Point of Gaze Analysis Reveals Visual Search Strategies

Umesh Rajashekar[a], Lawrence K. Cormack[b] and Alan C. Bovik[a]

[a]Dept. of Electrical and Computer Engineering; [b]Dept. of Psychology
The University of Texas at Austin, Austin, TX 78712, USA

ABSTRACT

Seemingly complex tasks like visual search can be analyzed using a cognition-free, bottom-up framework. We sought to reveal strategies used by observers in visual search tasks using accurate eye tracking and image analysis at point of gaze. Observers were instructed to search for simple geometric targets embedded in $1/f$ noise. By analyzing the stimulus at the point of gaze using the classification image (CI) paradigm, we discovered CI templates that indeed resembled the target. No such structure emerged for a random-searcher. We demonstrate, qualitatively and quantitatively, that these CI templates are useful in predicting stimulus regions that draw human fixations in search tasks. Filtering a $1/f$ noise stimulus with a CI results in a 'fixation prediction map'. A qualitative evaluation of the prediction was obtained by overlaying k-means clusters of observers' fixations on the prediction map. The fixations clustered around the local maxima in the prediction map. To obtain a quantitative comparison, we computed the Kullback-Leibler distance between the recorded fixations and the prediction. Using random-searcher CIs in Monte Carlo simulations, a distribution of this distance was obtained. The z-scores for the human CIs and the original target were -9.70 and -9.37 respectively indicating that even in noisy stimuli, observers deploy their fixations efficiently to likely targets rather than casting them randomly hoping to fortuitously find the target.

Keywords: Classification image paradigm, $1/f$ noise, eye tracking, saliency map, Kullback-Leibler distance, fixation prediction, visual search

1. INTRODUCTION

Visual search is a common yet important task that plays a significant role in the evolution of a species. For example, visual search tasks such as locating prey and detecting predators are important for the survival of a species. Despite the seemingly complex mechanisms underlying search, humans excel at search tasks. One particular aspect of the human visual system (HVS) that is critical to its success as an efficient searcher is its *active* nature of looking. The HVS is *foveated*, comprising of a very small, high resolution central region, the fovea and a low resolution surrounding region, the periphery. Foveated visual perception provides a large field of view without the accompanying data glut. In order to build a detailed representation of a scene from this variable resolution input, the HVS uses a dynamic process of actively scanning the visual environment using a suite of eye movements.[1] The eye movements that are most relevant to this discussion are the discrete *fixations* and the ballistic *saccades* used to jump to the next fixation. During visual search, humans analyze the visual world with the high resolution fovea and use the low resolution peripheral information and memory to guide the hunt. Understanding how the HVS selects and sequences image regions for scrutiny can help develop a deeper understanding of search strategies in humans, which can then be extended to the design of machine vision systems. In this paper, we describe how eye tracking, when combined with the a simple yet powerful image analysis technique - the classification image paradigm - can be used to reveal strategies used by humans in a simple search task and how the resulting classification image (CI) templates are useful in predicting human fixation patterns in similar search scenarios.

Traditionally, research on visual search has focused on attempting to identify if search is, cognitively, a serial or a parallel computation. Towards this end, recording the accuracy and reaction time with variable number

of distracters[2] have been used to indirectly quantify search strategies. Reaction times and accuracy results, however, do not reveal what image features draw the observer's fixations during the search task. Recently, the theory of classification images[3] has become a very popular tool in psychophysics to visualize strategies used by human observers in visual discrimination/detection tasks.[4] The basic idea behind the classification image paradigm is that external noise, when added to a discrimination task reduces the signal to noise ratio and induces subjects to err in classification tasks, which they would have otherwise executed correctly in the absence of the noise. Since the external noise influences the subject's response, statistics in the noise samples that affect subject responses to the stimuli can be used to reveal the features used by the observer to perform the task. Assuming that observers use linear time invariant strategies, the classification image paradigm analyzes correlation across the noise samples to investigate strategies used by the observer to perform the discrimination task. Typically, the external noise is selected to be gaussian, white noise and many trials (on the order of tens of thousands) are performed before a CI begins to evolve.

As mentioned earlier, eye movements are a direct consequence of foveation and are fundamental to the dynamic aspect of search. Not surprisingly, there has been considerable interest in modelling human eye movements, typically by investigating image features such as wavelet coefficients, symmetry and contrast,[5] object similarity[6] and combinations of randomized saliency and proximity factors[7] that attract fixations. A more recent trend in automatically predicting regions-of-interest has been the investigation of image statistics directly *at the point of gaze*. It is theorized[8] that early visual processing may exploit the statistics inherent in its environment to represent the input as efficiently as possible. In one reported work,[9] the human fixation regions were found to have higher spatial contrast and spatial entropy than the corresponding random fixation regions indicating that the human eye may be trying to select image regions that help maximize the information content transmitted to the visual cortex by minimizing the redundancy in the image representation. These results seem to indicate that, despite the large number of fixations made in the duration of most visual tasks, eye movements are not necessarily random but that fixations are possibly drawn to image features that are unique in some sense. We believe that analysis of the statistics of the stimulus at point of gaze in a search task will, similarly, reveal strategies employed in search and in essence motivates our approach.

