The LNL (linear, non-linear, linear) model has previously been successfully applied to the problem of texture segmentation. In this study we investigate the extent to which a simple LNL model can simulate human performance in a search task involving a target on a textured surface. Two different classes of surface are considered: 1/\(frak{n}\)-noise and near-regular textures. We find that in both cases the search performance of the model does not differ significantly from that of people, over a wide range of task difficulties.

**Keywords:** linear-nonlinear-linear model, visual search, texture


---

### Introduction

Previous work on modeling visual search can be divided into two classes: search among sets of search items; and search in more naturalistic, continuous stimuli (typically photographs). The search task we consider in this study is that of a single target on a textured surface (Clarke, Green, Chantler, & Emrith, 2008). These stimuli have several advantageous properties: they offer a more naturalistic search task than sets of abstract geometric items; yet they are also fully parameterized and created pseudo-randomly so we can create many trials of equivalent difficulty, something that is difficult to do with photographic stimuli.

Najemnik and Geisler have considered similar, related stimuli: a sinusoidal target embedded in \(1/\frak{n}\)-noise (Najemnik & Geisler, 2008). They found that an Ideal Observer, based on empirically derived visibility maps, gave a good account of human performance. The study presented here takes an alternative, complementary approach. Najemnik and Geisler assumed the existence of an image processing stage that provides noisy information about the location of the target, and then modeled the search strategy that uses this information. In contrast, our primary interest is in modeling the initial processing of the image, while assuming a very simple search strategy.

Wolfe’s Guided Search (GS) is commonly used as a framework when modeling visual search (Cave & Wolfe, 1990; Wolfe, 1994; Wolfe & Gancarz, 1996). Guided Search is a theoretical model that uses low-level, pre-attentive features (such as orientation, size, color, etc.) to construct feature maps. These feature maps are then combined with top-down knowledge about the target resulting in an activation map in which regions of high activation correspond to items which are similar to the target and vice versa. This activation map is used to guide a serial search process which passes items to an object recognition stage (also referred to as an attention gate). Two examples of models based on Guided Search are the Area Activation Model (Pomplun, Reingold, Shen, & Williams, 2000; Pomplun, Shen, & Reingold, 2003) and the Probabilistic Model (Rutishauser & Koch, 2007). Both of these models simulate human gaze patterns while searching for a target among distracters. As abstract stimuli are used, category based feature maps are trivial to construct.

When we move to the task of modeling search on more naturalistic stimuli we encounter the problem of deciding what features to use, and importantly, how to measure them. We know from visual search experiments using discrete items that a target’s saliency is governed by factors such as target-distracter similarity and the heterogeneity of the distracters in terms of basic features such as color, orientation and size (see Wolfe, 1998, for a review). However, while it is easy to measure these properties for sets of discrete search items, transferring them over to image based stimuli is not a trivial matter. A recent study by Pomplun (2006) has investigated the degree to which search on photographic stimuli is guided by four features: orientation, spatial scale, intensity and contrast. He found that there is significant guidance by all these features although bottom-up effects were also present.
A possible starting point for such work is to use a saliency algorithm (Gao, Mahadevan, & Vasconcellos, 2008; Itti & Koch, 2000). Unlike GS these saliency models are computational, in that they can be used to generate a saliency map for any given image. Pre-attentive feature maps (color, illumination contrast and local orientation) are computed over several spatial scales. A sequence of iterative inhibition algorithms and summations (cross-scale and cross-feature) are used to combine these maps, resulting in a two dimensional saliency map.

