Modelling visual search for surface defects

Alasdair Daniel Francis Clarke

All work carried out under the supervision of:

Mike J. Chantler and Patrick R. Green

Submitted in total fulfilment of the requirements of the degree of:

Doctor of Philosophy

June 28, 2010

Department of Computer Science,
School of Mathematics and Computer Science,
Heriot-Watt University,
Edinburgh

Abstract

Much work has been done on developing algorithms for automated surface defect detection. However, comparisons between these models and human perception are rarely carried out. This thesis aims to investigate how well human observers can find defects in textured surfaces, over a wide range of task difficulties. Stimuli for experiments will be generated using texture synthesis methods and human search strategies will be captured by use of an eye tracker. Two different modelling approaches will be explored. A computational LNL-based model will be developed and compared to human performance in terms of the number of fixations required to find the target. Secondly, a stochastic simulation, based on empirical distributions of saccades, will be compared to human search strategies.

Contents

List of Tables

List of Figures

List of Symbols

	Texture Synthesis
$i(x, y)$	pixel in an image
h	height map
$\underline{n}(x, y)$	unit surface normal
ρ	surface albedo
θ	elevation of illumination
ϕ	azimuth of illumination
	$1/f^{\beta}$**-noise**
f	frequency
β	frequency roll-off factor
σ_{RMS}	RMS Roughness
	***Near-Regular* Surfaces**
ρ	density - textons per row
σ_j	standard deviation of texton displacement

Table 1: Table of symbols: texture synthesis.

	Itti and Koch's Saliency Model
Pyr_i	level i from the Gaussian pyramid
c	centre pixel level in centre-surround
s	surround level in centre-surround
δ	difference between levels: $s = c + \delta$
$C_{c,s}$	intensity contrast feature map for centre c, surround s.
	LNL-based Search Model
G	Gabor filter
S	activation map
$d(x, y)$	Euclidean distance from the current fixation location to (x, y)
F_d	distance weighted activation map
F	distance and IOR weighted activation map
k	a constant, $k = 0.0013$
t	fixation number
$I_t(x, y)$	IOR mask for fixation number t, centred on (x, y)
p_i	probability of fixating maxima i
	Stochastic Search Simulation
$p = f(\beta, r)$	probability of detecting a defect for given β and r
r	distance from current fixation location to target
f	linear-regression model
x	uniformally distributed random variable, $x \in [0, 1]$
t	fixation number

Table 2: Table of symbols: models

Chapter 1

Introduction

Automatic surface defect detection is one of the main applications of computer vision and many different approaches and methods have been put forward. However, the ability of the human vision system to detect surface defects has not been studied in a rigorous way and little effort has been made to investigate how well computer vision algorithms can mimic human behaviour. Hence the aim of this thesis is to bring together relevant work on visual search, saliency, perception and texture discrimination for the purpose of analysing modelling human defect detection.

1.1 Motivation

The motivation for this thesis draws on two disparate areas of research: *automated defect detection* in computer vision, and *visual search* in psychology. Previous work on automated surface defect detection appears to have neglected consideration of human perception. While many different image processing techniques have been put forward to tackle the problem, different methods are rarely compared with each other or against human performance. Therefore we have little way of knowing which methods are most suitable for a particular task or how they measure up to human perception. It is common for only the overall accuracy of the proposed algorithm to be published: details on where and why certain algorithms succeed or fail are frequently omitted. Generally only a few example images from the test database are given and there is little indication of how the difficulty of one database of example defects compares to another.

In psychology, the study of how human observers search for a target is called visual search. Most work on visual search has used arrays of discrete, abstract items as stimuli. By using more naturalistic stimuli we can learn more about how human observers carry out their searches. One such task, with real-world relevance, is the problem of finding a defect in a complex surface.

1.2 Goals

The goal of this thesis is to provide a rigorous framework for investigating defect detection algorithms. This will involve using parameterised synthetic surface textures for test sets. Psychophysical experiments will be conducted to explore how well human observers can find defects and the use of an eye-tracker will allow for a greater understanding of human search strategies. A computational model will be developed and compared against human performance. Finally human search strategies will be analysed and compared against a stochastic simulation.

1.3 Scope

This thesis is restricted in scope to only considering visual defects on synthetic (computer generated) surfaces. The use of computer generated stimuli removes the problem of obtaining a large number of physical samples of defective surfaces. Perhaps more importantly, as all the synthetic surfaces used in this thesis are fully parameterised, surface and target properties can be controlled and modified to give a large range of task difficulties. Critical defects such as hair-line fractures in airplanes are not considered.

The majority of the modelling work will be evaluated using random phase, $1/f^{\beta}$-noise surface textures. These surfaces were chosen because of (a) their simplicity and (b) $1/f^{\beta}$-noise processes appear frequently in nature. This is discussed in Chapter 3. I will examine how well a computational model, and the human visual system, can find small indents in these surfaces over a range of target depths, orientations, and surface roughnesses. In order to assess the generality of the model, some experiments using near-regular surface textures will also be used. These surfaces differ greatly from the $1/f^{\beta}$-noise surfaces in that they have a high degree of periodicity and structural phase information.

The problem of *target identification* lies outside the scope of the modelling work in this thesis. The aim is not to construct an automated defect detection algorithm. Rather, the aim is to simuluate human search behaviour and develop a computational visual search model. Therefore, the models investigated in this thesis are not defect detection algorithms, as they cannot make a decision as to whether a surface is defective or not.

1.4 Contributions

The main contribution of this thesis is to bring together relevant work on visual search, saliency, perception, and texture discrimination for the purpose of modelling how human observers detect defects on textured surfaces. This involves the introduction of synthetic surface textures as stimuli for visual search. To the author's knowledge, defect detection algorithms have not previously been rigorously compared to human performance. Specific contributions are outlined below:

- A review of the visual search and perception literature which is relevant to the problem of surface defect detection;
- The introduction of synthetic surfaces as visual search stimuli and an investigation into how surface properties affect the ability of a human observer to locate a defect;
- A comparison between human observers and a computational saliency model in a demanding defect detection task;
- The development of an LNL[1]-based search model which models human perception in terms of how noticeable a surface defect is; and
- An investigation in human search strategies and memory. This involves a comparison between human search strategies and a stochastic search simulation.

1.5 Overview

This thesis is organised into eight chapters. **Chapter 2** is in two halves and contains a literature review of the relevant literature from the fields of computer vision and

[1] *linear-nonlinear-linear*

perception. The computer vision review is mainly focused on discussing some of the many different methods that have been used for surface defect detection while the second half of the chapter gives a general overview of human perception and visual search for readers unfamiliar with this area.

Chapter 3 is concerned with visual search stimuli and texture synthesis methods. The chapter starts with a literature survey of the different classes of stimuli that have been previously used in visusal search experiments. The rest of the chapter gives details of the two classes of surface texture that will be used throughout this thesis: $1/f^{\beta}$-noise and *near-regular* textures.

The key chapters of this thesis are Chapters 4, 6 and 7. **Chapter 4** contains a review of *visual saliency*, in particular, the computational model developed by Itti and Koch [2000]. This is followed by an investigation of how human observers conduct *visual searches for defects on textured surfaces*. A series of experiments is carried out exploring how features such as surface roughness, regularity, and the orientation and contrast of the target affect how salient it is. Human performance is compared with the output from the predominant visual saliency model.

Chapter 5 contains a review of visual search models such as Guided Search, the Area Activation Model, and the Target Acquisition Model. This is followed by **Chapter 6** which outlines the development of an image processing search model based on the LNL framework and compares it with human performance in a second set of visual search experiments.

Search strategies and the role of memory are discussed in **Chapter 7**. In particular, human scan-paths are analysed to find out how systematic their search strategies are. Two experiments are carried out. The first one uses the moving target paradigm [Horowitz and Wolfe, 1998] and attempts to assess if memory is used to guide search. The second experiment invovles a comparison between human scan-paths and a stochastic search simulation. The simulation is based on a random walk which makes saccades from empirically obtained distrubtions. The target detection part of the stochastic search model is based on empirical results from a signal-detection experiment.

Finally, **Chapter 8** contains the overall conclusions of this thesis and outlines potential future work.

Chapter 2

Literature Review

This chapter contains two literature reviews. Section 2.1 concerns computer vision and reviews some of the different approaches that have been used to tackle the problem of automated surface defect detection. The related problem of texture discrimination is also discussed. The second half of this chapter, Section 2.2, contains a general overview of the processes behind human perception and introduces the field of visual search.

There are several, more in-depth, literature reviews throughout this thesis. A discussion of the various stimuli that have been used in search experiments can be found in Section 3.1 while Section 4.1 introduces models of visual saliency and examines the role of bottom-up processes in perception. Chapter 5 is given over to a discussion of models of visual search and Section 7.2 discusses the role of memory in visual search and how systematic human observers appear to be.

2.1 Machine Vision and Defect Detection

This section is primarily aimed at readers coming from a perception/visual search background who may not be familiar with many of the concepts used throughout this thesis. Readers familiar with this area may wish to skip to Section 2.2.

The scope of this thesis only concerns non-critical surface defects. As such, critical, non-visible, defects such as hair-line fractures in airplanes [Drury, 2002] are outwith the scope of this work. Instead, I am interested in surface defects on textured

surfaces and how noticeable they are to human observers. Examples of this would be defects in textiles and fabric (see Kumar [2008] for a review), fresh produce, such as apples [Leemans and Destain, 2004, Throop et al., 2005] and oranges [Aleixos et al., 2002], and ceramic tiles [Boukouvalas et al., 1995, Xie and Mirmehdi, 2005b].

This is important in manufacturing as defects can have a large effect on the price of a good: cosmetic surface defects can decrease the likelihood of consumers purchasing a piece of fruit [Thompson and Kidwell, 1998] and even minor blemishes can put consumers off buying apples [Yue et al., 2009]. A defect in a fabric will reduce its price by 45% to 65% [Sengottuvelan et al., 2008, Srinivasan et al., 1992]. Another interesting example is facial scarring [Simmons et al., 2009], where the visibility of the scar can be very important to the patient.

With the exception of Simmons et al. [2009] none of the examples above have considered how well human observers can identify different surface anomalies and what features make some defects more noticeable than others. While there has been some research into how well quality control inspectors can identify defects, a large range of different figures have been given. Several studies have claimed skilled visual inspectors can find only 70% of defects [Mak and Peng, 2006, Sari-Sarraf and Goddard, 1999, Sengottuvelan et al., 2008], although no details on how this figure was reached are given. Similarly, Schicktanz [1993] suggests detection rates of 60%-75% of significant defects while Bodnarova et al. [2000] and Smith [1993] suggested figures of 80% and 90% respectively.

Khasawneh et al. [2003] carried out an eye-tracking experiment to investigate human search strategies in a visual inspection task. However, the task used stimuli made up of discrete items (letters) and the study is somewhat naive. As we will see below (see Section 2.2), visual search is a large and complex field and Khasawneh et al. [2003] make no reference to the vast literature on the topic. Their main conclusion appears to be that there is no correlation between the area searched and task accuracy.

2.1.1 Supervised and Unsupervised Surface Defect Detection

Automated surface defect detection algorithms can be classified as either supervised or unsupervised. While these terms do not appear to be consistently used, supervised detection usually involves training the algorithm on specific defect(s) to be found

[Kumar and Pang, 2002]. Manufacturing methods often produce different defects that fit into pre-defined categories and in these cases, supervised defect detection works well. An example would be in the manufacturing of ceramic tiles where Boukouvalas et al. [1995] have identified cracks, bumps, depressions, pin-holes, dirt, drops, undulations and colour defects. Similarly, in fabric manufacturing certain defects such as mixed filling, mispicks, kinky filling and misreed are common [Kumar and Pang, 2000].

Unsupervised detection is a more general, and hence difficult, task where the properties of the defect are not known in advance. Some studies count training an algorithm on a defect-free example surface as supervised defect detection [Gururajan and Sari-Sarraf, 2006], while others count this as unsupervised [Xie and Mirmehdi, 2005a,b].

As we will see later (Section 2.2), these two methods have broad parallels with human vision: in visual search tasks an observer is typically asked to find a pre-specified target where as the study of bottom-up, visual saliency is more concerned with finding image regions that stand out from their surroundings.

Although many different algorithms have been developed and applied to the problem of automated surface defect detection, most of them fall into one of two broad categories: filter based and statistical. (A recent review by Xie [2008] classified detection methods into four groups: statistical, structural, filter-based and model-based, with statistical and filter-based methods being the most popular.) Some examples of these methods will be outlined below in Section 2.1.3. Before that however, there is a short discussion of the related problem of texture discrimination.

2.1.2 Texture Discrimination and LNL-Methods

The problem of automatic defect detection is related to that of *texture discrimination*, also referred to as *texture segmentation* and *texture segregation* [Bergen and Julesz, 1983, Bergen and Landy, 1991, Julese, 1981]. Texture discrimination typically involves separating an image into foreground and background regions. This problem is almost effortless for human observers and researchers have investigated which image features facilitate this process. (This is an example of a *preattentive* process: it can be carried out very quickly, in parallel across the whole visual field and without the need of attention.) Much of the modelling work has made use of

LNL models (linear-nonlinear-linear; also referred to as filter-rectify-filter (FRF) and the backpocket model [Chubb and Landy, 1991]). These models are based on properties of the functional architecture of the primary visual cortex [Bovik et al., 1990, Malik and Perona, 1990, Morrone and Burr, 1988, Randen and Husoy, 1999a,b]. See Randen [1999] for a comparative study of filter based approaches to feature extraction for textures. See Landy and Graham [2004] for an excellent review of the visual perception of texture.

As the name implies, the basic LNL-model has three stages. The first stage is linear and models the output of simple V1 cells. Several different filter banks have been used for this stage, such as Gabor filters [Daugman, 1980], differences of offset Gaussians (DOOG) [Young, 1985] and differences of offset differences of Gaussians [Parker and Hawken, 1988]. These three families of linear filters are very similar and there is little reason to choose one over another [Malik and Perona, 1990]. (The modelling work in Chapter 6 of this thesis will make use of Gabor filters.)

This is followed by a non-linear stage, followed by the second linear filter [Unser and Edena, 1990]. Unlike the first-order filter, which is based on properties of simple V1 cells, less is known about how to go about modelling these stages. A non-linear stage such as a half- or full-wave rectifier is needed to distinguish textures with identical spatial averages. While a single rectifier allows the model to match human performance with many texture pairs, additional non-linearities are required to distinguish between textures composed of opposing textons [Malik and Perona, 1990]. Some form of adaptation has been put forward as an important mechanism in human vision [Graham et al., 1989, Sutter et al., 1989]. The second linear filter is usually taken to be some sort of energy pooling filter and is often implemented as either a Gaussian smoothing filter or a bandpass filter [Randen, 1999].

The LNL framework outlined above has proven very successful in modelling pre-attentive texture discrimination. [Malik and Perona, 1990] have compared an LNL-mode1 to psychophysical results by Krose [1986] and Gurnsey and Browse [1987]. Furthermore it is flexible, and easily expanded without being computationally intensive.

2.1.3 Filter-Based Defect Detection Methods

While a large variety of methods have been applied to the problem of automated surface defect detection, (such as statistical, morphological and model based) one

of the most popular types of tool is filter based algorithms [Xie, 2008]. Many of the algorithms share similarities with the LNL-framework discussed above, even if this connection is not always made explicitly. (For more comprehensive reviews of defect detection algorithms, see Chin [1988], Kumar [2008], Newman and Jain [1995], Song et al. [1992], Xie [2008].)

Sometimes a single optimal filter will be used for a supervised detection task, tuned to the properties of the defect [Bodnarova et al., 2002, Kumar and Pang, 2002, Mak and Peng, 2006]. Kumar and Pang [2002] have proposed an algorithm for selecting the optimal Gabor filter, from a bank of filters, based on the cost function used by Tang et al. [1995]. A test image containing a defect is divided into K non-overlapping squares of size $l \times l$. Each filter from the filterbank is applied to each of these subregions and the average output for the ith filter in the kth square is given by:

$$D_k^i = \frac{1}{l^2} \sum_{(x,y)\in k} I_i(x,y) \tag{2.1}$$

where D_{max}^i and D_{min}^i are defined as the maximum and minimum average outputs among the D_k^i for $k = 1, \ldots, K$. These are then used to define the cost function:

$$J(i) = \frac{D_{max}^i - D_{min}^i}{D_{min}^i} \tag{2.2}$$

The filter i that gives the biggest J is taken as the best filter to find the defect.

Hou and Parker [2005] propose a similar method which uses Support Vector Machines to pick an optimal set of Gabor filters. Another example is Sobral [2005] who based a detection algorithm on wavelet sub-band decompositions and applied it to the problem of leather inspection. An alternative method is to represent the defect-free woven fabric texture with a single optimal Gabor filter [Mak and Peng, 2006]. This works well for some fabrics as it exploits the highly regular periodic nature of the textile. The reconstructed image is then compared to the test image and any large differences are assumed to be due to defects. This method makes use of the highly regular nature of woven fabrics. A similar method has been put forward by Jasper et al. [1996].

Unsupervised methods typically use filter banks and compare feature vectors between a learnt defect-free surface and the test surface [Escofet et al., 1998, Kumar and Pang, 2000, 2002]. While providing more flexibility than optimal filters, filter-banks generate large amounts of data and can be computationally intensive. These methods appear to be common with fabric defect detection as the periodicity of the yarns provide valuable information [Chan and Pang, 2000].

Kumar and Pang [2000] have used a filter-bank of real Gabor filters (four orientations and four spatial scales). They use a nonlinear energy function based on a rectified sigmoidal function [Jain and Farrokhnia, 1991]. The 16 filter responses are then compared to the mean values from a defect-free training sample. These are then combined cross-orientation and then cross-scale using *image fusion* techniques [Casasent, 1997], before finally being binarised to give a segmentation of any defects in the test surface. The algorithm was tested on small photographs of fabrics, and simple synthetic images, and was found to perform well. However, no measurements of how noticeable the defects were was given.

2.1.4 Discussion

In the brief review above we can see that a large range of different approaches have been put forward for automatic surface defect detection. However, very little work has been carried out in comparing different computer algorithms to one another, or to human observers. Similarly, the problem of measuring how *noticable*, or *salient*, a defect is has largely been ignored. Several different performance rates are given for human performance but the details are not given. What affects human performance? Can we model this?

These questions have real importance in terms of manufacturing and commerce. People will not pay full price for defective goods, even if the defect is superficial. This leads to waste, especially if the goods are perishable. On the other hand, consumers will buy B grade stock if it is suitably reduced in price. However, it would be against a retailers interests to discount stock which has superficial defects that would go unnoticed by a consumer.

2.2 Human Perception and Visual Search

In the above section I have given an overview of how computer vision systems carry out defect detection tasks. The second half of this review chapter will give an introduction to *visual search*: how human observers carry out a search for a target. In this thesis, the target will be a defect on a textured surface. This section is primarily aimed at readers coming from a computer-vision background who may not be familiar with many of the concepts used through the rest of this thesis. As such, readers familiar with this area may wish to skip to Chapter 3.

2.2.1 Foveal Vision, Saccades & Fixations

The centre of our retina, the fovea, has greater acuity than the periphery [Findlay and Gilchrist, 2003, Chapter 2]. Because of this, we make many eye movements in order to bring relevant parts of the visual scene into the fovea. There are three different types of gaze shifting movements: *saccadic movements, smooth pursuit* and *vergence movements*. In this thesis, I will only be considering saccades: they are fast, ballistic, eye movements in which both eyes move simultaneously. Saccades are executed in order to fixate new regions of the visual scene. Over the past decade the use of eye-trackers has become more common. These track an observer's gaze, fixations and saccades. They have been used to give us a greater understanding of how vision works during reading [Rayner, 1998] and are increasingly being used in visual search and scene perception experiments. See Rayner [2009] and Findlay and Gilchrist [2003] for reviews.

2.2.2 Visual Search

In the psychology of perception, visual search usually involves human observers searching a display for a designated target. The task is usually to decide if the target is present in the display or not and observers are instructed to respond as quickly and accurately as they can. Displays consisting of a collection of discrete search items - typically letters or abstract shapes - are commonly used as stimuli and the target will be uniquely defined by one or more features such as colour, size and orientation. See Figure 2.1 for examples.

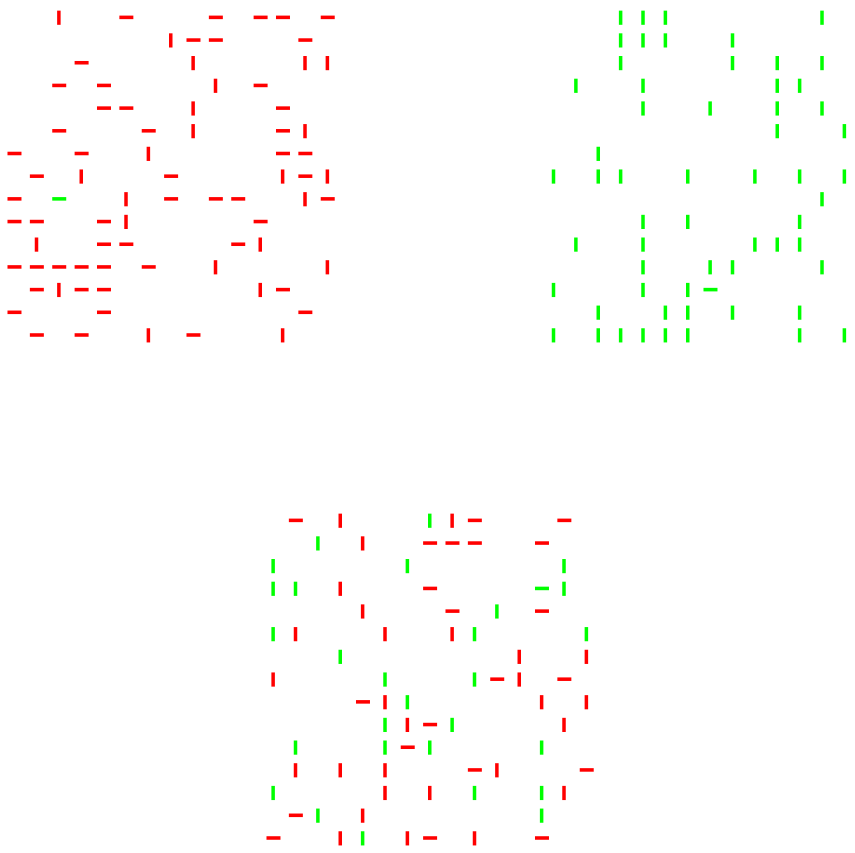

Figure 2.1: Examples of the type of stimuli traditionally used in visual search tasks. The top two images are examples of *feature searches*: the target is defined by a unique feature, colour and orientation respectively, and can be found with ease. (In the top left image the target is the green bar, while in the top right image, the target is the single horizontal bar.) These searches are sometimes referred to as *parallel* or *pop-out* searches. The bottom image is an example of a *conjunction search*. The target is defined in terms of both colour (green) and orientation (horizontal) and the search is harder. These are also referred to as *serial* searches.

One of the main analytical tools deployed in previous work on visual search is the gradient of the reaction time (RT) against set size slope. Early work by Treisman and Gelade [1980] classed visual searches as either *feature* searches or *conjunction* searches. This was based on the observation that if the target is defined by a unique feature (such as searching for a red item among green) then the time taken to find it appears not to be affected by adding more distracters. If the target is defined by a combination of features (such as in Figure 2.1) then observer reaction times increase with the addition of more distracters. Feature searches were said to be carried out in parallel while conjunction searches required a serial item-by-item search. Further evidence for this explanation comes from the fact that RT v Set Size slopes typically show a 2:1 ratio on average between *target absent* and *target present* trials [Chun and Wolfe, 1996, Kwak et al., 1991, Treisman, 1991].

This classification has been shown to be overly simplistic as there are many examples of conjunction searches that can be carried out faster than would be expected by a serial search [Cohen, 1993, Nagy and Sanchez, 1990]. A meta-study by Wolfe [1997] aggregated trials from 2500 experimental sessions and found no evidence for a bimodal distribution of RT v Set size slopes. This has led to the more relaxed statement that some searches are *efficient* while others are *inefficient* and task difficulty can be located anywhere on a continuous range from one extreme to the other. See Wolfe [1998] for an overview. Verma and McOwan [2008] have strengthened this argument by constructing an algorithm to generate stimuli for a texture discrimination task for a whole range of task difficulties.

Perhaps the most influential search model is Wolfe's Guided Search (GSM). This has been developed over the last 20 years and is now in its 4th version [Cave and Wolfe, 1990, Wolfe, 1994, 2007, Wolfe and Gancarz, 1996, Wolfe et al., 1989]. While this model will be discussed in detail in Chapter 5, a brief overview is given here. As the name implies, the model works on the assumption that search is *guided*. This means that the model will direct attention towards search items that have features in common with the target. So, in the conjunction search example in Figure 2.1 the model will search among the horizontal and green elements. This is achieved by first creating an activation map from the stimuli: search items that share characteristics with the target are given larger weights, while lower values are given to items that are dissimilar to the target. The activation map is assumed to be generated pre-attentively and is computed in parallel across the whole stimulus. Search items with high activation are then attended to in a serial manner until the target is identified. A simple inhibition of return process is used to stop the model repeatedly attending

to the same local maximum. In the case of feature (efficient) searches the target is very different from distracters and the activation map can give a large response to the target, allowing the serial attentional mechanism to go straight to the correct search item.

2.2.3 Serial Search versus Parallel Search and Signal Detection Theory

Unfortunately the issue of serial versus parallel search has been slightly confused by a split in methods in the literature. While the reaction time studies outlined above are typical, there is also a body of work on visual search that applies signal detection methods to the problem. These studies use short stimulus display times (typically around 200ms, which is slightly less than the duration of an average fixation) and measure an observer's accuracy. (See Palmer et al. [2000] and Verghese [2001] for reviews). Rather than assuming a two stage model (such as GSM), the signal detection theory (SDT) models propose a parallel stage followed by a decision rule. While these two camps are not mutually exclusive, the different assumptions have made comparisons problematic: GSM is geared towards explaining differences in reaction times in search tasks with an unconstrained time limit, whereas SDT accounts for varying accuracy in search tasks with a very short time limit. In order to account for reaction times in search with eye movements, SDT needs a number of additional assumptions.

In SDT models the search items in a display are represented as noisy random variables. The greater the difference between target and distracter the further the means of the random variables will be from each other, and the smaller the probability of a sample from the target distribution being smaller than a sample from the distracter distribution. The effect of increasing set size on accuracy elegantly drops out of this model, as increasing the number of distracters present in a display means we have to take more samples from the distracter distribution. Hence the probability of any one of them being greater than the target sample is increased.

Guided search models on the other hand, assume that items (or small subsets of items) are inspected in a serial manner. On the other hand, SDT accounts clearly show that target detection can, and does, occur in parallel across the whole stimuli. However, what is often overlooked by advocates of SDT accounts is that the nature of foveal vision enforces a serialism on any prolonged search process.

2.2.4 Overt versus Covert Attention

In a similar way to the split in the literature outlined above, there is also a distinction between studies investigating *covert* attention and those looking at *overt* attention. Overt visual attention is essentially the act of looking and fixating on a particular object or region of interest in the visual field. Covert attention on the other hand is the process of directing our mental attention to a region of the visual field without shifting our gaze, i.e., *'looking out the corner of the eye'* [Findlay and Gilchrist, 2003]. It has been shown that human observer can use covert attention independently from overt attention, [Posner, 1980, Sperling and Melchner, 1978], and even carry out visual searches without eye movements [Klein and Farrell, 1989]. Because of these findings there is a large body of literature concerning the deployment of attention, with little mention of foveal vision or eye movements.

For example, the influential Guided Search Model [Cave and Wolfe, 1990, Wolfe et al., 1989] is only concerned with the deployment of attention and makes no mention of saccades or foveal vision. In fact, Wolfe [2007] argues that modelling covert attention is more important than eye movements as eye movements are not required for visual search. Furthermore, once acuity factors have been controlled for, reaction times are comparable between search with and without eye movements. For reviews on attention, see Cave and Bichot [1999], Pashler [1998], Reynolds and Chelazzi [2004], Wright [1998].

However, in their book, *Active Vision*, Findlay and Gilchrist [2003] argue that not enough importance has been given to overt attention. (Similar arguments are also given by Rayner [2009] and Zelinsky [2008]). In most natural visual tasks, a shift of covert attention is usually followed by a shift of overt attention (a saccade) to the same region [Henderson, 1993, Irwin and Zelinsky, 2002, Rayner et al., 1978]. Motter and Holsapple [2007] have used a model of visual search to argue that there is no evidence of multiple shifts of covert attention during a single fixation.

2.2.5 Category Search

Surface defect detection can be thought of as an example of a category search task. In later chapters of this thesis observers will be asked to find a type of defect, such as an elongated indent, without knowing its exact appearence. It has been shown that search can be guided in a top-down manner if the observer knows the category,

but not the exact features, of the target [Bravo and Farid, 2009, Castelhano et al., 2008, Schmidt and Zelinsky, 2009, Yang and Zelinsky, 2009]. For example, Yang and Zelinsky [2009] investigated category search using images of teddy bears as targets and photographs of other common objects as distracters. They compared search guidance between specific searches (when the exact target is previewed) and category search (when the observer only knew that they were looking for a photograph of a teddy bear) and found that while specific search was faster, the category search still appeared to be guided. There were fewer fixations on distracters, and more initial saccades to the target, than would be expected with a random searcher. (Also see Oliva et al. [2003] and Zelinsky [2003].)

2.2.6 Finding Rare Targets

Finally, it is important to note that the prevalence of different target types has been shown to influence accuracy in visual searches [Fleck and Mitroff, 2007, Rich et al., 2009, van Wert et al., 2009, Wolfe, 2007, Wolfe et al., 2005]. Although this will not be discussed further in this thesis, it is important to note as in most real-life defect detection tasks we would expect the ratio of defective to defect-free surfaces to be very low. Visual inspection is difficult in practice because defects are generally rare. As targets become less common, people miss more. However, the implementation of a decision rule in a search model is outwith the scope of this thesis.

2.3 Conclusions

In this chapter I have reviewed how both humans and computer vision models search for defects on textured backgrounds. The main conclusion is, perhaps surprisingly, that the problem of finding a defect on an otherwise homogeneously textured surface has not been looked at in detail in the field of visual search: most previous work has concentrated on finding a target item amongst discrete distracters or searching photographic stimuli for a pre-defined target. Similarly, despite the fact that most computer defect detection systems are designed to replace human quality control inspectors, the issue of human performance in defect detection has been largely overlooked.

Chapter 3

Methods: Texture Synthesis

A problem with analysing the performance of various automated surface defect detection methods is that it is often difficult to find suitable surfaces on which to run the algorithms. For example, Iivarinen [2000] uses one image to train his algorithm, and only tests it on three examples of defective surfaces. Similarly, Amet et al. [2000] used a set of 36 fabrics, (256×256 pixels) to test an algorithm based on sub-band domain co-occurrence matrices. Hou and Parker [2005] used 50 images from the Brodatz [1966] Texture Database and Kumar and Pang [2000] only used 9 images to test their defect detection model. If a small set of test images is used then the algorithm's accuracy cannot be reliably measured and only a small range of defects can be tested.

Even if a large set of defective surfaces can be acquired, (such as Gururajan and Sari-Sarraf [2006], who used a test dataset of 360 fabric images; and Xie and Mirmehdi [2005b] who used 1500 defective ceramic tiles) there is still an issue with how an algorithm's performance should be evaluated. Currently defect detection algorithms are typically assessed by reporting the percentage of defects in the test set it can successfully find (along with the number of false-positives). Little indication is ever given of how varied the test set is, or how difficult the defects would be for a human observer to find. With no standards to measure these difficulties against, it is difficult to make an informed comparison between different algorithms without testing them all on the same dataset.

A solution to this problem is to use virtual, synthetic surface textures. This allows for the creation of a wide variety of test surfaces and defects. As the surfaces

can be created with a large, yet controlled, range of task difficulties, they have the potential to also be useful stimuli for the study of human perception and visual search. A review of stimuli that have previously been used in this field is given in Section 3.1, followed by Section 3.2 which identifies and specifies two very different classes of texture. While synthetic images have been used before to analyse defect detection algorithms [Kumar and Pang, 2000], previous work has not considered illumination models: Kumar and Pang simply used a grid as a test image and created a defect by making one of the lines thicker.

3.1 Stimuli used in Visual Search Experiments

The majority of previous work into visual search has used stimuli consisting of an array of discrete search items (see Section 2.1, Figure 2.1). However, more complex stimuli have increasingly been used in an effort to understand how search and object recognition are carried out in more naturalistic tasks. Several studies have used stimuli in which the abstract geometric shapes traditionally used are replaced with photographs of objects [Zelinsky, 1999] or simple line drawings [Biederman et al., 1988, Levin et al., 2001]. Levin et al. investigated which features are used by human observers when searching for a target from a broad category. To do this they carried out an experiment which used two categories of line drawings of either animals or everyday household objects. They found that observers were very good at this task and were employing some form of guidance. Newell et al. [2004] used rendered 3D objects in investigate memory processes and found evidence for object based memory during search.

A more common approach for creating naturalistic stimuli is to use photographs of scenes. This approach is proving to be increasingly popular [Aks and Enns, 1996, Biederman, 1973, Brockmole and Henderson, 2006, Henderson et al., 1999, 2008, McCarley et al., 2004, Neider and Zelinsky, 2006a, Oliva et al., 2004, Pomplun, 2006, Zelinsky, 2001, Zelinsky et al., 1997], However, analysing the results from experiments with photographic stimuli can prove complicated and the stimuli cannot be easily controlled or parametrised. Unlike experiments using arrays of search items, there is no easy way of creating different trials of similar difficulties when using photographs of scenes.

A related, simpler task is *search for a target on a background*. Perhaps surprisingly, this task has received comparatively little attention. Wolfe et al. [2002]

and Neider and Zelinsky [2006b] have investigated how the addition of a complex background affected reaction time versus set size slopes. They concluded that a complex background might slow the accumulation of information in the object identification stage, perhaps because the search items were not cleanly segmented from their surrounding backgrounds in the initial object segmentation phases. Only in Wolfe et al.'s final experiment, when the search items and background were designed to be very similar to each other was an increase in search slopes observed. Neider and Zelinsky [2006b] carried on this line of work with a series of experiments using more complex stimuli designed to investigate the effect of target-background similarity (TBS). They used photographs of children's toys as search items and constructed 'camouflage backgrounds' from the target item by tiling an $n \times n$ pixel patch from the target item. By increasing n, the TBS can be modified while leaving the distracter-background similarity constant. They carried out a series of eye tracking experiments but failed to find any conclusive results or pattern behind the scan-paths.

The use of noise has also been fairly common in visual search and psychophysical experiments [Burgess, 1985, Burgess and Colborne, 1988, Burgess et al., 1981, Levi et al., 2005, Myers et al., 1985, Najemnik and Geisler, 2005, 2008, 2009, Park et al., 2005, Rajashekar et al., 2002, 2004, 2006, Swensson and Judy, 1981, Tavassoli et al., 2007a,b,c, 2009]. The target is typically a simple geometric shape (square, dipole, Gabor patch) which is embedded in the noise. The task difficulty depends on the signal to noise level. Visual search in noise will be discussed further in Section 5.3.1.

3.1.1 Conclusion

This literature review provides a survey of the different stimuli that have been used in previous work on visual search. These stimuli can be broadly placed into two groups: abstract, synthetic images, and photographic images. While abstract stimuli, such as arrays of discrete items, have the advantages of being easy to create, control and analyse, they do not appear naturalistic and can lack the visual complexity that is present in many real world examples. On the other hand, photographic images provide rich and varied stimuli but sacrifice the ease of analysis and control that parametrised stimuli provide.

3.2 Texture Synthesis

One way to get round the problem of obtaining suitable stimuli for investigating surface defect detection is to create synthetic surfaces using computer algorithms. In order to create images for experiments a two step process is used. First the surface texture is created, represented as a $n \times n$ height map. Then an illumination model is used to render the image, giving it a naturalistic appearance.

Another advantage of using rendered surface textures as stimuli is that they allow for an investigation of the control of attention. By using a task that eliminates the possibility of top-down influences on visual search, bottom-up processes are isolated and can be tested against theoretical models. Finally they possess a natural appearance, yet have precisely controlled properties, allowing for many stimuli to be created with a given set of parameters.

Details are given below, first for the illumination model, and then for the construction of random-phase noise (Section 3.2.2) and height maps for *near-regular* textures (Section 3.2.3). Creating defects is discussed in Section 3.2.4.

