Perception & Psychophysics
2007, 69 (1), 103-113

An efficient technique for revealing visual search strategies with classification images

Abtine Tavassoli
University of Texas, Austin, Texas

Ian van der Linde
Anglia Ruskin University, Chelmsford, England

and

Alan C. Bovik and Lawrence K. Cormack
University of Texas, Austin, Texas

We propose a novel variant of the classification image paradigm that allows us to rapidly reveal strategies used by observers in visual search tasks. We make use of eye tracking, 1/*f* noise, and a grid-like stimulus ensemble and also introduce a new classification taxonomy that distinguishes between foveal and peripheral processes. We tested our method for 3 human observers and two simple shapes used as search targets. The classification images obtained show the efficacy of the proposed method by revealing the features used by the observers in as few as 200 trials. Using two control experiments, we evaluated the use of naturalistic 1/*f* noise with classification images, in comparison with the more commonly used white noise, and compared the performance of our technique with that of an earlier approach without a stimulus grid.

Understanding how humans search for objects in a visual scene is a fascinating yet poorly understood topic. The classification image paradigm, originally developed with 1-D signals for auditory psychophysics and later extended to images (2-D signals) for vision research, in order to study observer strategies in a vernier acuity task (Ahumada, 1996; Beard & Ahumada, 1998), is a potentially helpful tool for probing the search strategies used by human observers (Rajashekar, Cormack, & Bovik, 2004). In the original classification image paradigm, the observer responses, over numerous trials, in a psychophysical *yes–no* type experiment (e.g., detection of a target embedded in noise) are used to categorize the stimulus noise into four groups (hits, misses, false alarms, and correct rejections). The stimulus noise is then averaged within each category, and the combination of these averages produces what is referred to as the *classification image*. The obtained image shows how the observer weights individual pixels to make a decision, and the spatial patterns formed by these pixels can reveal the filters or features used by the observer.

Classification images have been used in several interesting areas: illusory contours (Gold, Murray, Bennett, & Sekuler, 2000), image feature detection and identification (Neri & Heeger, 2002), stereo (Neri, Parker, & Blakemore, 1999), and visual attention (Eckstein, Shimozaki, & Abbey, 2002).[1] A general limitation of this technique in its original form is that it required a large number of trials (often several thousand per observer). A variant of the classification image technique was developed in our lab for the study of visual search (Rajashekar, Cormack, & Bovik, 2002). In this technique, observers' eye movements were recorded while they searched for a target embedded in 1/*f* noise. By assuming that gaze would be drawn to points in the stimulus bearing some resemblance to the target, the noise at all fixations made during a trial was captured, and a large volume of data could thus be gathered in a short time.

In this article, we present a further variant of the classification image technique in the context of a simple visual search task. In this method, a 1/*f* noise mask is divided into discrete tiles (although other noise types may be used), and the target is embedded in one of them, selected at random. The eye movements of observers are then recorded while they search for the target, in order to determine the sequence of tiles fixated during the search.

1/*f* noise has an amplitude spectrum $A(f) = 1/f^a$, where a is near 1, which is similar to the amplitude spectra of natural images (Field, 1987), and it is used due to its appeal in simulating a realistic search environment. The additional power at low spatial frequencies (relative to white noise) results in rapid emergence of features in the classification image with our method, at the scale of reasonably sized targets, without the requirement for postprocessing. Several aspects of our technique allow it to rapidly reveal

A. Tavassoli, atavasso@ece.utexas.edu

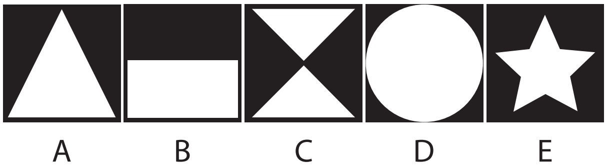

Figure 1. Targets used in the trials: (A) Triangle and (B) dipole. Additional shapes used in the analysis: (C) Bow tie, (D) circle, and (E) star.

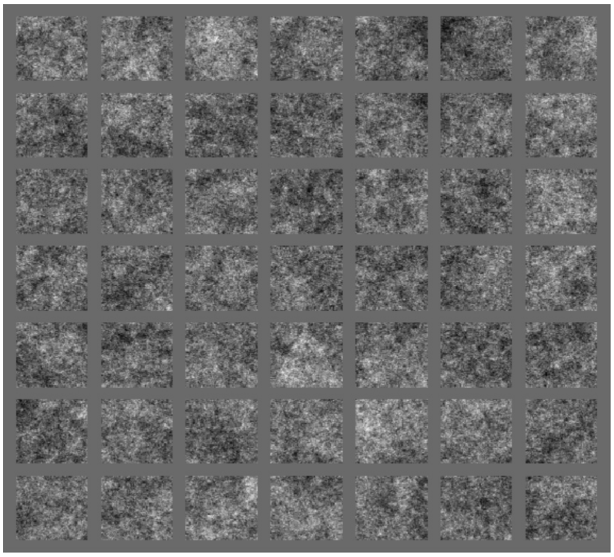

Figure 2. An example stimulus (with a higher signal-to-noise ratio than those used during the experiment). The target, a triangle, is in the tile immediately below the central one.

classification images. First, the use of eye tracking allows a large volume of data to be collected in a given time, in comparison with traditional psychophysical methods. Second, the use of discrete tiles makes the method more robust to saccadic inaccuracy, the tendency for observers to fixate different parts of the target, and the limited accuracy and precision of the eye movement recordings, all of which would ultimately result in loss of spatial precision (or blur) in the final classification images. Finally, our novel classification taxonomy provides several new categories for offline analysis, allowing us to differentiate foveal and nonfoveal aspects of the search process (see the Analysis Method section).[2] For comparison, we also ran two control experiments, identical to the main experiment but (1) without the grid, to demonstrate its efficacy in accelerating the convergence of classification image results, and (2) using uniform white noise, rather than 1/*f* noise, to demonstrate that the classification images produced with both noise types are similar whether search targets are embedded in 1/*f* noise directly or white noise is used and the noise tiles are pinkened prior to averaging, to amplify lower frequency information.

METHOD

Observers

Three of the authors (26, 28, and 40 years of age) served as observers. All had emmetropic vision. Two were experienced psychophysical observers, and all 3 were familiar with and comfortable in the eyetracker.

Apparatus

An SRI/Fourward Generation V Dual Purkinje eyetracker (Fourward Technologies, Buena Vista, VA) was used to record eye movements. This device has an accuracy of better than 10 min of arc, a precision of about 1 min of arc, and a response time of about 1 msec (although we would like to note that a principal advantage of our methodology is that it permits the use of a considerably less accurate tracker). A bite bar and forehead rest were used to minimize head movements. The continuous output voltage of the eyetracker was first passed through a hardware Butterworth low-pass filter (Krohn-Hite Corp., Brockton, MA) with a 100-Hz cutoff to eliminate extraneous high-frequency noise in the recording environment and then was sampled by the host computer at 200 Hz with a National Instruments data acquisition card (National Instruments Corp., Austin, TX).

