INFORMATION RECOVERY FROM RANK-ORDER ENCODED IMAGES

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Engineering and Physical Sciences

2008

By
Basabdatta B. Sen
School of Computer Science

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

(ERPJD)

Contents

List of Tables

List of Figures

Abstract

The time to detection of a visual stimulus by the primate eye is recorded at 100 – 150ms. This near instantaneous recognition is in spite of the considerable processing required by the several stages of the visual pathway to recognise and react to a visual scene. How this is achieved is still a matter of speculation. Rank-order codes have been proposed as a means of encoding by the primate eye in the rapid transmission of the initial burst of information from the sensory neurons to the brain. We study the efficiency of rank-order codes in encoding perceptually-important information in an image. VanRullen and Thorpe built a model of the ganglion cell layers of the retina to simulate and study the viability of rank-order as a means of encoding by retinal neurons. We validate their model and quantify the information retrieved from rank-order encoded images in terms of the *visually-important* information recovered. Towards this goal, we apply the 'perceptual information preservation algorithm', proposed by Petrovic and Xydeas after slight modification. We observe a low information recovery due to losses suffered during the rank-order encoding and decoding processes. We propose to minimise these losses to recover maximum information in minimum time from rank-order encoded images. We first maximise information recovery by using the pseudo-inverse of the filter-bank matrix to minimise losses during rank-order decoding. We then apply the biological principle of lateral inhibition to minimise losses during rank-order encoding. In doing so, we propose the Filter-overlap Correction algorithm. To test the perfomance of rank-order codes in a biologically realistic model, we design and simulate a model of the foveal-pit ganglion cells of the retina keeping close to biological parameters. We use this as a rank-order encoder and analyse its performance relative to VanRullen and Thorpe's retinal model.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

Acknowledgements

I am very grateful to my supervisor, Prof. Steve Furber, firstly, for giving me an opportunity to work with him. Secondly, for providing guidance, help and advice during regular weekly meetings. And last but not the least, for providing me with a flexibility on the job, which I required to look after my child, which in turn helped me immensely to work efficiently.

Also, I take this opportunity to acknowledge my heartfelt gratitude to: Dr. Amit Konar, my teacher at Jadavpur University, Kolkata (India), for motivating and guiding me in pursuing a PhD; Mr. U.K. Mishra, my reporting authority at Rourkela Steel Plant, Steel Authority of India Ltd., for providing help and advice during the course of my application for doctoral studies.

Thanks: to my colleagues, especially Viv Woods, Gavin Brown, Lilian Janin, Jeremy Singer, Louis Plana, not necessarily in that order, for being there when I needed advice, help, encouragement; to my husband, without whose encouragement and support I could not pursue higher studies; to my son, for typing a significant part of the thesis draft and for adjusting to my hectic schedules; to my parents and grandparents, for teaching me the values of honesty, sincerity and dedication.

I dedicate this work to my father, Shri Swapan Kumar Sen.

Chapter 1

Introduction

"A picture is worth a thousand words" — it is a well known saying, implying the speed and vividness with which our eyes help us to perceive our surroundings. Indeed we just have to glance at, for example, a picture on the wall to recognise whether it contains a landscape, a face of a girl or an aeroplane. Although it will take some time to see exactly how many trees the landscape contains, or to read the expression on the face, or to see how many windows the aeroplane has, yet, recognition is instantaneous. In fact, the survival of an animal often depends on the speed with which the animal can detect its prey or predator. It is based on this instant detection that the animal acts in accordance with the situation. Thus **rapid information recovery** from the outside world by our eyes or other sensory organs seems to be a result of evolution which has helped all the different species of living things to be fit to survive in a hostile environment, and thus to co-exist. So, how can the eye or the other sensory organs *process information* so **fast**? To this day, there is no definite answer to this query. Presently, this question poses a major problem for scientists and has encouraged a multi-disciplinary approach across different scientific communities such as Biologists, Computational Neuroscientists, Computer Scientists and Psychologists, to name a few, towards finding a solution.

In this thesis, we concentrate on **Rank-order code**, which is a hypothesis about how the eye achieves '*near-instantaneous*' recognition within a fraction of a second. More specifically, we investigate the **quantity** of *information* that can be *recovered* from a **rank-order encoded picture**, and **more importantly**, the **rate** and **time** of such recovery. The basic flow of the work is summarised in figure 1.1 and is listed below:

- An input image is presented to a rank-order encoder.
- The rank-order encoded image is decoded by reconstructing the original image from its encoded version.
- Our visual perception of the outside world follows a certain law, and cannot distinguish between certain aspects of an image (which will be dealt with in detail in chapter 5). We limit our present work to only those aspects of an image which are visually perceptible, and we term those aspects the **visually-important information**. Returning to the reconstructed image, we compare it with the original image by quantitatively measuring the *visually-important information* that is contained in the reconstruction as a percentage of that in the original, using an objective measure.
- Based on the objective measure, we study the information loss in the rank-order 'codec', shown in figure 1.1, i.e. losses incurred during both the processes of rank-order encoding and decoding and experiment with ways to
  - minimise the loss during decoding, and then
  - minimise the loss during encoding.

  We present the results and analyze them using the 'umbrella' phrase - *information recovery from rank-order codes.*

In this chapter, we start by giving a brief background to the the means of information processing in the nervous system, the **neural codes**, which will be elaborated upon in subsequent chapters. Following this, we discuss the inspiration behind carrying out the research presented in this thesis in relevance to the present day research in this field. Subsequently, we state our hypothesis in carrying out this work, followed by the research contributions made during its course. We conclude the chapter by presenting a chapter-wise structure of the thesis and the resulting publications.

1.1 On neural codes

The **retina** is the membrane lining the posterior wall of the eyeball where light must fall to enable vision. It has been a popular area for neuro-physiological research for reasons mentioned in chapter 3. The cells of the retina, as well as

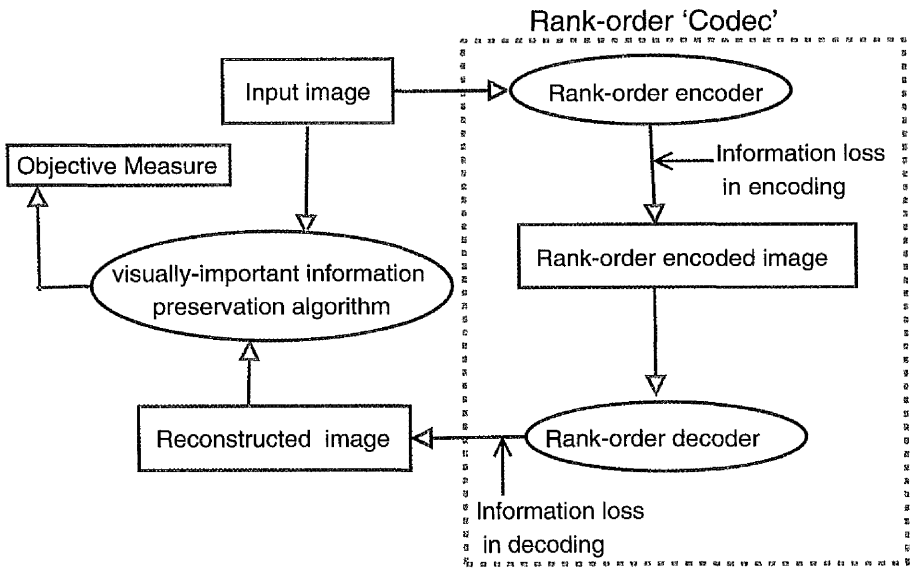


Figure 1.1: The flow of events during the coding and decoding of a picture using a rank-order **'codec'** — **co**ding-**dec**oding model.

those of the other parts of the central nervous system, are known as **neurons**. *Photoreceptors* are neurons in the retina which receive incident light from the external environment and convert it into electrical signals. These are then passed down several other layers of retinal neurons until they reach the **ganglion cell** layer. The ganglion cells in the retina differ from the other cells in that they convey information by generating sequences of voltage impulses, commonly referred to in the literature as *spike trains*, in response to different incoming signal strengths. Due to this characteristic, the ganglion cells are also known as **spiking neurons**, which **encode** information from the outside world into *trains of spikes*. This information is transmitted to the brain when the spike trains travel out of the ganglion cell layer and down the *optic nerve* to the brain. Such information about the external environment, encoded in trains of spikes generated and transmitted by spiking neurons, is known as a **neural code**.

The brain is analogous to a 'black box' where sequences of spikes enter from various sensory neurons, for example the neurons of the eye, the ear, the skin, etc. The primary function of the brain, then, is to interpret these spikes, decide upon the course of action based on such interpretation and act accordingly. So, the obvious presumption is that the brain must have an exhaustive dictionary,

whereby it maps a certain pattern of spikes to a particular stimulus. Alas, things are extremely complex where the brain is concerned, much to the chagrin of the people who are desperately trying to understand it. It has been observed experimentally that the same cell gives out different spike train patterns in response to the same stimulus, a fact which continues to baffle the neuroscience community to this day. Thus, one of the major questions that is still to be answered, even after years of research, is — how does the brain interpret the external world from a pattern of spikes? Another way to frame the question is —

- how do ganglion cells *encode* information from the external visual environment into trains of spikes?
- how does the brain *decode* these spikes to understand the coded information?

Answering either of these questions could be a strong indication, or even a solution, to the other. Experiments are continually being carried out towards that end. Work that starts by trying to answer the first question and thus works towards a solution to the second is known as a *'bottom-up approach'*. The reverse is known as a *'top-down'* approach.

1.2 Rank-order: a neural code?

Several theories about the neural codes have been proposed over the years, the basis of which is either the *frequency* or the *timing* of spikes in a spike train. The theory that the information is embedded in the frequency of firing of spikes by an individual or a group of neurons is the first and the oldest, and was proposed by E. D. Adrian based on empirical evidence [11]. Although evidence of such means of encoding exists throughout the nervous system, over the years there has been serious speculation about whether this is the only means of encoding being followed by all parts of the brain. In fact, it is now believed that there are a combination of various types of encoding techniques existent in the different parts of the brain, depending on the primary function of each part. Of these, the theory that **spike times** may play a role in *encoding information* is the most popular hypothesis.

The rank-order code theory is a hypothesis about the neural encoding technique based on the time to first spike of a neuron and is discussed at length in

chapter 2. Proposed by Thorpe et al, it suggests that information is encoded in the order in which a population of ganglion cells fire their first spikes [75]. In order to test the hypothesis, VanRullen and Thorpe designed and simulated a simplified model of the retina, which they refer to as the **retinal model**, and showed with empirical data that when a population of ganglion cells are stimulated by different parts of an image, the most important information about the image is encoded *in* the **timing** of the **first spikes** fired *by* the **first 10%** of the ganglion cells in the population. Further, the input image is recognisable by the time only the **first 1 – 2%** of the ganglion cells in the population have fired their first spikes [81]. Thus, the rank-order code hypothesis of neural encoding gives a biologically plausible explanation of the *ultra-rapid* processing of visual information in the retina. However, there is no firm evidence, to date, of rank-order codes from neuro-physiological studies made on the ganglion cells or the other neurons higher up in the visual pathway. Recently, though, empirical evidence of spike time encoding, especially first spike times, has been obtained in the somatosensory pathways, indicative of their playing a very important role in rapid encoding and transmission of information about external stimuli [53, 35].

We base our work on *VanRullen and Thorpe's retinal model*, and focus on **quantitatively** estimating the visually-important information that can be recovered from rank-order encoded images. More importantly, we study the rate of such recovery in terms of the percentage of information recovered, against the number of cells that fire their first spikes [66]. Such an approach is indicative of the time of information retrieval, an issue that is central to the rank-order code hypothesis. Based on our observation of such a measure, we introduce a novel decoding mechanism with the **aim** to achieve **rapid recovery** of as much information as possible, or in other words — to **maximise** information recovery from rank-order codes [67]. Subsequently, we **optimise** the information recovery from rank-order codes by settling on a *trade-off* between the **quantity** and **time** of information recovery. This we do by introducing the biological principle of lateral inhibition to reduce redundancy in the encoded data obtained by rank-order encoding using VanRullen and Thorpe's retinal model. Continuing on such a pursuit, we design and simulate a biologically realistic model of the retina for testing the performance of rank-order codes in a set-up that is constrained by the laws of biology. The model simulates the neurons of the **foveal pit** (defined in section 3.1.3) of the retina. Empirical data show that visual data encoded using

this model can be decoded more efficiently than VanRullen and Thorpe's model retina both in terms of time and quantity of information recovered.

So, why would we be interested in investigating the information recovery from rank-order codes? Or more specifically, why use rank-order codes at all? The reason and the motivation for doing so are discussed in the following section.

1.3 Motivation for studying rank-order codes

In spite of the speculation that revolves round rank-order codes, as to whether they are utilised at all by the retinal neurons, there has been no proposed hypothesis, nor any empirical evidence to date, that can provide an explanation for the 'near-instantaneous' visual perception that is observed by subjective empirical tests on the primate visual system. A picture flashed for just 20ms is enough for a human or a monkey to recognise the object in the picture [74, 19, 80]. The only empirical evidence of a neural code — the rate codes — cannot explain such speed of processing [22, 73]. Such a hiatus between 'effect' (the speed of visual processing) and 'cause' (the neural code that enable such efficient and rapid message-passing) is an intriguing problem and raises interest in further research on rank-order codes.

Our primary motivation for working on rank-order codes, though, is SpikeNET, a system designed by Thorpe et al in order to test the feasibility of using rank-order codes in simulating large-scale networks of *asynchronously-spiking* neurons [70]. SpikeNET consists of a layered architecture of spiking neurons to simulate a simplified version of the recognition circuit of the brain. With such a set-up, it can process and analyze a 150×150 pixel image in 5.7ms on a 2 GHz Pentium-IV machine [14]. Although the architecture of SpikeNET is far from biologically realistic, it still illustrates the viability of using rank-order codes in simulating hardware for fast processing of visual information [71]. Such models are an inspiration for the new and emerging field of 'biologically inspired computer architecture', whereby engineers are seeking to implement the principles of efficient and intelligent information processing used by the brain in developing novel computer hardware. Research on rank-order codes could, thus, be beneficial to the computer engineer in that, if it helps to engineer a power efficient computer architecture, it will be considered a significant achievement, even if rank-order codes are proved to be of limited applicability in biology.

Thus, having been motivated to study rank-order codes, our primary interest was to study the decoding mechanism. The motivation in choosing this particular aspect of rank-order codes is the need to estimate the potential of rank-order codes in encoding visually-important information **speedily**, and yet, **efficiently**. VanRullen and Thorpe decoded rank-order codes to test the performance of the codes in encoding visual information, both in terms of 'picture fidelity' as well as 'time-to-decoding' for achieving such fidelity. However, as mentioned in section 1.1, the behaviour of the human visual system is highly nonlinear and follows the characteristic of a band-pass filter (to be discussed in chapter 5). A more realistic approach to rank-order code performance evaluation with respect to vision would be to obtain a *quantitative estimate* of visually-important information that can be recovered from the codes. In order to do so, an important requirement is to minimise the loss incurred during decoding the codes, so that the information retained during rank order encoding can be recovered optimally. This would, in turn, give a true estimate of the loss in information suffered during rank-order encoding. The main motivation is, thus, to minimise the losses incurred with the rank-order 'codec', as shown in figure 1.1, to obtain an optimal objective measure. Moreover, such an estimate would be a relevant study for any other mode of neural encoding that may be discovered in the future.

1.4 Thesis hypothesis and questions

The primary hypothesis forming the basis of this work is that — *"Rank-order codes are a viable means of information encoding in applications that require fast and efficient information transmission"*. This hypothesis apparently overlaps with that of Thorpe's, when he proposed rank-order codes as a viable means of information encoding taking place in the visual system. However, in Thorpe's work, the main emphasis is on showing the superior performance of rank-order codes in providing an explanation of the tremendous speed of information processing taking place in the human visual system, as opposed to the widely accepted and empirically proved rate code theory of neural encoding [81]. We, on the other hand, study rank-order codes from a very different perspective — the thrust being to engineer a model inspired by the high efficiency and speed of working of the asynchronously firing biological neuronal populations. To that end, the main aim of the thesis is to find an answer to the query — *whether rank-order codes*

are, at all, able to represent input information efficiently, and if so, then how fast can that information be read from the codes? The presumption we made in framing the hypothesis is, in fact, a prediction towards an affirmative answer to this query, i.e. , saying — *"yes, it does"*. To prove our hypothesis, the first question would then be — *"How can we quantitatively measure the performance of rank-order codes with respect to vision?"*. Although quantitative measures using mean-square errors and information theory were produced by Thorpe et al in their study, the measures do not account for the non-linearity observed in the human visual system [81]. In our quest to explore the performance of rank-order codes in retrieving visually-important information, we found an objective measure, proposed by Petrovic and Xydeas, that suited our work [59]. Since such an *objective measure of visually-important information content in rank-order codes* was not carried out before, a corollary to our previous question would be — *"What effect does this measure have on the results of VanRullen and Thorpe?"*.

Based on the empirical results obtained while studying the above queries (elucidated in chapter 5), we pose a third question — *"Can the quantity of information retrieved from rank-order codes be maximised by improving the rank-order decoding techniques used by VanRullen and Thorpe?"*. Thus, our quest for a suitable decoding technique is to minimise the information loss incurred during decoding the rank-order encoded data, as indicated in figure 1.1, so that information recovery can be maximised. The main evaluating criterion here is the amount of information recovered in minimum time; the two extreme cases being either (a) perfect information recovery in a time which is beyond the permissible limits for rank-order codes, thus rendering them unsuitable for fast information recovery, or (b) very low information recovered within such time limits. A trade-off between the quantity of information recovered and the time to achieve such a recovery would be an ideal mechanism for a rank-order 'codec'. Thus the corollary to the third question above would be a fourth question — *"How can the information recovery from rank-order codes be optimised?"*.

Until this point in our work, we based our empirical studies on the performance of rank-order codes on the retinal model designed by VanRullen and Thorpe. However the model is far from being biologically realistic in terms of the design and layout of the basic components used to simulate visual processing. This brings us to a fifth and final question — *"Do rank-order codes perform better in a biologically realistic model?"*.

In the following chapters, we deal with the questions stated above in the same sequence as they are posed, and give a detailed empirical analysis of the performance of rank-order codes with respect to vision.

1.5 Thesis contribution

The overall contribution of the thesis is to provide a detailed analysis of *visually-important information recovery from rank-order codes*, supported with empirical data. The contributions made while working towards the larger aim are as listed below:

- Empirical work to validate the retinal model of VanRullen and Thorpe. This is the basic model used in the preliminary studies made during our work (chapter 4).
- The application of an objective measure to form a quantitative estimate of the information recovery from rank-order encoded images (chapter 5).
- Improving on the decoding mechanism used by VanRullen and Thorpe in their retinal model in order to retrieve the maximum possible information from a rank-order encoded image with respect to the input image. This is done by application of a pseudo-inverse algorithm to obtain the inverses of singular matrices encountered during decoding of the codes (chapter 6).
- Improving on the encoding mechanism used by VanRullen and Thorpe in order to retrieve maximum possible information in minimum possible time. Thus, we optimise the information recovery from rank-order codes. This is done by application of the biological principle of lateral inhibition to re-order the rank-ordered 'spikes' prior to decoding using VanRullen and Thorpe's method (chapter 7).
- The design and simulation of a *model of the foveal-pit of the retina* to test rank-order code performance in a biologically realistic environment (chapter 8).
- Empirical evidence of better performance of our FOVEAL-PIT MODEL in terms of information encoding using rank-order codes, as compared to the *retinal model* of VanRullen and Thorpe (chapter 8).

In addition, the thesis presents a literature review encompassing neuro-physiology (chapter 2), retinal-physiology (chapter 3), physiology and characteristics of ganglion cell (chapter 3) and neural codes (chapter 2).

1.6 Structure of the thesis

In chapter 2, we present a review of the basic structure of a neuron in section 2.1, and the mechanism of spike generation in section 2.2. Following this, in section 2.3, is a review of existing theories of neural encoding and an introduction to the rank-order codes.

In chapter 3, we review the basic structure of the retina. The main purpose of a literature review in this area is to become familiar with the neurons of the retina, especially the ganglion cells. The *retinal model* built by VanRullen and Thorpe consists only of layers of simulated ganglion cells. Thus, a preliminary review of these cells help in gaining a basic knowledge of their structure, characteristics and functioning, before proceeding towards the validation of VanRullen and Thorpe's retinal model. Further, to be able to design a model of the foveal-pit of the retina as presented in chapter 8, a wide literature review on the size and density of the cells is presented in this chapter. In this context, we also do a literature review on the principle of lateral inhibition used by the sensory neurons to reduce redundancy in the sensory input stimuli.

In chapter 4, we present the empirical results of our validation of VanRullen and Thorpe's retinal model, and show a qualitative estimate of the information recovered on decoding rank-order encoded images.

In chapter 5, we present an objective measure proposed by Petrovic and Xydeas, which we have used on VanRullen and Thorpe's retinal model to make a quantitative estimate of the performance of rank-order codes in information recovery. Petrovic's objective measure is based on subjective tests made on the non-linearity of the human visual system. Thus, the objective metric used on the information recovery is, in effect, a measure of the **perceptually-important** (which is same as *visually-important* and these terms are used interchangeably in this work) information content in the decoded picture with respect to the original. Such an approach to evaluate the performance of the rank-order codes is done for the first time.

In chapter 6, we present a novel method of using filter-banks and a pseudo-inverse technique for matrix inversion for decoding rank-order codes, the larger aim being to maximise the objective measure obtained while evaluating the rank-order code performance, as described in chapter 5. We present a detailed theoretical analysis of the method followed by empirical results.

Chapter 7 is a pursuit of means to optimise visually-important information recovery from rank-order codes. We introduce a novel algorithm — **Filter-overlap Correction algorithm** (FoCal) — based on the mechanism of lateral inhibition used by sensory neurons, in order to improve the method of rank-order encoding used by VanRullen and Thorpe. We present a theoretical and empirical analysis of the algorithm as applied to VanRullen and Thorpe's retinal model. Subsequently, we review the **Matching Pursuit** (MP) algorithm, which is observed to be mathematically the same as FoCal. This algorithm was used previously by Perrinet et al with the purpose of improving the quality of the decoded image from VanRullen and Thorpe's retinal model [54]. However, application of the MP algorithm to the retinal model as done by Perrinet et al is very different from that of applying the FoCal, the latter being very similar to the actual biological process of lateral inhibition.

In chapter 8, we present a new model, that of the *foveal pit of the retina*, with the motive to test the performance of rank-order codes when encoding is carried out with a biologically realistic retinal model. In other words, after settling on an optimal encoding technique using the biological principle of lateral inhibition, we now attempt to improve on the retinal model designed and used by VanRullen and Thorpe as a rank-order encoder by designing a FOVEAL-PIT MODEL on the basis of available factual data about the primate retina. In section 8.1, we review and indicate the actual biological parameters on which we base our choice of the simulation parameters. In section 8.2, we present the empirical results and analyse the performance of the FOVEAL-PIT MODEL with respect to that of VanRullen and Thorpe's retinal model.

Since evaluation of information recovery from rank-order codes using Petrovic and Xydeas's objective measure is a novel application used in this work, we deemed necessary a benchmarking of the above-mentioned measure against standard image fidelity measures in digital image processing. In chapter 9, we benchmark the objective measure with the more common methods of Root Mean Square Error and Fourier Transform in sections 9.1 and 9.3 respectively. Further,

we also obtain objective measures for out-of-sample images in section 9.2.

Finally, in chapter 10, we summarize the work presented in the earlier chapters, and elaborate upon the scope of work that can be carried out as a continuation of the investigation carried out in this thesis. Possible future research on related areas is also discussed.

1.7 Publications

The following publications are based on work presented in this thesis:

- [66] Basabdatta Sen and Steve Furber
  "Information Recovery from Rank-order Encoded Images",
  Proceedings of the International Workshop on Biologically Inspired Information Fusion, pp. 8–13,
  Guildford, U.K. August 2006.

- [67] Basabdatta Sen and Steve Furber
  "Maximising Information Recovery from Rank-order Codes", Proceedings of SPIE Defense and Security Symposium, Vol. 6570, pp. 65700C - 1 – 12, Orlando, Florida, U.S.A. April 2007.

Chapter 2

Neurons and Neural Communication

The brain is made up of two types of cell, viz. neuroglia and neurons. Neuroglia, meaning 'nerve glue', are also called glial cells and their population exceeds that of neurons. One of their main functions is to hold the neurons in place. Neurons are involved in information processing in the brain. Sensory neurons are the receptor neurons in the sensory organs and receive information from the outside world, for example the photoreceptors of the retina in the eye. This information is transmitted via other neurons to the brain. Conversely the brain sends messages to different parts of our body through cascades of neurons and ultimately on to the motor neurons from where the message is transmitted to the muscle fibres. Some neurons, such as the ganglion cells in the retina, have long axons to carry signals over long distances in the central nervous system and are called principal or relay neurons. These neurons propagate signals by generating voltage impulses and are also known as spiking neurons; other neurons such as the bipolar cells in the retina are concerned only with local processing and are called interneurons, some of which, like some amacrine cells in the retina, also spike.

A nerve comprises a bundle of nerve fibres, each of which is an axon of a spiking neuron. A nerve fibre is thus the basic conducting unit of the brain, transporting information by carrying voltage impulses generated in response to external stimuli or other neurons. There is an 'all or nothing' relationship between the stimulus and neural response. For example, if the stimulus is an electrical current applied to a single neuron, then, if that current is very weak, or if its duration is extremely short, it is unable to excite the neuron and no impulse is

generated. However, when the electric current is increased beyond a certain value, a voltage impulse is generated by the neuron. This is analogous to the pressure on the trigger of a rifle — either it is strong enough to fire the bullet or it is too weak to do anything. It is not possible to vary the intensity of the impulse by changing the strength of the stimulus. Further, if two consecutive stimuli are applied to the same neuron in rapid succession, the impulse generated in response to the first stimulus renders the fibre completely inexcitable to the second stimulus. This 'refractory' state lasts until the response to the first stimulus has subsided. The implication of this characteristic is that in response to a constant stimulus, a neuron transmits a series of impulses which cannot recur at more than a certain frequency. This phenomenon is analogous to a stream of bullets from a machine gun, rather than a continuous stream of water from a hose [11].

Then the question arises — how can we distinguish between the different strengths of stimuli, which can activate a nerve fibre, but do not affect the amplitude or strength of the reaction? In such a case, the pattern of spike generation could carry a message about the stimulus. Ideally, a dictionary for neural codes would relate a certain spike train pattern to a single input stimulus. In practice, however, there is nothing unique about a pattern of spikes generated by a neuron in response to a particular stimulus and there is no one-to-one mapping. Repeated presentations of the same stimulus give out different spike train patterns from a single neuron; however, the firing frequency has been found to vary with the stimulus strength and this concept gave rise to *rate code theory* (discussed in section 2.3.1). More recently, rate codes have been described as too complex and time consuming with respect to the tremendous *speed* of visual processing in the primate eye and this has led to a proposal of alternative coding schemes based on spike times to explain the fast processing techniques of the primate visual system (discussed in sections 2.3.2 and 2.3.3).

We present a brief description of the physiology of a neuron in section 2.1, followed by a discussion of how a neuron generates Action Potentials in section 2.2.1. In section 2.2.2, we present a brief overview of synapses used by neurons to transmit information for inter-neuron and neuro-muscular communication. In section 2.3, we discuss how a neuron encodes information about the external world in sequences of Action Potentials, thereby introducing the theory of rank-order codes in section 2.3.4.

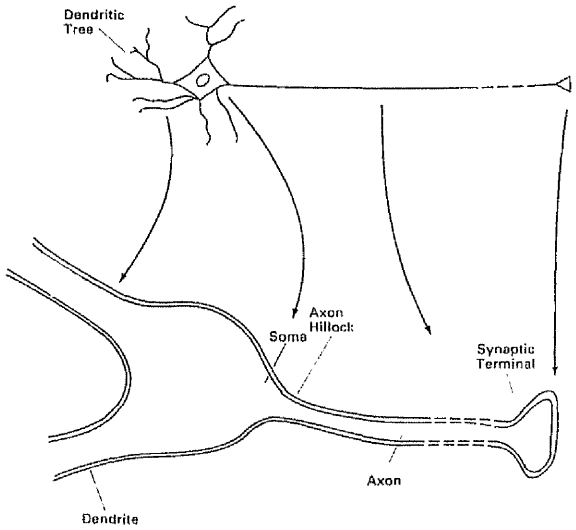


Figure 2.1: A neuron and its different parts [6].

2.1 Physiology of neurons

We start this section by a brief introduction into the basic structure of a neuron and the elementary role of each part in signal processing (section 2.1.1). This is followed (section 2.1.2) by an overview of the balance of electrical charges inside and outside the neuron, which is the cause of generation of neural spikes.

2.1.1 Structure of a neuron

Like all other cells in a living organism, neurons have a cell body called the **soma**, which consists of the *nucleus* and other substances required to maintain the metabolic activities of the cell. The *shell* of the soma is known as the **plasma membrane**. Additionally, there are some parts unique to neurons which enable them to function as the signal processors of the brain. These are shown in figure 2.1 and are listed below [6]:

- **Dendrites** are thin fibres extending from the soma, and are often highly branched so as to form a dense network known as a dendritic tree. The primary function of the dendrite is to receive and integrate information.
- An **axon** is a thin, tube-like fibre that originates from the soma and extends for distances ranging from microns to meters before branching out into

strands and sub-strands, each of which terminates at a *synapse* and is known as a *synaptic terminal* [6]. The main function of the axon is to propagate electrical signals from the soma to the synaptic terminals.

- An **Axon Hillock** is a specialised cone-shaped region adjacent to the exit point of the axon from the cell body. It is the site of generation of an Action Potential in spiking neurons.

- A **synapse** is a highly specialised structure to carry out the task of inter-cellular information transfer. The neuron which transmits information from its axon terminals is termed as **pre-synaptic**, while the neuron receiving the signal at its dendritic terminals is termed as **post synaptic**. The extracellular space separating the pre-synaptic and post-synaptic junctions is called a **synaptic cleft** and is 20–40nm in width. A synapse ensures *unidirectional* information transfer *from* the axon of the pre-synaptic cell *to* the dendrites of the post-synaptic cell. Synapses are further discussed in section 2.2.2.

2.1.2 Transmembrane potential of a neuron

The plasma membrane of a neuron acts as an impermeable barrier to ions found in the fluids inside and outside the membrane. However, the membrane has some pores that selectively allow certain ions to cross over from one side of the membrane to the other, and are commonly referred to as *ionic channels* in the literature [85]. These channels are characterised by the set of ions that can pass through them e.g. Potassium channels are those which allow movement of Potassium ions but are impermeable to any other type of ion. The main contributors to the extracellular charge are Sodium (Na^+), Calcium (Ca^{2+}) and Chlorine (Cl^-) ions, which exist in much higher concentration than in the cell interior. The cell interior has a higher concentration of Potassium (K^+) ions than that in its exterior, and a group of anions, denoted as A^-. The ionic channel is impermeable to A^- at all times. Because of the unequal distribution of the ions, there exists an electrical gradient as well as a concentration gradient across the plasma membrane. The electrical gradient is also known as transmembrane potential. The permeability of some of the ion channels is regulated by the transmembrane potential. In this sense, the ionic channels are referred to as 'gates', which are open or closed depending on the potential fluctuations across the cell membrane [15].

In normal conditions, the K^+ ion channels are open, i.e. they are permeable to K^+ ions. Due to the higher concentration of K^+ ions in the cell interior, the ions try to diffuse through these channels to pass over to the outside of the plasma membrane. As the K^+ ions leave the cell, the cell interior gets more negatively charged than normal. The excess negative charges attract the leaving K^+ ions, and this action slows down the rate of K^+ ion diffusion. At a certain point, the diffusion of K^+ ions due to the concentration gradient is exactly offset by the attraction of the negative charges on K^+ ions. This is known as the point of equilibrium for K^+ ions. Similar equilibrium potentials exist for other ions also, so that in normal conditions, the cell is in a state of equilibrium, and the cell interior is at a negative potential of approximately $-70mV$ with respect to the cell exterior. Under such a state, the net current flow across the plasma membrane is zero and the neuron is said to be in a **state of rest**. Thus the equilibrium potential is also known as the **resting potential**.

2.2 Signal processing in a neuron

In this section, we discuss how a neuron reacts to incoming signals from the outside world or from other neurons by generating spikes (section 2.2.1), which propagate down the axon before being transmitted on to other neurons or muscles using synapses (section 2.2.2).

2.2.1 Action Potential

Action Potentials are voltage impulses and are also known as 'spikes' because of their shape. These are used for signal transfer over long distances, for example in the axon of a ganglion cell of the retina which runs from the frontmost layer of the retina, through its whole depth down the optic nerve (shown in figue 3.1) to the Lateral Geniculate Nucleus (LGN) in the brain. The neuron generating the Action Potential is known as the spiking neuron; how a neuron generates a spike is described below.

Mechanism of Action Potential

As mentioned in section 2.1.2, in a state of rest, the interior of a neuron is negatively charged with respect to its exterior, maintaining a resting potential of

-70mV. The cell membrane is impermeable to the Na^+ ions and the gates of the Na^+ ion channels are closed.

When a cell is stimulated by an external signal, the conductance of the Na^+ ion channels increases and they open, and the cell becomes permeable to Na^+ ions, the flow of the ions being from the cell exterior to the cell interior. This increases the total positive charge inside the cell, and the phenomenon is known as **depolarisation**.

At the same time, the K^+ channel conductance also increases by a margin than during the state of equilibrium, making the efflux of K^+ ions more than the influx of Na^+ ions. The overall result of these changes due to the application of an external stimulus is a change in the transmembrane potential of the neuron. As long as this change in the transmembrane potential is below 15mV, the neuron regains its resting potential on removal of the stimulus. This phenomenon of going back to the neutral state is known as **repolarisation**.

However, if the external stimulus causes the transmembrane potential to rise from its resting potential of -70mV to around -55mV, there is a sudden increase in the Na^+ ion channel conductance, resulting in a heavy influx of Na^+ ions into the cell. The rate at which Na^+ ions enter the cell exceeds the rate at which K^+ ions leave the cell; this further increases the depolarisation, which in turn leads to a further increase in the channel conductance causing more Na^+ ions to enter the cell. The process is thus regenerative, giving rise to a huge surge in transmembrane potential that rises to around $+50mV$ *within 1ms.* Such a voltage surge of small duration is called an *Action Potential* and the threshold transmembrane potential of around -55mV at which the Action Potential is *initiated* is termed the **threshold voltage**.

Following this sudden surge in transmembrane potential, the cell membrane starts repolarising. The Na^+ ion channels are closed, and the cell membrane once again becomes impermeable to the Na^+ ions. The K^+ gates however remain in a state of increased conductance even when the cell is repolarising. Thus, with the Na^+ ion channels closed and the increased K^+ current flow towards the cell exterior, the transmembrane potential rapidly moves towards equilibrium.

However, the K^+ ion channel conductance remains high for longer than is required to reach equilibrium, causing more K^+ ions to leave the cell than under equilibrium conditions. This results in the transmembrane potential shooting below the resting potential and the phenomenon is known as **hyperpolarisation**.

Following the hyperpolarisation state, the K^+ ion channel conductance falls and the neuron eventually returns to its state of equilibrium, and the Action Potential is completed. The neuron is now ready to generate a second spike in response to an appropriate stimulus [15].

Refractory state

The time *from* when the cell membrane crosses the threshold voltage and the Action Potential is initiated, *until* the Action Potential reaches its peak within an interval of 1ms, is known as the **absolute refractory** period [15]. Any other stimulus acting on the neuron during this time will fail to generate an Action Potential. The time from when the cell moves towards repolarisation until it comes back to its neutral state is known as the **relative refractory** period. During this time, a stimulus that is larger than normal can elicit an Action Potential in the neuron.

Spike propagation

From the Axon Hillock, the Action Potential propagates down the axonal membrane by continuous regeneration of the Action Potential towards the direction of motion [15]. This is because when a portion of the membrane undergoes depolarisation and generates an Action Potential, the portion of the membrane adjacent to it also gets depolarized, and in turn generates an Action Potential, and so on. The membrane adjacent to the depolarising membrane in a direction opposite to the impulse propagation will not be depolarised as it will still be in the refractory state during that time. Therefore, the propagation of the Action Potential is unidirectional. Also, regardless of the axon length, the signal strength is constant from when it was initiated to when it terminates at the synapses.

2.2.2 The synapse

A synapse is always initiated only when the presynaptic neuron is depolarised. Depending on the method adopted by the synaptic junction to transmit a signal, synapses can be classified as either *electrical* or *chemical* synapses. The electrical synaptic cleft is commonly referred to as the **gap junction**. Intercellular communication through gap junctions is believed to be the simplest form of cell-to-cell interaction; small molecules and ions in one cell diffuse through pores

in the plasma membrane directly into the cell body of a neighbouring cell [15]. Chemical synapses are the more common synapses found in the brain, where a presynaptic neuron signal to the post-synaptic neurons by releasing certain molecules called neurotransmitters in the synaptic cleft. When these neurotransmitters reach the post-synaptic dendritic terminals, they cause the opening or closing of ionic channels in the post-synaptic cell membrane, which in turn causes a change in the transmembrane potential of the post-synaptic cell. If this change in transmembrane potential causes depolarisation in the post-synaptic cell, the synapse is excitatory and the change in potential is known as an **Excitatory Post-Synaptic Potential** (EPSP). Since a synapse is always initiated by a pre-synaptic terminal *depolarisation*, the excitatory synapse causing the post-synaptic cell to depolarise is also referred to in the literature as a *sign-conserving synapse.* On the other hand, a *hyperpolarisation* at the post-synaptic terminal is known as an **Inhibitory Post-Synaptic Potential** (IPSP) and the synapse is said to be inhibitory or *sign-inverting.* All synapses in the brain are either excitatory or inhibitory, and at every site of synaptic transmission, either an IPSP or an EPSP is elicited [85, 15].

There can be very high convergence of input information from hundreds or thousands of pre-synaptic neurons on to a single post-synaptic neuron. A pre-synaptic neuron, on the other hand, may synapse on up to hundreds of post-synaptic neurons due to its axonal branching [6]. On its own, an EPSP or an IPSP at a synaptic junction is too small to generate an Action Potential. But when all the potential changes caused by all the synapses at the dendritic terminals of the postsynaptic cell are summed, the transmembrane potential of the post-synaptic cell may change considerably so as to cross the threshold voltage and result in an Action Potential being generated. In such a case, the net effect of the potential changes at the input terminals of the post-synaptic cell is said to be excitatory, and the cell is said to have 'fired' (a 'spike'). If the net effect is inhibitory, the cell membrane potential falls below the resting potential and the cell is restricted from firing a spike; the cell is said to be inhibited. For non-spiking neurons, the axons are shorter, and there is no generation of spike. However, the mechanism of synaptic transmission is the same and it is the change in transmembrane potential that is passed on to the post-synaptic cell dendritic terminals as an EPSP or an IPSP.

Thus far, we have discussed the basic structure of a neuron that plays a role

in the generation, propagation and transmission of a signal. In section 2.3, we discuss how a neuron encodes information about the outside world into sequences of Action Potentials. We review the existing literature on the various hypothesis about neural codes and discuss their pros and cons (section 2.3.1 to section 2.3.3). Finally we introduce rank-order codes as a hypothesis about neural encoding and discuss the possibility of their playing a role in sensory pathways involved in fast information transmission (section 2.3.4).

2.3 Language of spikes

Every sensory organ of sight, sound, smell, etc. communicates to the brain in the same 'language' — a pattern of spikes. One would be tempted to presume that there is a dictionary where every pattern of spikes conveys a distinct meaning, yet there is no such dictionary. A neuron getting the same stimulus at different times generates different patterns. What then would be the 'language of the spikes', and how is information about the sensory world encoded in these spike patterns? This is a question which is yet to be answered in spite of ongoing research since the early twentieth century. Pioneering work in this regard was done by Adrian, where he proposed that a neuron encodes information about the external world in the frequency of occurrence of spikes in its spike train [11]. His observations were based on studies carried out on a sensory muscle of a frog which was connected to the sciatic nerve, such that any stretching or contraction in the muscle induces a response in the nerve. By stimulating the muscle with variable loads, he observed the patterns of spikes generated in the nerve. His observations are listed below:

- When there was no load on the muscle, there was no response from the nerve, as expected.
- When a load of 10 gm was put on the muscle for 10 sec, a train of voltage spikes was seen.
- When the load was maintained at 10gm but applied for 20 seconds; the number of spikes in the train increased.
- When the load was increased in steps up to 40gm and applied for varying times, for each increment in either load or time the number of spikes increased.

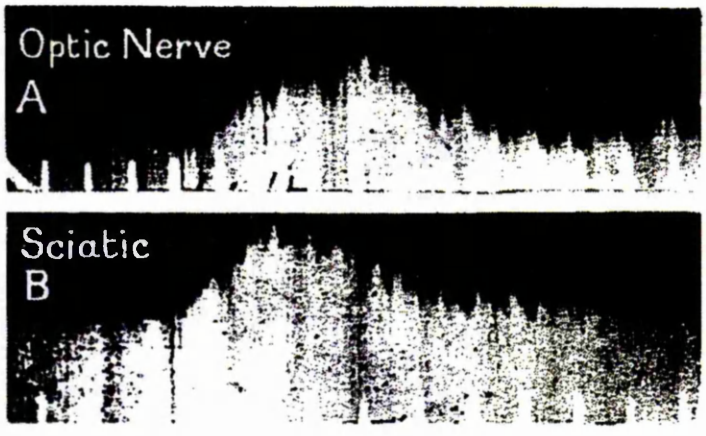


Figure 2.2: Neural response from (A) the optic nerve of a conger eel and (B) the sciatic nerve of a frog as recorded by E. D. Adrian [11].

Adrian performed similar pioneering experiments on the optic nerve of a conger eel with light as the input stimulus [11]. The records are of many fibres together, but they show well the general pattern of activity. In darkness, there is "a complete or almost complete absence of electric responses in the nerve". On illumination, a discharge which is irregularly oscillatory in nature and very similar to that of the sciatic nerve is observed and is shown in figure 2.2. Thus he observed that the message sent from the visual receptor through the optic nerve does not differ much from that sent out by the muscle through the sciatic nerve.

To summarise, Adrian observed that sensory messages travel to the brain at the onset of an external stimulus such as a muscle stretching for the sciatic nerve or an exposure to light for the optic nerve. The message consists of a succession of impulses. Although the *size* and *shape* of the impulses *do not vary*, the *frequency* with which they travel *depends* on the *strength of the stimulus*. Thus, a neuron encodes information about a stimulus by varying its rate of firing in response to the stimulus. This method of neural encoding came to be termed as the *rate code theory* and is defined below.

2.3.1 Rate code

Let a neuron be stimulated by an external stimulus, and its response be studied over a time window T. If N is the total number of spikes fired by the neuron in

this time window T, then the firing rate R of the neuron is:

$$R = \frac{N}{T}.$$

However, in certain sensory pathways, the information transmission from the sensory neurons to higher centres of the brain is very fast. To simulate such fast information transfer, the above method of spike rate calculation may be optimised by considering only two consecutive arriving spikes. In such a case, the firing frequency R will be $R = \frac{1}{t_{isi}}$, t_{isi} being the inter-spike interval.

Since single neuronal responses are typically random and noisy, a probabilistic treatment is more common, where the average firing rate over a number of *trials* (defined below) is considered. With passage of time, the theory has acquired different definitions depending upon the different averaging procedures [13]. Two methods commonly used in determining the neuronal response are [52]:

- Average firing rate: A neuron is subjected to the same stimulus more than once, each such occasion being termed a *trial*. The rate of firing in each trial is calculated. The firing rate of the neuron is expressed as an average of the average rate of firing R at each trial for the total number of trials.

- Time dependent firing rate: The stimulus in the external environment is mostly dynamic and changes with time. Since the firing frequency of a neuron varies according to the strength of a stimulus, a more practical approach is considered to be one which calculates the firing rate of a neuron over a number of trials in a short time window, and this in turn is done for several other time windows. The result is expressed as a post-stimulus time histogram, which gives the probability of occurrence of a spike in a certain time window in response to a certain stimulus.

A *Poisson model* is commonly used to simulate the firing rate model of a neuron. The model assumes that each spike in a time bin is independent of the past activities of the neuron. Although the model is not realistic, it is a good descriptive model as it can be used to generate spikes with exactly the same time dependent firing rate as the real data [52].

Based on evidence from ongoing research on neural encoding techniques it is speculated that in some parts of the brain, instead of a single neuron carrying all the message in its spike train response, spike trains from an *ensemble of*

neurons carry significant information about the stimulus [13]. Such a combined response by a population of neurons in encoding information is termed as the **populational code**. Several models exist that represent neuronal encoding using population codes, as well as several which are being suggested as a result of ongoing active research in this area. The models use firing rate as the basic spiking currency.

Even though firing rate code is an established neural encoding theory, there are other theories proposed, the most prolific of those being the concept of spike times. There is a strong lobby of scientists who believe that in certain parts of the brain, the population of neurons encode information in their relative spike timing. It is speculated that spike times may have important implications on the message conveyed by neurons. The bat auditory system and primate visual system are the two most cited areas [73].

In the following sections, we discuss several *temporal codes* (neural codes that are based on spike times) proposed as alternatives to rate codes. We start the discussion with the severe time constraints observed by subjective empirical tests on the primate visual system, which has led to these alternative theories.

2.3.2 Time constraints in vision

Pioneering work on the time taken for perception of a complex scene by humans was done during the 1970s. A novel technique of Rapid Sequential Visual Presentation (RSVP) using natural images was introduced whereby sequences of 16 colour images were presented to human subjects for 113ms per image [76]. The subjects were asked to press a button whenever they saw a certain picture, e.g. , 'a bat and a ball', placed randomly in the sequence. The results showed that 60% of the time the subjects were able to correctly respond to the target image within this 113ms. This near instantaneous identification of objects by human vision encouraged further work on this area.

Thorpe et al used event-related potential (ERP) for the first time to record response times of subjects in a scene categorization task [74]. Subjects were presented with images, about which they had no *a priori* information, at the rate of 20ms per image. They were then asked to identify if a picture contained an animal or not. This method was termed a 'go/no-go categorisation'. The results showed that approximately 95% of the subjects responded correctly with a mean

reaction time of 445ms, with a bias towards correct responses starting from 280–290ms after stimulus onset. Similar experiments on rhesus monkeys show even faster reactions of 250ms, with a bias towards correct response starting from 200ms. Categorisation tasks on other natural images such as 'food versus non-food' and 'trees versus non-trees' could be performed by trained monkeys within 250ms [72].

However, these reaction times involve (i) latency after stimulus onset, (ii) time taken by the signal to travel from the retina to the brain through the *afferent* nerves, (iii) decision making and (iv) time taken by the message to travel from the brain to the motor neurons through the *efferent* nerves. Thus the effective time for visual processing alone, which is the time needed in (ii) above, will be much less than 250ms [72]. ERP recordings, which reveal the signs of neural processing well before the motor output, showed that **visual processing** in the above task can be achieved **in under 150ms** [80, 72].