Saliency algorithms can be easily used as visual search models: in fact the initial empirical tests of Itti and Koch’s model were visual searches. Their model has been shown to exhibit human-like behavior for feature and conjunction searches using red/green and horizontally/vertically orientated bars (Itti & Koch, 2000). The model was also compared to human performance in a search task involving photographs of landscapes. As an eye-tracker was not employed, a conservative estimate of three saccades per second was used to compare human performance with the model. Itti and Koch found a poor correlation between human and computer performance with the model outperforming the human observers on the majority of the trials. A recent study (Navalpakkam & Itti, 2007) used Itti and Koch’s algorithm to investigate different top-down feature weighting mechanisms. They concluded that features should be weighted in order to maximize the target to distracter signal to noise ratio. This differs from most other models which weight features based on their similarity to the target.

A different approach was taken by Rao, Zelinsky, Hayhoe, and Ballard (2002). They used a Gaussian filter and its orientated derivatives to construct feature vectors for every location on the stimulus image. These vectors were compared to the response vector from the target and the L2 difference between the two was used as a saliency map. In order to simulate the human behavior reported in Zelinsky, Rao, Hayhoe, and Ballard (1997), the model accesses finer spatial scales on each fixation. This mechanism allows the model to generate scan-paths very similar to those seen in the human data. However, due to the nature of the search task, the human scan-paths show much less variation than is typically seen in visual search tasks. Furthermore the authors restricted their analysis to trials in which the target was found in three fixations.

Modeling visual search in naturalistic stimuli is a difficult problem. While abstract displays contain clearly defined sets of search items, more general stimuli do not. Even when photographic stimuli contain clear distinctions between search items and backgrounds, we do not yet understand the relationship between target-distracter and background similarities and how they affect task difficulty (Wolfe, Oliva, Horowitz, Butcher, & Bompas, 2002; Zelinsky et al., 1997). The problem becomes more difficult when we consider camouflage backgrounds (Neider & Zelinsky, 2006) where the distracter items are salient but the target is not.

In order to tackle this problem we suggest starting with simple stimuli consisting of a target on a textured background. This allows us to test features and algorithms on their own without having to consider high-level problems such as object recognition and target/background segmentation. In a previous paper we explored how human subjects perform in a search task involving rendered 1/fβ-noise surfaces (Clarke et al., 2008). In each trial the target—a small indent in the surface—was either present or absent and the observer had to determine which. We found that by changing the perceived roughness of the surfaces and the orientation of the target, the difficulty of the search task can by varied in a controlled way from pop-out (1–2 saccades required to find the target) to difficult (<50% accuracy). A comparison with Itti and Koch’s saliency algorithm was also carried out. While the absolute number of saccades between the model and the human observers did not match, both responded in a similar way to increasing surface roughness. However, the model was far more sensitive to changes in orientation than the human observers and had difficulty in finding the target when it was ±20° from the (vertical) illuminant vector, while people only had difficulty when the target was oriented ±5°.

In this study we investigate the extent to which a new search model, based upon the common LNL (linear-nonlinear-linear) framework, can match human performance. LNL models (also known as FRF or filter-rectify-filter-models) are based on properties of the functional architecture of primary visual cortex, and have been applied successfully to problems of texture discrimination and classification (Bergen & Landy, 1991; Bovik, Clark, & Geisler, 1990; Chantler, Petrov, Perischke, Schmidt, & McGunnigle, 2005; Landy & Oruc 2002; Malik & Perona, 1990; Morrone & Burr, 1988; Randen & Husoy, 1999). However our aim is to build a model of this kind with as simple properties as possible and to determine whether it responds to changes in the difficulty of search tasks in the same manner as human observers.

To avoid having to consider speed-accuracy trade-offs we will use a new search task in which a target is present on every trial and the observer simply has to press a button once they find it. This means we can measure performance just by examining the number of fixations required to find the target. In addition to comparing model and human performance with the stimuli used in the previous paper, we also introduce a second class of stimuli: near-regular textures (see Figure 1 for an example). The 1/fβ-noise surfaces have random phase spectra and are essentially structureless, and the target in the search task is a small ellipsoidal indent which can be recognized by a combination of shape-from-shading and the high contrast edge due to illumination. The difficulty of the search task is varied by changing the surface’s perceived roughness and the target’s orientation. The near-regular textures on the other hand are highly structured, consisting of an array of ellipsoidal textons. In this case we created a
search target by removing one of the textons. For these surfaces we investigate how human, and the model’s, performance responds to changes in regularity and density of the textons.