We have demonstrated earlier[10] that it is indeed possible to reveal low-level image features that are possibly used by observers in a search task by analyzing the stimulus at the point of gaze. In short, we recorded human eye movements in a visual search task where subjects looked for targets embedded in $1/f$ noise stimuli. Image patches at observers' point of gaze were analyzed using the classification image paradigm mentioned earlier to extract image properties that were most common in these patches. We found that the resulting CIs resembled attributes of the target for which the observer was searching. No such structure emerged for a random-searcher, indicating that even in very noisy displays, observers were not random-searchers but rather deployed their fixations to regions that resembled some aspect of the target.

In this paper, we extend our initial results to demonstrate, both qualitatively and quantitatively, the utility of these CIs in predicting fixations. Filtering an $1/f$ noise stimulus with an observer's CI results in an image that can be interpreted as a 'fixation prediction map'. Peaks in this map indicate a similarity in the stimulus to the CI and hence denote regions that should attract fixations (for the corresponding target). A qualitative measure of the prediction was obtained by overlaying observers' fixations (or a k-means cluster thereof) on the prediction map. We observed that fixations clustered around local maxima in the prediction map indicating that the CIs are good fixation predictors. To arrive at a quantitative measure, we compute a 'distance' between a pseduo-dense fixation map and the prediction map using the Kullback-Leibler distance - an information theoretic weighted correlation measure. We compare the performance of our fixation predictions to those of a random searcher and demonstrate that observers indeed utilize efficient search templates and are far from being random.

The rest of the discussion in this paper is organized as follows. In Section 2, we provide a brief overview of our previous work on revealing CIs using the classification image paradigm at point of gaze. Then, in Section 3 we demonstrate the efficacy of the CIs in predicting fixations. Finally, Section 4 presents the conclusions and some future directions of research.

(a) Circle (b) Dipole (c) Triangle (d) Bowtie

Figure 1. Targets used in the search task

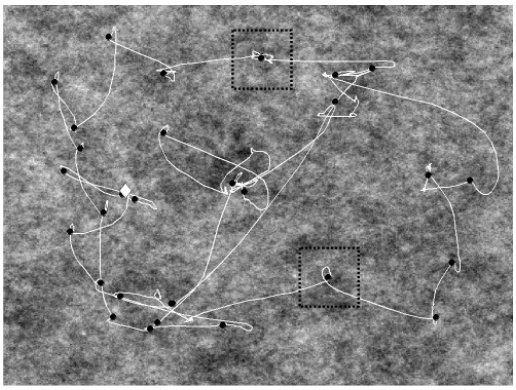


Figure 2. Sample recorded scanpath. Dots indicate fixations. Diamond shows location of 'Triangle' target

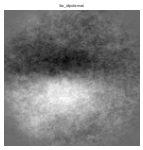
2. CLASSIFICATION IMAGE ANALYSIS AT POINT OF GAZE

In this section, we provide a quick overview of the methodology and the results of performing classification image analysis at observers' point of gaze. The experimental setup is presented in more detail in Ref. 10.

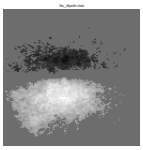
2.1. Experimental setup

Three observers were used for the experiments. Two of these observers (LKC and UR) were familiar with the experiments and the third (EM) was a naive observer. All observers either had normal or corrected-to-normal vision. Observers were instructed to locate simple geometric targets such as circles, dipoles, triangles and bowties (as shown in Fig. 1) that were embedded in stimuli of $1/f$ noise. This $1/f$ noise has an average Fourier magnitude that mimics the average spectrum of natural images[11] and makes an effective type of noise for obscuring (or 'masking') targets. The size of the $1/f$ noise stimuli was $640 * 480$ pixels and the size of the embedded target was $64 * 64$ pixels. The stimuli were displayed on a 21 inch, gamma corrected monitor at a distance of $180cm$ from the observer. The screen resolution was set at $640 * 480$ pixels. This set up corresponded to about 52 pixels/degree of visual angle.

Observers were shown a target image at the beginning of a block of trials. The observer's task during each presentation of the stimulus was to locate the target in the $1/f$ noise stimulus as soon as possible. Blocks of 50 trials with the target embedded randomly in space in the $1/f$ noise stimuli during each trial were used. During each trial, the $1/f$ noise image was selected randomly from a set of nine different patterns of $1/f$ noise to discourage the subject from remembering the structure from previous presentations. The signal-to-noise ratio of the target images was set so that the observers made many fixations (~ 20) to find the target. On finding the target (or what appeared to be a target), the observer pressed a button and proceeded to the next image. Feedback on the exact location of the target was provided before the observer proceeded to the next stimulus. The MATLAB psychophysics toolbox [12,13] was used for stimulus presentation. An example of one such stimulus is shown in Fig. 2. The location of the target (in this case a low contrast triangle) is indicated by the location of the diamond.



(a) Classification image



(b) Statistical thresholding

Figure 3. Statistical thresholding of CIs

Human eye movements were recorded using an SRI Generation V Dual Purkinje eye tracker. The output of the eye tracker (horizontal and vertical eye position signals) was sampled at $200Hz$ by a National Instruments data acquisition board in a Pentium IV host computer, where the data was stored for offline data analysis. During the experiment, periodic verifications of the eye tracker calibration were performed by displaying a dot on the display at the position of gaze in real-time and, if necessary, re-calibration was done.