Later, Fabre-Thorpe et al used the same procedure as above to test human subjects after training them with images that they would be showed for the trials [19]. Each subject was trained for 14 days with a subset of randomly selected images. Results showed that although such a 'familiarity training' improved the accuracy and overall reaction time, there is no improvement in the initial response time. Evidence of better performance in the trained case was observed only after 400ms. Thus, it can be said that there was no significant improvement in visual processing times even when subjects were looking at a data-set about which they had 'some' *a priori* knowledge. This specific mode of biological visual processing that cannot be improved in spite of extensive training was termed Ultra-Rapid Visual Categorisation (URVC) [74].

In the results presented above, the target was of high biological relevance to the behaviour of the subjects. Animals, food and trees have been present in the environment of primates for millions of years and it might be possible that sensory pathways are hard-wired to perceive such objects faster [80]. Thus the results above cannot be generalised to categorisation of all kinds of complex visual scenes. On such grounds, VanRullen and Thorpe performed URVC experiments on human subjects with coloured images of 'means of transport' as target objects [80]. Further, to compare the results of this task with 'natural' targets, they alternated these images with those of animals.

The results show that the percentage of correct responses in both tasks of

identifying natural and artificial targets was 95%, while the median reaction time was 350 ms. The primary demonstration of such a result is that visual categorisation of artificial objects in complex scenes is neither slower nor less accurate than categorisation of natural images. This implies that, contrary to speculation prior to the experiment, the time for visual categorisation of complex scenes can indeed be generalised, and that there is no prior 'hard-wiring' in favour of natural objects in the visual pathway. It can be reiterated here that the reaction times obtained in all of the above experiments include not only the visual processing time but also the latency to stimulus onset, the time for decision making, and the time to motor output.

The reaction times obtained from the above experiments provide severe constraints on the visual processing time in primates, considering that even if the visual information collected by the photo-receptors in the retina takes the shortest route to the brain, it will have to cross at least 10 synaptic stages [76]. Further, the response latency increases at each step through the pathway from the retina to the cortex, with the approximate shift in latency being 10ms at each consecutive stage [72]. Such data is also supported by the experimental results of ERP showing less than 150ms as the time for visual processing, which will allow roughly 10ms delay at each synaptic stage, presuming the shortest feed-forward path from retina to cortex. Again, it is not possible to rule out the use of feedback loops even though the use of feed-forward loops seems to be critical and more widely used by primates in early visual processing [19]. But then, a feedback loop processing time within 150ms will impose further constraints on the time of processing at each synaptic stage. On the other hand, neurophysiological data suggest that cortical neurons rarely fire at rates above 100 spikes/sec. This implies that within a time window of 10ms, a neuron will fire at most 1 spike, or will not fire at all. But to estimate the firing rate, at least 2 spikes are required in a time window . Thus, rate codes fail to provide an explanation of the ultra-rapid early visual processing times [76, 73].

However, there are situations where the use of rate coding would seem to be the best choice for the nervous system, e.g. the amount of force developed by a muscle will depend on the firing rate of the motor neurons. Yet, it looks highly unlikely, due to the various reasons mentioned above, that rapid sensory processing can be achieved using rate coding. Such observations prompted the need of alternative coding schemes that are perhaps being used in conjunction

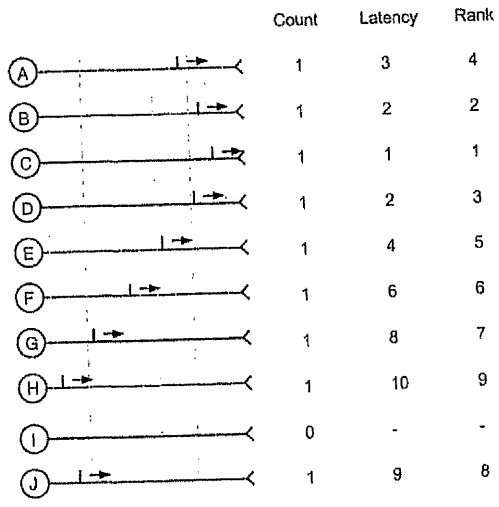


Figure 2.3: A population of 10 neurons that are subject to a stimulus. The information contained in the spike timings are encoded using the count, latency and rank-order codes [73].

with rate codes. For quite some time, it was speculated that spike times could be used by the neurons to encode information. Thus, temporal codes were proposed which opened up a broad scope for different novel coding schemes, some of which are discussed in the following sections.

2.3.3 Temporal codes

In this section, we discuss some of the viable options of coding using *spike times* as the basic coding currency as suggested by research in this area [73].

Let a stimulus be presented to a population of 10 neurons, and the response of each observed during a 10ms time window as shown in figure 2.3. Three of the different ways in which this population could encode and transmit information are [73]:

- **Count code**

  This is a population rate code as discussed in section 2.3.1. Since 9 of the 10 neurons fire during the 10ms time window, the rate will be 0.9 spikes/10ms, which is 90 spikes/s. Since the total number of possible states is 11 in this case, the amount of information that can be transmitted using 10 neurons will be $log_2(11) = 3.46$ bits. Thus, for N neurons, it will be $log_2(N + 1)$

bits.

- **Latency code**

  The maximum information of all the coding schemes proposed so far is obtained if the precise latency of each spike can be determined for each neuron in a population over the 10ms time window. For example if the spikes can be timed at a precision of 1ms, the maximum amount of information that could be transmitted in 10ms by a population of 10 neurons would be 10 * $log_2(10) = 33.2$ bits of information. Thus, for N neurons in a population over a time window of t ms, the information will be $N * log_2(t)$ bits. The drawback of this code is that it would be very difficult for an experimenter to determine the precise latency of firing in real neurons.

- **Relative latency code : Rank-order**

  The drawback of the above method may be removed if the relative latency of the spikes in a certain time window is considered, as shown in figure 2.3. This is the same as determining the rank of each firing neuron in a population. The code is termed as rank-order code. The order of firing of the cells is C, then B, D, A, E, F, G, J and finally H. For different stimulus intensity distributions, there could be 10! different ways in which the 10 cells can be rank-ordered. Thus the maximum information that can be transmitted using 10 neurons is $log_2(10!) = 21.79$ bits. For N neurons, $log_2(N!)$ bits of information can be transmitted.

Rank-order codes are discussed in more details in the following section, as abstracted from the works of Thorpe et al [75, 73].

2.3.4 Rank-order codes

Some sensory systems use the relative difference in stimuli arrival times or stimuli intensity at different receptors to process information [73]. For example we understand that a sound is coming from the left because the sound reaches the left ear before it reaches the right ear. In vision, the spiking neurons in the retina, the ganglion cells, have an onset latency that depends on the contrast[1] in the input stimulus, thus making our visual perception to be 'luminance invariant'.

[1] variation in the intensity of light about the mean luminance.

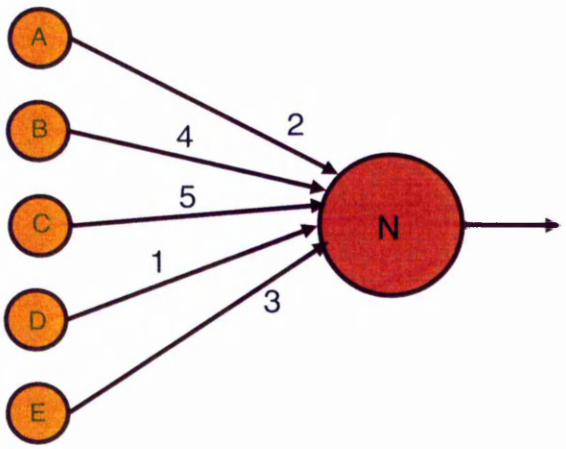


Figure 2.4: Decoding rank-order codes. The neurons A, B, C, D, E are pre-synaptic to the spiking neuron N.

The neurons corresponding to different intensity points in the input stimulus fire with different latencies; the one firing the earliest corresponding to the point of maximum intensity in the input stimulus.

Thorpe and Gautrais proposed that a simple way to encode this asynchrony in the firing times among the different neurons in a population is to use the order in which the neurons spike. The exact latency at which a neuron fires is not critical here. Rather, the relative rank of each spiking neuron in a population is important. Because of the intensity-to-latency transformation, the neurons will tend to generate spikes in an order which corresponds to the distribution of the applied stimulus. They termed this as the 'rank-order code'.

With respect to visual stimulus, although the absolute latency of the first spike of each neuron within a time window will vary with both luminance and contrast, there will be no change in the rank-ordering of the relative spiking latency. Thus, there will be an automatic normalisation of the inputs. The rank-order code is thus insensitive to variance in input luminance and contrast.

Decoding the codes.

The hypothesis of the rank-order code can only be tested if a rank-order encoded stimulus can be decoded. The purpose of decoding is multifarious, one of them being to enable the post-synaptic neurons to react to stimulus characteristics that are encoded in the firing patterns of the pre-synaptic neurons. Another

purpose, specific to application of rank-order codes in vision, is to be able to reconstruct the visual stimulus to evaluate the efficiency of the rank-order codes in encoding the stimulus. Moreover, the time to reconstruction and information retrieval from the rank-order encoded stimulus provides an estimation of the time to similar information recovery from a rank-ordered visual stimulus by neurons in subsequent heirarchy to the sensory neurons that are generating the codes. We will discuss more about rank-order codes with respect to vision in subsequent chapters of this thesis.

Basically, rank-order decoding is done by making a post-synaptic neuron sensitive to the order in which its pre-synaptic neurons fire, the sensitivity decreasing over time as a function of the rank of a spike [75]. One of the common methods cited in literature uses a decoding that has the sensitivity of the post-synaptic neuron decreasing monotonically in geometric progression [75, 73]. Let $\{n_{ai} \in N_a : i = 1 \ldots p\}$ be an ensemble of neurons which are pre-synaptic to a neuron n_f. Let the 'sensitivity factor' of the post-synaptic neuron n_f be a number $0 < S_{n_f} < 1$, and w_i be the weight of the connection between the post-synaptic neuron n_f and the i^{th} pre-synaptic neuron n_{ai} in the ensemble N_a, and $\{w_i \in W_n : i = 1 \ldots p\}$. The pre-synaptic neurons in N_a can fire in rank-order in $p!$ ways. The strength of activation produced as an output by the pre-synaptic neurons for a certain order of firing O_i is:

$$V_{O_i} = \sum_{i=1}^{p} S_{n_f}^{rank(n_{ai})} * w_i, \tag{2.1}$$

where $rank(n_{ai})$ is the firing rank of the neuron n_{ai} when the cells in N_a follow the firing order O_i. The magnitude of the activation will be maximum $V_{O_{max}}$ when the firing order O_{max} of the neurons in N_a is in the same as the order of the weights associated with them. Any other firing order will have a $V_{O_i} < V_{O_{max}}$.

The above decoding method may be described by an example shown in figure 2.4, where five pre-synaptic neurons $A - E$ make excitatory synapses on the post-synaptic neuron N, and have synaptic connections so that the order of firing $O_{max} = DAEBC$ will produce the maximal activation strength for N. If the sensitivity factor of the neuron N is 0.25, then:

$$V_{O_{max}} = (5 * 0.25^0) + (4 * 0.25^1) + (3 * 0.25^2) + (2 * 0.25^3) + (1 * 0.25^4) = 6.22$$

However, if the order of firing is $O_j = ABCDE$, the activation will be:

$$V_{O_j} = (4 * 0.25^0) + (2 * 0.25^1) + (1 * 0.25^2) + (5 * 0.25^3) + (3 * 0.25^4) = 4.61$$

The minimum activation will be produced for a firing order which is exactly reverse of O_{max}, i.e. $O_{min} = CBEAD$. Furthermore, by adjusting the threshold level of the post-synaptic neuron N, it can be made selective to the order of firing of the pre-synaptic neurons. In the case in figure 2.4, there are a total of $5! = 720$ ways in which the five pre-synaptic neurons can fire. By setting the threshold of N to 6.212, say, no ordering other than O_{max} can make N fire. Thus, the post-synaptic neuron can be made to be highly selective to the input stimulus features.

With respect to vision, a rank-ordered stimulus is decoded by assigning weights of decreasing magnitude to spikes firing later in time. The decreasing weights are observed to follow a power law up to the first 10% of the spikes (discussed in chapter 4). It has been demonstrated by Van Rullen and Thorpe that this mehod of rank-order decoding used for reconstruction of visual stimulus allows stimulus identification when only as few as 1% of the cells have fired their first spike [81]. Rank-order decoding with respect to vision is discussed in section 4.1.4.

2.3.5 Evidence of first-spike times

Recently, timing to first spike has been demonstrated to play a crucial role in whisker representation in the rat somatosensory cortex [53]. It is shown that many cells contain information in the timing of their spikes, most of this being in the time to first spike following each whisker deflection. Further, it was observed that patterns within the spike trains of cells corresponding to each whisker play a less important role. From such observations, it is inferred that the timing of the first spikes in the rat somatosensory cortex contributes to the coding of the spatially and behaviourally relevant features of the external environment. Recently, Johansson and Birznieks studied information conveyed by ensembles of afferent neurons about the direction of fingertip force and the shape of the surface contacting the fingertip [35]. They observed that the relative time of the first spikes evoked in a population of tactile afferent neurons contain rich and reliable information about the stimulus. Further, the information is transmitted much faster than it could be by rate codes. In another study made recently,

it was observed that retinal ganglion cells act as analog-to-delay units [79]. It was further observed that the first spike times make sense only with respect to a reference 'event', e.g. the first spike after stimulus onset. It is believed that the time of the very first spike following a certain 'event' is much more reliable than that of the following spikes in a spike train elicited by a certain sensory input.

2.4 Chapter summary

In this chapter we first reviewed the basic structure of a neuron, and made a study of its functioning in generating and propagating spikes. We reviewed the synaptic mechanisms of a neuron and then studied the encoding techniques used by the neurons to encode information from the external environment into trains of spikes. Further, we reviewed the literature on neural codes, and presented an analysis of their performance with relevance to the speed of information processing. Subsequently, we studied the rank-order codes proposed by Thorpe et al and presented a review of the encoding and decoding schemes.

Having reviewed neurons in general and some of the proposed neural encoding theories, we now move on to a review of the retina and its neurons. Our present work is based on the information processing by the spiking neurons in the retina, viz. the ganglion cells. In chapter 3, we study the structure and functioning of the neural and synaptic layers of the retina. This is followed by a review of the physiology and structure of the ganglion cells and their receptive fields in the primate retina. In chapter 8, we build a FOVEAL-PIT MODEL where we concentrate on simulating the ganglion cells corresponding to the cones of the foveal-pit in the primate retina. For the purpose of designing such a model, we review the foveal-pit and particularly the ganglion cell types and structures corresponding to the cones in the foveal-pit.

Chapter 3

Retina and Ganglion Cells

In the previous chapter, we discussed the basic structure of the neuron and how neurons communicate. As has been mentioned earlier, our present work is based on the *retinal model* by VanRullen and Thorpe. The model consists of several layers of *ganglion cells*, which are *spiking neurons* and the output cells of the retina. In this chapter we do a literature review on the retina, with special emphasis on the structure and functioning of the ganglion cells.

The retina is a thin transparent membrane forming an internal layer to the eyeball [65]. Due to its location, it is easily accessible for undertaking intracellular recordings to study and analyse neural mechanisms [90]. Further, the cells in the retina are arranged in distinct layers and in a regular fashion, which aids the study of the physiology and function of each cell type. Since the input to the retina is known, it is easy to understand what the cell outputs represent. Moreover, it is easy to stimulate the retina with a visual input, as this is the retina's natural stimulus [51]. It is for these and several other reasons that, for almost a century now, the retina has been a very popular area for scientists for carrying out histological analysis to investigate the principles of elementary information processing in the brain [85, 39, 25, 23].

We start by giving a brief introduction to the structure of the retina in section 3.1. We discuss briefly the several neural and synaptic layers in the retina that play a role in the passage of light from the photoreceptors to the ganglion cells. Such a discussion also helps in understanding the role of the other neurons in the retina in forming the basic structure of the ganglion cell receptive fields, discussed in section 3.3. Later in our work (chapter 8), we will describe a model of the foveal-pit in the retina, in anticipation of which we do a review

of the fovea and the arrangement of neurons in the foveal region in section 3.1.3. This is followed by a revisit of the ganglion cells in section 3.2 where we focus on the ganglion cells of the primate retina and discuss their morphology, layout and structure.

3.1 Structure of the retina

3.1.1 Retinal neural layers

A block diagram of the different layers of the retina is shown in figure 3.1. There are *five* main classes of neurons in the retina viz. **photoreceptor cells, horizontal cells, bipolar cells, amacrine cells** and **ganglion cells**. These cells are grouped into three distinct layers consisting of the cell bodies:

- The **outer nuclear layer** (ONL) contains the 'perikarya' [1] of the photoreceptor cells. The retina has two kinds of photoreceptors viz. **rods** and **cones**. Cones aid visual acuity, colour vision and vision in bright light. Rods are mainly responsible for vision in dim light, peripheral vision and motion detection. There are an estimated 6.3–6.8 million cones and 110–125 million rods in the retina [65, 85]. The density of rods and cones is not uniform across the retina and varies with retinal eccentricity. The cones show a peak spatial density at the foveal centre. For the purpose of this work, we ignore the rod and restrict our discussion to cone and its related pathways in the retina.

- **Horizontal cell** perikarya lie along the outer margin of the **inner nuclear layer** (INL). In the middle portion of the layer lie the **bipolar cell** perikarya, while the perikarya of **amacrine cells** lie along the inner margin. 80% of the total population of amacrine cells in the retina lie in this layer.

- The **ganglion cell layer** (GCL) consists of the perikarya of the **ganglion cells** as well as 20% of the total population of amacrine cells in the retina. Each axon of a ganglion cell is a nerve fibre. These fibres collectively form the optic nerve. Information collected by the dendrites of the ganglion cells is transmitted down the axon of the cells through the optic nerve on to

[1] cell body of a neuron, minus its synaptic terminals.

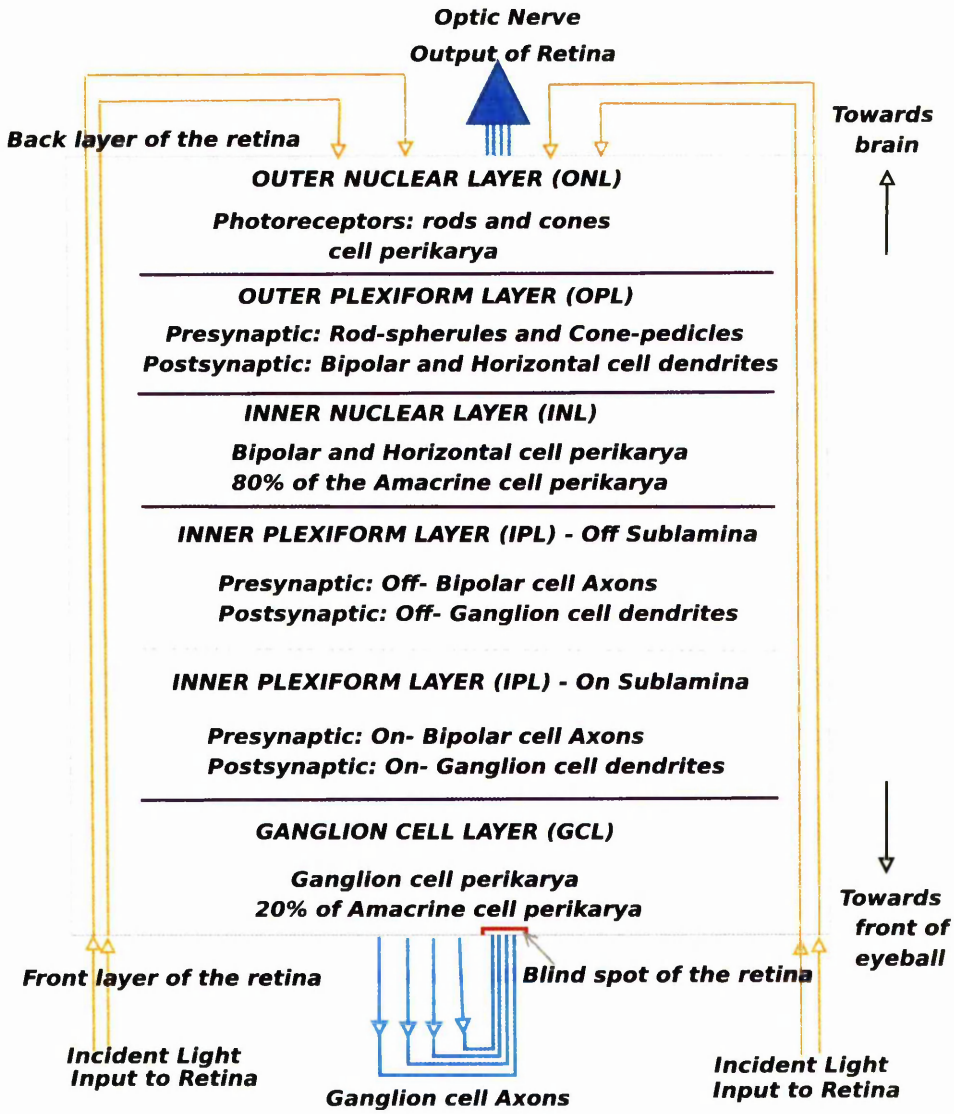


Figure 3.1: Layout of the neural and synaptic layers of the retina. The retina is the inner lining of the eyeball. Incident light passes through the thickness of the retina to stimulate the photoreceptor cells, which convert the light to electric signals. This transformed light energy then traverses the retina in the reverse direction, is processed at each step of its passage, and finally is transformed into electrical spikes by the ganglion cells, whose axons form the optic nerve. The optic nerve pierces the retina at the blindspot and travels through its entire depth before leaving the eyeball to transmit the information about the incident light to the brain.

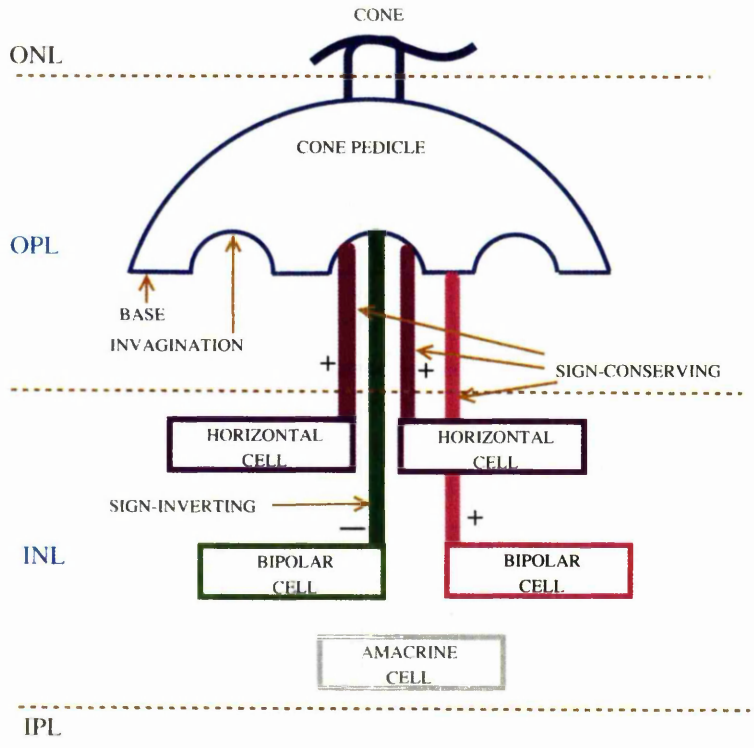


Figure 3.2: The cone-pedicle and its synapses to horizontal cell and bipolar cell dendrites in the OPL.

the brain in the form of electrical spikes. Thus, the ganglion cells form the output neurons of the retina.

We do a more detailed review of the ganglion cells, specifically the primate ganglion cells, in section 3.2.

3.1.2 Retinal synapses and synaptic layers

These three neural layers are interspersed with two plexiform layers constituting the sites of functional contact between the neurons as shown in figure 3.1 and described in the following sections.

Outer Plexiform Layer (OPL)

The **Outer Plexiform Layer** (OPL) consists of the connections between the photoreceptors in the ONL, which are the pre-synaptic neurons, and the bipolar and horizontal cells in the INL, which are the post-synaptic neurons. The synaptic terminal of a cone is known as **cone-pedicle** and it consists of multiple pocket-like openings at its base, which are known as *invaginations*, as shown in figure 3.2. Each of the invaginations of a cone-pedicle is associated with dendrites of horizontal cells that surround the dendrites of bipolar cells. The *invaginating* dendrites of the bipolar cells are known as invaginating processes and may be from the same or different cells in a single invagination. Similar for the horizontal cells. Bipolar cells which do not penetrate the invaginations are associated with the **base** of the cone-pedicle. We follow the behaviour of the synapses in the OPL with diagrams shown in figures 3.3 and 3.4. Cones hyperpolarize by the presence of light, i.e. when illuminated, and are said to be in OFF state, while they depolarize with an absence of light, i.e. in darkness, and are said to be in ON state [85], as shown in figure 3.3. A cone makes a sign-conserving synapse (defined in section 2.2.2) with the horizontal cell dendrites in its pedicles as shown in figure 3.2. Thus the invaginating horizontal cells connected to a cone-pedicle are hyperpolarised (OFF) by the presence of light and depolarised (ON) by the absence of light, shown in figure 3.3. A post-synaptic horizontal cell provides a *negative feedback* to its presynaptic cone, i.e. the horizontal cell is presynaptic in a sign-inverting manner to the cone from which it obtained its input. The exact mechanism and reason for the feedback connection is still a matter of debate [39, 85]. It is speculated that this feedback modulates the cone output so that it responds to the contrast rather than the wide variations of light intensity in our natural surroundings. In this manner, the horizontal cell output indirectly affects the bipolar cells that are fed by the cone, which in turn modulate their outputs to the ganglion and the amacrine cells in the IPL. Further, the horizontal cell is also known to be presynaptic in a sign-inverting manner to the bipolar cells in its vicinity, which has a direct effect on the bipolar cell outputs.

It is experimentally observed that *invaginating* bipolar cells make sign-inverting synapses with their respective cones (shown in figure 3.2) and are depolarised with an increase in light, [91, 85, 39]. These cells are called ON bipolar cells, and constitute the ON cone-pathway of light (figures 3.3 and 3.4). On the other hand, bipolar cells which are associated with cone-pedicle bases make sign-conserving

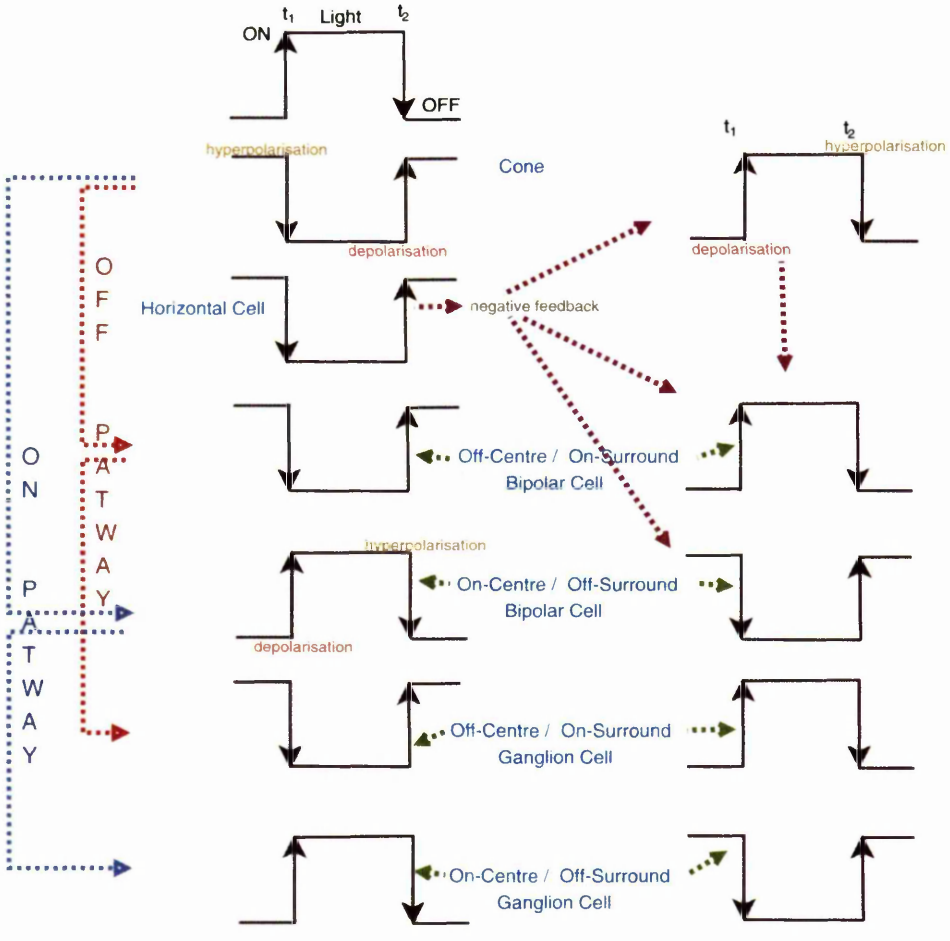


Figure 3.3: A simplified timing diagram of the effect of light on the various retinal neurons, and an antagonistic effect on the bipolar and ganglion cells due to direct and indirect negative feedback from the horizontal cells on to the bipolar cells. The ON-centre/OFF-surround and OFF-centre/ON-surround receptive field structures are discussed in section 3.3.1. The amacrine cells are not mentioned here but are shown in figures 3.4 and 3.8.

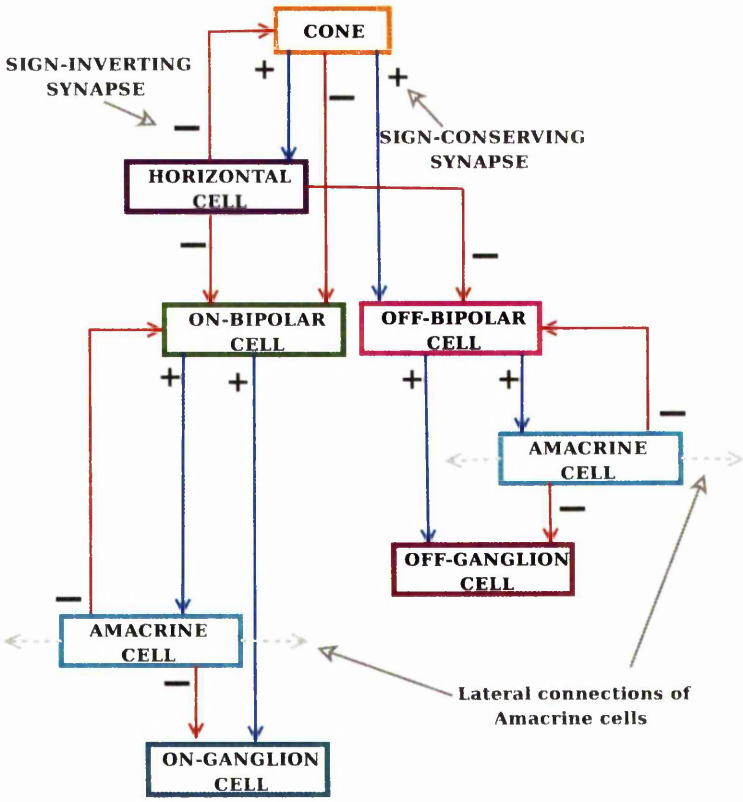


Figure 3.4: The synaptic connections in an ON- and OFF- cone pathway. The sign-conserving synaptic connections are shown with a '+' and a blue arrow while the sign-inverting connections are indicated by a '–' and a red arrow.

synapses with their respective cones, and are hyperpolarised by an increase in light. These are called the OFF bipolar cells and constitute the OFF cone-pathway of light [39].

Inner Plexiform Layer (IPL)

The **Inner Plexiform Layer** (IPL) consists of the connections between pre-synaptic bipolar cells and post-synaptic ganglion and amacrine cells. This layer has two distinct strata as shown in figure 3.1. The axons of ON and OFF bipolar cells terminate at different layers in the IPL, viz. the on-sublamina and the off-sublamina respectively, and are pre-synaptic to the ganglion cells. The bipolar cells make sign-conserving synapses to the ganglion cells, shown in figures 3.3 and 3.4. Thus, ON ganglion cells are post synaptic to ON bipolar cells and their dendrites arborise in the ON sublamina of the IPL. Similarly, the OFF ganglion cell dendrites arborise in the OFF sublamina of the IPL and are postsynaptic to the OFF bipolar cells. The behaviour of the ganglion cells to bipolar cell inputs is shown in figure 3.3.

The amacrine cells are postsynaptic in a sign-conserving manner to the bipolar cells, while they are presynaptic in a sign-inverting manner to the ganglion cells, shown in figure 3.4. The amacrine cells are also pre-synaptic to the bipolar cells in a sign-inverting manner, and thus provide a negative feedback. This modulates the direct input of the bipolar cell to the ganglion cell. Further, the amacrine cells also have sign-conserving lateral connections with other amacrine cells. The exact chemistry of the amacrine cell synapses and their varied functionalities are still in active research domain; much is now known about them, but much more still needs to be known [91, 90, 39, 25, 85].

In the following section we discuss the foveola, also known as the foveal pit, its structure and location with respect to the retina. This helps in forming a background to the discussion of the ganglion cells in section 3.2.

3.1.3 Foveal pit

At the centre of the retina on the posterior wall of the eyeball is a depressed area of 1.5mm diameter, which is responsible for the most distinct vision. The floor of the depression is termed the **foveola** or the **foveal pit**, and the sides of the depression are called the **clivus** or the **foveal slope**. The rim of the foveal slope

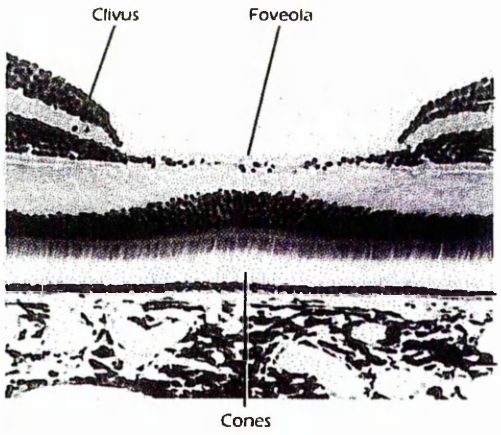


Figure 3.5: The foveal pit and the clivus. The outer rim of the foveal slope is termed the parafovea. All these regions constitute the fovea which is a part of the central retina [65].

is called the **parafovea**. Together, the whole area is termed the **fovea** [65].

The foveola is a circular region of about $400\mu m$ in diameter (1.4 degrees of visual angle[2]) shown in figure 3.5 [65]. It is devoid of any blood vessels, and the only photoreceptors available here are the cones. The retina is at its minimum thickness in this region. Such an arrangement provides a direct access of the photoreceptors to the incoming light. This is unlike elsewhere in the retina, where the light has to pass through all the other neural layers of the retina to reach the photoreceptors, since the retina is inside out. This helps make the fovea the region of highest visual acuity. The cone-pedicles connected to the cones of the foveola are displaced radially outwards from the foveal centre, so that they are closer to the capillaries of blood vessels surrounding the fovea to meet their nutritional requirements. The cone-pedicles of the cones within $300\mu m$ of the centre of the fovea are found to be displaced by about $350\mu m$, while the bipolar cells and ganglion cells connected to these pedicles are displaced by another $50\mu m$ [92].

The primate foveal cones are arranged in a very dense and fairly regular triangular mosaic. The minimum centre-to-centre spacing between cones in the human fovea is found to be $2.8–3.0\mu m$, while the maximum density of cones at the foveola is found to be $225,000/mm^2$ [92].

[2]1 degree of visual angle is equal to 288 microns on the retina [40].

3.2 Ganglion cells in the primate retina

The ganglion cells are spiking neurons and are the output cells of the retina. There are around 10–15 different morphological types of ganglion cell in the mammalian retina. The classification is based mainly on the size and the branching pattern of the dendritic trees [90]. Empirical data obtained from neuro-physiological experiments indicate that all the ganglion cell types provide complete coverage of the retina (except the blindspot[3]) with their dendritic trees. Consequently, a light spot projected onto the retina can stimulate at least one ganglion cell of any given type. Since the different types of cell process different aspects of the visual stimulus, such as contrast, size and movement, there is simultaneous processing of the stimulus by 10–15 parallel channels [91].

Pioneering work on the mammalian retina was done by Kuffler when he studied the characteristics of the ganglion cell receptive field of the cat retina [83]. Subsequently, ganglion cells in the retina of primates and non-primates such as mud-puppy, rabbit, frog, turtle, rat, were studied. The ganglion cell types across these retinas are observed to vary according to the food habits and survival strategies of the animal.

The main types of ganglion cell in the primate retina are the **midget** and the **parasol** ganglion cells which make up 80% and 10% of the total population of ganglion cells respectively. These two cells have been distinguished in the retinas of all primates studied so far. The very high spatial density of these two cell groups is the most distinctive feature of the primate retina. The following sections give a brief description of the morphology, spatial distribution and some functional properties of the midget and parasol ganglion cells of the primate retina. *The contents are abstracted from the published works of Dacey and others* [47, 41, 46].

3.2.1 Midget ganglion cells

Midget ganglion cells constitute 80% of the total ganglion cell population in the human retina. In the foveal region, they constitute around 95% of the ganglion cell population, whereas in the peripheral region, 45% of the ganglion cells are midget cells, as shown in figure 3.6. Compared to other types of ganglion cell in the

[3]The region where the optic nerve leaves the retina is devoid of any retinal photoreceptors, making the region insensitive to light (shown in figure 3.1).

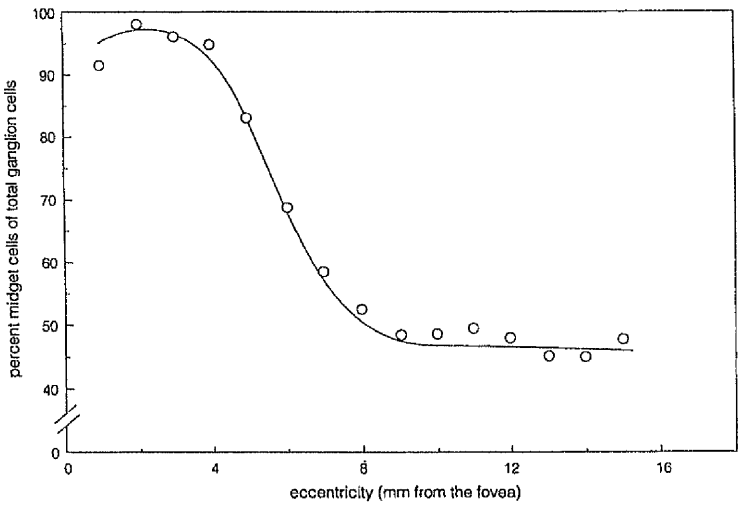


Figure 3.6: Percentage of midget cells at different retinal eccentricities [47].

primate retina, midget cells have the smallest dendritic field sizes. Consequently, they have a very high density in the central retina. They make up the dominant projection to the parvocellular layers of the lateral geniculate nucleus (LGN) in the brain and are also known as P cells. These cells are associated with colour vision and define the limit of spatial resolution across the visual field.

Depending on the layer of stratification of the dendritic trees, the midget cells are classified as ON- or OFF-centre cells. ON-centre cells branch out in the inner section of the IPL and receive their input from the ON midget bipolar cells, while the OFF-centre cells are post-synaptic to the OFF midget bipolar cells and stratify at the outer half of the IPL. At a certain retinal eccentricity, the ON-centre cells have an average dendritic field diameter 30% greater than their OFF-centre counterparts, and thus a greater dendritic overlap between the neighbouring cells. These characteristics are found to be consistent across the whole of the retina.

The ON- and OFF-centre cell mosaic are independent of one another because of their different layers of stratification. The arrangement of the cell bodies in each mosaic is highly regular, with neighbouring dendritic trees showing very little overlap in each independent mosaic. There is an increase in dendritic field size with increasing retinal eccentricity which is more or less matched by a decrease in spatial density, keeping the 'dendritic overlap' (defined as dendritic field area $\times$ cell density [46]) approximately constant over most of the retina. For example if dendritic overlap is to be maintained, then a reduction in dendritic field size

by a factor of 3 implies an increase in spatial density by a factor of 9. Thus evolutionary pressure for higher spatial acuity can result in a great increase in the number of cells that subserve this aspect of vision.

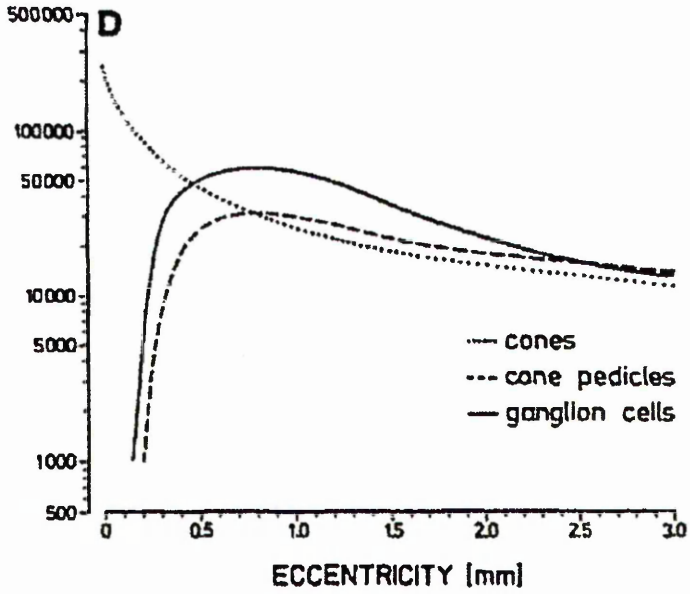
3.2.2 Parasol ganglion cells

The parasol ganglion cells constitute 10% of the total ganglion population in the primate retina. Also known as the M cells, the parasol ganglion cells project on to the magnocellular layer of the LGN.

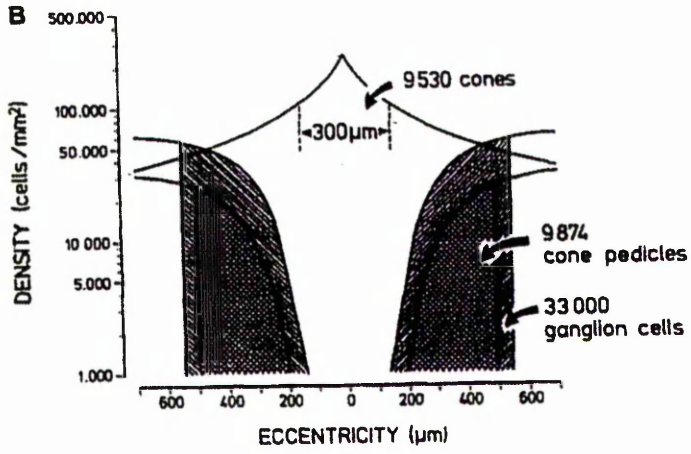
The mosaic of the ON-centre and the OFF-centre parasol ganglion cells are independent and they can be distinguished by their distinct stratification layers as with the midget cells. The ON-centre parasol cells have $30-50\%$ larger dendritic fields and thus a lower spatial density than those of the OFF-centre cells. This trait is found to be consistent at all retinal eccentricities. The dendritic fields of all the cells increase with eccentricity, but in midget cells this increase is steeper than that of the parasol cells. Consequently, the difference in size between the midget and parasol cells increases towards the fovea, thus maintaining a constant dendritic overlap, which appears to be a characteristic feature of the ganglion cell mosaics. At 50 degree retinal eccentricity, the ratio of the dendritic tree diameters of the parasol and the midget cells is recorded as $\sim 3:1$ while the ratio is $\sim 10:1$ at 3 degrees from the fovea. Thus, the corresponding ratio of dendritic field area will range from 9:1 to 100:1. The dendritic overlap for parasol cells in the monkey retina is recorded as 3.4, while that for the midget cells is 1, the same as in the human retina. Thus the ratio of the densities of the parasol to that of the midget cells ranges from 1:3 to 1:30 from the retinal periphery towards the fovea.

3.2.3 Ganglion cell to cone ratio

The data presented in this section is taken from a study carried out by Wassle et al on a region of up to 3mm eccentricity in the monkey retina to measure the densities of cone, cone-pedicles and ganglion cells in the region [92]. In a $25\mu m$ sampling window, the density of cones at the central fovea was found to be $250,000/mm^2$ as shown in the plot in figure 3.7(a). Since the retina is very thin at the foveal pit, cones packed in high density are the only neurons found in this region. The cone-pedicles corresponding to these cones are shifted by about $350\mu m$ away from the centre of the fovea. The location of the ganglion



(a)



(b)

Figure 3.7: (a) The plots show the density of cones, cone-pedicles and ganglion cells at different eccentricities in the central retina. (b) The displacement of the cone-pedicles and the ganglion cells corresponding to cones that lie within an eccentricity of $150\mu m$ radius of the centre of the fovea. Both the pictures are taken from [92].

cells connected to these cone-pedicles are further shifted by about $50\mu m$ due to oblique layout of intermediate bipolar cells and ganglion cell dendrites. In spite of this shift, the layout of the cone to the ganglion cell connection has a very orderly mapping. In a circular region of $150\mu m$ radius around the fovea, 9530 cones were found. Connected to these cones were approximately the same number of cone pedicles (9874) in a circle of $500\mu m$ radius ($150\mu m$ radius $+350\mu m$ shift) around the fovea. Within a further shift of $50\mu m$, i.e. in a circle of radius $550\mu m$ round the fovea, the total number of ganglion cells was found to be 33000. From this data, the ganglion cell to cone ratio is calculated to be $3.34 : 1$. This displacement is shown in figure 3.7(b). A smaller circle round the fovea with $50\mu m$ radius showed a ganglion cell to cone ratio of $4 : 1$. Thus, it is anticipated that the ratio may be even higher in the centre of the fovea.

3.3 Ganglion cell receptive field

The term **receptive field** was coined by Hartline, who first discovered them while experimenting on ganglion cells of the vertebrate retina. He defined receptive field of a ganglion cell as the area of the retina which must receive illumination in order to cause a discharge in the particular ganglion cell or nerve fibre [64]. Study on visual receptive fields in mammals was first done by Kuffler on the cat retina [83]. The physical dimension of the receptive field centre is defined by the dendritic field size of a ganglion cell [92]. The surround dimension, on the other hand, depends on the horizontal and amacrine cells and their synapses. The properties of the receptive field essentially determine the role that a ganglion cell might play in processing visual information.

Below, we present our interpretation of how the ganglion cell receptive fields are formed in the cone-pathway, based on factual data collected by reviewing various sources.

