Improving Search Task Performance Using Subtle Gaze Direction

Ann McNamara*
Saint Louis University

Reynold Bailey†
Rochester Institute of Technology

Cindy Grimm‡
Washington University in Saint Louis

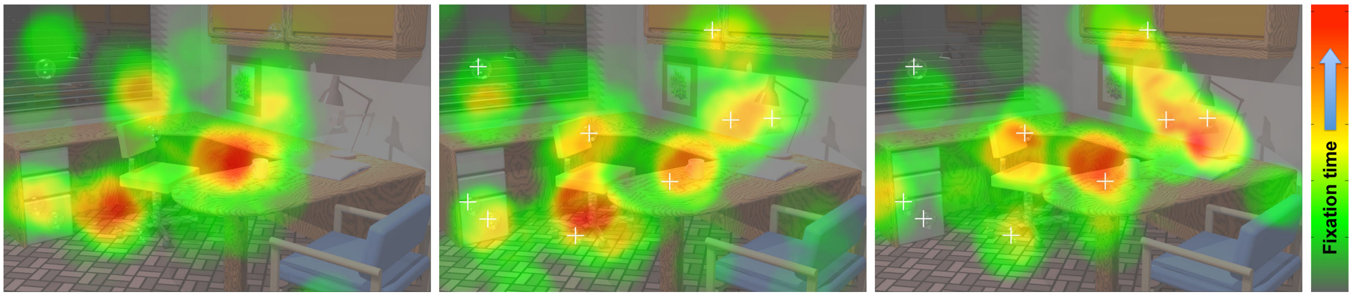

Figure 1: *From left to right: distribution of fixation time under normal viewing conditions, using Subtle Gaze Direction modulations, and obvious modulations*

Abstract

A new experiment is presented which demonstrates the usefulness of an image space modulation technique called Subtle Gaze Direction (SGD) for guiding the user in a simple searching task. SGD uses image space modulations in the luminance channel to guide a viewer's gaze about a scene without interrupting their visual experience. The goal of SGD is to direct a viewer's gaze to certain regions of a scene without introducing noticeable changes in the image. Using a simple searching task we compared performance using no modulation, using subtle modulation and using obvious modulation. Results from the experiments show improved performance when using subtle gaze direction, without affecting the user's perception of the image. Results establish the potential of the method for a wide range of applications including gaming, perceptually based rendering, navigation in virtual environments and medical search tasks.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Display algorithms;

Keywords: eye-tracking, psychophysics, gaze direction, luminance

1 Introduction

Humans routinely perform visual search tasks such as searching for a familiar face in a crowd or scanning a document for some important information. Although such searches are a natural part of our visual processing, there are situations in which the task becomes quite complex and demanding. There are numerous factors which impact the difficulty of a visual search. For example, size of the scene, size of the target, subtlety of the target, contrast, number of objects in the scene, etc. Some types of visual searches may even require specialized training and significant experience in order for the viewer to become proficient. In the medical profession, for example, deciphering x-rays while searching for abnormalities is a demanding search task [Schwaninger et al. 2007] [Schwaninger et al. 2004].

Figure 2: *Example of an image used in this study. The search targets are the transparent spheres "bubbles" in the image.*

One way to improve performance in such tasks is to develop a technique to guide the viewer's gaze toward the regions of a scene that are important for successful completion of the task. To date several researchers have focused on *following* the viewer's gaze pattern to gain efficiencies in rendering and presentation [Luebke et al. 2002] [Duchowski 2002] [O'Sullivan et al. 2003]. However, research is beginning to emerge which looks at *directing* a viewer's gaze about a scene [Kim and Varshney 2006] [Mitchell 2004] [Kosara et al. 2001] [DeCarlo and Santella 2002]. This paper focuses on the simple task of counting targets in an image (see Figure 2). Accurately counting targets efficiently is a necessary task for many applications. For example air traffic controllers need to accurately monitor all aircraft in their vicinity. The gaze directing technique used in this paper, which we call Subtle Gaze Direction,

*e-mail: mcnamaam@slu.edu
†e-mail:rjb@cs.rit.edu
‡e-mail:cmg@cse.wustl.edu

APGV 2008, Los Angeles, California, August 9–10, 2008.
© 2008 ACM 978-1-59593-981-4/08/0008 $5.00

combines real-time analysis of eye movement data with subtle image space modulation to *direct* the viewer's gaze towards selected targets of known location.

