How Chromaticity Guides Visual Search in Real-World Scenes

Alex D. Hwang (ahwang@cs.umb.edu)
Emily C. Higgins (emilychiggins@gmail.com)
Marc Pomplun (marc@cs.umb.edu)
University of Massachusetts at Boston, Department of Computer Science
100 Morrissey Boulevard, Boston, MA 02125-3393, USA

Abstract

To study 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. Recently, a first study on saccadic selectivity in real-world scenes was conducted, demonstrating visual guidance by low-level features such as intensity and spatial frequency (Pomplun, 2006). However, that study only used grayscale displays, because chromaticity is assumed to dominate search behavior in such a way that it might be difficult to measure concurrent guidance by other dimensions. To test this assumption and to assess the effects of chromaticity on visual search performance, time course, and feature guidance in real-world displays, the present study measured eye movements during two versions of a set of 160 real-world displays: One version contained its natural colors, whereas the other version was converted to grayscale. The results indicate that the hue dimension indeed dominates search at the cost of other dimensions, and that chromaticity information leads to faster target detection without influencing the high-level, strategic control of visual attention.

Keywords: Visual search; eye movements; saccadic selectivity; visual guidance; chromaticity

Introduction

When visually searching a scene for a target item, we move our eyes from one fixation point to the next, selectively attending to some sequence of locations in the landscape. What determines which elements of the scene will draw our attention in the course of a given search task? How is our visual attention guided during search?

One of the most prominent research efforts, the Guided Search Theory (see, e.g., Wolfe, 1998), focuses upon the guidance of 'lower-level', perceptual processes during search. The Guided Search Theory posits two stages in the process of visual search. During the first stage, what is known as an 'activation map' of a scene is developed. In this stage, all locations within a scene are assigned activation values. The total activation of a given scene location is the result of the combined effects of (1) top-down (or task-driven) activation, which will rise with increased similarity to the target, and (2) bottom-up (or stimulus-driven) activation, which is independent of the target, but varies according to the location's distinctiveness within the scene. Then, in the second stage of visual search, we serially attend to those regions of the map with highest activation values.

The Guided Search Theory has been shown to be consistent with a wide variety of psychophysical visual search data (e.g., Brogan, Gale, & Carr, 1993). Besides the standard measures of response time and error rate, these data also encompass more fine-grained measures, most importantly eye-movement patterns. 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 type of non-target item (distractor), 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 distractors sharing a certain feature such as color or shape with the target item received a disproportionately large number of saccadic endpoints (e.g., Findlay, 1997; Hooge & Erkelens, 1999). Almost all of these studies employed simple, artificial search displays for better experimental control and simpler data analysis. Our visual system, however, evolved in and is trained for real-world visual input.

Therefore, the recent increase in using real-world visual search stimuli is not surprising. One line of inquiry, for example, investigates how the semantic context of a scene and our expectations about where things belong in the world might help to guide our attention during search (e.g., Neider & Zelinsky, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006). In Neider & Zelinsky (2006), participants were asked to search a landscape for objects, such as jeeps and helicopters, which we expect to find in certain fixed environments. Using eye-movement recording and reaction time measures, the researchers gauged the improvements in search efficiency under conditions in which actual target positions conform to our expectations.

The first research effort to measure saccadic selectivity for basic visual features in real-world images was recently conducted by Pomplun (2006). In this study, significant visual guidance by features such as intensity and contrast were found during the search of complex scenes. However, all images were presented in grayscale. This was done so that effects along other dimensions, with potentially weaker guidance, could be discerned. Here, we build upon this foundation and proceed to compare these processes in color and grayscale viewing conditions.

In the present study, we record participants' eye movements to investigate how lower-level features, such as

intensity and intensity contrast, might guide visual attention differently in grayscale and color versions of real-world scenes. In addition, guidance by chromaticity features is measured in order to test whether they – especially along the hue dimension – indeed exert stronger guidance than others. For abstract color search displays using a large set of 64 colors it was shown that search is most strongly guided by the hue dimension, followed by intensity and saturation (Xu, Higgins, Xiao & Pomplun, to appear). In the same study, through various approaches to modeling color guidance, it was also suggested that the HSI (hue, saturation, intensity) color space is especially well-suited for describing such guidance effects. Therefore, the guidance analysis in the present study is also based on the HSI color space.