3.2.1 Illumination

The three dimensional height maps that represent the surface textures are rendered to generate images of surfaces under a specified illumination. This stage is important, as a surface can have drastically different appearances under different illumination conditions [Chantler, 1995]. I 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, and it is modelled by the dot product:

$$i(x, y) = \lambda \rho(x, y) \underline{n}(x, y).\underline{l} \tag{3.1}$$

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 and the strength of the light source is denoted by λ. The surface normal, $\underline{n}$, is estimated by:

$$p(x,y) = h(x,y) - h(x-1,y) \qquad q(x,y) = h(x,y) - h(x,y-1) \tag{3.2}$$

$$\underline{n} = \frac{1}{\sqrt{1+p^2+q^2}}[p,q,1]^T \tag{3.3}$$

where p and q are discrete estimators of the surface's partial derivatives in x and y respectively. The illumination vector, $\underline{l}$, is given by:

$$\underline{l} = (sin\theta cos\phi, sin\theta sin\phi, cos\theta) \tag{3.4}$$

where $\theta = 60°$ and $\phi = 90°$ are the elevation and azimuth of the illumination direction. These are held constant throughout this thesis. Self-shadowing is implemented by setting all negative values of i to zero. Cast shadows are ignored for a variety of reasons. Firstly, cast shadows are rarely produced by surfaces with low relief unless the elevation of illumination is very low. Also, without a more sophisticated illumination and rendering model (incorperating inter-reflections, diffuse directional light sources, etc) cast shadows will look unrealistic. Finally, the inclusion of cast shadows would complicate both the analysis of human performance and the development of a computational model.

3.2.2 $1/f$-Noise Textures

The main class of texture that will be considered in this thesis is $1/f^{\beta}$-noise (where f is frequency). The process of $1/f^{\beta}$-noise occurs frequently in nature and provides a good approximation to the power spectra of many images of natural scenes [Field, 1987, van der Schaaf and van Hateren, 1996, Voss, 1988]. Balboa and Grzywacz [2003] have measured and compared the power spectra of atmospheric and underwater natural images and found that they have frequency fall-offs of around -2 and -2.5 respectively.

These surface textures are produced by using $1/f^{\beta}$-noise to model the *height function* of surfaces. It is important to emphasise that the stimuli are not created directly from $1/f^{\beta}$-noise, as has been done in other studies [Kayser et al., 2006, Rajashekar et al., 2002] but are instead created by constructing height maps which are then rendered using a lighting model that implements Lambert's Cosine Law. See Figure 3.1 for an example of a height map and the corresponding rendered image.

Early work was by Voss [1985, 1988] was based on the work by Mandelbrot [1983] and used simple fractal algorithms to generate mountainous landscapes. Although these surfaces will be modelled in the frequency domain, they can also be implemented in the spatial domain: Perlin noise is a widely used computer graphics technique that is very similar to $1/f$-noise processes [Perlin, 2002, Perlin and Hoffert, 1989].

These surfaces are specified with two parameters: the spectral roll-off β and the RMS-roughness, σ_{RMS}. (The RMS-roughness is the *root mean squared*: $\sigma_{RMS} = \sqrt{\frac{1}{n}\sum h(x,y)^2}$.) They can be generated in the Fourier domain with power spectrum given by:

$$S(u,v) = \frac{k}{(\sqrt{u^2+v^2})^{\beta}} \tag{3.5}$$

where k is the scaling factor used to give the required σ_{RMS}, and u and v are the Cartesian coordinates of the power spectrum [Linnett, 1991, McGunnigle, 1998].

Padilla et al. [2008] have explored the effect these parameters have on the perceived surface roughness of $1/f^{\beta}$-noise surfaces. Observers were shown pairs of animated, rendered surfaces and a method of adjustment was used to determine the relationship between β, σ_{RMS} and perceived roughness. This allowed them to construct an estimator for perceived roughness, based on the variance of a bandpass filter specified as a Gaussian in the frequency domain. (Also see Padilla [2008].)

3.2.3 Near-Regular Textures

A second class of surface texture will also be considered in this thesis: *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 positions of the texton elements [Liu et al., 2004a,b]. In contrast with the $1/f^{\beta}$-noise textures, these surfaces are highly structured and periodic.

As with the $1/f^{\beta}$-noise surfaces, these surfaces are created by first synthesising a height map and then rendering it using Lambert's cosine law. The height map is

created by placing ellipsoidal shaped textons (Equation 3.6) at regular intervals on a flat surface.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{3.6}$$

Two texton densities were used: $\rho = 1.875$ and $\rho = 2.461$ textons per degree of visual angle. The textons were randomly varied in two ways to give a near-regular texture: size and σ_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 3.6. A random offset was applied to each texton by adding a normally distributed error to its centre point. By varying the standard deviation of this error, $\sigma_p \in \{0, 1/2, 1, 2\}$, the regularity of the underlying lattice can be varied. 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. Example textures can be seen in Figure 3.2.

3.2.4 Creating Defects

Given how different the two classes of textures discussed above are from each other, a different type of defect will be used for each of them. For the near-regular textures, a simple way of adding a defect to a surface is to remove a texton from it. For the $1/f^{\beta}$-noise surfaces a small indent will be made in the surface by subtracting a three dimensional ellipsoid from the height map (See Equation 3.6). To make the indent appear more realistic it was first convolved with a two dimensional smoothing filter, B, to soften the hard vertical edges:

$$B = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

It was then cut out of the three dimensional surface, with the depth being adjusted so that a small hole with volume $\approx 10mm^3$ was created. (Some examples are given in the following chapter: see Figures 4.1 and 4.2 for an example of a smooth and rough surface respectively.)

3.3 Conclusions

In this chapter I have identified two classes of synthetic surface texture. These will be used in the following chapters to investigate how human observers search for surface defects and how computational models perform in comparison. The main advantage of using synthetic surfaces over photographs of surface textures is one of control: as the surfaces are fully parametrised, the degree to which the defect is noticeable can be easily controlled and many different, yet equivalent trials can be created for any given task difficulty.

In the next chapter I will explore how human observers cope with a series of defect detection/visual search tasks and will compare human performance with a popular model of low-level vision [Itti and Koch, 2000].

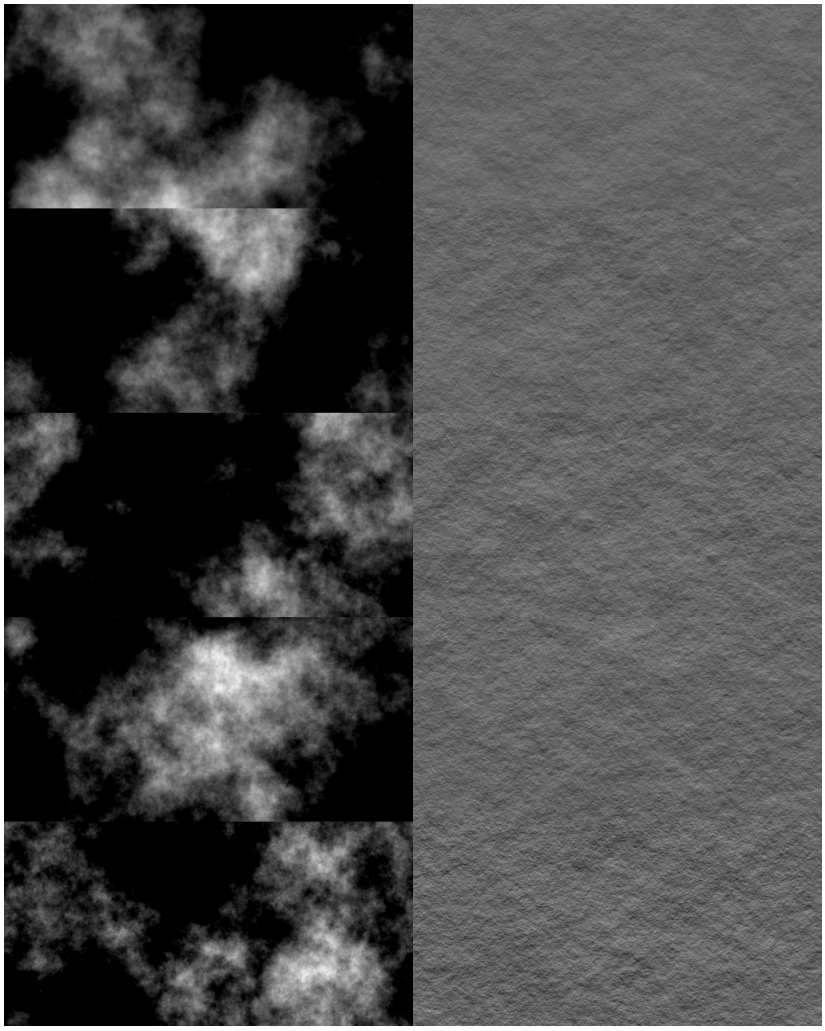

Figure 3.1: Examples of $1/f^{\beta}$-noise surfaces for $\beta \in \{1.55, 1.60, 1.65, 1.70, 1.75\}$ (ordered bottom to top) and $\sigma_{RMS} = 1$. Height maps are shown on the left, while the right column shows the corresponding rendered images. As β decreases the surface becomes rougher. As with all stimuli in this thesis, the illumination conditions were held constant with $\theta = 60°$ and $\phi = 90°$.

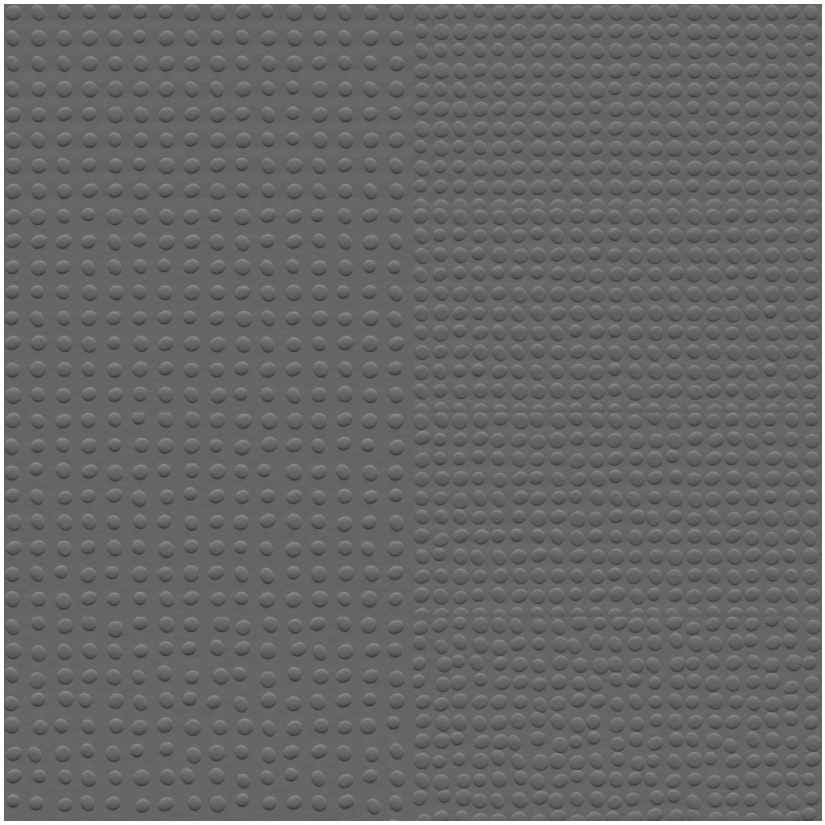

Figure 3.2: Examples of a near-regular texture for $\sigma \in \{0, 1/2, 1, 2\}$ (top to bottom) and $\rho = \{1.875, 2.461\}$ (left to right).

Chapter 4

Visual Search for a Defect on a Homogeneous Surface

In this chapter I will explore how human observers perform in a search task involving a target on a homogeneous surface texture. I will investigate how surface and target properties affect our ability to find a defect in a forced choice target absent/present task using the two different surface textures described in the previous chapter. For the $1/f^{\beta}$-noise surface textures the effect of varying surface roughness, along with the depth and orientation of the target, will be explored, while for the near-regular surfaces I will vary texton density and the degree of regularity. Altogether, four experiments will be carried out. The use of an eye-tracker will allow for search strategies to be investigated and a computational saliency model [Itti and Koch, 2000, Walther and Koch, 2006] will be run on the experimental stimuli and compared to the results from the psychophysical experiment.

This chapter starts with a discussion of the concept of visual saliency, computed by bottom-up visual processes (Section 4.1). This review is centred around Itti and Koch's [2000] saliency model and the extent to which it can explain visual search paths. This is followed by a methods section in which the procedures for the experiments in this, and Chapter 6, are given (Section 4.2). The core of this chapter contains a series of four experiments designed to investigate how well human observers can carry out a visual search task on the textural stimuli discussed in the preceding chapter (Section 4.3). Finally, results from the human observers are compared to the performance of the saliency model in Section 4.4.

4.1 Literature Review: Visual Saliency

Models of visual saliency attempt to capture the effect of bottom-up processes on the allocation of visual attention, and hence fixations and saccades. The concept of a saliency map was introduced by Koch and Ullman [1985] and further developed by Itti et al. [1998] and Itti and Koch [2000]. In a review of computational modelling of visual attention Itti and Koch [2001] identified five components deemed essential for such models: the computation of visual features pre-attentively across the entire visual scene; the integration of these feature maps resulting in a single attentional control command; generating attentional scan-paths; the interaction between covert and overt attentional deployment; and the interplay between scene perception and attention. However, much of the work on visual saliency has equated covert with overt attention (either implicitly or explicitly) and the relationship between scene perception and attention has frequently been overlooked in preference for studying the effects of low level, bottom-up features such as contrast, orientation and colour (although see Tatler [2009] for recent developments).

There have been several attempts to formalise the concept of saliency in terms of theoretical computational principles: for example, Itti and Baldi [2006, 2009] propose a definition of saliency in terms of Bayesian surprise and suggest that attention is attracted to visually surprising signals. An alternative approach has been taken by Bruce and Tsotsos [2006, 2009] who explore the hypothesis that saliency computation aims to maximise the amount of information available to the visual system while a third hypothesis is based on the idea that saliency should be optimal in a decision-theoretic sense. This idea was put forward by Gao and Vasconcelos [2005], who called it discriminant saliency, and further developed by Gao et al. [2008].

4.1.1 Saliency in Search: Attentional Capture

Although visual search is a top-down task, as we are looking for a predefined target, most theoretical search models contain a significant bottom-up contribution. (This is discussed in greater detail in Chapter 5.) Furthermore, in some ways defect detection can be thought of as a bottom-up task: a surface defect is not necessarily pre-defined. Rather, we are given some homogeneous surface and we want to know if there is a conspicuous anomaly present.

The effect of task irrelevant bottom-up information, such as a non-target singleton in an array of search items, is referred to as attentional capture (see Egeth and Yantis [1997], Rauschenberger [2003], Theeuwes [1993] for reviews). For example, Theeuwes et al. [1998] have shown that a highly salient coloured singleton will capture attention in a search unless the colour of both the singleton and target are known in advance. However the degree to which top-down processing can override stimulus driven saliency is still an open question and an active research area [Hickey et al., 2006, Lamy and Zoarisa, 2009, Sawaki and Katayama, 2008].

4.1.2 Itti and Koch's Visual Saliency Model

Itti and Koch's [2000] saliency model has been widely used since publication and it has been the foundation for many other models [Frintrop, 2006, Gao et al., 2008, Navalpakkam and Itti, 2007, Peters et al., 2005]. The model works on the assumption that our attention is attracted to regions in the visual field which are salient, i.e. they differ from their surrounds in terms of low-level features such as luminance and orientation. There are many examples of this theory in action in the natural world such as animals using bright colours to attract mates. In urban environments traffic lights and emergency exit signs are brightly coloured to stand out from their backgrounds.

The model is based on three visual features - luminance contrast, orientation and colour- with each feature computed over a multi-scale Gaussian pyramid, Pyr_s, where s is a scale level of the pyramid [Burt and Adelson, 1983]. These features are assumed to be computed pre-attentively and when normalised and combined give the resulting saliency map which is then used to control shifts of attention and saccades.

In order to mimic biological vision each feature is computed in a centre-surround structure. In practice this involves computing differences between fine and coarse scale feature responses. If the centre pixel is at scale c then the surround scales are taken to be $s = c+\delta$ with $c \in \{2,3,4\}$ and $\delta \in \{3,4\}$. (Note: in the implementation used below I take $c \in \{1,2,3,4\}$ and $\delta \in \{1,2,3\}$ - see Section 4.2.4) Seven features are computed in total: one for intensity (luminance) contrast, four orientation contrast channels and two for red/green and blue/yellow double-opponent channels. There is a variety of evidence for each of these: see Leventhal [1991] for intensity

contrast; DeValois et al. [1983], Tootell et al. [1988] for orientation and Engel et al. [1997], Luschow and Nothdurft [1983] for detail on colour channels.

The simplest feature is intensity contrast which is defined as the absolute difference between pixels in pyramid levels: $C_{c,s} = |Pyr_c - Pyr_s|, \forall c, s$. While this is conceptually the same as a difference of Gaussian bandpass filter, the implementation in the Matlab version developed by Walther and Koch [2006] relies on interpolation due to the sub-sampling introduced by the Gaussian pyramid. Local orientation is computed by simply applying Gabor filters to the different levels of the Gaussian pyramid, orientated at $0°, 90°, 180°$ and $270°$. The use of colour is outwith the scope of this thesis so the colour channels will not be discussed.

The model now faces a large signal-to-noise problem with 42 feature maps potentially containing information on salient signals. (1 contrast + 4 orientations + 2 opponent-colours = 7 feature channels, and 6 spatial scales were used, giving a total of 42 maps.) Simply normalising each feature map before summing has been shown to perform poorly [Itti and Koch, 1999]. The saliency model deals with this problem by means of a within feature spatial competition scheme. In practise this entails iteratively convolving each feature map with a large difference of Gaussian filter to simulate excitation and inhibition, adding the result to the original feature map and setting negative values to zero. This has the effect of reducing the weight of feature maps with numerous local maxima while exciting maps with only a few isolated peaks.

After this normalisation process has been applied to each feature map, a cross-scale summation is carried out resulting in three conspicuity maps, which are further normalised before being summed to give the final saliency map. The focus of attention (FOC) is assumed to be directed towards the most salient image location, which corresponds to the maximum of the saliency map, as determined by a winner-takes-all neural network (WTA) [Koch and Ullman, 1985]. In order to allow for shifts of attention, an inhibitory feedback from the WTA array to the saliency map is introduced. Excitatory lobes (half width of four times the radius of the FOA) are also included in order to model the visual system's tendancy to favour the closer of two equally conspicuous objects. A sequence of attentional shifts can be generated by repeated application of the WTA and inhibition of return (IOR) algorithms. Note, while the saliency algorithm is initially described as modelling visual attention it has frequently been taken as a model of saccades and fixations. It also fails to take foveal vision into account.

4.1.3 What behaviour can the model explain?

Since the publication of the model it has been implemented and compared to human performance on a wide variety of visual stimuli. Itti and Koch [2000] showed that it responded to visual search arrays (red/green, horizontal/vertical bars) in the same manner as a human observer with pop-out search exhibited in feature searches while conjunction searches required serially stepping through search items. Itti and Koch also carried out an experiment comparing human and model performance in a visual search task involving landscape photographs with a military vehicle acting as the target. In order to compare the model's performance with human reaction times the model was assumed to make three saccades a second. However there the correlation between human and model performance was poor with the model outperforming human observers in the majority of trials.

Parkhurst et al. [2002] carried out a more extensive comparison using a very similar model. The scan-paths of human observers were recorded during free-viewing of stimuli consisting of photographs of home interiors, natural landscapes, cityscapes and computer generated fractals.(Note: they used highly structured fractal iamges, such as the Mandelbrot Set 1983, rather than the random phase, noise-based methods used in this thesis.) They found that the saliency at fixation locations was significantly above chance and that saliency had a stronger effect on the initial fixations. Furthermore, the effect was strongest for the fractal images and weakest with the photographs of home interiors. A separate study, by Einhäuser and Koing [2003], challenged the role of luminance contrast in saliency. Photographs of outdoor scenes were used as stimuli and image regions were modified to increase or decrease luminance contrast in order to investigate what effect contrast has on attracting attention. The results found that while there was a correlation between initial fixation locations and luminance contrast, moderate modifications of the contrast had no significant effect on attracting or repelling fixations. These findings were rebutted by Parkhurst and Niebur [2004] who cited several methodological problems in Einhäuser and Koing's study and went on to describe a model based on luminance contrast and texture contrast that could account for the empirical data in the earlier paper.

A more recent study by Elazary and Itti [2008] looked at the correlation between salient regions in the visual field and interesting objects. This involved running the model on thousands of images from the LabelMe database [Russell et al., 2005]. This is a dynamic dataset containing over 150000 images with over 250000 labelled

objects of interest. The results from running the saliency algorithm on this dataset showed the most salient image region, as defined by the model, correlated with an interesting object 43% of the time, and one of the three most salient regions was an object 76% of the time. Both these values are higher than the chance levels, of 21% and 43% repectively.

4.1.4 Further Development

There are a number of aspects of human vision that are not represented in the saliency model. Peters et al. [2005] investigated non-linear interactions between orientation features at short range (for clutter reduction) and long range (for contour facilitation) along with the addition of a detailed model for eccentricity dependent processing which takes the foveal structure of the retina into account. They found that all three modifications significantly improved the model's performance, although interestingly there was no difference between the full eccentricity-dependent processing model (where higher frequencies fall off faster with eccentricity) and a simple approximation which is applied to the final saliency map. Walther and Koch [2006] developed a Matlab implementation for the algorithm and outlined a proto-object recognition algorithm which allows object-based inhibition of return to be implemented. This involves finding the feature map which contributes most to the current focus of attention in the saliency map. The model has also been modified to include some top-down elements based on a prespecified target [Navalpakkam and Itti, 2005, 2007].

While these related visual saliency algorithms have been widely used in the past decade they have also been criticised. Henderson et al. [2007] argues that the model does a very poor job of accounting for attention and that top-down factors are far more important than the basic bottom-up features that are used by the model. Especially during visual search, scan-paths are often very task dependant and observers will search in different regions of the scene depending on what they are looking for [Neider and Zelinsky, 2006a].

Additionally, several different systematic tendencies have been shown to be present in scan-paths which can not be explained using bottom-up saliency models. For example, human observers have been shown to make more horizontal saccades than vertical, even when isotropic, structureless stimuli are used [Gilchrist and Harvey, 2006]. Dragoi and Sur [2004] examined the eye movements of rhesus monkeys

free-viewing natural scenes and found relationships between the image structure (in terms of local orientation) at successive fixation locations: a fixation on an image region is likely to be followed by a fixation on a near by region with similar orientation or a further away region with a different orientation. However Dorr et al. [2009] have challenged this claim. They carried out an experiment using dynamic scenes (videos) and human observers and examined the relationship between local features (colour, orientation, motion and geometric invariants) at successive fixations. While there were differences between natural and artificial (random) scan-paths, Dorr et al. argue that these can be attributed to spatio-temporal correlations in natural scenes and a target selection bias.

Observers have also been shown to prefer making fixations in the centre of an image [Tatler, 2007, Tatler and Vincent, 2008]. Interestingly, this bias appears not to be an artefact of the central placement of interesting objects in most photographic stimuli, but is present even when the object of interest is away from the centre. Theoretical criticisms of saliency models [Baddeley and Tatler, 2006, Vincent et al., 2007] have argued that their internal structure is only loosely based on properties of the primary visual cortex and that most of the design choices are fairly arbitrary. Baddeley and Tatler [2006] claimed that high frequency edges, rather than contrast, provide the important information in such models and that including other features does not provide any additional benefit. Similarly, Yanulevskaya et al. [2008] found differences between fixated and non-fixated regions in terms of edges.

4.1.5 Conclusions

This literature survey has reviewed previous work on computational visual salency. In particular, one of the most widely used saliency models has been detailed and discussed. While its success in explaining human behaviour is mixed it has been shown to work well with simple stimuli which contain little in the way of high level information. Therefore, we would expect the model to cope well with the task of finding an anomaly on an otherwise homogeneous surface texture.

4.2 Experimental Methods

The experiments described below are designed to investigate how well human observers can find a defect in a textured surface. Two types of surface texture will be used and a variety of parameters will be investigated. I will then compare the performance of the observers with that of the Matlab implementation of Itti and Koch's visual saliency algorithm. I expect human observers and the saliency model to find the defect more easily on a smooth $1/f^{\beta}$-noise surface than on a rough one. For the near-regular surfaces, I expect that increasing the regularity will make the search task easier.

4.2.1 Set-up

A Tobii x50 eye-tracker was used to record observers' gaze patterns. The Tobii is based on PCCR (pupil centre corneal reflection), which involves illuminating the eye with near infrared light and recording the resulting refelections from the cornea and the pupil. The fixation filter was set to count only those fixations lasting longer than 100ms within an area of 30 pixels. This means that a fixation was only registered if an observer's gaze stayed within a circle with radius 30 pixels, for at least 100ms. The accuracy of the eye-tracker was 0.5° - 0.7° and the spatial resolution was 0.35°. Each trial started with a central fixation cross, which was displayed for 2 seconds, followed by the stimuli, which was displayed until the observer responded via a keypress, pressing 'd' for a target present response, 'k' for target absent.

4.2.2 Stimuli

Stimuli were created as described in Section 3.2.2. All stimuli were 1024×1024 pixels in size and displayed on a calibrated NEC LCD2090UXi monitor. The pixel dimensions were 0.255mm by 0.255mm resulting in images with physical dimensions 261mm by 262mm. The monitor was linearly calibrated, gamma = 1, with a Gretag-MacBeth Eye-One, with the maximum luminance set at $120cd/m^2$. This resulted in the rendered images appearing as if they were being lit under bright room lighting conditions.

Stimulus presentation was controlled by Clearview (Tobii Technology Inc). The viewing distance was controlled by use of a chin rest, placed 0.87m away from the

display monitor. At this distance one pixel is approximately $1'$ of visual angle and the stimuli were 16.7° across. A target was added to half the images at a random location between 6° and 7.5° from the centre. The targets in the $1/f^{\beta}$ experiments, and the textons in the *near-regular* experiment, subtended approximately 0.66° of visual angle. The *surface roughness, target depth* and *near-regular* experiments comprised of 300 trials while the *target orientation* experiment contained 240 trials.

4.2.3 Observers

Five observers were used for each experiment: all had normal or corrected to normal vision and were between 21 and 30 years old. Some observers participated in multiple experiments. Observers were given several practice trials. They were told that the target would be present on half the trials and for the $1/\beta$-noise surfaces it would always be an indent in the surface of the same size and shape, while for the *near-regular* surface it would be a missing texton. They were instructed to decide whether the target was present or not and to respond with a key press for target present or absent as quickly and accurately as they could. No time limit was imposed on the task. Observers were allowed to take a short rest every hundred trials (120 trials for the *target orientation* experiment).

4.2.4 Visual Saliency Model

The Matlab Saliency Toolbox [Walther and Koch, 2006] (an implementation of Itti and Koch's model) was used with only minor changes made to the default parameters specifying the resolutions of the Gaussian pyramid. Since the model was originally designed and tested on photographs containing macroscopic objects the resolution settings are quite low, i.e. the image is blurred and reduced in size a lot. While this works well with photographs (where we measure average contrast over fairly large areas) the stimuli used in the following experiments contain a lot of very fine, high frequency information.

In order to accommodate this level of detail I have changed parameters relating to the levels of the Gaussian pyramid to be used: $c \in \{1, 2, 3, 4\}$ and $\delta \in \{1, 2, 3\}$ (see Section 4.1.2). This corresponds to the following constants in the Matlab toolbox:

```
params.minLevel = 1;
params.maxLevel = 4;
params.minDelta = 1;
params.maxDelta = 3;
params.mapLevel = 2;
```

I will use the same method of comparison with human observers as used by Itti and Koch [2000], and consider the number of fixations required to find the target. A maximum limit for the number of fixations was set as equal to the inter-subject mean number of fixations taken on target absent trials for a given set of parameters. These data can be seen in Figure 4.13, Figure 4.14, and 4.15. This provides a measure of how many fixations a human is prepared to make before giving up a search and making a negative response, and allows for the model's accuracy to be expressed as the proportion of trials in which it fixates the target before the maximum number of fixations is reached.

Comparing the number of fixations made by the model and human observers is reliable as long as accuracy rates are high, as in the *surface roughness* experiment. Where they are lower, in the *target depth* and *orientation* experiments, we also use accuracy rate (the proportion of trials on which an observer finds a target, or the model fixates it within the maximum number of fixations) as a comparison. A common means of comparing human fixation data with model predictions is to compare fixation locations [Parkhurst et al., 2002, Peters et al., 2005, Tatler, 2007, Tatler et al., 2005]. However, this method is not appropriate with the stimuli used here, because the statistics of the image vary little across the background and are only distinct at the target. There would therefore be no reason to expect any correlation in the locations of fixations on the background texture made by observers and by the model.

4.3 Results From Experiments 1 - 4

4.3.1 Experiment 1: $1/f^{\beta}$-noise: Surface Roughness

This initial experiment is designed to investigate how varying surface roughness can cause a target to become more or less conspicuous. The apparent conspicuity of

a defect can be measured in terms of reaction times and the number of fixations required to find it. Stimuli were created as detailed in Section 3.2.2 with $\beta \in \{1.55, 1.6, 1.65, 1.7, 1.76\}$ and $\sigma_{RMS} \in \{0.8, 1.0, 1.2\}$. The target was an ellipsoid with a = b = 10, c = 2, and was subtracted from the surface texture so that it created a hole with volume 10mm^3. These parameters were selected to provide a good range of task difficulties, based on data from a pilot experiment. Example stimuli are shown in Figures 4.1 and 4.2.

Results and Discussion

Overall, observers' accuracy was high, and for the target absent trials 99.5% of responses were correct. This suggests that the search target was well defined and easily identifiable: observers had no trouble in rejecting background patches. The few false positives that did occur can likely be attributed to observers accidentally pressing the wrong response key. There was no indication that increasing surface roughness had any effect on the number of false positives. Table 4.1 shows overall accuracy for each observer on the target present trials, and Figure 4.4 (left) the effect of the two surface roughness parameters on accuracy in these trials. A two way repeated measures ANOVA (analysis of variance) gives significant effects ($p < 0.05$) of β, ($F(4, 16) = 79$), σ_{RMS}, ($F(2, 8) = 58$), and the interaction ($F(8, 32) = 13$) on the mean inter-subject accuracy.

Participant	GW	HW	LM	JF	PS	Overall
Accuracy for target present trials	87%	78%	81%	87%	83%	83%
Accuracy for target absent trials	98%	99%	100%	100%	100%	99%

Table 4.1: Table showing accuracy for each observer in Experiment 1. Observers clearly had no difficulty in rejecting patches of the background as there are very few false positives.

Figure 4.3 shows the mean reaction time data on correct trials for each individual observer while Figure 4.4 (Top Left) shows the inter-subject mean reaction times. The pattern of variation between individuals suggests speed/accuracy trade-offs: comparing Figure 4.3 with Table 4.1 shows that observer 1 (GW) was the slowest but the (joint) most accurate (12.67% of targets missed) while observers 2 and 3 (HW and LM) were the fastest and also missed a greater number of targets(22% and 19.33%). Despite these differences, all observers were affected by surface roughness in the same way, with longer reaction times when searching on rougher surfaces. Inter-subject mean reaction time is shown in Figure 4.4 (Top-Right) and a two-way

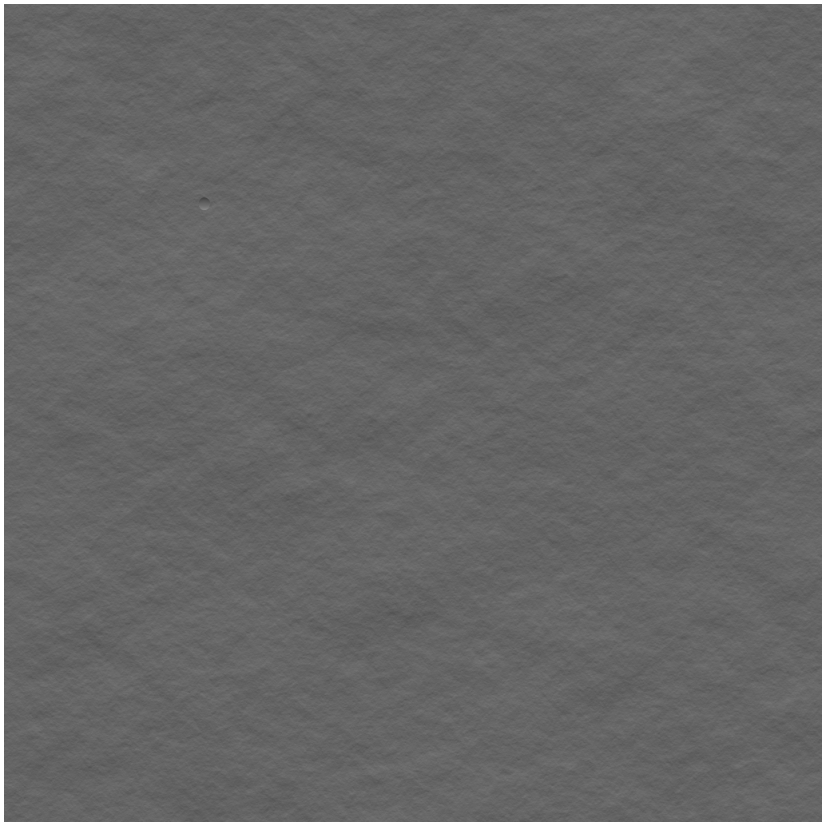

Figure 4.1: An example of a smooth $1/f^{\beta}$-noise surface. In this example $\beta = 1.75$ and $\sigma_{RMS} = 0.8$. The defect can be easily located in the upper left quadrant of the surface.

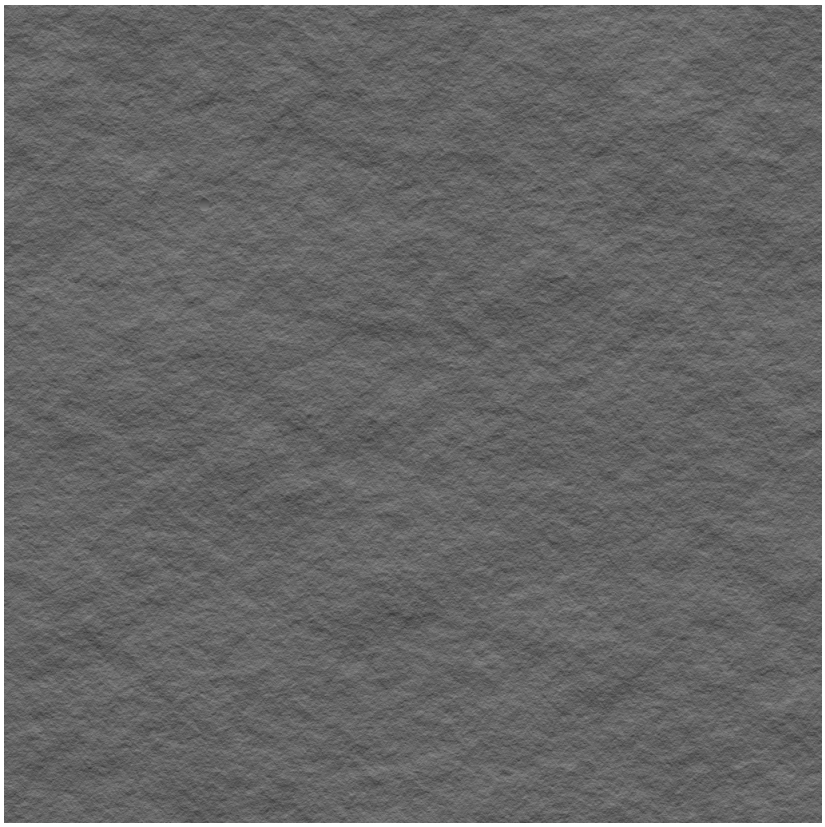

Figure 4.2: An example of a rough $1/f^{\beta}$-noise surface. In this example $\beta = 1.6$ and $\sigma_{RMS} = 1.2$. The defect is in the same location as with the previous example, but is now much harder to identify.