2.2. Image Analysis at Point of Gaze

To build a CI at the point of gaze, the sampled voltages from each trial were first converted to gaze coordinate position on the image. Next, the path of the observer's gaze was divided into fixations and the intervening saccadic eye movements using spatio-temporal criteria derived from the known dynamic properties of human saccadic eye movements.[14] The resulting patterns of fixations and saccades for a single trial are shown in Fig. 2. We defined a region-of-interest (ROI) of $128 * 128$ pixels around each fixation (two examples of which are shown in Fig. 2 by the boxes). To form a CI for an observer searching for a particular target, the noise (*only*) in all the ROIs corresponding to the observer searching for that target were averaged together. The resulting CI for one observer (LKC) searching for the 'Dipole' target is show in Fig. 3(a). Gray denotes zero, white corresponds to positive pixel values and black to negative pixel values. To enhance the details in the results, the CIs were then thresholded for statistical significance. In other words, only the pixels whose magnitudes exceeded 1 standard deviation of the distribution of pixel intensities in the CI were retained and the rest were set to zero. The statistically thresholded CI corresponding to the CI in Fig. 3(a) is shown in Fig. 3(b). The statistically thresholded results for all observers and targets are shown in Fig. 4. The first row in Fig. 4 illustrates the targets that the observers were looking for. Each of the other rows show the CIs for the three observers (LKC, UR and EM) for each of these targets.

2.3. Discussion

The observer, LKC, for example, while searching for the dipole, seemed to attend to a small, central portion of dipole, perhaps weighting the lower white portion more. For the case of a circle, the observer seemed to be fixating at points that have a bright region while for the triangle, the observer seemed to be searching for a white region and the right diagonal edge of the triangle (indicated by the appearance of the dark region towards the top right). Darker regions in the image seemed to attract observer LKC's attention in the search for the 'Bowtie'. The CIs for other observers indicates that different observers use idiosyncratic search strategies.

Traditionally, white gaussian noise has been used as the masking signal in the classification image paradigm. To evaluate the effect of the $1/f$ nature of the noise on our classification results, we randomly sampled the nine $1/f$ noise stimuli to simulate a random searcher who casts fixations randomly, hoping to find the target by chance. The result of adding the ROIs at the random fixation points is shown in far right in Fig. 5 (displayed using the same range of gray scales to highlight the relative pixel magnitudes across CIs). The lack of any image structure in the random sampling case and the ability to generate many CIs across subjects from the same set of nine $1/f$ noise stimuli indicates that while the $1/f$ noise is a great masking stimulus, it does not significantly influence the resulting CI.

