

A Stochastic Search for a Target on a Textured Background

Alasdair D. F. Clarke, P. R. Green and M. J. Chantler

November 9, 2009

Abstract

We present a comparison between human search performance and that of a stochastic model. The model uses the results of a Signal-Detection experiment to model target detectability and makes random saccades, weighted by empirically derived saccade amplitude and direction distributions. We compare the model against human performance in terms of the number of saccades required to find the target and the spatial distribution of the fixations. We find that the model compares well with human performance despite not integrating information across successive fixations or holding previously fixated image regions in any form memory.

1 Introduction

Visual search is a common task which we encounter daily in real life. Within the confines of the laboratory search tasks frequently involve searching within an image for a designated target item. A complete computational model of visual search should contain two parts: a *feature extraction mechanism* and a *search strategy*. The feature extraction stage takes information from the stimulus and processes it in order to produce an activation map. This process should take both top-down guided search [Wolfe, 2007, Zelinsky, 2008] and bottom-up saliency effects [Gao et al., 2008, Itti and Baldi, 2009, Itti and Koch, 2000] into account. For the sets of abstract, discrete search items commonly used as visual search stimuli, categorical features such as colour, orientation, shape and size are often used. Simple qualitative comparisons between the search items and the target can be used to model top-down guidance [Pomplun et al., 2003, Rutishauser and Koch, 2007]. For more complex stimuli, such as a target hidden in

image noise or in a photograph of a natural scene, we no longer have a discrete set of items to consider and more sophisticated image processing techniques are required [Hwang et al., 2009, Pomplun, 2007, Rao et al., 2002, Tavassoli et al., 2009, Zelinsky, 2008].

This study, however, is concerned with search strategies: the part of a model that uses an activation map to generate successive saccades. While a number of different mechanisms have been put forward, the most commonly implemented has been the the maximum a priori (MAP) Observer. This strategy simply directs saccades to the current maximum of the activation map. A simple inhibition of return (IOR) mechanism is used to stop the model returning to previously fixated maxima and depending on the model, the maxima will either represent a search item, or the centre of gravity of a number of search items. As most previous computational models have primarily been interested in the feature extraction stage of search, this strategy has often been used for simplicity [Clarke et al., 2009, Itti and Koch, 2000, Pomplun et al., 2003, Rao et al., 2002, Rutishauser and Koch, 2007, Zelinsky, 2008].

An alternative to the MAP Observer is the Ideal Observer which Najemnik and Geisler [2005, 2008] have derived for a search task involving a target hidden in $1/f$ -noise. Their model is based on visibility maps calculated from empirical data collected during a signal detection experiment. These maps are used to control how much noise is added to potential target locations (with more noise at higher eccentricities) in the simulations. The Ideal Observer then makes saccades to the target location that will maximise the likelihood of it being able to identify the target during the following fixation. This contrasts with strategies where the model makes a saccade to the location which is currently most likely to be the target. In a second experiment

observers carried out a prolonged search using similar stimuli. The results showed that over a range of task difficulties human observers take a similar number of saccades to the Ideal Observer. Furthermore, when averaged over all trials the Ideal Observer matched the spatial distribution of fixations: both the model and human observers exhibited a preference for fixating above and below the centre of the image.

Models like these are examples of systematic search. Broadly, systematic search can be made up from two types of behaviour. Firstly, the search could be controlled by image properties, in either a bottom-up or top-down manner, or a combination of both. Both the Guided Search [Wolfe, 2007] and saliency [Itti and Koch, 2000] models fall into this category. Secondly, search strategies can also involve some serial dependence between one fixation and the next. This can include both memory processes (for example, inhibition of return) and regular scanning. However, some models, such as Rutishauser and Koch’s probabilistic model [2007] introduce a random element. The aim of this paper is to investigate whether, in a particular type of search task, an entirely random model of saccade selection can account for empirically recorded scan paths.

Several general tendencies observed in scan-paths have been taken to be indicative of systematic search strategies. Gilchrist and Harvey [2006] argue that the presence of horizontal bias in saccade directions indicates a systematic component in visual search and suggest that systematic tendencies can be hard to detect in scan-paths because of the interaction with salience-based object selection. Similarly, Over et al. [2007] have shown that search scan-paths exhibit coarse-to-fine structures, i.e. observers make shorter saccades as a trial progresses. Over et al. [2003] found that saccade directions are influenced by the edges of the search image and reported a preference for making saccades parallel to the boundaries of the stimuli.

Aks et al have argued that the presence of $1/f$ dynamics in saccade-time series is evidence of a systematic component in visual search that relies on memory of previous fixation locations [Aks, 2006, Aks et al., 2002]. They carried out the same time-series analysis on a random walk and found that it did not exhibit the same properties. However it is possible that Aks’ result is an artefact of studying

the compound time-series of large number of visual searches, one after another. If we were to look at the saccade amplitude time-series of several individual searches, each with a coarse-to-fine dynamic, then we would expect to see a strong low frequency component which could, at least partially, explain Aks et al.’s result. Similarly, *distance-to-target* dynamics have been put forwards as evidence for some systematic component in visual search [Tseng and Li, 2004]. These dynamics suggest that there are two phases to the search process: an ineffective stage, followed by an effective stage in which the distance from the current fixation location to the target decreases monotonically. However Greene has shown that these dynamics also arise in simple random walk simulations [Greene, 2008]. These simulations generated saccades at random long. The target was assumed to be detected if it was within 20 pixels of the current fixation, and the model was given a maximum of 15 saccades to find the target. Further support for the idea of unsystematic visual search comes from several studies which show that memory only plays a small role in visual search tasks [Horowitz and Wolfe, 1998, 2001, Kuna et al., 2008, Wolfe et al., 2000].

However chance has been demonstrated to play a significant role in visual search performance. Using saccade amplitude distributions Motter and Holsapple [2001] calculated the probability of fixating on the target by chance under different conditions. While this chance component decreases as the number of distracters increase, it continues to account for a sizeable fraction of performance. Random walks have been successfully used to model an observer’s speed and accuracy in present/absent forced choice experiments [Reeves et al., 2005, Stone, 1960]. Rather than model the spatial distribution of fixations these models simulate the observer’s decision making process. The random walk occurs between two boundaries one for a target present response and one for target absent, and is governed by a drift and bias.

Overall, while there is strong evidence for systematic search in many tasks, the random element is stronger than we might expect in many others. However there is no clear evidence yet for purely random-walk search. In this paper we investigate just this, and ask if there are situations and stimuli for which

a random-walk is a good model for human observers. A good place to look for such behaviour would be with simple stimuli (to avoid high-level semantic information from being used) containing a difficult to find target, such as a target embedded in noise [cite some relevant papers]. We have chosen to use fractal surfaces, which when illuminated give the appearance of a naturalistic surface texture [Clarke et al., 2008]. See Figure 1 for an example. These surfaces have two advantages over traditional visual search stimuli. Firstly, unlike the arrays of abstract discrete geometric objects that are often used, they appear naturalistic. Furthermore, unlike photographs of natural scenes, they are fully controlled and parametrised, and by controlling the seed of the random number generator many different, yet equivalent, textures can be created.