3.3.1 Receptive field structure

The receptive field of a cell is defined as the area on which light must fall to elicit response from the cell [39]. The receptive fields of cones are very narrow and react to light directly over them. Generally, there are no cones dedicated to particular bipolar cells, and the latter connect to the cone-pedicles of more than one neighbouring cone, exceptions being some bipolar cells in the midget pathway

of the primate fovea, which have a one-to-one connection with the corresponding cone. Thus the bipolar cells generally have a wider receptive field than the cones, as shown in figure 3.8. An ON-bipolar cell depolarises when light falls on cones that fall within this receptive field. Similarly, the OFF-bipolar cells hyperpolarise to incident light on any cone within its receptive field. It is observed that the horizontal cells connect to all cones within their dendritic field [91, 39]. Moreover, the horizontal cells also communicate laterally in a sign conserving manner through gap junctions, shown in figure 3.8, and are therefore affected indirectly by synapses from neighbouring cones. Those cones also form a part of the horizontal receptive field. Thus, the horizontal cells have very large receptive fields. The sign-inverting synapses made by a horizontal cell on the connected cones affect all the cones that lie within its receptive field due to the lateral connections. This feedback is fed forward by each cone to their respective bipolar cells. This has two implications on the bipolar cell receptive fields:

- Firstly, this provides an 'antagonistic' signal to the bipolar cells. This can be explained with an example thus — When a spatially uniform light is incident on a certain area encompassing some cones, the cones get hyperpolarised, as shown in figures 3.3 and 3.8. This makes the horizontal cells connected to these cones (directly and indirectly as explained above) hyperpolarized and the corresponding ON-bipolar cells depolarised, while the OFF-bipolar cells gets hyperpolarised. However, the cones get a sign-inverting feedback from the horizontal cells, which has the effect of reducing the hyperpolarisation in the cone. This 'negative' hyperpolarisation, when fed to the ON-bipolar cells in a sign-inverting manner, gives rise to a 'positive' hyperpolarisation signal. This affects the bipolar cell by reducing the depolarisation caused by the incident light on the cones on its receptive field. On the other hand, the 'negative' hyperpolarisation, when fed to the OFF-bipolar cells in a sign-conserving manner, gives rise to a 'negative' hyperpolarisation signal. This affects the OFF-bipolar cell by depolarising it partially so as to reduce the hyperpolarisation caused in it by the incident light. Thus, the feedback signal provides antagonistic inputs to the bipolar cells.

- Secondly, since the horizontal cells integrate light from a large area as mentioned above, the bipolar cell receptive fields have a centre-surround structure, where the surround is antagonistic to the centre. This can be explained thus — Let us refer to the cones which provide direct input to

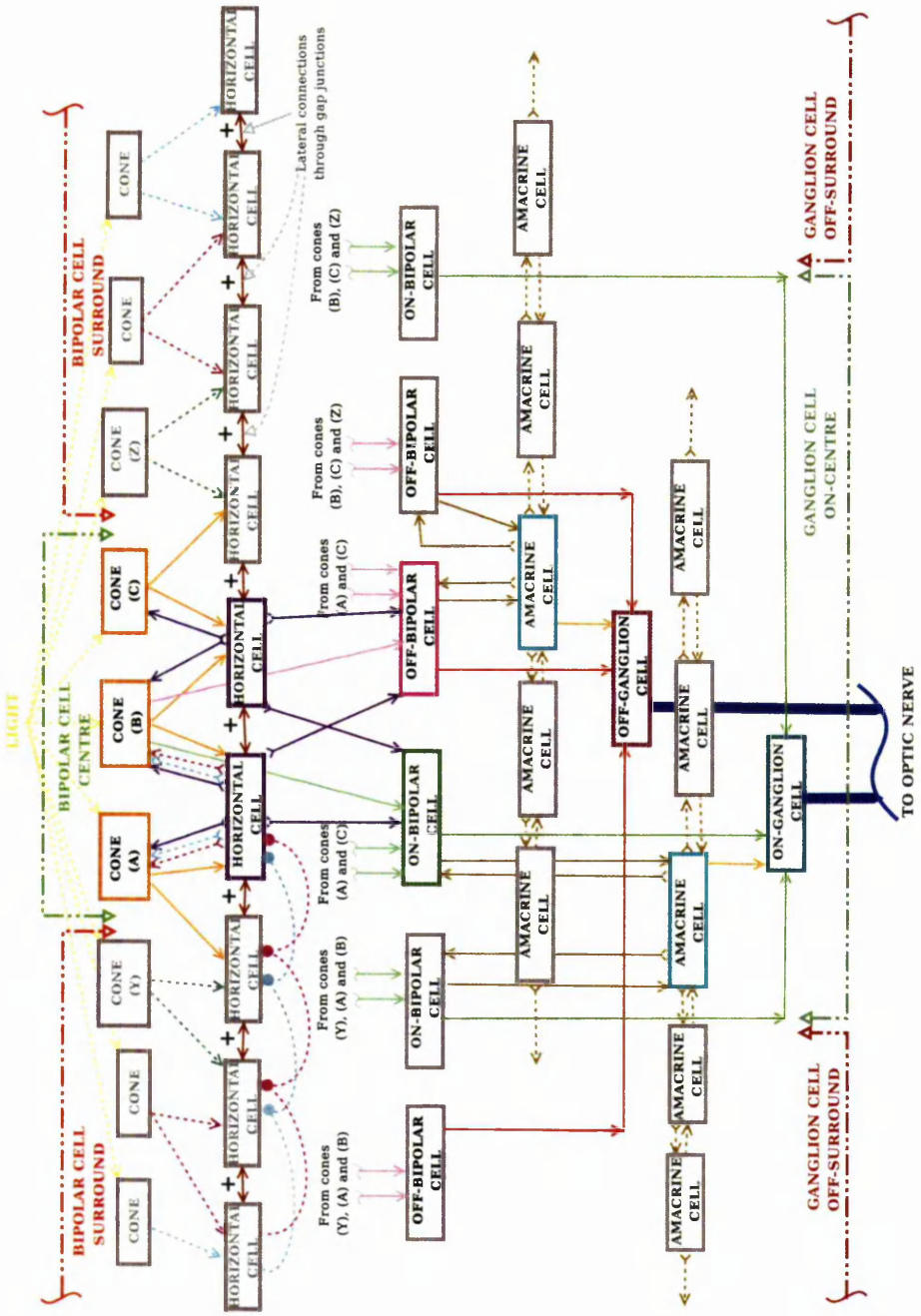


Figure 3.8: A simplified block diagram of the synaptic connections of the retinal neurons, and the formation of the centre-surround structure of the bipolar and ganglion cell receptive fields. The right hand side horizontal cell feedback connections from the surround on the cones in the centre are not shown explicitly for clarity of the diagram.

(a) (b)

Figure 3.9: (a) ON-center/OFF-surround structure, (b) OFF-centre/ON-surround structure of ganglion cell receptive fields.

the bipolar cells as 'the cones of the central area'. Since direct inputs of the cones of the central area to the bipolar cells are much stronger than their feedback signals, light falling on these cones depolarises the bipolar cells and the area covered by these cones is known as the ON-centre of the ON-bipolar cell receptive field. However, if light falls on the surrounding areas of these cones, the cones in these surrounding areas will antagonise the horizontal cells which are also connected to the cones of the central area (since horizontal cells have a much wider reach). These horizontal cells in turn provide a negative feedback to the cones of the central area. This will elicit a hyperpolarising response from the bipolar cells, which is 'antagonistic' to their normal behaviour. This surrounding area that elicits antagonistic response from the bipolar cells is called the OFF-surround of the ON-bipolar cell receptive field. Thus, the bipolar cell receptive fields have a **centre-surround structure**; the ON-centre cells have an ON-centre/OFF-surround receptive field (figure 3.9(a)), while the OFF-centre cells have an OFF-centre/ON-surround receptive field. (figure 3.9(b)).

The ganglion cell receptive fields are direct 'descendents' of the bipolar cells; the ON-ganglion cells inherit the ON-centre/OFF-surround structure of the ON-bipolar cells, while the OFF-ganglion cells inherit the OFF-centre/ON-surround structure of the OFF-bipolar cells. However, the centre and the surround of the ganglion cell receptive fields are wider than those of the corresponding bipolar cells. This is because the dendrites of the ganglion cells receive inputs from several bipolar cells. The combined effect of the receptive field centres of all these bipolar cells constitutes the centre of the ganglion cell receptive field. Again, each of these bipolar cells has a surround region, which adds up in constituting the receptive field surround of the ganglion cells. It is speculated that the inhibitory synapse from the amacrine cells on the ganglion cells helps in sharpening the antagonistic

surround of the ganglion cell receptive field [39]. Furthermore, due to the lateral synapses provided by amacrine cells upon themselves, the negative feedback of an amacrine cell on a ganglion cell is affected by bipolar cell signals from a much wider area, which further adds to the width of the receptive field surround of the ganglion cell, shown in figure 3.8.

The centre-surround structure of the receptive fields due to the negative feedback of the postsynaptic cells on presynaptic cells give rise to a very interesting phenomenon in the retina of the vertebrates that is termed 'lateral inhibition'. In section 3.4, we introduce this mechanism used by the sensory neurons to reduce data-redundancy in the input stimulus. We have later used this principle to rank-order encode visual stimulus efficiently (chapter 7). Prior to that, we discuss below some data regarding the actual dimensions of the ganglion cell receptive field centre and surround as found in the mammalian and primate retina, as a background to our work in chapter 8.

3.3.2 Receptive field centre

It has been observed that in the primate retina, the ratio of the average radius of the midget cell receptive field to the average radius of their respective dendritic trees is 1.65 at $0-10°$, 1.46 at $10-20°$, and 1.06 at $20-30°$ eccentricity. A possible reason for the receptive field being actually greater than the dendritic field of a ganglion cell is that these cells are not directly connected to the receptors. There are three layers of cells, the horizontal, bipolar and amacrine cells which might add to the spatial extent of ganglion cell dendrite tree to form the centre of the receptive field. As seen from the data above, this difference in size is greatest near the fovea and decreases with increasing eccentricity. A possible reason could be that, due to an increase in the inter-receptor space with increasing eccentricity, cells connected to these receptors may be less likely to contact one another via gap junctions. This could result in the narrowing of the area from which signals are drawn into bipolar cell, thus narrowing their receptive fields [89].

3.3.3 Receptive field surround

It is observed that there is a significant scatter in the surround sizes of neighbouring midget and parasol cells at a certain eccentricity in the primate retina. Also, unlike the central regions, the surrounds of both cell types cover the same

range of sizes. On average, the ratio of the radii of a midget cell dendritic tree to its receptive field centre will be 0.15, so that the surround of the cell is on average 6.7 times wider than the centre of the same cells, and is thus about 45 times larger in area. A similar ratio for the parasol cells is recorded as 0.21, so that the surround radius is on an average 4.8 times that of the centre, and is about 23 times larger in area [89].

3.4 Linking redundancy and lateral inhibition

In this section, we give a brief overview of the redundancy in our natural environment and how it is handled by the retinal cells using the lateral inhibition technique.

3.4.1 Redundancy reduction: goal of sensory processing

Gibson suggested that to understand the nature of visual perception, we must understand the nature of the environment around us. He suggested that the perceptual capabilities in an organism have evolved with the purpose of exploiting the laws of its surrounding environment in order to take decisions and act accordingly for its survival [34]. In other words, to understand fully the principles behind our visual perceptual mechanism and why it behaves the way it does, we need to understand the laws of the surrounding environment that is being sensed by our retinal cells.

Following such ideas, Attneave used Shannon's information theory to quantify the information in a visual stimulus [3]. He found that while the information carrying capacity of the optic nerve from the retina is constrained by the number of nerve fibres, yet the combination of the range of luminances in our visual environment is astronomical. At the same time, he demonstrated that certain visual stimuli are such that information obtained from them are highly redundant, i.e. portions of a stimulus are highly predictable from knowledge of its other portions. Any such physical invariance in a scene is a source of redundancy for the eye because the eye is capable of abstracting the invariance and extrapolating the redundant information. Homogeneous brightness, symmetry, a constantly changing contour, are all examples of redundant impact as these can be extrapolated. Thus, according to Attneave, "it appears likely that a major function of the perceptual machinery is to strip away some of the redundancy of stimulation,

to describe or encode incoming information in a form more economical than that in which it impinges on the receptors."

Continuing on such lines, Barlow postulated the redundancy-reducing hypothesis, whereby he claims that the sensory neurons code the input stimulus in a manner such that they preserve all the important information in the stimulus while relaying it to later stages and, yet, do it by the smallest average expenditure of energy [4]. This goal is achieved by discarding the redundancy in the incoming information and thus 'economising' the use of neural impulses in transmitting the information. He concludes by stating that the design of sensory organs reflect their usage in perceiving the external environment. We observe a reiteration of Gibson's idea in such a statement.

Following such views, independent research has been carried out on the statistics of natural images. In Field's attempt to provide an explanation of the relation between the statistics of natural images and the functioning of the cells in our sensory pathways, he observed that images from the natural environment, far from being random, show very strong correlation among neighbouring pixels [32]. Representation of this correlation in the spatial-frequency power spectrum shows that the amplitude of the Fourier coefficients falls with frequency, f, by a factor of $1/f$ approximately, indicating a scale invariance in natural images. The high degree of correlation in natural images gives rise to redundancy. So, in a region of uniform contrast, if the grey level of a certain pixel is known, that of the others in its near vicinity can be fairly estimated or predicted. Subjective tests of such predictive abilities in humans show that redundancy ranges from 46% for complex natural scenes to 74% for a human face [38].

We summarize the foregoing discussion thus: the environment around us is apparently chaotic and to represent it would be a very complex task for our sensory organs [28]. With the aid of Shannon's information theory, it could be deduced that the narrow dynamic range of the optic nerve fibre is incapable of dealing with the enormous luminance range of our visual environment. Clearly there is some sort of optimisation being done by the early stages in our visual pathway, so that all the visual information from the outside world is passed on without loss, yet with the expenditure of minimum neural resources [29, 28, 94, 27, 95]. This understanding gave rise to the speculation that the environment around us follows an organised structure, and it is to this structure that the sensory systems adapt [30, 2, 55]. The rest of the chaotic information in the environment

is unnecessary and redundant and the sensory organs ignore it. Independent research on the statistics of natural images has shown that, indeed, only a small subset of all the possible combinations of luminances in our natural environment actually makes sense to our perception [42, 43, 82, 1]. We have not learnt to recognise the rest of the set of luminance combinations, possibly because it is not required for our survival. In other words, the information contained in the natural environment around us is highly redundant, and it is up to the sensory organs to get rid of the redundancy, so that the information is passed on to higher levels with the greatest efficiency and using minimum resources [68, 62, 96, 17].

In the next section we talk about lateral inhibition, which is thought to be a means by which our sensory systems deals with the redundancy in the incoming information from the environment around us.

3.4.2 Lateral inhibition: means of redundancy reduction

While experimenting on the eyes of a horseshoe crab (Limulus), Hartline noticed that the sensory elements exert an influence on one another through lateral interconnections, the nature of such an interaction being purely inhibitory [36]. This means that the frequency of discharge of impulses in a randomly selected optic nerve fibre is either decreased or stopped by illuminating neighbouring areas of the eye. Because of its nature, the mechanism came to be known as lateral inhibition. Subsequent experiments by Hartline et al brought forward several characteristics of lateral inhibition [37, 5]. They observed that, firstly, the degree of lateral inhibition depends on the intensity of illumination of the inhibiting source, i.e., the greater the intensity, the greater is the degree of inhibition. Secondly, it depends on the location of the inhibitory source and decreases with increasing distance of the latter. Thirdly, simultaneous observation of the responses from two receptor units showed that nearby units inhibit one another mutually. This was termed as 'mutual inhibition' which is a special case of lateral inhibition. In later works, the mutual inhibition was described quantitatively by two simultaneous equations [63]:

$$r_1 = e_1 - K_{1,2}(r_2 - r^0_{1,2}) \tag{3.1}$$

$$r_2 = e_2 - K_{2,1}(r_1 - r^0_{2,1}) \tag{3.2}$$

where r is the response of a receptor unit and e is the excitation supplied by the external stimulus on the receptor. The second term on the right hand side of both equations represents the inhibitory influence of the second cell on the first in equation 3.1, and vice-versa in equation 3.2. K denotes the coefficient of inhibitory influence of one receptor on the other, and r^0 represents the threshold frequency that must be exceeded before one receptor can exert any inhibition on the other. So, $r^0_{1,2}$ is the frequency of receptor 2 at which it begins to inhibit receptor 1, and the reverse for $r^0_{2,1}$.

It was further observed that the inhibitory effect of several cells upon a certain cell is additive. This means that the total inhibition exerted on any one receptor by other receptors can be expressed as an arithmetic sum of the individual inhibition of each of the inhibiting receptors. This is expressed quantitatively as:

$$r_p = e_p - \sum_{j=1}^{n} K_{p,j}(r_j - r^0_{p,j}) \tag{3.3}$$

where $p = 1, 2, \ldots, n$; $j \neq p$; and $r_j \geq r^0_{p,j}$. The summation term on the right hand side of equation 3.3 represents the magnitude of the total inhibitory influence exerted on the excitation corresponding to the cell p.

The above results were indicative of lateral inhibition and how it works. Hartline et al concluded that an important effect of lateral inhibition is to enhance the visual contrast at regions of sharp spatial gradient and discontinuities in an image [37]. This leads to an accentuation of the information that is important to an organism, i.e. the edges, at the expense of less significant information, which are the regions of uniform luminance and can be predicted by the brain as discussed in the section 3.4.1. So, it can be said that lateral inhibition removes redundancy from a visual input by acting as a filter for enhancing the edges. This allows efficient encoding of the incoming information by the retinal neurons. This conclusion fits very well with the studies on the statistics of natural images and the nature of the processing in the early part of the visual pathway as discussed in section 3.4.1.

We summarise the discussion in sections 3.4.1 and 3.4.2 thus — Although the natural environment around us has a vast range of luminance levels, only a small subset of this is actually needed by a living organism to be able to survive. In other words, only a small part of this information is useful for our visual perception. The rest of the information is redundant. Years of evolution have

taken advantage of this redundancy by designing the retinal cells so that they can filter out the redundancy while encoding the visual information from the environment before passing it onto higher levels of visual processing. To achieve this, the retinal neurons use the lateral inhibition mechanism. Lateral inhibition enhances the edges in a scene and gets rid of redundancy in visual information. The message is thus coded in an optimal number of impulses. Thus it can be said that lateral inhibition is a means adopted by our visual system to adapt to the surrounding environment and strip the information collected from the surroundings of its highly redundant contents to pass on all the vital information to the brain using minimal available resources.

In chapter 7, we apply the lateral inhibition principles to VanRullen and Thorpe's retinal model for removing redundancy in the coefficients of filtering, introduced due to over-sampling of the input image.

3.5 Chapter summary

In this chapter, we have done a brief review on the different neural and synaptic layers of the retina and their inter-connections. A more detailed review on the ganglion cells and their receptive fields is done as a background to the validation of *VanRullen and Thorpe's retinal model* in chapter 4. For the designing of a more biologically realistic model of the primate retina, we have reviewed the midget and the parasol ganglion cells, the two primary types of ganglion cells found in the primate retina. We have focused our discussion on the size and density of these cells in the foveal region as a background to the FOVEAL-PIT MODEL which is discussed in chapter 8.

In the next chapter, we give a detailed account of how we have validated VanRullen and Thorpe's retinal model, along with the empirical results from the validation.

Chapter 4

Validating the Retinal Model

In previous chapters, we reviewed the physiology and functioning of the neurons, particularly the retinal spiking neurons, viz. the ganglion cells. We also discussed the several theories that have been proposed over time in an attempt to discover a dictionary that translates the stimulus from our external environment to individual and/or collective neuronal firing patterns. In this context, and as has already been discussed in section 2.3.4, rank-order codes seem to be a reasonable hypothesis in explaining the ultra-rapid speed with which the initial burst of spikes from the ganglion cells enable us to perceive our surroundings. VanRullen and Thorpe designed a simple model of the retina to test the performance of rank-order codes in encoding images [81]. This they called the *retinal model*, which has a layered organisation consisting of several ON- and OFF-centre ganglion cells of various sizes, each size tiling the retina uniformly in a separate layer. The results of their simulation show that a rank-order encoded image can be decoded with reasonable fidelity so as to enable the recognition of the object in the image by the time only 1% of the ganglion cells of the model retina have fired their first spikes [81].

In this chapter, we discuss the procedures followed in our validation of the retinal model of VanRullen and Thorpe. We present the empirical results of our simulation and analyse and discuss the implications of the results.

4.1 VanRullen and Thorpe's retinal model

In this section, we first discuss the simulation of the basic structure of the ON- and OFF-centre ganglion cell receptive fields. Next, we discuss the simulation of the different sizes of the cell receptive fields, and how a picture is processed with the

different layers of such cells. We then show how the processed image is encoded with rank-order codes to resemble a population of neurons asynchronously firing their first spikes. To test the (a) efficiency and (b) speed with which the rank-order codes can encode information about an image, the image is reconstructed progressively with each incoming spike and the results are evaluated with respect to the information content of the input image.

4.1.1 Simulation of the ganglion cells

Rodieck proposed that, for "mathematical simplicity", a "convenient function to choose" for defining the centre-surround structure of the ganglion cell receptive fields is the "sum of two Gaussian functions, a positive one and a wider negative one", as shown in figure 4.1 [84]. This function is commonly known as the *Difference of Gaussians* (*DoG*). Saying mathematically, if $\mathbf{x}$ is a two-dimensional vector defined as:

$$\mathbf{x} = (x, y), \text{ then}$$

$$\|\mathbf{x}\|^2 = x^2 + y^2,$$

and the equation for the DoG function is:

$$\Phi(\mathbf{x}) = \frac{1}{2\pi\sigma_1^2} exp\left[\frac{-\|\mathbf{x}\|^2}{2\sigma_1^2}\right] - \frac{1}{2\pi\sigma_2^2} exp\left[\frac{-\|\mathbf{x}\|^2}{2\sigma_2^2}\right], \tag{4.1}$$

where the first term represents the centre of the receptive field with σ_1 being the measure of its width, and the second term represents the surround with a width of σ_2. The factors $\pi\sigma_1^2$ and $\pi\sigma_2^2$ are used for normalisation.

In their model, VanRullen and Thorpe used Field's generalisation where the width of the surround is set at three times the width of the centre of the receptive field [33]. Thus, in equation 4.1, $\sigma_2 = 3\sigma_1$, so that the expression for the DoG that is used in the simulation is:

$$\Phi(\mathbf{x}) = \frac{1}{2\pi\sigma_2^2}\left(9exp\left[\frac{-\|\mathbf{x}\|^2}{2\sigma_1^2}\right] - exp\left[\frac{-\|\mathbf{x}\|^2}{2\sigma_2^2}\right]\right). \tag{4.2}$$

The salient points of the model design are as follows:

- Ganglion cells of eight different sizes tile the retina. Their receptive fields are represented by eight different sizes of DoG functions.

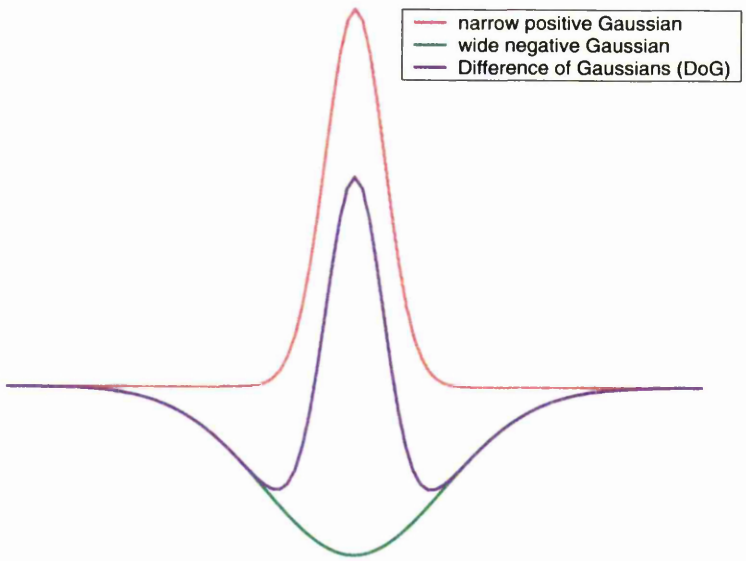


Figure 4.1: One dimensional Difference of Gaussians (DoG) function as a difference of a narrow positive Gaussian and a wide negative Gaussian function.

s	n	σ_1
1	5	0.5
2	11	1.0
3	23	2.0
4	47	4.0
5	95	8.0
6	191	16.0
7	383	32.0
8	767	64.0

Table 4.1: A table showing the size $n \times n$ of a matrix representing a DoG function at a particular scale $s \in \mathcal{S}$, where $\mathcal{S} = \{1, 2, \ldots, 8\}$. The standard deviation of the centre Gaussian of each DoG is given in the corresponding σ_1 column.

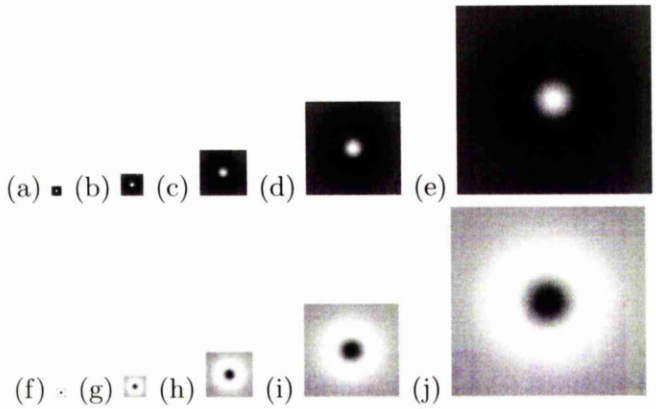


Figure 4.2: (a–e) scale 1 to scale 5 of ON-centre/OFF-surround DoG functions. (f–j) scale 1 to scale 5 of OFF-centre/ON-surround DoG functions.

- At each size, there are two types of DoG, each representing either an ON-centre or an OFF-centre ganglion cell receptive field.
- Each size and type of DoG function tiles the retina in independent layers. Thus there are a total of sixteen independent layers of ganglion cells.
- The resolution of tiling the retina at a certain layer decreases with increasing size of cell. Thus the number of DoG functions at a larger scale will be fewer than that at a smaller scale.

A total of eight different sizes of ganglion cell receptive field are simulated by scaling up the smallest DoG by powers of 2 as tabulated in table 4.1. At each scale $s \in \mathcal{S}$, where $\mathcal{S} = \{1, 2, \ldots, 8\}$, the size of the matrix representing the DoG function is $n \times n$, where $n = 3 \times 2^s - 1$. Thus, the DoG function at scale $s = 1$ has a matrix size of 5×5 with a standard deviation of the centre Gaussian as $\sigma_1 = 0.5$. This function corresponds to the the smallest size of both ON- and OFF-centre ganglion cells as shown in figure 4.2 (a) and (f). Five sizes of DoG and two types at each size are shown in figure 4.2 (a) to (j) (The remaining three larger sizes are not shown here due to space constraints). The first layer of ganglion cells is a uniform tiling by cells of size and type shown in figure 4.2 (a), while the second layer is a uniform tiling of cell type and size shown in figure 4.2 (f). The third and the fourth layers are tiled by the size and type of cells shown in figures 4.2 (b) and (g) respectively, and so on. All the DoG matrices are normalised such that the convolution of a DoG at a certain scale s, referred to in this thesis as

Φ_s, with itself is unity:

$$\langle \Phi_s, \Phi_s \rangle = \sum_{\mathbf{x} \in \mathbb{Z}^2} \Phi_s(\mathbf{x}) \Phi_s(\mathbf{x}) = 1 \tag{4.3}$$

4.1.2 Simulation of image processing in retina

Once the ganglion cell receptive fields are simulated, the retinal model is ready to process an input image. To simulate the activation of a set of ganglion cells with the contrast at each point of an input stimulus, an input image $\mathcal{I}$ of size $m \times p$ is *filtered* with all sixteen layers of DoG functions. *Filtering* refers to the *convolution* of the image with the ON- and OFF-centre DoG functions, such that convolution by placing Φ_s at a certain pixel at spatial location (k, l) of the picture generates a coefficient $C_s(k, l)$ of filtering given by:

$$C_s(k, l) = \sum_{x=1}^{n} \sum_{y=1}^{n} \mathcal{I}(k - x)(l - y).\Phi_s(x, y) \tag{4.4}$$

As mentioned in section 4.1.1, the resolution of filtering is decreased with increasing scale of DoG function. At a scale s, Φ_s is placed at every 2^{s-1} pixels on the image. This makes the number of coefficients of filtering at scale s a quarter of that at scale $s - 1$. For the image $\mathcal{I}$ of size $m \times p$, the total number of coefficients of filtering will be

$$M = 2mp \sum_{s=1}^{8} \frac{1}{4^{s-1}} = \frac{8}{3} mp = \frac{8}{3} N,$$

where $N = mp$ represents the total number of pixels in the image. This implies that a DoG function of eight sizes, with two types at each size, is placed on various pixels of the images a total of M times. As each DoG represents a ganglion cell, the image can be thought of as being processed by a total of M ganglion cells. In our validation, we have used images of size 128×128 so that the number of coefficients generated on filtering an image is 43,691 (rounded). A data-set of sixty-five images, of both natural objects in our immediate surroundings and man-made structures, are used to evaluate the performance of the codes in *simulating* the *rapid visual processing of the retina.* Some of the images from our data-set are shown in figure 4.3.

Figure 4.3: A sub-set of the images used during the various experiments described in this thesis.

4.1.3 Rank-order encoding

The magnitude of each coefficient $C_s(k,l)$ in equation 4.4 corresponds to the strength of activation of a cell produced by the contrast at location (k,l) of $\mathcal{I}$ with respect to its neighbourhood, such that the larger of two coefficients corresponds to a cell which fires earlier than that corresponding to the smaller coefficient. In other words, $C_{s_1}(k,l) > C_{s_2}(k+1,l)$ (say) implies that the cell corresponding to $\Phi_{s_1}(k,l)$ fires earlier than that corresponding to $\Phi_{s_2}(k+1,l)$, where s_1 and s_2 may be the same or different scales. Thus, the relative magnitude of the coefficients is simulating the latency in the time to firing of its corresponding cell. The larger the magnitude of the coefficient, the shorter is its latency of firing and, therefore, the higher it is in the rank-order of the incoming spikes with respect to time. Thus, rank-order encoding of the spikes is simulated by arranging the coefficients obtained in equation 4.4 in descending order.

The DoG corresponding to the largest coefficient is the cell which is the first to fire its first spike in the population of M ganglion cells. The ON-centre and OFF-centre Φ_s are mathematical complements of each other so that if the ON-centre Φ_s at spatial location (k,l) generates a positive coefficient, the OFF-centre Φ_s at that location will generate a negative coefficient, and vice-versa. This is a situation where at a certain spatial location (k,l), either the OFF-centre or the ON-centre cell fires, but not both. This is true for all spatial locations across all the scales. Therefore, in a population of M ganglion cells in *VanRullen and Thorpe's retinal model*, only 50% fire, giving rise to $P = 0.5M$ positive coefficients of filtering. Thus, we have an array $\{r_i \in R : i = 1 \ldots P\}$ of rank-ordered coefficients, simulating rank-ordered spikes fired asynchronously by P ganglion cells in a population of M.

4.1.4 Rank-order decoding

To test the performance of rank-order codes in encoding information about an image efficiently, VanRullen and Thorpe decoded the rank-ordered coefficients to reconstruct the input stimulus. The resolution of DoG filter functions in the *retinal model* is such that the neighbouring filters of a certain layer overlap as shown in figure 4.4. When an input image is processed by such a layer of cells, a point at a certain spatial location in the image is sampled by multiple filters of the same layer. Again, the filters at larger scales also sample each part of

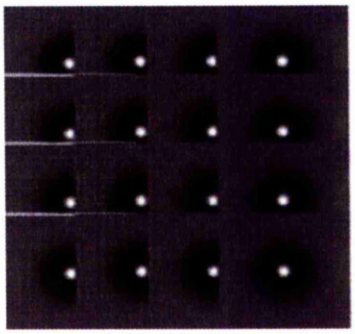


Figure 4.4: Overlap among neighbouring DoG filters sampling the input image.

the spatial location that has already been sampled by filters lower down in scale. Thus the neighbouring filters across the layers are also overlapping. Further, $\langle \Phi_{s_1}(k,l), \Phi_{s_2}(k+1,l) \rangle \neq 0$. However, because of the normalisation in equation 4.3, $\langle \Phi_{s_1}(k,l), \Phi_{s_2}(k+1,l) \rangle = \epsilon$, where $|\epsilon| \ll 1$. Further, instances (say when $s_1 = 1$ and $s_2 = 8$) when the above approximation does not hold true are very few due to sparse sampling by the larger filters. Therefore, for all practical purposes, the DoG filters are considered to be an *approximately orthogonal* set of basis vectors that tile the retina uniformly [81]. Further, the DoG functions are circularly symmetric. Therefore, under the assumption that the DoG filters form a set of orthogonal basis functions, the inverse of a DoG matrix will be the matrix itself. VanRullen and Thorpe use this approximation to reconstruct the input image as discussed below. Later we will improve on this approach by identifying ways to correct for the non-orthogonality (discussed in chapters 6 and 7).

Reconstruction using coefficients

From equation 4.4, and using the approximation that the DoG are orthogonal filters, the input image $\mathcal{I}$ can be reconstructed as:

$$\mathcal{I}_{rec}(k,l) = \sum_{s} \sum_{x} \sum_{y} C_s(x,y).\Phi_s(x-k)(y-l) \tag{4.5}$$

Initially, $\mathcal{I}_{rec}$ is an empty matrix of the same size as the input image $\mathcal{I}$. Let $\Phi^1_{s_i}$ be the filter at scale s_i, which when placed at the spatial location (k,l) of the image $\mathcal{I}$ produces the largest coefficient r_1 in the rank-ordered coefficient set R defined in section 4.1.3. To start with, the $\Phi^1_{s_i}$ is scaled up by r_1 and plugged into $\mathcal{I}_{rec}$ with the centre of the scaled-up DoG at the pixel (k,l). This is done sequentially for

each consecutive rank-ordered coefficient r_i of the set R and the process is referred to as the **progressive reconstruction** of the input image. Reconstructed images using up to the first 1%, 5%, 10% and 20% of the coefficients of filtering are shown in figure 4.5. The reconstruction process is similar to the perception of visual messages by the brain as it receives an asynchronous stream of spikes from a population of asynchronously firing ganglion cells. As shown in VanRullen and Thorpe's work, we observe in figure 4.5 that by the time 1% of the cells fire their first spikes, the main subject of the input image is recognisable from the reconstruction. A qualitative evaluation of the reconstructed images show that by the time 10% of the total population of ganglion cells have fired their first spikes, the corresponding reconstructed images appear to contain all the perceptually important information present in the original.

Reconstruction with a Look-Up-Table

In an earlier section, we discussed that the value of a certain coefficient C_s corresponding to a certain spatial location (k, l) represents the strength of the stimulation that drove the ganglion cell at that location above threshold and caused it to fire a spike at a certain rank r_i with respect to other cells in a population. While decoding the rank-order encoded image, each ganglion cell is assigned a weight depending on its rank of firing. So far, we have used the coefficients of filtering of an image as weights for the purpose of decoding and reconstruction. Yet, the brain does not know anything about the input stimulus. All it can see is the order of arrival of spikes, irrespective of the activation that caused a ganglion cell to fire a spike. Thus, true rank-order decoding can be simulated if there is a common LUT of weights, where each entry is associated with a certain rank of a spike, irrespective of the corresponding input stimulus contrast levels. VanRullen and Thorpe generated an LUT by averaging the rank-ordered coefficients of around 3000 images [81]. If T is the total number of images in a data-set, then a weight in the LUT associated with a spike at rank i is:

$$LUT(i) = \frac{1}{T} \sum_{u=1}^{T} r_{iu}, \tag{4.6}$$

where u represents the u^{th} image in the data-set. In our simulation, $T = 65$. The LUT weights are plotted against the corresponding rank in figure 4.6(a) as

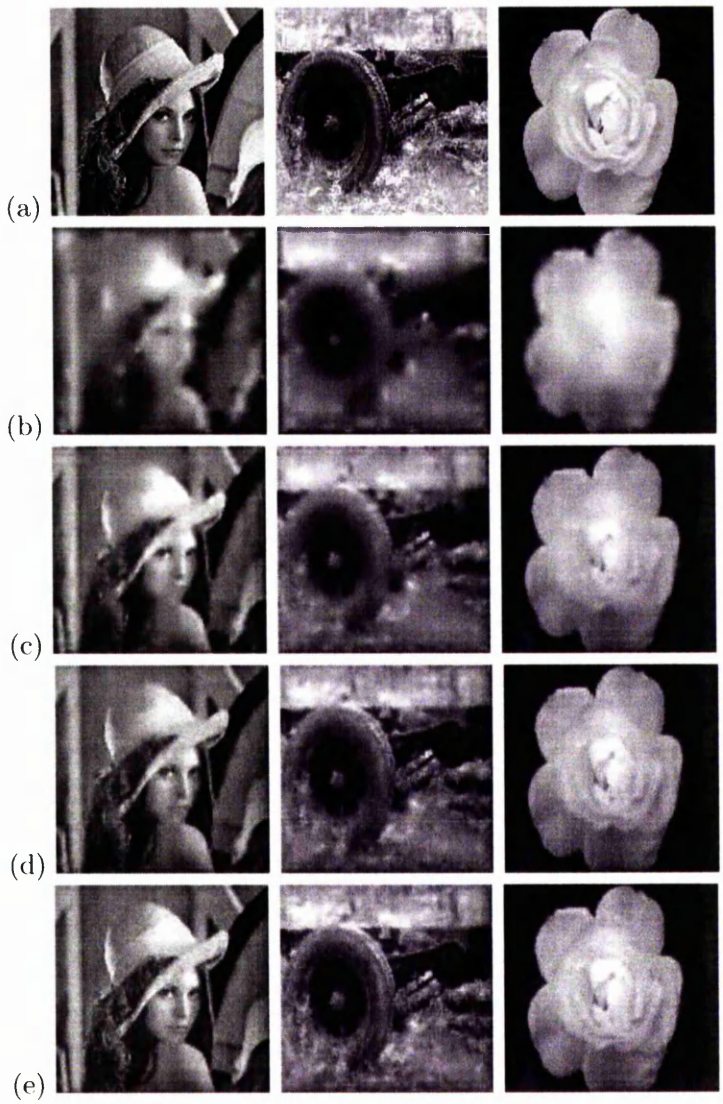
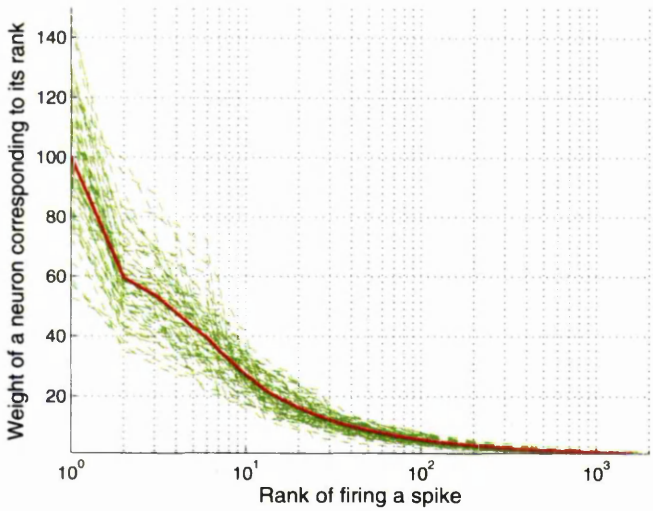
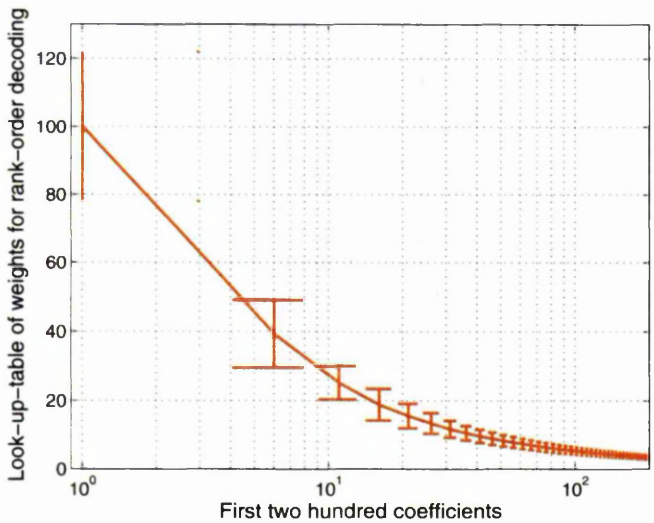


Figure 4.5: (a) input images. (b–e) Reconstructed images using the first (b) 1%, (c) 5%, (d) 10% and (e) 20% respectively of the rank-ordered coefficients of filtering the image with the DoG filter-set.



(a)



(b)

Figure 4.6: (a) A semi-log plot of a look-up-table (LUT) of weights used during decoding the rank-order codes to weigh each neuron depending on its order of firing a spike in a population of asynchronously firing neurons. (b) The standard deviation of the rank-ordered coefficients of filtering for all sixty-five images about the LUT. Each point on the LUT corresponding to a certain rank represent the mean of the coefficients at that rank.

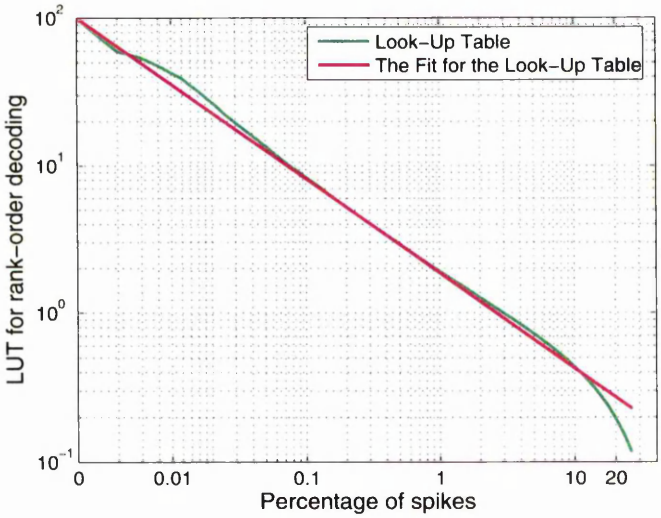


Figure 4.7: A log-log plot of the LUT up to the first 20% of the coefficients.

a semi-log plot. The weights are normalised so that the largest weight has a value of 100. A sharp drop in the magnitude of the second weight is observed, from where the slope is almost linear until about the 20^{th} coefficient. The slope again falls and gradually becomes flatter until it becomes very close to 0 by the time about 1000 ganglion cells have fired their first spikes. This corresponds to $\simeq$ 2% of the total number of cells. The implication of such a nature of the plot is: the vital information in the rank-order encoded form of an image can be retrieved by the time 2% of the ganglion cells in a population have fired their first spikes. The rank-ordered coefficients of filtering of each individual image are shown as a spread about the mean, which is the LUT. The standard deviation of the spread about the LUT is shown in the errorbars in figure 4.6(b). The deviation is significant for the first 10 spikes, after which it diminishes. This high deviation of the data in the upper regions of the LUT plot conforms with the wide variety of images used to generate the coefficients, with widely varying contrast values at similar spatial locations.

The progressively reconstructed images using the first 1%, 5%, 10% and 20% of the weights from the LUT are shown in figure 4.8. Indeed, the reconstruction with the first 1% of the rank-ordered spikes does enable us to recognise the object in a picture, while the reconstruction with the first 20% of the spikes does not seem to hold more information than that with the first 10% of the spikes. This

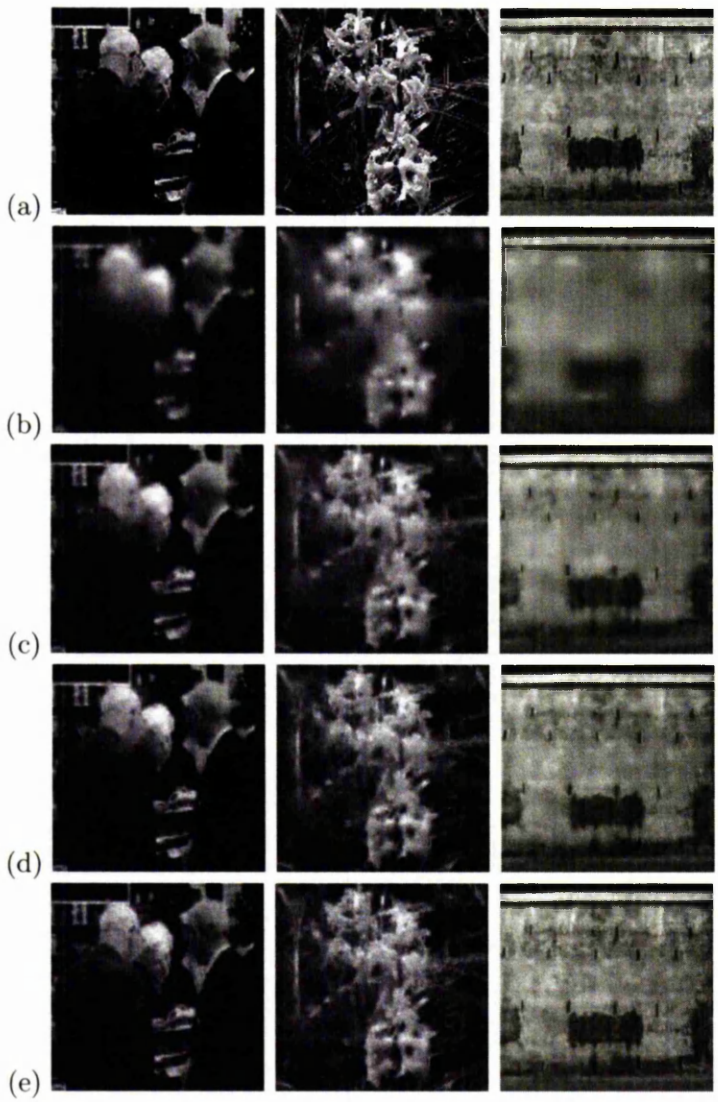


Figure 4.8: (a) input images. Reconstructed images using the first (b) 1%, (c) 5%, (d) 10% and (e) 20% respectively of the weights from the LUT.

conforms with the behaviour of the LUT plot as shown in figure 4.7. There is a fall in the slope of the LUT after the first 10% of the coefficients, indicating a fall in the rate of information recovery after this point. A sharp fall of the slope after the arrival of the first 20% of the spikes indicates that the later spikes add very little information for the decoding and that the most important information is carried by the **first 10 − 20%** of the spikes. Our validation results, thus, confirm those of VanRullen and Thorpe [81].

Power Law fit for the LUT

The LUT plot in figure 4.7 is fitted with an equation of the form

$$f(r) = Zr^{-\gamma},$$

where $\gamma = lna$, f is some function of r, and Z and a are constants with values of 100 and 0.63 respectively. Here, f represents the LUT weights and r represents the rank of the spikes. This is similar to the power law equation given by [16]:

$$p(x) = Cx^{-\alpha},$$

where $p(x)$ is the probability of measuring x, and α and C are constants. A quantity is said to follow a power law when the probability of measuring the quantity varies inversely as a power of that value. As observed in figure 4.7, the LUT follows a power law for the first 10% of the coefficients, after which it deviates. In our application, that would mean that for the first 10% of the coefficients, the **importance** of the information from the first few spikes varies as the power of the rank of the coefficients.