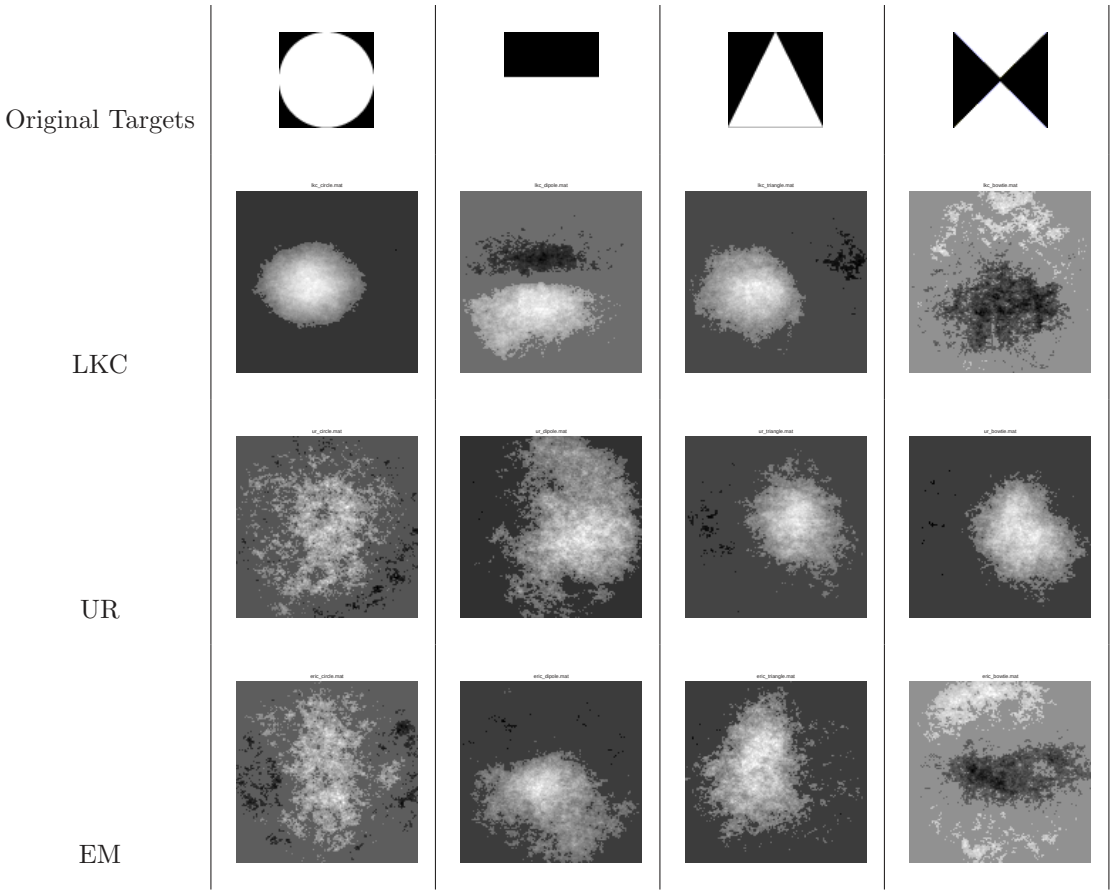


Figure 4. Statistically filtered CIs for 3 observers (LKC, UR and EM)

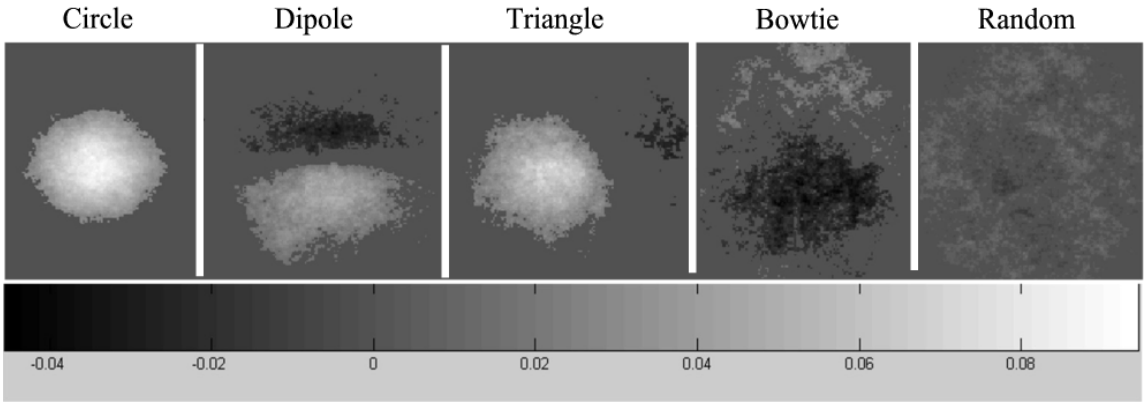


Figure 5. Comparing classification images for humans vs random fixations

In conclusion, unlike the CIs for observers, the CI for random sampling tends towards an image with no specific structure, indicating that observers are not random searchers, but that they actually are directing their fixations to regions that resemble the target.

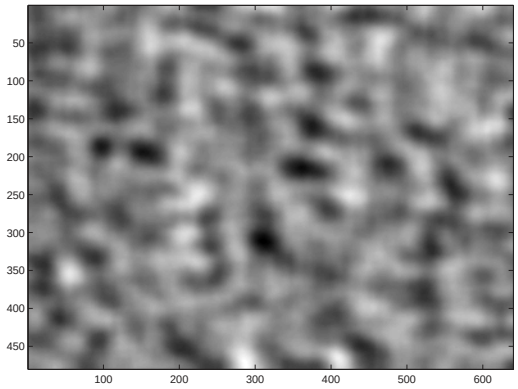


Figure 6. Prediction obtained by convolving LKC's triangle CI with a noise image

3. PREDICTING OBSERVERS' FIXATIONS

The CIs shown in Fig. 4 were obtained by averaging only the noise at each of the observer's fixation. The CIs, therefore, reveal image features that were common in many of the observer's fixation. These pixel correlations reveal the image features that observers looked for during the search task. Since the CI templates in Fig. 4 were derived from human eye movement data, these CIs (before statistical filtering) can be used as fixation prediction kernels to predict observers' fixations in a new search scenario. Note that while the original target templates shown in Fig. 1 are the optimal search templates, they do not necessarily predict observers' fixations. To predict an observers' fixations in a new noise stimulus, we use a CI template as a search template and convolve the noise image using this CI. A resulting filtered image using the observer LKC's triangle CI (Fig. 4) is show in Fig. 6. Peaks (indicated by bright regions) in this image corresponds to a good match between regions in the noise image and the observers's search template and hence signify a region that would attract the observer's fixation. The resulting image can be visualized as a likelihood map where the peaks indicate regions that will draw fixations with a high probability. We now demonstrate both qualitatively and quantitatively that these prediction maps are indeed good indicators of regions that will draw fixations.

3.1. A qualitative comparison between fixation predictions and recorded fixations

A good fixation prediction map would require that a majority of the recorded fixations cluster around the local maxima in the prediction map. Therefore, to evaluate the utility of the fixation prediction map in actually predicting human fixations in a search task, we first collect the fixations of all observers looking for a particular target in the given noise image as shown in Fig. 7(a). Next, we simplify the visualization process by grouping these fixations into different clusters using the popular k-means clustering algorithm.[15] Given the total number of clusters, N, the k-means algorithm begins by randomly selecting N cluster centers. All points nearest to a given cluster center are grouped together, and the cluster center is updated by computing the mean of each cluster. This process continues until the locations of the centers stabilize.