Besides analyzing visual guidance, we use the obtained eye-movement data to gain insight into possible differences in the time course of search processes between grayscale and color displays. The variables we examine include "standard" eye-movement variables such as fixation duration, saccade amplitude, initial saccadic latency, but additionally we use the gaze-position information to determine how quickly participants approach the target region. This set of analyses is aimed at studying the attentional effect of chromaticity - which is widely assumed to dominate visual search behavior - in a quantitative manner in real-world displays.

Method

Participants

Thirty participants performed this experiment. All were students or faculty members at the University of Massachusetts, Boston. Each was entitled to a $10 honorarium.

Apparatus

Eye movements were tracked using an SR Research EyeLink-II system. After calibration, the average error of visual angle in this system is 0.5°. Its sampling frequency is 500 Hz. Stimuli were presented on a 19-inch Dell P992 monitor. Its refresh rate was set to 85 Hz and its resolution to 1280×1024. Participant responses were entered using a handset or game-pad.

Materials

A total of 160 photographs (resolution 800×800 pixels) of real-world scenes, including landscapes, home interiors, and city scenes, were selected as stimuli (see Figure 1, left column). They were presented in grayscale and in color conditions (as described below). From each scene, a cutout of 64×64 pixels was selected as a target. Targets were chosen randomly, but were inspected and in several cases rejected and newly chosen to exclude uninformative (e.g., completely black or white), ambiguous, or semantically rich locations. In order to further minimize semantic effects during search - while conserving the type and distribution of the natural low-level visual features - the scenes were rotated randomly by 90, 180, or 270 degrees. The previewed target element of the scene was in each case likewise rotated. A central screen region of 192×192 pixels was excluded as a possible source of target locations. Target locations were otherwise distributed approximately evenly across the display area. Participants sat approximately 60 cm from the screen. The horizontal and vertical viewing angle was about 1° for the target image and about 13° for the search display.

Procedure

We began each trial by providing the participant with a short set of instructions and fitting him/her with the eye-tracking headset. A nine-dot calibration was then performed. An additional single-dot, drift-correcting calibration was also performed before each trial.

The set of stimulus photographs was divided, by random selection, into 2 groups of 80, Set A and Set B. Participants were also evenly divided into 2 groups of 15. One half of participants viewed Set A of photographs in color and Set B in grayscale, while the other half of participants viewed Set A in grayscale and Set B in color. All participants viewed 4 blocks of 40 photographs, in an alternating sequence of grayscale and color blocks. One participant group first viewed a block of photographs presented in color while the other first viewed a block of scenes presented in grayscale. The order of presentation individual scenes was held constant across the participant groups, but whether the scene was presented in grayscale or in color differed between groups.

After each block of trials, the participant was given the opportunity to take a break from the experiment. Before beginning each of the 160 search trials, a participant was to fixate upon a central marker while pressing a button on the game pad to correct for drift. The target element of the scene would then appear at the center of the screen for 2 seconds. After this preview of the target, the full scene was presented. Participants were to search the scene and, when they believed they had found the target, to press a button on the game-pad while fixating on this location. If the participant did not press this button within 6 seconds of the onset of the photograph, the trial would time-out and terminate and the next trial would begin. After each trial, a small white or yellow square indicated the correct target location (see Figure 1, left column).

Data Analysis

While the analysis of standard eye-movement variables in this study is straightforward, the computation of visual guidance is more complex, and sometimes rather arbitrary decisions need to be made. Following Pomplun (2006), we first compute an "attentional landscape" for each search display by centering a Gaussian function (standard deviation 1° to approximate the human fovea size) on each fixation made by any of the participants who viewed that display. These functions were summed across the display, creating a smooth landscape indicating fixation density, which is

closely correlated with the distribution of attention in the display (Findlay, 2004). By multiplying the intensity of each pixel in a display with the corresponding value of the landscape, a visualization of attention can be created such as in Figure 1 (right column).

