

Visual search in noise: Revealing the influence of structural cues by gaze-contingent classification image analysis

Umesh Rajashekar

Department of Electrical and Computer Engineering and
Center for Perceptual Systems, The University of Texas at
Austin, Austin, TX, USA



Alan C. Bovik

Department of Electrical and Computer Engineering and
Center for Perceptual Systems, The University of Texas at
Austin, Austin, TX, USA



Lawrence K. Cormack

Department of Psychology and Center for Perceptual
Systems, The University of Texas at Austin, Austin, TX, USA



Visual search experiments have usually involved the detection of a salient target in the presence of distracters against a blank background. In such high signal-to-noise scenarios, observers have been shown to use visual cues such as color, size, and shape of the target to program their saccades during visual search. The degree to which these features affect search performance is usually measured using reaction times and detection accuracy. We asked whether human observers are able to use target features to succeed in visual search tasks in stimuli with very low signal-to-noise ratios. Using the classification image analysis technique, we investigated whether observers used structural cues to direct their fixations as they searched for simple geometric targets embedded at very low signal-to-noise ratios in noise stimuli that had the spectral characteristics of natural images. By analyzing properties of the noise stimulus at observers' fixations, we were able to reveal idiosyncratic, target-dependent features used by observers in our visual search task. We demonstrate that even in very noisy displays, observers do not search randomly, but in many cases they deploy their fixations to regions in the stimulus that resemble some aspect of the target in their local image features.

Keywords: classification images, visual search, eye movements, $1/f$ noise

Introduction

Visual search is a common yet important task that plays a significant role in the survival of a species. Tasks such as locating prey or ripe fruit are some examples of real-world visual search. Despite the seemingly complex mechanisms underlying search, humans excel at search tasks. One particular aspect of the human visual system that is critical to its success as an efficient searcher is its active nature of looking. The human visual system uses variable resolution sampling to capture the image on the retina. The resolution is highest in the central (foveal) region and rapidly falls toward the periphery. This provides both high acuity and a large field of view without the accompanying data glut.

For a foveated visual system to be effective, however, it must be able to deploy a suite of eye movements to scan the visual field. To build a detailed representation of the image, the human visual system uses a dynamic process of actively scanning the visual environment using discrete fixations linked by saccadic eye movements (Yarbus, 1967). The eye

gathers most information during the fixations whereas little information is acquired during the saccades due to saccadic suppression and motion blurring (e.g., Burr, Morrone, & Ross, 1994). During each fixation, humans presumably analyze the scene with the high-resolution fovea and use the low-resolution peripheral information and memory to plan subsequent fixation and search for specific targets. An understanding of how the human visual system selects and sequences image regions for scrutiny is not only important to better understand biological vision but is also the fundamental component of any foveated, active, artificial vision system. The instantiation of automatic search models into the next generation of efficient, foveated, active vision systems (Klarquist & Bovik, 1998) could be potentially applied to a diverse array of problems including automated pictorial database query and data mining; image understanding; autonomous vehicle navigation; real-time, foveated video compression (Lee, 2000); and automated visual search in, for example, cancer detection.

Previous visual search experiments that involved the detection of targets in the presence of distracters support

the hypothesis that saccadic programming during a visual search task is not random but can be influenced and guided by many visual cues (Wolfe, 1998, Wolfe & Horowitz, 2004), such as color (Williams, 1967) and shape (Findlay, 1997; Murray, Beutter, Eckstein, & Stone, 2003), in the periphery and hence affect search performance. For example, by measuring the accuracy of the first saccade in a 10-AFC letter discrimination task, Murray et al. (2003) demonstrate that saccadic targeting is not random but uses shape cues in the search task. The degree to which these features, which are generally salient and against a blank background, affect multiple fixation search strategies is usually measured using reaction times and detection accuracy. Reaction times and accuracy results, however, do not reveal what image features or structural cues are attracting the observer's gaze.