Power laws are characterised by a property of scale invariance, whereby a distribution looks the same regardless of the scale that is considered. In later chapters, we show that the distribution of the LUT is scale invariant and does not change with change in image size.

4.2 Chapter summary

In this chapter, we validated *VanRullen and Thorpe's retinal model*, and showed that we derived similar empirical results from our simulation. We also showed that the LUT approximately follows a power law fit up to the first 10% of the

coefficients. On visual inspection of the quality of the reconstructed image, we observe that with the first 1–2% of the ganglion cells firing their first spikes, the information is retrievable up to the point where the input picture is recognisable. By the time the first 10–20% of the cells fire their first spikes, most of the visible information seems to be recovered. However, a quantification of the information content in the reconstructed image with respect to the original is desirable for precise evaluation of rank-order codes in terms of how much information is recovered and the time to such recovery. Further, a quantitative measure which takes into account the contrast sensitivity of the eye is desirable. This is because, after all, the rank-order codes are proposed as a means of information transmission by the eyes. So, measuring the visually-important information content in the reconstruction, rather than a direct comparison of the pixel values of the respective images, makes sense.

In the next chapter, we introduce an objective measure, which was originally proposed and used by Petrovic and Xydeas in measuring perceptually-important information content in fused images with respect to the original. We adapt the algorithm to measure the perceptually-important information in a rank-order decoded image with respect to the original.

Chapter 5

Quantitative Evaluation of Rank-order Codes

In the previous chapter, we validated and presented an empirical analysis of the *retinal model of VanRullen and Thorpe* which they designed for testing the performance of rank-order codes in the rapid encoding of information about a visual stimulus. From our empirical results, we observe that, indeed, by the time only 1% of the ganglion cells in the population have fired their first spikes, objects in the reconstructed image are recognisable. This is by visual inspection and gives a qualitative estimate of the performance of rank-order codes in early encoding of a visual stimulus. At this point we propose to quantitatively evaluate the performance of the codes in terms of *perceptually-important* information recovered from the rank-order codes as well as time to such recovery [66]. We turn to Digital Image Processing techniques existing in literature to obtain a suitable method for applying to our work.

Image quality measurements are vital for monitoring information loss during image processing and benchmarking various image processing algorithms. Subjective evaluations has been a common way of evaluating the quality of reconstructed images, but such methods are expensive and time consuming [77, 93, 87]. Further, there is no consistency in the data available, as subjective judgements may vary depending on the ambient light conditions for example, or many other trivia [78, 87]. Such conditions necessitate quantitative measures, referred to as *objective measures* in literature, to evaluate image quality degradations post-processing [86, 56, 93]. Most of the algorithms proposed initially were computationally simple, the most widely used among these metrics being the Mean

Squared Error and Signal to Noise Ratio [49, 48, 86, 87]. However, on subjective evaluation of images evaluated objectively with Mean Squared Error (MSE) for example, they were found to be significantly different in terms of perception by the human eye [12, 87]. Thus, the objective measures were found to be lacking in consistency in correlating with subjective quality measures [48]. This prompted research on objective measures based on the Human Visual System (HVS) characteristics [7, 8]. Daly proposed the Visible Difference Predictor (VDP) algorithm using a simplified model of the HVS to determine the probability of an observer noticing the difference between two images [12]. However, such measures of image fidelity by *error estimation* has been criticised in favour of *structural distortion measurement*, which is described as a good approximation to the image distortion perceived by the human eye [87]. On such basis, the Universal Quality Index (UQI) as an image quality assessment method was proposed, which was later extended to propose the Structural SIMilarity Measurement (SSIM) [86, 87, 88]. Studies show that UQI perform better than error estimates such as MSE, PSNR and even VDP, while the SSIM performs better than UQI [86, 88, 10]. These algorithms were evaluated against the mean opinion score (MOS) (subjective quality measurement), and yet, the algorithms themselves were not validated or optimised with subjective trials. We found that an algorithm proposed by Petrovic and Xydeas used extensive subjective trials for validating and optimising an objective metric that they proposed to estimate the perceptual information preserved (PIP) in a fused image with respect to the original images that were fused [93, 59, 57]. Also, this algorithm falls under the category of 'structural distortion measurement' rather than 'error estimation' [87]. Although UQI, VDP, Mutual Information (MI), and other algorithms were proposed for image quality assesment in image fusion, none of the algorithms were subjectively validated [78, 24, 60, 56]. Further, while image quality assesment of fused images using VDP and UQI were better than that using MI, those using PIP are better than all the others mentioned above, except a combined model using the VDP and PIP in cascade (vdPIPp), which again might be computationally inefficient, given considerable computation requirements in each constituent method [58].

Quantitative evaluations using MI and MSE have already been shown to provide positive results in favour of rank-order codes with respect to rate-codes [81]. Yet, in terms of visual perception, the absolute difference in the grey level values in each and every pixel of a reconstructed and input image is not important.

This is because the *contrast sensitivity of the human eye* has upper and lower thresholds and is a non-linear function of such absolute differences in grey levels. Besides, there are many discrepancies in 'error estimate' algorithms as noted earlier. Rather, an objective metric quantifying *perceptually-important* structures retained in a reconstructed image with respect to the input image would be a more appropriate evaluation of rank-order codes. Since subjective evaluations are expensive, we would prefer a meaure that has been extensively validated and/or parameterised and/or optimised with subjective trials.

The *raw primal sketch* is considered by Marr to be a rich description of an image since it contains virtually all the perceptually-important parts in the image [50]. The *primitives* of such a sketch are edges, bars, blobs and terminations, each of which have *attributes* of orientation, contrast, length, width and position. The exact estimation and comparison of such parameters are beyond the scope of this work. For our current purpose of evaluating the performance of rank-order codes, a comparison of the '*edge*' *primitive* and its *attributes* of *contrast, orientation* and *location* would be a very good place to start comparing the perceptual similarity between two images.

Petrovic and Xydeas used a *Perceptual Information Preservation* algorithm to quantify the success of 'information fusion', which is measured in terms of *accuracy* in 'transfer of local gradient information' *to* the fused image *from* two or more input images that are to be fused [59]. Further, as has been mentioned earlier, Petrovic and Xydeas validated and optimised their objective metric against extensive subjective tests [57]. In their work, the *goal* of 'image fusion' is stated to be "faithful representation of the most important input information in the fused image". This is very similar to the goal of several neural hypotheses — a code which can faithfully represent the perceptually-important information about the outside world. Indeed, it is with such a motive that we have come this far in our work, where we ask — *What percentage of the perceptually-important information in an image is contained in its rank-order encoded form?* A solution is to suitably adapt the Perceptual Information Preservation (*PIP*) Algorithm put forward by Petrovic and Xydeas. How we do so, the results thereof and their implications, are the topics of discussion in the rest of this chapter.

5.1 Perceptual information preservation

An *edge* in an image corresponds to a discontinuity in its intensity surface. It can be approximated by a piecewise linear curve composed of short, linear edge elements, known as *edgels*. Each edgel can be considered as a vector characterized by *strength*, *orientation* and a *spatial location*. As stated earlier, the human visual system (*HVS*) perceives an image by extracting information contained in the variations in grey levels, which are maximum near the edges, rather than absolute grey level values. Applied to image fusion evaluation, this means that an ideally fused image would contain all the perceptually-important edge information present in the input images [59]. Applied to our work, an ideally decoded (i.e. reconstructed) image would contain all the perceptually-important edge information contained in the input image. Again, a fusion algorithm that correctly fuses the most important edges performs better than one that only preserves the less significant ones. Similarly, rank-order codes would be considered to be performing well if the most important edge information in an input image is retained while encoding, which could be termed as a '*perceptually lossless*' encoding of an image. Furthermore, a quantitative evaluation of image fusion would therefore measure (a) how well the fused edges represent those in the inputs and (b) how important are those edges in terms of perception by the HVS. Similar measures would apply for quantitative evaluation of the performance of rank-order codes — how well do the edges in the reconstructed image represent those of the input, and how important are those edges in visual perception?

The PIP algorithm as adapted in our work is shown in figure 5.1 and discussed in the following sections. The empirical results of applying the algorithm to reconstructed images are also shown. The algorithm gives an objective measure of the performance of rank-order codes in achieving a 'perceptually lossless' encoding of an image.

5.1.1 Normalising images

We apply the algorithm to an input image I_0 as shown in figure 5.2(a) and its reconstruction R_0. Reconstruction with 20% of the coefficients is shown in figure 5.2(c). It is observed that there is a wide deviation in the pixel contrast values in I_0 and R_0, which is very apparent from their respective histograms in figures 5.2(e) and 5.2(f) respectively. Empirical results show that the maximum

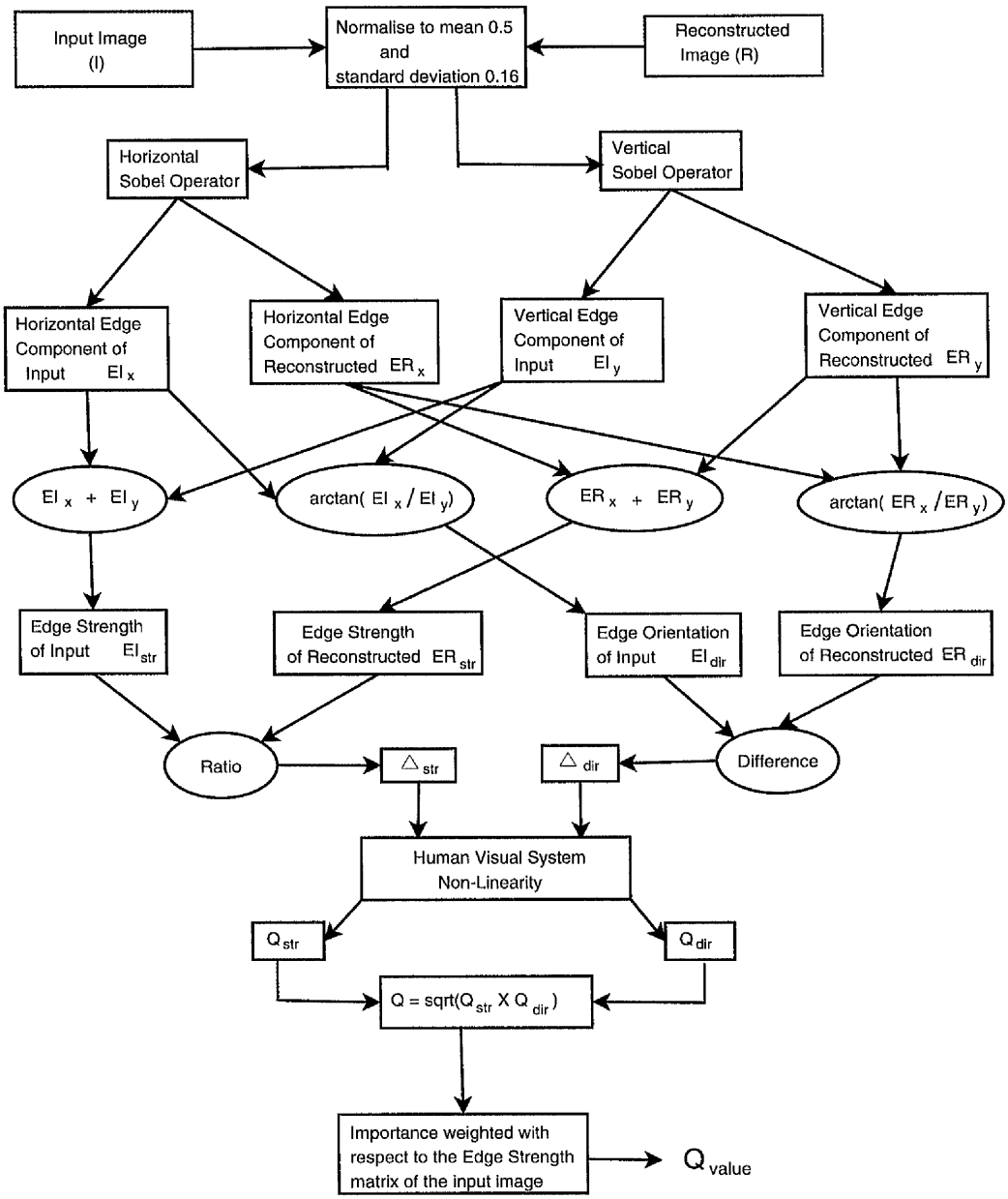


Figure 5.1: Flowchart of Petrovic and Xydeas's perceptual information preservation algorithm, as adapted in the current work for quantitative evaluation of information recovery from rank-order encoded images.

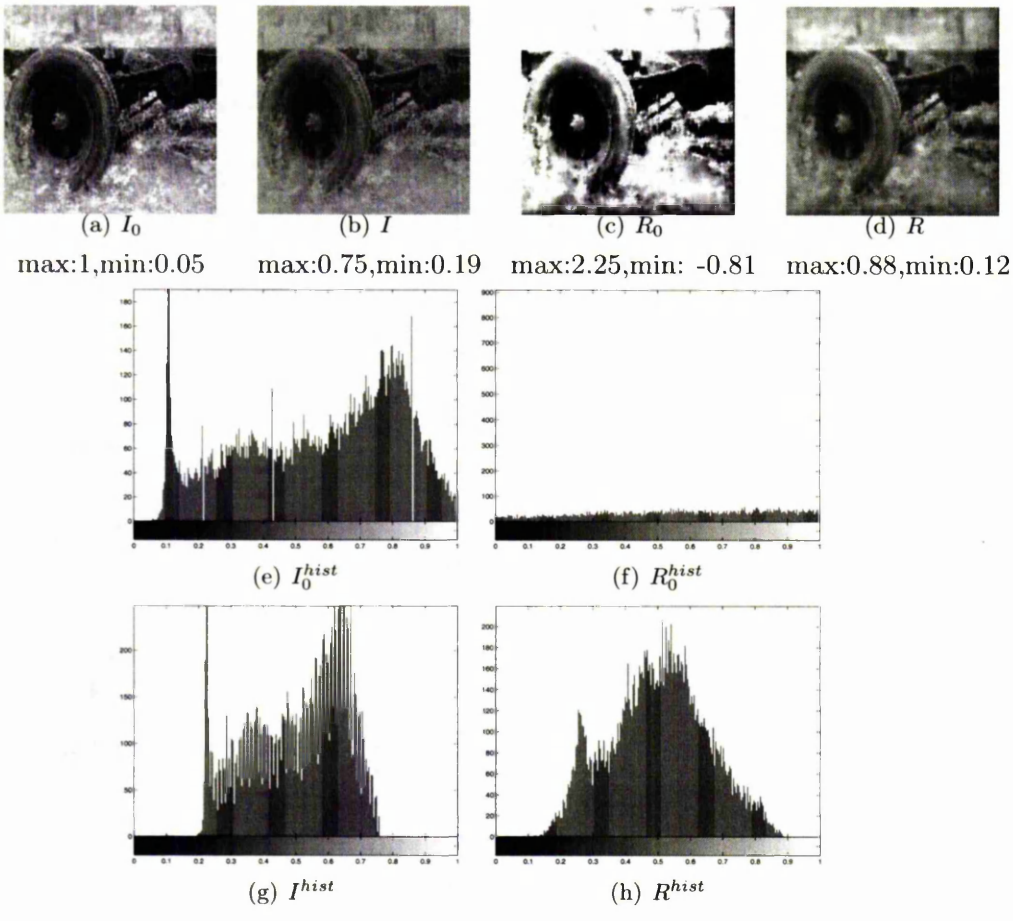


Figure 5.2: (a) An input image I_0 and its maximum and minimum pixel values. (c) Reconstruction of I_0 using 20% of the coefficients of filtering, and the maximum and minimum pixel value in the reconstructed image R_0. Histogram (e) I_0^{hist} of I_0 and (f) R_0^{hist} of R_0, indicating the wide deviation of the pixel values in the two images, and thus a wide variation in the image contrast. I_0 and R_0 are normalised to a mean of 0.5 and a standard deviation of 0.16 to get (b) I and (d) R respectively. Their respective maximum and minimum pixel values are also shown. Histogram (g) I^{hist} of I and (h) R^{hist} of R indicates the decrease in deviation of the overall pixel values as a result of normalisation, and thus a decreased image contrast.

1	2	1
0	0	0
-1	-2	-1

(a) horizontal

1	0	-1
2	0	-2
1	0	-1

(b) vertical

Figure 5.3: (a) Horizontal and (b) vertical Sobel operator templates used for detecting the horizontal and vertical components of the edges in an image.

contrast value of a pixel in I_0 is 1, while the minimum is 0.05. Similar measures for the reconstruction R_0 are 2.25 and -0.81 respectively. This wide deviation is prone to produce below optimal results in fidelity measures based on the magnitude of each pixel at corresponding locations of interest in two images. The distortion may be attributed to the process of encoding and decoding the original pixel values in I_0.

To correct the anomaly, we normalise both I_0 and R_0 to a common mean of 0.5 and standard deviation of 0.16, to get I and R, as shown in figures 5.2(b) and 5.2(d) respectively. The corresponding histograms of the normalised images are also shown in figure 5.2(g) and figure 5.2(h). The maximum and minimum pixel values of I are 0.75 and 0.19 respectively while the corresponding values in R are 0.89 and 0.12. We observe that the deviation of the pixel values in the input and the reconstructed images are lessened after normalisation.

5.1.2 Edge detection and comparison

Two normalised images I and R are passed through a *Sobel operator*, which is a first order differentiator, to obtain the horizontal and vertical components of each edgel in both images. The Sobel operator constitutes two components:

Horizontal operator, shown as a 3×3 template in figure 5.3(a). Filtering an image with the horizontal template detects the horizontal components EI_x and ER_x of each edgel in the images I and R respectively.

Vertical operator, shown in figure 5.3(b), which detects the vertical component of each edgel, viz. EI_y and ER_y in I and R respectively.

From these two components, as shown in figure 5.1, the strength of an edgel at a spatial location (k, l) for each of the images I and R can be calculated as

$$EI_{str}(k,l) = |EI_x(k,l)| + |EI_y(k,l)|$$

and

$$ER_{str}(k,l) = |ER_x(k,l)| + |ER_y(k,l)|,$$

while the direction of each edgel is obtained as

$$EI_{dir}(k,l) = \arctan\left(\frac{EI_x(k,l)}{EI_y(k,l)}\right).$$

and

$$ER_{dir}(k,l) = \arctan\left(\frac{ER_x(k,l)}{ER_y(k,l)}\right).$$

The strength of an edgel in an image is the contrast at the pixel position corresponding to the edgel. However, it is the relative difference in contrast values of neighbouring pixels, expressed as ratio of the contrast values and known as the *contrast ratio*, that is more important in visual perception than the absolute difference. The relative difference between the contrast values at (k,l) of I and R can be expressed as the contrast ratio $\Delta_{str}(k,l)$, where

$$\Delta_{str}(k,l) = \frac{\min(EI_{str}(k,l), ER_{str}(k,l)}{\max(EI_{str}(k,l), ER_{str}(k,l)},$$

where EI and ER correspond to the edgel parameters of the input and the reconstructed images respectively. Similarly if $\Delta_{dir}(k,l)$ is the difference in orientation at (k,l), then

$$\Delta_{dir}(k,l) = \frac{||EI_{dir}(k,l) - ER_{dir}(k,l)| - \frac{\pi}{2}|}{\frac{\pi}{2}}.$$

Both Δ_{str} and Δ_{dir} are normalised to lie in the range 0 to 1. These are linear parameters, and are modulated to represent the contrast sensitivity of the HVS. In the following section, we discuss the HVS non-linearity and how the linear functions Δ_{str} and Δ_{dir} are transformed into non-linear functions to conform with the contrast sensitivity of the HVS.

5.1.3 Human Visual System non-linearity

In this section, we discuss briefly the terms *contrast sensitivity* and *detection threshold* of the human eye. We then continue with Petrovic's PIP algorithm from where we left off in section 5.1.2. We describe and validate the psychometric function used by Petrovic and Xydeas, based on subjective tests, to simulate the non-linearity of the HVS.

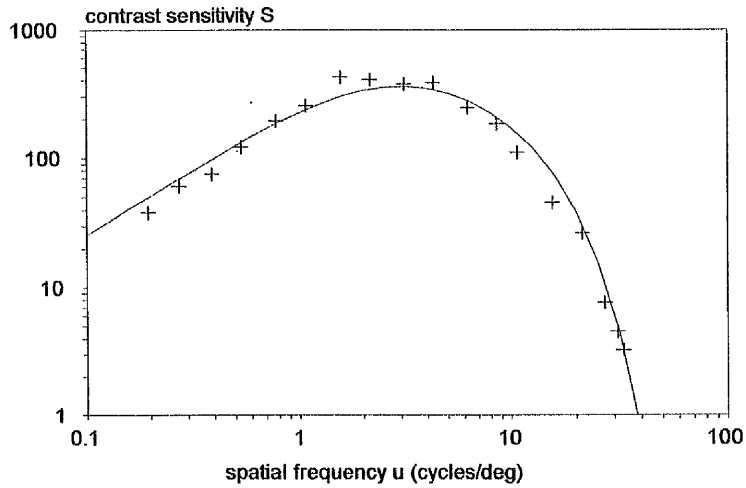


Figure 5.4: Contrast sensitivity from subjective tests shown as a scatter, which is fitted with a plot that follows the characteristics of a bandpass filter [20].

Contrast sensitivity and detection threshold

The **Contrast sensitivity** of the human eye is defined as the reciprocal of the minimum contrast required for detection [20]. It is mathematically defined as the *reciprocal* of the **modulation threshold**, also known as the **detection threshold**, of the eye for sinusoidal gratings. The modulation threshold of a sinusoidal grating is the minimum *modulation depth* required for detection of the pattern by the human eye (appendix A). Subjective tests show that the characteristic contrast sensitivity function of the eye resembles a bandpass function as shown in figure 5.4. It is usually expressed as a function of the spatial frequency of a sinusoidal grating and is seen from figure 5.4 to be most sensitive to frequency changes at 10 cycles per degree. The shape of the curve implies that the *more sensitive* the eye is to a certain frequency, the lower the *detection threshold* at which it can detect changes. In other words, certain spatial frequencies can be better detected by the eye than others, and at those frequencies, the eye has a high sensitivity and can detect contrast with a very low modulation threshold. But for spatial frequencies that fall at the edge of the sensitivity range of the eye, the detection threshold is much larger, and the eye can detect changes only with a large variation in the contrast.

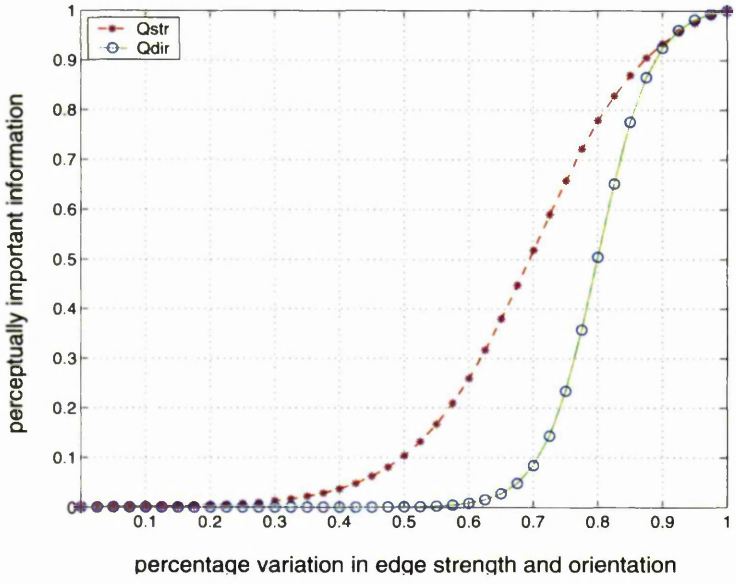


Figure 5.5: The Human Visual System non-linearity shown as psychometric functions, plotted with parameters decided by results of subjective trials [59]. The steeper fall in the curve for orientation show a greater sensitivity of the eye to changes in orientation of the edges in an image than to changes in the contrast of individual pixels.

Applying HVS non-linearity

When comparing two images, the detection of the difference in edge luminosity and orientation at a certain position by the human eye depends on the detection threshold of the eye. The detection threshold is different for contrast and orientation, and is defined as the point at which the probability of detection of the difference by the human eye is 50%. With reference to this algorithm, this threshold is the point at which the edge strength or orientation has perceptually degraded to 50% of its original quality. The nonlinear behavior of the eye is defined as a psychometric function

$$f(x) = \frac{K}{1 + exp^{-s(x-d)}}, \tag{5.1}$$

where d is the detection threshold such that $f(x) = 0.5$ at $x = d$, s is the steepness parameter, and K is a constant so that $f(x) = 1$ at $x = 1$. Petrovic and Xydeas use results of subjective trials to set the value of the parameters $[d, s]$ to $[0.7, 11]$

and $[0.8, 24]$ for the contrast ratio and the orientation respectively [59]. Putting $x = \Delta_{str}$ and $[d_1, s_1] = [0.7, 11]$ in equation 5.1,

$$Q_{str} = \frac{K_1}{1 + exp^{-s_1(\Delta_{str} - d_1)}}, \tag{5.2}$$

and putting $x = \Delta_{dir}$ and $[d_2, s_2] = [0.8, 24]$,

$$Q_{dir} = \frac{K_2}{1 + exp^{-s_2(\Delta_{dir} - d_2)}}. \tag{5.3}$$

The plots of Q_{str} and Q_{dir} are shown in figure 5.5.

5.1.4 Objective measure of information recovered

Modulating Δ_{str} and Δ_{dir} using equation 5.1 adapts the linear parameters to the nonlinearity in the HVS. The matrices Q_{str} and orientation Q_{dir} in equations 5.2 and 5.3 respectively contain the modulated value of corresponding pixel locations in the matrices Δ_{str} and Δ_{dir} and reflect the perceptual information of each edgel in terms of its contrast and orientation (figure 5.5). The geometric mean of these two component matrices produces the matrix Q, which is a measure of the preservation or degradation of the perceptually-important edges that are present in the reconstructed image with respect to the input.

$$Q = \sqrt{Q_{str} \times Q_{dir}}.$$

Q is then importance weighted with the edge-strength matrix of the input picture EI_{str}. This is expressed as a normalised sum to give a single objective measure for the information recovered during stimulus reconstruction when decoding rank-order codes.

$$Q_{value} = \frac{\sum(EI_{str} \times Q)}{\sum EI_{str}},$$

such that $0 \leq Q_{value} \leq 1$; $Q_{value} = 1$ when the same two images are being compared, while $Q_{value} = 0$ when an image is being compared with a matrix with the elements initiated to 0.

The flowchart in figure 5.1 is applied for comparison of an input image and its reconstruction using 1% of the coefficients of filtering and is shown as another flowchart in figure 5.6. The results of applying the algorithm to progressively reconstructed images using 5%, 10% and 20% of the coefficients of filtering are

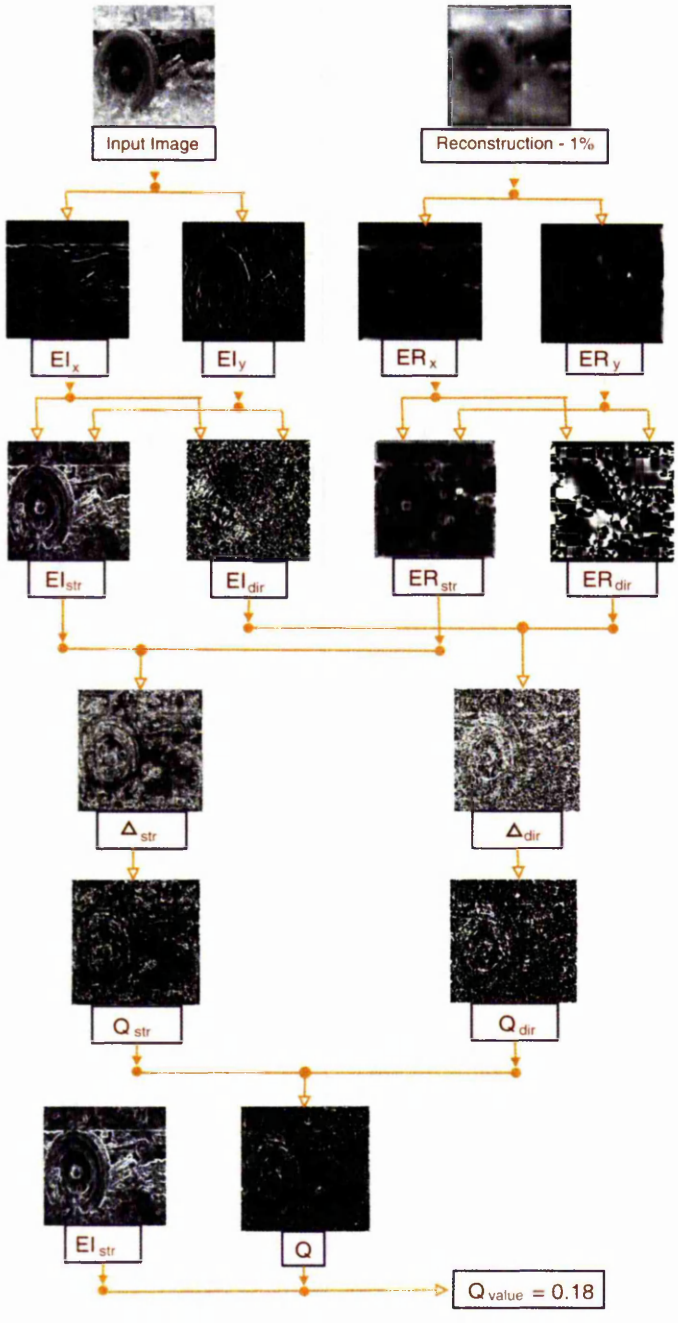


Figure 5.6: Implementing perceptual information preservation algorithm to estimate the perceptually important information content in an image reconstructed using first 1% of the true coefficients of filtering as compared to the input image. The final objective measure is obtained as Q_{value}.

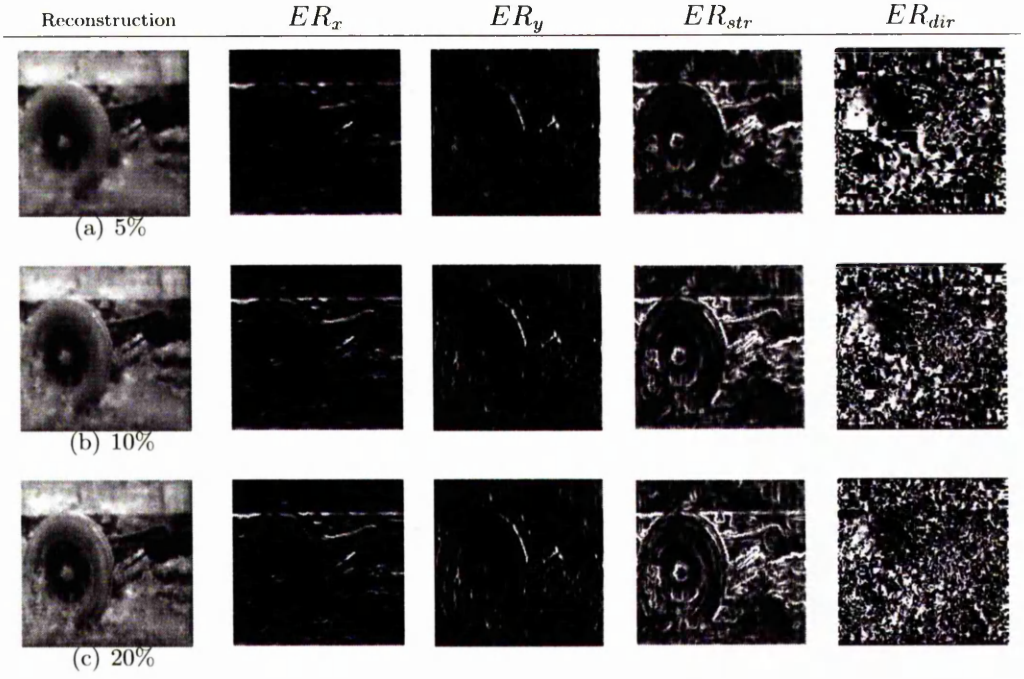


Figure 5.7: Reconstructed image using (a) 5% (b) 10% (c) 20% of the coefficients of filtering. The results of using the Sobel operator to get the horizontal and vertical edge components are shown as ER_x and ER_y respectively, whereby the magnitude and orientation of the edges are obtained and are shown as ER_{str} and ER_{dir} respectively. The matrices obtained from subsequent steps through the algorithm and finally the Q_{value} for each of the reconstructed images in (a)–(c) are shown in figure 5.8.

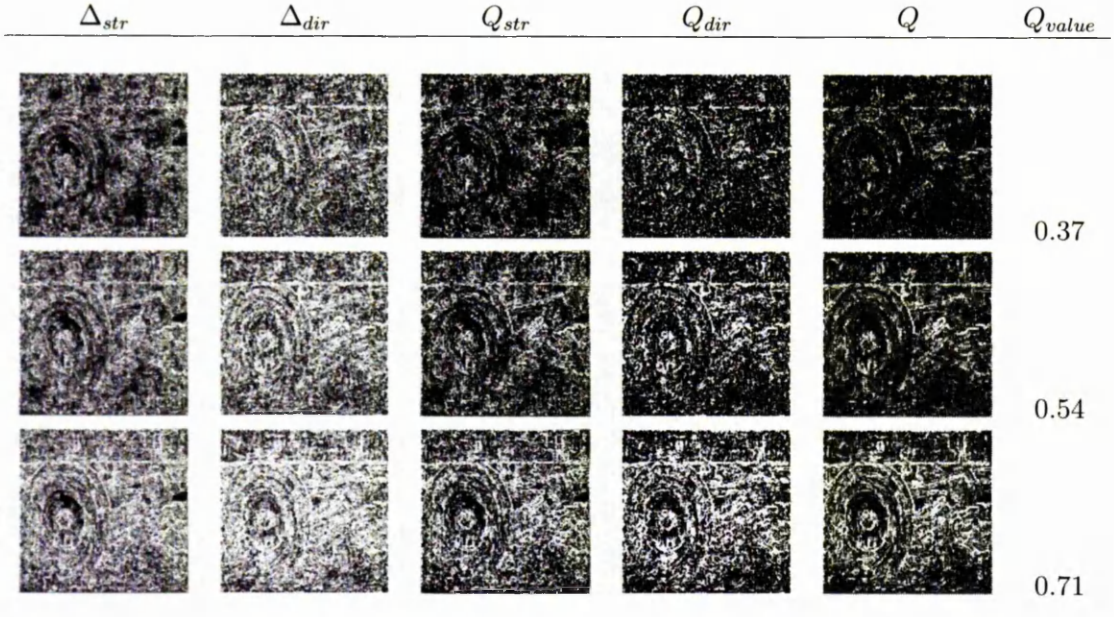


Figure 5.8: The top, middle and bottom rows correspond to the images (a), (b) and (c) respectively in figure 5.7. Each of the rows are a continuation from the rows of figure 5.7 as subsequent steps through the PIP algorithm. The relative magnitude of the edges in each of the reconstructed images (as in columns ER_{str} and ER_{dir} in figure 5.7) with respect to the input (EI_{str} and EI_{dir} in figure 5.6) are shown in column Δ_{str}, while the difference in orientation of the same is shown in Δ_{dir}. These two parameters are then modulated according to the HVS non-linearity to obtain Q_{str} and Q_{dir}. The matrix Q is obtained as a geometric mean of Q_{str} and Q_{dir}, which is then importance weighted with EI_{str} to get the objective metric Q_{value}.

shown in figures 5.7 and 5.8.

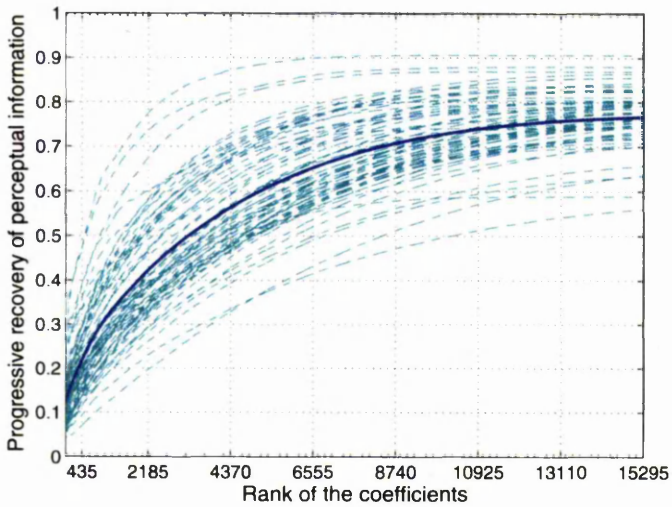
5.2 Information recovery

Having obtained an objective measure Q_{value} for comparing two images in terms of perceptually-important information content in section 5.1, we now proceed to apply it to quantitatively evaluate information recovery from rank-order codes using this measure. Empirical results of information recovery when the coefficients of filtering are used as weights in rank-order decoding are discussed in section 5.2.1, while that of using the weights from the LUT for decoding are discussed in section 5.2.2.

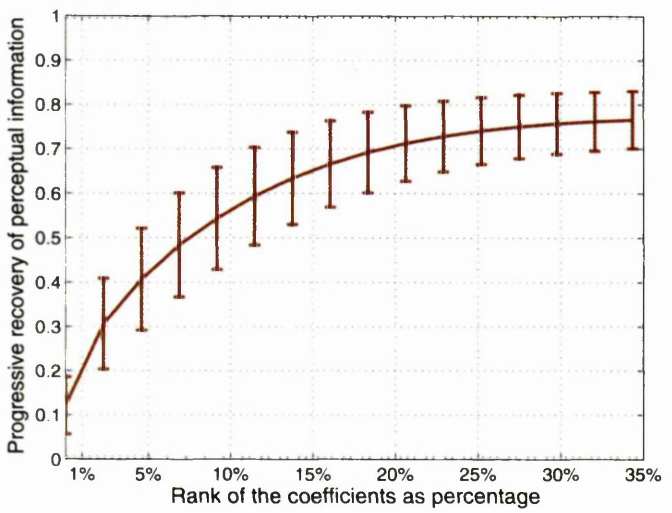
5.2.1 Information recovery with coefficients

Figure 5.9(a) shows the information recovery plot for each of the sixty-five individual images in the data-set as a spread about the mean Q_{value} at each point in the progressive reconstruction. From the mean plot, it is observed that around 70% of the information can be recovered by the time 20% of the coefficients are used for reconstruction. After this point the curve rises very slowly, indicating that there is no significant contribution from the following coefficients to perceptual information content in the image. By the time around 30% of the coefficients are used for reconstruction, a little more than 75% of the information is recovered, and the curve looks saturated. From the above observations, it can be said that using the respective coefficients of filtering for decoding rank-order encoded images, 75% of the perceptually-important information contained in an input image can be recovered on average, using the first 30% of rank-ordered coefficients of filtering. Furthermore, 70% of the information is recovered by the time only around 20% of the coefficients are used for reconstruction. Thus, up to 93% of retrievable information is recovered with reconstruction using around 20% of the coefficients of filtering.

Figure 5.9(b) shows the standard deviation of the spread of individual information recovery plots about the mean information recovered. The deviation is observed to be the most in the interval from when 5% of the coefficients are used for reconstruction to the point where around 10% of the coefficients are used. This indicates a significant variation in the rate of information recovery for the various images. This is discussed more in section 5.2.2. The deviation at the



(a)



(b)

Figure 5.9: (a) Spread of the of the respective information recovery plots, shown as dashed cyan lines, during progressive reconstruction of each of the sixty-five images in our data-set using coefficients of filtering. The average of the spread is shown as a mean information recovery plot in the solid blue line. (b) Standard deviation of the spread shown in '(a)' about the mean information recovery plot is shown as errorbars.

tail of the plot shows that around 70 – 85% of the information can be recovered on decoding a rank-order encoded image, selected from a random data-set of monochrome images of size 128×128.

In the next section we do a similar discussion for rank-order decoding using LUT weights.

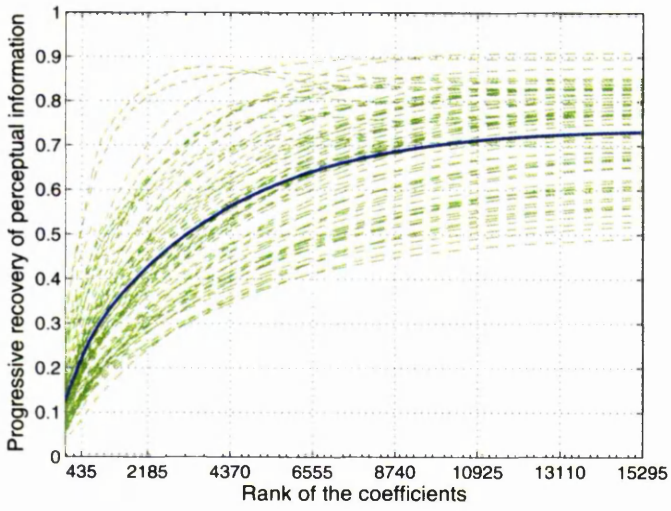
5.2.2 Information recovery with LUT

The information recovery plots for progressive reconstruction of each of the sixty-five individual images in our data-set using the LUT is shown in figure 5.10(a) as a spread about the mean information recovery plot. Firstly, it is observed that the total amount of perceptually-important information that can be recovered from rank-order codes is approximately 72 – 73% of that contained in the input image and by the time only up to 30% of the LUT weights are used for reconstruction. Secondly, a total of around 65% of the information can be retrieved by the time around 15% of the LUT weights are used for reconstruction. Thus, more than 90% of the retrievable information from rank-order codes is recovered by the time 15% of the ganglion cells have fired their first spikes. Thirdly, there is a drop of about 2 – 3% in the information recovered as compared to that in figure 5.9(b) as a result of using approximate weights from the LUT for rank-order decoding.

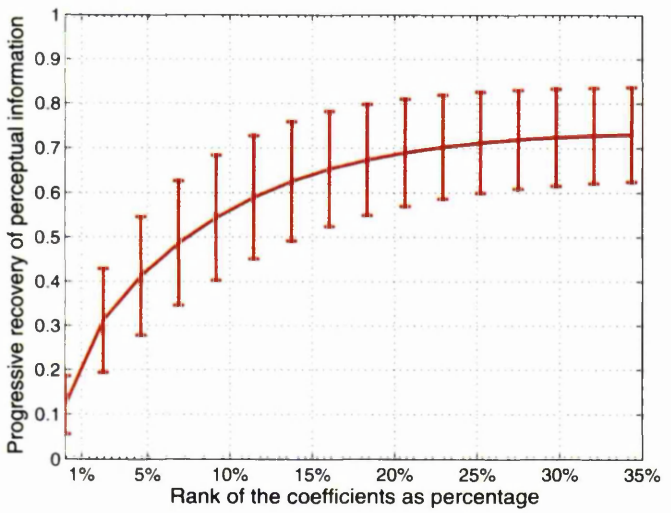
The standard deviation of the spread about the mean information recovery plot is shown in figure 5.10(b). Comparing with figure 5.9(b), we see an increase in the deviation of the data about the mean. However, the behaviour of the plot is essentially the same, with the deviation maximum between the time when 5 – 10% of the ganglion cells have fired their first spikes.

5.2.3 More empirical results

Three input images (a), (b) and (c), shown in figures 5.11, 5.12 and 5.13 respectively, are rank-order encoded using VanRullen and Thorpe's retinal model. The rank-order encoded images are then progressively reconstructed using up to the first 30% of their respective coefficients of filtering and are also shown in the respective figures. The objective measure Q_{value} obtained at every point of the progressive reconstruction for each of the images is plotted in figure 5.14. This is the **perceptually-important information recovery plot**, or simply, the **information recovery plot** for reconstruction using up to the first 35% of the



(a)



(b)

Figure 5.10: (a) Spread of the of the respective information recovery plots, shown as dashed green lines, during progressive reconstruction of each of the sixty-five images in our data-set using LUT. The average of the spread is shown as a mean information recovery plot in the solid blue line. (b) Standard deviation of the spread shown in '(a)' about the mean information recovery plot is shown as errorbars.

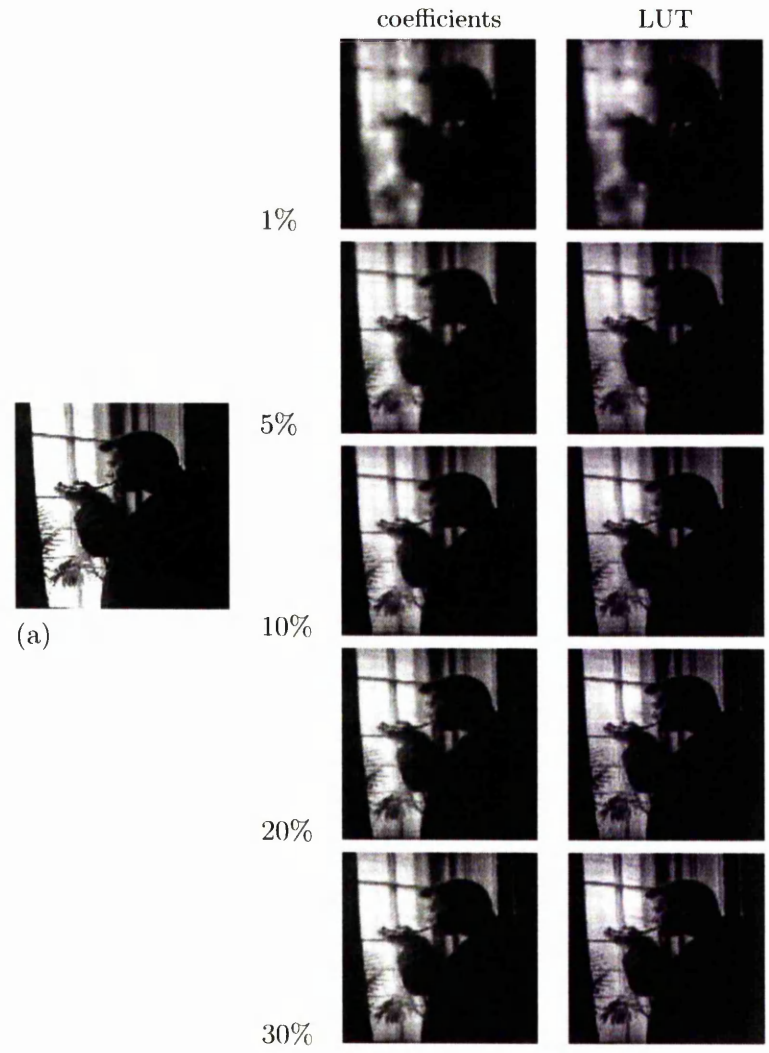


Figure 5.11: An input image (a) reconstructed from its rank-order encoded form using 1%, 5%, 10%, 20% and 30% of the coefficients of filtering and LUT values.