The basic idea underlying our variant of computing visual guidance for a particular stimulus dimension is as follows: We determine for each position in every display how similar the local features in that dimension are to the target features. For example, let us consider intensity. If the target is mostly bright, and a local display area is also bright, then there is great similarity. If we have a similarity measure to quantify this, we can compute the similarity of local areas with the target across the display, thereby obtaining a feature similarity landscape. If intensity guides attention, we would expect that areas of higher similarity (e.g., bright areas when the target is bright) receive more saccadic endpoints than other, less similar areas. We can exploit this fact by computing the correlation (Pearson's r) between the attentional landscape and the feature similarity landscape. If r = 1, it then means that search is entirely guided by intensity, and if r = 0, intensity does not guide search at all. This is our operational definition of visual guidance.

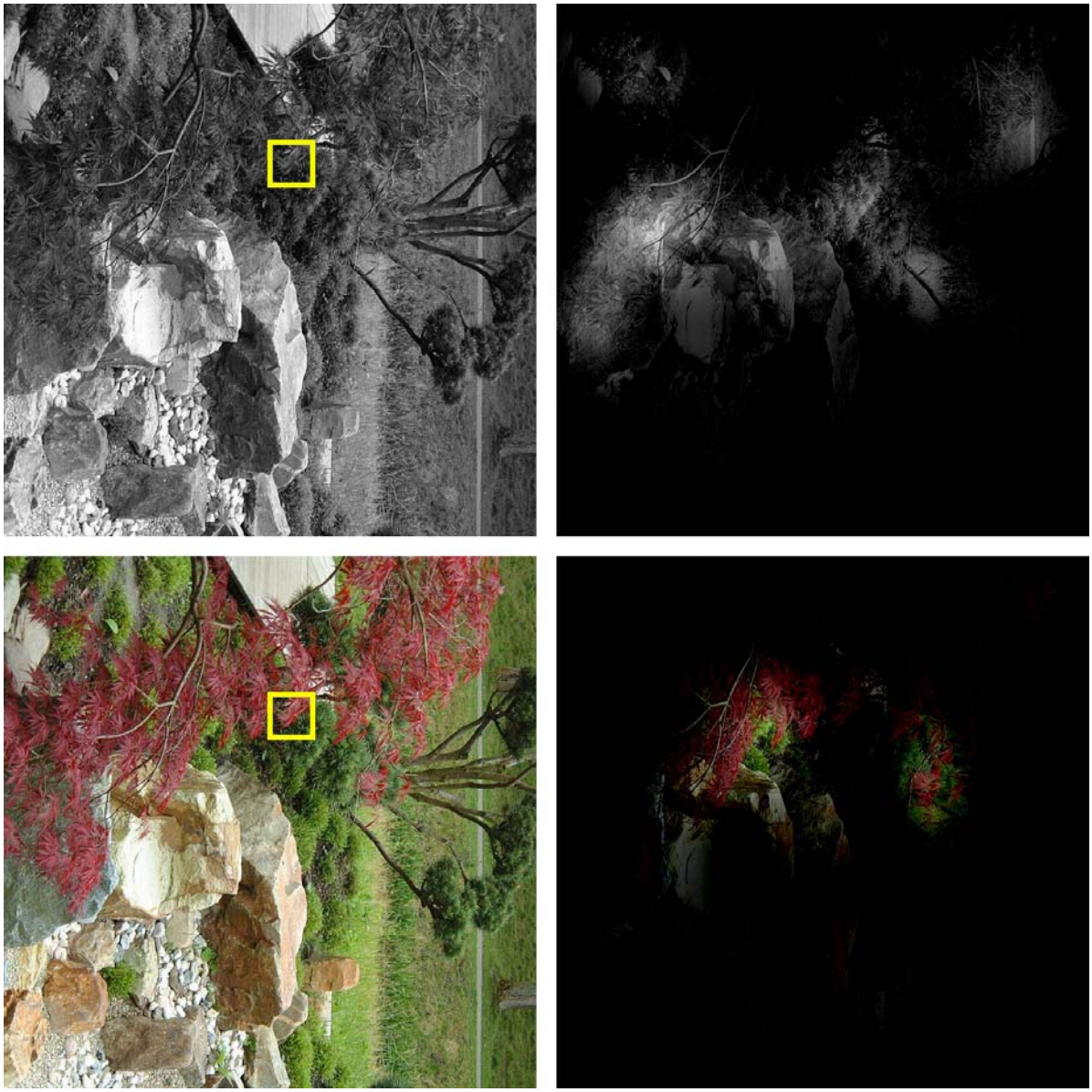

Figure 1: Left column: Sample display (top panel: grayscale version, bottom panel: color version) with the target "cutout" marked by a yellow square, as it was indicated to the participants after their response or time-out. Right column: Visualization of the "attentional landscape", i.e., the distribution of saccadic endpoints for each version of the display. In the grayscale image (top panel), it can be seen that participants consider a larger area - containing leaves - for their search, whereas in the color display they focus on areas containing both red and green hues.

However, how shall we define similarity along a given stimulus dimension in such complex stimuli? For example, in the color example shown in Figure 1 (bottom), it does not seem to be one dominant or average hue that guides search, but rather a combination of two hues, namely red and green. To account for this complexity, we compute the similarity of histograms for the given stimulus dimension rather than the similarity of average values. Features along every dimension whose guidance we would like to study are divided into ten intervals of equal size. Then histograms – the number of pixels falling into each of the ten intervals – are computed for the target and for local 64×64 pixel areas in each display. Adjacent areas are placed in such a way that they overlap by 50%, which leads to the computation of 676 histograms pre display.