Previous attempts to model search on such surfaces have used filter based image processing models [Clarke et al., 2009]. However, while this model gives a good account of human performance, in terms of the number of saccades required to find the target, it fails to account for the pattern of fixations made by human observers and there was no apparent relationship between human fixation locations and (non-target) local maxima in the activation map: only 22% of the model’s potential saccade targets were located within 1° of the saccade target chosen by a human observer. Further analysis (see Appendix A) shows the model’s saccade target can be expected to coincide with the human saccade target 19% of the time if saccades are made at random. Furthermore, human observers often make long saccades that cannot be explained using the eccentricity dependant fall-off in activation.

In this study, in order to investigate search strategy only, we will assume a feature extraction mechanism and use an empirical visibility model based on the results of a signal detection experiment as the starting point for the search strategy. This mirrors Najemnik and Geisler’s methods [2005]. However, we will use the visibility model in conjunction with a stochastic process which draws saccade amplitudes and directions from empirically derived distributions. This will allow us to determine whether a simple random walk model is sufficient to account for human saccade paths in our task.

Note that there are two differences between the stimuli used by Najemnik and Geisler and

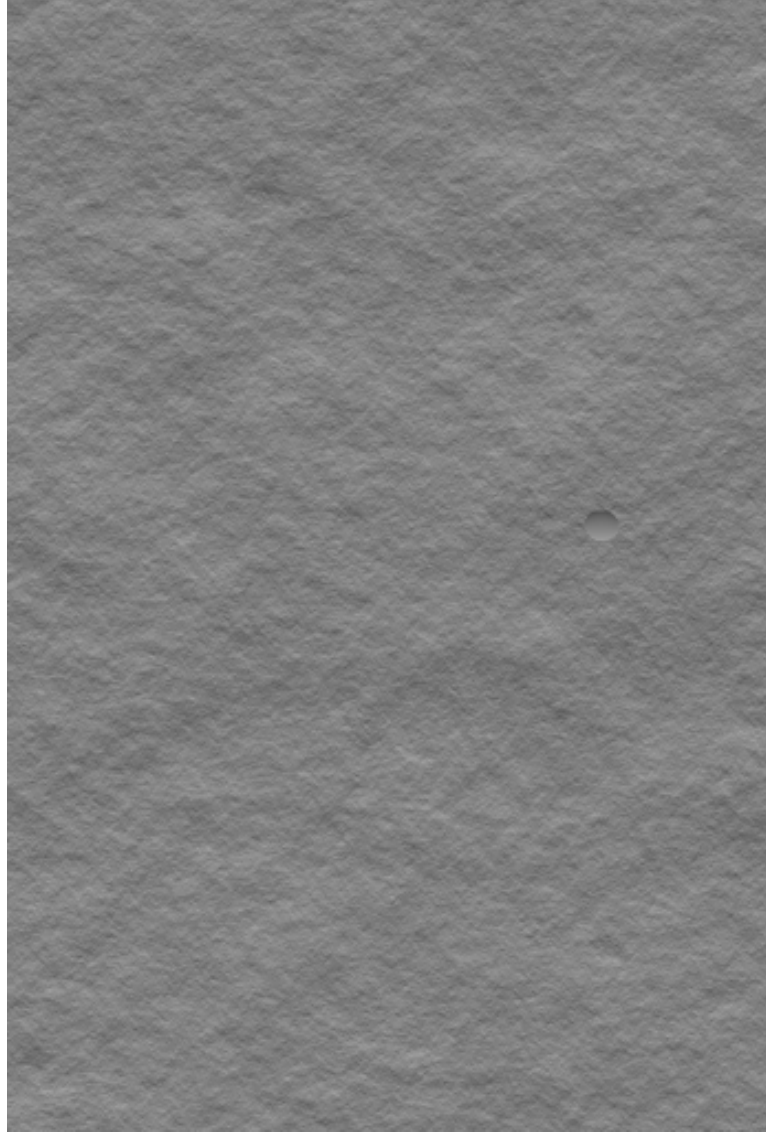


Figure 1: An example $1/f^\beta$ -noise surface. Note: this example is only 512×512 pixels, the stimuli used in the experiment where 1024×1024 , while the target dent stayed the same size.

ours. Firstly they used $1/f$ -noise as their stimuli whereas we have combined $1/f$ -noise with an illumination model to give surfaces that appear naturalistic [Clarke et al., 2008]. Secondly, they consider only 85 potential target locations, whereas in our stimuli the target can be centred at any pixel within a given distance from the centre. While we would not expect human observers to notice this difference, the lower number did considerably simplify the derivation of the Ideal Observer. If the target is allowed to be positioned anywhere in the stimuli, then the independence assumption required by the Ideal Observer model is no longer valid.

2 Experiment One: Signal Detection

The aim of this experiment is to measure the probability of target detection for different eccentricities and surface roughness combinations. This will then give us a visibility map upon which we can base a simple model for the probability of target detection at different eccentricities and surface roughnesses.

In order to carry out the 2AFC design used by Najemnik and Geisler [2008], the location of the target is cued for blocks of trials. This allows for (covert) attention to be deployed away from the current fixation location and to the target’s location. To avoid this priming effect we will use a target/absent task where the location of the target is not known before the trial.

2.1 Stimuli

A range of rough surfaces were generated by applying Lambert’s cosine law to height maps generated by a $1/f^\beta$ -noise process. For full technical details see Appendix B. The surface roughness is governed by $\beta \in \{1.60, 1.65, 1.70\}$ and a scaling factor, RMS roughness, which was kept constant, $\sigma_{RMS} = 1.1$. For the target present trials, the target was located at one of 72 potential locations: nine different eccentricities were used $0.84^\circ \leq r \leq 7.5^\circ$, and eight evenly spaced orientations. The target was made by subtracting an ellipsoid from the three dimensional surface and subtended 0.66° of visual angle. For an example, see Figure 1.

For each parameter combination, twenty different trials were created (by changing the

random seed used to create the noise we can create different, yet statistically identical textured surfaces). Additionally, 160 target absent trials were included for each value of β . This gave a total of 2160 target present trials and 480 target absent. (The number of target absent trials was based on pilot results and ensured that observers made roughly equal numbers of positive and negative responses. As a large number of the target present trials were answered incorrectly we do not need so many target absent trials.)

2.2 Method

Two participants carried out all the trials, split into twenty subgroups, over a number of days. They were paid £50 each. Within each subgroup of 132 trials there were 33 runs of 4 trials. During each run the participants were instructed to keep their eyes fixated on the centre of the image. Each trial consisted of a fixation cross (500ms), stimulus (200ms), white noise mask (500ms), and finally a fixation cross was displayed until a target present or absent response was given.