One might argue that if the information regarding important regions is available why not simply present that to the user. We agree that in many cases this would be the correct solution, however, cases exist where an algorithm can be beneficial in "suggesting" places to look without disturbing the visual experience of the viewer.

Subtle Gaze Direction depends on the well established fact that the peripheral vision processes stimuli more quickly than the foveal vision [Ogden and Miller 1966]. When viewing a scene for the first time, the low acuity peripheral vision of the Human Visual System (HVS) locates areas of interest. The slower, high-acuity foveal vision is then involuntarily directed to fixate on these regions. By modulating regions of the scene that appear only to the peripheral vision we can force the peripheral vision to locate areas of interest, which are subsequently focused on. This causes the eyes to move involuntarily (saccade) to focus the foveal vision on the modulated region in an attempt to identify the stimuli detected. Luminance modulations were chosen because the HVS is very sensitive to luminance changes [Spillmann 1990].

The modulations are made by alternately blending the pixels in a small area with some amount of black, then some amount of white. The modulation blends from black to white at a rate of 10 Hz. We use a Gaussian falloff function with a radius of 32 pixels, which in our setup corresponds to approximately a $1cm$ diameter circular region on screen.

A small pilot study was conducted to determine the amount of black/white for which the modulations are just intense enough to be detected by the peripheral vision. Three participants were involved in this pilot study. They were each presented with five randomly selected images from the complete test set. They were instructed to fixate on a cross in the center of the image while modulations were presented in random peripheral regions. Using the keyboard (+/-), they adjusted the black and values in step sizes of 0.005 until the modulations were just noticeable. The final values (0.095 percent maximum of black/white) were obtained by averaging the results of the three participants.

Additionally, the viewer's foveal vision is never allowed to fixate on the modulated region. This is achieved by monitoring the direction component of the saccade velocity vector, to determine if the foveal vision is about to enter the modulated region. If this is the case, the modulation is immediately terminated.

Figure 1 illustrates the results of introducing modulations into an image on the viewer's gaze pattern. The Figure shows a heat map of the average scan patterns over the image for 6 observers. The image on the left shows the scan pattern resulting from normal viewing. The middle and right images were altered by adding modulations at the target regions indicated by white crosses. In the middle image we applied the subtle modulations, in the right image larger modulations were used.

This paper presents a psychophysical experiment that explores the impact of Subtle Gaze Direction on performance during a visual search task. The results show that this method works well without introducing noticeable artifacts into the image.

2 Experiment

Twenty-four images served as stimuli for the experiment (see Figure 3). Six environments were chosen and populated with four different target counts, ranging from 4 targets to 12 targets for a total of 24 images. The targets were small transparent spheres roughly uniformly distributed within the scene. Some spheres were deliberately placed so as to be difficult to resolve. The reason for this was to allow us to investigate if those hard to see targets were more easily resolved using modulation. All of the models used to create the images were taken from the RADIANCE web site [Larson and Shakespeare 2004]. Presentation order was randomized to eliminate any learning effects. Images were presented for 14 seconds. A black screen with a white cross at the center was presented between each image to allow the participant to refocus on the center of the screen. This also means the initial viewing position for each image is the same i.e. viewing begins in the center of the images. An example of an image used in this study is shown in Figure 2. Image sizes varied as shown in Figure 3. In cases where image size was smaller than the viewing screen the image was centered on a black background.

Figure 3: *Scenes used in the experiments. All images were 693 by 1024 pixels except the bathroom scene and the interior scene which were 691 x 1024 and 797 x 1024 respectively.*

Participants were seated in front of the computer screen in a well lit room with their chin comfortably resting on a chin-rest to reduce head movement. Using an infrared camera-based eye-tracking system[1], data pertaining to fixation position and saccades were recorded for the dominant eye of each participant. A fixation is defined as any pause in gaze $\geq 150ms$. Participants were instructed to remain as still as possible while the eye-tracker was calibrated and the experiment was conducted. The chin-rest was positioned $75cm$ from the screen. At this distance, the actual perceptual span (area of high acuity) of the observer occupies a circular region of diameter $5cm$ on the screen [Rayner 1975]. The subtle modulations were presented in a smaller ($1cm$ diameter) circular region.

Eye-tracking was employed to record the viewer's fixation and saccades while counting targets in the various images. Eye-tracking information also served as input to trigger the modulations on targets that were not attended to, in an effort to highlight them so the user could identify them and include them in their count. Image complexity varied as did the number of targets. The behavior of the targets was also varied as follows:

[1] ViewPoint EyeTracker® by Arrington Research, Inc.