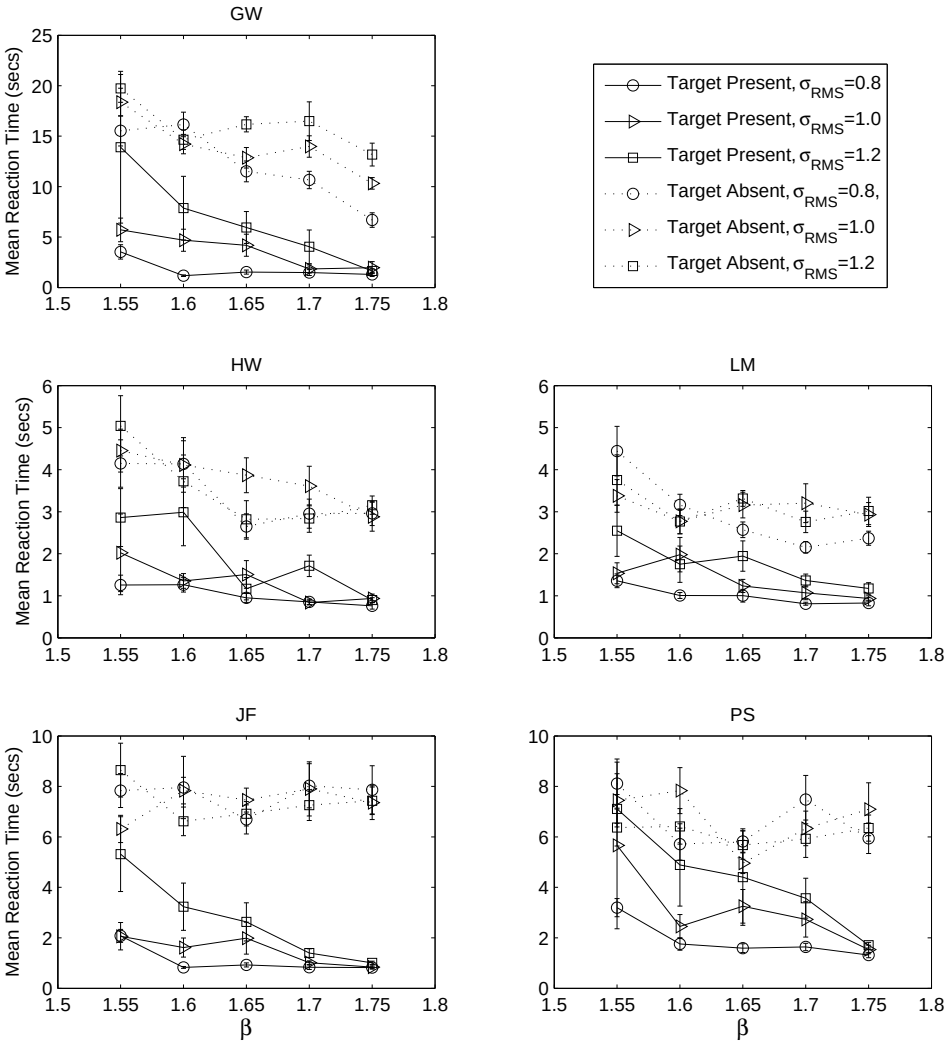

Figure 4.3: Mean reaction time plotted against surface roughness for each observer in Experiment 1. When reaction times are compared to the accuracy results (shown in Table 4.1) we see evidence for a classic speed-accuracy trade-off.

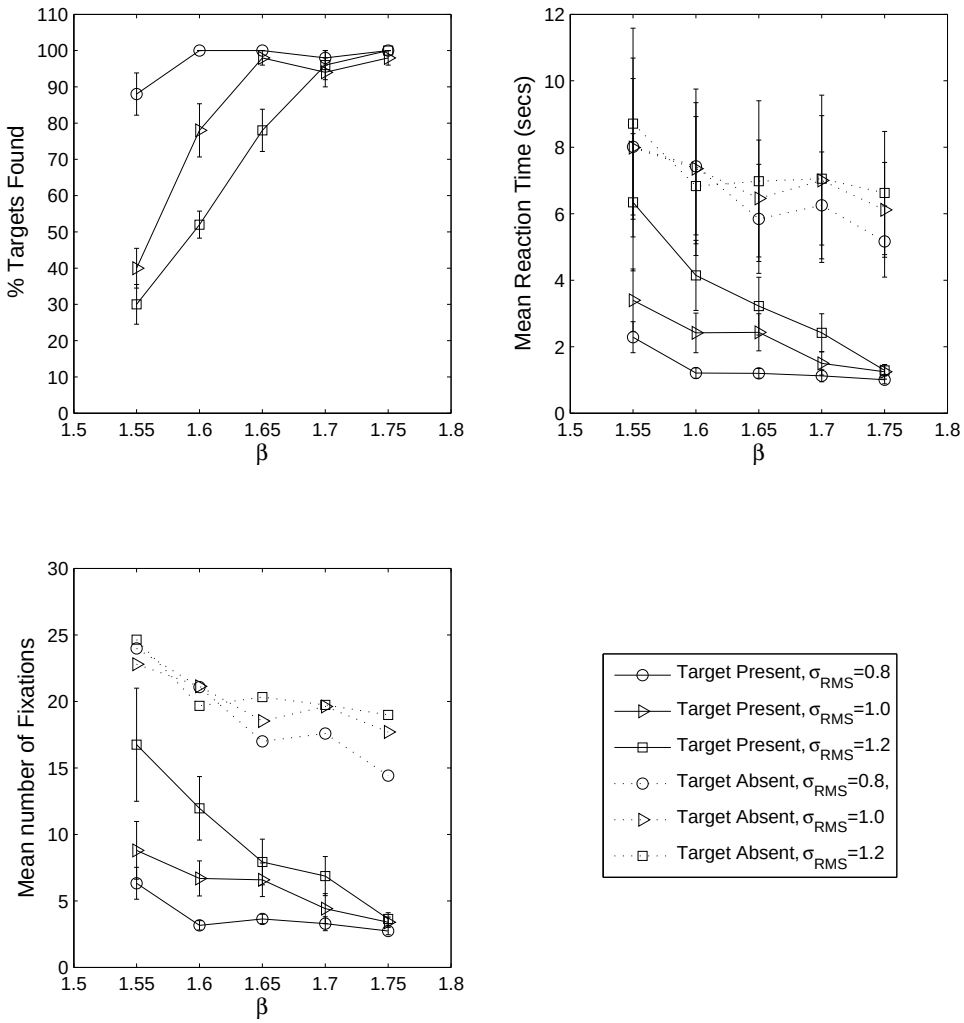

Figure 4.4: Results from Experiment 1. (Top Left) Mean accuracy for *target present* trials plotted against surface roughness. (Accuracy for target absent trials was near 100% correct.) (Top Right) Inter-subject mean reaction times plotted against surface roughness. The large error bars are due to the different speed-accuracy trade-offs used by the individual observers. (Bottom) Inter-subject mean number of fixations required to find the target.

repeated measures ANOVA gives a significant effect ($p < 0.05$) of β ($F(4,16) = 8.8$), σ_{RMS}, ($F(2,8) = 9.0$) and the interaction ($F(8,32) = 4.5$). Surprisingly, surface roughness does not have a significant effect on the reaction times in the target absent trials ($F(4,16) = 2.745, p = 0.065$ and $F(2,8) = 0.729, p = 0.512$ for β and σ_{RMS} respectively). Results from canonical target present/absent experiments usually find that target absent times are approximately twice those of target present times. Here however, the difference between reaction times for target present and target absent actually decreases with increasing difficulty. The low accuracy (30%) suggests that this is due to the difficulty of the search task: when the surface is rough observers are probably close to responding *target absent* when they eventually find the target.

The relationship between the number of fixations made on each trial and surface roughness is shown in Figure 4.4 (Bottom). The effects of both variables and their interaction are significant($p < 0.05$ for β ($F(4,16) = 8.1$), σ_{RMS} ($F(2,8) = 8.1$) and their interaction ($F(8,32) = 4.4$). The implication that most variance in reaction time is due to variance in number of fixations, rather than duration, is confirmed by significant correlations between reaction time and number of fixations on each trial (values of r for individual observers range from 0.899 to 0.971, all $p < 0.0001$). As with reaction times, surface roughness did not have as large an influence on the number of fixations on target absent trials. A two-way repeated ANOVA gives a significant effect for β with $F(4,16) = 4.131, p = 0.17$, but not for σ_{RMS} ($F(2,8) = 0.895, p = 0.446$) or the interaction ($F(8,32) = 1.423, p = 0.2254$).

Do observers have to fixate on the target to be able to identify it? In order to investigate this I first looked at the distance on each trial from the target to the centre of the fixation when the response key was pressed. Unfortunately these data is somewhat noisy, due to motor-latencies, and occasional saccades away from the target after identification but before the key press response is given. Instead, as it is not possible to define exactly the time at which the decision to press the key is made, the *final fixation to target distance* is defined as the distance from the target to either the fixation during which the response key was pressed, or the fixation before it, whichever was smaller. This criterion allowed for some variation in the time between the decision to respond and initiation of a saccade away from the target. Specifically, it meant that when an observer made a saccade away from the target after fixating it but before responding with a key press, the shorter distance was counted provided that the response occurred during the next fixation.

Figure 4.5 (left) shows the distribution of final fixation to target distances over

all trials and observers. (Two trials in which the response key was pressed several saccades after the target was fixated were removed from the analysis.) It appears that the distances fit a Poisson distribution, although there is a slight hint that a bimodal distribution might be present. While the majority of trials, 82%, have a final fixation to target distance of 1° or less, there also appears to be a second, smaller set of trials with a larger final fixation to target distance. These account for 5% of all correct target present trials and indicate that the target was identified without fixation. it is possible that this behaviour is more common than the data suggests, as observers may be detecting the target without fixation, and then making a saccade to the target before making a response. Figure 4.5 (right) shows how the mean final fixation to target distance changes with surface roughness. A two-way repeated measures ANOVA gives a significant effect only of β ($F(4, 16) = 3.05, p = 0.048$). As β increases, and the surface appears less rough, mean distance from final fixation to target increases, as identification without fixation becomes more frequent. The lack of an effect of σ_{RMS} is probably due to a lack of data for the rougher surfaces, where the proportion of correct responses is small. This question of the effect of eccentricity on target detectability will be further investigated in Chapter 7, Section 7.5.

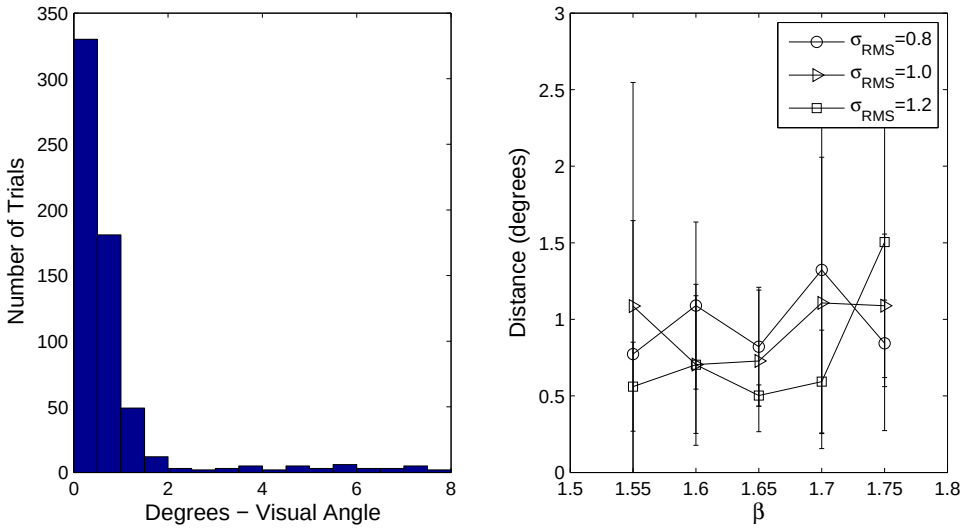

Figure 4.5: Distance from final fixation to target in Experiment 1. (Left) shows the histogram over all trials with *final fixation to target distance* of less than 8°. [There were four trials with distances larger than this which are not shown in the histogram.] (Right) The effect of roughness on the mean final-fixation-to-target distance. (Legend as in Figure 4.3.)

The large majority, 82%, of correct responses occurred when fixating within 1° of the target. Considering the trials on which the target was present but missed, we

can use the figure of 1° as a criterion to determine how often the target was fixated but not identified. This happened in 20% of the target missed trials.

Finally, Figure 4.6 shows the distribution of fixations over all the target absent trials. As can be seen, the large central bias in fixation placement reported by Tatler [2007] is not present in this experiment. Overall, the fixations are well spread out with a slight bias towards the peripheral regions of the stimuli. This suggested that over the course of the experiment observer learnt that the target was always located away from the centre of the stimulus. Further details are shown in Figure 4.7.

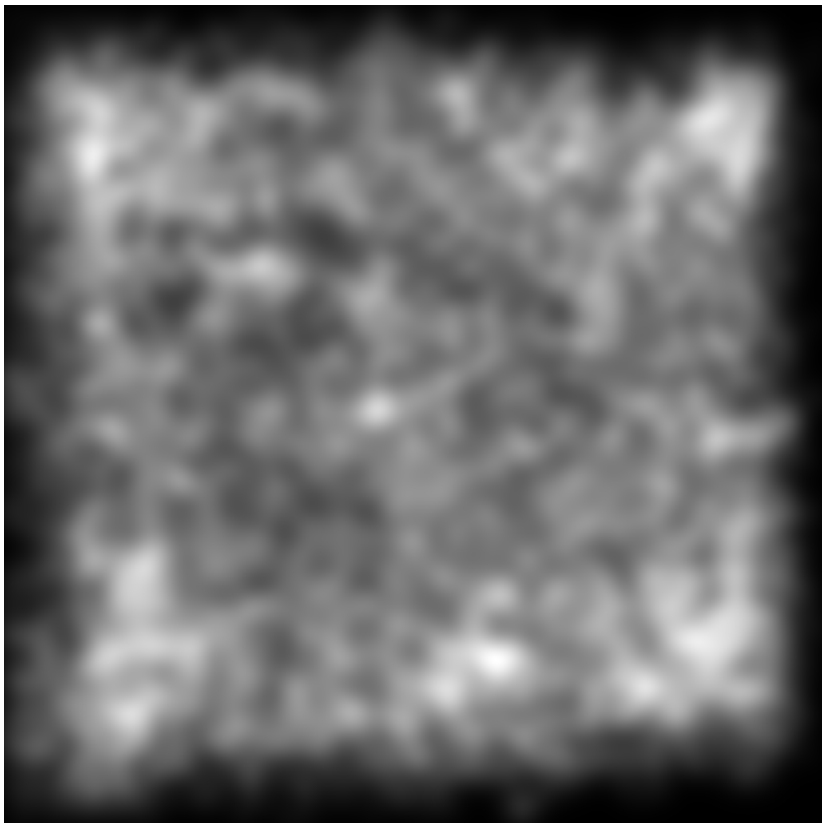

Figure 4.6: Hotspot map containing fixations from all the target absent trials in Experiment 1. The initial two fixations in each trial were not included in order to minimise any bias introduced by the fixation cross which is displayed before the trial starts.

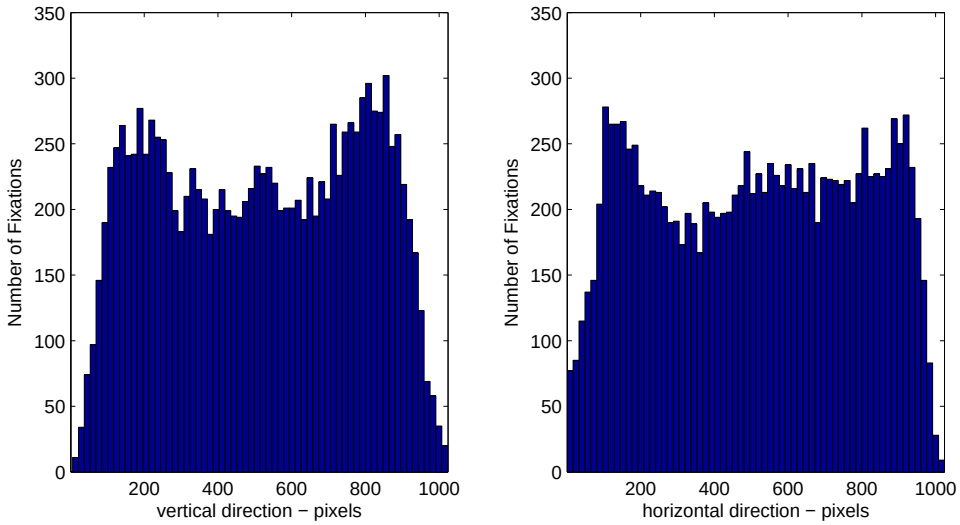

Figure 4.7: Histogram showing the distribution of fixations in the horizontal and vertical directions for target absent trials in the $1/f^{\beta}$-noise: surface roughness Experiment 1. (x-axis units are in pixels).

4.3.2 Experiment 2: $1/f^{\beta}$-noise: Target Depth

The previous experiment investigated the effects of varying properties of the background surface on visual search. In the next two experiments I will consider the effects of changing target properties, starting with target depth. As the depth of the target is reduced, the gradient at its edges, and hence the contrast created by the illumination process, decreases, and the target should become harder to find. To vary target depth, a target was created in the same way as in the previous experiment, and then its depth was reduced scaling factor, z_k. Setting $z_k = 1$ gives the same depth (and hence level of contrast) as used previously.

Pilot studies showed that people had difficulty in identifying the target for $z_k = 0.5$, even when its location was known. Therefore, the following values of z_k were used: $0.6, 0.7, 0.8, 0.9$ and 1.0. The target was placed on a subset of images from the preceding experiment: $\beta \in \{1.6, 1.65, 1.7\}$ and $\sigma_{RMS} = 1.0$. These values were chosen as they give a range of roughness over which target detection is neither too hard nor too easy. For each value of β, ten surfaces were created and each surface was displayed five times as a target absent trial, and once as a target present trial for each of the five values of z_k. Target locations were determined in the same way as before.

Results

Accuracy for each observer is shown in Table 4.2. While the number of false positives are similar to those seen in the *surface roughness* experiment, the hit-rate is noticeably lower. Accuracy of target detection fell as the target was made shallower, to the extent that when z_k was 0.6 or 0.7 the level of accuracy fell considerably below those found in the first experiment (Figure 4.8). Both surface roughness β and target depth z_k have significant effects on accuracy (repeated measures ANOVA: $F(2,8) = 89.5, p < 0.05$ for β; $F(4,16) = 146.6, p < 0.05$ for z_k; $F(8,32) = 3.5, p < 0.05$ for the interaction).

As there are very few correct target present trials for $z_k = 0.6$ and 0.7, reaction times and numbers of fixations are unreliable measures for these cases. Therefore, only the reaction times for $z_k = 1, 0.9$ and 0.8 are shown in 4.8. Over this range, surface roughness and target depth both have significant effects on accuracy (repeated measures ANOVA: $(F(2,8) = 7.0, p = 0.049$ for β; $F(2,8) = 10.3, p = 0.026$ for z_k; $F(2,16) = 2.9$, N.S. for the interaction). As in Experiment 1a, the results for numbers of fixations follow a similar pattern to reaction time (see Figure 4.8 below).

Participant	MK	LM	PS	LC	RL	Overall
Accuracy for target present trials	42%	50%	51%	49%	53%	49%
Accuracy for target absent trials	100%	95%	99%	93%	91%	96%

Table 4.2: Table showing accuracy for each observer in Experiment 2. While this experiment was more difficult than Experiment 1, the accuracy for the target absent trials is still very high.

4.3.3 Experiment 3: $1/f^{\beta}$-noise: Target Orientation

In this experiment an elongated target is used and its appearance can be changed by varying its orientation. As an elongated target is rotated, the angle that its long axis makes with the incoming light varies, resulting in variation in the contrast at its edges (see illustration in Figure 4.9). Therefore as the target is rotated towards the light source it should become harder to detect.

The target used in this experiment was an ellipse with axes subtending approximately 0.7° by 0.2°. The volume of the indent, its location, and the illumination conditions were the same as in the first experiment. Unlike the previous two experiments, the roughness parameters were kept constant (β and σ_{RMS} were 1.65 and

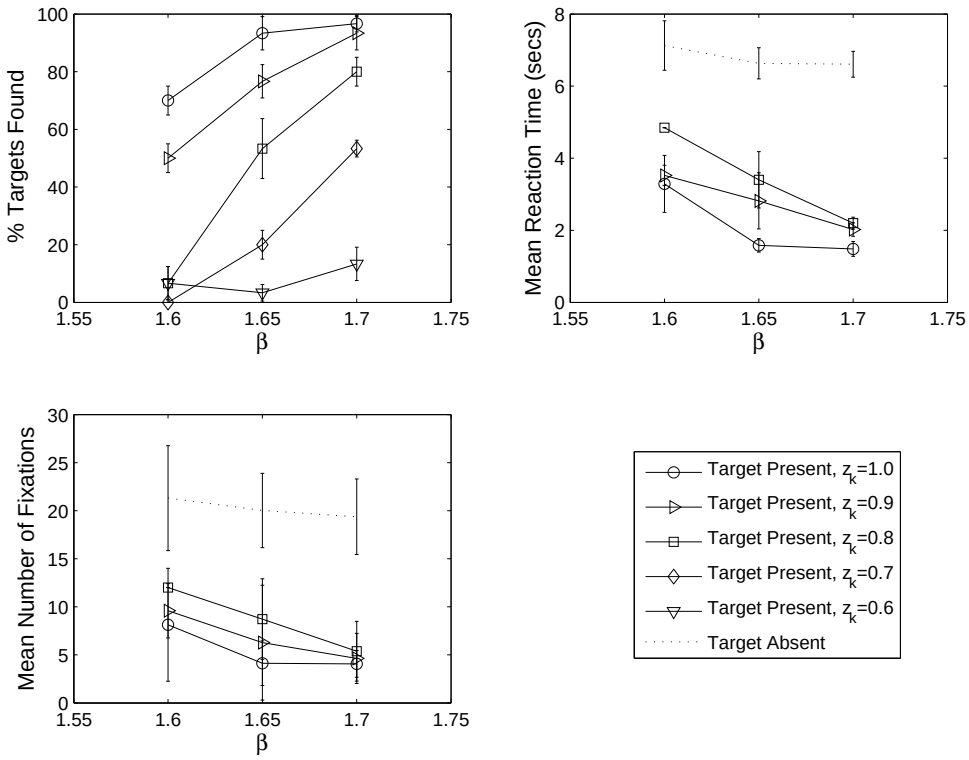

Figure 4.8: Results from Experiment 2. (Top Left) Mean accuracy for *target present* trials, (Top Right) mean reaction times for cases $z_k = 0.8 - 1.0$ and (Bottom) inter-subject mean number of fixations to target. (Note: as target absent trials have no target, the z_k parameter has no effect on the surface.)

1.0 respectively). The variable in this experiment was the orientation of the target. 12 orientations were used: $\theta \in \{90° \pm \phi | \phi = 0°, 5°, 10°, 20°, 30°, 45°, 90°\}$, where 90° corresponds to the direction of illumination and, due to symmetry, $0° = 180°$.

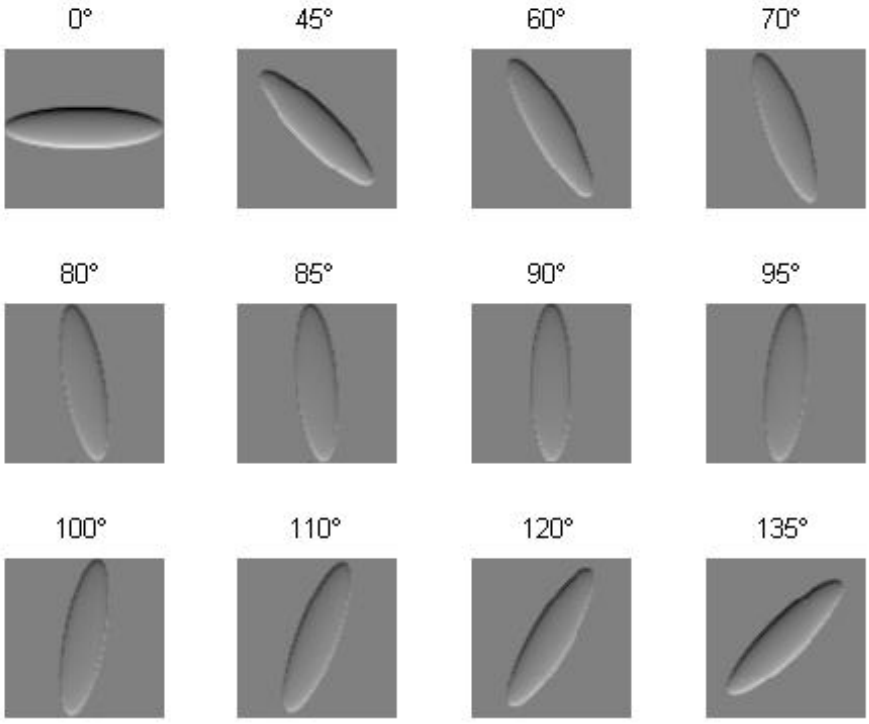

Figure 4.9: The effect of rotating an elongated target relative to the direction of illumination (from above). Orientations are in degrees relative to the horizontal. Note how contrast at the edges of the target changes with the orientation, reaching a minimum at 90°

Results

Overall accuracy for each observer is given in Table 4.3 and, interestingly, there is a significant number of false positives. This is possibly due to the increased uncertainty about the target's appearance. The relationships between target orientation and both accuracy and reaction time are shown in Figure 4.10. It is clear that there is a sharp drop in accuracy rates as target orientation approaches vertical, and all observers found the search task very difficult for targets orientated at $90° \pm 5°$. Again the number of fixations per trial followed a similar pattern to the reaction times.

As we would expect, target detection is hardest when it is oriented parallel to the illumination, but it is important to note that the effect is not linear. Instead, there is a narrow band in which orientation has a strong effect on search performance.

Participant	LM	JF	LC	HW	SP	Overall
Accuracy for target present trials	65%	80%	70%	60%	84%	72%
Accuracy for target absent trials	94%	100%	76%	94%	73%	88%

Table 4.3: Table showing accuracy for each observer in Experiment 3.

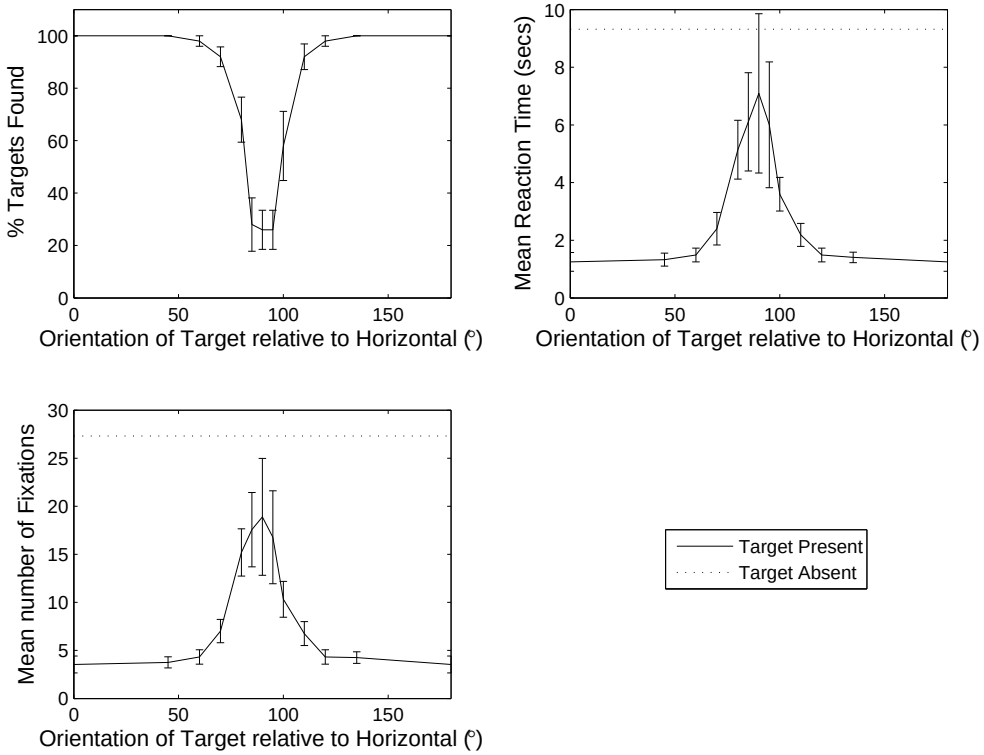

Figure 4.10: Results from Experiment 3. (Top Left) Mean accuracy for *target present* trials, (Top Right) mean reaction times and (Bottom) inter-subject mean number of fixations to target. (Note: as target absent trials have no target, the θ parameter has no effect on the surface.)

4.3.4 Experiment 4: Near-Regular Textures

This last experiment is designed to investigate how well human observers can find a missing texton in a near-regular texture. The stimuli are created as described in Section 3.2.3 with the texton density, $\rho \in \{1.875, 2.461\}$ textons per degree and

degree of regularity, $\sigma_j \in \{0, 0.5, 1, 1.5, 2\}$. The size of the textons was randomly varied with $a, b \in \{8, 9, 10, 11\}$ and $c = 1$.

Results

The overall accuracy of each observer is given in Table 4.4 and the individual results are shown in Figure 4.11. As can be seen, the accuracy for trials with no defect is extremely high. There are large differences between observers, both in terms of the number of defects found and the time in which it took to find them. However these are broadly consistent with speed-accuracy trade-offs. The mean inter-subject results can be seen in Figure 4.12. Neither σ_j or ρ had a significant effect on accuracy with $F(4, 16) = 2.524$ and $F(1, 4) = 3.634$ respectively. Interestingly though, both parameters had a significant effect on reaction time in the target present trials with $F(4, 16) = 9.702, p < 0.001$ for σ_j and $F(1, 4) = 8.256, p = 0.45$ for ρ. Reaction times for target absent trials were also significant ($F(4, 16) = 8.2, p = 0.001$ and $F(1, 4) = 20.0, p = 0.11$ for σ_j and ρ respectively). There was also a significant interaction $\sigma_j \times \rho$ for target absent reaction times with $F(4, 16) = 4.331, p = 0.015$.

Participant	LM	MK	MY	NC	FH	Overall
Accuracy for target present trials	63%	93%	81%	72%	58%	73%
Accuracy for target absent trials	100%	100%	100%	98%	100%	99%

Table 4.4: Table showing accuracy for each observer in Experiment 4.

4.4 Comparison with Model

A comparison, in terms of the number of saccades needed to find the target, between the model and human performance for the *surface roughness* experiment is given in Figure 4.13. Overall the model outperforms human observers, taking fewer saccades than human subjects in order to fixate the target. This agrees with the results reported by Itti and Koch [2000]. The graph shows that both humans and the model respond to increasing roughness in a similar way: more fixations are required to find the target on a rougher surface than on a smoother one.

The results for the *target depth experiment* can be seen in Figure 4.14. The model now out-performs human observers, both in terms of accuracy and the number of fixations required to find the target. Figure 4.15 shows the results for the *target*

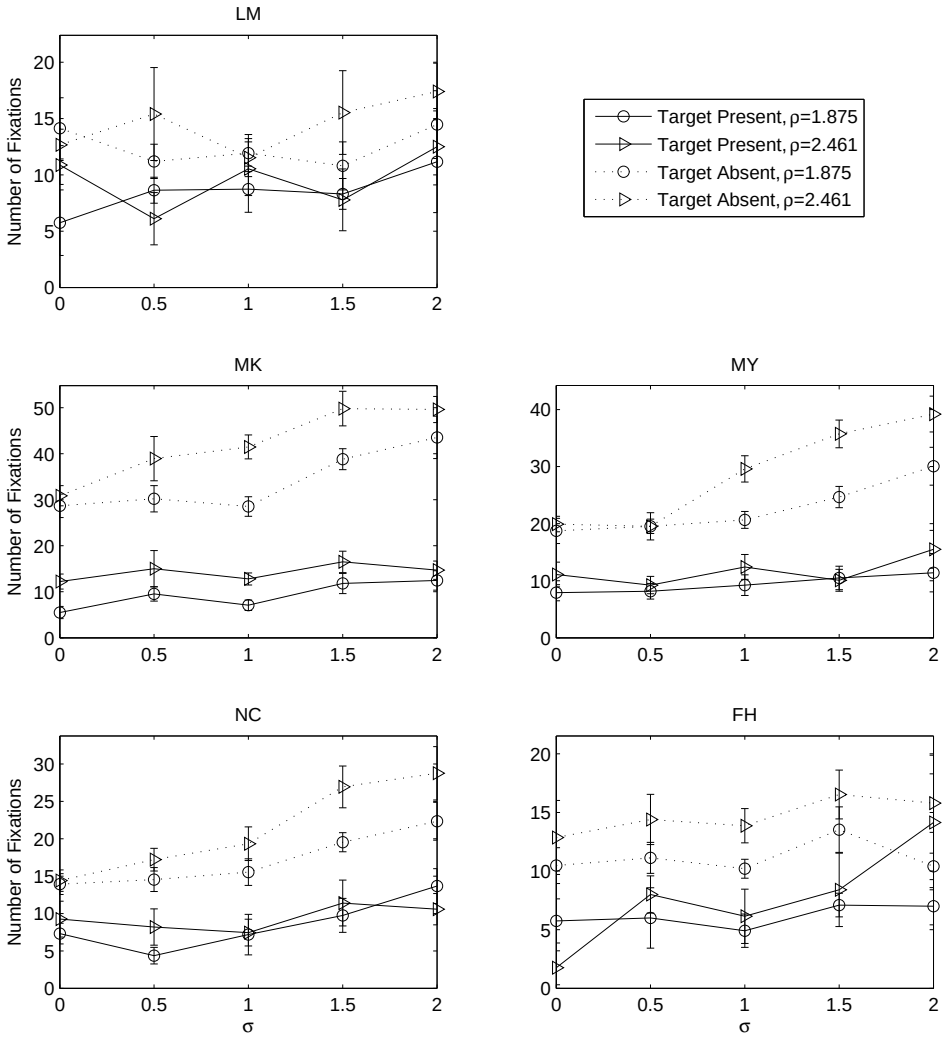

Figure 4.11: Results for each individual observer in Experiment 4. Only correct trials are shown. As can be seen, the different observers make different speed-accuracy trade-offs.

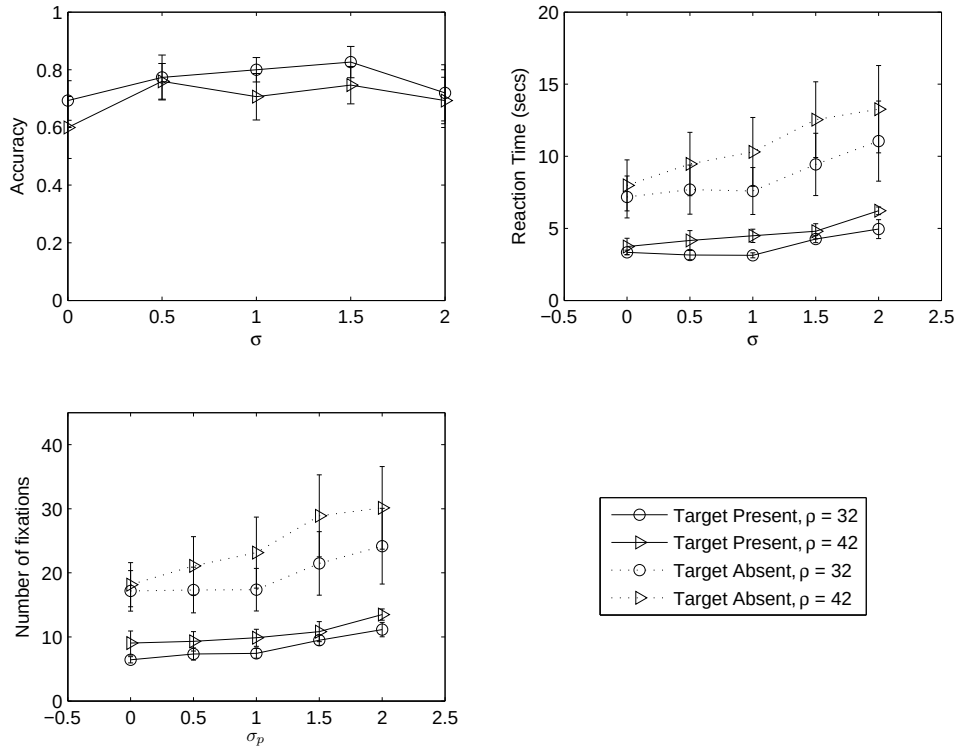

Figure 4.12: Results from Experiment 4. (Top Left) Accuracy, (Top Right) mean reaction times and (Bottom) inter-subject mean number of fixations to target.