A Tobii x50 eyetracker was used to track the participants’ gaze. A fixation filter was used so that only trials in which the mean distance from the central fixation cross to the observers gaze was less than 1° were included in the analysis (Additionally, only trials in which the standard deviation of the gaze’s x and y components was less than $2/3^\circ$ were used.)

2.3 Results

Trials in which fixation was not held at the centre of the image (14%) were removed from further analysis. The results for the two individual participants are shown in Figure 2. For all cases, the accuracy for the target absent trials is $>90\%$, and hence false positives will not be included in any further analysis.

The two subjects performed similarly and the mean target present performance is shown in Figure 3. We will model this by using a simple multi-linear, with thresholding, model:

$$f(\beta, r) = 4.09\beta - 0.11r - 5.97$$

$$p(T|\beta, r) = \begin{cases} f(\beta, r) & (f \geq 0) \\ 0 & (f < 0) \end{cases}$$

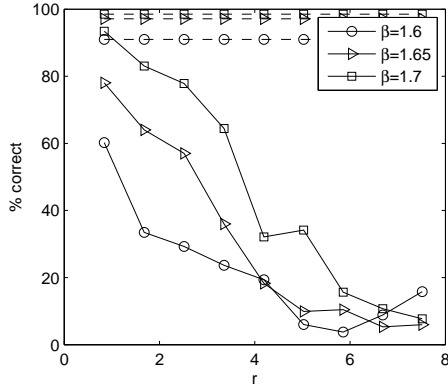
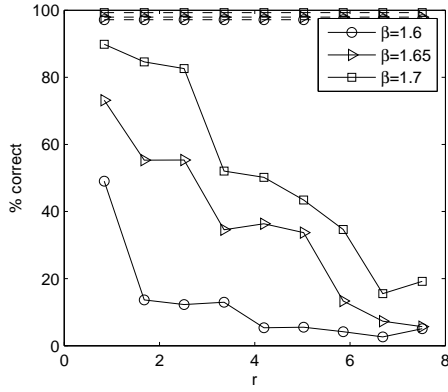


Figure 2: Results for the two individuals from the signal detection experiment. The solid lines the accuracy results for the target present trials, while the dashed lines shows accuracy for target absent.

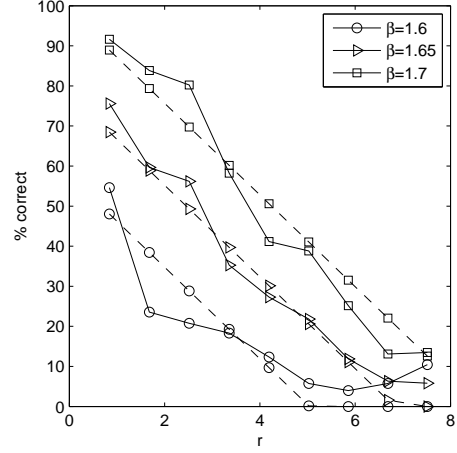


Figure 3: Solid lines show the inter-subject mean over β and r while the dashed line shows the best fit multi-linear regression.

This regression model gives $R^2 = 0.934$. Note, the points $\beta = 1.6, r = 5.86, 6.69, 7.51$ were not including in the regression model.

3 Simulation

The simulation used the linear-regression detection model, based on the results of the signal detection experiment, together with data on human saccade distributions taken from previous work, to construct a stochastic model of saccade paths during visual search.

3.1 Human Saccade Distributions

In this section we will present a new analysis of data from the first of Clarke et al.'s [2009] visual search experiments. This experiment involved similar stimuli to those described above, although only 3 different target eccentricities were used. Seven observers completed the experiment and the task was simply to press a button on the keyboard once the target had been found. They were given unlimited time and an eye-tracker was used to monitor fixation locations. As the surface was made rougher (decreasing β and increasing σ_{RMS}) a greater number of saccades were required to find the target. We will now carry out some further analysis of this data set.

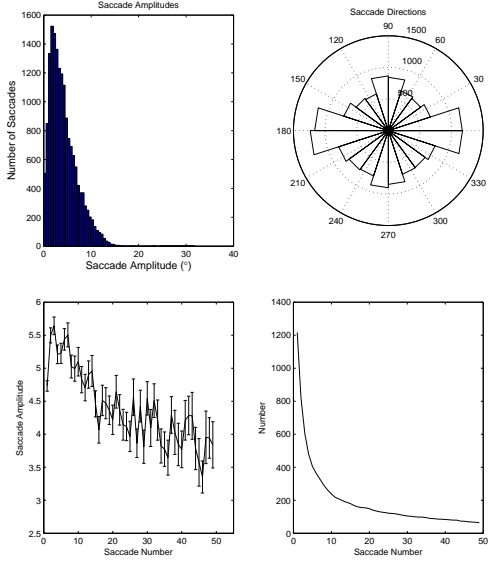


Figure 4: (Top Left) Saccade amplitude histogram, (Top Right) Saccade direction rose plot, (Bottom Left) Saccade amplitude plotted against time, and (Bottom Right) The number of the fixations included in the analysis on the left.

The global saccade amplitude and direction distributions are shown in Figure 4 (top left and top right). As we can see, the saccade directions show the same horizontal bias as reported by Gilchrist and Harvey [2006] and Najemnik and Geisler [2008]. The saccade amplitude time series (Figure 4 (bottom left)) shows evidence for a coarse-to-fine search strategy, as reported by Over et al. [2007].

Figure 5 shows the distributions of saccade amplitude and directions according to the location of the starting fixation, in relation to the stimuli boundaries. For example, the largest saccade amplitudes appear to occur after a fixation is made near the corner of the stimuli early on in the search. Again, this agrees with the coarse-to-fine strategy described by Over et al.

3.2 Algorithm

3.2.1 Inputs

The model is given the size of the search area, N , the roughness of the surface β , and the target location (chosen randomly for a given eccentricity r). The initial fixation is set to the centre of the search area.

3.2.2 Target Detection

On each fixation the probability that the model detects the target is given by $p = f(\beta, r)$ where f is the linear regression model described above. For each fixation we generate a random number $x \in [0, 1]$ and check to see if $x \leq p$, in which case the model detects the target, makes a saccade to the target’s location, and the search is terminated. If the model does not detect the target (i.e. $x > p$) then a random saccade is made to a new location.

3.2.3 Generating a Saccade

In order to generate scan-paths we will use the empirical distributions of saccades (Figure 5) as probability density functions. Since the amplitude distribution varies with the position of a saccade in the sequence made during a search trial we use separate empirical distributions for fixations $1 \leq t \leq 5$, $6 \leq t \leq 10$, $11 \leq t \leq 15$ and $t > 15$ (where t is the fixation number).

3.2.4 Results

Considering first the performance of the model in terms of the number of fixations required to find the target, we see that it performs in a similar way to the human observers in Clarke et al. [2009] (Figure 6). While the model slightly overestimates the number of saccades needed on easy surfaces, and underestimates the amount needed for harder surfaces, it remains within the 95% confidence interval of the mean human observer.

