

## **Measurment of spatial orientation using a biologically plausible gradient model**

Sherwood, Adam

For additional information about this publication click this link.

<http://qmro.qmul.ac.uk/jspui/handle/123456789/5038>

Information about this research object was correct at the time of download; we occasionally make corrections to records, please therefore check the published record when citing. For more information contact [scholarlycommunications@qmul.ac.uk](mailto:scholarlycommunications@qmul.ac.uk)

ISSN 1470-5559

# Measurement of spatial orientation using a biologically plausible gradient model

Adam Sherwood



RR-05-01

April 2005



# Measurement of spatial orientation using a biologically plausible gradient model

Adam Sherwood

A Thesis submitted for the degree of Doctor of Philosophy

Department of Computer Science  
Queen Mary, University of London  
April 2005





## **Abstract**

This thesis presents a biologically plausible computational model for measurements of local spatial orientation in the human visual system — named ‘GROM’. A gradient framework is used to model cortical receptive fields. As background to this work relevant neurophysiology, key psychophysical paradigms and results are detailed.

The performance and parameter dependency of the model is examined using a range of key visual stimuli:

- simple stimuli: gratings of variable orientation and spatial frequency
- a natural image: a commonly used dense natural scene
- illusions of orientation: Café Wall illusion and contrast-modulated gratings.

Model performance on illusory stimuli is compared to human psychophysical performance. Results show that the model is in agreement with tendencies found psychophysically. This supports GROM’s biological plausibility. Heuristics are developed to aid an understanding of how the model generates these illusionary tendencies.

Finally, a novel genetic-algorithm approach was employed to breed textures with maximally illusory structure. Psychophysical testing of these textures shows their illusory nature. This methodology and the results it generates represent a novel, experimentally verified prediction of the model.

# Declaration

I declare that this thesis has been composed by myself and that it describes my own work. It has not been accepted in any previous application for a degree. All verbatim extracts are distinguished by quotation marks and that all sources of information have been specifically acknowledged.

Adam Sherwood

# Acknowledgements

At the start of this journey Peter McOwan showed great faith in me by taking me under his wing to be his research student. He then guided me through the research process in an informal, well judged fashion, performing deft managerial manoeuvres often without me realising. More recently he held my hand through the painful writing up phase. Through it all he remained good humoured, enthusiastic and the most intellectually inspiring figure of my life. I owe big, yet I know nothing is expected in return.

Derek Arnold and Alan Johnston both helped enormously during my research. Fabrizio Smeraldi and Graeme Andrews both helped in mathematical discussions. Jason Dale, Andrew Graves and Lukas Zalewski all helped out with various programming issues.

Systems support were ever patient in the face of my continued badgering. In particular Tim Kay, Matt Bernstein, Derek Coppen, Simon Boggis, Keith Clarke, Dave Hawes and Colin Powel. The Administrative staff should also be acknowledged as a well oiled team and factory of classic emails, heckles and banter.

My original crew, office mates and friends, Andrew Anderson and Keith Anderson, supposedly not related, created a happy working atmosphere and many amusing distractions.

Friends and cohorts of the department provided moral support and trips to the greasy spoon or hospital cafeteria for lunch, they include; Hayley Hung, Alessio Del Bue, Dave Russel, Alex Leung, Melanie Aurnhammer and Tao Xiang (Tony).

Friends outside academia have also been caught in the vortex of thesis writing, putting up with my flakiness and mood swings. Cécile, Adam, Rob, Andy, Fred, etc.

My parents also put up with my lack of visits, frequent excuse making and surprisingly encouraged me to start a PhD in the first place! My understanding girlfriend Esther performed multiple roles, including grinch support and proof reader general.

Thanks to them all.

# Contents

<b>Declaration</b>	<b>1</b>
<b>Acknowledgements</b>	<b>2</b>
<b>1 Introduction</b>	<b>8</b>
1.1 Road map . . . . .	8
1.1.1 Background introduction . . . . .	8
1.1.2 Model development . . . . .	9
1.1.3 Model results . . . . .	9
1.1.4 The model applied to two illusions of orientation . . . . .	10
1.1.5 Automatic Generation of Illusory Stimuli - AGIS . . . . .	10
1.2 Background introduction . . . . .	10
1.2.1 The problem . . . . .	10
1.2.2 Mode of exploration . . . . .	11
1.2.3 The Plenoptic function and early vision . . . . .	12
1.3 Biology of the visual system . . . . .	13
1.3.1 Pre-cortical areas . . . . .	13
1.3.2 Primary visual cortex . . . . .	15
1.3.3 Orientation selectivity . . . . .	16

1.3.4	Cortical receptive fields . . . . .	18
1.4	Cortical modelling . . . . .	20
1.4.1	Acuity . . . . .	20
1.4.2	Receptive field fitting . . . . .	20
1.5	Motion perception and the second-order paradigm . . . . .	22
1.6	Spatial orientation . . . . .	28
1.6.1	Psychophysical orientation measures . . . . .	28
1.6.2	Texture based investigation . . . . .	29
1.6.3	Second-order spatial stimuli . . . . .	30
1.6.4	Population coding . . . . .	33
1.6.5	Mathematical orientation models . . . . .	35
1.6.6	Computer vision approaches to measuring orientation . . . . .	37
1.7	Modelling . . . . .	39
1.7.1	Basic modelling paradigm . . . . .	39
1.7.2	Occam's razor and second-order stimuli . . . . .	40
<b>2</b>	<b>Model development</b>	<b>41</b>
2.1	Finding an image derivative . . . . .	41
2.2	Taylor/Jet representation . . . . .	43
2.3	Two-dimensional Tangent Measure . . . . .	45
2.4	Contrast invariance . . . . .	48
2.5	Multi-scale tangent estimation . . . . .	48
2.6	Tangent measures at many orientations . . . . .	49
2.7	Favouring oblique measures . . . . .	52
2.8	Deducing a final single measure . . . . .	53
2.9	Summary . . . . .	53

<b>3</b>	<b>Model Results</b>	<b>60</b>
3.1	Basic Stimuli . . . . .	60
3.2	Natural Images . . . . .	68
3.3	Machine vision comparison . . . . .	74
<b>4</b>	<b>GROM applied to contrast-modulated gratings</b>	<b>77</b>
4.1	Introduction . . . . .	77
4.2	GROM orientation map results . . . . .	78
4.3	Population based analysis . . . . .	82
4.4	Conclusions . . . . .	87
<b>5</b>	<b>GROM applied to the Café Wall illusion</b>	<b>88</b>
5.1	Introduction . . . . .	88
5.2	GROM and the Café Wall . . . . .	94
5.3	The effect of brick offset in the Café Wall . . . . .	97
5.4	Understanding GROM's performance with the Café Wall . . . . .	100
5.4.1	Estimation of direction of maximum gradient . . . . .	100
5.4.2	Inclusion of multiple derivatives — extending the k-vector . . . . .	103
5.4.3	Difference in zero-crossings of $G_{2,0}$ and $G_{1,0}$ . . . . .	106
5.5	Conclusion . . . . .	109
<b>6</b>	<b>Automatic generation of illusory stimuli</b>	<b>111</b>
6.1	Introduction to Evolutionary algorithms . . . . .	111
6.2	Applications of Genetic Algorithms . . . . .	113
6.3	AGIS: Stimulus generation . . . . .	116
6.3.1	AGIS: A broad outline . . . . .	116
6.3.2	The goal of the Genetic Algorithm used in AGIS . . . . .	116

6.3.3	The details of the Genetic Algorithm . . . . .	117
6.3.4	Parameters . . . . .	121
6.4	Technique validation . . . . .	122
6.4.1	Evolution of textures . . . . .	123
6.4.2	Psychophysical testing . . . . .	123
6.4.3	Psychophysical results . . . . .	130
6.4.4	Preferential model selection . . . . .	131
6.5	AGIS Summary . . . . .	131
<b>7</b>	<b>Conclusions</b>	<b>135</b>
7.1	Main model results . . . . .	135
7.2	Illusory stimuli . . . . .	137
7.2.1	The contrast-modulated stimulus . . . . .	138
7.2.2	The Café Wall stimulus . . . . .	139
7.3	AGIS . . . . .	141
7.4	Areas of novelty . . . . .	141
7.5	Future work . . . . .	142
7.5.1	Sparse sampling . . . . .	142
7.6	Textures of Roncato . . . . .	144
7.6.1	Grouping on orientation . . . . .	145
7.7	Concluding remarks . . . . .	146
<b>A</b>	<b>Gaussian derivative synthesis</b>	<b>151</b>
<b>B</b>	<b>Image derivatives</b>	<b>154</b>
<b>C</b>	<b>Gaussian derivatives</b>	<b>155</b>

<b>D</b>	<b>Convolution and Separability</b>	<b>157</b>
<b>E</b>	<b>Steerable Filtering</b>	<b>159</b>
<b>F</b>	<b>Steerable convolution</b>	<b>161</b>
<b>G</b>	<b>Reference frame rotation</b>	<b>162</b>
<b>H</b>	<b>Simple Fourier-energy model implementation</b>	<b>165</b>
<b>I</b>	<b>Minimisation</b>	<b>167</b>
I.1	An example in one dimension . . . . .	167
I.2	An example in two dimensions . . . . .	169
	<b>Bibliography</b>	<b>180</b>



# Chapter 1

## Introduction

How does the human visual system measure the spatial orientation of detail observed in a scene? This is the principal research question this thesis sets out to explore.

This chapter is designed to give background to the research question. The first section is designed to act as a guide to the thesis or ‘road map’, detailing the contents of each of the subsequent chapters; sub-section titles roughly follow the subsequent chapter titles.

The computational model presented in this thesis is referred to as: the Gradient based Robust Orientation Model, for convenience, this is abbreviated to ‘GROM’. It is termed ‘gradient based’ as the model utilises gradient information and ‘robust’ as it is designed to estimate orientation in a variety of poorly-conditioned cases.

### 1.1 Road map

#### 1.1.1 Background introduction

This section contains a review of the background biological and computational context that GROM is based on. The *Plenoptic Function* rationalisation of the visual world is presented and is followed by an introduction to the biological architecture and processes that interpret low-level visual information in the visual cortex. Particular attention is paid to the nature, and modelling of, orientation selectivity in V1 and the associated cortical receptive fields therein.

As a well studied and debated problem similar in nature to spatial orientation modelling, a section is dedicated to introducing motion perception and the second-order paradigm.

The section entitled ‘Psychophysical orientation measures’ (see page 28) contains a discussion of the experimental methodologies that relate to measures of orientation in the human visual system. It covers texture segmentation models followed by more sophisticated and explicit Fourier-based models. The conceptual linkage between Fourier-based spatial orientation models and motion models is discussed. Spatial orientation models based on population coding techniques are also discussed. Finally in this section, models from machine vision, unconstrained by biological plausibility, are discussed.

For completeness a brief discussion of the philosophical framework behind the biological modelling process is included. This includes a discussion of Marr’s modelling approach [Marr, 1982] and Occum’s razor in relation to second-order stimuli.

### 1.1.2 Model development

Chapter 2 details the mathematical steps of the model. The model relies on gradient measures. These gradients are found by taking the derivative of image structure as detailed in ‘Finding an image derivative’ (see page 41). The section entitled ‘Taylor/Jet representation’ (see page 43) illustrates the concept of ‘jets’ - groups of multiple derivatives taken at each image point. ‘Jets’ are used in the tangent estimation described in ‘Two-dimensional Tangent Measure’ (see page 45). The section entitled ‘Multi-scale tangent estimation’ (see page 48) describes a multi-scale extension to the least-squares tangent measure explained in the section prior.

In modelling the orientation selective structures of the human visual cortex, estimates of tangent are made in multiple rotated reference frames, this procedure is described in the section entitled ‘Tangent measures at many orientations’ (see page 49).

A further step that increases robustness in the model is described in ‘Favouring oblique measures’ (see page 52) and the final single measure of orientation is deduced in ‘Deducing a final single measure’ (see page 53).

### 1.1.3 Model results

This chapter contains three interlinked sections, in each, GROM output is examined with different stimulus types. These sections address the following questions:

- what is the effect of manipulating model parameters with simple sine-wave stimuli and circular sine-wave stimuli?
- what is the effect of manipulating model parameters with a commonly used unscaled natural image?

- How does GROM compare to a machine vision algorithm for measuring orientation?

#### 1.1.4 The model applied to two illusions of orientation

In order to test GROM’s efficacy as a perpetual model, results from it must be compared to human psychophysical performance in relevant tasks. It is argued that, not only should the model perform as expected on simple stimuli, but it should also replicate the phenomena observed in illusions of orientation. The stimuli tested are the contrast-modulated grating (a second-order stimulus) in Chapter 4 and the Café Wall illusion (a geometric illusion) in Chapter 5.

#### 1.1.5 Automatic Generation of Illusory Stimuli - AGIS

In the previous chapters results from GROM are shown to reflect human psychophysical performance in two well-studied illusions of orientation. Chapter 6, entitled ‘Automatic generation of illusory stimuli’ (see page 111), details a novel method of generating new illusions of orientation named Automatic Generation of Illusory Stimuli (abbreviated to AGIS). The method presented in this chapter uses a Genetic Algorithm technique combined with GROM to ‘breed’ new illusions of orientation. The bred stimuli are tested psychophysically and shown to contain illusions of orientation. This section in effect provides an experimental prediction of the model, which is then psychophysically validated.

## 1.2 Background introduction

### 1.2.1 The problem

Study of the human visual system is the vanguard in the field of neuroscience. Understanding how vision works gives insight into the mechanisms of the brain while at the same time enabling us to harness its robust evolved algorithms to build artificial intelligence.

The human visual system extracts many different and varied qualities from a visual scene, including depth, motion and colour. Arguably one of the most basic of these is orientation. Orientation is a so called ‘low-level’ quality as it is thought to be the precursor to higher-level functions such as form and consequently object recognition [Marr, 1982].

This thesis aims to describe a model for how the human visual system measures the orientation of detail in a scene. The meaning of orientation here, is the alignment of detail at a point in the foveal visual field, for example, straight lines whose luminance varies vertically, are horizontal.

Applying GROM to any point on those lines should result in a measure that indicates horizontally aligned detail.

The rest of this chapter is dedicated to introducing the problem of measuring orientation from various perspectives. The different fields that are relevant to the study of this problem are put into context, then the generally accepted ‘scientific knowledge’ and current state of the literature in each field is presented; finally, the philosophy of computational modelling approach is discussed.

### 1.2.2 Mode of exploration

The subject of this thesis is interdisciplinary and therefore draws on the results of scientific study from various backgrounds. These include:

- **Neurobiology** — modelling of any biological mechanism must first establish the behaviour of the most fundamental elements of the system. This includes organisation in the visual cortex and cortical receptive field types.
- **Psychology** — a system as complex as the human brain is difficult to model based only on the study of its fundamental elements – Psychology aims to understand the living human brain by measuring its behaviours. The cells and pathways studied on a neurophysiological level are often not human but from primates, cats or even rats. Vertebrate mammals have similar visual systems but the human visual system is arguably the most advanced — with such a broad range of abilities [Bruce *et al.*, 1996, Zeki, 1993].
- **Functional imaging** — it is possible to measure activity in parts of the human brain using non-invasive imaging tools such as *functional magnetic resonance imaging* (fMRI) or *positron emission topography* (PET), these provide a resolution down  $\sim 2mm$  cubed [Dougherty *et al.*, 2003]. Function can be attributed to neural architecture at a coarse scale using *transcranial magnetic stimulation* (TMS) whereby function is temporarily disabled and locatable.
- **Cell recordings** — On a much finer scale, it is also possible to measure single-cell activity in humans [Lenz *et al.*, 1988]. Techniques are also being developed that will give access to multiple single-cell recordings in humans [Vetter *et al.*, 2004]. These will potentially overcome the subjectivity of single-cell recordings — as they can potentially be atypical cell types. Examining groups of cells *in vivo* in visual cortex could potentially feedback into any modelling endeavour.
- **Psychophysics** — as a complimentary exploration tool, psychophysical experiments are

performed on human subjects. The difficulty with this approach is that it doesn't attribute overall system performance — that is, human behaviour — to specific areas of the brain.

- **Modelling** — various models discussed in this thesis aim to bring together understanding gained from the aforementioned disciplines. Understanding how other models work and fail to work feeds into the model described here.

### 1.2.3 The Plenoptic function and early vision

The word Plenoptic comes from plenus meaning full or complete, and optic. The plenoptic function is a concept used to concisely detail what is measurable from a visual scene [Adelson and Bergen, 1991]. There follows an attempt to describe the different aspects of the plenoptic function and how it relates spatial orientation and motion.

Placing a pinhole camera at a point in three-dimensional space would result in an inverted image being produced on the back of the camera apparatus, on the 'image plane'. This projected image is a multi-dimensional measurement of the scene, at a time  $t$ . Contained within the projected image is information about the colour of the objects in the scene, related to the wavelength,  $\lambda$ , and a representation of spatial distribution of the luminance within the scene from the point of view of the camera.

To place a pinhole camera in space to capture the image, a restriction has been placed on the total amount of visual information that is captured, compared to that which is available in all space. The placing of the pinhole camera itself, irrespective of its orientation toward a particular scene, (dealt with in  $\phi$  and  $\theta$ ) has, in effect, the total visual information available. Put simply, scenes look different when viewed from physically different places. This introduces three final variables to the plenoptic function;  $V_x$ ,  $V_y$  and  $V_z$  — the positioning of the camera in an arbitrary three dimensional Cartesian coordinate system.

In descriptions of human vision,  $V_x$  is normally taken to be the line determined by both eyes. The slightly different scenes that are seen from each eye can be understood as two discreet sampling points along  $V_x$ . Variation along this line can be described as parallax. Variations over time in parallax can indicate looming, where objects approach the subject.

Bringing all these variables together any optical scene should be describable by a function,  $p$ , which varies  $p(x, y, \lambda, t, V_x, V_y, V_z)$ .

This conceptualisation is useful for simplifying investigations into vision, as pointed out by De Yoe and Van Esen [De Yoe and Van Esen, 1988], the basic strategy for measuring any of these qualities is the same - measuring orientation within a projection of this space. Most attempts to describe the human visual system recognise the variables of the plenoptic function as the varying

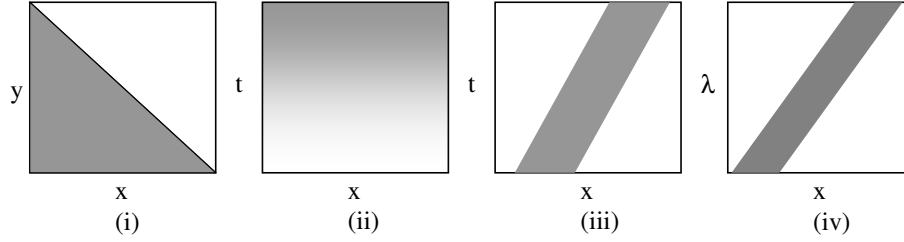


Figure 1.1: This illustrates how simple features in space are represented in a collection of two dimensional plenoptic functions. The measured variable is response, indicated by the grey level, e.g. brightness at a particular point in space or time, or at a particular frequency of light. [i] Image plane plenoptic function showing region of darkness in bottom left. [ii] Plenoptic projection showing effect of increase in brightness, gradual in time, uniform in x direction. [iii] Shows a bar moving uniformly in x direction [iv] A rainbow varying in the x direction.

attributes of a scene, then reduce any given problem to a useable sub-set of those attributes. Some examples of this are shown in Figure 1.1 (see page 13).

Although measuring gradients in plenoptic projections is a useful in concept, the biological approaches will differ according to the physics of each projection. For example, McOwan and Johnston [McOwan and Johnston, 1996] performed psychophysical investigation of analogous illusions in the space-space and space-time projections. Their findings, replicated by GROM, suggest different encoding stratagem in different projections.

## 1.3 Biology of the visual system

### 1.3.1 Pre-cortical areas

In order to model a low-level visual process, it is first necessary to understand how light information is transformed and subsequently processed in the visual parts of the brain. This section will briefly describe the biology of the visual system from transmission of information from the eye, through the visual pathways to the primary visual cortex.

The eye is the transducer between the light information of the outside world and information processing axons firing inside the brain. A general good description of the eye can be found in the following references [Tov  , 1996, Bruce *et al.*, 1996, Zeki, 1993].

Kuffler discovered that some retinal cells have concentric receptive field patterns [Kuffler, 1953], where light illuminating a circular central region made the cell fire above the background level whilst illuminating the annular surround of the central region made the cell fire below

background. These are named *centre-on* receptive fields. Cells with receptive fields of the opposite polarity were also found; these are named *centre-off*.

There is concern about the over-simplification of the eye as a simple transducer, devoid of processing; this largely surrounds the role of *amacrine cells*, whose scheme of connection varies widely [Vaney and Hughes, 1990]. However, the majority of cells represent the widely accepted view of the eye, as being responsible for transduction, light-level adaptation and some simple spatial processing via centre-surround receptive fields.

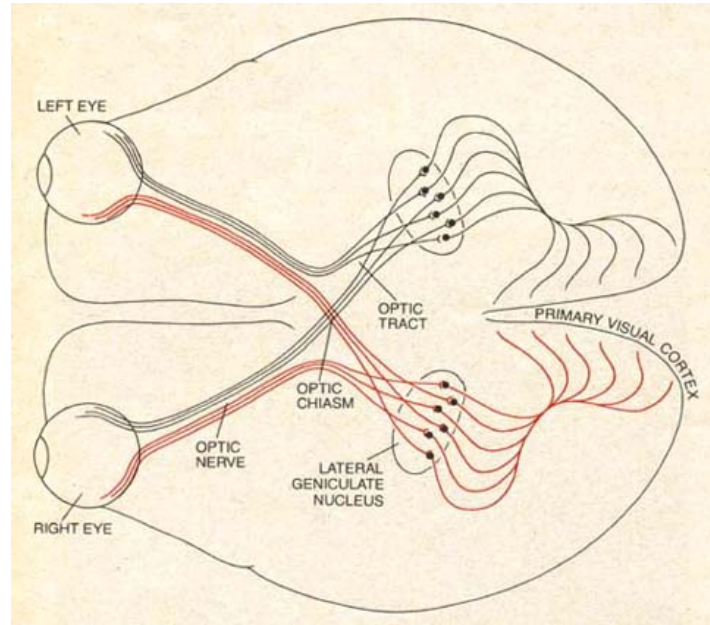


Figure 1.2: A diagram showing the primary optic pathways leading from the eye, taken from the original Hubel and Wiesel publication in Scientific American [Hubel and Wiesel, 1979].

Axons from retinal cells in both eyes cross-over at the *optic chiasm*, where the left-hand visual field of view is routed to the right hemisphere and vice-versa, as in Figure 1.2 (see page 14). After the chiasm, the axons terminate at the *lateral geniculate nucleus* (LGN). The LGN is formed of six layers, interleaved of inputs from opposing eyes. The distribution of cells within the layers is still *retinotopic* — meaning that the relative spatial distribution of activity due to light at the retina is preserved in these cells. From here, the principal pathway, the *optic radiation*, leads to the *primary visual cortex* at the back of the brain [Zeki, 1993].

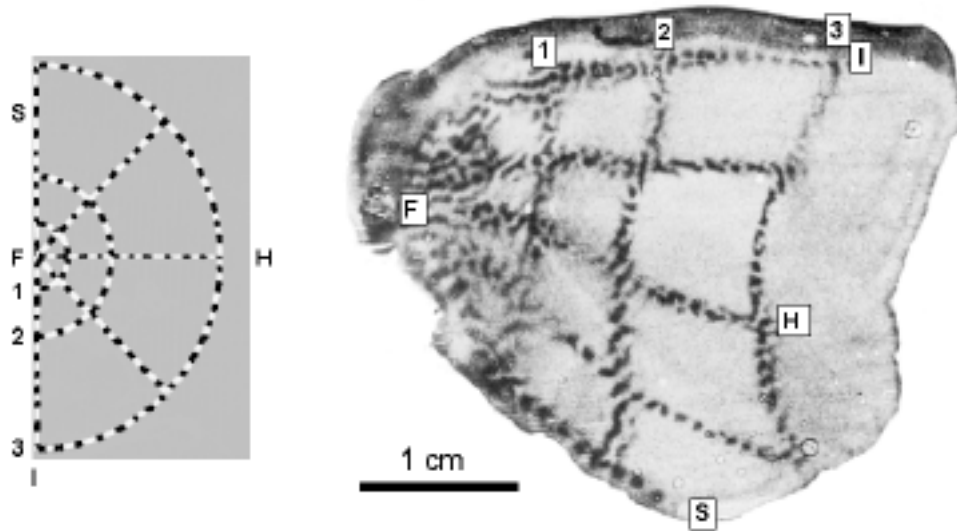


Figure 1.3: Adapted from a diagram by Tootell [Tootell *et al.*, 1988], this diagram shows a visual stimulus (left) presented to a sectioned macaque monkey with a flattened, deoxyglucose autoradiograph of V1 (right) showing the distribution of metabolic activity in darkened regions. The tangential lines labelled 1, 2, 3, are at  $1^\circ$ ,  $2.3^\circ$  and  $5.4^\circ$ . The line from the fovea, F to label H demarcates the horizontal meridian, I, the lower visual field while S indicates the upper.

### 1.3.2 Primary visual cortex

We now consider the role of the primary visual cortex in visual processing. Through a mixture of neuro-physiological and anatomical investigation, the emerging evidence suggests that this area of cortex is the first and primary visual area in which the visual scene is ‘broken-down’ for subsequent processing by higher visual areas. Investigations detailed below, have shown a strong and intriguing relationship between neural architecture and function.

The primary visual area is known as *V1*, particularly in relation to humans and the *striate cortex*, due to the notable striping parallel to the cortical surface when stained to reveal *cytoarchitecture*, or anatomical cell nucleus distribution. The optic radiation terminates intricately in different layers, best divided into their originating pathways from the LGN [Tovée, 1996]:

- *Parvocellular* layers of the LGN terminate in layers 4a and 4c $\beta$ , which lead to V1 layers 2 and 3.
- *Magnocellular* layers of the LGN terminate in layers 4c $\alpha$  and layer 6 of V1.

The amount of cortex dedicated to a part of the visual field varies with eccentricity. This is



known as the *cortical magnification factor*. fMRI studies have explored this function in primary visual cortex [Engel *et al.*, 1997, Popovic and Sjostrand, 2001] and later visual areas [Dougherty *et al.*, 2003]. Foveal vision clearly occupies more of the visual cortex than peripheral, and consequently has greater acuity due to this [Duncan and Boynton, 2002] Figure 1.3 (see page 15).

Organisation and cell connectivity up until this area is comparatively well known and understood. However, specific accounts of the organisation of V1 vary. There are features including receptive fields, orientation columns and ocular dominance columns, that are widely agreed upon and give insight into the possible function of this area. These features will form the basic elements of the biologically plausible computation model presented below.

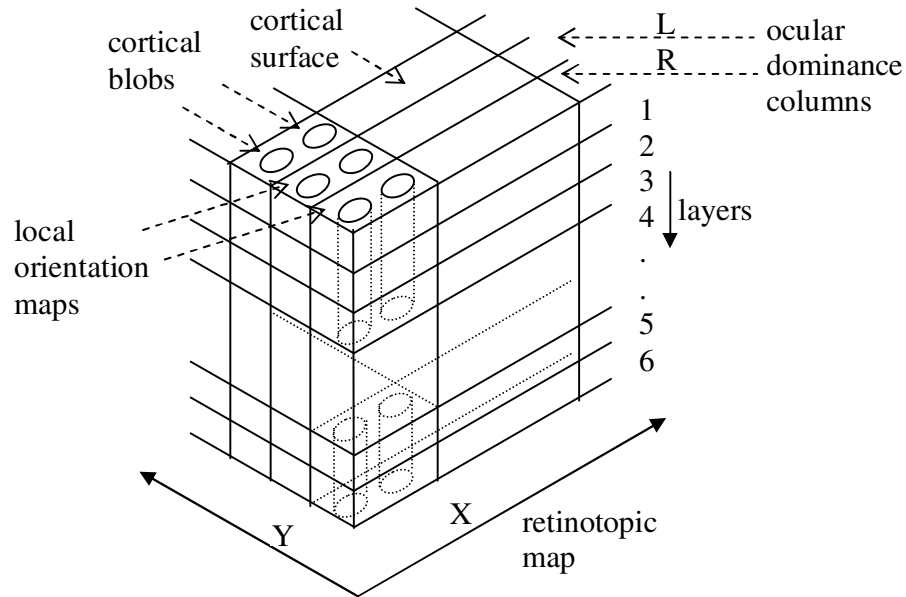


Figure 1.4: A schematic diagram showing the modular organisation of the primary visual cortex. Whilst macroscopically retinotopic, it has microscopically orientation maps and ocular dominance columns. Adapted from a diagram in Hubel and Livingstone [Livingstone and Hubel, 1988].

### 1.3.3 Orientation selectivity

After introducing the basic biology of the visual system prior to and including primary visual cortex, it is necessary to describe the facets most relevant to orientation perception in more detail. Prior to V1, none of the cells have orientation selective receptive fields but are circularly symmetric. The structure described below consists largely of cells which respond differently depending on the orientation of the presented stimulus.

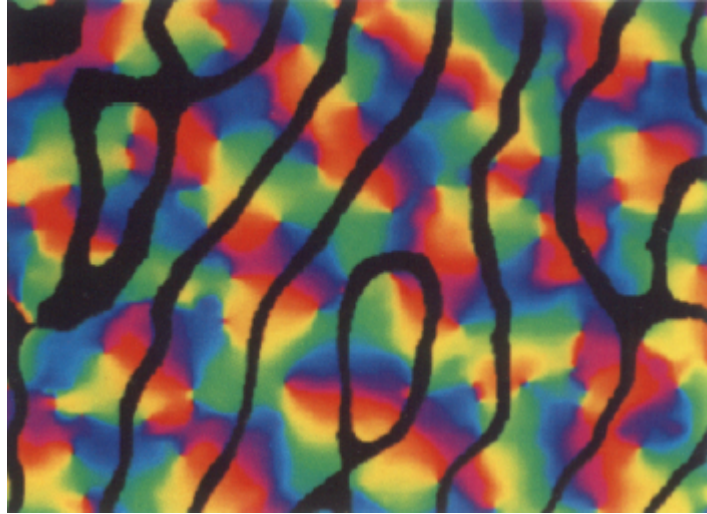


Figure 1.5: Blasdel introduced an optical technique for exploring orientation selective structures in macaque monkeys *in vivo*. The area of cortex under examination is stained with cytochrome-oxidase so that it darkens on cell metabolic activation. Gratings are presented monocularly and the optical changes are recorded. Image processing techniques are then used to produce coloured maps showing orientation preference across the cortex, according to which gratings produced darkening where. The example above masks the boundaries between ocular dominance columns in black, with orientation selective areas in between; each module is demarcated by two pinwheels, easily identified as having all colours around a singularity, which is coincident with cortical CO-blobs [Blasdel, 1992].

Early single cell investigations of V1 revealed a highly detailed anatomical structure thought to be based around straight lines, modules<sup>1</sup> and perpendicular intersections [Hubel and Wiesel, 1979], whereas the sophisticated optical techniques investigating activity, developed by Blasdel [Blasdel, 1992], have revealed a more complicated picture.

The primary visual cortex is retinotopic at one scale, as seen in Figure 1.3 (see page 15), but a smaller scale examination reveals a series of self-similar modules, as seen in Figure 1.4 (see page 16), each containing an intricate orientation-selective map Figure 1.5 (see page 17). Each map has within it and underneath it, running perpendicular down into the cortical surface, cells that respond preferentially to like orientations and directions [Blasdel, 1992, Zeki, 1993, Swindale, 1996, Tovée, 1996, Bruce *et al.*, 1996, Ferster and Miller, 2000], for example the simple cells described below. The distribution of these orientation selective areas is repetitive and has several notable characteristics. These include:

---

<sup>1</sup>areas distinct in function

- *blobs* when stained with a cytochrome-oxidase (CO) marker, two blobs are found in the centre of each module that are not selective for orientation.
- *ocular dominance columns* where one eye preferentially activates a local area, these are found paired in bands.

Various accounts exist of how the blob and column structures may have developed. They fall into two camps, that are not necessarily mutually exclusive. It has been proposed that the structure developed as a result of exposure to information in the optic array. Hebbian learning rules have been cited as a plausible mechanism by which these structures can spontaneously develop in order to efficiently extract information from the visual scene [Erwin *et al.*, 1995, Swindale, 1996]. The basic premise of a Hebbian rule is that if two neurons fire simultaneously, ie. they correlate, then their connection is strengthened.

More recently, it has been proposed that the finer distinctive features of these structures arise in order to minimise the axonal ‘wire’ length, giving rise to the distinctive orientation selective regions [Koulakov and Chklovskii, 2001] and the ocular dominance columns [Chklovskii, 2000].

It has been found that the minimum-sized bar required to indicate an orientation increases with visual field eccentricity [Zanker, 1998]. This is intuitively obvious as magnification decreases with eccentricity, thereby making more difficult the task of identifying oriented structure on the periphery.

### 1.3.4 Cortical receptive fields

As previously discussed, different layers of V1 contain cells that have different behaviours. These behaviours can be classified by examining their response to selective illumination. By illuminating precise and specific sections of the visual field that are slightly spatially displaced a map of a cell’s response can be built up. These are known as *receptive fields*. Receptive fields can have both spatial and temporal properties as some react transiently to illumination. Generally, the main receptive field types fall into the following categories; although nomenclature varies [Skottun *et al.*, 1991]. Hubel and Wiesel [Hubel and Wiesel, 1968] identified:

- *Centre-surround cells* are found in layer 4c, one of the termination points for the LGN; like those found in the retina and LGN, they have an inhibitory surround with an excitatory centre.
- *Simple cells* are defined as having two or more alternately excitatory and inhibitory elongated parallel lobes. These are found predominantly in layers 4 and 6, which receive direct inputs from the LGN. At minimum, the lobes have similar widths to the centres of

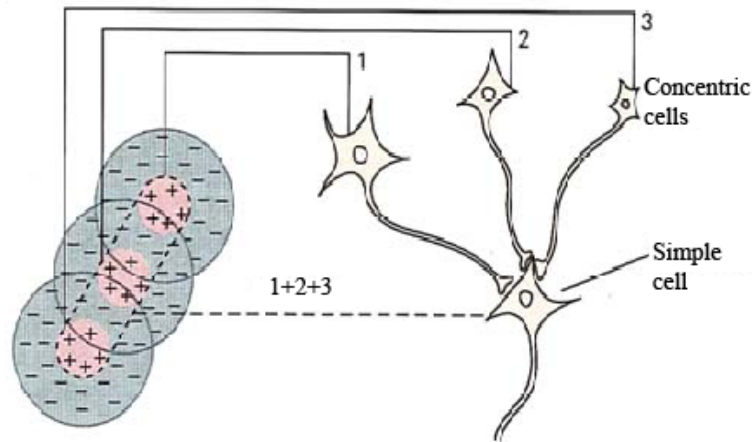


Figure 1.6: A schematic diagram showing a possible composition for the simple cell, originally proposed by Hubel and Wiesel [Hubel and Wiesel, 1962].

the centre-surround receptive fields found in the LGN and retinal ganglion cells, suggesting they synapse from them. The asymmetrical nature of the lobe arrangement means they respond preferentially to certain orientations of stimulus.

- *Complex cells* have been found predominantly in layers 2, 3 and 5. These cells do not have the distinct lobes of simple cells but still display the same orientation and direction sensitivity of simple cells.
- *Hyper-complex cells* still have orientation or direction selectivity but prefers a short bar ending within its RF - these are also termed 'end-stopped' cells. This end-stopped cell type can be classified as having two distinct sub-types; those with distinct lobes like simple cells, sometimes termed *Hypercomplex I* [Kato *et al.*, 1978]; and those with no distinct lobes, like complex cells, termed *Hypercomplex II*.

An interesting account of functional organisation can be found by DeAngelis *et al.*, where the results of many single cell recordings from cats are examined in concert to ascertain the main degrees of variation between cells in the visual cortex [DeAngelis *et al.*, 1999]. Models of orientation mechanisms need to account for the properties of these cells types; combinations of Gaussian and difference of offset Gaussian profiles are used below to account for these cell types.

After the initial discovery of these cells types within the primary visual cortex, sought to replicate the results. However, this was difficult without a more definite method of identifying cell types. Such a method, described in a review by Skottun *et al.* [Skottun *et al.*, 1991], involves using a drifting sine-wave stimulus within the RF of an unknown cell type. If the time domain

Fourier transform shows frequency components at the same frequency as the stimulus, e.g. that the cell fires in time with the passing stimulus or in time but phase shifted, then it is a simple cell type. However, if the cell first predominantly at even multiples of the stimulus spatio-temporal frequency then it is considered complex. These two classifications derive from the assumptions about linear summation over the area of the RF of the field types, but form a more uniform scientific regime for classification.

## 1.4 Cortical modelling

### 1.4.1 Acuity

It is important to have some understanding of the acuity of the visual system, and its contributing factors when trying to model low-level behaviour. In order to accurately reproduce the systems behaviour, a model must aim to relate its own internal scales with those of the visual system. At the lower end of that scale, the model should attempt to recreate the performance of the human visual system in the limits of recognising detail ie. the model should have the same acuity as the visual system.

As mentioned above, the visual system is composed of many stages, for modelling purposes, we are interested in the acuity of the visual system overall. This has been measured by presenting high contrast gratings to subjects. This process has yielded estimates for acuity in the region of 55 to 60cpd [Campbell and Green, 1965, Miller, 1979]. This measure implies a remarkable overall conservation of detail through the visual system, nearing the upper-limit set by the eye.

Acuity can also be tested in vernier stimulus experiments. Typically two similar sine wave stimuli are placed next to each other vertically and adjustably offset. For both luminance [Whitaker, 1993] and contrast defined stimuli [Volz and Zanker, 1996], it is found that the limit of being able to perceive an offset is smaller than the overall acuity limit for that stimulus. This phenomena is referred to as hyper-acuity and is attributed to neural sampling in each stimulus, where the effect of the offset is essentially amplified by the pooling.

### 1.4.2 Receptive field fitting

Marcelja suggested that Gabor functions can be used to model receptive field profiles [Marcelja, 1980] as they satisfy two conditions simultaneously. They are; the localization of detail in space and the simultaneous analysis of spatial frequency.

This means that a linear combination of different Gabor functions can represent a local patch of

space and simultaneously indicate the spatial frequency of the structure present. Marcelja goes on to show several examples of how receptive field profiles fit Gabor functions. Yet more extensive fitting of Gabor functions to simple and complex receptive fields was performed by Pollen and Ronner [Pollen and Ronner, 1983].

After an extensive study of cell properties in macaque monkeys De Valois et al. found that many striate cells have narrow spatial bandwidths. De Valois argued that these cells, considered as two-dimensional spatial filters critical for detailed spatial analysis [De Valois *et al.*, 1982] should be included into models of spatial vision.

De Valois went on to identify a narrowly tuned subgroup, having a spatial frequency bandwidth of 1.07 octaves and orientation tuning bandwidth of  $26^\circ$ , making up 31.5% of the striate cell population.

Koenderink proposed an alternative view of the function of receptive field profiles [Koenderink and van Doorn, 1987]. Instead of a local spatial frequency analysis via the Gabor function presented previously, an extensive ‘fuzzy-derivative’ framework was proposed. By viewing RFs as  $n^{th}$ -order differentials of Gaussians (DOGs), a *local jet* structure can be built up which is a local Taylor expansion. It was proposed that this bank of derivatives can then be utilised by different biological *programs* for different purposes.

Koenderink sets out a framework for the mathematical description of differently oriented simple cells by the simple steering of coefficients of the Taylor expansion, this is similar to that used in GROM and the motion model of Johnston et al. [Johnston *et al.*, 1992]. Indeed, this framework is the basis of GROM and of the corresponding multi-channel gradient model of motion (MCGM) [Johnston *et al.*, 1992]. Assuming RFs as DOGs gives direct access to gradient measures in the space-space and space-space-time projections of the plenoptic function.

Young et al. performed a series of studies into the fitting of RFs to Gaussian and Gaussian-derivative profiles [Young *et al.*, 2001]. The main conclusions of these were:

- Gaussian-derivatives accounted better than Gabor functions for the low- and high-frequency portion of spatial-frequency tuning
- Gaussian-derivatives account for the non-uniform spacing between zero-crossings found in biological data. Gabor patches have regular zero-crossings determined by their spatial frequency
- the Gabor model simplifies to the Gaussian-derivative model for a simple Gaussian and a two-lobed receptive field.

This equivalence can be demonstrated, in the one-dimensional case of the two-lobed simple-cell

CRF model, termed  $G_1$ , as:

$$G_1 = \frac{\partial G_0}{\partial x} = \frac{-x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}} \quad (1.1)$$

and correspondingly, where  $\lambda$  is the wavelength of the sinusoidal envelope, the Gabor function can be written:

$$Gb_1 = \sin\left(\frac{x}{\lambda}\right) G_0 = \sin\left(\frac{x}{\lambda}\right) e^{\frac{-x^2}{2\sigma^2}} \quad (1.2)$$

It can be seen by inspection that Equation 1.1 and Equation 1.2 can be equivalent where:

$$\sin\left(\frac{x}{\lambda}\right) \sim \frac{-x}{\sigma^2} \quad (1.3)$$

this occurs where  $\lambda$  is large compared (and negative) to sigma, as generally,  $\sin x \simeq x$  for small  $x$ .

Above all, the novelty of the work by Young et al. is the presentation of an explicit mechanism by which Gaussian derivative profiles can be synthesized biologically by finding the difference between locally offset Gaussian profiles, termed *difference of offset Gaussians* (DOOG). These offer a plausible biological mechanism for the synthesis of higher-order DOG profiles. This work is summarised in ‘Gaussian derivative synthesis’ (see page 151).

The differing choices of mathematical function for fitting to biological data represent two approaches; local spectral analysis and gradient analysis. The dispute will no doubt continue as fitting functions to biological data is unreliable. The difficulty is caused by the following factors:

- any individual cell can be an anomaly among a population
- the rate coding is based on an intrinsically random firing process

Which function is ultimately accepted as the best fit will arguably be determined by results of modelling based on each approach.

## 1.5 Motion perception and the second-order paradigm

Motion in a visual scene can be defined as the gradient of contours in a space-time plenoptic projection and is hence a very similar problem to measuring orientation in the space-space projection. It is also valuable to study motion modelling as ideas have been carried over to the

orientation domain

[Adelson and Bergen, 1985, Bergen and Landy, 1991, Landy and Bergen, 1991]. Debate in motion modelling led to the formulation of second-order stimuli [Chubb and Sperling, 1988] which has analogies in texture and orientation in the spatial plenoptic projection. Therefore, useful insight is gained by examining the development of motion models that aid understanding of the development of GROM.

In general, all models of motion require a non-linear combination of two spatially separated detectors [Borst and Egelhaaf, 1989]. This is a consequence of the time averaging required in motion detection. If two simple detectors, A and B are combined in a linear way to form a motion estimation C, then the time average of the inputs at A and B are the same as the time averaged output at C. To avoid this, a non-linear scheme of combination is required. All the models discussed below contain some form of non-linearity.

Reichardt [Reichardt, 1959, Reichardt, 1969] proposed a simple *correlator* method to explain an insect motion perception. This was the first in a class of correlation-type motion models that was eventually adapted to a human perceptual motion model by van Santen and Sperling [van Santen and Sperling, 1985].

Such a correlator is typically constructed ‘detectors’ displaced by a distance,  $\Delta$  that are connected together with a temporal low-pass filter, which approximates to a temporal delay,  $\epsilon$  [Reichardt, 1986]. If the detectors fire separated in time by  $\epsilon$ , it is likely that an image of the same object has moved between the detectors at the speed defined by  $\frac{\Delta}{\epsilon}$ . This makes each detector highly tuned to a certain velocity (speed and direction), thereby requiring many detectors to account for the range of perceivable velocities, and being potentially corruptable if the image contains spatial frequencies higher than the spatial displacement of the detectors.

Barlow and Levick introduced a motion model [Barlow and Levick, 1965], similar to the correlator, with simple detectors of illumination at two discrete locations plus an inhibitory non-linear combination of the two inputs, rather than a facilitatory combination as in the correlator.

The details of detector and temporal delay stages vary in complexity throughout the range of correlator type models. A good review can be found by Borst and Egelhaaf [Borst and Egelhaaf, 1989].

The simplicity of the design does not allow for various facets of human visual perception, chiefly contrast sensitivity. By combining output from simple filters with a time-delay, the signal is dependant on the contrast of the underlying texture.

Adelson and Bergen proposed a more sophisticated approach to explain human motion perception, based around the concept of *motion energy* [Adelson and Bergen, 1985], see Figure



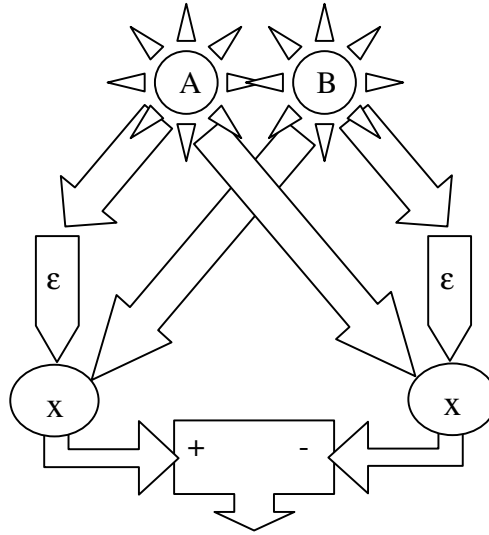


Figure 1.7: A schematic showing the layout of a Reichardt correlator [Reichardt, 1959, Reichardt, 1969]. Input signals arrive at two ‘detectors’, A and B, a fixed distance apart. Each side is in multiplicative combination with a temporal low-pass filter that introduces an effective delay,  $\epsilon$ , the result is the difference of the two sides. A light signal moving from right to left will produce a positive signal, left to right, a negative signal.

1.8 (see page 24). The output of spatial and temporal filters is combined to form spatio-temporal filters oriented in space-time so as to detect contours oriented in that projection.

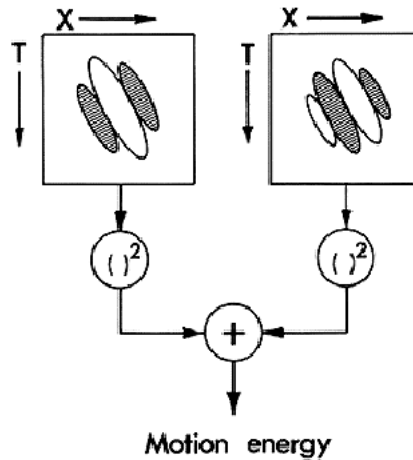


Figure 1.8: A diagram, adapted from Adelson and Bergen [Adelson and Bergen, 1985], showing the combination of two spatio-temporal filters added in quadrature. These form sub-units in the overall Motion Energy model presented by Adelson and Bergen.

Motion energy is defined as: the constructed, same oriented, two and three lobe spatio-temporal filters added in quadrature. The model uses filters constructed with sine and cosine Gabor functions filtered with two different temporal kernels. Where varying the number of spatial lobes makes the code invariant to the phase of the structure passing beneath the spatially coincident filters.

This generates a filter set, argued as physiologically plausible by Adelson and Bergen, of four spatio-temporal filters. In their description these are named, A, A', B and B', where A and B denote the sine/cosine Gabor spatial phase and the prime denotes the type of temporal filter used.

A measure of opponent energy is deduced by forming four additive combinations of these filter outputs, A-B', A'+B, A+B' and A'-B, named 'Oriented linear responses'. These are added together in quadrature to form two measures of 'Oriented energy',  $(A - B')^2 + (A' + B)^2$  and  $(A + B')^2 + (A' - B)^2$ . The two oriented energies are then added to form a measure of *opponent energy*,  $4(AB' - A'B)$ .

In order to perceive all directions of motion at a range of spatial and temporal frequencies, many motion energy detectors must be employed. Taking a very convenient mathematical tool, and an obvious analogy from the study of auditory perception, motion perception is often thought of as being an attempt to perform a Fourier analysis of spatio-temporal space [Watson and Ahumada, 1985, Heeger, 1987]. Motion models employing such a strategy have become known as *Fourier Energy Models* (FEMs). Stimuli defined by movement of smooth variations of luminance, luminance-defined stimuli, have distinct Fourier power-spectra oriented about the origin of Fourier space, and are readily distinguishable with FEMs. These are commonly termed *first-order stimuli*. Stimuli defined by the variation of a property other than luminance, typically contrast-defined stimuli, are known as *second-order stimuli*.

Chubb and Sperling explored the psychophysical effect of moving contrast-modulated stimuli and defined how they should be constructed in order to be invisible to conventional FEMs [Chubb and Sperling, 1988]. This was important to the development of motion models as it led to a second motion channel being proposed. This additional channel rectifies the initial spatial signal.

The effect of this is to transform what is measurable at the luminance extremes of the stimulus - lightest and darkest sections become the same. Maximum positive and minimum negative extremes of any contrast-defined stimulus would be set to maximum in the rectified output. The values around the mean, or zero (depending on the scale) become a minimum, and therefore, any structure at the mean luminance values is detectable by another level of ordinary spatial filtering.

In terms of the Fourier space distribution of such a stimulus, this makes the velocity of a contrast-modulation, normally displaced from the origin, become centred at the origin. FEMs

readily detect structure that is origin-centred in spatio-temporal Fourier-space, therefore, a another FEM is applied on top of the filter outputs of the rectifying stage.

Zhou and Baker [Zhou and Baker Jr., 1993] made single-cell recordings in cat cortices named areas 17 and 18. They concluded that the distinct differences in the spatial tuning curves in response to the presentation of second-order stimuli implied the existence of two distinct populations, which they proposed was evidence of the two functional channels defined above.

Various psychophysical stimuli have been suggested that reportedly support separate channel motion processing. The following example stimuli were chosen in order give a reasonable coverage of second-order stimuli and also both gradient and non-gradient models have made a reasonable attempt at explaining the phenomena observed during psychophysical investigation of the stimuli.

- Ledgeway and Smith [Ledgeway and Smith, 1994] examined *interleaved stimuli*, where a psychophysical stimulus is constructed of one first-order and one second-order stimulus with the same spatial frequency, in alternate frames with a variable delay between. The phase of both types of stimulus was adjusted in order to create the sensation of motion if the two stimuli were processed in a coherent integrated fashion. Interleaving two different second-order stimuli created the sensation of motion, whereas first-order interleaved with second-order stimuli did not. This has been proposed as further evidence that the two types of stimulus are processed separately.
- Hammett et al. [Hammett *et al.*, 1993] proposed a stimulus containing summed sine waves with relative frequencies  $3f$  and  $4f$ . The result was a contrast modulation travelling in one direction and a luminance ‘beat’ travelling in the opposite direction. Psychophysical performance seemed to suggest a feature-based model of motion, as both velocities, forward and back, could be perceived in an apparently transparent affect.
- Zanker [Zanker, 1993] coined the term *theta motion* for a stimulus whose motion was defined by the movement of boundaries of subsections, where each subsection consists of random-dot kinematograms. Movement of the boundary is clearly perceived yet the motion is non-Fourier.

Lu and Sperling experimented using complex stereoscopic stimuli with variation defined in depth and texture; both second-order stimulus attributes [Lu and Sperling, 1995a]. Subjects were trained to attend to certain ‘salient’ features and report direction change depending on the feature being attended to. They argue that as first- and second-order channels were defined as monocular, stereoscopic perception of motion necessitates the existence of the third channel responsible for velocity computation after the selection of a salient feature.

Johnston et al. developed a unified gradient model, the *multi-channel gradient model* (MCGM)

to explain motion perception that does not require a separate mechanism to detect second-order stimuli [Johnston *et al.*, 1992]. The two-dimensional version of this model [Johnston *et al.*, 1999] is able account for results from the interleaved stimuli of Ledgeway and Smith [Ledgeway and Smith, 1994].

Benton argues that the shift in balance between forwards and reversed motion reflects the operation of a transient temporal filter in human motion perception, rather than a switch in computational strategy [Benton *et al.*, 2000]. In relation to the beat pattern stimuli of Hammett *et al.* [Hammett *et al.*, 1993] Benton explains that a percept of transparency is obtainable with the MCGM, with both forwards and reversed motion present in the velocity field output [Benton *et al.*, 2000].