The model

The model presented here consists of two parts: an LNL-model, which is used to generate an activation map, and a search algorithm which uses the activation map to generate saccades.

Activation map

The first stage is linear and consists of a bank of Gabor filters. As our stimuli are grayscale we do not consider color channels. A dedicated contrast channel was not used, as summing the Gabor responses for a given spatial scale approximates the response of the band-pass filter, which is commonly used to compute illumination contrast. Based on the results of pilot studies we used eight equally spaced orientation channels.

Gabor filters have parameters $\sigma_u, \sigma_v, u_0$ and $\varphi$:

$$G(u, v) = \exp\left(-\frac{1}{2} \left(\frac{(u - u_0)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}\right)\right), \tag{1}$$

where $u_0 = (60 \cdot 1.8^{n+3})/1024$ cpd, $\sigma_u = \sigma_v = 2'$ and $r = 1, \ldots, 7$ is the spatial scale. Eight orientation channels were used: $\varphi = n \cdot \pi/8$, where $n = 0, \ldots, 7$.

The second stage is non-linear. First, it rectifies the negative responses from the filter and second, it applies a weighting/scaling to the feature maps to deal with signal to noise problems. We square each filter response before normalizing to [0, 2]. Each resulting map is then divided by its median pixel intensity. This means that maps with a small number of local maxima will have a relatively low median value, $<1$, and hence their peaks will be emphasized relative to maps containing no strong peaks.

Finally a smoothing filter is applied to each map. If the smoothing filter is weak then the model will consider local maxima as saccade targets; if a stronger smoothing filter is used then the model will fixate centers of gravity. We used a two dimensional Gaussian with $\sigma_u = \sigma_v = 3.75$ cpd. Finally all the feature maps are summed across scales and orientations to give the activation map, $S$. Note that if we were to remove the normalization and the median division function from the 2nd stage to leave only the square then this would be equivalent to a simple local energy estimator. However, we are only interested in detecting small regions differing in energy content from their backgrounds, hence the use of the [0,2] scaling and median division (Figure 2).

Selecting where to fixate next

The model generates saccades using a three stage process. First a negative exponential mask is used to weight the baseline activation map to give:

$$F_d(x, y) = S(x, y) \cdot e^{-kd}, \tag{2}$$
where \( d \) is the Euclidean distance between \((x, y)\) and the current fixation location (in pixels), and \( k \) is a constant \((k = 0.0013)\). This process is a simple approximation of foveal vision. A more rigorous implementation would involve weighting each spatial scale separately, although work by Peters, Iyer, Itti, and Koch (2005) suggests this offers little improvement over the method used here.

This is followed by a simple inhibition of return (IOR) mechanism, implemented as a series of two-dimensional masks. Gaussian masks are used for this, which decay over time:

\[
F(x, y) = F_d(x, y) \prod_{t=1}^{n} \left( 1 - \frac{2}{t+1} I_t(x, y) \right),
\]

where \( t \) is the fixation number; \( I_t(x, y) \) is a two dimensional Gaussian mask (normalized to \([0,1]\)) centered at \((x_t, y_t)\) with \( \sigma = 45 \) pixels.

Results from a pilot study suggested, unsurprisingly, that a deterministic model that fixates the maxima of the feature map (after applying the fixation dependant processing) performs too well and is not able to respond to the stimuli in the same way as human subjects. Therefore we use a simple stochastic process: we consider the three largest local maxima in the resulting fixation map, \( F(x_i, y_i) \), and assign them probabilities:

\[
p_i = \frac{F(x_i, y_i)}{\sum_{i=1}^{3} F(x_j, y_j)}.
\]

These are then used to choose which of the three largest maxima will be selected as the target for the next saccade. The model will continue to make saccades until either it is fixating within 1° of the target or a maximum cut-off limit, of 300 fixations, is reached.

