it is the guidance of visual attention by low-level visual features of the target such as its color or texture (for instance, restricting search to just the blue items in the scene when the target is known to be blue) that makes search so efficient (e.g., Chen & Zelinsky, 2006; Rutishauser & Koch, 2007; Wolfe, 1994; 1998). However, since targets usually contain a variety of features along several feature dimensions, how do we weight these dimensions for guiding attention?

Two major factors need to be considered: First, the physiology of the visual system - features in individual dimensions are processed by distinct neural pathways, leading to differences in the latency, resolution, and accuracy of their perception (e.g., Moutoussis & Zeki,1997; Hegdé, 2008). Attentional control may rely on the dimensions that can be processed most efficiently in order to maximize search performance. Second, the informativeness of feature dimensions - for instance, when searching for a blue object, color is highly informative if there is only one blue object in the scene and is less informative if most objects in the scene are blue. Previous studies have suggested that in a given display observers facilitate search by attending less to a more frequent target feature, despite the need for assessing cross-dimensional informativeness at search onset (Pomplun, 2006; Shen, Reingold & Pomplun, 2000).

In order to study how the visual system tunes guidance across feature dimensions, we had human subjects perform visual search in real-world displays. The influence of contextual factors was reduced by rotating the displays and using randomly chosen cutouts from the displays as search targets (Figure 1a), while preserving the natural low-level features (see Pomplun, 2006). For each image, subjects first memorized the target and then searched for it in the large display while their eye movements were recorded. During search tasks, eye movements closely reflect shifts of attention (Findlay, 2004; Motter & Holsapple, 2007) and can reveal their bias toward visual features in the display that match those in the target, thereby indicating visual guidance (Findlay, 1997; Navalpakkam & Itti, 2007; Pomplun, 2006; Rutishauser & Koch, 2007; Shen, Reingold & Pomplun, 2000).

----- Insert Figure 1 about here -----

Method

We used 160 colored photographs of real-world scenes (800×800 pixels, 13° visual angle), randomly rotated by 90°, 180°, or 270°, as search displays. Search targets (64×64 pixels, 1°) were chosen randomly from the rotated displays, excluding the central screen region of 3°×3°. Several targets were newly chosen to avoid uninformative or semantically rich locations.

Thirty subjects aged 19 to 35 viewed the stimuli on a 19-inch Dell P992 monitor (1280×1024 pixels at 85 Hz). Their eye movements were tracked using an EyeLink-II (SR Research Ltd., Canada) system with a sampling frequency of 500 Hz and accuracy of approximately 0.5°. Subjects performed four blocks of 40 trials in which they first viewed the target at the center of the screen for two seconds, followed by the search display for a maximum duration of six seconds. The subjects' task was to find the target, fixate on it and press a button on a game-pad to terminate the trial. Fixation-density maps were generated by convolving the distribution of fixations with a 2D Gaussian kernel ($\sigma = 1°$ to approximate the human fovea size). For each trial, we excluded from analysis the initial and final three fixations due to their strong bias toward central, conspicuous image features and toward the search target, respectively, which would have diluted the measurement of feature guidance (Pomplun, 2006).

Low-level visual features along eight dimensions, chosen for their relevance to the processing of texture, shape, and color, were measured within a 64×64 pixel window at 48×48 evenly spaced display positions and in the target. Six of these dimensions were represented by eight-bin feature histograms: The red-green, blue-yellow, and luminance dimensions of the Derrington-Krauskopf-Lennie (DKL) color space (Derrington, Krauskopf & Lennie, 1984), elevation of spatial frequency bands, orientation of edges (cf. Pomplun, 2006), and luminance gradient (average luminance differences between neighboring pixels in all eight directions). Target-similarity of a local display position along these dimensions was computed by intersecting the display and target histograms after scaling them linearly to set their maximum values to one. Two additional dimensions, luminance contrast (standard deviation of luminance) and luminance entropy (Shannon entropy of luminance) were computed as scalar variables. Their target similarity was computed as the negative absolute difference between their values for the target and the local display area.

Visual guidance by a given dimension was obtained as the Pearson correlation between target similarity and fixation density at all measurement positions across all displays. As a control measure, the Receiver Operating Characteristic (ROC) was computed by finding a set of thresholds with 0%, 1%, …, 100% of the display area having target-similarity values above them (Tatler, Baddeley & Gilchrist, 2005). For each threshold, the proportion of all subjects' fixations hitting above-threshold display areas was measured. When plotting this fixation proportion as a function of the above-threshold display proportion, the area below the function is the ROC measure. A value of 0.5 indicates prediction of fixation density by target similarity at chance level (no guidance), whereas 1 is the theoretical maximum for prediction and guidance.

