Visual Attention and Change Detection

Ty W. Boyer (tywboyer@indiana.edu)
Thomas G. Smith (thgsmith@indiana.edu)
Chen Yu (chenyu@indiana.edu)
Bennett I. Bertenthal (bbertent@indiana.edu)
Department of Psychological & Brain Sciences, Indiana University
1101 East Tenth Street
Bloomington, IN 47405 USA

Abstract

Studies suggest that visual attention, guided in part by features' visual salience, is necessary for change detection. An image processing algorithm was used for measuring the visual salience of the features of scenes, and participants' ability to detect changes made to high and low salience features was measured with a flicker paradigm while their eye movements were recorded. Changes to high salience features were fixated sooner, for shorter durations, and were detected faster and with higher accuracy than those made to low salience features. The implications of these results for visual attention and change detection research are discussed.

Keywords: change detection; visual salience; eye tracking

Introduction

Attention is the process of allocating perceptual and cognitive resources to select which information in the environment will enter consciousness. Finding a person in a crowd, locating one's car in a parking garage, identifying features that distinguish predator from prey, or finding a red circle amongst red and blue squares in a laboratory experiment all require control of visual attention (Kahneman & Henik, 1981; Triesman & Gelade, 1980; Wolfe, 2003). Attention contributes to an ability to make sense of our rich visual world and learn from experience.

The degree to which attention is guided by the features of the viewed stimulus versus the viewer's goals, expectations, and subjective evaluations are of paramount importance to researchers who study visual attention (Egeth & Yantis, 1997; Torralba, Oliva, Castelhano, & Henderson, 2006; Treue, 2003). Recently, computational saliency models have been developed to analyze visual scenes in terms of their stimulus properties (Itti, Koch, & Niebur, 1998; Koch & Ullman, 1985; Parkhurst, Law, & Niebur, 2002). These models have been used to predict viewers' fixation patterns as they view images, providing support for the suggestion that bottom-up visual saliency contributes to the guidance of visual attention (Itti & Koch, 2001; Foulsham & Underwood, 2008).

Other research shows that in the absence of visual attention we are particularly poor at detecting changes made to the features of scenes, a phenomenon known as change blindness (Hollingworth, 2006; Rensink, 2000a; 2000b; Rensink, O'Regan, & Clark, 1997; Simons & Levin, 1997). Changes that occur during saccades, eye-blinks, interleaved frames, cuts, and pans largely escape perceptual awareness. In experimental change detection tasks, visual salience is one factor guiding the direction of attention to features in the scene, and thus it is conjectured that salience contributes to whether and how quickly the changing feature will be detected (Kelley, Chun, & Chua, 2003; Simons & Ambinder, 2005). This is suggested under the assumption that viewers must direct visual attention to the feature that is changing, and are unlikely to do so if it is less salient than other features competing for visual attention.

Yet, contrary to this prediction, two recent studies, Stirk and Underwood (2007) and Wright (2005), report that the visual salience of stimulus image features, determined with formal salience algorithms, *does not* predict response times in a change detection task. By contrast, both of these studies found that the higher level semantic characteristics of changing features influenced their detection speeds (i.e., the changing feature's congruence with the theme of the scene, or whether it had been subjectively rated as high or low salience by independent viewers). Neither of these studies directly measured visual attention (i.e., eye movements), however, so it is uncertain how salience may have affected visual attention. As such, why these previous studies failed to find a relation between salience and change detection in requires closer scrutiny.

In the current research, eye movements are used to systematically assess the distribution of attention across each change detection trial. The primary goals of the current study are to examine: 1) whether stimulus feature salience predicts visual attention in a change detection paradigm, and 2) whether change detection requires overt visual attention, or whether covert attention suffices.

Computational saliency maps were used to identify the visual salience of the features within a set of images (Itti et al., 1998). Changes were applied to features identified as either high or low salience, and participants viewed these modified images interleaved with the originals in a flicker paradigm (Rensink et al., 1997). Visual attention was measured with a remote eye-tracking system that enabled examination of the fixation sequences that index overt visual attention, as well as the fixation durations that index the amount of cognitive processing of features during the search process. In Experiment 1 participants viewed scenes until they executed a manual response indicating change detection. Since the amount of time they had to view the scenes was open-ended, this experiment is limited in its ability to determine whether overt attention is necessary or whether covert attention is sufficient for change detection. In Experiment 2, participants viewed the scenes for a fixed percentage of the time that participants took to identify the change in Experiment 1, and their accuracy in localizing and identifying the change was measured. By limiting search time, there will be more trials in which subjects never fixate the change location, and it will be possible to test whether changes are detected even if they are not fixated, which would suggest that covert attention to the changed feature is sufficient for detecting the change.