Next we compare how efficiently human observers and the model cover the search area during search. If human search has systematic properties we would expect the fixation locations to be approximately uniformly distributed: as there are no salient regions in the image apart from the target there is no reason to fixate any one area more than any other (except for a potential innate central fixation bias [Tatler, 2007]). The hotspot maps for the model and human observers are shown in Figure 7. For the human data, these included the first 30 fixations (excluding the initial fixation) from all trials in which at least 32 fixations were made (to minimise the effect of detecting the target). The model data was generated by simply running the model 105 times without the target detection component. The two hotspot maps were normalised. As we

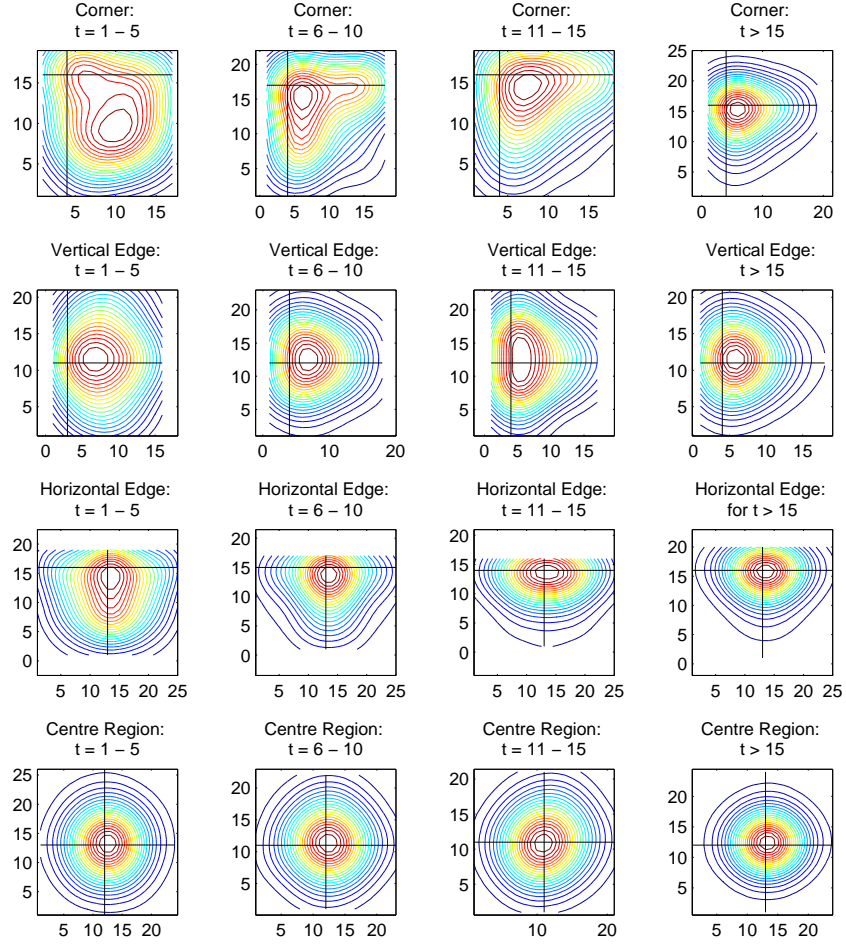


Figure 5: The effect of time and fixation location on saccades. These were created by dividing the search area into 5×5 equally sized subregions. Horizontal and vertical axis of symmetry were assumed and used to combine information from the four corners regions. Similarly, the three remaining subregions along the left edge were reflected and combined with those along the right edge (similar for the top and bottom edges). Finally the nine central regions were grouped together.

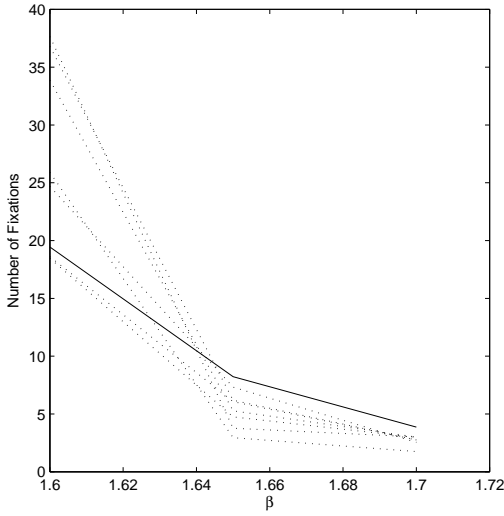


Figure 6: Comparison between human observers and the model. The top graph shows the performance, in terms of the number of fixations required to find the target, for each of the seven participants (dashed line), while the solid line shows the model’s performance.

can see, the model appears to exhibit more of a central bias than the human data.

Another way to compare how systematic human observers are is to look at the number of re-fixations. We have done this by checking each fixation to see if it is located within 30pixels ($1/2^\circ$ of visual angle) of any of the previous three fixations. For each trial we computed the number of re-fixations per fixation. This number gives us an indication of how strong inhibition of return (IoR) is in the search task considered here and the results (see Figure 8) show that humans appear to be no more systematic than the stochastic simulation. In fact, the model actually makes less re-fixations per fixation on the harder trials.

Finally, we will use the Voronoi method proposed by Over et al. [2006] to compare the distribution of saccades from the human and model. This method allows us to study the uniformity of fixation density and involves computing the bounded Voronoi cells [Voronoi, 1907] for a set of fixation coordinates and looking at the distribution of cell areas. Some examples of Voronoi diagrams can be seen in Figure 9. However, instead of looking at the skewness and fitting curves to the distribution

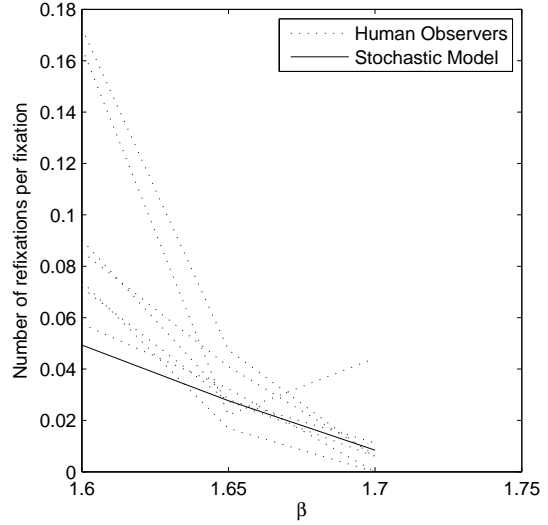


Figure 8: Number of times a saccade is directed to the location of one of the previous three fixations, divided by total number of fixations for that trial.

of cell areas as Over et al. did, we will look at how the cells change over time, as successive fixations are made. A simple way of doing this is to look at how the area of the largest cell changes with fixation number. If search is systematic then we would expect observers to direct fixations towards regions of the display area which they have not already attended to and the size of the largest cell will fall more steeply than for a stochastic search.