Recently, the theory of classification images (Eckstein & Ahumada, 2002; Simoncelli, 2003)—a noise-based reverse correlation technique—has been extensively used in psychophysics to reveal auditory (Ahumada & Lovell, 1971), spatial (Beard & Ahumada, 1998), and spatiotemporal (Neri & Heeger, 2002) mechanisms used by human observers in detection tasks, discrimination tasks, or both. The basic idea underlying the classification image paradigm is that external noise added to a discrimination task reduces the signal-to-noise ratio and therefore influences the subject's response to the stimulus. Assuming that subjects use linear time-invariant strategies, this paradigm analyzes correlations across the noise samples to reveal strategies used by the subjects to perform the classification task. In this paper, we use the classification image paradigm as the image analysis routine to investigate search strategies.

Eye movements are necessary for any foveated visual system and are a salient feature of (natural) visual search. Naturally, there has been considerable interest in investigating image features that influence human eye movements in natural viewing tasks (Parkhurst & Niebur, 2003; Reinagel & Zador, 1999). The result of analysis at the point of gaze reveals that eye movements are not really random (despite the staggering number of fixations made in the duration of the task) but are possibly drawn to image patches with high contrast and low interpixel correlation. An investigation into these point-of-gaze statistics in a visual search task has the potential to reveal strategies (Rajashekar, Cormack, & Bovik, 2002) employed by observers and motivates the approach used in this paper.

The human visual system has probably evolved multiple mechanisms for controlling gaze. These mechanisms differ in the amount of image processing and interpretation they require, and the relative importance of each of them is probably situation dependent. Because mechanisms that require relatively little image interpretation are likely to be most relevant for current work in artificial vision, our goal is to develop an image-based theory of human eye movements to isolate and understand the data-driven



Figure 1. Targets used in the search task.

mechanisms that guide eye fixations. In this paper, using a combination of eye tracking and classification image analysis, we extracted low-level image features used by observers in a search task.

We recorded the eye movements of three observers as they searched for simple geometric targets (shown in Figure 1) embedded in a noise stimulus. The noise had a Fourier magnitude that was inversely proportional to the spatial frequency. An example of the stimulus with an observer's eye movements superimposed on it is shown in Figure 2. A quick examination of the stimulus might help the reader to appreciate the difficulty of the search task. Further, the pattern of fixations in Figure 2—which was fairly typical—may even suggest that in a stimulus with such low signal-to-noise ratios, observers might be simply deploying their fixations uniformly across the stimulus, without respect to any spatial structure, until the fovea fortuitously landed close enough to the target for detection to occur. However, as we will demonstrate, a simple reverse correlation analysis of the noise stimuli at the point of gaze reveals quite a different strategy.

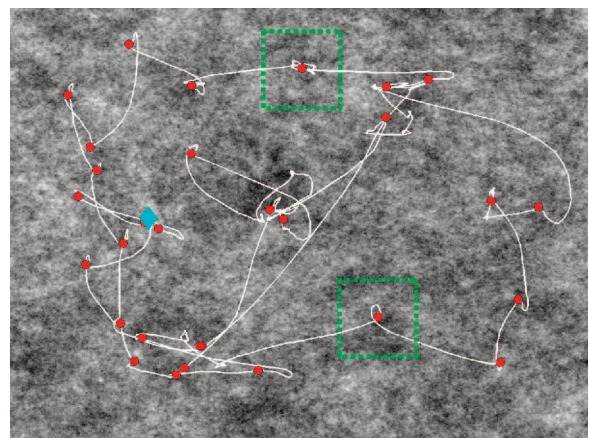


Figure 2. An example of the stimulus with superimposed eye movements from a single trial. The white line indicates the gaze position recorded from the eye tracker. The red dots indicate the computed location of fixations. The cyan diamond shows the location of the target (in this case, the triangle). The dashed green rectangles show the size of the noise patches extracted around each fixation locus for analysis.

Methods

Observers

Three observers were used for the experiments. Two of these observers (L.K.C. and U.R.) were authors, and the third (E.M.) was naive as to the purpose of the experiment. All observers either had normal or corrected-to-normal vision.