Experiment 1
Method

Participants Thirty-five volunteers participated. They had normal or corrected to normal vision and were awarded course credit for participating.

Stimuli and Apparatus The stimuli were 28 digital bitmap photographs of outdoor scenes. The Itti, Koch, and Nieber (1998) saliency map algorithm was used to identify the relative visual salience of the features of each image. This algorithm makes center-surround comparisons at numerous scales on color, intensity, and orientation channels. See Figure 1 for the model and mathematical formula for the saliency maps, which were used to identify one high and one low salience feature in each image (see Figure 2 for an example of the composite saliency map output that was used to visually guide selection of features). Two modified versions of each image were produced in Adobe Photoshop by changing the color of or removing and background filling the identified high and low salience features. Figure 2 also shows the high and low salience modified versions of the image. Areas of change were defined with rectangles outlining the features that were modified. The saliency map algorithm has pixel level resolution, and on average 66.4% of the pixels in the high salience areas registered as salient on at least one of the channels, compared with an average of 3.6% of the pixels in the low salience areas, although high salience areas (138 x 154 pixels) were on average spatially slightly smaller than the low salience areas (208 x 133 pixels). The experiment was conducted on a Tobii 2150 eye tracking system, with gaze sampled at 50 Hz. The images were presented in 1024 x 768 resolutions on a 21.3” (54.1 cm) screen, from an average distance of 73cm. E-Prime software (Psychology Software Tools, Pittsburgh, PA) was used for presentation and data collection.

Design and Procedure Participants completed 56 trials; one low and one high salience feature change trial with each image. Each trial began with a cross at center screen, for a 1-sec fixation. Next, the original image was presented for 300 ms, followed by a blank gray screen for 100 ms, then the modified image for 300 ms, then the blank screen for 100 ms. This original-blank-modified- blank cycle repeated until the participant indicated detection of the change, or a maximum of 60 repetitions. The participant then typed a response to describe the change.

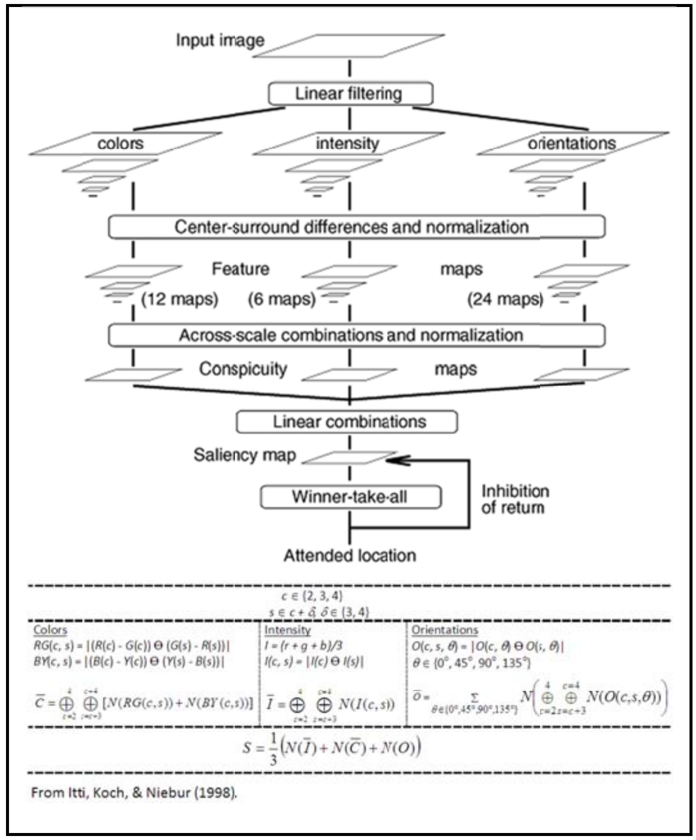

Figure 1. The saliency map algorithm used to identify high and low salience features in each of the stimulus images.

