JAMES COOK UNIVERSITY
AUSTRALIA

Ross Glenn Raymond Marchant, B.E. (Elec.)

Local Feature Analysis Using Higher-Order Riesz Transforms

Thesis submitted for the

Doctor of Philosophy

at the

College of Science, Technology and Engineering
James Cook University

June, 2016

Supervised by:

Prof. Peter Ridd

Dr. Paul Jackway

Statement of the Contribution of Others

The ideas in this thesis and their development are my own work. The contribution of others is acknowledged below:

Intellectual support

- Prof. Peter Ridd was my primary supervisor at James Cook University. He provided general academic guidance and reviewed various papers and this thesis.
- Dr. Paul Jackway was an external supervisor from the Commonwealth Scientific and Industrial Research Organisation (CSIRO). He suggested the "monogenic signal" as a research topic and provided guidance and suggestions specific to my research during our discussions. Together we co-authored five conference papers and one journal paper. The subject and content of the papers was my own work. Paul provided detailed reviews and advised me of any changes that were necessary.
- Prof. Bruce Litow provided some mathematical assistance regarding linear algebra in the early stages of the PhD research.

Financial support

I received financial support for the first 3.5 years of the PhD:

- Australian Post-graduate Award (APA).
- James Cook University top-up stipend.
- CSIRO top-up stipend and travel allowance.

I also worked part time at the James Cook University Marine Geophysical Laboratory which allowed my to continue working on the PhD research once the stipends had finished.

Acknowledgements

I would like to thank my supervisor at James Cook University in Townsville, Prof. Peter Ridd, who gave me a job at the university after a chance encounter, then later encouraged me to start the PhD. Working with Peter on a diverse range of topics has been a joy.

I must thank my supervisor Dr Paul Jackway from the Commonwealth Scientific and Industrial and Research Organisation who agreed to supervise this research from Brisbane. Without Paul I would not have been able to pursue an image analysis topic and remain living in Townsville. Apart from his scientific advice, he also helped introduce me into the wider imaging community.

None of this would have been possible without years of patience and emotional support from Melissa, who put up with living in the tropics far longer than she intended. She kept me healthy and on track and stopped me from throwing it all in on multiple occasions.

I thank Jonathan Kane and Hannah Hamilton for their friendship and support and giving me a cheerful place to stay every time I visited Brisbane.

I also acknowledge my mother and brothers for putting up with one topic of conversation for way too many years, Emeritus Prof. Trevor Heath for his friendship and mentoring since I was a student at the University of Queensland, and the staff of the College of Science, Engineering and Technology at James Cook University who go out of their way to help.

Abstract

Rich descriptions of local image structures are important for higher-level understanding of images in computer vision. Phase-based representations allow the discrimination of symmetric features, such as lines, and anti-symmetric features, such as edges, independent of their strength. Methods to obtain phase information include quadrature filters using the Hilbert transform, spherical quadrature filters using the Riesz transform, and 2D analytic signals such as the monogenic signal and signal multi-vector.

This thesis develops a new local image descriptor, called the *circular harmonic vector*, consisting of the higher-order Riesz transforms of an image. The circular harmonic vector describes the symmetries of the local image structure. It extends previous analytic signals, and is formulated in the context of 2D steerable wavelet frames. Methods are introduced to solve for the parameters of a general signal model by splitting the circular harmonic vector into model and residual components. In particular, the super-resolution method, normally used for the resolving of spike trains, can be applied.

The methods are applied to estimating the parameters of sinusoidal, multi-sinusoidal and half-sinusoidal phase-based image models. The sinusoidal model describes lines and edges in terms of amplitude, phase and orientation. Using higher-order Riesz transforms in the circular harmonic vector gives better parameter estimates, and the residual component is used to develop a new detection measure for junctions and corners. The multi-sinusoidal model is applied to coral core x-ray analysis, from which separate reconstruction of features is possible as a result of the wavelet basis. The half-sinusoidal model is used to obtain the amplitudes and orientations of the line and edge segments in junctions and corners, with phase discriminating their type.

Finally, a new representation of local image structure through scale is introduced. It describes the continuous response of the circular harmonic vector response shifted though scale in the form of a quaternion-valued matrix. The matrix is derived from the higher-order Riesz transforms of an isotropic wavelet frame given by Fourier series basis functions in the logarithmic frequency domain. New measures for scale selection are developed, along with a continuous version of phase congruency that is combined with previous image models to detect and discriminate image features in an illumination invariant way.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Local Image Descriptors

The advent of ubiquitous digital cameras, high-speed computers and fast internet has seen an increase in use of the digital image as a medium for communicating information. Whereas the human visual system can extract information from images easily, thanks to the resources of a massively parallel organic computer with millions of years of development, coaxing a normal computer to perform the same task is not easy, and has spawned an entire field of research known as *image analysis*.

