ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Vision Research xxx (2006) xxx–xxx

Vision Research

www.elsevier.com/locate/visres

Saccadic selectivity in complex visual search displays

Marc Pomplun*

Department of Computer Science, University of Massachusetts at Boston, 100 Morrissey Boulevard, Boston, MA 02125-3393, USA

Received 7 September 2005; received in revised form 2 December 2005

Abstract

Visual search is a fundamental and routine task of everyday life. Studying visual search promises to shed light on the basic attentional mechanisms that facilitate visual processing. To investigate visual attention during search processes, numerous studies measured the selectivity of observers' saccadic eye movements for local display features. These experiments almost entirely relied on simple, artificial displays with discrete search items and features. The present study employed complex search displays and targets to examine task-driven (top-down) visual guidance by low-level features under more natural conditions. Significant guidance by local intensity, contrast, spatial frequency, and orientation was found, and its properties such as magnitude and resolution were analyzed across dimensions. Moreover, feature-ratio effects were detected, which correspond to distractor-ratio effects in simple search displays. These results point out the limitations of current purely stimulus-driven (bottom-up) models of attention during scene perception.

Keywords: Eye movements; Visual search; Visual attention; Saccadic selectivity; Scene perception

1. Introduction

Whenever we look for the mouse pointer on our computer screen, get a bottle of beer from the refrigerator, or try to find our car in a parking lot, we perform visual search. Our ability to efficiently locate a visually distinctive item in a given scene is crucial for performing most of our everyday tasks. Understanding the attentional processes underlying visual search behavior thus promises to shed light on the mechanisms that enable us to process complex visual information with seemingly little effort. Consequently, for several decades visual search has been one of the most thoroughly studied paradigms in vision research. In a well-studied version of the visual search task, participants have to decide as quickly and as accurately as possible whether a visual scene, composed of multiple search items, contains a pre-specified target item. Many of these studies analyzed the dependence of response times and error rates on the number of search items in the scene. Although this set of variables was rather sparse, it led to the development of numerous theories of visual search. These theories differ most significantly in the function they ascribe to visual attention and its control in the search process (for a review see Wolfe, 1998).

One of the currently most influential theories of visual search is the Guided Search Theory (e.g., Cave & Wolfe, 1990; Wolfe, 1994, 1996; Wolfe, Cave, & Franzel, 1989). According to this theory, visual search proceeds in two consecutive stages: an initial stage of pre-attentive processing that guides a subsequent stage of serial search. After stimulus onset, a parallel analysis is carried out across the display, and pre-attentive information is extracted from it to generate an "activation map" that indicates likely target positions. The activation for each search item consists of a top-down and a bottom-up component. The top-down (task-driven) activation of an item increases with greater similarity of that item to the target, whereas its bottom-up (stimulus-driven) activation increases with lower similarity to other items in its neighborhood. This activation map guides shifts of attention during the subsequent serial search process so that the most promising items are checked first.

* Fax: +1 617 287 6433.
E-mail address: marc@cs.umb.edu.

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.12.003

Data obtained from a large number of visual search studies support the Guided Search Theory. These measurements do not only include ''standard'' psychophysical variables such as response times and error rates, but also more detailed data such as eye-movement trajectories. Analyzing the features of the inspected items and relating them to the features of the target item can provide valuable insight into the search process. Based on this idea, several visual search studies have examined saccadic selectivity, which is defined as the proportion of saccades directed to each distractor (non-target item) type, by assigning each saccadic endpoint to the nearest item in the search display. The Guided Search Theory received support from several of these studies which revealed that those search items sharing a certain feature such as color, shape, or orientation with the target item attracted a disproportionately large number of saccadic endpoints (e.g., Findlay, 1997; Hooge & Erkelens, 1999; Motter & Belky, 1998; Pomplun, Reingold, & Shen, 2001; Scialfa & Joffe, 1998; Williams & Reingold, 2001; but see Zelinsky, 1996).

In general, visual search literature indicates that distractor types that are more similar to the search-target receive greater saccadic selectivity (e.g., Shen, Reingold, Pomplun, & Williams, 2003). If the display items vary along multiple dimensions, features from more than one dimension can guide the search process. This cross-dimensional pattern of saccadic selectivity adapts to the informativeness of features in each dimension: If we make the features along one dimension more similar to each other, saccadic selectivity is likely to shift toward other dimensions (cf. Williams & Reingold, 2001). The flexibility of visual guidance in search tasks was further demonstrated by the distractor-ratio effect (Bacon & Egeth, 1997; Egeth, Virzi, & Garbart, 1984; Kaptein, Theeuwes, & van der Heijden, 1995; Poisson & Wilkinson, 1992; Shen, Reingold, & Pomplun, 2000; Zohary & Hochstein, 1989). In an eye-movement study on this effect by Shen et al. (2000), participants had to detect a target item among two types of distractors, each of which shared a different feature with the target. While the total number of search items was held constant, the ratio between the two distractor types was varied across trials. Saccadic selectivity for one of the two target features was found to increase with fewer display items sharing this feature with the target, indicating that participants tended to search along the stimulus dimension shared by fewer distractors. Such findings indicate that observers are able to change their pattern of visual guidance to take advantage of more informative dimensions.

To date, saccadic selectivity and visual guidance have almost exclusively been investigated in artificial, specifically designed search displays containing discrete search items. These search items typically consisted of combinations of features that varied in two or more discrete levels along a well-defined set of dimensions. Besides better experimental control, another reason for using artificial displays instead of natural images is the elimination of high-level, semantic information and its influence on the scanning patterns (cf. Henderson, Weeks, & Hollingworth, 1999). On the other hand, it is obvious that the full range of capabilities and characteristics of our visual system can only be studied in more complex, natural scenes, for which it has evolved and been trained.

Surprisingly, only one visual search study has measured saccadic selectivity in real-world scenes (Rao, Zelinsky, Hayhoe, & Ballard, 2002). These researchers proposed a computational eye-movement model for visual search tasks that uses local scene representations derived from oriented spatiochromatic filters (derivatives of Gaussians) at multiple scales. To test this biologically plausible model, participants of an eye-movement study were presented with natural search scenes in which they had to find a designated target item. However, throughout the experiment, the search items were arranged in the same semicircle, they did not overlap, and the lighting conditions were constant. So while this experiment overcame the constraints of explicit stimulus dimensions and features, the complexity and variance of the search displays was otherwise still comparable to simple, artificial displays. In the Rao et al. (2002) study, this setup was necessary and successful at yielding highly efficient searches and obtaining clear evidence for the temporal eye-movement characteristics predicted by the model.

To investigate visual guidance in complex search displays, the present study employed 200 grayscale images (see Fig. 1A for an example). For the current exploratory approach, it seemed prudent to eliminate color information in order to avoid the strong attentional capture by color features. Although color is an important search dimension that most likely guides visual search in everyday tasks, it was not included in these displays in order to facilitate the assessment of other—possibly less guiding—dimensions. In each trial, participants were first presented with a small-target image, which they had to memorize. This target image was a cutout (see yellow square in Fig. 1A) of the larger search display, which was subsequently shown to the participants. Their task was to determine the position of the target image in this display while their eye movements were monitored. Eye-movement recording made it possible to determine the ''attentional landscapes'' (cf. Pomplun, Ritter, & Velichkovsky, 1996), that is, the distribution of saccadic endpoints in the search displays (see Fig. 1B). Although the spatial distributions of visual attention and saccadic endpoints are not identical, they are known to be closely coupled during visual search tasks (Findlay, 2004). Therefore, in the present article these two terms will be used interchangeably.

Since the local information in the search displays did not only vary along explicitly defined dimensions, a set of four dimensions was chosen for the study of visual guidance. These dimensions were intensity, contrast, predominant spatial frequency, and predominant orientation of edges.

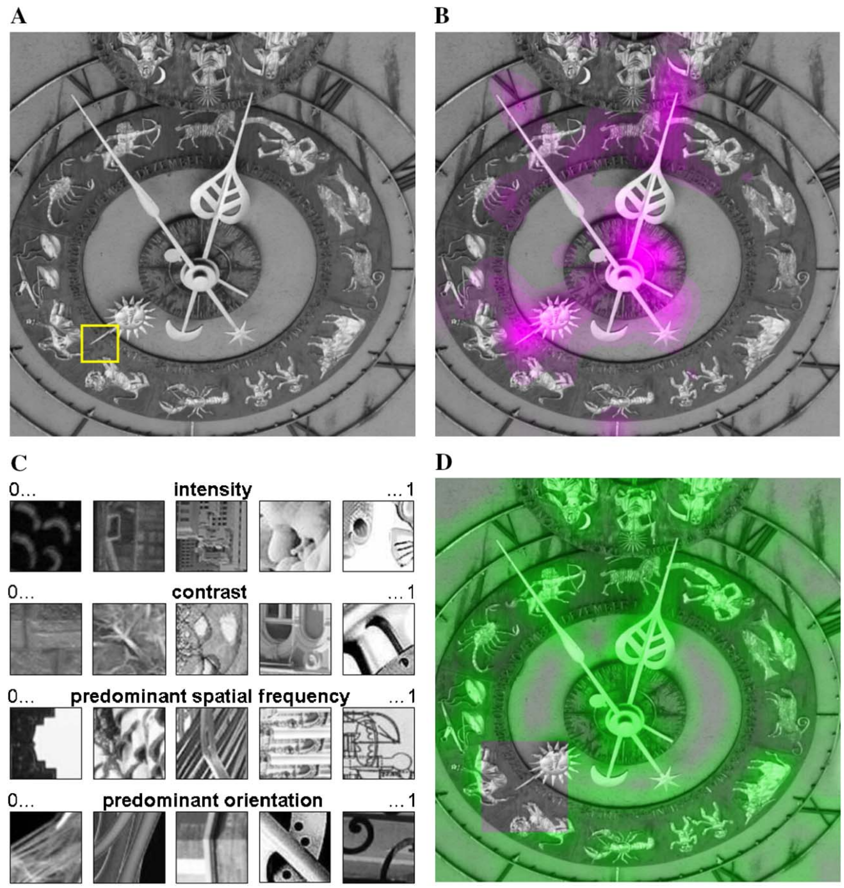

Fig. 1. (A) Sample search display with target area marked by a yellow square (shown as post-trial feedback); (B) distribution of saccadic endpoints during search as indicated by the amount of pink coloring; (C) sample cutouts illustrating the variation of local display information along the four stimulus dimensions; (D) local contrast in the sample image, with more saturated green corresponding to higher contrast.