**Methods**

In order to test the model against human performance we carried out a target always present visual search task. How long people are prepared to search difficult target present/absent trials varies from person to person and depends on factors such as observer tiredness and the ratio of target present to target absent trials. By only considering target present trials we reduce the interpersonal variance and we can consider modeling the search process separately from the decision process.

**Stimuli—Experiment 1a**

We used rendered uniform albedo \( 1/f^\beta \) random phase surfaces for stimuli (Chantler, 1995). These surfaces are controlled by two parameters: frequency roll-off \( \beta \) and RMS-roughness \( \sigma_{\text{RMS}} \). Together these two parameters control the perceived roughness of the surface (Padilla, Drbohlav, Green, Spence, & Chantler, 2008). As we decrease \( \beta \) the amount of high frequency information in

![Figure 2. A small section of a surface texture with target, and the corresponding activation map. In this case, the global maximum of the activation map corresponds to the target. Further examples, including examples of the model’s intermediate stages, can be found in the additional materials.](https://jov.arvojournals.org/pdfaccess.ashx?url=data/journals/jov/933534/)
the height map increases, which makes it appear rougher. \( \sigma_{\text{RMS}} \) is a scaling parameter and higher values lead to rougher surfaces. The height maps were treated as Lambertian surfaces with constant albedo (0.8), similar to that of matt white paint. The light source is kept constant with elevation 30° and azimuth 90° (i.e. lit from above). For further technical details of the texture and rendering model used, see Clarke et al., 2008).

An ellipsoid was used to create an indent in the surface:

\[
\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1. \tag{5}
\]

In this experiment the target was a circular indent with \( a = b = 10, c = 2 \) pixels, and a volume of 50 pixels\(^3\). The perceived roughness of the surface was varied to change the difficulty of the search task (\( \beta = 1.6, 1.65, 1.7; \sigma_{\text{RMS}} = 0.9, 1.1 \)). For each trial a target was positioned randomly on a circle, centered on the middle of the image, with radius 1.7° ± 0.7°, 3.8° ± 0.7° or 5.9° ± 0.7° visual angle. [A Matlab .m file containing the texture synthesis model can be found at [1].]

### Stimuli—Experiment 1b

For the second part of this experiment we investigate how well the search model can simulate human search for an elongated target, \( a = 20, b = 5, c = 2 \), in a \( 1/\beta \)-noise surface. In this case, the orientation of the target with respect to the direction of illumination is important: as the target’s orientation changes from horizontal to vertical, the contrast decreases, causing the search task to increase in difficulty (see Figure 3). The target orientations used with respect to the horizontal axis were \( \theta = 0°, ±30°, ±45°, ±60°, ±75°, ±80° \). The illumination vector was kept constant at azimuth 90°. Target orientations 80° < \( \theta < 90° \) were not included as observers have difficulty identifying these targets (Clarke et al., 2008). Two different roughnesses were used: \( \sigma_{\text{RMS}} = 1, \beta = 1.6, 1.7 \) and the target was randomly placed with eccentricity \( 5° < r < 6.7° \).

### Stimuli—Experiment two

In order to test the generality of our model we also applied it to a different class of texture: near-regular textures. A regular texture is one which consists of a regularly repeating pattern, and a near-regular texture is a regular texture with a degree of randomness added. This can either be in size, shape and/or position of the textron elements (Liu, Collins, & Tsin, 2004; Liu, Lin, & Hays, 2004). We will also use a different type of target: a missing textron in the lattice. In contrast with the \( 1/\beta \)-noise textures used earlier, these surfaces are highly structured, periodic and the target is now defined by the lack of a high frequency edge. An example texture can be seen in Figure 1.