Computational image analysis involves the transformation of an image, which typically exists as a 2D intensity signal, into useful information. This could be identification of a face in passport photograph, confirmation of the presence of a tumour in a medical imaging scan, or location of micro-fossils in a sediment sample, for example. Many tasks that humans currently perform could be automated or augmented by computer vision systems, freeing up human labour for other endeavours and thus increasing productivity. Research into image analysis is therefore important for future technological development.

Often the first step in image analysis is to obtain a useful representation of the local image structure. *Local* means in a restricted area around a particular location in the image, while *structure* refers to the pattern and strength of the variation in pixel intensity in this area. Of particular interest is representations from which the parameters of local image *features* can be obtained. Image *features* are structures such as lines, edges, corners and junctions, which carry useful information about the patterns and shapes within an image. For example, edges and corners typically denote the boundaries of 3D objects as they appear in 2D images, while lines (also called roof edges) and junctions also indicate structures, such as the blood vessels in a retinal scan. The *parameters* of these features could be the orientation of a line, the angle of a corner, the number of line segments in a junction, and so on. Apart from delivering feature parameters, the local description can also be used to differentiate between feature types, *detect* the location of features in an image according to their *strength*, or simply augment the original image signal with generic local geometric information.

Information from these local descriptors is fed into higher-level algorithms to perform tasks such a pattern recognition or image classification. For example, one such set called *basic image features* (BIFs) [42] is the input to perform natural character recognition [92] and texture classification

[20], while the well known *scale invariant feature transform* (SIFT) [69] has seen wide adoption and use in image registration (matching areas between two images of the same object but with different viewpoints) and identifying objects in images. The success of higher-level methods is partly determined by the local image description that feeds it. It is therefore important to have a representation of local image structure that provides both a rich description of the features of interest, and is robust to common image transformations such as noise, illumination, contrast, rotation, viewpoint and scale, among others.

The achieve these goals, we may employ *models* of various features. A model is a functional description of the local image structure that has parameters describing its shape in terms of the feature components. The models and their parameters are chosen so they encode the information that is useful for image interpretation. For example, an edge and its orientation, a corner with variable angle, or a line with width parameters. No one model can encompass all features, but more general models are useful as they provide information about multiple patterns or features in the one representation. Obtaining robust estimations of the model parameters requires appropriate measurements of the local image structure. Developing better performing measurement methods has been the subject of considerable research. Common approaches are based on linear operators such as filters, non-linear operators such as using mathematical morphology, or more signal dependent methods such as machine learning. It is desirable that estimated parameters are either invariant to certain image transformations, or equi-variant, meaning that the value changes monotonically with a parameter of the transformation. Having a set of model parameters that give a unique description of the signal and are either invariant or equi-variant achieves a *split of identity* [29] of the original signal into parts that can be separately analysed.

1.1.1 Sinusoidal Models

Phase-based representations of local image structure are one of the most useful and widely used descriptions. The idea is to model the local signal structure as a sinusoid in terms of amplitude, which represents feature strength, and phase, which represents symmetry, that is, feature type. In 1D the model parameters are given by the analytic signal [39]. In 2D, there is an extra orientation parameter in the model to describe the direction of symmetry. As such, there are a multitude of different approaches to estimate the parameters, such as 2D steerable quadrature filters constructed using the Hilbert transform, [37, 38, 98–100], 2D analytic signals constructed using the Hilbert transform [10, 12, 13, 114], 2D analytic signals constructed using the Riesz transform [29, 135, 136, 141, 144], and 2D steerable wavelets [94, 117, 123, 124].

The orientation parameter of the sinusoidal model appears as a useful measure for the analysis of images. Ideally it should be invariant to both illumination (amplitude) and symmetry type (phase). Exploratory work was performed on estimating the growth direction in coral core x-ray images and micro-crystal geological images. However, limitations were found in recent 2D analytic signal and wavelet approaches to generating a sinusoidal model. Addressing these problems was the motivation behind this thesis.

Initially the focus was on improving sinusoidal model estimation using higher-order Riesz transforms (RTs) of an image. Later the project evolved into creating a general method of modelling local image structure. The approach presented in this thesis bridges previous 2D analytic signal and 2D steerable wavelet approaches into a complete framework for analysis. A rich general geometric representation of local image structure is developed, called the *circular harmonic vector*, which consists of the responses of an image to the higher-order RTs of an isotropic wavelet. The representation is used to find the amplitude and orientation parameters of line, edge, junction and corner models, in order to describe a wide range of image features in an illumination and rotationally invariant way. It also features a method of detecting corners and junctions and is extended to give a continuous representation of image structure over scale.