Results & Discussion

Trials where participants were unable to detect the change (3.77% of all trials) or misidentified what had changed (0.87% of all trials) were excluded from all analyses.

Response Times Figure 3 summarizes the response times per condition. A 2 x 2 repeated measures ANOVA with response time as the dependent measure revealed a significant main effect for salience (high vs. low), $F(1, 34) = 106.39$, $p < .0001$, $\eta_p^2 = .76$, a marginally significant effect for change type (color vs. presence/absence), $F(1, 34) = 3.79$, $p = .06$, $\eta_p^2 = .10$, and a salience x change type interaction, $F(1, 34) = 20.32$, $p < .001$, $\eta_p^2 = .37$. Changes made to high salience features were detected faster than those made to low salience features, and color changes were detected faster than presence/absence changes.

Continuous salience scores were computed for each area of change of each image, as above, in terms of the proportion of pixels within the area identified as salient by the saliency map algorithm. These salience scores were significantly correlated with mean response times, $r = -0.28$, $p = .04$, further indicating that response times decreased as the salience of the region of change increased. A partial correlation controlling for the surface area of the change revealed an even more significant relation between salience and response time, $r = -0.34$, $p = .01$.

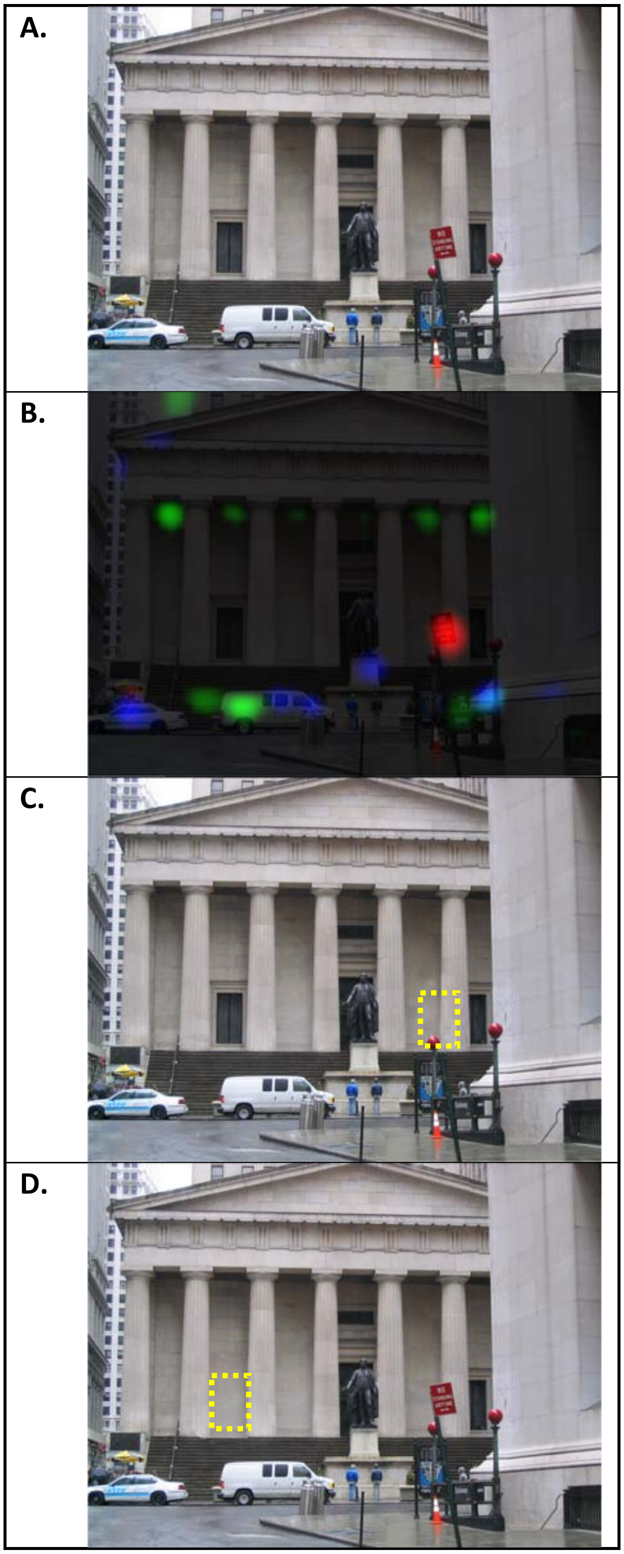

Figure 2. Example images: A) original; B) composite saliency map used to identify regions of high and low visual salience, with red indicating color salience, green intensity salience, and blue orientation salience; C) high salience feature change; D) low salience feature change. Note: Changed feature areas are illustratively outlined in C and D.