Stimuli and tasks

The stimuli consisted of simple geometric targets embedded in noise. The noise had a Fourier magnitude that was inversely proportional to the spatial frequency. This type of noise, generally referred to as $1/f$ noise, mimics the average spectrum of natural images (Field, 1987). The presence of more salient large-scale structures in this type of noise (as compared with white noise) makes $1/f$ noise a desirable stimulus in these types of experiments. A set of nine noise images was generated offline and used throughout the experiment. The targets were simple shapes, viz., circles, dipoles, triangles, and bow ties (abutting triangles), as shown in Figure 1. The signal-to-noise ratio of the stimuli was set so that the observers had to make approximately 20 fixations to find the target (this corresponded to signal a Michelson contrast of 10% and a noise RMS contrast of 23%). The target and the noise stimuli had approximately the same mean luminance such that local luminance by itself was not a cue to the target location. The stimulus size was 640×480 pixels and the size of the embedded target was 64×64 pixels (1.2×1.2 deg). The stimuli were displayed on a 21-in., gamma-corrected monitor at a distance of 180 cm from the observer. The screen resolution was set at 640×480 pixels (12×9 deg), corresponding to about 52 pixels/deg of visual angle. The MATLAB psychophysics toolbox (Brainard, 1997; Pelli, 1997) was used for stimulus presentation.

The observer's task during each trial was to locate the target as quickly as possible. At the beginning of each block of 50 trials, observers were told which target would be presented. Each observer was presented with 250 trials per target. Very rarely, when the eye tracker lost track of the observer's gaze for the entire period of a trial, the trial was discarded. On each trial, one of the nine noise images was selected, and the target was added to the noise at a random spatial location (with the constraint that the target did not overlap the edge of the display). Every trial had a hidden target; there were no catch trials in which the target was omitted. However, because the size of the target was small compared with that of the stimulus, one could consider all stimulus regions not containing the target to resemble catch trials in the traditional psychophysics experiments. On finding the target (or what appeared to be a target),

the observer pressed a button. Following the button press, the observer was given feedback about the correct target location and then proceeded to the next trial. Because the signal-to-noise ratio was very low, on average, observers were able to locate the hidden target only in 50% of the trials. A trial was considered a hit if the observer's last fixation landed anywhere in the 64×64 pixel area of the target. An example of the stimulus is shown in Figure 2, with the location of the target (a triangle in this case) indicated by the cyan diamond. Also shown are an observer's recorded sequence of eye movements and computed fixations, the details of which will be discussed below.

Eye tracking

Human eye movements were recorded using an SRI Generation V Dual Purkinje eye tracker. It has an accuracy of <10 arcmin, a precision of ~ 1 arcmin, a response time of under 1 ms, and a bandwidth of DC to >400 Hz. The output of the eye tracker (horizontal and vertical eye position signals) was sampled at 400 Hz by a National Instruments data acquisition board in a Pentium IV host computer, where the data were stored for offline data analysis. The effective spatial resolution of the data acquisition board (corresponding to the 16-bit resolution of the input channels) was over an order of magnitude higher than the precision of the eye tracker, guaranteeing that no information was lost in the sampling process.

A bite bar and a forehead rest were used to restrict the subject's head movement. The subject was first positioned in the eye tracker and a system lock established onto the subject's eye. A linear interpolation on a 3×3 calibration grid was then done to establish the transformation between the output voltages of the eye tracker and the position of the subject's gaze on the computer display. During the experiment, periodic verifications (every 10 trials) of the calibration were done by displaying a dot on the display at the computed position of gaze in real time and, if necessary, recalibration was done (although this was rarely required).

Image data acquisition

The sampled voltages from each trial were converted to gaze coordinates (i.e., position of gaze on the image in pixels). Next, the path of the subject's gaze was divided into fixations and the intervening saccadic eye movements using spatiotemporal criteria derived from the known dynamic properties of human saccadic eye movements. A sequence of eye position recordings was considered to constitute a fixation if the recorded gaze coordinates remained within a stimulus diameter of 0.5 deg visual angle for at least 100 ms. If in the 50-ms window following a fixation the sequence of gaze recordings was found to

have occurred at a distance greater than 1 deg from the previous fixation point, a saccade was assumed to have been executed and the previous fixation was terminated. The exact algorithm (adapted from Applied Science Laboratories, 1998) accommodated for drifts, blinks, and microsaccadic eye movements.