Results and Discussion

We collapsed all subjects' gaze fixations to create fixation-density maps (Figure 1b). Furthermore, for each display we created eight target-similarity maps indicating the similarity of each display location to the search target along eight selected dimensions (see Methods). All dimensions were found to exert guidance (Figure 2a), defined as the spatial correlation between their target-similarity maps and the corresponding fixation-density maps (all rs > 0.18, ps < 0.0001). As discussed above, a more informative dimension is indicated by a smaller proportion of the display being similar to the target in that dimension – or a larger proportion differing from it. We thus defined informativeness as the proportion of the display that differs from the target, i.e., whose target similarity is below 50% of its maximum. Across dimensions, mean informativeness and mean guidance showed a very strong positive correlation ($r = 0.97$, $p < 0.0001$). To demonstrate the independence of this result from any specific guidance measure, we also applied the common ROC measure (cf. Tatler, Baddeley & Gilchrist, 2005) and obtained similar results ($r = 0.96$, $p < 0.0005$).

----- Insert Figure 2 about here -----

Within each dimension, guidance was positively correlated with informativeness (all *r*s ranging from 0.18 to 0.42, all *p*s < 0.05), indicating some adaptation of guidance to informativeness in individual trials. To test the hypothesis that guidance by a feature dimension entirely depends on its informativeness in a given display, we analyzed guidance for each dimension in those 80 displays in which it was most informative. If guidance were completely determined by informativeness in individual displays, we would expect the informativeness-guidance markers to lie on the same regression line as in Figure 2a. However, as shown in Figure 2b, guidance in these displays remained significantly below this prediction, $t(7) = 4.16$, $p < 0.005$. Guidance due to informativeness is therefore accounted for partially, but not entirely, by informativeness within the context of specific search scenes.

The present data indicate that visual guidance it is heavily determined by the informativeness of feature dimensions for real-world search, conceivably through long-term adaptation. While guidance can be adapted to exploit informativeness in individual search displays, this mechanism cannot entirely override the long-term bias. To allow efficient search, the assessment of informativeness in individual tasks may be rather coarse, and guidance may rely to a large extent on the long-term heuristic. Since average real-world informativeness of a feature dimension is a powerful predictor of its contribution to guiding search, we conclude that either guidance is largely independent of physiological bias, or our visual system evolved to exploit informativeness for efficient search.

Acknowledgments

The project was supported by Grant Number R15EY017988 from the National Eye Institute to M. P.

References

Chen, X. & Zelinsky, G. J. (2006). Real-world visual search is dominated by top-down guidance. *Vision Research*, 46, 4118–4133.

Derrington, A. M., Krauskopf, J. & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *Journal of Physiology,* 357, 241-265.

Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research*, 37, 617–631.

Findlay, J. M. (2004). Eye scanning and visual search. In J. M. Henderson & F. Ferreira (Eds.), *The Interface of Language, Vision, and Action: Eye Movements and the Visual World,* New York: Psychology Press, 135–159.

Hegdé, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology*, 84, 405-439.

Motter, B. C., & Holsapple, J. (2007). Saccades and covert shifts of attention during active visual search: Spatial distributions, memory, and items per fixation. *Vision Research*, 47, 1261–1281.

Moutoussis, K. & Zeki, S. (1997). Functional segregation and temporal hierarchy of the visual perceptive systems. *Proceedings of the Royal Society London B*, 264, 1407–1414.

Navalpakkam, V. & Itti, L. (2007). Search goal tunes visual features optimally. *Neuron*, 53, 605–617.

Neider, M. B. & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, 46, 614–621.

Pomplun, M. (2006). Saccadic selectivity in complex visual search displays. *Vision Research*, 46, 1886-1900.

Rutishauser, U. & Koch, C. (2007). Probabilistic modeling of eye movement data during conjunction search via feature-based attention. *Journal of Vision*, 7, 1–20.

Shen, J., Reingold, E. M., & Pomplun M. (2000). Distractor ratio influences patterns of eye movements during visual search. *Perception*, 29, 241-250.

Tatler, B. W., Baddeley, R. & Gilchrist, I. (2005). Visual correlates of fixation selection: effects of scale and time, *Vision Research*, 45, 643-659.

Torralba A., Oliva A., Castelhano, M. & Henderson J., (2006), Contextual guidance of attention in natural scenes: The role of global features on object search. *Psychological Review*, 4, 766-786.

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*. Howe, UK: Psychology Press, 13-71.

Figure Captions

Figure 1. (**a**) Example of a search display and its target. (**b**) The same display with its luminance multiplied by the fixation-density map, highlighting the subjects' combined distribution of gaze fixations during search.