A calibration routine was run at the beginning of each session and after every 25 trials during a session in order to establish the linear relationship between output voltage and monitor coordinates. For the calibration, the observer fixated each of nine points in a 3 × 3 grid spanning a visual angle of 7° × 7° on the display. The average horizon-

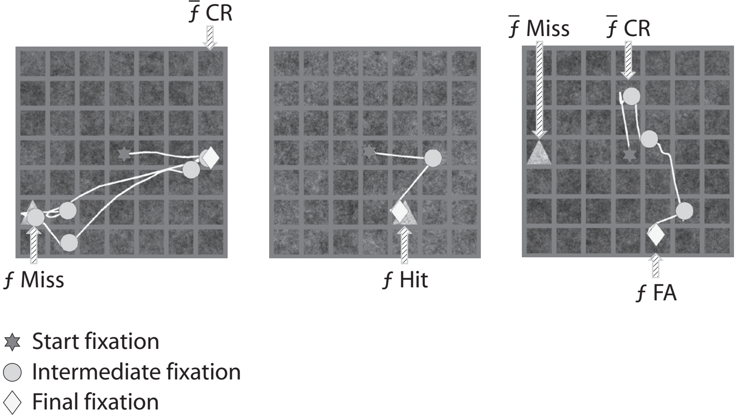

Figure 3. Examples of scanpaths and tile categories. The signal-to-noise ratio has been increased for illustration purposes.

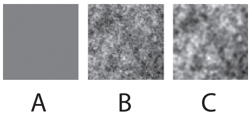

Figure 4. The expected image from randomly sampling tiles: (A) Raw, (B) contrast stretched, and (C) low-pass filtered (using a 3 × 3 Gaussian mask, with $\sigma = 0.9$).

tal and vertical voltages were then fit (separately) to the three unique horizontal and vertical screen positions (corrections were performed for the small cross-talks). Afterward, a dot was superimposed on the computed gaze position in real time so the observer could immediately verify that calibration was successful. In addition to the mandatory recalibration every 25 trials, the calibration was automatically checked at the beginning of each trial. This was done by requiring that the computed fixation be within ±0.25º of the center of the fixation mark for 500 msec at the beginning of each trial. If 5 sec elapsed before this requirement was met, recalibration was automatically initiated.

The observers viewed the stimuli on an Image Systems 21-in. grayscale monitor (Image Systems Corp., Minnetonka, MN) driven by a Matrox Parhelia graphics card (Matrox Graphics, Dorval, Quebec) at a screen resolution of 1,024 × 768 pixels, a grayscale resolution of 8 bits per pixel, and a refresh rate of 60 Hz. The screen was placed 134 cm from the observer and subtended a visual angle of 16º × 12º, giving approximately 1 min of arc per screen pixel. The luminance output was linearized by putting the inverse of the monitor's measured gamma function in the display look-up table. The ambient illumination in the laboratory was kept constant for all the observers, and there was a minimum of 5 min to adapt to the ambient illumination and screen luminance while the eyetracker was calibrated.

The experimental software was written in MATLAB (The MathWorks, Natick, MA), and the stimulus presentation itself was controlled using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Gaze positions were calculated in real time so that feedback could be provided after each trial. Fixation points and the intervening saccades were discriminated offline, on the basis of the spatiotemporal properties of human eye movements, by using an adaptation of an ASL fixation detection algorithm (Applied Science Laboratories, Bedford, MA). This three-stage algorithm was robust with respect to small drifts, blinks, and microsaccades.

Stimuli

The stimulus consisted of a single 64 × 64 pixel target embedded in a 7 × 7 mosaic of 64 × 64 pixel tiles containing 1/*f* masking noise, where $a = 0.8$. The two targets used are shown in Figures 1A and 1B (the shapes shown in panels C–E were used in data analysis, but not in the experiment per se; see below). One hundred mosaics were generated offline by creating one hundred 544 × 544 pixel 1/*f* noise images and then superimposing the 12-pixel-wide gray borders. On each trial, the target was added to a randomly selected tile in the noise mosaic. An example stimulus in which a triangle is embedded in the tile immediately below the center one is shown in Figure 2. The signal-to-noise ratio (SNR) in this example is somewhat higher than those used in the actual experiment, where the SNR was determined at the beginning of each session as described below.

Procedure

Each observer ran four sessions for the main experiment: two sessions of 100 trials for each of the two target types. Before every session, the SNR yielding 68% correct target detection was determined using the QUEST adaptive procedure (Watson & Pelli, 1983). Note that this is effectively a contrast threshold, but we covaried the contrast of the target and of the noise so that the entire grayscale was used but never exceeded. This SNR threshold was determined using the same procedure as that in the experiment itself. In other words,