The MCGM gives accounts of other illusions [Johnston and Clifford, 1995] and responds well to second-order stimuli such as contrast-modulated gratings and noise. Johnston *et al.* [Johnston *et al.*, 1999] discussed the similarities between FEMs and the MCGM, originally highlighted by Bruce *et al.* [Bruce *et al.*, 1996], and showed their cut-down model producing FEM-like results for contrast modulated noise. The MCGM also predicted the existence of induced motion, where a static random dot background is seen to move as a contrast envelope moves over it [Johnston *et al.*, 1999].

Using *gradient-plots*, Benton showed that there is sufficient information for gradient based motion analysis to be able to extract ‘non-Fourier’ motion from the *sampled phase shift* stimuli of Taub [Taub *et al.*, 1997, Benton, 2002]. Gradient-plots are found by performing convolutions with a temporal DOG ( $G_1(t)$ ) and spatial DOG ( $G_1(x)$ ) filter in the direction of motion. A distribution of speed can be found by plotting the each pixel’s response to these filters on orthogonal axes. The results showed clear peaks that corresponded to the speed of the second-order contrast-modulation in the stimuli. This work showed that second-order information is available in the most simple gradient based model.

The proliferation of additional channels to account for motion phenomena in FEMs creates the need for an increasingly complex integration of channels to produce a single coherent motion percept. Furthermore, as channels have only been identified using bespoke lab-based stimuli, it is arguable that they have no evolutionary reason to exist.

For reasons of parsimony it would be preferable to explain motion phenomena via a single unified model; FEM based approaches have fail to do this. This failure suggests the possibility that a different approach might be the basis of a more parsimonious model. As a possible alternative, the MCGM could arguably be labelled complex in it’s construction. In the absence of a rigorous framework to compare the complexity of the FEMs and the MCGM, the number of free parameters used in each could be examined as an alternative measure. Details of the implementation of FEMs rarely include the total number of input parameters, examples from

those that do are; the model presented by Perrone with 9 parameters [Perrone, 2004] and Wilson et al. with  $\sim 50$  in [Wilson *et al.*, 1993]. By comparison, the MCGM of Johnston et al. has 6, each based on measurements from the biological system, rather than chosen to fit experimental data.

Considering the unified explanation of existing first- and second-order motion stimuli, the small number of model parameters and the prediction and verification of new stimuli, the model of Johnston et al. warrants further examination and represents a good basis for model development in the spatial orientation domain.

## 1.6 Spatial orientation

### 1.6.1 Psychophysical orientation measures

In the section entitled ‘Motion perception and the second-order paradigm’ (see page 22) the development of motion processing models was covered. Motion is related to measures of static spatial orientation via the plenoptic function. To measure orientation in a space-space-time plenoptic projection is to measure motion; to measure orientation in a space-space plenoptic projection is to measure static *spatial orientation*. Inevitably approaches to the two problems are similar; seminal work in the motion field [Adelson and Bergen, 1985] has been adapted, and forms the basis of early models of how human spatial orientation is computed [Bergen and Landy, 1991, Landy and Bergen, 1991].

These same early models for spatial orientation are unable to account for perception of second-order textures. Texture with an average mean luminance whose variation was defined by changes in contrast, orientation or some other higher level statistic. Consequently models for spatial orientation have developed (in parallel with their motion counterparts) separate channels to account for unexplainable ‘second-order’ phenomena, which involve the variation of a stimulus attribute other than luminance [Lu and Sperling, 1995c, Smith *et al.*, 2001, Atherton, 2002].

An important area where orientation can be studied macroscopically, is *textures*; where groups of elements or non-local regions of structure form percepts based on their statistics. Textures consist of elements with varying orientation [Nothdurft, 1985a], but also spatial frequency, contrast, luminance, etc. Texture based investigation is useful for examining grouping behaviour and the non-local statistical rules that may govern spatial perception [Julesz, 1984, Nothdurft, 1985a], including those which pertain to orientation.

Spatial orientation mechanisms can also be probed with more simple *grating* stimuli, often having variation in luminance (first-order) and/or contrast (second-order). Gratings and other more

simple stimuli aim to uncover the more elemental rules relating to specific local computations. For example, in early psychophysical work by Blakemore et al. [Blakemore *et al.*, 1970] it was found that the perceived angle between straight lines joined at one end appeared more acutely angled as they neared parallel, at both the acute and obtuse extremes. They proposed a ‘lateral inhibitory’ interaction in the orientation domain to explain their findings.

They argued that, lateral inhibition, in the case of orientation selectivity of CRFs, occurs where neighbouring cells with different orientation preference interact with each other when stimulated. For each cell, structure oriented at its preferred orientation makes it fire maximally, whereas, structure markedly different from its preferred orientation makes it inhibit the reaction of neighbouring cells [Blakemore *et al.*, 1970]. Where inhibitory and excitatory reactions sum linearly, this effectively accentuates the reaction to a single line. Additionally, they argue, this increases the perceived subtended angle between coincident similarly oriented lines - a ‘repulsion effect’.

### 1.6.2 Texture based investigation

By using textures constructed of oriented finite lines, Nothdurft explored texture segmentation neurophysiologically [Nothdurft, 1985c] and psychophysically [Nothdurft, 1985a, Nothdurft, 1985b], see Figure 1.9 (see page 29). To explain his findings, he proposed that texture segmentation is performed using a *structure gradient*,  $\frac{\Delta\theta}{\Delta x}$ , as segmentation is aided by a greater orientation difference,  $\Delta\theta$ , and inhibited by greater inter-element distance,  $\Delta x$ . Landy and Bergen [Landy and Bergen, 1991] investigated and supported the structure gradient hypothesis using stimuli constructed from the oriented filtering of noise.

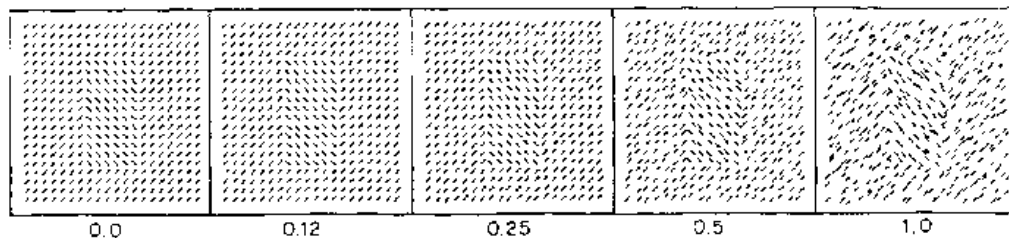


Figure 1.9: Several examples of Nothdurft’s texture segmentation stimuli. Left to right, the variation of the individual line segments position increases, whilst the clarity of the segmenting central section decreases. Adapted from Nothdurft [Nothdurft, 1985a]

### 1.6.3 Second-order spatial stimuli

Dakin et al. psychophysically explored the interaction between carrier and envelope for Gabor patches and attempted to explain the results using variations of two-channel orientation models [Dakin *et al.*, 1999]. Despite significant variations of fundamental filtering characteristics and their subsequent combination, Dakin et al. failed to account for various attractive and repulsive interactions between the envelope and carrier orientations. It was generally noted that two channel Orientation FEMs, also referred to as *filter-rectify-filter* (FRF) models, accurately model the veridical orientation within a stimulus but fail to account for illusory effects within the psychophysical results, see Figure 1.10 (see page 30).

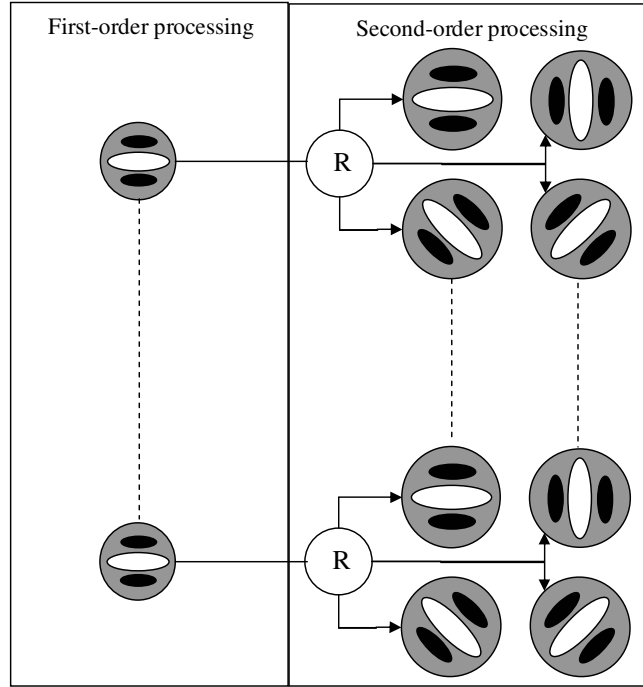


Figure 1.10: Adapted from Dakin et al., this figure shows the two separate processing streams of an FRF model. The ‘R’ inside the second-order sub-box means ‘rectification’ [Dakin *et al.*, 1999].

Smith et al. presented a modification of the FRF model to account for measured repulsion and attraction effects in orientation tasks for first- and second-order stimuli [Smith *et al.*, 2001]. The experiments were performed with a centre and surround stimulus and judgments were made on the orientation of the centre CM or LM grating, while an outer annular surrounding CM or LM grating was rotated between trials. The modification entailed centring and scaling operations being performed on the outputs of the filter populations within each channel to account for data from grating psychophysics. Coefficients for the two fitting operations were varied to account for results from different stimuli.

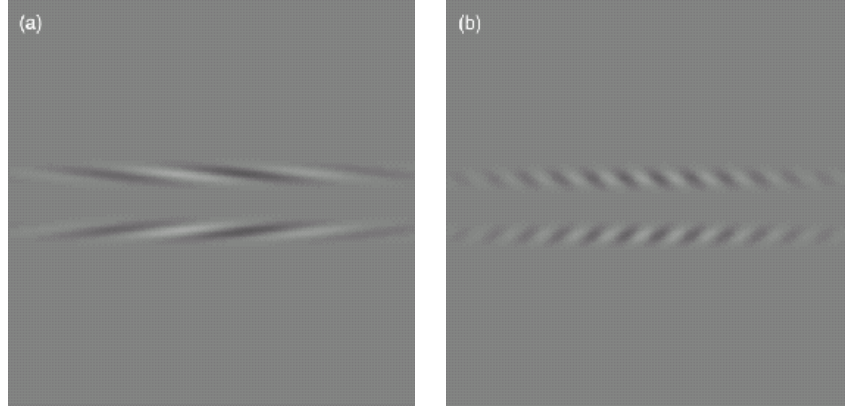


Figure 1.11: An example of two twisted-chord stimuli taken from Skillen et al. [Skillen *et al.*, 2002]. Upper and lower chords in both are horizontal — they have horizontal envelopes. (a) shows the Fraser effect where the angle between carrier and envelope appears less (more acute) than it is, causing an apparent attraction between the two cords at the right hand side, as the carriers have opposite sense. (b) shows the Zöllner effect where the angle between carrier and envelope appears more (more obtuse) than it is, causing a repulsion between the two cords at the right hand side, as the carriers have opposite sense.

Skillen et al. [Skillen *et al.*, 2002] investigated the interaction between carrier and envelope spatial scales and relative orientation in the twisted-cord illusion Figure 1.11 (see page 31). Despite referring to two-channel FRF models they found significant evidence of interaction between the first- and second-order information of the stimulus. Twisted-cord displays attraction, termed *Fraser effect*, between the envelope (held horizontal) and carrier orientations, making the subtended angle appear less (more acute) than it is, causing apparent attraction between the two cords at the right hand side. At larger subtended angles between carrier and envelope, that effect becomes repulsive, termed *Zöllner effect*, meaning the subtended angle appears more (more obtuse) than it is, causing an apparent repulsion between the two cords at the right hand side. The extent of these effects was found to be dependant on the relative, rather than absolute, scales between the carrier and envelope.

Adaptation and threshold experiments have been used to indicate the existence of a separate channel for processing second-order texture phenomena. Broadly speaking, adaptation experiments aim to fatigue cells responsible for perceiving specific aspects of a stimulus, for example direction of movement, spatial frequency or spatial orientation. After prolonged exposure to fatiguing stimuli, perception of similar stimuli is impeded. This impediment makes it more difficult to perceive the same or like stimuli, therefore, the thresholds for just detecting such stimuli is increased. It is extrapolated that minima and maxima in these thresholds for detection indicate the presence of dissimilar or similar stimuli respectively.



Langley et al. performed adaptation experiments where the threshold of detection was measured for the envelope angle of a contrast-modulated (CM) grating after the subject had adapted to a luminance modulated (LM) grating of a particular orientation and spatial frequency [Langley *et al.*, 1996]. Results showed that adaptation to LM gratings most like the first-order information of the subsequent CM grating increased the threshold of (75%) correct envelope angle estimation. A preceding LM grating with orientation and spatial frequency characteristics most like the envelope of the CM grating had little effect on the threshold of correct envelope estimation.

It was argued that this implied that the filtering stage precedes non-linear processing. Spatial filtering is thought to be the stage where adaptation occurs [Blakemore and Campbell, 1969], as it is thought to be caused by the fatiguing of individual RFs, therefore, if a rectification occurred before filtering and divided into two channels of first- and second-order information, there should be no interaction between the initial adaptive first-order LM grating and the perception of the CM grating's second-order envelope angle.

Langley et al. performed a further experiment, without the initial adaptive grating, to determine the threshold of (75%) correct discrimination of envelope orientation for differing carrier and envelope spatial frequencies. Results showed a distinct minima for the correct discrimination. It is argued that this is contrary evidence to the theory that orientation is measured using local spatial changes in contrast. As one would expect that thresholds for second-order envelope detection should decrease (ie. they get easier to detect) as the carrier frequency increases. This is because, as the carrier frequency increases, so does the local gradient which would yield larger local contrast variations making the envelope easier to detect.

It should be noted that, as found in psychophysical experiments on CM gratings by McOwan and Johnston [McOwan and Johnston, 1996], if a carrier is oriented at  $45^\circ$  relative to an envelope, the envelope orientation is misperceived. This may have affected judgment for the discrimination tasks used by Langley et al., where the relative angle between carrier and envelope was also  $45^\circ$ . Compounding this, the frequency of both envelope and carrier are similar to those where McOwan and Johnston found maximal misperception.

Schofield and Georgeson also proposed that LM (first-order) and CM (second-order) gratings do not facilitate each other's detection at threshold and proposed a two channel mechanism to explain their results [Schofield and Georgeson, 1999]. Similarly, Kingdom et al. [Kingdom *et al.*, 2003] failed to find sub-threshold facilitation between different types of second-order textures, they were orientation, spatial-frequency and contrast modulated. Kingdom argues this could suggest a separate mechanism detects each stimulus type.

However, a common thread to each of these findings is the underlying assumption that threshold facilitation should work between first- and second-order stimuli, or indeed between more diverse

stimuli types, though there is intermodal evidence suggesting this. For example, Stein et al. showed the facilitation of visual stimuli with accompanying audio stimuli [Stein *et al.*, 1996]. If the threshold-facilitation experimental paradigm works on high-level, holistic ‘channels’, then the paradigm is likely valid. The existence of discrete channels could arguably account for the lack of facilitatory interaction between first- and second-order stimuli. However, if threshold-facilitation is actually a consequence of low-level fatigue of individual receptive fields, then there is no reason to suppose the paradigm valid between diverse stimulus types as local information changes between stimuli.

Spatial orientation experiments have been examined in this section. With the exception of McOwan and Johnston’s [McOwan and Johnston, 1996] orientation task, these experiments have sought to find evidence of a two-channel mechanism for determining orientation of carrier and modulation separately. Some of the evidence presented in this thesis will show that it is not necessary to use separate mechanisms to measure orientation of first- and second-order stimulus attributes. Indeed, being able to model local orientation perception with one unified mechanism may mean that the ‘2<sup>nd</sup>-order’ nomenclature is unnecessary.

#### 1.6.4 Population coding

There has been much investigation into the nature of the coding of information in the brain. An early investigator, Adrian [de Charms and Zador, 2000], proposed that information, transmitted along axons, was encoded with the rate of the firing, termed a *rate code*. This is an attractive proposition since, in visual science, if a spatial-receptive field represents a parameter locally encoding a stimulus, then the mean firing rate indicates the amount of that parameter present at that point. How information can be encoded and decoded using this hypothesis is intuitively obvious. All the models for the measure of orientation presented above rely on the rate coding assumption.

There is an alternate view, termed *temporal coding*, that information is encoded in the timing of the individual spikes. This hypothesis seems to be plausible in relation to the encoding of non-static stimuli [Dayan and Abbot, 2001]. Supporting this hypothesis experimentally, Buracas et al. [Buracas *et al.*, 1998, de Charms and Zador, 2000] showed that the neurons in the middle temporal region of the macaque spike in time with the spatio-temporal structure of a presented stimulus.

These apparently conflicting views can be seen as being at opposite ends of a graduated scale. Decreasing the time window over which the mean rate is calculated produces a coarse *rate code* at one extreme, through to an individual spike-time graph at the other, which is equivalent to a temporal code.

The orientation-tuning of individual simple cells in V1 is far broader than the *just noticeable difference* (JND) between two oriented lines [Orban, 1984, Swindale, 1996]. This also applies to the control of movement, where precision in movement is far better than can be accounted for given the width of tuning of cells in monkey motor cortices [Georgeopoulos *et al.*, 1986]. To attempt an explanation, Vogels utilised the same population coding approach used in motor control to model spatial orientation [Vogels, 1990].

In the Vogels model, each individual cell represents a vector, the preferred orientation of the cell represents direction and the activity of the cell represents the magnitude. The vector sum of a population of cells is its resultant orientation; this is known as the *population vector*. The results of a simulation showed that the overall deviation in the final orientation measure was of the same magnitude as the JND found in psychophysical discrimination tasks. The most noticeable flaw of the model is its inability to account for the robustness of the psychophysical JNDs to contrast changes [Skottun *et al.*, 1987].

Changes in contrast cause a change in the response of the entire population, the response of the population in low-contrast situations make the orientation harder to discriminate above the noise level — the population vector becomes unreliable. This can be overcome by increasing the population of cells, thus making the number of cells at the preferred orientation significant above the noise, making the population vector result more clear. However, the size of the population required to replicate the psychophysically measured JNDs is greater than neurophysiological investigations suggest.

Extending the model used by Vogels [Vogels, 1990], Zohary proposed a population code to model spatial orientation and frequency selectivity [Zohary, 1992]. An implementation of the model conformed well to JNDs from orientation and spatial frequency discrimination tasks in humans. The size of the population required to mimic human performance was found to be  $10^3 - 10^4$  neurons.

Zohary generalised the model beyond selectivity to spatial orientation and frequency — two dimensions in the multidimensional space of possible stimulus attributes. Cells found in neurophysiological investigation can be selective over many stimulus attributes (dimensions) such as colour, form, stereo and motion processing. However, Zohary pointed out that the maximum number of dimensions which such a code can plausibly model is three to five, given the limitation on the number of cells in a hypercolumn and the requirement of 1% selectivity in each dimension.

In the case of motion, the population of speed sensitive cells found in neurophysiological investigation of V5-MT was found skewed towards speeds similar to those found in human experience, e.g. walking speed [Cheng *et al.*, 1994]. This would make any vector based method of reading the code difficult as they rely on a uniform response to all possible rates of the quality they measure [Johnston *et al.*, 1992].

Gilbert and Wiesel investigated the effect of ‘contextual stimuli’ on neuronal responses in cat striate cortex [Gilbert and Wiesel, 1990]. After selecting a cell and mapping out its receptive field characteristics, they measured the orientation tuning with and without surrounding oriented bars. It was found that the presence of the bars affected the width and centre of orientation tuning of the selected cells. It was proposed that this result represented direct evidence of the effect of horizontal connections in striate cortex. Gilbert and Wiesel attempted a local model to explain this result, using a population of coincident differently oriented cells. The height of peaks in the population’s orientation responses were reduced to simulate the presence of nearby similarly oriented structure. The population’s overall resulting orientation is similar to the vector sum scheme presented by Vogels [Vogels, 1990]. Possible explanations for various attractive and repulsive interactions between oriented structure were presented, although the results of the experiments, didn’t necessarily require this model to explain them as they were based upon a small sample of cells.

The current debate centres on the nature of the mechanism which reads a population code [Pouget *et al.*, 2000]. Populations vary in size, noise characteristics and tuning, and each of these facets have an impact on any final reading of the code. There are a variety of ways to read the final result implied by a code’s response to stimulus [Seung and Sompolinsky, 1993, Deneve *et al.*, 1999], including the aforementioned population vector [Deneve *et al.*, 1999], and a *maximum likelihood* (ML) estimator. ML estimators require extensive knowledge of the statistics of neurons in the population and have been shown to be the optimum solution for reading large populations [Seung and Sompolinsky, 1993]. Seung and Sompolinsky implemented ML estimators based on neural networks and found extensive re-training was required for accurate results on new stimuli, ultimately questioning their biological plausibility.

More recently, Deneve et al. have implemented a ‘recurrent network of non-linear neurons’ which they claim comes close to implementing ML. Furthermore, they argue that despite extensive use of parameter fitting, their model is biologically plausible given that only 2-3 training iterations are required to reach a steady solution.

### 1.6.5 Mathematical orientation models

Atherton proposed a framework for extracting energy-based and phase-based orientation information [Atherton, 2002]. This is built on an energy-based method of measuring orientation in machine vision [Langely and Atherton, 1991], which in turn relied on the texture segmentation approach of Landy and Bergen [Landy and Bergen, 1991]. Energy models of orientation perform local Fourier analysis and therefore correctly measure the orientation of sinusoidal gratings, yet one of their failings is that they are unable to correctly account for the position of edges within

gratings [Georgeson and Freeman, 1997]. When prompted, human subjects highlight the step between light and dark in a sinusoidal grating as an edge, however this does not correspond to a minima or maxima of energy.

Atherton's proposed extension to this simple energy-based orientation model affords sensitivity to both bars and edges, the components of a sinusoidal grating. However, the extension is not unified as a final measure of orientation, rather it is divided into  $n$  separate channels, each of which examines the orientation of  $n$ -fold symmetric structure.

How many times can a feature be rotated in  $360^\circ$  and still look the same? — this is  $n$ -fold symmetry. For example, a light to dark edge, can only be rotated by  $360^\circ$  to look the same and is therefore one fold symmetric. A bar, however, is the same if it is rotated by  $180^\circ$  or  $360^\circ$  and is therefore two-fold symmetric. This point is illustrated in Figure 1.12 (see page 36).

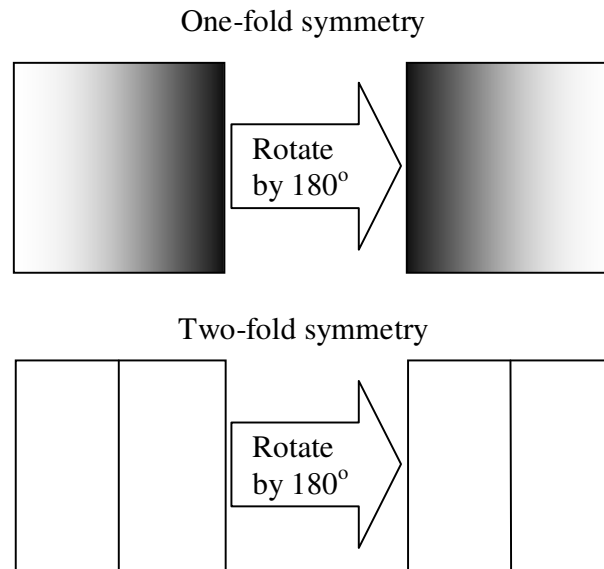


Figure 1.12: This figure demonstrates the nomenclature of  $n$ -fold symmetry. In the upper case it is shown that a light to dark edge looks different when rotated through  $180^\circ$ , in fact it only looks the same again when rotated through  $360^\circ$ . In the lower case, a bar looks the same when rotated through  $180^\circ$  - making it two-fold symmetric.

The orientation within these separate mechanisms is expressed as a phase shift of the Fourier transform, relative to an arbitrary reference frame. The evidence for multiple-fold symmetric orientation mechanisms is called into question by the findings of Wolfe et al. [Wolfe *et al.*, 1999], who showed that pre-attentive orientation segmentation is two-fold symmetric. This is in keeping with neurophysiological investigations which have revealed orientation selectivity spans  $180^\circ$  [Blasdel, 1992, Swindale, 1998].

Atherton also discusses how the energy-based mechanism can be used to extract ‘envelope’ orientation, as it is rectified, and the phase based mechanism can be used to extract ‘carrier’ orientation due to it’s sensitivity to phase. Unfortunately no explicit evidence is presented of the model’s ability to detect the orientation of second-order structure.

### 1.6.6 Computer vision approaches to measuring orientation

Estimating spatial orientation without the restrictions of biological plausibility would be much easier. Spatial orientation gives access to the most basic structures of a scene and is therefore an important pre-processing step in computer vision. Local orientation gives access to local shape, which if in a continuous contour, most likely indicates the presence of an object. Through this logic, a scene description can be attempted.

Among the earliest and most influential works on spatial orientation in ‘computer vision’ are those of Bigün and Granlund [Bigün and Granlund, 1987]. The approach is described as least-squares fitting of principal axes in the Fourier-domain. The local spatial Fourier-domain contains a description of the Fourier components that includes orientation information. This is similar to Fourier-energy models that attempt to find alignment within the Fourier-domain, the principal difference is the exact technique used to deduce the principal components (or axes) within the Fourier-domain description of the local spatial patch being analysed. In this case, the fitting is achieved by finding Eigenvectors of an ‘inertia-matrix’ that is defined by the following elements:

$$J_{ii} = \frac{1}{4\pi^2} \sum_{j \neq i} \int_{E_n} \left( \frac{\partial f}{\partial x'_j} \right)^2 dE_n \quad (1.4)$$

and for the diagonal elements:

$$- J_{ij} = -\frac{1}{4\pi^2} \int_{E_n} \frac{\partial f}{\partial x'_i} \frac{\partial f}{\partial x'_j} dE_n \quad (1.5)$$

where  $f$  is a function describing the image,  $x'_j$  is the Fourier-domain coordinate associated with the spatial dimension,  $x_j$ ,  $dE_n = dx'_1 dx'_2 \dots dx'_n$  and  $n$  is the number of spatial dimensions.

The final measure of orientation,  $\phi$  is given by:

$$2\phi = \arctan(J_{22} - J_{11}, 2J_{12}) \quad (1.6)$$

Although the earlier presentation did generalise to an  $n$ -dimensional case, the original work of Bigün and Granlund didn't consider the three-dimensional case of optic flow (space-space-time). Later work by Bigün extended the two-dimensional case to measures of optic flow [Bigün, 1991].

More recently, van den Boomgaard and van de Weijer presented a technique that applies a robust least-squares approach on top of a *structure tensor* (a covariance matrix of gradient vectors) [van den Boomgaard and van de Weijer, 2002].

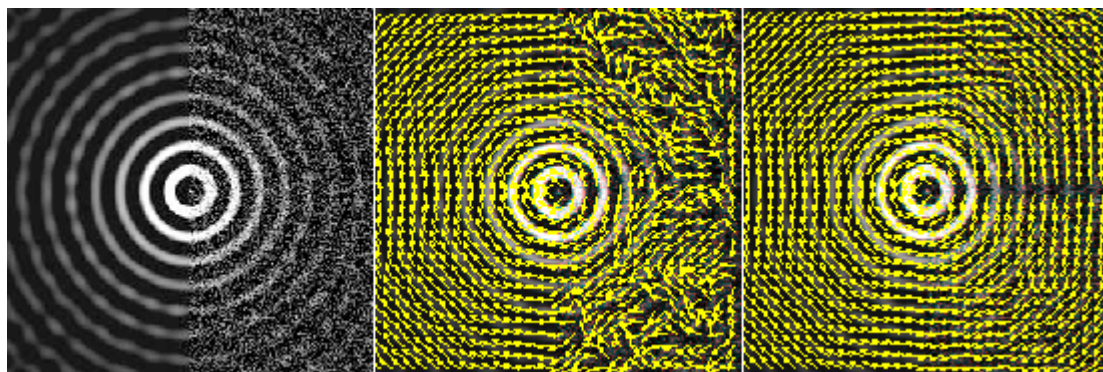


Figure 1.13: Adapted from a paper by Feng and Milanfar [Feng and Milanfar, 2002]. A circular stimulus (left), the right half of which has added noise, with the corresponding orientation results, at a single scale (middle), and using the multiscale version of their method (right). Yellow arrows indicate orientation. The increased robustness to noise of the multiple- over the single-scale result is clearly noticeable.

Another new technique presented by Feng and Milanfar combines multiple techniques for accuracy and robustness to noise [Feng and Milanfar, 2002], see Figure 1.13 (see page 38). Local to any one pixel, gradient vectors are found on many scales using *Kalman filtering*; a covariance matrix is formed and Eigenvectors are deduced using *Principal components analysis* (PCA). PCA is a method of extracting the vectors from a space that contain the most variation in that space — the principal components. The operation is equivalent to finding the Eigenvectors from a covariance matrix. Each element in a covariance matrix is an expression of the variance between two variables for all the measurements of those variables. The process of finding Eigenvectors from large matrices can be computationally expensive, in the Feng and Milanfar method, they use Single Value Decomposition to find these in the optimal time [Deprettere, 1988].

The Eigenvectors of the covariance matrix, the principal components, each have an Eigenvalue associated with them. This value is a measure of the amount of variance that is contained in the direction of the vector. In the Feng and Milanfar method, this gives a way of ranking the certainty of each orientation measure over different scales, essentially comparing local orientation over many scales. They argue that this optimally solves the problem of deducing local orientation

of smooth structure while remaining robust to noise.

## 1.7 Modelling

### 1.7.1 Basic modelling paradigm

There are limitations to each of the modern techniques of physical measurement of activity in the visual cortex. If visual science is to truly understand the human visual system, or the problem of how to progress from a measure of visual information to visual awareness, then more techniques are required than biological subject measurement.

Marr [Marr, 1982] formalised a modelling approach to understanding the human visual system by stating the problem as an information-processing task. He argued that the visual system could be seen as converting physical measurement of the surroundings via an algorithm into understanding the visual scene. To wholly understand the visual process, or part thereof, three separate levels of understanding must be attained:

- *computational theory* — to wholly understand what the goal of the computation is, this sets up the framework in which the elements, the algorithms, might be understood.
- *representation and algorithm* — essentially the design phase, to specifying exactly how the computational theory, set out above, is implemented. This stage specifies the nature of the input, output and transformation process between the two. In the case of understanding the human visual system, useful approximations for the input can easily be made with analogy to the modern camera and pixel array; specifying the output is more difficult.
- *hardware implementation* — to specifying how the above design can be physically implemented. In the case of human vision, this means understanding which parts, i.e. visual areas, receptive field types, etc. are doing which transformations, and how.

To better understand the human visual process we can undertake psychophysical and neurophysiological studies of primates, and from the results we can then formulate explicit mathematical models for the visual cortex. The computational model discussed below provides a framework that permits the biologist to attribute mathematical functionality to neurophysiological structures, and gives the psychologist an understanding of the processes underpinning a subject's psychophysical performance. Robotics researchers gain an algorithm for a perceptual process that can be implemented in suitable hardware.

Production of an explicit mathematical model also generates predictions that can subsequently be tested experimentally. Feedback may then be exploited to modify the model to better



approximate the experimental neurophysiological and psychophysical data. This incremental modelling approach should converge on a realistic and biologically plausible solution. Moreover, study of a fundamental process such as orientation measures, inevitably leads to questions about the nature of higher visual areas such as grouping [Julesz, 1984, Prins *et al.*, 2003]. The model presented in this thesis aims to specify how orientation is measured at the smallest scale. Psychophysical experiment generally relies on judgment that is dependant on the whole hierarchy of human capability, possibly described as awareness. This includes the perception of fundamental qualities from local orientation, through grouping of orientation, memory, decision making and the final physical communication of the ‘result’.

### 1.7.2 Occam’s razor and second-order stimuli

The principle of Occam’s razor states that [Russel, 1954] (page 494);

“It is vain to do with many what can be done with few.”

With this logic it can be argued that, in application to scientific problems, the shortest explanation that fits the data is the most correct. This idea can be argued in many areas of science, particularly in biology, where evolution dictates that mechanisms are only as complicated as they need be to survive. There is arguably a tendency to over complicate explanations of visual perception phenomena with the introduction of additional channels to account for new phenomena which contradict the old doctrine.

In relation to the construction of new perceptual models, Occum’s razor is relevant to recent discussions of second-order phenomena [Clifford *et al.*, 1998, Johnston *et al.*, 1992]. Traditional models of texture and motion perception involve non-linear rectification stages that essentially dismiss half of the available information by making positive and negative filter responses equally important in an energy measure. First-order stimuli are defined as varying locally in luminance. Second-order stimuli vary in some other aspect of a stimulus, such as orientation or contrast. Energy-based models fail to detect second order stimuli because light and dark regions are treated the same after the rectification stage [Atherton, 2002, Bergen and Landy, 1991, Landy and Bergen, 1991].

This shortcoming in energy-based models is normally corrected by the introduction of a non-rectified channel that can perceive the stimulus variation. It is often pointed out that second-order stimuli are non-existent in nature; therefore, the evolution of a channel specifically for measuring second-order stimuli is unlikely. More likely is that it is a clue that gives insight into the human visual system’s orientation-measuring hardware.

## Chapter 2

# Model development

This section will describe the methods in finding a robust orientation measure using GROM, a summary of the contents can be found in Table 2 (see page 42).

### 2.1 Finding an image derivative

Blurred derivatives of an image,  $I$ , can be found by convolving them with derivative of Gaussian (DOG) kernels, according to:

$$I_{n,m}(x,y) = G_{n,m}(x,y) * I(x,y) \quad (2.1)$$

where  $I_{n,m}$  is the result of a two-dimensional convolution ( $**$ ) between a Gaussian kernel  $G_{n,m}$  with the image, and  $n$  is the  $n^{th}$  derivative in an arbitrary x-direction, and  $m$  the  $m^{th}$  derivative in an orthogonal y-direction. A detailed explanation of this can be found in the appendices [‘Image derivatives’ (see page 154)].

Figure 2.1 (see page 43) shows an example of a one-dimensional convolution, where a first derivative Gaussian kernel,  $G_1(x)$ , is convolved with a sine-wave function, using the same standard deviation ( $\sigma = 1.5pixels$ ) and filter support (23 pixels) as the model parameters. Figure 2.2 (see page 43) shows a two dimensional convolution in graphical form, this time, to highlight the effect, using different parameters to those used in the model,  $\sigma = 6.0pixels$  and a filter support of 51 pixels. The DOG filter shown in Figure 2.2 (see page 43) has a profile like receptive fields found in visual cortex [Hubel and Wiesel, 1968]. Neurophysiological investigation [Hubel and Wiesel, 1968] has reported these profiles and profiles like higher-order DOG filters. It has also been shown how these filters can be synthesized by finding the difference between the

‘Finding an image derivative’ (see page 41)	Shows how images are filtered with derivatives of Gaussians to find their derivatives.
‘Taylor/Jet representation’ (see page 43)	Shows how multiple derivatives can be combined to make local jets.
‘Two-dimensional Tangent Measure’ (see page 45)	Derives a contrast insensitive tangent measure using a least squares fit approach.
‘Multi-scale tangent estimation’ (see page 48)	Describes a multi-scale extension to the tangent measure.
‘Tangent measures at many orientations’ (see page 49)	Describes the measurement of tangent in multiple rotated reference frames, using directional image derivatives.
‘Favouring oblique measures’ (see page 52)	Describes how and why oblique measures are favoured in the final orientation estimation from the various rotated reference frames.
‘Deducing a final single measure’ (see page 53)	Describes the last step of the model where orientation is deduced with an accompanying variance measure.

Table 2.1: A table of chapter contents

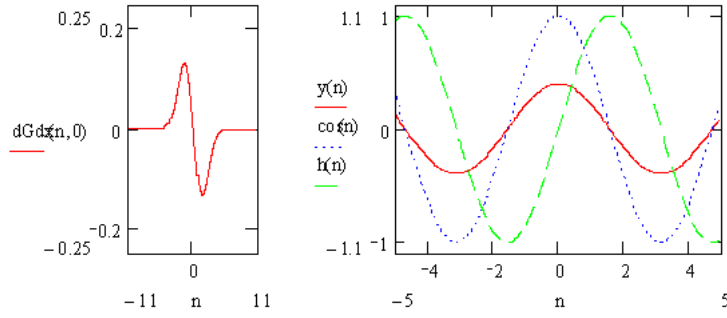


Figure 2.1: A first derivative of Gaussian kernel (left) is convolved with a sine function (right, green) resulting in a cosine function (right, red) which has a reduced amplitude but the same phase as a plotted cosine function (right, blue). The first derivative of sine is cosine.

filter outputs of two slightly offset Gaussian distribution profiles [Young *et al.*, 2001]  
‘Gaussian derivative synthesis’ (see page 151).

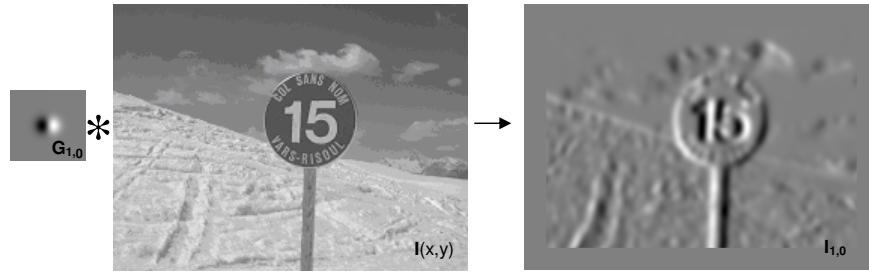


Figure 2.2: A ‘visual equation’ analogous to Equation 2.1, showing a Gaussian kernel with  $\sigma = 6.0pixels$  and a filter support of 51pixels (left), differentiated once in the horizontal direction, therefore a  $G_{1,0}$ , convolved with an image  $I$  (middle) and showing the resulting image derivative  $I_{1,0}$  (right). Note the horizontal changes in the image are extracted with polarity showing up as black or white, depending on the local slope, vertical changes are not detected.

## 2.2 Taylor/Jet representation

The Taylor series (2.2) is a widely used approximation method [Weisstein, 2004, Boas, 1993]. Based on various measurements of a function at one point, it provides an estimation of the value of the same function at another point displaced from the original. The series can be used to estimate the luminance in an image, such that the image brightness at a point  $P(a, b)$  can be estimated as:

$$f(x, y) = f(a + p, b + q) = \sum_{n=0}^{\infty} \frac{1}{n!} \left[ p \frac{\partial}{\partial x} + q \frac{\partial}{\partial y} \right]^n f(a, b) \quad (2.2)$$

where the vector,  $\mathbf{H}$ , with components (p, q), is an offset-vector from the original point P, coordinates (a, b), to a point  $\mathbf{P}+\mathbf{H}$ , coordinates (x, y), where the approximation is required.

The quality of the approximation is determined by:

- The distance from the point of measurement. The further away from a centre of measurement an approximation is made, the less accurate the approximation will be. Noisy or detailed images, whose Taylor expansion contains high amounts of higher-order derivatives can introduce mis-approximation at small distances from the measurement point
- The number of terms in the Taylor series approximation. The lower the number of terms, the less accurate any approximation will be. The number of terms contained in a Taylor series approximation at any one point in an image has been defined as the ‘jet’ order [Koenderink and van Doorn, 1988]. Where ‘jet’ indicates a collection of derivatives, the order indicates the maximum order of derivatives contained in the jet.

In the section entitled ‘Finding an image derivative’ (see page 41) a biologically plausible mechanism for the estimation of image derivatives by convolution with receptive field like DOG filters is presented. It not only allows image differentiation but can also allow the visual system access to jets and therefore to extrapolate away from the centre of a jet. Effectively, p and q are weights of the outputs of the receptive fields that enable approximation of the visual field at any displaced point  $\mathbf{P}+\mathbf{H}$ . This approximation has been used to give a way of combining the output from higher-order DOG filters.

The truncated Taylor series expansion or Jet [Koenderink and van Doorn, 1988] can be viewed as a form of image representation. Rather than considering an image as a collection of grey-scaled values at each pixel point, the jet representation encourages a scene to be thought of as a collection of increasing orders of two-dimensional spatial derivatives at each point in image-space. Figure 2.4 (see page 56) shows an example of interpolating into areas of an image that are not at the centre of a measured point. The main advantages of this representation are that:

- images can be sampled sparsely. Equation 2.2 provides an explicit method for estimating the image structure at a point displaced from the centre of a jet measure
- the space can be transformed with ease. It is often desirable to transform an image space from a Cartesian grid to a different shape in order to more efficiently extract information. The value of points in the new ‘distorted’ space can be computed using jets, dense or sparse, from the original image; this approach relieves the problems of over and under sampling associated with any transformation [Bolduc and Levine, 1998, Wallace and D.J., 2003].

For simplicity, and in order to examine the behaviour of the local computational model in

abstraction from the effects that sampling might introduce, interpolation is not used in GROM. All results are based on forming an image jet at each point on the rectangular pixel grid. The subject of exploring the effects of interpolation from sparse-sampling is discussed in the section entitled ‘Sparse sampling’ (see page 142).

## 2.3 Two-dimensional Tangent Measure

As discussed previously, in the section regarding the plenoptic function ‘The Plenoptic function and early vision’ (see page 12), information about the visual scene is normally deduced by measuring orientation in a particular plenoptic projection, for example speed is the orientation in a chosen space-time projection. In this case, the desired measure is the local orientation in the x-y image plane, which is the direction of greatest local luminance change at a point in that plane. Figure 2.6 (see page 57) shows an example image of a Gaussian distribution as a grey scale plot (left), the direction of maximum luminance change indicated by needles in each pixel square (middle) and the direction of the contour orientation (right).

The local measure of luminance change is the derivative, the component of a derivative in the direction  $\theta$  can be written:

$$\nabla (I(x, y)) \cdot \begin{bmatrix} \sin \theta \\ \cos \theta \end{bmatrix} = \begin{bmatrix} \frac{\partial I(x, y)}{\partial x} \\ \frac{\partial I(x, y)}{\partial y} \end{bmatrix} \cdot \begin{bmatrix} \sin \theta \\ \cos \theta \end{bmatrix} \quad (2.3)$$

where  $\theta$  is measured clock-wise from vertical. To find the direction of the maximum change, one must maximise Equation 2.3 with respect to  $\theta$ , to find  $\theta_{max}$  the direction of maximum change. Where a function reaches a maximum, its first derivative becomes zero as that function reaches horizontal, equating the first derivative of a function to zero is to minimise it. In depth coverage of this mathematical procedure can be found in the appendix entitled ‘Minimisation’ (see page 167).

such that:

$$\begin{aligned} & \frac{\partial}{\partial \theta} \left( \begin{bmatrix} \frac{\partial I(x, y)}{\partial x} \\ \frac{\partial I(x, y)}{\partial y} \end{bmatrix} \cdot \begin{bmatrix} \sin \theta \\ \cos \theta \end{bmatrix} \right)_{\theta_{max}} = 0 \\ \Rightarrow & \frac{\partial I(x, y)}{\partial x} \cos \theta_{max} - \frac{\partial I(x, y)}{\partial y} \sin \theta_{max} = 0 \\ \Rightarrow & \cot \theta_{max} = \frac{\frac{\partial I(x, y)}{\partial y}}{\frac{\partial I(x, y)}{\partial x}} \end{aligned} \quad (2.4)$$

The result shown in Equation 2.4 shows the direction of maximum luminance change expressed as a cotangent function. The result is dependant on the sense of the orientation measure; if, for example, orientation is measured positive anti-clockwise from the horizontal, the direction of maximum luminance change is expressed in terms of tangent. For simplicity, going forwards the

algorithm will use orientation expressed in terms of tangent, but the final orientation deduction will take account of this using the relation  $\tan \theta = \cot(\frac{\pi}{2} - \theta)$ .

One can approximate  $\tan \theta$  using higher orders of differentiation, by differentiating Equation 2.4 with respect to x:

$$\frac{\partial}{\partial x} \left( \frac{\partial I(x,y)}{\partial x} \tan \theta_{\max} \right) = \frac{\partial^2 I(x,y)}{\partial x^2} \tan \theta_{\max} + \frac{\partial I(x,y)}{\partial x} \frac{\partial}{\partial x} (\tan \theta_{\max}) = \frac{\partial}{\partial x} \frac{\partial I(x,y)}{\partial y} \quad (2.5)$$

Where we assume  $\frac{\partial}{\partial x} (\tan \theta_{\max}) \approx 0$ , that is, the variation of the direction of maximum change,  $\tan \theta_{\max}$  is small with respect to a change in position, x, Equation 2.5 becomes:

$$\tan \theta_{\max} = \frac{\frac{\partial}{\partial x} \frac{\partial I(x,y)}{\partial y}}{\frac{\partial^2 I(x,y)}{\partial x^2}} \quad (2.6)$$

This process can be repeated, by taking further derivatives of Equation 2.5, such that:

$$\tan \theta_{\max} = \frac{\left( \frac{\partial I(x,y)}{\partial y} \right)}{\left( \frac{\partial I(x,y)}{\partial x} \right)} \simeq \frac{\left( \frac{\partial}{\partial x} \frac{\partial}{\partial y} (I(x,y)) \right)}{\left( \frac{\partial}{\partial x} \frac{\partial}{\partial x} (I(x,y)) \right)} \simeq \frac{\left( \frac{\partial^{n-1}}{\partial x^{n-1}} \frac{\partial}{\partial y} (I(x,y)) \right)}{\left( \frac{\partial^n}{\partial x^n} (I(x,y)) \right)} \quad (2.7)$$

The assumption made above, that,  $\frac{\partial}{\partial x} (\tan \theta_{\max}) \approx 0$ , is only true for those parts of a scene that vary smoothly.

As demonstrated in the section entitled ‘Finding an image derivative’ (see page 41), derivatives of an image at a particular point are considered here as the result of the application of DOG filters at that point. This implies that the use of several different image measurements can give additional numerical estimations of orientation. These estimations can be combined using a least-squares fit approach.

For simplicity, it will become clear that it is convenient to consider two vectors, one with an extra derivative in the x-direction, and a second with an extra derivative in the y-direction:

$$k_x = \begin{pmatrix} \frac{\partial}{\partial x} (I(x,y)) \\ \vdots \\ \frac{\partial^n}{\partial x^n} (I(x,y)) \end{pmatrix} = \begin{bmatrix} I_{1,0} \\ \vdots \\ I_{n,0} \end{bmatrix} \quad (2.8)$$

$$k_y = \begin{pmatrix} \frac{\partial}{\partial y} (I(x,y)) \\ \vdots \\ \frac{\partial}{\partial y} \frac{\partial^{n-1}}{\partial x^{n-1}} (I(x,y)) \end{pmatrix} = \begin{bmatrix} I_{0,1} \\ \vdots \\ I_{n-1,1} \end{bmatrix} \quad (2.9)$$

where n is the highest image derivative under consideration, values for n are based on the highest order of DOG-like receptive fields reported in the Neuroscience literature, in this case, a double end-stopped receptive field [Young *et al.*, 2001]. Such a receptive field has an inhibitory lobe at each end and an excitatory centre, this is equivalent to the fourth spatial derivative of Gaussian.

Using the notation from Equation 2.8, Equation 2.7 can be re-written:

$$\tan \theta_{\max} = \frac{I_{0,1}}{I_{1,0}} \simeq \frac{I_{1,1}}{I_{2,0}} \simeq \frac{I_{2,1}}{I_{3,0}} \simeq \frac{I_{3,1}}{I_{4,0}} \quad (2.10)$$

where each term is a different measure of the stimulus, consequently they can be treated as a population of measurements. Minimising least-squares gives a good single measure of local orientation, such that  $\tan \theta_{\max}$  satisfies:

$$\frac{\partial}{\partial \tan \theta_{\max}} \left( \sum_n (K_{y_n} - (\tan \theta_{\max}) K_{x_n})^2 \right) = 0 \quad (2.11)$$

where  $K_{x_n}$  and  $K_{y_n}$  are the  $n^{th}$  elements of the vectors  $K_x$  and  $K_y$ , described in Equation 2.8 and Equation 2.9. This can be rewritten using the vector notation:

$$\min_{(\tan \theta_{\max})} \left[ (K_y - (\tan \theta_{\max}) \cdot K_x)^2 \right] \Rightarrow \frac{\partial}{\partial (\tan \theta_{\max})} (K_y - (\tan \theta_{\max}) \cdot K_x)^2 = 0 \quad (2.12)$$

Solving for  $\tan \theta_{\max}$ :

$$\tan \theta_{\max} = \frac{K_x \cdot K_y}{K_x \cdot K_x} = \frac{I_{1,0}I_{0,1} + I_{2,0}I_{1,1} + \dots + I_{n,0}I_{n-1,1}}{I_{1,0}I_{1,0} + I_{2,0}I_{2,0} + \dots + I_{n,0}I_{n,0}} \quad (2.13)$$

The arguments above can be repeated using the reciprocal of Equation 2.13, as  $\tan \theta = \frac{1}{\cot \theta}$ ;

$$\cot \theta_{\max} = \frac{K_x \cdot K_y}{K_y \cdot K_y} = \frac{I_{1,0}I_{0,1} + I_{2,0}I_{1,1} + \dots + I_{n,0}I_{n-1,1}}{I_{0,1}I_{0,1} + I_{1,1}I_{1,1} + \dots + I_{n-1,1}I_{n-1,1}} \quad (2.14)$$

It should be noted that Equation 2.11 has been arranged to minimise to the better conditioned ratios of Equation 2.13 and Equation 2.14. These are considered *well-conditioned* as the denominator cannot readily be zero. If Equation 2.11 were stated as:

$$\min_{(\tan \theta_{\max})} \sum_n \left( \tan \theta_{\max} - \frac{K_{y_n}}{K_{x_n}} \right)^2 \quad (2.15)$$

this would lead to a least-squares estimation of  $\tan \theta_{\max}$  of:

$$\tan \theta_{\max} = \sum_n \frac{K_{y_n}}{K_{x_n}} = \overline{\left( \frac{K_{y_n}}{K_{x_n}} \right)} \quad (2.16)$$

where the over-line indicates a mean of the ratios. Were any element of  $K_x$  to be zero, this would lead to the estimate being undefined. The measure minimised in Equation 2.12,  $(K_y - (\tan \theta_{\max}) \cdot K_x)$  is chosen to make this possibility very unlikely, as all image derivatives (up to n=4) are only zero in the case of completely uniform or flat structure.



As individual measures of a stimulus,  $I_{n,m}$ , and therefore  $K_{x_n}$ , can be zero, the orientation estimation in Equation 2.16 would be ill-defined as a result of any of the elements of  $K_x$  being zero. The same would not occur with the ratios in Equation 2.13 and Equation 2.14 as every element of the vector in the denominator would need to be zero to result in an ill-defined estimation. The form of Equation 2.13 and Equation 2.14 behaves better where filter measures are zero, and is consequently termed a *well-conditioned ratio*.

## 2.4 Contrast invariance

The human visual system is known to be very perceptually robust to changes in contrast; even in extreme lightness or darkness, detail can be perceived [Rodieck, 1988]. Filters used in the least-squares tangent measure are coincident at the same point, therefore they measure the same local luminance change. This luminance change is of equal magnitude in the numerator and denominator. These tangent based measures are insensitive to local luminance magnitude but sensitive to the direction of luminance change. They are termed therefore, *contrast invariant*.