They were selected because they are known to be relevant to the early stages of the human visual processing hierarchy and are common dimensions in artificial search displays (see Wolfe, 1998). Instead of using large feature vectors such as the ones computed by Rao et al. (2002), the present study focused on only four fundamental dimensions to make the data analysis more transparent and comparable to other saccadic selectivity studies. For each target image and for any position in a given search display, the value of four local stimulus variables, referring to the four dimensions, were calculated on a scale from 0 to 1 (see Fig. 1C). This computation resulted in four additional types of "landscapes" such as the contrast landscape shown in Fig. 1D. Correlating these four with the attentional landscapes made it possible to investigate the following questions: Is there visual guidance in complex search displays along the four chosen dimensions? What are the quantitative and qualitative differences across these dimensions in guiding the observers' attention? What is the resolution of visual guidance, that is, how similar does a feature have to be to the target feature in order to guide eye movements? Are there "feature-ratio effects" in analogy to the distractor-ratio effects found in simple search displays?

2. Method

2.1. Participants

Sixteen students of the University of Massachusetts at Boston (seven female, nine male, aged 19–35) participated in the experiment. They had normal or corrected-to-normal vision and were naïve with regard to the purpose of the study. Each of them received an honorarium of $10 for their participation.

2.2. Apparatus

Stimuli were presented on a 21-in. Dell P1130 monitor using a screen resolution of 1152 by 864 pixels and a refresh rate of 85 Hz. Eye movements were measured with an SR Research EyeLink-II system that provides an average error of 0.5° of visual angle and a sampling frequency of 500 Hz.

2.3. Materials

The experiment encompassed 200 search displays. Each display showed a grayscale bitmap of 800×800 pixels with 256 gray levels, subtending a visual angle of about 25° horizontally and vertically. Out of these 200 displays, 120 showed real-world images of landscapes, gardens, city scenes, buildings, and home interiors (see Fig. 1A for a sample display). These images were randomly rotated by 90°, 180°, or 270° to limit the influence of high-level, semantic information on the participants' scanning patterns. The other 80 displays presented complex artificial images such as fractals or abstract mosaics. From each bitmap, a small, square-shaped cutout of 64×64 pixels ($2° \times 2°$) was chosen to serve as the search target for that display. The target locations were selected manually to avoid uninformative (e.g., completely black) and highly ambiguous targets that had multiple matches in the display. Furthermore, the target areas were chosen to contain only minimal semantic information. The resulting 200 target positions were distributed approximately homogeneously across the display area. All 16 participants were tested with the same set of targets.

2.4. Procedure

Each participant performed 200 experimental trials, one for each search display, which were administered in random order. Every trial started with a 4-s presentation of the 64×64 pixel search target at the center of the screen. The participants' task was to memorize this image. Immediately afterward, the target image was replaced by the 800×800 pixel search display. The participants were told that the previously shown small image was contained somewhere in this large display. Their task was to find the position of the small image within the large one as quickly and as accurately as possible. Moreover, they were instructed to fixate their gaze on that target position and press a designated button on a game pad as soon as they were certain that they had found the target. This button press terminated the trial. If participants did not press the button within 5 s after the onset of the large display, the trial was ended automatically (timeout). In either case, participants received visual feedback about the actual position immediately after the end of the trial (see Fig. 1A).

2.5. Data analysis

For the investigation of saccadic selectivity, the local density of saccadic endpoints across each display ("attentional landscape") was calculated as follows: For each experimental trial, every fixation in the display was associated with a Gaussian function, which was centered at the fixation position and had a standard deviation of one degree of visual angle. This value was chosen to match the approximate size of the human fovea. Fixation duration did not enter the analysis, as it is more likely to be correlated with the effort of memory retrieval and comparison processes than with saccadic-target selection (Hooge & Erkelens, 1999; Shen et al., 2003). All Gaussian functions for the same trial were summed across the display area, and the resulting function was normalized to have an average value of 1 over this area. This ensured that the data obtained from each display and each participant had the same weight in the data analysis. The average, rather than the volume covered by the function, was chosen to be normalized to make individual local values of the resulting measure more easily interpretable. Finally, all 16 participants' functions for the same display were averaged to generate a smooth attentional landscape that indicated the amount of visual attention—measured by the density of saccadic endpoints—across positions in the display. Fig. 1B illustrates this function for the sample stimulus in Fig. 1A. The more saccadic endpoints a local region in the image received, the more strongly it is overlaid with the color pink. As can clearly be seen, most of the regions that attract the greatest amount of saccadic endpoints are similar to the target area in that they show one or more elongated bright structures on a dark background. Interestingly, participants' eye movements were also attracted by some areas that show an elongated structure in an orientation that is substantially different from the one in the target area.

To analyze saccadic selectivity during the search process, four appropriate local stimulus variables were computed across all of the 200 displays, each of them measured within a local area of 64×64 pixels. These were the intensity L_i (average brightness of local pixels), the contrast L_c (standard deviation of local brightness), predominant spatial frequency L_f (most elevated frequency band in local area as compared to baseline data), and predominant orientation L_o (angle of predominant orientation of local edges). Fig. 1C shows sample areas with features varying along these four variables, and in Appendix A a mathematical definition of the variables is provided. Each of them was computed once for every target bitmap and for 74×74 positions in every search display. The horizontal and vertical offset between neighboring positions was ten pixels. This offset led to substantial overlap between measurement areas and to six-pixel-wide margins on the right and bottom sides of the search displays that were excluded from all measurements. Fig. 1D shows an example of a contrast landscape. Notice that, to avoid artifacts in the analysis of saccadic selectivity, no feature information was computed or analyzed within a $5° \times 5°$ square centered at the target positions. This was due to the fact that whenever participants detect a target, they are likely to look at it for an extended duration before they terminate the trial. Including these target area fixations in the selectivity analysis would have led to an elevated number of saccadic endpoints aimed at the target features, indicating visual guidance towards those features, regardless of whether such guidance was actually exerted during the search process.

Table 1
Correlation coefficients (Pearson's r^2) between pairs of local stimulus variables

Stimulus dimension	Intensity	Contrast	Spatial frequency	Orientation
Intensity	1.000	0.016	<0.001	<0.001
	1.000	0.056	0.001	0.009
Contrast	0.016	1.000	0.007	0.010
	0.056	1.000	0.011	0.017
Spatial frequency	<0.001	0.007	1.000	0.012
	0.001	0.011	1.000	<0.001
Orientation	<0.001	0.010	0.012	1.000
	0.009	0.017	<0.001	1.000

In each cell, the upper value refers to the entire set of values (across all positions in all search displays), while the lower value refers to only the values at the target positions.

One problem with the use of complex search displays—and certainly an important reason for the dominance of artificial displays in visual search literature—is the difficulty of defining local stimulus variables that vary independently of each other. Such independence is desirable for the analysis of visual guidance. Let us assume that we have two local stimulus variables A and B that are strongly correlated within the displays, including the target areas. If we measure visual guidance by both A and B, we cannot rule out that only one of them, say A, guides the search, and guidance by B is measured just because of B's correlation with A. Unfortunately, the local variation in real-world images is not homogeneously random but follows certain statistical patterns, causing correlations between stimulus variables (e.g., Baddeley, 1997; Torralba & Oliva, 2003). Although there are decorrelation algorithms such as the Mahalanobis Transform (e.g., Therrien, 1992), it is difficult to interpret the obtained values of the transformed variables. Therefore, in the present study, the local stimulus variables were chosen in such a way that they capture fundamental local display properties in a straightforward manner with minimal correlations between pairs of variables. Table 1 presents the resulting correlation coefficients (Pearson's r^2) across all display positions, and also separately for all target positions.