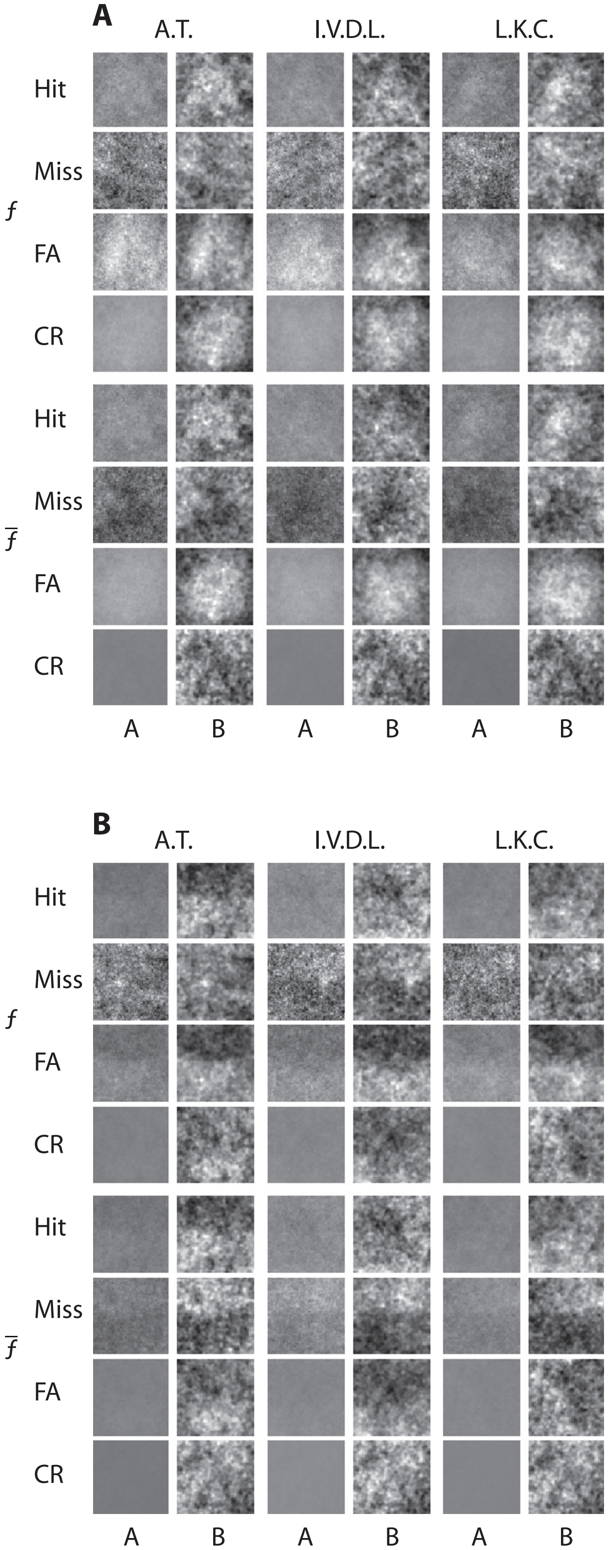

Figure 5. The average images for (A) triangle and (B) dipole target search are shown for 3 observers. Columns labeled "A" contain the raw average images collectively scaled to a single common grayscale color map, and columns labeled "B" contain the raw images after low-pass filtering and individual contrast enhancement.

Table 1
The Noise Tile Taxonomy

	Category		
	Nonfoveal	Foveal	Observer Response
Target present	$\bar{f}$ Hit	f Hit	Maintained fixation on target
		f Miss	Continued search
	$\bar{f}$ Miss		Not fixated
Target absent	$\bar{f}$ FA	f FA	Maintained fixation on target
		f CR	Continued search
	$\bar{f}$ CR		Not fixated

a trial during the threshold determination was exactly the same as a trial during the experiment, except that, in the former, the SNR was varied to find the 68% correct point, whereas in the latter, the SNR was fixed at that point. Since the first several trials of the QUEST are necessarily done at a relatively high SNR, these trials served to familiarize the observers with the task.

At the beginning of each trial, a fixation mark appeared at the center of the display for a maximum of 5 sec. As was described earlier, if the observer's computed fixation was within our error tolerance, the trial continued. Next, the fixation mark was replaced by the stimulus for 5 sec, and the observer searched for the target with the goal of having his fixation on the correct tile when the trial ended. The computer provided audio feedback ("correct" or "incorrect") after each trial.

The use of a common initial fixation point and a fixed, 5-sec trial duration ensured a somewhat consistent strategy and criterion across observers that yielded several fixations per trial. To wit, if we had used a very short duration, the experiment would effectively have become a 49-alternative forced choice yielding few fixations per trial. If we had used long or unlimited durations, different response criteria could have resulted in very different strategies, including exhaustive search. We chose 5 sec as a compromise that would allow the observers to visit several (five to six, on average) likely tiles without the search becoming exhaustive (resulting in fixations on very unlikely tiles). Post hoc analyses (see the Results section) suggested that the compromise was an acceptable one. It would also be possible, of course, to use a variable payoff matrix (for example), instead of imposing a time limit, but we chose the simpler option in order to demonstrate our basic method. The small number of fixations that fell between the tiles in the stimulus grid were not included in our analysis.

Analysis Method

Classification taxonomy. In a *yes–no* detection experiment, responses can be categorized into hits, misses, false alarms, and correct rejections, depending on the observer's response and whether the target was actually present. In the psychophysical classification image paradigm, the stimulus noise is averaged within each category, and these averages are combined to form the classification image. For example, the average for the hits and false alarms can be subtracted from the average for the misses and correct rejections, under the assumption that if a given pixel inclines the observer to say "target present" when bright (say), it should also incline the observer to say "target absent" when dark (Ahumada, 1996). The fidelity of the image from each category will actually depend on the observer's sensitivity and bias, but the fidelities seem to be about equal in the simple psychophysical situation, so combining the averages with equal weight is close to optimal (Ahumada, 2002).

In this study, we simply extended the categorization above to accommodate eye movements. Consider that each fixation (excluding the initial fixation at stimulus onset) involves two decisions: the decision to fixate a certain tile (and not the others) and the subsequent decision to either remain on that tile or continue searching. The presumption in defining our taxonomy is that the former is based primarily on nonfoveal information and the latter is based primarily on foveal information. Consider the left panel in Figure 3. The first fixation is to a tile on the far right, which does not contain the target. This tile can thus be labeled a *nonfoveal false alarm* ($\bar{f}$ FA), since the incorrect decision that the target was in that tile was (presumably) based on peripheral information. Also, each tile except the central one and the one containing the target can be labeled a *nonfoveal correct rejection* ($\bar{f}$ CR), since the correct decision that the target was not in those tiles was also based on peripheral information, and the tile actually containing the target can be labeled as a *nonfoveal miss* ($\bar{f}$ Miss). Finally, when the eye moves to the subsequent tile (in the lower left), the tile at the first fixation can be labeled a *foveal correct rejection* (fCR), since the decision to reject this tile and continue searching was based on foveal information. Later in the trial, the observer actually fixates the tile containing the target, making that tile an $\bar{f}$ Hit, but then continues searching, so that tile also becomes an fMiss. If the observer had decided to remain on the tile containing the target, instead of continuing his search, this tile would have become an fHit. Trials in which the observer quickly finds the target and in which the observer never fixates the target are shown in the center and right panels, respectively.

Tiles were categorized postexperiment according to Table 1 for analysis. Note that for a given trial, each tile can belong to more than one category. As is shown in the table, each fixated tile was classified

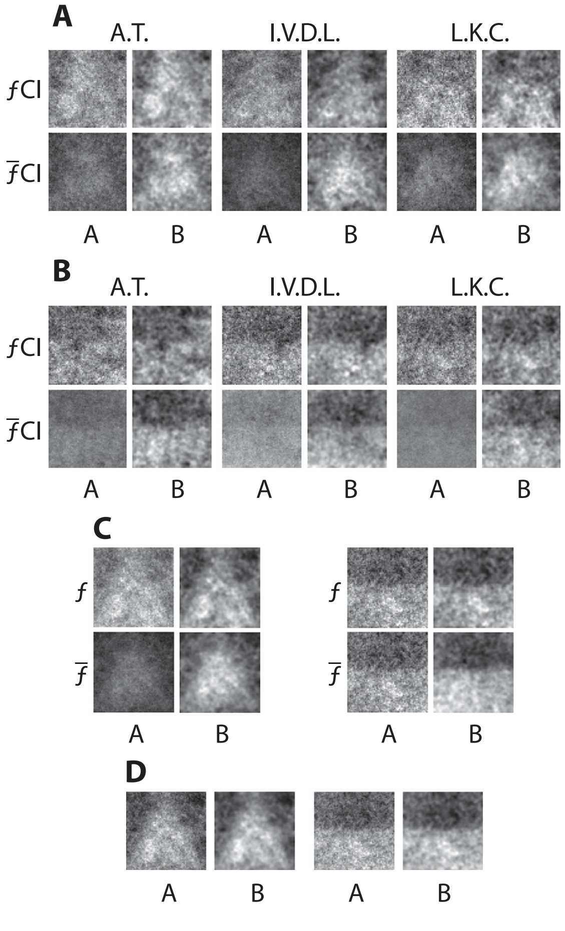

Figure 6. Classification images for (A) triangle and (B) dipole target search are shown for 3 observers. (C) Foveal and nonfoveal classification images combined across observers. (D) Classification images combined across foveal and nonfoveal categories and across observers. Columns labeled "A" contain the raw images collectively scaled to a single common grayscale color map, and columns labeled "B" contain the raw images after low-pass filtering and individual contrast enhancement.