For each fixation f_t , ($t = 1, \dots, n - 2$, where n is the number of fixations in the trial), the Voronoi cells for fixations f_1, \dots, f_t , are created and the area of the largest cell, A_t , is computed. The mean A_t , over all trials, is shown in Figure 10 (left). As we can see, the human observers outperform the stochastic simulation and the size of the largest Voronoi cell decreases rapidly during the first 5 fixations. However if we look at the speed at which the observers reduce the size of the cell ($A_t - A_{t+1}$) we see that after the initial five fixations the stochastic simulation matches human performance (Figure 10 (right)).

4 General Discussion

The results suggest that, for visual searches involving locating a target on an otherwise ho-

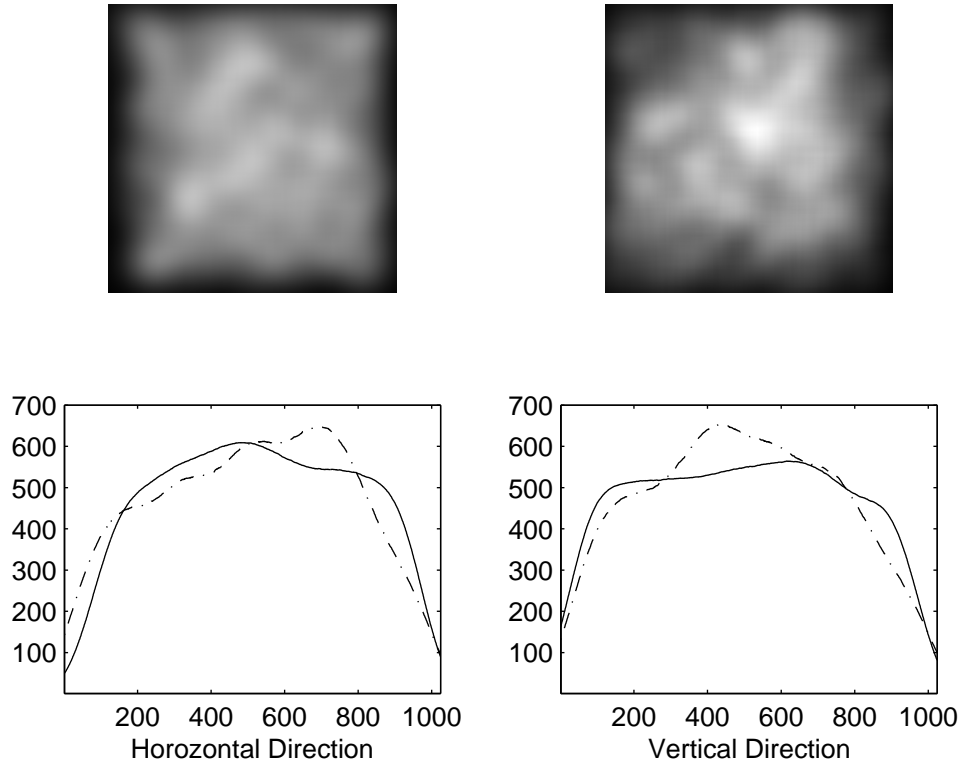
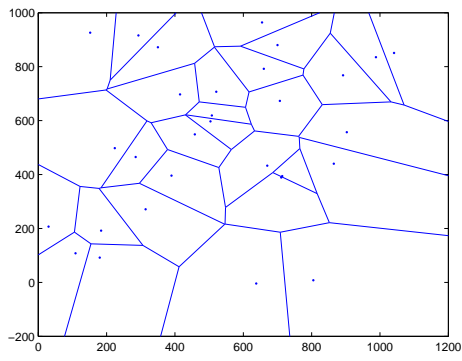
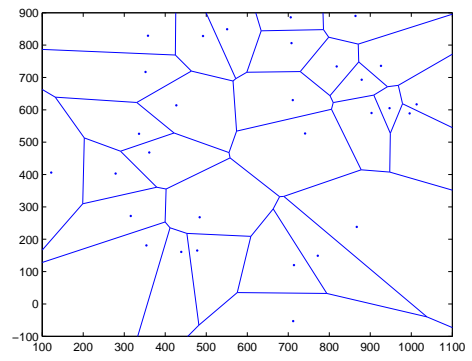


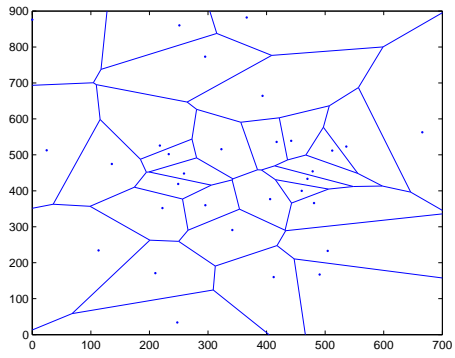
Figure 7: (top left) Hotspot map of first 30 fixations over all trials (containing over 32 fixations) and observers. (top right) Hotspot map generated by the stochastic model. The stochastic model appears to be more biased towards the centre of the image. (bottom left/right) Density of fixations along the horizontal/vertical directions. Legend: (solid line) human observers (dashed line) stochastic model.



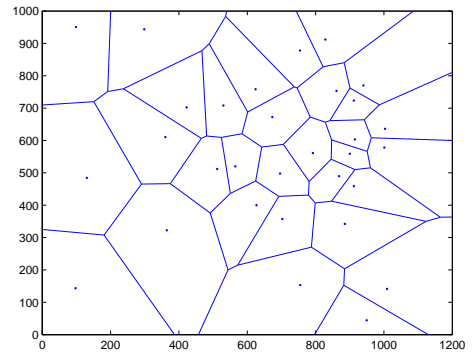
(a) Human Observer



(b) Human Observer



(c) Model



(d) Model

Figure 9: Examples of Voronoi plots from both the human observers and the stochastic model.