The resulting pattern of fixations for a single trial is shown by the red dots in Figure 2. The white lines show the eye movement trajectories linking the fixations. We defined a region of interest (ROI) as a square region of size 128×128 pixels (twice the size of the target) around each fixation (two examples of which are shown by the dashed green boxes in Figure 2). The noise in the ensemble of these ROIs around the fixation points was then analyzed to determine if it contained any statistical relationships to the targets for which the subject was searching.

Results

We extended the classification image technique by incorporating eye movements as follows. For a given observer and target, noise pixels in an ROI (twice the size of the target) around each fixation were averaged across trials to yield classification images. Every fixation made by the observer, including the final fixation and those that landed on the target, was used in the analysis. A vast majority of these fixations, however, did not land on the target. A resulting classification image for one observer (L.K.C.) searching for the dipole target is shown in Figure 3A. In this image (and all further classification images), gray denotes mean luminance, white corresponds to pixel values above the mean luminance, and black corresponds to pixel values below the mean luminance. To enhance the details in the classification images, a statistical thresholding was performed by setting pixel intensities within one standard deviation of the mean equal to the mean (Beard & Ahumada, 1998). The statistically thresholded classification image corresponding to the classification image in Figure 3A is shown in Figure 3B. The statistically thresholded results for all observers and targets are shown in Figure 4. The first row in Figure 4 illustrates the

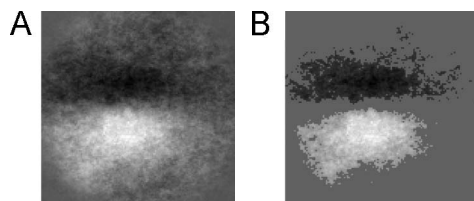


Figure 3. Statistical thresholding of classification images. (A) Original classification image; (B) classification image after statistical thresholding.

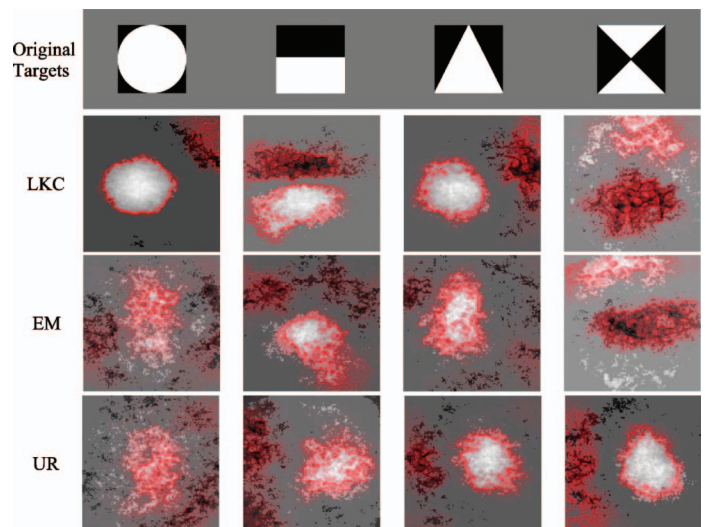


Figure 4. Classification images at point of gaze: Statistically filtered classification images for three observers (L.K.C., U.R., and E.M.) and the four targets (shown in the top row against a neutral gray background to highlight the edges of the target). The red outlines indicate the aggregate of the edges of the classification images from 200 bootstrap replications of the noise averaging. A well-defined boundary indicates that the structure in the classification image is robust, whereas spatially diffused edges indicate that the shape of the classification image in that region varies across trials.

four targets (circle, dipole, triangle, and bow tie) that observers were instructed to find in the stimulus. Each of the other rows shows the classification images for the three observers (L.K.C., U.R., and E.M.) for these targets. Each subject made an average of 3,000 fixations per target. Thus, each classification image is the result of averaging around 3,000 noise patches.