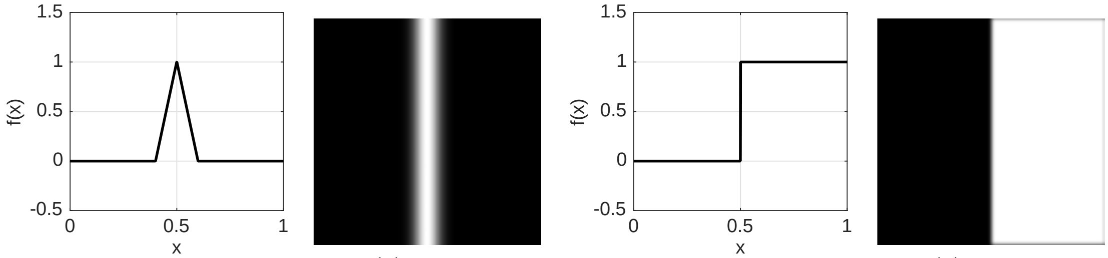
To place the research in context it is necessary to give an overview of phase-based signal representations and the development of the RT for use in deriving image models. The following sections review the different analytic signal and steerable filter approaches to representing local image structure, with a particular focus on the sinusoidal model. The various drawbacks to each method are discussed, which then motivates the development of the new framework. Finally, an overview of each chapter of the thesis is given.

1.2 1D Phase

Phase-based representations of the local structures of both 1D and 2D signals are useful as they describe both the magnitude (strength) and *linear symmetry* of the local structure. Linear symmetry refers to symmetry of the local signal structure along a particular axis. In 1D there is only one axis, so the linear symmetry is the symmetry at a point. Thus a 1D signal, $f(x)$, is *locally symmetric* at a point of interest located at $x = 0$ if $f(x) = f(-x)$ in the region around $x = 0$, and *anti-symmetric* if $-f(x) = f(-x)$.

In 2D images, linear symmetry refers to the symmetry along a particular axis. Let us represent this axis by the orientation vector $\mathbf{o} = [\cos\theta, \sin\theta]$ and denote the image coordinates as $\mathbf{z} = [x, y]$. An image, $f(\mathbf{z})$, is locally symmetric at a point of interest located at $\mathbf{z} = 0$ along an orientation $\theta \in [0, 2\pi)$ if $f(\langle \mathbf{z}, \mathbf{o} \rangle) = f(-\langle \mathbf{z}, \mathbf{o} \rangle)$ and anti-symmetric if $-f(\langle \mathbf{z}, \mathbf{o} \rangle) = f(-\langle \mathbf{z}, \mathbf{o} \rangle)$. Note, in this thesis we shall always assume a local coordinate system where the point of interest is at the centre of the local area at $\mathbf{z} = 0$.

A cursory look at some typical 1D signals and 2D images shows that the type of linear symmetry differentiates simple structures. For example, a spike in a 1D signal and a line in an image are locally symmetric, while a step in a 1D signal and an edge in an image are anti-symmetric (Figure 1.1). A signal that is symmetric at a point is called an *even* signal, or *odd* if it is anti-symmetric. These terms shall be used interchangeably.



(a) 1D spike (b) 2D line (c) 1D step (d) 2D edge

Figure 1.1: Examples of even, symmetric features and odd, anti-symmetric features in both 1D and 2D.

1.2.1 Analytic Signal

The term *phase* comes from the definition of the 1D sinusoidal signal given by $f(x) = A\cos(wx+\phi)$, where A is the amplitude, w is the frequency, and ϕ is phase. The *instantaneous* phase at location x of this signal is

$$\phi(x) = wx + \theta \pmod{2\pi}. \tag{1.1}$$

The instantaneous phase describes what angle the sinusoid is at in its cycle, and thus the shape of the sinusoid at that point. For example, a phase of 0 is the peak, π is the trough, and $\pm\pi/2$ are the points of maximum change. These also correspond to points of symmetry and anti-symmetry respectively. Furthermore, the phase is invariant to the amplitude of the sinusoid.