## 2.5 Multi-scale tangent estimation

The k-vectors in Equation 2.8 and Equation 2.9 can be extended to include more than one scale:

$$k_x = \left[ \begin{array}{c} \left[ \begin{array}{c} \frac{\partial}{\partial x} (I(x, y)) \\ \vdots \\ \frac{\partial^n}{\partial x^n} (I(x, y)) \end{array} \right]_{\sigma_1} \\ \vdots \\ \left[ \begin{array}{c} \frac{\partial}{\partial x} (I(x, y)) \\ \vdots \\ \frac{\partial^n}{\partial x^n} (I(x, y)) \end{array} \right]_{\sigma_s} \end{array} \right] = \left[ \begin{array}{c} \left[ \begin{array}{c} I_{1,0} \\ I_{2,0} \\ \vdots \\ I_{n,0} \end{array} \right]_{\sigma_1} \\ \vdots \\ \left[ \begin{array}{c} I_{1,0} \\ I_{2,0} \\ \vdots \\ I_{n,0} \end{array} \right]_{\sigma_s} \end{array} \right] \quad (2.17)$$

$$k_y = \left[ \begin{array}{c} \left[ \begin{array}{c} \frac{\partial}{\partial y} (I(x, y)) \\ \vdots \\ \frac{\partial}{\partial y} \frac{\partial^{n-1}}{\partial x^{n-1}} (I(x, y)) \end{array} \right]_{\sigma_1} \\ \vdots \\ \left[ \begin{array}{c} \frac{\partial}{\partial y} (I(x, y)) \\ \vdots \\ \frac{\partial}{\partial y} \frac{\partial^{n-1}}{\partial x^{n-1}} (I(x, y)) \end{array} \right]_{\sigma_s} \end{array} \right] = \left[ \begin{array}{c} \left[ \begin{array}{c} I_{0,1} \\ I_{1,1} \\ \vdots \\ I_{n-1,1} \end{array} \right]_{\sigma_1} \\ \vdots \\ \left[ \begin{array}{c} I_{0,1} \\ I_{1,1} \\ \vdots \\ I_{n-1,1} \end{array} \right]_{\sigma_s} \end{array} \right] \quad (2.18)$$

where the scales  $\sigma_1$  to  $\sigma_s$  are the standard deviation of the Gaussian distribution of the DOG filter. The arguments made in ‘Two-dimensional Tangent Measure’ (see page 45) still apply; each stimulus measure can be considered as part of a population of estimations which are then combined with the same least-squares fit approach previously described. The multi-scale versions of  $K_x$  and  $K_y$  from Equation 2.12 are used in Equation 2.13 and Equation 2.14 to obtain a multi-scale estimate of orientation.

$$\tan \theta_{\max} = \frac{K_x \cdot K_y}{K_x \cdot K_x} = \frac{[I_{1,0}I_{0,1} + \dots + I_{n,0}I_{n-1,1}]_{\sigma_1} + \dots + [I_{1,0}I_{0,1} + \dots + I_{n,0}I_{n-1,1}]_{\sigma_s}}{[I_{1,0}I_{1,0} + \dots + I_{n,0}I_{n,0}]_{\sigma_1} + \dots + [I_{1,0}I_{1,0} + \dots + I_{n,0}I_{n,0}]_{\sigma_s}} \quad (2.19)$$

## 2.6 Tangent measures at many orientations

Figure 2.1 (see page 43) shows a derivative of a sine wave in the direction of the x-axis of the graph. However, in two dimensions, a derivative can be taken in any direction, not necessarily parallel to a fixed axis. Figure 2.7 (see page 57) shows orientation selective filters at a range of orientations, mimicking the known structure of the visual cortex, see ‘Orientation selectivity’ (see page 16). Using rotated filters in the tangent measure is equivalent to rotating the reference frame. As discussed in the introduction, the visual system has many orientation-specific receptive fields to perform oriented filtering.

This stage of the model is implemented using steered filtering. It is not suggested that ‘steering’ procedure is implemented biologically. More details on the implementation of this can be found in ‘Steerable Filtering’ (see page 159).

Equation 2.13 and Equation 2.14, expressed in tangent and cotangent, vary between plus and minus infinity. The least-squares measures of  $\tan \theta$  and  $\cot \theta$ , can be converted using trigonometric identities to measures that vary between known limits:

$$\cos^2 \theta = \frac{1}{1 + \tan^2 \theta} = \frac{1}{2} (1 + \cos(2\theta)) \quad (2.20)$$

and:

$$\sin^2 \theta = \frac{1}{1 + \cot^2 \theta} = \cos^2 \left( \frac{\pi}{2} - \theta \right) = \frac{1}{2} (1 - \cos(2\theta)) \quad (2.21)$$

Rotating the X-Y reference frame about each pixel — the centre of application — yields multiple measures of orientation, as in Figure 2.7 (see page 57) (b). These multiple measures of tangent and cotangent are transformed to sinusoidal functions using Equation 2.20 and Equation 2.21, The pattern of responses will contain estimates of image structure orientation plus bias introduced by the rotational offset of the reference frame rotation. For N frames, the two sets of measures yield the following equations:

$$\begin{bmatrix} \tan(\theta) \\ \tan(\theta - \phi) \\ \vdots \\ \tan(\theta - (N-1)\phi) \end{bmatrix} \rightarrow \begin{bmatrix} \cos(2\theta) \\ \cos(2\theta - 2\phi) \\ \vdots \\ \cos(2\theta - 2(N-1)\phi) \end{bmatrix} \quad (2.22)$$

and:

$$\begin{bmatrix} \cot(\theta) \\ \cot(\theta - \phi) \\ \vdots \\ \cot(\theta - (N-1)\phi) \end{bmatrix} \rightarrow \begin{bmatrix} -\cos(2\theta) \\ -\cos(2\theta - 2\phi) \\ \vdots \\ -\cos(2\theta - 2(N-1)\phi) \end{bmatrix} \quad (2.23)$$

where each vector element is from a different rotated frame,  $\phi$  is the offset between each of the rotated reference frames and  $\theta$  is the image structure's alignment, relative to the fiducial frame. Note that the constant terms and multiplying terms introduced in Equation 2.20 and Equation 2.21 have been eliminated. It is useful to consider one set of cosine measures of image structure orientation, as in Equation 2.22, and another in terms of sine measures. This requires a further phase-shift transformation of the measures in Equation 2.23 only.

The following transformation is required to convert the measures, where  $\alpha$  is an arbitrary angle:

$$-\cos \left( 2 \left( \alpha + \frac{\pi}{2} \right) \right) = \sin(2\alpha) \quad (2.24)$$

The phase shift is achieved by taking the first measure from a different reference frame, one offset by  $\frac{\pi}{2}$ . In practice, the measures are stored as vectors to introduce the required phase-shift; a fixed amount is added to the vector's index. Hence Equation 2.23 becomes:

$$\begin{bmatrix} \cot(\theta) \\ \cot(\theta - \phi) \\ \vdots \\ \cot(\theta - (N-1)\phi) \end{bmatrix} \rightarrow \begin{bmatrix} \sin(2\theta) \\ \sin(2\theta - 2\phi) \\ \vdots \\ \sin(2\theta - 2(N-1)\phi) \end{bmatrix} \quad (2.25)$$

Figure 2.6 (see page 58) shows how the sinusoidal distribution of measures from different reference frames phase shift, as the underlying image structure rotates. This rotation of underlying image structure causes a translation of the sinusoidal distribution. The overall orientation can be estimated by finding the phase-shift between the measured estimations and a fiducial ‘reference’ sine wave. This sine-wave is chosen arbitrarily and defines zero orientation.

The phase shift  $\theta$ , can be deduced by multiplying the measures with two reference sine waves, such that:

$$\sin(2(i\phi + \theta)) \cdot \sin(2i\phi) = \frac{\cos(2\theta) - \cos(4i\phi + 2\theta)}{2} \quad (2.26)$$

$$\cos(2(i\phi + \theta)) \cdot \cos(2i\phi) = \frac{\cos(2\theta) + \cos(4i\phi + 2\theta)}{2} \quad (2.27)$$

$$\cos(2(i\phi + \theta)) \cdot \sin(2i\phi) = \frac{\sin(4i\phi + 2\theta) - \sin(2\theta)}{2} \quad (2.28)$$

$$\sin(2(i\phi + \theta)) \cdot \cos(2i\phi) = \frac{\sin(4i\phi + 2\theta) + \sin(2\theta)}{2} \quad (2.29)$$

these equations can then be combined to eliminate terms containing the offset bias,  $4i\phi$ :

$$\text{Equation2.26} + \text{Equation2.27} = \cos(2\theta) \quad (2.30)$$

$$\text{Equation2.28} - \text{Equation2.29} = \sin(2\theta) \quad (2.31)$$

Another way to deduce the orientation of the image structure from the rotated measures is to consider the geometry of rotating a reference frame. This approach is less intuitive than observing a phase-shift of the measures, but each mathematical step is justifiable. See ‘Reference frame rotation’ (see page 162).

## 2.7 Favours oblique measures

Poor tangent and cotangent estimates of orientation can occur for a particular rotated reference frame when local image structure is aligned to either X or Y in that frame. In these frames values from all derivatives aligned to either X or Y can cause ill-conditioning in either the tangent or cotangent estimate as there effectively no variation of stimulus luminance profile for the oriented filters to measure. To overcome this issue GROM uses a weighting factor, combining additional measures taken from the stimulus, to favour local orientation estimates from reference frames that are oblique to the local image structure — so avoiding the degenerate cases. This weighting can be justified, using the simple tangent measure case of  $\tan \theta_{max} = \frac{I_{0,1}}{I_{1,0}}$  and considering only one reference frame, by minimising the error propagated to the orientation measure,  $\sigma_\theta$ , from the variance of the filter measures,  $\sigma_{I_{1,0}}$  and  $\sigma_{I_{0,1}}$ :

$$\sigma_\theta^2 = \frac{\partial \theta}{\partial I_{1,0}} \sigma_{I_{1,0}}^2 + \frac{\partial \theta}{\partial I_{0,1}} \sigma_{I_{0,1}}^2 = \frac{I_{1,0} \sigma_{I_{0,1}}^2 - I_{0,1} \sigma_{I_{1,0}}^2}{I_{1,0}^2 + I_{0,1}^2} \quad (2.32)$$

where  $I_{1,0}$  and  $I_{0,1}$  are the measures from orthogonal first order DoG filters. By inspection, it is easy to see that, as the variance of the filter measures is roughly equal, ie.  $\sigma_{I_{1,0}} \simeq \sigma_{I_{0,1}}$ ,  $\sigma_\theta$  is minimum where  $I_{1,0} = I_{0,1}$ , at  $45^\circ$  or oblique to the direction of the local  $x$  and  $y$ .

The weighting measures used to favour the oblique frames are derived from a further set of rotated measurements taken on the stimulus. These additional measures in effect form a multiplicative ‘comb’ to remove measures that would cause ill-conditioning; in effect the weighting function is near-zero where the corresponding measure of tangent or cotangent is ill-conditioned. For the measures deriving from the tangent estimates these weightings are:

$$\frac{1}{\cos(2\theta)} = \frac{\sum_N (K_x \cdot K_y)_\phi (K_x \cdot K_y)_{\phi+\frac{\pi}{2}} \cos(2\theta)_\phi}{\sum_N (K_x \cdot K_y)_\phi (K_x \cdot K_y)_{\phi+\frac{\pi}{2}}} \quad (2.33)$$

and:

$$\frac{1}{\sin(2\theta)} = \frac{\sum_N (K_x \cdot K_y)_\phi (K_x \cdot K_y)_{\phi+\frac{\pi}{2}} \sin(2\theta)_{\phi+\frac{\pi}{2}}}{\sum_N (K_x \cdot K_y)_\phi (K_x \cdot K_y)_{\phi+\frac{\pi}{2}}} \quad (2.34)$$

where the subscript of  $+\frac{\pi}{2}$  indicates that values from columns offset by this angle are used. This is introduced as the measures deduced in Equation 2.25 are phase shifted, i.e. picked from offset reference frames, in order to access a pure sine measure.

## 2.8 Deducing a final single measure

In order to verify the model, it is necessary to deduce a final measure of orientation as a number between  $-\frac{\pi}{2}$  and  $+\frac{\pi}{2}$ . However, it is not clear how orientation is represented in the brain, so this stage of the model is not intended to model biology. Orientation being expressed in a *dual rail*<sup>1</sup> fashion is more biologically plausible, as a comparison between the two rails could uniquely represent an orientation.

One can then combine the weighted averages from Equation 2.33 and Equation 2.34 to a final measure of image structure orientation, suitable for display and validation only:

$$\tan \theta = \frac{\overline{\sin 2\theta}}{1 + \overline{\cos 2\theta}} = \frac{1 - \overline{\cos 2\theta}}{\overline{\sin 2\theta}} \quad (2.35)$$

Equation 2.33 and Equation 2.34 have corresponding variance measures. A measure for the variance of the population that is used to deduce the weighted average is given by:

$$\sigma^2 = \frac{1}{N} \frac{\sum_N (K_x \cdot K_y)_\phi (K_x \cdot K_y)_{\phi+\frac{\pi}{2}} \left( \left( \overline{\sin(2\theta)} - \sin(2\theta)_\phi \right)^2 + \left( \overline{\cos(2\theta)} - \cos(2\theta)_\phi \right)^2 \right)}{\sum_N (K_x \cdot K_y)_\phi (K_x \cdot K_y)_{\phi+\frac{\pi}{2}}} \quad (2.36)$$

## 2.9 Summary

A model has been developed that shows how responses from simple and higher-order filters can be combined to arrive at a contrast-invariant robust measure of orientation. In this context, ‘robust’ should be taken to mean that the orientation measure is independent of ill-conditioning represented in the single filter inputs — it will be demonstrated in the subsequent results that GROM gives an accurate measure of orientation where single filter inputs are zero.

A least-squares fit approach was proposed for the combination of filter responses arriving at a well-conditioned ratio. This approach is extended to include filter responses over many scales. The action of a ratio is to make the estimate of orientation invariant to contrast, where the local luminance of stimuli — the amplitude of the signal — is cancelled out between the numerator and denominator.

The plausibility of the filters used in GROM are discussed in ‘Receptive field fitting’ (see page 20) and ‘Gaussian derivative synthesis’ (see page 151). Koenderink proposed the idea that local

---

<sup>1</sup>a dual rail code is one which is singly redundant, if one rail fails, there is enough information in the remaining rail to deduce an answer

jets effectively decompose the ‘sensorium’ for use by different biological programs [Koenderink and van Doorn, 1987]. These local jets are equivalent to a local Taylor expansion, therefore they have implicit within them the ability to approximate luminance at a point displaced from their centre of measurement. The filters that make up the jets are accounted for by the work of Young et al. [Young *et al.*, 2001]. This shows that the effect of higher-order derivative filters can be synthesised by spatially-offset lower-order DOG filters.

Orientation selective structures of the brain are modelled in the section concerning rotated reference frames ‘Tangent measures at many orientations’ (see page 49). This work demonstrates that local jets are available at many orientations.

The model of the DoG filters, that make up the jets, that in turn form orientation selective columns of the visual cortex — combine to form a coherent biologically plausible basis for GROM. In order to form the least-squares fit ratio for contrast invariant estimation of orientation use in GROM, these structures need to be connected together in a certain way. Filters contributing to the numerator must in some way be excitatory, and those contributing to the denominator must in turn contribute inhibitory signal to a result. These networks have not been verified, however, contrast invariance in orientation selective structures strongly suggests an inhibitory feedback mechanism [Cheng *et al.*, 1994, Carandini and Sengpiel, 2004], which can be modelled as a ratio. These factors all contribute to claims that GROM is biologically plausible.

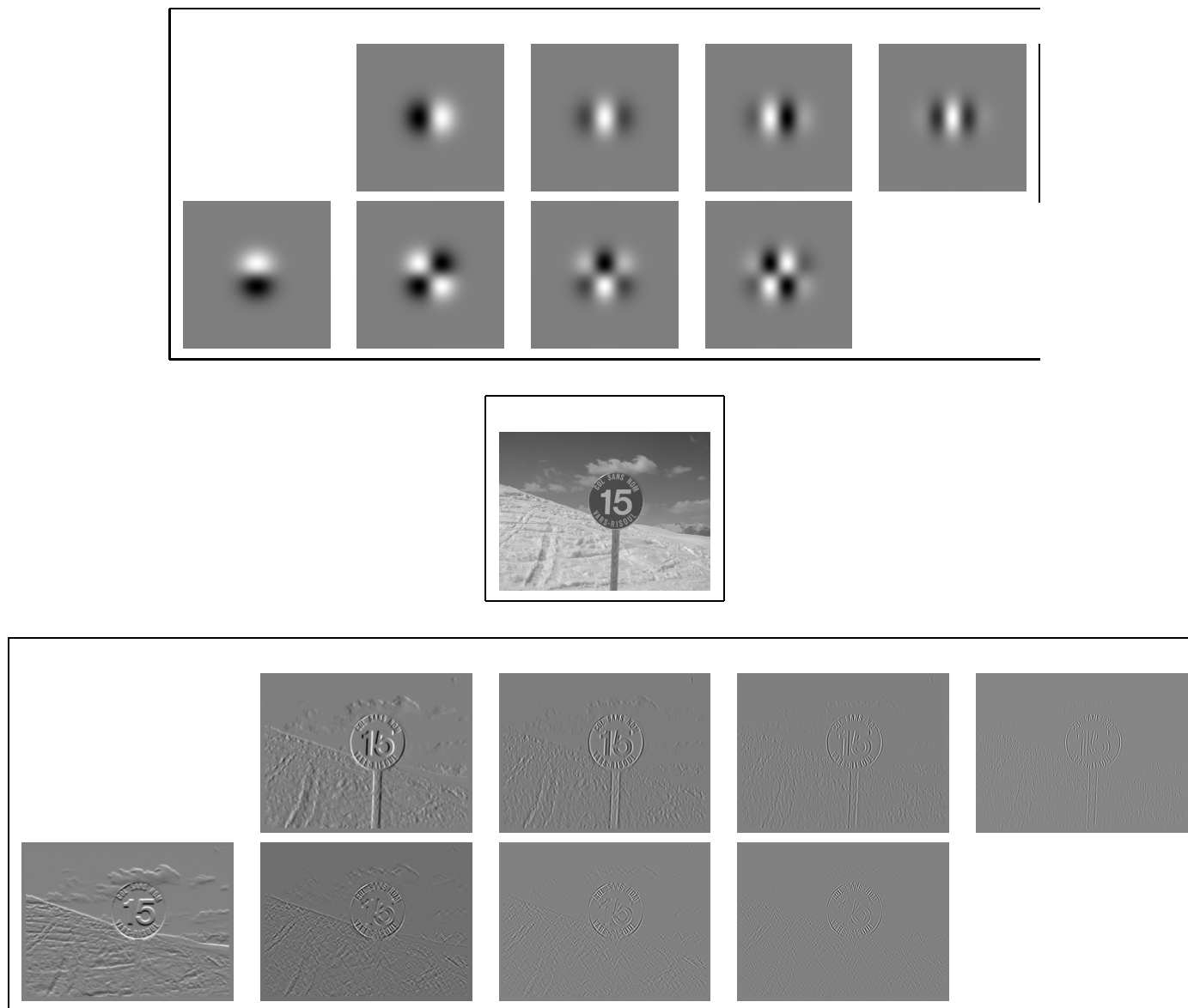


Figure 2.3: A filter bank  $G_{n,m}$ , where  $n$  is the column and  $m$  is the row, of the higher order partial derivatives used GROM (top box). This is convolved with the image  $I$  (middle). Resulting in a bank of correspondingly differentiated convolution results  $I_{n,m}$  (bottom box). The image  $I$ , the corresponding filter  $G_{n,m}$  and convolved image  $I_{n,m}$  are the visual equivalent of Equation 2.1.





Figure 2.4: (left) A sample colour image. (right) An image reconstructed using a Taylor expansion based on a sparsely sampled version of the original. Sampling points here were five pixels apart and each of the three colour channels was treated as a separate surface, the jet went up to the fourth partial derivative.

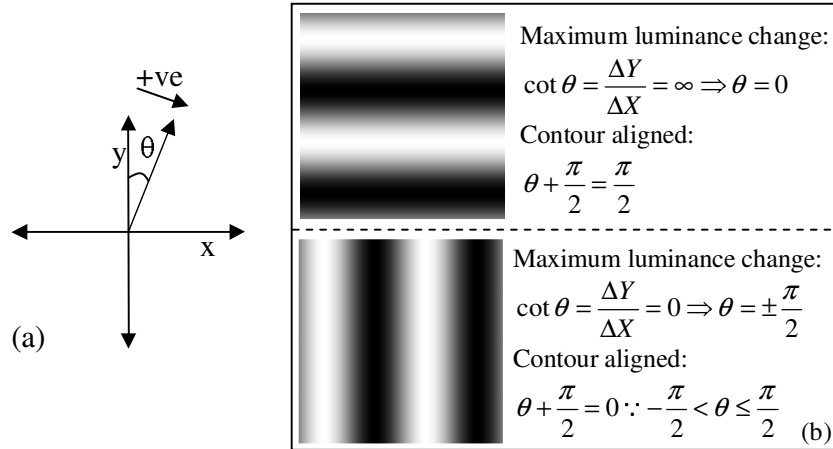


Figure 2.5: (a) A Cartesian frame showing how orientation is defined with respect to the axes. (b) Two examples of oriented texture with the deduction of their orientation beside. In these examples,  $\Delta Y$  is the difference in the luminance value in the Y direction and  $\Delta X$  the same in the X direction. The ratio of these two differences has been shown to be the cotangent of the angle of the direction of maximum luminance change. In the upper section of (b), the luminance changes maximally in the vertical direction, oriented at  $\theta = 0$ , meaning the orientation along the contour, oriented at  $\frac{\pi}{2}$  to the direction of maximum luminance change is  $\theta = 0 + \frac{\pi}{2} = \frac{\pi}{2}$ . The same logic follows for the lower case, however this demonstrates that the contour orientation must lie within  $-\frac{\pi}{2} \leq \theta \leq +\frac{\pi}{2}$ ; this illustrates the fact that orientation has twofold rotational symmetry — meaning that, for this definition of orientation, the orientation is the same when the figure is rotated through  $180^\circ(\pi)$ .

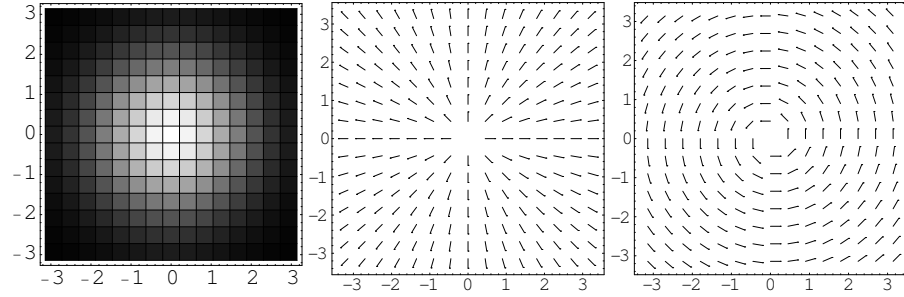


Figure 2.6: (left) A grey scale plot showing a Gaussian distribution with standard deviation of 1.5 pixels. (middle) A ‘needle’ plot showing the direction of the maximum change in luminance of the grey scale plot. (right) A ‘needle’ plot showing the direction perpendicular to the maximum change in luminance of the grey scale plot — referred to later as the *contour orientation*.

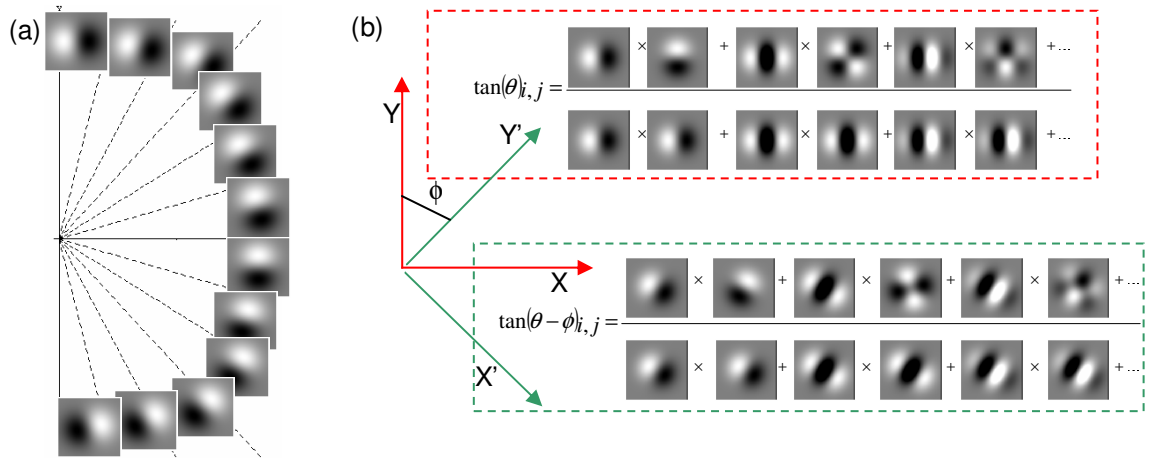


Figure 2.7: (a) A schematic showing simple filters ( $G_{1,0}$ ) in reference frames rotated about the local origin, modelling orientation hyper-columns found in the visual cortex. (b) A graphical schematic of Equation 2.13, the tangent measure of local orientation, in the X-Y reference frame (red) and a frame X'-Y', rotated about the origin by  $\phi$  (green).

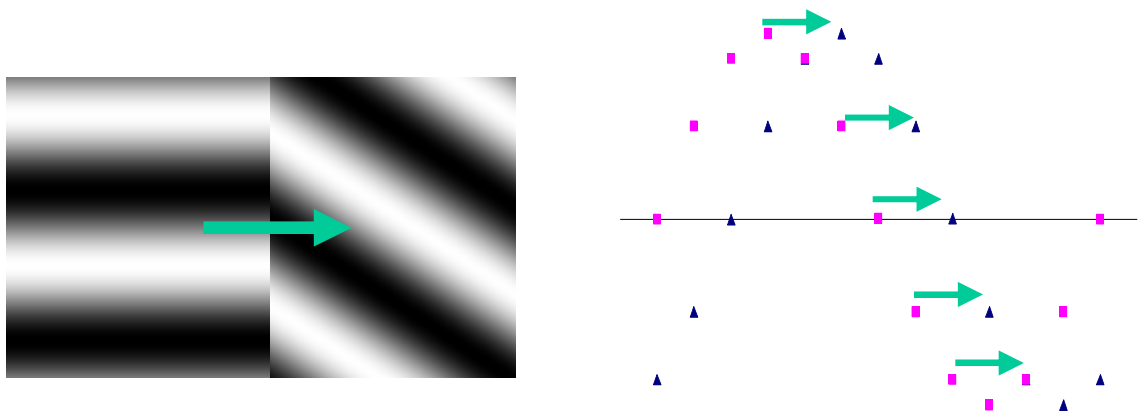


Figure 2.8: A schematic diagram showing what happens to the rotated measures (points in graph, right) from an arbitrary pixel when the image structure (left) is rotated by  $\theta$ . Image structure rotation is seen to cause a phase shift in the measures — the points in the right hand plot. A rotation of image structure of  $180^\circ$  would lead to the phase shift wrapping round to the same position.

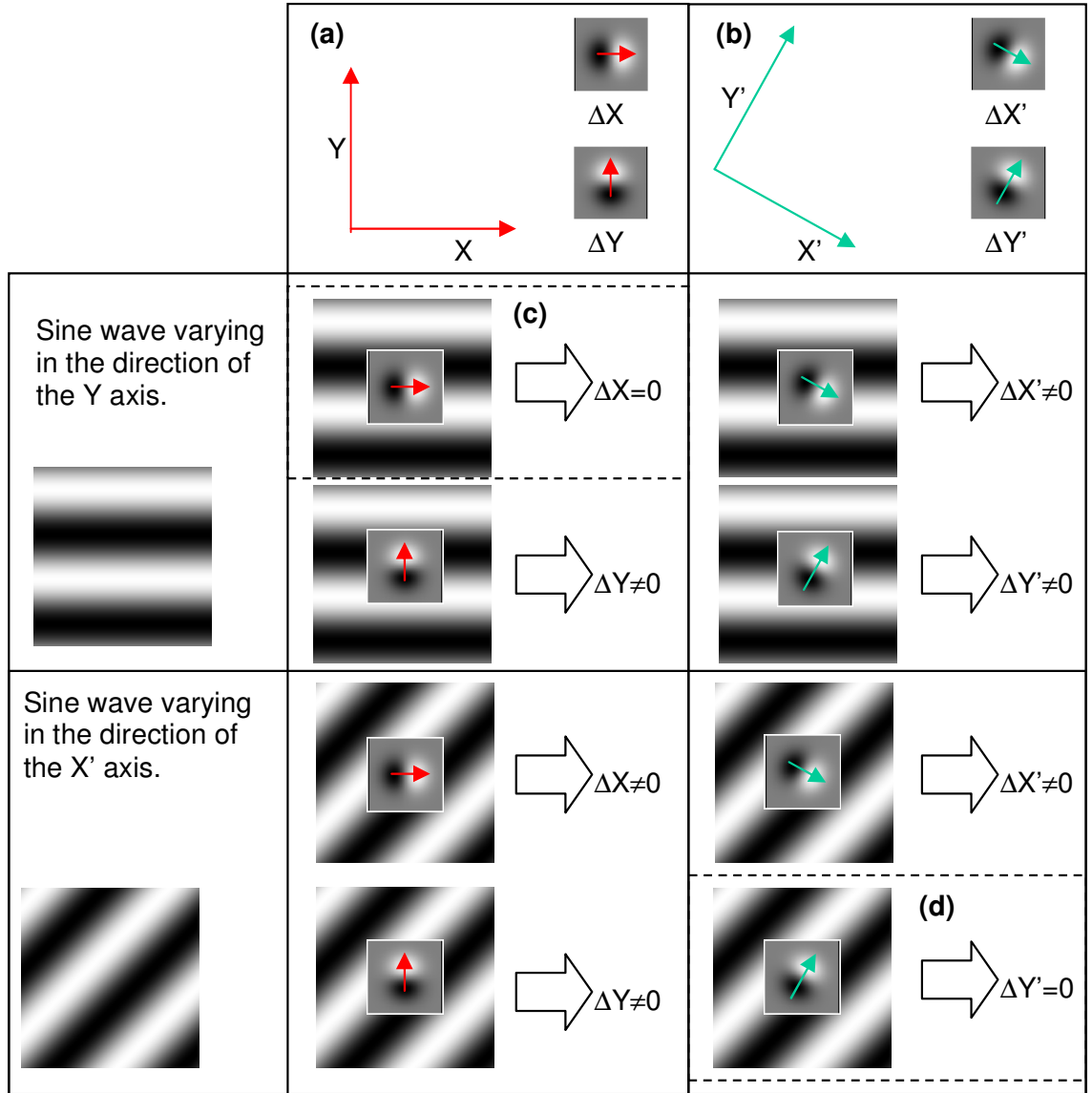


Figure 2.9: A figure showing the problem that occurs when a particular reference frame is aligned with image structure. (a) A fiducial reference frame  $X$ - $Y$ , with correspondingly aligned filters,  $\Delta X$  and  $\Delta Y$ , showing how filters in that column are aligned. (b) A reference frame  $X'$ - $Y'$  rotated through a non-zero offset, with its rotated filters  $\Delta X'$  and  $\Delta Y'$ , again showing the alignment of the frame for its column. (c) The zero offset frame shows no response from a filter,  $\Delta X = 0$ , where image structure (the sine wave) is invariant in the direction of the filter's variation. (d) Another case where the diagonal image structure is aligned with the filter in the non-zero offset frame giving  $\Delta Y' = 0$ .

## Chapter 3

# Model Results

This chapter is dedicated to the model results for simple stimuli. The effect of the variation of model parameters is examined, along with the overall performance of the orientation measure for simple stimuli. Additionally, results from this model are compared to a recently presented machine vision model.

### 3.1 Basic Stimuli

This section explores the effect of the variation of model parameters on basic stimuli. The model parameters include:

- the number of derivatives in the tangent measure — length of the k-vector.  
‘Two-dimensional Tangent Measure’ (see page 45)
- the number and size of scales of the DOG Kernels used in the orientation measure.  
‘Multi-scale tangent estimation’ (see page 48)
- the favouring of measures from reference frames oblique to the image structure.  
‘Favouring oblique measures’ (see page 52).

Basic stimuli are a good initial test of a model as the orientation output is comparable to a known ‘ground truth’. Unless otherwise stated, the model is applied to images that are generated as 256 by 256 pixels square, Figure 3.1 (see page 61) is a schematic showing an example output and its interpretation.

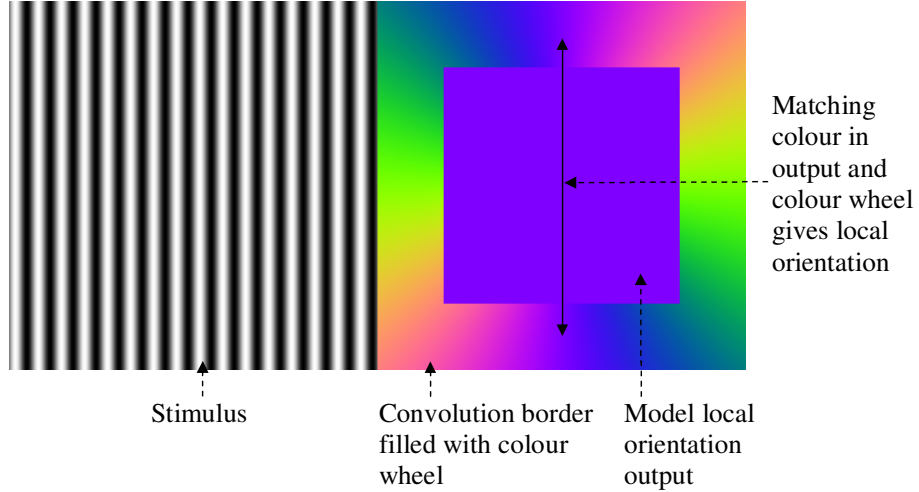


Figure 3.1: A figure showing a typical input and output of the orientation model. The stimulus (left) plotted here in 256 levels of grey, originates as a two-dimensional array of floating point numbers, from which the orientation map (right) is calculated using the method described in this thesis. This display was selected as it gives the orientation in physical, verifiable sense but is still local to each pixel. The orientation is the contour orientation — taken to mean the alignment of detail. In the figure, vertical bars result in a vertical read-out on the colour wheel. This is also perpendicular to the direction of greatest gradient change.

The *orientation map* output of the model is smaller by a homogenous border incurred by convolution. The border size varies according to the largest scale kernel used. The larger the standard deviation of the scale, the larger *filter support* required to contain the detail of the filter such that there is a near-zero value at the periphery of the kernel as in Figure 3.1 (see page 62). A table of the filter supports used can be found in Table 3.1 (see page 67).

Where multiple scales are used, the variation between scales is in octaves, i.e. the standard deviation increases by a factor of two between scales and the filter support size also increases in octaves, as in Table 3.1 (see page 67), though it is set to the nearest odd-number as kernels used in discrete convolution must have an odd numbered side length [Bracewell, 1986]. Unless otherwise stated, for example in cases where the effect of model parameter variation is being demonstrated the parameters have the standard values given in table Table 3.2 (see page 67).

The effect of  $k$ -vector length is shown on an 8cpd sine-wave in Figure 3.3 (see page 63) and on a 2.5cpd circular sine-wave in Figure 3.4 (see page 64). The circular sine wave also shows the effect of not favouring measures from reference frames oblique to the local structure. Generally, it is seen that where  $k > 1$ , the problems of local flatness are overcome in both the simple and circular sine waves — the orientation map output corresponds well to the stimulus in these cases. Where

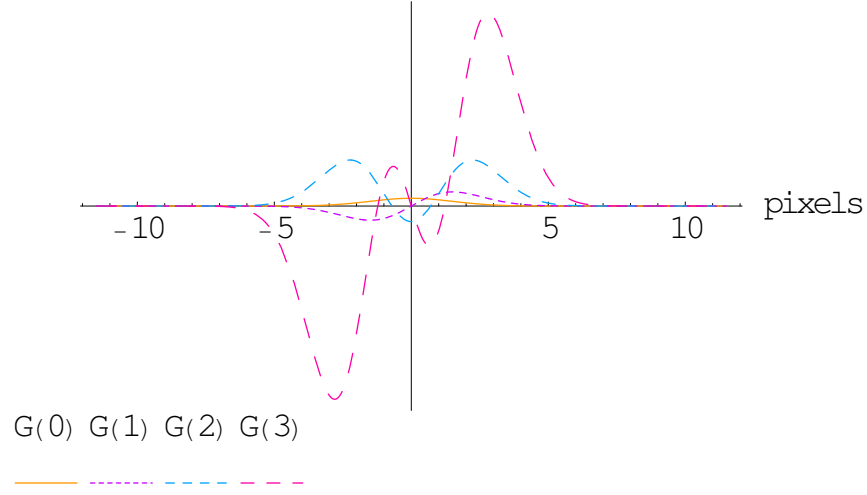


Figure 3.2: A plot showing a Gaussian distribution where  $\sigma = 1.5$ , with the corresponding first, second and third derivatives labelled  $G(0)$ ,  $G(1)$ ,  $G(2)$  and  $G(3)$  respectively. This plot shows that at  $x = 11.5$  pixels (equating to a filter support of 23 pixels) the Gaussian and its derivatives are small in comparison to their maxima.

oblique measures are not favoured, the results show orientation mis-measure along lines aligned to the reference frames.

The effect of increasing the number of derivatives and scales in the ratio can be seen in Figure 3.5 (see page 65). Increasing the length of the k-vector and number of scales decreases the amount of variance. Not favouring of oblique measures, as in the plot ‘s4k4 equal’ produces mis-measure at angles aligned to the 12 equally spaced reference frames — this can be seen to give markedly higher variance.

Figure 3.6 (see page 66) shows the mean orientation of an 8cpd sine-wave stimulus oriented at  $-\frac{\pi}{7}$  when its spatial frequency is varied, for various model parameters. The orientation is correctly measured up to the Nyquist limit of spatial frequency. As one degree of visual angle scales to 128 pixels; this implies a theoretical maximum perceivable spatial frequency of 64cpd. Above this spatial frequency aliasing occurs; this in itself creates textures with orientation information, as is demonstrated by the section of the graph with a spatial frequency above 64cpd. Possible aliasing effects in human vision have been proposed by Willams et al. [Williams, 1985].

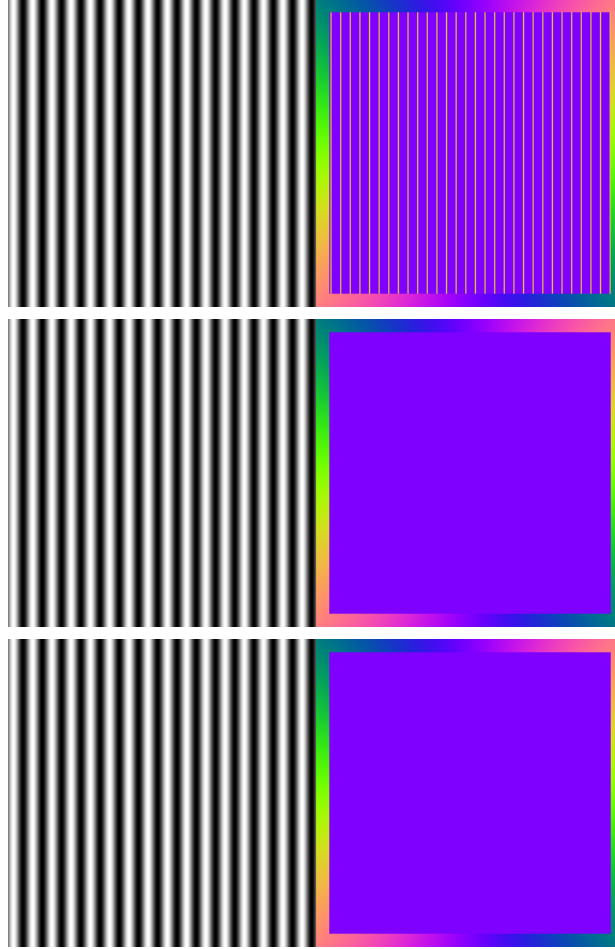


Figure 3.3: The effect of k-vector length on a simple 8cpd sine-wave. [top row]  $k = 1$  [middle row]  $k = 2$  [bottom row]  $k = 4$ . These results show the effect of increasing the number of ratios in the least-squares fit. Orientation is mis-measured by the simple ratio, where  $k = 1$ , as relative minima and maxima in the sine-wave stimuli cause the ratio to be ill-conditioned. This is obviated by using higher-derivatives, where  $k > 1$ , as these are non-zero at minima and maxima [Weisstein, 2004, Boas, 1993] and produce well-conditioned ratios at all points. The variance for these orientation maps is contained in the graph in Figure 3.5 (see page 65).



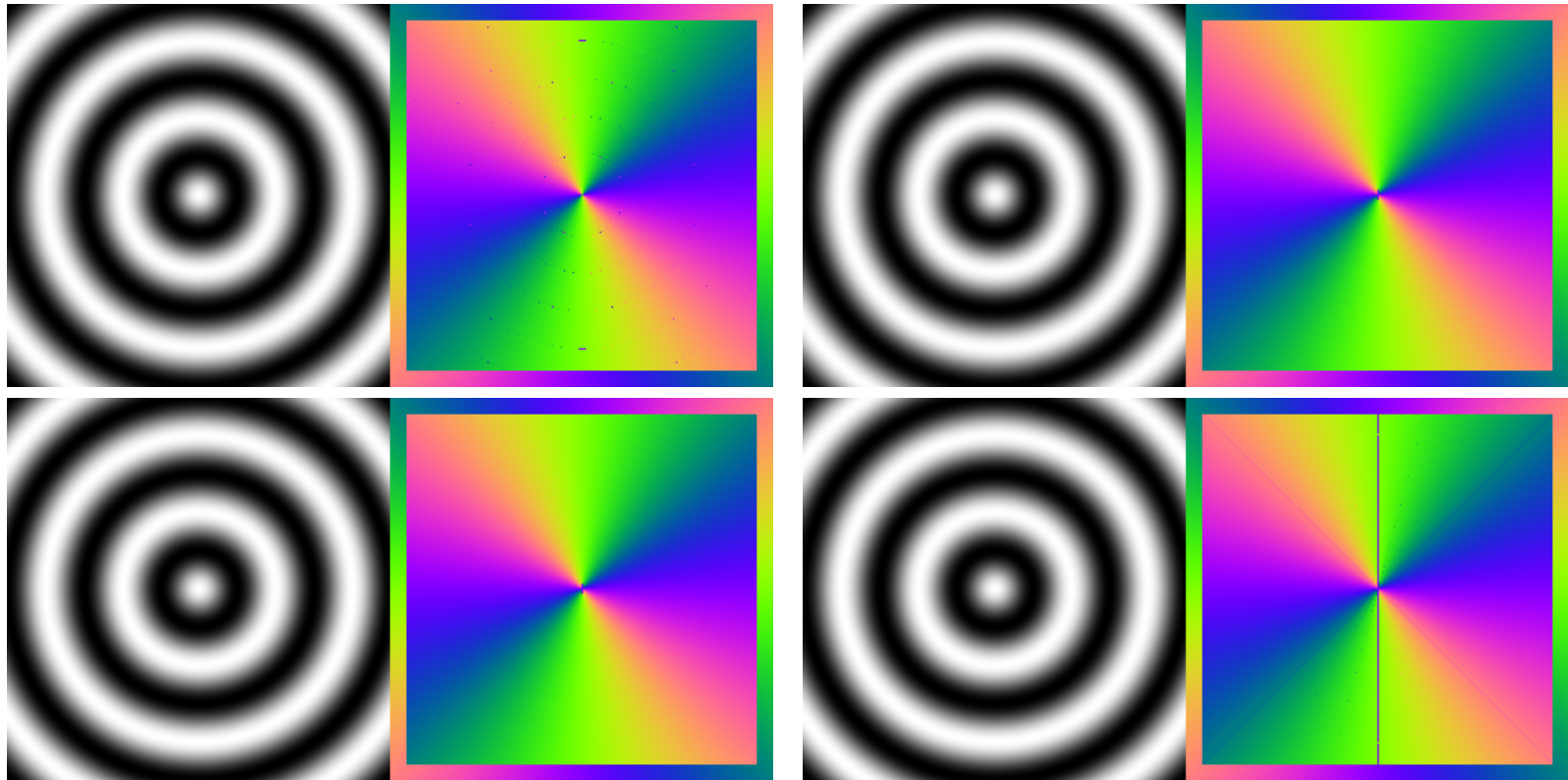


Figure 3.4: The effect of  $k$ -vector length on a 2.5cpd circular sine-wave stimulus. [top left]  $k = 1$  (top right)  $k = 2$  (bottom left)  $k = 4$  (bottom right)  $k = 4$  and measure from oblique reference frames are not favoured. The problems associated with local minima and maxima in the  $k = 1$  case are also seen in the circular sine wave stimulus above; again this is solved by the inclusion of higher-order derivatives, where  $k > 1$ . Where structure is aligned with any of the 12 reference frames — which must happen at some point in a circular stimulus — the orientation is mis-measured. The effect of this can be seen in the bottom right orientation map above. At orientations equal to one of the orientations of the reference frames, the orientation of structure is mis-measured due to ill-conditioning of the ratio. This problem is solved by favouring measures from reference frames not aligned to the local structure. ‘Favouring oblique measures’ (see page 52).

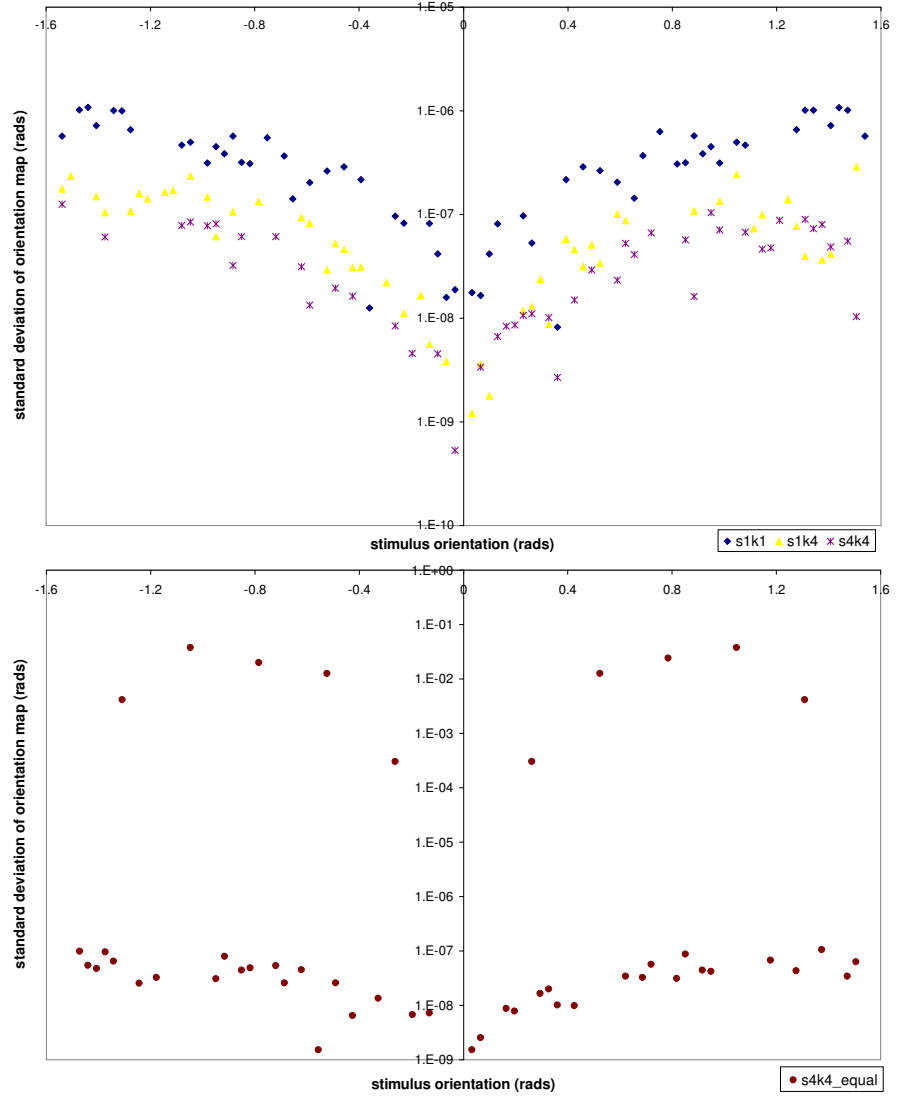


Figure 3.5: Graph showing the effect of scale and k-vector length on the variance of an orientation map, plotted on a logarithmic scale. The stimulus used was an 8cpd rotated sine wave, as seen in Figure 3.3 (see page 63)). The orientation of the sine wave was changed and the standard deviation of values representing orientation were measured. (top) ‘s1k1’ implies one scale, with a simple ratio, i.e.  $\sigma_1 = 1.5$  pixels and  $k = 1$ . The effect of increasing the number of derivatives and scales in the ratio can be seen in the figure. Increasing the length of the k-vector and number of scales decreases the amount of variance. (bottom) By not favouring of oblique measures, as in the plot ‘s4k4 equal’, the 12 equally spaced orientations aligned to reference frames can be seen to have markedly higher variance. Note that the variance has a tendency towards zero for stimulus orientations measuring toward zero. This is an artefact of the radian value used to represent orientation. Towards zero it is further from  $\pm \frac{\pi}{2}$  discontinuity, and is therefore less likely to span it, which would incur a larger variance value.

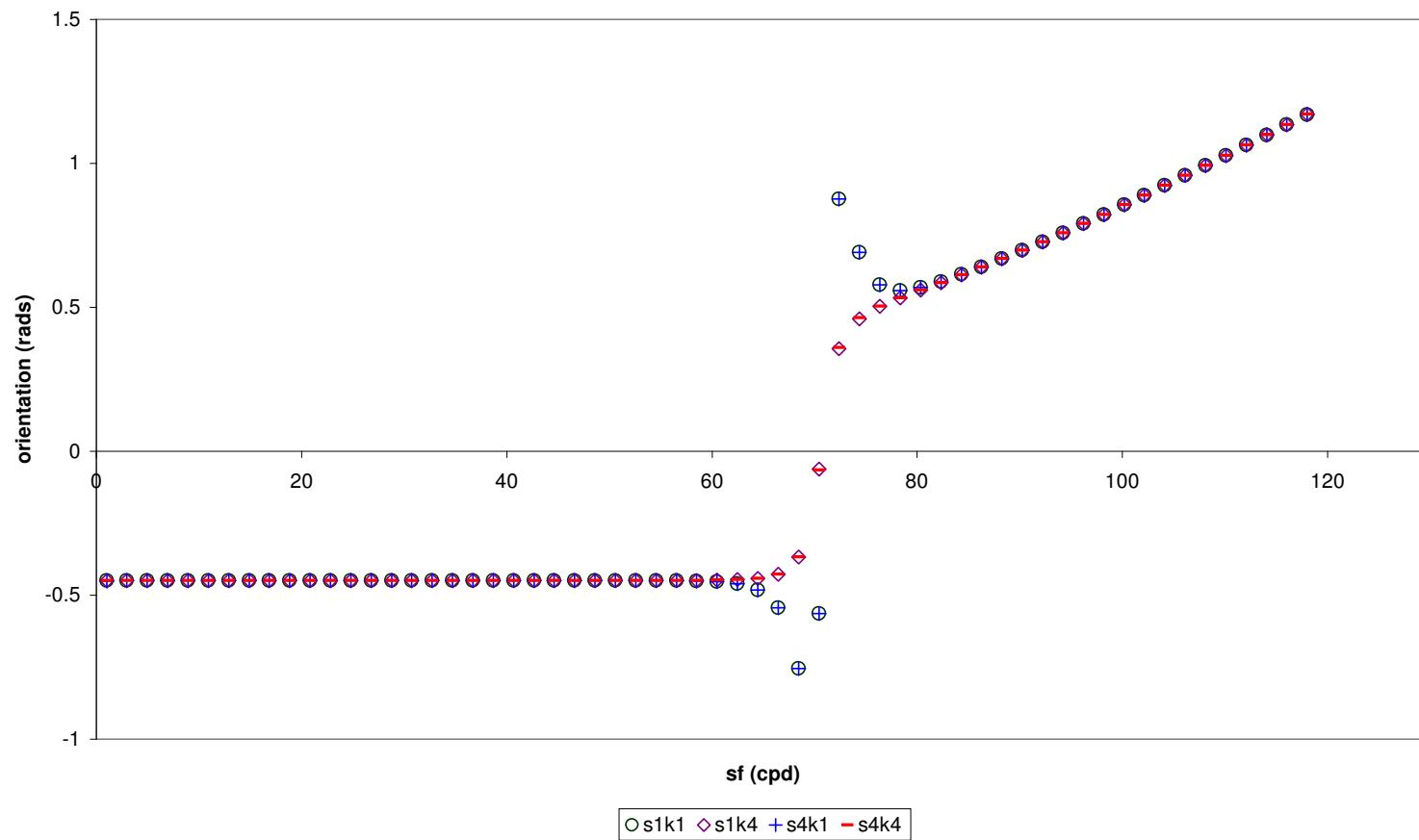


Figure 3.6: A graph showing the mean orientation of an 8cpd sine-wave stimulus oriented at  $-\frac{\pi}{7}$  when its spatial frequency is varied, for various model parameters. The orientation is correctly measured up to the Nyquist limit of spatial frequency. One degree of visual angle scales to 128 pixels; this implies a theoretical maximum perceivable spatial frequency of 64cpd. Above this spatial frequency aliasing occurs; this in itself creates textures with orientation information, as is demonstrated by the section of the graph with a spatial frequency above 64cpd. Possible aliasing effects in human vision have been proposed by Willams et al. [Williams, 1985].