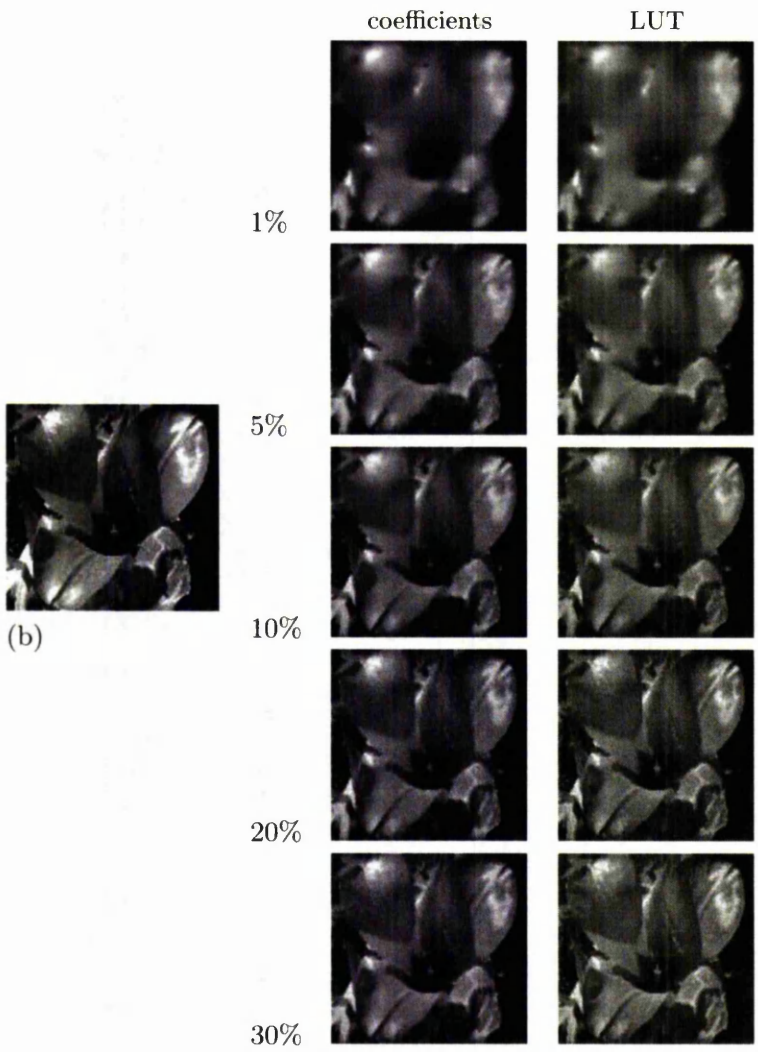


Figure 5.12: An input image (b) reconstructed from its rank-order encoded form using 1%, 5%, 10%, 20% and 30% of the coefficients of filtering and LUT values.

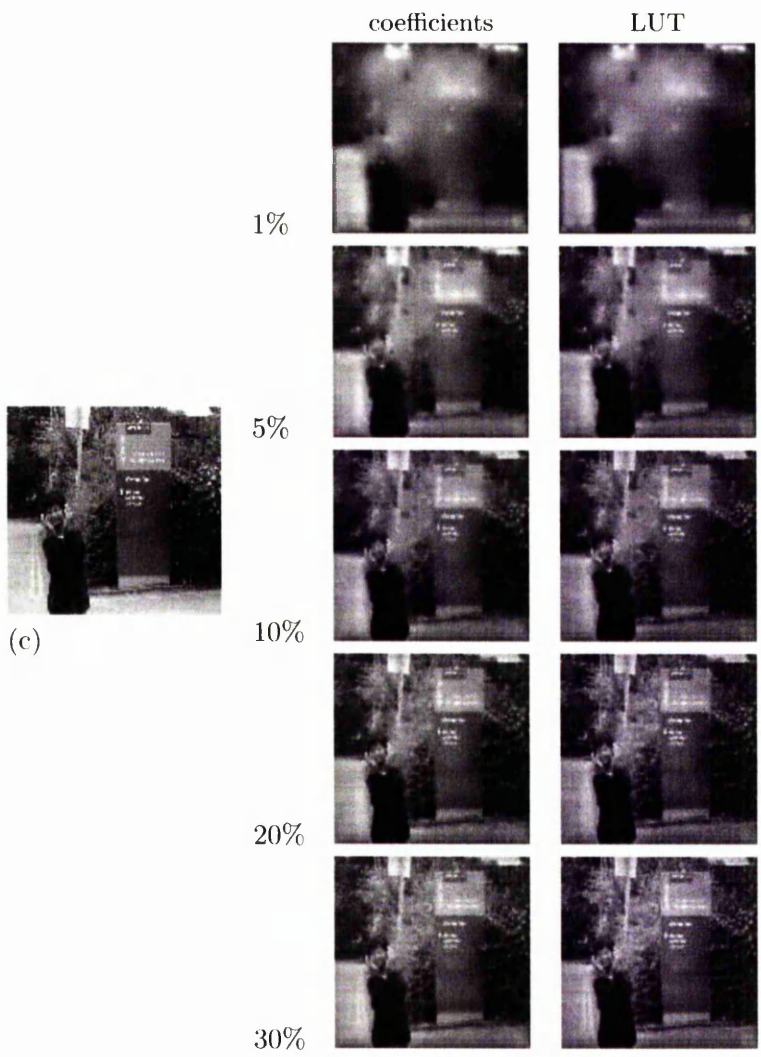


Figure 5.13: An input image (c) reconstructed from its rank-order encoded form using 1%, 5%, 10%, 20% and 30% of the coefficients of filtering and LUT values.

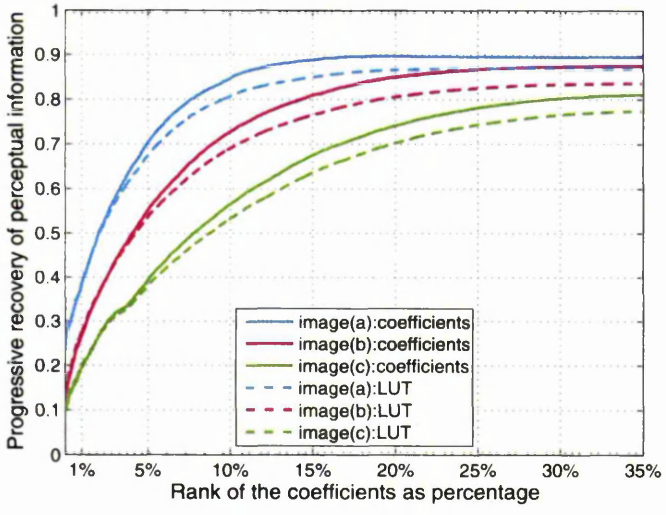


Figure 5.14: Information recovery plots for three images (a), (b) and (c) shown in figures 5.11, 5.12 and 5.13 respectively when reconstructed using (i) their respective coefficients of filtering (shown in solid lines), (ii) values from the LUT (shown as dashed lines).

coefficients of filtering [66]. We observe that around 70 – 90% of information is retrieved by the time 20% of the coefficients are used for reconstruction. Further, we also observe that the rate of information recovery is faster for the image (a) that contain fewer edges compared to the others. This is intuitively obvious as less edge information has to be recovered from the image, and hence the curve builds up at a faster rate.

The plots for information recovery for these images using the weights from the LUT are also shown in figure 5.14. Comparing these plots with the reconstruction using the respective coefficients of filtering, we observe that the information recovery for individual images is more when reconstructed using coefficients of filtering as compared with the values from LUT. On visual inspection of the three images reconstructed using the LUT, also shown in figures 5.11 –5.13, we can make a similar observation. We can thus comment that the objective information recovery measure, Q_{value}, gives an appropriate relative measure for information recovery from rank-order codes. Further, according to VanRullen and Thorpe's observation, by the time 1% of the cells fire their first spikes, we can indeed recognise the images in figures 5.11 and 5.12, which, from figure 5.14, consists of around

40% and 30% of the information respectively compared to the respective input images. However, the image in figure 5.13 is not recognisable for reconstruction with 1% of the spikes, and where the quantity of information recovered with respect to the original image is 20% as seen in the plot in figure 5.14. The image is recognisable by the time 5% of the cells have fired their first spikes and the information recovered with respect to the original is 40%.

5.3 Chapter summary

In this chapter, we have discussed an objective measure, Q_{value}, originally proposed by Petrovic and Xydeas for the evaluation of the performance of an image fusion algorithm, to evaluate the perceptually-important information that can be recovered from rank-order encoded images, and the rate of such a recovery. We then apply this measure to quantify the information recovered during each step of progressive reconstruction of a rank-order encoded image. This is done using coefficients of filtering and LUT for rank-order decoding. Continuing thus, we obtain an information recovery plot for each incoming spike. We first observe the steepness of the curve, indicating the rate of information recovery. We find that 90% of the information that can be recovered is retrieved by the time only 15% of the spikes have arrived. By the time 20% of the spikes arrive, almost all the retrievable information is recovered. On average, approximately 70% of the original information can be recovered from the codes. Further, we see that quantitative measures of perceptual information recovery using the objective measure agrees with visual inspection of the reconstructed images. Also, information recovery for images with fewer edges is faster than those having more detailed edge information. Although recognition with the first 1% of the spikes is generally true, for more detailed images with multiple objects of interest like that in figure 5.13, recognition is delayed and is obtained by the time first 5% of the spikes have been fired. However, this is also quite early on in time and supports the rank-order encoding hypothesis.

Thus obtaining a quantitative measure of the visually-important information recovered from rank-order encoded images, we proceed to optimise the method of decoding, so as to minimise the information loss suffered during decoding rank-order codes, and thus maximise information recovery from rank-order encoded images. In the next chapter, we discuss about how we use the Singular Value

Decomposition of matrices to obtain the pseudo-inverse of a DoG filter-bank, and thereby optimise the decoding method of rank-order codes.

Chapter 6

Maximising Information Recovery

In the previous chapter we discussed an objective metric Q_{value} for the perceptually-important information that can be recovered from a rank-order encoded image. By measuring Q_{value} at each step during the progressive reconstruction of a rank-order encoded image, we could plot the *perceptually-important information* recovered. An observation of the plot in figure 5.9(a) shows that on average, around 70% of the perceptually-important information can be recovered by the time around 20% of the coefficients of filtering are being used for rank-order decoding, i.e. for image reconstruction. However, the decoding method used in VanRullen and Thorpe's retinal model is lossy because of the approximation of the DoG filters as an orthogonal set of basis vectors. Ideally, a true estimation of the performance of rank-order codes in encoding visually-important information can be obtained if there is no information loss during decoding. In this chapter, we use decoding methods to deal with this approximation, and thus try to minimise information loss during decoding so that information recovery from rank-order codes can be maximised [67].

An ideal decoding method appropriate for the retinal model would be to take the inverse of the DoG matrices and obtain a perfect reconstruction with the coefficients of filtering, so that $Q_{value} = 1$. However, the DoG matrices are singular, and their inverses can only be obtained by using the pseudo-inverse method of matrix inversion, as discussed in section 6.1.3. We, therefore, take the pseudo-inverse of a **filter-bank**, consisting of the DoG matrices, and obtain a perfect decoding of a rank-order encoded image, i.e. a **perfect reconstruction** of

an input image using its respective *coefficients of filtering*. From those coefficients, we construct a look-up-table of weights for true rank-order decoding. We recover information from the rank-ordered spikes using the weights from the LUT and observe a substantial increase in perceptually-important information recovery. The empirical data from these studies are discussed in section 6.2. In the following section 6.1, we do a mathematical analysis of the approximation during rank-order decoding as done in VanRullen and Thorpe's retinal model.

6.1 Optimal rank-order decoding: theoretical analysis

Although using more than 30% of the coefficients in decoding would defeat the very essence of rank-order codes — rapid information recovery — yet, we observe that the information recovery plots, shown in figure 5.9(a), tend to saturate by the time 30% of the cells are supposed to have fired their first spikes. This implies that for all practical purposes, there is no further information retrieval beyond this point. Such a behaviour may be attributed to the approximation of the set of DoG functions as orthogonal in VanRullen and Thorpe's retinal model, thus rendering the decoding method of rank-order codes as non-optimal. In section 6.1.2, we do a mathematical analysis of this approximation, followed by a discussion on optimal decoding in section 6.1.3 and subsequent sections.

6.1.1 DoG filter-bank

We refer back to equation 4.4 in chapter 4, and repeat some of the points already discussed there, albeit with a different perspective.[1] Here, we consider image $\mathcal{I}^{m\times p}$ as a row vector $\mathcal{I}^{1\times N}$, where $N = mp$. Each DoG matrix $\Phi_s^{n\times n}$ is considered as a column vector $\Phi_s^{N\times 1}$, where the number of non-zero elements will be n^2. The total number of DoGs in the retinal model is $M = \frac{8}{3}N$. On concatenating the M DoG column vectors, we have a DoG filterbank $F^{N\times M}$:

$$F^{N\times M} = \left((\Phi_s^1)^{N\times 1}|(\Phi_s^2)^{N\times 1}|\cdots|(\Phi_s^i)^{N\times 1}|\cdots|(\Phi_s^M)^{N\times 1}\right) \tag{6.1}$$

[1]Some of the points in the following sections are repetitions of those discussed in chapter 4, in order to provide clarity and readability to the text.

The array of coefficients of filtering $\{C_i \in C : i = 1 \ldots M\}$ is then obtained as:

$$C^{1\times M} = \mathcal{I}^{1\times N} F^{N\times M} \tag{6.2}$$

This array C is then sorted in descending order to obtain the rank-ordered array of coefficients of filtering.

6.1.2 Non-optimal decoding of rank-order codes

Once we have the rank-ordered array of coefficients, we now want to decode the rank-order code to recreate $\mathcal{I}$, knowing C and F. We start with equation 6.2, which represents a set of M equations (matrix F) in N unknowns (the elements in the array $\mathcal{I}^{1\times N}$) , and thus represents an *overdetermined system.* The solution to $\mathcal{I}$ from such a set of equations, when F and C are provided, can be obtained thus:

$$\mathcal{I} = C\,(F)^{-1} \tag{6.3}$$

However, obtaining an inverse of the matrix F is a non-trivial problem because:

1. F is a rectangular matrix.
2. the DoG matrices are singular, such that their determinants $|\Phi_s^i| = 0$. Therefore, the filter-bank F is a rectangular, near singular matrix.

To deal with such a situation, VanRullen and Thorpe made an approximation while decoding rank-order encoded images, when they assumed the DoG filters to be an approximately orthogonal set of basis functions. This is explained in the following section.

DoG as approximately orthogonal function

The overlap of the DoG functions during filtering the image while rank-order encoding can be expressed by the inner product of the DoG functions in any k^{th} and l^{th} column of the filter-bank F in equation 6.1:

$$\langle (\Phi_{s_1}^k)^{N\times 1}, (\Phi_{s_2}^l)^{N\times 1} \rangle = \epsilon, \tag{6.4}$$

where s_1 and s_2 may be same or different scales and $\epsilon \ll 1$. Thus, for all practical purposes, it is assumed that $\epsilon \simeq 0$ so that the filter-bank F in equation 6.1 is an

approximately orthogonal matrix. Therefore, its transpose is approximately its own inverse:

$$(F)^{-1} \simeq F^T \tag{6.5}$$

Therefore from equations 6.5 and 6.3, we have

$$\mathcal{I} \simeq CF^T \tag{6.6}$$

Error due to approximation

Equation 6.6 gives a trivial solution for a non-trivial problem, and provides for significant information recovery from the rank-order codes, as discussed earlier. However, the error due to the approximation discussed above leads to non-optimal information recovery, and we rewrite equation 6.6 as:

$$\hat{\mathcal{I}} = CF^T, \qquad \text{where} \qquad \hat{\mathcal{I}} \neq \mathcal{I}.$$

The error of approximation is:

$$\mathcal{E} = \hat{\mathcal{I}} - \mathcal{I} \tag{6.7}$$

Due to this error, the images in figures 5.11 – 5.13 appear blurred, which agrees with an objective metric $Q_{value} < 1$ as obtained in the plots in figure 5.14, implying imperfect reconstruction. The optimal decoding of the rank-order codes can be achieved by making $\mathcal{E} = 0$ or $\mathcal{E} \rightarrow 0$. For this, we have to obtain a perfect or near perfect solution for $(F)^{-1}$ in equation 6.3.

In the next section, we discuss how to deal with this approximation using **Singular Value Decomposition** (SVD) and the *Moore-Penrose* method of computing the **Pseudo-inverse** of matrices. We start by giving a brief description of these two techniques followed by a theoretical as well as empirical analysis of how we have used the algorithms for our goal of obtaining an optimal decoding of the rank-order codes.

6.1.3 Optimal decoding using pseudo-inverse

The pseudo-inverse of a **rectangular matrix** or a **square singular matrix** is defined as:

$$F^+ = (F^T F)^{-1} F^T \tag{6.8}$$

Let $\hat{C}$ be a set of values which replace C in equation 6.3, so that in equation 6.7, the error $\mathcal{E} = 0$ and

$$\mathcal{I} = \hat{C}F^T \tag{6.9}$$

$$\Rightarrow \mathcal{I}F = \hat{C}F^TF$$

$$\Rightarrow C = \hat{C}F^TF \qquad [\text{ from equation 6.2}]$$

$$\Rightarrow \hat{C} = C(F^TF)^{-1} \tag{6.10}$$

Putting this value of $\hat{C}$ in equation 6.9,

$$\mathcal{I} = C(F^TF)^{-1}F^T$$

$$= CF^+ \tag{6.11}$$

[from equation 6.8]

6.1.4 Dealing with ill-conditioned matrices

However, in practical applications, there are often situations where we do not have ideal input conditions. Under such circumstances, the stability of the solution to a problem often depends on the sensitivity or **condition** of the matrix. A matrix is said to be **well conditioned** if small perturbations in the input data cause reasonably proportional changes in the solution. Mathematically, the **condition number** of a problem f with an input x is defined as [69]:

$$Cond(f) = \frac{|\text{relative change in solution}|}{|\text{relative change in input data}|} = \frac{\left|\frac{[f(\hat{x})-f(x)]}{f(x)}\right|}{\left|\frac{(\hat{x}-x)}{x}\right|}, \tag{6.12}$$

where $\hat{x}$ is a close approximation to the input x, resembling non-ideal input conditions. The problem f is said to be well conditioned if the condition number $Cond(f)$ is close to unity. On the other hand, if the relative changes in the solution to equation 6.12 is much larger than that of the deviation of the input from ideal, the condition number $Cond(f) \gg 1$, and f is said to be **ill-conditioned**.

In the case of matrices, the farther the condition number of a matrix is from unity, the nearer it is to **singularity**. Singular matrices have a condition number equal to infinity. For reasons discussed in section 6.1.6, we focus on the condition number of diagonal matrices, which is defined as [69]:

For any diagonal matrix $D = diag(d_i)$,

$$cond(D) = \frac{max|d_i|}{min|d_i|} \tag{6.13}$$

To deal with ill-conditioned matrices corresponding to ill-conditioned linear systems, a robust and powerful technique of **Singular Value Decomposition** (SVD) exists. SVD is used for dealing with **rectangular, singular** or **numerically close to singular** matrices, and provides the **least squares solution** for a system of equations which has an infinite number of solutions or no solutions at all. In the following sections is a definition of SVD (as extracted from [26]), and a discussion of its application in this work in order to find the solution to $\mathcal{I}$ in equation 6.11. This is followed by a discussion of computing the pseudo-inverse of F in equation 6.11 using SVD.

6.1.5 Singular Value Decomposition (SVD)

Defining SVD

Any matrix $A^{M\times N}$, where $M \geq N$, M being the number of rows and N the number of columns, can be written as the product of an $M \times N$ column orthogonal matrix U, an $N \times N$ diagonal matrix W with positive or zero elements, and the transpose of an $N \times N$ orthogonal matrix V as shown below [26]:

$$A^{M\times N} = U^{M\times N} \cdot W^{N\times N} \cdot (V^T)^{N\times N} \tag{6.14}$$

and

$$W^{N\times N} = \begin{pmatrix} w_1 & & & & \\ & w_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & w_N \end{pmatrix}$$

The diagonal elements of the matrix W are known as the singular values and this decomposition of the matrix A is known as the Singular Value Decomposition. The matrices U and V are orthogonal matrices, while their columns are orthonormal.

For the case where $M < N$ in the above matrix A, the decomposition in

equation 6.14 holds with the condition that

$$w_j = 0 \qquad \forall j = (M+1) \cdots N.$$

The corresponding columns of U are also zero, while the condition of orthogonality of the columns hold for all the rows up to the M^{th} row.

Relevance of SVD to rank-order codec

While decoding the rank-order codes, we do a progressive reconstruction of image $\mathcal{I}$ in equation 6.11, making the set of equations pass through all the three possible phases, whereby, at first, $M < N$, and the number of equations (M) is less than the number of variables (N). This is a case of an **undetermined** set of equations with $(N - M)$ different solutions, thus giving rise to an uncertainty. In such cases, SVD gives a solution that is 'almost unique' for similar permutations of the columns U, elements of W and columns of V [26].

The number of equations continues to increase with progressive reconstruction using each incoming spike, while the number of variables, N, remains constant. For the case when $M = N$, the set of equations should have a unique solution. However, if the matrix F in equation 6.11 is ill-conditioned, there may be deviations in the solution, which may be further magnified with inexact input values in the matrix C. SVD deals with such ill-conditioning in the matrix F by giving a **least squares solution** to the array $\mathcal{I}$.

As the number of equations exceed the number of variables, $M > N$. This is the case where we have more equations than unknowns and is known as the **overdetermined** set of equations. Ideally, there might be linear dependencies among the set of equations, and then, any N of the total M equations may give an exact solution. However, in most real applications, there is no linear dependency and the set of equations are inconsistent, having no definite solution. Further, with a non-ideal input matrix C in equation 6.11, SVD ensures a least square solution to $\mathcal{I}$.

SVD makes the inversion of a rectangular non-singular matrix straightforward and yields the pseudo-inverse of the matrix. For singular or near singular matrices, the pseudo-inverse gives the least squares solution to a set of equations as discussed in the following sections.

6.1.6 Pseudo-inverse using SVD

Definition

If A is an $M \times N$ matrix, then there exists an $N \times M$ matrix A^+ , known as the pseudo-inverse of A , which is derived by applying an inverse operation on the SVD of A. Therefore, from equation (6.14) we have

$$\begin{aligned} A^+ &= (U \cdot W \cdot V^T)^{-1} \\ &= (V^T)^{-1} \cdot diag(1/w_j) \cdot U^{-1} \\ &= V \cdot diag(1/w_j) \cdot U^T \qquad (6.15) \\ &\qquad \text{[since } U \text{ and } V \text{ are orthogonal]} \end{aligned}$$

Relevance to rank-order codec

Under conditions of non-singularity of the matrix A, the solution in equation 6.15 is trivial, as the matrices U and V are orthogonal, their inverses are their own transpose, and the matrix W is diagonal so that its inverse is the matrix itself with its diagonal elements reciprocated.

However, if A is ill-conditioned (as is F in our case, discussed in section 6.2), the situation becomes non-trivial. As defined in equation 6.13, the condition number of the matrix W in equation 6.15 is the ratio of the maximum to the minimum singular value in the diagonal matrix W. Thus, if the minimum singular value $w_{min} \simeq 0$ the condition number of the matrix will be infinity. Similarly, if $w_{min} < \gamma$, where γ is the machine's floating point precision, then this will give rise to round-off error problems. In both these cases, the condition number will be near infinity [26]. Both these cases will lead to erroneous solutions. In the case where $A^{M \times N}$ is an ill-conditioned matrix, if $M = N$, some of the elements of the matrix W are zero, or near zero, so that their condition numbers tend to infinity. Further, for the case when $M < N$, many of the singular numbers will be zero. Under such conditions, the set of equations are inconsistent and give rise to multiple solutions. In such cases, SVD is used to find the solution by applying the rule in equation 6.16: When a singular value w_j is zero, its inverse, which is ∞, is replaced by zero. Further, if w_j is so small that it approaches the machine's floating point precision, it is replaced by zero, and is then treated as

in equation 6.16.

$$\begin{aligned} &\text{if} \quad w_j = 0, \\ &\text{or if} \quad w_j < \gamma, \\ &\qquad \text{where } \gamma \text{ is the machine's floating point precision (say),} \\ &\text{then} \quad \frac{1}{w_j} = 0. \end{aligned} \tag{6.16}$$

By doing this, SVD finds the closest solution from the multiple solutions for the set of inconsistent equations and, thus, provides the least squares solution to the problem. However, there is no fixed rule as to 'how small' a singular value has to be before it can be replaced by zero. In real applications, some discretion is exercised in deciding at what threshold to zero the small w_j's, depending on the acceptable deviations in the solution given by the residual. Thus, SVD cannot be applied blindly, and we will discuss the implications of doing so in our empirical analysis in section 6.2. Subsequently, in the same section, we discuss about how we set a threshold parameter that is suitable to our problem, and the results obtained thereby. The general procedure that we follow while finding the pseudo-inverse of the matrix F during progressive image reconstruction is described below.

Procedure

For a matrix $A^{M \times N}$

1. if $M \geq N$:
   Apply SVD to obtain
   $$A = UWV^T \tag{6.17}$$
   else
   Apply SVD to A^T to obtain $A^T = UWV^T$,
   so that
   $$A = VWU^T \tag{6.18}$$
2. Set a threshold parameter Γ for w_{min} so that the condition number n_c is close to unity:

$$n_c = \frac{w_{max}}{\Gamma} \simeq 1, \tag{6.19}$$

where Γ is a constant and is applied using discretion to manipulate the least square error within acceptable limits.

3. Define

$$z_j = \begin{cases} 1/w_j & \text{if} \quad w_j > \Gamma \\ 0 & \text{otherwise} \end{cases} \tag{6.20}$$

4.

$$A^+ = \begin{cases} VZU^T & \text{if} \quad M \geq N \\ UZV^T & \text{otherwise,} \end{cases} \tag{6.21}$$

where $Z = diag(z_j)$.

In the next section, we use the principle of matrix pseudo-inversion using SVD as described in steps 1 – 4 to solve equation 6.11.

6.2 Optimal rank-order decoding: empirical analysis

In this section, we describe how we use the pseudo-inverse of a DoG filter-bank to get $Q_{value} = 1$, i.e. a perfect stimulus reconstruction with the coefficients of filtering. As expected, the Q_{value} for the decoded stimulus using the approximate coefficient values from the LUT also increases for every image compared to the earlier method of reconstruction. In our analysis, we used images of size 32×32 due to computational constraints, which will be discussed more later. Therefore, in our empirical analysis, N = 1024, while M = 1365(rounded). The total number of ganglion cells will be $2M = 2730$, since M represents only the positive coefficients, and hence those cells which fired in the whole population of ganglion cells.

6.2.1 Reconstruction using exact data

We return to where we left off in equation 6.11, and calculate the least squares estimate of $\mathcal{I}$, viz. $\hat{\mathcal{I}}$, by taking the pseudo-inverse of F, which represents the set

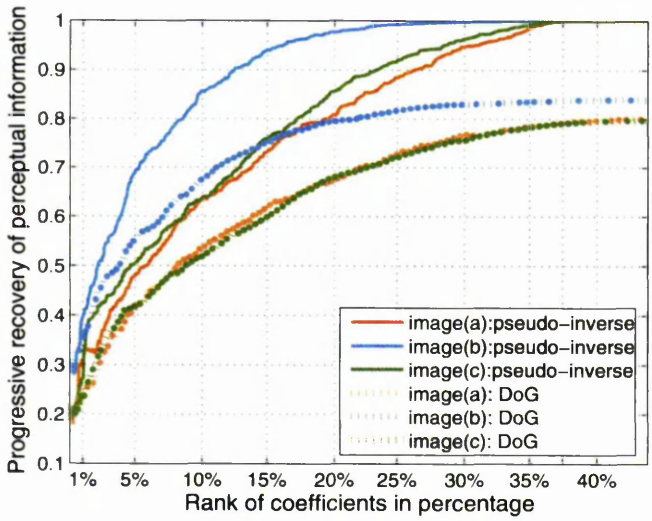
of linear equations in our system, using SVD as in equations 6.17 – 6.21. The set of coefficients of filtering for each image, C, represents the exact input to obtain a solution to the set of equations represented by F. Initially, to observe the system response, we do not throw away any singular values, i.e. we set $\Gamma = 0$ in equation 6.19, and therefore we obey equation 6.16 only for singular values which are zero. We now order the coefficients in C according to their rank and perform a progressive reconstruction of $\mathcal{I}$ as described below:

The elements in C are arranged in descending order of their magnitudes, to represent a progressively decreasing activation level and hence progressively increasing latency of firing. The columns of F are permuted so as to exactly match the columnwise location of $\Phi_s(k,l)$ to that of the coefficient $C_s(k,l)$ in the rank-ordered array C. Let G be a matrix having the same number of columns as the number of rows in F, and whose rows grow progressively. G is initialised as: $G_{row1} = F_{col1}$, where F_{col1} is the first column of F. We want a solution to $G\mathcal{I} = C$, which corresponds to an underdetermined set of equations with $M < N$. First, we apply equations 6.18 – 6.21 (with $\Gamma = 0$ in equation 6.19) to get $\hat{\mathcal{I}}(1) = G^{+}C$. This will correspond to the image reconstruction when only the most stimulated cell has fired one spike. Using Petrovic's method, a quantitative measure $Q_{value}(1)$ of the information content in $\hat{\mathcal{I}}(1)$ with respect to $\mathcal{I}$ is obtained. Next, we assign $G_{row2} = F_{col2}$. Now G has two non-zero rows, and $\hat{\mathcal{I}}(2) = G^{+}C$, whence we obtain $Q_{value}(2)$. On proceeding further, the number of rows in G exceeds N, i.e. now $M > N$, and the pseudo-inverse is calculated using equation 6.17 and equations 6.19 – 6.21 (with $\Gamma = 0$ in equation 6.19). We observe that for a certain number of coefficients κ, where κ is some fraction of M, there is a perfect reconstruction of the input image indicated by $Q_{value}(\kappa) = 1$. The magnitude of κ depends on the image characteristics. We discuss the empirical results of such a reconstruction in the following section.

Results

The plot for progressive recovery of perceptually-important information by this method is shown in figure 6.1(d) for the three images shown in figure 6.1 (a) – (c). Firstly, we observe that we obtain a perfect information recovery using the method of pseudo-inverses to decode the rank-order codes. The full information recovery occurs when around 30 – 35% of the coefficients are used for

(a) (b) (c)



(d)

Figure 6.1: (a)–(c) Three input images. (d) Progressive recovery of perceptual information for the three images when reconstructed with their respective coefficients of filtering using the pseudo-inverse method of rank-order decoding, shown in solid line, and VanRullen and Thorpe's method of decoding, shown in dashed lines.

reconstruction. The variation in the number of coefficients required for full information recovery may be attributed to image characteristics, where images with fewer edges are recovered faster (for example the image in figure 6.1 (b)). Information recovery plots using VanRullen and Thorpe's method of decoding the same three images in figure 6.1 (a) – (c) are also shown in figure 6.1(d) for comparison of the performance of the two methods of decoding. We observe that, secondly, using VanRullen and Thorpe's method of decoding, information recovery is around 15 – 20% lesser at the time of full information recovery using the method of pseudo-inversion for decoding. Thirdly, the initial rises of the plots in figure 6.1(d) imply that the rate of information recovery is much faster for all the three images when decoded using the pseudo-inverse method as compared to VanRullen and Thorpe's method.

A full information recovery using the coefficients of filtering indicates that the loss in information recovery during decoding (please refer to figure 1.1) is completely removed by using the pseudo-inverse method. We now apply this method in rank-order decoding where the value of the coefficients are lost and the input image is reconstructed using an LUT for assigning weights to each incoming spike in decreasing sensitivity. This is discussed in the next section.

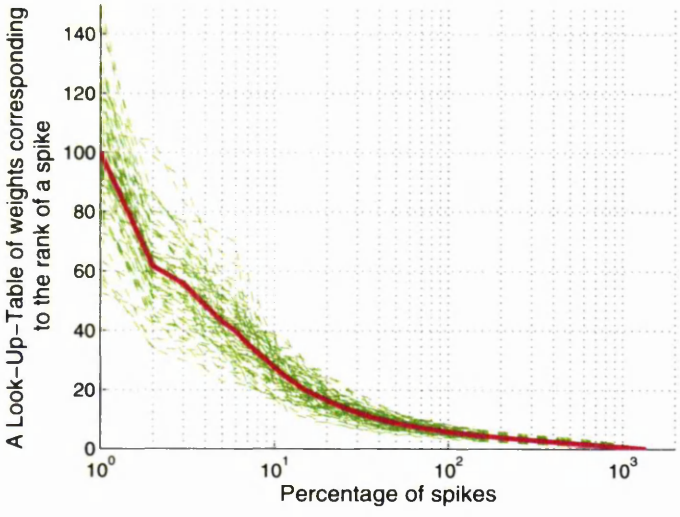
6.2.2 Reconstruction using approximate data

Constructing a LUT

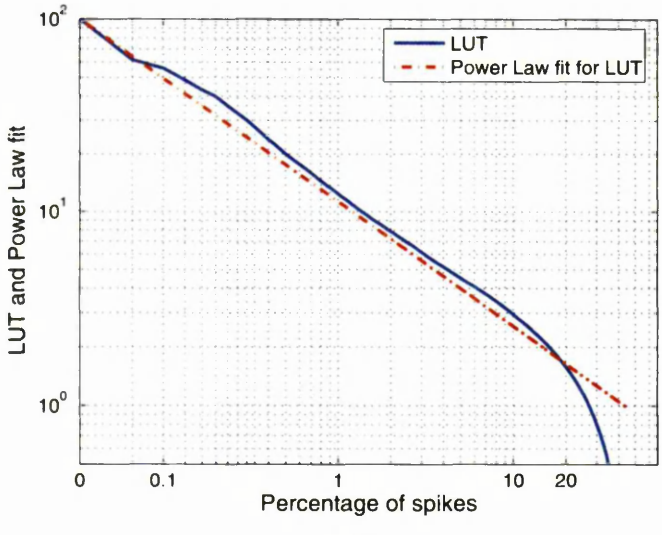
We now move on to the true rank-order decoding of the visual stimulus, where we have lost the coefficients of filtering and prepare a look-up-table (LUT) of coefficients in a similar manner to that described in section 4.1.4. The LUT for this system of decoding is shown in figure 6.2. Our observation from the LUT is that the characteristics of the distribution is unchanged from that shown in figures 4.6(a) and 4.7, and which were generated using images of size 128 × 128. Thus, the LUT distribution is scale-invariant for any given set of images, which is a characteristic for a distribution following the power law (discussed in section 4.1.4).

Results

We use the values from the LUT to do a progressive reconstruction as mentioned in section 6.2.1, by progressively adding rows to G and doing $\hat{\mathcal{I}}(t) = G^{+}LUT(t)$.



(a)



(b)

Figure 6.2: A look-up-table of weights when rank-order decoding is being done using the pseudo-inverse of the filter-bank of DoG matrices.

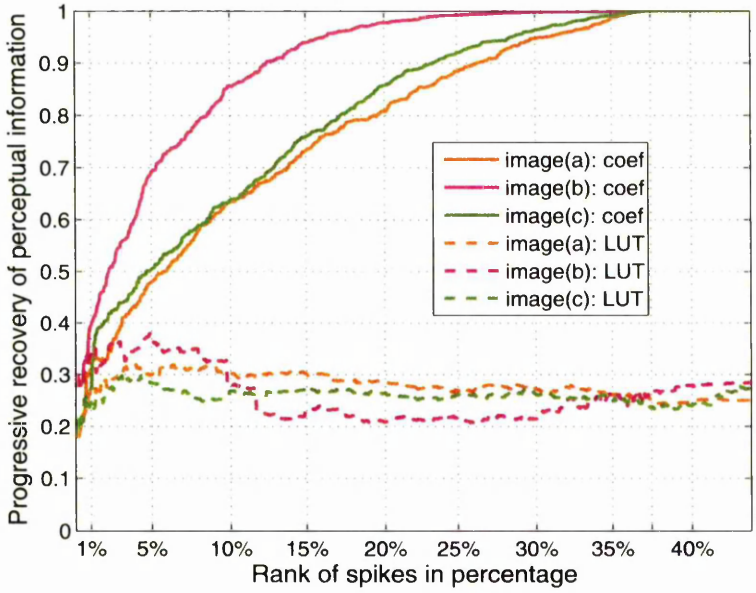


Figure 6.3: Progressive recovery of perceptually-important information for the three images, shown in figure 6.1 (a) – (c), when reconstructed with the LUT and using the pseudo-inverse method of rank-order decoding, shown in dashed lines, compared to the information recovery using the same method of decoding with the respective coefficients of filtering of the images, shown in solid lines.

The progressive recovery plot for three images is shown in figure 6.3 with a comparison of the plot using the coefficients of filtering for the respective images. The poor recovery using the LUT may be attributed to the ill-condition of G. From the plot characteristics and referring to the discussion in section 6.1.6, we observe that the solution to $\hat{\mathcal{I}}$ is far from the exact or least squares solution, the exact solution at a certain rank being indicated by the corresponding point in the recovery plot using the coefficients of filtering in figure 6.3. This is the result of applying SVD blindly in solving G by setting the parameter $\Gamma = 0$ in equation 6.19. In the following section, we apply some discretion in deciding the value for the threshold parameter Γ so that it follows the condition set in equation 6.19.

Deciding on threshold parameter Γ

Table 6.1 shows the condition number of the matrix G during the progressive recovery of two images, while figure 6.4 shows the plot of the singular values during the several stages of the progressive reconstruction of the images. For both the images, the deviation of the condition number from the ideal value of unity is very high in regions of $400 < M < 1200$. At the two extremes, the condition number is seen to be closer to unity. Further, the degree of ill-condition varies with the permutation of the equations, i.e. the choice of columns in the matrix F. In our case, we have a fixed permutation of the columns of the matrix F — that of the rank-ordered coefficients. Any other order would not make sense to our application. Within such constraints, some of the images perform worse than the others as the particular permutation of columns make G more ill-conditioned. As seen in table 6.1, the condition number of image (a) are much more fluctuating than that of image (b).

It is under these conditions that we apply discretion and set a threshold value Γ which makes the information recovery plot for a certain image close to its own least squares solution plot. For doing this, we study the nature of the singular values of G during progressive reconstruction of images which is shown in figure 6.4. The plots for the singular values of image(b) show that on average, around 100 or more singular values have values less than 0.05 for the underdetermined case, whereas the figure is 50 or more for the overdetermined case. The figures are much improved for image (a). However, the maximum singular value has a consistent range of around 1.8 for both the images. Further, most of the singular values for each of the overdetermined and the underdetermined cases,

Condition number of $G^{M\times N}$ during progressive recovery of image(a)

case 1: Underdetermined (no. of rows (M) < no. of columns (N = 1024))

M	950	850	750	650	550	450	50
w_{max}	1.7656	1.7515	1.7463	1.7186	1.7002	1.6919	1.4356
w_{min}	9.58×10^{-5}	8.8×10^{-4}	1.6×10^{-3}	2.3×10^{-3}	0.0031	0.0041	0.0309
$cond(G^{M\times N})$	1.8×10^{4}	1.98×10^{3}	1.11×10^{3}	735	554	410	46.44

case 2: Overdetermined (no. of rows (M) > no. of columns (N = 1024))

M	1050	1100	1150	1200	1250	1300	1350
w_{max}	1.7734	1.7800	1.7973	1.7984	1.8132	1.8149	1.8244
w_{min}	5.38×10^{-6}	7.03×10^{-5}	1.6×10^{-4}	4.7×10^{-3}	6.7×10^{-4}	0.0640	0.2875
$cond(G^{M\times N})$	3.2918×10^{5}	2.53×10^{4}	1.12×10^{4}	3.8×10^{3}	2.67×10^{3}	28.34	6.34

Condition number of $G^{M\times N}$ during progressive recovery of image(b)

case 1: Underdetermined (no. of rows (M) < no. of columns (N = 1024))

M	950	850	750	650	550	450	50
w_{max}	1.816	1.812	1.81	1.8	1.79	1.78	1.63
w_{min}	3.1×10^{-5}	2.8×10^{-4}	0.001	0.003	0.004	0.005	0.1
$cond(G^{M\times N})$	5.8×10^{4}	6.2×10^{3}	1.63×10^{3}	634.76	473.9	316.3	14.95

case 2: Overdetermined (no. of rows (M) > no. of columns (N = 1024))

M	1050	1100	1150	1200	1250	1300	1350
w_{max}	1.8183	1.8196	1.8209	1.8231	1.8237	1.8246	1.8251
w_{min}	9.3×10^{-8}	1.09×10^{-5}	2.3×10^{-5}	6.6×10^{-5}	1.2×10^{-4}	0.002	0.03
$cond(G^{M\times N})$	1.9×10^{7}	1.6×10^{5}	7.7×10^{4}	2.7×10^{4}	1.4×10^{4}	897.6	55.64

Table 6.1: Table showing the condition of DoG filter-bank matrix F for both overdetermined and underdetermined cases during progressive recovery of two images.

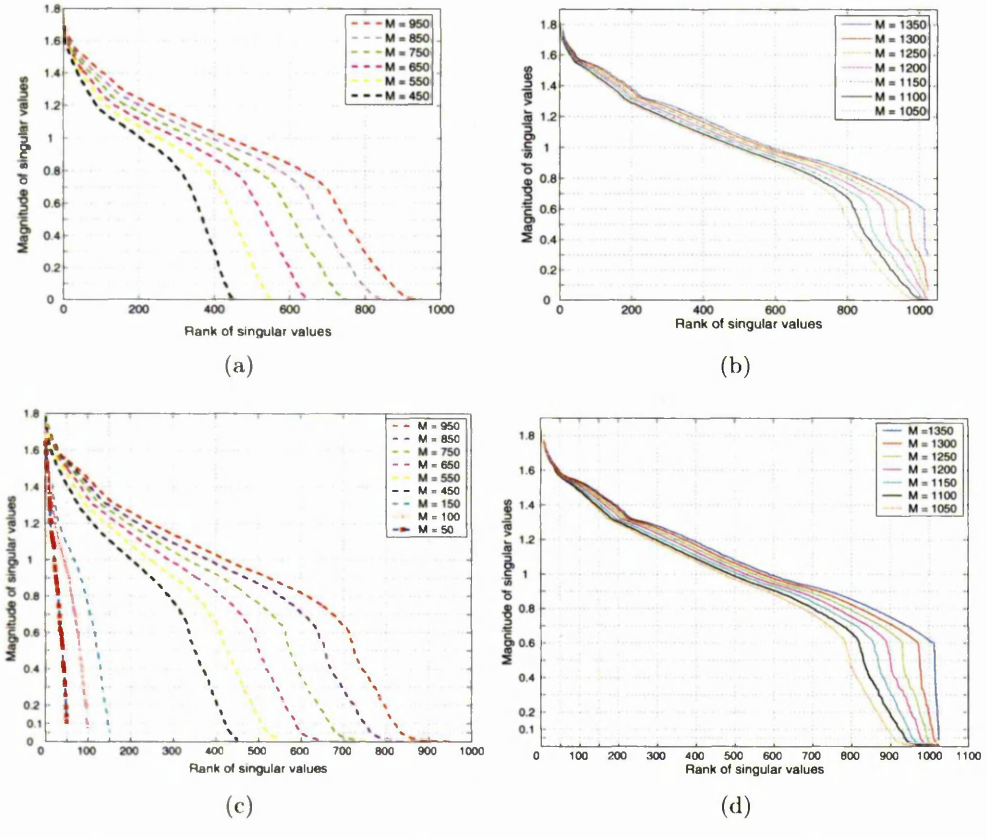


Figure 6.4: Plots showing the singular values of the filter-bank matrix F during progressive recovery of images (a) and (b) in figure 6.1. The singular values for the **underdetermined** case during progressive reconstruction of images (a) and (b) in figure 6.1 are shown here in (a) and (c) respectively. The singular values for the **overdetermined** case during progressive reconstruction of images in figure 6.1 (a) and (b) are shown here in (b) and (d) respectively.

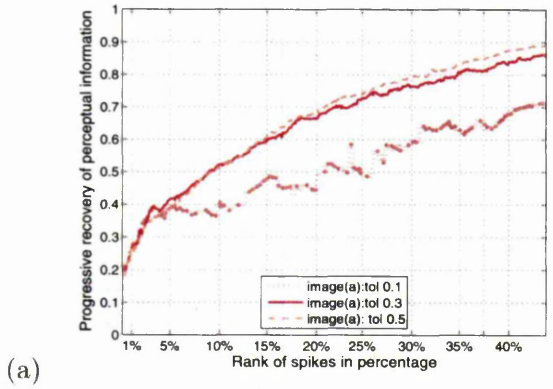
for both the figures are greater than 0.6, after which there is a steep descent of the curves to 0, indicating that very few singular values lie within this range and can be ignored. Thus, we may decide on a value of $\Gamma \leq 0.6$, giving a condition number $n_c \geq \frac{1.8}{0.6} = 3$ which is fairly close to unity.

Based on such a study, we set the threshold parameter Γ to three values viz. 0.1, 0.3 and 0.5 and observe the information recovery plot for the three images shown in figure 6.1 (a) – (c). The condition number n_c in each of the above cases will be 18, 6 and 3.6 respectively. Thus, for each case, all singular values less than Γ are set to zero prior to calculating G^+ using equations 6.17–6.21. The plots for the progressive reconstruction of each of the three images, and for each of the three values of Γ, are shown in figure 6.5. We do this for all the images in our data-set of sixty-five images. From our results, we observe (as is also seen in figure 6.5) that using $\Gamma = 0.3$ gives a better information recovery than $\Gamma = 0.1$. Further using $\Gamma = 0.5$ does not produce significant improvement in the information recovered than using $\Gamma = 0.3$, and rather shows a lesser efficiency in information recovery for the image in figure 6.1 (b). Based on these observations, we use a value of $\Gamma = 0.3$ as it is observed to give optimal information recovery for almost every image in our data-set.

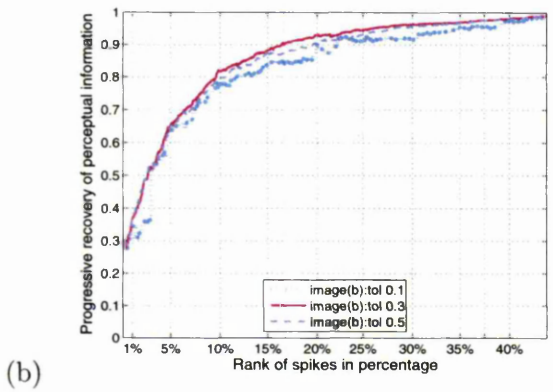
Results

Information recovery plots for the three images in figure 6.1 (a) – (c) using a threshold of $\Gamma = 0.3$ and the method of pseudo-inverse for rank-order decoding is shown in figure 6.6. In our simulation we use images of size 32×32 due to time and memory constraints. Information recovery plots for the three images using VanRullen and Thorpe's method of decoding are also shown in the same figure for comparison. We observe that the perceptual information recovered is on average $10 - 20\%$ higher using the pseudo-inverse method of decoding as compared to VanRullen and Thorpe's method of decoding. Further, the rate of information recovery is higher using the pseudo-inverse method of decoding than by VanRullen and Thorpe's method. Scaling down the image sizes leads to essentially the same behaviour although the exact value of information content may vary for different sizes of a certain image (not shown here). This is expected to hold even when the images are scaled up.