Figure 2. Informativeness and guidance for eight different feature dimensions. Error bars indicate standard error. (**a**) Values across all 160 displays with linear regression (guidance = 0.72·informativeness + 0.04); (**b**) values for the 80 most informative displays in each dimension with same line as in (a).

Figure 1

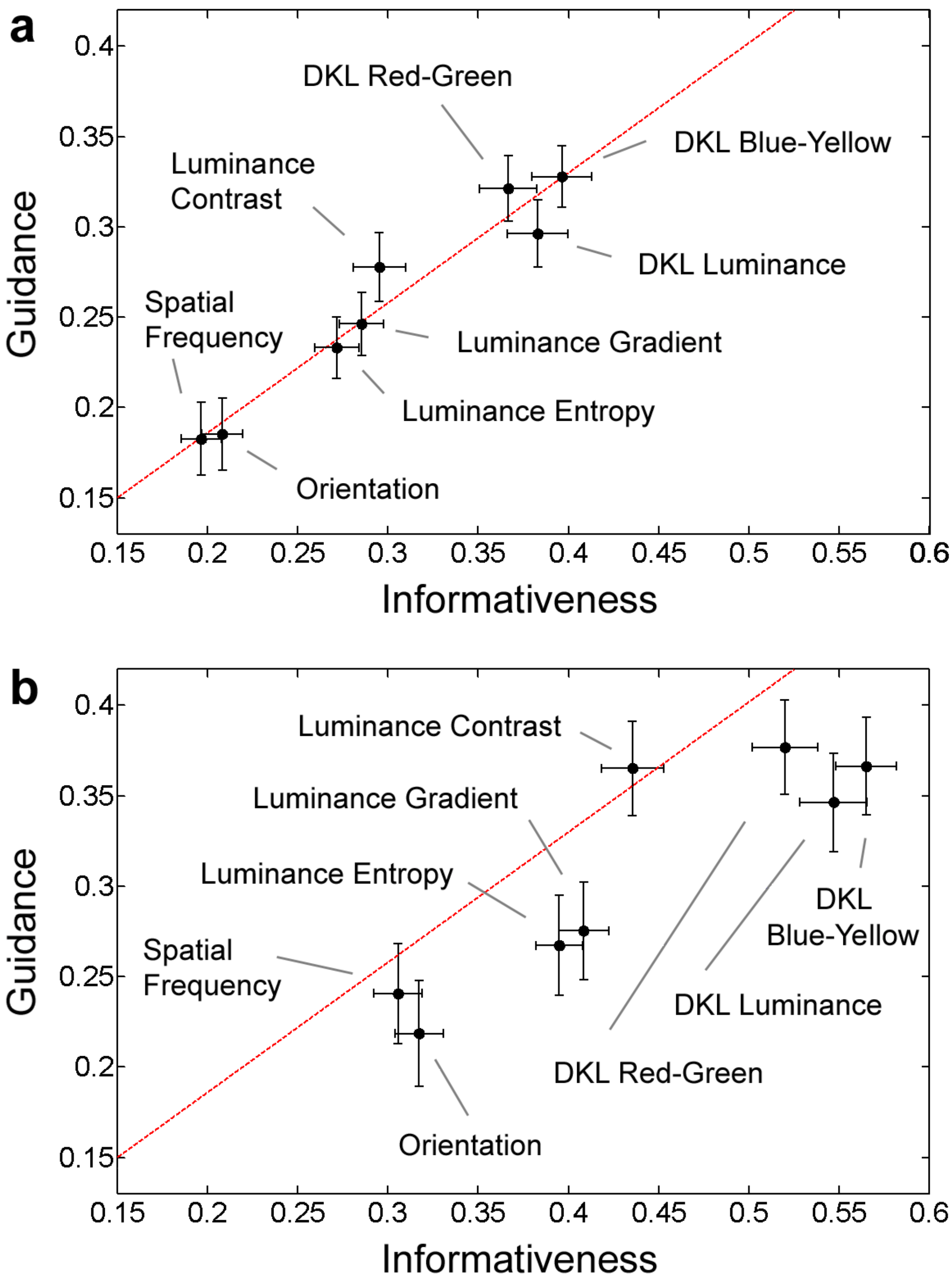
a
Guidance
0.15
0.2
0.25
0.3
0.35
0.4
Informativeness
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
DKL Red-Green
DKL Blue-Yellow
Luminance Contrast
DKL Luminance
Spatial Frequency
Luminance Gradient
Luminance Entropy
Orientation
b
Guidance
0.15
0.2
0.25
0.3
0.35
0.4
Informativeness
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
Luminance Contrast
Luminance Gradient
Luminance Entropy
Spatial Frequency
DKL Blue-Yellow
DKL Luminance
DKL Red-Green
Orientation

Figure 2