To quantify the uncertainty associated with the shapes of these classification images, we bootstrapped (Efron, 1994) the averaging procedure and computed the boundaries of each of the resulting thresholded classification images as follows. To detect the boundaries of the statistically filtered classification images, the two largest regions in the statistically filtered image were detected using a connected-components algorithm. The outlines of each of these regions were then used to represent the boundary of the classification image. Bootstrapping was then used to verify the robustness of these boundaries. First, the ensemble of noise patches at the observer's point of gaze was resampled (with replacement) 200 times. The image patches in each bootstrap sample (3,000 patches, on average) were then averaged together, statistically filtered, and processed to detect the boundaries. The boundaries of the resulting 200 classification images were then added together and superimposed on the classification images to reflect the stability of the boundary. The aggregate of all of the bootstrapped boundaries is shown superimposed in red in Figure 4.

Discussion

The most striking result of this analysis is the emergence of classification images (in Figure 4) that resemble spatial features of the target. Notice that the boundaries of most classification images from the bootstrap procedure are well defined, indicating that the shape information for these classification images is indeed reliable. (In fact, the full width of a bootstrapped contour at a given location gives an upper bound on the width of the 99.5% confidence interval about that location). This indicates that although gaze is being rapidly shifted about the image in an effort to find the target as quickly as possible, saccadic programming is being clearly influenced by spatial features in the noise with sufficient precision to generate these robust classification images.

Also worthy of note is the fact that these classification images vary in a target-dependent manner within an observer (to varying degrees), and they also vary across observers for a given target. The data of L.K.C. and E.R., for example, show fairly dramatic changes across the targets, indicating that the gaze of these observers was attracted by features unique to a particular target. For example, this effect is most pronounced for observer L.K.C. who, for the case of the circle, seemed to be fixating at regions of high luminance but modifies the search template for the dipole search by fixating at regions that have a luminance profile that matches the horizontal edge of the dipole. This adaptation of the classification image to match some feature of the target is evident, albeit to a lesser extent, for observer E.M. who changes his strategy for the dipole, the triangle, and the bow tie. Observer U.R., in contrast, seemed to use a simpler heuristic, consisting of fixating bright regions of roughly the appropriate size.

The interobserver variability is evident for most of the shapes. Even in the case of the triangle, in which the classification images look fairly similar, closer inspection reveals some subtle but reliable differences. Observer L.K.C.'s gaze tends to land on the right side of bright regions that have a dark region further to the right, whereas observer U.R.'s gaze lands on the left side of bright regions that have dark regions further to the left. Both

of these observers, however, are attracted to roughly circular bright areas on average, whereas observer E.M. clearly favors a more angular, elongated structure.

We also simulated an observer who randomly looks about the stimulus hoping to find the target by chance (this behavior was largely consistent with our own introspection of behavior in this very difficult task; observers reported often being surprised when their gaze landed on the target). We did this by simply selecting random spatial coordinates from a uniform distribution with the constraint that the entire ROI surrounding the fixation point had to be within the image. The number of random fixations used was approximately equal to that comprising the samples from the human observers (around 3,000). The result of adding the noise pixels at the random fixation points is shown in the far right in Figure 5, where all the classification images in the figure are displayed using the same range of gray scales to highlight the relative pixel magnitudes across classification images. The lack of any image structure in the random sampling case and the ability to generate many classification images across subjects from the same set of $1/f$ noise stimuli indicates that observers are not random searchers but are actually directing their fixations to regions that resemble some feature of the target.

To verify that observers were indeed trying to follow our instructions to find the target quickly, we analyzed the histograms of the fixation duration (time per fixation) and saccadic magnitudes for observers' eye movements for this experiment. A relatively short fixation dwell time (mean = 0.25 s) and a wide range of saccadic magnitudes (mean = 3 deg, $SD = 1.8$ deg) confirmed that our observers were indeed trying to survey the range of the stimulus area quickly and were not fixating on particular regions for an inordinately long time. These statistics reflect previously reported measures of fixation durations and saccade magnitudes in natural viewing tasks (Duchowski, 2002; Rayner, 1998; Yarbus, 1967).