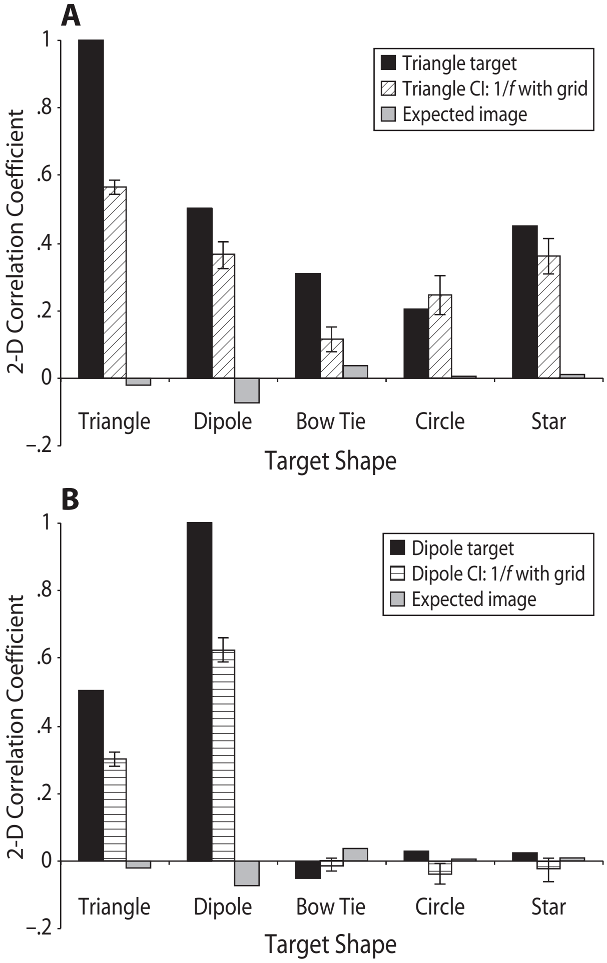

Figure 7. Zero-lag 2-D correlation coefficients showing the structural similarity (A) between the classification images for the triangle search and each of the test shapes and (B) between the average classification image for the dipole search and each of the test shapes. Error bars show the standard errors of the correlations across observers and categories (foveal and nonfoveal).

as an $\bar{f}$ Hit or an $\bar{f}$ FA, depending on whether the tile contained the target or not. The tile was then additionally classified as belonging to one of the foveal categories, depending on the observer's response: either maintaining fixation on the tile, indicating that he thought the target was there, or continuing the search, indicating that he thought the target was elsewhere. Tiles not fixated were classified as $\bar{f}$ Miss (target present) or $\bar{f}$ CR (target absent).

Generating the average images and the classification images. Pixel-by-pixel averaging of images within each category was used to obtain the average noise images corresponding to that category. It is important to keep in mind that only the noise patches are used as input to this process, and *not* the target. Any structure revealed through these methods therefore originates from the influence of particular samples of noise on the observers' responses.

The average noise tiles were combined in the usual manner (Hit + FA − Miss − CR; Ahumada, 2002) to create the classification images, but this was done separately for our foveal and nonfoveal categories.

Because we used a finite number of noise tiles (49 × 100 = 4,900), the expected average image that would result by randomly sampling tiles is not uniformly zero but, rather, is the average of all the tiles. This expected image, corresponding to a null hypothesis that an observer does not use spatial structure in the tiles to select fixation points, is shown in Figure 4A. As one might expect, it is very flat (with a standard deviation of just .0015 on a 0-to-1 scale) but does contain some spatial structure, which can be made clearer by contrast stretching (Figure 4B), and blurring (Figure 4C). This overall average can be thought of as the bias each pixel has as a result of using a finite number of noise samples. Although the spatial structure in this overall average does not closely resemble the search targets (and we have quantified this assertion by calculating comparative 2-D correlation coefficients for each of our experimental targets; see Figure 7), we must be aware that any average noise image or classification image resembling this expected image does not possess a significant structure of its own.

RESULTS

Average/Classification Images

The pixel-by-pixel averages of the noise tiles in each of the eight categories are shown for each observer in Figure 5. Columns labeled "A" contain the raw average images collectively scaled to a single common grayscale color map, and columns labeled "B" contain the raw images after low-pass filtering (using a 3 × 3 pixel Gaussian mask with $\sigma = 0.9$ pixel) and individual contrast enhancement. The former shows the relative fidelity of the average image from each category, and the latter reveals possible structures present in each of the classification images. All the categories presented some target-dependent spatial structure, except for $\bar{f}$ CR, which converged to the overall average shown in Figure 4. $\bar{f}$ Hit, $\bar{f}$ FA, fHit, fFA, and fCR all show features associated with the target, whereas both $\bar{f}$ Miss and fMiss present features anticorrelated with the target. In the future, more accurate pixel weights could be obtained by applying a foveation algorithm (e.g., Geisler & Perry, 1998; Lee & Bovik, 2003) to the stimuli at each fixation point prior to computing the nonfoveal average images and classification images, to attenuate higher spatial frequencies outside the acuity range of the visual system in a space-variant fashion at each fixation. In this article, however, we will confine ourselves to simple averaging of unfoveated patches, in order to illustrate our basic method.

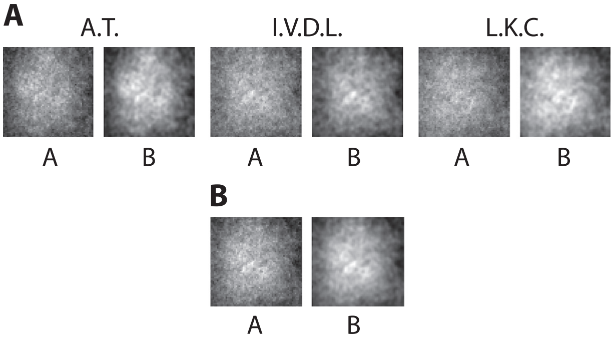

Figure 8. Classification images for triangle target search in 1/*f* noise, without a stimulus grid, are shown for 3 observers (panel A), and the combined classification images are also presented (panel B). Columns labeled "A" contain the raw classification images after individual contrast enhancement, and columns labeled "B" contain the raw images after low-pass filtering and individual contrast enhancement.