3. Results

The proportion of trials that were ended through manual response, that is, before the 5-s timeout, was 55.8%. In these trials, participants searched for an average duration of 3.25 s before pressing the button. These numbers suggest that the 5-s timeout made the target detection task difficult, which was intended since the present study focused on the search process rather than the target detection process. The average duration of all 35,945 fixations that entered the analysis, i.e., that ended before the manual response or timeout, was 257 ms. To assess the accuracy of the participants' performance, the Euclidean distance of the last fixation in each trial to the center of the target area was computed. As a baseline measure, the distance of this fixation to the target in the following trial was also computed (for the last fixation of trial 200, its distance to the target in trial 1 was measured). This way the fixation and target positions were decorrelated, while their statistical distributions remained unchanged. If the participants' eye movements were guided towards the search targets, the distance between their last fixation and the target should be smaller for the actual data than for the decorrelated ones. In those trials with manual response, the actual mean distance was found to be 126 pixels, which was significantly smaller than for the decorrelated data (314 pixels), $t(199) = 20.38$, $p < 0.001$. Regarding the timed-out trials, the actual mean distance was 200 pixels, which was larger than for the manually terminated trials, $t(195) = 11.23$, $p < 0.001$, but still smaller than for the decorrelated data, $t(195) = 12.50$, $p < 0.001$. These findings indicate that the participants' gaze was guided towards the target area in both trials with and—to a smaller extent—without manual response. Notice that, since the computation of the attentional landscapes required the cumulative eye-movement data of all participants, throughout this article the variance in the data was calculated across search displays and not across participants. Some of the analyses below included missing values because, for instance, several displays did not contain the whole range of intensity features. These missing values are indicated by lower degrees of freedom in the statistical tests.

In analogy to studies using discrete search items, the first aim of the eye-movement analysis was to establish the existence of visual feature guidance in complex search images along the four chosen dimensions. The basic idea underlying this analysis is the following: If there is visual guidance by, for example, the intensity dimension, then in those trials with a dark target bitmap, dark areas in the search display should receive greater saccadic selectivity than bright areas. Accordingly, bright targets should strengthen saccadic selectivity for bright areas. To conduct this analysis, the values of each of the four local stimulus variables were divided into three intervals, e.g., low versus medium versus high-intensity. This division was performed in such a way that, across all positions in all search displays, the same number of values fell into each interval. We can think of these intervals as the visual search features, in analogy to features such as red, green, horizontal, or vertical that are typically used in artificial search displays. The choice of three features was made to provide a first, clear insight into visual guidance patterns. The effect of varying the number of features per dimension will be discussed later in this section.

For each of the four dimensions, the 200 search trials were classified according to the interval in that dimension to which the search target belonged. Then, saccadic selectivity for each class of trials (e.g., trials with low versus medium versus high-intensity target) was analyzed separately. In the present context, saccadic selectivity was operationally defined as the average density of saccadic endpoints that areas sharing a particular feature—e.g., medium intensity—received across all positions in the relevant group of trials. Notice that these values only measure selectivity and are independent of the proportion of a feature in the search displays.

Fig. 2 presents the results of this analysis. For the intensity dimension (Fig. 2A), the pattern of results is clear: Low-intensity areas received the greatest saccadic selectivity if the target was also of low intensity, for medium-intensity areas this happened for medium-intensity targets, and for high-intensity areas it was true for high-intensity targets. A two-way analysis of variance (ANOVA) of saccadic selectivity using the between-trial factor display intensity (low, medium, and high) and the within-trial factor display intensity (low, medium, and high) revealed a significant interaction between the two factors, $F(4, 344) = 15.04$, $p < 0.001$. This result is evidence for the presence of visual guidance by intensity in complex search displays, as it demonstrates that the intensity of the target influences the pattern of saccadic selectivity along the intensity dimension. The factor target intensity did not exert a significant effect on saccadic selectivity, $F(2, 172) = 1.73$, $p > 0.1$, which was expected because the average saccadic selectivity for each display was normalized to 1. Finally, the factor display intensity did not have a significant effect, $F(2, 344) = 2.50$, $p > 0.05$, indicating that participants divided their attention evenly across all levels of local intensity in the displays.

The pattern of results for the contrast dimension (Fig. 2B) was similar. An ANOVA analogous to the one above showed the interaction between target contrast and display contrast to be significant, $F(4, 386) = 14.29$, $p < 0.001$. While the main effect of target contrast was not significant, $F(2, 193) = 1.22$, $p > 0.2$, the main effect of display contrast was, $F(2, 386) = 61.57$, $p < 0.001$. This finding reveals that participants did not distribute their attention evenly across the three feature intervals. Areas of low, medium, and high contrast received saccadic selectivity values of 0.536, 1.216, and 1.431, respectively. This result provides evidence for a bottom-up effect with regard to local contrast in search displays: Regardless of the target contrast, high-contrast areas receive more attention than low-contrast ones. This is an intuitive finding, because low-contrast areas simply contain less information or information that is harder to process, so observers prefer to attend to higher-contrast areas in order to perform efficient search.

The analysis of predominant spatial frequency (Fig. 2C) yielded results that were similar to the intensity data, although somewhat less pronounced. As for the intensity and contrast parameters, the two-way ANOVA revealed a significant interaction between target and display frequency, $F(4, 392) = 7.73$, $p < 0.001$. Target frequency had no main effect, $F(2, 196) < 1$, whereas display frequency

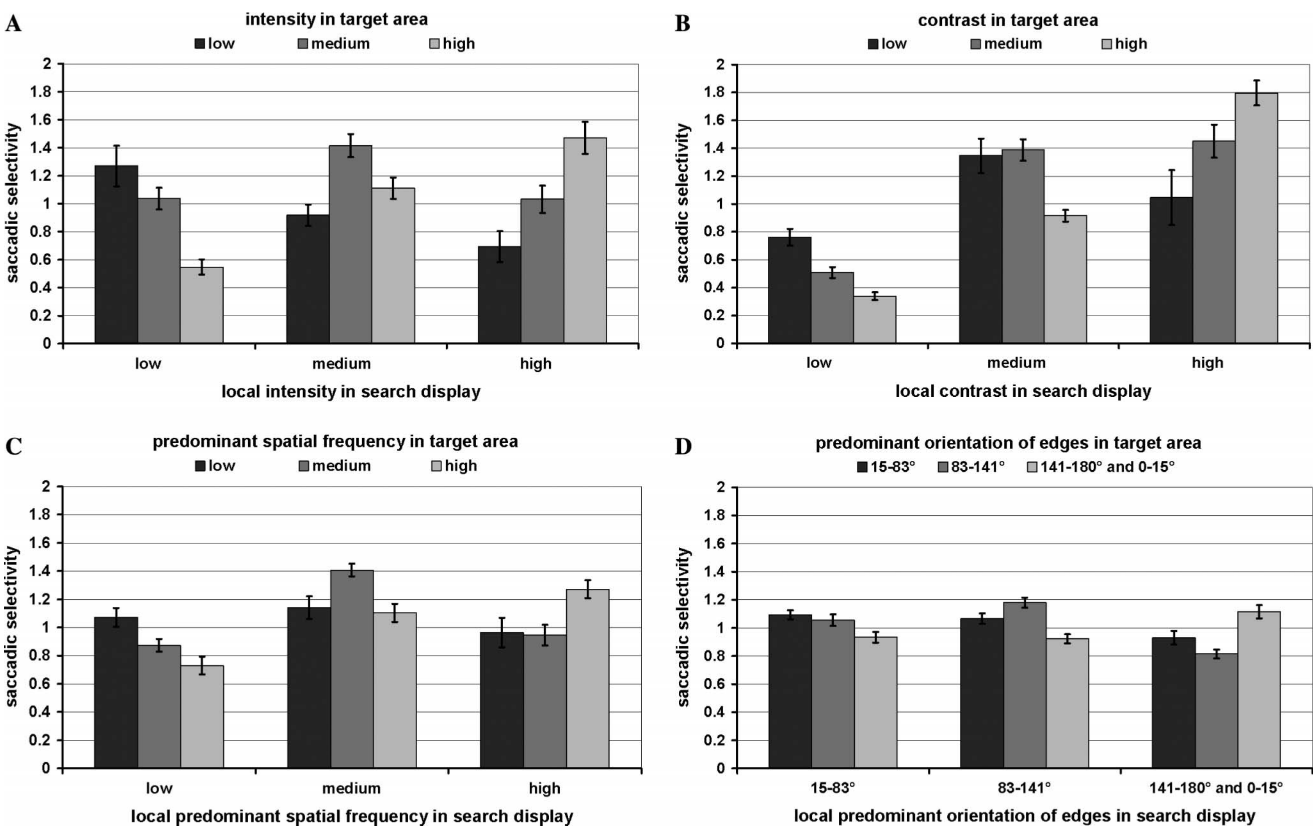

Fig. 2. Saccadic selectivity as a function of the four local stimulus variables across the display and in the target area. (A–D) Refer to the variables intensity, contrast, predominant spatial frequency, and predominant orientation, respectively. In all figures, error bars indicate the standard error across displays.

had a significant main effect, $F(2,392) = 47.61$, $p < 0.001$, signifying a bottom-up effect for spatial frequency. The saccadic selectivity values for low-, medium-, and high-frequency areas were 0.891, 1.217, and 1.060, respectively. A possible explanation for this finding is that areas near the extremes of the spatial frequency spectrum are less informative and harder to visually process.