Scale	$\sigma$ (pixels)	filter support (pixels)
1	1.5	23
2	3.0	47
3	6.0	69
4	12.0	93

Table 3.1: A table showing the size of the standard deviation of Gaussians in each scale, and the required filter support, according to the tolerance to near-zero values used by Johnston et al. [Johnston *et al.*, 1999].

Parameter	symbol	value
k-vector length	k	4
number of scales	s	1
filter support	–	23 pixels
standard deviation of first scale	$\sigma$	1.5 pixels
favouring oblique measures	–	on
number of columns	N	12

Table 3.2: A table showing the value and status of the main model parameters. Unless otherwise stated, they should be assumed to be these values throughout.

## 3.2 Natural Images

This section shows the performance of the model when applied to natural images. There is no way of comparing the results to the veridical orientation; it is generally for the reader to assess the performance by comparing detail in the image to the orientation output in the same region. The interactions between noise and scale can be seen; areas with scant detail are problematic as pixel noise in the luminance can dominate the orientation measure. Any single speck in luminance incurs local gradient information and hence, orientation information.

$$\frac{\partial I_{x,y}}{\partial x}^2 + \frac{\partial I_{x,y}}{\partial y}^2 \quad (3.1)$$

This problem is lessened by using a Sobel style gradient measure (Equation 3.1) to form a gradient map. This is different from traditional Sobel operator implementation as it has a far bigger kernel and standard deviation, however, the theory remains the same: to indicate the magnitude of gradient information at a pixel point. Note that as both terms are squared, this gradient measure is independent of rotation.

The two partial derivatives of the image required to form this measure are already calculated in the estimate of orientation. It is possible to ‘gate’ the orientation output by selecting an arbitrary minimum gradient map value, below which the local orientation measure can be ignored as insignificant and set to black in the orientation map. The minimum gradient map magnitude is chosen to be the mean value of the gradient map magnitude for each stimuli. This preserves the orientation measure in areas of the image with detail, but removes orientation information in areas where the image is slow changing. This is intended only as an aid for display though it is worth noting that, as this orientation measure relies on gradient information, this measure could potentially be employed as a measure of salience for orientation information in the visual system. It should be noted that where multiple scales are used, the smallest scale is used for this gating.

Natural images are captured under a variety of lighting conditions. The example test image used in this section, ‘lena’, a common image used to demonstrate various models and image processing techniques, and measures 512 pixels square. This size is modelled as 4 degrees of visual angle, hence in order to scale the image to the equivalent proportion of visual angle, it should be viewed from a distance 14 times it’s width on the page ( $\frac{1}{\tan 4^\circ} \simeq 14$ ). The reader should note that this is far smaller than normal viewing distance, this should be taken into account when assessing results.

Figure 3.7 (see page 70) shows the effect of including more derivatives in the tangent estimate on a natural image. Results appear best where  $k = 4$ , however, only on close examination can it be seen as less noisy from the  $k = 1$  case.

Figure 3.8 (see page 71) shows the effect of including multiple scales on model results. It can be seen that the inclusion of larger scale DOG filters gives less noise in the orientation output, but generally the results for the first three rows are similar. The orientation map in the last row — where only the higher scale filters ( $\sigma_2 = 3.0$ ,  $\sigma_3 = 6.0$  and  $\sigma_4 = 12$  pixels) are used — seems to have more orientation information on the detail. However, edges of the features are broadened, this is expected as the larger DOG kernels include a larger blur.

Figure 3.8 (see page 71) shows the effect of changing the threshold of the gating on the gradient measure on model results. The threshold is varied according to:

$$threshold = x \left( \overline{\frac{\partial I_{x,y}}{\partial x}^2 + \frac{\partial I_{x,y}}{\partial y}^2} \right) \quad (3.2)$$

where  $x$  is the factor varied between 0.1 and 5.0 and the over bar indicates averaging of this value over the entire gradient map. Changing this parameter varies the amount of orientation information on the clearest lines in the orientation output.

The effect of favouring oblique measures on a natural image as seen in Figure 3.5 (see page 65). Favouring oblique measures reduces the problem caused by detail-alignment with a column. The result suggests that a directional bias is introduced when oblique measures are not favoured. This demonstrates that favouring oblique measures gives better results for natural images.

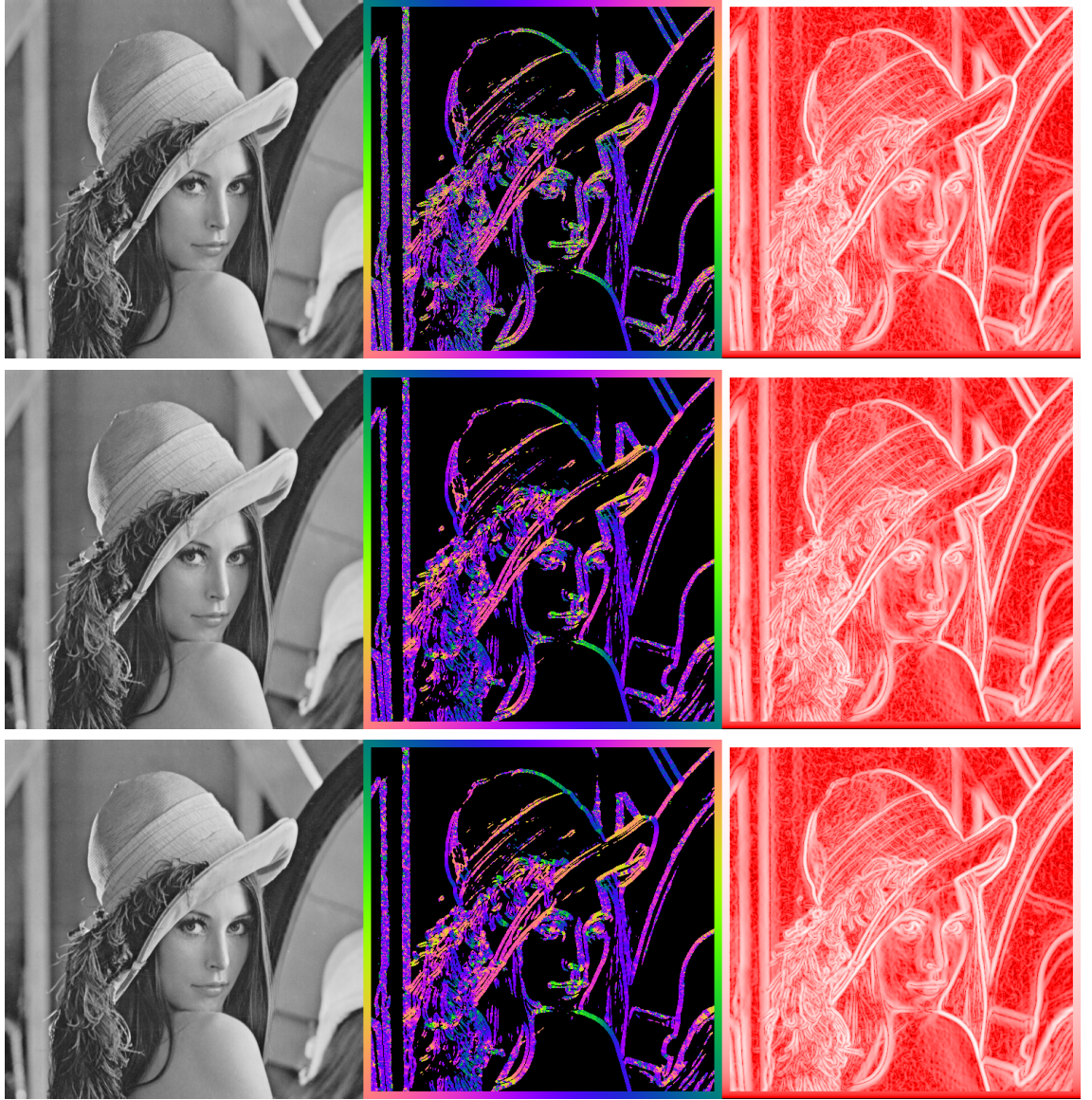


Figure 3.7: A figure showing the effect on a natural image of including more ratios in the least squares fit, this is similar to the experiment shown in Figure 3.3 (see page 63). The input image (left column), is processed resulting in a masked orientation map (middle column) and gradient map (right column) with which the orientation map in the same row is masked where the gradient map's corresponding pixel is above the overall mean value. (top row) A simple ratio in each reference frame,  $k = 1$ . (middle row)  $k = 2$ , showing a an improved, less noisy masked-orientation output. (bottom row)  $k = 4$  on close examination, this appears to be the best orientation output for a single scale, less noisy than where  $k = 1$ .



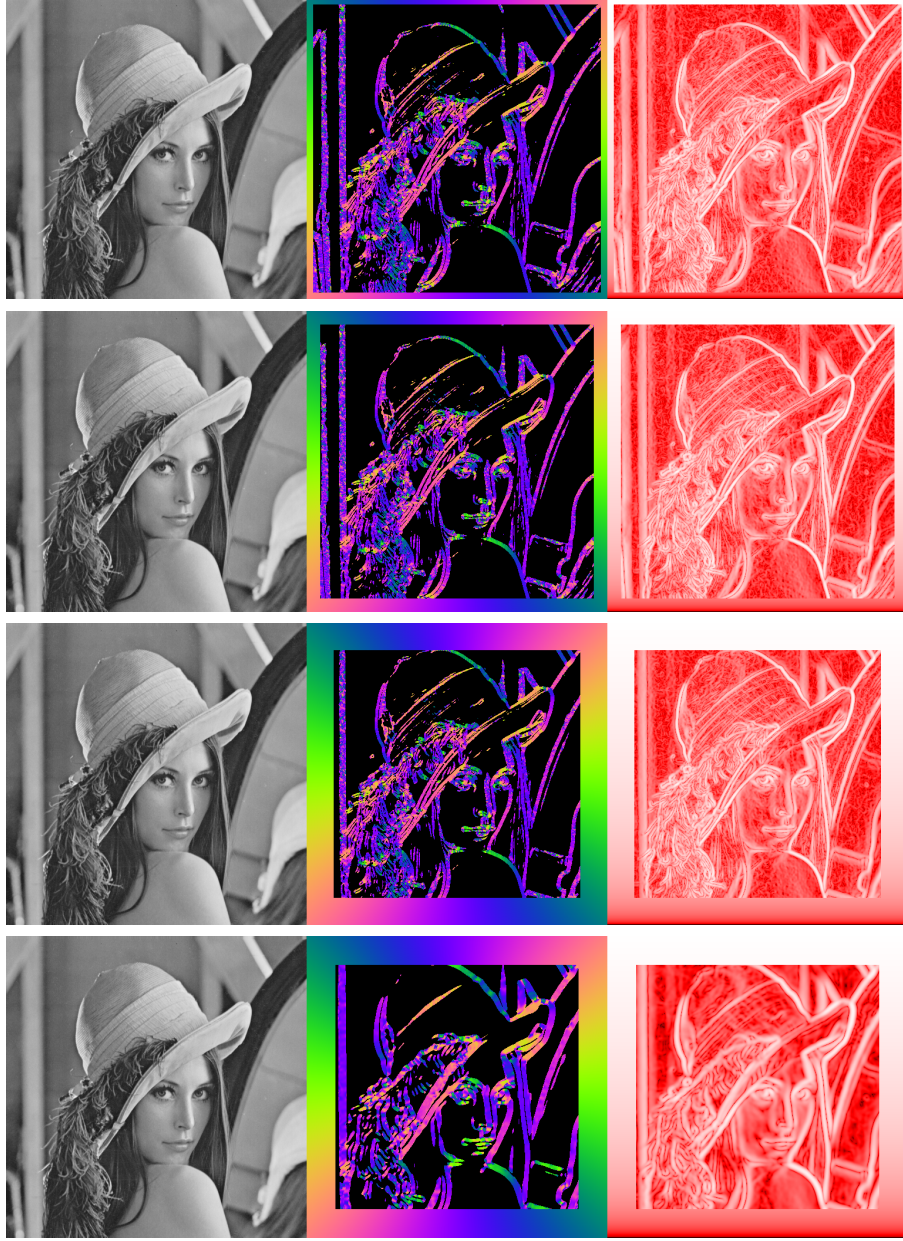


Figure 3.8: A series of model results showing the effect of including multiple scales. [top row] One scale ( $\sigma = 1.5$  pixels). [second row] Two scales ( $\sigma_1 = 1.5$  and  $\sigma_2 = 3.0$  pixels). [third row] Four scales ( $\sigma_1 = 1.5$ ,  $\sigma_2 = 3.0$ ,  $\sigma_3 = 6.0$  and  $\sigma_4 = 12$  pixels). [fourth row] Three upper scales ( $\sigma_2 = 3.0$ ,  $\sigma_3 = 6.0$  and  $\sigma_4 = 12$  pixels). It can be seen that the inclusion of larger scale DOG filters gives less noise in the orientation output, but generally the results for the first three rows are similar. The orientation map in the last row — where only the higher scale filters ( $\sigma_2 = 3.0$ ,  $\sigma_3 = 6.0$  and  $\sigma_4 = 12$  pixels) are used — seems to have more orientation information on the detail. However, edges of the features are broadened, this is expected as the larger DOG kernels include a larger blur.



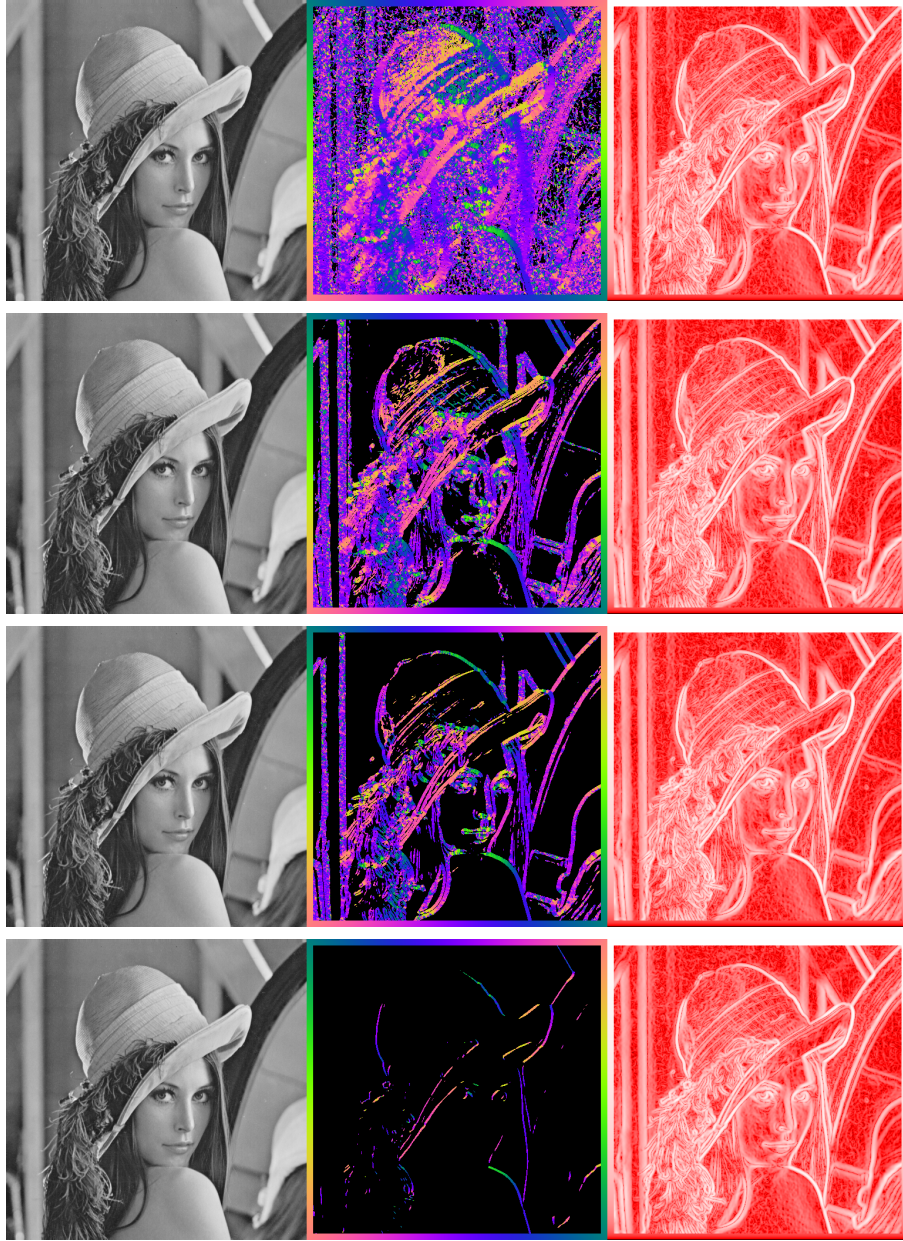


Figure 3.9: A series of results, with a single scale  $\sigma_1 = 1.5$  pixels and  $k = 4$ , showing the effect of changing the threshold of the gating on the gradient measure given by Equation 3.1. (top) The threshold is 0.1 of the mean gradient measure over the stimulus. (middle top) 0.5 of mean. (middle bottom) The mean value. (bottom) 5.0 times the mean value. Changing this parameter varies the amount of orientation information on the clearest lines in the orientation output.

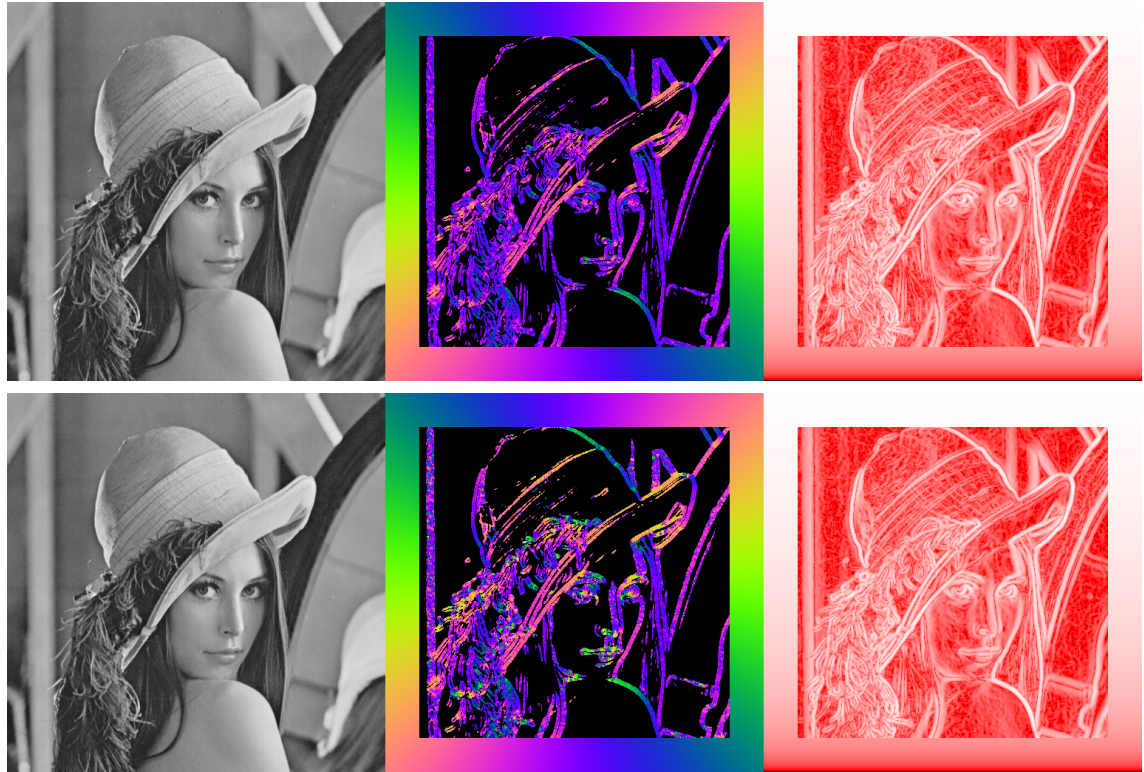


Figure 3.10: The effect of favouring oblique measures on a natural image as seen in this figure. Favours oblique measures reduces the problem caused by detail-alignment with a column. [top] Without favouring oblique measures. [bottom] Favours oblique measures. It seems that the orientation map from the bottom row includes a wider range of colours, as the detail in the scene is randomly aligned one would expect the range of colours in the orientation map to be equivalent to the range available. This result suggests that a directional bias is introduced when oblique measures are not favoured. This demonstrates that favouring oblique measures gives better results for natural images.

### 3.3 Machine vision comparison

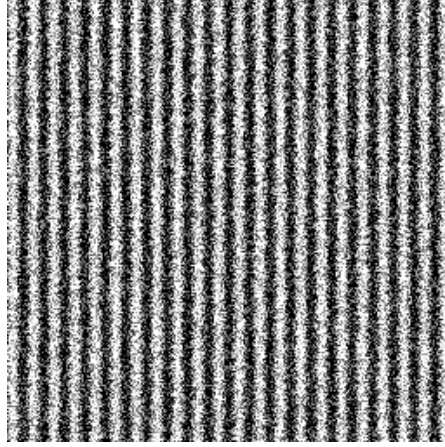


Figure 3.11: An example stimulus used in an experiment to compare noise susceptibility between the model presented in this thesis and that presented by Feng and Milanfar. The variance of the additive Gaussian noise in this stimulus is 0.58, and the angle is arbitrarily set to vertical (zero).

As machine vision models are designed for use on natural scenes, their noise performance is of particular importance. Although machine vision techniques are not restricted by biological plausibility, it is still interesting to compare results between Computer Vision and Biological models of vision, ie. GROM. This section shows a comparison between the model and a machine vision technique’s susceptibility to noise. Feng and Milanfar tested their model’s (‘Computer vision approaches to measuring orientation’ (see page 37)) susceptibility to noise by performing an additive-noise experiment on a sine-wave stimulus [Feng and Milanfar, 2002]. This model was chosen as it is relatively recent and covers an aspect of the performance of GROM otherwise undiscussed — susceptibility to noise.

The test stimulus is based on a zero-mean noise term added to a two-dimensional sine-wave stimulus. The noise distribution of variance,  $\sigma$ , means the probability of a noise value  $N$  is given by:

$$P_{\sigma}(N) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-N^2/2\sigma^2} \quad (3.3)$$

This generates a perturbation for each pixel  $N_{x,y}$ , giving a resulting stimulus  $f(x, y)$  of:

$$f(x, y) = N_{x,y} + \sin(2\pi S(x \sin \theta + y \cos \theta)) \quad (3.4)$$

where  $S$  is the spatial frequency of the stimulus and  $\theta$  its orientation. The spatial frequency was set to 12.04cpd, based on empirical estimation from stimuli presented in figures by Feng and Milanfar [Feng and Milanfar, 2002]. The orientation is set to vertical for simplicity; this does not influence the result as the stimulus orientation is arbitrary. An example stimulus is shown in Figure 3.11 (see page 74), where the variance  $\sigma$  is 0.58, the maximum variance used in Feng and Milanfar's experiment and here.

In generating results for comparison, a stimulus size of 256 pixels square was used to ensure the resulting orientation map contained a good population of pixels. In the absence of noise, the orientation map is zero in value for each pixel, the average error in angle is the sum of the absolute values of the orientation map, divided by the number of pixels, or:

$$\text{Average Error Angle} = \frac{\sum_{y=0}^{(256-\text{border})} \sum_{x=0}^{(256-\text{border})} |O_{x,y}|}{(256 - \text{border})^2} \quad (3.5)$$

A graph of the results is shown superimposed with results of Feng and Milanfar in Figure 3.12 (see page 76). As the comparison is with a machine vision technique, the model need not be restricted to parameters dictated by biology. Hence, the model's susceptibility to noise was reduced by increasing the standard deviation of the DOGs.

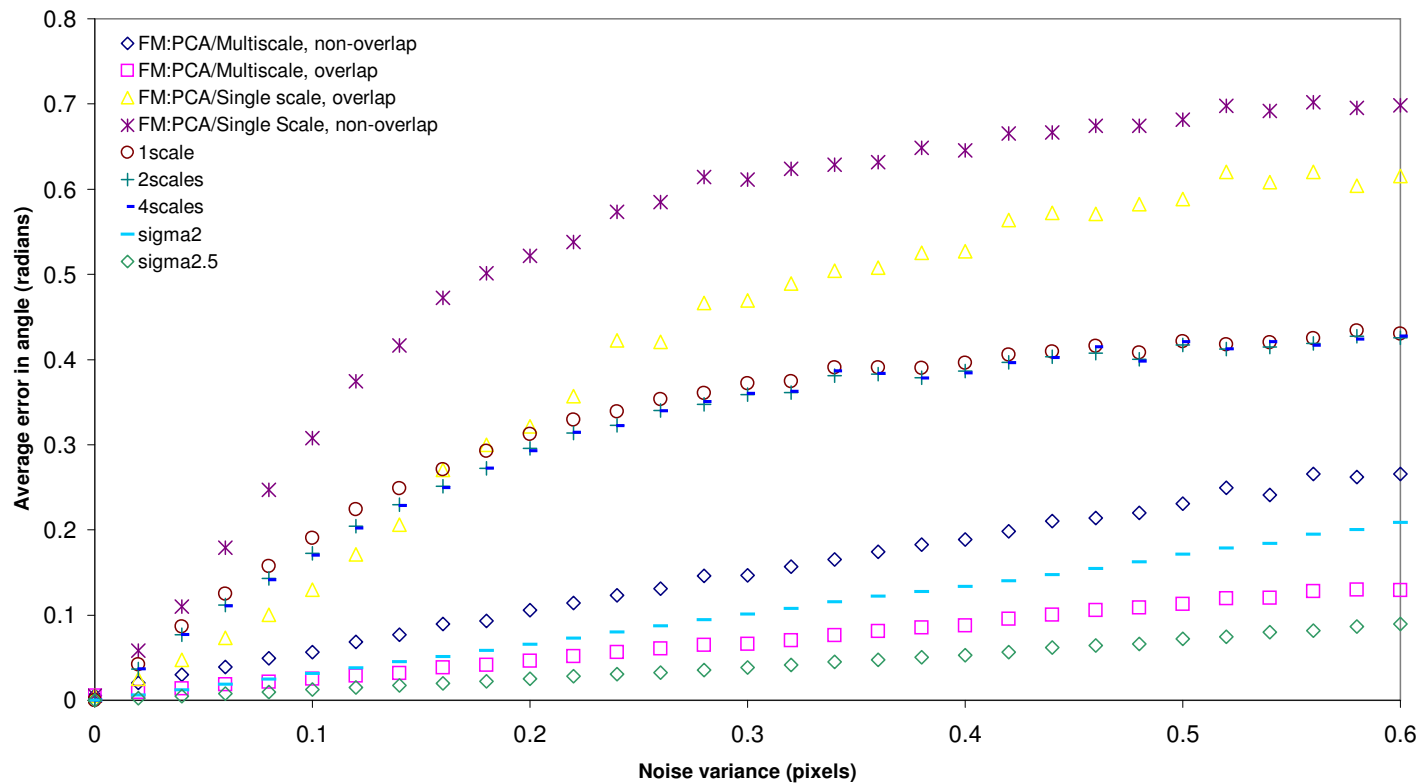


Figure 3.12: A graph showing the results of an additive noise experiment for the model with the results of Feng and Milanfar superimposed [Feng and Milanfar, 2002]. Zero-mean noise was added to a simple sine-wave stimulus. The variance of the noise was increased as the mean error in angle over the stimulus was measured. The comparison results were generously supplied by the original authors.

## Chapter 4

# GROM applied to contrast-modulated gratings

This chapter concerns the performance of GROM when applied to contrast-modulated gratings. Results are compared with psychophysical results for perceived orientation reported in experiments performed by McOwan and Johnston [McOwan and Johnston, 1996]. GROM's performance is shown to correspond to misperception found psychophysically, whereas the simple Fourier-energy model (SFEM) corresponds more closely with actual stimulus orientation. Unless otherwise stated, the model parameters used were as those stated in Table 3.2 (see page 67). Where multiple scales have been used, the standard deviation increased by a factor of two between scales.

### 4.1 Introduction

This section shows the performance of the model on arguably one of the most basic of second-order stimuli, the contrast-modulated (CM) grating. As described earlier, second-order stimuli are defined as those which have a mean grey luminance over the entire stimulus, where detail is expressed as spatial variation in the local variance from mean luminance [Chubb and Sperling, 1988].

In an experiment by McOwan and Johnston, subjects were required to compare the perceived orientation of the modulation of the grey bars in a CM grating to the orientation of a simple sine-wave grating, held at a fixed  $45^\circ$  angle (similar to Figure 4.1 (see page 79)). The point of subjective equality was found by determining at what modulation orientation,  $f_m$ , the

probability of estimating the tilt as above or below the orientation of the reference grating fell to chance. This task necessitates that subjects estimated the orientation of the entire grey bar.

The apparent orientation of the grey modulation envelope of a CM grating was found to be misperceived depending on the orientation and spatial frequency of the carrier. McOwan and Johnston measured the extent of the misperception as these parameters were varied [McOwan and Johnston, 1996]. The main two findings were that the misperception of the orientation of the grey bar was:

- at the lowest carrier spatial frequency examined (0.63cpd)
- when the carrier and envelope were at a relative orientation of  $45^\circ$ .

In both these cases, the grey bar appeared oriented towards the carrier orientation. Hence for the psychophysical investigation performed by Johnston and McOwan, the *maximal illusory case* is where  $f_m = 0.84cpd$ ,  $\theta_m = 45^\circ$ ,  $f_c = 0.63cpd$  and  $\theta_c = 0^\circ$ .

Equation 4.1 gives the equation of the surface created by sinusoidally modulating the contrast of an oriented sine grating (the carrier). The parameters used are the modulation spatial frequency  $f_m$  angled at  $\theta_m$  and the carrier spatial frequency  $f_c$  angled at  $\theta_c$ .

$$I(x, y) = (1 + \sin(2\pi f_m (\sin(\theta_m)x + \cos(\theta_m)y))) \sin(2\pi f_c (\sin(\theta_c)x + \cos(\theta_c)y)) \quad (4.1)$$

## 4.2 GROM orientation map results

Figure 4.1 (see page 79) shows two stimuli and their resulting orientation maps. The upper is a simple sine-wave grating, oriented at the same angle as the grey bar modulation in the CM grating below ( $\theta_{\text{simplesine}} = \theta_m$ ). More information on second-order stimuli can be found in ‘Second-order spatial stimuli’ (see page 30).

Figure 4.3 (see page 81) shows the result of changing various model parameters on the orientation and gradient maps while the CM grating is in the maximum illusory case. That is with  $f_m = 0.84cpd$ ,  $\theta_m = 45^\circ$ ,  $f_c = 0.63cpd$  and  $\theta_c = 0^\circ$ . The carrier is at  $45^\circ$  to the envelope. The orientation map changes markedly between the  $k = 1$  and  $k = 4$  cases, but not with the variation of the number of scales,  $s$ . Consequently, the  $k = 4$  and  $s = 1$  case is examined further.



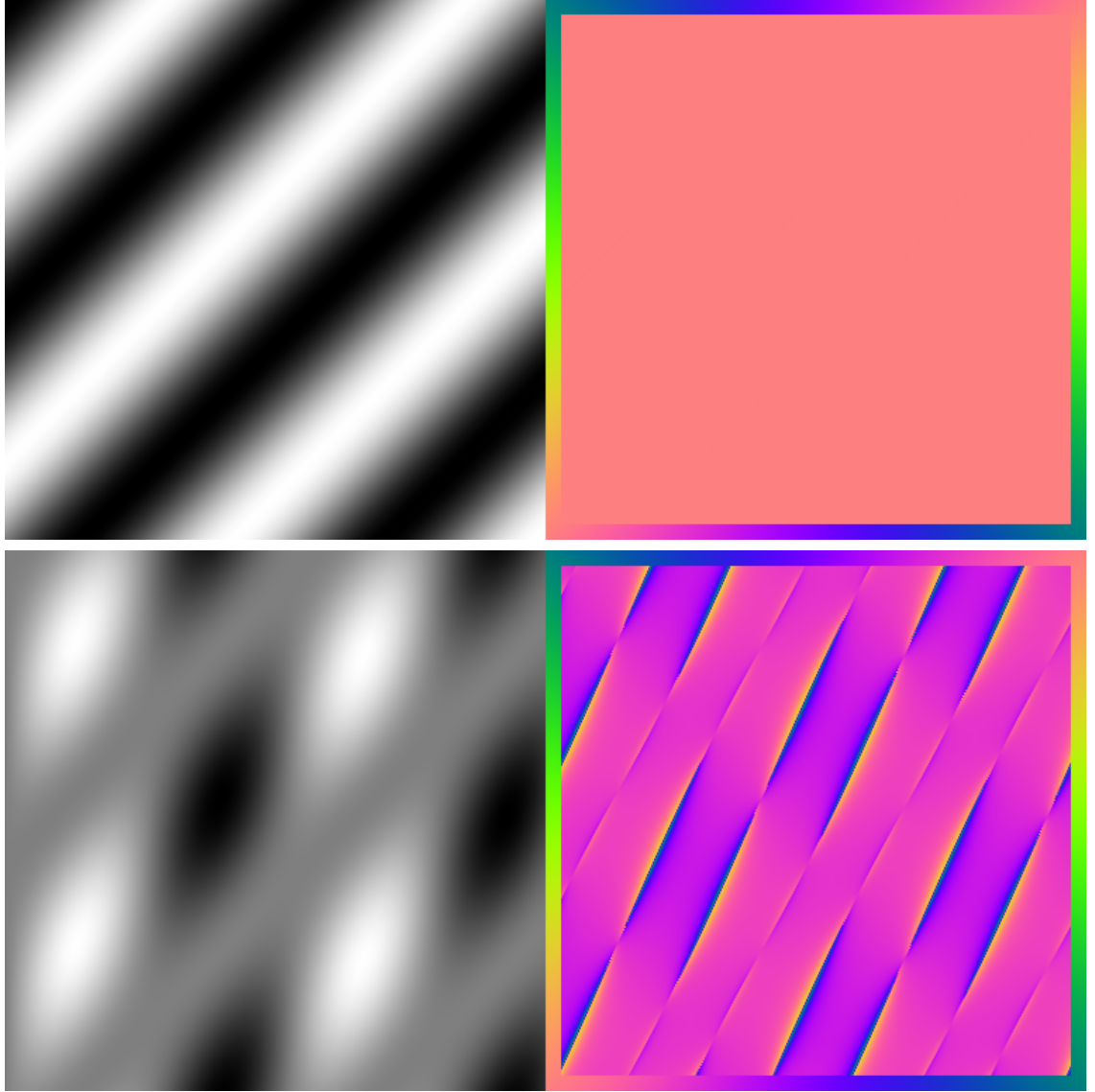


Figure 4.1: This figure shows two 384 pixels square stimuli with corresponding orientation maps, (top) a simple sine-wave grating with a spatial frequency of  $f_c = 0.63cpd$ , showing black and white bars oriented at  $\theta_c = 45^\circ$ . (bottom) a CM grating in the maximal illusory case, whose modulation envelope (seen as grey bars) is at the same orientation as the stimulus above, with  $f_m = 0.84cpd$ ,  $\theta_m = 45^\circ$ ,  $f_c = 0.63cpd$  and  $\theta_c = 0^\circ$ . The grey bars in the CM grating are perceived as tilted towards the vertical carrier vertical. Examining the orientation map for the simple sine-wave stimulus (top) reveals a uniform orientation map, whereas the CM grating (bottom) has a more complex spatial structure warranting further analysis. The aim is to use this model, which measures local orientation, to show how this is affecting the misperception of the global orientation of the grey bars.



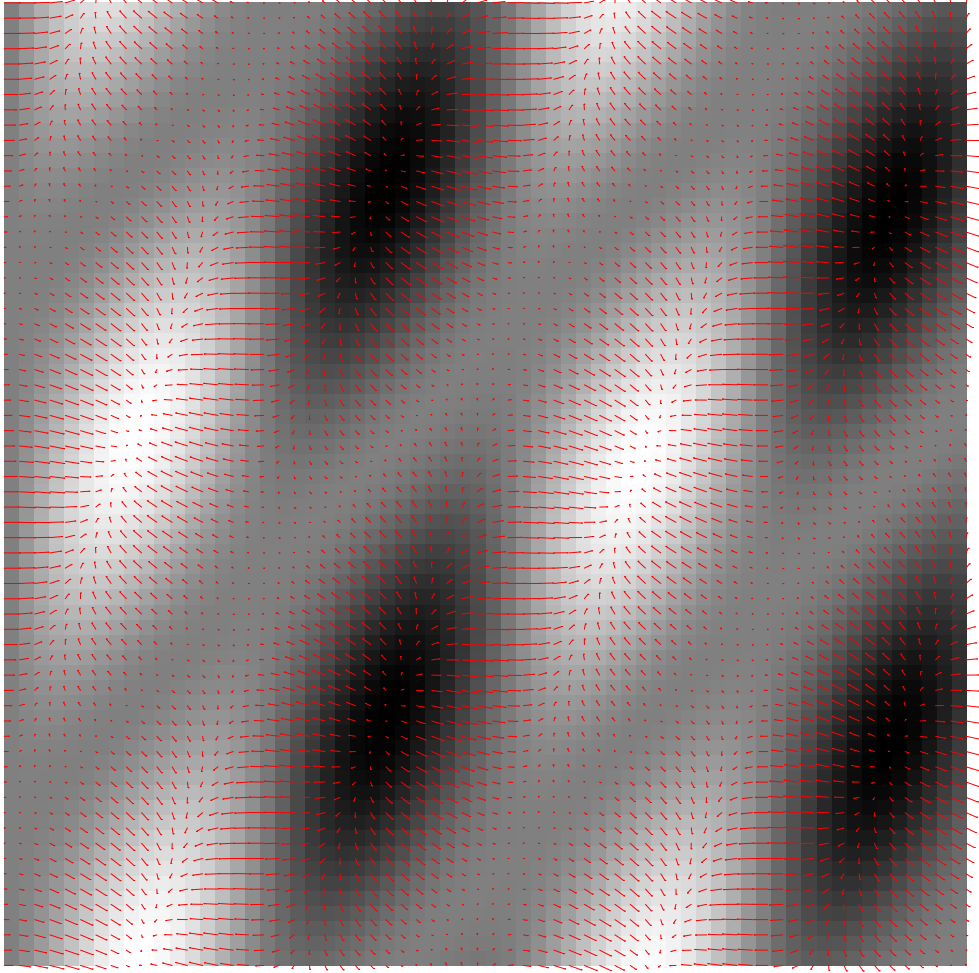


Figure 4.2: This Figure shows a 384pixels square CM grating in the maximal illusory case ( $f_m = 0.84cpd$ ,  $\theta_m = 45^\circ$ ,  $f_c = 0.63cpd$  and  $\theta_c = 0^\circ$ ) with a red overlay showing the direction of the maximum change in luminance value. The overlay is calculated by applying the ‘grad operator’ ( $\nabla$ ) in Mathematica to the analytical function describing the surface. The size of the red elements indicates the magnitude of the gradient. One disadvantage of this type of gradient indicator — spaced directional bars — is that they are sparse, and could therefore hide spurious results in the space between the centre of indicators.

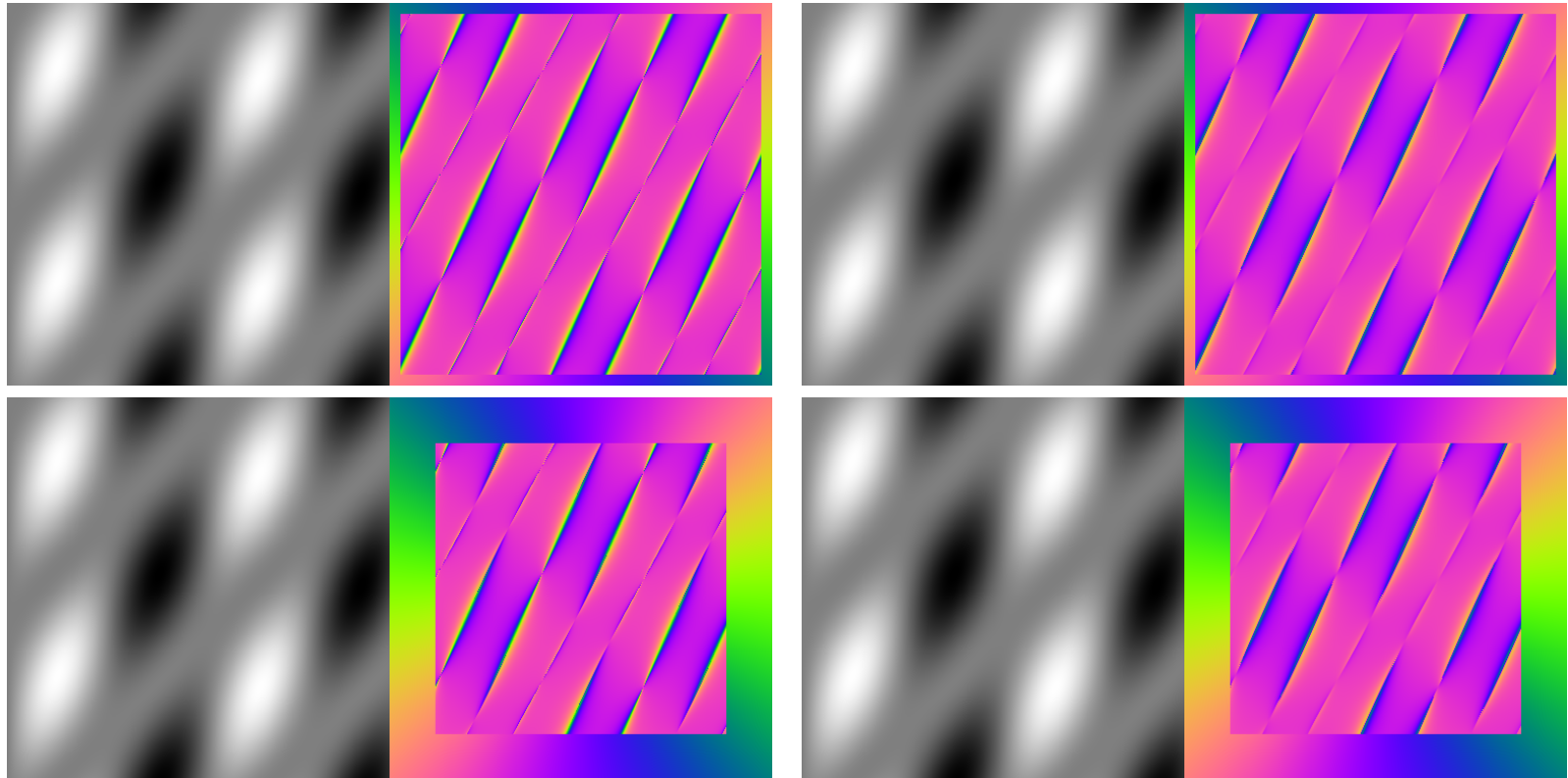


Figure 4.3: The results above are intended to show the difference in the orientation output with the variation of model parameters for the CM grating in the maximum illusory case ( $f_m = 0.84cpd$ ,  $\theta_m = 45^\circ$ ,  $f_c = 0.63cpd$  and  $\theta_c = 0^\circ$ ), where the carrier is at  $45^\circ$  to the envelope. In each, the left hand pane is the same 384pixels square stimulus and right is the orientation output. (top left) One scale and a simple ratio,  $k = 1$ . (top right) one scale and  $k = 4$ . (bottom left) Four scales,  $k = 1$ . (bottom right) Four scales,  $k = 4$ .

### 4.3 Population based analysis

As evidenced in Figure 4.1 (see page 79), further analysis of the orientation map is required in order to qualify the result of this model in relation to the CM grating, as the orientation map has detailed spatial structure. The model measures local orientation whereas the result of psychophysical investigation into the ‘tilt’ of the envelope in a CM grating reveals a misperception of the entirety of the grey bars [McOwan and Johnston, 1996].

The general experimental paradigm is as follows; the low-contrast regions of the CM grating are misperceived [McOwan and Johnston, 1996, Smith *et al.*, 2001]; hence only the local orientation measures from these regions need be examined — shown as non-black in Figure 4.4 (see page 84). The orientation measure is local, i.e. the model does not prescribe how a local orientation measure is grouped in order to arrive at a high-level percept of orientation of an oriented structure<sup>1</sup>. The local orientation information in the region of the grey bars is put into ‘angular bins’, which contain the population of pixels having the orientation of that bin, within a certain tolerance defined by the number of bins. A graph of the results of this analysis can be found in Figure 4.5 (see page 85).

The stimulus used was a 384 pixels square stimuli were used to allow for enough of the low-contrast, i.e. grey bar, region to be processed in the lowest spatial frequency case. As the psychophysics indicate, the largest illusion is observed at the lower spatial frequency of the carrier, hence  $f_c = 0.63cpd$  stimuli are used. 37 angular bins were used so that bins were centred every  $10^\circ$  over the  $180^\circ$  gamut.

This experiment relies on selecting the ‘low contrast regions’ — the grey bars, the stimuli and orientation map are masked according to where the envelope:

$$env = (1 + \sin(2\pi f_m(\sin(\theta_m)x + \cos(\theta_m)y))) < 0.87 \quad (4.2)$$

Equation 4.2 shows the envelope (grey bar) part of the equation for a contrast modulated grating, see Equation 4.1, with the tolerance used in this experiment, 0.87. This tolerance value was adjusted, by eye, to create a mask that only selected the grey low-contrast regions of the envelope. Orientation information outside these regions is ignored as it is assumed that it does not contribute to the illusory nature of the grey bars, as per the experimental paradigm assumed above.

Figure 4.4 (see page 84) shows the results of the population based analysis of the orientation map for a range of carrier conditions. Peaks can be seen at orientations displaced from the veridical envelope orientation of  $+45^\circ$ , towards the vertical at  $0^\circ$  at the carrier orientations,  $f_c = -22.5^\circ$ ,

---

<sup>1</sup>defined as an identifiable structure in an image or stimulus such as the grey bars

$f_c = 0^\circ$  and  $f_c = 22.5^\circ$ . These represent tilts in the same sense as those found psychophysically. These results are summarised in Table 4.1 (see page 83), where the significance of the peaks is also examined. This shows that a significant portion of the pixels lie in an illusory range for the illusory carrier conditions mentioned above.

The same experiment was performed for the SFEM, as seen in Figure 4.6 (see page 86). They show a peak at the veridical carrier orientation rather than suggesting any illusory tilt.

Carrier orientation (deg)	Peak (deg)	Avg. clearance (pixels)	Avg. clearance (%)
-45	45	4987	8.2
-22.5	30	5770.5	9.6
0	30	22993	38.3
22.5	35	28655.5	47.6
45	45	60094	100

Table 4.1: This table shows a summary of the results from the graph in Figure 4.5 (see page 85). The peak the orientation-pixel population for each carrier orientation is shown. The average clearance for each peak is also included, defined as the average of the difference between the peak point and the two neighbouring points, and is expressed in pixels and as a percentage of the non-masked pixel population. This measure is intended to be an indicator of the significance of the peaks.

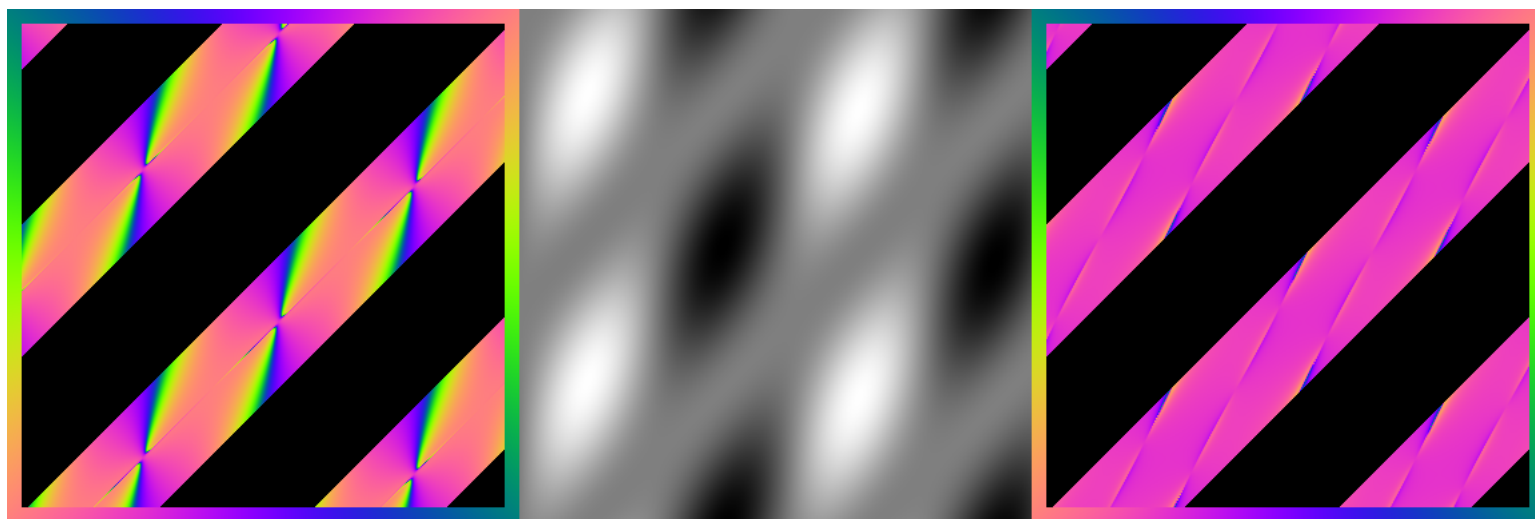


Figure 4.4: This figure shows the orientation map output from the SFEM (left) — from a CM grating with grey bars at  $45^\circ$  — the maximal illusory case (centre). Also shown is the corresponding orientation map output from GROM (right) using one scale and  $k = 4$ . The low-contrast regions in the stimulus (centre), seen as diagonal grey bars, are the only unmasked regions in the orientation maps (left and right). Orientation information from each pixel in these regions is ‘binned’ to produce the following graphs; Figure 4.5 (see page 85) and Figure 4.6 (see page 86)

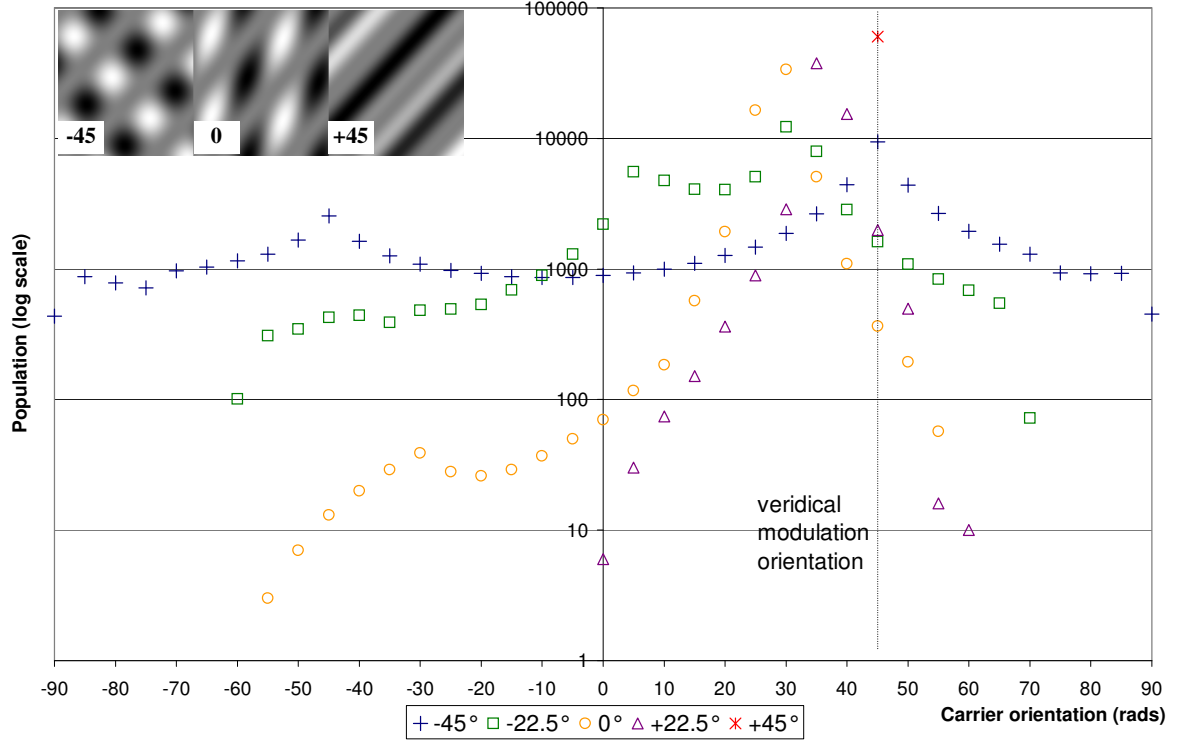


Figure 4.5: This graph (top) shows the population of bins against an orientation range for pixels in the low contrast region of a CM grating as shown in Figure 4.4 (see page 84) (right), where the modulation frequency,  $f_m = 0.84cpd$  and the carrier frequency,  $f_c = 0.63cpd$ . The gradient model parameters used were; one scale and  $k = 4$ , as it was found that increasing the number of ratios had no obvious effect on the results. This graph shows the five angles of carrier used in psychophysics by McOwan and Johnston [McOwan and Johnston, 1996]. Where the carrier is aligned parallel to the modulation envelope at  $+45^\circ$  (inset right), the only data point is for the angular bin that straddles the angle of the carrier, centred at  $+45^\circ$ . For the orthogonal case —  $-45^\circ$  (inset left) — there are two main peaks at  $+45^\circ$  and  $-45^\circ$ , the orientations of the carrier and modulation respectively, though the modulation peak is slightly higher.  $+22.5^\circ$  has a well populated peak centred at  $+37.5^\circ$ , suggesting a misperception of  $7.5^\circ$  towards the vertical. Interestingly the  $-22.5^\circ$  case, has a peak much further from the veridical orientation but is far less populated than the  $+22.5^\circ$  peak. Most importantly, the case with the vertical carrier of  $0^\circ$  (inset middle) has two ‘peaks’ with an average of  $+26.25^\circ$ , suggesting a tilt toward the vertical of  $18.75^\circ$ , though, as the  $-22^\circ$  case suggests, less populated peaks incur less of a misperception than their centres imply.