The phase concept can be extended to arbitrary 1D signals that are square-integrable, that is, $f(x) \in L_2(\mathbb{R})$. These signals can be represented by a convergent series of sine and cosine functions, the Fourier series,

$$f(x) = f_e(x) + f_o(x), \tag{1.2}$$

where

$$f_e(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(nx), \tag{1.3}$$

$$f_o(x) = \sum_{n=1}^{\infty} b_n \sin(nx). \tag{1.4}$$

Thus we can write the signal in the form of a sinusoid with varying amplitude and orientation,

$$f(x) = A(x)\cos(\phi(x)), \tag{1.5}$$

where $A(x) = |f_e(x) + \mathrm{i}f_o(x)|$ and the phase is given by $\phi(x) = \arg(f_e(x) + \mathrm{i}f_o(x))$. The signal is thus *modelled* by a sinusoid with given phase and amplitude. Since $f_e(x)$ is even, and $f_o(x)$ is odd at the origin, the phase describes how even or odd the signal is at a particular location x.

1.2.2 Analytic Signal

The *analytic signal* proposed by Gabor [39] extends this representation to the continuous frequency domain given by the Fourier transform. It is given by the signal and its *Hilbert transform*,

$$f_a(x) = f(x) + \mathrm{i}\mathcal{H}[f](x), \tag{1.6}$$

where $f(x)$ is the original signal and $\mathcal{H}[f](x)$ is the Hilbert transform of $f(x)$ given by

$$\mathcal{H}[f](x) = \frac{1}{\pi} P.V. \int_{-\infty}^{\infty} \frac{f(\tau)}{\tau - x}\, d\tau. \tag{1.7}$$

In the spectral domain the Hilbert transform is a Fourier multiplier,

$$\mathcal{H}[f](x) \quad \overset{\mathcal{F}}{\longleftrightarrow} \quad -\mathrm{i}\frac{\omega}{|\omega|}\mathcal{F}[f](\omega), \tag{1.8}$$

which is very similar to the derivative operator,

$$\mathcal{D}[f](x) \quad \overset{\mathcal{F}}{\longleftrightarrow} \quad \mathrm{i}\omega\, \mathcal{F}[f](\omega). \tag{1.9}$$

The Hilbert transform is equivalent to inverting the amplitude of the negative frequencies of a 1D signal. The analytic signal can therefore be constructed by setting all of the negative frequencies to zero. The Hilbert transform can also be thought of as performing a phase shift of $\pi/2$ on a signal, and the analytic signal is often referred to as the original signal being *in quadrature* with itself.

Using $A(x) = |f_a(t)|$ and $\phi(t) = \arg(f_a(t))$ it can be expressed in the form of a complex exponential,

$$f_a(x) = A(x)e^{i\phi(x)}, \tag{1.10}$$

and the original signal recovered as $f(x) = A(x)\cos(\phi(x))$, where $A(x)$ is called the *instantaneous amplitude*, $\phi(x)$ is the *instantaneous phase*, and the derivative of the phase, $\frac{d}{dt}\phi(x)$, is defined as the *instantaneous frequency* [43], although this measure is not really meaningful [68]. The phase of the analytic signal therefore represents the signal symmetry at a point, while the amplitude independently represents the strength of that symmetry.

Thus the analytic signal also provides a *sinusoidal model* of the local 1D signal structure and performs a *split of identity* into instantaneous amplitude, phase and frequency components. This representation of a signal has found many uses in areas such as radio communications [6] and quantum mechanics [127]. Interestingly, the Hilbert transform is the only 1D integral operator that commutes with both translation and dilation [78]. The analytic signal of a simple 1D signal (Figure 1.2) shows that the amplitude is high at the location of both steps and spikes, while a phase of $-\pi/2$ indicates a rising edge, $\pi/2$ indicates a falling edge, and 0 indicates a peak.

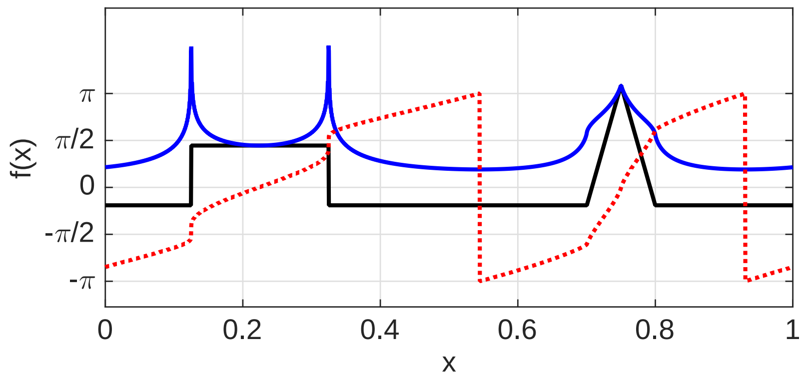


Figure 1.2: Amplitude (blue) and phase (red) of the analytic signal of a 1D signal (black) with zero mean.