Finally, the results for the predominant orientation as illustrated by Fig. 2D show a weaker, but still highly significant pattern of visual guidance. For the predominant angle α, the first two feature intervals were defined by $15° \leqslant \alpha < 83°$ and $83° \leqslant \alpha < 141°$, and the third interval was composed of the intervals $141° \leqslant \alpha \leqslant 180°$, and $0° \leqslant \alpha < 15°$. These range definitions were selected in order to have the same number of samples in each interval and to avoid interval boundaries at horizontal or vertical orientation, as these orientations occur disproportionately frequently in natural images (e.g., Baddeley, 1997). The two-way ANOVA yielded a significant interaction between target orientation and display orientation, $F(4,392) = 10.90$, $p < 0.001$. While there was no significant main effect of target orientation, $F(2,197) = 2.32$, $p > 0.05$, the main effect of display orientation did reach significance, $F(2,394) = 4.26$, $p < 0.05$. The features one, two, and three received selectivity values of 1.027, 1.056, and 0.953, respectively, indicating a slight bottom-up effect that is hard to explain intuitively.

To validate these results, the same computation of saccadic selectivity was performed again, but this time with the decorrelated fixations introduced at the beginning of this section. For these data, none of the four variables showed an effect of visual guidance, that is, an interaction between the target feature and the display feature, all Fs < 1.96, ps > 0.1. While this analysis supports the conclusion that eye movements are guided by target features, one can still think of one possible confound in the data: It is conceivable that the $5° \times 5°$ square around the target that was excluded from eye-movement analysis was chosen too small. Then local scanning behavior may have caused elevated density of fixations in the neighborhood of the square, and this neighborhood may still have been more likely than the rest of the display to contain target features. If this was the case, then the measured saccadic selectivity pattern may be an artifact caused by display properties rather than an indicator of guided visual search. To investigate this possibility, the fixation positions in each display were rotated by 180° around the target position. Those fixations that the rotation moved outside the 800×800 pixel display area were "wrapped around" this area in horizontal and vertical directions. For example, a rotated fixation with an x-coordinate of 830 would have been set to 30. Through this manipulation, fixations and display features were decorrelated, while the relative fixation density near the target remained the same. Once again, none of the four ANOVAs showed a significant effect of visual guidance, all Fs < 2.36, ps > 0.05, which rules out this possible confound.

When we compare the results across dimensions, do they tell us that visual guidance by the orientation of local edges in the search display is clearly weaker than by the other three stimulus dimensions? Such a finding would be in line with the results of a study by Gilchrist, Heywood, and Findlay (2003), which found that saccades were less sensitive to stimulus differences in orientation than to differences in contrast or spatial frequency. However, it must be stated that the present study treats the dimensions spatial frequency and orientation rather ''unfairly'' with regard to the measurement of visual guidance. Intensity and contrast values are defined in a straightforward and robust way: Changing the brightness of a few pixels will lead to only very small changes in the local intensity and contrast values. This situation is different for the predominant spatial frequency and orientation, where the features ''compete'' against each other and the winner determines the value of the variable. A change in the brightness of only a few pixels can dramatically alter the value of local spatial frequency or orientation; for example, it could even switch from vertical to horizontal orientation. Consequently, measurements of frequency and orientation are by definition more noisy than for the other variables, which reduces the measured effect of visual guidance by these dimensions.

Furthermore, in display regions of very low contrast, neither spatial frequency nor orientation can be perceived and thus cannot exert any visual guidance. This adverse effect on guidance could be compensated by adding ''no frequency'' and ''no orientation'' features, which encompass all low-contrast areas. It was found that this method does increase the measured visual guidance, but only by a few percent, because low-contrast areas attract only very few saccadic endpoints (see above). On the other hand, the problems with this method are an increased correlation of spatial frequency and orientation with contrast as well as the difficulty to appropriately set the size of the new feature intervals. Therefore, it was decided not to apply this method, but to treat all four dimensions in the same way.

When comparing the results across dimensions, we should also consider the possibility that the choice of only three intervals was too coarse for some of the dimensions to show their entire potential for visual guidance. To investigate this hypothesis, the amount of visual guidance as a function of the number of features per dimension was analyzed. Visual guidance by a particular dimension was operationally defined as the average amount of saccadic selectivity that each feature received when the target shared the same feature, divided by the average amount that the feature received regardless of the target feature. A higher value of visual guidance by a certain dimension thus indicates a stronger ability of the dimension to bias attention towards display areas that share the same feature along that dimension with the target, whereas a visual guidance value of 1 means that there is no such bias.

Fig. 3 displays visual guidance in each of the four dimensions as a function of the number of features per dimension. As before, interval boundaries were computed in such a way that each interval contained the same number of data points, and for the orientation dimension no interval boundaries were placed at horizontal or vertical orientation. As can clearly be seen, visual guidance generally increases with a growing number of features. While this increase is steep for a small number of features, it levels off for greater numbers and finally seems to asymptote to a dimension-specific maximum guidance value.

How does this pattern of results arise? To explain this, let us first introduce the term ''saccadic selectivity bias.'' It describes the quotient of saccadic selectivity for a particular value of a stimulus variable in a given trial, divided by the average saccadic selectivity for that value across all trials. Thus, in a given trial, a bias smaller or greater than 1 for a particular value of a stimulus variable indicates that areas of this value receive a smaller or greater density of saccadic endpoints, respectively, than they usually do. Notice that for each stimulus dimension the average saccadic selectivity bias for values within the target feature interval is identical to visual guidance as defined above. Now let us take intensity guidance as an example and assume that we are searching for a target of intensity 0.5. Then, by the principle of visual guidance, areas in the search display that have the same intensity 0.5 should receive the greatest saccadic selectivity bias. However, areas of neighboring intensities 0.49 and 0.51 are hardly perceptually distinguishable from the ones of intensity 0.5 and should therefore receive almost the same bias. More distant intensities such as 0.4 or 0.6 will presumably receive a significantly smaller bias, which may still be greater than 1, depending on how ''fine-grained'' intensity guidance is. At any rate, we would expect intensities like 0 and 1, which

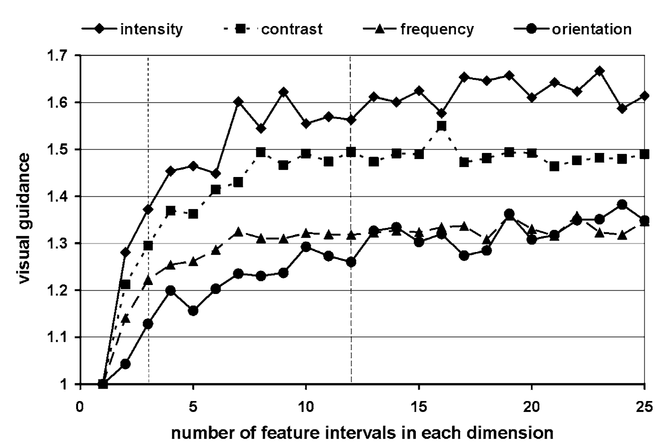

Fig. 3. Visual guidance by each of the four stimulus dimensions as a function of the chosen number of feature intervals per dimension. The left vertical dashed line indicates guidance for three features per dimension, as chosen for the analysis shown in Fig. 2, while the right one refers to 12 features, which was selected as the optimal number of features for guidance measurement for the other analyses.

are far away from the target intensity, to receive a bias below 1.

Let us further assume that for each dimension this pattern of saccadic selectivity bias across stimulus values follows a normal distribution centered at the target value, with a dimension-specific standard deviation. For each dimension, this standard deviation would indicate the "spread" of visual guidance, that is, the extent to which areas that are similar, but not identical to the target still receive elevated saccadic selectivity. With this model in mind, let us look at Fig. 3 once again. By definition, using only one feature interval per dimension must lead to a visual guidance value of 1. With two features, the target always falls into one of two large intervals of width 0.5, and visual guidance is computed as the average saccadic selectivity bias within this interval. Thus, if the spread of visual guidance is small, averaging over such large intervals will lead to a measured visual guidance that is much lower than the saccadic selectivity bias at the target value. For a large spread, this discrepancy will be smaller. With an increasing number of features, the intervals will become narrower so that the averaging of saccadic selectivity bias will be limited to a range closer to the target value, that is, the peak of the assumed normal distribution. Therefore, the measured visual guidance will increase with a greater number of features—at first quickly and then more slowly, as the interval width is inversely proportional to the number of features—and finally asymptote to the saccadic selectivity bias for the target value.

However, Fig. 3 contains more information than this general pattern of measured visual guidance: It shows that the relative differences between the intensity, contrast, and spatial frequency dimensions remain approximately constant, regardless of the number of features. Visual guidance by orientation, however, is clearly the lowest for three features (see left vertical dashed line) but strongly increases with a greater number of features to reach a level comparable to spatial frequency guidance. According to the model outlined above, a possible explanation would be that spatial frequency and orientation evoke the same saccadic selectivity bias at the target value, but the spread of visual guidance is smaller for orientation. When using only a few features, this would lead to a measurement of weaker orientation guidance, but with an increasing number of features orientation could "catch up" with spatial frequency.