As with the \( 1/\beta \)-noise surfaces, the stimuli used in the second experiment are created by first synthesizing a height map and then rendering it using Lambert’s cosine law. The height map is created by placing ellipsoidal shaped textons at regular intervals. Two densities were used, \( \rho = 1.875 \) and \( \rho = 2.461 \) textons per degree. These textons were randomly varied in two ways to give an irregular texture: size and \( \sigma_p \), the amount of random offset from the lattice point. To vary the size, \( a \) and \( b \) were randomly set to 8, 9, 10 or 11 pixels in Equation 5. Offset was applied to each textron by adding a normally distributed error to its center point. By varying the standard deviation of this error, \( \sigma_p \), we can vary the regularity of the underlying lattice. Finally, a small amount of Gaussian noise (std. =0.25) was added to the phase spectrum in order to make the images appear more naturalistic.

Unlike the first experiment, where the target was a dent, in this experiment the target is a missing textron. As in Experiments 1a and 1b, the target was randomly located with eccentricity constrained to either 3.33° or 6.67°.

### Setup

Stimulus presentation was controlled by Clearview (Tobii Technology Inc). All stimuli were 1024 × 1024 pixels in size and displayed on a calibrated NEC LCD2090UXi monitor. The pixel dimensions were

![Figure 3](https://example.com/figure3.png)  
**Figure 3.** The orientation of the target, with respect to the illumination, affects its appearance. In all cases the target is a small indent, illuminated from above. When it is close to perpendicular to the illumination direction there is a high contrast edge. The contrast diminishes as we rotate the target towards vertical. **Note:** The image has been scaled up for illustrative purposes.
0.255 mm by 0.255 mm resulting in images with physical dimensions 261 mm by 261 mm. The monitor was linearly calibrated with a Gretag-MacBeth Eye-One; maximum luminance set at 120 cd/m². This results in the rendered images appearing as if they were being lit under bright room lighting conditions.

A Tobii ×50 eye-tracker was used to record observers’ gaze patterns. The fixation filter was set to count only those fixations lasting longer than 100 ms within an area of 30 pixels. The accuracy of the eye-tracker was 0.5°–0.7° and the spatial resolution was 0.35°. The viewing distance was controlled by use of a chin rest placed 0.87 m away from the display monitor. At this distance, one pixel is approximately 1° of visual angle; images subtend 16.7° and the targets subtend 0.66° of visual angle.

Running the search model

As our model of visual search is stochastic we ran it seven times to obtain a measure for the average number of fixations required to find the target. The same stimuli were used for both the human and computer vision experiments. The maximum number of saccades allowed for the model was set to 300; this allowed the model to find over 99% of the targets in both Experiments 1a and 2: which was comparable with human performance. The small number of trials on which the model failed were not included in any further analysis.

Results

Experiment 1a: Varying roughness

The aim of this experiment was to compare how well human observers and our LNL-based search model can find a small indent over a range of surface roughnesses.

The results show that the number of saccades required to find the target increases with roughness (i.e. increasing σRMS or decreasing β) and with target eccentricity (Figure 4). Observers managed to find over 99% of the targets. In the small number of trials in which the observer gave up, they had spent at least 2 minutes searching for the target.

All three variables had a significant effect on the mean number of saccades: for β, F(2) = 43.1, p = 0.001; for σRMS, F(1) = 71.8, p < 0.001; and for r, F(2) = 14.6, p = 0.008.

Figure 4. Comparison between human observers (blue lines) and the LNL search model (red lines). (left) σRMS = 0.9, (right) σRMS = 1.1.
Additionally, there was a significant interaction between the two parameters that controlled roughness, $\beta$ and $\sigma_{\text{RMS}}$: $F(2,1) = 30.6$, $p = 0.002$. There was no significant interaction between target eccentricity and either of the two roughness parameters. The distribution of saccade amplitudes and orientations is shown in Figure 5. (These are discussed later).