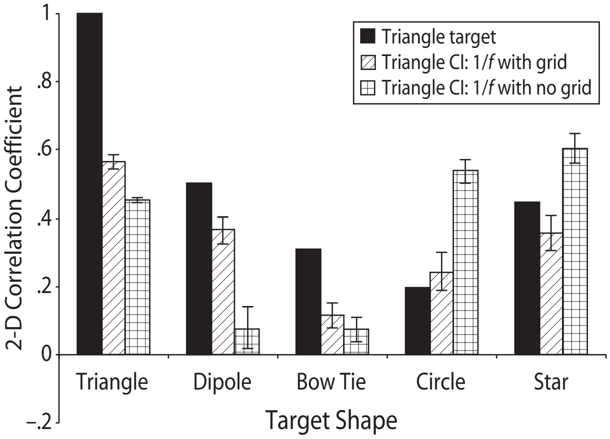

Figure 9. Zero-lag 2-D correlation coefficients showing the structural similarity between the classification images for the triangle search and those for each of the test shapes, comparing the main experiment with the no-grid control experiment. Error bars show the standard errors of the correlations across observers.

The foveal and nonfoveal classification images, fCI and $\bar{f}$CI, obtained by linearly combining the average images in the four response categories (defined in our classification taxonomy) in both the foveal and the nonfoveal cases, are shown in Figures 6A and 6B. Both foveal and nonfoveal classification images were created for each observer and each target. As is shown in Figure 5, columns labeled "A" contain the raw average images collectively scaled to a single common grayscale color map, and columns labeled "B" contain the raw images after low-pass filtering and individual contrast enhancement. These foveal and nonfoveal classification images provide cleaner target-like features.

Average images for both target types in the foveal and nonfoveal categories, averaged across all 3 observers, are shown in Figure 6C. The combined classification image, obtained by averaging the foveal and nonfoveal classification images across observers, is shown for each target type in Figure 6D. These combined classification images obviously show a strong resemblance to the sought targets.

The level of structural similarity between the classification images (shown in Figures 6A and 6B) and the search targets was quantified by computing the zero-lag 2-D correlation coefficients between them and the set of shapes in Figure 1. The correlation coefficients obtained, averaged across observers and the categories (foveal and nonfoveal), are shown by the hatched bars in Figures 7A and 7B for both the triangle and the dipole classification images. Also shown are the coefficients obtained by computing the correlation between the search target and each of the shapes (black bars) and the coefficients obtained by computing the correlation between the expected image (shown in Figure 4) and each of the shapes (gray bars). The error bars show the standard errors of the coefficients across observers and categories (foveal and nonfoveal). Note that the correlations are highest when computed between a classification image generated from a particular target (the triangle in panel A and the dipole in panel B) and that target itself. Moreover, the patterns of the experimental correlation coefficients (hatched bars) are virtually identical to those obtained using the targets themselves, rather than the classification images (black bars). These results show that our technique produces classification images that rapidly converge to relatively high fidelity representations of the pixel weights used by the observers and that, in this case, these weights strongly resemble the actual targets.

Control Experiments

Implementation without a grid. To show the effect of dividing the stimulus into a grid of tiles and using the accompanying taxonomy, we simply repeated the experiment without the grid, as was done in earlier work pioneering the use of eye tracking with classification images (Rajashekar et al., 2002). In this version of the technique, the actual location of each fixation is computed, and the 64 × 64 pixel patch of the stimulus noise surrounding each fixation is sampled and stored. The resulting set of noise patches is then simply averaged to form the *classification images* for each observer.[3] These are shown in Figure 8A; columns labeled "A" contain the classification images for the triangle target search after individual contrast enhancement, and columns labeled "B" contain the raw images after low-pass filtering and individual contrast enhancement. The combined classification image obtained by averaging the classification images across observers, is shown in

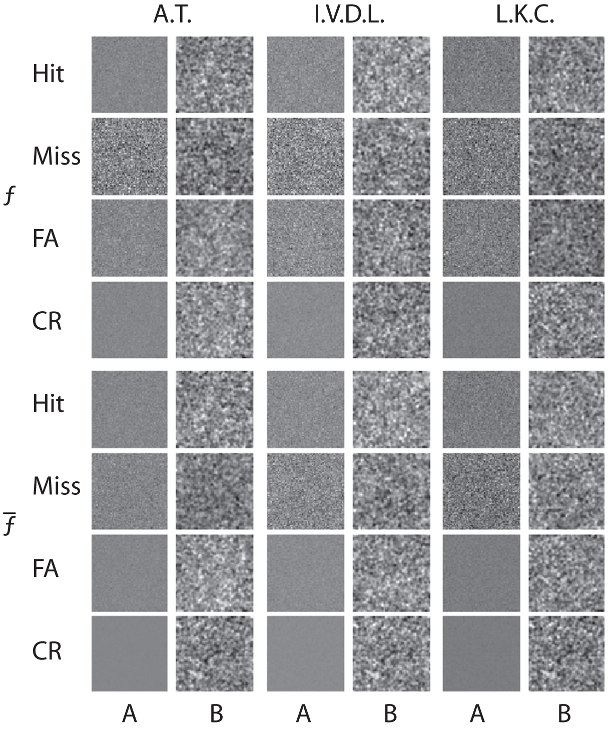

Figure 10. The average images for the triangle target search in white noise are shown for 3 observers. Columns labeled "A" contain the raw average images collectively scaled to a single common grayscale color map, and columns labeled "B" contain the raw images after low-pass filtering and individual contrast enhancement.

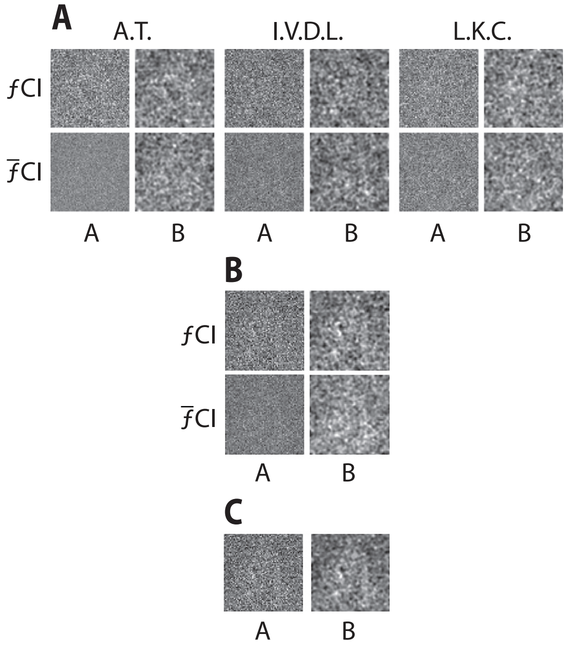

Figure 11. Classification images for the triangle target search in white noise are shown (A) for 3 observers, (B) combined across observers for foveal and nonfoveal categories, and (C) combined across observers and foveal and nonfoveal categories. Columns labeled "A" contain the raw images collectively scaled to a single common grayscale color map, and columns labeled "B" contain the raw images after low-pass filtering and individual contrast enhancement.