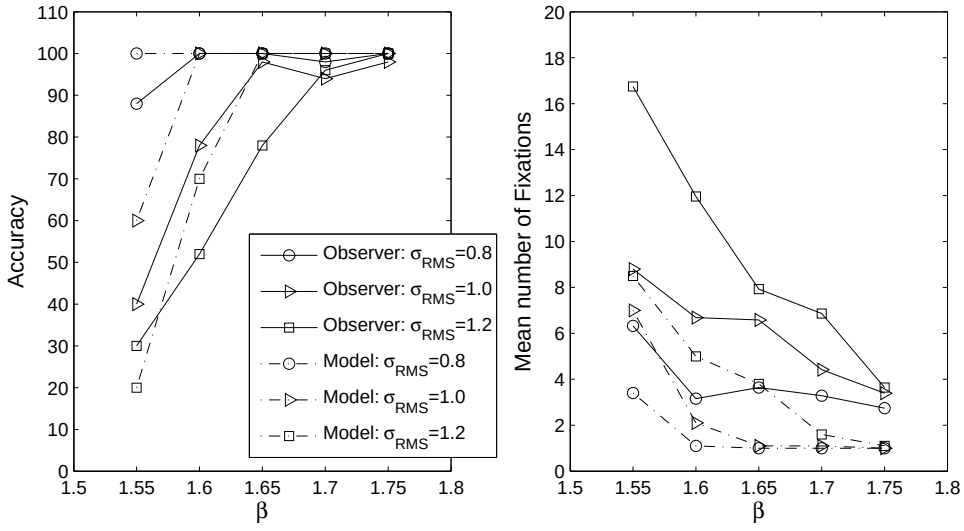

Figure 4.13: Comparison between human observers and the saliency model for Experiment 1. (Left) the number of targets found and (Right) the number of fixations required to find the target.

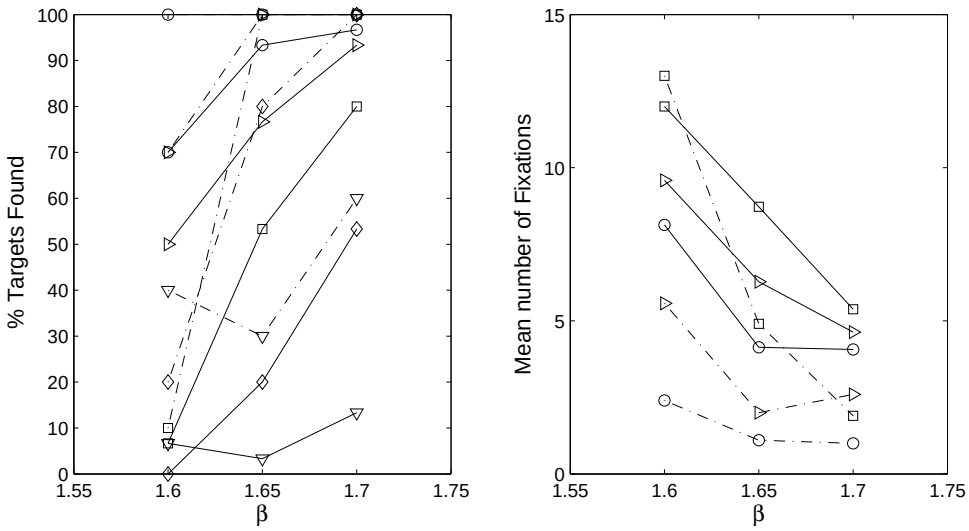

Figure 4.14: Comparison between human results and the saliency model for Experiment 2. The mean number of fixations on target absent trials was 21.3, 20.0 and 19.4 for $\beta = 1.6, 1.65$ and 1.7 respectively. Legend as in Figure 4.8. Dashed line shows the model's performance.

orientation experiment. As can be seen, the model performs poorly when the target is near vertical, much worse than the human observers. Furthermore, while humans had difficulty over $90° \pm 5°$ the model performed badly over a much wider range, around $90° \pm 20°$. These results indicate that the orientation channel in the saliency model does a poor job of matching human perception for an elongated target.

The results for the *near-regular* experiment from the saliency model can be seen in Figure 4.16. As we can see, the model performs poorly in terms of accuracy, finding less than 40% of the targets accross the range of parameters used.

4.4.1 Discussion

Comparison between the experimental results and the performance of Itti and Koch's saliency model suggests that the features used by the model, while capturing some aspects of human behaviour, are not sufficient to give an adequate simulation of visual search for a target on a rough surface. The closest match between human and model search performance occurred with the set of stimuli used in the surface roughness experiment, where the two parameters of surface roughness were varied. Although there were discrepancies in the absolute number of fixations by humans and model, the model reproduced all the effects of background roughness parameters.

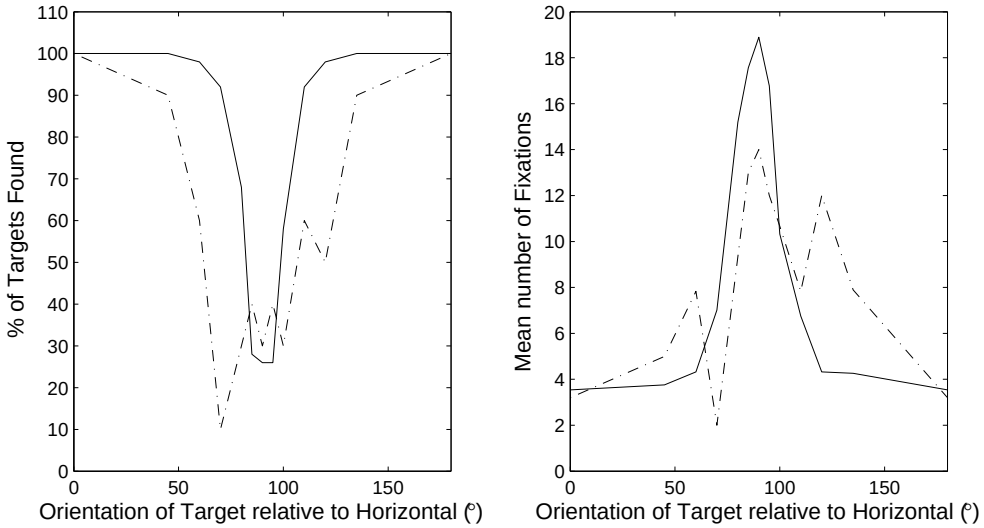

Figure 4.15: Comparison between human results and the saliency model for Experiment 3. Solid line shows the human results while the dashed line shows the saliency algorithm.

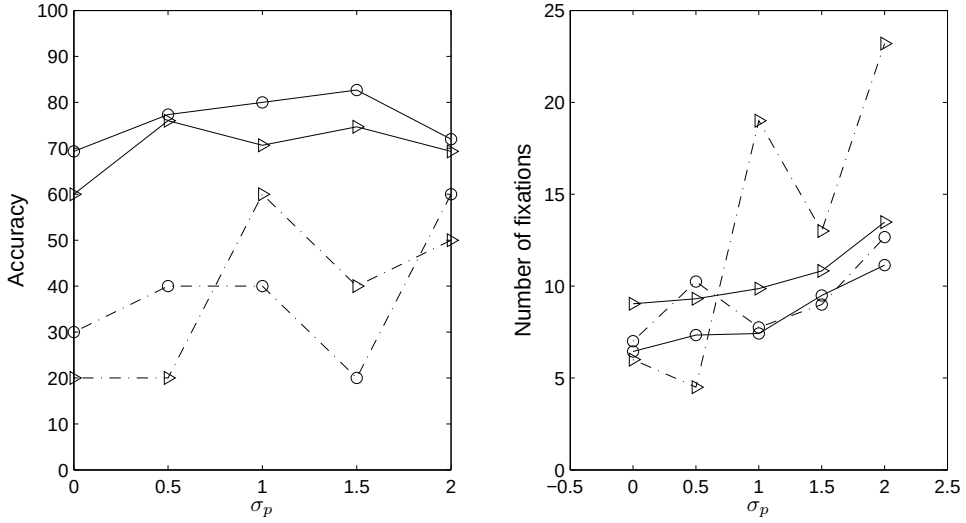

Figure 4.16: Comparison between human observers and the saliency model for Experiment 4. Legend as in Figure 4.11. Dashed line shows the model's performance.

This is a surprisingly good match, given that the model was not developed to work on such stimuli and has not been assessed in such a way before. When search performance with an elongated target was considered in Section 4.3.3, however, there was not only a difference in absolute levels of performance but also in the effect of target orientation, with the ability of the model to detect the target falling to low levels over a considerably wider range of orientations than in the case of human observers.

The model was tested several times with varying numbers of spatial scales and orientations in the filter bank and its performance was found to be robust, as long as the spatial scale which best matches the scale of the target was present. The conclusion is therefore that there is a clear discrepancy in the case of oriented targets, with the model unable to match human performance when they are oriented close to the direction of illumination. What could the cause of this discrepancy be? The saliency model that I used is likely to diverge from human performance because it does not incorporate eccentricity dependent processing [Peters et al., 2005, Vincent et al., 2007]. However this gives the model constant spatial resolution at all distances from fixation, while human resolution falls, and so the model would be expected to perform better with all targets. Similarly, the model does not incorporate any process of extracting solid shape from shading, which is known to contribute to efficient detection of targets in human visual search [Braun, 1993], but this feature would result in poorer model performance across all targets, which is not the case.

Another possible reason for poor performance of the model with elongated targets is that, when the target is oriented close to the vertical, the contrast decreases. If the model is generally less sensitive to low contrast than humans, the result would be poorer performance. However, there is no evidence for such a difference in the target depth experiment, Section 4.3.2, where contrast at the target is reduced by making it shallower. The two results together cannot be explained by a difference between humans and model in sensitivity to contrast, and imply that the results arise specifically because the saliency model, despite having a dedicated orientation channel, is failing to take advantage of the directional nature of the elongated targets in Section 4.3.3.

Despite the shortcomings of the saliency model, it does manage to give a broad approximation to human performance over a wide variety of difficulties.

4.5 Conclusions

These initial experiments were designed to investigate human performance in a search task using computer generated surface textures as stimuli. The effects of roughness on visual search performance in this experiment, as measured both by accuracy and reaction time, are closely similar to the effects of set size in conventional visual search tasks using arrays of discrete items. When $\sigma_{RMS} = 0.8$, the reaction time vs β slope is near horizontal implying that search is efficient. As σ_{RMS} increases the magnitude of the gradient increases implying that search is less efficient. At the rough end of the range, the task became very difficult with target hit rates far lower than those commonly encountered in visual search tasks (see Figure 4.4). There was a similar pattern of results with the near-regular textures, with both the density and regularity of textons affecting reaction times. While neither parameter had a significant effect on accuracy this is likely to be due to the use of a small number of human observers with large inter-subject differences in terms of speed-accuracy trade-offs. One way to get round this problem is to use a *target always present* experimental design (as will be done in Chapter 6).

I therefore conclude that it is possible to change the parameters of these continuous, synthetic surface textures in ways that have systematic effects on ability to identify a small anomalous region in the surface. The very small number of false positives recorded in the experiment indicates that observers did not have any trouble in identifying the target once they fixated it; rather, the increase in difficulty with rough surfaces came from an inability to identify the target pre-attentively based on the contrast information present. Observers have to switch from using pre-attentive vision to carrying out some sort of serial search strategy, leading to an increase in both the mean number of fixations and the variation (Figure 4.13).

Analysis of distances between target and fixation when targets are identified demonstrates two patterns; on the majority of trials, fixation is within 1° of the target when it is recognised, but on others it falls in a higher range centred around 6°, indicating recognition of the target in peripheral vision. There is some evidence that the second pattern is more common when surfaces are smoother, which would be expected as the demands placed by the task on acuity of visual processing will be lower on smoother surfaces. These data is likely to be biased towards smaller distances. It is likely, especially for the smoother, easier trials, that the target is identified while fixation is elsewhere but a saccade is made to the target in the time it takes to execute a keypress. This hypothesis is followed up and investigated in

Section 7.5. Fixation of a target does not ensure that it will be recognised; on 20% of trials when the target was missed, the target was fixated, but not detected, at some point during the search.

This chapter also contained a comparison between human performance and Itti and Koch's saliency algorithm, in terms of the number of fixations required to find the target (Section 4.4). The model proved to be a good fit for human data in both Experiment 1 (surface roughness) and Experiment 2 (target depth) over a wide range of task difficulties. This is perhaps surprising since the model has been designed to be a general purpose saliency model and has not been tuned to the task. However, Experiment 3 shows that the model does not cope with changes in the orientation as well as human observers and it struggles with some targets which humans can find within a couple of seconds. This suggests that the model's orientation channel is not functioning in the same way as in human visual search.

Chapter 5

Models of Visual Search

In the previous chapter I compared Itti and Koch's visual saliency algorithm to human performance in a series of visual search experiments involving a defect on an otherwise homogeneous textured surface. While the model proved a good match for the human data when a circular indent was used, it failed to mimic human behaviour when an elongated target was used. In this short chapter I will give a comprehensive review of previous work on modelling visual search before designing my own model in Chapter 6.

A complete, computational model of visual search should possess three components. Firstly some form of image-processing is needed for generating an activation map. This map should reflect the foveated nature of human vision (see Section 2.2.1) as the difficulty of a visual search task depends not only on the properties of the target and background, but also the distance the target is from the point of fixation and/or centre of the image [Motter and Holsapple, 2007, Najemnik and Geisler, 2008]. Secondly, the model needs to possess a method for generating saccades and choosing where to fixate next. Finally, a decision rule is needed, which can not only recognise and identify the target, but also terminate a trial if it cannot find the target. Most work on modelling search has concentrated on the first of these three components. (Note: the design and implementation of decision rules is outwith the scope of this thesis).

The aim of this review chapter is to identify what work has been done on computational models of visual search. In particular, what approaches are best suited to modelling human performance in the defect detection task in the previous chapter. This literature review is split into three sections, starting with a discussion of

theoretical models and the importance of guidance in search. This is followed by a discussion of computational models, firstly for stimuli consisting of arrays of search items, and then for more general, naturalistic stimuli.

5.1 Theoretical Models

Theoretical models accounting for visual search performance have mainly been concerned with attentional processes rather than eye-movements. However, this has not stopped these models being used to model fixations and saccades. (See Section 2.2.4 for a discussion of overt and covert attention and arguments for why eye movements should be considered and included in search models). There have been three main models of visual search: feature integration theory (FIT), its successor, guided search (GS) and the signal detection theory based model (SDT). FIT and GS are discussed below, while a discussion on SDT can be found in Section 2.2.3.

Feature integration theory was developed by Treisman and Gelade [1980] and attempts to explain the apparent distinction between fast parallel and longer serial searches. Feature searches are typically associated with searches in which the target is defined in terms of a unique feature, such as colour or orientation while targets defined by a conjunction of features tend to require longer, serial searches. FIT makes the assumption that only a small set of primitive features can support pre-attentive parallel search and if these fail then a serial process will take place, searching through all the items one at a time. The Guided Search Model was developed by Wolfe et al. [1989] and improves on FIT by allowed the serial process to be guided. Wolfe et al. argue that the existence of conjunction searches with shallow search slopes cause a problem for FIT. To solve this problem, GS allows the serial search mechanism access to the pre-attentive feature maps in order to facilitate guided search. For example, Figure 5.1 shows a classic example of a conjunction search. Neither the colour or orientation feature maps can locate the green, horizontal, target on their own, but they can help the serial component of search, allowing all the horizontal and green items to be picked out, reducing the number of items that need to be inspected.

Guided Search has continued to be developed over the last two decades [Cave and Wolfe, 1990, Wolfe, 1994, Wolfe and Gancarz, 1996, Wolfe et al., 1989] and is currently in its 4th incarnation, GS4 [Wolfe, 2007]. While equations and parameters

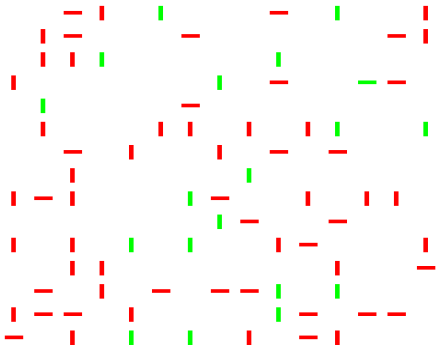

Figure 5.1: Exampe of a conjunction search. The search target is the green horizontal bar.

have been put forward for some aspects of the model's behaviour, it's computational aspect is still very limited in scope. Wolfe admits that the model is still limited to object based search (such as in Figure 5.1) and has no mechanisms for partitioning a continuous image into objects of interest [Wolfe, 2007].

GS4 allows both top-down and bottom-up information to influence the deployment of attention. In search, the bottom-up, saliency aspect is also referred to as attentional capture. See Egeth and Yantis [1997], Yantis [2000] and Rauschenberger [2003] for reviews. In GS, saliency is calculated by comparing the orientation and colour between different elements. If a search item differs in these dimensions from other nearby items it it considered to be salient, while differences between further apart items are considered less important. The top-down, guided, part of the model is assumed to work on broad categorical representations. For example, if the target is near horizontal, then attention will be directed towards search items with shallow slopes, while near-vertical slopes will be neglected. The effects of top-down guidance and bottom-up saliency are computed for each feature (colour and orientation) and then combined to give an activation map.

Despite Wolfe's stated preference for studying and modelling covert attention over scan-paths (see Section 2.2.4), saccadic selectivity has become a common way of assessing the degree to which searches are guided [Findlay, 1997, Hooge and Erkelens, 1999, Motter and Belky, 1998, Pomplun et al., 2001, 2003, Scialfa and Joffe, 1998, Shen et al., 2000, Tavassoli et al., 2009, Williams and Reingold, 2001, Williams, 1967]. This involves assigning the endpoint of each saccade to a display item. If search is guided then we would expect to see more saccades assigned to

display items which share features with the target. For example, Motter and Belky [1998] carried out an an analysis of scan-paths from rhesus monkeys during a search task and found that they used colour to guide their search. Similarly, Shen et al. [2000] carried out an experiment looking at saccadic selectivity during a conjunction search. Their target was a red X, among red O's and green X's. They varied the ratio of the two types of distracters and found that as the number of distracters sharing the same colour as the target decreased, saccadic selectivity, in terms of colour, increased.

There are some alternative ways to implement the concept of guidance. For example, top-down processes could modulate the search features to maximise the target's response, or to maximise the signal-to-noise ratio [Navalpakkam and Itti, 2005, 2007, Rutishauser and Koch, 2007]. Furthermore, although GS only considers guidance in terms of top-down, low-level processing, other types of guidance can influence search. For example Neider and Zelinsky [2006a] have shown that high-level factors such as scene gist and context can influence search. They carried out an experiment in which observers searched a photograph for either a jeep, blimp or a helicopter and found that more fixations were directed towards the ground when searching for the jeep, the sky when looking for a blimp and helicopter. Eckstein et al. [2006] have produced similar results.

5.2 Computational Models of GS

In the previous section Wolfe's Guided Search model was discussed. This section is concerned with computational implementations of GS. While GS is often taken as the starting point for computational models, many details vary from model to model. Two examples of computational models are the Area Activation Model [Pomplun et al., 2000, 2003] and the Probabilistic Model [Rutishauser and Koch, 2007]. Both of these models simulate human gaze patterns while searching for a target among distracters. As abstract stimuli are used the feature extraction stage of these models is fairly trivial as simple category labels (such as "'red"' and "'vertical"') are used. However, both models diverge in their saccade selection algorithms and the analysis used to compare the model to human performance.

Pomplun's Area Activation Model is based on the assumption that observers will direct saccades towards regions of the stimuli that will give them maximum task

relevant information. This means that not only are an individual item's features important, but also its spatial location in relation to other task relevant items. The model achieves this by applying a large Gaussian filter (referred to as the fixation field) to the feature responses to generate the activation map. Fixation locations are assumed to correspond to the local maxima of the activation map and the size of the fixation field is iteratively fitted to pilot data so that the model matches the same number of fixations as human observers. The model uses a deterministic "local minimisation of scan path length" algorithm [Pomplun, 1998, Zelinsky, 1996]: make a saccade to the nearest local maxima that has not already been visited.

Target detection mechanisms are not considered and the model is only compared to human observers on target absent trials. Furthermore, the Area Activation Model does not allow for inhibition of return to decay over time and hence the model is not capable of making refixations: once it has fixated each local maxima in the activation map the search terminates. Because of this, only the first five saccades made by human observers where analysed, in terms of saccade amplitude and saccadic selectivity. The model was found to be a good approximation to human behaviour under both measures.

While Rutishauser and Koch's [2007] model is also based on Guided Search, it is different to the Area Activation model in several respects: saccades are assumed to be made to individual objects rather than centres of gravity and the saccade target is determined in a more sophisticated way. Rather than use a deterministic rule, an element of randomness is introduced and the activation values for each search item are taken as the mean values for a Poisson process. For each saccade, the following is computed for each search item x:

$$F(x) = \lambda(x) + E(r)k_1 + \delta \tag{5.1}$$

where $\lambda(x)$ is a sample from the associated Poisson process, and r is the distance from the currently fixated item to x. The function E is set so that the model is more likely to make saccades of a similar length to those made by human observers. Finally, δ makes the model more likely to make a saccade in roughly the direction as the previous saccade. The model chooses to make a saccade to the item which gives the largest value for F. A simple inhibition of return mechanism means that the last n fixated search items are not considered as potential saccade targets. Also, unlike Pomplun's model, target detection is considered: on each fixation the model

considers a set number of search items with the highest activation within $D°$ of the currently fixated target. The effect of varying the different parameters is investigated and the model is fitted to human data. It does a good job of matching human behaviour in terms of the number of saccades required to find the target, saccadic selectivity and saccade amplitudes.

5.3 Naturalistic Stimuli

The search models discussed so far have been limited in scope to discrete item search. However, modelling visual search in naturalistic stimuli is a more 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 et al., 2002, Zelinsky et al., 1997]. The problem becomes more difficult when we consider camouflage backgrounds [Neider and Zelinsky, 2006b] where the distracter items are salient but the target is not.

Furthermore, when we move to the task of modelling search on more naturalistic stimuli we encounter the problem of deciding what features to use, and importantly, how to measure them. From visual search experiments using discrete items we know 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 colour, orientation and size. 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. In these, even simple low level features such as local contrast, colour and orientation can be measured in many different ways. Zelinsky [2008] sums it up well:

> Complicating the extension of a search theory to realistic contexts is the selection of an appropriate representational space. The problem is that the dimensions of this space are largely unknown. Although most people would agree that a coffee cup consists of more visual features than a coloured bar, it is not apparent what these features are.

5.3.1 Guidance and Saccadic Selectivity in Natuarlistic Stimuli

Pomplun [2006] investigated how top-down knowledge about the target influenced fixation locations in a visual search task using greyscale photographs of natural scenes (landscapes, gardens, city scenes, buildings, and home interiors). On each trial, participants were shown a different target region to find in a photograph. Hotspot maps, referred to as *attentional landscapes* by Pomplun, were constructed for each image by centring two dimensional Gaussian distributions, with standard deviations of one degree of visual angle, on each fixation location, normalising, and summing across all fixations and observers. Intensity, contrast, orientation and spatial frequency features were constructed and the target's features were compared to the image regions fixated using the attentional landscape. Significant interactions between target and fixated features were found for all four features, however, only the intensity feature provided a strong main effect for the target's feature indicating top-down guidance. The other three features all exhibited significant main effects on the fixation regions indicating a bottom-up component. Chen and Zelinsky [2006] carried out a similar experiment using photographs of everyday items as stimuli. They used colour saturation as a way of controlling the items' saliency and compared reaction times and saccadic behaviour between trials with and without target preview. They found that when a preview was available observers' reaction times were faster. Additionally, a higher percentage of initial saccades were directed towards the target and the salient colour singleton only attracted attention when a target preview was not given.

Saccadic selectivity has also been investigated in $1/f$-noise stimuli [Rajashekar et al., 2002, 2004, 2006, Tavassoli et al., 2007a,b,c, 2009]. The classification image (CI) paradigm is commonly used in visual discrimination tasks [Ahumada, 1996, Beard and Ahumada, 1998] and Rajashekar et al. [2002] applied it to the analysis of fixations made during visual search. This involves calculating the mean local neighbourhood across all fixations made during a search for a simple geometric shape (a circle, triangle, dipole) embedded in $1/f$ noise. With this stimuli the classification images resemble the search target [Rajashekar et al., 2002, 2004, 2006]. A simple predictive model was implemented which involved filtering the search stimuli with the relevant CI. A k-means clustering algorithm was then applied to the empirical fixations and the cluster centres were found to correspond to local maxima in the prediction map. This method will be discussed further in 7.7.

Tavassoli et al. [2007a] carried out a variant of this procedure that involved dividing the $1/f$ stimuli into discrete subregions, with the target centred, and occupying most of a region. The new method was found to be more efficient than Rajashekar et al. [2004]'s, requiring fewer empirical fixations to give target-like classification images. It has also been used to investigate orientation anisotropies at fixated regions in $1/f$-noise [Tavassoli et al., 2007b]. A further study [Tavassoli et al., 2009] used Gabor patches as the target and used spectral analysis to investigate fixated image regions. They found evidence for both top-down and bottom-up guidance.

Najemnik and Geisler have taken a different approach and have derived the Ideal Observer for a search task involving a target embedded on noise Najemnik and Geisler [2005, 2008, 2009] . This model is based on a detection rule, empirically obtained from a signal detection experiment. The authors point out that the model is theoretical and while it can make predications about human behaviour, it is not a computational, image-processing model. It will be discussed in more detail in Chapter 7.

5.3.2 Modelling Search with Naturalistic Stimuli

A possible starting point for a visual search model is to use a saliency algorithm [Gao et al., 2008, Itti and Koch, 2000] (see Section 4.1). Unlike Guided Search these saliency models are computational, in that they can be used to generate a saliency map for any given image. Pre-attentive feature maps (colour, 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 behaviour for feature and conjunction searches using red/green and horizontally/vertically orientated bars [Itti and 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 and 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 maximise 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 et al. [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 behaviour reported in Zelinsky et al. [1997], the model accesses successively 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, the task used by Zelinsky et al. [1997] was relatively easy for participants to carry out: they typically only needed to make 3 fixations to find the target. Furthermore, the search items where arranged on an semicircle, with the fixation cross placed below the items. This arrangement of search items appears to have encouraged observers to show much less variation in their scan-paths than is typically seen in more general visual search tasks and the first fixation was directed towards the centre of the semicircle in most cases. These two factors encouraged the human observer to behave systematically. While this model shares some similarities with the saliency model discussed above, the fact that it incorporates information about the search target is a crucial difference.

More recently, Zelinsky [2008] has introduced his Target Acquisition Model (TAM), which builds on the earlier work by Rao et al. [2002]. TAM is a computational model and has been designed to fulfil three criteria that Zelinsky suggests a general search theory should meet. Firstly, Zelinsky argues that a model should be computationally explicit. The model should also be able to operate over different classes of stimuli without needing different parameters and should be able to work on stimuli of different complexities. Finally, any model of eye-movements should incorporate a model of the foveated retina, as without one, eye-movements would not be needed. In order to meet this last condition Zelinsky uses Geisler and Perry's model [Geisler and Perry, 1998, 2002, Perry and Geisler, 2002].

TAM is an image based model and as inputs it is given an image of the target (taken from the search stimulus) along with the search stimulus. It uses three channels: luminance and two opponent colour channels. Each of these channels are convolved with a bank of 1st and 2nd order Gaussian derivatives resulting in a 72

dimensional feature vector for every pixel location in the retina-transformed search image. A feature vector is also taken from one pixel in the target image, and this is correlated with the feature vectors from the search image to give a *target map*. An *inhibition map* is added to the target map in order to inhibit rejected distracter objects. (The inhibition does not decay with time but this is suggested for future work.) TAM includes a simple target detection rule which reports the target as found if the maximum of the target map is greater than 0.995. If the target is not found then the model makes a saccade, either to the current maxima (if it is not already fixating it), or the spatial average of the target map.

Another recent computational model has been developed by Hwang et al. [2009], which builds on earlier work by Hwang et al. [2007], Pomplun [2006] and the Area Activation Model. The model used eight features: two opponent colour channels, luminance, intensity, two features for direction, and two features for complexity. Feature vectors from the target and the image were compared using the Histogram Interaction Similarity Method [Swain and Ballard, 1991] and used to create *similarity landscapes* and compared to the attentional landscapes obtained empirically. These similarity landscapes are combined, although Hwang et al. argue that a weighted product is more suitable for modelling top-down guidance than the more commonly used weighted sum. They found that their model predicated fixation locations of an observer as well as another observer's scan-path.

Finally, the computer vision community has also developed some models of visual search. These models are slightly different from the ones discussed above in that they are not trying to model human behaviour and experiments are rarely carried out. An example is VOCUS, developed by Frintrop et al. [2005] (also see Frintrop [2006]). The model builds on earlier work by Backer et al. [2001] and Itti et al. [1998] and applies a top-down modulation of feature maps. VOCUS has been shown to work well in a variety of search tasks , successfully finding the target in three or fewer fixations.

5.4 Conclusions

In this chapter I have given an overview of the different approaches that have been used to design computational visual search models. As I have shown, these models can be broadly split into two categories: those which are designed for stimuli consisting of discrete search items and those which use image processing techniques and

can be applied to more general (photographic) stimuli. The first group of models [Pomplun et al., 2003, Rutishauser and Koch, 2007] will not work with the textural stimuli used in this thesis as they contain no well defined search items.

A common feature of the search models that work on naturalistic stimuli [Hwang et al., 2009, Rao et al., 2002, Zelinsky, 2008] is that they generate a feature vector from a pre-defined target, and use this to weight the activation map. This requirement of having a perfect copy of the target as an input, without which the model can not generate a valid feature vector for the target, does not hold for the surfaces and defect targets considered in this thesis. Even the relatively small changes between circular indents can cause large changes in filter responses. However as the previous chapter showed, human observers are quite capable of finding the defect under a wide range of conditions. To get around this problem, I will use a similar approach as Itti and Koch's visual saliency model and search for image regions which differ from their surroundings, rather than regions which are similar to a pre-defined target.

Chapter 6

An LNL-Based Search Model

6.1 Introduction

In Chapter 4 I investigated how human performance in a defect detection task varies with surface and target properties such as regularity and orientation. Human performance was also compared to a bottom-up visual saliency algorithm [Itti and Koch, 2000, Walther and Koch, 2006]. The results showed that the model only partially fitted the human data: in particular there is a discrepancy between the performance of the model and human observers when searching for an elongated defect (Experiment 3).

Chapter 5 contained a comprehensive review of how the problem of modelling visual search has been tackled in the past and in this chapter I will attempt to construct a search model which can simulate human behaviour in an unsupervised surface defect detection task. The model is based on an LNL-framework (see Section 2.1.2) and will be compared to human performance in a series of experiments, using a *target always present* design. This will remove the need for considering speed/accuracy trade-offs. Furthermore, it avoids the problem of defining a decision rule for the model for target absent trials. Instead, the model is assumed to find the defect when it fixates on it and only one measure, the number of saccades needed to find the target, needs to be used in order to compare performance between the search model and the human observers.

6.2 Texture Discrimination and LNL Models

The problem of texture segregation, segmentation and discrimination has been tackled by both the fields of computer vision and perception [Bergen and Julesz, 1983, Bergen and Landy, 1991, Julese, 1981]. Much of the modelling work has made use of LNL models (linear-nonlinear-linear, also referred to as FRF[1] and the backpocket model) which are based on properties of the functional architecture of the primary visual cortex [Bovik et al., 1990, Malik and Perona, 1990, Morrone and Burr, 1988, Randen and Husoy, 1999a,b]. (See Section 2.1.2 for more details.)

As the LNL framework is both biologically inspired and an effective model for texture segmentation, it will be used as the basis for the construction of a search model in this chapter. Several other search models also make use of this framework [Itti and Koch, 2000, Zelinsky, 2008] however they are rarely expliticity described as such.

6.3 Model Design

The model has two parts. The first part takes the form of an LNL (or FRF) process and is used to generate an activation map, A. One of the principles behind the implementation of the LNL process is to keep it as simple as possible. For example, unlike the visual saliency algorithm that was used in Chapter 4, the model detailed here only contains an orientation channel (as again, colour is outside the scope of this thesis). The contribution to saliency from image contrast is assumed to be picked up in the orientation channel: the sum of the Gabor filters over a single scale gives an approximation to the bandpass filters commonly used for contrast features (see Figure 6.1). Secondly, rather than use an iterative inhibition-excitation process to weight the individual feature responses, a far simpler non-linear process is used in which the filter response map is divided by its mean. Details are given in Sections 6.3.2 to 6.3.4.

The second part of the model is an algorithm for generating saccades. This involved taking foveal vision into account, and including an inhibition of return (IOR) mechanism so that the model does not immediately re-fixate previously fixated image regions. In order to improve the fit with human observers, the model includes

[1] filter-rectify-filter

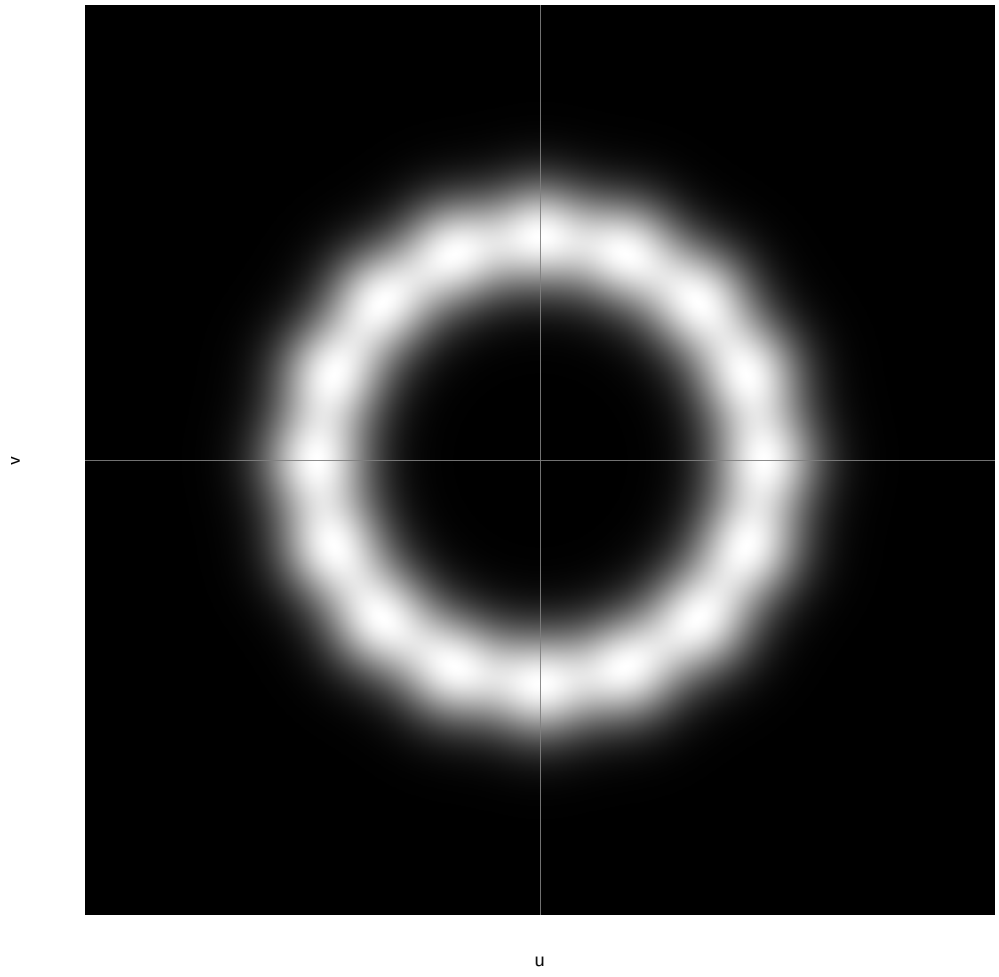

Figure 6.1: The image above shows frequency domain representiations of all eight Gabor filters used in the model for a given spatial scale ($r = 6$). Note how it approximates a DoG bandpass filter

a stochastic component and randomly selects one of the most salient regions in the activation map to fixate.

6.3.1 Scope

The problem of target identification is not considered here: the model will keep making saccades until it has fixated on, or near, the target. As is common with several other models of visual search, search behaviour in target absent trials is not considered [Rutishauser and Koch, 2007, Zelinsky, 2008].

The model receives the test image, I, and the target's location as inputs, and it is a *guided search model* in the sense that it directs saccades to local maxima of the activation map. This guidance takes place in parallel with the whole stimuli being considered on each fixation. However, as it is an unsupervised defect detection model, it is not trained on the target or surface and hence does not incorporate any top-down information about the properties of the background or target. Although the model is not given any explicit top-down knowledge of the target's properties, the defect is assumed to be detectable in terms of bottom-up features alone as it is a unique anomaly on an otherwise homogeneously textured surface.