The mean number of saccades over all trials can be seen in Table 1. An independent $t$-test gives $t(12) = -1.381$, $p = 0.192$. Therefore, the null hypothesis, that the means of the human and model populations are equal, is accepted. **Figure 4** shows how the mean number of saccades taken by the human subjects and the model to find the target varies with surface roughness. We carried out a 4 way mixed model ANOVA on $\beta$, $\sigma_{\text{RMS}}$, $r$ and $\delta$ (which distinguishes between instances of the model and human subjects). As above, the ANOVA shows significant effects for $\beta$, $\sigma_{\text{RMS}}$, $r$. However, $\delta$ does not have a significant effect, $F(1) = 1.8944$, $p = 0.194$; and neither do its interactions: $\beta \times \delta$ has $F(2,1) = 0.427$, $p = 0.658$; $r \times \delta$ has $F(2,1) = 0.190$, $p = 0.829$; $\sigma_{\text{RMS}} \times \delta$ has $F(1,1) = 0.01$, $p = 0.921$. Similarly, three and four way interactions involving $\delta$ are also non-significant. Hence there is no evidence that the human observers and the model are affected differently by any of these parameters.

**Experiment 1b: Varying target orientation**

The overall means are shown in Table 1. Again, the means from each run of the model are close to the human means and an independent $t$-test does not detect any differences between the populations: $t(12) = -2.250$, $p = 0.64$. Neither the human observers nor the computer model are affected by changing the target’s orientation until it nears vertical, the direction of the illumination vector (see Figure 6). Once the target’s orientation is greater than $75^\circ$ there is a rapid rise in the number of saccades required to find the target. The effect is greater for $\beta = 1.6$ than for $\beta = 1.7$. As the effect of $\theta$ is not linear, and the variance is not uniform, it is not appropriate to carry out an ANOVA. However, it is clear from **Figure 6** that both the model and the human observers perform equally well when the target is easy to find, and both respond to increasing task difficulty in a similar manner. The only discrepancy between the two occurs when $\theta = 80^\circ$ at which point the model fails to match the mean human performance. This difference is not great and there is a large amount of variance between the human observers. We conclude that there is no evidence the model and human observers respond differently to changes in $\theta$, the target’s orientation, in this search task.

**Discussion**

The results from the two psychophysical experiments above agree with those in Clarke et al. (2008). As well as investigating human search on $1/f^\beta$-noise surfaces Clarke et al. also compared human performance with Itti and Koch’s saliency model. They found that the saliency model are affected by changing the target’s orientation until it nears vertical, the direction of the illumination vector (see Figure 6). Once the target’s orientation is greater than $75^\circ$ there is a rapid rise in the number of saccades required to find the target. The effect is greater for $\beta = 1.6$ than for $\beta = 1.7$. As the effect of $\theta$ is not linear, and the variance is not uniform, it is not appropriate to carry out an ANOVA. However, it is clear from **Figure 6** that both the model and the human observers perform equally well when the target is easy to find, and both respond to increasing task difficulty in a similar manner. The only discrepancy between the two occurs when $\theta = 80^\circ$ at which point the model fails to match the mean human performance. This difference is not great and there is a large amount of variance between the human observers. We conclude that there is no evidence the model and human observers respond differently to changes in $\theta$, the target’s orientation, in this search task.

**Table 1.** Mean number of fixations (SD) required to find the target, averaged over all trials.

<table>
<thead>
<tr>
<th></th>
<th>Exp 1a</th>
<th>Exp 1b</th>
<th>Exp 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observers</td>
<td>7.02 (1.33)</td>
<td>5.99 (2.01)</td>
<td>11.25 (2.57)</td>
</tr>
<tr>
<td>Model</td>
<td>7.80 (0.67)</td>
<td>7.71 (0.30)</td>
<td>10.24 (1.01)</td>
</tr>
</tbody>
</table>

**Figure 5.** Histogram showing saccade amplitudes and rose plot showing saccade directions for the perceived roughness experiment.
model responded to increasing roughness in a similar manner to the human participants, although the absolute number of saccades did not match. However, in the orientation experiment the performance saliency model fell steeply before that of humans as the target’s orientation approached vertical. The results presented here suggest that a simple LNL-based model, using a Gabor filter bank, offers a better match with human performance in this search task. The model manages to match the number of fixations needed by human observers to locate the target, and it successfully does this over a wide range of surface roughnesses and target orientations and eccentricities.