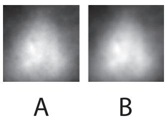

Figure 12. Combined classification images for uniform noise (A) after being pinkened and (B) after being pinkened, low-pass filtered, and contrast stretched.

Figure 8B. Although there does appear to be some spatial structure in these images, it seems less specifically triangular than that seen in Figure 6A. This was confirmed by doing the same correlation analysis as that just described, the results of which are shown in Figure 9. The black and hatched bars show the target/shape and raw classification-image/shape correlations replotted from Figure 7A, and the quilted bars show the correlations obtained using the classification images shown in Figure 8. Not only is the pattern of correlations across shapes different, but also the actual target used (the triangle) produced a substantially lower correlation with the classification images than did two of the other shapes (the circle and the star).

Implementation with white noise. 1/*f* noise approximates the spectral distribution of natural scenes, making it a valuable tool for probing search behavior within a statistically natural visual environment. Despite this important benefit, the presence of spatial correlation in 1/*f* noise leads to classification images that do not correctly estimate the linear *independent* contribution of each pixel to an observer's behavior, since the noise itself is already spatially correlated. In this control experiment, we showed that because information actually determining the observer's behavior exists predominantly at low spatial frequencies (presumably), the classification images converge to a similar degree, regardless of whether 1/*f* noise is used or whether another noise type (such as white noise) is postprocessed to amplify lower frequencies after the experiment has been completed.

To compare the classification images derived from 1/*f* noise with those derived from white noise, we simply repeated our procedure, using uniform white noise,[4] with 200 trials and the same 3 observers. Figure 10 shows the resulting data in the same format as that in Figure 5. Visual comparison of the two figures indicates an apparent lack of spatial structure in the white noise case when processed for viewing as before, with low-pass filtering and contrast enhancement. The foveal and nonfoveal classification images are shown for each observer in Figure 11A and averaged across observers in Figure 11B, and the combined classification image is shown in Figure 11C. Features of the triangle target are present but comparatively faint in the average images. Some spatial structure emerges in the combined classification image, but it is unclear without further processing (see below).

To effect a fairer comparison, we pinkened our white noise stimuli and recalculated the classification images. We then compared these pinkened classification images with those obtained directly from the 1/*f* noise stimuli. The pinkening procedure was derived from the computation of the unbiased estimate described by Abbey and Eckstein (2002). This procedure involves multiplying each noise image by the covariance matrix of the 1/*f* noise (computed to within an arbitrary scaling factor) given by $B * B^T$, where B represents the 1/*f* blurring filter and ^{T} the matrix transposition. The classification images combined across foveal and nonfoveal categories and across observers are shown in Figure 12. Panel A shows

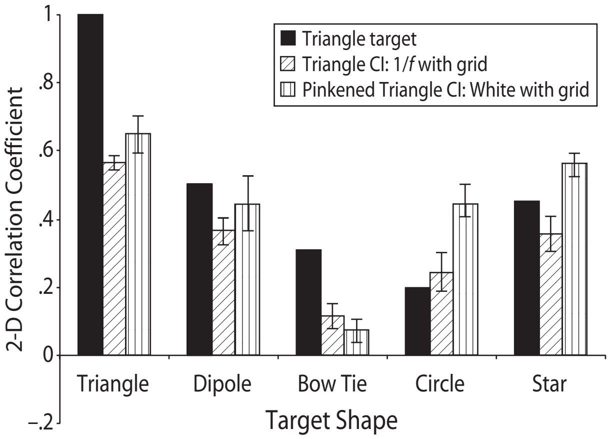

Figure 13. Zero-lag 2-D correlation coefficients showing the structural similarity between the classification images for the triangle search and each of the test shapes, comparing the main experiment with the (pinkened) white noise control experiment. Error bars show the standard errors of the correlations across observers.

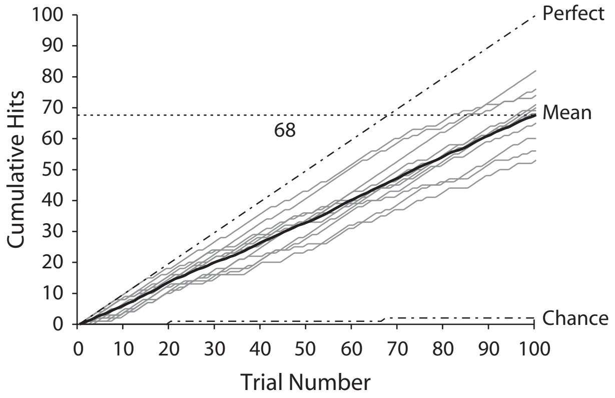

Figure 14. Graph of observer performance over time, measured as cumulative number of hits.

the raw result obtained with the pinkened white noise, and panel B shows the low-pass filtered and contrast stretched version. Again, our results indicate that preblurring (using 1/*f* noise stimuli) or postblurring (pinkening uniform noise post hoc) produces closely comparable results, evidenced by the correlation analysis shown in Figure 13. The black and hatched bars show the target/shape and raw classification images/shape correlations replotted from Figure 7A, and the striped bars show the correlations obtained using the classification images shown in Figure 11.

These results demonstrate that the use of 1/*f* noise (effectively, the preblurring of stimuli to make them more naturalistic) does not impact the resultant classification images too dramatically, although we might expect that the strategies human observers employ in different noise conditions may be modulated accordingly. White noise has the benefit of uncorrelated pixel values, leading to an unbiased estimate of each pixel's importance, but the disadvantage of possessing statistics that deviate from natural images, in comparison with 1/*f* noise. Therefore, important insights may be gained using other noise types, particularly 1/*f*, when we seek to study more naturalistic behavior.

Performance Measures

In general, classification images are valuable insofar as observers do the same thing on each trial. If an observer switches back and forth between two strategies, say, the pixel weights will reflect the linear combination of the two, with no way to disentangle them. Our task is slightly more complicated than those used in traditional psychophysics. We therefore wanted to ensure that the observers' performance remained roughly constant across trials and did not depend on the target location (i.e., the initial target eccentricity). Although this is not a direct measure of strategy, a change in strategy would probably be accompanied by a change in performance.

Performance over location and time. Figure 14 shows the cumulative number of hits as a function of trial number, obtained for the 3 observers with two sets of 100 trials for each of the two search tasks (triangle and dipole target search) in the basic 1/*f* noise, with grid, experiment. The mean cumulative hit number is represented by the thick black curve, and it reaches the 68% rate sought during the QUEST procedure at the final trial. These performances are compared with that of a perfect observer (dashed curve labeled as perfect) and with that of a random observer (dashed curve labeled as chance).

Each set of 100 trials yielded a slope between roughly 0.5 and 0.8, and for each set, this slope was roughly constant throughout. Again, this is not direct evidence that the observers did not change strategies, but it does indicate that a constant level of performance was maintained within a session.