Effects of $1/f$ noise on the analysis

Traditionally, experiments using the classification image analysis have used additive white Gaussian noise as

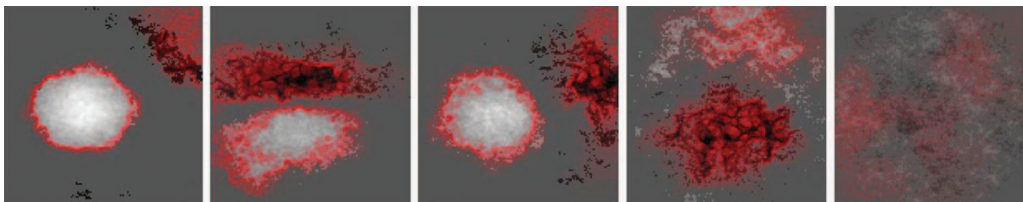


Figure 5. Human versus random search. The first four classification images (from left to right) show the classification images for observer L.K.C. across the four targets. To simulate a searcher who looks about randomly hoping to find the target by chance, the noise stimuli were sampled randomly. The result of averaging the noise pixels at these random fixation points is shown in the far right. The lack of significant structure in the classification image for the random searcher sharply contrasts with the obvious target-like structures generated by the observer. All the classification images are displayed using the same range of gray scales to highlight the relative pixel magnitudes across classification images.

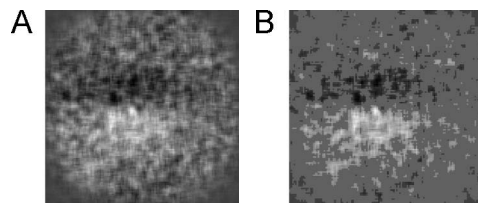


Figure 6. Effect of prewhitening filter on classification image. (A) Original classification image (processed with average filter) and (B) classification image after statistical thresholding.

the masking stimulus. White noise images, being spatially uncorrelated by design, do not contribute to any artificial structure to the classification images. In our experiments, we used $1/f$ noise as the stimulus because the presence of many large-scale, target-like salient features inherent in the noise structure made it an effective masker. Due to the correlated nature of $1/f$ noise, the resulting classification images are not unbiased linear templates. To obtain the true unbiased linear estimator, we can apply a prewhitening filter (Abbey & Eckstein, 2000) to the classification images obtained using $1/f$ noise. An example of the dipole classification image for observer L.K.C. before and after prewhitening is shown in Figure 6. However, even the classification images obtained using the prewhitening filter does not reflect the true unbiased template in our experiments. This is because, unlike the experimental situation in psychophysics, the contribution of each pixel to the creation of the classification image is shift variant across trials (fixation points in this case) due to several factors. First, oculomotor precision and measurement errors inherent in eye movements and their recording result in spatial uncertainty in the exact location of an observer's fixation. Second, even if we ignore errors in recording fixations and simply assume that an observer was using only a single visual feature to succeed in the search task, there is no guarantee that the observer will precisely fixate the same location on this feature every time. For example, assume that the observer always looked for black triangles in the case of the "bow-tie" search. During search, the observer could decide to fixate the left triangle in some trials and the right triangle on others. Thus, the noise samples extracted around these fixations are not necessarily perfectly aligned across trials, resulting in spatially blurred classification images. However, this does not imply that observers are unable to use precise shape information. In a related study (Beutter, Eckstein, & Stone, 2004), subjects performed an 8-AFC contrast discrimination task and classification images were generated using a saccade-contingent analysis and an 8-AFC perceptual decision framework. The resulting classification images for these two cases were not found to be significantly different from each other, indicating that saccade mechanisms can indeed use precise shape information to guide search.