**Experiment 2: Near-regular textures**

In order to test the generality of our model we tested it on a different class of texture: near-regular textures. The results are shown in Figure 7. All three parameters affect the mean number of fixations required to find the target:

![Figure 6](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933534/) Mean number of saccades required to find the target for (left) $\beta = 1.7$ and (right) $\beta = 1.6$.

![Figure 7](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933534/) The results from the near-regular experiment for (left) $\rho = 1.875$ textons per degree and (right) $\rho = 2.461$ textons per degree.
for $\sigma_p$ we have $F(3) = 41.629, p < 0.001$; for $r$ we have $F(1) = 28.309, p = 0.002$ and finally for $\rho$ we have $F(1) = 25.698, p = 0.002$. There is also an interaction between the two parameters controlling the surface’s appearance, $\sigma_p$ and $\rho$, with $F(3,1) = 5.890$ and $p = 0.006$. As with the $1/f^\beta$-noise surfaces interactions involving $r$ are not significant.

We also investigated how well our search model works on these stimuli and found that the LNL-based model could successfully identify the target. Given that the search model was designed and tested for a very different class of stimuli, this is an encouraging result in itself. Comparing the mean number of saccades required for the model and human observers over all trials we find that there is no statistical difference, $t(12) = 0.826, p = 0.372$. (See Table 1.) However as we can see from Figure 7 the model behavior is different as we vary $\sigma_p$. While a 4-way between subjects ANOVA does not show a significant main effect for $\delta$, $F(1) = 2.702, p = 0.131$, the $\delta \times \sigma_p$ interaction is significant ($p < 0.001$).

**Discussion**

Across all conditions, our model makes a similar number of saccades as a human observer in our experiments. In both model and observers, there is a large increase in the number of fixations when high texton density is combined with high variability in texton placement. However, there are also differences in performance as the task parameters are varied; in particular, the observers were sensitive to the eccentricity of the target, requiring more fixations to find a more eccentric target, whereas this parameter does not affect the model. If we compare the saccade amplitude histograms for the $1/f^\beta$-noise experiment with the near-regular texture experiment, we see that the human observers are somehow changing their search behavior and are making far more saccades with amplitude 0.5–1 degrees (compare Figures 5 and 8). This difference appears to be independent of the parameters $\sigma_p$, $r$, $d$, and is exhibited for all subjects. This suggests that some feature of the stimuli not captured in our activation map is causing a change in search patterns, shown as an increased number of very short saccades. This pattern of search may be responsible for the effect of target eccentricity on human performance.

The aim of the above experiments was to determine whether a simple LNL-based search model can give a good account of human performance over a range of task difficulties. The first experiment involved searching for an indent on a rough, $1/f^\beta$-noise surface. In this case the model was able to find the target in the same number of saccades as human observers over a wide range of task difficulties. The model responded to changes in background roughness and in target orientation (with respect to the illumination direction) in the same way as the human observers. We then evaluated our model with a quite different surface-target combination: a near-regular texture with a missing lattice point. Here, there were differences in the way observers and model responded to...
some parameters of the task, particularly the eccentricity of the target, but the mean number of fixations to find the target, across all conditions, was the same in both cases.

While our model was designed to find targets in naturalistic images it can also be applied to search tasks where the targets and distractors are discrete items. Figure 9 shows an example of a standard pop-out effect in visual search where the target differs from the distractors in a simple feature. The activation map generated by the model from this image shows a strong peak at the location of the target. This demonstrates that our search model is not limited to finding targets in surface textures.