Target detection is assumed to either require a fixation or, alternatively, detecting the target away from fixation will cause a saccade to be made towards the target: this distinction is outside the scope of the model. This seems a reasonable assumption as over 80% of human reponses in Chapter 4, Section 4.3 occured when they were fixating within 1° of the target.

6.3.2 1st Linear Stage

The first stage is linear and consists of a bank of Gabor filters. As the stimuli are greyscale, colour channels are not considered. 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. (See Figure 6.1 for an example.) Based on the results of pilot studies eight equally spaced orientation channels were used.

Gabor filters have parameters σ_u, σ_v, u_0 and ϕ:

$$G(u,v) = e^{-\frac{1}{2}\left(\frac{(u-u_0)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}\right)} \tag{6.1}$$

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

6.3.3 Non-Linearity

The second stage is non-linear and it has two aims: first, it rectifies the negative responses from the filters and second, it applies weights to the feature maps to increase the target's signal against the background noise. (Note: this noise is the filter's response to non-defective patches of the surface texture.) This weighting is achieved by first normalising each response map to $[0, 2]$ before then dividing each map 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 emphasised relative to maps containing no strong peaks. A simple example is shown in Figure 6.2.

6.3.4 2nd Linear Stage

The second linear stage consists of a smoothing filter and response map summation. If the smoothing filter is weak then the model will consider local maxima, corresponding to small salient objects, as saccade targets. If a stronger smoothing filter is used then the model will direct its attention towards centres of gravity. This allows the same model to potentially behave in a similar way to both models which fixate centres of mass between interesting objects [Pomplun et al., 2000, Rao et al., 2002] and to models which fixate individual search items, such as Rutishauser and Koch's, [2007]. A two dimensional Gaussian with $\sigma_u = \sigma_v = 3.75$cpd is used for this task.

Finally, all the feature maps are summed across scales, passed through the non-linear operator again, before being summed across orientations to give the activation map, S. Note that if I were to remove the normalisation 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, I am only interested in detecting small

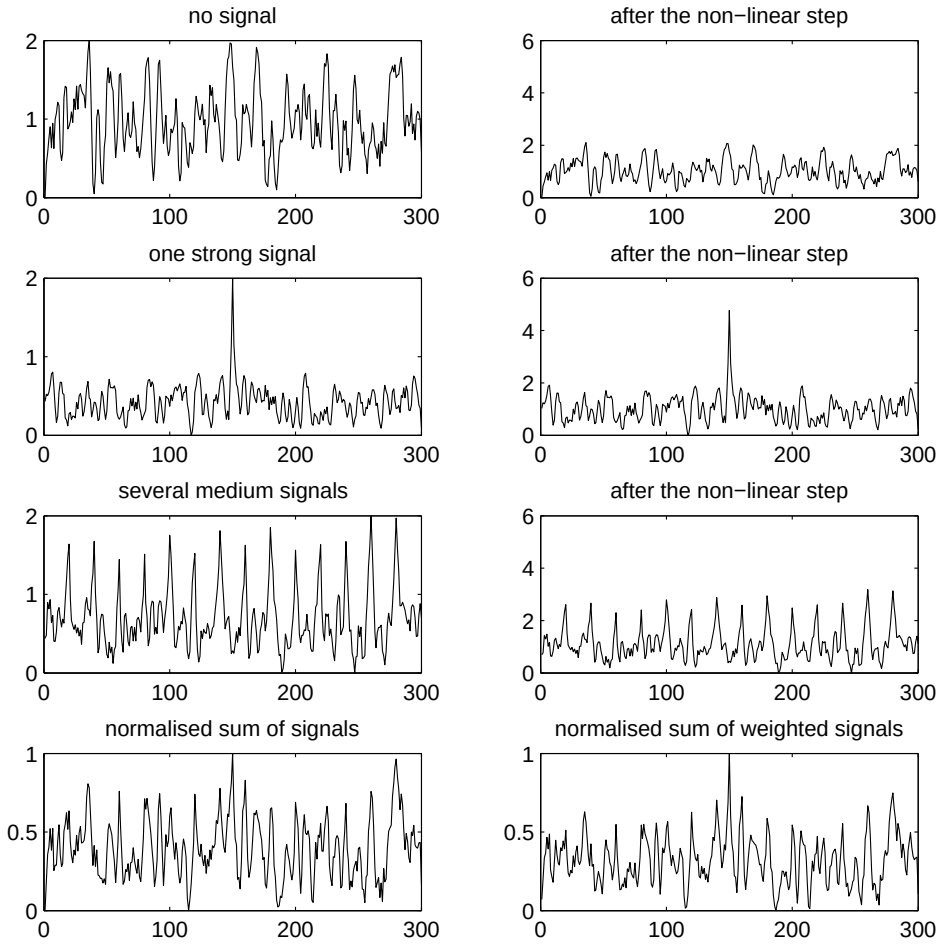

Figure 6.2: Example of the non-linear step. (Note: this is a simple example with dummy data to illustrate the non-linear step. No filtering is applied.) The top three plots on the left show noisy signals, normalised to $[0, 2]$. The top figure contains no signal, second top has one signal, while third top contains many spikes. The bottom-left figure contains the result when these signals are added together. The corresponding figures on the right show the result of dividing each signal by its median. This has the effect of giving greater emphasis to signals with a strong peak: the maximum for three signals, after the non-linear step, are 2.10, 4.77 and 3.18 respectively.

regions differing in energy content from their backgrounds, hence the use of the $[0, 2]$ scaling and median division.

The end result of this process is the activation map, $S(x, y)$. This gives an indication of how different a region of texture is from its surroundings. This activation map is then used to generate a sequence of saccades.

6.3.5 Generating Saccades

The model generates saccades using a three stage process. First a negative exponential mask is used to weight the activation map to take foveal vision into account:

$$F_d(x, y) = S(x, y) \cdot e^{-kd} \tag{6.2}$$

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 et al. [2005] suggests this offers little improvement over the method used here. They compared a model based on previously published contrast-detection and orientation-discrimination thresholds [Virsu and Rovamo, 1979] with simpler approximations based on a two dimensional Gaussian fall-off [Parkhurst et al., 2002] and a negative exponential. They found that all three models offered an improvement against a baseline model with no eccentricity effects, and that the simple exponential method works at least as well as the more rigorous model, which in turn out-performed the Gaussian mask.

Finally, a inhibition of return (IOR) mechanism is applied, implemented as a series of two dimensional masks centred on the location of previous fixations. Gaussian masks are used for this, with the strength of inhibition decaying over time:

$$F(x, y) = F_d(x, y) \cdot \prod_{t=1}^{n} (1 - \frac{2}{t+1} I_t(x, y)) \tag{6.3}$$

where t is the fixation number and $I_t(x, y)$ is a two dimensional Gaussian mask (normalised to $[0, 1]$) centred at (x_t, y_t) with $\sigma = 45$ pixels. This method was chosen due to its simplicity. For more information on models of inhibition of return see Klein [2000].

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; taking too few saccades to locate the target. To tackle this I will use a similar approach as Rutishauser and Koch [2007] and use a simple stochastic process is used: the k (= 3) largest local maxima in the resulting fixation map are considered as potential saccade targets, $F(x_i, y_i)$, and are assigned probabilities:

$$p_i = \frac{F(x_i.y_i)}{\sum_{i=1}^{k} F(x_i, y_i)} \tag{6.4}$$

These are then used to choose which of the k 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.

6.4 Methods

In order to test the model against human performance a *target always present* visual search task was carried out. 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 considering only the target present trials the interpersonal variance can be reduced and the search process can be modelled separately from the decision process. The experiments were set-up in the same way as Experiments 1-4, see Section 4.2 for more details.

6.4.1 Stimuli

The stimuli were created in the same way as those in Experiments 1, 3 and 4. In addition to looking at surface properties, and target orientation, the target's eccentricity (distance from the centre of the stimulus) was varied. Specific results are given in Section 6.5.

6.4.2 Observers

Seven subjects were used for each experiment: some subjects took part in more than one experiment, all had normal or corrected to normal vision, and all were between 18 and 30 years old. Subjects were given several practice trials and they were informed that the target would be present in all trials and would always be an indent in the surface of the same size and shape (or a missing texton, in the case of the near-regular experiment). They were instructed to respond by pressing the space bar on the keyboard once they had found the target. No time limit was imposed on the task. Subjects were told to inform the supervisor if they were having great difficulty in finding the target, in which case they were allowed to skip the trial (in practice this accounted for less than 1% of trials).

6.4.3 Search Model

As the visual search model is stochastic it was run 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 the experiments detailed below which is comparable with human performance. This maximum limit is somewhat arbitary and is only included to stop the model endlessly looping when it can not find the target. The small number of trials on which the model failed were not included in any further analysis.

6.5 Results from Experiments 5 - 7

The aim of these experiments is to compare how well human observers and the LNL-based search model can find a small defect in an otherwise homogeneous surface texture. In each experiment, the observers managed to find over 99% of the targets. In the small number of trials in which they could not locate the target, they had spent at least 2 minutes searching. These trials were not included in any further analysis.

6.5.1 Experiment 5: $1/f^{\beta}$-noise: Surface 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.

Stimuli

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 ($\beta = 1.6, 1.65, 1.7$ and $\sigma_{RMS} = 0.9, 1.1$). For each trial a target was positioned randomly on a circle, centred on the middle of the image, with radius $1.7° \pm 0.7°, 3.8° \pm 0.7°$ or $5.9° \pm 0.7°$ visual angle.

Results

The psychophysical results in this experiment are shown in Figure 6.3. Comparing these results with those in Section 4.3, Figure 4.4 we see that the mean number of saccades to find the target is larger in the current experiment. This is because the mean includes some difficult trials which, if given the option, as in Experiment One, the observer would likely have responded *target absent*. However, due to the different task they had to carry on searching.

All three variables had a significant effect on the mean number of saccades: for β, $F(2,5) = 43.1$, $p = 0.001$; for σ_{RMS}, $F(1,6) = 71.8$, $p < 0.001$; and for r, $F(2,5) = 14.6$, $p = 0.008$. Additionally, there was a significant interaction between the two

parameters that controlled roughness, β and σ_{RMS}: $F(2,5) = 30.6$, $p = 0.002$. There was no significant interaction between target eccentricity and either of the two roughness parameters. The distributions of saccade amplitudes and orientations are shown in Figure 6.4 and the distribution of saccade directions shows the same horizontal bias as reported by Gilchrist and Harvey [2006].

6.5.2 Experiment 6: $1/f^{\beta}$-noise: Target Orientation

Stimuli

This experiment is similar to Experiment 3, Section 4.3.3 and the target is again an elongated indent. As observers have to carry on searching for the target and are not given the choice of giving up and responding *target absent* , target orientations close to vertical were not used: $\theta \in \{90° \pm \phi | \phi = 10°, 15°, 20°, 30°, 45°, 90°\}$, where 90° corresponds to the direction of illumination and, due to symmetry, 0° = 180°. Two values of β were used (1.6 and 1.7) and the target was randomly located with the constraint that it was between 5° and 6.7° away from the centre of the stimulus.

Results

The results are shown in Figure 6.5 and, again, they agree with those in Chapter 4. Orientation has very little effect on the visibility of the target until it is near vertical. Once the target's orientation is greater than 75° there is a rapid rise in the number of saccades required to find the target. The effect is greater for the rougher surface, ($\beta = 1.6$), than for the smoother ($\beta = 1.7$).

6.5.3 Experiment 7: Near-Regular Textures

Stimuli

In order to test the generality of the model it was also applied to a different defect detection problem: finding a missing lattice point in near-regular textures. An example is shown in Figure 3.2. Two texton densities where used, $\rho = 1.875$, and $\rho = 2.461$ textons per degree. The near regular lattice governing texton placement

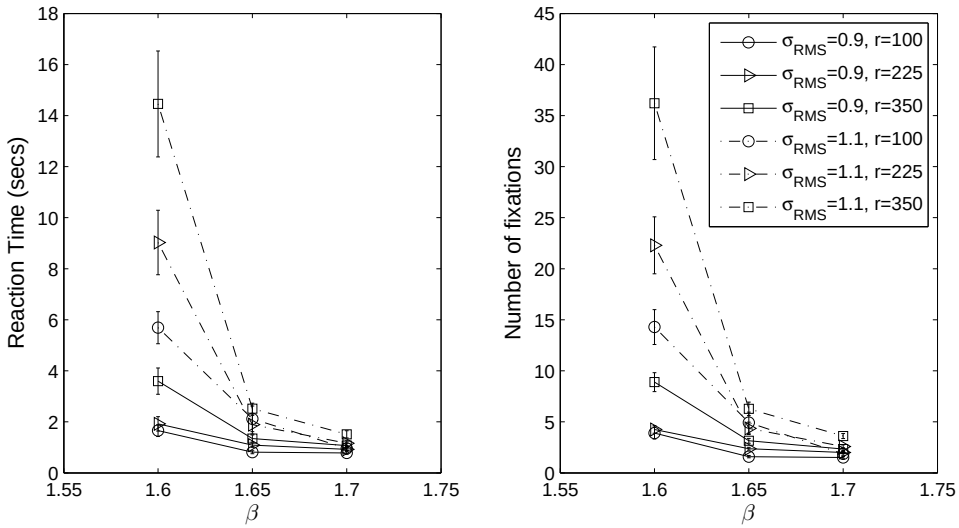

Figure 6.3: Inter-observer mean reaction time (Left) and number of saccades to target (Right) plotted against surface roughness for Experiment 5.

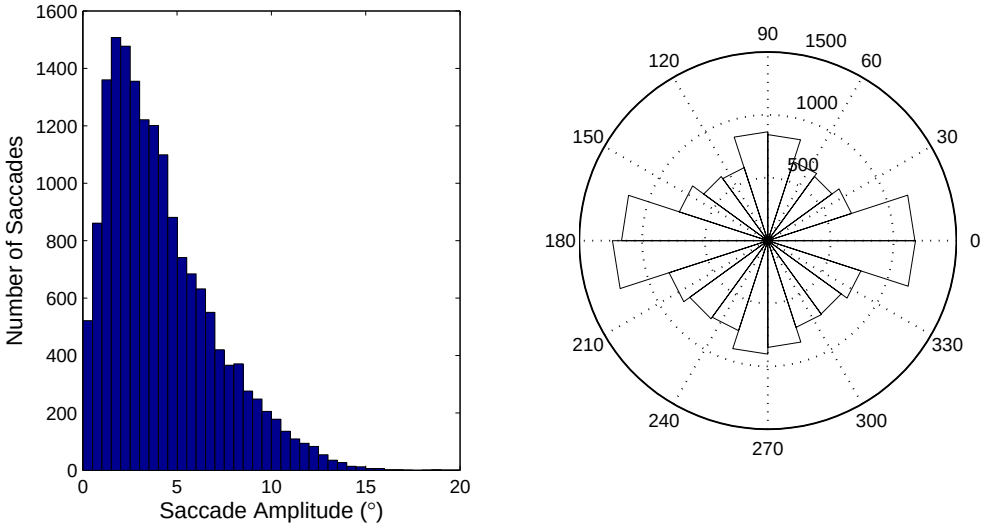

Figure 6.4: Saccade amplitude histogram and saccade direction rose plot for Experiment 5.

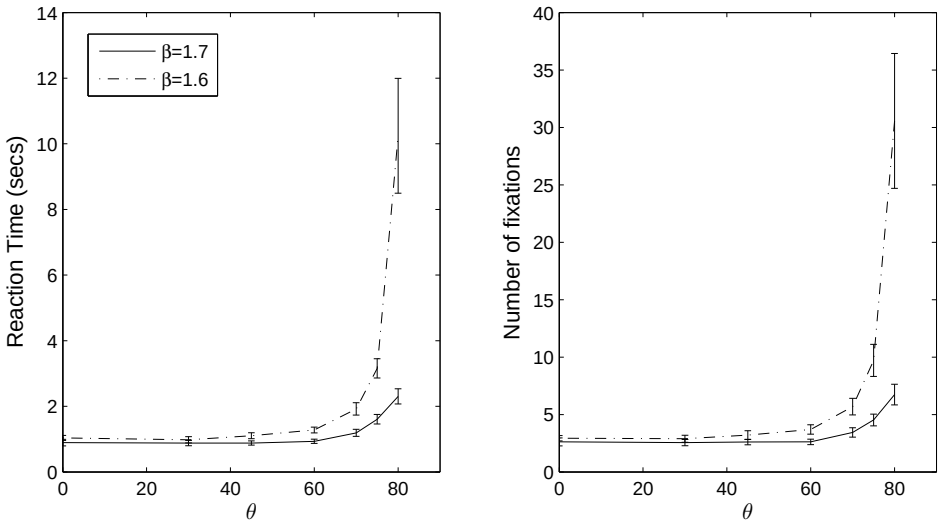

Figure 6.5: Inter-observer mean reaction time (Left) and number of saccades to target (Right) plotted against target orientation for Experiment 6.

was deformed by applying a normally distributed error to the placement of each texton. By varying the standard deviation of this error, the regularity of the near-regular texture can be controlled. This experiment used $\sigma_p \in \{0, 1/2, 1, 2\}$. See Section 3.2.2 for further details. The target was randomly locations with eccentricity $\approx 3.33°$ or $\approx 6.67°$.

Results

The results from the psychophysical experiment are shown in Figure 6.6. All three parameters affect the mean number of fixations required to find the target: for σ_ρ we have $F(3,4) = 41.629$, $p < 0.001$; for r we have $F(1,6) = 28.309$, $p = 0.002$ and finally for ρ we have $F(1,6) = 25.698$, $p = 0.002$. There is also an interaction between the two parameters controlling the surface's appearance, σ_ρ and ρ, with $F(3,4) = 5.890$ and $p = 0.006$. As with the $1/f^\beta$-noise surfaces interactions involving r are not significant.

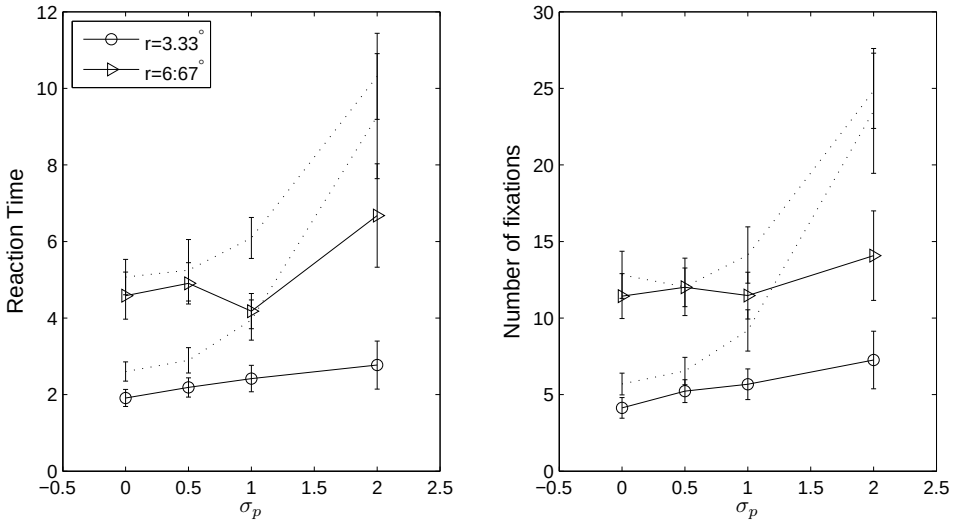

Figure 6.6: Inter-observer mean reaction time (Left) and number of saccades to target (Right) plotted against surface regularity for Experiment 7. The solid line shows the results for $\rho = 1.875$ while the dashed line shows $\rho = 2.461$.

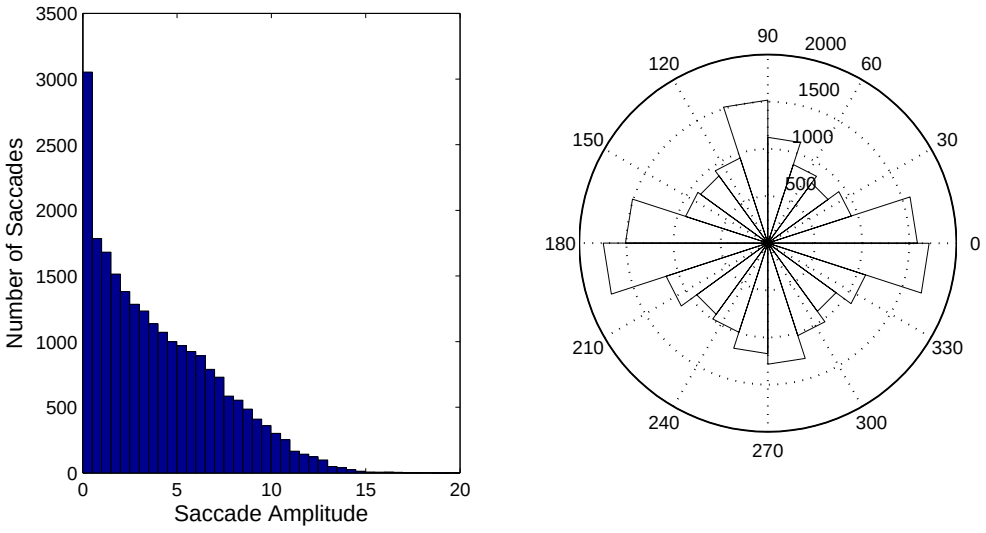

Figure 6.7: Saccade amplitude distribution and direction rose plot for human results on the near-regular textures, Experiment 7.

Experiment	Human Observers (std)	LNL-based Model (std)
5: Roughness	7.02 (1.33)	7.80 (0.67)
6: Orientation	5.99 (2.01)	7.71 (0.30)
7: Near-regular	11.25 (2.57)	10.24 (1.01)

Table 6.1: The mean number of fixations required to find the target for Experiments 5-7, compared to the mean number of fixations required by the LNL-based model.

6.6 Comparison with Model

6.6.1 Surface Roughness

The mean number of saccades over all trials can be seen in Table 6.1. An independent t-test comparing the human results with those from running the model seven times gives $t(12) = -1.381$, $p = 0.192$. Therefore, the null hypothesis, that the means of the human and model populations are not significantly different, is not rejected. Figure 6.8 shows how the mean number of saccades taken by the human subjects and the model to find the target varies with surface roughness. A four way mixed model ANOVA was carried out on β, σ_{RMS}, r and δ (which distinguishes between instances of the model and human subjects). As above (Section 6.5.1) the ANOVA shows significant effects for β, σ_{RMS}, r. However, δ does not have a significant effect ($F(1,12) = 1.8944$, $p = 0.194$) and neither do its interactions: $\beta \times \delta$ has $F(2,1) = 0.427$, $p = 0.658$; $r \times \delta$ has $F(2,1) = 0.190$, $p = 0.829$; $\sigma_{RMS} \times \delta$ has $F(1,1) = 0.01$, $p = 0.921$. Similarly, three and four way interactions involving δ are also non-significant. Hence there is no evidence that the human observers and the model are affected differently by any of these parameters.

6.6.2 Target Orientation

In Section 4.3.3 it was shown that Itti and Koch's 2000 saliency algorithm does not give similar performance to human observers in a target absent/present forced choice task. In particular, it appears to be too sensitive to changes in the elongated target's orientation. In this section the LNL-based search model is compared to human results in the *target always present* experiment.

The overall means are shown in Table 6.1. Again, the means from each run of the model are close to the human means and an independent t-test does not detect

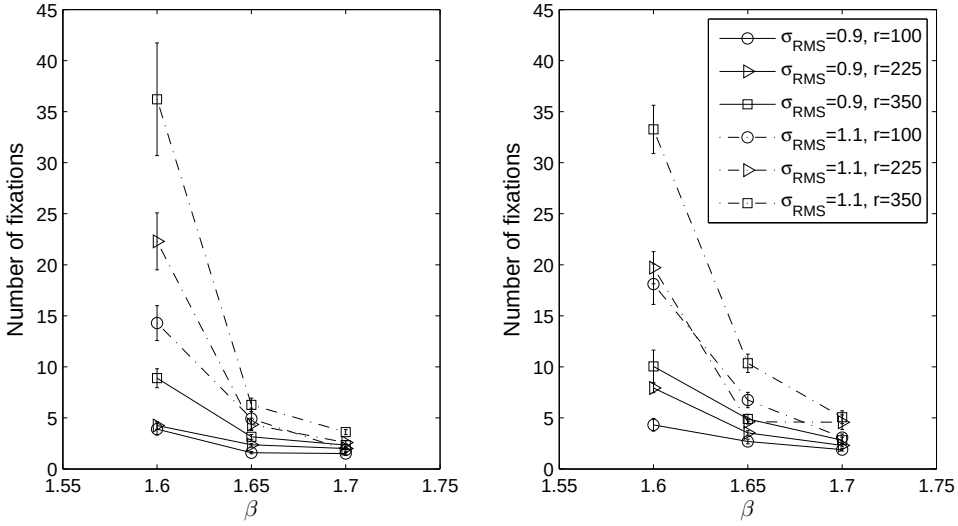

Figure 6.8: Comparison between human results (Left) and the LNL-search model (Right) in terms of the number of saccades required to find the target for Experiment 5.

any differences between the populations: $t(12) = 1.67$, $p = 0.13$. 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 6.9).

As the effect of θ is not linear, and the variance is not uniform, it is not appropriate to carry out an ANOVA. However, it is clear from Figure 6.9 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 $\theta = 80°$ at which point the model fails to match the mean human performance. However, this difference is not great and there is a large amount of variance between the human observers. I conclude that the model responds to changes in the target's orientation, θ, in a similar way as the human observers.

6.6.3 Near-Regular Textures

Finally, the search model was tested on near-regular textures and the results showed 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 6.1.) However as can be seen from Figure 6.10 the model's behaviour is different as σ_ρ is varied. While a four way, between subjects ANOVA does not detect a significant main effect for δ, $F(1,4) = 2.702$, $p = 0.131$, the $\delta \times \sigma_\rho$ interaction is significant ($p < 0.001$).

With both the model and the human 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 behaviour and are making far more saccades with amplitude $0.5° - 1°$ (compare Figures 6.4 and 6.7). This difference appears to be independent of the parameters σ_ρ, r, ρ, and is exhibited by all subjects. This suggests that some feature of the stimuli not captured in the 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.

6.6.4 Saccade Statistics

While the LNL-based search model succeeds in modelling human performance (in terms of the number of saccades required to find the target), it does 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 6.11 for an example). As the example shows, human observers often make long saccades that cannot be explained using the eccentricity dependant exponential fall-off. While one possibility would be that the fall-off function is too strong, this suggestion can be discarded 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 6.11 is typical of the model's behaviour the saccade targets for the model are compared with those chosen by human observers in Experiment 5, (Section 6.5.1). Over all the trials and all non-target fixations, only 22% of

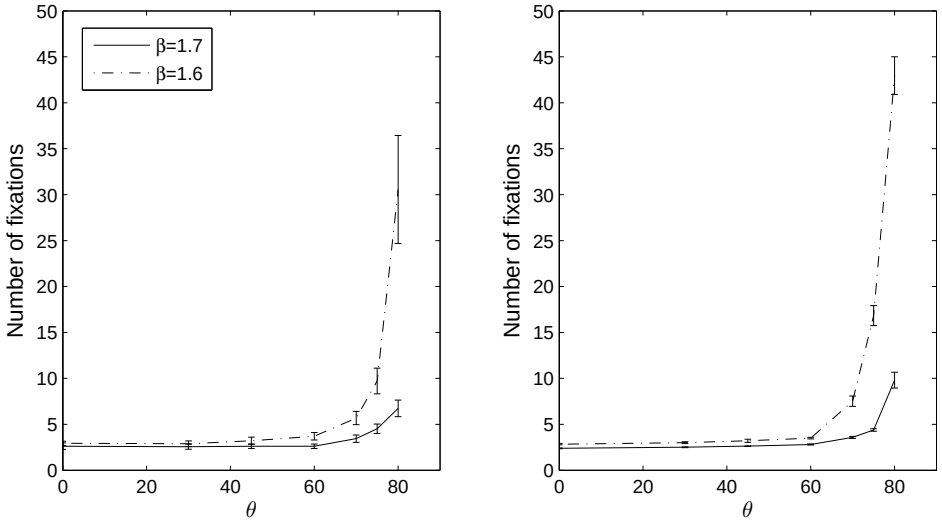

Figure 6.9: Comparison between human results (Left) and the LNL-search model (Right) in terms of the number of saccades required to find the target for Experiment 6.

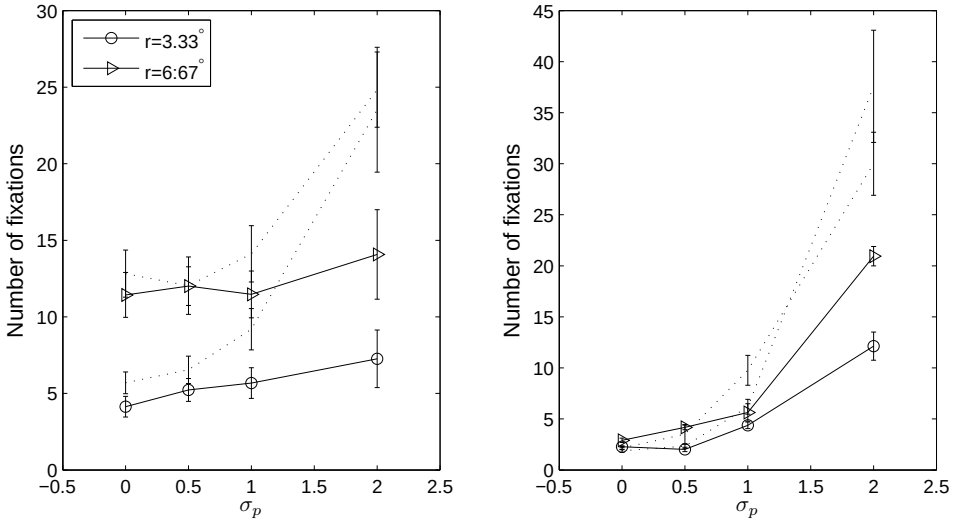

Figure 6.10: The results from Experiment 7 for (Left) $\rho = 1.875$ textons per degree and (Right) $\rho = 2.461$ textons per degree.

the saccades made by human observers were directed to within 1° of one of the three saccade targets considered by the model. Furthermore, over 25% fell more than 4.3° (equivalent 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. For example, 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°. In this case the fixations will occur somewhere within a circle with area $A = r^2$ and hence the probability of the human saccade landing within 1° of one of the model's saccades is:

$$p = \frac{3\pi}{A} = \frac{3}{r^2} \tag{6.5}$$

If we take $r = 4°$ (over half of the human saccades are under 4° in amplitude) then human observers would be expected to fixate within 1° of one of the fixation locations considered by the model 19% of the time. This is close to the 22% obtained from the empirical comparison.

This analysis suggests that the LNL-based search 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.

6.7 General Discussion

The results from the three psychophysical experiments above agree with those from Chapter 4. As well as investigating human search on $1/f^{\beta}$-noise surfaces the earlier chapter also compared human performance with Itti and Koch's saliency model (Section 4.4) and found that the saliency 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 in this chapter suggest that a simple LNL-based model, using a Gabor filter bank, offers a better match with human performance in this

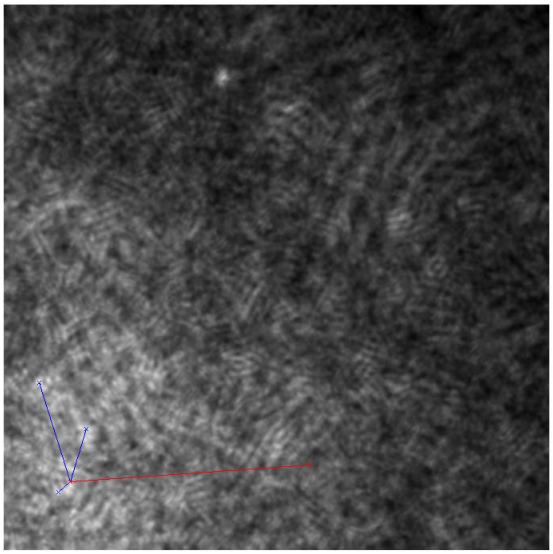

Figure 6.11: Comparison between model and human saccade selection. The red line shows the saccade made by a human observer while the blue lines show the three saccades considered by the LNL search model.

search task. The first experiment involved searching for an indent on a rough, $1/f^{\beta}$-noise surface (Section 6.5.1). 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 also responded to changes in background roughness and in target orientation (with respect to the illumination direction) in the same way as the human observers (Section 6.5.2) Finally the model was evaluated with a quite different surface-target combination: a near-regular texture with a missing lattice point (Section 6.5.3). 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.

6.7.1 Discrete Item Visual Search Stimuli

While the LNL-based search 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 6.12 shows an example of a standard pop-out effect in visual search where the target differs from the distracters 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 the search model is not limited to finding targets in surface textures.

Most models that simulate search tasks among discrete items [Pomplun et al., 2003, Rutishauser and Koch, 2007] depend on feature labels such as red/green, or horizontal/vertical, rather than using statistics measured directly from the stimulus. While the LNL-based search 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 it can fixate on centres of gravity, in a similar way to the Area Activation model [Pomplun et al., 2003]. As human observers make fixations on both items and centres of gravity, a model of visual search should incorporate both types of behaviour. Since the parameters of the LNL-based search model were identical in both the texture experiments, and in the search for a discrete item, (Figure 6.12), these results suggest a generality beyond the context of $1/f^{\beta}$-noise surfaces in which the model was developed.

6.7.2 Comparison with other Search Models

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 and Koch, 2000] and the $1/f^{\beta}$-noise stimuli used here (Chapter 4). Although Rao et al.'s [2002] model offers a very good simulation of human search, 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 the LNL-based search model covers a much wider range of task difficulty, with easy trials requiring only one or two fixations while the difficult trials need over 40.

The search task used by Najemnik and Geisler [2008] is similar to the one considered here, although they used $1/f$-noise directly whereas here, a rendering model was used to produce naturalistic images of textured surfaces. Their aim was to derive the theoretical ideal observer, in terms of search strategies, and compare it to human behaviour. While their model gives a good account of human search strategies they do not propose what image features or filters should be used to generate the activation map. More importantly, they only consider a finite number of possible target locations which has the effect of simplifying the derivation of the Ideal Observer. While this gives a more elegant model, it means that their search strategy can not easily be applied to my stimuli in which the target can be located at any location (down to pixel level): Najemnik and Geisler consider 85 potential target locations while we have 1024×1024 pixels to consider. Not only does the large increase in potential target locations create computational problems, moving to the pixel level causes some of the underlying assumptions of the model break down. In particular, the activation at a particular pixel cannot be assumed to be independent from its neighbours. 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 neighbours. For this reason, the LNL model cannot be used directly as a front end to Najemnik and Geisler's ideal search strategy. The same conclusion applies to any image processing model that generates an activation map (such as those of Itti and Koch [2000], 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.

6.7.3 Conclusions

I have shown that a model based on an LNL filter bank can successfully model human performance, in terms of the number of fixations required to find the target, 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). Two different classes of surfaces were used as stimuli and the model gave a good account of human behaviour over a range of surface roughnesses, regularities and target orientations. The aim was to determine whether the information extracted from stimuli by the model is sufficient to account for human search, and, to this end, the search strategy was modelled as a simple stochastic process constrained by inhibition of return. However, the model does not scan the stimulus in the same was as the human observers. The following chapter investigates human search strategies on these homogeneous surfaces in greater detail.

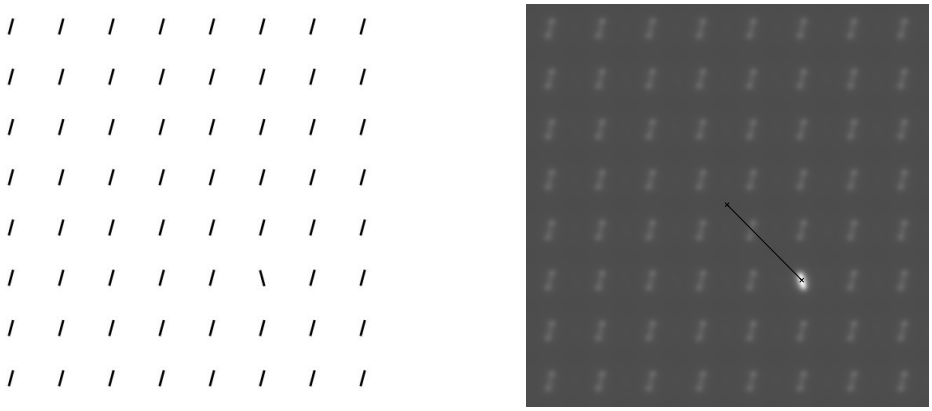

Figure 6.12: Performance of the LNL-based search model on a discrete item search. Left: array of search items. Right: Activation map. As can be seen, the target produces a large response in the activation map and the model's first saccade is directed towards the target.