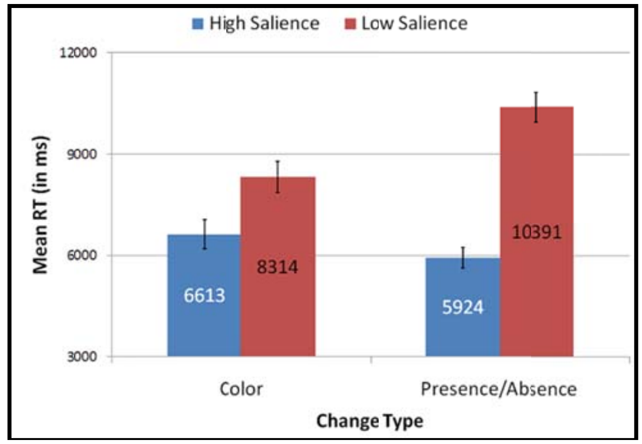

Figure 3. Mean RTs (with SEMs).

Eye Tracking Measures Figure 4 illustrates the median fixation sequence for the high and low salience version of one image. The high salience profile shows relatively few fixations (4) and a rapid convergence on the high salience change, whereas the low salience profile shows a large number of fixations (38), distributed across numerous features before fixation on the change, and termination with response. Figure 5 shows the median distances from the current fixation to the area of change over time, and illustrates that participants converged more quickly on high than low salience areas of change, and, moreover, that the distributions were much more skewed for low than high salience changes. We quantified the eye tracking measures two ways: 1) Time to Fixation was calculated as the time from the beginning of the trials until the first fixation on the change area. 2) Fixation Duration was the cumulative duration in the change area before a response, as a proxy for the relative difficulty of detecting the change.

A 2 x 2 repeated measures ANOVA with time to fixation as the dependent measure revealed significant main effects of salience, $F(1, 34) = 20.76$, $p < .001$, $\eta_p^2 = .38$, and change type, $F(1, 34) = 18.42$, $p < .001$, $\eta_p^2 = .35$, and a salience x change type interaction, $F(1, 34) = 16.35$, $p < .001$, $\eta_p^2 = .33$. Participants shifted gaze to the high salience change regions (M = 4273 ms) faster than low salience change regions (M = 5918). Time to fixation was significantly correlated with response time, $r = 0.88$, $p < .0001$, and the continuous salience measure of the change, $r = -0.25$, $p = .06$, also when controlling for surface area with partial correlation, $r = -0.36$, $p < .01$. A second 2 x 2 repeated measures ANOVA with fixation duration as the dependent measure revealed significant main effects of salience, $F(1, 34) = 13.52$, $p = .001$, $\eta_p^2 = .29$, and change type, $F(1, 34) = 13.92$, $p = .001$, $\eta_p^2 = .29$, and a non-significant salience x change type interaction, $F = 2.61$, *ns*. Fixation durations were lower for high salience changes (M = 674 ms) than low salience changes (M = 805).