Before this hypothesis could be tested, Fig. 3 was used to answer the following question: Which number of features is most appropriate for the analysis of visual guidance? As shown above, using only a few features does not adequately measure the visual guidance that a given dimension can produce. However, due to fewer data points for each feature, using a greater number of features increases the noise in the measurement. It was therefore decided to use 12 features for each dimension for further analysis, which seemed to provide the best compromise between sensitivity and precision of measurement (see right vertical dashed line in Fig. 3).

To test the hypothesis about orientation and to examine the resolution of visual guidance, the spread of guidance across features was analyzed. Twelve features per dimension were used to compute the average saccadic selectivity bias for features relative to the target feature along a given stimulus dimension. For example, considering the intensity dimension, let us assume that in a given trial the target is of intensity 5. Then all display areas of intensity 4—which are only slightly darker than the target—are of relative intensity -1, and areas of intensity 7 are of relative intensity 2. Fig. 4 shows the result separately for each dimension. In each panel, a vertical dashed line marks the saccadic selectivity bias value at relative feature 0—the target feature itself. By definition, this bias value is identical to the visual guidance value for 12 features as shown in Fig. 3. Only the relative features from -8 to 8 are shown, because relative features outside this interval can only occur if the target is at one of the extremes for the respective variable, and therefore the statistical power is very low.

As expected, for the intensity and contrast dimensions (Figs. 4A and B, respectively), the bias clearly decreases with greater feature distance from the target and reaches below-average values (less than 1) within a distance of about 2–4 features from the target. For the spatial frequency dimension (Fig. 4C), the pattern is shallower, but nevertheless elevated for near-target features in a way similar to intensity and contrast. The data suggest that, outside the elevated interval from relative features -2 to 2, below-target frequency has a slightly greater saccadic selectivity bias than above-target frequency. A statistical analysis of the bias for this below-target frequency (1.01) versus the above-target frequency (0.82) across all 200 displays supported this assumption, $t(111) = 4.70$, $p < 0.001$. This may indicate that participants' attention is only as "fine-grained" as the target, that is, its spatial resolution is set to the maximum value that is necessary to compare the current memory content to the local display information. However, at this point this is pure speculation and requires further investigation.

Because orientation is a clock variable, its relative features were defined differently than for the other dimensions: They were measured in both clockwise and counterclockwise directions. For instance, local orientation 1 for target orientation 5 entered the analysis as both relative orientation -4 and relative orientation 8. The results for orientation (Fig. 4D) show an elevated bias near the target, but also local bias maxima at relative orientations -6 and 6, which are approximately perpendicular to the target orientation. A statistical analysis revealed that the average bias value for the perpendicular orientations was significantly greater (1.10) than for orientations -3 and 3 (0.93), which correspond to angles of approximately $-45°$ and $45°$, $t(199) = 2.85$, $p < 0.01$. A plausible explanation for this pattern is that natural images contain disproportionately many $90°$ angles between intersecting edges, which is a consequence of the dominance of horizontal and vertical orientations (e.g., Baddeley, 1997). As

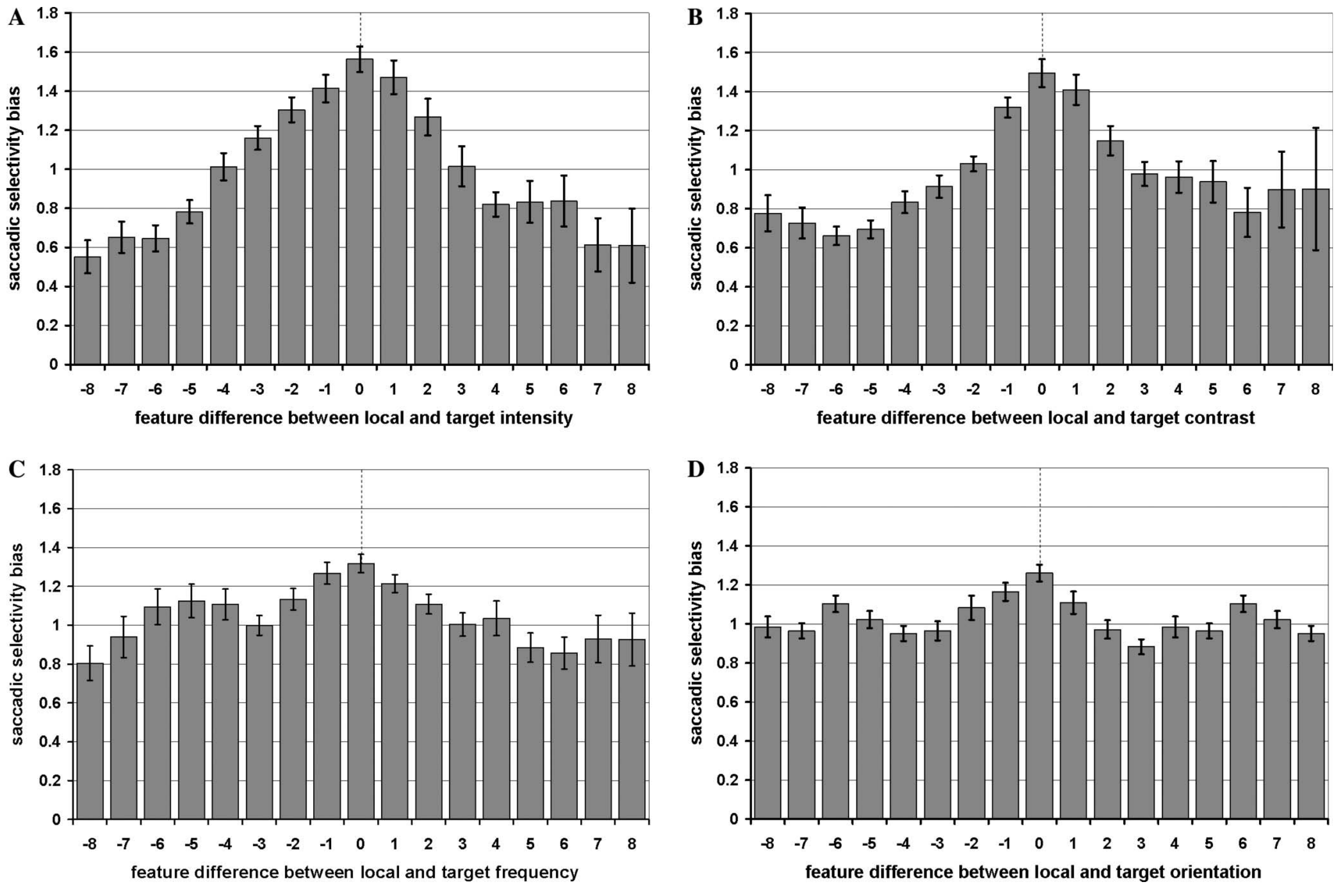

Fig. 4. Saccadic selectivity bias for local display features as a function of their relative position to the target feature along the respective stimulus dimension. Twelve features per dimension were used. (A–D) Refer to the variables intensity, contrast, predominant spatial frequency, and predominant orientation, respectively. The vertical dashed lines mark relative position 0, which stands for the target feature itself. Greater horizontal distance from this line corresponds to greater difference between local and target feature.

discussed above, the local orientation measured at such a 90° intersection is drawn towards one of the two edges, even if they are equally pronounced. However, both edges may exert visual guidance, which possibly causes the observed effect of saccadic selectivity bias. This extreme spread of visual guidance across the orientation spectrum is especially detrimental to the measurement of guidance with only a few feature intervals, which would also explain the orientation data shown in Fig. 3.

For a statistical analysis of visual guidance and its spread across dimensions, one-way ANOVAs with the factor dimension were conducted for each of these two variables, followed by pairwise t-tests with Bonferroni-adjusted probabilities. Fig. 5A presents the visual guidance values. All four mean values were significantly greater than 1, all $t(199) > 6.04$, $p < 0.001$, showing that each of the four dimensions played a role in guiding the participants' eye movements. The ANOVA revealed a significant effect by the factor dimension, $F(3, 591) = 7.43$, $p < 0.001$. Guidance by intensity (1.56) did not statistically differ from contrast (1.50), $t < 1$, but was significantly greater than for spatial frequency (1.31), $t(199) = 3.28$, $p < 0.01$, and orientation (1.26), $t(199) = 4.02$, $p < 0.001$. Visual guidance for contrast did not differ from spatial frequency, $t(199) = 2.29$, $p > 0.1$, but was greater than for orientation, $t(199) = 2.89$, $p < 0.05$. The difference between spatial frequency and orientation was not significant, $t < 1$.

The spread of visual guidance in a display was operationally defined as the root-mean-square distance from the target feature to all features whose saccadic selectivity bias exceeded 1, weighted by the amount by which each one exceeded it. For each dimension, this measure describes the range of features around the target feature that receive elevated saccadic selectivity (see Fig. 5B). The ANOVA showed a significant effect by the factor dimension, $F(3, 597) = 65.46$, $p < 0.001$. For intensity, the spread of visual guidance was smaller (2.33 features) than for contrast (3.43), spatial frequency (3.70), and orientation (4.37), all ts > 7.59, ps < 0.001. The spread of contrast guidance was not statistically different from the one for spatial frequency, $t(199) = 1.48$, $p > 0.5$, but significantly smaller than the one for orientation, $t(199) = 5.95$, $p < 0.001$. Spatial frequency guidance revealed a smaller spread than did orientation guidance, $t(199) = 4.73$, $p < 0.001$.