Chapter 7

Stochastic Search Strategies

7.1 Introduction

In the previous chapter a LNL-based search model was shown to provide a good prediction of the difficulty of finding a defect in a rough surface. The model was tested over a wide range of perceived surface roughnesses and task difficulties and made a similar number of saccades to human observers in all cases. However, the model did a poor job of accounting for human scan-paths and search strategies and only predicted human fixation locations at chance levels. This suggests that signal-to-noise ratio in the activation map generated by the LNL-based model is a good model for human performance, but the local maxima in the activation map do not provide a good predication of fixation locations.

In this chapter I will explore search strategies and how much of a role visual memory has in determining search performance. The first experiment (Section 7.3) will investigate memory using a moving target paradigm [Horowitz and Wolfe, 1998]. This is followed by a comparison between human performance and a stochastic search simulation. Unlike the previous model which was concerned with feature extraction, the stochastic search simulation only attempts to model search strategy and saccade choice. The model is outlined in Section 7.4 and uses the results of a signal detection experiment (Section 7.5) for the target detection model. The model is compared with human observers in Section 7.6.

7.2 Literature Review

A complete computational visual search model contains two parts: a feature extraction mechanism and a search strategy. In the previous chapter we explored the feature extraction stage and modelled it using an LNL algorithm. While this model provided a good fit with human data in terms of task difficulty (the number of fixations required to find the target), it does a poor job of making human-like scan paths.

The search strategy part of a model typically uses an activation map to generate successive saccades. While a number of different mechanisms for this have been put forward, the most commonly implemented is the MAP Observer [Najemnik and Geisler, 2005]. This strategy directs saccades to the local maxima of the activation map, and a simple inhibition of return mechanism is used to stop the model returning to previously fixated maxima. As most previous computational models have primarily been interested in the feature extraction stage of search, the MAP (maximum a posteriori) strategy has often been assumed for simplicity [Clarke et al., 2009, Itti and Koch, 2000, Pomplun et al., 2003, Rao et al., 2002, Rutishauser and Koch, 2007, Wolfe, 2007]. Memory is usually only implemented as an inhibition of return process, and is frequently assumed to be perfect [Pomplun et al., 2003] in which case the model never makes re-fixations. Wolfe [2007] suggests a more sophisticated model for IOR but it is still tied to individual, discrete search items.

7.2.1 Systematic Search

Many models assume that search is systematic, or guided, in some sense. For example, the Area Activation Model [Pomplun et al., 2003] visits local maxima in its activation map using the rule 'make a saccade to the nearest maxima that has not been visited yet'. While the guided search hypothesis has been shown to be valid for many kinds of search, due to the homogeneous nature of the stimuli considered in this thesis it is unlikely to have as strong a role to play. (See Section 5.3.1 for a discussion of guided search.)

Several general tendencies have been taken to be indicative of systematic search strategies. For example, Gilchrist and Harvey [2006] argue that the presence of horizontal bias in saccade directions indicates a systematic component in visual

search. They suggest that systematic tendencies can be hard to detect in scan-paths because of the interaction with salience-based object selection. 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 the details of the precise nature of the random walk and the following comparison were not included in the paper. Furthermore, it is possible that Aks' result is an artifact of studying the compound time-series of large number of visual searches, one after another. It has been shown that a coarse-to-fine dynamic is often present in saccade patterns during search [Over et al., 2007]. If we were to look at the saccade amplitude time-series of a continuous sequence of individual searches, each with its own 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.

An alternative to the MAP searcher is the Bayesian Ideal Observer. Najemnik and Geisler [2005, 2008] have derived the ideal observer for search for a small Gabor patch, at one of 85 pre-determined locations in $1/f$-noise. Unlike the MAP Observer which makes a saccade to where it calculates that the target is currently most likely to be, the Ideal Observer makes saccades in order to maximise its chance of being able to find the target with the next fixation. Najemnik and Geisler conducted a signal detection experiment and derived visibility maps from the empirical results. These visibility maps were used to determine the probability of identifying the target for given signal-to-noise levels and the distance from the current fixation location to the target. These were then used to construct the Bayesian Ideal Observer for a search with eye movements. The model was compared to human observers and was found to offer a good match in terms of the number of saccades to find the target, saccade direction and fixation distribution. However, the Ideal Observer makes a number of assumptions, such as the potential target locations being independent, and it cannot be applied to search stimuli in the same way as the computational search models discussed above [Hwang et al., 2009, Zelinsky, 2008]. Najemnik and Geisler describe it as:

> ...complementary to existing computational models of visual search. Unlike these previous approaches, it is not a heuristic model that can be applied to arbitrary stimuli but a formal, parameter-free analysis for a particular class of naturalistic stimuli. The ideal observer is not meant to be a plausible model of human visual search...

7.2.2 Memory in Search

A related issue to the question of guidance is that of the importance of memory in visual search. Do observers remember the regions of the stimuli which they have already searched? Do they remember what, and where they have previously fixated? Again, as the stimuli under consideration here are homogeneous there are no salient visual landmarks which can be used as a reference point by memory. Of course, the boundaries of the stimuli can still be used as reference points.

Search can be systematic without relying on visual memory. For example, an observer might decide to search for the target in a left-right, top-bottom manner. In doing so, assuming a perfect target detection mechanism, the observer will visit each image location at most once, without needing any memory of where she has already searched.

Separating the effects of a search strategy from memory is difficult. It is further complicated by the fact that the human visual system is far from perfect and can quite easily miss salient visual cues. For example, most of us have learnt from experience that it is worth double checking places you have previously searched when searching for a missing item. Eye-tracking data does not distinguish between saccades made to previously visited areas that we remember searching earlier, and those that we do not remember searching.

In the literature it appears that visual memory only has a weak role to play in visual search [Boot et al., 2004, Horowitz and Wolfe, 1998, 2001, 2003, Kunar et al., 2008, Wolfe et al., 2000]. This is not to say that we do not have a memory of where we have previously searched, but that we do not utilise that knowledge to search more efficiently. Much of the recent research on memory in search stems from a paper by Horowitz and Wolfe [1998] which suggested that visual search appears to be amnesic. They carried out a visual search experiment using discrete search items which were randomly relocated every 111ms. Perhaps surprisingly, they found no difference in search efficiency (RT v set size slopes) between this task and (normal) static search. Although reaction times were slightly longer in the dynamic case this was ascribed to decreased observer confidence. This conclusion was further supported by Gilchrist and Harvey [2000] who analysed scan-paths and compared the number of re-fixations in the human data to the number predicted by simple *sampling with,* and *sampling without replacement* models. They found that except

for a short lived inhibition of return effect, there was little evidence of a memory process.

McCarley et al. [2003] have found similar results using a gaze-contingent display. In their search task observers were forced to choose between making a saccade to either of two search items: one which had been previously inspected, and one that had not. They found that while the re-fixation rate was tied to the number of fixations since the item had last been fixated, there was only an effect for the first 4 search items after which the re-fixation rate was close to chance levels. Another study by Horowitz and Wolfe [2001] employed a multiple-target search paradigm (using normal, static displays). They found that a memory-free model gave a better explanation of the data than a memory-driven model.

Wolfe et al. [2000] investigated 'post-attentive vision' by carrying out a visual search experiment in which the same search array was used from one trial to the next. For each trial the observers were asked if a given letter was absent or present. The results showed that search was no more efficient in the repeated stimulus search than when a different stimulus was used for each trial. A similar experiment was carried out by Körner and Gilchrist [2007] who asked observers to search for two successive targets in the same display. They found that reaction times were shorter in the second task, but only if the second target had been fixated recently during the first search. This was further investigated by Kunar et al. [2008] who tried to find out why observers do not use their memory of the search display to guide search and if they can be encouraged to do just this under certain conditions. They found that, although search performance was relatively inefficient, it was more efficient to carry it out visually than to do a memory recall search. However, if only a subset of the search items are ever relevant, observers can learn this and use it to guide their search. However, search within this subgroup is still inefficient.

Horowitz and Wolfe's [1998] conclusion was challenged by Kristjansson [2000] who repeated their experiment, but instead of randomly relocating all search items, the location of the target was swapped with that of a randomly determined distracter every 110ms. Kristjansson found that under this condition, there was a significant difference between the dynamic and static conditions. He also carried out a second experiment using the same random relocation paradigm as Horowitz but with a larger range of set sizes: as set size increases, the proportion of search items that are randomly relocated to a location that was previously occupied by an item increases. The results showed that as the number of search items increased search became less

efficient in the dynamic case, while staying the same for the static trials. Kristjansson concludes that memory process in search are tied to location. This interpretation is supported by Beck et al. [2006b] who carried out a similar experiment using a gaze-contingent display and found that changing the features of the distarcter items had little effect on search, but changing their location did. A review by Shore and Klein [2000] also supports the hypothesis that visual search makes use of several memory processes: perceptual learning across blocks of trials, trial-to-trial priming, and within trial tagging (IOR). They argue that Horowitz and Wolfe's conclusion, that search is anemsiac, depends on two assumptions: that the same search strategy is used in both the dynamic and static case and that search efficiency is equivalent in the two conditions.

7.2.3 Stochastic Search

Perhaps surprisingly, there does not appear to be a lot of literature on stochastic search processes. Morawski et al. [1980] and Arani et al. [1984] derived a series of models with varying degrees of systematicness and randomness. However, no visual search experiments were carried out and the models were not fitted to human behaviour. (Also see Melloy et al. [2006] for a similar study.) More recently Motter and Holsapple [2001] used saccade distributions to calculate the probability that an observer fixates on a target by chance. While this probability decreases with the number of discrete search items, it continues to account for a sizeable fraction of search performance. Greene [2008] used a random walk to investigate the *distance to target dynamics* reported by Tseng and Li [2004] (who suggested that scan-paths could be split into an ineffective and effective stage, where the effective stage was characterised by a monotonically decreasing distance to target). Greene's results showed that an unguided random walk exhibited the same behaviour.

7.2.4 Conclusions

The above literature review suggests that the evidence for systematic search in homogeneous stimuli is mixed. Memory for previously fixated locations does not appear to play a strong role in search strategies. Evidence for systematic search patterns is limited to statistical regularities, such as the horizontal bias in saccade

direction [Gilchrist and Harvey, 2006] and the coarse-to-fine dynamic in saccade amplitude [Over et al., 2007].

These ideas are explored in the rest of this chapter. First, in Section 7.3, the moving target paradigm is used with the $1/f^{\beta}$ surfaces. This is followed by a development of a *Stochastic Search Simulation* which is compared against human search strategies. This model is based on empirical saccade distributions and hence incorporates the horizontal bias and coarse-to-fine dynamics described above.

7.3 Experiment 8: Moving Target

The aim of this experiment is to use the defect detection task developed here to test Horowitz and Wolfe's conclusion that visual search has no, or little, memory with the defect detection task considered here [Horowitz and Wolfe, 1998, 2001, 2003]. If this holds true then we would expect that frequently changing the location of the target would have little effect on search times.

7.3.1 Methods

Six participants took part and they were asked to locate a target as quickly and accurately as possible. No maximum time limit was imposed on the task. The experiment consisted of 225 trials, split into three blocks of 75. Three different surface roughnesses were used, $\beta \in \{1./59, 1.62, 1.65\}$, $\sigma_{RMS} = 1.1$. The Matlab Psychophysical Toolbox [Brainard, 1997, Pelli, 1997] was used to display stimuli as the software for the Tobii eye-tracker (Clearview and Tobii Studio) is not currently capable of running more sophisticated experiments. This means that the eye-tracker was not used in this experiment.

The target was located randomly on a lattice with spacing 0.84°, although not within 1.88° of the stimulus edge. Additionally, the target was not allowed to be at one of the central nine locations at the start of a trial. Figure 7.1 shows all possible target locations.

There were two *dynamic flash* (moving target) conditions, $s \in \{0.75, 1.25\}$, where s denotes the display time for each target location. After the image had been

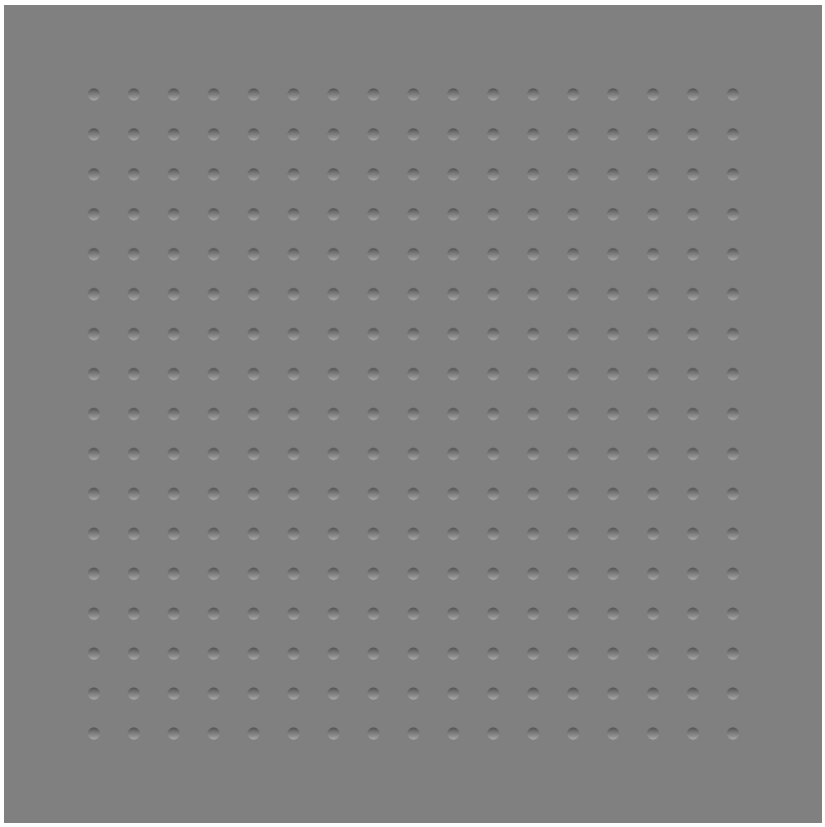

Figure 7.1: All the potential target locations for Experiment 8. Note: The target was not allowed to start located in one of the nine central locations, but it could however move there during a trial.

displayed for s seconds, there was a brief mask (100ms) and then a new image was shown, with the target at a new, randomly determined location on the same background.

For each value of s there was a set of control trials. These were identical to the dynamic trials except that the target did not move after each flash. There are referred to as *static flash*. Finally, there was a *static control* case in which the target did not move, and there were no flashes.

Once the observers had pressed a button to indicate that they had found the target and the reaction time was saved by the computer. A static image was then shown with two targets - one at the current target location and one at the previous target location. The observers were required to mouse click on the target to confirm they had indeed found it. Two targets were displayed to get around the problem of observers identifying the target just before it moved, and responding with a key press just after it had moved.

7.3.2 Results

The mean and median reaction times for each of the six observers are shown in Figure 7.2. There does not appear to be any consistent difference between the control, static and dynamic conditions. This is similarly true for the inter-subject mean and median (see Figure 7.3).

7.3.3 Discussion

The results show that moving the target every $\approx$ 1000ms does not increase the difficulty of the search. This suggests that the search process is not led by a memory-based process. If it were then we would expect observers to find the static targets more easily than the dynamic ones which can move to a location that has already been searched.

However, it seems more likely that there are some problems with the design of the experiment. Sudden stimulus onsets are known to be salient [Theeuwes, 2004, Tsotsos et al., 1995, Yantis and Jonides, 1984] and the masking flash may have encouraged the observers to change their search strategy. Rather than searching

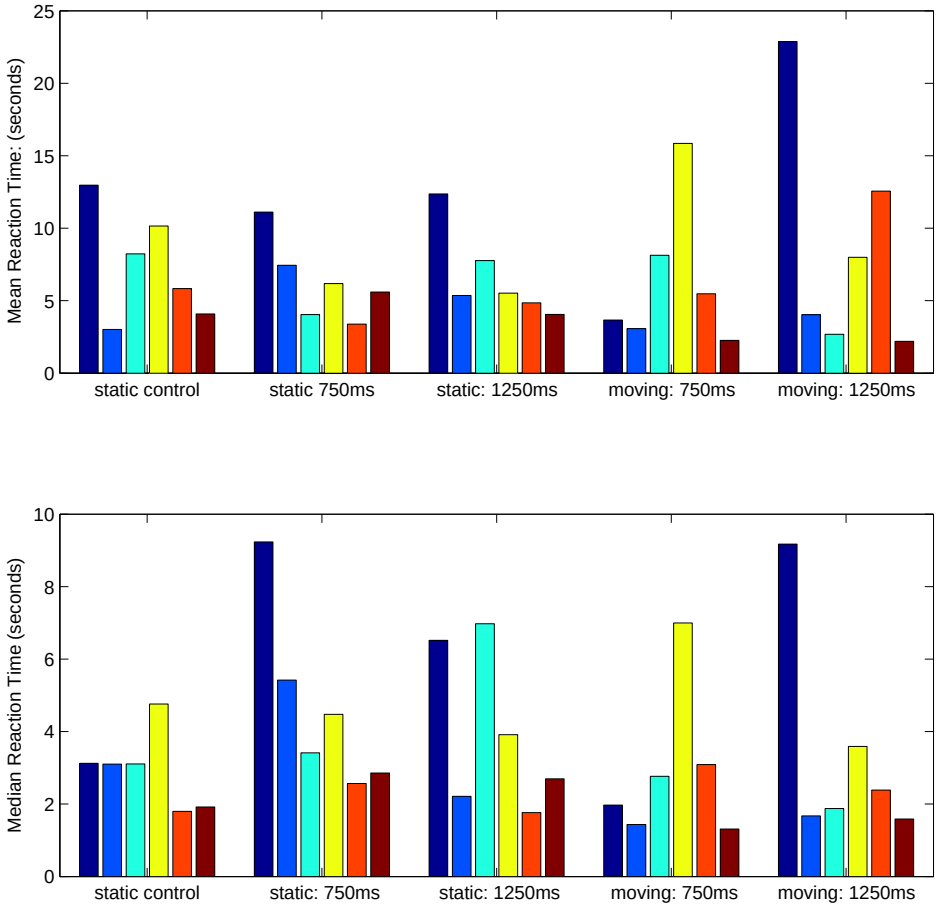

Figure 7.2: Mean and median reaction times for each of the five individual observers in Experiment 8. As can be seen, there are large inter-personal differences.

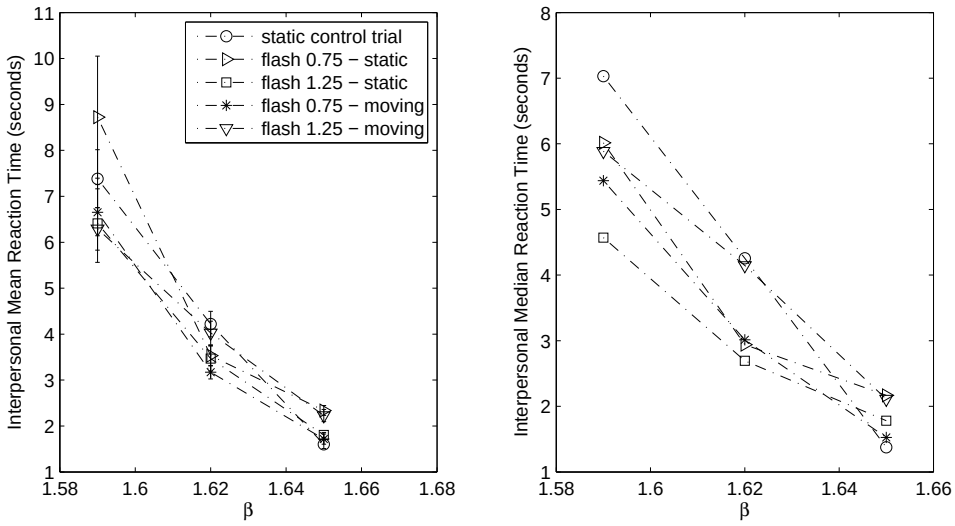

Figure 7.3: Inter-personal mean and median across all six observers in Experiment 8. There do not appear to be any differences between the five experimental conditions.

around the stimuli, observers might instead keep their attention directed towards the centre of the stimuli and wait for the target to appear in their field of view [Shore and Klein, 2000]. Unfortunately, as it was not possible to use an eye-tracker in this experiment (due to software limitations and the lack of support for the Tobii eyetracker in the Matlab Psychophysical Toolbox), this hypothesis cannot be investigated further.

7.4 Stochastic Scan-Path Simulation

While the literature reviewed above (Section 7.2.2) is mixed, the overall picture, together with the above experiment (Section 7.3) suggests that memory plays, at best, a limited role in visual search. As there are no discrete search items, the locational memory processes described by Kristjansson [2000] and Beck et al. [2006a] have no search items to be based on: there are not previous items to remember having visited, so observers only have a memory of 'where' has been previously searched to guide them. Similarly, while the LNL-model in the previous Chapter took the same number of saccades to find the target, the results from 6.6.4 suggest that human fixations are not correlated with maxima in the activation map.

If human search strategies are not heavily dependant on memory processes, or guided by image statistics, then perhaps a random walk will provide a good expla-

nation of human performance and scan-paths. To explore this further, the scan-path data collected in Chapter 6 will be further analysed and compared with a stochastic process.

As in Chapter 6, simplicity will be a guiding principle for the design of this model. The simulation will use a target detection model based on the results from a signal-detection experiment (see Section 7.5, below). This will provide us with a model for the probability of an observer detecting a defect for a given surface roughness and target eccentricity.

If the target is not detected in a given fixation, a saccade will be chosen from the empirical data collected in Experiment 5, Section 6.5.1. This is done as it is the simplest way to to make the stochastic simulation make human-like saccade distributions. Separate distributions are used for different fixation numbers and location. The aim is to investigate if a model which does not incorporate any explicit memory of previously fixated regions can mimic human search behaviour. The resulting, randomised scan-paths will then be analysed in terms of performance (number of saccades required to find the target), hotspot maps (how well distributed are the fixation locations), re-fixations (is there evidence for inhibition of return in surface search?), and Voronoi cells (how well distributed are the fixations on each trial, and how efficiently is the search conducted with time?)

7.4.1 Scope of the Search Simulation

Unlike the LNL-based model, the stochastic search simulation is not based on image processing methods and is not given the image as an input. Instead, it is simply given $N = 1024$ pixels (the size of the search area), the roughness of the surface, β (used to determine the likelihood of detecting the target), and the target's location (chosen randomly for a given eccentricity r). The initial fixation is set to the centre of the search area.

7.4.2 Target Detection

On each fixation the probability that the model detects the target is given by $p = f(\beta, r)$ where r is the distance from the current fixation to the target and β governs how rough the surface is. f is a linear regression model (Equation 7.1 below) based

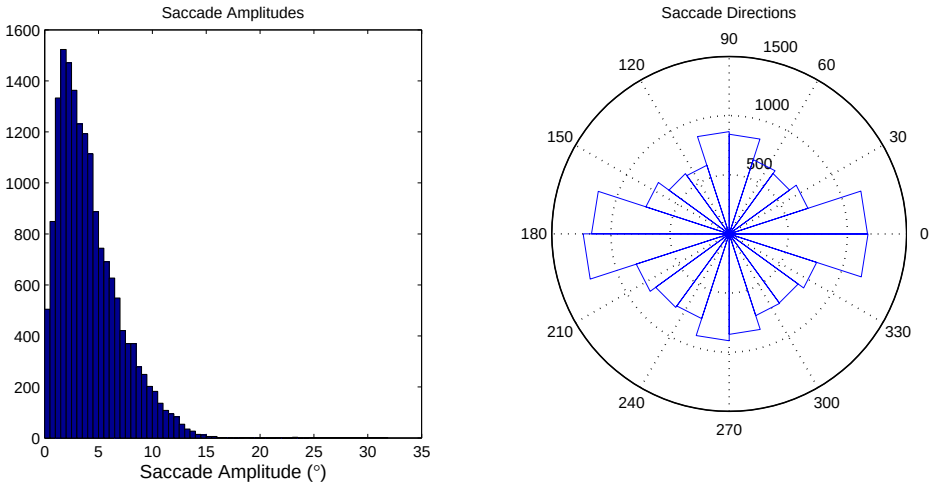

Figure 7.4: Saccade distributions from Experiment 5. (Left) A histogram showing the distribution of saccade amplitudes, over all observers and trials. (Right) A rose plot of saccade directions.

on the results from the Signal Detection experiment detailed below (Section 7.5). For each fixation a random number $x \in [0, 1]$ is generated. If $x \leq p$ then the simulation detects the target, makes a saccade to the target's location, and the search is terminated. If the simulation does not detect the target (i.e. $x > p$) then a random saccade is made to a new location.

7.4.3 Generating Saccades

Version One

Empirical distributions of saccade directions and amplitudes obtained from Experiment 5, Section 6.5, will be used to generate human-like scan paths. These are shown in Figure 7.4 (see Section 6.5.1 for details of the experiment). The relationship between saccade amplitude and direction is shown in Figure 7.5. As saccades are not evenly distributed, in terms of direction, the data for this figure has been normalised in terms of saccade direction. As can be seen, there does not appear to be a strong dependence between saccade amplitude and direction and hence version one of the stochastic search simulation treated the two as independent distributions.

Figure 7.6 shows how mean saccade amplitude decreases with fixation number. This coarse-to-fine pattern agrees with previous findings [Over et al., 2007] and was incorporated into the first version of the stochastic search model: for each saccade

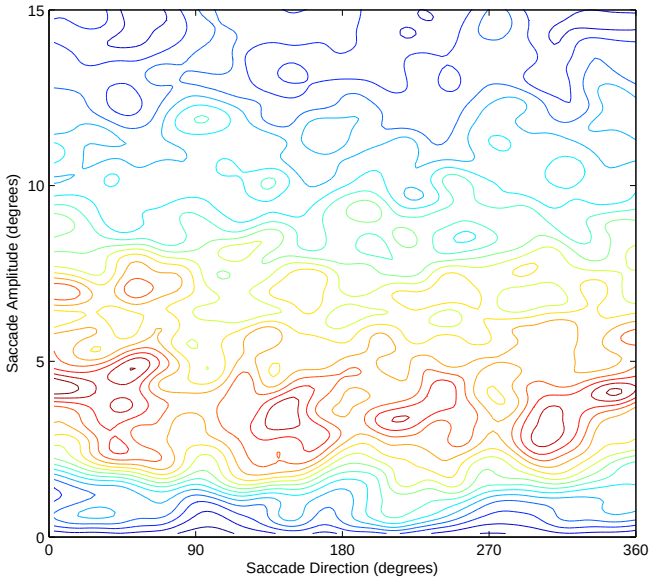

Figure 7.5: Contour plot showing saccade direction against amplitude. As observers exhibit a preference for some saccade directions over others, the data has been normalised by saccade direction.

number t, $1 \leq t \leq 50$ an amplitude from the distribution of saccade amplitudes made by human observers on the tth saccade of a search trial was selected. If the simulation needs to make more than 50 saccades to find the target, it draws amplitudes from the distribution for $t = 50$. If the model chose a saccade that would take it outside the search area, then it simply picked another saccade, until it chose one that would keep it within the search boundaries.

While this simulation initially look promising it occasionally performed poorly. In particular, the simulation would sometimes 'get stuck in a corner or an edge.' An example is given in Figure 7.7.

This appears to be caused by the simulation's overly-simplistic behaviour at the boundary of the search area. The closer the current fixation location is to the search boundary, the more likely it is that a potential saccade could take it outwith the search area. In particular, long saccades are more likely to pass over the boundary than short saccades. This means that the closer the model is to an edge, the more likely it is to make a short saccade, which means the following fixation will also be close to the edge. This behaviour does not appear to be present in the human data, and will be investigated below.

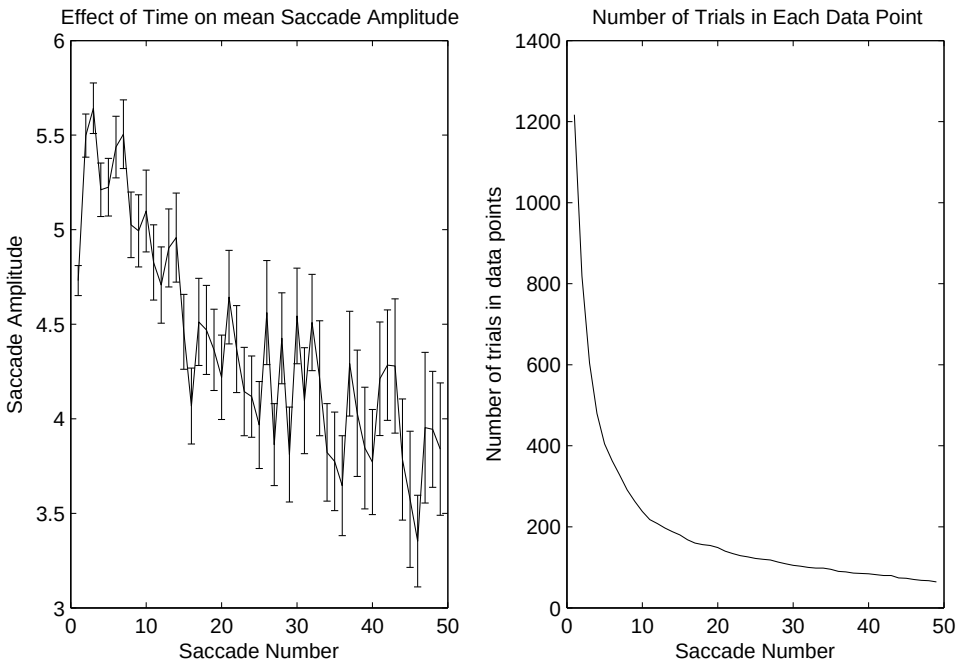

Figure 7.6: (Left) As the search progresses the mean of the human saccade amplitudes decreases. (Right) The number of saccades involved in each data point in the amplitude graph.

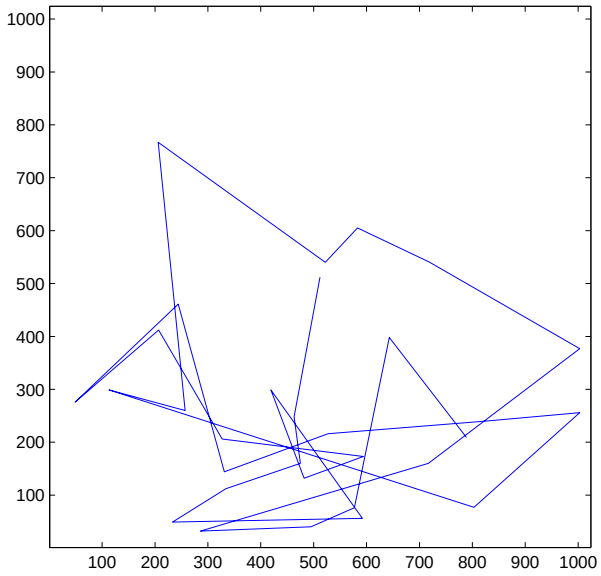

Figure 7.7: A typical example of version 1 of the Stochastic Search Simulation. As can be seen, the fixation locations are not distributed very evenly.

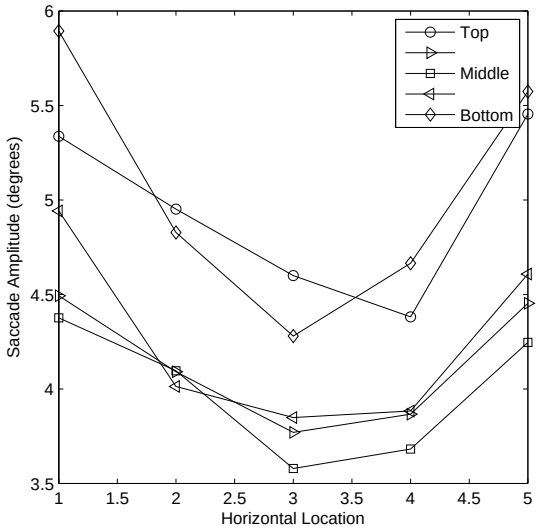

Figure 7.8: The effect of fixation location on mean saccade amplitude. As can be seen, observers make longer saccades when they are fixating near the edge of the stimulus.

Version Two

The second version of the stochastic search simulation takes the location of the current fixation into account by using different saccade amplitude and direction distributions at the edges and centre of the display. The search area was divided into 5×5 equally sized sub-regions. Figure 7.8 clearly shows that the horizontal fixation location has an effect on saccade amplitude (a similar result was obtained for the vertical location). In particular, the shortest saccades are made away from fixations located in the centre of the search stimuli while the longest are made away from fixations in the corners of the display. This is likely to be at least partly down to the fact that when we are fixating in the middle of the search area, the longest saccade we can make, while staying inside the stimulus boundary is just under 12°. However, if we are fixating in a corner then it is possible to make a saccade through 22° of visual angle and still be within the boundary. As explained above however, Version 1 of the stochastic search model actually behaved in the opposite way.

Figures 7.9 and give 7.10 more detail of how the saccade amplitudes depend on both time and fixation location. However, now that we are dividing the data into a large number of small subsets, it is become somewhat sparse and hence noisy. To get round this, some of the subsets will be merged together. In particular, the corner distributions will be merged together, after the equivalent reflections in the

horizontal and vertical axis. Likewise, all the edge regions will be merged to give a distribution of edges from horizontal edges and vertical edges. Finally, the middle nine subregions will be merged. Additionally, the data for fixation number will be binned for $1 \leq t \leq 5$, $5 < t \leq 10$, $10 < t \leq 15$ and $t > 15$. Contour plots of the final distributions used in the simulation are shown in Figure 7.11.

7.5 Experiment 9: 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 a visibility map which will give a simple model for the probability of target detection at different eccentricities and surface roughnesses.

The experimental task is a *target present/absent forced choice*. This differs from the task used by Najemnik and Geisler [2008], which was a 2IFC (two interval forced choice) blocked for spatial position. This meant that observers knew in advance which location the target would appear in, and had to determine whether the first or second stimulus of a pair contained it.

There are advantages and disadvantages of both methods. The target absent or present approach allows us to determine how well observers can identify the target when *they do not know in which location it will appear*, or even whether it is present. However, this causes a problem with false hits as there is no way of knowing what the observer incorrectly judged to be the target. This problem is solved using the 2IFC task, as the observer has to make a judgement about a specific location on the stimulus. However, for target present trials we are no longer measuring how well an observer can identify the target for different roughnesses and eccentricities, but *how well can an observer detect the target if they know where it will appear*. As discussed in Section 2.2.4 it is known that attention can be deployed away from fixation and in doing so increases our ability to discriminate the target in the visual field.

7.5.1 Methods

For the target present trials, the target was located at one of 72 potential locations: nine different eccentricities were used $0.84° \leq r \leq 7.5°$, and eight evenly

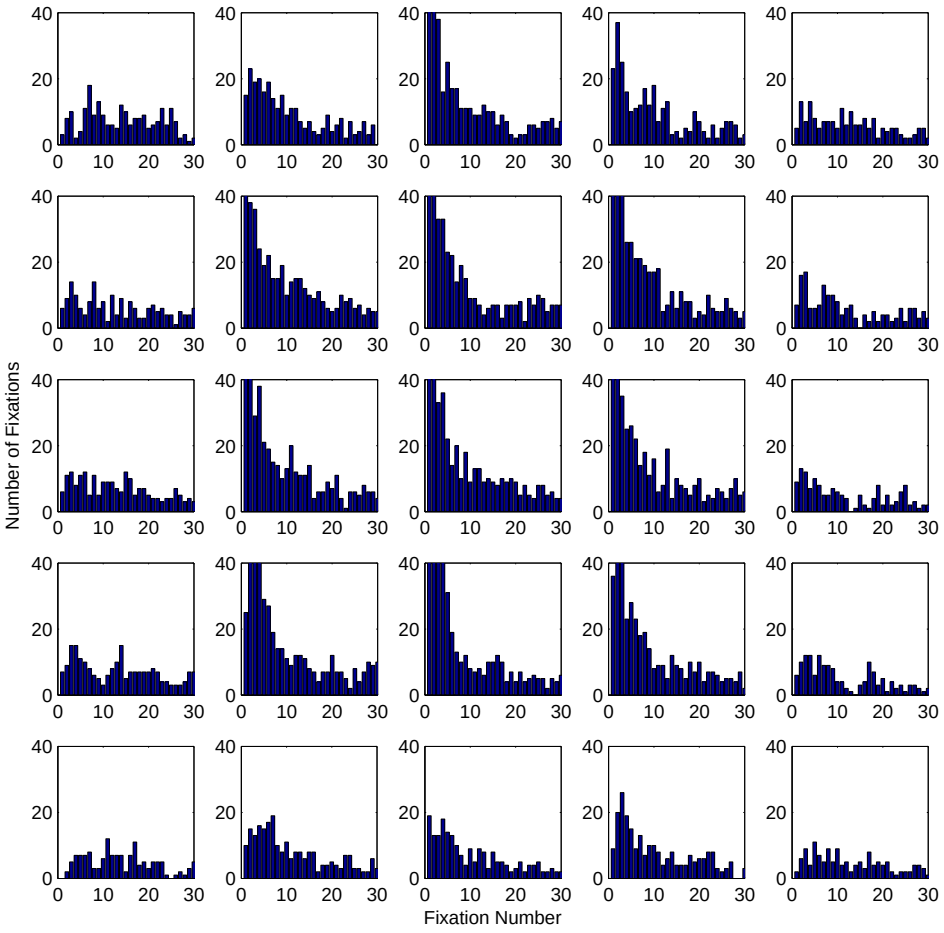

Figure 7.9: Each of the 5×5 subplots shows the number of fixations made in the corresponding subregion of the stimuli. The x-axis is the ordinal fixation number, while the y-axis shows how many saccades were made, over all trials and all observers. We can see that most saccades originate from the central subregions. Note: the y-axis has been truncated at 40 fixations to improve the comparison between histograms.