Conclusions

Analysis of stimuli at the observer's point of gaze can provide an understanding of strategies used by observers in visual tasks. Although existing proof for the guided saccade-targeting hypothesis is supported using efficiency of task performance or reaction times, in this paper we demonstrated that classification image analysis and accurate eye tracking can be used in conjunction to reveal shape cues that guided saccades in a difficult visual search task. Our results indicate that even in very noisy stimuli, human observers are not random in their search strategy; instead, they make directed eye movements to regions in the stimuli that resemble some structural feature of the target. Given the difficulty of the task and the rapidity at which fixations are made, we find it remarkable that the visual system seems to analyze spatial structure in the low-resolution periphery and direct fixations to regions that resembled some spatial feature of the target.

The goal of this study was to investigate the influence of structural cues in visual search. However, reverse correlation analysis of noise stimuli at the point of gaze of observers has also been used to provide valuable information on the time course of how visual information is gathered in the course of visual search (Caspi, Beutter, & Eckstein, 2004). Using a 5-AFC search for a bright Gaussian target among four dim distracters in the presence of dynamic noise, it was shown that visual information gathered before the initiation of a saccade is used to plan future saccades even if the information cannot be used for the current saccade. Recently, it has been shown that human observers nearly achieved optimal search performance in a similar search task (Najemnik & Geisler, 2005). Yet, truly, optimal performance is computationally expensive. The classification images obtained in our experiments suggest that subjects do not necessarily use all the target features to succeed in the search task, but often they only use a subset of the target features. Moreover, the differences we find across observers, even when the targets are simple geometric shapes, indicate that observers adopt idiosyncratic heuristics even if, in the end, they all approach optimal behavior.

The analysis procedure presented in this paper is not without its shortcomings. The classification images resulting from this analysis lack in the spatial detail that usually results from psychophysical trials, where the contribution of each pixel to the creation of the template is shift invariant across the trials. With the added dimension of eye movements, the noise pixels across trials are no longer guaranteed to be perfectly aligned, and it is therefore very likely that many image features are blurred or even lost in the averaging process. This issue of spatial uncertainty introduced by eye movements was recently addressed by constraining the locations of both the target and fixations to 1 of 49 square regions on a rectangular grid (Tavassoli, van der Linde, Cormack, & Bovik, 2004).

Fixations in any location within a tile were automatically used to select the noise within that tile. By cleverly combining the speed of data collection of the eye tracking methodology with the spatial accuracy of the psychophysical setup, this approach produces classification images of higher spatial resolution than reported here using fewer trials than traditional 2-AFC psychophysical visual tasks.

We are investigating other shift-invariant algorithms, such as the magnitude of the Fourier transform, and more sophisticated data analysis algorithms, like principal component analysis (Duda, Hart, & Stork, 2000) and independent component analysis (Hyvärinen, Karhunen, & Oja, 2001), to analyze the ROIs. Also, given that an observer is fixating at a point, the selection of the next fixation point is based on the low-resolution periphery. To address this issue, we are attempting a multiresolution analysis where the image patch at the next fixation point is filtered using established models of resolution fall off (Geisler & Perry, 1998) to simulate the variable resolution perception of the human visual system.

Acknowledgments

This research was supported by a grant from the National Science Foundation (ITR-0427372) and by ECS-0225451.

Commercial relationships: none.

Corresponding author: Umesh Rajashekar.

Email: umesh@ece.utexas.edu.

Address: The University of Texas at Austin, Center for Perceptual Systems, College of Liberal Arts, 1 University Station A8000, Austin, TX 78712.