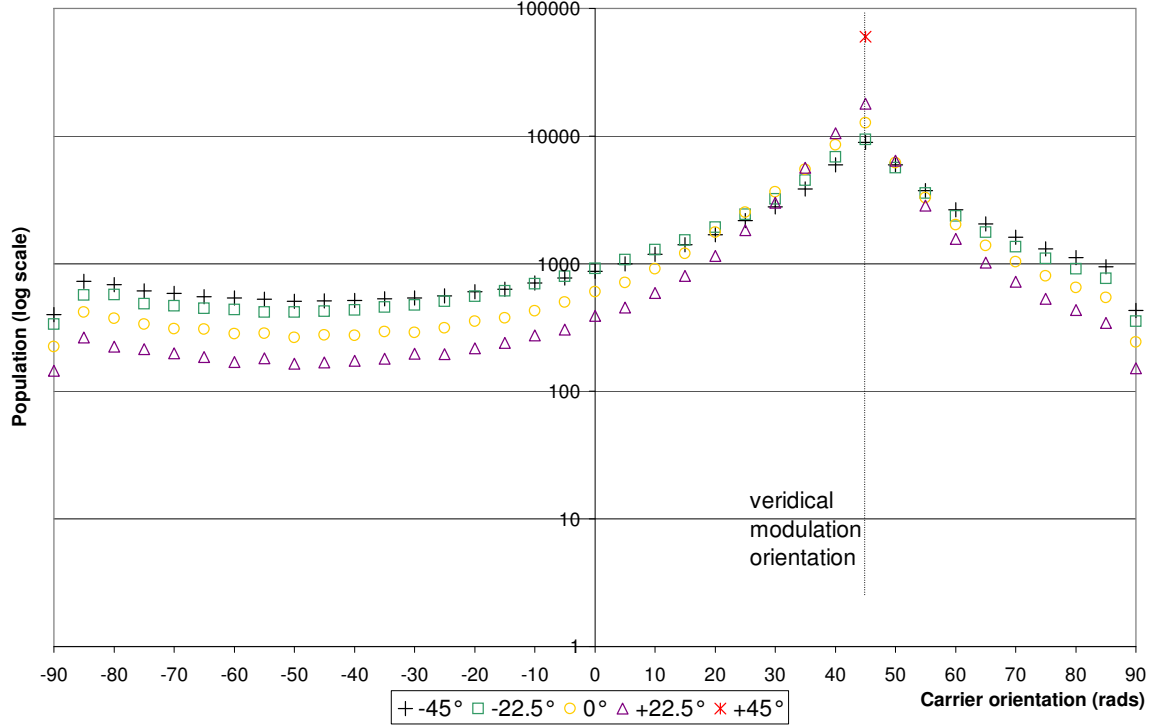


Figure 4.6: The corresponding graph to Figure 4.5 (see page 85) but calculated with a SFEM, see Figure 4.4 (see page 84) (left). This graph shows the population of bins against an orientation range for pixels in the low contrast region of a CM grating, where the modulation frequency,  $f_m = 0.84cpd$  and the carrier frequency,  $f_c = 0.63cpd$ . This graph shows a population peak at the veridical angle of the grey bars,  $+45^\circ$ , rather than indicating the presence of other orientations that could lead to the systematic illusory misperception found psychophysically [McOwan and Johnston, 1996] and with GROM as seen in Figure 4.5 (see page 85).

Angle of carrier (degrees)	McOwan and Johnston re- sult	Position of peak from GROM	relative height of peak
$-45^\circ$	$+.16^\circ$	$0^\circ / 90^\circ$	medium / medium
$-22.5^\circ$	$-3.4^\circ$	$-15^\circ$	low
$0^\circ$	$-6.7^\circ$	$-20^\circ$	high
$+22.5^\circ$	$-2.6^\circ$	$-7.5^\circ$	high
$+45^\circ$	$-0.16^\circ$	$0^\circ$	high

Table 4.2: This table shows a summary of the results of the psychophysics of McOwan and Johnston [McOwan and Johnston, 1996] (estimated from the graphs) and GROM over the range of orientations. The psychophysical results have been adjusted so that PSE orientation has been changed to the same sense as the illusion (counter-clockwise is negative). Although it is difficult to deduce an absolute value for the size of the tilt implied by GROM, the trend is in the correct direction — towards vertical, and the relative size of the tilt between carrier orientations is also in agreement with the psychophysics.

## 4.4 Conclusions

The results shown in Figure 4.5 (see page 85) make use of the following two assumptions:

- that a peak in the population of pixels at a certain orientation indicates the likely orientation of the grey bar
- that the stronger the peak relative to the population of other ‘orientation bins’, the stronger this orientation appears relative to the rest.

Given the two assumptions above, the results summarised in Figure 4.5 (see page 85) are in agreement with the psychophysical investigation of the same stimulus. Table 4.2 (see page 87) shows a comparison of the results between the psychophysical investigation by McOwan and Johnston and the experimental results presented in Figure 4.5 (see page 85). This table demonstrates that the trends in the misperception of the low-contrast regions found psychophysically are also shown in this experiment.



## Chapter 5

# GROM applied to the Café Wall illusion

This chapter shows results from GROM applied to the Café Wall illusion. It is shown that GROM performance when applied to the Café Wall compares favourably with the psychophysical findings of Gregory and Heard [Gregory and Heard, 1979].

### 5.1 Introduction

Gregory and Heard first described the Café Wall illusion in 1979 [Gregory and Heard, 1979], although they attributed the original earlier ‘discovery’ to another lab member. The pattern was noticed on the outside wall of a local café, constructed of black and white half-offset bricks horizontally interleaved with a grey line — the mortar line, as in Figure 5.1 (see page 89) (top). Alternate mortar lines are seen to slant in opposite directions, appearing either slightly clockwise from horizontal, or slightly anti-clockwise. This makes the rows of bricks appear wedge like. Gregory and Heard measured the slant as being up to  $3^\circ$  horizontal in the maximal case. The illusion in the Münsterberg, where the mortar lines are the same luminance as the black tiles (see Figure 5.1 (see page 89), bottom), was measured as  $1^\circ - -2^\circ$  [Gregory and Heard, 1979].

Gregory and Heard suggested a system of ‘border locking’ as a possible mechanism explaining the illusion. Gregory and Heard’s assumption was that *spatial registration*<sup>1</sup> is normally maintained using border-luminance differences for locking features together — so that features are normally aligned at borders. Hence, they argue, systematic distortions are expected where locking occurs

---

<sup>1</sup>assumed to mean the alignment of the border of the tiles and mortar

across narrow gaps or lines of neutral luminance.

However, they concede that this proposition merely gives a possible ‘modus operandi’ of the mechanism, rather than a complete functional explanation of the illusion. This, they argue, requires greater knowledge of the physiology behind it.

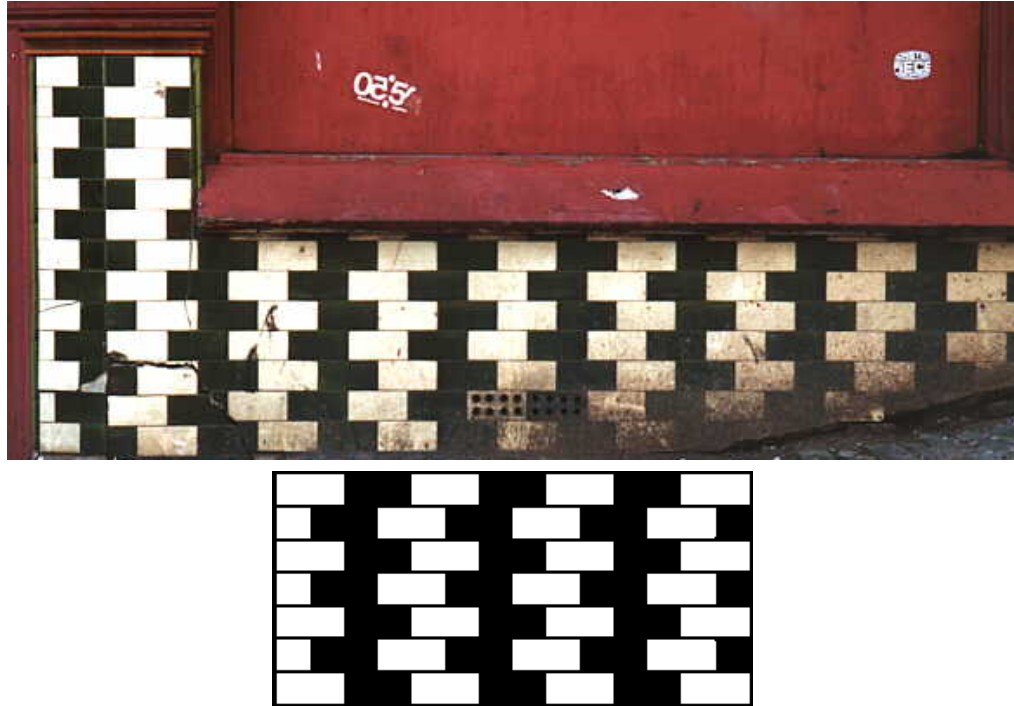


Figure 5.1: (top) A modern photograph showing the original Café Wall. The illusion is still clear; the rows of bricks appear wedge shaped despite the fact they’re horizontal [Ray]. (bottom) A similarly scaled Münsterberg figure, having black rather than grey mortar lines. Taken from [Gregory and Heard, 1979], this shows the same tendency tilts as the Café Wall illusion but in the black lines, though the effect is not as strong.

McCourt offered an alternate explanation of the Café Wall illusion [McCourt, 1983] by calling on a previously described theory concerning ‘brightness induction’ [McCourt, 1982]. In simple terms, the grey mortar lines of the pattern are ‘induced’ to the opposite colour that they border, for example, where the mortar line is adjacent to a black brick, it appears white; and adjacent to a white brick, black. This gives the impression of a ‘twisted cord’ in the mortar regions [Morgan and Moulden, 1986, Fraser, 1908], as seen in an example in Figure 5.2 (see page 90) (top).

In a more recent study, Kitaoka examined the Münsterberg and Café Wall illusions [Kitaoka, 1998], attributing their slants to so called *corner effects* (*c.f.*

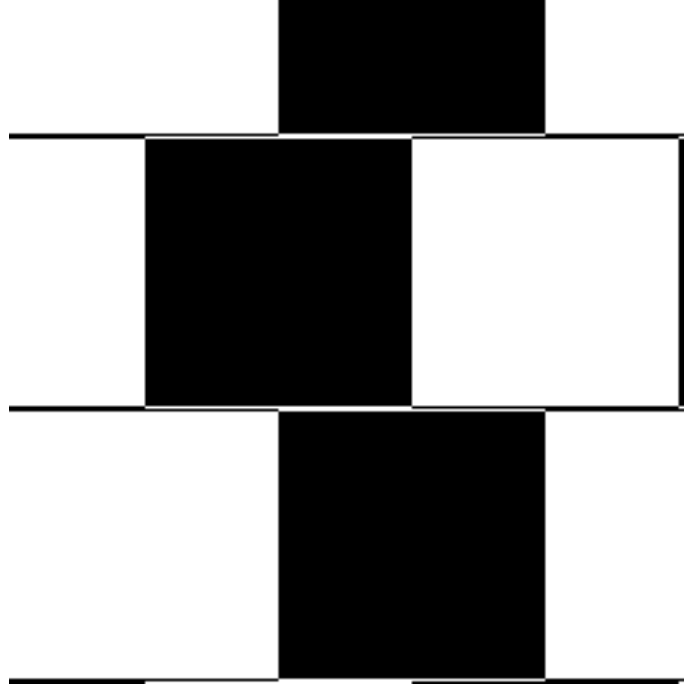


Figure 5.2: A Café Wall with the grey mortar lines replaced with the inducted colours, as per the theory of McCourt [McCourt, 1983]. This reportedly creates the effect of slanted lines, ‘twisted cords’ within the mortar regions.

[Moulden and Renshaw, 1979]). Kitaoka offers a slightly more explicit heuristic than that of Gregory and Heard or Moulden and Renshaw. He argues that the basic unit of these illusions is the corner: where two solid regions — one black, the other white — meet at two perpendicular edges which come together to one point, as in Figure 5.3 (see page 91) (top). Corners in this arrangement have been observed as being perceived as acute to perpendicular — less than  $90^\circ$  [Moulden and Renshaw, 1979]. Kitaoka uses this observation to account for the tilts in the black and grey lines in the Münsterberg figure and Café Wall illusion respectively. A schematic taken from Kitaoka shows this argument in (top) Figure 5.3 (see page 91).

Kitaoka offers another heuristic related to the relative luminance of the lines and tiles; the magnitude of illusion at line elements produced by corner effects decreases according to the difference in luminance between the mortar line and corners.

This additional assertion can be used to account for the difference in the magnitude of the illusory tilts in the Münsterberg figure and the Café Wall illusion. In the Café Wall this implies a reduction in the illusory effect of the black corners — which all have grey horizontal borders. This luminance heuristic also provides for an additional illusory effect from the white corners

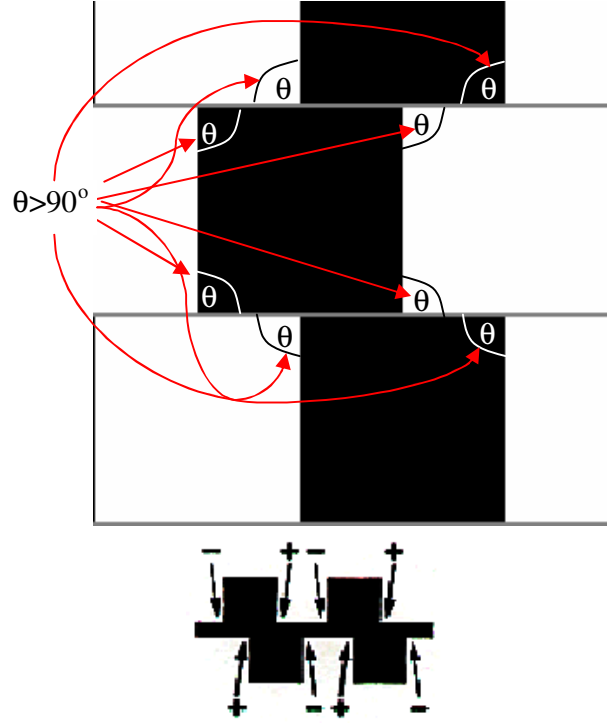


Figure 5.3: (top) A Café Wall stimulus showing corners surround by their opposite colour as being perceived acutely (hence the inset ' $\theta > 90^\circ$ '), as per Kitaoka [Kitaoka, 1998]. This causes an overall tilt in the continuous horizontal line going through the brick rows. (bottom) A Münsterberg figure with a thick centre line. Kitaoka argues that the tilt caused by the corner effects from both corners of each white cell (highlighted with '+' and '-') counteract each other. This results in no perceived overall tilt of the centre line if the interleaving black line is too thick.

(vertical side black, horizontal side grey).

Half of the white corners are bordered by white-grey-white, the other half by white-grey-black. The white-grey-white and white-grey-black corners (lying on the same mortar line) are both perceived as acute, causing contradictory tilt in the mortar as they are at horizontally opposing corners of the square. However, according to Kitaoka, the white-grey-black has a stronger effect than the white-grey-white. Hence the acuteness caused by this corner dominates, causing an overall line tilt.

The preceding argument can also be applied to the Münsterberg figure: the white-black-white and white-black corners are more equal and therefore counteract each other more — causing an overall weaker illusory tilt in the continuous interleaving black line. Observations from Gregory and Heard support the corner effect hypothesis as they show that increasing the thickness and

darkening the mortar line decreases the strength of the illusory tilts in the Café Wall [Gregory and Heard, 1979].

Kitaoka developed the ‘illusion of shifted gradations’ based on the previous *corner effect* assumptions. A typical ‘shifted gradations’ figure is shown in Figure 5.4 (see page 92) (c.f. [Morgan and Moulden, 1986]). Slight differences in phase between two parallel sine-wave based luminance-stepped lines are used to create an illusion of tilt based on the same corner effects, but with a range of shades in discrete steps of luminance.

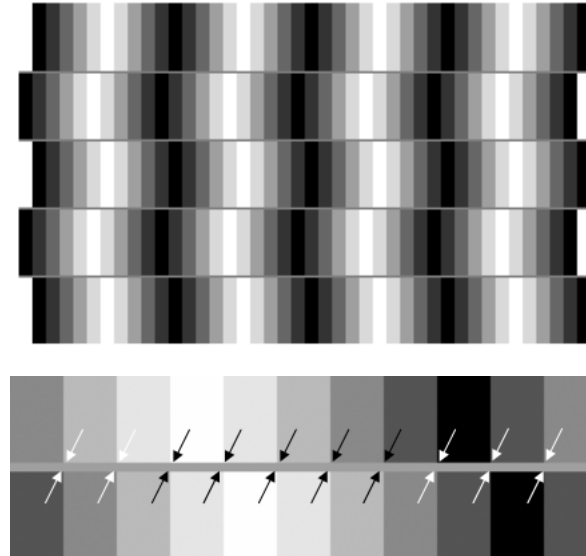


Figure 5.4: (top) An ‘illusion of shifted gradations’ taken from Kitaoka [Kitaoka, 1998], this shows the same wedge-like horizontal patterns induced by a misperception of the border lines, which Kitaoka claims are caused by successive *corner effects*. (bottom) The accompanying explanatory diagram, also taken from Kitaoka, showing a two line subset of the above, with arrows indicating the acutely perceived corners.

The ‘border locking’ of Gregory and Heard, the ‘brightness induction’ of McCourt and the ‘corner effect’ of Kitaoka (c.f. from Morgan and Moulden) all rely on an empirical assessment of the stimulus to account for it. They are each rules or heuristics specific to one or a few illusions, rather than being an attempt to form an explicit model based explanation of a set of illusions. None of them are the basis for a specific theory concerning the general group of geometric illusions, of which the Café Wall and Münsterberg are members. Of the three potential explanatory accounts, only Kitaoka goes on to use his heuristics to form a prediction — Kitaoka used the theory that the corner effect causes the misperception of the mortar line by adding similar offset corners to regular patterns, as seen in Figure 5.5 (see page 93). However, no

psychophysical investigation of this stimulus was performed.

In contrast to the models mentioned above, GROM was not developed to explain the Café Wall specifically, but was designed as a robust gradient based model of local orientation measure in the human visual system. The Café Wall was chosen as a stimulus as it is a much studied geometrical illusion of orientation that has already been psychophysically investigated. Consequently, data exists that can be compared to GROM's performance.

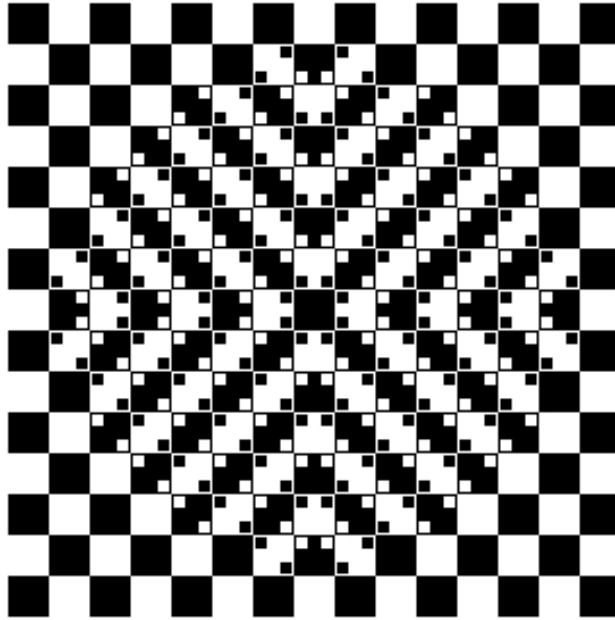


Figure 5.5: This figure shows an example pattern of Kitaoka that gives a three dimensional impression [Kitaoka, 1998]. The pattern is formed from black and white offset and overlapping squares. Kitaoka formed this based on predictions coming from the proposed *corner effect*, whereby corners composed of black and white are generally perceived acutely. Being able to predict the direction of slant led to a stimulus being created with slants diverging towards the centre of the figure — giving an illusory bulge. This is an interesting demonstration of local orientations being manipulated to create a three-dimensional effect.

In the later section entitled ‘GROM and the Café Wall ’ (see page 94), the gradient model presented here is applied to the Café Wall. In the section entitled ‘The effect of brick offset in the Café Wall ’ (see page 97), results of an experiment where mortar slant is examined as alternate rows of bricks are offset is presented. In the section ‘The effect of brick offset in the Café Wall ’ (see page 97), an attempt is made to formulate a ‘general rule’ for predicting the slants in similar illusions.

Parameter	Gregory and Heard value (degrees of visual angle)	Computational value (degrees of visual angle) [pixels]
Black/white square brick dimension	2.18°	0.78° [100]
Minimum mortar thickness	0.017°	0.016° [2]

Table 5.1: A table showing the value of the two main parameters used in the Café Wall stimulus, both in the original experiment and in the experiment presented here.

In the section entitled ‘Understanding GROM’s performance with the Café Wall ’ (see page 100), a simple-filter is rotated about various image points to show whether or not local gradient information is the basis of the slants. This hypothesis is shown to be invalid. By examining the number of derivatives used in the ratio, use of the simple filter is shown to not display any clear slants. Further discussion examines the nature of the illusory slants in terms of the underlying behaviour of the model at regions near the corners in the mortar.

## 5.2 GROM and the Café Wall

Gregory and Heard found that the alternate horizontal mortar lines of the Café Wall illusion were perceived as having a slant in opposite directions [Gregory and Heard, 1979]. This causes a wedge-like, trapezoid distortion of each row of bricks. Here, the gradient model for orientation is applied to a scaled Café Wall stimulus and the results in the mortar regions examined. Table 5.1 (see page 94) shows a list of the parameters used in the original experiments by Gregory and Heard and their analogue in the stimulus used in the computational experiment.

The table shows that the brick size used in this experiment is smaller than that used by Gregory and Heard. This value was used so that two alternate lines could be included in a reasonable stimulus size of 256 pixels square — equivalent to 2° of visual angle. An equivalent-scaled version would require overall dimensions of at least 640 pixels square to include two mortar lines. It is not thought that the difference in this parameter is particularly significant to the results. All the published explanations of the illusion indicate that the illusion is generated at the mortar line, the thickness of which has been shown psychophysically to be important to the strength of the tilt. Also, as the filter-support of the single scale model is 23 pixels, stimulus features separated by more than this distance will not interact.

The orientation model was applied (at one scale, where  $k = 4$ ) to a 256 pixels square Café Wall

stimulus with the brick offset at a maximum half-brick width between alternate rows. The resulting orientation map is masked to removed the output signals outside the mortar regions, as in Figure 5.6 (see page 95). Spots of different colour in the mainly green (indicating horizontal) lines indicate that the local orientation in the mortar is non-uniform. Further analysis of the distribution of orientations in the mortar lines is required in order to interpret this output more rigorously.

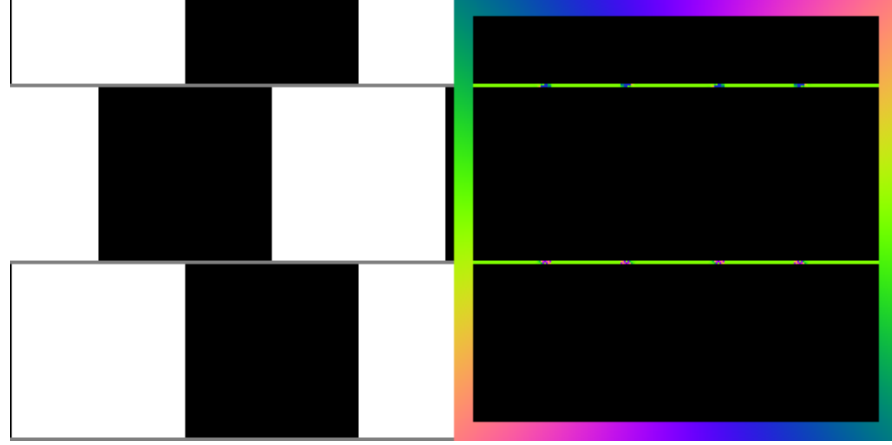


Figure 5.6: 256pixels square stimulus (left) and orientation map (right) at 1 scale,  $k = 4$ . The regions outside the mortar lines are masked out to black in order to highlight the region of interest — the orientation map of the mortar region. The local orientation measure in each line is clearly different at certain points, one a green (horizontal) bar with spots of blue (with the sense  $\backslash$ ), the other a green (horizontal) bar with spots of pink (with the sense  $/$ ).

Employing the same methodology to the ‘binning’ approach taken with the contrast-modulated grating experiments (in particular Figure 4.5 (see page 85)), the local orientation measures from the 2 pixel wide mortar line were counted between 180 discrete, equal, consecutive angular limits dividing the range:  $0^\circ$  to  $180^\circ$  ( $\pi$ ), where orientation increases clockwise from vertical.

The usual output orientation display was modified so that horizontal was in the centre of the range to make the display more intuitive.  $180^\circ$  was added to orientations less than  $0^\circ$  (anti-clockwise of vertical). The resulting data was plotted on a polar grid. The top mortar line is plotted on the left and the bottom on the right, in Figure 5.7 (see page 96). The red dots are oriented at the centre of each bin. The distance of the red dots from the origin is proportional to the  $\log$  of the population of pixels within the orientation limits of that bin. The results are summarised in the figure.

The results show bias in the directions expected, the top mortar has the sense  $\backslash$  and the bottom  $/$ . This is established via the relative populations in the angular bins above and below horizontal.



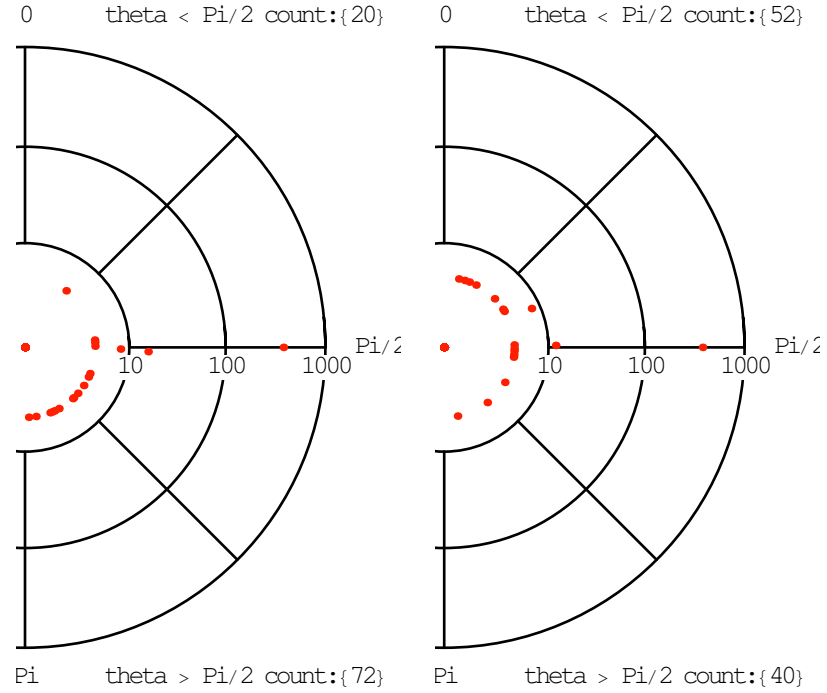


Figure 5.7: Two log-polar plots each showing the ‘binned’ data from the orientation output of one mortar line. The colour map of the same information is shown in Figure 5.6 (see page 95). In the plots, the data is presented in the sense of contour orientation of the original stimulus structure is preserved (i.e. vertical on the stimulus is vertical on the plot, horizontal on the stimulus is horizontal on the plot). (left-hand figure) The ‘binned’ data of the orientation map of the upper mortar line. (right-hand figure) The ‘binned’ data of the orientation map of the lower mortar line. As expected, in each plot, the most populated bin is horizontal ( $90^\circ$  or  $\frac{\pi}{2}$ ). The left-hand plot is more populated in the bins below horizontal (with the sense  $\backslash$ ). This is demonstrated by the inset counts, with 20 pixels above, versus 72 pixels below horizontal, where ‘theta’ is the angle of the centre of each bin. The right-hand plot is more populated in the bins above horizontal (with the sense  $/$ ). This is again born out by the counts inset, showing 52 pixels with orientation above horizontal and 40 pixels with orientation below. This finding corresponds with the sense of the slants found psychophysically by Gregory and Heard [Gregory and Heard, 1979].

The results in Figure 5.7 (see page 96) are in agreement with earlier psychophysical investigations [Gregory and Heard, 1979]. Indeed, informal observation of the stimuli also supports a correlation between GROM orientation distribution and the perceived slant. The ‘population-bin’<sup>2</sup> approach has been used because it is unclear how the human visual system groups local orientation measures to arrive at an overall perceived orientation of a feature such as a mortar line. Moreover, it can be argued that the ‘population-bin’ approach demonstrates the correct tendency of the perception of the mortar line slants without recourse to a specific model of how the local orientation information is grouped.

Figure 5.7 (see page 96) also shows a population of pixels with orientations above and below horizontal. One would expect the number of pixels with orientations above horizontal in the upper mortar line to be the same as the number below horizontal in the lower mortar line, as the slants reverse between the two. This equality in populations is not observed, an offset is apparent in the results. It is proposed that this offset is caused by a population of pixels that does not change with a reversal of slant. This can be tested by examining what happens to the aggregate above/below populations (pixels above horizontal minus pixels below horizontal) as the bricks are offset through a whole cycle as the slant reverses for the same line as the bricks offset. This is tested in the following section entitled ‘The effect of brick offset in the Café Wall ’ (see page 97).

### 5.3 The effect of brick offset in the Café Wall

This section presents an experiment where the horizontal brick-offset<sup>3</sup> of alternate rows in a Café Wall was changed over the entire range of possible values; two brick widths, one pixel towards the left at a time. This covers a state where the stimulus appears as alternating vertical black and white columns (save for the grey mortar lines), *brickoffset* = 0 pixels, to the intermediate state where the stimulus appears as a checkerboard (save for the grey mortar lines), *brickoffset* = 100 pixels, to the final state, the same as the initial condition, *brickoffset* = 200 pixels.

---

<sup>2</sup>defined as the process of examining the population distribution of orientation map pixels

<sup>3</sup>the offset of the bricks in alternate rows

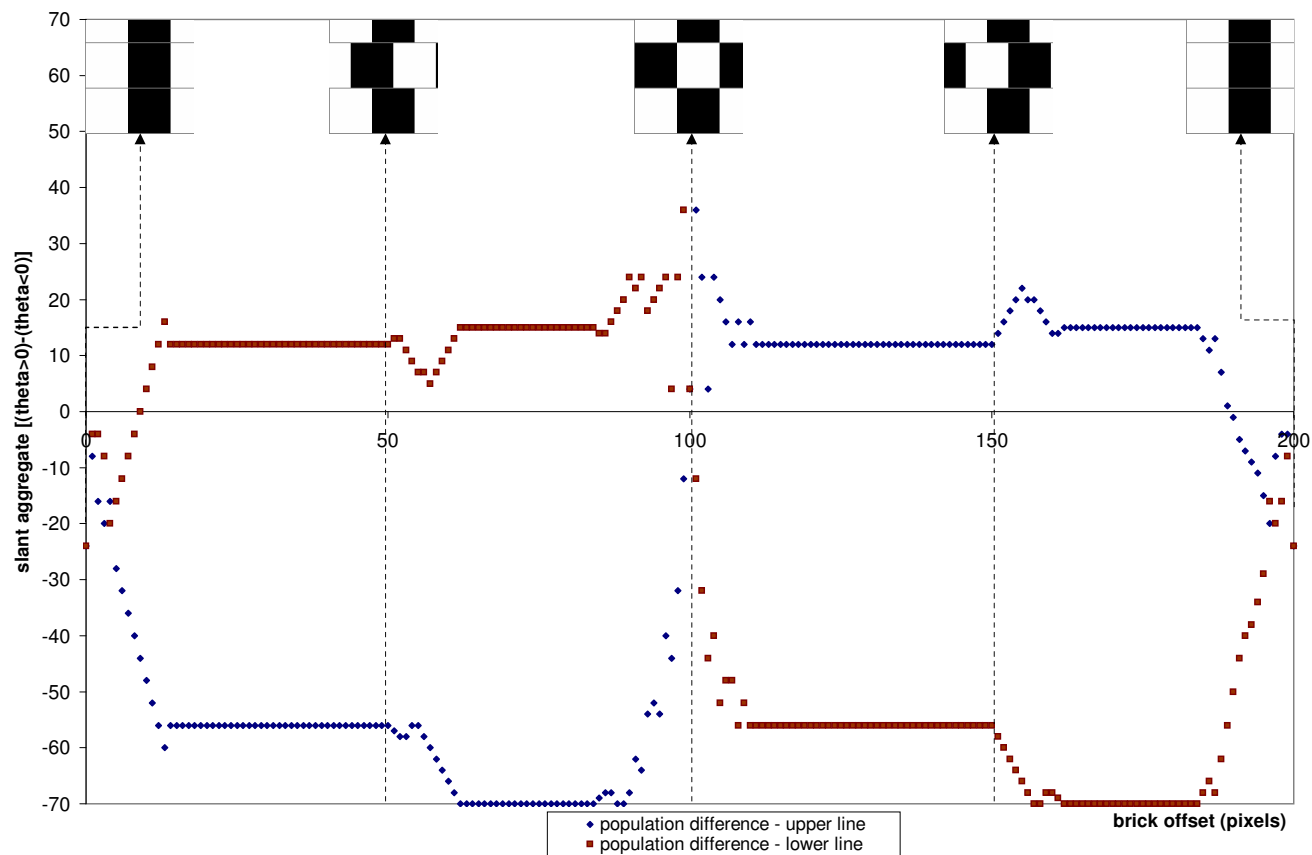


Figure 5.8: A graph showing the population aggregate — population above horizontal less the population below horizontal — for two alternate mortar lines, against brick-offset. The aggregate, and hence tendency, is shown to repeat over the entire brick offset cycle possible — 200 pixels. The aggregate for the two lines crosses at *brickoffset* = 100 pixels, the checkerboard case. Note that there is a vertical offset apparent in the aggregate; this implies there is orientation structure that doesn't change with brick offset. In the column and checkerboard cases, the aggregates are equal — implying no slant, yet they still have an offset.

The graph shown in Figure 5.8 (see page 98) is based on the same population-bin technique used in Figure 5.7 (see page 96). The population used was one mortar line. An aggregate<sup>4</sup>. For the two cases where the stimulus is symmetrical, this graph shows the following:

- that the ‘tendencies’ in both lines cross (and are equal) at the checkerboard condition<sup>5</sup> where *brickoffset* = 100 pixels— meaning the slant in each line becomes equal and then with increasing offset turns to the opposite direction. Although the checkerboard case was not psychophysically investigated by Gregory and Heard, this does agree with informal observation of the stimulus
- that the ‘tendencies’ are equal in the column conditions<sup>6</sup>, where *brickoffset* = 0 pixels and *brickoffset* = 200 pixels— meaning that any slant is equal. Again, although the column cases were not psychophysically investigated by Gregory and Heard, this does agree with informal observation of the stimulus
- the graph shows that there is an offset caused by a population of pixels that does not change as the slant reverses.

Figure 5.7 (see page 96) shows a single case (*brickoffset* = 50 pixels) of the aggregate graph in Figure 5.8 (see page 98). In the polar plots, some of the structure well below horizontal is common to both the upper and lower lines. Also, taking into account the log scale, it can be seen that the main variation between the mortar lines is near the horizontal. This observation can be used to argue for there being structure common to both lines, far from horizontal, that does not affect a slant on that line. The offset in Figure 5.8 (see page 98) also supports this observation, showing a non-zero aggregate even in symmetrical cases (checkerboard and columns) where there is no illusion.

The graph also displays peaks around the highlighted conditions, these could conceivably be caused by the corners interacting as they become closer, although the exact cause is unclear. This hypothesis is potentially supported by the fact that these artefacts are seen on either side of the checkerboard condition, at these brick offsets the corners of the central offsetting strip get closer to the upper and lower strips of bricks. A more concrete explanation might follow for further investigation.

---

<sup>4</sup>number of pixels with orientation above horizontal minus pixels with orientation below

<sup>5</sup>where the pattern appears checkered, bar the mortar lines

<sup>6</sup>where the black bricks are aligned in columns separated by columns of white bricks

## 5.4 Understanding GROM’s performance with the Café Wall

As seen in ‘GROM and the Café Wall ’ (see page 94), explicitly finding the local orientation of each pixel requires further processing on an orientation map in order to reason the tendency of the overall slant in mortar line. However, this does not offer any information regarding what facets of GROM are enabling this behaviour. A better understanding of the exact mechanisms in the model that enables GROM to show the same tendency as psychophysical results could be useful for the following reasons:

- to predict the behaviour of the model with different illusions
- to understand better the exact nature of illusions and identify common features between them
- to understand exactly what parts of the model are essential to good performance with illusory stimuli enables the model to be as parsimonious as possible
- quickly identify illusory stimuli without the need to resort to detailed and laborious investigations with GROM.

The following sections explore facets of GROM that might offer explanations of the performance of the model specifically in relation to the Café Wall, but it is hoped that these findings can be generalised to other illusory stimuli.

### 5.4.1 Estimation of direction of maximum gradient

As the local orientation model presented here is based on a least-squares estimation of the direction of maximum gradient, one possible rule could be based on a crude estimate of the direction of maximum gradient; in which direction does luminance change most steeply? In order for this rule to be applied elsewhere, it must first be validated as a good approximation to, for example, the full analysis applied to the Café Wall above.

In order to investigate this approach, a stimulus is convoluted with a simple filter,  $G_{1,0}$  (c.f. ‘Finding an image derivative’ (see page 41)), that is rotated through a range of angles, and the maximum response is found, as in Figure 5.9 (see page 102). Where the filter is a maximum, the gradient is a maximum. However, this is 90° out of phase with the orientation of the structure, referred to as the *contour orientation*.

Point	max filter response	model orientation
1	26	159
2	51	151
3	154	66
4	129	169
5	90	90
6	90	90

Table 5.2: A table summarising Figure 5.9 (see page 102), detailing results for each labelled point (left column) with the contour orientation at which the response to the rotated simple-filter is at a maximum (middle column) and the contour orientation given by GROM at the same point (right column).

The simple filter used has the same parameters as the standard model, a 23 pixels square filter support for a  $G_{1,0}$  with a standard deviation of 1.5 pixels at it's centre. The Café Wall stimulus used was the same dimensions as those used elsewhere in this section; 256 pixels square in total with 100 pixels square bricks interleaved with 2 pixels thick grey mortar.

The graph in Figure 5.9 (see page 102) and Table 5.2 (see page 101) show that the horizontal structure of the mortar, at points 5 and 6, is detected as horizontal by both models. However, as can be seen in Table 5.2 (see page 101), the orientation given by the simple filter approach does not compare to the orientation given by the gradient model for points 1 to 4. This result shows that the illusion cannot be explained by merely examining the first derivative gradient information.

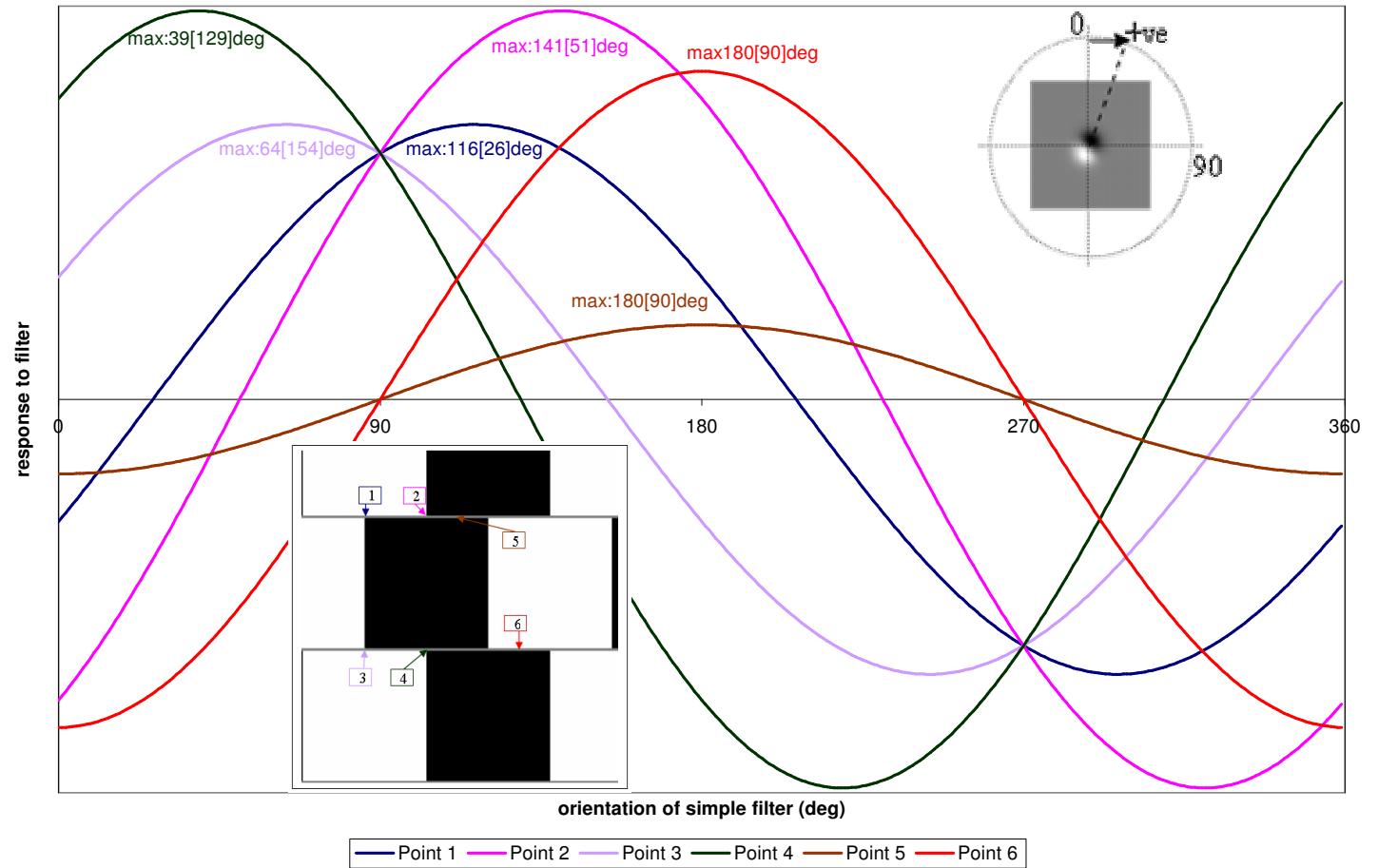


Figure 5.9: A graph showing the response to a rotated simple filter (shown top right inset) at six points on a 256pixels square Café Wall stimulus. An inset (bottom left) indicates the location of these points on a Café Wall stimulus [Gregory and Heard, 1979] measuring 256pixels square, equivalent to  $2^\circ$  of visual angle. Four of the points (1, 2, 3 and 4) are concentrated on unique configurations of corner and brick, whereas the two other points (5 and 6) are on clear sections of the mortar away from the corners. The angle of rotated simple-filter giving the maximum convolution result is given for each plot. The number in brackets that shows the contour orientation equivalent for this maximum.

### 5.4.2 Inclusion of multiple derivatives — extending the k-vector

As  $k$  increases, higher order derivatives are included in the tangent ratio — making it more robust. This is shown in Figure 3.3 (see page 63), where the effect of including more derivatives is to make the output for a sine-wave stimulus much more uniform.

Figure 5.10 (see page 104) shows the Café Wall stimulus and orientation map pairs with increasing length of  $k$ -vector. This figure gives further confirmation that a first-derivative gradient based deduction of the direction of maximum gradient does not account for the illusion. The  $k = 1$  case shows little difference in the orientation map between the two mortar lines, especially if it were subjected to the grouping-independent ‘population-bin’ approach. The figure also indicates that the correct overall tendency in the results comes from the inclusion of higher-order filters ( $G_{1,1}$ ,  $G_{2,0}$ , etc), ie. where  $k > 1$  in the ratios.

Figure 5.11 (see page 105) explores the filter responses at two selected points, 1 and 6 (taken from the labelling scheme used in Figure 5.9 (see page 102)). The  $k = 1$  and  $k = 2$  cases for points 5 (not shown) and 6 are the same. Both show a tangent function which goes to infinity at filter rotations of  $90^\circ$ . This is caused by the response to the  $G_{1,0}$  and  $G_{2,0}$  filters going to zero in the denominator of the ratio, as in Equation 2.13.

In the case of points 1 to 4 (only the graph from point 1 is shown as 1 to 4 exhibit the same behaviour), this does not occur; there are no common zero-crossings<sup>7</sup> in the response to the rotation of the  $G_{1,0}$  and  $G_{2,0}$  filters. The function mapped out by the ratio where  $k = 1$  has the form of a tangent, whereas in the  $k = 2$  case, the function does not have tangent form.

This shows why the final orientation results where  $k = 1$  and  $k > 1$  are different depending on the stimulus. This difference comes from the inclusion of higher order derivatives, i.e.  $G_{2,0}$ , which, at points showing illusory tendency, have different zero-crossings to  $G_{1,0}$ .

---

<sup>7</sup>where the graphed quantity passes through zero



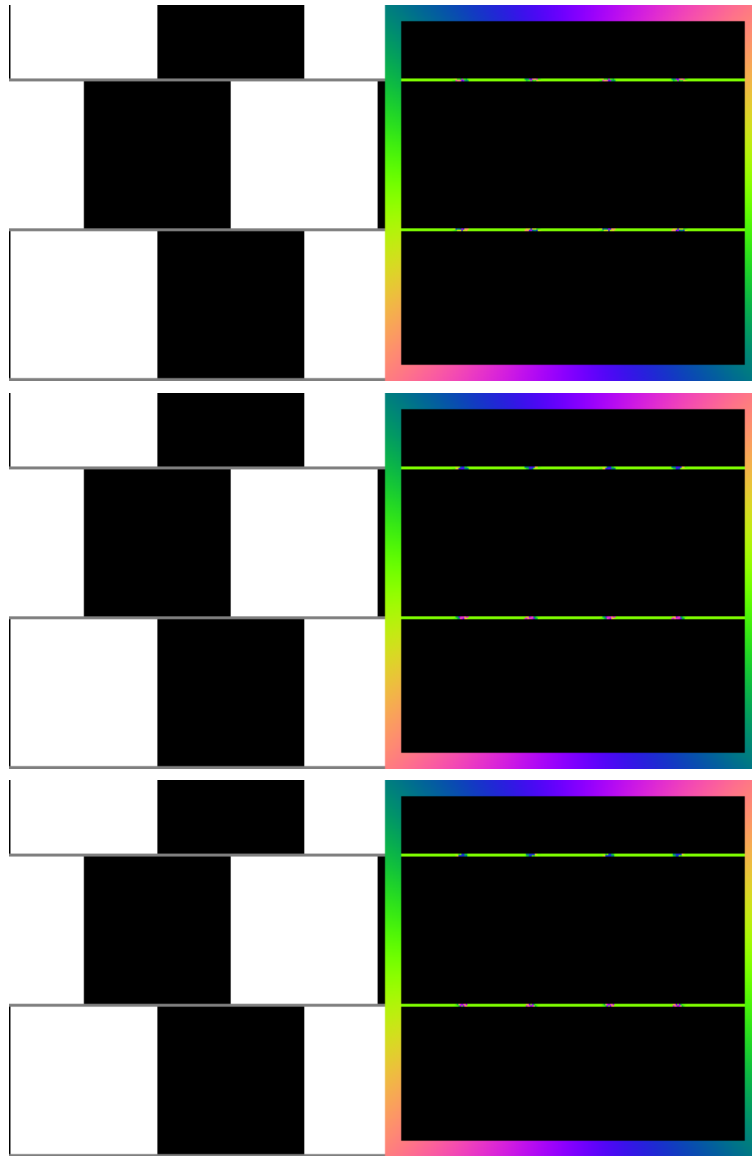


Figure 5.10: These figures show the effect of increasing the value of  $k$  for the ratios, equivalent to increasing the number of derivatives in the ratio. (top) A Café Wall stimulus and masked orientation output computed using  $k = 1$  at one scale. (middle) A Café Wall stimulus and masked orientation output computed using  $k = 2$  at one scale. (bottom) A Café Wall stimulus and masked orientation output computed using  $k = 4$  at one scale. The orientation output in the corner regions of the mortar in each of the cases above is different. Clear slant tendencies, which follow psychophysical trends examined above in ‘GROM and the Café Wall ’ (see page 94) can only be seen middle and bottom where  $k > 1$ .