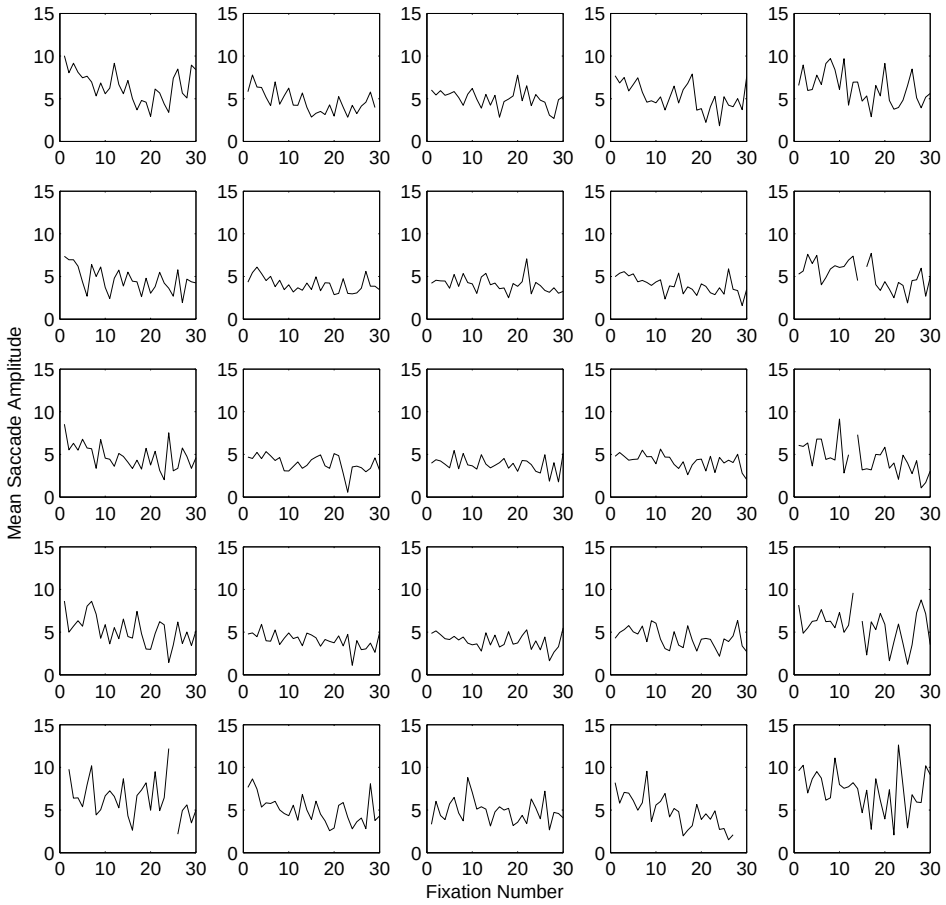

Figure 7.10: Each of the 5×5 subplots shows how the amplitude of the saccades made from the corresponding stimulus region changes with fixation number.

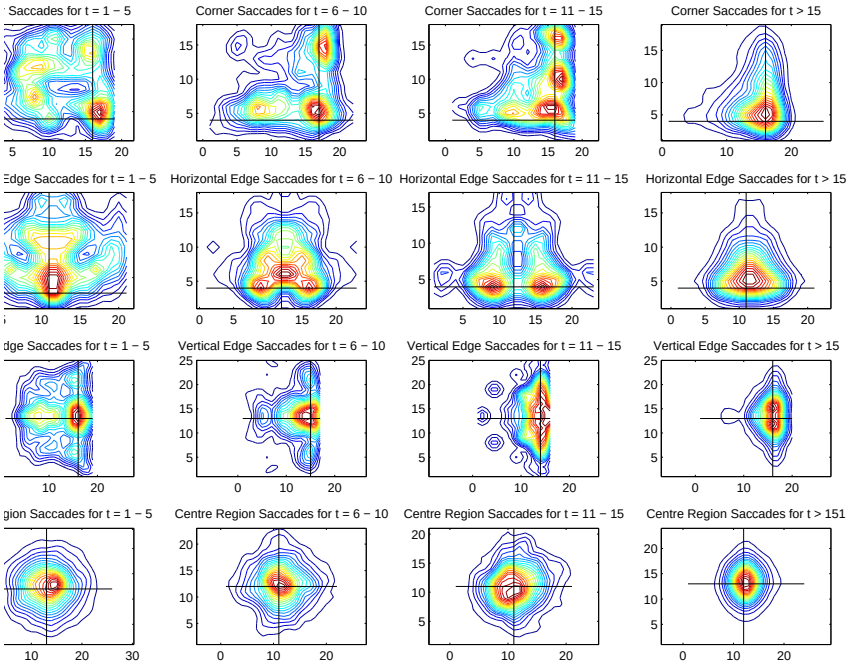

Figure 7.11: Human saccade statistics by position and time. Separate distributions are given for the corners (Top Row); horizontal edges,vertical edges, and the centre region (Bottom Row) of the stimuli. Subplots along each row show how the distributions change as more saccades are made.

spaced orientations. The target was made by subtracting an ellipsoid from the three dimensional surface and subtended 0.66° of visual angle. Surface roughness was controlled by $\beta \in \{1.6, 1.65, 1.7\}$.

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.)

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 four 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.

7.5.2 Results

The observers' gaze location was sampled every 20ms and trials were included in the analysis only if they satisfied these conditions:

- The mean distance between the observer's gaze location and the centre of the stimulus was less than 1°
- The x and y gaze components had a standard deviation of less than 0.67°

Trials in which fixation was not held at the centre of the image (14%) were removed. The results for the two individual participants are shown in Figure 7.12. 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 directional data were very noisy and hence all further analysis will assume an isotropic visibility map.

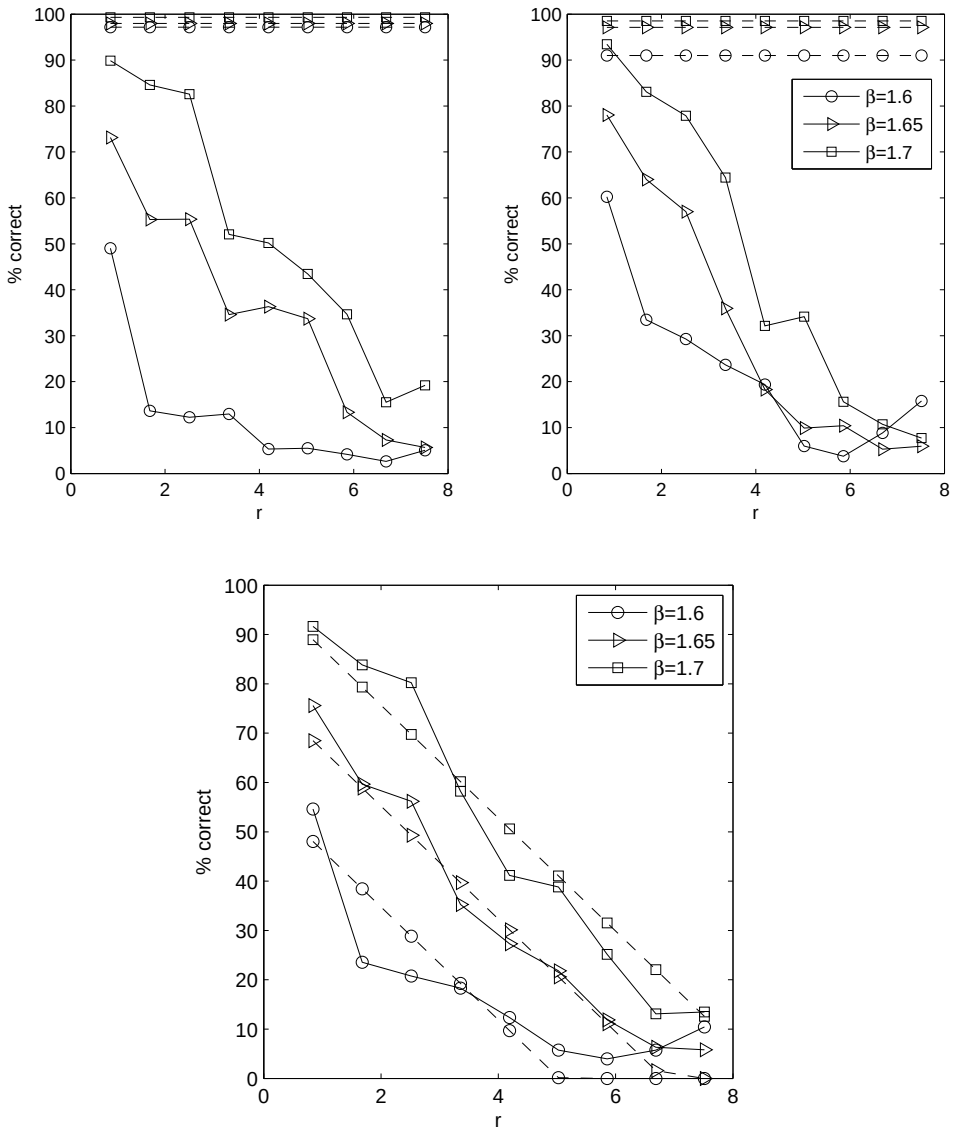

Figure 7.12: Results from Experiment 9. (Top) Individual results for each observer. (Bottom) Mean observer accuracy (solid lines) and the multi-linear regression model (dashed).

The two subjects performed similarly and the mean target present performance is shown in Figure 7.12. This will be approximated by a simple multi-linear regression model:

$$p(T|\beta, r) = 4.09\beta - 0.11r - 5.97 \tag{7.1}$$

This regression model gives $R^2 = 0.934$. Note: due to the obvious thresholding effects, the data points for $\beta = 1.6$, $r = 5.86, 6.69, 7.51$ and $\beta = 1.65$, $r = 7.51$ where not included in the linear regression. Furthermore, $p(T|\beta)$ is set to 0 for these parameter values.

7.5.3 Conclusion

The two observers performed similarly in the task and a linear regression model is a good fit with the results. This linear regression model will be used in the target detection part of the Stochastic Search Simulation, as detailed above in Section 7.4.2.

7.6 Evaluating the Stochastic Search Simulation

Now that I have described a target detection mechanism (depending on r and β), and the saccade distributions (depending on the fixation number, t, and location (x_f, y_f)) we will compare the stochastic search simulation with the scan-path data from the experiment in Section 6.5.1. As the simulation is also based on these data, it follows that it will have similar saccade statistics to those described in Figures 7.4, 7.5 and 7.11. The interesting questions are:

- will the stochastic search simulation take the same number of fixations to find the target as human observers do?
- will the simulation locate its fixations as efficiently as human observers, despite having no concept of memory, including IOR?

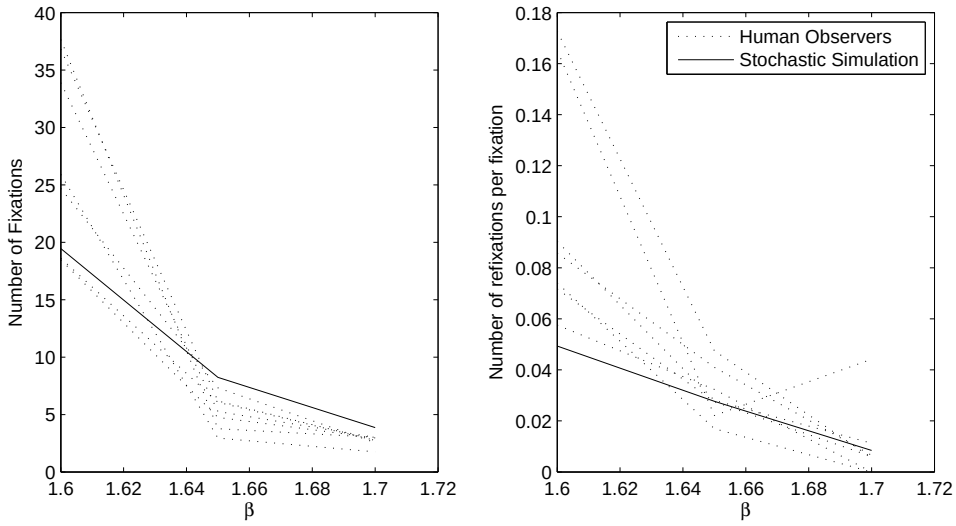

Figure 7.13: (Left) Number of fixations required by the human observers and the stochastic simulation to find the target. (Right) The number of re-fixations per fixation made by the human observers and the stochastic simulation.

7.6.1 Number of Saccades

Considering first the performance of the simulation in terms of numbers of fixations required to find targets, we see that it performs in a similar way to the human observers in the previous chapter (Figures 7.13 (left)). While it finds the target in fewer fixations than the mean human observer, when we compare the individual observers with the simulation in each condition we see that the model is within the range we would expect from a person for $\beta = 1.6$. For $\beta = 1.65$ and $\beta = 1.7$ the model performs slightly worse than human observers. However, if human observers conducted systematic searches the opposite result would be expected: that the simulation's performance would become even poorer relative to human performance as more saccades were required to find the target.

7.6.2 How Systematic are People?

Next, I compare how efficiently human observers and the simulation cover the area of the stimulus during search. If human search has systematic or memory-based processes one would expect the stochastic search simulation to make more re-fixations than the human observers. Indeed, as discussed above, previous studies have shown this is the case in discrete item search [Klein, 2000].

For the current comparison, the concept of a re-fixation is not as well defined: we cannot talk in terms of discrete items, so instead, a re-fixation will be defined as fixating within $r = 1/2^\circ$ of one of the previous n fixations. For each trial we computed the number of re-fixations per fixation with $n = 3$, in order to investigate short term IOR processes. 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 7.13(Right)) show that humans appear to be no more systematic than a stochastic process. (A larger value of n would of indicated a longer term memory process.)

This is not to say that there is no IOR in human search in more general, item-based search tasks. Firstly, as our stimuli contain no search objects, any IOR process would have to be operating spatially rather than being applied to search objects. Secondly, the stochastic search model implicitly contains an IOR component as it draws saccade amplitudes from the empirical distribution. However, it can see from Figure 7.13 human observers do not appear to re-fixate recently fixated regions any more or less than we would expect a stochastic, memory-less process to; if anything, they make more re-fixations than we would expect. Moreover, this difference cannot be attributed to human observers making two successive fixations at the same point, as by design, the model makes these small saccades as frequently as the human observers.

Another way to look for differences between the human observers and the stochastic search simulation is to look at the overall hotspot maps of fixations (Figure 7.14). Here we can see that both hotspot maps are similar, although the stochastic model is slightly more biased towards the centre of the search stimuli, both in the horizontal and vertical directions. Interestingly though, the human hotspot map does not show the same pattern as Najemnik and Geisler [2008]. They found that human observers showed a preference for making fixations slightly above and below the central fixation point. They used this to argue that their Ideal Observer was a better fit than the MAP model.

Finally I will compare how systematic human observers and the stochastic search simulation are by using the Voronoi method, proposed by Over et al. [2006]. If human search has systematic properties, we would expect the proportion of the stimulus area searched (i.e. falling within some criterion distance of a fixation) to increase more rapidly over the course of a human search trial than a random walk made by the model. The Voronoi method allows us to study the uniformity of fixation density and involves computing the bounded Voronoi cells [Voronoi, 1907]

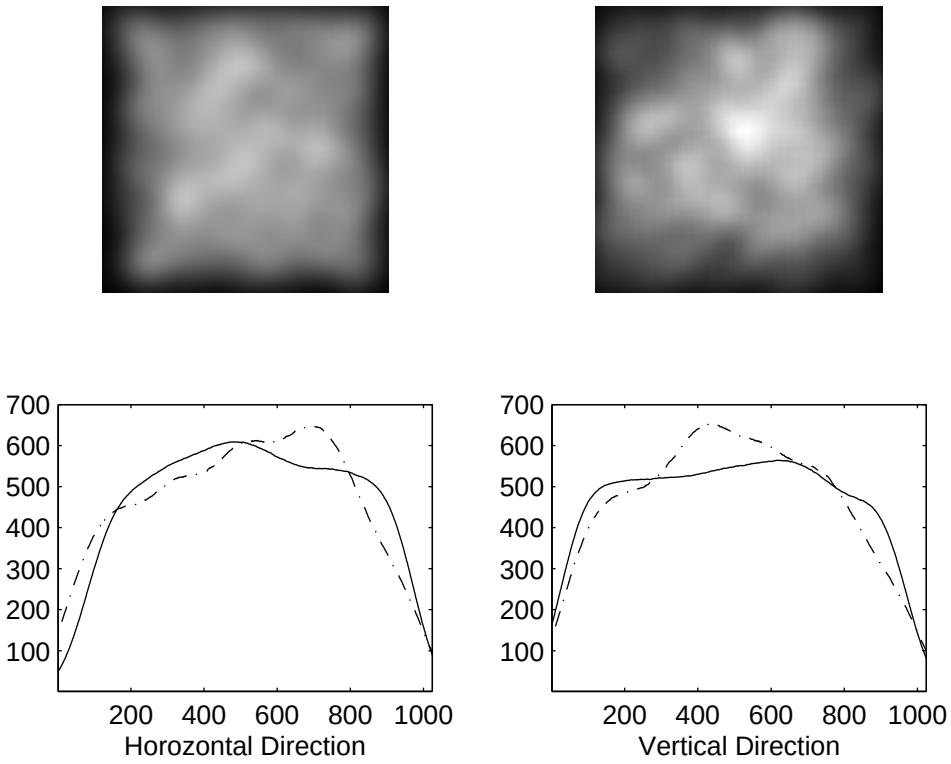

Figure 7.14: (Top) Hotspot maps for human observer (Left) and the stochastic search simulation (Right). Both appear to be well distributed around the search area. (Bottom) Graphs showing how the density of fixations change in the (Left) horizontal and (Right) vertical directions. The solid line shows human fixation density while the dashed lines shows the results from the simulation.

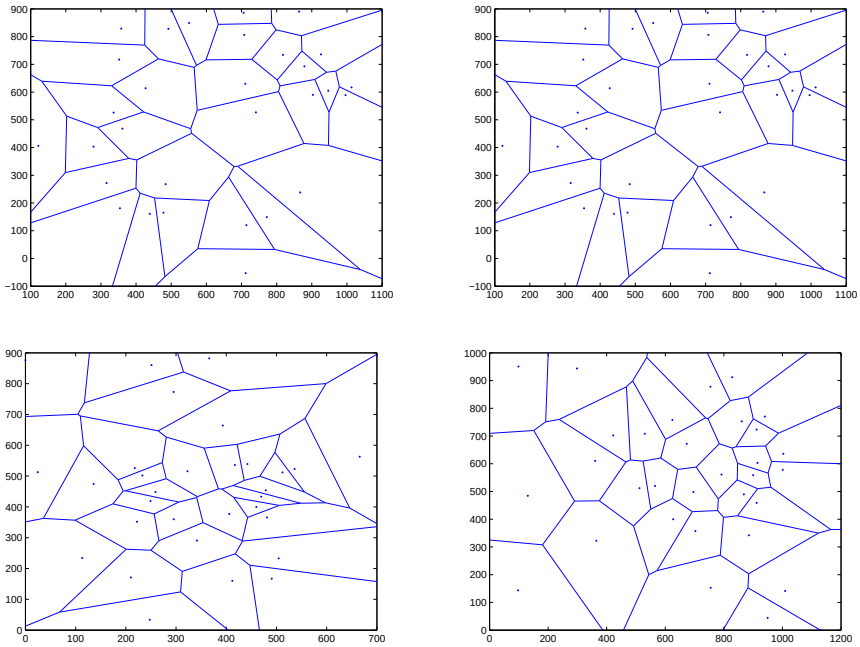

Figure 7.15: (Top) Example scan paths and related Voronoi plots from human observers. (Bottom) From the model.

for a set of fixation coordinates and looking at the distribution of cell areas. Some examples are shown in Figure 7.6.2. In order to compare the stochastic simulation with human observers for each fixation in each trial the Voronoi cells were computed. For each fixation f_t in a trial, the Voronoi cells made by the fixations f_i, $1 \leq i \leq t$ were created and the area (in pixels) of the largest cell was computed. See Figure 7.16 for an example of the first 10 fixations in a trial.

Figure 7.17 (left) shows how the mean maximum Voronoi cell over all trial decreases with successive fixations. As can be seen, the data from the human observers quickly diverges away from the stochastic simulation. This means that the simulation does not direct its attention to unexamined regions of the search area as quickly as the human observers. However, the gap in performance does not appear to get larger over time. To investigate this further, the rate of change was calculated, to give an idea of how quickly the search area is covered. This is shown in Figure 7.17 (Right) and we can see that the difference in maximum Voronoi cell size occurs in the first five fixations. After that, the human observers reduce the size of the largest Voronoi cell at the same rate as a random walk.

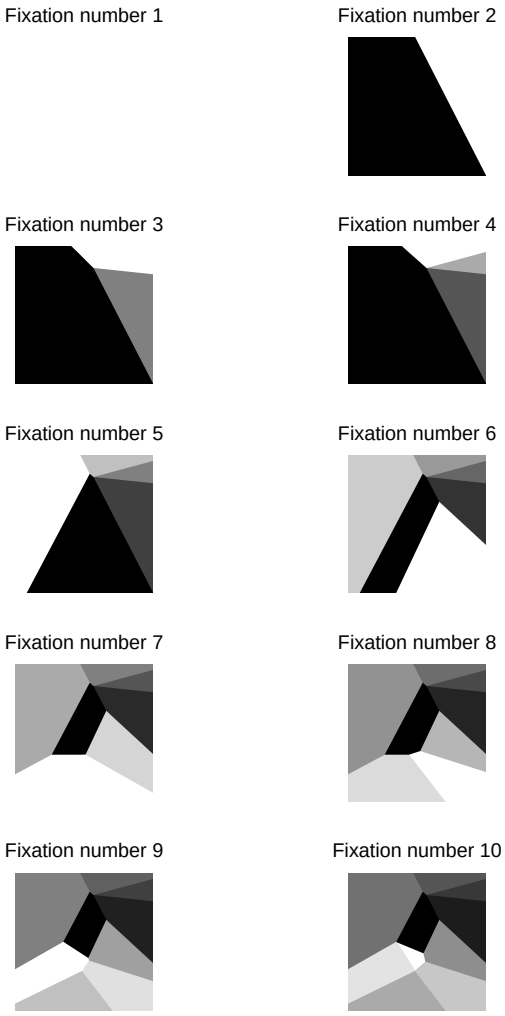

Figure 7.16: Example of Voronoi cells for the first 10 fixations of a trial.

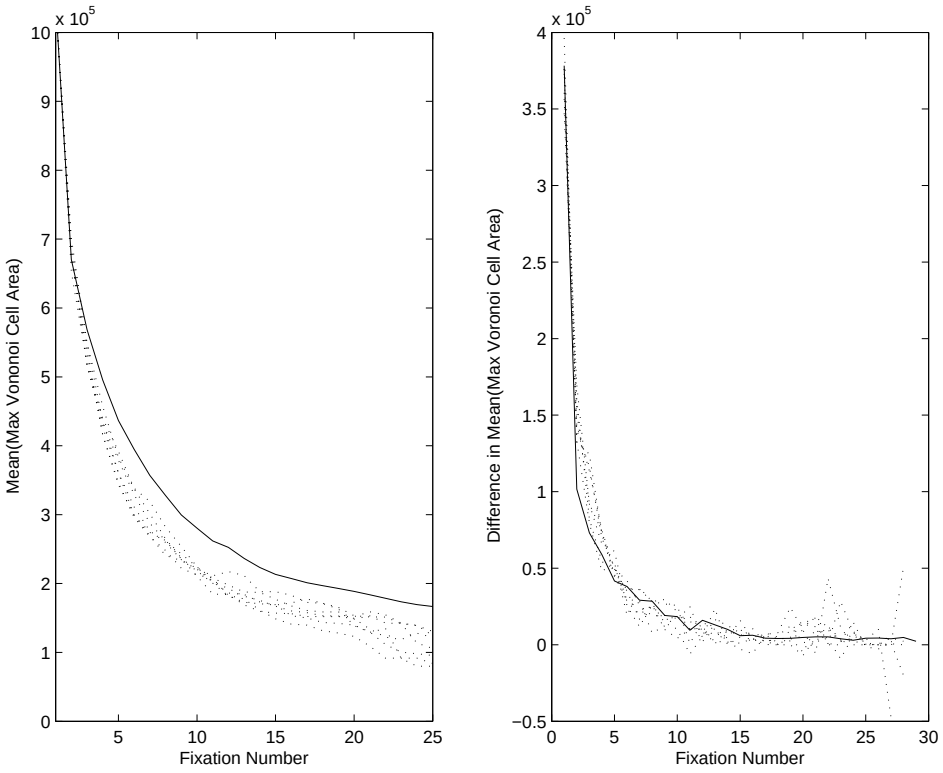

Figure 7.17: (Left) How the maximum Voronoi cell area changes with time. The dotted lines show the seven human observers while the solid line shows the stochastic model. (Right) This shows the derivative of the graph on the left: a measure of how quickly the search area is covered. The main difference between the model and human observers occurs during the first few fixations. After these initial few fixations the human observers appear to be no more systematic than the simulation.

7.7 General Discussion

The results suggest that for visual searches involving locating a target on an otherwise homogeneous surface texture human behaviour can be modelled closely by a stochastic process. The random walk simulation 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 simulation also makes a similar number of re-fixations and, except for the initial few fixations, appears to search as efficiently as a human observer.

However, this is not to say that the results challenge Guided Search: in fact the two models could quite easily work together. In fact, the model presented here *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, then a random walk model could be used to choose which item should be fixated next. Also, although the stochastic search model did not need any form of memory or inhibition of return, this does not mean that there is no inhibition of return in any form of human search. Indeed, as the stimuli used in this thesis contain no search objects, the results may imply that IOR processes cannot operate with respect to spatial coordinates defined with respect to the stimulus boundaries, but only with respect to discrete search objects. However, as can see from Figure 7.13 (Right) human observers do not appear to re-fixate recently fixated regions any more or less than a stochastic, memory-less process would be expected 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 [2005, 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 the experiments presented here: the human observers show no preferences for fixating any particular regions (see Figure 7.14).

There have been a series of studies using Classification Images to investigate guidance in a search task involving targets embedded in $1/f$-noise [Rajashekar et al.,

2002, 2006, Tavassoli et al., 2009]. This involves 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 7.18. This could be because either guidance does not play a significant role with these stimuli or, due to the small size of the target, there is a larger degree of variance in fixation placement with respect to the intended point of interest. This means that even if the observers were directing their attention to regions of the surface that resembled the target, they would be unlikely to be visible in the classification images.

7.7.1 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 chapter 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 I 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 as it is 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.

The 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

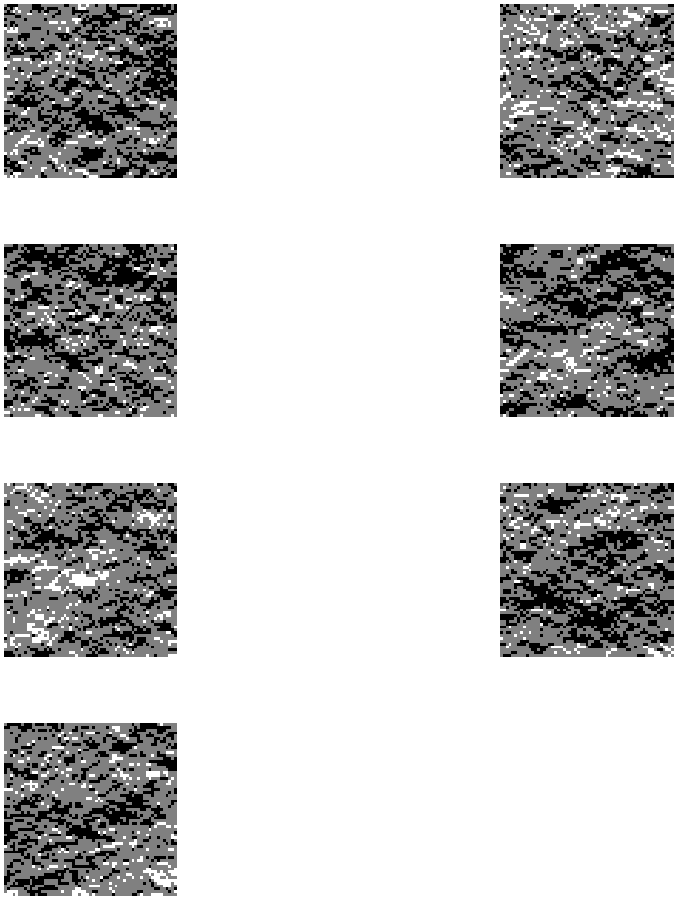

Figure 7.18: Classifications from the experiments performed in Section 6.5.1. Each classification image shows the mean image patch fixated by each individual observer.

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.

Chapter 8

Conclusions

The motivation behind this thesis was to conduct a rigorous investigation into perceptual defect detection. As discussed in Chapter 2, previous work on defect detection algorithms has neglected comparing human and computer performance. Similarly, the problem of finding an anomaly on a homogeneous surface has received very little attention in the field of visual search. The main contribution of this thesis is to bring together relevant work on visual search, saliency, perception and texture discrimination for the purpose of modelling human defect detection.

8.1 Contributions

8.1.1 A Review of Visual Search Literature Relevant to Surface Defect Detection

This thesis brings together relevant work from the fields of computer vision and perception. As stated above, the performance of human observers is rarely considered when designing visual defect detection systems. A general overview of visual search is provided in Section 2.2 while more detailed discussions on visual saliency and computational models of visual search are given in Section 4.1 and Chapter 5 respectively. Finally, Section 7.2 reviews previous work on the role of memory and systematicness in visual search.

8.1.2 Synthetic Surfaces as Visual Search Stimuli

I have introduced rendered textured surfaces as novel stimuli for visual search experiments (Chapter 3). These textures have a number of advantages over more traditional stimuli. Unlike arrays of discrete search items, textured surfaces are naturalistic in appearance. Unlike photographs of natural scenes, these stimuli are created on a computer with controlled parameters. By controlling the seed used by the pseudo-random number generator many different, yet equivalent, textures can be created for use in psychophysical experiments. Furthermore, they do not contain high level, semantic information that is often present in photographs and can have a strong influence on scan-paths.

Chapter 4 is given over to an investigation of how well human observers can find surface anomalies and performance is found to vary systematically with surface roughneess and indent depth and orientation (with respect to the illumination direction). A widely used computational saliency model has been shown to only offer a partial explanation of human performance when searching for a defect on a rough surface (Section 4.4). As discussed in Section 5.3, the nature of these stimuli - continuous, with a category-defined target - means that existing computational models of visual search can not be readily applied to the problem.

8.1.3 An LNL-based model for Visual Search

I have shown that an LNL-based search model can successfully model the perceived difficulty of visual search for a small indent on a $1/f^{\beta}$-noise surface (Chapter 6). This model has been designed with simplicity in mind and hence only uses a single bank of Gabor filters, rather than seperate contrast and orientation filter banks (6.3). With the $1/f^{\beta}$ stimuli the model takes a similar number of saccades to find the target as a human observer, and this holds over a range of task difficulties. In particular, unlike Itti and Koch's sialiency algorithm, the LNL-model can match human performance for a search for an elongated target over a range of orientations (Section 6.5.2). The model also copes well with the *near-regular* surfaces.

8.1.4 Memory and the Stochastic Search Simulation

To investigate the role of memory in visual searches for a target on a homogeneous textured surface a *Stochastic Search Simulation* was designed (Section 7.4). This simulation uses an empirically based linear-regression model for target detection (Section 7.5) and uses the saccade distributions from Experiment 5 (Section 6.5.1). The Stochastic Search Simulation is used to analyse human performance in Section 7.6 and appears to give a good account of human search behaviour in terms of the number of fixations required to find the target; number of re-fixations; hotspot maps and Voronoi cells. This suggests that memory does not play an important role in visual searches for a target on a homogeneous textured surface.

8.2 Future Work

This thesis has provided a rigorous framework for investigating automated surface defect detection and there are numerous ways in which this research could be continued and expanded. For example, it would be interesting to try and combine the stochastic search simulation with an image processing model (which would replace the linear-regression model currently used to model the probability of detecting the target on a given fixation). Unfortunately the current version of the LNL-based model is not suitable for this task: as the target is not always the activation map's global maximum (particularly for rough surfaces) the model cannot distinguish between the target and false positives. Common image processing methods such as matched filters do not appear suitable for this task either as it is unclear how surface roughness would have an affect on task difficulty.

Another important problem, (which has been outside the scope of this thesis) is modelling why - and when - human observers decide that a surface is defect free. This is a challenging problem as it can depend on the interpersonal differences and the rarity of defective trials. However it is a vital part of any real-life defect detection system as most of the samples on a production line will be defect free. Related to this is the issue of texture homogeneity: by how much can different image statistics vary before observers decide that there is an anomaly in the surface? Instead of localised anomalies (defects), texture gradients could be introduced. While texture discrimination has received a lot of attention in the vision sciences, most work has investigated what features facilitate pre-attentive texture discrimination. The

question of modelling attentive discrimination, using tasks as difficult as the visual searches in this thesis, has received far less attention.

There is also a wealth of surface textures that can be used to explore how models, and human observers, cope with different visual stimuli. There are many different texture features and dimensions [Emrith, 2008] and it would be interesting to attempt to model how the saliency of textural anomalies varies with other surface features. Illumination conditions can also be changed and are likely to have an effect on performance.

Publications by the Candidate

Journal Papers

- A. D. F. Clarke and P. R. Green and M. Chantler and K. Emrith, Visual Search for a Target Against a $1/f^{\beta}$ Continuous Textured Background, *Vision Research*, 40:21, pp. 2193-2203, 2008
- A. D. F. Clarke and P. R. Green and M. Chantler, Modelling Visual Search on a Rough Surface, *Journal of Vision*, 9:4(11), pp. 1-12, 2009

Conference Talks

- A. D. F. Clarke and P. R. Green and M. Chantler, Visual Search for a Target Against a Continuous Textured Background, *4th Scottish Perception Meeting*, Stirling, Scotland, 7th December 2007
- A.D.F. Clarke and P.R. Green and M.J. Chantler and K. Emrith, Modelling Visual search for a Target Against a $1/f^{\beta}$ Continuous Textured Background, *European Conference on Visual Perception*, Utrecht, the Netherlands, 24th-28th, August 2008
- A. D. F. Clarke and P. R. Green and M.J.Chantler, Stochastic Search on a Homogeneous Surface Texture, *AVA Easter Meeting and AGM*, 31st March 2009

Bibliography

Jr. Ahumada, A. J. Perceptual classification images from vernier acuity masked by noise. *Perception*, 25, 1996.

D. Aks. $1/f$ dynamic in complex visual search: Evidence for self-organized criticality in human perception. In M. A. Riley and G. C. Van Orden, editors, *Tutorials in comtemporary nonlinear methods for the behavioral sciences*, pages 329–359. NSF (web book: http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp, 2006.

D. Aks and J. Enns. Visual search for size is influenced by a background texture gradient. *Journal of Experimental Psychology: Human Perception and Performance*, 22:1467–1481, 1996.