We also measured the success rates of the observers in four different initial eccentricity regions covering the full stimulus, the center tile (Zone 1) and three concentric square annuli surrounding the center tile (Zones 2–4), to see whether the location of the target had any influence on the performance. Because these zones were square, they in-

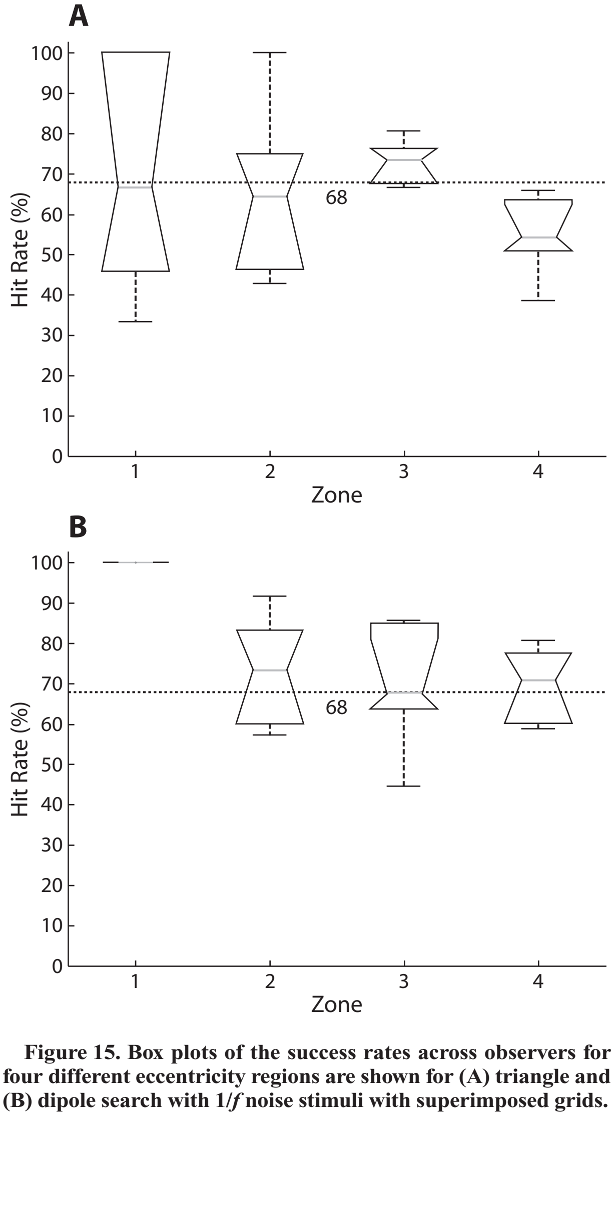

Figure 15. Box plots of the success rates across observers for four different eccentricity regions are shown for (A) triangle and (B) dipole search with 1/*f* noise stimuli with superimposed grids.

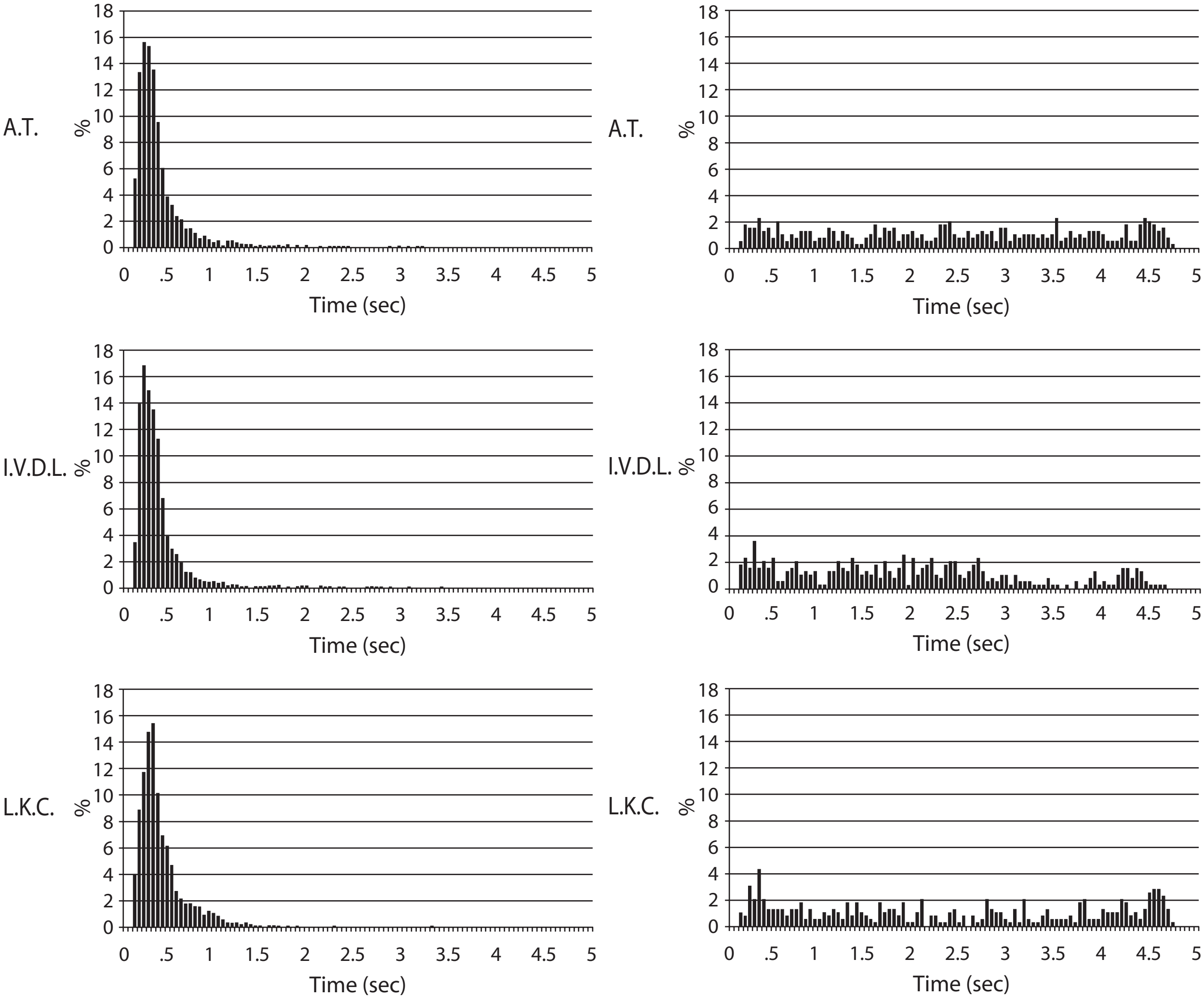

Figure 16. Dwell time distribution for each observer: (A) Nonfinal and (B) final fixations.

clude tiles centered at eccentricity ranges of 0º, 1.19º–1.68º, 2.38º–3.36º, and 3.56º–5.04º. Figure 15 shows the comparison of box plots of the success rates in the four different initial eccentricity regions, for sets of 100 trials performed by the observers for both targets: (A) triangle and (B) dipole. The only obvious aberration in the data is that the dipole target was always detected when presented in the central tile, presumably because this target at this location results in the edge's being presented directly to the foveola. The triangle target was also more difficult to detect when presented in the outermost tiles, but not dramatically so.