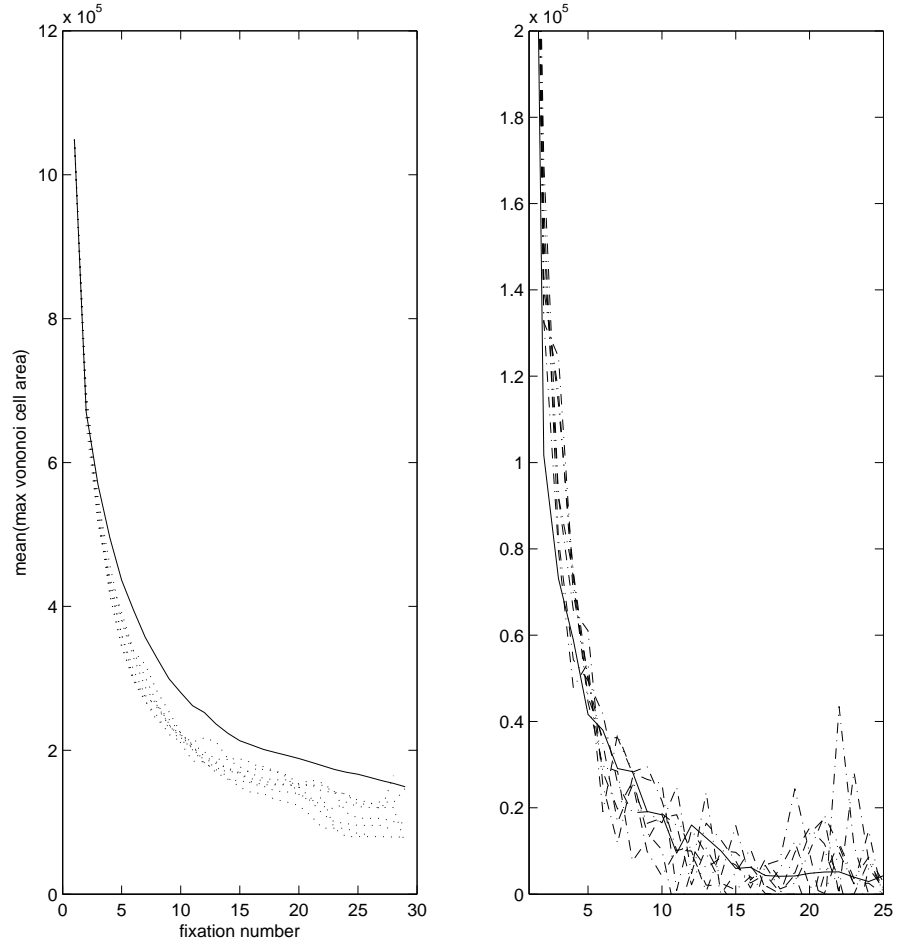


Figure 10: (left) How the maximum Voronoi cell area changes with time. Dotted lines show the seven human observers while the solid line shows the stochastic model. (right) This shows the derivative of the graph on the left: a measure of how quickly the search area is covered. The main difference between the model and human observers occurs in the first few fixations. After these initial fixations human observers appear to be no more systematic than the model.

homogeneous surface texture, human behaviour can be modelled closely by a stochastic process. The random walk model finds the target in a similar number of fixations as a typical human observer and produces scan-paths that are spatially distributed in a similar way to human scan-paths. The random walk model also makes a similar number of re-fixations and, except for the initial few fixations, searches as efficiently as human observers.

This pattern could be indicative of differing fields of view between search with and without eye-movements [Motter and Simoni, 2008].

However, this is not to say that our results challenge Guided Search: in fact the two models could quite easily work together. In fact, the model presented in this paper *is* guided. When the target is salient against the textured background then the model is likely to make a saccade to that location. One could imagine that if there are several search items that could potentially be the target, as there are in most typical visual search experiments, that a random walk model could be used to choose which item should be fixated next. Also, although our model did not need any form of memory or inhibition of return, we do not mean to suggest that there is no inhibition of return in human search. Indeed, as our stimuli contain no search objects, any IoR process would have to be operating with respect to spatial coordinates defined with respect to the stimulus boundaries, rather than being applied to discrete search objects. However, as we can see from Figure 8 human observers do not appear to re-fixate recently fixation regions any more or less than we would expect a stochastic, memory-less process to. Further research and experiments would be needed to characterise this behaviour more thoroughly.

This is a somewhat surprising result given that Najemnik and Geisler [2008] have shown that human observers appear to be near optimal in their search strategy. Najemnik and Geisler also compared the spatial distributions of the fixations chosen by their ideal observer, a MAP model, and human subjects, and found that both human subjects and the ideal observer show a clear preference for fixation on small regions above and below the centre of the image. However, there is no evidence for this distribution in our data, which show no preferences for fixating any particular regions

(see Figure 7).

There have been a series of papers using Classification Images to investigate guidance in a search task involving targets embedded in $1/f$ -noise [Rajashekar et al., 2002, 2006, Tavasoli et al., 2009]. This involves taking recording all the fixation locations and computing the mean region that is fixated on. For a search task involving a geometric shape embedded in $1/f$ noise, these classification images have been shown to resemble the target which is being searched for. However these results do not appear to hold when the analysis methods are applied to the data from Clarke et al. [2009]. The classification images obtained from the seven observers are shown in Figure 11. It could be that the target is too small when compared to the background and does not show up due to spatial uncertainties (need a better way of saying this).

4.1 Conclusions

As stated above, a complete, computational visual search model should possess two parts: a feature extraction front end and a search strategy. The aim of this paper was to explore to what extent a search strategy based on a random walk could account for human performance. Previously implemented search strategies have generally worked in a serial manner, checking items one at a time, with some form of imperfect memory [Melloy et al., 2006, Rutishauser and Koch, 2007]. One search model that makes use of parallel target detection over a serial sequence of fixations is the Ideal Observer. One problem with this approach is that it assumes that the target will be located at one of a predefined independent finite number of potential target locations. Unfortunately, this assumption breaks down when image processing techniques are used, as the activation at any pixel is likely to be correlated with its neighbours. Hence we have explored an alternative explanation of human search strategies: a random walk. While the use of a random walk to explain patterns of fixations is not new [Aks et al., 2002, Greene, 2008, Morawski et al., 1980], our model is unlike earlier ones in being strongly based on empirical data. We find that a random walk behaves in a similar way to human observers, both in terms of the number of saccades required to find the target, and the spatial distribution of fixations.

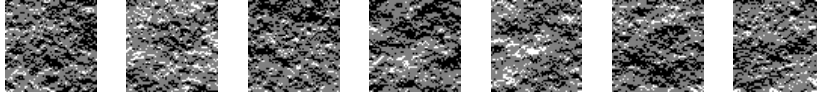


Figure 11: Classification Images obtained for each of the seven observers in the first visual search experiment in Clarke et al. [2009]

Our results here suggest that inhibition of return; integration of information across fixations, and more general memory based processes do not have a large role to play in at least one type of search task (search for an inconspicuous target on a continuously textured surface). Future testing of models of visual search should consider not only possible differences between search strategies on different types of stimuli, but also variation between observers in their strategies. It may be possible to obtain evidence for more than one model of search strategy depending on the observers tested.

A Appendix: LNL model comparison

In our previous work we have applied an LNL-based model to the problem of modelling search for a target on a rough surface (Clarke, et al., 2009). A bank of Gabor filters was applied to the input image and then passed through a non-linearity. This nonlinear processing strengthened the signal of filter response maps containing a small number of strong local maxima (as opposed to maps which contained a large number of local maxima). Finally these feature maps were passed through a 2nd order linear filter (local energy pooling) before being summed together to give an activation map. This activation map was then passed to a simple saccade selection algorithm. For each fixation an exponential distance dependant fall-off was applied to the activation map along with a simple inhibition of return process. The model would then randomly select one of the n ($=3$) largest local maxima as its saccade target.