While Figs. 4A and B suggest that visual guidance is more focused for contrast than for intensity, the analysis of spread yielded the opposite result. The reason is that

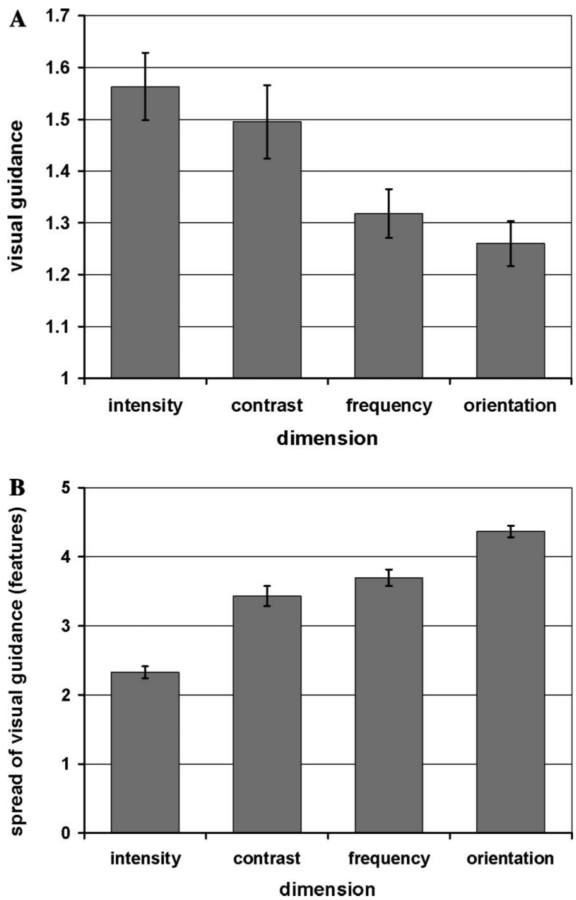

Fig. 5. (A) Visual guidance and (B) its spread across features for each of the four stimulus dimensions.

areas with intensities that substantially differ from the target very reliably receive a below-average saccadic selectivity bias, whereas there is much more uncertainty with regard to the contrast dimension. Neither visual guidance nor its spread showed a significant difference between the contrast and spatial frequency dimensions, which is interesting given the above mentioned problems with the measurement of guidance by spatial frequency and orientation. Orientation had the greatest spread, presumably due to the disproportionately high occurrence of 90° angles as discussed above.

The final analysis addressed the question of whether there are feature-ratio effects in complex search displays in analogy to the distractor-ratio effects found in simple displays. In other words, is the degree of visual guidance by a target feature inversely related to its occurrence within the search display? This analysis was conducted separately for each of the four stimulus dimensions. For each search trial and each dimension, the proportion of the 74×74 points in the search display sharing the target feature in that dimension was computed. Then for each dimension the 200 trials were sorted by that proportion in ascending order and divided into four groups: the 50 displays with the lowest proportion (group 1, percentiles 0–25), the displays ranked 51–100 (group 2, percentiles 25–50), the displays ranked 101–150 (group 3, percentiles 50–75), and the 50 displays with the highest proportion (group 4, percentiles 75–100). For each of the four groups of trials, the mean visual guidance by the same dimension was computed using 12 features per dimension. If there is a feature-ratio effect for that dimension, then guidance for the lower groups—i.e., those trials in which the target feature appeared less frequently in the search display—should be greater than for the higher ones.

As can be seen in Fig. 6, all four variables reveal a very similar pattern of visual guidance for groups 1–4 (intensity: 2.64, 1.42, 1.28, and 1.34; contrast: 2.20, 1.57, 1.29, and 1.03; frequency: 1.66, 1.28, 1.17, and 1.17; orientation: 1.70, 1.28, 1.19, and 1.20): Mean visual guidance is greater for group 1 than for the other groups, which indicates the presence of feature-ratio effects. For each variable, a one-way ANOVA with the factor proportion was computed, demonstrating significant effects on visual guidance for intensity, contrast, and spatial frequency, all $Fs(3,196) > 7.66$, $ps < 0.001$. The orientation dimension showed only a tendency towards such an effect, $F(3,196) = 2.21$, $p = 0.089$. All of the first three variables

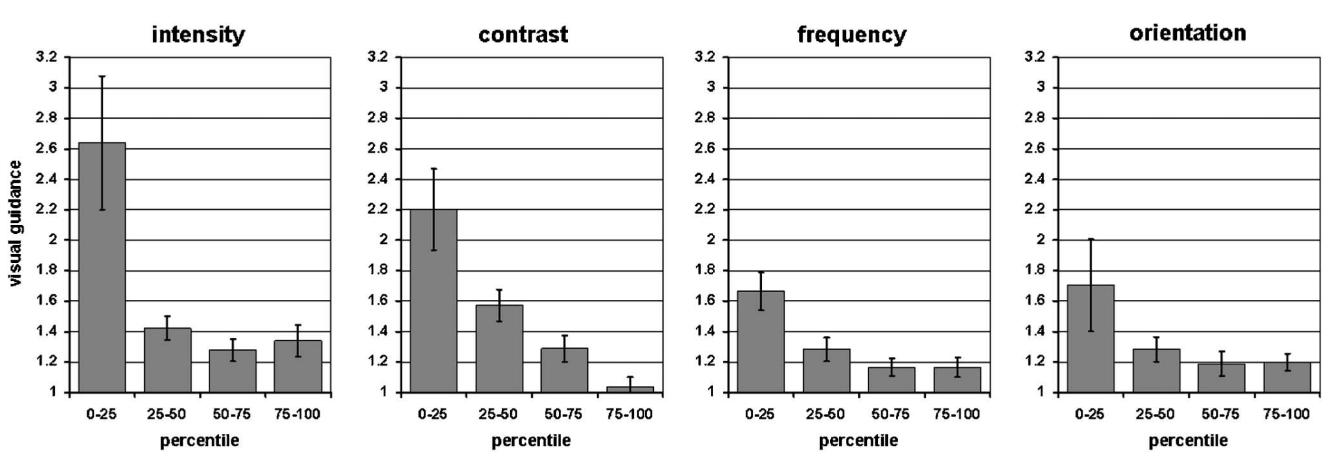

Fig. 6. Visual guidance as a function of the proportion of the target feature in the search display, separated into four percentile groups. Results are shown separately for each of the four stimulus dimensions.

demonstrated significantly greater guidance for group 1 than for each of the groups 2–4, all $ts(196) > 2.91$, $p < 0.05$ (Bonferroni-adjusted). There were no significant differences between groups 2, 3, and 4, all $ts(196) < 1.31$, $ps > 0.5$, except for the contrast dimension, which revealed a tendency towards greater guidance by group 2 than by group 4, $t(196) = 2.48$, $p = 0.084$. The mean guidance values suggest that contrast guidance decreases throughout groups 1–4, whereas for the other variables the only difference is between group 1 and the other groups. However, further research is necessary to statistically substantiate this assumption.

4. Discussion

The present study built upon the concept of visual guidance in search processes as described within the Guided Search Theory (e.g., Wolfe, 1998). With the help of eye-movement recording, the current work extended this concept towards a quantitative analysis of visual guidance in complex images that were not specifically designed for experimentation. To do this, four sample dimensions of local visual content in the display were defined, namely intensity, contrast, predominant spatial frequency and predominant orientation. Moreover, a method for measuring guidance along these dimensions was developed and applied. The results showed that all four dimensions significantly guided the search process, as saccadic selectivity was biased towards those features in a search display that were also features of the search target. This finding demonstrates that even in complex displays without discrete search items and explicit features there is visual feature guidance. We still do not know the exact nature of the guiding features and their dimensions and how they are selected for a particular target. Nevertheless, it was demonstrated that whatever the guiding dimensions are, they correlate with all of the four stimulus dimensions chosen for the present study.

The amount and pattern of guidance were found to vary considerably across dimensions. Caution, however, must be exercised when interpreting these differences, since they may depend on the definition of the local stimulus variables. Whereas the definition of intensity and contrast in the present study was straightforward, scalar variables describing spatial frequency and orientation properties could be defined in numerous ways and inevitably involve greater noise. Consequently, it is not surprising that guidance was quantitatively greater for intensity and contrast than for frequency and orientation. Analyzing the spread of guidance, however, gave more detailed insight into the saccadic selectivity data. While the amplitude of the saccadic selectivity bias was most pronounced for intensity and contrast, the width and slope of its elevated area in feature-space around the target feature were found to be similar for all four dimensions. The pattern of results for orientation indicated that its measurement was noisy and distorted by the disproportionate occurrence of 90° angles in the search displays. Accordingly, the spread of orientation guidance clearly exceeded the values for the other three dimensions. Spatial frequency revealed an interesting asymmetry in its spread, showing a saccadic preference for below-target rather than for above-target frequency. A possible interpretation is that attention is tuned to disregard details that are more fine-grained than those in the memorized target pattern. This explanation is in line with the ideas presented by Rao et al. (2002) that such an attentional mechanism would increase search efficiency; however, further research is required to investigate this speculation.