In our case, however, it is not possible to guess the number of fixation clusters beforehand. To address this problem, we first replace each fixation in the noise image (Fig. 7(a)) by a 2D Gaussian envelope. An illustration of this process is shown in Fig. 9. The variance of the Gaussian envelope was chosen so that the full width at half-max (FWHM) of the Gaussian reflected the diameter of the foveola (about 1^o visual angle). This approximation of a fixation point accounts for the uncertainty of the exact location of the fixation point. By spreading a fixation around, this approximation also encompasses any small errors in calibration, accuracy and/or precision of the eye movement measurement. Finally, replacing each fixation point by a 2D Gaussian envelope reflects the probability that a neighborhood region around the fixation could have been selected as a fixation point. We then found the local maxima in this map and the locations of these local maxima were selected as the initial guess for the cluster centers for the k-means algorithm.

The result of applying the k-means algorithm on the fixations of Fig.7(a) is as shown in Fig. 7(b). To aid visualization, each cluster was represented by an ellipse centered on the cluster mean. The standard deviation of

the cluster along principal orthogonal directions was used to compute the major and minor axes of the ellipse. Clearly, even though the k-means clustering algorithm does a good job at grouping nearby fixations, the density of fixations in a cluster is often sparse. The reason for this is that in the classic k-means implementation there is no control over the growth of the cluster size. Points that are closest to a cluster mean, even if the distance of the point to the cluster is much larger than the average distance of the points in the cluster, are included in the cluster.

A simple solution to this problem is the idea of a density-constrained clustering wherein the growth of the cluster is constrained by a minimum density requirement. In other words, the cluster is allowed to grow in size only if the new cluster contains a minimum number of fixations per unit area. Details of the implementation of a density-based algorithm, DBSCAN, can be found in Ref. 16. To prevent the k-means clusters from becoming too sparse, we processed each k-means cluster using the above mentioned density-constrained clustering algorithm. This breaks a sparse cluster into smaller but denser clusters (if possible). The result of applying the density-constrained clustering on the k-means clusters of Fig. 7(b) is shown in Fig. 7(c). Notice that many of the sparse clusters are now represented by a group of smaller, dense clusters. As before, each resulting cluster was then represented by an ellipse.

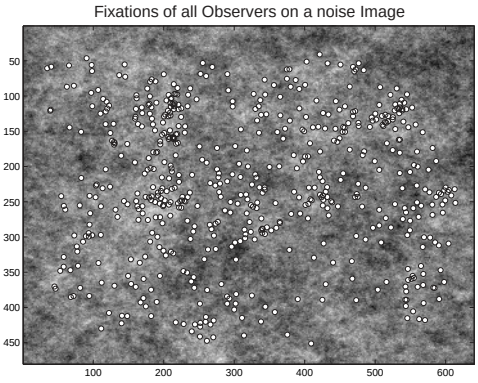
Finally, to visualize the efficacy of the fixation prediction map, these ellipses (representing a cluster of fixations) were overlaid on the (thresholded) fixation prediction map (of Fig. 6) as illustrated in Fig. 7(d). The local maxima in the prediction map have been labelled in the decreasing order of magnitude and hence decreasing order of probability of attracting a fixation. A 'good' (quantified in section 3.2) overlap between the local maxima and the fixation ellipses indicates that the CIs are indeed good fixation prediction kernels. The results for a few more noise images is shown in Fig. 8.

3.2. Comparing fixation predictions and recorded fixations - A quantitative measure

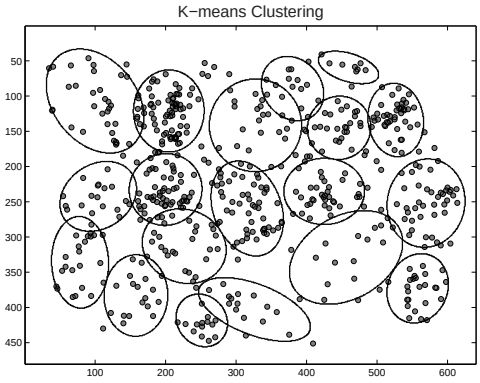
Evaluating the performance of fixation predictions quantitatively involves comparing a sparse set of discrete observer fixations (as shown in Fig. 9(a)) to a dense fixation likelihood map (as shown in Fig. 6). The comparison would be easier if both the recorded fixations and predictions were either dense or sparse. In one approach,[5] the dense prediction map is made sparse by detecting a certain number of local maxima (equal to the number of observer fixations for that image, for example) and clustering nearby local maxima into regions-of-interest. Each region is then labelled alphabetically for identification. This clustering and labelling process is repeated for observers' fixations. Overlapping regions are given the same alphabet identifier. A string matching/editing algorithm is then used to compare ROIs predicted by algorithms and those actually fixated by humans. Other methods of comparing human fixations to predictions are discussed in Ref. 17.