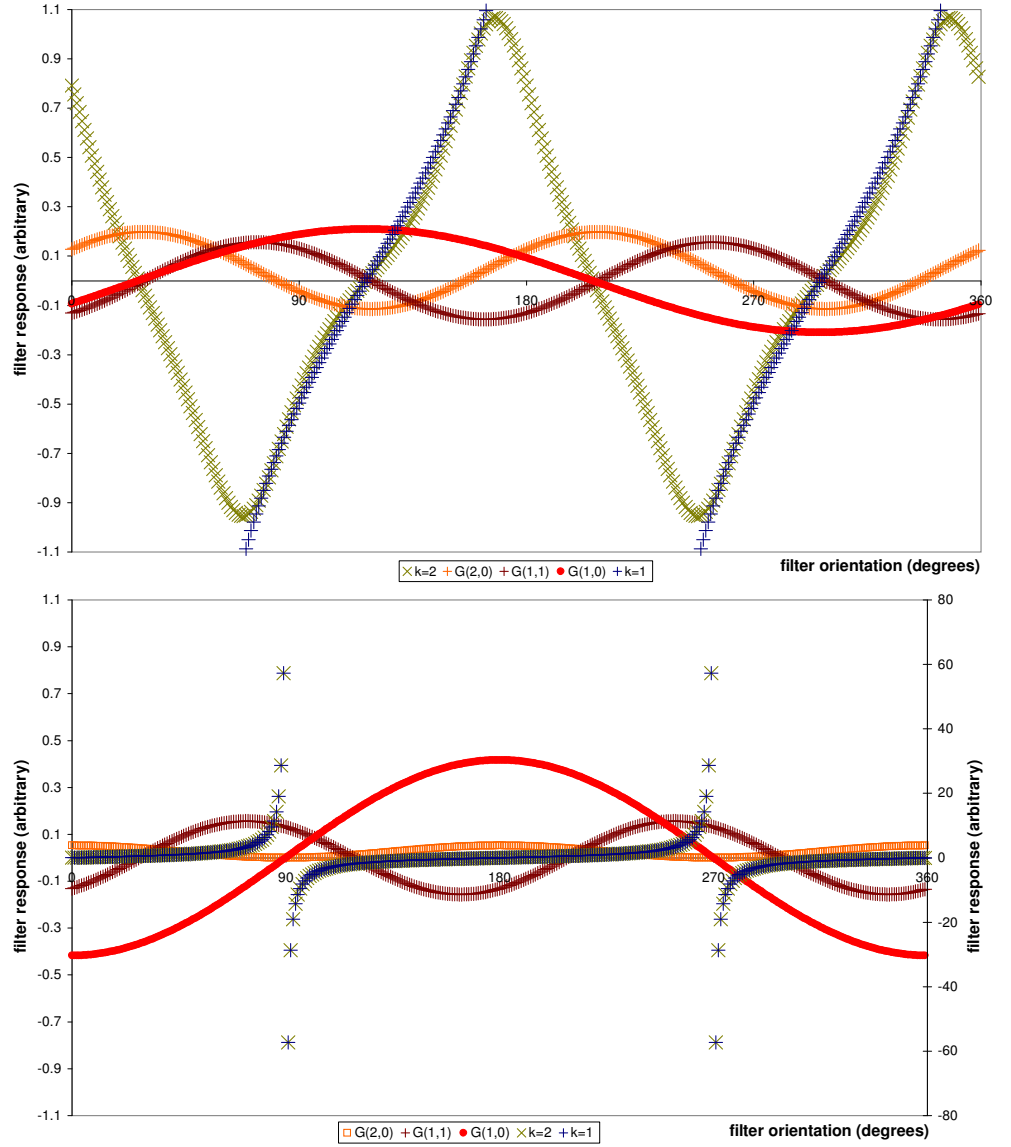


Figure 5.11: These graphs show the response of the Café Wall at two points, 1 (top) and 6 (bottom) (same labelling as in Figure 5.9 (see page 102)), to convolution with  $G_{1,1}$  and  $G_{2,0}$  as they are rotated through  $360^\circ$ . The resulting tangent ratio where  $k = 1$  and  $k = 2$  is also plotted. Where  $k = 1$ , a normal tangent function is mapped out at points 1 and 6. However, for  $k = 2$  a non-tangent function is clearly mapped out for point 1 — this is seen in the top graph as a green ‘zig-zag’ function. The implication is that this will ultimately effect the deduction of a final measure of orientation.

### 5.4.3 Difference in zero-crossings of $G_{2,0}$ and $G_{1,0}$

The difference in zero-crossings has been shown to affect the type of function mapped out with the variation of orientation of filters.

Figure 5.12 (see page 107) shows the  $G_{1,0}$  and  $G_{2,0}$  superimposed on the two points chosen for closer examination, 1 and 6. For points 5 and 6, simple vertical luminance steps, zero-crossings occur at the same  $90^\circ$  rotation. However, as explained in the figure, in the more complicated case of points 1 to 4, zero crossings occur at different orientations.

This leads to the non-tangential functions seen in Figure 5.11 (see page 105) (bottom), which in turn has an effect on the final estimate of orientation. Table 5.2 (see page 101) shows this effect on the final estimate for points 1 to 4, where GROM output on the horizontal ( $90^\circ$ ) mortar line is not shown as horizontal. As these points are a small sample of those in the corner regions, they do not quite show the same tendencies as the ‘binning’ shown in Figure 5.7 (see page 96). Despite this, points 1 and 2 are below horizontal (with the sense  $\searrow$ ), where as point 3 is above (with the sense  $\nearrow$ ) and point 4 is below (with the sense  $\searrow$ ).

The ratio will not become vertically asymptotic<sup>8</sup> wherever  $G_{1,0}$  and  $G_{2,0}$  do not have a zero response at the same time. Hence the ratio does not map out a tangent function. However, the argument presented above suggests that grey areas near corner configurations would likely result in this behaviour. Indeed, it can be concluded from the slants at the corner regions in Figure 5.6 (see page 95) that the local slant tends towards the apex of the corner. Applying this as a heuristic results in the predicted slant in the angle shown by the sense of the red lines in Figure 5.13 (see page 108).

---

<sup>8</sup>approach vertical

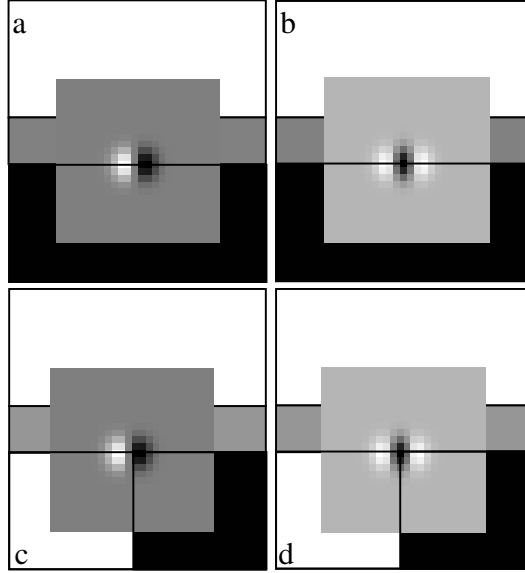


Figure 5.12: The diagram shows a  $G_{1,0}$  and a  $G_{2,0}$  filter overlaid, rotated to  $90^\circ$  at points 1 and 6 on the Café Wall. (a) Shows a  $G_{1,0}$  at point 6. (b) Shows a  $G_{2,0}$  at point 6. (c) Shows a  $G_{1,0}$  at point 1. (d) Shows a  $G_{2,0}$  at point 1. At a filter rotation of  $90^\circ$  filters  $G_{1,0}$  and  $G_{2,0}$  have zero response at point 6. The structure underneath the filter is horizontally symmetric, so a horizontally symmetric zero-sum filter, such as  $G_{1,0}$  and  $G_{2,0}$ , will have a zero response. This is demonstrated in the graph resulting from point 6 in Figure 5.11 (see page 105). For point 1 (the lower row) neither filter will have a zero response at rotation of  $90^\circ$ . The sum over the lobes is uneven due to the black brick in the lower right corner. In the case of a corner with three different levels of luminance, estimating the rotation required for the sum of the luminance over the positive and negative lobes to sum to zero is non-trivial. However, it is possible to envisage how the zero-response angle might not be the same for the two-lobed  $G_{1,0}$  and three-lobed  $G_{2,0}$  filters. The graph in Figure 5.11 (see page 105) for point 1 demonstrates that the zero-crossings (where the filter has zero response) are not equal for  $G_{1,0}$  and three-lobed  $G_{2,0}$ .

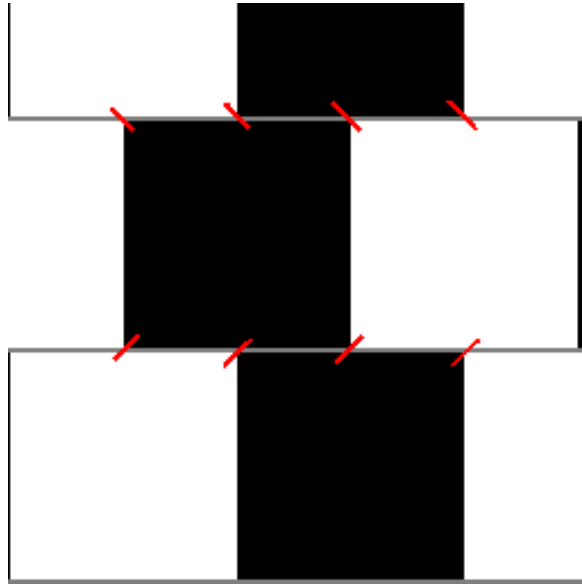


Figure 5.13: A Café Wall stimulus marked with red lines in the mortar at the corners of the bricks. The sense of the slants shown follow from the earlier arguments which suggest that at these points there is a difference in zero response angles of the  $G_{1,0}$  and  $G_{2,0}$  filters. It was demonstrated that on the simple horizontal structure away from the corners, this does not occur.

## 5.5 Conclusion

This chapter presented GROM applied to the Café Wall stimulus. After showing the orientation map in the mortar regions, further analysis was warranted to interpret whether a slant in the same sense as those measured psychophysically was implied by the local orientation information. An attempt was made to avoid assumptions as to how local orientation information is grouped upon in coming to a global perception of the slant of the continuous mortar structure. This was achieved by examining the distribution of orientations over the entire population of pixels in the mortar region. It was shown that the relative population of pixels above and below the horizontal do infer the sense of slants perceived in the mortar lines.

The change in the aggregate was examined as one row of bricks was offset over the entire range. This range includes two key conditions where there is no perceived illusory slant:

- black and white columns
- checker board.

The slant at these configurations is perceived as equal; this is shown in the measure from the model results via the equal population aggregate in these conditions. The graph of the (above minus below) population aggregate against brick offset shows a vertical offset, supporting the idea that some pixels remain unchanged by the brick offset and hence do not affect the slant perceived in the illusion.

As explored in the introduction, various empirical rules have been proposed that attempted to explain the slants in the Café Wall illusion. These rules are attractive as they offer intuitive accounts for the illusory effect. The model based investigation above suggests a more complex reasoning to the illusory slants. However, this need not obviate the possibility of there being a valid heuristic that could be employed to predict the illusory tendencies of such illusions.

The simple ‘estimation of direction of maximum gradient’ rule was investigated and disproved, presenting a more complicated set of reasons behind the misperception of the slants. A heuristic was deduced from the orientation map results and applied wherever a grey region meets a corner configuration. This seemed to predict the slants in the Café Wall illusion. This explanation, based on GROM, does agree with the ‘corner effect’ [Morgan and Moulden, 1986, Kitaoka, 1998]. The slant towards the corner apex given by GROM provides a mechanism to make the angle subtended at the corner acute.

A more recent study by Roncato concentrates on the individual elements of Café Wall-like illusions. Roncato examined the shape deformation of a rectangular border with various simple

internal and external texture patterns. A simple gradient based explanation of the findings is proposed in the section entitled ‘Textures of Roncato’ (see page 144).

## Chapter 6

# Automatic generation of illusory stimuli

This chapter describes a novel approach to Automatic Generation of Illusory Stimuli (AGIS). A general introduction to similar techniques is presented. This is followed by an explanation of the Genetic Algorithm based computational method and then by an examination of results of the implementation. The method and results of psychophysical testing of the textures are described, the general aim of which is to show that textures generated with this technique contain illusions of orientation.

### 6.1 Introduction to Evolutionary algorithms

An *optimization technique* is defined as the method of searching for a solution to a problem by the automatic manipulation of data used to find a solution. Usually, optimization techniques find solutions to a problem that maximise or minimise the value of a certain function. Optimization techniques based on the evolution of many possible solutions to a problem were introduced by Holland [Holland, 1975]. Holland formalised the approach as a study of the process of *adaptation*. By drawing analogy to the Darwinian evolutionary process that occurs in living organisms, Holland specified the basic components that an algorithm would require in order to harness naturally occurring adaptation to solve optimization problems.

The ideas presented by Holland were developed into a computational technique described by Goldberg [Goldberg, 1989]. Goldberg illustrated their potential use in solving problems computationally by using examples. One such problem solving approach, termed a *Genetic*



*Algorithm* (GA), had the following characteristics:

- *population* - a pool of possible solutions to the problem was maintained for a *generation*
- *structure* - each member of the population had data, analogous to the genotype found in Nature, for storing the fundamental attributes required for a potential solution to the problem
- *life cycle* - the operations in each generation, based on survival of the ‘fittest’:
  - *reproduction* - the instantiation of new members of a population using *crossover*, the mixing of structures of successful members of the previous generation
  - *mutation* - the perturbation of randomly chosen single elements of the structure of each member of the population. This allows new possible solutions to emerge
  - *fitness* - the relative success, and therefore likelihood of survival, of a member of the population in solving the problem is judged by the score that its structure achieves when a *fitness function* is applied.

Goldberg originally uses examples called *Simple Genetic Algorithms*, where the start of each generation involves a complete replacement of the population. In this case, the population is said to be *non-overlapping*, where each member is newly bred through crossover of the previous members [Goldberg, 1989]. A simple adaptation to this approach, proposed by De Jong [De Jong, 1975], called a *Steady-state Genetic Algorithm*, involves keeping a number of the best individuals of the previous population to compete against the newly bred members of the new generation. The populations are said to be *overlapping* in this case. This process is summarised in the schematic diagram in Figure 6.1 (see page 113).

Many variations on this basic framework have been used for numerous applications. Bäck discusses the nomenclature of *Evolutionary Algorithms*, of which GAs are a subset [Bäck, 1996], and points out that ad hoc modifications lead to a wide variety of algorithms significantly different from the classic Goldberg ‘Genetic Algorithm’.

The GA approach used in AGIS is a heavily customised version of the Simple Genetic Algorithm described by Goldberg [Goldberg, 1989]. After a description of some relevant applications of GAs, there follows a description of the intended goal of AGIS, and how operators have been developed to reach this goal.

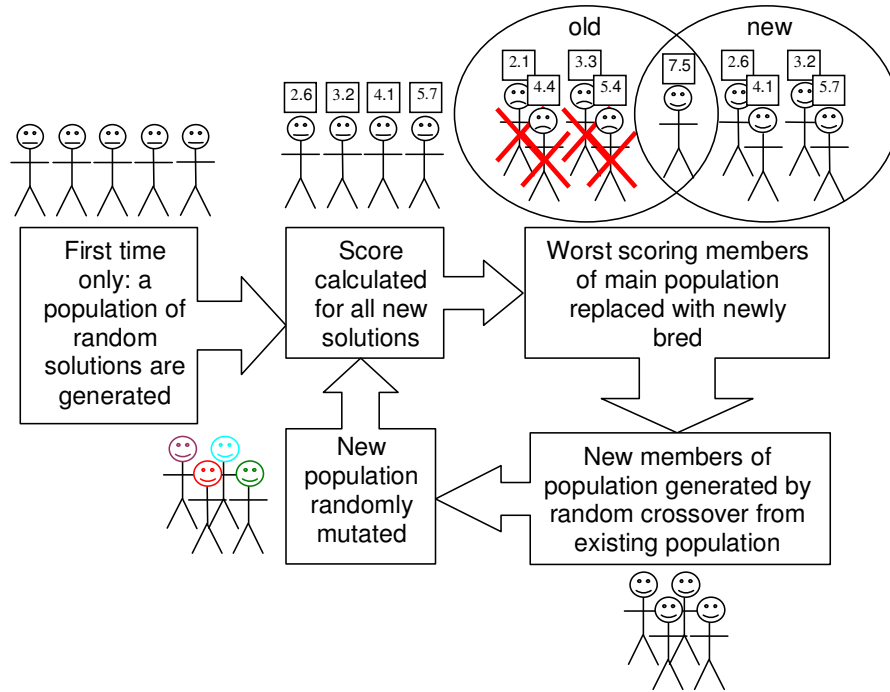


Figure 6.1: A diagram showing the steps in a steady-state GA.

## 6.2 Applications of Genetic Algorithms

As an optimization technique, GAs have been used to solve various problems. They are a commonly used ‘out of the box’ technique for solving problems. This has given rise to a plethora of different applications in numerous fields, a summary of which in its entirety is not within the scope of this thesis. Two examples have been selected, based on the similarity of some aspects to AGIS.

Sims presented a novel technique for creating simulated ‘structures, textures and motions’ for use in computer graphics and animation [Sims, 1991]. The key novelty of the technique is what Sims termed *interactive perceptual selection*, where a traditional GA’s fitness function was replaced by a human perceptual judgment. During each generation, a user was required to select several preferred individuals which were then used as parents for the next generation.

One particular novelty of the approach is the use of symbolic expressions in the genotype; these enabled limitations of traditional fixed-length structures of traditional Goldberg-like genotypes to be overcome. Sims used a low-level language to generate images, rather than using a fixed length structure that would potentially have been a bigger space in which to find solutions. The use of a description language enables potentially complex changes to be made to the generated image

using only a small amount of data. Rather than evolving fixed length structure, Sims was therefore evolving expressions.

Using this approach gave rise to artificial content for virtual environments, an example of which is shown in Figure 6.2 (see page 114), for this image, the evolved expression was:

```
(sin (+ (- (grad-direction (blur (if (hsv-to-rgb (warped-colour-noise
#(0.57 0.73 0.92) (/ 1.85 (warped-colour-noise x y 0.02 3.08)) 0.11 2.4))
#(0.54 0.73 0.59) #(1.06 0.82 0.06)) 3.1) 1.46 5.9) (hsv-to-rgb
(warped-colour-noise y (/ 4.5 (warped-colour-noise y (/ x y) 2.4 2.4))
0.02 2.4))) x))
```



Figure 6.2: An example of an image bred by Sims. The method is based on a Simple Genetic Algorithm but the fitness stage is performed via user choice - termed *perceptual selection*. Preferred structures are selected by a user and are subsequently used as the basis for the proceeding generation. Sims used a low-level description language to generate images, rather than a fixed length structure. The image presented is adapted from Sims [Sims, 1991]

More recently, Bond and Kamil presented a Genetic Algorithm technique for testing an ecological theory of a selection mechanism [Bond and Kamil, 2002]. They attempted to test the hypothesis that *frequency-dependent predation* may be a factor contributing to *visual polymorphism*<sup>1</sup> in

<sup>1</sup>where a species has a variety of outer appearances

camouflaged prey species. It was originally proposed by Poulton [Poulton, 1890] that frequency-dependent predation was a result of how predators searched for prey. Searching is easier with a narrow criteria, and most successful when that criteria encompasses as large a segment of a varied prey species as possible.

This predator behaviour will tend to suppress the most common prey morphs and encourage rarer types. Hence the ability for a species to be varied in outer appearance is an advantage. Bond and Kamil showed that a species can evolve polymorphism as a survival tactic. By presenting artificial camouflaged moths on peckable monitors they simulated predation, and by inferring that pecked moths were more visible, this was used as a fitness test in a GA. As well as testing an ecological theory, this approach yielded a semi-automated method of generating textures that are more camouflaged to this species of predator.

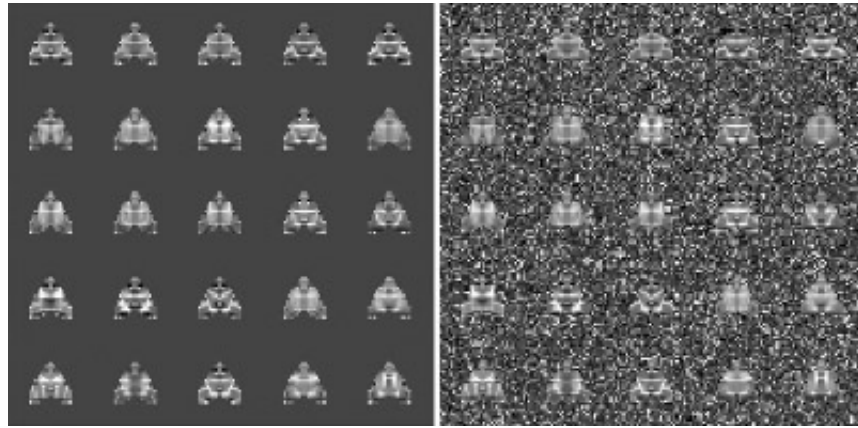


Figure 6.3: Two example stimuli as presented to the Blue Jays, it shows several generated moths in rows on two different backgrounds. The more discernable the moth, the more likely it was to be pecked. Adapted from Bond and Kamil [Bond and Kamil, 2002].

The experiments described above represent different levels of interactivity for GAs in the production of patterns. Both of the algorithms discussed above require pauses for a judgment of fitness — via interaction, either with a human or a blue jay. AGIS obviates the need for an interactive fitness judgment as it utilises perceptual models for this stage. This automation hastens the speed of evolution and therefore improves the overall solution-search bandwidth while at the same time comparing two perceptual models.

## 6.3 AGIS: Stimulus generation

### 6.3.1 AGIS: A broad outline

AGIS aims to generate illusory stimuli by searching a conditioned space using a GA with a fitness function that relies on two models of the human perception of local orientation. As suggested in chapter 4, for misperceived contrast-modulated gratings, GROM produces results more like human perception than the SFEM (see ‘Simple Fourier-energy model implementation’ (see page 165)). Therefore, maximising a fitness function that sums the difference between these measures over randomly generated stimuli should produce solutions that contain differences between the orientation of perceived and veridical structure, and therefore illusions of orientation.

There follows an explicit description of the GA-based AGIS algorithm. An empirical incremental approach was used to adapt the behaviour of the operators used and the values of the corresponding parameters, see Table 6.1 (see page 122).

In order to validate AGIS and the above assumptions on the performance the two orientation models, a selection of ‘best solutions’ were tested psychophysically. A description of this testing and the results from it are presented below. Additional questions are also examined in order to attempt to validate the technique.

### 6.3.2 The goal of the Genetic Algorithm used in AGIS

The goals of the algorithm are discussed and motivated here, these dictate the implementation of the operators, particularly the fitness function.

The overall goal of the algorithm is to find textures that maximize the difference in GROM and SFEM orientation over the entire texture. Final evolved textures should, judged perceptually and structurally, contain contours that have illusions of orientation.

This approach relies on the assertion that the difference in these models is: that one measures misperceptions better than the other - which is more veridical; and that this is true for any stimulus type. Additionally, by summing the difference over the entire stimulus, textures that contain large regions of orientation difference should be favoured over those with more local differences. This motivates the form of the fitness function, which includes a product term that is the sum of the difference of the local orientation model outputs over the entire stimulus.

Ideally, the generated stimuli should have high-gradient edge-like structures in order that they can be fixated upon during psychophysical testing. This motivates the inclusion of a product term in the fitness function that sums a non-direction selective first-derivative based measure.

To form large stimuli for psychophysical testing, another condition is that the textures tessellate with like textures meaning they are periodic over the 128 pixel square. Tessellating patterns are a simple byproduct of the structure chosen, discrete Fourier-space (DFS), when used without a luminance offset, produces textures constrained to being regular and tessellating.

### 6.3.3 The details of the Genetic Algorithm

The GA is based on a steady-state approach [De Jong, 1975], which allows a specified number of the top scoring individuals, from a generation to be preserved in the next, as seen in Figure 6.1 (see page 113). This enables good solutions to be maintained so that, should crossover and mutation not produce better individuals, breeding in the next generation can attempt to modify the exact same successful individuals.

The size of the population used,  $N_p$ , was empirically determined by the computational constraints of processing time and memory. Generally, a broader search is performed by using a larger population size, however this utilises more memory and increases the computational time per generation and therefore, the overall run-time. The number of low-scoring individuals replaced in each generation is given by,  $N_r$ , therefore, the number of high-scoring individuals maintained is given by,  $N_p - N_r$ .

The number of generations over which textures were evolved,  $N_g$ , was also determined empirically. The progression of top-fitness score per generation was examined in ‘test-runs’ of the implementation. Plotting these progressions in score, as seen in Figure 6.9 (see page 126), seemed to suggest the number of generations used as reasonable as the scores become horizontally asymptotic. Obviously, this parameter was also constrained by time.

#### The structure

The structure of an individual contains the data required to form a texture which the perceptual models are applied to to calculate a fitness score. Two-dimensional DFS was chosen as the data representation that would most likely produce periodic textures. In practice, DFS is divided into a two-dimensional grid where each point is separated by one cycle per degree. A reverse discrete Fourier-transform is used with the DFS structure to generate individual stimuli size of 128 pixels square. This is scaled to one degree of visual angle in both the orientation models used.

Each structure is constructed of two same sized DFSs, one representing the *real*, or *cosine*, coefficients and the other for the *complex*, or *sine* coefficients. This is equivalent to one DFS data set of complex numbers [Weisstein, 2004].

The structure was restricted such that no values above the maximum spatial frequency,  $F_{\max}$ , were used. This meant that a wide range of periodic textures with continuous contours can be generated with a relatively small amount of data ( $\sim F_{\max}^2$  values). By contrast, if textures were generated by direct pixel manipulation, the amount of data would be large ( $\sim 128^2$  values) and there would be no obvious way to enforce a restriction on the continuity of the structures when tessellated.

Below  $F_{\max}$ , frequencies are selected using a probability distribution based on the Contrast Sensitivity Function. Contrast sensitivity in humans varies with spatial frequency. Manno and Sakrison [Manno and Sakrison, 1974] fitted a function to psychophysical data from contrast sensitivity experiments which has the form:

$$\text{csf}(f) = 2.6(0.0192 + 0.114f)e^{-0.114f^{1.1}} \quad (6.1)$$

where  $f$  is the spatial frequency in cycles per degree of visual angle. Figure 6.4 (see page 118) shows a graph of the contrast sensitivity function (CSF) covering the frequency space used in the structure.

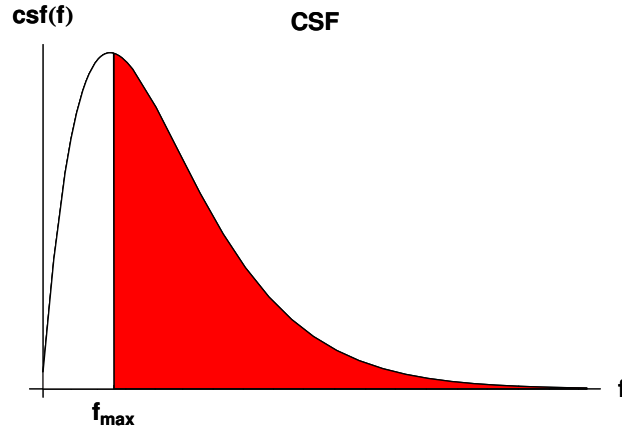


Figure 6.4: A graph showing the Manno and Sakrison's fit to the human contrast sensitivity function, [Manno and Sakrison, 1974], which has the form:  $2.6(0.0192 + 0.114f)e^{-0.114f^{1.1}}$ . Below  $F_{\max}$ , this function is the probability distribution for mutation and initialisation of an individual's structure. Elements in the structure at frequencies above this value (coloured in red) are not used.

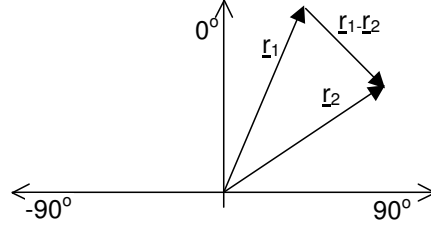


Figure 6.5: A diagram representing the output of GROM and SFEM, each model result is represented as a unit vector  $\underline{r}_n$ . The magnitude of the difference between the two is given by  $|\underline{r}_1 - \underline{r}_2|$ .

### The fitness function

The fitness function is the part of the algorithm that directs the ‘solution search’ towards an optimal solution (described in ‘The goal of the Genetic Algorithm used in AGIS’ (see page 116)). For AGIS, the two main goals taken into account in calculating the score returned by the fitness function are:

- the maximisation of the local orientation difference between the two models
- the maximisation of the amount of local gradient information.

Figure 6.5 (see page 119) shows how the local orientation difference is calculated, based on the magnitude of the vector difference of the two orientation measures,  $|\underline{r}_1 - \underline{r}_2|$ . For the same pixel, the gradient measure is calculated based on:

$$\text{gradient measure} = GM_{x,y} = \sqrt{\left(\frac{\partial I_{x,y}}{\partial y}\right)^2 + \left(\frac{\partial I_{x,y}}{\partial x}\right)^2} \quad (6.2)$$

The overall goal is to maximise the (model determined) orientation difference and gradient measure over the whole stimulus. To obtain this, the two measures are combined to give the following per-pixel measure:

$$\text{Fitness score} = \frac{\sqrt{\left(\sum_{y=0}^{(256-\text{border})} \sum_{x=0}^{(256-\text{border})} |\underline{r}_1 - \underline{r}_2|_{x,y}^2\right) \left(\sum_{y=0}^{(256-\text{border})} \sum_{x=0}^{(256-\text{border})} GM_{x,y}^2\right)}}{(256 - \text{border})^2} \quad (6.3)$$



## The crossover operator

The crossover operator is used to produce replacement members of the population. During crossover, members of the whole population are selected successively in pairs, and based on a probability,  $P_c$ , are used to breed new stimuli; else they are left unchanged to pass through to the replacement population - they effectively become their own offspring if they are not crossed over.

Crossover involves four individuals: two parents and two offspring, as seen in Figure 6.6 (see page 120). Each offspring is the opposite mix of the two parents. When the crossover operation is applied it is multi-point, meaning an offspring's value at each point in the structure is inherited directly from one of two parents at that point. Which parent is the source of each inherited DFS value is based on an equal probability *coin flip*, the other offspring gets the value from the other parent.

Good textures would likely have harmonics<sup>2</sup> that form edges. Harmonics can be encoded by values in several points in the structure. Multi-point crossover enables inheritance of harmonics from one parents or a mix of harmonics from both parents.

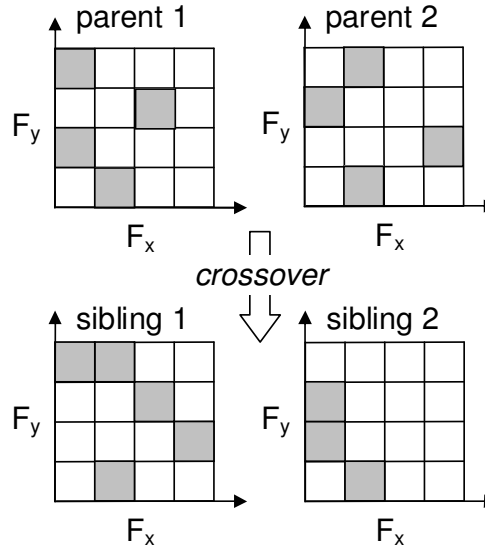


Figure 6.6: A schematic diagram showing crossover when it takes place. Two parents generate two siblings, each the opposite mix of their parents. Here, the grids represent DFS, where a grey box indicates a non-zero value at that spatial frequency,  $(F_x, F_y)$ .

---

<sup>2</sup>comprised of whole-number multiples of a base frequency

## The mutation operator

The mutation operator acts on every individual in the population per generation. It changes values in both the real and complex DFS structures below the maximum frequency,  $F_{\max}$ . The probability of changing a value in either structure is based on the CSF and an empirically determined probability. The probability was manipulated to ensure that approximately 1 value in the structures, real and complex included, was changed per mutation operation. As the probability of mutation is a non-uniform function — according to the CSF distribution — a fixed probability variable is not a valid description of mutation in itself. However, an effective value,  $P_m$ , can be estimated by averaging the number of mutated values in the structure over a large number of generations, see Table 6.1 (see page 122).

When an element of either structure is selected for mutation, it can take a range of values, based on a random integer value,  $n$ , from 0 to 8:

$$\text{new element value} = 2^n \quad (6.4)$$

this gives the element a range of possible values, similar to the binary scheme devised in original GAs described by Goldberg [Goldberg, 1989].

## Initialisation of population

For the first generation of the GA, the population need their DFS structures initialising in order to generate a valid stimulus (one with a non-zero value in its structure) a fitness score for each member. This initialisation is selected to be the same as the mutation process in order to reduce the over all complexity of the development of the operators. Hence over all pixels, there is an effective probability  $P_m$  that a pixel is set to a value:

$$\text{initial pixel value} = 2^n \quad (6.5)$$

where  $n$  ranges from 0 to 8. However, as with the mutation operator, the true probability at an individual pixel level is based on the same CSF probability distribution.

### 6.3.4 Parameters

A series of operators have been described that are integral to AGIS' working. However, also crucial are the values of the variables used, see Table 6.1 (see page 122). The value of the

Name	symbol	impact	value
Overall population	$N_p$	the number of individuals in the population	50
Replacement	$N_r$	the number of lowest scoring individuals replaced each generation	40
Number of generations	$N_g$	the number of generations in the evolution	1000
Mutation rate	$P_m$	the effective probability that a element of the structure would be randomly changed during mutation	0.014
Crossover rate	$P_c$	the probability that a pair of individuals will contribute crossed-over individuals to the next generation	0.90
Frequency-space restriction	$F_{\max}$	the maximum frequency used in the Fourier-space structure	8.32cpd

Table 6.1: A table showing the variables and values used in the operators in AGIS.

variables chosen are similar to those mentioned for Steady-State Genetic Algorithms in Bäck [Bäck, 1996], however, they were determined via empirical methods to give appropriate performance in AGIS. The final values used were those found to best satisfy the goal of the algorithm, though these values may not be optimal.

## 6.4 Technique validation

This section is dedicated to examining the performance of the algorithm, principally by asking the following questions:

- did the fitness function show signs of progression, and therefore evolution?
- are the initial textures significantly different from the final textures?
- does a hand-selected subset of the textures generated contain illusions of orientation?
- misperceived orientation in textures should be detectable by one model preferentially. Does one model contain more local orientation signals at the misperceived angle?

The first two are dealt with in the first subsection entitled ‘Evolution of textures’ (see page 123), the details of the psychophysical testing can be found in the section entitled

‘Psychophysical testing’ (see page 123). The final question is discussed in ‘Preferential model selection’ (see page 131).

#### 6.4.1 Evolution of textures

A gradual increase in the best *score* - the fitness value of the best individual, should be observed during the computation. Without a progression in score over many generations, it cannot be argued that an evolution has occurred as the stimuli could be a facet of the initial randomised structure data. Without evidence of an increase in fitness score, evolution cannot have occurred.

Typically, initialising the population gives a varied series of textures, ranging from completely blank, through simple sine wave gratings to more complicated composite gratings. Given that the initialisation process involves putting values in the DFS structure, this result is unsurprising. A typical range of structures from initialised individuals can be seen in Figure 6.7 (see page 124).

In all, the algorithm ran only 19 times. This was sufficient to get 3 candidate textures for psychophysical testing. The criteria for selection was that their be some definite vertical or horizontal structure that, when tessellated with itself would form a continuous contour to enable orientation judgment. Candidates were also judged, by eye to contain possible illusions of orientation. All 19 final textures can be seen in Figure 6.8 (see page 125). The textures in the figure are more complicated in nature than the initialised textures.

In order to be sure that the selected 3 candidate textures were evolved using AGIS, rather than being the product of chance in the initialisation, the highest fitness score at each generation from the population they originated from was recorded. This is seen in the graph in Figure 6.9 (see page 126). The graph clearly shows that the highest score for each of the populations (from which the highest was chosen as a candidate) increases towards the final 1000<sup>th</sup> generation.

#### 6.4.2 Psychophysical testing

Out of approximately 19 of the ‘best solutions’ evolved in different instances of the algorithm, three candidate stimuli were chosen for psychophysical testing. Best solutions that lacked the continuous structure required for psychophysical testing were discarded. Of the remaining stimuli, three were selected based on an initial test: that successful textures, viewed on a normal VDU monitor using a straight edge, contain apparent illusions of orientation. The selected stimuli were then submitted to a psychophysical test, the method and results of which follow.

The centre patch of the psychophysical test stimuli in Figure 6.11 (see page 128) were selected out of the textures resulting from AGIS. The psychophysical task was to align the outer ring,

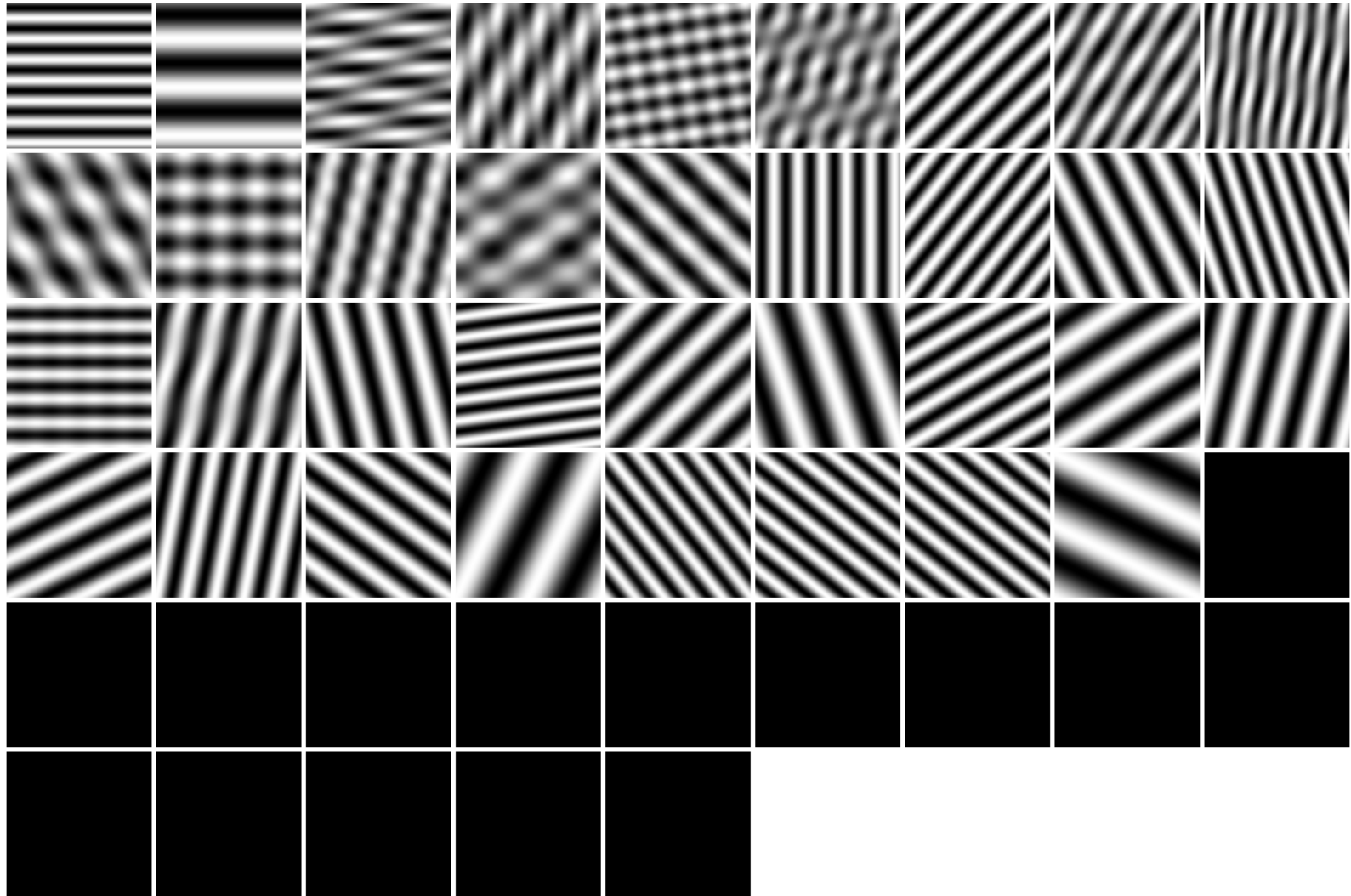


Figure 6.7: This figure shows a typical population of initial textures, each separated by a white border. The textures are ordered according to their score, from the highest scoring (top left) to the lowest scoring (bottom right). Note the lowest scoring textures are blank - as textures are initialised based on a probability of mutation, which in some cases means a zero-filled DFS structure.

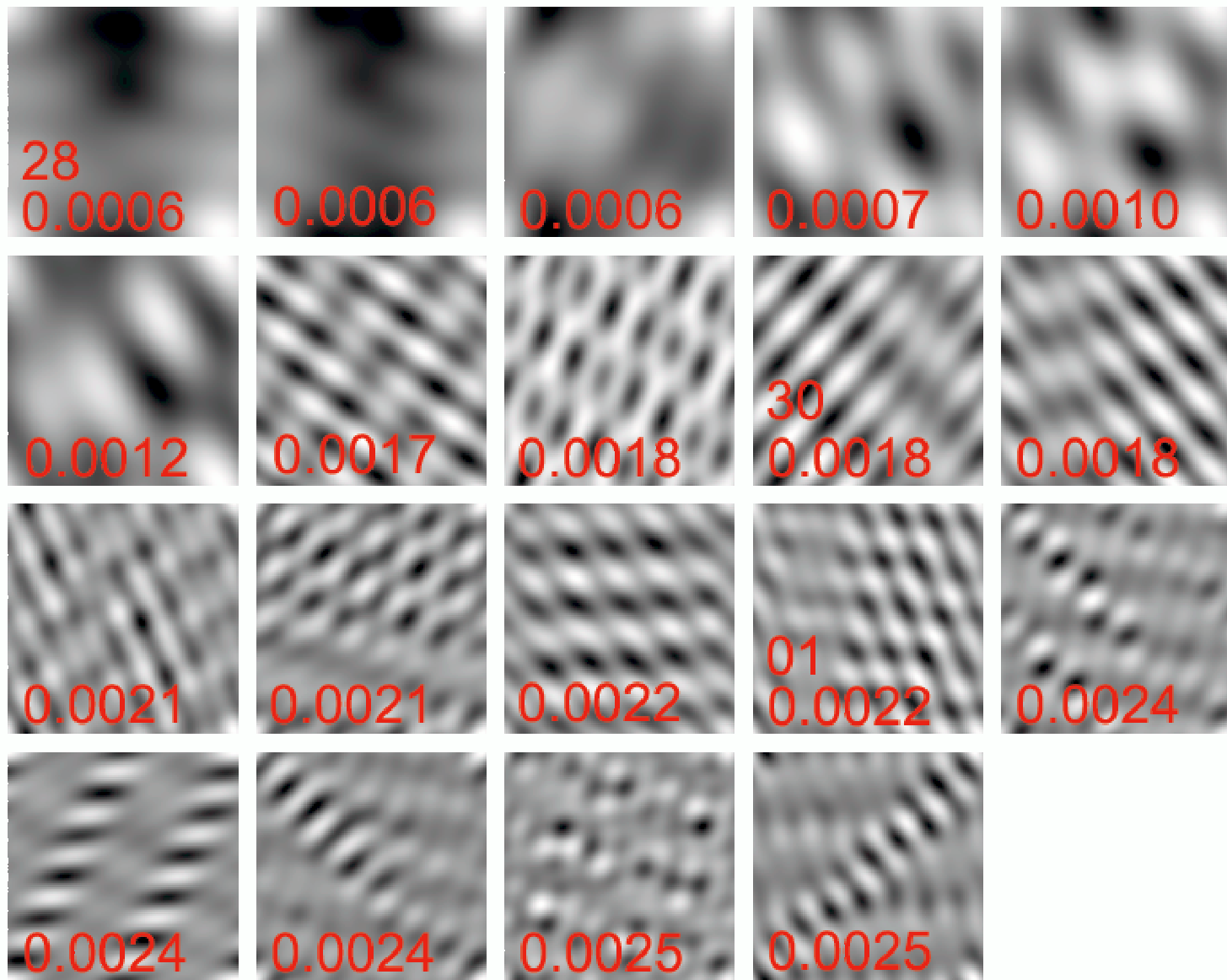


Figure 6.8: In total, the algorithm was run 19 times. The highest scoring texture each run can be seen in the figure above. A subset of 3 were chosen based on their appropriateness for testing, ie they contain coherent vertical or horizontal structure. They are indicated as 01, 28 and 30 above.

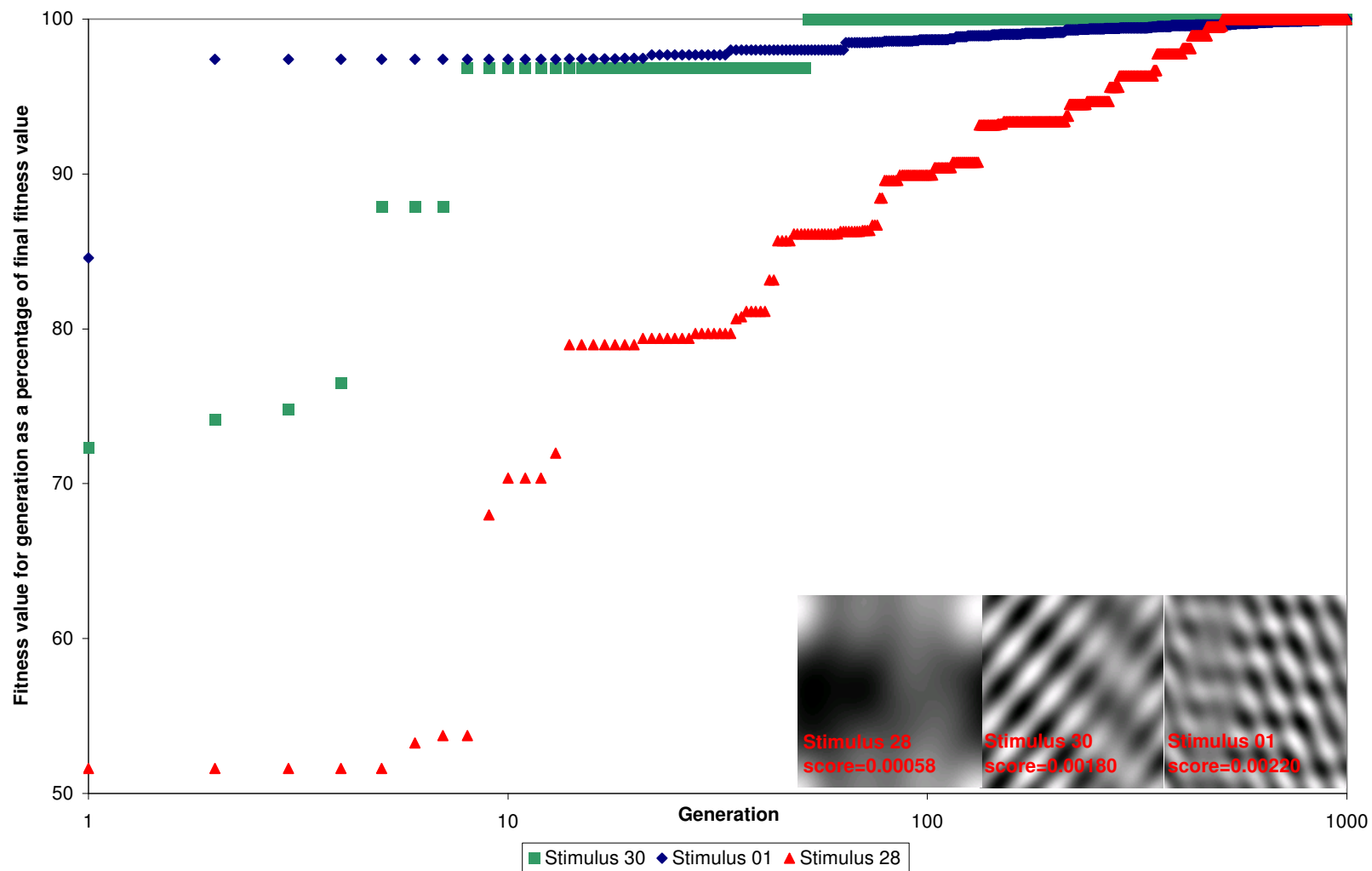


Figure 6.9: This graph shows the progression of best fitness score per generation as a percentage of the final top fitness value for each of the selected stimuli. The three candidate textures originated from the three populations shown here. Each selected stimulus is also shown (inset right) with it's final score calculated as per Equation 6.3.

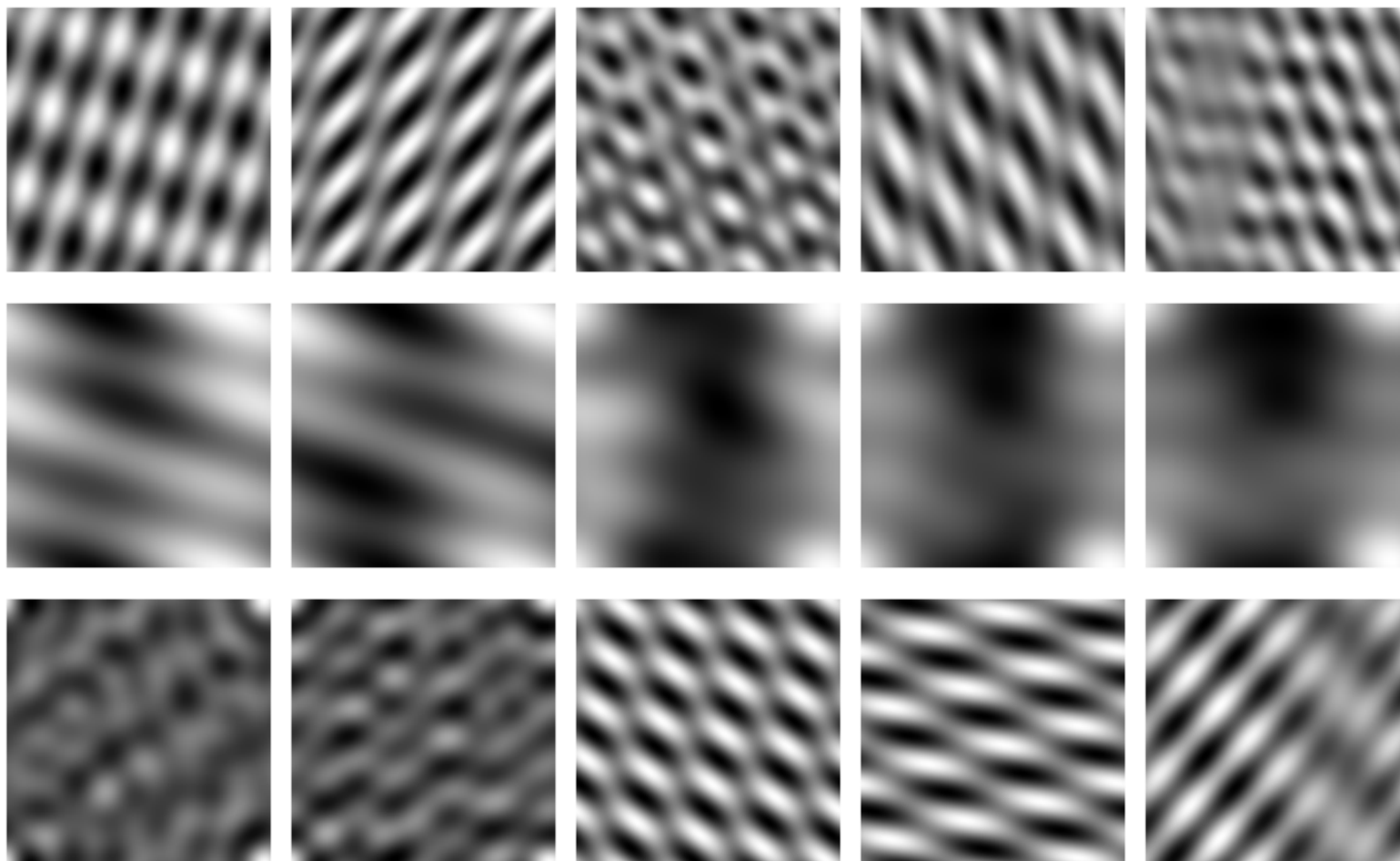


Figure 6.10: This figure shows a selection of the top scoring individuals in intermediary populations — while the algorithm was still running. In the case of 01 (top row) and 30 (bottom row), intermediary best scoring stimuli were only stored when the score changed — they are shown above. For stimulus 28, a selection was made from best scoring stimuli from each significant step in the fitness score shown in Figure 6.9 (see page 126).