Overall, the spread analysis gave us an estimate for the resolution of visual guidance. Leaving aside the noise in the measurement of frequency and orientation guidance, it seems that across all dimensions about 30–50% of the features that are closest to the target features receive—at least slightly—elevated saccadic selectivity. This finding indicates that visual guidance is not confined to the precise target features but is tuned rather broadly. However, the observed spread patterns also suggest that the actual resolution of visual guidance is clearly finer than indicated by its feature range. For all of the four chosen dimensions, it seems that a feature will receive less elevated saccadic selectivity if it differs from the target feature by more than ≈8% of the entire feature range along the respective dimension. It is important to notice, though, that this is only a rough estimate based on the saccadic selectivity data for 12 features per dimension. Further studies are necessary to statistically analyze these effects. Once confirmed, the current estimates will add new information to the discussion of the maximum number of distinctive visual search features along a given dimension (e.g., Wolfe, 1998).

The four stimulus dimensions also showed quantitative but few qualitative differences with regard to their feature-ratio effects. Displays with a lower proportion of a target feature in any dimension induced stronger visual guidance by that feature. Even though orientation demonstrated only a statistical tendency towards such an effect, the known problems with orientation measurement suggest that greater statistical power would reveal a significant effect as well. Another cross-dimensional difference was indicated by the data, namely that contrast guidance decreased rather gradually with increasing feature-ratio, whereas for the other dimensions this decrease occurred abruptly between proportion groups 1 and 2. However, only a statistical tendency towards a gradual pattern for contrast was found.

All of the present results, but especially the feature-ratio effects, illustrate the fact that the current study did not just replicate known effects for yet another set of visual search stimuli. Instead, the study showed that top-down control based on low-level search features guides visual search, even in complex and natural displays. While the use of simple, artificial search displays eliminates the high-level influence of semantic context on visual guidance, it introduces the possibility for another type of high-level influence—the conscious learning of task-specific search strategies. This possibility arises from the dramatic reduction, as compared to randomly selected natural scenes, in the variance of simple displays across trials. For example, let us consider a typical visual search experiment that employs two different distractor types to study the distractor-ratio effect. Most likely, the only variables that vary from trial to trial are the positions of search items, the ratio between the two distractor types, and the presence of the target. After a few practice trials, most participants will understand this variance at least roughly. Because they are instructed to perform their task as quickly and as accurately as possible, they are likely to think about how to take advantage of this variance knowledge to speed up their searches. Eventually, many of them will find out that they can determine the presence of a target faster if they search through the distractors of the less represented type. Then these participants only have to identify this type quickly after display onset to decide about their search strategy for the current trial. Consequently, their behavior during the experiment may be influenced by task-specific, high-level decision processes, which most likely differ from the mechanisms that guide search during everyday tasks.

In contrast, the variance in the complex displays used in the present study is comparable to the variance in real-world scenes that we face every day. Therefore, the participants in the present study already possessed highly optimized search mechanisms that allowed them to reach their maximum performance in the given task. It was neither necessary nor possible for them to learn a specific, artificial search strategy for more efficient search. This way it was ensured that the participants' search was guided by pre-attentive processes that are similar to the ones that they use so efficiently in everyday visual tasks. It must be noted, though, that the complex displays introduced the possibility of semantic influence on visual attention, even though the experimental method aimed at minimizing this effect. Regardless of this possible semantic influence, the results show the presence of visual guidance by low-level features in complex displays. The finding of feature-ratio effects for these low-level features further supports the view that pre-attentive, parallel processes guide our search even in complex, natural scenes. It is implausible that participants consciously analyze the proportions of different target features in the display before deciding on a search strategy or on the next saccadic target.

The present results also relate to current research on eye movements during scene perception (see Henderson, 2003, for a review). Most of these studies have examined free viewing of natural scenes. Based on the results, several computational models of eye movements and visual attention have been proposed. Probably, the most prominent models are the ones by Itti and Koch (2000) and by Parkhurst, Law, and Niebur (2002). Both models employ computational schemes for deriving a saliency map from local display features such as color, brightness, and orientation. This saliency map guides visual attention and eye movements during scene perception. Both approaches predict eye-movement patterns that correlate significantly with actual human scan paths for the same displays. It is characteristic for this line of research that both models only consider bottom-up influences on saccadic target selection, but do not account for top-down guidance. The current work demonstrates the extent to which top-down processes determine saccadic selectivity in natural scenes, even with regard to low-level features. These findings point out some significant limitations of models that use low-level salience as the only predictor of saccadic target selection.

Future research needs to address the role of color information in the cross-dimensional guidance of visual search, which was excluded from the current exploratory study. Another restriction of the present study was the constant size of the search targets as well as the screen units for the measurement of local variables. They were simply set to diameters of $\approx 2°$ of visual angle, which approximates the size of the human fovea. Further studies should account for the variable size of the visual span, that is, the area around fixation from which task-relevant information can be extracted (Bertera & Rayner, 2000). Such empirical data could be used to advance existing computational models of visual guidance such as the one by Rao et al. (2002) or the Area Activation Model (Pomplun, Reingold, Shen, & Williams, 2000; Pomplun, Shen, & Reingold, 2003) towards complex search displays.

Acknowledgments

I thank May Wong, Zhihuan Weng, and Chengjing Hu for their help with setting up the experiment. I am also grateful to Jiye Shen and two anonymous reviewers for their helpful comments on earlier versions of the manuscript and to Michelle Umali for her editorial assistance.

Appendix A.

Each of the four local stimulus variables is computed for a 64×64 pixel area centered at pixel (x, y) in a given search display. For each variable, after its computation as described below, its values are scaled to range from 0 to 1 across the 200 search displays. The local-contrast variable $L_i(x, y)$ is simply defined as the average brightness within the 64×64 pixel input square (see Fig. 1C, first row):

$$L_i(x,y) = \frac{1}{64^2} \sum_{x_p=x-32}^{x+31} \sum_{y_p=y-32}^{y+31} I(x_p, y_p),$$

where $I(x_p, y_p)$ is the intensity at pixel (x_p, y_p) in the display. Similarly, the local-contrast variable $L_c(x, y)$ is computed as the standard deviation of intensity within the specified square (see Fig. 1C, second row):

$$L_c(x,y) = \frac{1}{64}\sqrt{\sum_{x_p=x-32}^{x+31}\sum_{y_p=y-32}^{y+31}[I(x_p,y_p) - L_i(x,y)]^2}.$$

To determine spatial frequency and orientation variables, a discrete Fourier transform of the input square is performed. The result of this computation, which is implemented as a fast Fourier transform (FFT), is a 64×64 array of complex numbers, with the real (Re) and imaginary (Im) components specifying the coefficients of cosine and sine functions, respectively, of different spatial frequencies that add up to match the 64×64 input data (e.g., Brigham, 1988). It is important to notice that the Fourier transform is periodic in nature and thus "assumes" that the intensity pattern it receives repeats infinitely in the horizontal and vertical directions. Consequently, mismatches between the left and right sides of the input square, or between its top and bottom, will have the same effect as vertical or horizontal edges, respectively, on the result of the Fourier transform. Therefore, studies on image statistics that do not compensate for this effect will inevitably find an artificially elevated proportion of horizontal and vertical edges. In the present study, a Blackman window function (see Gonzalez & Woods, 2002) is applied to minimize the discontinuities at the borders of the input image and thereby avoid such artifacts. For a given offset (x_0, y_0) from the center of the input square, within a radius of 31 pixels, the window function W is computed as follows:

$$W(x_0,y_0) = 0.42 - 0.5\cos\left[\pi\left(1-\frac{\sqrt{x_0^2+y_0^2}}{31}\right)\right] + 0.08\cos\left[2\pi\left(1-\frac{\sqrt{x_0^2+y_0^2}}{31}\right)\right].$$

For greater radii, W is set to zero. As shown in Fig. 7A, W is a smooth function that gently approaches zero as it reaches the borders of the input square. If we take the deviation of each input pixel's intensity from the average intensity within the input square and multiply it with its associated value of W, the resulting pattern can be Fourier transformed with only minimal border artifacts or other distortions. Taken together, for an input square centered at the display coordinates (x,y), the complex coefficients for the frequency domain coordinates (u,v) with $-32 \leqslant u < 32$ and $-32 \leqslant v < 32$ are computed as:

$$F(x,y;u,v) = \sum_{x_0=-32}^{31}\sum_{y_0=-32}^{31}[I(x+x_0,y+y_0) - L_i(x,y)] \cdot W(x_0,y_0)\cdot e^{-j\frac{2\pi}{64}(ux_0+vy_0)}.$$