In our experiments, since there were substantially many fixations per noise image (about 400 fixations per image), we opted to extrapolate this human eye fixation data to a pseudo-dense prediction map similar to that method used in Ref. 18. To achieve this, we replaced each fixation by a 2D Gaussian envelope as described in Section 3.1. The motivation, as explained earlier in this section, for substituting each fixation point by a 2D Gaussian envelope was to build a pseudo-dense fixation map of observers' fixations (Fig. 9(c)) which could then be compared with the dense prediction map (Fig. 6) obtained by convolving the stimulus with a prediction kernel. The pseudo-dense fixation map and the prediction map can now be visualized as probability distributions. Peaks in the prediction map indicate regions in the stimuli that are likely to draw fixations and peaks in the eye fixation map denote regions where many observers actually made fixations. The goal then is to compare these two dense probability density functions. The distance between two probability distributions can be quantified by a well known information theoretic measure - the Kullback Leibler distance.[19] The Kullback-Leibler distance or the relative entropy between two probability density functions $p(x)$ (say, the fixation map in this case) and $q(x)$ (the pseudo-dense prediction map) is given by $\mathscr{D}(p||q) = \sum_{x \in \mathscr{X}} p(x) log\left(\frac{p(x)}{q(x)}\right)$.

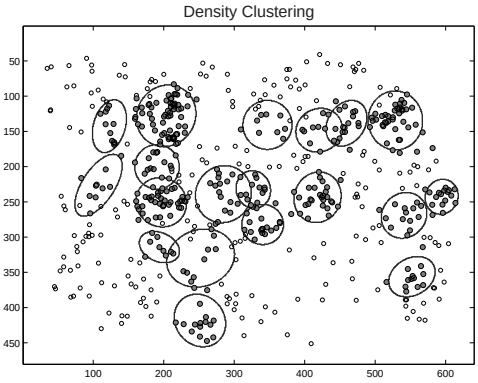
The KLD is analogous to a weighted correlation measure and measures the 'closeness' between two probability density functions. From an entropy perspective, the KLD can also be interpreted as the inefficiency incurred by assuming the distribution of a random variable to be q when the true distribution is actually p. Note that $\mathscr{D}(p||q) \geq 0$ and is $= 0$ if and only if $p = q$. Though it is convenient to think of $\mathscr{D}$ as 'distance' between distributions, $\mathscr{D}$ is not a distance metric because it asymmetric and does not satisfy the triangle inequality



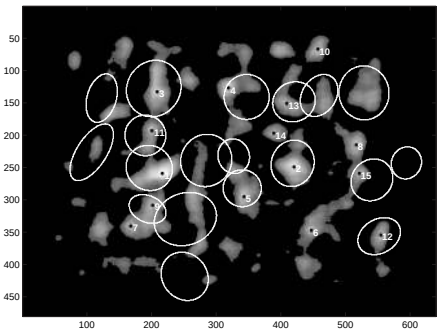
(a) Collection of fixations of all observers looking for a triangle in this noise image



(b) Result of K-means clustering



(c) Results of Density clustering the K-means clusters



(d) Overlaying clusters on fixation prediction map

Figure 7. Illustrating the intermediate steps in clustering the fixations

required of distance metrics. The symmetrization of the KLD can be achieved easily (such as the average of $(\mathscr{D}(p||q) + \mathscr{D}(q||p))$). We chose to use the harmonic distance $1/\left[1/\mathscr{D}(p||q) + 1/\mathscr{D}(q||p)\right]$ due to the advantages described in Ref. 20.

We used this symmetric KLD to quantify the distance between our fixation predictions and the recorded eye fixations (now represented by the pseudo-dense fixation map). Before computing the distance, both these maps were first normalized to be probability density functions. The average KLD results for all subjects and targets are shown in Fig. 10. Each figure represents the average KLD results for all the observers looking for one particular target in nine different $1/f$ noise images. The black bars in these figures denote the KLD between the eye fixations recorded for a particular observer (indicated on the abscissa) looking for a particular target and the predictions obtained using that observer's CI corresponding to the target under consideration. Also shown (by the gray bar) is the distance between the recorded eye fixations and the prediction obtained using the actual target templates shown in Fig. 1. The resulting prediction map would indicate the regions an ideal searcher would fixate at while searching for the target. The original target template, while the optimal search template, does not necessarily predict the human eye fixations as well the corresponding CI for the observer. Finally, using the random-fixation CIs in Monte Carlo simulations, a distribution of the KLD (shown by the white bars) was

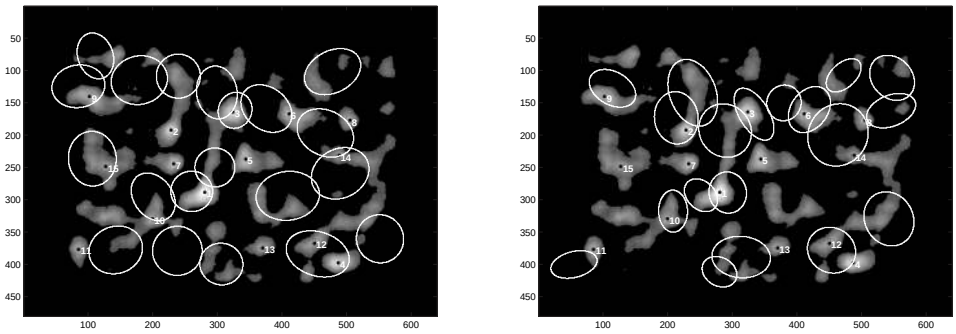


Figure 8. More Clustering Examples

obtained. The average z-score for the human CI was -9.70 indicating that the observers were far from being random searchers. Further, the z-score for the original target was -9.37 indicating that observers used efficient search templates (since the target templates of Fig. 1 would be used by the ideal observer).