- **GROUP 1, NO MODULATION:** no behavioral actions applied to the targets, so images were viewed normally with no modulations.
- **GROUP 2, SUBTLE MODULATION:** subtle image modulations were used to highlight the target regions in an effort to aid in counting. Modulation was never applied to targets while they were being directly viewed. Any modulation was applied in the periphery only, and modulations were terminated as the user moved their gaze toward the modulated region. Thus the viewer was never allowed to directly view the modulated region. A modulation radius of 0.04 degrees of visual angle was defined to ensure subtlety.
- **GROUP 3, OBVIOUS MODULATION:** subtle behavior was exaggerated so that the modulations were clearly visible by increasing the size of the modulation. Modulation was similar to the modulation applied in Group 2, however, in this condition the modulations were deliberately set to be more obvious. A modulation radius of 0.125 degrees of visual angle was defined to ensure visibility.

The targets subtended visual angles ranging from 0.05 to 0.08 degrees, depending on their location in the scene. Therefore subtle modulations subtended ≤ 0.5 the size of the targets, while the obvious modulations subtended a visual angle of between 1.5 to 2 times the size of the target.

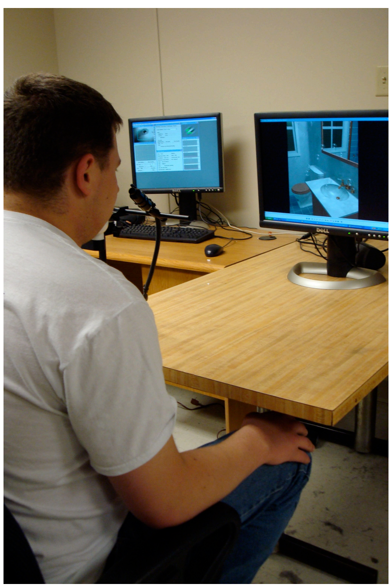

Figure 4: *Experimental Set Up*

Eighteen participants were assigned randomly to one of the three groups. Participants volunteered from a group of undergraduates. All had normal or correct-to-normal vision and were naive to the purpose of the experiment. Participants viewed the images on a 22" LCD Screen at a resolution of 1200 X 1600 from a distance of 75cm. Head position was held constant using a chin rest for support, as shown in Figure 4. The eye movements of each participant were recorded along with a count of the targets found and the time to respond for each image. An informal exit interview questioned the participants about the quality of the images to determine if the modulations in conditions 2 and 3 were disturbing to the viewer. Participants in condition 2 reported nothing unusual, whereas in condition 3 participants reported seeing the modulations.

The task involved viewing each scene and counting the number of targets present. Participants verbally reported the number of targets counted on completing the task.

3 Results and Discussion

IMAGE	NONE	SUBTLE	OBVIOUS
Image A: Soda Hall	25.0%	54.2%	66.0%
Image B: Conference	79.2%	79.2%	66.7%
Image C: Interior	12.5%	29.2%	58.3%
Image D: Office	20.8%	62.5%	54.2%
Image E: Bathroom	37.5%	50.0%	29.2%
Image F: Counter	70.8%	62.5%	66.7%
AVERAGE	40.97%	56.25%	56.94%

Table 1: *This table shows the percentage of accurate detection of all the targets in an image. 100% means all of the targets were found. Each column is the average over 4 cases. Standard Deviations were of the order of 2%.*

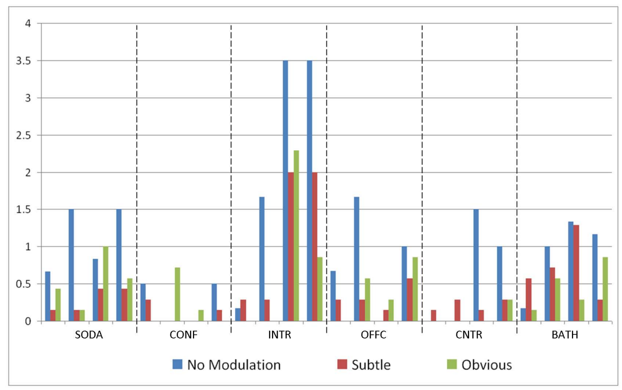

Figure 6: *Experimental Results: This chart shows the sum of differences between the actual number of targets and the number of targets reported for each condition.*

The number of targets reported and the time to respond was recorded for each image. The average response times were consistent across all three conditions, 6.272, 6.495 and 6.570 seconds (with standard deviations of 0.55, 0.94 and 0.96) for the no modulation, subtle modulation and obvious modulation conditions respectively.