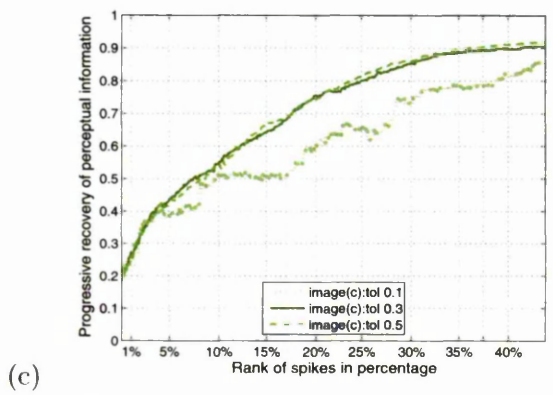
The information recovery plot for all the images in our data-set using the pseudo-inverse method of rank-order decoding discussed so far in this chapter is



(a)



(b)



(c)

Figure 6.5: (a) – (c) Information recovery plot for the three images shown in figure 6.1 (a) – (c) respectively, using the pseudo-inverse method of decoding and with values from the LUT. For each image, reconstruction is done using the LUT and with the threshold parameter Γ (referred to as 'tol' in the plots) set at 0.1 (shown in dotted lines), 0.3 (shown in solid lines) 0.5 (shown in dashed lines).

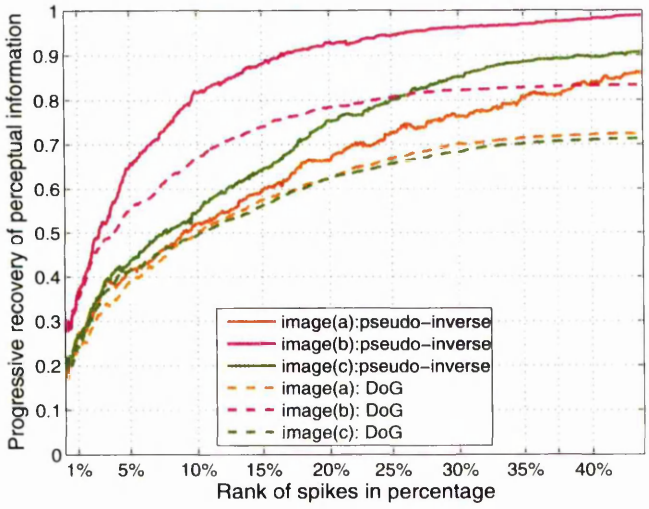
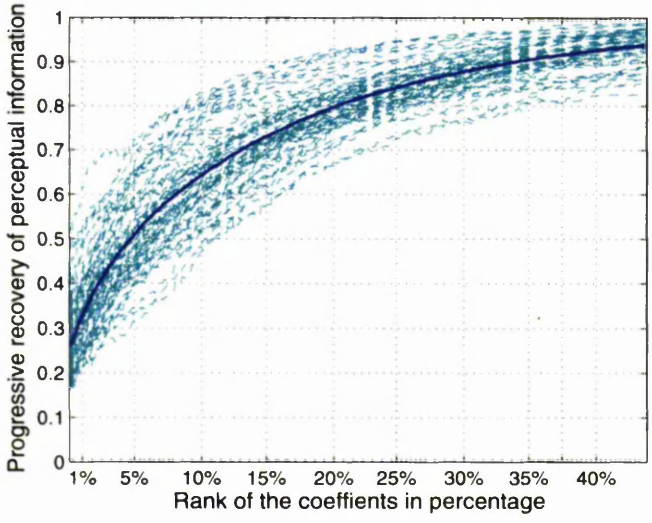


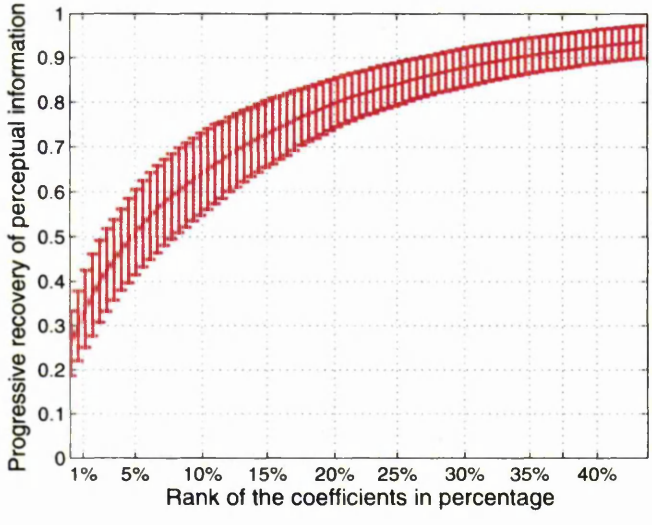
Figure 6.6: Information recovery plot for three images, shown in figure 6.1 (a) – (c), using the pseudo-inverse method of decoding, the threshold set as $\Gamma = 0.3$, using the LUT, shown in solid lines. The information recovery plots for the same images using VanRullen and Thorpe's method of decoding with the LUT is shown in dashed lines for comparison.

shown in figure 6.7(a), as is the mean information recovery plot. The plot of the same data as the standard deviation about the mean is shown in figure 6.7(b). Similar plots for VanRullen and Thorpe's method of decoding using images of size 32×32 are shown in figure 6.8(a) and 6.8(b) respectively for comparison. We observe that on average more than 95% of the information in an input image can be recovered using the pseudo-inverse method of decoding, the deviation being 90 – 98% across various images in the data-set of sixty-five images. For VanRullen and Thorpe's method of decoding, the total information that can be retrieved is 80% after which the curve saturates. The deviation across the images at saturation is 75 – 85%. Thus, the total information recovery from rank-order encoded images is increased by more than 15% by improving on decoding technique. Comparing figures 6.7(b) and 6.8(b), we observe that the standard deviation of the information recovery from individual images is much more using VanRullen and Thorpe's method of decoding than that using the pseudo-inverse method.

A comparison of the average rate of information retrieval using the two methods of decoding is shown in figure 6.9. We observe that the initial rate of recovery

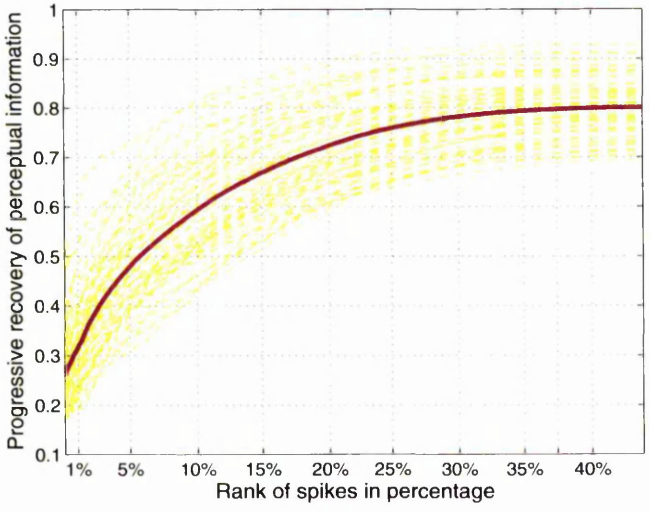


(a)

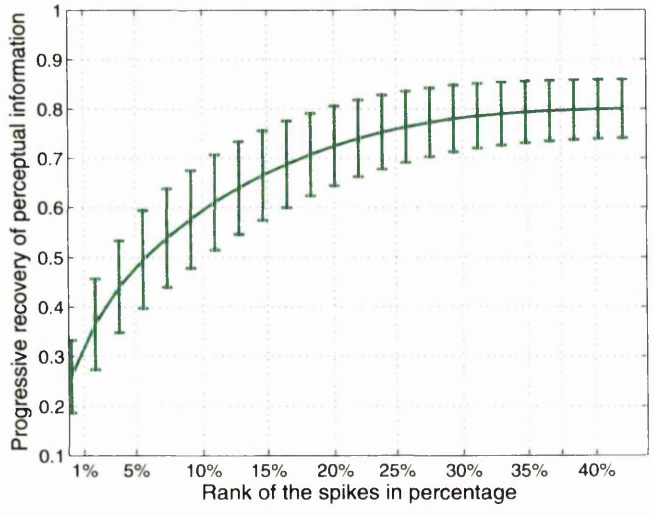


(b)

Figure 6.7: (a) Individual information recovery plots for the sixty-five images in our data-set (shown in dashed lines) as a spread about the mean information recovery plot (shown as solid line) using the pseudo-inverse method of decoding with LUT, and with the threshold parameter set at $\Gamma = 0.3$. (b) The same data presented as a mean, with the standard deviation shown as errorbars.



(a)



(b)

Figure 6.8: (a) Individual information recovery plots for the sixty-five images in our data-set (shown in dashed lines) as a spread about the mean information recovery plot (shown in solid line) using VanRullen and Thorpe's method (with images of size 32×32) of decoding with LUT (b) The same data presented as a mean, with the standard deviation shown as errorbars.

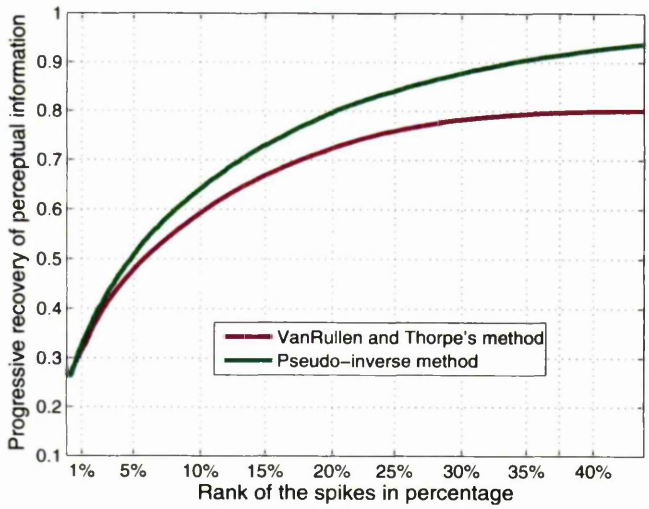


Figure 6.9: (a) Comparison of the mean information recovery using VanRullen and Thorpe's method and pseudo-inverse method of decoding a rank-order encoded stimulus. The decoding is done using the LUT.

till about the first 1 – 2% of the spikes is same for the two methods. Thereafter, the information recovery plot for the pseudo-inverse method of decoding gets steeper so that by the time 5% of the spikes have arrived, there is a 1% increase in the information recovered using this method as compared to VanRullen and Thorpe's method of decoding. By the time 10% spikes have fired, the difference in information retrieval increases to 5%, and steadily increases until when 30% of the spikes have fired, there is an increase of 10% in the recovered information using pseudo-inverse method of decoding compared to that of VanRullen and Thorpe. Overall, using the pseudo-inverse method of rank-order decoding has improved the quantity as well as the rate of information recovery from the rank-order codes.

6.3 Chapter summary

In this chapter, we discuss how VanRullen and Thorpe's method of stimulus reconstruction is sub-optimal for information recovery from a reconstructed image. We go on to describe how we obtain perfect reconstruction using the coefficients of filtering and the pseudo-inverse of a DoG filter-bank, which is not possible

using VanRullen and Thorpe's method where the DoG is used as its own inverse as an approximation. The difficulty of using the pseudo-inverse method of rank-order decoding with approximate data from the LUT is discussed. We describe how application-specific discretion is required while applying the Singular Value Decomposition technique for obtaining the pseudo-inverse of an ill-conditioned matrix, and how we have solved the problem successfully.

A comparison of the information build-up using both VanRullen and Thorpe's method and the pseudo-inverse method of rank-order decoding is shown. The results show that there is a $10-15\%$ increase in the information recovery from a reconstructed stimulus by the time $35-40\%$ of the cells fire their first spikes. We also observe that the information recovery does not saturate even after 40% of the cells have fired their first spikes, indicating that the total information recovery may well be nigh 99% for most of the images by the time the curve saturates.

The simulation is done using images of size 32×32. The drawback of the method is that it needs a large amount of memory for storing and working with the filter-banks. Also, the computation time of the pseudo-inverse of a matrix is very long. Moreover, with our progressive recovery, we have to calculate pseudo-inverses of the DoG filter-bank by progressively adding columns. The total time for plotting the information recovery curve for such a progressive recovery takes more than 12 hours for a 32×32 image on a desktop computer.

Progressive information recovery plots for images of size 8×8 and 16×16 show similar relative information recovery among the images, indicating that the method is scalable and produces consistent results for images of different sizes although the absolute values of the information recovered varies. It is thus speculated that these characteristics will also hold for scaled-up images. This is left to be done as a future work.

Since we get a perfect reconstruction using the coefficients of filtering, we can say that by minimising the information loss during decoding, we have maximised the information recovery from rank-order codes. From this data, we get an estimate of the information loss suffered during encoding. However, such a recovery is delayed until approximately 37% or more cells have fired their first spikes. Since rank-order code hypothesis is essentially based on fast information recovery, such delayed recovery does not hold in favour of rank-order codes. Moreover, the pseudo-inverse method of decoding is computationally expensive both in terms of time and machine memory. These factors led us to explore ways

of **optimising** the information recovery, whereby the information recovered will be maximum possible in minimum time — a deterioration from perfection using the pseudo-inverse method of decoding, but an improvement on VanRullen and Thorpe's lossy rank-order 'codec' — as discussed in the following chapters.

Chapter 7

Optimising Information Recovery

In the previous chapter, we used the pseudo-inverse method of rank-order decoding to obtain a perfect reconstruction of the input image using the coefficients of filtering of the input image with a DoG filter-bank, thereby maximising the information recovery. For true rank-order decoding with LUT, there was a 10% increase in the overall quantity of information recovered from the rank-order encoded images as compared to that with VanRullen and Thorpe's method of decoding. Moreover, the rate of information recovery with the first few spikes increased with such an improvement in the decoding method. However, the time to maximum information recovery is delayed until approximately 37% or more cells have fired their first spikes. Such delayed information recovery defeats the essence of the rank-order codes — rapid information recovery — and is therefore outside the permissible 'time-limits' of rank-order decoding; the ideal would be (say) the time until around 20% of the cells have fired their first spikes. Moreover, the pseudo-inverse method of decoding is very demanding computationally, and is difficult to implement with image sizes larger than 32×32 on a desktop computer. This led us to find ways of optimising the information recovery where the information loss suffered would be minimised in such a manner as to optimise the information recovery in the sense that maximum 'possible' information is recovered in minimum 'possible' time.

The DoG filters in VanRullen and Thorpe's retinal model were assumed to be orthogonal in spite their non-orthogonality. We suggest that this approximation can be rectified by correcting the coefficients of filtering in equation 4.4, so that they are stripped of their effects of filter overlap. In doing so, we would also act in accordance with the laws of nature. This is because, due to the overlap of

the neighbouring filters while sampling the image, the information from a certain locality is picked out by many filters at once. This gives rise to redundancy in the data by representing a piece of information using more coefficients of filtering than are actually required. When rank-ordered, these coefficients will tend to appear consecutively, thus occupying the higher ranks with redundant information, while pushing the coefficients from a different locality, which are next in order in terms of importance of information content, down the rank-order. Thus, although the local retrieval of data is of high fidelity, the overall data recovery is bound to be delayed. This contradicts the motive of the rank-order encoding which is fast information retrieval, i.e. getting the maximum information from each additional spike. Such an over-representation of data is very similar to the redundancy observed in natural images, which our sensory systems handle using the mechanism of lateral inhibition [68, 2, 82, 55, 42]. Therefore, we anticipate that applying the principles of lateral inhibition will take away the redundancy in our coefficient set too, thus aiding a proper rank-ordering of the coefficients, so that the coded information can be retrieved optimally.

In fact, the pseudo-inverse method of decoding essentially corrects the coefficients of filtering for the non-orthogonality of the filters. This it does while finding a least squares solution for the set of equations, and thus is done as a part of rank-order decoding. Alternatively, we can improve VanRullen and Thorpe's method of rank-order encoding by using the principles of lateral inhibition as argued above. In this chapter, we present such an improvement of VanRullen and Thorpe's method of rank-order encoding, and in doing so, we propose a novel algorithm — **Filter-overlap Correction algorithm** (FoCal) — to correct the coefficients of filtering in VanRullen and Thorpe's retinal model for overlap of the DoG filters. We call such an encoder the 'Filter-overlap Corrected Rank-order encoder' (FoCRen) and is shown in figure 7.1. The decoding used is the same as used by VanRullen and Thorpe in their retinal model. Thus the filter-overlap corrected rank-order 'codec' shown in figure 7.1 is essentially the same as the rank-order 'codec' in figure 1.1, with a modification in the rank-order encoder block by the addition of FoCal.

We introduced the concepts of redundancy and lateral inhibition in section 3.4. In section 7.1, we do a theoretical analysis of FoCal and its relation to the lateral inhibition mechanism. In section 7.2, we present the empirical analysis of the performance of FoCRen. In section 7.3, we present the Matching Pursuit algorithm,

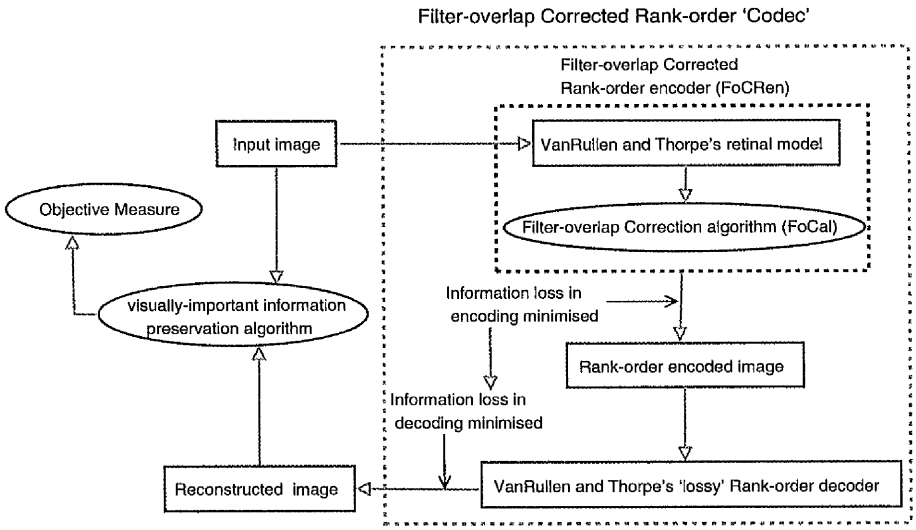


Figure 7.1: Optimising information recovery by modifying the rank-order encoding mechanism of VanRullen and Thorpe with the addition of a Filter-overlap Correction algorithm.

which is mathematically similar to FoCal, yet very different in its application to image reconstruction using VanRullen and Thorpe's retinal model.

7.1 Lateral inhibition for rank-order encoding

Having reviewed the lateral inhibition mechanism used as a redundancy reduction technique by the retinal cells in sections 3.4.1 and 3.4.2, we now proceed to apply this concept to correct the redundancy introduced due to overlap of neighbouring DoG functions in VanRullen and Thorpe's retinal model, and thus try to improve their rank-order encoding mechanism.

Since the neighbouring filters in VanRullen and Thorpe's retinal model overlap, they sample the same information to a considerable extent. Thus, these neighbouring cells tend to fire at around the same time, which means that the coefficients of filtering are ranked in close vicinity to one another. As such, there is more information from a locality in the picture than is required, which is, therefore, redundant. This redundancy causes vital information from another pixel and its locality, which have a lesser contrast than the previous locality, to be pushed further down in rank in the ordered array of coefficients. The rank-order

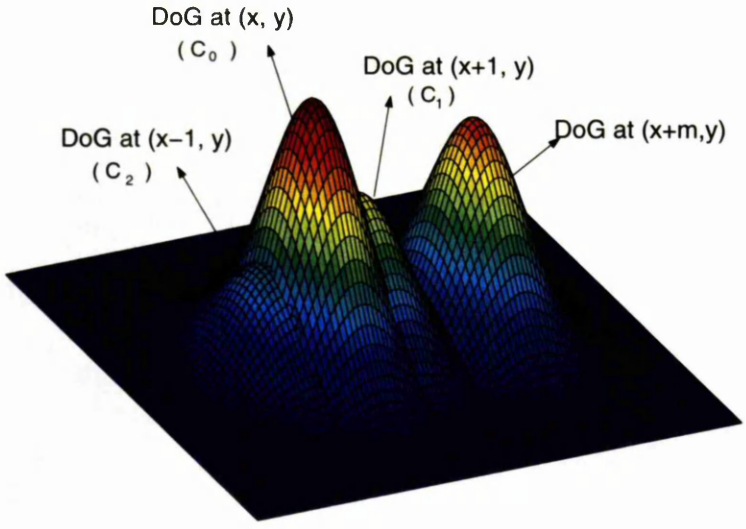


Figure 7.2: The effect of suppression by the earliest firing spike in a locality on those fired later, and thus suppressing redundant information. This inhibitory influence of the largest on the later firing spikes enhances the order of other cells in another locality which might be carrying more important information.

encoding is thereby distorted, and the main benefit of this type of neural encoding, viz. retrieving the perceptually-important information in minimum time, is lost.

To correct this discrepancy, we introduce the biological principle of lateral inhibition to the model. The idea is that the largest coefficient in a locality corresponds to a neuron which is stimulated by the point of maximum contrast in the locality and hence fires first. In doing so, it imposes the strongest inhibition on its neighbouring cells, and suppresses their response. This suppression allows cells corresponding to important information from a distant locality to fire earlier. The corresponding coefficients of filtering will thus move higher up in rank.

The concept is illustrated in figure 7.2. Let (x, y) be the spatial location corresponding to the strongest stimulus in its neighbourhood. We denote the DoG centred at (x, y) by Φ_0, with two neighbouring and highly overlapping filters centred at $(x + 1, y)$ and $(x - 1, y)$ and denoted by Φ_1 and Φ_2 respectively (only two are shown here for clarity). Considering t_0 as a certain initial time, the cell corresponding Φ_0 will fire earliest in time after t_0, compared to those corresponding to Φ_1, Φ_2, and other neighbouring cells. Let C_0 be the coefficient of filtering by Φ_0 at (x, y). Since the cells corresponding to Φ_1 and Φ_2 fire later

in time, both C_1 and C_2 are less than C_0 in magnitude. Figure 7.2 shows the DoG at a certain location scaled up by the corresponding coefficient of filtering so that their peaks represent the latency of firing of the corresponding cells in reverse order. Due to their high degree of overlap with Φ_0, the filters Φ_1 and Φ_2 pick up almost the same information from the image and thereby tend to fire soon after Φ_0 and within short intervals of one another. This means that even though the coefficients C_1 and C_2 carry redundant information, they will be very close in magnitude to C_0, and thus will tend to be grouped together with C_0 in the rank-ordered data-set. This, in turn, will cause the cell corresponding to the DoG Φ_m, say, at the spatial location (x_m, y_m) to fire later in time, in spite of it carrying information that is more important than C_1 and C_2, which are carrying redundant and hence unnecessary data.

Let us consider the filters Φ_0 and Φ_1 for the time being. The overlap between the filters can be represented by the the degree of correlation $\langle \Phi_0, \Phi_1 \rangle$ between them. Due to its overlap with filter Φ_0, the coefficient C_1 carries redundant information which is equivalent to the correlation term, mentioned above, scaled up by the coefficient of filtering C_0. Thus, the additional information r_1 of the DoG at the location $(x+1, y)$ would be written as:

$$r_1 = C_1 - C_0.\langle \Phi_0, \Phi_1 \rangle. \tag{7.1}$$

By a similar argument, the additional information r_0, corresponding to the DoG Φ_0 would be written as:

$$r_0 = C_0 - C_1.\langle \Phi_0, \Phi_1 \rangle. \tag{7.2}$$

Interestingly, we observe that equations 7.1 and 7.2 resemble equations 3.1 and 3.2. The coefficients of filtering C_1 and C_0 are directly proportional to the input contrast and represent the input stimulus terms e_1 and e_2 in equations 3.1 and 3.2 respectively. The correlation term $\langle \Phi_0, \Phi_1 \rangle$ corresponds to K, the coefficient of inhibition. The threshold frequency r^0 is irrelevant here, as the method of rank-order encoding depends on the firing latency, rather than the firing frequency. These equations can be interpreted as the two cells corresponding to the filters Φ_0 and Φ_1 mutually inhibiting one another, indicating that the principle of lateral inhibition can be incorporated in VanRullen and Thorpe's retinal model by correcting the coefficients of filtering for the effects of filter-overlap.

Now, since $\langle \Phi_0, \Phi_1 \rangle > 0$ due to overlap, and $C_0 > C_1$, therefore $r_1 < C_1$, resulting in a suppression of C_1 by C_0. Although there is a collective inhibitory effect on the cell corresponding to the DoG Φ_0 similar to that shown in equation 3.3, resulting in a deduction of the absolute value of the coefficient C_0, which might influence its ranking, yet, for the time being we ignore the inhibitory effects on the largest coefficient of filtering in a locality, and consider the suppression by the largest of all the others smaller than it. This is similar to the 'Winner Take All' mechanism and gives a credibility to our approach. However, due to this simplification, the effect of filter-overlap is not completely eliminated and some redundancy is left in the data.

Thus, effectively, by virtue of being the largest among all the coefficients that correspond to filters overlapping with the filter Φ_0, C_0 applies a lateral inhibition on all these other coefficients. As this causes a reduction in the effective response r for the individual cells, it enables the uninhibited cell corresponding to the DoG Φ_m, say, at the spatial location (x_m, y) to fire after C_0. A similar suppression occurs for the co-efficients surrounding the location (x_m, y) and overlapping with the DoG Φ_m, allowing a cell at (x_n, y_n), say, to fire next, and so on. This brings a fairness to the rank-ordering of the coefficients by preserving the priority of the information carried by them, so that most of the perceptually-important information can be retrieved from the first few coefficients, corresponding to the neurons which fire their first spikes within a very short time, and thus simulating the fast information processing by the retina.

In the next section, we introduce the Filter-overlap Correction algorithm, and discuss how we have used it to improve the rank-order encoding mechanism used by VanRullen and Thorpe for their retinal model, thereby introducing FoCRen.

7.1.1 Filter-overlap Correction algorithm (FoCal)

We apply FoCal to the set of rank-ordered coefficients of filtering $\{r_i \in R : i = 1 \ldots P\}$ of VanRullen and Thorpe's retinal model, introduced in section 4.1.3. Let $\xi = \emptyset$ be initially an empty set that grows with each iteration of the algorithm. The steps through the algorithm are as listed below:

1. The largest coefficient r_1 of the array R is taken out and added to ξ:

$$R = R - r_1$$

$$\xi = \xi \cup r_1$$

2. Each of the remaining elements in $\{r_i \in R : i = 2 \ldots P\}$ is now corrected for overlap of its corresponding DoG filter with the DoG filter corresponding to r_1. Thus, if Φ_s^1 is the filter corresponding to r_1 and Φ_s^2 is the filter corresponding to r_2, then r_2 is corrected for filter-overlap thus:

$$r_2 = r_2 - r_1.\langle \Phi_s^1, \Phi_s^2 \rangle$$

   This is done for each of the remaining coefficients in the set R.

3. The elements in R are now re-arranged in descending order of magnitude.

4. Steps 1 to 3 are repeated $P - 1$ times.

5. The algorithm is stopped when there is only one coefficient left in R at step 3, which will be after $P - 1$ iterations. This element is added to the set ξ, which now has P elements and is the set of filter-overlap corrected coefficients.

In the next section, we discuss the empirical results of applying FoCal to VanRullen and Thorpe's retinal model, thereby getting an improved rank-order encoder, which we call 'Filter-overlap Corrected Rank-order encoder' (FoCRen), shown in figue 7.1.

7.2 Filter-overlap Corrected Rank-order encoder (FoCRen)

Each of the sixty-five images in our data set is processed using VanRullen and Thorpe's retinal model and their respective coefficients of filtering corrected for filter-overlap using FoCal introduced in section 7.1.1. Thus, we obtain a rank-order encoded image using FoCRen. As we shall see in the following sections, this method of encoding minimises the information losses suffered during encoding.

The decoding in the filter-overlap corrected rank-order 'codec' is being done using VanRullen's method of rank-order decoding, which is lossy due to the approximation of decorrelation of filters. However, the improvement in rank-order encoding using FoCal actually minimises the correlation among the filters, so

that the error due to approximation during decoding is minimised, and hence the information loss during decoding is also minimised. Thus, using FoCRen, the overall information loss in the filter-overlap corrected rank-order 'codec' is minimised.

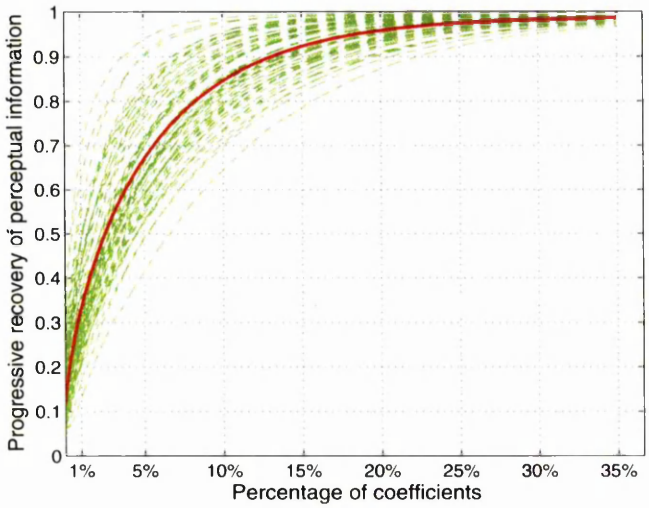
Each of the images rank-ordered encoded using FoCRen is progressively reconstructed using rank-order decoding as used by VanRullen and Thorpe in their retinal model. The results of reconstruction using coefficients and LUT are discussed in the following sections.

7.2.1 Reconstruction using coefficients

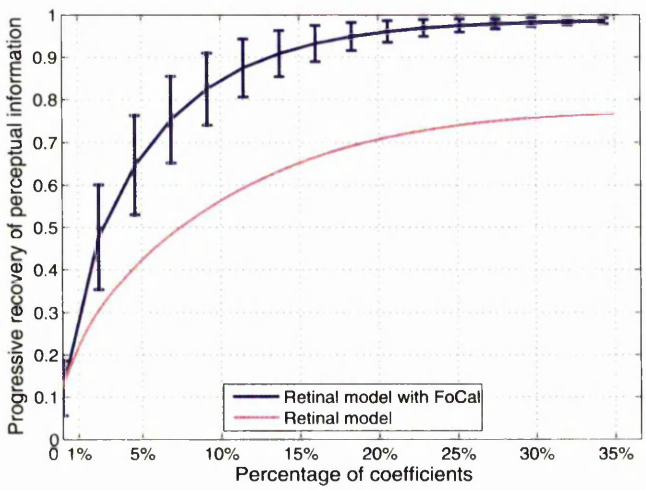
Plots of information recovery for each of the images are shown in figure 7.3(a) as a spread about the mean information recovery plot. It is observed that, firstly, around 99% of the information contained in the input image can be retrieved on average from images rank-order encoded using FoCRen. Moreover, this recovery is made by the time 30% of the coefficients are used in reconstruction, after which the curve saturates. Secondly, more than 95% of the information is retrieved on average by the time 20% of the coefficients are used in reconstruction, and around 85% of the information is retrieved by the time 10% of the coefficients are used for image reconstruction. Thirdly, for some images (images not shown here) with fewer edge details, 99% of the information is recovered with only 10% of the coefficients of filtering used for reconstruction.

The standard deviation of the spread about the mean is shown using error bars in figure 7.3(b). The total information recovered with 25% or more of the coefficients is nearly the same for all the pictures shown by the very small standard deviation of the data in that area. The average information recovery plot of the retinal model prior to applying FoCal is shown in the same figure for comparison. It is observed that there is an increase of 20% on average in the total amount of information recovered using FoCRen. Further, the rate of information recovery is much faster, with a 10% difference in information recovery by the time the first 1% coefficients are used for image reconstruction. By the time 5% of the coefficients are used for image reconstruction, there is a difference of 25% in the quantity of information recovered. This difference starts diminishing after the first 20% coefficients have been used for reconstruction and is observed to settle at around 20% for the first 30-35% of the coefficients, when both plots saturate.

The above results indicate the improvement in the quantity of information



(a)



(b)

Figure 7.3: (a) The green lines show the progressive recovery of the perceptually-important information in the reconstruction of each of the sixty-five images in our data set which were rank-order encoded using FoCRen. The red line is the mean of the individual information recovery plots. (b) The standard deviation of the individual information recovery plots about the mean in (a) is shown using errorbars. The average information recovered prior to applying FoCal in the retinal model is shown here for comparison.

retrieved by the time 20% of the cells have fired their first spikes over that using the pseudo-inverse method of decoding. Further, this study is conducted on a desktop and with images of size 128 × 128. Thus, although the image reconstruction using coefficients of filtering is not perfect, yet, we have optimised the information recovery by improving both the quantity and time of information recovery.

For subjective evaluation of the information recovered, the reconstruction of two rank-order encoded images using FoCRen for the first 1%, 5%, and 10% of the coefficients are shown in figure 7.4. A comparison is made with the reconstruction of the same images rank-order encoded using VanRullen and Thorpe's method of encoding. Significant improvement in information recovery using FoCRen is apparent from visual inspection of the reconstructed images.

7.2.2 LUT for FoCRen

After applying FoCal to the coefficients of filtering of all the sixty-five images in our data-set with VanRullen and Thorpe's retinal model, we prepare a Look-up-table for FoCRen, which is shown in figure 7.5(a). A comparison with the LUT of the retinal model prior to applying FoCal is also shown in the same figure. The plots are normalised so that the maximum value of the LUT is 100. The LUT for FoCRen plotted on a log-log scale is shown in figure 7.5(b), fitted with a plot that follows the power law, as in figures 4.7 and 6.2(b). Thus the LUT for FoCRen also follows the power law until the first 10% of the cells have fired their first spikes.

7.2.3 Reconstruction using LUT

Having obtained the LUT, we now progressively reconstruct each of the sixty-five images in our data-set, rank-order encoded using FoCRen, with the LUT in figure 7.5. The perceptual information recovered for each of the individual images is shown in the figure 7.6(a) as a spread about the average information recovered. The standard deviation of this data-spread is shown as error-bars in the plot shown in fig 7.6(b), along with a similar plot VanRullen and Thorpe's method of encoding for comparison. It is observed that, using FoCRen, there is an overall increase of 15% in the information retrieved until the time when 20% of the cells have fired their first spikes, compared to that using VanRullen

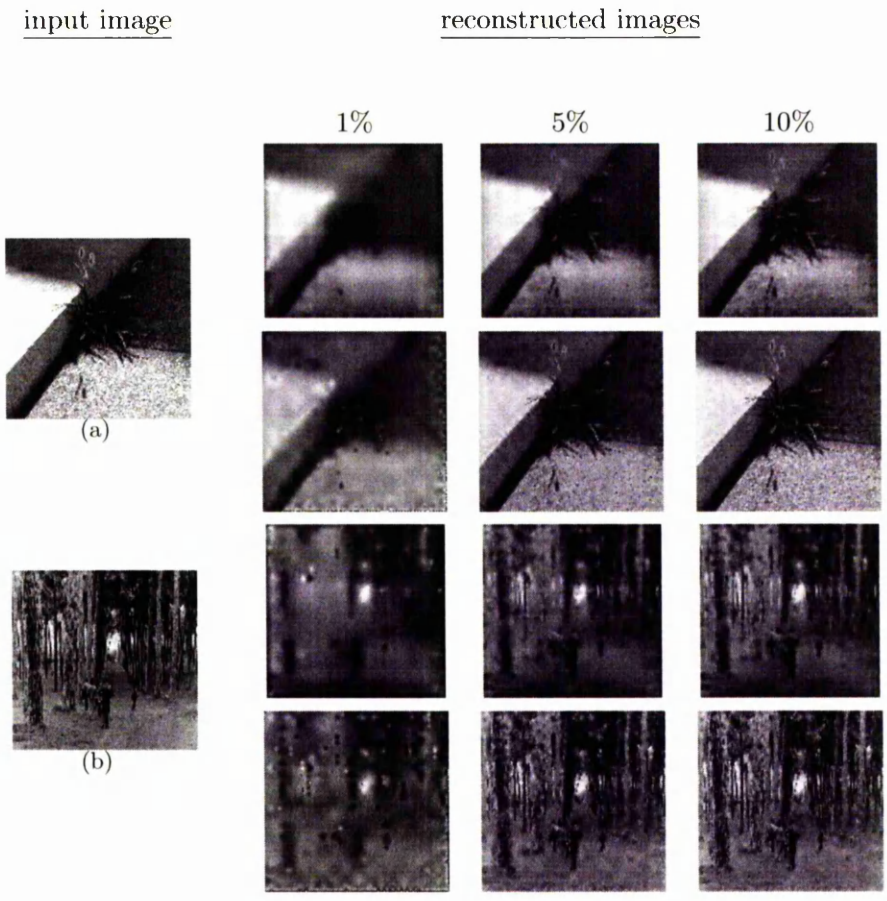
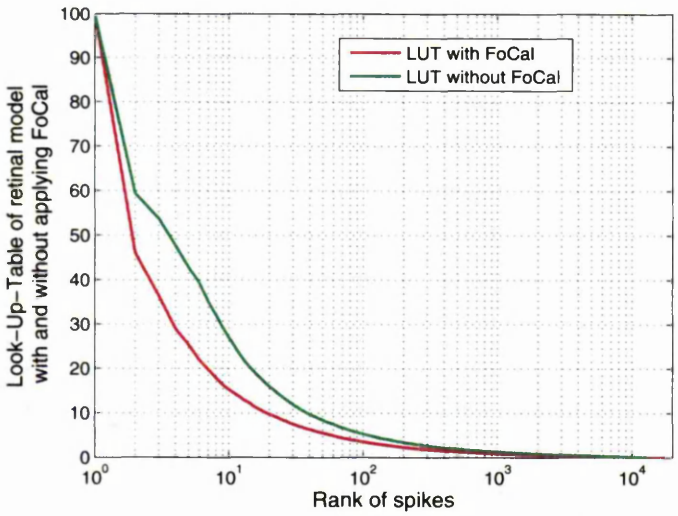
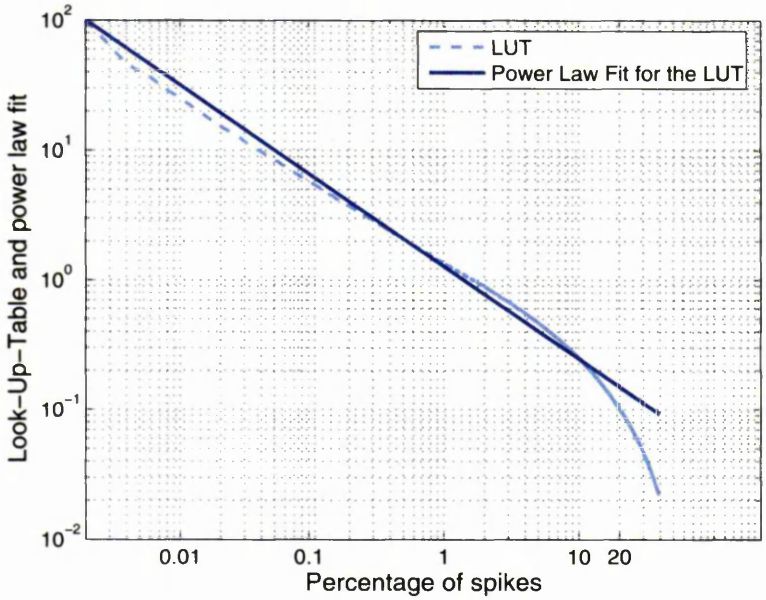


Figure 7.4: Reconstruction of two input images in (a) and (b) with 1%, 5% and 10% of the total number of coefficients, and rank-order encoded using (top) VanRullen and Thorpe's method of encoding and (bottom) FoCRen.

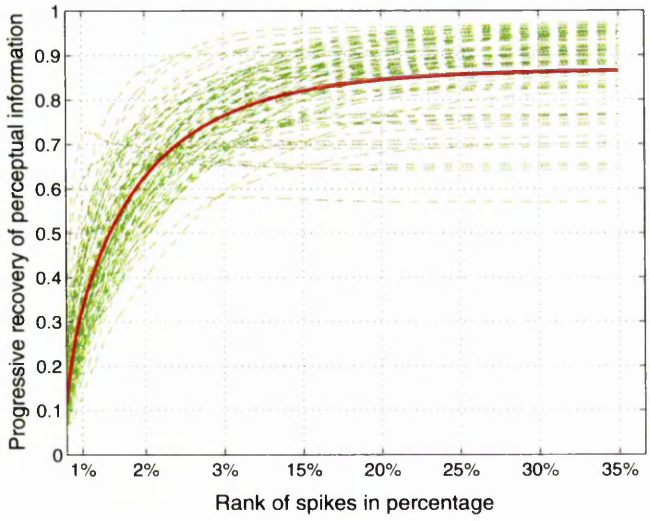


(a)

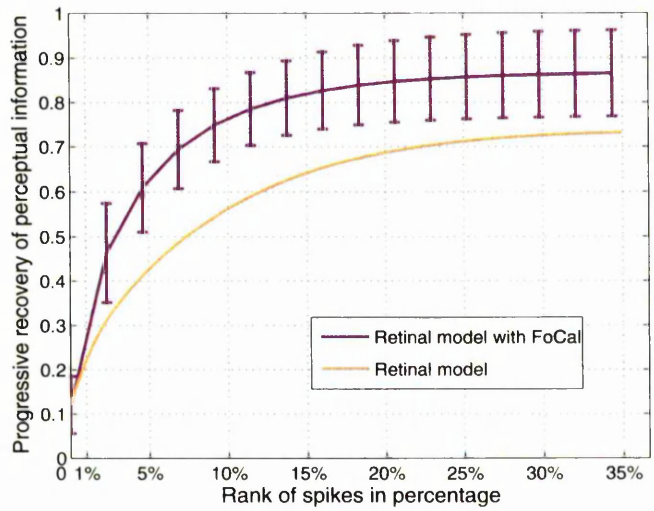


(b)

Figure 7.5: (a) Comparison of LUT for retinal model (blue line) and LUT for FoCRen (pink line). (b) LUT for FoCRen (dashed line) fitted with a curve that follows the power law (blue line) until the first 10% of spikes.



(a)



(b)

Figure 7.6: (a) The information recovery plot for all the sixty-five images shown as a spread about the mean information recovery plot for progressive reconstruction of images using FoCRen, and the LUT for FoCRen in figure 7.5. (b) Standard deviation of the individual plots in (a) about the mean plot. The mean information recovery plot using VanRullen and Thorpe's method of rank-order encoding is also shown for comparison.

and Thorpe's method of encoding. The rate of information recovery from rank-order encoded images using FoCRen is also observed to be consistently better than that using VanRullen and Thorpe's method of encoding. A comparison of images reconstructed after encoding using the two methods of encoding and their respective LUTs is shown in figure 7.7. The improvement in image quality for images reconstructed using FoCRen for rank-order encoding images is apparent on visual inspection, and agrees with the information recovery plots for the same.

In the next section, we discuss the Matching Pursuit algorithm, which ismathematically the same as FoCal, although the method of application to VanRullen and Thorpe's retinal model is different.

7.3 Matching Pursuit algorithm

Let f be a signal that has to be decomposed into a linear expansion of waveforms whose properties are adapted to the local structure of the signal f. Let $\mathcal{D}$ be a redundant dictionary of functions, some of which match the local structure of f. The Matching Pursuit (MP) algorithm decomposes any signal f into waveforms which are selected from the redundant dictionary $\mathcal{D}$ to "best-match" the signal structures. Introduced by Mallat and Zhang, MP is a greedy algorithm that chooses the most suitable waveform from $\mathcal{D}$ to approximate part of a signal f in each iteration [21]. Thus the algorithm isolates the coherent structures of a signal f with respect to the dictionary $\mathcal{D}$. Bergeaud and Mallat applied this algorithm in image analysis to obtain an efficient representation of low level structures in an image [9]. With the image as a signal, and a redundant dictionary of Gabor functions, they could represent the signal by selectively picking the Gabor functions that best approximate a local feature of the image at each iteration.