Even though the KLD does not give an intuitive feel for the closeness of prediction to actual eye fixation (say, in terms of percentage match), it is a convenient measure to compare and contrast the performance of one prediction kernel again another and a convenient way to quantify the deviation of an observer from being a random searcher.

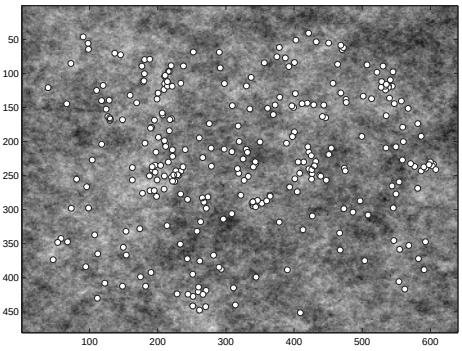
4. CONCLUSION

Analysis of stimuli *at the observers' point of gaze* can provide an understanding of strategies used by observers in visual tasks. In this paper, we demonstrated that combining the classification image paradigm with accurate eye tracking can be used to reveal strategies used by observers in visual search. The results indicate that even in simple search tasks like the one described above, human observers are not random in their search task but rather make directed eye movements to regions in the stimuli that resemble the target they are looking for. Also, we demonstrated both qualitatively and quantitatively (using the Kullback-Leibler measure) that the CIs are indeed good fixation prediction kernels.

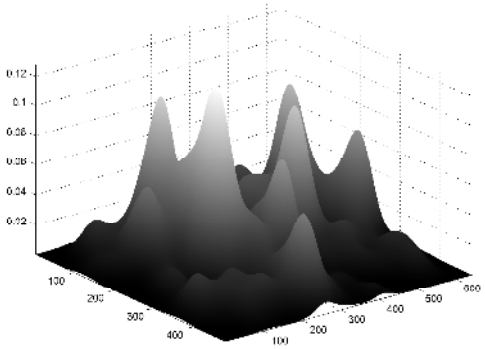
The analysis procedure presented in this paper is not without its shortcomings. The classification image paradigm has been popular in psychophysics where the contribution of each pixel to the creation of the template is shift-invariant across the trials. With the added dimension of eye movements, the noise pixels in the image patches are no longer guaranteed to be aligned across image ROIs as in the psychophysics experiments. It is therefore quite possible that even though interesting features are revealed in Fig. 4, many image features might be getting 'washed out' in the averaging process. We are investigating shift invariant algorithms (such as the magnitude of the Fourier transform) and more sophisticated data analysis algorithms (like Principal Component Analysis[15] and Independent Component Analysis[21]) to analyze the ROIs. Finally, 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 multi-resolution analysis where the image patch at the next fixation point is filtered using established models of resolution fall off[22] to simulate the variable resolution perception of the HVS.

ACKNOWLEDGMENTS

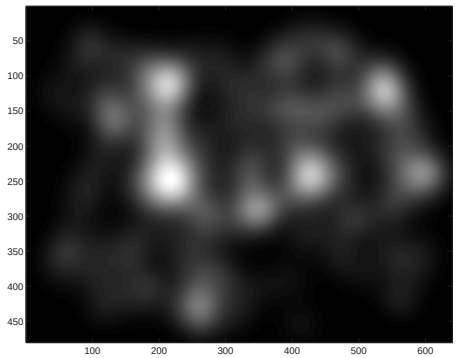
This research is supported by financial support from the Texas Advanced Research Program and the National Science Foundation.



(a) Collection of all fixations for one observer searching for the dipole target



(b) 3D visualization of dropping Gaussians on eye fixations



(c) 2D visualization of dropping Gaussians on eye fixations

Figure 9. Generating a pseduo-dense fixation map from discrete fixations

REFERENCES

1. A. L. Yarbus, *Eye Movements and Vision*, Kluwer Academic Publishers, January 1967.
2. J. M. Wolfe, "Visual search," in *Attention*, H. Pashler, ed., University College London Press, 1998.
3. M. P. Eckstein and J. Albert J. Ahumada, "Analysis of a complex of statistical variables into principal components," *Journal of vision* **2**(1), p. 1x, 2002.
4. B. L. Beard and A. J. Ahumada, Jr., "A technique to extract relevant image features for visual tasks," *SPIE Proc. Human Vision and Electronic Imaging III* **Vol. 3299**, pp. 79–85, 1998.
5. C. M. Privitera and L. W. Stark, "Algorithms for defining visual regions-of-interest: comparison with eye fixations," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **Volume: 22**, pp. 970–982, Sept 2000.
6. B. Moghaddamand and A. Pentland, "Probabilistic visual learning for object detection," *Fifth Int. Conf. Computer Vision* , pp. 786–793, June 1995.
7. W. Klarquist and A. C. Bovik, "Fovea: a foveated vergent active stereo system for dynamic three-dimensional scene recovery," *IEEE Tran. on Robotics and Automation* **14**, pp. 755–770, October 1998.