To compute the similarity between the target and a local area for a given dimension, we use the histogram intersection distance (Swain & Ballard, 1991), which has been successfully applied to image retrieval problems. Basically, for each bin, the smaller value in the two histograms is taken. These values are summed across all ten bins, resulting in the measure of similarity. Notice that the eye-movement data of all participants need to be accumulated for a useful attentional landscape. Therefore, in this study the standard error and statistical significance are computed across the 160 displays instead of the 30 subjects (cf. Pomplun, 2006).

Results and Discussion

As expected, participants reported target detection faster in the color condition (4068 ms) than in the grayscale condition (4772 ms), $t(159) = 8.25$, $p < 0.001$. In this analysis, for timed-out trials (14.5% of color trials and 25.8% of grayscale trials) the total trial duration of 6 seconds was entered. In order to assess the accuracy of the participants' performance, we distinguished between two variables: First, we measured the variable *response accuracy*, defined as the proportion of trials in which the participants' gaze position during manual response was closer than 2° of visual angle to the center of the target area. Second, we measured *cover accuracy* as the proportion of trials during which participants fixated at least once within a radius of 2° from the center of the target area. The motivation for introducing this second variable was that the task was so difficult that participants often saw the target quite quickly but could not decide – or not decide correctly – which of two or more similar areas contained the target.

A two-way analysis of variance (ANOVA) with the factors chromaticity (grayscale vs. color) and measurement (response vs. cover accuracy) revealed that accuracy was greater for color (response: 0.317; cover: 0.543) than for grayscale (response: 17.2; cover: 0.403), $F(1; 159) = 83.93$, $p < 0.001$ (see Figure 2). Expectedly, cover accuracy was greater than response accuracy, $F(1; 159) = 625.63$, $p < 0.001$, while there was no interaction between the two factors, $F(1; 159) < 1$. This finding demonstrates that chromaticity information facilitates both the determination of relevant display areas and the decision about the target location.