Let g_γ be a set of basis vectors. Let f be the input signal that has to be decomposed using MP. Often, f is represented as

$$f = \sum_{i=0}^{n-1} \lambda_i g_{\gamma_i},$$

where $\lambda_i = \langle f, g_{\gamma_i} \rangle$. This is only an approximation if the vectors in g_γ are not

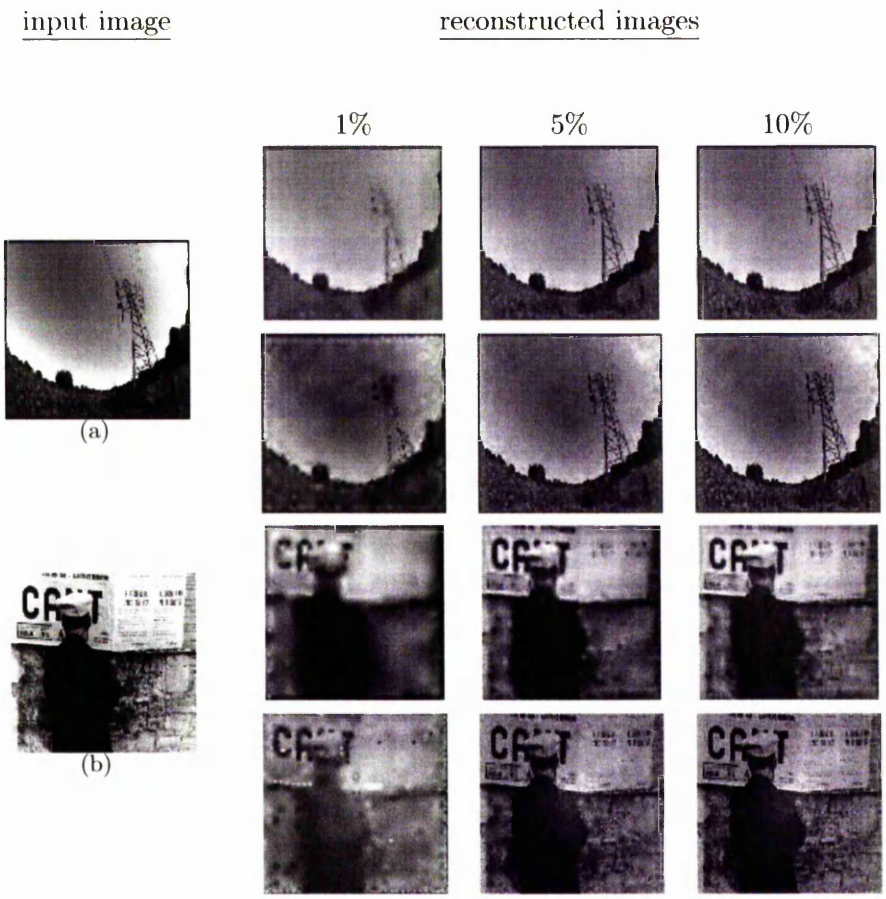


Figure 7.7: Reconstruction of two input images in (a) and (b), rank-order encoded using (top) VanRullen and Thorpe's original method and (bottom) FoCRen by the time 1%, 5% and 10% of the total number of ganglion cells have fired their first spikes. Both reconstructions are done using the respective LUTs.

orthogonal. The exact representation of f would be written as

$$f = \sum_{i=0}^{n-1} \lambda_i g_{\gamma_i} + \mathcal{R}^n f, \tag{7.3}$$

where $\mathcal{R}^n f$ is the residual in a sum of n terms corresponding to the decomposition of f into n components. Equation 7.3 can be written as

$$f = \sum_{i=0}^{n-1} \langle f, g_{\gamma_i} \rangle g_{\gamma_i} + \mathcal{R}^n f, \tag{7.4}$$

For the first iteration, $i = 0$ in equation 7.4,

$$f = \langle f, g_{\gamma_0} \rangle g_{\gamma_0} + \mathcal{R}^1 f$$

$$\Rightarrow \mathcal{R}^1 f = f - \langle f, g_{\gamma_0} \rangle g_{\gamma_0} \tag{7.5}$$

For a perfect representation of f, the aim will be to minimize the residual term $\mathcal{R}^1 f$ in the l.h.s. of equation 7.5. This can be done by minimizing the difference in the r.h.s. of equation 7.5, i.e. the term $\langle f, g_{\gamma_0} \rangle$ needs to be maximised. This in turn can be done by choosing g_{γ_0} from the whole set g_γ such that

$$|\langle f, g_{\gamma_0} \rangle| \geq |\langle f, g_\gamma \rangle| \tag{7.6}$$

Now, $\langle f, g_{\gamma_0} \rangle$ is the similarity measure between the function f and the basis vector g_{γ_0}. So, by choosing the g_γ for which this similarity is maximum, the dictionary element which is most similar to the signal structure at this stage is chosen. This is done iteratively for the components along the other vectors in the dictionary $\mathcal{D}$ until the residual at the $(n+1)^{th}$ iteration drops below an arbitrarily chosen threshold ϵ.

This strategy was later used by Perrinet et al in trying to improve on the tight constraint of orthogonality of the filters in the reconstruction method of the retinal model [54]. They show that rank-order encoding with MP in the retinal model gives a better performance in retrieving the information about the input image than that of the original model. The MP algorithm as applied to rank-order encoding in the retinal model is given below:

Let I^0 be an image and $\Phi_{\sigma\lambda}$ be a DoG in a redundant dictionary $\mathcal{D}$ of DoG functions at various scales σ and spatial locations $\lambda = (u, v)$ in $2D$ space. Let

$C_{\sigma\lambda} = \langle I, \Phi_{\sigma\lambda} \rangle$ be the set of coefficients obtained by filtering I with a set of filters $\Phi_{\sigma\lambda} \in \mathcal{D}$.

Initialize

Let

$$I^0 \leftarrow I.$$

$$C^0_{\sigma\lambda} = \langle I^0, \Phi_{\sigma\lambda} \rangle$$

Match

Find $\sigma^0\lambda^0$ that corresponds to the largest coefficient $C_{\sigma^0\lambda^0} \in C^0_{\sigma\lambda}$.

$$(\sigma^0\lambda^0) = Argmax_{\sigma\lambda} \left(|C^0_{\sigma\lambda}| \right),$$

$$C_{\sigma^0\lambda^0} = \langle I, \Phi_{\sigma^0\lambda^0} \rangle$$

Update

Therefore, as in equation 7.3, and for $i = 0$,

$$I^0 = C_{\sigma^0\lambda^0}.\Phi_{\sigma^0\lambda^0} + I^1, \tag{7.7}$$

where I^1 is the residue of the first iteration.

$$\Rightarrow I^1 = I^0 - C_{\sigma^0\lambda^0}.\Phi_{\sigma^0\lambda^0} \tag{7.8}$$

Let

$$C^1_{\sigma\lambda} = \langle I^1, \Phi_{\sigma\lambda} \rangle \tag{7.9}$$

Putting equation 7.8 in equation 7.9,

$$C^1_{\sigma\lambda} = \langle (I^0 - C_{\sigma^0\lambda^0}.\Phi_{\sigma^0\lambda^0}), \Phi_{\sigma\lambda} \rangle$$

$$= \langle I^0, \Phi_{\sigma\lambda} \rangle - C_{\sigma^0\lambda^0}.\langle \Phi_{\sigma^0\lambda^0}, \Phi_{\sigma\lambda} \rangle$$

$$\Rightarrow C^1_{\sigma\lambda} = C^0_{\sigma\lambda} - C_{\sigma^0\lambda^0}.\langle \Phi_{\sigma^0\lambda^0}, \Phi_{\sigma\lambda} \rangle \tag{7.10}$$

Since

$$C_{\sigma^0\lambda^0} - C_{\sigma^0\lambda^0}.\langle \Phi_{\sigma^0\lambda^0}, \Phi_{\sigma^0\lambda^0} \rangle = 0,$$

$C^1_{\sigma\lambda}$ is calculated for all the elements in $(C^0_{\sigma\lambda} - C_{\sigma^0\lambda^0})$.

Repeat Match and Update for the Second Iteration
Find $\sigma^1\lambda^1 = Argmax_{\sigma\lambda}(|C^1_{\sigma\lambda}|)$ such that

$$C_{\sigma^1\lambda^1} = \langle I^1, \Phi_{\sigma^1\lambda^1} \rangle$$

is the largest coefficient in $C^1_{\sigma\lambda}$.

$$I^2 \leftarrow I^1 - C_{\sigma^1\lambda^1}.\Phi_{\sigma^1\lambda^1},$$

where I^2 is the residue of the second iteration. Let

$$C^2_{\sigma\lambda} = \langle I^2, \Phi_{\sigma\lambda} \rangle$$

$$\Rightarrow C^2_{\sigma\lambda} = C^1_{\sigma\lambda} - C_{\sigma^1\lambda^1}.\langle \Phi_{\sigma^1\lambda^1}, \Phi_{\sigma\lambda} \rangle \quad \text{(from equation 7.10)}$$

Stop
The algorithm is stopped at the $(l+1)^{th}$ iteration, when
the largest coefficient $C_{\sigma^l\lambda^l}$ is less than an arbitrarily chosen threshold ϵ.

$$C_{\sigma^l\lambda^l} < \epsilon, \text{where} C_{\sigma^l\lambda^l} \in C^l_{\sigma\lambda}$$

7.3.1 FoCal vs. MP

The mathematical analysis of the two algorithms presented in sections 7.3 and 7.1.1 show them to be similar. However, the fundamental difference with respect to our work is in the application of the algorithms to retrieve information from rank-order encoded images. We have developed FoCal with the aim of reducing redundancy in the data introduced by filter-overlap during rank-order encoding. We use the principle of lateral inhibition, a mechanism used by sensory neurons to reduce redundancy in sensory input data. The similarity of application of FoCal with the application of lateral inhibition by sensory neurons is drawn in

section 7.1.

On the other hand, Perrinet et al applied MP iteratively on the whole image whereby at every step, all the coefficients of filtering are calculated and the largest coefficient is matched with its corresponding filter and spatial location. This filter is scaled up with the largest coefficient and subtracted from the image, thereby obtaining a residue image, which is the input image for the next iteration. Thus, at every iteration, there is a *pursuit* to find the *matching* filter for the largest coefficient, and hence the nomenclature. Evidently, the application is from a very different perspective than FoCal, although the rank-order encoded data obtained using the two methods are the same.

7.4 Chapter summary

In this chapter, we introduce a novel encoding method, viz. the Filter-Overlap Correction algorithm (FoCal) for rank-order encoding images using VanRullen and Thorpe's retinal model. We name this improved rank-order encoder FoCRen. We show the similarity between FoCal and the redundancy reducing technique of lateral inhibition used in our biological sensory neurons. The results of applying FoCRen to rank-order encoded images are presented along with a comparison with the results using VanRullen and Thorpe's method of encoding. Finally, we present the Matching Pursuit algorithm which is mathematically similar to FoCal and was applied by Perrinet et al to improve rank-order encoding using VanRullen and Thorpe's retinal model. However the basis of the two algorithms are very different; while FoCal is based on the principle of lateral inhibition, MP is based on a 'match and update' mechanism.

In the next chapter, we propose a biologically realistic model of the retina for rank-order encoding images and discuss and compare results with rank-order encoding using VanRullen and Thorpe's retinal model and that using FoCRen.

Chapter 8

Towards Biological Realism

In chapters 6 and 7, we proposed ways to minimise information loss during decoding and encoding of rank-order codes and thus maximise information recovery from rank-order encoded images. So far in our work, we have used VanRullen and Thorpe's retinal model as the basic rank-order encoder, which uses sixteen layers of ganglion cells tiling the retina independently as was discussed in chapter 4. This is a simplification of the complexities in the retina, and the model has been used to simulate rank-order encoding 'successfully' in terms of information recovery from rank-order encoded images. However, using sixteen layers of ganglion cells is biologically unrealistic. This is because firstly, the retina is an inside-out coating on the eyeball with the ganglion cells forming the frontmost layer of the retina in the direction of incident light. As a result, the incident light has to travel through the whole thickness of the retina to reach the photo-receptors cell layer, the photons getting scattered at each layer. Thus sixteen layers of ganglion cells would be very inefficient in terms of the amount of photons that actually reaches the photoreceptors, which would in turn reduce the photo-sensitivity of the photo-receptors, and thus affect the efficiency and accuracy of vision. In our review of the ganglion cells in chapter 3, we observed that there are mainly four types of ganglion cells in the primate retina. Each of these arborize in different layers. Thus each layer has a particular cell size. Moreover, ON- and OFF-centre cells of each type are not complementary, contrary to those used in VanRullen and Thorpe's retinal model.

Based on such factual data, we propose to use biologically realistic parameters for simulating the cells of the retina. In chapter 3, we saw that the depth of the retina is minimum at the foveal-pit, which is devoid of any other cell types

except the photoreceptors. This makes this region most accessible to incoming light. Thus, this is the area of the retina which is most sensitive to light, and is responsible for the ability of the primate eye to distinguish very fine details in a scene [18]. We propose to simulate the cells of the foveal-pit, thereby deviating from VanRullen and Thorpe's retinal model, towards a *biologically realistic retinal model*, based on the physiology of ganglion cells found in the primate retina. We call our model — the FOVEAL-PIT MODEL.

In section 8.1, we discuss the design of the FOVEAL-PIT MODEL and briefly reminisce the salient points of the physiology of ganglion cells presented in section 3.2 in the context of our model design. In section 8.2, we do an empirical analysis of the performance of the model as a rank-order encoder in terms of information recovery from rank-order encoded images.

8.1 Simulating foveal-pit model

The limited width of the foveal pit demands provisions for motion in our visual system [23]. To see an object clearly, its retinal image should fall on the foveal pit, and should be kept there long enough for its finest details to be discriminated [45, 44]. To accomplish this task, the primate eye makes fast, voluntary movements. **Saccades** are periods of time when the eyes are rotating to shift the centre of vision, i.e. the foveal pit, from one spatial position on the object to another. **Fixations** are periods when the gaze is held at a point of interest on the object to expose it to the fovea long enough for the details to be picked up. Information processed by the cones of the foveal pit during a certain fixation is passed on to the ganglion cells that are connected to the cones in the area.

We simulate the receptive fields of the ganglion cells corresponding to the foveal-pit with the DoG functions as used in VanRullen and Thorpe's retinal model. However, the size of the centre and surround of the DoG functions, as well as their physical layout is in keeping with the actual physiology of those found in the primate retina, and which we will talk about in the following sections. As was done in VanRullen and Thorpe's retinal model, we tile an input image homogeneously with these DoG functions, which is a simplified simulation of the various fixations of the eye as it makes several saccades over the image. Since time is not being considered and all the points are assumed to be processed simultaneously, this would mean that the most salient information gathered out

of all fixations will be the first to cause a ganglion cell to fire a spike. In this work, our motive is to study the efficiency of rank-order codes in transmitting perceptually-important details in a scene. This we do by decoding the codes, and evaluating the information retrieved both in terms of the quantity of perceptually-important information retrieved and the time taken for this retrieval.

8.1.1 Eccentricity of the cells

As one moves concentrically outwards from the centre of the fovea, a consistent change in the size and density of the ganglion cells is observed. The first requirement in designing the FOVEAL-PIT MODEL will be to decide on a certain eccentricity in the fovea. We have already reviewed the ganglion cell to cone ratio in section 3.2.3. Based on data presented in the review, let us take a window of $25\mu m$ around the centre of the fovea, which is a circular region of radius $12.5\mu m$ and area $491\mu m^2$. Since the density of cones in the region is $250,000/mm^2$ (discussed in section 3.2.3), this area of $491\mu m^2$ will have approximately 123 cones. Keeping in view the cone to ganglion cell ratio of 4 : 1, we can assume that these 123 cones are connected to a total of around 492 ganglion cells.

A simulation of this region consisting of 123 cones could be made with a region of $11 \times 11 = 121 (\simeq 123)$ pixels in the image, where each pixel represents a cone. This region could be thought of as a point of fixation where the fovea stops between two saccades to gather information. Thus, we could divide a 128×128 image homogeneously along the rows and columns into 11×11 matrices, resembling the foveal excursions made during the saccades over the whole image.

Having decided upon the eccentricity, we now review the different types of ganglion cells present in the fovea at that eccentricity. The dendritic tree sizes of these cells and the overlap of their dendritic trees with those of the neighbouring cells are also discussed.

8.1.2 A review of foveal ganglion cells

In this section, we review some of the points discussed in section 3.2 on which we base our simulation of the ganglion cell receptive fields.

The foveal ganglion cells are primarily of midget and parasol varieties. The midget ganglion cells constitute around 95% of the ganglion cell population in the region. The mosaic of the ON- and OFF-centre midget cells are independent

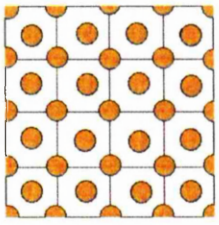


Figure 8.1: An example of the layout of the midget ON-centre and OFF-centre ganglion cells for sampling a 4×4 image raster.

of each other with the neighbouring dendritic trees in each mosaic showing very little overlap. On average, the dendritic field diameter of an ON-centre midget cell is 30% greater than that of the OFF-centre cell across all eccentricities in the retina.

Similar to the midget cells, the ON- and OFF-centre parasol cell mosaics are independent, with a neighbouring overlap of dendritic trees (defined in section 3.2.1) of neighbouring cells recorded at 3.4. The ratio of the dendritic tree diameter of the parasol cell to that of the midget cell of each type is found to be 10 : 1 in the foveal region. Since each type of ganglion cell homogeneously tiles the retina, a point of space is sampled by several cells at the same time.

The diameter of the receptive field centre of a ganglion cell is on average 1.5 times larger than that of its dendritic tree. The surround of a midget cell receptive field is on average 6.7 times wider than its centre, while that of a parasol cell is 4.8 times wider than its centre.

8.1.3 Simulating foveal ganglion cells

The array of cones in the fovea is represented by a raster where each pixel corresponds to a cone, as shown in figure 8.1. An image is sampled by each of the midget ON- and OFF-centre cells (shown as orange balls in figure 8.1) centred at every pixel along the column and every half pixel along the rows. The total number of cells of each type will, therefore, be double the number of pixels. For example, in the 4×4 raster of figure 8.1, there are a total of 16 circles placed at the centre of each pixel. The number of circles centered at the junction, considering the half circles at the edges and $1/4^{th}$ circles at the corners is also 16. Each pixel represents a cone and each circle is an OFF-centre midget cell. Therefore for 16 pixels, there will be 32 midget OFF-centre cells. The midget ON-centre cells

Ganglion Cell Types		Receptive Field Simulation Parameters					
		matrix size (n)	std. dev. centre (σ_c)	centre width in pixels (w_c)	std. dev. surround (σ_s)	sampling resolution	total cells for 121 pixels
midget	OFF-centre	3×3	0.8	3	$6.7 \times \sigma_c$	col: 1 row: $\frac{1}{2}$	484
	ON-centre	11×11	1.04 $\simeq$ (0.8 + 30% of 0.8)	5 $\simeq$ (3 + 30% of 3)			
parasol	OFF-centre	61×61	8 $\simeq (10 \times 0.8)$	33 $\simeq (10 \times 3)$	$4.8 \times \sigma_c$	col: 5 row: $2\frac{1}{2}$	18
	ON-centre	243×243	10.4 $\simeq (10 \times 1.04)$	53 $\simeq (5 \times 10)$			

Table 8.1: Table showing the various parameters of the simulation of the ganglion cells of the foveal pit at an eccentricity of $12.5\mu m$.

are placed at a similar resolution, so there will be a total of 64 midget cells which is four times the number of cones. Thus, in a 11×11 raster with 121 pixels, there will be a total of 484 midget cells.

Table 8.1 shows the various specifications that we use in simulating the foveal ganglion cells. As shown in the table, the OFF-centre midget cell is simulated with a DoG of matrix size 3×3. The standard deviation of the centre Gaussian of the DoG is set at 0.8, so that the centre of the DoG will span 3 pixels. This causes it to overlap with the centers of 8 neighbouring cells of the same variety. Although the midget cells in the primate fovea do not show such high overlap, a DoG with its centre-surround structure cannot be represented with a centre-width smaller than 3 pixels, as a limitation of representing a function in a digital raster. Although the overlap could be avoided by reducing the sampling resolution, we choose to maintain the high sampling resolution in our simulation as a trade-off for maintaining ganglion cell to cone ratio of $\simeq 4 : 1$ as found in regions close to the foveal centre. The dendritic tree diameter of an ON-centre midget cell is considered to be 30% larger than that of an OFF-centre cell. The DoG corresponding to a midget ON-centre cell in our simulation would therefore span 5 pixels, (to maintain odd-numbered width for symmetry) obtained by setting the standard deviation of the centre Gaussian of a DoG matrix of size 11×11 to 1.04, which in turn is 30% more than 0.8, the standard deviation of the midget OFF-centre cells.

Coming to the parasol cells, the ratio of the parasol centre diameter to that of the midget centre diameter of both the ON- and OFF-centre cells in our simulation is maintained at 10:1. The OFF-centre parasol is simulated with a DoG matrix of

size 61×61, with the standard deviation of the centre Gaussian as $0.8 \times 10 = 8$. This gives the centre-width of the DoG as 33 pixels which is approximately 10 times the centre-width of the midget OFF-centre DoG. Similarly, the ON-centre parasol cell is simulated with a 243×243 matrix, with the standard deviation of the centre DoG as 10.4, which is 10 times that of the midget ON- cells, giving a centre-width of 53 pixels.

The standard deviation of the surround is maintained at 6.8 times and 4.7 times those that of the centre for the midget and parasol cells respectively. The sampling resolution is $2\frac{1}{2}$ along the rows and 5 along the columns. In an 11×11 raster, the total number of parasol cells will be 18, the number of cells of each type being 9. Therefore the midget cells will constitute 96% of the total ganglion cell population in our simulation, which is fairly close to the biological estimate of 95% in the foveal region. The ganglion cell to cone ratio for our 11×11 raster is thus $502{:}123 = 4.08{:}1 \simeq 4{:}1$.

In the next section, we do an empirical analysis of the performance of this model in rank-order encoding an input image.

8.2 Empirical analysis of foveal-pit model

We filter the 128×128 image shown in figure 8.2(a) using the FOVEAL-PIT MODEL in the same manner as in equation 4.4, to obtain a total of 73728 coefficients of filtering. Since, we are imitating biology, intuitively, we must use the biological principle of lateral inhibition to do away with the redundancy in the data before rank-order encoding. We have already done this (albeit approximately, as explained in sections 7.1) by applying FoCal (introduced in section 7.1.1) to the retinal model. Conversely, we encode the image without removing redundancy, using VanRullen and Thorpe's method of encoding, and observe the results before using FoCal for rank-order encoding.

The encoded image is reconstructed using VanRullen and Thorpe's method of decoding. The plot for the progressive recovery of perceptually-important information is shown in figure 8.2(d), and the corresponding reconstructed picture is shown in figure 8.2(b). The information recovery plot for the same image encoded using VanRullen and Thorpe's retinal model is shown the same figure, while the reconstructed image is shown in figure 8.2(c). The quantitative recovery agrees with the qualitative evaluation of the two reconstructed images.

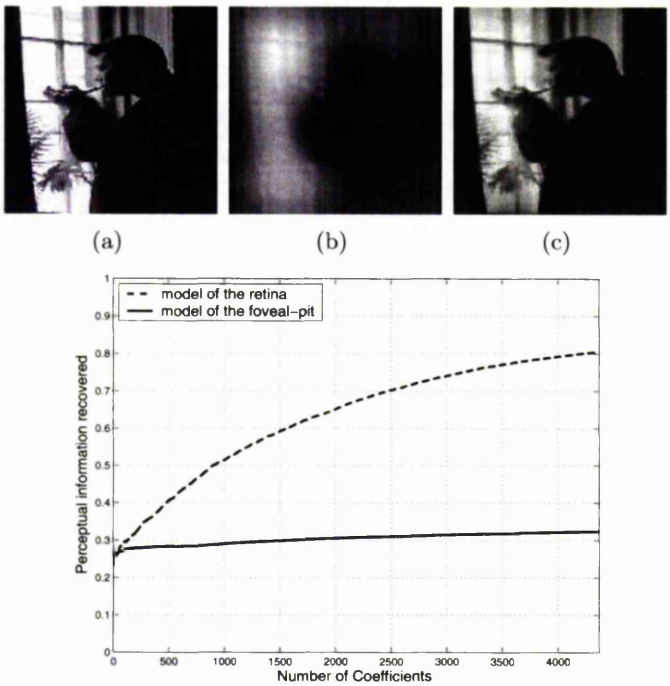


Figure 8.2: (a) Input image. (b) Reconstructed image using the first 6% ($\simeq$ 4400) coefficients of filtering of the FOVEAL-PIT MODEL. (c) Reconstructed image using the first 10%($\simeq$ 4400) coefficients of filtering of VanRullen and Thorpe's retinal model. (d) Information recovery plot for the progressive reconstruction of images shown in (b) and (c).

The failure of the method in reconstruction may be ascribed to the assumption of orthogonality for the filters. The centre widths and the sampling resolution of the image, as shown in table 8.1, suggest that there is a very high overlap among the neighbouring filters, making them unsuitable to be approximated as orthogonal functions. Speaking biologically, this high overlap of the filters suggests significant redundancy in the neural encoding as a point in space is being sampled by at least five filters at a time. While redundancy takes care of the noise in the input data and is useful from such a perspective, yet, while rank-order encoding images, it gives rise to an over-representation of the information contained in the input stimulus. Clearly, this form of redundancy is hindering the performance of our model, as has already been anticipated by us in the above paragraph, and should be removed. Towards that goal, we now incorporate FoCal to the FOVEAL-PIT MODEL for encoding images in rank-order and discuss the results.

8.2.1 Reconstruction using coefficients of foveal-pit model

From the results presented in figure 8.2(d), it is evident that the coefficients of filtering of the images in the FOVEAL-PIT MODEL have to be corrected for filter-overlap by applying FoCal prior to image reconstruction using VanRullen and Thorpe's method of decoding. Thus FoCal will be an inherent part of rank-order encoding using the FOVEAL-PIT MODEL, rather than an enhancement, as was for VanRullen and Thorpe's retinal model, taking our encoder closer to biology. The information recovery from images rank-order coded using such a biologically realistic encoder is discussed in the sections below.

Reconstruction Using Coefficients of Filtering

The plot for the average information recovery for all the sixty-five images using the FOVEAL-PIT MODEL is shown in figure 8.3(a). The errorbars indicate the standard deviation of the information recovered from the individual images about the mean information recovered.

Firstly, we compare the plot in figure 8.3(a) with that in figure 8.2(d) and observe that when FoCal is applied to the coefficients of filtering with the FOVEAL-PIT MODEL, the reconstruction works. This indicates that, by incorporating the

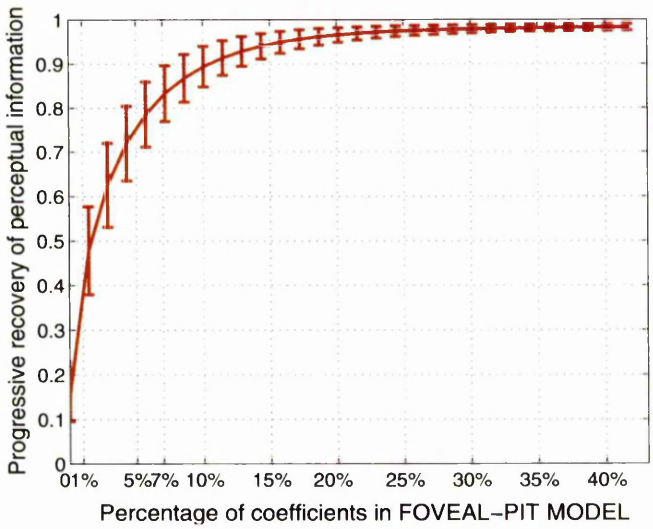
biological mechanism of lateral inhibition in the simulated cells of the FOVEAL-PIT MODEL, we decorrelate each cell from its spatial neighbourhood. In other words, making the model more biologically accurate enhances its performance.

On taking a closer look at the plot in figure 8.3(a), we observe that up to 90% of the perceptual information can be recovered by the time only 10% of coefficients of filtering are used for reconstruction. On average, the initial information recovery is very fast, and thus conforms to the requirements of rank-order method of neural encoding. A total information recovery of 98 - 99% is observed, which is achieved by the time 20% of the coefficients of filtering are used for reconstruction. After this point, the information recovery plot reaches saturation, and very little additional information is recovered. In figure 8.3(b), we compare the average perceptual-information recovery plots from images rank-order encoded using FOVEAL-PIT MODEL, VanRullen and Thorpe's method applying FoCal (FoCRen) and without applying FoCal. On average, rate of information recovery is slightly higer for FoCRen compared to the FOVEAL-PIT MODEL. However both these methods are much better in encoding information than VanRullen and Thorpe's original encoding method. The progressive reconstruction of an image rank-order encoded using the three methods as mentioned above is shown in figure 8.4 while the progressive information recovery plots for the same are shown in figure 8.5. The rapid information recovery using FOVEAL-PIT MODEL, with only four layers of filters compared to sixteen layers in FoCRen, confirms the fact that rank-order hypothesis indeed works well for models that are designed in accordance with those of evolution.

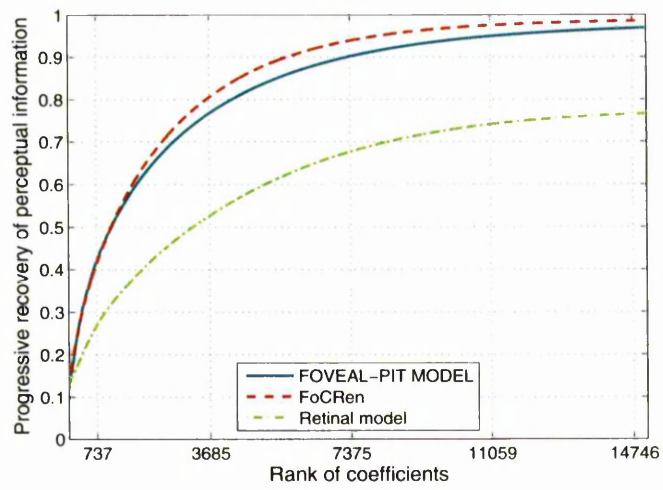
In the next section, we construct the Look-Up-Table for the rank-ordered weights of the FOVEAL-PIT MODEL and then reconstruct an input image using this LUT.

LUT of the foveal-pit model

We generate the LUT of the FOVEAL-PIT MODEL in the same manner as we have been doing for the retinal model. In figure 8.6(a), the LUT is shown as a mean of the rank-ordered coefficients of filtering of each of the sixty-five images in our data-set with the FOVEAL-PIT MODEL. In figure 8.6(b), we show a comparison of the LUT of the FOVEAL-PIT MODEL with those generated using FoCRen and VanRullen and Thorpe's method of rank-order encoding. The log-log plot shown in figure 8.7 indicates that the LUT of the FOVEAL-PIT MODEL does not follow



(a)



(b)

Figure 8.3: (a)Mean information recovery plot for progressive reconstruction using the coefficients of filtering of sixty-five images rank-order encoded using the FOVEAL-PIT MODEL. The error bars show the standard deviation of the information retrieved for each individual image about the mean information recovery plot. (b) A comparison of the mean information recovery plots for images rank-order encoded using FOVEAL-PIT MODEL, FoCRen and VanRullen and Thorpe's retinal model.

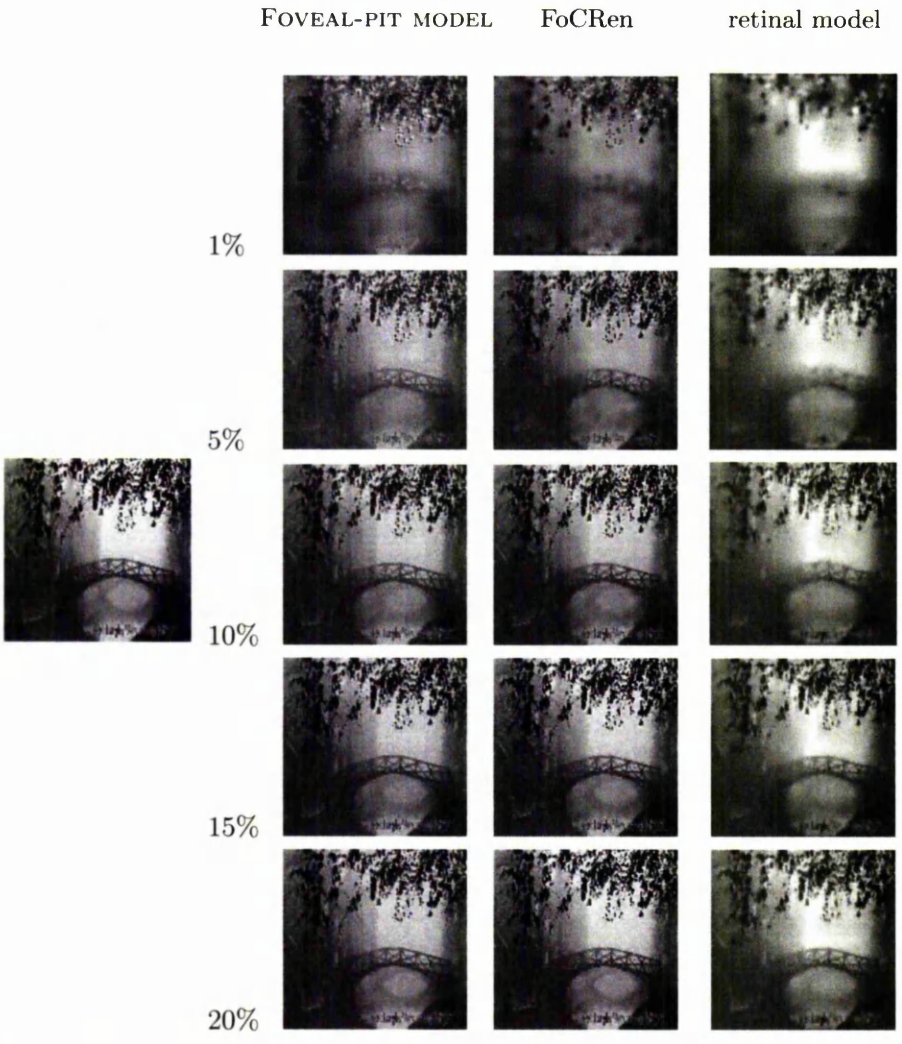


Figure 8.4: An input input image and its reconstruction using 1%, 5%, 10%, 15% and 20% of the coefficients of filtering of (left) FOVEAL-PIT MODEL, (middle) FoCRen and (right) VanRullen and Thorpe's retinal model. The progressive reconstruction of the image using each of these three methods of rank-order encoding is shown in figure 8.5.

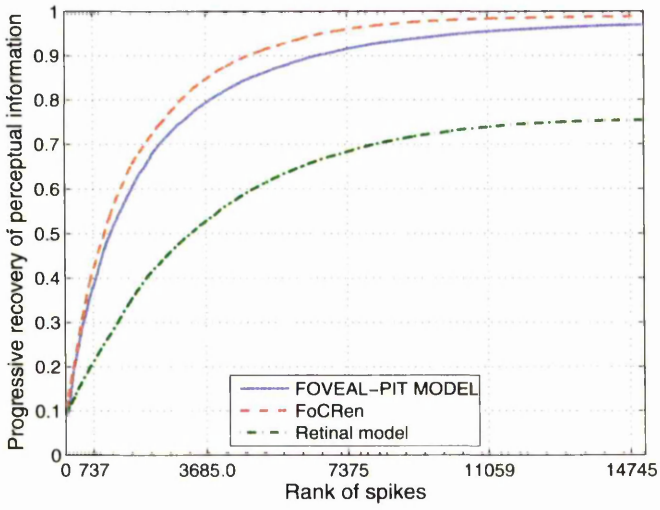


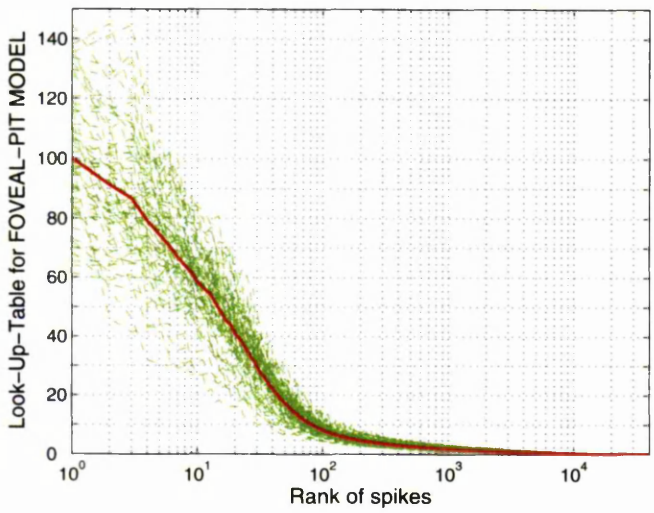
Figure 8.5: Perceptual information recovery plot for progressive reconstruction of an input image shown in figure 8.4 using the rank-ordered coefficients of filtering of the FOVEAL-PIT MODEL, FoCRen, and VanRullen and Thorpe's retinal model.

the power law.

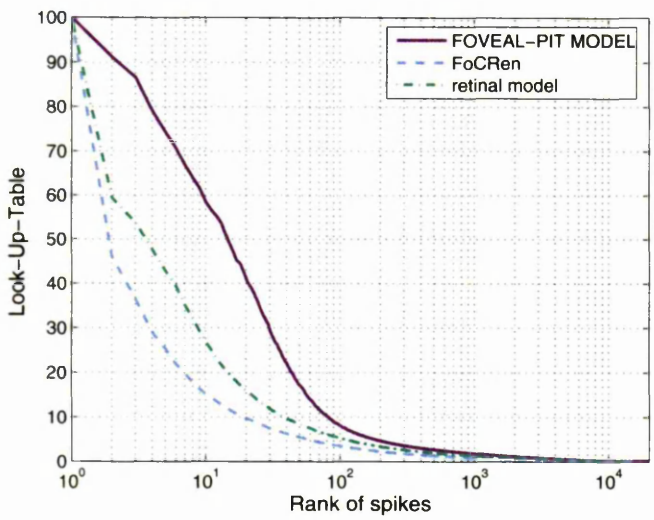
Reconstruction with LUT of foveal-pit model

The results of information recovery using the LUT of the FOVEAL-PIT MODEL are shown in figure 8.8(a). We observe that 75 – 80% of the perceptually-important information about an input image is recovered by the time 7 – 10% of the neurons of the FOVEAL-PIT MODEL have fired their first spike. Further, a total information recovery of more than 85% on average is observed, which is obtained by the time 15% of the ganglion cells of the FOVEAL-PIT MODEL have fired their first spikes, after which the curve goes into saturation and there is no further information recovery. This rapid build-up of information is very much in agreement with the basic concept of the rank-order code hypothesis in vision.

The information recovery plot of the FOVEAL-PIT MODEL is compared with similar plots of FoCRen and VanRullen and Thorpe's retinal model in figure 8.8(b). Since the number of coefficients in the FOVEAL-PIT MODEL ($\simeq$ 73728) is greater than that of VanRullen and Thorpe's retinal model ($\simeq$ 43691), the information recovery plot is shown against similar *number* of spikes rather than the percentage. The progressive reconstruction of an input image rank-order encoded using



(a)



(b)

Figure 8.6: (a) Look-up-table for decoding rank-order encoded images using the FOVEAL-PIT MODEL, shown as a mean of the plots of rank-ordered coefficients for each of the sixty-five images in our data-set. (b) Comparison of the LUTs of the FOVEAL-PIT MODEL, FoCRen and VanRullen and Thorpe's retinal model.

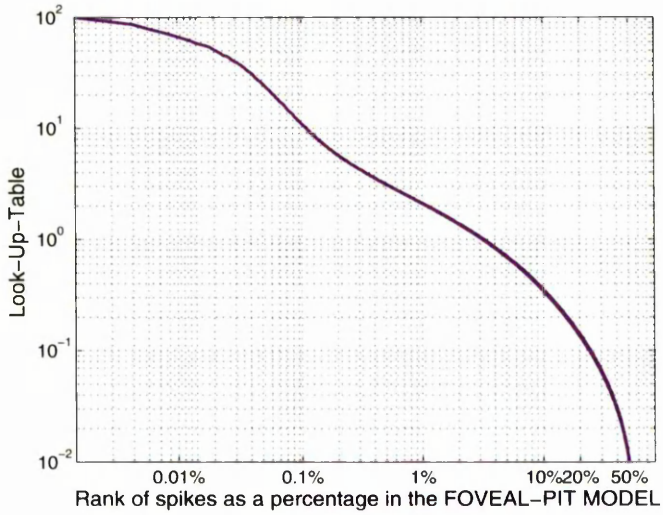
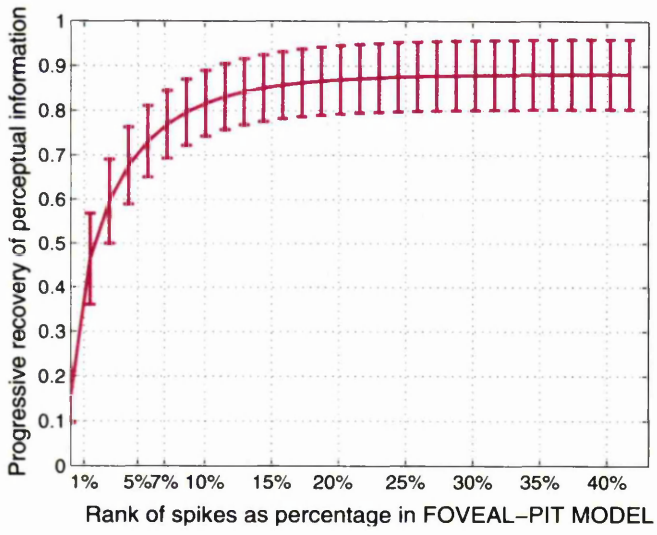


Figure 8.7: Look-up-table for decoding rank-order encoded images using FOVEAL-PIT MODEL.

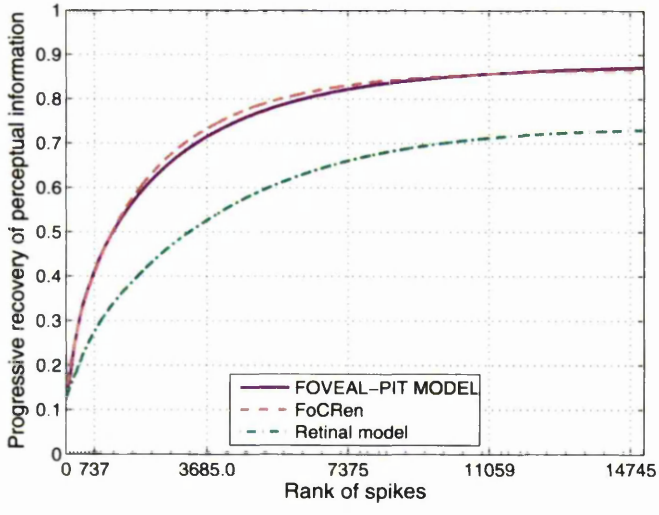
the three methods of encoding viz. the FOVEAL-PIT MODEL, FoCRen and VanRullen and Thorpe's retinal model, is presented for a qualitative evaluation in figure 8.9, while the progressive information recovery plots for the same is shown in figure 8.10. We observe that the rate and quantity of information recovery is almost the same for the FOVEAL-PIT MODEL and FoCRen. This in spite of the fact that the FOVEAL-PIT MODEL has only four layer (and sizes) of filters, while the retinal model has sixteen layers. Thus, the FOVEAL-PIT MODEL can be said to be performing more efficiently in rank-order encoding visual stimuli than VanRullen and Thorpe's retinal model. Further, both the models far out-perform VanRullen and Thorpe's original method of encoding. We may thus say that by implementation of biologically realistic *principles*, viz. FoCal with retinal model, and *models*, viz. FOVEAL-PIT MODEL, the performance of rank-order codes is much improved.

As a comparison of the two models, viz. the FOVEAL-PIT MODEL and VanRullen and Thorpe's retinal model, in light of the above results, we make a few observations:

Firstly, the number of cells in the FOVEAL-PIT MODEL, which approximately simulates only a very small eccentricity of the fovea, is much larger ($\simeq 30,000$) than that in VanRullen and Thorpe's retinal model, which is a simplified simulation



(a)



(b)

Figure 8.8: (a) The mean information recovery plot for progressive reconstruction of an input image using the LUT of the (a) FOVEAL-PIT MODEL. The standard deviation of the information recovery plot for individual images and all the sixty-five images in our data-set about the mean information recovery plot shown as error-bars. (b) Comparison of the information recovery plots from images rank-order encoded using the FOVEAL-PIT MODEL vs. FoCRen and VanRullen and Thorpe's retinal model.

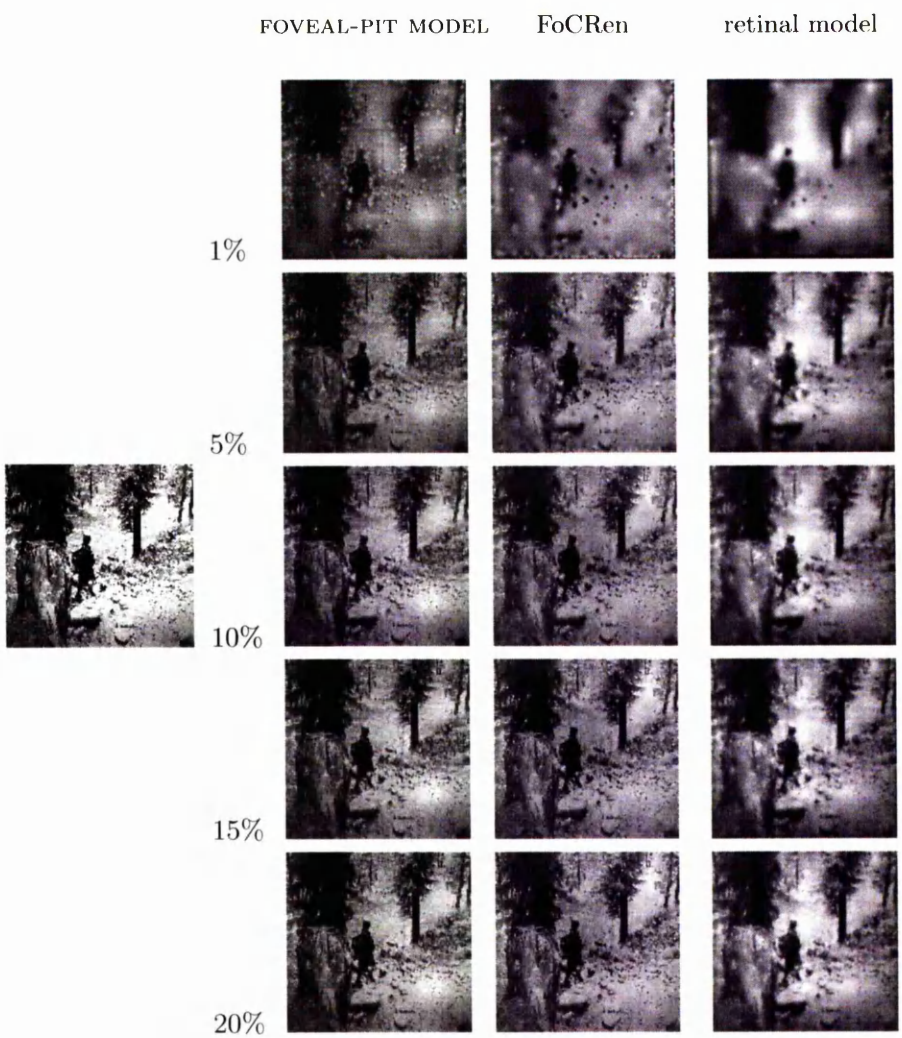


Figure 8.9: An input input image and its reconstruction using 1%, 5%, 10%, 15% and 20% of the LUT weights of (left) the FOVEAL-PIT MODEL, (middle) FoCRen and (right) VanRullen and Thorpe's retinal model.

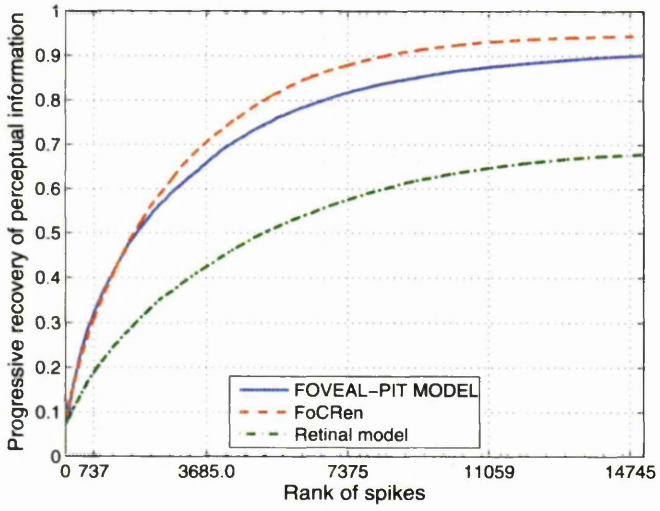


Figure 8.10: Perceptual information recovery plot for an input image shown in figure 8.9 using the LUT weights of the FOVEAL-PIT MODEL, FoCRen and VanRullen and Thorpe's retinal model.

of the various cell sizes across the whole of the retina. We have discussed earlier that accuracy of vision is the 'primary' function of the fovea, different parts of the retina having different specific 'primary' functions. The results above indicate that the visual accuracy required for recognition can be achieved with only four different sizes of cells. The addition of further increased cell sizes does not contribute to the information that is being transmitted to the brain.
Secondly, cells in sixteen independent layers (VanRullen and Thorpe's retinal model) parallelises the information retrieved from a scene four times more than that done by only four layers of cells (FOVEAL-PIT MODEL). However, keeping in view the similarity in performance of the two models, we may comment that a 'shallow' model with high density of cells in each layer nullifies the effect of 'deep' models with sparse population of cells in each independent layer, and provides a similar efficiency in information transmission.