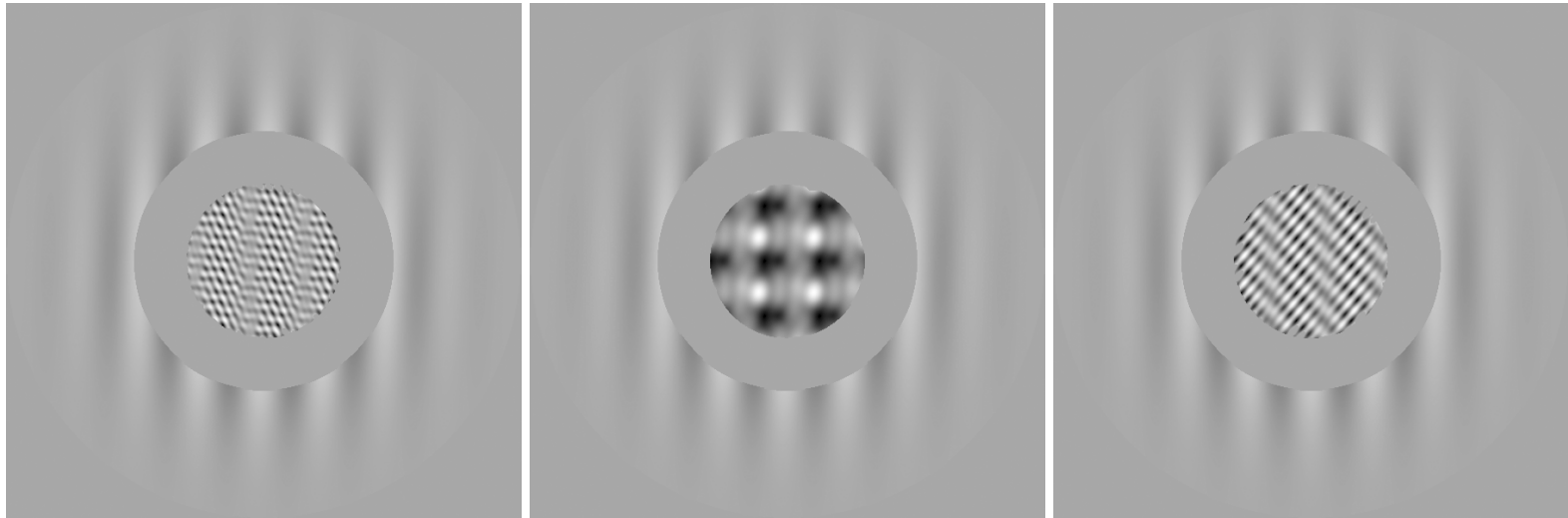


Figure 6.11: Three candidate stimuli chosen for psychophysical testing based on the possibility, judged by eye, that they contain illusions of orientation. Named after the shortened date of their generation, they are (left) stimulus 01, (middle) stimulus 28, and (right) stimulus 30. During testing, the surround was rotated using a control box, with four buttons. Subjects were instructed to use three of them, the *left-hand button* rotated the outer ring counter-clockwise, whereas the *right-hand button* rotated it clockwise. The *top button* confirmed that the subject was satisfied with the alignment of near-vertical central texture and the lines in the outer ring. In each example stimulus above, the outer ring has been adjusted to the mean orientation found in the psychophysics. The intended effect of this is that the structure in the stimulus and outer ring are aligned. As each stimulus contains at its centre a maximum of 3 tessellated versions of the original generated stimuli. Therefore, it cannot be argued that the non-vertical psychophysical orientation-match is caused by a true rotation of stimulus structure.

containing a Gabor, to the ‘near-vertical’ structure of the test stimuli. The task was specified with the following spoken instructions:

“Whilst fixating on the red spot in the centre of the screen, adjust the orientation of the outer ring, using the left and right buttons on the controller, such that the bars of the ring line up with the near-vertical structure of the central patch. When you are happy with the alignment, press the top button on the controller to proceed.”

Each of the 6 subjects then familiarised themselves with the experimental set up by fixating on the spot on the screen and practicing the task with one random example. When ready, the subject then completed 5 trials in which each texture was judged once, in random order. The number of textures presented in each trial was 6, double the number of the originally selected textures, as each texture was presented as AGIS had generated it and also as a mirror image (horizontally-flipped) version of itself. The horizontal-flip would align all structure, including illusory, the other side of the vertical meridian of the stimulus. The half difference between the judged orientations of the original and flipped stimuli pairs is therefore a bias-free estimate of the perceived alignment of the near-vertical structure.

The equation of the rotatable Gabor patch, at a point  $I(x, y)$  in the image, is given by:

$$I(x, y) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{(x-x_c)(y-y_c)}{2\sigma^2}} \sin(2\pi f(\sin(\theta)x + \cos(\theta)y)) \quad (6.6)$$

where  $\theta$  is the adjustable orientation,  $\sigma = 1.56cpd$  the standard deviation of the Gaussian envelope,  $f = 0.5cpd$  is the spatial frequency of the sine wave structure and  $I(x_c, y_c)$  is the centre of the stimulus. A 1cpd thickness annular border is placed radially between the tessellated stimulus patches and the rotatable Gabor patch. This begins at 1.5degrees from the central spot and ends at 2.5degrees from the central spot.

Before any testing of subjects commenced, the monitor was gamma-corrected in order to ensure that the linear grey-scale used in the encoding of bitmapped textures was represented by a linear scale of luminosity of the monitor. All stimuli were displayed on a 19” Sony Trinitron Multiscan 500PS monitor with a refresh rate of 100 Hz driven by a VSG 2/5 from Cambridge Research Systems. The background of the display was grey (Commission Internationale d’Eclairage (CIE) 1931 chromaticity chart:  $x = 0.329$ ,  $y = 0.329$   $z = 50.813$ ).

The subjects were situated at the correct distance (1.91m) in order to ensure that the central patch (0.1m on the screen) of generated texture subtended 3 degrees of visual angle. This meant the central circular stimulus measured 3 repeating texture patches high and wide at the most. Stimuli used in the experiment can be seen in Figure 6.11 (see page 128).

Due to the constraints on the size of laboratory room available, stimuli were sub-sampled to 64 pixels square. Halving the resolution of the stimulus reduces the maximum spatial frequency in the stimulus by half, from 64cpd to 32cpd. As the maximum spatial frequency of stimulus is already restricted below this value,  $F_{\max}$ , sub-sampling should not have impacted the results significantly.

Of the selected stimuli generated, one (so called ‘28’) contained a clear structure that seemed continuous and near-horizontal. This structure was rotated through 90 degrees in order that the psychophysical task be the same for each texture.

### 6.4.3 Psychophysical results

Table 6.2 (see page 130) shows the data collected from 6 subjects. Apart from the author, prior to the experiment, each subject was naive to the stimuli, task and aim. Most subjects had some previous psychophysical experiment experience. No subject has been discounted from the data presented.

Results are generally considered as significant if the two-tailed p-Value is below 0.05. As seen in Table 6.2 (see page 130), two of the AGIS stimuli meet this condition and therefore were found to contain illusions of orientation.

Observer	Stim 01	Stim 28	Stim 30
EKC	0.5	1.3	0.8
DHA	1.5	1.8	0.3
ARK	5.1	1.8	0
AJA	2.1	0.4	-0.4
ST	0.6	3.5	0.9
ACS	1.5	0.65	-0.15
Average:	1.9	1.6	0.24
Standard error:	0.69	0.45	0.21
Two-tailed p-Value:	0.042	0.018	0.31
Student t-score:	2.7	3.5	1.1
Significant?	yes	yes	no

Table 6.2: A table showing the individual average half difference in judged tilt between reversed stimuli over 5 trials. In the 8 lines of the table, values are expressed in degrees, increasing positive clockwise from vertical. The average over individuals and sample standard deviation are shown along with the two-tailed p-Value of the distribution and the Student t-score. These statistics indicate a significant illusory tilt for stimuli 01 and 28.

#### 6.4.4 Preferential model selection

AGIS maximises the absolute value of the local *difference* in orientation summed over the entire orientation map, without regard for which model correctly measures the perceived local orientation of the generated stimuli. Therefore, a further test must be performed to ascertain, if illusory, which model correctly measures the illusory components of a psychophysically tested stimuli.

The stimuli generated have been processed by two perceptual models of orientation. The goal of the algorithm is to find textures that maximise the difference in orientation between GROM and SFEM. Therefore, as highlighted in the section

‘The goal of the Genetic Algorithm used in AGIS’ (see page 116), the stimuli generated will be ‘correctly’ misperceived by one model preferentially over another.

In order to validate GROM it must be shown that its orientation map indicates the psychophysical results and that the SFEM does not. However, this validation hypothesis is not related to validation of AGIS per se, as results shown above indicate that texture with illusory contours have been generated successfully.

Figure 6.13 (see page 133) shows the orientation map pixel count at every degree in the orientation range, between  $\pm 90^\circ$ . The orientation map from both the models for the illusory stimuli generated by AGIS are shown in Figure 6.12 (see page 132). It is proposed that the most generic measure is the comparison of the number of pixels at the illusory orientation. These results can be seen by examining the counts compared for stimulus 01 and 28 in Figure 6.13 (see page 133). We assume the illusions are not caused by flanking behaviour which would not be demonstrated with this approach.

The results in Figure 6.13 (see page 133) show that GROM has a much larger population of pixels — for both stimuli GROM has 45 pixels, while SFEM has 13 at the illusory orientation for both illusory stimuli ( $1^\circ$  and  $2^\circ$ ). Therefore, GROM correctly measures the illusory nature of the selected stimuli and the SFEM is unable to account for its misperception.

### 6.5 AGIS Summary

AGIS, a novel method of generating textures that contain illusions of orientation has been presented in this chapter. Instances of an implementation of the method have generated multiple stimuli, of which a subset, three textures, were selected for psychophysical testing.

Selected stimuli were presented to naive subjects in original and horizontally flipped<sup>3</sup> form and

---

<sup>3</sup>equivalent to a horizontal reflection

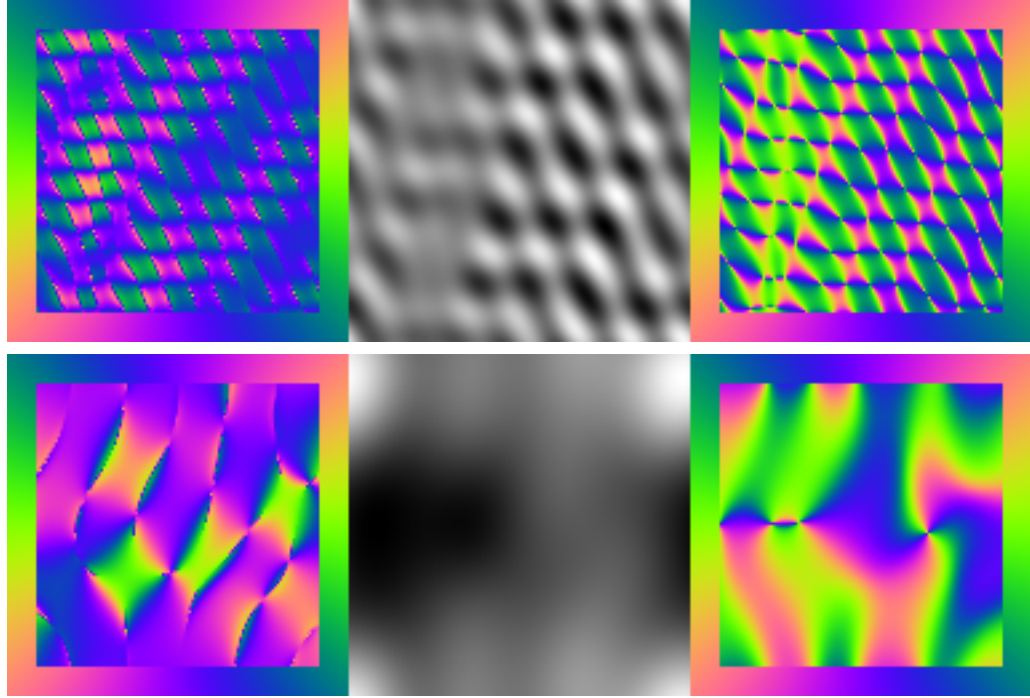


Figure 6.12: A figure showing the orientation maps for the AGIS generated 128 pixels square stimuli 01 (top) and 28 (bottom). The figures show GROM output (left) the stimulus (middle) and the result of the SFEM (right). Simple inspection of these maps does not easily show which model is best able to account for the illusory nature of the stimuli and which has not.

were asked to judge the alignment of the near-vertical structure. For two of the stimuli, 01 and 28, results of psychophysical testing showed that they contain illusions of orientation.

In the centre circle, each psychophysical stimulus contains a maximum of 3 tessellated versions of the generated stimuli. Therefore, it cannot be argued that the non-vertical psychophysical orientation match is caused by non-vertical stimulus structure - as they tessellate on a rectangular grids. This argument is demonstrated in Figure 6.14 (see page 134).

In the section entitled ‘Preferential model selection’ (see page 131) data was presented which indicates that GROM detects the illusory contours in stimulus 01 and 28, whereas the SFEM does not.

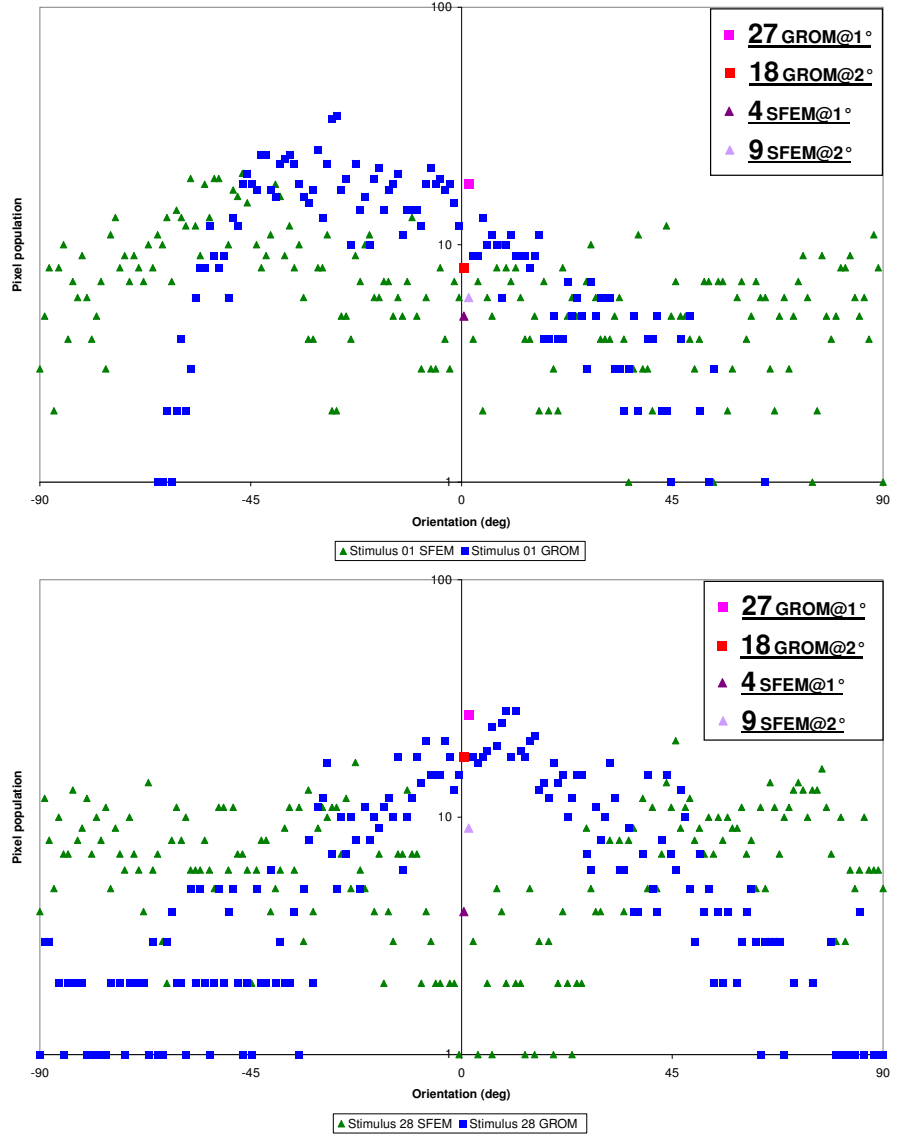


Figure 6.13: Two graphs showing the number of pixels at each angle in the orientation map output of the illusory stimuli generated by AGIS. The abscissa, measured on a log scale, indicates the population at a pixel and the ordinate indicates the orientation measured clockwise from the vertical. The orientation range,  $-90^\circ$  to  $+90^\circ$  is divided into 180 population bins, making the first two bins, highlighted with count in the inset in each case,  $1^\circ$  and  $2^\circ$  respectively. Both stimuli 01 (top) and 28 (bottom) show a higher pixel count for GROM at the illusory orientation found in the psychophysics,  $\approx 1^\circ - 2^\circ$ . Results have not been shown for stimulus 30 as it was not found psychophysically to be illusory.

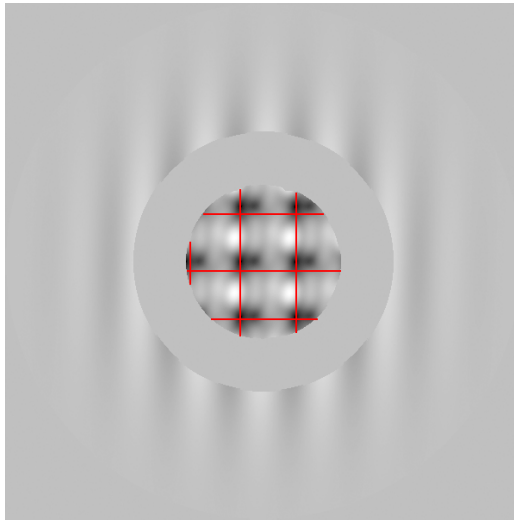


Figure 6.14: This figure shows the psychophysical stimulus with the outer matching ring aligned with the average orientation found in the psychophysics. Overlaid in red are the lines showing the where the original generated stimuli are joined. This demonstrates that the overall structure in the stimuli is vertically aligned and that a veridical slant in the structure is not responsible for the perceived slant.

## Chapter 7

# Conclusions

This chapter contains a summary of the main points arising from the different research themes detailed in this thesis. There is also a comprehensive list of the novel aspects of the research. The chapter is arranged in a series of sections dealing with the main research areas individually. This is followed by a Future Work section that discusses some of the interesting and potentially novel work that could follow from this research. Remarks that conclude the thesis are found at the end.

### 7.1 Main model results

The chapter entitled ‘Model Results’ (see page 60) covers the model’s results when applied to elementary visual stimuli. These results also show the effect of varying model parameters. The main conclusions that can be drawn from results where the model is applied to these simple stimuli are as follows:

- as demonstrated in Figure 3.3 (see page 63) and Figure 3.4 (see page 64), increasing the length of the  $k$ -vector - increasing the number of derivatives used in the least-squares fit - improves performance with simple sine-wave stimuli. The orientation map is non-uniform in the case where  $k = 1$  with orientation mis-measure evident where the sine-wave stimulus is at minima and maxima. This is due to the ill-conditioning of the simple ratio in the  $k = 1$  case. Where  $k > 1$ , this problem does not occur. This provides a possible functional explanation for the existence of cortical cells implementing higher-order derivatives found in the visual system [Young *et al.*, 2001].
- as demonstrated in Figure 3.4 (see page 64) and Figure 3.5 (see page 65), favouring measures from reference frames oblique to the underlying structure improves results. Where



oblique measures are not favoured local orientation mis-measure is evident along reference frame aligned radial lines in the orientation map. This property of the model can only be observed in cases where degenerate 1D local stimulus structure is aligned with one of the 12 reference frames. The model overcomes this problem by using a weighting function that favours least-squares estimates (and therefore filter measures) taken in reference frames oblique to the structure itself.

- Figure 3.6 (see page 66) shows the GROM output orientation as a function of spatial frequency for a sine-wave stimulus with an arbitrary orientation. Below the Nyquist frequency of the model ( $\frac{128}{1.5}$ ), the orientation computed by the model is identical to the stimulus. However, above this spatial frequency, the orientation varies, an effect probably due to aliasing. It is not yet clear how this relates to aliasing in human vision.

This chapter also shows results from the model when applied to a natural image. The conclusions that can be drawn from these results are less quantitative, as there is no ‘ground truth’ to compare the model results to. There are also issues correctly scaling a natural image to the model, as  $1^\circ$  of visual angle only corresponds to a small area of a natural scene. However, it is still useful to undertake an informal observation of the results in Figure 3.7 (see page 70), and to note that increasing the length of the k-vector,  $k$  from  $k = 1$  to  $k = 2$  has an effect on the orientation map, whereas increasing to  $k > 2$  does not seem to effect the quality of the model output. It is possible that this reflects the distribution of the amount of high-order derivative information in this image, i.e. that this image contains information adequately encoded using mainly  $G_1$  and  $G_2$  filters.

As the stimulus used is a real image, it contains real image noise, rather than being synthetically generated and therefore noiseless, as with the stimuli discussed above. One might expect the use of higher-order differentials to adversely affect the orientation output as they are regarded as sensitive to noise. These results suggest this is not the case.

The effect of varying the model scale parameters for natural images is shown in Figure 3.8 (see page 71). From observation, increasing the scale parameter seems to ‘clean up’ the model result. Orientation detail can be observed around the major image detail rather than in the slowly changing regions. This result is to be expected considering the scaling of the image, using larger scales of filters in the model has the same overall effect as changing the zoom on the original image - smaller details, and therefore their orientation information, are lost in this process. Another observation from this figure is that increasing the largest scale in the multiple scale version of the model has little effect. This suggests, for GROM, that the lower scale is more dominant in determining the overall orientation. This could be related to the way the scales are combined in the proposed multiscale version of the model.

The effect of favouring oblique measures with a natural image can be seen in Figure 3.9 (see page 72). A comparison of the two computed orientation maps reveals a much wider variety of colours - meaning orientations - where oblique measures are favoured as demonstrated in the lower orientation map. Whereas the upper orientation map, where oblique measures are not favoured, seems to contain a smaller range of colours - and therefore orientations. This shows that favouring oblique measures is not only beneficial with synthetic stimuli, but also with natural images. Further investigation is required to ascertain the exact reasoning for this outcome.

In the subsection of the ‘Model Results’ (see page 60) chapter entitled ‘Machine vision comparison’ (see page 74), GROM’s noise performance is compared to an example of an orientation model from the field of Machine Vision - where models are not constrained by the requirements of biological plausibility. As shown in Figure 3.12 (see page 76), using the same test stimuli GROM compares favourably with the model from Feng and Milanfar with respect to noise susceptibility. The results show the average error in the orientation output caused by an increasing amount of random noise in the stimulus.

Where GROM’s parameters are constrained to those dictated by biology, the model shows a greater susceptibility to noise. However, Feng and Milanfar do not associate their model with any specific overall scaling - it is pixel based - this means they were free to find the parameters that work best with a particular test stimulus. Correspondingly, where the parameters of GROM are not constrained to those dictated by the scaling to the human visual system, with base scales of, at best, 2.5 pixels, GROM out-performs that presented by Feng and Milanfar in terms of noise performance.

Overall, this chapter shows GROM performs well over a range of simple synthetic and natural image stimuli. The results also provide some justification for higher-order cortical filters. However, in order to support specific claims of biological plausibility and perceptual authenticity it is necessary to apply the model to more complex stimuli that the human visual system is known mis-measure, ie. visual illusions.

## 7.2 Illusory stimuli

Chapters 4 and 5 detail experiments where GROM is applied to illusions of orientation comparing GROM results with results from psychophysical testing on the same illusory stimuli.

Such comparison between local computational model and psychophysical results is difficult. Subject judgments even in low-level visual task can still be influenced by higher-order processes, for example attending to particular areas of a stimulus to ascertain the local or global orientation of structures may involve higher-order attentional segmentations. Therefore, correlating the

results of psychophysical experiments (dealing with stimulus structure) to the value in GROM's computed orientation maps is non-trivial.

Testing of GROM, that is comparing local model and psychophysical orientation results allows some discussion of the assertion that the misperception of structure (e.g. slants in mortar or grey bars) is caused by the local computational strategy. In order to test this assertion, the relationship between local computed and perceived structural orientation should be based on as few assumptions as possible. Examining the distribution of local orientations within the area of a misperceived structure is one way of achieving this without resorting to a model of local orientation grouping.

### 7.2.1 The contrast-modulated stimulus

It is easily seen from the performance of GROM with simple sine wave stimuli that have uniform orientation, the local orientation is uniform and non-structured (see Figure 4.1 (see page 79) (top)). Where the local orientation is non-uniform, as for example in the circular sinusoidal stimuli in Figure 3.4 (see page 64), the local orientation reflects the orientation of the structure. Hence, for illusory structure, if GROM is to account for subject performance, it is assumed that the mis-perception is caused by local orientation mis-measure within that structure. Ergo, the specific orientation (or tendency thereof) of the structure measured psychophysically must be a component of the distribution of local orientations within that part of the stimulus structure found to be misperceived (e.g. slants in mortar or grey bars). This means, that in the case of the contrast-modulated grating (CM grating), the distribution of local orientations within low-contrast regions, as in Figure 4.1 (see page 79) (bottom), must be examined.

To ascertain the validity of this claim, the local orientations are 'binned' over a number of discrete orientation ranges in order to find the orientation distribution over the area of low-contrast region, the section entitled 'Population based analysis' (see page 82) details this procedure. Figure Figure 4.5 (see page 85) shows how the distribution of orientations within the low-contrast region of the CM grating varies with the orientation of the carrier. These results show the same tendency as the psychophysical results of McOwan and Johnston [McOwan and Johnston, 1996] - as the distribution contains a greater population of measure corresponding to the psychophysically measure illusory orientations. The way this distribution changes with the orientation of the carrier also corresponds with the tendencies in the psychophysical findings. Indeed, these results are supported by the fact that; for the two carrier orientations where no illusion of orientation is observed psychophysically, at parallel and perpendicular carrier orientations, the 'binned' distribution of orientations shows clear peaks at the modulation orientation, suggesting no mis-measure in the model output.

By comparison, ‘binning’ local orientations from the simple Fourier-energy model (SFEM) returns a distribution that is heavily populated at the orientation of the modulation - the veridical orientation of the stimulus, and so does not support an explanation of the measured illusion.

These results support a claim that GROM explains illusionary tendencies found in the contrast-modulated grating. The results suggest that measurements of local orientations within the structure of the contrast-modulated grating give rise to elements that bias the perception away from simply recovering the orientation of the modulation envelope.

### 7.2.2 The Café Wall stimulus

The Café Wall illusion was also examined using the same ‘binned’ distribution of orientations. Psychophysical results of Gregory and Heard show that the Café Wall stimulus is perceived as having slants in the horizontal mortar lines that alternate in sense between alternate mortar lines. Figure 5.7 (see page 96) shows how the tendency of these alternate sense slants is observed in the binned distribution.

As the brick offset between alternate rows is changed, informal observation shows that the slant of an arbitrary mortar line changes gradually between extremes over the cycle of two brick widths as in Figure 5.8 (see page 98). The illusory tilts change in the following fashion:

- maximal slant in one sense  $\nearrow$
- checkerboard with no perceived slant
- maximal slant in the opposite sense  $\searrow$
- alternate columns of black and white with no perceived mortar slant
- maximal slant in the original sense  $\nearrow$
- ...

The configuration of the bricks and mortar between alternate rows has the same arrangement - though translated horizontally - as between the two extremes of brick offset.

The aggregate population of pixels with local GROM orientations (with a sense) above horizontal, minus those with orientations (with a sense) below, varies as expected by observation and crosses over as expected between the brick offset extremes detailed above. A constant population (seen as a vertical offset in Figure 5.8 (see page 98)) can be seen in the aggregate. This points to the existence of pixels with orientations that do not vary with brick-offset and hence, perceived slant.

This supports the result shown in Figure 5.7 (see page 96) where the aggregate population is not seen to be equal and opposite between alternate rows in the maximal illusory case.

As a plausible explanation and at least in order to gain further insight into the Café Wall illusion, an attempt is made to understand the internal behaviour of the model at the illusory points identified in Figure 5.6 (see page 95).

Figure 5.9 (see page 102) and Table 5.2 (see page 101) show that rotating a simple filter at certain key points to show the variation of local gradient information with alignment angle is insufficient in explaining the correct illusory tendencies. Figure 5.10 (see page 104) further demonstrates that the correct and identifiable tendencies in the orientation map are only seen where  $k > 1$ , i.e. higher-order derivatives are included in the least squares fit.

Closely examining the behaviour of the least-squares ratio as it is rotated at key points, shows that points displaying the correct illusory tendency (illusory points) respond differently from those showing the veridical horizontal stimulus contour orientation in the mortar line (horizontal points). ‘Illusory points’ change as the length of the k-vector is changed; ‘horizontal points’ do not. For the ‘illusory points’, where  $k = 1$ , the least-squares ratio maps out a tangent function, whereas where  $k > 1$ , a tangent function is not observed (as in Figure 5.11 (see page 105)). This implies that the inclusion of higher-order derivatives is mediating GROM’s mis-measure of local orientation.

More insight into the source of the stimulus mis-measure is gained by examining what happens to the response to  $G_{1,0}$  and  $G_{2,0}$  filters used in the divisor of the ratio. In the case of the ‘horizontal points’, they are seen to be zero simultaneously, causing the ratio to be vertically asymptotic - the spikes found in the tangent function mapped out - as expected. However, the response to the  $G_{1,0}$  and  $G_{2,0}$  filters at the ‘illusory points’ are never simultaneously zero. Therefore no vertical asymptotes results, hence a tangent function is not generated over the different reference frames and final local orientation measure is slanted, which causes the mis-measurement and so arguably the illusion.

This argument is also presented diagrammatically in Figure 5.12 (see page 107), where a Café Wall stimulus marked with the tendencies implied by the local mis-measure of the model is shown in Figure 5.13 (see page 108). It is suggested that this same low-level reasoning can also be applied to other stimuli as a potential avenue for future work, as discussed in the following section entitled ‘Textures of Roncato’ (see page 144).

### 7.3 AGIS

AGIS details the development and testing of a novel method for automatically generating illusory stimuli using GROM. As asserted in the previous section entitled ‘The contrast-modulated stimulus’ (see page 138), the performance of the SFEM and GROM are different with the contrast-modulated stimulus. The local orientation distribution in the low-contrast regions of the CM grating for the SFEM is more veridical (like the stimulus orientation) whereas the for GROM it is mis-measured with the same tendency found psychophysically. Using an optimization technique, stimuli can be found that maximise the difference in local orientation measure between GROM orientation and that of the SFEM. Constraints placed on the space used by the optimization technique meant the type of stimuli generated could be governed; meaning only low spatial frequency tessellating patterns were generated.

The generated stimuli were tested psychophysically for illusions of orientation. For a selection of the generated stimuli a psychophysical stimulus was constructed. This tested stimulus consisted of multiples of the generated stimulus tessellated on a 3 by 3 square lattice in the centre - surrounded with an outer ring containing oriented sinusoidal structure (as seen in Figure 6.11 (see page 128)). The psychophysical task was to align the oriented structure in the outer ring to the AGIS generated structure in the centre. Two out of the three selected generated stimuli showed clear illusions of orientation.

These illusions were generated by maximising the difference between the output of GROM and the SFEM. That subjects do indeed perceive these generated stimuli as illusions gives further support to the psychophysical validity of GROM. Arguably, a requirement of any model seeking validity is that it is able to correctly predict previously unknown phenomena. These experiments add another aspect to the current second-order paradigm discussion in the literature as highlighted previously. The illusions exist because the SFEM is unable to correctly account for orientations in low-contrast regions of the contrast-modulated structure. It is in effect this baseline that facilitates the optimization technique in generating these illusions.

### 7.4 Areas of novelty

The broad ideas for GROM are based on the motion model developed by Johnston et al. [Johnston *et al.*, 1992]. However, GROM contains novel adaptations developed in order for it perform robustly in the space-space domain. Table 7.1 (see page 142) details the novel aspects of GROM.

Along with the novelties of GROM itself there are also novel aspects to the experimental

Section title	Function
‘Two-dimensional Tangent Measure’ (see page 45)	Justification of least-squares approach using approximations
‘Favouring oblique measures’ (see page 52)	Favouring least-squares measures from oblique reference frames
‘Multi-scale tangent estimation’ (see page 48)	Least-squares fit of tangent using filters with multiple scales
‘Deducing a final single measure’ (see page 53)	Trigonometric manipulation of the tangent measures from the different reference frames arriving at a final value of orientation.

Table 7.1: This table details the steps in the model that have been developed over the course of this research to make GROM robust. A logical consequence of this approach is that the developed GROM is capable of predicting a range of psychophysical phenomena.

framework and results presented in this thesis. These can be found listed in Table 7.2 (see page 143).

## 7.5 Future work

It is useful at this stage to highlight some areas of interest that could follow on from the research presented in this thesis.

### 7.5.1 Sparse sampling

As mentioned in the section entitled ‘Taylor/Jet representation’ (see page 43), GROM utilises image derivatives. The Taylor expansion is a convenient method for interpolating image surface values at a displaced position from the point where filters are applied. As per Konderink and van Doorn [Koenderink and van Doorn, 1988], the conceptualisation of a point in an image as a set of derivatives at a that point is referred to as a ‘Jet’.

One interesting property of image jets, that could be the subject of further study, is that the jet representation allows images to be sampled sparsely, and then reconstructed by a process of interpolation at all original image points based on the derivative information from a sparse selection of sample points. Image derivatives can also be computed at any point displaced from the centre of a sparse measure. GROM can be configured to operate on interpolated image derivatives rather than measured image derivatives. As the human visual system is known to rely

Section title	Function
‘GROM applied to contrast-modulated gratings’ (see page 77)	Successfully validating an explicit perceptual model (GROM) against psychophysical performance on an illusion of orientation (CM grating)
‘GROM applied to the Café Wall illusion’ (see page 88)	Successfully validating an explicit perceptual model (GROM) against psychophysical performance on a geometric illusion (the Café Wall illusion)
‘GROM applied to the Café Wall illusion’ (see page 88)	Giving a plausible low-level explanation of the illusory nature of a geometric illusion (the Café Wall illusion) using filters modelled on those found in the human visual system
‘Automatic generation of illusory stimuli’ (see page 111)	Comparing two perceptual models to automatically generate stimuli with illusions of orientation
‘Automatic generation of illusory stimuli’ (see page 111)	Automatic Generation of Illusory Stimuli. Developing a technique that simultaneously tests two perceptual models (GROM and the SFEM) and generates predictions as a result.

Table 7.2: This table details novel aspects of the experimental framework and results presented in this research.



on a mosaic of retinal photoreceptors feeding onto cortical filters, positioning within the early visual pathway is subject to biological variability rather than a regular array of filters based on a rectangular grid. By introducing the more biological constraint of sparseness and varying the distribution it could be possible to construct a more realistic model. Such a model might allow for the patching together and correlating of near-local information, and hence could give further insight into mechanisms supporting local orientation measure and grouping.

## 7.6 Textures of Roncato

Roncato studied larger geometrical illusions like the Café Wall by selecting individual illusory elements - effectively dividing up the Café Wall [Roncato, 2000]. A brief summary of this work is given here to motivate a further area of study that could prove worth investigation in future. In the first experiment performed, rectangular borders divide an inner and outer region of texture, as in Figure 7.1 (see page 148) (left). The textures were varied, as was the luminance of the rectangular border. The resulting psychophysical phenomena was a deformation of the rectangular border into a trapezoid. This was found to depend on the textures and the luminance of the border itself.

Based on these findings for simple elements, Roncato deduces the *overflow rule*. This rule states that the contour and neighbouring regions can be thought of as a stream of water and the banks it flows between. Where the luminance is lower than the darker region or higher than the lighter region, the water level can be thought of as being contained within the banks. When the contour luminance is between the neighbouring luminances the analogous water can be thought of as having burst its banks - therefore shifting into surrounding areas.

Roncato used these tendencies to construct more complicated illusions, as seen in Figure 7.1 (see page 148) (right). Which were also examined psychophysically, the basic finding being that the concave and convex tendencies in the patterns were not predicted based by the overflow rule.

A system of *local integrations* was proposed to explain the two sets of results. This led to a reasonably complicated overall framework to explain the results.

As potential future work, the ‘general rule’ proposed earlier in the Café Wall investigation (c.f. Figure 5.13 (see page 108)) could also be applied to the Roncato textures. Figure 7.2 (see page 149) shows the slant tendencies suggested by the ‘general rule’, overlaid as a red line. This predicts the nature of the trapezoidal perception found by Roncato for the two most illusory textures in the first experiment.

Figure 7.3 (see page 150) shows the same rule in application to the more complicated figures constructed by Roncato. Again, the ‘general rule’ (mentioned earlier in

‘Difference in zero-crossings of  $G_{2,0}$  and  $G_{1,0}$ ’ (see page 108)) seems to predict the tendency found psychophysically.

The heuristic (‘general rule’) derived in this thesis in the earlier investigation of a different stimulus seems able to predict the behaviour of the Roncato stimuli. Future work could systematically investigate the validity of this ‘rule’ and its accuracy in application to these stimuli, thus expanding the range of stimuli that can be handled by GROM.

### 7.6.1 Grouping on orientation

The difference between local results of GROM and psychophysical results on the estimates of structural orientation could be the subject of further research to investigate orientation grouping mechanisms. The elemental textures, such as those of Roncato discussed above in ‘Textures of Roncato’ (see page 144), could offer one method of approaching this important area. The informal predictions presented previously, based on GROM applied to these textures, if verified, might offer a method of explicitly examining the relationship between local mis-measure of orientation and global misperception of image structure (e.g. the low-contrast regions of the CM grating, the mortar lines of the Café Wall illusion).

Gaining understanding of the grouping process could be achieved by incrementally increasing the size of the image structures whilst, with a low-level understanding of the cause of local-orientation mis-measure, manipulating the local misperceptions of orientation. These results could map out points in a parameter-space with the following potential dimensions:

- distance between, and distribution of points of orientation mis-measure
- strength of orientation mis-measure
- tendency of orientation mis-measure
- size of structure that contains points of orientation mis-measure
- luminance of structure that contains point of orientation mis-measure
- resulting overall orientation of image structure

Experimenting with these parameters could give an indication of how these stimulus properties interact in the visual system to generate perceived orientation and stimulus distortion.

Thus far, GROM has only been able to show tendency that agrees with psychophysical results for a number of types of stimuli. An explicit model for grouping on local orientation measures - how

these local measures combine, would provide a computational model for structural orientation enabling direct comparison with psychophysical results.

The work in this thesis has shown that within the population of low-level local orientation measurements exists the signal that would produce the illusions. It is an as yet unanswered question of how and why the visual system extracts this signal, as suggested, modelling Roncato stimuli may provide a useful probe into this area.

## 7.7 Concluding remarks

In this thesis a novel model for local orientation measure in the human visual system has been developed, implemented, tested and discussed. The model, based on a contrast-invariant gradient-based measure, was shown to perform favourably when applied to a range of important visual stimuli. The results from GROM have been compared to psychophysical results. It is important to reiterate that a single model has accounted for all these results, and that parameters were held constant, rather than being tuned to explain particular psychophysical phenomena.

Although the model has scaling consistent with the human visual system, the current lack of an explicit model of local orientation grouping means that it was not possible to directly estimate the structural orientation measured psychophysically. However, the tendency within populations of local orientations calculated by the model was able to successfully predict the known perceptual phenomena.

The low-level reasoning behind local orientation mis-measure is explored by comparing illusory and non-illusory stimulus points in a Café Wall illusion. Results from this support the assertion that these illusions of orientation can be explained by the low-level local computational strategy employed in the human visual system rather than higher-level cognitive effects.

Claims of validity for a computational model should need to pass a number of hurdles. The results mentioned above form part-validation of GROM and its strategies, however, as an additional ‘model-predict-verify’ validation cycle, illusory stimuli were bred automatically (without human intervention or judgemental) using a Genetic Algorithm to breed stimuli based on the difference in orientation output between GROM and an SFEM. Psychophysical assessment of AGIS generated stimuli found that they contain illusions of orientation. GROM was found to successfully predict this behaviour, whilst the SFEM did not.

The model presented has been shown to be a good candidate for human perception of local orientation. In conjunction with the gradient based model for motion perception [Johnston *et al.*, 1992] GROM supports the argument that gradient based approaches are indeed the method employed in the human visual system, rather than the alternate Fourier methods.

GROM is arguably able to account for first- and second-order orientation stimuli, with one common mechanism. This is more parsimonious than approaches that rely on additional channels, such as specialist rectifying second-order channels that exist in Fourier-based motion and orientation models [Lu and Sperling, 1995b, Smith *et al.*, 2001].

GROM also could also be used in machine vision applications - it has been shown that it compares favourably with a current machine vision algorithm for orientation extraction. This supports the idea that a computational study of biological vision systems, in addition to giving insight into the human visual system and neuroscience in general, can also assist in developing better computer vision technologies.

The results of further investigation using GROM may well give insight into grouping mechanisms and lead on to a more complete perceptual model. Furthermore, the robustness afforded by the biological basis of GROM will hopefully lead to viable real-world applications.

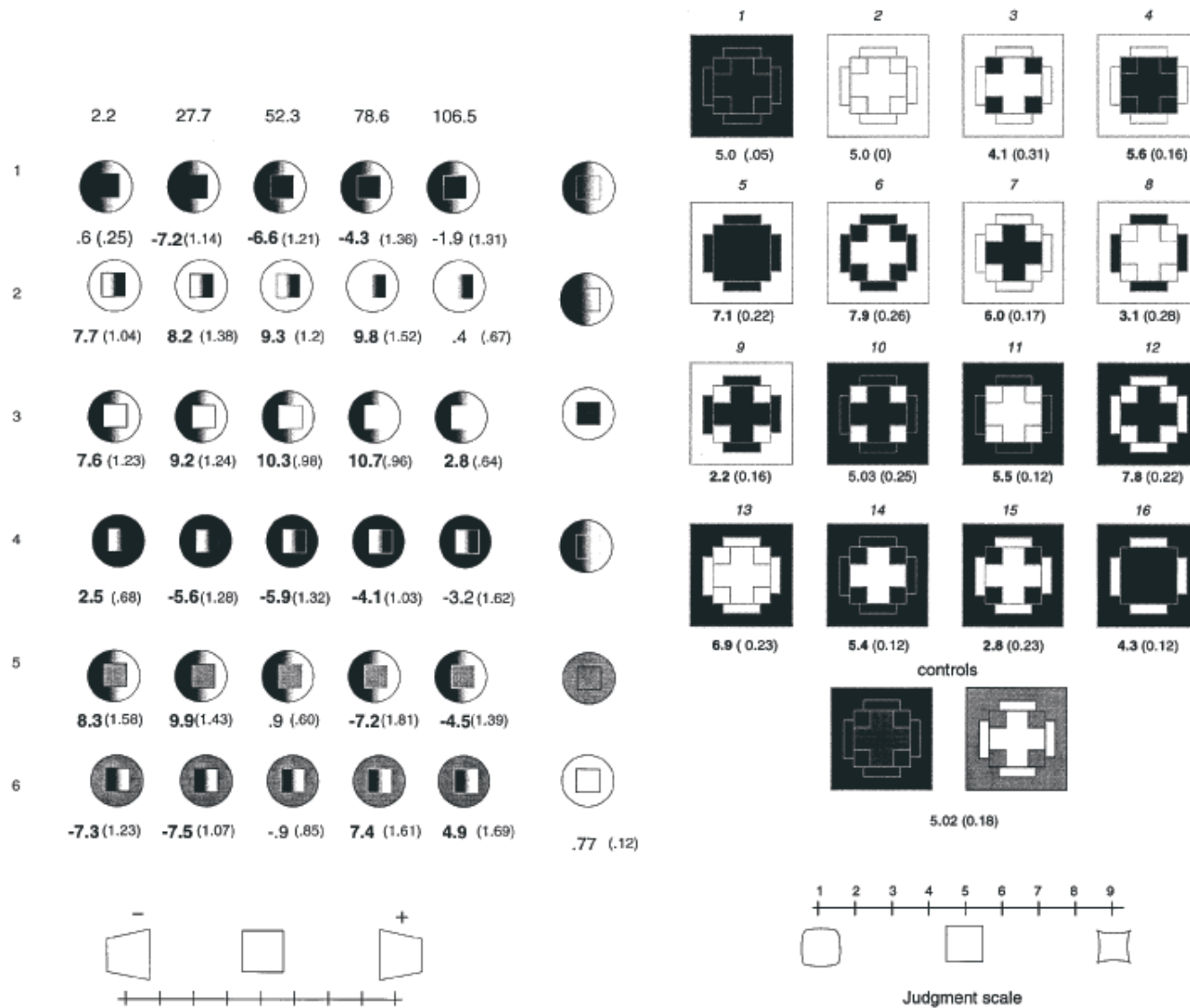


Figure 7.1: The two main figures from Roncato showing trapezoidal illusions (left) and concave/convex illusions right. The values in bold are the psychophysical estimates of strength of each illusion, based on the scale at the bottom of each figure. Adapted from Roncato [Roncato, 2000].

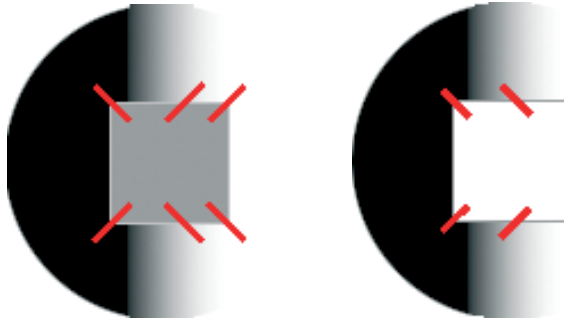


Figure 7.2: (left) A pattern taken from the column 4 of row 5 of Figure 5 in Roncato [Roncato, 2000]. This pattern was found through a psychophysical matching experiment to have a trapezoidal perceived shape, small at the left vertical edge, large at the right - negative on the scale shown in Figure 7.1 (see page 148). The overlaid red lines, were applied according to a possible ‘general rule’ proposed earlier (c.f. Figure 5.13 (see page 108)). In the centre of the top horizontal bar of the rectangular border a red line with the sense ‘/’ was chosen as there is a lighter border than there would be in the path of a line drawn with the other sense, ‘\’. It is proposed that the addition of the two-like (middle and right-hand corner) tendencies counteracts the one opposite tendency, but weakens the overall effect. (right) A pattern taken from column 3 of row 3 in Figure 5 from the same source. This pattern was found to be perceived as having positive trapezoidal shape (opposite to the figure left). The overlaid red lines, were applied according to the same rule discussed above. The slant implied by the two lines causes a large trapezoidal deformation in this case.

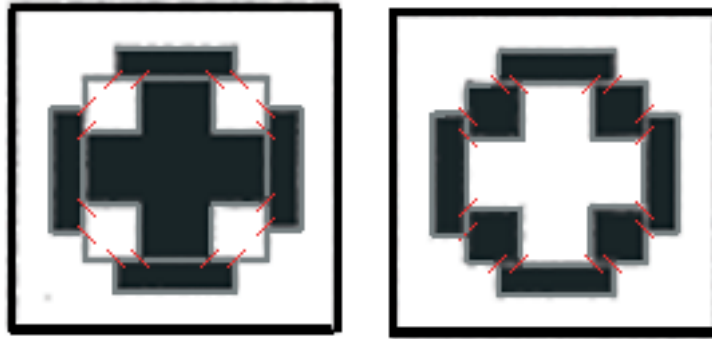


Figure 7.3: (left) Pattern 9 from Figure 9 - an experimental pattern constructed by Roncato, confirmed psychophysically as being perceived convex. (right) Pattern 6 from Figure 9 - an experimental pattern constructed by Roncato [Roncato, 2000], found in psychophysical investigation to be perceived concave. The red bars overlaid are from the application of a ‘general rule’ derived from work where GROM was applied to the Café Wall illusion. These red-line overlays predict the correct tendencies. It is argued that the rule proposed above is particularly relevant in relation to these patterns as subsections of the grey border lines in both patterns are equivalent to subsections of the Café Wall. The lines correctly predict the concavity and convexity in each case.

# Appendix A

## Gaussian derivative synthesis

Simple cortical receptive fields (RFs) are a basic measurement unit of the visual system [Hubel and Wiesel, 1968]. Banks of RFs are known to exist in the early stages of human visual cortex. Young described a biologically plausible method of how higher DOGs can be implemented by taking weighted differences between variably offset simple Gaussian profiles [Young *et al.*, 2001]. Figure A (see page 152) shows some examples of this method, with analytic and DOOG methods compared on the same axis. Equations A.1 show how the first four Gaussian derivative profiles are implemented.

$$\begin{aligned}
 G_0(x) &= DOOG_0 = G(x) \\
 G_1(x) &\cong DOOG_1 = \frac{G(x+\Delta x) - G(x-\Delta x)}{\Delta x} \\
 G_2(x) &\cong DOOG_2 = \frac{G(x+2\Delta x) - 2G(x) + G(x-2\Delta x)}{\Delta x} \\
 G_3(x) &\cong DOOG_3 = \frac{G(x+3\Delta x) - 3G(x+\Delta x) + 3G(x-\Delta x) - G(x-3\Delta x)}{\Delta x} \\
 G_4(x) &\cong DOOG_4 = \frac{G(x+4\Delta x) - 4G(x+2\Delta x) + 6G(x) - 4G(x-2\Delta x) + G(x-4\Delta x)}{\Delta x}
 \end{aligned} \tag{A.1}$$

$$\begin{aligned}
 &\int_{-\infty}^{\infty} \left| G_1(x) - \frac{(G_0(x+\frac{\tau}{2}) - G_0(x-\frac{\tau}{2}))}{\tau} \right| dx \\
 &= 2 \int_0^{\infty} \left( \frac{e^{-\frac{(x-\frac{\tau}{2})^2}{2\sigma^2}}}{\tau\sigma\sqrt{2\pi}} - \frac{e^{-\frac{(x+\frac{\tau}{2})^2}{2\sigma^2}}}{\tau\sigma\sqrt{2\pi}} - \frac{xe^{-\frac{x^2}{2\sigma^2}}}{\sigma^3\sqrt{2\pi}} \right) dx \\
 &= 2 \left( \frac{1}{\sigma\sqrt{2\pi}} - \frac{\frac{2}{\sqrt{\pi}} \int_0^{\frac{\tau}{2\sigma\sqrt{2}}} -x^2 dx}{\tau} \right)
 \end{aligned} \tag{A.2}$$

Equation A.2 shows the working for the difference between one-dimensional, first DOG and DOOG filters summed between plus and minus infinity. This shows the dependency of the ‘error’,



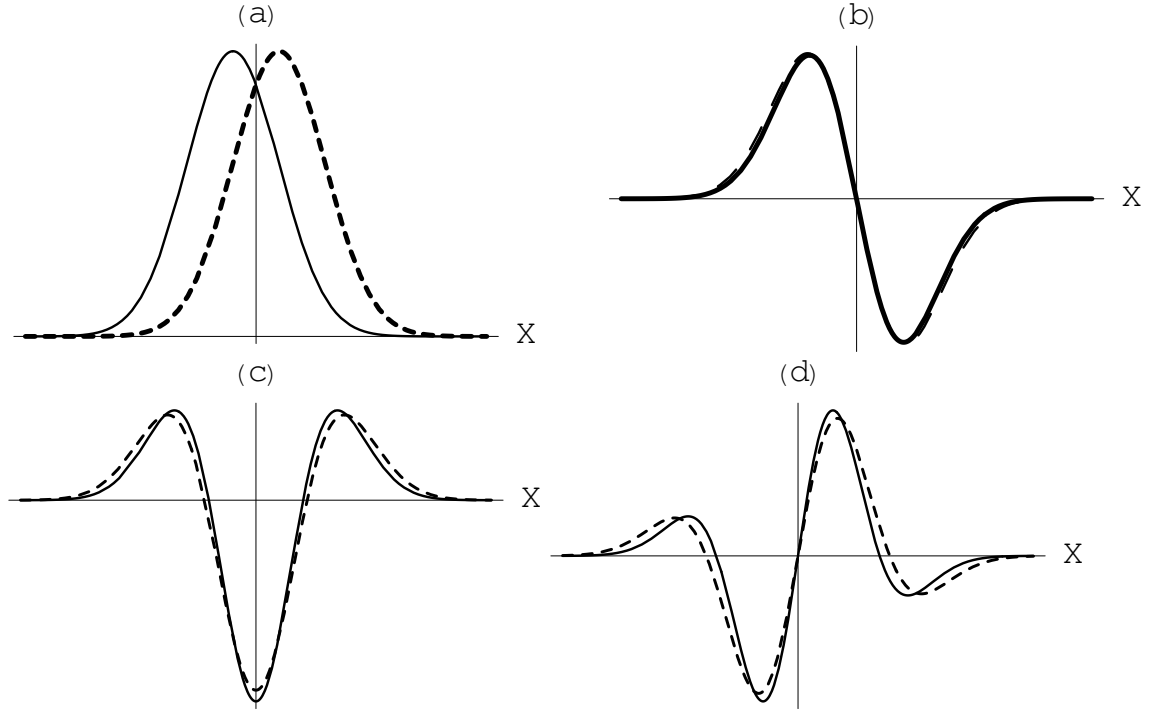


Figure A.1: (a) Two Gaussian profiles offset. (b) The result of taking the difference between the two offset profiles in (a) (dashed line) and a first derivative of Gaussian formed using the analytic equation (solid line). (c) A second derivative of Gaussian,  $G_2$ , formed using the equation (solid line) and DOOG method. (d) Analytic (solid line) and DOOG (dashed line) method  $G_3$ .

i.e. the sum of the difference between the two functions, on the offset,  $\tau$ , and the standard deviation,  $\sigma$ , of the Gaussian distributions Figure A (see page 153). This shows that as the offset  $\tau$ , i.e. separation between the centres, of the distributions increases, the error increases; and as the standard deviation,  $\sigma$ , increases the error decreases.