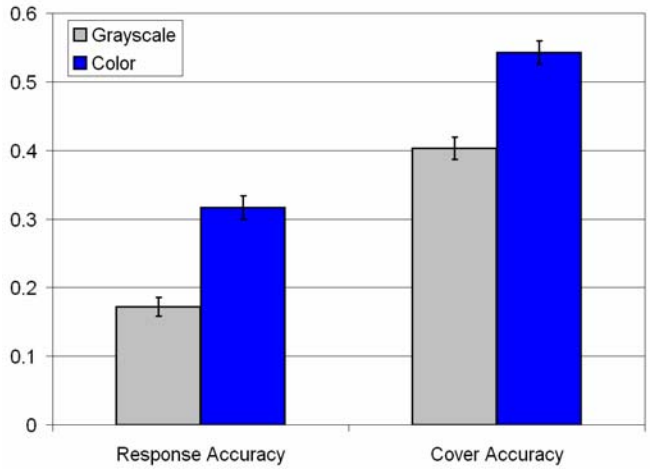

Figure 2: Response accuracy (proportion of trials with correct target report) and cover accuracy (proportion of trials that included at least one target fixation). In all figures, error bars indicate standard error across the 160 displays.

In order to study the temporal aspects of this facilitation, we computed histograms of the number of fixations that participants made before their gaze entered the target area (2° radius) for the first time. Figure 3 shows these histograms for the grayscale and color conditions. The frequency in each bin was measured as the proportion with regard to the entire set of trials, including those without manual response. However, trials without manual response were not included in the rightmost bin. This arrangement allows a "fair" comparison of absolute values between grayscale and color trials.

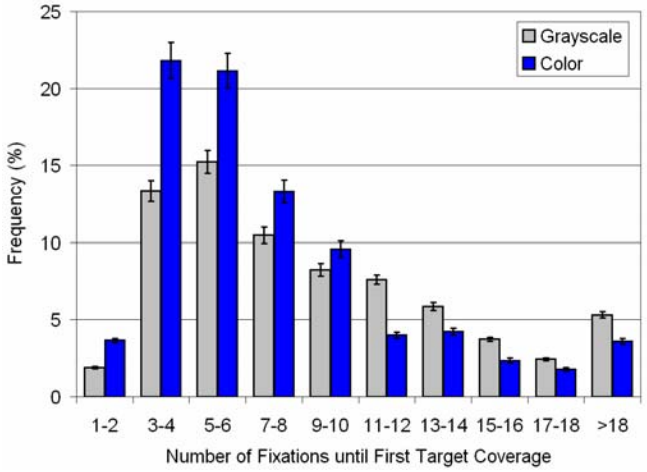

Figure 3: Histograms of the number of fixations made until entering the target area for the first time

While the target was unlikely to be hit within the first two fixations in either condition, there was a dramatic increase in this probability for fixations number 3 to 6. For color

displays, this increase was especially strong, followed by a steep decrease after fixation number 6 and very small frequencies after fixation number 10. Grayscale displays exhibited a much more gradual decrease so that after fixation number 10, initial target coverage was more likely for grayscale displays than for color displays. This difference in the shapes of the two histograms was reflected by a significant interaction of the factors *fixation bin* and *display type* in a two-way ANOVA, $F(9, 1431) = 460.89$, $p < 0.001$. In summary, the data suggest that chromaticity does not usually guide attention to the target area right from the start, but typically allows participants to attend to the target after only a few additional saccades, which is clearly less likely without chromaticity information.

Another way of studying cognitive effects by chromaticity is the analysis of basic eye-movement variables. One of these variables is initial saccadic latency, measured as the time between stimulus onset and the execution of the first saccade. Longer initial latency is typically associated with a more strategic selection of first fixation targets, i.e., increased top-down control of attention. Initial saccadic latency was found to be greater for grayscale images (391 ms) than for color images (371 ms), $t(159) = 3.04$, $p < 0.005$ (see Figure 4, left). This result suggests that top-down control plays a more important role in grayscale search, whereas bottom-up, "automatic" guidance is more emphasized in color search.