We compared the reported number of targets to the actual number of targets and used this to define a correlation. These correlations are graphed in Figure 5. The correlation values represent how close the reported number of targets were to the actual number of targets. Higher correlation corresponds to more accurate task performance. Correlation is higher for the modulated conditions than in the static image with values of 0.80, 0.89 and 0.90 for groups 1, 2 and 3 respectively. This indicates that participants did slightly better on task with the aid of modulation as opposed to normal viewing.

Another interpretation of the results is to simply compare the number of targets missed during counting in each case. This data is shown in Figure 6. Each bar represents the absolute difference between the actual number of targets and the number reported. Smaller bars indicate more accurate counts, no bar indicates 100% precision. The blue bars show the differences in the normal no modulation viewing condition, while the orange and green bars show the modulated images, subtle and obvious respectively. The four bars represent the number of targets for each image (the number of targets ranged from four to 12). Each cluster is one image, as labeled

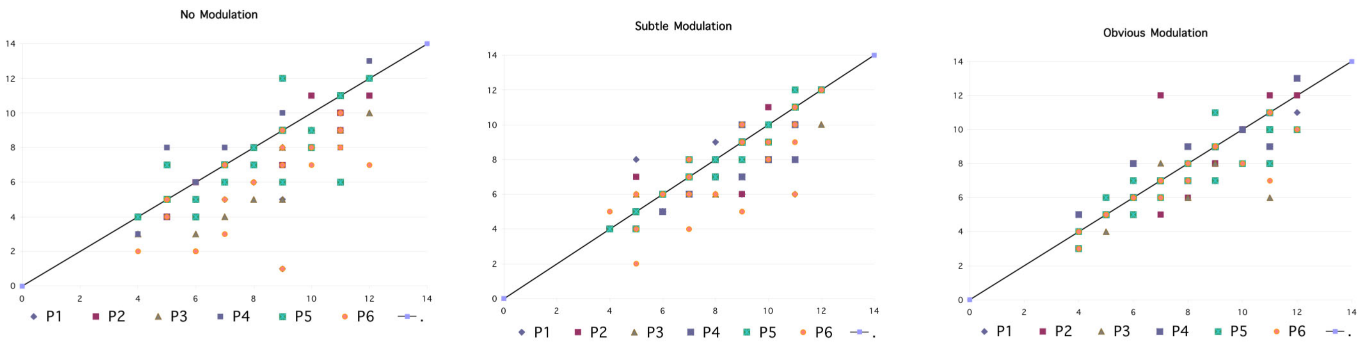

Figure 5: *The graphs show the correlations between the actual number of targets present in each image and the number of targets reported by the observers for groups 1, 2 and 3 respectively.*

on the x-axis. As the data shows, in most images, the modulation aids in the accuracy of the results. In some cases the number of targets reported does not increase as a function of the number of targets, one reason for this may be the participant's failure to see the targets, and subsequent failure to include them in their count.

The percentage of correct counts reported also reveals that a higher percentage of counts returned in the modulated imagery were accurate compared to the imagery with no modulation. What is interesting in this analysis is that the percentage correct in both modulated cases is higher than in the static imagery. However, it is important to note that while no obvious distractions were noted by participants viewing the subtle modulation case, all participants in the obvious modulation condition reported seeing the modulations. They noted that while the modulations did distract their gaze, it also helped them to identify targets they may not have otherwise counted. This indicates that simple region highlighting, even if noticeable, can contribute to improving task accuracy. The data suggests that subtle gaze direction, where the highlighting is sufficiently faint so as to go unnoticed, successfully guides the viewer's gaze to the target regions, thereby improving task performance. Data is tabulated in Table 1.

The highest discrepancies occurred in the image of the interior scene (Image C). Here the composition of the scene may have influenced the visibility of the bubbles, with only one person getting 100% accuracy. The placement of the targets was also made deliberately difficult. One reason for the poor results in the bathroom scene may be due to the fact that participants reported being confused regarding the inclusion or exclusion of targets reflected in the mirrors.

Further statistical analysis of the data was conducted. A between-subjects ANOVA over the correlations resulted in $F(23; 15) = 64.04; p <= 0.001$. This gives evidence that a significant effect of condition is present between the tasks i.e. performance differed in each group. This can also be seen in Figure 1 which compares distribution of fixation time across images.

In summary the results show slight improvement in task performance when modulation is employed to direct gaze to target regions. This seems to hold true whether or not the viewer notices the modulations. Some applications may elect to include obvious modulations whereas in other applications subtlety may play a key role.