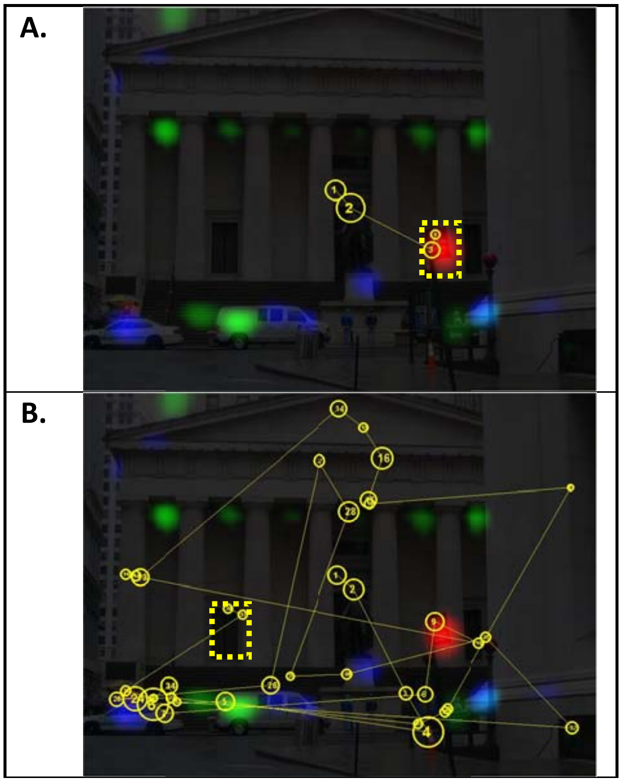

Figure 4. Fixation path for the participant with the median number of fixations prior to change detection for the outlined high (A) and the low salience change (B).

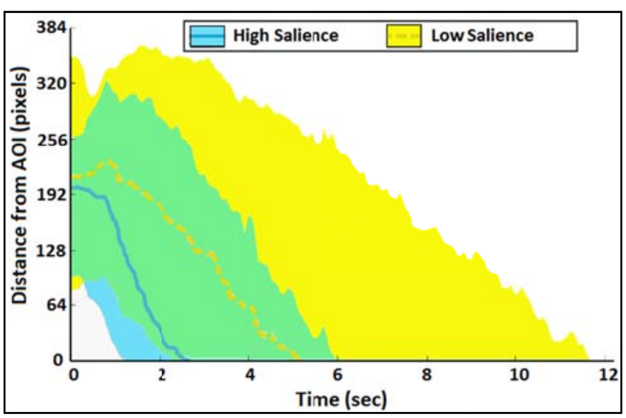

Figure 5. The solid blue and dotted orange lines are median distances from high and low salience areas of change. Shading indicates the 25th and 75th percentiles (light blue and yellow for high and low salience, green for the overlap).

Participants were faster to fixate and required less fixation time to identify changes made to features identified as highly salient by the saliency map algorithm. This indicates that the relative visual salience of changed features influenced the speed with which they were attended, as well as the cognitive load that was necessary to process them.

The eye movement measures indicate that visual salience guided where participants directed attention. This, in turn, raises the question of whether changes were covertly detected (i.e., without direct fixation), or whether overt visual attention is necessary, and whether this is affected by feature salience. Although the strong correlation between time to fixation and response times suggests that overt visual attention may be necessary, this remains an empirical question, because participants had an open-ended timeframe to scan the scene, and even if they had covertly detected the change they may have hesitated to respond until they had overtly verified it. Experiment 2 follows up on this by limiting the inspection times to a percentage of the median time needed by participants to identify the changes in this first experiment. This will reduce the likelihood that participants will fixate the changing features, and if covert attention is sufficient for change detection, then even in these impoverished conditions they will identify the changes.

Experiment 2
Method

Participants Twenty volunteers, with normal or corrected to normal vision, participated for course credit.

Stimuli and Apparatus All stimuli and equipment were the same as in Experiment 1.

Design and Procedure Participants completed 56 trials; one low and one high salience feature change trial with each image. A flicker paradigm with the same within trial timing parameters as Experiment 1 was used; however, rather than terminate the trial with a response, each stimulus was shown for a fixed amount of time. One half of the sample was shown the stimuli for 70% of the median times needed to detect the changes in Experiment 1, the other half was shown the stimuli for 90% the median times. Once the flicker cycle ended, participants localized the changed feature by mouse clicking on a blank gray screen and identified the change by typing a response.

Results & Discussion

Figure 6 summarizes the localization and identification accuracy data. Localization and identification rates were highly correlated, r = .89, p < .0001, and the statistical comparisons for each mirrored the other; therefore we only report analyses of the localization data here. A 2 x 2 x 2 mixed model ANOVA with localization accuracy as the dependent measure revealed significant main effects of salience, $F(1, 18) = 21.19$, $p < .0001$, $\eta_p^2 = .54$, with higher accuracy in the high than low salience trials (M = .66 and .54), and condition, $F(1,18) = 7.69$, $p = .01$, $\eta_p^2 = .30$, with higher accuracy in the 90% than 70% condition (M = .67 and .53). The effect of change type and all interactions were non-significant, all $F \leq 2.19$, all $p \geq .15$.