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.

N. Aleixos, J. Blasco, F. Navarrónc, and E. Moltó. Multispectral inspection of citrus in real-time using machine vision and digital signal processors. *Computers and Electronics in Agriculture*, 33:121–137, 2002.

A. L. Amet, A. Ertüzün, and A. Erçil. Subband domain co-occurrence matrices for texture defect detection. *Image & Vision Computing.*, 2000.

T. Arani, M. H. Karwan, and C. G. Drury. A variable-memory model of visual search. *Human Factors*, 26:631–639, 1984.

G. Backer, B. Mertsching, and M. Bollmann. Data- and model-driven gaze control for an active-vision system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1415–1429, 2001.

R. J. Baddeley and B. W Tatler. High frequency edges (but not contrast) predict where we fixate: a bayesian system identification analysis. *Vision Research*, 46: 2824–2833, 2006.

R. M. Balboa and N. M. Grzywacz. Power spectra and distribution of contrasts of natural images from different habitats. *Vision Research*, 43:2527–2537, 2003.

B. L. Beard and A. J. Ahumada. A technique to extract relevant image features for visual tasks. In *Human Vision and Electronic Imaging III*, volume 3299, pages 79–85. SPIE-International Society for Optical Engineering, 1998.

M. R. Beck, M. S. Peterson, W. R. Boot, M. Vomela, and A. F. Kramer. Explicit memory for rejected distracters during visual search. *Visual Cognition*, 14:150–174, 2006a.

M. R. Beck, M. S. Peterson, and M. Vomela. Memory for where, but not what, is used during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 32:235–250, 2006b.

J. Bergen and B. Julesz. Rapid discrimination of visual patterns. *IEEE Trans. Syst. Man Cybern*, 13:857–863, 1983.

J. R. Bergen and M. S. Landy. Computational modelling of visual texture segregation. In *Computational Models of Visual Processing*. Cambridge, MA: MIT Press, 1991.

I. Biederman. Searching for objects in real-world scenes. *Journal of Experimental Psychology*, 97:22–27, 1973.

I. Biederman, T. W. Blickle, R. C. Teitelbaum, and G. J. Klatsky. Object search in nonscene displays. journal of experimental psychology. *Journal of Experimental Psychology*, 14:456–467, 1988.

A. Bodnarova, M. Bennamoun, and K. Kubik. Suitability analysis of techniques for flaw detection in textiles using texture analysis. *Pattern Analysis and Applications*, 2000.

A. Bodnarova, M. Bennamoun, and S. Latham. Optimal gabor filters for textile flaw detection. *Pattern Recognition*, 35:2973–2991, 2002.

W. R. Boot, J. S. McCarley, and A. F. Kramerand M. S. Peterson. Automatic and intentional memory processes in visual search. *Psychonomic Bulletin & Review*, 5:54–861, 2004.

C. Boukouvalas, J. Kittler, R. Marik, M. Mirmehdi, and M. Petrou. Ceramic tile inspection for colour and structural defects. In *Proceedings of AMPT95*, pages 390–399, 1995.

A. C. Bovik, M. Clark, and W S Geisler. Multi-channel texture analysis using localised spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:55–73, 1990.

D. H. Brainard. The psychophysics toolbox. *Spatial Vision*, 10:433–436, 1997.

J. Braun. Shape-from-shading is independent of visual attention and may be a 'texton'. *Spatial Vision*, 7:311–322, 1993.

M. J. Bravo and H. Farid. The specificity of the search template. *Journal of Vision*, 9:1–9, 2009.

J. R. Brockmole and J. M. Henderson. Using real-world scenes as contextual cues for search. *Visual Cognition*, 13:99–108, 2006.

P. Brodatz. *Textures: a Photographic Album for Artists and Designers.* Dover Publications, 1966.

N. D. B. Bruce and J. K. Tsotsos. Saliency based on information maximization. *Advances in Neural Information Processing Systems*, 18:155–162, 2006.

N. D. B. Bruce and J. K. Tsotsos. Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, 9(3):1–24, 2009.

A. E. Burgess. Visual signal detection. iii. on bayesian use of prior knowledge and cross correlation. *Journal of the Optical Society of America A*, 2:1498–1507, 1985.

A. E. Burgess and B. Colborne. Visual signal detection. iv. observer inconsistency. *Journal of the Optical Society of America A*, 5:617–627, 1988.

A. E. Burgess, R.F. Wagner, R. J. Jennings, and H. B. Barlow. Efficiency of human visual signal discrimination. *Science*, 214:93–94, 1981.

P. J. Burt and E. H. Adelson. The laplcaian pyramid as a compact image code. *IEEE Transactions on Communications*, 31:532–540, 1983.

D. Casasent. Detection filters and algorithm fusion for atr. *IEEE Transactions on Image Processing*, 6:114–125, 1997.

M. S. Castelhano, A. Pollatsek, and K. R. Cave. Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review*, 15:795–801, 2008.

K. R. Cave and N. P. Bichot. Visuospatial attention: beyond a spotlight model. *Psychonomic Bulletin & Review*, 6:204–223, 1999.

K. R. Cave and J. M. Wolfe. Modelling the role of parallel processing in visual search. *Cognitive Psychology*, 22:225–271, 1990.

C. Chan and G. Pang. Fabric defect detection by fourier analysis. *IEEE Transactions on Industry Applications*, 36:1267–1276, 2000.

M. J. Chantler. Why illuminant direction is fundamental to texture analysis. *IEE Pro-ceedings Vision, Image and Signal Processing*, 142:199–206, 1995.

X. Chen and G. J. Zelinsky. Real-world visual search is dominated by top-down guidance. *Vision Research*, 46:4118–4133, 2006.

R. T. Chin. Automated visual inspection: 1981 to 1987. *Computer Vision Graphics Image Process*, 41:346–381, 1988.

C. Chubb and M. S. Landy. Orthogonal distribution analysis: A new approach to the study of texture perception. In M. S. Landy and J. A. Movshon, editors, *Computational Models of Visual Processing*, pages 291–301. Cambridge, MA: MIT Press., 1991.

M.M. Chun and J.M. Wolfe. Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology*, 30:39–78, 1996.

A. D. F. Clarke, P. R. Green, and M. Chantler. Modelling visual search on a rough surface. *Journal of Vision*, 9:1–12, 2009.

A. Cohen. Asymmetries in visual search for conjunctive targets. *Journal of Experimental Psychology: Human Perception and Performance*, 19:775–797, 1993.

J. D. Daugman. Two dimensional spectral analysis of cortical receptive field profiles. *Vision research*, 20:847–856, 1980.

R. L. DeValois, D. G. Albrecht, and L. G. Thorell. Spatial-frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22:545–559, 1983.

M. Dorr, K.R. Gegenfurtner, and E. Barth. The contribution of low-level features at the centre of gaze to saccade target selection. *Vision Research*, 2009.

V. Dragoi and M. Sur. Image structure at the center of gaze during free viewing. *Journal of Cognitive Neuroscience*, 18:737–748, 2004.

C. G. Drury. Visual inspection reliability: What we know and why we need to know it. *16th Human Factors in Aviation Maintenance Symposium*, pages 348–363, 2002.

M. P. Eckstein, B. Drescher, and S. S. Shimozaki. Attentional cues in real scenes, saccadic targeting and bayesian priors. *Psychological Science*, 17:973–980, 2006.

H. E. Egeth and S. Yantis. Visual attention: Control, representation, and time course. *Annual Review of Psychology*, 48:269–297, 1997.

W. Einhäuser and P. Koing. Does luminance-contrast contribute to a saliency map for overt attention? *European Journal of Neuroscience*, 17:1089–1097, 2003.

L. Elazary and L. Itti. Interesting objects are visually salieny. *Journal of Vision*, 8(3)(3):1–15, 2008.

Khemraj Emrith. *Perceptual Dimensions for Surface Texture Retrieval (PhD Thesis)*. Heriot-Watt University, 2008.

S. Engel, X. Zhang, and B. Wandell. Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature*, 388:68–71, 1997.

J. Escofet, R. Navarro, M. S. Millan, and J. Pladelloreans. Detection of local defects in textiles webs using gabor filters. *Optical Engineering*, 37:2297–2307, 1998.

D. J. Field. Relations between the statistics of natural images and the response profiles of cortical cells. *Journal of the Optical Society of America*, A:2379–2394, 1987.

J. M. Findlay. Saccade target selection during visual search. *Vision Research*, 37: 617–631, 1997.

J. M. Findlay and I. D. Gilchrist. *Active Vision: The Psychology of Looking and Seeing*. Oxford Psychology Series, 2003.

M. S. Fleck and S. R. Mitroff. Rare targets are rarely missed in correctable search. *Psychological Science*, 18:943–947, 2007.

S. Frintrop. *VOCUS: A Visual Attention System for Object Detection and Goal-directed Search (PhD Thesis)*. Springer Berlin/Heidelberg, 2006.

S. Frintrop, G. Backer, and E. Rome. Goal directed search with a top-down modulated computational attention system. In *In Proceedings of the Annual Meeting of German Association for Pattern Recognition*. Wein, Austria, 2005.

D. Gao and N. Vasconcelos. Discriminant saliency for visual recognition from cluttered scenes. In *Advances in neural information processing systems*, volume 17, pages 481–488. Cambridge, MA: MIT Press, 2005.

D. Gao, V. Mahadevan, and N. Vasconcelos. On the plausibility for the driscriminant center-surround hypothesis for visual saliency. *Journal of Vision*, 8(7)(13):1–18, 2008.

W. S. Geisler and J. S. Perry. A real-time foveated multiresolution system for low-bandwidth video communication. In B. E. Rogowitz and T. N. Pappas, editors, *Human vision and electronic imaging III*, pages 294–305. Bellingham, WA: SPIE., 1998.

W. S. Geisler and J. S. Perry. Real-time simulation of arbitrary visual fields. In A. T. Duchowski, editor, *Eye Tracking Research & Applications Symposium: Proceedings ETRA*, pages 83–87. New York: Association for Computing Machinery, 2002.

I. D. Gilchrist and M. Harvey. Refixation frequency and memory mechanisms in visual search. *Current Biology*, 10:1209–1212, 2000.

I. D. Gilchrist and M. Harvey. Evidence for a systematic component within scan paths in visual search. *Visual Cognition*, 14:704–715, 2006.

N. Graham, J. Beck, and A. Sutter. Two nonlinearities in texture segregation. *Investigative Ophthalmology & Visual Science*, 30, 1989.

H. H. Greene. Distance-from-target dynamics during visual search. *Vision Research*, 48:2476–2484, 2008.

R. Gurnsey and R. Browse. Micropattern properites and presntation conditions influencing visual texture discrimination. *Perception and Psyhophysics*, 41:239–252, 1987.

A. Gururajan and H. Sari-Sarraf. Statistical approach to unsupervised defect detection and multiscale localization in two-texture images. *Optical Engineering*, 47, 2006.

J. M. Henderson. Eye movement control during visual object processing: Effects of initial fixation position and semantic constraint. *Canadian Journal of Experimental Psychology*, 47:498–504, 1993.

J. M. Henderson, P. Weeks, and A. Hollingworth. The effects of semantic consistency on eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 25:210–228, 1999.

J. M. Henderson, J. R. Brockmole, M. S. Castelhano, and M. Mack. Visual saliency does not account for eye movements during visual search in real-world scenes. In R. v. Gompel, M. Fischer, W. Murray, and R. Hill, editors, *Eye Movement Research: Insights into Mind and Brain*, pages 537–562. Oxford: Elsevier., 2007.

J. M. Henderson, C. L. Larson, and D. C. Zhu. Full scenes produce more activation than close-up scenes and scene-diagnostic objects in parahippocampal and retrosplenial cortex: An fmri study. *Brain and Cognition*, 66:537–562, 2008.

C. Hickey, J. J. McDonald, and J. Theeuwes. Electrophysiological evidence of the capture of visual attention. *Journal of Cognitive Neuroscience*, 18:604–613, 2006.

I. Th. C. Hooge and C. J. Erkelens. Peripheral vision and oculomotor control during visual search. *Vision Research*, 39:1567–1575, 1999.

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.

T. S. Horowitz and J. M. Wolfe. Memory for rejected distractors in visual search? *Visual Cognition*, 10:257–298, 2003.

Z. Hou and J. M. Parker. Texture defect detection using support vector machines with adaptive gabor wavelet features. In *Seventh IEEE Workshops on Application of Computer Vision*, volume 1, pages 275–280, 2005.

A. D. Hwang, E. C. Higgins, and M. Pomplun. How chromaticity guides visual search in real-world scenes. In *In Proceedinds of the 29th Annual Cognitive Science Society*, pages 371–378. Austin, TX: Cognitive Science Society, 2007.

A. D. Hwang, E. C. Higgens, and M. Pomplun. A model of top-down attentional control during visual search in complex scenes. *Journal of Vision*, 9(5):1–18, 2009.

J. Iivarinen. Surface defect detection with histogram-based texture features. In *Proceedings of SPIE 4197*, pages 140–145, 2000.

D. E. Irwin and G. J. Zelinsky. Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics*, 64:882–895, 2002.

L. Itti and P. F. Baldi. Bayesian surprise attracts human attention. In *Advances in Neural Information Processing Systems*, volume 19, pages 547–554. Cambridge, MA:MIT Press, 2006.

L. Itti and P. F. Baldi. Bayesian surprise attracts human attention. *Vision Research*, 49:1295–1306, 2009.

L. Itti and C. Koch. A comparison of feature combination strategies for saliency-based visual attention systems. In *SPIE human vision and electronic imaging IV (HVEI'99), San Jose, CA*, pages 473–482. Springer Verlag, 1999.

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.

L. Itti and C. Koch. Computational modelling of visual attention. *Nature Reviews: Neuroscience*, 2:1–10, 2001.

L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, Nov 1998.

A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, 24:1167–1169, 1991.

W. J. Jasper, S. J. Garnier, and H. Potlapalli. Texture characterization and defect detection using adaptive wavelets. *Optical Engineering*, 35:3140–3149, 1996.

B. Julese. Textons, the elements of texture perception, and their interactions. *Nature*, 290:91–97, 1981.

C. Kayser, K. L. Nielsen, and N. K. Logothetis. Fixations in natural scenes: Interaction of image structure and image content. *Vision Research*, pages 2535–2545, 2006.

M. T. Khasawneh, S. Kaewkuekool, S. R. Bowling, R. Desai, X. Jiang, A. T. Duchowski, and A. K. Gramopadhye. The effects of eye movements on visual inspection performance. *Proceedings of industrial engineering research conference*, 2003.

R. Klein and M. Farrell. Search performance without eye movements. *Perception & Psychophysics*, 46:476–482, 1989.

R. M. Klein. Inhibition of return. *Trends in Cognitive Science*, 4:138–147, 2000.

C. Koch and S. Ullman. Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 4:219–227, 1985.

A. Kristjansson. In search of remembrance: evidence for memory in visual search. *Psychological Science*, 11:328–332, 2000.

C. Körner and I. D. Gilchrist. Finding a new target in an old display: Evidence for a memory recency effect in visual search. *Psychonomic Bulletin & Review*, 14: 846–851, 2007.

B. J. Krose. *A Description of Visual Structure (PhD Thesis)*. PhD thesis, Delft University of Technology, 1986.

A. Kumar. Computer-vision-based fabric defect detection: A survey. *IEEE Transactions on Industrial Electronics*, 55:348–363, 2008.

A. Kumar and G. K. H. Pang. Fabric defect segmentation using multichannel blob detectors. *Optical Engineering*, 39:3176–3190, 2000.

A. Kumar and G. K. H. Pang. Defect detection in textured materials using gabor filters. *IEEE Transactions on Industry Applications*, 38:425–440, 2002.

M. A. Kunar, S. Flusberg, and J. M. Wolfe. The role of memory and restricted context in repeated visual search. *Perception and Psychophysics*, 70:314–328, 2008.

H. W. Kwak, D. Dagenbach, and H. Egeth. Further evidence for a time-independent shift of the focus of attention. *Perception & Psychophysics*, 49:473–480, 1991.

D. Lamy and L. Zoarisa. Task-irrelevant stimulus salience affects visual search. *Vision Research*, 49:1472–1480, 2009.

M. S. Landy and N. Graham. Visual perception of texture. In L. M. Chalupa and J. S. Werner, editors, *The Visual Neurosciences*, pages 1106–1118. Cambridge, MA: MIT Press., 2004.

V. Leemans and M. F. Destain. A real-time grading method of apples based on features extracted from defects. *Journal of Food Engineering*, 61:83–89, 2004.

A. Leventhal. The neural basis of visual function. *In Vision and visual dysfunction*, 4, 1991.

D. M. Levi, S. A. Klein, and I. Chen. What is the signal in noise? *Vision Research*, 45:1835–1846, 2005.

D. T. Levin, Y. Takarae, A. Miner, and F. C. Keil. Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. *Perception & Psychophysics*, 63:676–697, 2001.

L. M. Linnett. *Multi-Texture Image Segmentation (PhD Thesis)*. Heriot-Watt University, 1991.

Y. Liu, R. T. Collins, and Y. Tsin. A computational model for periodic pattern perception based on frieze and wallpaper groups. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:354–371, 2004a.

Y. Liu, W-C. Lin, and J. Hays. Near-regular texture analysis and manipulation. *ACM Transactions on Graphics (SIGGRAPH)*, 23:368–376, 2004b.

A. Luschow and H. C. Nothdurft. Pop-out of orientation but no pop-out of motion at isoluminance. *Vision Research*, 33:91–104, 1983.

K. L. Mak and P. Peng. A survey of automated visual inspection. *Detecting Defects in Textile Fabrics with Optimal Gabor Filters*, 1:274–282, 2006.

J. Malik and P. Perona. Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A*, 7:993–, 1990.

B. B. Mandelbrot. *The Fractal Geometry of Nature*. W.H.Freeman & Co Ltd, 1983.

J. S. McCarley, R. F. Wang, A. F. Kramer, D. E. Irwin, and M. S. Peterson. How much memory does oculomotor search have? *Psychological Science.*, page 2, 2003.

J. S. McCarley, A. F. Kramer, C. D. Wickens, E. D. Vidoni, and W. R Boot. Visual skills in airport security screening. *Psychological Science*, 15:302–306, 2004.

G. McGunnigle. *The Classification of Textured Surfaces Under Varying Illuminant Direction (PhD Thesis)*. Heriot-Watt University, 1998.

B. J. Melloy, S. Das, A. K. Gramopadhye, and A. T. Duckiwski. 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.

M. C. Morrone and D. C. Burr. Feature detection in human vision: a phase-dependent energy model. *Proc. R. Soc. Lond. B*, pages 221–245, 1988.

B. C. Motter and E. J. Belky. The guidance of eye movements during active visual search. *Vision Research*, 38:1805–1815, 1998.

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 J. W. Holsapple. Saccades and covert shifts of attention during active visual search: Spatial distributions, memory, and items per fixation. *Vision Research*, 47:1261–1281, 2007.

K. J. Myers, H. H. Barrett, M. C. Borgstrom, D. D. Patton, and G. W. Seeley. Effect of noise correlation on detectability of disk signals in medical imaging. *Journal of the Optical Society of America A*, 2:1752–1759, 1985.

A. L. Nagy and R. R. Sanchez. Critical color differences determined with a visual search task. *Journal of the Optical Society of America A*, 7:1209–1217, 1990.

J. Najemnik and W. S. Geisler. Optimal eye move ment 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.

J. Najemnik and W. S. Geisler. Simple summation rule for optimal fixation selection in visual search. *Vision Research*, 49:1286–1294, 2009.

V. Navalpakkam and L. Itti. Modeling the influence of task on attention. *Vision Research*, 45:205–231, 2005.

V. Navalpakkam and L. Itti. Search goal tunes visual features optimally. *Neuron*, 53(4):605–617, Feb 2007.

M. B. Neider and G. J. Zelinsky. Scene context guides eye movements during search. *Vision Research*, 46:614–621, 2006a.

M. B. Neider and G. J. Zelinsky. Searching for camouflaged targets: Effects of target-background similarity on visual search. *Vision Research*, 46:2217–2235, 2006b.

F. N. Newell, V. Brown, and J. M. Findlay. Is object search mediated by object-based or image-based representations. *Spatial Vision*, 17:511–541, 2004.

T. S. Newman and A. K. Jain. A survey of automated visual inspection. *Computer Vision and Image Understanding*, 61:231–262, 1995.

A. Oliva, A. Torralba, M. Castelhano, and J. M. Henderson. Top-down control of visual attention in object detection. In *Proceedings of the International Conference on Image Processing*, volume 1, pages 253–256. Los Alamitos, CA: IEEE, 2003.

A. Oliva, J. M. Wolfe, and H. Arsenio. Panoramic search: The interaction of memory and vision in search through a familiar scene. *Journal of Experimental Psychology: Human Perception and Performance*, 30:1132–1146, 2004.

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. *Mathematical Models for Perceived Roughness of Three-Dimensional Surface Textures*. Ph.D. Thesis, Heriot-Watt University,, 2008.

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.

J. Palmer, P. Verghese, and M. Pavel. The psychophysics of visual search. *Vision Research*, 40:1227–1268, 2000.

S. Park, E. Clarkson, M. A. Kupinski, and H. H. Barrett. Efficiency of the human observer detecting random signals in random backgrounds. *Journal of the Optical Society of America A*, 22:3–16, 2005.

A. Parker and J. Hawken. Two-dimensional spatial structure of receptive fields in monkey striate cortex. *Journel of the Optical Society of America A*, pages 598–605, 1988.

D. J. Parkhurst and E. Niebur. Texture contrast attracts overt attention in natural scenes. *European Journal of Neuroscience*, 19:783–789, 2004.

D. J. Parkhurst, K. Law, and E. Niebur. Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42:107–123, 2002.

H. Pashler. *Attention*. Philadelphia: Taylor & Francis Press, 1998.

D. G. Pelli. The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10:437–442, 1997.

K. Perlin. Improving noise. *ACM Transactions on Graphics (TOG)*, 21:681–682, 2002.

K. Perlin and E. M. Hoffert. Hypertexture. *Proceedings of the 16th annual conference on Computer graphics and interactive techniques*, 23:253–262, 1989.

J. S. Perry and W. S. Geisler. Gaze-contingent real-time simulation of arbitrary visual fields. In B. E. Rogowitz and T. N. Pappas, editors, *Human vision and electronic imaging VII*, pages 294–305. Bellingham, WA: SPIE., 2002.

R. J. Peters, A. Iyer, L. Itti, and C. Koch. Components of bottom-up gaze allocation in natural images. *Vision Research*, 45(8):2397–2416, 2005.

M. Pomplun. *Analysis and Models of Eye Movements in Comparative Visual Search (Ph.D. thesis)*. Cuvillier, 1998.

M. Pomplun. Saccadic selectivity in complex visual search displays. *Vision Research*, 46:1886–1900, 2006.

M. Pomplun, E. M. Reingold, J. Shen, and D. E. Williams. The area activation model of saccadic selectivity in visual search. In L. R. Gleitman and A. K. Joshi, editors, *Proceedings of the twenty-second annual conference of the cognitive science society*, pages 375–380. Mahwah, NJ Erlbaum, 2000.

M. Pomplun, E. M. Reingold, and J. Shen. Peripheral and parafoveal cueing and masking effects on saccadic selectivity in a gaze-contingent window paradigm. *Vision Research*, 41:2757–2769, 2001.

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.

M. I. Posner. Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32:2–25, 1980.

U. Rajashekar, L. K. Cormack, and A. C. Bovil. Visual search: Structure from noise. In *In Proceedings of the eye tracking research & applications symposium*, pages 119–123. New Orleans: ACMPress, 2002.

U. Rajashekar, L. K. Cormack, and A. C. Bovik. Point of gaze analysis reveals visual search strategies. In Bernice E. Rogowitz, editor, *Human Vision and Electronic Imaging IX*, volume 5292. SPIE, Bellingham WA, E, 2004.

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.

T. Randen. Filtering for texture classification: a comparative study. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(4):291–310, 1999.

T. Randen and J.H. Husoy. Optimal filter-bank design for multiple texture discrimination. *International Conference on Image Processing*, pages 215–218, 1999a.

T. Randen and J.H. Husoy. Texture segmentation using filters with optimized energy separation. *Image Processing, IEEE Transactions on*, 8(4):571–582, 1999b.

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.

R. Rauschenberger. Attentional capture by auto- and allo-cues. *Psychonomic Bulletin & Review*, 10:814–842, 2003.

K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 85:618–660, 1998.

K. Rayner. The 35th sir frederick bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62:1457–1506, 2009.

K. Rayner, G. W. McConkie, and S. E. Ehrlich. Eye movements and integrating information across fixations. *Journal of Experimental Psychology: Human Perception and Performance*, 4:529–544, 1978.

J. H. Reynolds and L. Chelazzi. Attentional modulation of visual processing. *Annual Review of Neuroscience*, 27:647, 2004.

A. Rich, M. A. Kunar, M.J. Van Wert, B. Hidalgo-Sotelo, T.S. Horowitz, and J.M. Wolfe. Why do we miss rare targets? exploring the boundaries of the low prevalence effect. *Journal of Vision*, 8:1–17, 2009.

B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. Labelme: A database and web-based tool for image annotation., 2005.

U. Rutishauser and C. Koch. Probabilistic modeling of eye movement data during conjunction search via feature-based attention. *Journal of Vision*, 7(5):1–20, 2007.

H. Sari-Sarraf and J. S. Goddard. Vision systems for on-loom fabric inspection. *IEEE Transactions on Industry Applications*, 35:1252–1259, 1999.

R. Sawaki and J. Katayama. Top-down directed attention to stimulus features and attentional allocation to bottom-up deviations. *Journal of Vision*, 8:1–8, 2008.

K. Schicktanz. Automatic fault detection possibilities on nonwoven fabrics. *Mellrand Textilberrchie*, pages 294–295, 1993.

J. Schmidt and Gregory J. Zelinsky. Search guidance is proportional to the categorical specificity of a target cue. *Quarterly Journal of Experimental Psychology*, 19: 1–11, 2009.

C. T. Scialfa and K. Joffe. Response times and eyemovements in feature and conjunction search as a function of eccentricity. *Perception & Psychophysics*, 60: 1067–1082, 1998.

P. Sengottuvelan, A. Wahi, and A. Shanmugam. Automatic fault analysis of textile fabric using imaging systems. *Research Journal of Applied Sciences*, 3:26–31, 2008.

J. Shen, E.M. Reingold, and M. Pomplun. Distractor ratio influences patterns of eye movements during visual search. *Perception*, 29:241–250, 2000.

D.I. Shore and R.M. Klein. On the manifestations of memory in visual search. *Spatial Vision*, 14:59–75, 2000.

D. R. Simmons, A. Bell, A. Bowman, D. Brown, R. L., K. Millar, J. P. Siebert, M. Xiao, and A. Ayoub. Measurement of facial scarring in children with cleft lip/cleft lip and palate: A preliminary study. In *Proceedings of the European Conference on Visual Perception*, 2009.

B. Smith. Making war on defects. *IEEE Spectrum*, 30:43–50, 1993.

J. L. Sobral. Optimised filters for texture defect detection. *IEEE International Conference on Image Processing, 2005*, 3:565–568, 2005.

K. Y. Song, M. Petrou, and J. Kittler. Texture defect detection: A review. In *Applications of Artificial Intelligence X: Machine Vision and Robotics*, volume 1708, pages 99–106, 1992.

G. Sperling and M. J. Melchner. The attention operating characteristic: Examples from visual search. *Science*, 202:315–318, 1978.

K. Srinivasan, P. H. Dastoor, P. Radhakrishnaiah, and S. Jayaraman. Fdas: A knowledge-based framework for analysis of defects in woven textile structures. *Journal of the Textile Institute*, 83:431–448, 1992.

A. Sutter, J. Beck, and N. Graham. Contrast and spatial variables in texture segregation: testing a simple sptial-frequency channel model. *Perception and Psychophysics*, 46:312–332, 1989.

M. Swain and D. Ballard. Indexing via color histograms. *International Journal of Computer Vision*, 7:11–32, 1991.

R. G. Swensson and P. F. Judy. Detection of noisy visual targets: models for the effects of spatial uncertainty and signal-to-noise ratio. *Perception & Psychophysics*, 29:521–34, 1981.

H. W. Tang, V. Srinivasan, and S. H. Ong. Texture segmentation via nonlinear interactions among gabor feature pairs. *Optical Engineering*, 34:125–134, 1995.

B. Tatler, editor. *Eye Guidance in Natural Scenes: A Special Issue of Visual Cognition*. Psychology Press, 2009.

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.

B. W. Tatler and B. T. Vincent. Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, 2(2)(5):1–18, 2008.

B. W. Tatler, R. J. Baddeley, and I. D. Gilchrist. Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, 45:643–659, 2005.

A. Tavassoli, I. van der Linde, A. C. Bovik, and L.K. Cormack. An efficient technique for revealing visual search strategies with classification images. *Perception & Psychophysics*, 69:103–112, 2007a.

A. Tavassoli, I. van der Linde, A. C. Bovik, and L.K. Cormack. Orientation anisotropies in visual search revealed by noise. *Journal of Vision*, 7:1–8, 2007b.

A. Tavassoli, I. van der Linde, A. C. Bovik, and L.K. Cormack. Foveated analysis of image features at fixations. *Vision Research*, 47:3160–3172, 2007c.

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.

J. Theeuwes. Visual selective attention: a theoretical analysis. *Acta Psychologica*, 83:93–154, 1993.

J. Theeuwes. Top-down search strategies cannon override attention capture. *Psychonomic Bulletin & Review*, 11:65–70, 2004.

J. Theeuwes, A. F. Kramer, S. Hahn, D. E. Irwin, and G. J. Zelinsky. Attentional control during visual search: The effect of irrelevant singletons. *Journal of Experimental Psychology: Human Perception and Performance*, 24:1342–1353, 1998.

G. D. Thompson and J. Kidwell. Explaining the choice of organic produce: Cosmetic defects, prices, and consumer preferences. *American Journal of Agricultural Economics*, 80:277–287, 1998.

J. A. Throop, D. J. Aneshansley, W. C. Anger, and D. L. Peterson. Quality evaluation of apples based on surface defects: development of an automated inspection system. *Postharvest Biology and Technology*, 36:281–290, 2005.

R. B. Tootell, S. L. Hamilton, M. S. Silverman, and E. Switkes. Functional anatomy of macaque striate cortex. i. ocular dominance, binocular interactions, and baseline conditions. *Journal of Neuroscience*, 8:1500–1530, 1988.

A. Treisman. Features and objects: The fourteenth bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology*, 40A:201–237, 1991.

A. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.

Y. C. Tseng and C. S. Li. Oculomotor correlates of context-guided learning in visual search. *Perception & Psychophysics*, 66:1363–1378, 2004.

J.K. Tsotsos, S.M. Culhane, and Winky Yan Kei Wai. Visual attention: detecting abrupt onsets within the selective tuning model. *Computer Architectures for Machine Perception*, pages 76–87, 1995.

M. Unser and M. Edena. Nonlinear operators for improving texture segmentation based on features extracted by spatial filtering. *IEEE Transactions on Systems, Man and Cybernetics*, 20:804–815, 1990.

A. van der Schaaf and J. H. van Hateren. Modelling the power spectra of natural images: statistics and information. *Vision Research*, 28:2759–2770, 1996.

M. J. van Wert, T. S. Horowitz, and J. M. Wolfe. Even in correctable search, some types of rare targets are frequently missed. *Attention, Perception & Psychophysics*, 7:541–553, 2009.

P. Verghese. Visual search and attention: A signal detection theory approach. *Neuron*, 31:523–535, 2001.

M. Verma and P. W. McOwan. Generating customised experimental stimuli for visual search using genetic algorithms shows evidence for a continuum of search efficiency. *Vision Research*, 2008.

B. T. Vincent, T. Troscianko, and I. D. Gilchrist. Investigating a space-variant weighted salience account of visual selection. *Vision Research*, 47(10-12):1809–1820, 2007.

V. Virsu and J. Rovamo. Visual resolution, contrast sensitivity, and the cortical magnification factor. *Experimental Brain Research*, 37:475–494, 1979.

G. Voronoi. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. *Journal für die Reine und Angewandte Mathematik*, 133:97–178, 1907.

R. F. Voss. Random fractal forgeries. In *Fundamental Algorithms For Computer Graphics*, pages 805–835. Berlin: Springer-Verlag, 1985.

R. F. Voss. Fractals in nature: from characterisation to simulation. In *The Science of Fractal Images*, pages 21–70. New York: Springer-Verlag, 1988.

D. Walther and C. Koch. Modeling attention to salient proto-objects. *Neural Networks*, 19:1395–1407, 2006.

D. E. Williams and E. M. Reingold. Preattentive guidance of eye movements during triple-conjunction search tasks. *Psychonomic Bulletin and Review*, 8:476–488, 2001.

L G Williams. The effects of target specification on objects fixated during visual search. *Acta Psychologica*, 27:355–360, 1967.

J. M. Wolfe. Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, 1:202–238, 1994.

J. M. Wolfe. What can 1 million trials tell us about visual search. *Psychological Science*, 9:33–39, 1997.

J. M. Wolfe. Visual search. In Pashler H., editor, *Attention*. London UK: University College London Press, 1998.

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 and G. Gancarz. Guided search 3.0. In V. Lakshminarayanan, editor, *Basic Clinical Applications of Vision Science*, pages 189–192. Dordrecht, Netherlands: Kluwer Academic, 1996.

J. M. Wolfe, K. R. Cave, and S. L. Franzel. Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15:419–433, 1989.

J. M. Wolfe, N. Klempen, and K. Dahlen. Postattentive vision. *Journal of Experimental Psychology: Human Perception and Performance*, 26:693–716, 2000.

J. M. Wolfe, A. Oliva, T.S. Horowitz, S. J. Butcher, and A. Bompas. Segmentation of objects from backgrounds in visual search tasks. *Vision Research*, 42:2985–3004, 2002.

J. M. Wolfe, T. S. Horowitz, and N. M. Kenner. Cognitive psychology: Rare items often missed in visual searches. *Nature*, 435:439–440, 2005.

R. D. Wright. *Visual Attention*. Oxford University Press, 1998.

X. Xie. A review of recent advances in surface defect detection using texture analysis techniques. *Electronic Letters on Computer Vision and Image Analysis*, 7:1–22, 2008.

X. Xie and M. Mirmehdi. Localising surface defects in random colour textures using multiscale texem analysis in image eigenchannels. In *Proceedings of the 12th IEEE International Conference on Image Processing*, pages 1124–1127, 2005a.

X. Xie and M. Mirmehdi. Texems: Texture exemplars for defect detection on random textured surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:1454–1464, 2005b.

H. Yang and G. J. Zelinsky. Visual search is guided to categorically defined targets. *Vision Research*, 49:2095–2103, 2009.

S. Yantis. Goal directed and stimulus driven determinants of attentional control. In S. Monsell and J. Driver, editors, *Attention and Performance*, volume 18, pages 73–103. Cambridge: MIT Press, 2000.

S. Yantis and J. Jonides. Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 10:601–621, 1984.

V. Yanulevskaya, J. M. Geusebroek, J. B. C. Marsman, and F. W. Cornelissen. Natural image statistics differ for fixated vs non-fixated regions. In *Perception ECVP Abstract Supplement*, volume 37, page 56, 2008.

R. Young. The gaussian deriviative theory of spatial vision: analysis of cortical cell receptive field line-weighting profiles, 1985.

C. Yue, F. Alfnes, and H. H. Jensen. Discounting spotted apples: Investigating consumers' willingness to accept cosmetic damage in an organic product. *Journal of Agricultural and Applied Economics*, 41:29–46, 2009.

G. J. Zelinsky. Using eye saccades to assess the selectivity of search movements. *Vision Research*, pages 2177–2187, 1996.

G. J. Zelinsky. Precueing target location in a variable set size "'nonsearch"' task: Dissociating search-based and interference-based explanations for set size effects. *Journal of Experimental Psychology: Human Perception and Performance*, 25: 875–903, 1999.

G. J. Zelinsky. ye movements during change detection: Implications for search constraints, memory limitations, and scanning strategies. *Perception & Psychophysics*, 63:209–225, 2001.

G. J. Zelinsky. Detecting changes between real-world objects using spatio-chromatic filters. *Psychonomic Bulletin and Review*, 10:533–555, 2003.

G. J. Zelinsky. A theory of eye movements during target acquisition. *Psychological Review*, 115:787–835, 2008.

G. J. Zelinsky, R. P. N. Rao, M. M. Hayhoe, and D. H. Ballard. Eye movements reveal the spatiotemporal dynamics of visual search. *Psychological Science*, 8: 448–453, 1997.