4 Conclusions and Future Work

We presented an experiment to compare task performance in digital images across three sets of stimuli. In summary the results indicate that using either subtle or obvious image modulations on the target regions improves the precision of a simple counting task. The difference between using subtle and obvious modulations is the level of disruption to image viewing. With subtle image modulation none of the participants reported noticing the modulations, whereas with the obvious image modulation all participants reported seeing the modulations.

In this experiment we only modulated the luminance channel. It may be that the modulations would be more successful in certain conditions if we used other channels. We have experimented with the warm-cool channel, but found luminance to be slightly more efficient [Bailey et al. 2008]. It may be that the most effective modulation should be a function of the image itself, and may be dynamic depending on the behavior in the scene.

The results from this initial study are promising and several follow up experiments are imminent. There are several avenues open for future investigation. In this experiment participants were asked to identify targets and there were no other distracters in the image. Often in visual search the task involves discriminating targets from non-targets (enemy versus friendly for example). One possible line of inquiry would be to examine the usefulness of Subtle Gaze Direction in imagery where the task involves identifying or separating targets from non-targets. By applying Subtle Gaze Direction in target regions gaze could be directed only to the targets, making them more distinguishable from non-target regions. Another interesting problem is that of moving targets, or moving imagery generally. Future experimentation will focus on the performance of subtle gaze direction in dynamic environments, such as animations and interactive environments. Subtle Gaze Direction could be used to help guide a user's navigation, or to highlight those parts of an animation that are more relevant to the application. We would expect that stronger modulations would be necessary in order for Subtle Gaze Direction to be effective in dynamic scenes.

We have shown that Subtle Gaze Direction can improve people's performance on a counting task without noticeably changing the image. Example uses of this technique might be to help guide gaze in more complex visual search tasks where targets are numerous or difficult to identify.

References

BAILEY, R., MCNAMARA, A., SUDARSANAM, N., AND GRIMM, C. 2008. Subtle gaze direction. *ACM Transactions on Graphics*. To appear.

DECARLO, D., AND SANTELLA, A. 2002. Stylization and abstraction of photographs. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, ACM Press, New York, NY, USA, 769–776.

DUCHOWSKI, A. T. 2002. A breadth-first survey of eye-tracking applications. *Behav Res Methods Instrum Comput 34*, 4 (November), 455–470.

KIM, Y., AND VARSHNEY, A. 2006. Saliency-guided enhancement for volume visualization. *IEEE Trans. on Visualization and Computer Graphics (IEEE Visualization 2006) 12, No. 5*, 925–932.

KOSARA, R., MIKSCH, S., AND HAUSER, H. 2001. Semantic depth of field. In *IEEE Symposium on Information Visualization 2001 (InfoVis 2001)*.

LARSON, G. W., AND SHAKESPEARE, R. 2004. *Rendering With Radiance: The Art And Science Of Lighting Visualization*. Booksurge Llc.

LUEBKE, D., WATSON, B., COHEN, J. D., REDDY, M., AND VARSHNEY, A. 2002. *Level of Detail for 3D Graphics*. Elsevier Science Inc., New York, NY, USA.

MITCHELL, G. E. 2004. *Taking Control Over Depth of Field: Using the Lens Blur Filter in Adobe Photoshop CS*. The Light's Right Studio.

OGDEN, T. E., AND MILLER, R. F. 1966. Studies of the optic nerve of the rhesus monkey: Nerve fiber spectrum and physiological properties. *Vision Research 6*, 485–506.

O'SULLIVAN, C., DINGLIANA, J., AND HOWLETT, S. 2003. Eyemovements and interactive graphics. *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movement Research*, 555–571. J. Hyona, R. Radach, and H. Deubel (Eds.).

RAYNER, K. 1975. The perceptual span and peripheral cues in reading. *Cognitive Psychology 7*, 65–81.

SCHWANINGER, A., HARDMEIER, D., AND HOFER, F. 2004. Measuring visual abilities and visual knowledge of aviation security screeners. *Proceedings IEEE ICCST 38*, 258–264.

SCHWANINGER, A., MICHEL, S., AND BOLFING, A. 2007. A statistical approach for image difficulty estimation in x-ray screening using image measurements. In *APGV '07: Proceedings of the 4th symposium on Applied perception in graphics and visualization*, ACM, New York, NY, USA, 123–130.

SPILLMANN, L. 1990. *Visual Perception: The Neurophysiological Foundations*. Academic Press, San Diego.