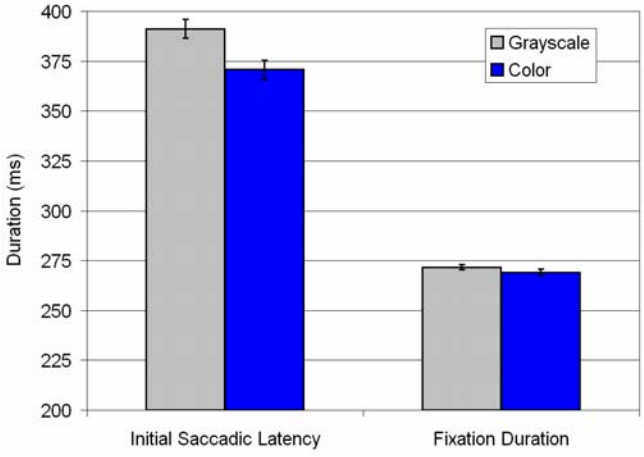

Figure 4: Initial saccadic latency and fixation duration

One of the most basic eye-movement variables is *fixation duration*, informing us about how long it takes an observer to retrieve task-relevant information from the area around the current gaze position. Longer fixation duration is thought to reflect more demanding tasks or higher levels of cognitive processing (e.g., Rayner, 1998). Interestingly, there was no significant difference in mean fixation duration between grayscale (272 ms) and color displays (269 ms), $t(159) = 1.52$, $p > 0.1$ (see Figure 4, right). Thus, there is no evidence from this variable that the presence of chromaticity reduces cognitive demands. The data analysis so far seems to indicate that adding chromaticity information does not induce higher-level cognitive, i.e., strategic changes, but rather improves visual guidance toward the most relevant display areas.

If this assumption is correct, then we would expect another basic eye-movement variable, *saccade amplitude*, to show no difference between the color and grayscale conditions either. In search tasks, longer saccades are typically found in less difficult tasks that require less systematic search, for example in a "pop-out" displays in which the target item is blue and all distractor items are red. Saccade amplitude is known to vary over the time course of viewing a real-world image (e.g. Henderson, 2003). When searching a complex display, the first saccade is often rather short in order to give the observer an initial "overview" of the entire display, followed by longer search saccades and finally very short saccades during the decision about the target and the manual response. To analyze this time course and determine possible differences between grayscale and color search, we computed saccade amplitude as a function of its temporal position within a trial. The time course of every trial that contained a button response was divided into five intervals of equal duration.

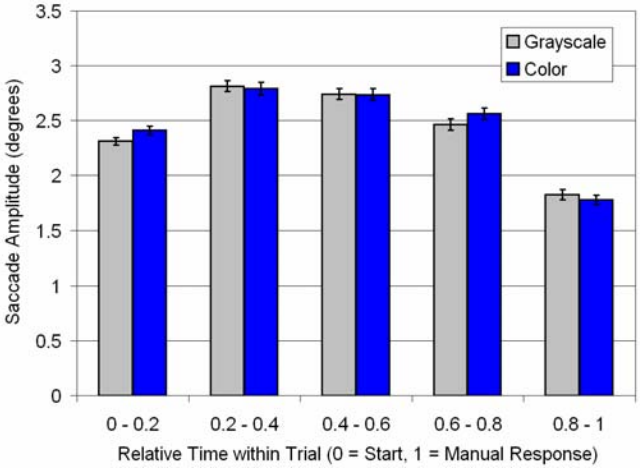

Figure 5: Saccade amplitude during the time course of a search trial.

Figure 5 illustrates the result of this computation. A two-way ANOVA with the factors *time interval* and *display type* showed that, as predicted, saccade amplitude differed across time intervals, $F(4; 636) = 189.90$, $p < 0.001$. Interestingly, though, there was no significant difference in mean saccade amplitude between grayscale (2.51°) and color displays (2.50°), $F(1; 159) < 1$. Moreover, there was no interaction between the two factors, $F(4; 636) = 1.58$, $p > 0.1$. This pattern of results indicates that chromaticity affected neither mean saccade amplitude nor its variation over the time course of a trial. It provides further evidence for the hypothesis that the availability of chromaticity does not lead to high-level, strategic changes in task performance.

Since the crucial difference between grayscale and color search then appears to be in the guidance of visual attention, it is important to study the factors underlying this guidance

more closely. We already know that the intensity and intensity contrast information strongly guide attention during search in grayscale displays (Pomplun, 2006). In the current experiment, these two variables are present in both types of display. Thus, it is informative to study the differences in the amount of visual guidance between display types. Do observers keep the same level of intensity and intensity contrast guidance when chromaticity is available, or will they shift their attention toward chromatic features at the cost of intensity and intensity contrast guidance? How does the strength of guidance by chromatic features compare to non-chromatic features? Is hue still the dominant dimension to guide search in real-world scenes, like it is in abstract displays (Xu et al, to appear)?