While the algorithm succeeded in modelling human performance (in terms of the number of saccades required to find the target), it did not account for the selection of individual fixation points on each saccade, in trials where

more than one or two saccades were required to find the target. There was no apparent relationship between human fixation locations and (non-target) local maxima in the activation map (see Figure 2 for an example). As the example shows, human observers often make long saccades that cannot be explained using our eccentricity dependant exponential fall-off. While one possibility would be that the fall-off function is too strong, we can discard this suggestion as weakening the activation fall-off function would cause the model to diverge from human performance in terms of number of saccades to targets at high eccentricities.

To explore whether Figure 2 is typical of the model's behaviour we compared saccade targets for the model with those chosen by human observers in the experiment reported by Clarke et al. [2009]. Over all observers, trials and fixations, 22% of saccades were directed to within 1° of one of the three saccade targets considered by the model, and over 25% fell more than 4.3° (equal to a quarter of the display's length) away from the nearest point considered by the model. The LNL model is therefore able to predict the locations of only a small proportion of non-target fixations during visual search. Furthermore, most of the successful cases can be accounted for by chance. Let us assume (a) that all (both the model's and human) saccades are no more than r in amplitude; and (b) the three potential fixations considered by the model are separated from each other by at least 2°

Then the fixations will occur somewhere within an circle with area $A = \pi r^2$. Therefore the probability of the human saccade landing within 1° of one of the model's saccades is $p = 3\pi/A = 3/r^2$. If we take $r = 4^\circ$ (over half of the human saccades are under 4° in amplitude) then we would expect human observers to fixate within 1° of one of the fixation locations considered by the model 19% of the

time, which is close to the 22% we obtained from the empirical comparison.

The above analysis suggests that the LNL model, while offering a good prediction of the difficulty of the search task, does not succeed in modelling saccade selections any better than if it did not possess an activation map.

B Appendix: Texture Synthesis

For the experiments in this paper a very simple model referred to as $1/f^\beta$ -noise will be used. We can represent a surface by an $n \times n$ matrix h . This matrix is referred to as a height map and for any $(x, y) \in \{Z \times Z | 0 < x, y < n\}$, $h(x, y) = z$ gives us the height of the surface. The $1/f^\beta$ noise has only two parameters: the frequency roll-off, β , and the RMS-roughness, σ_{RMS} . The RMS-roughness is the standard deviation of the surface's height map and acts as a scaling factor in the z -axis. The roll off factor β controls the amount of high frequency information in the surface: increasing β causes the high frequencies to drop off more quickly, so the texture appears smoother [Padilla et al., 2008]. Note that we use β here to denote the roll-off of the magnitude of the inverse discrete Fourier transform of the height map. The same term and symbol also sometimes refers to the roll off factor in the power spectrum of an image. See Chantler et al. [2005] for a model relating these two parameters.

The surface is generated in the Fourier domain with the magnitude spectrum given by:

$$S(u, v) = \frac{k}{\sqrt{u^2 + v^2}} \quad (2)$$

where k is the scaling factor required to give the desired σ_{RMS} . The phase spectrum is randomised and by using different values to seed the random number generator we can create many different surfaces with the same properties. Taking the inverse discrete Fourier transform of the magnitude and phase spectra gives us our height map h .

The two dimensional height map that represents our surface texture is then rendered to generate an image of the surface under specified illumination. This stage is important, as a surface can have drastically different appearances under different light conditions

[Chantler, 1995]. We will use one of the simplest models, known as Lambert's Cosine Law. This treats the surface as a perfectly diffuse reflector, i.e. it reflects the same amount of light in all directions. It is easily modelled by the dot product:

$$i(x, y) = \lambda \cdot \rho(x, y) \underline{n}(x, y) \underline{l} \quad (3)$$

where i is the image we are creating, \underline{n} is the unit surface normal to the height map and \underline{l} is the unit illumination vector. The albedo, ρ , determines how much light is reflected by the surface. The strength of the source light is denoted by k . The surface normal \underline{n} is estimated by:

$$p(x, y) = h(x, y) - h(x-1, y) \quad q(x, y) = h(x, y) - h(x, y-1) \quad (4)$$

$$\underline{n} = \frac{1}{\sqrt{1 + p^2 + q^2}} [p, q, 1]^T \quad (5)$$

Self shadowing occurs when $i(x, y) < 0$ i.e. the surface is orientated such that its normal makes an obtuse angle with the illumination vector. Self shadowing is implemented by setting all negative values to 0. Cast shadows are not implemented in this model. The illumination conditions will be kept constant throughout this paper with elevation = 60° , azimuth = 90° and the strength of the source light being set at 150 cd/m^2 . The albedo value will be kept constant at 0.8 throughout all the experiments, which is approximately the value of matte white paint.

To use rendered surfaces in visual search experiments, we need to choose a target. Rather than using a target with an artificial appearance, we create an anomaly in the surface texture, in the form of a small pothole in the surface. We create these targets by subtracting the lower half of a small three dimensional ellipsoid from the surface. Our ellipsoid defect is defined by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \quad (6)$$

To make the indent appear more realistic it is convolved with a two dimensional smoothing filter to soften the hard vertical edges.

References

- D. Aks. $1/f$ dynamic in complex visual search: Evidence for self-organized criticality in human perception. In M. A. Riley, editor, *Tutorials in contemporary nonlinear methods for the behavioral sciences*, pages 329–359, 2006.
- D. J. Aks, G. Zelinsky, and J. C. Sprott. Memory across eye-movements: $1/f$ dynamic in visual search. *Journal of Non-linear Dynamics in Psychology & the Life Sciences*, 6:1–15, 2002.
- M. J. Chantler. Why illuminant direction is fundamental to texture analysis. *IEEE Proceedings Vision, Image and Signal Processing*, 142:199–206, 1995.
- M. J. Chantler, A. Petrou, M. and Penirschke, M. Schmidt, and G. McGunnigle. Classifying surface texture while simultaneously estimating illumination. *International Journal of Computer Vision (VISI)*, 62:83–96, 2005.
- A. D. F. Clarke, P. R. Green, M. Chantler, and K. Emrith. Visual search for a target against a $1/f^\beta$ continuous textured background. *Vision Research*, 48:2193–2203, 2008.
- A. D. F. Clarke, P. R. Green, and M. Chantler. Modelling visual search on a rough surface. *Journal of Vision*, 9x:1–12, 2009.
- D. Gao, V. Mahadevan, and N. Vasconcelos. On the plausibility for the discriminant center-surround hypothesis for visual saliency. *Journal of Vision*, 8(7)(13):1–18, 2008.
- I. D. Gilchrist and M. Harvey. Evidence for a systematic component within scan paths in visual search. *Visual Cognition*, 14:704–715, 2006.
- H. H. Greene. Distance-from-target dynamics during visual search. *Vision Research*, 48:2476–2484, 2008.
- T. S. Horowitz and J. M. Wolfe. Visual search has no memory. *Nature*, 357:575–577, 1998.
- T. S. Horowitz and J. M. Wolfe. Search for multiple targets: Remember the targets, forget the search. *Perception and Psychophysics*, 63:272–285, 2001.
- Alex D. Hwang, Emily C. Higgins, and Marc Pomplun. A model of top-down attentional control during visual search in complex scenes. *Journal of Vision*, 9(5):1–18, 2009.
- L. Itti and P. F. Baldi. Bayesian surprise attracts human attention. *Vision Research*, 49:1295–1306, 2009.
- L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12):1489–1506, May 2000.
- M. A. Kuna, S. Flusberg, and J. M. Wolfe. The role of memory and restricted context in repeated visual search. *Perception and Psychophysics*, 70:314–328, 2008.
- B. J. Melloy, S. Das, A. K. Gramopadhye, and A. T. Duckowski. A model of extended, semi-systematic visual search. *Human Factors*, 48:540–554, 2006.
- T. Morawski, C. G. Drury, and M. H. Karwan. Predicting search performance for multiple targets. *Human Factors*, 22:707–718, 1980.
- B. C. Motter and J. W. Holsapple. Separating attention from chance in active visual search. In J. Braun, C. Koch, and J. Davis, editors, *Visual attention and neural circuits*, pages 159–175. Cambridge, MA: MIT Press, 2001.
- B. C. Motter and Diglio A. Simoni. Changes in the function visual field during search with and without eye movements. *Vision Research*, 2008.
- J. Najemnik and W. S. Geisler. Optimal eye movement strategies in visual search. *Nature*, 434:387–391, 2005.
- J. Najemnik and W. S. Geisler. Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, 8(3):1–14, 2008.
- E. A. Over, I. T. Hooge, and C. J. Erkelens. Visual search: saccade parameters depend on the shape of the search area [abstract]. *Journal of Vision*, 3(429):251–261, 2003.
- E. A. Over, I. T. Hooge, and C. J. Erkelens. A quantitative measure for the uniformity of fixation density: The voronoi method. *Behaviour Research Methods*, 38:251–261, 2006.