8.3 Chapter summary

In this chapter, we have built a FOVEAL-PIT MODEL in order to study the effects of using a biologically realistic model for simulating rank-order encoding in vision.

Further, we simulate the biological mechanism of lateral inhibition to do away with the redundancy due to oversampling of the input by the cells of the model. This is done after the spikes are generated, which is an approximation of the processes used in biology, where such redundancy reducing technique is implemented prior to spike generation, i.e. at the photoreceptor layers. In spite of such an approximation, the results are the same as those of VanRullen and Thorpe's retinal model, which uses sixteen layers of filters to sample input data, compared to four layers of such cells in the FOVEAL-PIT MODEL. Thus, the FOVEAL-PIT MODEL is a 'shallow' but densely packed model, and resembles the retina more closely than a 'deep' and sparse model such as VanRullen and Thorpe's retinal model. We may thus say that rank-order codes perform efficiently in computational models that closely approximate the laws of biology.

In this work, we have used Petrovic's objective metric for studying the information recovery from rank-order codes with respect to vision. In the next chapter, we benchmark this measure with some commonly used measures in image processing such as Root Means Square Error and Fourier Analysis.

Chapter 9

Benchmarking

So far in our work, we have been quantitatively evaluating information recovery from rank-order encoded images using Q_{value}, an adapted version of Petrovic and Xydeas's objective measure [59]. It is a novel method of measuring the image fidelity using the rank-order 'codec' based on human visual system non-linearity. However, as with all novel methods, it needs to be benchmarked against an existing technique for image quality assessment. We do such a benchmarking in section 9.1 using Root Mean Squared Error (RMSE). We further validate Q_{value} in section 9.2 by observing the information recovery from data which is not a part of our data-set, thus doing out-of-sample testing for the objective metric. Finally, in section 9.3, we use Fourier Analysis to rank-order encode and decode images and obtain information recovery plot to further benchmark Q_{value} against an existing technique of image coding.

9.1 RMSE as information recovery measure

The objective metric Q_{value} we use in evaluating the information recovery from rank-order codes shows (figure 8.8(b)) that the FOVEAL-PIT MODEL and the FoCRen perform almost similarly, while both show a substantial improvement from the retinal model. Subjective evaluation of the image reconstruction quality shown in figure 8.9 conforms with the above results. At this point, it would be interesting to observe the quality of the reconstructed images during progressive recovery using an objective measure commonly used in image processing domain. Although Mean Square Error (MSE) estimates are found to be poorly correlated with subjective evaluation results [48, 87], our purpose here is to *compare* the

performance of the three models viz. retinal model, FoCRen and FOVEAL-PIT MODEL, rather than quantifying information recovery. We use the square root of Mean Square Error to compare the performance of the three models of rank-order 'codec' in information recovery with respect to the original image.

Root Mean Square Error (RMSE)

Let $\mathcal{I}$ and $\mathcal{R}$ be the original and reconstructed image respectively. Then the Mean Square Error between the two images is defined as [8, 49, 48]:

$$MSE(\mathcal{I},\mathcal{R}) = \frac{1}{V_T H_T}\sum_{x=1}^{V_T}\sum_{y=1}^{H_T}[\mathcal{I}(x,y) - \mathcal{R}(x,y)]^2,$$

where V_T and H_T are the total number of rows and columns in $\mathcal{I}$ and $\mathcal{R}$, and (x,y) represents a spatial location in both the image matrices. We calculate the Root of the Mean Square Error (RMSE):

$$RMSE(\mathcal{I},\mathcal{R}) = \sqrt{MSE(\mathcal{I},\mathcal{R})}.$$

Results

The RMSE for all the images in our data set was obtained during progressive recovery of each image using each of the three LUTs viz. that of the retinal model, the FoCRen and the FOVEAL-PIT MODEL. The average RMSE plot for each method is shown in figure 9.1. Comparing with figure 8.8(b), we find that the essential behaviour of the perceptual information recovery plot and the RMSE plots are the same. However, the performance of the FOVEAL-PIT MODEL is deteriorated with respect to FoCRen in figure 9.1 compared to figure 8.8(b). This may be attributed to the fact that the perceptual edge information measure takes into account only the edge information in the image, and measures this in terms of HVS contrast sensitivity. Thus, we might comment that although the image fidelity is better for FoCRen as compared to the FOVEAL-PIT MODEL as seen in figure 9.1, the perceptually important information content is almost the same as observed from figures 8.9 and 8.8(b).

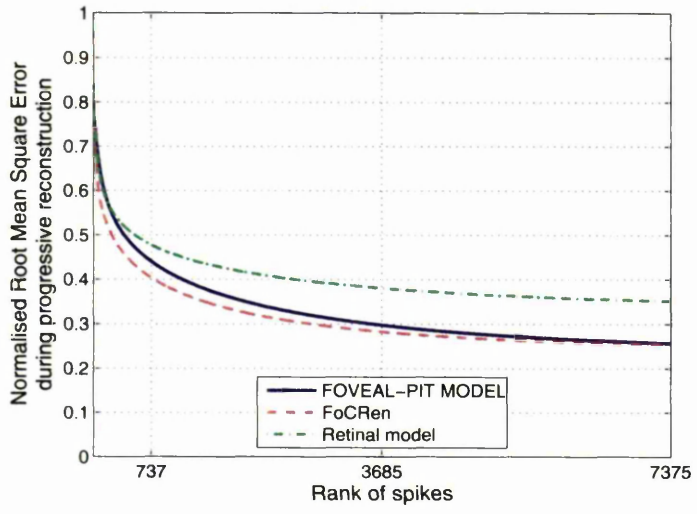
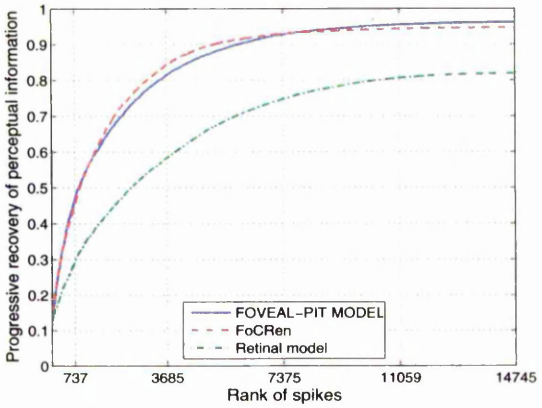


Figure 9.1: Root Mean Square Error plots during progressive reconstruction of images rank-order encoded using the retinal model, FoCRen and FOVEAL-PIT MODEL.

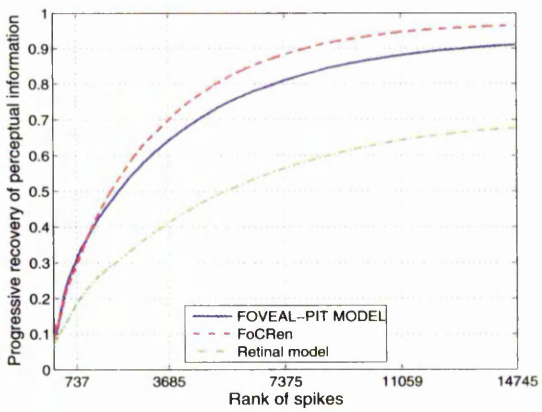
9.2 Out-of-set sample testing

In figure 4.3, we have presented some of the sample images from our data-set of sixty-five images used in this work. Using the objective metric Q_{value}, an average of approximately 75% information can be recovered from a rank-order encoded image with respect to the original when the retinal model is used as a rank-order 'codec', while approximately 85% average information can be recovered using FoCRen and the FOVEAL-PIT model. A subjective evaluation of reconstructed images also conforms with the quantitative results as has been discussed in sections 8.2.1 and 7.2.3. Further in section 9.1, we have seen that the results obtained using Q_{value} as the objective metric conforms with those obtained using RMSE, an objective measure used widely in image processing. These results speak optimistically about using Q_{value} as a standard measure for measuring perceptually-important information in future studies of neural codes with respect to vision. At this point, one might raise a question as to "How does Q_{value} perform as an objective measure for images which are not a part of the data-set, and thus do not contribute to the Look-Up-Table used for rank-order decoding?" A positive result in this regard would indicate the generic use of Q_{value} as an image quality metric to evaluate the preservation of perceptually-important information

(a)



(b)



(b)

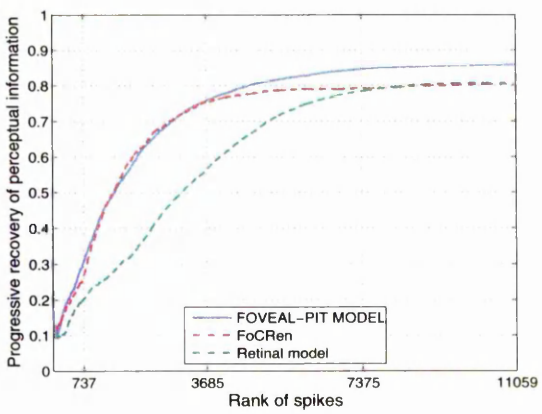


Figure 9.2: (a)– (c) Three images which are not a part of our data-set of sixty-five images and their perceptual information recovery plots using the LUT weights of the FOVEAL-PIT MODEL, FoCRen and the Retinal model.

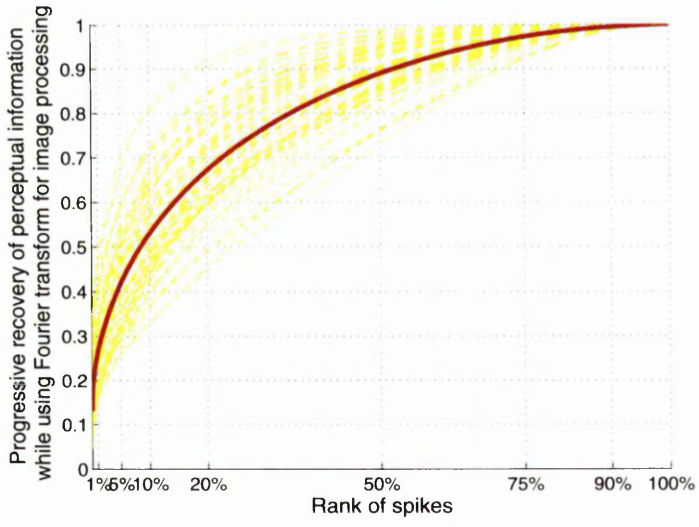
in one image with respect to another.

To test the information recovery from rank-order encoded images that are not a part of our data-set of sixty-five images, we obtain three images at random and observe the perceptually important information recovery plot using retinal model, FoCRen and FOVEAL-PIT MODEL. The images and their respective information recovery plots are shown in figure 9.2. We observe that the information recovery plot works well for images which are not a part of the data-set. Further to this we also observe that for the image in figure 9.2(c), which mainly consists of text[1], the information recovered using FOVEAL-PIT MODEL is approximately 5% more than that using FoCRen. Also, the total quantity of information recovered using the FoCRen and retinal model is the same, although the rate of information recovery is faster for FoCRen. An extensive search for image types for which Q_{value} may not work well may be done as future work. Further, in a recent study, it has been observed that Petrovic's objective metric works poorly in noisy images [56]. This aspect has not been dealt with in this work, and may be a hindrance when noise is incorporated in rank-order codes, and may also be dealt with in future work.

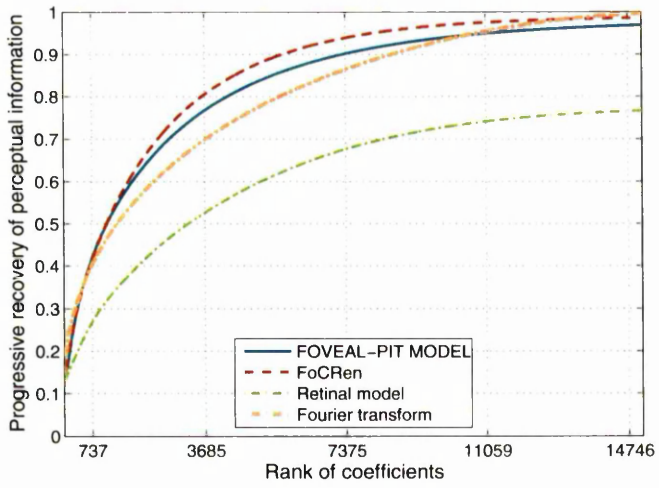
9.3 Fourier 'codec' and information recovery

There are several methods used in image processing applications which give perfect reconstruction of encoded images for example Fourier transforms, wavelet transforms, Laplacian pyramid, etc. [31, 61, 8]. Here, we choose to show a perfect image reconstruction obtained using Fourier transform, which is widely used to deal with various aspects of image processing. However, it might be reiterated here that in this work, we are simulating biological visual processing, and hence our transform is the DoG, which resembles ganglion cell receptive fields. These transforms are singular, and hence the use of pseudo-inverse method of decoding (chapter 6) to obtain a perfect reconstruction. We use Fourier transform to encode an image, rank order the Fourier coefficients and then use inverse Fourier transform to reconstruct the image progressively using the rank-ordered Fourier coefficients. We obtain Q_{value} at each step of progressive reconstruction and the information recovery plot thus obtained is compared with similar plots using the coefficients of the DoG filter of the three rank-order 'codec's presented in this

[1]Due to image size restrictions, a partial view of the nameplate showing occupants of IT302, IT Building, University of Manchester, is obtained for this study.



(a)



(b)

Figure 9.3: (a) The progressive information recovery from all the sixty-five rank-order encoded images when the Fourier transform is used as a basic filter instead of the DoG, shown as a spread about the mean. (b) The mean information recovery plot of '(a)' compared with the mean information recovery plots that were shown earlier in figure 8.3(b).

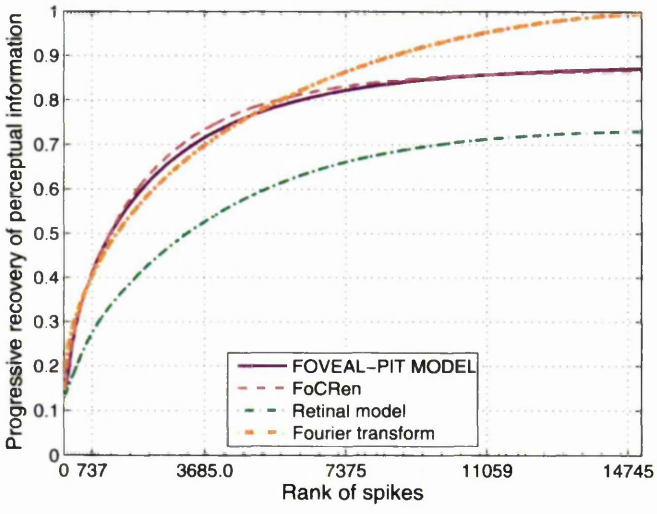


Figure 9.4: The mean information recovery plot of figure 9.3(a) compared with the mean information recovery plots that were shown earlier in figure 8.8(b).

work so far.

In figure 9.3(a), we show the perceptual information recovery plot for all the images in our data-set as a spread about the mean plot when an image is Fourier transformed, rank-order encoded, and then reconstructed using the inverse Fourier transform. The mean information recovery plot is compared with similar plots of the retinal model, FoCRen and FOVEAL-PIT MODEL in figure 9.3(b). We observe that the rate of information recovery is faster for the FOVEAL-PIT MODEL and FoCRen compared to that using Fourier transform. Although there is a full information recovery when Fourier transform is used for image encoding and decoding, it is only around 1 – 2% more than than using FOVEAL-PIT MODEL and FoCRen. Further, we compare the Fourier information recovery plot with the mean information recovery plots using the LUT of retinal model, FoCRen and FOVEAL-PIT MODEL in figure 9.4. We observe that even without using the true coefficients of filtering, the FOVEAL-PIT MODEL and FoCRen works better than the information recovery plot using Fourier Transform in terms of rate of information recovery for the first 5% coefficients. These observations open up the viability of rank-order 'codec' being used as an image compression model, and can be considered as a future work.

9.4 Conclusion

In this chapter, we have benchmarked Q_{value} against Root Mean Square Error as an image quality assessment metric. Further, we have studied the performance of this metric when Fourier transform is used for image processing in lieu of DoG filter. We also study the behaviour of Q_{value} for the retinal model, FoCRen and FOVEAL-PIT MODEL for images which are not a part of our data-set. Based on the observations, we have mentioned certain work that could be undertaken as future work.

In the next chapter, we discuss the conclusions that can be drawn from the work presented in this thesis, and also scope for future work.

Chapter 10

Conclusion and Future Work

To conclude, we review the work and results presented in this thesis. We restate our thesis questions from the Introduction chapter and examine how, and to what extent, the results of our work have been successful in answering these. In doing so, we draw conclusions on the implications of the results with respect to the thesis hypothesis. The conclusions drawn from the thesis leave a wide scope in carrying this research forward. We conclude the thesis by suggesting some viable research possibilities based on the work presented here.

10.1 Conclusion on hypothesis

The primary question that this thesis has tried to answer is *"Are rank-order codes, at all, able to represent input information efficiently, and if so, then how fast can that information be read from the codes?"* An answer to this query will be a conclusion about the thesis hypothesis. Towards that end, we investigate some sub-queries in the same sequence as stated in section 1.4.

We base our work on *VanRullen and Thorpe's retinal model* as a rank-order encoder of visual input. We start by validating the works of VanRullen and Thorpe, presented in chapter 4, whereby a static monochrome image is rank-order encoded using the model. The rank-order is then decoded by reconstructing the image and the image fidelity of the reconstruction with respect to the input is inspected visually. The purpose behind decoding is to test (i) the performance of rank-order codes in encoding visual information (ii) how fast perceptually-important information can be recovered from rank-order encoded data. From these perspectives, a qualitative evaluation of rank-order code performance is

presented as a part of the validation which confirms the results obtained originally by VanRullen and Thorpe. Having done so, we now proceed towards answering the thesis queries, the primary requirement and motive of such validation.

10.1.1 On quantifying information recovery

Question 1

The first query in trying to answer the primary question of the thesis was to ask *"How can we quantitatively measure the performance of the rank-order codes with respect to vision?"*. Since the contrast sensitivity of the human eye is nonlinear, as discussed in chapter 5, a least mean square or information theoretic measure of the information recovered from rank-order encoded images, as has been done in previous research [81], seems inappropriate. In chapter 5, we propose a novel way of quantitatively measuring the perceptually-important information in rank-order decoded images — by adapting an objective metric proposed by Petrovic and Xydeas which they used for measuring perceptual information preservation in image fusion [59]. In doing so, we come to our second query:

Question 2

"What effects does the measure have on the results of VanRullen and Thorpe's simulation?". We present the answer to this query based on our empirical results obtained in chapter 5:

- On average around 70% of the perceptually-important information with respect to the original image can be retrieved from rank-order encoded images by the time 20% of the ganglion cells in VanRullen and Thorpe's retinal model have fired their first spikes.
- Very little contribution to information recovery is made beyond this point, with on average 72–73% information recovered by the time 30% of the ganglion cells have fired their first spikes.
- Further, information recovery is faster for images with fewer edges than for those with more detailed content.
- Such quantitative comparisons agree with a visual inspection and comparison of the input and decoded, i.e. reconstructed, images.

- Decoding with the coefficients of filtering, instead of true rank-order decoding using a generic Look-Up-Table of weights, does not show a perfect reconstruction, which is in agreement with visual inspection of the reconstructed images. This observation leads us to a third query and is discussed below.

With such empirical results, we have obtained a satisfactory answer to our second query, so that we now proceed to the third query, concerning ways of improving the decoding of rank-order encoded images.

10.1.2 On optimising information recovery

Question 3

Based on the observation that the technique used by VanRullen and Thorpe in decoding rank-order encoded images is lossy, as has been confirmed with the quantitative measure, we come to our third query — *"Can the quantity of information retrieved from rank-order codes be improved by improving on the decoding techniques used by VanRullen and Thorpe?"*. In search of the answer to this query we introduce a pseudo-inverse method of rank-order decoding, whereby we obtain a substantial improvement in the amount of information recovered from rank-order encoded images, and also in the time to such recovery. The results from the empirical analysis as presented in chapter 6 are summarised as below:

- Perfect reconstruction of input image is obtained when coefficients of filtering are used for rank-order decoding.
- Total information recovery is on average 20% more than that with VanRullen and Thorpe's method of decoding.
- Rate of information recovery is faster compared with VanRullen and Thorpe's method of decoding.

Such empirical results indicate that the quantity and rate of information recovery from rank-order encoded images can indeed be improved by improving on decoding techniques. The novelty in the method of rank-order decoding is one of the important contributions of the thesis and has the potential of leading on to interesting future research, which will be discussed further in section 10.2.

Question 4

Although the rate and quantity of information recovery is improved using the pseudo-inverse method of decoding compared to VanRullen and Thorpe's method of decoding, the time to maximum information recovery is not until on average approx 37% of the ganglion cells have fired their first spikes. Moreover, the pseudo-inverse method of decoding is computationally expensive in terms of resources and time. Consequently empirical analysis with images of size larger than 32×32 has not been possible. A study with images of size 128×128 would be desirable. On such grounds, we propose to explore other viable options of optimising the rank-order code performance, the goal being to recover the maximum possible visually-important information in minimum time, the time until 20% of the cells have fired their first spikes being the ideal permissible 'time-limit' for information recovery. This leads us to our fourth query — *"How can the information recovery from rank-order codes be optimised?"*.

We fall back on intelligent and efficient technique of lateral inhibition, adopted by sensory evolution, to provide us with a novel algorithm, viz. Filter-overlap Correction algorithm (FoCal), and use it to improve the rank-order encoding used by VanRullen and Thorpe in their retinal model. In doing so, we do away with the approximation of VanRullen and Thorpe's method of rank-order decoding, and use it in image reconstruction. We obtain, not a perfect, but 'near' perfect recovery of perceptually-important information in the reconstructed image with respect to the original. However, the time to information recovery is much improved than that using VanRullen and Thorpe's rank-order 'codec' or the pseudo-inverse method of decoding. Thus, we have a trade-off between the quantity of information recovered and the time to information recovery — an optimum performance from the rank-order codes.

The empirical results of using FoCal as presented in chapter 7 are summarised below:

- On average around 15% increase in the total information recovery is obtained compared to encoding using VanRullen and Thorpes method.
- However, the rate of information recovery is very fast compared to the other two methods used so far viz. the pseudo-inverse method and VanRullen and Thorpes method. On average 75% of the information is recovered by the time only 10% of the cells have fired their first spikes, while a total

information recovery of around 85% on average is achieved by the time around 20% of the cells have fired their first spikes.

- Reconstruction of the input stimulus using coefficients of filtering is near perfect with respect to the original, with a total of around 99% perceptually-important information recovered from the rank-order codes.

We consider FoCal as the answer to our fourth query, and proceed to the fifth and final question in the next section.

10.1.3 On biologically realistic rank-order encoding

Question 5

The fifth and final question that we try to answer is *"Do rank-order codes perform better in a biologically realistic model?"*. In trying to answering this query, we make a major contribution to the thesis in designing and simulating the FOVEAL-PIT MODEL, which is a biologically realistic model of the foveal-pit in the retina. We use this model to process images and use FoCal to encode the information in rank-order. Thus, we have a biologically realistic rank-order encoder, whereby rank-order encoding is being done using models and principles based closely on biological parameters. We use VanRullen and Thorpe's method of decoding to retrieve information from encoded images using the above encoder. The results of our empirical analysis are summarised below:

- The total information retrieved from the codes is almost the same as that obtained using FoCal for rank-order encoding (FoCRen).

- This, is in spite the FOVEAL-PIT MODEL having only four layer of filters compared to sixteen in VanRullen and Thorpe's retinal model. This is a very high optimisation on the rank-order encoding mechanism.

- However, the rate of information recovery, on average, is slightly better for the FoCRen.

From the observation that the quantity and rate of information recovery in both the FOVEAL-PIT MODEL and FoCRen is almost same in spite of the former having quarter of the number of cell layers as that of the latter indicates a very

efficient rank-order encoding mechanism in the FOVEAL-PIT MODEL. Such observations lead to a very important conclusion — that a model designed closely on biological parameters performs optimally in encoding the maximum information with fewer resources (fewer ganglion cell layers). Thus, to our fifth query we can say that the rank-order codes indeed perform efficiently in a biologically realistic model.

Such answers to our sub-queries have answered our primary thesis question in the affirmative, so that we can restate our thesis hypothesis as a thesis conclusion: **"Rank-order codes are a viable means of information encoding in applications that require fast and efficient information transmission"**.

10.2 Future Work

In the following sections, we list some of the possible directions that can be adopted by future research in this area.

10.2.1 Towards realistic simulation

So far in our simulation, we have used the phrase 'by the time x% (say) of ganglion cells have fired their first spike' to indicate the time to information recovery. However, there are two presumptions here which are stated below, along with the possibilities of extension of the simulation presented in this thesis in doing away with these presumptions.

Incorporating time

The time to recovery is actually counted by the number of ganglion cells that have fired their first spikes, rather than the actual time to such firing. The rate of information recovery, thus, is quantified with respect to rank of spiking, and thus has a regularly spaced scale. However, a population of ganglion cell is unlikely to fire at uniform time intervals. Thus, quantifying the rate of information recovery with respect to time of firing — in other words, the latency of firing of the ganglion cells — will be a more realistic evaluation. Again, a realistic way of doing this would be as discussed below.

Incorporating neural models

That the ganglion cells are firing is only a presumption made based on the sign and magnitude of the coefficients of filtering an input image with simulated ganglion cell receptive fields. However, in real time, some neurons may fail to fire. Thus a more realistic simulation of firing ganglion cells would be to incorporate some model of a neuron using existing and available software. This would give a more realistic picture of rank-order encoding using the two models discussed in this thesis. Further, using such models, the time to spiking, as suggested above, will be incorporated by default.

10.2.2 Towards realistic neural coding

A cascaded model

Empirical evidence suggests that rate codes are ubiquitous in the central nervous system. Although recent research have shown evidence of time to first spikes being used as neural codes in the somatosensory pathways [53, 35], an independent existence of rank-order code seems highly unlikely. One interesting extension to the work in this thesis would be to incorporate a cascaded model whereby the first spikes are rank-order encoded, but subsequent spikes from each ganglion cell in a population are 'rate encoded'. Information retrieval using such a combined model could shed light on the biological plausibility of such a co-existence.

Noisy codes

So far, we have considered a time $t = 0$ when a population of cells is initiated with a stimulus and they start firing. On the contrary, empirical evidence suggests a low sustained rate of firing even when there is a total absence of input stimulus. Thus, on application of stimulus, the neurons are at different levels of membrane potential. Consequently, their firing latency may not vary linearly with the strength of the input stimulus. For example, a cell which is on the verge of reaching the threshold voltage will fire with very low input stimulus strength. On the other hand, a cell which is in an absolute refractory state will not fire even on application of a strong stimulus. Such non-linearity incorporates noise in the model and would be a very desirable addition to the existing simulation results in this thesis.

Encoding motion

We have already introduced eye saccades in chapter 8. This work could be carried further with eye saccades in real time. A different perspective would be to incorporate dynamic stimulus and study real-time encoding techniques of the eye, and thereby, the viability of using rank-order encoding techniques with motion.

10.2.3 Redundancy reduction pre spiking

The phototransduction taking place in the photoreceptors and subsequent transmission of electrical signals through the other neural layers of the retina is analogue in behaviour. The ganglion cells act as analogue to digital converters — converting the analogue variations in their respective cell membranes into voltage spikes at discrete intervals. In chapter 7, we introduced the Filter-overlap Correction algorithm, whereby we modify the firing order of the ganglion cells to simulate the technique of lateral inhibition, used by the photoreceptors for redundancy reduction in the input stimulus. However, the algorithm is only an approximation; in every iteration the spikes firing later in time are corrected with respect to the earlier firing spikes, while the earlier firing spikes remain uncorrected. This renders the algorithm a one-way correction for redundancy with a bias towards the earlier firing spikes. Lateral inhibition, on the other hand, as discussed in section 3.4.2, is a two-way process whereby each cell can be thought of as having a feedback from its neighbouring cell. Thus two neighbouring cells form a closed loop, and both get corrected for redundancy in data and continue to do so until a balance is reached. The process is analogous to an electrical power grid, where the distribution of electrical power to individual nodes is automatically balanced by interconnections between the nodes. Interestingly, the results of such an iterative correction is similar to that of the method of pseudo-inverses (discussed in chapter 6), which gives the least squares solution for a set of ill-conditioned equations. Moreover, such a correction is done by the photoreceptors, prior to passing on the electrical signal to subsequent layers. Thus, redundancy is reduced pre spiking, unlike in our simulation. Once percolated down to the ganglion cell layer for spike generation, we argue that a perfect stimulus reconstruction might be obtained by rank-order decoding, without going through the computationally expensive method of pseudo-inverses.

The arguments presented here are only a speculation, and the viability of

extending our FOVEAL-PIT MODEL to accommodate such a mechanism leaves interesting space for future research in this area. In fact, such an argument has the potential to be an independent research topic in itself, and information processing along the cone pathway can be simulated prior to feeding the above signals to ganglion cell layers.

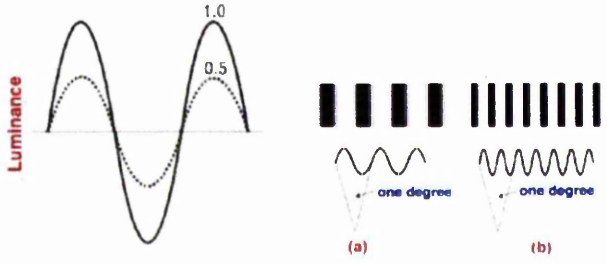
Appendix A

Contrast Sensitivity of Human Eye

The human eye can better distinguish between two objects or between an object and its background if the difference in luminance is large. In practice, relative difference in luminance is more important than the absolute difference [20]. The relative difference can be expressed by the *ratio between two luminance values*, known as the *contrast ratio*.

$$contrast\ ratio = \frac{L_{max}}{L_{min}},$$

where L_{max} and L_{min} are the maximum and minimum luminances in a certain image. The **contrast** between two objects is the difference between two luminances



(a) (b)

Figure A.1: (a) Contrast in a sinusoidal pattern. (b) A sinusoidal pattern and spatial frequency.

divided by the sum of them.

$$contrast = \frac{L_{max} - L_{min}}{L_{max} + L_{min}}.$$

If L_{mean} is the average luminance, and x is an arbitrary deviation of the amplitude about the mean luminance L_{mean}, then,

$$L_{min} = L_{mean} - x,$$

and

$$L_{max} = L_{mean} + x$$

$$\Rightarrow L_{max} - L_{min} = 2x$$

and

$$L_{max} + L_{min} = 2L_{mean}$$

Therefore,

$$contrast = \frac{x}{L_{mean}} \tag{A.1}$$

The reciprocal of the minimum contrast required for detection is called the **contrast sensitivity** of the eye.

Contrast for a sinusoidal luminance pattern is measured by the **modulation depth** of the pattern, defined as the amplitude of the sinusoidal variation divided by the average luminance, expressed in equation A.1 and as shown in figure A.1(a). The minimum modulation depth required for the detection of this pattern by the human eye is called the **modulation threshold** or the **detection threshold** of the eye. The **Contrast sensitivity of the human eye** is usually measured with sinusoidal luminance variations and is defined as the reciprocal of the *modulation threshold* as in equation A.2.

$$contrast\ \ sensitivity = \frac{1}{modulation\ \ threshold}. \tag{A.2}$$

The modulation threshold depends on the wavelength of the sinusoidal luminance variation.The reciprocal of the wavelength of a pattern as shown in figure A.1(b) is called its spatial frequency.

Bibliography

[1] B. A. Olshausen and David J. Field. Natural image statistics and efficient coding. *Network: computation in neural systems*, 7:333–339, 1996.

[2] B. A. Olshausen and David J. Field. Vision and coding of natural images. *American Scientist*, 88:238–245, 2000.

[3] Fred Attneave. Some informational aspects of visual perception. *Psychological Review*, 61(3):183–193, 1954.

[4] Horace B. Barlow. *Possible Principles Underlying the Transformation of Sensory Messages.* MIT press, Cambridge, Massachusetts, 1961.

[5] Horace B. Barlow. *Three Points About Lateral Inhibition.* MIT press, Cambridge, Massachusetts, 1961.

[6] Irwin B. Levitan and Leonard K. Kaczmarak. *The Neuron: Cell and Molecular Biology.* Oxford University Press, 198 Madison Avenue, New York, 10016, 2002. ISBN 0-19-514522-4.

[7] Andrew B. Watson. Efficiency of an image code based on human vision. *Journal of the Optical Society of America A*, 4(12):2401–2417, 1987.

[8] Andrew B. Watson. *Digital Images and Human Vision.* The MIT Press, Cambridge, Massuchesets and London, England, 1993. ISBN 0-262-23171-9.

[9] Francois Bergeaud and Stephane G. Mallat. Matching pursuit of images. *SPIE conference, Orlando, Florida, U.S.A.*, 2491:2–13, April 1995.

[10] Martin Cadik and Pavel Slavik. Evaluation of two principal approaches to objective image quality assessment. *Eighth International Conference on Information Visualization*, 4:513–518, 2004.

[11] E. D. Adrian. *Basis of Sensation.* Haffner Publishing Company, London, 1928.

[12] Scott Daly. *Digital Images and Human Vision, ed. Andrew B. Watson.* The MIT Press, Cambridge, Massuchesets and London, England, 1993. ISBN 0-262-23171-9.

[13] Peter Dayan and L. F. Abbot. *Theoritical Neuroscience: Computational and Mathematical Modeling of Neural Systems.* MIT Press, Cambridhe, Massachusetts, 2001. ISBN 0-262-04199-5.

[14] A. Delorme and S. Thorpe. Spikenet: an event-driven simulation package for modelling large networks of spiking neurons. *Network: Computational Neural Systems*, 14(4):613–628, 2003.

[15] John E. Dowling. *Neurons and Networks: An Introduction to Neuroscience.* The Belknap Press of Harvard University Press, Cambridge, Massachusets, 1992. ISBN 0-674-60820-8.

[16] M. E. J. Newman. Power laws, pareto distributions and zipf's law, January 2005.

[17] Koch Kristin et al. How much the eye tells the brain. *Current Biology*, 16:1428–1434, July 2006.

[18] Michael F. Land. *Eye Movements in Daily Life.* The MIT Press, Cambridge, Massuchesets and London, England, 2004. ISBN 0-262-03308-9.

[19] M. Fabre-Thorpe, A. Delorme, C. Merlot, and S. Thorpe. A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience*, 13:171–180, 2001.

[20] Peter G. J. Barten. *Contrast Sensitivity of the Human Eye and its Effects on Image Quality.* SPIE Optical Engineering Press, Bellingham, Washington, U.S.A., 1999.

[21] Stephane G. Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions On Signal Processing*, 41(12):3397–3415, 1993.

[22] J. Gautrais and Simon Thorpe. Rate coding versus temporal order coding: A theoretical approach. *Biosystems*, 48:57–65, 1998.

[23] Mitchell Glickstein. *The Visual Neurosciences volume-1, ed. L. M. Chalupa and H. S. Werner.* The MIT Press, Cambridge, Massuchesets and London, England, 2004. ISBN 0-262-03308-9.

[24] Qu Guihong, Dali Zhang, and Pingfan Yan. Information measure for performance of image fusion. *Electronics Letters*, 38(7):313–315, March 2002.

[25] Richard H. Masland. The fundamental plan of the retina. *Nature Neuroscience*, 4(9):877–886, 2001.

[26] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C.* Cambridge University Press, Cambridge, 1992.

[27] J. H. van Hateren. Real and optimal neural images in early vision. *Nature*, 360:68–70, November 1992.

[28] Joseph J. Atick. Could information theory provide an ecological theory of sensory processing? *Network*, 3:213–251, 1992.

[29] Joseph J. Atick and A. Norman Redlich. Towards a theory of early visual processing. *Neural Computation*, 2:308–320, 1990.

[30] Joseph J. Atick and A. Norman Redlich. What does the retina know about natural scenes? *Neural Computation*, 4:196–210, 1992.

[31] Peter J. Burt and Edward H. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4):532–540, April 1983.

[32] David J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of Optical Society of America A*, 4(12):2379–2394, 1987.

[33] David J. Field. What is the goal of sensory coding. *Neural Computation*, 6:559–601, 1994.

[34] James J. Gibson. *The Perception of the Visual World.* Houghton Mifflin Company, The Riverside Press, Cambridge, Massachusetts, 1950.

[35] Roland Johansson and Ingvars Birznieks. First spikes in ensembles of human tactile afferents code complex spatial fingertip events. *Nature Neuroscience*, 7(2):170–177, February 2004.

[36] H. K. Hartline. Inhibition of activity of visual receptors by illuminating nearby retinal elements in the limulus eye. *Fed. Proc.*, 69(8), 1949.

[37] H. K. Hartline, H. G. Wagner, and F. Ratliff. Inhibition in the eye of limulus. *Journal of General Physiology*, 39:651–673, 1956.

[38] Daniel Kersten. Predictability and redundancy of natural images. *Journal of Optical Society of America A*, 4(12):2395–2400, 1987.

[39] Helga Kolb. How the retina works. *American Scientist*, 91:28–35, 2003.

[40] Helga Kolb, Eduardo Fernandez, and Ralph Nelson. Organisation of the retina and visual system. `http://webvision.med.utah.edu/`, 2005.

[41] Helga Kolb and David Marshak. The midget pathways of the primate retina. *Documenta Ophthalmologica*, 106:67–81, 2003.

[42] Daniel L. Ruderman. Statistics of natural images. *Network: Computation in Neural Systems*, 5:517–548, 1994.

[43] Daniel L. Ruderman and William Bialek. Statistics of natural images: scaling in the woods. *Physical Review letters*, 73(6):814–818, August 1994.

[44] A. Lewis, R. Garcia, and Li Zhaoping. The distribution of visual objects on the retina: connecting eye movements and cone distributions. *Journal of Vision*, 3:893–905, December 2003.

[45] A. Lewis, R. Garcia, and Li Zhaoping. Understanding cone distributions from saccadic dynamics. is information rate maximised? *Neurocomputing*, 58–60:807–813, 2004.

[46] D. M. Dacey and M. R. Peterson. Dendritic field size and morphology of midget and parasol ganglion cells of the human retina. *Proceedings of the National Academy of Sciences of the USA*, 89:9666–9670, October 1992.

[47] Dennis M. Dacey. The mosaic of midget ganglion cells in the human retina. *The Journal of Neuroscience*, 13(12):5334–5355, December 1993.

[48] Ahmet M. Eskicioglu and Paul S. Fisher. Image quality measures and their performance. *IEEE Transactions on Communications*, 43(12):2959–2965, 1995.

[49] Rogerio M. Kinape and Mardson F. Amorim. A study of the most important image quality measures. *Proceedings of the* 25^{th} *Annual International Conference of the IEEE EMBS*, 1:934–936, September 2003.

[50] D. Marr. *Vision.* W. H. Freeman and Company, New York, 1982.

[51] Markus Miester and Michael J. Berry. The neural code of the retina. *Neuron*, 22:435–450, March 1999.

[52] Stefano Panzeri. *Course Lecture Slides of Computational Neuroscience.* Faculty of Life Sciences, University of Manchester, U.K., 2006.

[53] Stefano Panzeri, Rasmus Petersen, Simon Schultz, and Michael Lebedev. The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron*, 29:769–777, March 2001.

[54] L Perrinet, S Samuelides, and S Thorpe. Coding static natural images using spiking event times: do neuron cooperate? *IEEE Transactions On Neural Networks*, 15:1164–1175, 2004.

[55] Yuri Petrov and Li Zhaoping. Local correlations, information redundancy, and sufficient pixel depth in natural images. *Journal of Optical Society of America A*, 20(1):56–66, January 2003.

[56] Vladimir Petrovic. A human perception inspired quality metric for image fusion based on regional information. *Information Fusion*, 8(2):193–207, 2007.

[57] Vladimir Petrovic. Subjective tests for image fusion evaluation and objective metric validation. *Information Fusion*, 8(2):208–216, 2007.

[58] Vladimir Petrovic and Costas Xydeas. Evaluation of image fusion performance with visible differences. *Proceedings of European Conference on Computer Vision*, LNCS 3023:380–391, May 2004.

[59] Vladimir Petrovic and Costas Xydeas. Objective evaluation of signal-level image fusion performance. *Optical Engineering*, 14(8):1–8, August 2005.

[60] Gemma Piella and Henk Heijmans. A new quality metric for image fusion. *Proceedings International Conference on Image Processing*, 3:173–176, September 2003.

[61] Robi Polikar. The engineers ultimate guide to wavelet analysis. `http://users.rowan.edu/~polikar/WAVELETS/WTtutorial.html`, 1999.

[62] J. Puchalla, E. Schneidman, R. Harris, and Michael Berry. Redundancy in the population code of the retina. *Neuron*, 46:493–504, May 2005.

[63] Floyd Ratliff. *Inhibitory Interaction and the Detection and Enhancement of Contours.* MIT press, Cambridge, Massachusetts, 1961.

[64] G. S. Brindley. *Physiology of the Retina and Visual Pathway.* Edward Arnold Limited, London, 1970.

[65] Richard S. Snell and Michael A. Lemp. *Clinical Anatomy of the Eye.* Blacwell Science, Inc., Commerce Place, 350 Main Street, malden, MA, 1998. ISBN 0-632-04344-X.

[66] B. Sen and S. Furber. Information recovery from rank-order encoded images. *Proceedings of International Workshop on Biologically Inspired Information Fusion, Guildford, U.K.*, pages 8–13, August 2006.

[67] B. Sen and S. Furber. Maximising information recovery from rank-order codes. *Proceedings of SPIE Defense and Security Symposium, Orlando, Florida, U.S.A.*, 6570:65700C–1–12, April 2007.

[68] M Srinivasan, S. Laughlin, and A. Dubs. Predictive coding: a fresh view of inhibition in the retina. *Proceedings Royal Society London B*, 216:427–459, 1982.

[69] Michael T. Heath. *Scientific Computing: An Introductory Survey.* The McGraw-Hill Companies, Inc., London, U.K., 1997.

[70] S. Thorpe, R. Guyonneau, N. Guilbaud, J. Allegraud, and R. VanRullen. Spikenet: a simulator for modeling large networks of integrate and fire neurons. *Neurocomputing*, 26–27:989–996, 1999.

[71] S. Thorpe, R. Guyonneau, N. Guilbaud, J. Allegraud, and R. VanRullen. Spikenet: real-time visual processing with one spike per neuron. *Neurocomputing*, 58–60:857–864, 2004.

[72] Simon Thorpe. Ultra-rapid scene categorization with a wave of spikes. *Proceedings of the Second International Workshop on Biologically Motivated Computer Vision, Tübingen, Germany*, 2525:1–15, November 2002.

[73] Simon Thorpe, Arnold Delorme, and Rufin VanRullen. Spike-based strategy for rapid visual processing. *Neural Networks*, 14:715–725, 2001.

[74] Simon Thorpe, D. Fize, and C. Marlot. Speed of processing in human visual system. *Nature*, 381:520–522, 1996.

[75] Simon Thorpe and J. Gautrais. Rank order coding. *Computational Neuroscience: Trends in Research*, 13:113–119, 1998.

[76] Simon Thorpe and Imbert Michael. *Biological Constraint on Connectionist Modelling*. Elsevier, Amsterdam, 1989.

[77] Alecander Toet and Eric M. Franken. Perceptual evaluation of different image fusion schemes. *Displays*, 24:25–37, 2003.

[78] Alexander Toet and Maarten A. Hogervorst. Performance comparison of different graylevel image fusion schemes through a universal image quality index. *Proceedings of SPIE, Signal Processing, Sensor Fusion and Target Recognition XII*, 5096:552–561, August 2003.

[79] R. VanRullen, R. Guyonneau, and S. Thorpe. Spike times make sense. *TRENDS in Neurosciences*, 28(1):1–4, January 2005.

[80] Rufin VanRullen and Simon Thorpe. Is it a bird? is it a plane? ultra-rapid visual categorisation of natural and artifactual objects. *Perception*, 30:655–668, 2001.

[81] Rufin VanRullen and Simon Thorpe. Rate coding versus temporal order coding: What the retinal ganglion cells tell the visual cortex. *Neural Computation*, 13:1255–1283, 2001.

[82] Dawei W. Dong and Joseph J. Atick. Statistics of natural time-varying images. *Network: Computation in Neural Systems*, 6:345–358, 1995.

[83] Stephen W. Kuffler. Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16:37–68, 1953.

[84] R. W. Rodieck. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5:583–601, 1965.

[85] R. W. Rodieck. *The First Steps in Seeing.* Sinauer Associates, Inc., Sunderland, Massachusets, P.O.Box 407, 1998. ISBN 0-87893-757-9.

[86] Zhou Wang and Alan C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, March 2002.

[87] Zhou Wang, Alan C. Bovik, and Lu Ligang. Why is image quality assessment so difficult? *IEEE International Conference on Acoustics, Speech and Signal Processing*, 4:3313–3316, May 2002.

[88] Zhou Wang, Alan C. Bovik, Rahim R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

[89] H. Wassle and L. Peichl. Size, scatter and coverage of ganglion cell receptive field centres in the cat retina. *Journal of Physiology*, 291:117–141, 1979.

[90] Heinz Wassle. Parallel processeng in the mammalian retina. *Neuroscience*, 5:1–10, October 2004.

[91] Heinz Wassle and Brian B. Boycott. Functional architecture of the mammalian retina. *Physiological Reviews*, 71(2):447–480, April 1991.

[92] Heinz Wassle and et al. Retinal ganglion cell density and cortical magnification factor in the primate. *Vision Research*, 30(11):1897–1911, 1990.

[93] Costas Xydeas and Vladimir Petrovic. Objective pixel-level image fusion performance measure. *Proceedings of SPIE, Sensor Fusion: Architectures, Algorithms and Applications IV*, 4051:89–98, April 2000.

[94] Li Zhaoping. Different retinal ganglion cells have different functional goals. *International Journal of Neural Systems*, 3(3):237–248, 1992.

[95] Li Zhaoping. *The handbook of brain theory and neural networks, the second edition, ed. A. Arbib, Michael.* The MIT Press, Cambridge, Massuchesets and London, England, 2002.

[96] Li Zhaoping. Theoretical understanding of the early visual processes by data compression and data selection. *Network: Computation in Neural Systems*, 17(4):301–334, December 2006.