Observer dwell times. As was discussed in the Method section, the observers were given 5 sec to find the target, in order to ensure a fairly consistent strategy across observers, allowing several fixations to be made per trial but precluding the possibility of an exhaustive search. Figure 16A shows the distribution of the dwell times from the main experiment for all the fixations, excluding the initial and the final ones for each trial, for all the observers and both target types. It can be seen that the dwell times are concentrated mainly between 200 and 600 msec, in accordance with previous studies (Jacob, 1995). Figure 16B shows the distribution of the dwell times for only the final fixations. Over 83% of the dwell times observed for the final fixations are equal to or longer than 600 msec, the upper bound on typical fixation durations, reaching 95% for cases in which the target is actually found. We interpret this observation as indicating that, in our experiment, search was fairly naturalistic and that there was enough time for the observers to deliberately select a single tile as containing the target on most trials. For greater rigor in ensuring that the final fixation categories (*f*Hit, *f*FA) do not contain search fixations, one could eliminate those with dwell times below a threshold.

DISCUSSION AND CONCLUSIONS

In this article we have demonstrated a technique for expediting the convergence of classification images in visual search experiments. In fact, for each of the 3 observers and two target types, and with only 200 trials per observer, we see that the classification images obtained with our method closely resembled the target sought (Figures 5–7). Although the number of tiles falling into many of the categories was small, we still managed to obtain fairly distinctive average images and, hence, convincingly robust classification images. Stronger classification images were obtained than were obtained with a nongrid control experiment (this claim is supported by both a visual inspection of the results and the strength of the correlation coefficients obtained between the classification images and the targets). The use of naturalistic 1/f masking noise was evaluated with a second control experiment in which white noise was used. Visual inspection and correlation coefficients indicate that there is a minimal difference between classification images generated with either noise type if we either pinken white noise tiles and compare with 1/f noise tiles or whiten 1/f noise tiles and compare with white noise tiles.

In addition, we have introduced a new taxonomy for the categorization of results from each fixation during a trial. This new taxonomy simply extends the conventional signal detection theory categories to distinguish foveal and nonfoveal processes. However, this extension should allow us and others to characterize the kinds of information used in the fovea and periphery during naturalistic visual search. For instance, Figure 5A shows blob-like average images across observers for the nonfoveal category $\bar{f}$ FA—hence, characterizing the features that attracted observer fixations to tiles not containing the target. But as outlined in our taxonomy, noise images in the $\bar{f}$ FA category are divided into two foveal categories: fFA (corresponding to the observer's final selection of a wrong candidate) and fCR (corresponding to a rejection of a wrong candidate). In fact, fFA presents sharper target-like features, in comparison with fCR and $\bar{f}$ FA. Although preliminary, such results hint at the difference between the foveal and the nonfoveal selection processes. Moreover, stimuli could be filtered to take into account the eccentricities of the tiles with regard to the fixation points (foveation), prior to averaging, thus eliminating any contribution of spatial frequencies outside the pass band of the visual system at a given eccentricity.

AUTHOR NOTE

This research was funded by NSF Grants ECS-022545 and ITR-0427372. We acknowledge Wilson S. Geisler and Eyal Seidemann for their useful suggestions. We also thank Umesh Rajashekar for his valuable comments and for providing several of the MATLAB functions on which the data capture software was built. We gratefully acknowledge the helpful and constructive comments made by Richard F. Murray and two other anonymous reviewers. Correspondence concerning this article should be addressed to A. Tavassoli, Center for Perceptual Systems, 1 University Station, A8000, University of Texas, Austin, TX 78712 (e-mail: atavasso@ece.utexas.edu).

REFERENCES

Abbey, C. K., & Eckstein, M. P. (2002). Classification image analysis: Estimation and statistical inference for two-alternative forced-choice experiments. *Journal of Vision*, **2**, 66-78.

Ahumada, A. J., Jr. (1996). Perceptual classification images from Vernier acuity masked by noise. *Perception*, **25**, 18.

Ahumada, A. J., Jr. (2002). Classification image weights and internal noise level estimation. *Journal of Vision*, **2**, 121-131.

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*: *Proceedings of SPIE* (Vol. 3299, pp. 79-85). Bellingham, WA: SPIE.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10**, 433-436.

Eckstein, M. P., & Ahumada, A. J., Jr. (2002). Classification images: A tool to analyze visual strategies. *Journal of Vision*, **2**, 1.

Eckstein, M. P., Shimozaki, S. S., & Abbey, C. K. (2002). The footprints of visual attention in the Posner cueing paradigm revealed by classification images. *Journal of Vision*, **2**, 25-45.

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*, **4**, 2379-2394.

Geisler, W. S., & Perry, J. S. (1998). A real-time foveated multiresolution system for low-bandwidth video communication. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human Vision and Electronic Imaging III: Proceedings of SPIE* (Vol. 3299, pp. 294-305). Bellingham, WA: SPIE.

Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioral receptive fields for visually completed contours. *Current Biology*, **10**, 663-666.

Jacob, R. J. K. (1995). Eye tracking in advanced interface design. In W. Barfield & T. A. Furness (Eds.), *Virtual environments and advanced interface design* (pp. 258-288). New York: Oxford University Press.

Lee, S., & Bovik, A. C. (2003). Fast algorithms for foveated video processing. *IEEE Transactions on Circuits and Systems for Video Technology*, **13**, 149-162.

Neri, P., & Heeger, D. J. (2002). Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nature Neuroscience*, **5**, 812-816.

Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature*, **401**, 695-698.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, **10**, 437-442.

Rajashekar, U., Cormack, L. K., & Bovik, A. C. (2002). Visual search: Structure from noise. In *Proceedings of the 2002 Symposium on Eye Tracking Research and Applications* (pp. 119-123). New York: ACM Press.

Rajashekar, U., Cormack, L. K., & Bovik, A. C. (2004). Point of gaze analysis reveals visual search strategies. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human Vision and Electronic Imaging IX*: *Proceedings of SPIE* (Vol. 5292, pp. 296-306). Bellingham, WA: SPIE

Simoncelli, E. P. (2002). Seeing patterns in noise. *Trends in Cognitive Sciences*, **7**, 51-53.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, **33**, 113-120.

NOTES

[1] For an explanatory discussion of the classification image technique, see Eckstein and Ahumada (2002) and Simoncelli (2002).

[2] We use the term *foveal* to refer to a central patch 1º of visual angle across and *nonfoveal* to refer to regions outside this patch.

[3] This is not strictly a classification image but can be thought of as the average spatial structure that was fixated by the observer.

[4] We used uniform, rather than Gaussian, noise because of higher RMS contrast; at a 68% correct SNR in our task, Gaussian noise would have been substantially clipped at the tails.

(Manuscript received September 16, 2004;
revision accepted for publication April 5, 2006.)