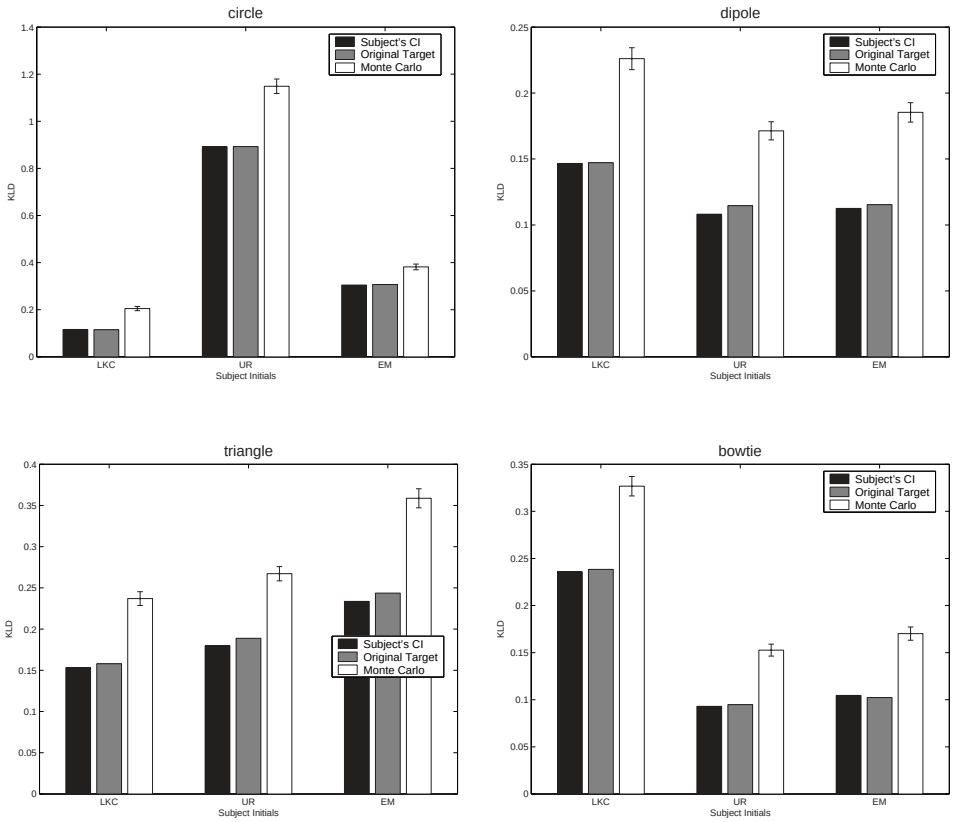


Figure 10. KLD for all subjects and targets

8. H. B. Barlow, *Possible principles underlying the transformation of sensory messages*, pp. 217–234. M.I.T. Press, Cambridge MA, 1961.
9. P. Reinagel and A. M. Zador, "Natural scene statistics at the center of gaze," *Network: Computation in Neural Systems* **10**(1-10), 1999.
10. U. Rajashekar, L. K. Cormack, and A. C. Bovik, "Visual search: Structure from noise," *Proceedings of the Eye Tracking Research and Applications Symposium* , pp. 119–123, 2002. New Orleans, LA, USA, March 25-27.
11. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am.* **A**(4(12)), pp. 2379–2394, 1987.
12. D. H. Brainard, "The psychophysics toolbox," *Spatial Vision* **10**, pp. 433–436, 1997.
13. D. G. Pelli, "The videotoolbox software for visual psychophysics: Transforming numbers into movies," *Spatial Vision* **10**, pp. 437–442, 1997.
14. Applied Science Laboratories, "Eye tracking system instruction manual." Ver 1.2, 1998.
15. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Harcourt Brace Jovanovich, San Diego, Second ed., November 2000.
16. J. S. Martin Ester, Hans-Peter Kriegel and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *Proc. of 2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD-96)* , pp. 226–231, 1996.
17. D. S. Wooding, "Eye movements of large populations: Ii. deriving regions of interest, coverage, and similarity using fixation maps," *Behavior Research Methods, Instruments and Computers* **34**(4), pp. 509–517, 2002.

18. D. S. Wooding, "Fixation maps: quantifying eye-movement traces," *Proc. of the Eye Tracking Research and Applications Symposium* , pp. 31 – 36, 2002. New Orleans, LA, USA, March 25-27.
19. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 1991.
20. D. Johnson and S. Sinanovic, "Symmetrizing the kullback-leibler distance," March 2001. http://cmc.rice.edu/docs/docs/Joh2001Mar1Symmetrizi.pdf (Last viewed Dec. 9, 2003).
21. A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 1 ed., May 2001.
22. W. S. Geisler and J. S. Perry, "A real-time foveated multiresolution system for low-bandwidth video communication," *Human Vision and Electronic Imaging, SPIE Proceedings* **3299**, pp. 294–305, 1998. B. Rogowitz and T. Pappas (Eds.).