- E. A. Over, I. T. Hooge, B. N. Vlaskamp, and C. J. Erklens. Coarse-to-fine eye movements strategy in visual search. *Vision Research*, 47:2272–2280, 2007.
- S. Padilla, O. Drbohlav, P.R. Green, A.D. Spence, and M.J. Chantler. Perceived roughness of $1/f^\beta$ noise surfaces. *Vision Research*, 48:1791–1797, 2008.
- M. Pomplun. Advancing area activation towards a general model of eye movements in visual search. In W. D. Gray, editor, *Integrated Models of Cognitive Systems*, pages 120 – 131. New York Oxford University Press, 2007.
- M. Pomplun, J. Shen, and E. M. Reingold. Area activation: A computational model of saccadic selectivity in visual search. *Cognitive Science*, 27:299 – 312, 2003.
- U. Rajashekar, L. K. Cormack, and A. C. Bovik. Visual search: Structure from noise. In *In Proceedings of the eye tracking research & applications symposium*, pages 119–123. New Orleans: ACM Press, 2002.
- U. Rajashekar, A. C. Bovik, and L. K. Cormack. Visual search in noise: Revealing the influence of structural cues by gaze-contingent classification image analysis. *Journal of Vision, Special Issue on Finding visual features: Using stochastic stimuli*, 6: 379–386, 2006.
- P. N. Rao, G. J. Zelinsky, M. M. Hayhoe, and D. H. Ballard. Eye movements in iconic visual search. *Vision Research*, 42:1447–1463, 2002.
- A. Reeves, N. Santhi, and S. DeCaro. A random-ray model for speed and accuracy in perceptual experiments. *Spatial Vision*, 115: 73–83, 2005.
- Ueli Rutishauser and Christof Koch. Probabilistic modeling of eye movement data during conjunction search via feature-based attention. *Journal of Vision*, 7(5):1–20, 2007.
- M. Stone. Models for choice-reaction time. *Psychometrika*, pages 251–260, 1960.
- B. W. Tatler. The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7:14(4):1–17, 2007.
- A. Tavassoli, I. van der Linde, and A. C. Bovik. Eye movements selective for spatial frequency and orientation during active vision search. *Vision Research*, 49:173–181, 2009.
- Y. C. Tseng and C. S. Li. Oculomotor correlates of context-guided learning in visual search. *Perception & Psychophysics*, 66: 1363–1378, 2004.
- Georgy Voronoi. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. *Journal für die Reine und Angewandte Mathematik*, 133:97–178, 1907.
- J. M. Wolfe. Guided search 4.0: Current progress with a model of visual search. In W. Gray, editor, *Integrated Models of Cognitive Systems*, pages 99–119. New York: Oxford, 2007.
- J. M. Wolfe, N. Klempen, and K. Dahlen. Postattentive vision. *Journal of Experimental Psychology: Human Perception and Performance*, 26:693–716, 2000.
- Gerg J. Zelinsky. A theory of eye movements during target acquisition. *Psychological Review*, 115:787–835, 2008.

List of Figures

1	An example $1/f^\beta$ -noise surface. Note: this example is only 512×512 pixels. the stimuli used in the experiment where 1024×1024 , while the target dent stayed the same size.	3
2	Results for the two individuals from the signal detection experiment. The solid lines the accuracy results for the target present trials, while the dashed lines shows accuracy for target absent.	5
3	Solid lines show the inter-subject mean over β and r while the dashed line shows the best fit multi-linear regression.	5
4	(Top Left) Saccade amplitude histogram, (Top Right) Saccade direction rose plot, (Bottom Left) Saccade amplitude plotted against time, and (Bottom Right) The number of the fixations included in the analysis on the left.	6
5	The effect of time and fixation location on saccades. These were created by dividing the search area into 5×5 equally sized subregions. Horizontal and vertical axis of symmetry were assumed and used to combine information from the four corners regions. Similarly, the three remaining subregions along the left edge were reflected and combined with those along the right edge (similar for the top and bottom process). Finally the nine central regions were grouped together.	7
6	Comparison between human observers and the model. The top graph shows the performance, in terms of the number of fixations required to find the target, for each of the seven participants (dashed line), while the solid line shows the model's performance.	8
7	(top left) Hotspot map of first 30 fixations over all trials (containing over 32 fixations) and observers. (top right) Hotspot map generated by the stochastic model. The stochastic model appears to be more biased towards the centre of the image. (bottom left/right) Density of fixations along the horizontal/vertical directions. Legend: (solid line) human observers (dashed line) stochastic model.	8
8	Number of times a saccade is directed to the location of one of the previous three fixations, divided by total number of fixations for that trial.	8
9	Examples of Voronoi plots from both the human observers and the stochastic model.	9
10	(left) How the maximum Voronoi cell area changes with time. Dotted lines show the seven human observers while the solid line shows the stochastic model. (right) This shows the derivative of the graph on the left: a measure of how quickly the search area is covered. the main difference between the model and human observers occurs in the first few fixations. After these initial fixations human observers appear to be no more systematic than the model.	10
11	Classification Images obtained for each of the seven observers in the first visual search experiment in Clarke et al. [2009]	11
		13