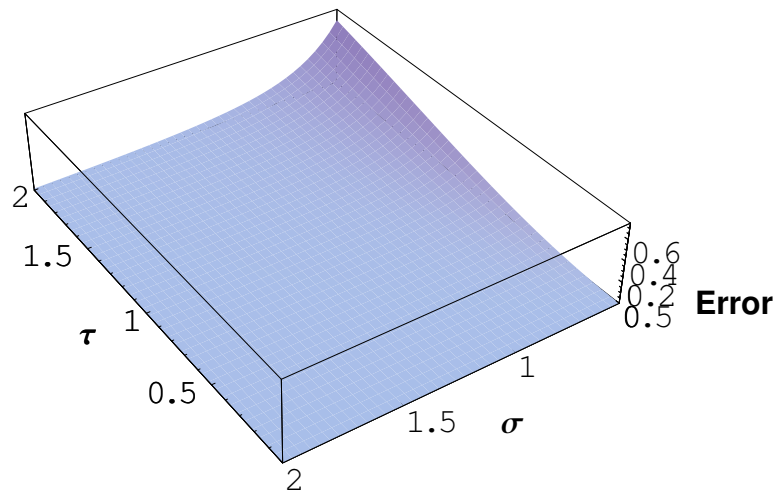


Figure A.2: A graph showing the dependence of the ‘error’ between simple first-order DOG and DOOG filters on the separation,  $\tau$ , and standard deviation,  $\sigma$ , of the distributions.

## Appendix B

### Image derivatives

It can be stated that, where an image  $I$  is convolved with a Dirac delta function  $\delta$  [Boas, 1993, Weisstein, 2004]:

$$I \rightarrow \delta * I \quad (\text{B.1})$$

where the symbol  $*$  represents a continuous one-dimensional convolution. A Gaussian distribution,  $G(\sigma) \rightarrow \delta$  as  $\sigma \rightarrow 0$ . So, at a scale  $\sigma$ , an image can be approximated such that, as  $\sigma \rightarrow 0$ :

$$I \simeq G(\sigma) * I \quad (\text{B.2})$$

where  $G(\sigma)$  represents a Gaussian distribution with a non-zero standard deviation,  $\sigma$ , this operation results in a blurring of the image. Differentiating both sides with respect to a spatial dimension  $x$ , gives [Bracewell, 1986]:

$$I' = \frac{\partial}{\partial x} (G(x, \sigma) * I) = \frac{\partial}{\partial x} G(x, \sigma) * I \quad (\text{B.3})$$

this means that the derivative of any function can be found by convolving the function with the derivative of a Gaussian distribution, however there is a blur incurred, introduced by the finite width of the Gaussian,  $\sigma$ .

## Appendix C

### Gaussian derivatives

In one dimension, a Gaussian or Normal distribution has the form:

$$G(x, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (\text{C.1})$$

where  $\sigma$  is the standard deviation and  $\mu$  is the mean of the distribution. The first derivative of a Gaussian distribution is:

$$\frac{\partial}{\partial x} (G(x, \sigma)) = \frac{x}{\sigma^2 \sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (\text{C.2})$$

and generally a derivative order  $n$ , has the following form:

$$\begin{aligned} G_n(x, \sigma) &= \left(\frac{-1}{\sigma\sqrt{2}}\right)^n H_n\left(\frac{x}{\sigma\sqrt{2}}\right) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ &= \left(\frac{-1}{\sigma\sqrt{2}}\right)^n H_n\left(\frac{x}{\sigma\sqrt{2}}\right) G_0(x, \sigma) \end{aligned} \quad (\text{C.3})$$

where  $H_n$  is a Hermite polynomial [Boas, 1993, Weisstein, 2004] of order  $n$ , given by:

$$H_n(x) = (-1)^n e^{x^2} \frac{\partial^n}{\partial x^n} (e^{-x^2}) \quad (\text{C.4})$$

the first five Hermite polynomials are, where  $n$  is zero to four:

$$\begin{aligned}
H_0(x) &= 1 \\
H_1(x) &= 2x \\
H_2(x) &= 4x^2 - 2 \\
H_3(x) &= 8x^3 - 12x \\
H_4(x) &= 16x^4 - 48x^2 + 12
\end{aligned} \tag{C.5}$$

For two dimensions a Gaussian centred at the origin can be written as:

$$G(x, y, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{C.6}$$

and a Gaussian of derivative order  $n$  in the  $x$ -direction,  $m$  in the  $y$ -direction becomes:

$$G_{n,m}(x, y, \sigma) = \left( \frac{-1}{\sqrt{2\pi\sigma^2}} \right)^{n+m} H_n \left( \frac{x}{\sqrt{2\sigma^2}} \right) H_m \left( \frac{y}{\sqrt{2\sigma^2}} \right) G_0(x, y, \sigma) \tag{C.7}$$

## Appendix D

# Convolution and Separability

By inspection, Equation C.6 and Equation C.7 are x-y separable, meaning that they can be divided into factors that are variables in either x or y exclusively. This section will show that the properties of convolution and the separability of the two-dimensional Gaussian can lead to computational simplification of the convolution process.

Where  $**$  represents a continuous two-dimensional convolution, in full it can be written:

$$f ** g = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x', y') g(x - x', y - y') dx' dy' \quad (D.1)$$

Convolving with a separable function (Equation C.7) we have:

$$\begin{aligned} f(x, y) ** G_{n,m}(x, y, \sigma) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x', y') G(x - x', y - y', \sigma) dx' dy' \\ &= -1^{n+m} \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^{n+m+1} \int_{-\infty}^{\infty} H_m \left( \frac{y-y'}{\sqrt{2\sigma^2}} \right) e^{-\frac{(y-y')^2}{2\sigma^2}} \left[ \int_{-\infty}^{\infty} f(x', y') H_n \left( \frac{x-x'}{\sqrt{2\sigma^2}} \right) e^{-\frac{(x-x')^2}{2\sigma^2}} dx' \right] dy' \\ &= G_{0,m}(0, y) * [G_{n,0}(x, 0) * f(x, y)] \end{aligned} \quad (D.2)$$

where  $*$  is a one dimensional convolution.

It can be seen from Equation D.2 that the two-dimensional convolution process can be split into two parts; firstly a convolution in one direction, arbitrarily chosen as x, with a Gaussian derivative of order n then the result can be convolved in the other direction, y, with a derivative of order m, varying only in the y-direction.

This separability is useful in implementation, it means that the vertical and horizontal filtering

processes can be separated. Rather than performing a two-dimensional convolution with a square kernel, side length  $n$ , where the number of operations would be  $O(n^2)$ ; two one-dimensional convolutions can be performed, one on the original image and another in the orthogonal direction on the result, with the number of operations reduced to  $O(2n)$ .

## Appendix E

# Steerable Filtering

Freeman and Adelson [Freeman and Adelson, 1991] define the term ‘Steerable Filter’ as follows:

“... a class of filters in which a filter of arbitrary orientation is synthesized as a linear combination of a set of basis filters.”

The scalar product is used to find the component of one vector in the direction of another. To find the component of differentiated image in the direction of  $\bar{u}$ , a unit vector aligned at an angle  $\theta$  to the horizontal, the following must be computed:

$$\begin{aligned}\nabla\phi(x, y) \cdot \bar{u}(\theta) &= \begin{bmatrix} \frac{\partial\phi(x, y)}{\partial x} \\ \frac{\partial\phi(x, y)}{\partial y} \end{bmatrix} \cdot \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \\ &= \frac{\partial\phi(x, y)}{\partial x} \cos \theta + \frac{\partial\phi(x, y)}{\partial y} \sin \theta\end{aligned}\tag{E.1}$$

This can be rewritten in terms of Gaussian image derivatives as:

$$D_\theta(\phi(x, y)) = G_x \cos \theta + G_y \sin \theta\tag{E.2}$$

where D represents a derivative operator aligned at an angle  $\theta$  and  $G_x$  and  $G_y$  represent the image convolved with separable rectangularly aligned derivative of Gaussian filters. To find higher order derivatives, one can recursively apply the logic of Equation E.1, such that:

$$\begin{aligned}\nabla(D_\theta) \cdot \bar{u}(\theta) &= \nabla(G_x \cos \theta + G_y \sin \theta) \cdot \bar{u}(\theta) \\ &= \begin{bmatrix} G_{2x} \cos \theta + G_{xy} \sin \theta \\ G_{xy} \cos \theta + G_{2y} \sin \theta \end{bmatrix} \cdot \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \\ &= G_{2x} \cos^2 \theta + 2G_{xy} \cos \theta \sin \theta + G_{2y} \sin^2 \theta\end{aligned}\tag{E.3}$$



where the final step in Equation E.3 equates  $G_{xy}$  and  $G_{yx}$  as partial differentiation is commutative [Boas, 1993, Weisstein, 2004]. Similarly to the step in Equation E.1, the component of a rectangularly aligned derivative parallel to a unit vector  $\bar{u}_\perp$  can be found, such that if:

$$u_\perp(\theta) = \begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix} \quad (\text{E.4})$$

then:

$$\nabla\phi(x, y) \cdot \bar{u}_\perp(\theta) = \begin{bmatrix} G_x \\ G_y \end{bmatrix} \cdot \begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix} = -G_x \sin \theta + G_y \cos \theta \quad (\text{E.5})$$

Applying the scalar product to the result given in Equation E.2, the rotated equivalent of a crossed derivative would be:

$$\begin{aligned} & \nabla(\nabla\phi(x, y) \cdot \bar{u}(\theta)) \cdot \bar{u}_\perp(\theta) \\ &= \nabla(G_x \cos \theta + G_y \sin \theta) \cdot \bar{u}_\perp(\theta) \\ &= \begin{bmatrix} G_{2x} \cos \theta + G_{xy} \sin \theta \\ G_{xy} \cos \theta + G_{2y} \sin \theta \end{bmatrix} \cdot \begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix} \\ &= \cos \theta \sin \theta (G_{2y} - G_{2x}) + G_{xy} (\cos^2 \theta - \sin^2 \theta) \end{aligned} \quad (\text{E.6})$$

## Appendix F

# Steerable convolution

It is demonstrated above in ‘Convolution and Separability’ (see page 157) that a set of steered filters can be generated from linear combinations of filters in a rectangularly aligned ‘basis set’. It is however possible to implement a further computational saving by utilising the distributive nature of convolution:

$$\begin{aligned} I_{n,m}^\theta &= G_{n,m}^\theta \otimes I \\ &= (aG_{x,0}^0 + bG_{0,y}^0 + \dots) \otimes I \\ &= aG_{x,0}^0 \otimes I + bG_{0,y}^0 \otimes I + \dots \\ &= aI_{x,0}^0 + bI_{y,0}^0 + \dots \end{aligned} \tag{F.1}$$

where  $I_{n,m}$  is the resulting image filtered by  $G_{n,m}$ , a Gaussian filter of derivative order  $n$  in the  $x$ -direction and  $m$  in the  $y$ -direction, steered to an angle  $\theta$ ;  $a$  and  $b$  are the steering parameters,  $I$  is the original image and  $\otimes$  represents a discrete two-dimensional convolution.

The last step of Equation F.1 demonstrates that linear combinations of the image convolved with the basis set can be used to synthesize a final convolution result. The computational saving is maximised in situations where images are convolved with filters at many angles. These can be synthesized from the result of the original image convolved with the basis set.

## Appendix G

### Reference frame rotation

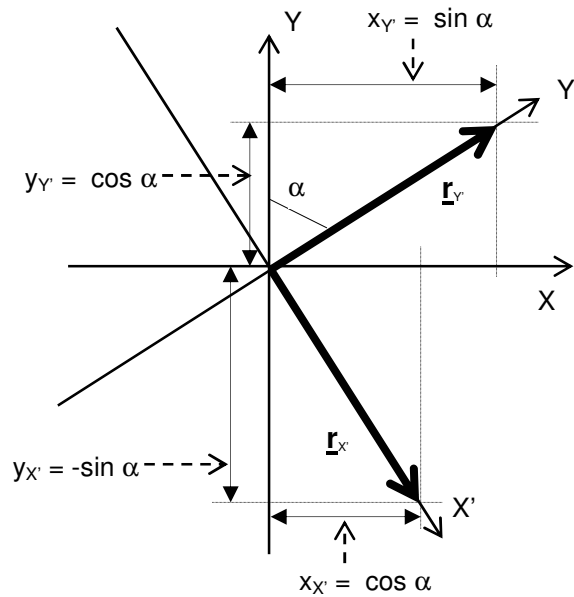


Figure G.1: A figure showing two Cartesian reference frames,  $X-Y$  and  $X'-Y'$  rotated through an angle  $\alpha$ .

Figure G.1 (see page 162) shows two Cartesian reference frames,  $X-Y$  and a reference frame rotated by an arbitrary angle  $\alpha$ , clockwise from vertical,  $Y$ . Equation G.1 describes unit vectors lying on the  $X'-Y'$  axes when mapped to the  $X-Y$  reference frame. Where  $R$  is a vector in the

fiducial frame and  $R'$  is a vector in a rotated frame:

$$\begin{aligned}
 \underline{R} &= \begin{bmatrix} x \\ y \end{bmatrix}_{X-Y} = ar_{x'} + br_{y'} \\
 &= a \begin{bmatrix} \cos \alpha \\ -\sin \alpha \end{bmatrix} + b \begin{bmatrix} \sin \alpha \\ \cos \alpha \end{bmatrix} \\
 &= \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \\
 &= \underline{C}_\alpha \underline{R}'
 \end{aligned} \tag{G.1}$$

Using the rotation matrix,  $C_r$ , it is possible to find the equation of a line in a rotated reference frame in the co-ordinate system of the X-Y reference frame. As a result of the trigonometric transformations performed in Equation 2.20 and Equation 2.21 columns have an angular separation of  $2\phi$ , where  $\phi$  is the separation between orientation hyper-columns. The process of repeating measures in multiple reference frames described above leads to a set of measures made in rotated reference frames. Using Equation G.1 these measures can be compared in a reference Cartesian frame by applying the rotation  $C_{2i\phi}$  to each measure.

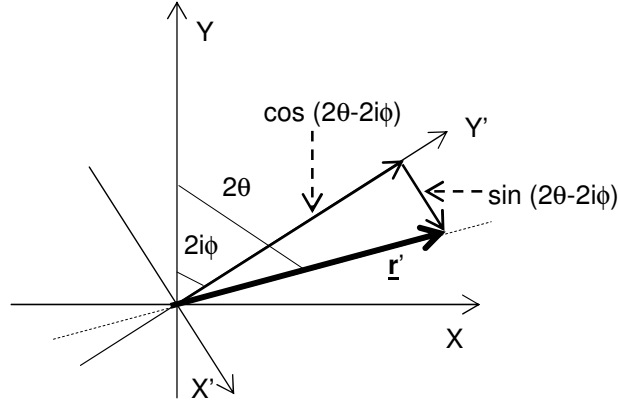


Figure G.2: A diagram showing two Cartesian frames; one reference, one rotated to  $2i\phi$  and a unit vector formed of the two measures derived in Equation 2.22 and Equation 2.25.

A unit vector in a rotated X'-Y' reference frame can be constructed from the two measures deduced above in Equation 2.22 and Equation 2.25:

$$\underline{R}'_{2i\phi} = \begin{bmatrix} \cos(2\theta - 2i\phi) \\ \sin(2\theta - 2i\phi) \end{bmatrix} \tag{G.2}$$

This vector can be rotated into the standard X-Y reference frame using Equation G.1, where  $i$  is the  $i^{th}$  measure performed in a reference frame rotated to  $2i\phi$ :

$$\begin{aligned} \underline{R} = \begin{bmatrix} \cos(2\theta) \\ \sin(2\theta) \end{bmatrix} &= \underline{C}_{2i\phi} \underline{R}'_{2i\phi} = \begin{bmatrix} \cos(2i\phi) & \sin(2i\phi) \\ -\sin(2i\phi) & \cos(2i\phi) \end{bmatrix} \begin{bmatrix} \cos(2\theta - 2i\phi) \\ \sin(2\theta - 2i\phi) \end{bmatrix} \\ \begin{bmatrix} \cos(2\theta) \\ \sin(2\theta) \end{bmatrix} &= \begin{bmatrix} \cos(2i\phi) \cdot \cos(2\theta - 2i\phi) + \sin(2i\phi) \cdot \sin(2\theta - 2i\phi) \\ -\sin(2i\phi) \cdot \cos(2\theta - 2i\phi) + \cos(2i\phi) \cdot \sin(2\theta - 2i\phi) \end{bmatrix} \end{aligned} \quad (\text{G.3})$$

Using Equation G.3, the multiple measures in different reference frames, derived in Equation 2.22 and Equation 2.25, estimations of orientation derived above can be compared in the same reference frame;

$$\begin{bmatrix} \cos(2\theta) \\ \cos(2\theta - 2\phi) \\ \vdots \\ \cos(2\theta - 2(n-1)\phi) \end{bmatrix} \rightarrow \begin{bmatrix} \cos(2\theta) \\ \cos(2\theta) \\ \vdots \\ \cos(2\theta) \end{bmatrix} \quad (\text{G.4})$$

and;

$$\begin{bmatrix} \sin(2\theta) \\ \sin(2\theta - 2\phi) \\ \vdots \\ \sin(2\theta - 2(n-1)\phi) \end{bmatrix} \rightarrow \begin{bmatrix} \sin(2\theta) \\ \sin(2\theta) \\ \vdots \\ \sin(2\theta) \end{bmatrix} \quad (\text{G.5})$$

## Appendix H

# Simple Fourier-energy model implementation

This chapter gives the details of the simple Fourier-Energy model (SFEM) used for comparison purposes in ‘GROM applied to contrast-modulated gratings’ (see page 77) and later in the chapter about AGIS, entitled ‘Automatic generation of illusory stimuli’ (see page 111). The implementation detailed below comes from Atherton [Atherton, 2002], which is carried forward from Langley and Atherton [Langley and Atherton, 1991].

This model for finding orientation works by filtering an image,  $I(x, y)$ , with a series of rotated Gabor filters,  $\psi_k(x, y)$ , given by:

$$\psi_k(x, y) = (\cos(2\pi u(\cos(\theta_k)x + \sin(\theta_k)y)) + i \sin(2\pi u(\cos(\theta_k)x + \sin(\theta_k)y))) e^{-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}\right)} \quad (\text{H.1})$$

leading to the energy response,  $E_k(x, y)$ , given by:

$$E_k(x, y) = I(x, y) \otimes \psi_k(x, y) \quad (\text{H.2})$$

These M energy responses are combined to arrive at a single measure,  $\theta_{E2}$  for orientation at each point (x, y), given by:

Parameter	symbol	value
Number of orientations	M	16
Spatial frequency	u	$\frac{128}{1.5}$ cpd
Standard deviation of Gaussian envelope	$\sigma_x, \sigma_y$	1.5 pixels

Table H.1: This table details the parameters and their values for the SFEM.

$$\theta_{E2}(x, y) = \frac{1}{2} \tan^{-1} \left( \frac{\sum_{k=1}^M E_k(x, y) \sin(2\theta_k)}{\sum_{k=1}^M E_k(x, y) \cos(2\theta_k)} \right) \quad (\text{H.3})$$

where  $\theta_k$  is the orientation of the reference frame.

Table H.1 (see page 166) details the parameters used in the SFEM. these values are largely in keeping with those used by Langely and Atherton [Langely and Atherton, 1991] and later by Atherton [Atherton, 2002]. However, the Gaussian envelope used has the same standard deviation as that used in the GROM.

# Appendix I

## Minimisation

This section describes the mathematical process of *minimisation*. Use is made of this process to in the derivation of the model to find the direction of maximum change on an arbitrary surface. This process will first be described via a simple one dimensional example and then further expanded upon with an example of known surface in two dimensions.

Minimisation is the generic term given to this process, however, this process can also be used to find the maximum points in an equation. Hence this process can interchangeably be referred to as *maximisation*.

### I.1 An example in one dimension

The red curve in Figure I.1 (see page 168) has the following equation:

$$y = (x - 1)^2 \tag{I.1}$$

If continuous, where a function reaches a minimum or maximum it becomes horizontal. The gradient of a horizontal part of a function is zero. Therefore, finding the points where the gradient, or derivative of a function is zero is where the function is a maximum or minimum. This procedure can be applied to find the maximum or minimum of Equation I.1:



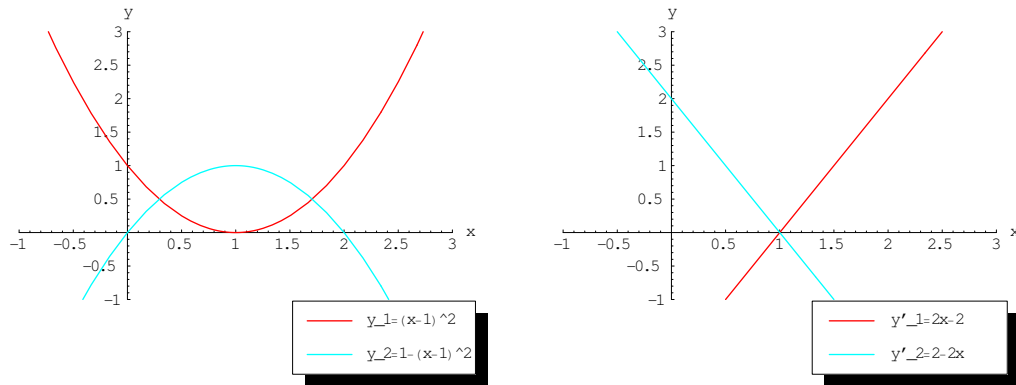


Figure I.1: (left) This figure shows two quadratic curves plotted on an axis,  $(x - 1)^2$  in red and  $1 - (x - 1)^2$  in light blue. (right) This figure shows the first derivative of the same two quadratic curves, plotted on the same axis.

$$\begin{aligned}
 \frac{\partial y}{\partial x} &= 0 \\
 \Rightarrow \frac{\partial((x-1)^2)}{\partial x} &= 0 \\
 \Rightarrow 2x - 2 &= 0 \\
 \Rightarrow x &= 1
 \end{aligned}
 \tag{I.2}$$

One can find out if this result is for a minima or maxima by finding the sign of the second derivative:

$$\frac{\partial^2 y}{\partial x^2} = \frac{\partial^2 ((x-1)^2)}{\partial x^2} = 2
 \tag{I.3}$$

If the result is positive, the point found is a minima, and if negative, the point is at a maxima. This can also be understood by looking at the graph of the first derivative,  $y'_1$ , in Figure I.1 (see page 168) (right). This shows the gradient of the red curve going from negative to positive via zero, that can only mean that the original function came down to a minima at  $x = 1$  and went back up again.

The same arguments can be applied to the function mapped out by the blue curve in Figure I.1

(see page 168), where the minima or maxima can be found:

$$\begin{aligned}
\frac{\partial y}{\partial x} &= 0 \\
\Rightarrow \frac{\partial(1-(x-1)^2)}{\partial x} &= 0 \\
\Rightarrow 2 - 2x &= 0 \\
\Rightarrow x &= 1
\end{aligned} \tag{I.4}$$

Examining the progression of the first derivative of the blue curve in Figure I.1 (see page 168) (right) shows that the gradient starts positive and goes via zero to negative. This can only mean that as  $x$  increases, the function increases (positive gradient) to a point where it is a maximum (at  $x = 2$ ) and then decreases again. Therefore, this point is a maxima. Also, the negative second derivative indicates that it must be at a maximum, rather than minimum:

$$\frac{\partial^2 y}{\partial x^2} = \frac{\partial^2(1-(x-1)^2)}{\partial x^2} = -2 \tag{I.5}$$

## I.2 An example in two dimensions

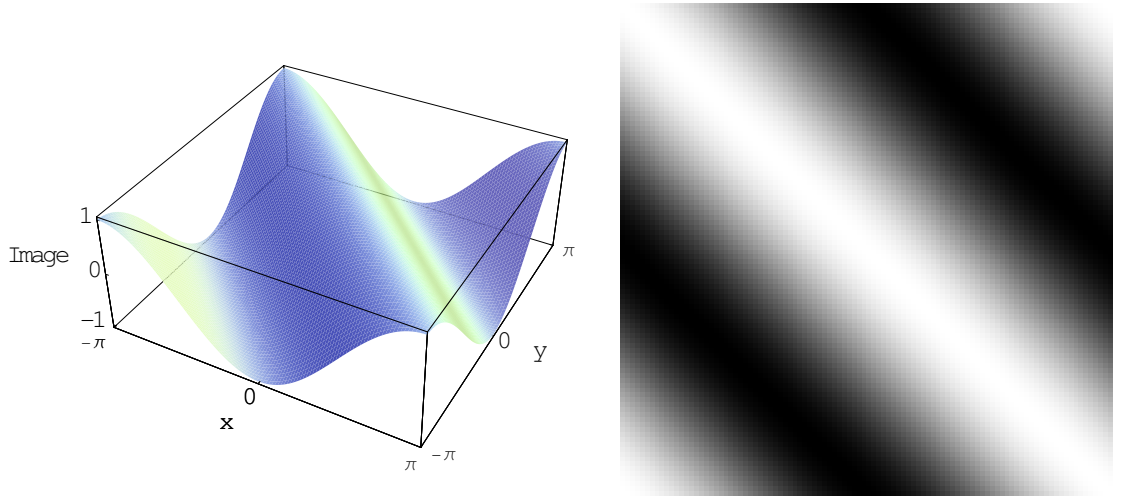


Figure I.2: (left) This surface has the equation  $I(x, y) = \cos(x + y)$ . (right) The same surface plotted on a conventional grey scale.

The model presented in this thesis effectively aims to find the direction of maximum change on a surface. This can be achieved in part by using maximisation, and also by referring to vectors. In this thesis, orientation is measured positive clockwise from vertical, therefore, a unit vector,  $\underline{r}_\theta$  in the direction  $\theta$  is written:

$$\underline{r_\theta} = \begin{bmatrix} \sin \theta \\ \cos \theta \end{bmatrix} \quad (\text{I.6})$$

The gradient on a surface in two dimensions is found in individual components aligned to the x and y axis using the  $\nabla$  or ‘Grad’ operator against the equation for the image surface,  $I(x, y)$ . The expression for the amount of gradient, a single-valued scalar, is the scalar product of ‘Grad’ and  $\underline{r_\theta}$ :

$$\nabla (I(x, y)) \cdot \underline{r_\theta} = \nabla (I(x, y)) \cdot \begin{bmatrix} \sin \theta \\ \cos \theta \end{bmatrix} \quad (\text{I.7})$$

Using the example function  $I(x, y) = \cos(x + y)$ , see Figure I.2 (see page 169), the maximisation procedure can be used to find the direction of maximum gradient, expressed in terms of  $\theta$ . Once an expression for the gradient has been derived, as in Equation I.7, the same maximisation procedure as in the one-dimensional case is used, namely solving:

$$\begin{aligned} & \frac{\partial}{\partial \theta} (\nabla(\cos(x + y)) \cdot \underline{r_\theta}) = 0 \\ \Rightarrow & \frac{\partial}{\partial \theta} \left( \begin{bmatrix} -\sin(x + y) \\ -\sin(x + y) \end{bmatrix} \cdot \begin{bmatrix} \sin \theta \\ \cos \theta \end{bmatrix} \right) = 0 \\ \Rightarrow & -\sin(x + y) \cos \theta + \sin(x + y) \sin \theta = 0 \\ & \Rightarrow \cos \theta = \sin \theta \\ & \Rightarrow \theta = \frac{\pi}{4} \pm n\pi \end{aligned} \quad (\text{I.8})$$

Finding the second derivative of this function, in order to deduce if the vector  $\underline{r_\theta}$  is at a maxima or minima is unnecessary. Just from examination of any continuous surface it is easy to envisage that the minima and maxima point in opposite directions up and down the slope. This is reflected in the finding that the solutions in Equation I.8 are separated by  $\pi$  radians.

# Bibliography

- [Adelson and Bergen, 1985] E.H. Adelson and J.R. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2:284–299, 1985.
- [Adelson and Bergen, 1991] E.H. Adelson and J.R. Bergen. The plenoptic function and the elements of early vision. In M.S. Landy and J.A. Movshon, editors, *Computational models of visual processing*. MIT, 1991.
- [Atherton, 2002] T.J. Atherton. Energy and phase orientation mechanisms: a computational model. *Spatial Vision*, 15(4):415–441, 2002.
- [Bäck, 1996] T. Bäck. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford University Press, UK, 1996.
- [Barlow and Levick, 1965] H. B. Barlow and W. R. Levick. The mechanism of directionally selective units in rabbit’s retina. *Journal of physiology*, 178:477–504, 1965.
- [Benton *et al.*, 2000] C.P. Benton, A. Johnston, and P.W. McOwan. Computational modelling of interleaved first- and second-order motion sequences and translating 3f-4f beat patterns. *Vision research*, 40:1135–1142, 2000.
- [Benton, 2002] C.P. Benton. Gradient-based analysis of non-Fourier motion. *Vision research*, 42:2869–2877, 2002.
- [Bergen and Landy, 1991] J.R. Bergen and M.S. Landy. Computational modelling of visual texture segregation. In M.S. Landy and J.A. Movshon, editors, *Computational Models of Visual Processing*. MIT, 1991.
- [Bigün and Granlund, 1987] J. Bigün and G.H. Granlund. Optimal orientation detection of linear symmetry. In *Proceedings of the IEEE first international conference on computer vision*, pages 433–438, London, 1987.
- [Bigün, 1991] J. Bigün. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE-PAMI transactions on pattern analysis and machine intelligence*, 13(8):775–790, 1991.

- [Blakemore and Campbell, 1969] C. Blakemore and F.C. Campbell. On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal changes. *Journal of physiology*, 203:237–260, 1969.
- [Blakemore *et al.*, 1970] C. Blakemore, R.H.S. Carpenter, and M.A. Georgeson. Lateral inhibition between orientation detectors in the human visual system. *Nature*, 228:37–39, 1970.
- [Blasdel, 1992] G.G. Blasdel. Orientation selectivity, preference and continuity in monkey striate cortex. *Journal of neuroscience*, 12(8):3139–3161, 1992.
- [Boas, 1993] Mary L. Boas. *Mathematical methods in the physical sciences*. John Wiley and Sons, 1993.
- [Bolduc and Levine, 1998] M. Bolduc and M.D. Levine. A review of biologically motivated space variant data reduction models for robotic vision. *Computer Vision and Image Understanding*, 69(2):170–184, 1998.
- [Bond and Kamil, 2002] A.B. Bond and A.C. Kamil. Visual predators select for crypticity and polymorphism in virtual prey. *Nature*, 415:609–613, 2002.
- [Borst and Egelhaaf, 1989] A. Borst and M. Egelhaaf. Principals of visual motion detection. *Trends in neuroscience*, 12:297–306, 1989.
- [Bracewell, 1986] Ronald Newbold Bracewell. *The Fourier transform and its applications*. McGraw-Hill, New York, 3 edition, 1986.
- [Bruce *et al.*, 1996] V. Bruce, P.R. Green, and M.A. Georgeson. *Visual Perception*. Psychology press, East Sussex, UK, 3rd edition, 1996.
- [Buracas *et al.*, 1998] G.T. Buracas, A.M. Zador, M.R. DeWeese, and T.D. Albright. Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron*, 20:959–969, 1998.
- [Campbell and Green, 1965] F.W. Campbell and D.G. Green. Optical and retinal factors affecting visual resolution. *Proceedings of the Royal Society of London B*, 166:576–593, 1965.
- [Carandini and Sengpiel, 2004] M. Carandini and F. Sengpiel. Contrast invariance of functional maps in cat primary visual cortex. *Journal of Vision*, 4:130–143, 2004.
- [Cheng *et al.*, 1994] K. Cheng, T. Hasegawa, K.S. Saleem, and K. Tanaka. Comparison of neuronal selectivity for stimulus speed, length and contrast in the prestriate visual cortical areas v4 and mt of the macaque monkey. *Journal of physiology*, 71:2269–2280, 1994.
- [Chklovskii, 2000] D.B. Chklovskii. Binocular disparity can explain the orientation of ocular dominance stripes in primate primary visual area (v1). *Vision research*, 40:1765–1773, 2000.

- [Chubb and Sperling, 1988] C. Chubb and G. Sperling. Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A*, 5(11):1986–2007, 1988.
- [Clifford *et al.*, 1998] C.C. Clifford, J.N. Freedman, and L.M. Vaina. First- and second-order motion perception in gabor micropattern stimuli: psychophysics and computational modelling. 1998. <http://www.bu.edu/bravi/research/120rdGabor/120rdGabor.htm>.
- [Dakin *et al.*, 1999] S.C. Dakin, C.B. Williams, and R.F. Hess. The interaction of first- and second-order cues to orientation. *Vision research*, 39:2867–2884, 1999.
- [Dayan and Abbot, 2001] P. Dayan and L.F. Abbot. *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. MIT Press, 2001.
- [de Charms and Zador, 2000] R.C. de Charms and A. Zador. Neural representation and the cortical code. *Annual review of neuroscience*, 23:613–647, 2000.
- [De Jong, 1975] K.A. De Jong. *Analysis of the behaviour of a class of genetic adaptive systems*. PhD thesis, University of Michigan, 1975. studies overlapping vs. non-overlapping populations in genetic algorithms.
- [De Valois *et al.*, 1982] R.L. De Valois, D.G. Albrecht, and L.G. Thorell. Spatial frequency selectivity of cells in macaque visual cortex. *Vision research*, 22:545–559, 1982.
- [De Yoe and Van Esen, 1988] E.A. De Yoe and D.C. Van Esen. Concurrent processing streams in monkey visual cortex. *Trends in Neuroscience*, 11:219–226, 1988.
- [DeAngelis *et al.*, 1999] G. C. DeAngelis, G. M. Ghose, I. Ohzawa, and R. D. Freeman. Functional micro-organization of primary visual cortex: Receptive field analysis of nearby neurons. *Journal of neuroscience*, 19(9):4046–4064, 1999.
- [Deneve *et al.*, 1999] S. Deneve, P.E. Latham, and A. Pouget. Reading population codes: a neural implementation of ideal observers. *Nature neuroscience*, 2(8):740–745, 1999.
- [Deprettere, 1988] F. Deprettere. *SVD and signal processing: Algorithms, applications and architectures*. Elsevier science publishing company, 1988.
- [Dougherty *et al.*, 2003] R.F. Dougherty, V.M. Koch, A.A. Brewer, B. Fischer, J. Modersitzki, and B.A. Wandell. Cortical magnification factor in human primary visual cortex correlates with vernier acuity thresholds. *Journal of Vision*, 3(10):586–598, 2003.
- [Duncan and Boynton, 2002] R.O. Duncan and G.M. Boynton. Cortical magnification factor in human primary visual cortex correlates with vernier acuity thresholds. *Journal of Vision*, 2(7):129a, 2002.

- [Engel *et al.*, 1997] S.A. Engel, G.H. Glover, and B.A. Wandell. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex*, 7:181–192, 1997.
- [Erwin *et al.*, 1995] E. Erwin, K. Obermayer, and K. Schulten. Models of orientation and ocular dominance columns in the visual cortex: a critical comparison. *Neural Computation*, 7:425–468, 1995.
- [Feng and Milanfar, 2002] X. Feng and P. Milanfar. Multiscale principal components analysis for image local orientation estimation. In *Proceedings of the 36th Asilomar Conference on Signals, Systems and Computers.*, Pacific Grove, California, 2002.
- [Ferster and Miller, 2000] D. Ferster and K.D. Miller. Neural mechanisms of orientation selectivity in the visual cortex. *Annual review of neuroscience*, 23:441–471, 2000.
- [Fraser, 1908] J. Fraser. A new visual illusion of direction. *British Journal of Psychology*, 2:307–320, 1908.
- [Freeman and Adelson, 1991] W.T. Freeman and E.H. Adelson. The design and use of steerable filters. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(9):891–906, 1991.  
<http://dx.doi.org/10.1109/34.93808>.
- [Georgeopoulos *et al.*, 1986] A.P. Georgeopoulos, A.B. Schwartz, and R.E. Kettner. Neuronal population coding of movement direction. *Science*, 233:1416–1419, 1986.
- [Georgeson and Freeman, 1997] M.A. Georgeson and T.C.A. Freeman. Perceived location of bars and edges in one-dimensional images: Computational models and human vision. *Vision research*, 37(1):127–142, 1997.
- [Gilbert and Wiesel, 1990] C.D. Gilbert and T.N. Wiesel. The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision research*, 30(11):1689–1701, 1990.
- [Goldberg, 1989] D.E. Goldberg. *Genetic Algorithm in Search, Optimization and Machine Learning*. Addison Wesley, 1989.
- [Gregory and Heard, 1979] R.L. Gregory and P. Heard. Border locking and the Café Wall illusion. *Perception*, 8:365–380, 1979.
- [Hammett *et al.*, 1993] S. T. Hammett, T. Ledgeway, and A.T. Smith. Transparent motion from feature- and luminance-based processes. *Vision research*, 33(8):1119–1122, 1993.
- [Heeger, 1987] D.J. Heeger. Model for the extraction of image flow. *Journal of the Optical Society of America A*, 4:1455–1471, 1987.

- [Holland, 1975] J.H. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, 1975.
- [Hubel and Wiesel, 1962] D.H. Hubel and T.N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cats visual cortex. *Journal of physiology*, 160:106–154, 1962.
- [Hubel and Wiesel, 1968] D.H. Hubel and T.N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *Journal of physiology*, 195:215–243, 1968.
- [Hubel and Wiesel, 1979] D.H. Hubel and T.N. Wiesel. Brain mechanisms of vision. *Scientific american*, 241(3):130–144, 1979.
- [Johnston and Clifford, 1995] A. Johnston and C.W.G. Clifford. A unified account of three apparent motion illusions. *Vision research*, 35:1109–1123, 1995.
- [Johnston *et al.*, 1992] A. Johnston, P.W. McOwan, and H. A. Buxton. A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *Proceedings of the Royal Society of London B*, 250:297–306, 1992.
- [Johnston *et al.*, 1999] A. Johnston, P.W. McOwan, and C.P. Benton. Robust velocity computation from a biologically motivated model of motion perception. *Proceedings of the Royal Society of London B*, 266:509–518, 1999.
- [Julesz, 1984] B. Julesz. A brief outline of the texton theory of human vision. *Trends in neuroscience*, 6:41–45, 1984.
- [Kato *et al.*, 1978] H. Kato, O. Bishop, and G.A. Orban. Hypercomplex and simple/complex cell classifications in cat striate cortex. *Journal of neurophysiology*, 41:1071–1095, 1978.
- [Kingdom *et al.*, 2003] F.A.A. Kingdom, N. Prins, and A. Hayes. Energy and phase orientation mechanisms:a computational model. *Visual neuroscience*, 20:65–76, 2003.
- [Kitaoka, 1998] A. Kitaoka. Apparent contraction of edge angles. *Perception*, 27:1209–1219, 1998.
- [Koenderink and van Doorn, 1987] J.J. Koenderink and A.J. van Doorn. Representation of local geometry in the visual system. *Biological cybernetics*, 55:367–375, 1987.
- [Koenderink and van Doorn, 1988] J.J. Koenderink and A.J. van Doorn. Operational significance of receptive field assemblies. *Biological cybernetics*, 58:163–171, 1988.
- [Koulakov and Chklovskii, 2001] A.A. Koulakov and D.B. Chklovskii. Orientation preference patterns in mammalian visual cortex: A wire length minimization approach. *Neuron*, 29:519–527, 2001.
- [Kuffler, 1953] S.W. Kuffler. Discharge patterns and functional organisation of mammalian retina. *Journal of neurophysiology*, 16:37–63, 1953.



- [Landy and Bergen, 1991] M.S. Landy and J.R. Bergen. Texture segregation and orientation gradient. *Vision research*, 31(4):679–691, 1991.
- [Langely and Atherton, 1991] K. Langely and T.J. Atherton. The inference of structure in images using multi-local quadrature filters. In *Proceedings of the British machine vision conference*, pages 111–117, Glasgow, 1991.
- [Langley *et al.*, 1996] K. Langley, D.J. Fleet, and P.B. Hibbard. Linear filtering precedes nonlinear processing in early vision. *Current biology*, 6(7):891–896, 1996.
- [Ledgeway and Smith, 1994] T. Ledgeway and A.T. Smith. Evidence for separate motion-detecting mechanisms for first- and second-order motion in human vision. *Vision research*, 34(20):2727–2740, 1994.
- [Lenz *et al.*, 1988] F. A. Lenz, J. O. Dostrovsky, R. R. Tasker, K. Yamashiro, H. C. Kwan, and J. T. Murphy. Single-unit analysis of the human ventral thalamic nuclear group: somatosensory responses. *J Neurophysiol*, 59(2):299–316, 1988.
- [Livingstone and Hubel, 1988] M.S. Livingstone and D.H. Hubel. Segregation of form, colour, movement and depth: anatomy, physiology and perception. *Science*, 240:740–749, 1988.
- [Lu and Sperling, 1995a] Z. Lu and G. Sperling. Attention-generated apparent motion. *Nature*, 377:237–239, 1995.
- [Lu and Sperling, 1995b] Z. Lu and G. Sperling. The functional architecture of human visual motion perception. *Vision research*, 35:2697–2722, 1995.
- [Lu and Sperling, 1995c] Z. Lu and G. Sperling. Three-systems theory of human visual motion perception: review and update. *Journal of the Optical Society of America A*, 377:237–239, 1995.
- [Manno and Sakrison, 1974] J.L. Manno and D.J. Sakrison. The effects of a visual fidelity criterion on the encoding of images. *IEEE Transactions of Information Theory*, 20(4):525–535, 1974.
- [Marcelja, 1980] S. Marcelja. Mathematical description of the response of simple cortical cells. *Journal of the Optical Society of America*, 70(11):1297–1300, 1980.
- [Marr, 1982] D. Marr. *Vision*. W. H. Freeman, 1982.
- [McCourt, 1982] M.E. McCourt. A spatial frequency dependend grating induction effect. *Vision research*, 22:119–134, 1982.
- [McCourt, 1983] M.E. McCourt. Brightness induction and the Café Wall illusion. *Perception*, 12(2):131–142, 1983.

- [McOwan and Johnston, 1996] P.W. McOwan and A. Johnston. A second order pattern reveals separate strategies for encoding orientation in two-dimensional space and space-time. *Vision research*, 36(3):425–430, 1996.
- [Miller, 1979] W.H. Miller. Ocular optical filtering. In H. Autrum, editor, *Handbook of sensory physiology*. Berlin: Springer Verlag, 1979.
- [Morgan and Moulden, 1986] M.J. Morgan and B. Moulden. The Münsterberg figure and twisted cords. *Vision research*, 26(11):1973–1800, 1986.
- [Moulden and Renshaw, 1979] B. Moulden and J. Renshaw. The Münsterberg illusion and ‘irradiation’. *Perception*, 8:275–301, 1979.
- [Nothdurft, 1985a] H.C. Nothdurft. Orientation sensitivity and texture segmentation in patterns with different line orientation. *Vision research*, 25:551–560, 1985.
- [Nothdurft, 1985b] H.C. Nothdurft. Sensitivity for structure gradient in texture discrimination tasks. *Vision research*, 25:1957–1968, 1985.
- [Nothdurft, 1985c] H.C. Nothdurft. Texture discrimination: representation of orientation and luminance differences in cells of cat striate cortex. *Vision research*, 25:99–113, 1985.
- [Orban, 1984] G.A. Orban. *Neuronal operations in the visual cortex*. Springer, Heidelberg, 1984.
- [Perrone, 2004] J. A. Perrone. A visual motion sensor based on the properties of v1 and mt neurons. *Vision research*, 44:1733–1755, 2004.
- [Pollen and Ronner, 1983] D.A. Pollen and S.F. Ronner. Visual cortical neurons as localized spatial frequency filters. *IEEE Transactions on systems, man and cybernetics*, 13(5):907–915, 1983.
- [Popovic and Sjostrand, 2001] Z. Popovic and J. Sjostrand. Resolution, separation of retinal ganglion cells, and cortical magnification in humans. *Vision research*, 41:1313–1319, 2001.
- [Pouget *et al.*, 2000] A. Pouget, P. Dayan, and R. Zemel. Information processing with population codes. *Nat Rev Neurosci.*, 1(2):125–132, 2000.
- [Poulton, 1890] E.B. Poulton. *The colours of animals*. Appleton, New York, 1890.
- [Prins *et al.*, 2003] N. Prins, N.K. Nottingham, and A.J. Mussap. The role of local grouping and global orientation contrast in perception of orientation-modulated textures. *Vision research*, 43:2315–2331, 2003.
- [Reichardt, 1959] W.E. Reichardt. Autocorrelation and the central nervous system. In W.A. Rosenblith, editor, *Sensory Communication*, pages 303–318. MIT Press Cambridge, 1959.

- [Reichardt, 1969] W.E. Reichardt. Movement perception in insects. In W. Reichardt, editor, *Processing of optical data by organisms and by machines*. New York: Academic Press, 1969.
- [Reichardt, 1986] W.E. Reichardt. Processing of optical information by the visual system of the fly. *Vision research*, 26(1):113–126, 1986.
- [Rodieck, 1988] R. Rodieck. The primate retina. *Comparative Primate Biology*, 4 Neurosciences:203–278, 1988.
- [Roncato, 2000] S. Roncato. The effect of luminance variation on the apparent position of an edge. *Perception and Psychophysics*, 62(4):762–785, 2000.
- [Russel, 1954] B. Russel. *History of western philosophy*. George Allen and Unwin LTD, 1954.
- [Schofield and Georgeson, 1999] A. Schofield and M.A. Georgeson. Sensitivity to modulations of luminance and contrast in visual white noise: separate mechanisms with similar behaviour. *Vision research*, 39:2697–2716, 1999.
- [Seung and Sompolinsky, 1993] H.S. Seung and H. Sompolinsky. Simple models for reading neuronal populations codes. *Proceedings of the national academy of science*, 90:10749–10753, 1993.
- [Sims, 1991] K. Sims. Artificial evolution for computer graphics. *Computer Graphics*, 25(4):319–328, 1991.
- [Skillen *et al.*, 2002] J. Skillen, D. Whitaker, A.V. Popple, and P.V. McGraw. The importance of spatial scale in determining illusions of orientation. *Vision research*, 42:2447–2455, 2002.
- [Skottun *et al.*, 1987] B. C. Skottun, A. Bradley, G. Sclar, I. Ohzawa, and R. D. Freeman. The effects of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behavior. *Journal of physiology*, 57(3):773–786, 1987.
- [Skottun *et al.*, 1991] B.C. Skottun, R.L. De Valois, D.H. Grosf, J.A. Movshon, D.G. Albrech, and A.B. Binds. Classifying simple and complex cells on the basis of response modulation. *Vision research*, 31:1079–1086, 1991.
- [Smith *et al.*, 2001] S. Smith, C.W.G. Clifford, and P. Wenderoth. Interaction between first- and second-order orientation channels revealed by the tilt illusion: psychophysics and computational modelling. *Vision research*, 41:1057–1071, 2001.
- [Stein *et al.*, 1996] B.E. Stein, N. London, L.K. Wilkinson, and D.D. Price. Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of cognitive neuroscience*, 8:497–506, 1996.

- [Swindale, 1996] N.V. Swindale. The development of topography in the visual cortex: a review of models. *Network: Computation in Neural Systems*, 7:161–247, 1996.
- [Swindale, 1998] N.V. Swindale. Orientation tuning curves: empirical description and estimation of parameters. *Biological cybernetics*, 78:45–56, 1998.
- [Taub *et al.*, 1997] E. Taub, J. D. Victor, and M. M. Conte. Nonlinear preprocessing in short-range motion. *Vision research*, 37:1459–1477, 1997.
- [Tootell *et al.*, 1988] R.B.H. Tootell, E. Switkes, M.S. Silverman, and S.L. Hamilton. Functional anatomy of macaque striate cortex. *Journal of neuroscience*, 8:1531–1568, 1988.
- [Tovée, 1996] M.J. Tovée. *An introduction to the visual system*. Cambridge university press, Cambridge, UK, 1996.
- [van den Boomgaard and van de Weijer, 2002] R. van den Boomgaard and J. van de Weijer. Robust estimation of orientation for texture analysis. In *The 2nd international workshop on texture analysis and synthesis*, Heriot-Watt University, 2002.
- [van Santen and Sperling, 1985] J.P.H. van Santen and G. Sperling. Elaborated reichardt detectors. *Journal of the Optical Society of America A*, 2:300–321, 1985.
- [Vaney and Hughes, 1990] D.I. Vaney and A.A. Hughes. Is there more than meets the eye? In C. Blakemore, editor, *Vision: Coding and efficiency*. Cambridge, 1990.
- [Vetter *et al.*, 2004] R.J. Vetter, J.C. Williams, J.F. Hetke, E.A. Nunamaker, and D.R. Kipke. Chronic neural recording using silicon-substrate microelectrode arrays implanted in cerebral cortex. *IEEE Trans Biomed Eng.*, 51(6):896–904, 2004.
- [Vogels, 1990] R. Vogels. Population coding of stimulus orientation by striate cortical cells. *Biological cybernetics*, 64:25–31, 1990.
- [Volz and Zanker, 1996] H. Volz and J. M. Zanker. Hyperacuity for spatial localization of contrast-modulated patterns. *Vision research*, 36(9):1329–1339, 1996.
- [Wallace and D.J., 2003] A.M. Wallace and McLaren D.J. Gradient detection in discrete log-polar images. *Pattern Recognition Letters*, 24(2):2463–2470, 2003.
- [Watson and Ahumada, 1985] A.B. Watson and A.J. Ahumada. Model of human visual-motion sensing. *Journal of the Optical Society of America A*, 2:322–341, 1985.
- [Weisstein, 2004] E.W. Weisstein. Mathworld. *MathWorld*, 2004.  
<http://mathworld.wolfram.com/>.
- [Whitaker, 1993] D. Whitaker. What part of vernier stimulus determines performance? *Vision research*, 33(1):27–32, 1993.

- [Williams, 1985] D.R. Williams. Aliasing in human foveal vision. *Vision research*, 25(2):195–205, 1985.
- [Wilson *et al.*, 1993] H. R. Wilson, V. P. Ferrera, and C. Yo. A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, 9(1):79–97, 1993.
- [Wolfe *et al.*, 1999] J.M. Wolfe, N.L. Klempe, and E.P. Shulmen. Which end is up? two representations of orientation in visual search. *Vision research*, 39:2075–2086, 1999.
- [Young *et al.*, 2001] R.A. Young, R.M. Lesperance, and W.W. Meyer. The gaussian derivative model for spatial-temporal vision: I. cortical model. *Spatial Vision*, Vol. 14(No. 3,4):261–319, 2001.
- [Zanker, 1993] J.M. Zanker. Theta motion: a paradoxical stimulus to explore higher order motion extraction. *Vision research*, 33(4):553–569, 1993.
- [Zanker, 1998] J.M. Zanker. Limiting factors for the detection of orientation. *Perception*, 27:167–181, 1998.
- [Zeki, 1993] S. Zeki. *A vision of the human brain*. Blackwell scientific publications, Oxford, UK, 1993.
- [Zhou and Baker Jr., 1993] Y.X. Zhou and C.L. Baker Jr. A processing stream in mammalian visual cortex neurons for non-Fourier responses. *Science*, 261(5117):98–101, 1993.
- [Zohary, 1992] E. Zohary. Population coding of visual stimuli by cortical neurons tuned to more than one dimension. *Biological cybernetics*, 66(3):265–272, 1992.