Most models that simulate search tasks among discrete items (e.g. Pomplun et al., 2003; Rutishauser & Koch, 2007) depend on feature labels such as red/green, or horizontal/vertical, rather than using statistics measured directly from the stimulus. While our model has been primarily designed to find a single target on a continuous, textured background, it can also be used as a model of discrete item search. In this context, the LNL model has the advantage that by varying the second linear filter it can be made to carry out either an item-wise search or to fixate on centers of gravity, in a similar way to Pomplun et al.’s (2003) Area Activation model. As human observers make fixations on both items and centers of gravity, a model of visual search should incorporate both types of behavior.

Since the parameters of the model were identical in both experiments, and in the search for a discrete item (Figure 9) these results suggest a generality beyond the context of 1/$f^\beta$-noise surfaces in which the model was developed.

The model also compares well with other computational search models. Itti and Koch’s saliency model has been shown to provide a poor correlation with human performance in search tasks using both landscape photographs (Itti & Koch, 2000) and the 1/$f^\beta$-noise stimuli used in this study (Clarke et al., 2008). Although Rao et al.’s (2002) model offers a very good simulation of human search for a set search task, both in the number of saccades and the locations of fixations, it has only been tested for one specific search task, in which human observers needed only three saccades to fixate on the target. The search task used to assess our model covers a much wider range of task difficulty, from 1 or 2 fixations to over 40.

The search task used by Najemnik and Geisler (2008) is similar to ours, although they used 1/$f$-noise directly whereas we use a rendering model to produce naturalistic images. Their aim is to derive the theoretical ideal observer, in terms of search strategies, and compare it to human behavior. While their model gives a good account of human search strategies they do not propose what features and filters should be used to generate the activation map, and, more importantly, they only consider a finite number of possible target locations. This has the effect of simplifying the derivation of the Ideal Observer, and gives a more elegant model. However, it means that their search strategy can not be applied to our stimuli, in which the target can be located at any location (down to pixel level). Not only does the large increase in potential target locations (Najemnik and Geisler consider 85 potential target locations while we have 1024 × 1024 pixels to consider) create computational problems, moving to the pixel level causes some of the underlying assumptions of the model break down. In particular, we cannot assume that the activation at a particular pixel is independent from its neighbors. In fact, due to the 2nd order linear smoothing filter used in the LNL model, every pixel will be correlated to some extent with its neighbors. For this reason, our LNL model cannot be used directly as

Figure 9. Performance of the LNL model on a discrete item search. Left: array of search items. Right: Activation map.
a front end to Najemnik and Geisler’s search model, and the same conclusion applies to any image processing model that generates an activation map (such as those of Itti & Koch, 2000 and Rao et al., 2002). Additional processing would be required in order to reduce the information present in the activation map to a small number of independent potential target locations.

Conclusions

We have shown that a model based on an LNL filter bank can successfully model human performance in a visual search task involving a target on a complex background. These stimuli are naturalistic and allow us to create trials with a large range of task difficulties, from easy (1–2 fixations to target) to difficult (30+ fixations). We used two different classes of surfaces as stimuli and the model gave a good account of human behavior over a range of surface roughnesses, regularities and target orientations. Our aim was to determine whether the information extracted from stimuli by the model is sufficient to account for human search, and, to this end, we modeled search strategy as a simple stochastic process constrained by inhibition of return. In future work, we intend to test the adequacy of these assumptions about search strategy.

Acknowledgments

We would like to acknowledge the support of EPSRC grants EP/F02553X/1 and EP/D059364/1.

Commercial relationships: none.
Corresponding author: Alasdair D. F. Clarke.
Email: adfcl@hw.ac.uk.
Address: The Texture Lab, School of Mathematics and Computer Science, Heriot-Watt University, Riccarton Campus, Edinburgh, Scotland, EH14 4AS, UK.

References