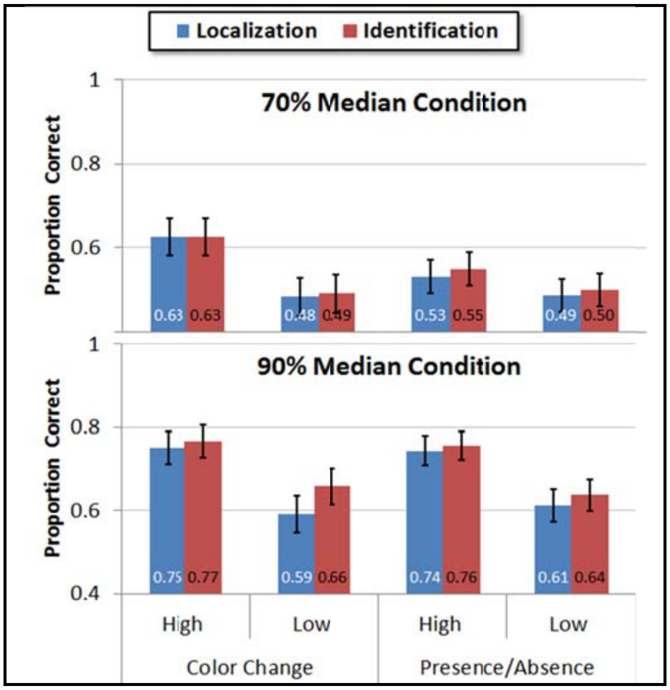

Figure 6. Proportion correct (SEMs).

Participants fixated the changing feature on 59.6% of trials. Localization accuracy was significantly correlated with whether the area of change had been fixated, $r = .51$, $p = .001$ and $r = .56$, $p < .001$, for the 70% and 90% conditions, respectively. Table 1 summarizes the relation between changed feature fixations and localization success. Two mixed model ANOVA were run, one with the proportion of trials participants fixated the change and the other with the time to fixation as the dependent measure. The proportion of trials where the change was fixated did not differ with salience, $F = .03$, *ns*, although high salience changes were fixated faster than low salience changes, $F(1,18) = 11.18$, $p < .01$, $\eta_p^2 = .38$ ($M = 2518$ and 3302 ms, respectively). Participants were more likely to fixate color than presence/absence changes, $F(1,18) = 10.72$, $p < .01$, $\eta_p^2 = .37$ ($M = .64$ and .56 of trials, respectively), and were faster to do so, $F(1,18) = 6.76$, $p = .02$, $\eta_p^2 = .27$ ($M = 2807$ and 3165 ms, respectively). The fixation proportions and times did not differ with whether participants were in the 70% or 90% condition, both $F \leq 2.8$, *ns*, although condition interacted with salience for fixation proportions, $F(1, 18) = 5.54$, $p = .03$, $\eta_p^2 = .24$, due to a condition difference for high but not low salience trials, suggesting presentation time modulated visual search.

Table 1: Changed feature fixations and localization accuracy

		Localized Change	
		No	Yes
Fixated Change	No	28.3%	12.1%
	Yes	11.5%	48.1%

Consistent with Experiment 1, stimulus salience influenced visual attention and change detection. This is noteworthy since the design effectively controlled for visual salience, as participants had even less time to identify high salience feature changes because participants in Experiment 1 took less time to identify those changes. It is also especially noteworthy that participants were able to localize the change on a fair proportion of trials where they had not fixated it (12.1%), which provides tentative support that covert attention may be sufficient for change detection (the change areas, on average, occupied only 3% of the entire image, suggesting a localization-without-fixation rate about four times what might be expected purely by chance). The correlation between change fixation and localization, and the consistency between the two (76.4% contingency), however, suggests that, even if not absolutely necessary, overt visual attention plays a major role in change detection.

General Discussion

The results of two experiments suggest that the visual salience of features, viewers' visual attention, and the ability to detect changes to those features are related. The findings suggest that the salience of features within a scene guide attention (measured with fixation patterns), as well as the cognitive processes required to identify changes (measured with fixation durations). Participants fixated high salience features faster than low salience features, particularly when those features changed through manipulation of presence/absence, most likely because the changing features were not visible half of the time in this condition. Also, participants required shorter fixation durations for high salience feature changes, particularly color changes, perhaps because color changes involved less global feature transformation (i.e., spatial properties are maintained, requiring more overt attention to notice the change). Thus, the data suggest that visual salience guides eye movements, affecting how quickly changed features are fixated, and that overt attention on features is maintained to verify changes.