References

- Abbey, C. K., & Eckstein, M. P. (2000). Estimates of human observer templates for simple detection tasks in correlated noise. *Proceedings of SPIE*, 3981, 70–77.
- Ahumada, A. J., Jr., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America*, 49, 1751–1756.
- Applied Science Laboratories. (1998). *Eye tracking system instruction manual* (Version 1.2). Bedford, MA: Applied Science Laboratories.
- Beard, B. L., & Ahumada, A. J., Jr. (1998). A technique to extract relevant image features for visual tasks. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human Vision and Electronic Imaging III. SPIE Proceedings* (Vol. 3299, pp. 79–85). Bellingham, WA: SPIE. [Article]
- Beutter, B. R., Eckstein, M. P., & Stone, L. S. (2004). Classification images reveal that saccades and perception use similar shape information in a visual search task [Abstract]. *Journal of Vision*, 4(8), 672a, <http://journalofvision.org/4/8/672/>, doi:10.1167/4.8.672.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436. [PubMed]
- Burr, D. C., Morrone, M. C., & Ross, J. (1994). Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Nature*, 371, 511–513. [PubMed]
- Caspi, A., Beutter, B. R., & Eckstein, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 13086–13090. [PubMed] [Article]
- Duchowski, A. T. (2002). *Eye tracking methodology: Theory and practice*. Springer-Verlag, London, UK (ISBN: 1-85233-666-8).
- Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern classification* (2nd ed.). San Diego: Harcourt Brace Jovanovich.
- Eckstein, M. P., & Ahumada, A. J., Jr. (2002). Classification images: A tool to analyze visual strategies. *Journal of Vision*, 2(1), i–i, <http://journalofvision.org/2/1/i/>, doi:10.1167/2.1.i. [PubMed] [Article]
- Efron, B. (1994). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A, Optics, and Image Science*, 4, 2379–2394. [PubMed]
- Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research*, 37, 617–631. [PubMed]
- Geisler, W. S., & Perry, J. S. (1998). A real-time foveated multiresolution system for low-bandwidth video communication. In B. Rogowitz & T. Pappas (Eds.), *Human Vision and Electronic Imaging, SPIE Proceedings* (Vol. 3299, pp. 294–305). Bellingham, WA: SPIE. [Article]
- Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis* (1st ed.). New York, NY: John Wiley & Sons.
- Klarquist, W., & Bovik, A. C. (1998). Fovea: A foveated vergent active stereo system for dynamic three-dimensional scene recovery. *Proceedings of the IEEE International Conference on Robotics and Automation*, 14, 755–770.
- Lee, S. (2000). *Foveated video compression and visual communications over wireless and wireline networks*. Unpublished doctoral dissertation, University of Texas at Austin, TX.
- Murray, R. F., Beutter, B. R., Eckstein, M. P., & Stone, L. S. (2003). Saccadic and perceptual performance in

- visual search tasks: II. Letter discrimination. *Journal of the Optical Society of America A, Optics, Image Science and Vision*, 20, 1356–1370. [[PubMed](#)]
- Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, 434, 387–391. [[PubMed](#)]
- Neri, P., & Heeger, D. J. (2002). Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nature Neuroscience*, 5, 812–816. [[PubMed](#)] [[Article](#)]
- Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, 16, 125–154. [[PubMed](#)]
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442. [[PubMed](#)]
- Rajashekar, U., Cormack, L. K., & Bovik, A. C. (2002, March 25–27). Visual search: Structure from noise. *Proceedings of the Eye Tracking Research and Applications Symposium* (pp. 119–123). New Orleans, LA.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422. [[PubMed](#)]
- Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network*, 10, 341–350. [[PubMed](#)]
- Simoncelli, E. P. (2003). Seeing patterns in the noise. *Trends in Cognitive Sciences*, 7, 51–53. [[PubMed](#)]
- Tavassoli, A., van der Linde, I., Cormack, L. K., & Bovik, A. C. (2004, December 16). *The efficient use of classification images for the psychophysical investigation of visual search*. 9th AVA Christmas Meeting (UK): Images, Perception & Psychophysics, Birmingham, UK.
- Williams, L. G. (1967). The effects of target specification on objects fixated during visual search. *Acta Psychologica*, 27, 355–360. [[PubMed](#)]
- Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention*. East Sussex, UK: University College London Press.
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews: Neuroscience*, 5, 495–501. [[PubMed](#)]
- Yarbus, A. L. (1967). *Eye movements and vision*. New York, NY: Plenum Press.