From these cosine and sine coefficients, we can derive the amplitude spectrum $|F(x, y; u, v)|$ that measures the contribution of particular 2D spatial frequencies to the local display content at position (x, y):

$$|F(x,y;u,v)| = \sqrt{\mathrm{Re}(x,y;u,v)^2 + \mathrm{Im}(x,y;u,v)^2}.$$

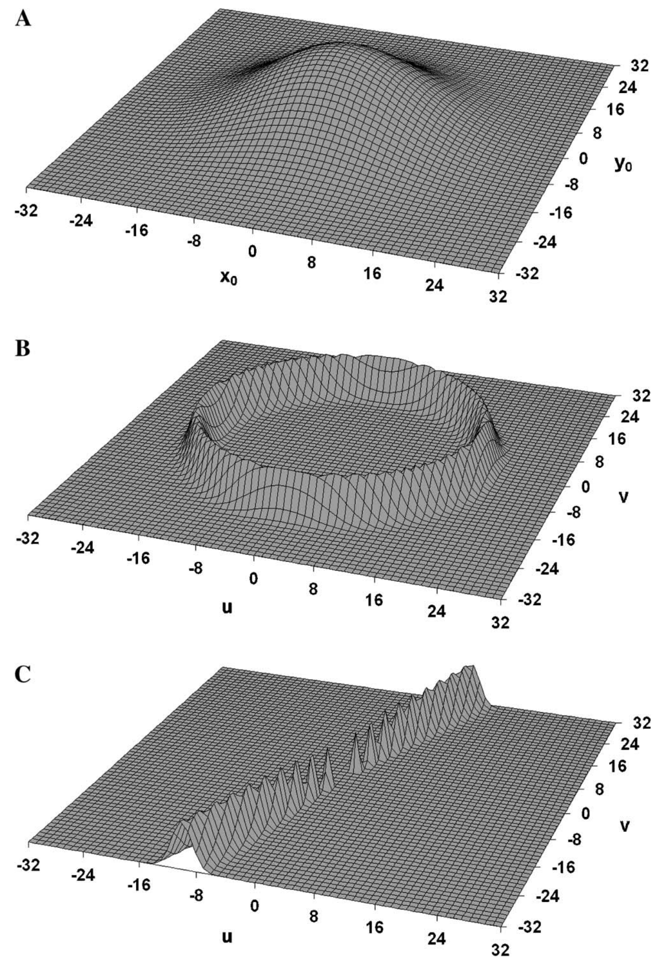

Fig. 7. Supporting 2D functions for the accurate measurement of local predominant spatial frequency and orientation of edges. (A) Blackman window function; (B) smooth weighting function for the computation of spatial frequency, shown for $r = 20$; (C) smooth weighting function for the computation of orientation, shown for $\theta = 163°$.

In natural images, the statistical distribution of these amplitudes is not homogeneous across frequencies (e.g., Torralba & Oliva, 2003). To compensate for that, the average amplitude spectrum $|\overline{F}(u,v)|$ across all positions in all 200 search displays is computed, and then the relative amplitude spectrum $A(x,y;u,v)$ at position (x,y) in a given display is defined as

$$A(x,y;u,v) = \frac{|F(x,y;u,v)|}{|\overline{F}(u,v)|}.$$

The absolute frequency represented at position (u,v) in the relative amplitude spectrum increases with the distance of (u,v) from the center of the spectrum. Thus, the basic idea for determining the value of the local predominant frequency variable $L_f(x,y)$ is to find the radius r of the circle centered at $(0,0)$ in the spectrum that touches the greatest average relative amplitude. In order to reduce the noise in the measurement, a radial Gaussian distribution is used to generate a smooth weighting function $f(u,v;r)$:

$$f(u,v;r) = \mathrm{e}^{-\frac{\left(r-\sqrt{u^2+v^2}\right)^2}{2\sigma_f^2}}.$$

Throughout this study, the standard deviation σ_f was set to 1.2. Fig. 7B illustrates the weighting function f for $r = 20$. Using this function, the local predominant frequency $L_f(x, y)$ can finally be computed as

$$L_f(x,y) = \underset{0<r\leqslant 31}{\arg\max} \sum_{u=-32}^{31} \sum_{v=-32}^{31} A(x,y;u,v) \cdot f(u,v;r).$$

The computation of local predominant orientation follows the same principle; it just uses a different weighting function $o(u, v; \theta)$. Edges of a given orientation θ—measured as the deviation ($0° \leqslant \theta < 180°$) from a horizontal line in counterclockwise direction—in the input square will elevate the values in the amplitude spectrum at a polar angle that is perpendicular to this orientation. Once again with the help of a Gaussian function, the weighting function is computed by

$$o(u,v;\theta) = \mathrm{e}^{-\frac{(\theta-[\arctan(v,u)+90^\circ]\mathrm{mod}180^\circ)^2}{2\sigma_o}},$$

where the modulo function is defined for real numbers, and the standard deviation σ_o is set to 2.2°. Fig. 7C shows the function o for $\theta = 163°$. Based on this function, $L_o(x, y)$ is defined by

$$L_o(x,y) = \underset{0^\circ\leqslant\theta<180^\circ}{\arg\max} \sum_{u=-32}^{31} \sum_{v=-32}^{31} A(x,y;u,v) \cdot o(u,v;\theta).$$

The third and fourth rows in Fig. 1C present sample areas with varying predominant spatial frequency and orientation, respectively. In the present study, 1000 different values of r and θ were used for each computation of L_f and L_o. Notice that due to the symmetry properties of the Fourier transform only one half of the amplitude spectrum needs to be evaluated. In the above description, the entire spectrum was considered for illustrative purposes.

References

Bacon, W. F., & Egeth, H. E. (1997). Goal-directed guidance of attention: Evidence from conjunctive visual search. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 948–961.

Baddeley, R. (1997). The correlational structure of natural images and the calibration of spatial representations. *Cognitive Science, 21*, 351–372.

Bertera, J. H., & Rayner, K. (2000). Eye movements and the span of the effective visual stimulus in visual search. *Perception & Psychophysics, 62*, 576–585.

Brigham, E. O. (1988). *The fast Fourier transform and applications*. Englewood Cliffs, NJ: Prentice Hall.

Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology, 22*, 225–271.

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 32–39.

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* (pp. 135–159). New York: Psychology Press.

Gilchrist, I. D., Heywood, C. A., & Findlay, J. M. (2003). Visual sensitivity in search tasks depends on the response requirement. *Spatial Vision, 16*, 277–293.

Gonzalez, R. C., & Woods, R. E. (2002). *Digital image processing* (2nd Ed.). Upper Saddle River NJ: Prentice Hall.

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

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

Hooge, I. T., & Erkelens, C. J. (1999). Peripheral vision and oculomotor control during visual search. *Vision Research, 39*, 1567–1575.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489–1506.

Kaptein, N. A., Theeuwes, J., & van der Heijden, A. H. C. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 1053–1069.

Motter, B. C., & Belky, E. J. (1998). The guidance of eye movements during active visual search. *Vision Research, 38*, 1805–1815.

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

Poisson, M. E., & Wilkinson, F. (1992). Distractor ratio and grouping processes in visual conjunction search. *Perception, 21*, 21–38.

Pomplun, M., Reingold, E. M., Shen, J., & Williams, D. E. (2000). The area activation model of saccadic selectivity in visual search. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the Twenty Second Annual Conference of the Cognitive Science Society* (pp. 375–380). Mahwah, NJ: Erlbaum.

Pomplun, M., Reingold, E. M., & Shen, J. (2001). Peripheral and parafoveal cueing and masking effects on saccadic selectivity in a gaze-contingent window paradigm. *Vision Research, 41*, 2757–2769.

Pomplun, M., Ritter, H., & Velichkovsky, B. M. (1996). Disambiguating complex visual information: Towards communication of personal views of a scene. *Perception, 25*, 931–948.

Pomplun, M., Shen, J., & Reingold, E. M. (2003). Area activation: A computational model of saccadic selectivity in visual search. *Cognitive Science, 27*, 299–312.

Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research, 42*, 1447–1463.

Scialfa, C. T., & Joffe, K. (1998). Response times and eye movements in feature and conjunction search as a function of eccentricity. *Perception & Psychophysics, 60*, 1067–1082.

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

Shen, J., Reingold, E. M., Pomplun, M., & Williams, D. E. (2003). Saccadic selectivity during visual search: The influence of central processing difficulty. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eyes: Cognitive and applied aspects of eye movement research* (pp. 65–88). Amsterdam: Elsevier Science Publishers.

Therrien, C. W. (1992). *Discrete random signals and statistical signal processing*. Englewood Cliffs: Prentice-Hall.

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

Williams, D. E., & Reingold, E. M. (2001). Preattentive guidance of eye movements during triple conjunction search tasks. *Psychonomic Bulletin and Review, 8*, 476–488.

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

Wolfe, J. M. (1996). Extending guided search: Why guided search needs a preattentive "item map". In A. F. Kramer, M. G. H. Coles, & G. D.

Logan (Eds.), *Converging operations in the study of visual attention* (pp. 247–270). Washington, DC: American Psychological Association.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–71). England UK: Hove.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 419–433.

Zelinsky, G. J. (1996). Using eye saccades to assess the selectivity of search movements. *Vision Research, 36*, 2177–2187.

Zohary, E., & Hochstein, S. (1989). How serial is serial processing in vision? *Perception, 18*, 191–200.