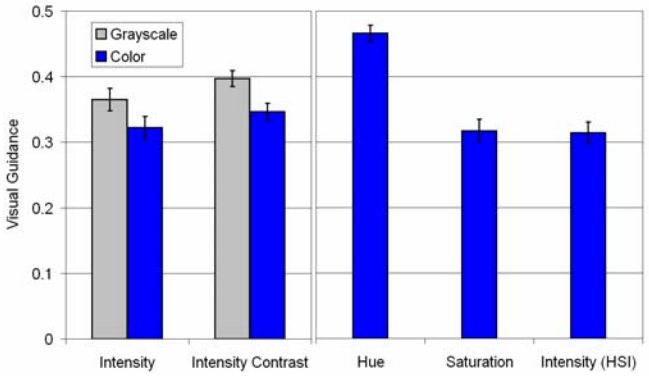

Figure 6: Visual guidance by intensity and intensity contrast in grayscale displays, and by intensity, intensity contrast, hue, saturation, and HSI intensity in color displays. HSI intensity slightly differs from perceptual intensity, because in the HSI space, red, green, and blue are equally weighted.

The results of this analysis are shown in Figure 6 (for a definition of visual guidance see Methods section). Intensity guidance was found to be smaller for color displays (0.32) than for grayscale displays (0.37), $t(159) = 3.27$, $p < 0.005$. Similarly, intensity contrast guidance was smaller for color displays (0.35) than for grayscale displays (0.40), $t(159) = 4.71$, $p < 0.001$, As expected, guidance by hue (0.47) was greater than guidance by intensity, intensity guidance, saturation (0.32), and HSI intensity (0.31), all ts$(159) > 8.07$, ps < 0.001. No other difference reached significance, all ts$(159) < 1.68$, ps > 0.05. These findings can be seen as evidence for the availability of chromaticity information, especially with regard to the hue variable, to dominate visual guidance and to reduce, at least slightly, the guidance by other stimulus dimensions.

Conclusions

The present study has demonstrated that in real-world scenes the presence of chromaticity allows us to direct our attention more quickly to a visually distinguished search target and to recognize this target with greater certainty. The present eye-movement analysis allowed us to study the mechanisms underlying this difference in task performance. It seems that the availability of chromatic features does not lead to higher-level cognitive, strategic changes in the way the search is performed, as suggested by the invariance of fixation duration and saccade amplitude. Even the time course of saccade amplitude, a fine-grained indicator of the different cognitive stages during a search process, did not differ between the color displays and their grayscale counterparts. However, the finding of shorter saccadic latency for color displays indicates that chromaticity information is likely to increase the contribution of low-level processes to attentional control.

Accordingly, the enhancement of search performance by chromaticity seems to be almost entirely based on the increased effectiveness of low-level visual guidance. It is especially the hue dimension that guides attention so quickly - typically within a few saccades - to the task-relevant display areas; observers rely on it so heavily that guidance by other dimensions decreases. Through these insights, the present study has provided a first glimpse at the mechanisms underlying chromaticity guidance of search in real-world scenes. It is just a very small step on the way towards understanding the control of visual attention.

Acknowledgments

This research was funded in part by an Undergraduate Research Award to Emily C. Higgins by the University of Massachusetts at Boston.

References

Brogan, D., Gale, A., & Carr, K. (1993). *Visual search 2*. London: Taylor & Francis.

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 Henderson, J.M. & Ferreira, F. (Eds.), *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*, pp. 135-159. New York: Psychology Press.

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

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

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.

Rayner, K. (1998). Eye Movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372-422.

Swain, M.J. & Ballard, D.H. Color indexing. International *Journal of Computer Vision*, 7:1 1991.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113, 766-786.

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

Xu, Y., Higgins, E., Xiao, M., & Pomplun, M (to appear). Mapping the color space of saccadic selectivity in visual search. *Cognitive Science*.