The systematic relation between fixation patterns and feature salience is proof positive that low-level features contribute to the search for changes in visual scenes. The relations between visual salience and fixation patterns suggest that features were pre-attentively selected, and low-level visual properties guided visual search. The results from Experiment 2, in particular, suggest that changes may be detected before confidence is sufficient for explicit change detection, but that responses are delayed while overt visual attention is held to confirm the change. That is, participants in Experiment 2 were given only a fraction of the time participants in Experiment 1 required to detect the change, and yet they were able to localize and identify the change with reasonable accuracy.

Some previous research has also found that visual salience predicts change detection speed (e.g., Rensink et al., 1997; Simons & Ambinder, 2005); however, visual salience was not precisely defined in any of these accounts. By contrast, Stirk and Underwood (2007) and Wright

(2005) examined whether visual saliency predicts performance on a change detection task using a formal approach similar to that of the current study. They found, however, that visual salience failed to predict change detection response times. These null effects are inconsistent with our findings. There were, however, a number of differences between studies. Most notably, Stirk and Underwood (2007) manipulated the semantic role of the changed object in the scene, revealing that participants were faster to detect changes made to scene incongruent than congruent features. Similarly, Wright (2005) found that, although formally defined visual salience failed to predict response times, independent participants' subjective evaluations of feature salience predicted change detection.

Relatedly, some theorists caution against over-emphasizing the role of salience at the expense of higher-level subjective processes that affect visual guidance (Neider & Zelinsky, 2006; Oliva & Torralba, 2007; Torralba et al., 2006). We do not disagree with this assertion, and indeed our data showed that visual salience was by no means a perfect predictor of eye movements or response times. We assume that the viewers' interpretation of the scenes and the features that changed varied across images, but that these effects were randomly distributed across conditions. As such, in the very least, our results indicate that feature salience plays a role in the guidance of visual attention in change detection. Perhaps more importantly, our analysis of eye movements suggests that low-level visual saliency plays dual roles in change detection tasks, affecting both the rapidity of directing visual attention to the changing features and the amount of overt visual attention necessary before confidence is sufficient for response.

In conclusion, the current study showed that stimulus salience contributes to the detection of changes in visual scenes. Participant's eye movements revealed that feature salience influenced visual attention, which in turn affected change detection. The results of Experiment 2 provide additional support that low-level visual salience guided visual attention and participants' ability to accurately identify changes, even with an impoverished time limit on the presentation of the stimulus. These results thus suggest that change detection typically involves pre-attentive as well as attentional processes that are systematically related to stimulus salience.

Acknowledgments

This research was supported by NSF BCS 0924248.

References

Egeth, H. E. & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology, 48,* 269-297.

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

Hollingworth, A. (2006). Visual memory for natural scenes: Evidence from change detection and natural search. *Visual Cognition, 14,* 781-807.

Itti, L. & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience, 2,* 1-11.

Itti, L., Koch, C., & Niebur, E. (1998). A model of salience-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 1254-1259.

Kahneman, D. & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.) *Perceptual Organization* (pp. 181-211). Hillsdale, NJ: Erlbaum.

Kelley, T. A., Chun, M. M., & Chua, K-P. (2003). Effects of scene inversion on change detection of targets matched for visual salience. *Journal of Vision, 2,* 1-5.

Koch, C. & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4,* 219–227.

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

Oliva, A. & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences, 11,* 520-527.

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

Rensink, R. A. (2000a). The dynamic representation of scenes. *Visual Cognition, 7,* 17-42.

Rensink, R. A. (2000b). Seeing, sensing, and scrutinizing. *Visual Research, 40,* 1469-1487.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8,* 368-373.

Simons, D. J. & Ambinder, M. S. (2005). Change blindness: Theories and consequences. *Current Directions in Psychological Science, 14,* 44-48.

Simons D. J. & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1,* 261-267.

Stirk, J. A. & Underwood, G. (2007). Low-level visual salience does not predict change detection in natural scenes. *Journal of Vision, 7,* 1-10.

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.

Treisman, A. M. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97-136.

Treue, S. (2003). Visual attention: The where, what, how and why of saliency. *Current Opinion in Neurobiology, 13,* 428-432.

Wright, M. J. (2005). Salience predicts change detection in pictures of natural scenes. *Spatial Vision, 18,* 413-430.