The Hilbert transform has an infinite impulse response and therefore the value of the analytic signal at a point requires the entire signal to compute. However, normally we are interested in the amplitude (strength) and phase (symmetry) of the signal in a local area. Typically this is achieved by *localising* the analytic signal by first convolving the original signal with a band-pass filter centred at the frequency (scale) of interest (Figure 1.3). Since this convolution and the Hilbert transform can be performed efficiently in the Fourier domain, the operations are combined. The set of the filter and its Hilbert transform is referred to as a *quadrature filter pair*. The isotropic part of the quadrature filter is an even function and thus has a high response at the symmetric parts of the signal, such as roof edges and peaks. The Hilbert transformed part is an odd function and thus has a high response at the locally anti-symmetric parts, such as steps.

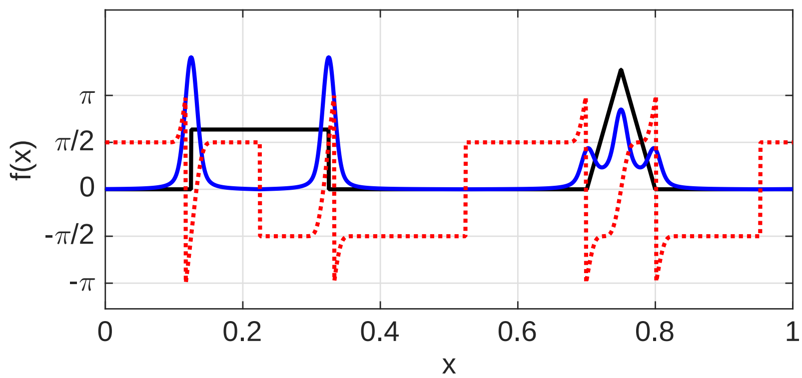


Figure 1.3: Amplitude (blue) and phase (red) of the analytic signal of a 1D signal (black) localised with a band-pass filter.

It has been shown that the amplitude of the analytic signal is high at locations of common 1D signal features, as well as their 2D image equivalents, and that the phase can be used to discriminate their type. [83, 84, 100, 128]. The *sinusoidal model* from the analytic signal therefore provides a description of local signal structure that is useful for local feature analysis.

1.3 2D Phase Using the Hilbert Transform

The analytic signal has been shown to be a powerful method of describing local symmetries in a 1D signal, enabling the identification of features such as steps, peaks, and ramps. Local symmetries also describe features in higher dimensional signals, in particular 2D images. For example, lines and edges are analogous to peaks and steps in a 1D signal, while complex features such as crossed lines or corners can be described by the superposition of multiple components with individual axes of symmetry.

Extending the phase-based representation of the analytic signal to 2D images is thus an attractive proposition. For example, multiple amplitude values could describe the strength of a feature and its components, while multiple phase values could independently describe their shape. However, in moving to two dimensions there is an additional complexity. Unlike the 1D case, each symmetry is also associated with a particular *direction*. For example, at the centre of a local image structure consisting of two crossed lines, there is the superposition of two symmetric components with different orientations. This section reviews some of the approaches to 2D phase-based image representation and their limitations

1.3.1 Intrinsic Dimension

To understand current approaches to generalising the analytic signal we must review the concept of *intrinsic dimension*. In 1D, signal symmetry is restricted to one variable; however, higher dimensional signals may have multiple symmetries in multiple orientations at the same location. The intrinsic dimension of a signal at a point differentiates between single and multiple symmetries. It refers to the dimension of the subspace required such that the error between the original signal and the signal projected onto the subspace is within a certain tolerance [8]. In other words, if we have a signal of dimension n which can be expressed using m orthogonal basis vectors ($0 \leq m \leq n$), the intrinsic dimension of the signal is m. This type of signal is often referred to as an imD signal.

For image analysis and feature detection we are interested in the local intrinsic dimension at each point in an image. For 2D signals this may be defined spatially as [145]:

- i0D - The image is a constant - $f(\mathbf{z}) = c$.
- i1D - The image is constant along one direction and therefore be completely characterised by its variation along the orthogonal direction - $f(\mathbf{z}) = f(\langle \mathbf{z}, \mathbf{o} \rangle)$ where $\mathbf{o} = [\cos\theta, \sin\theta]$ and θ is the orientation.
- i2D - Everything else.

Likewise the intrinsic dimension can be defined in the spectral domain by considering the Fourier transform of the local signal [61]:

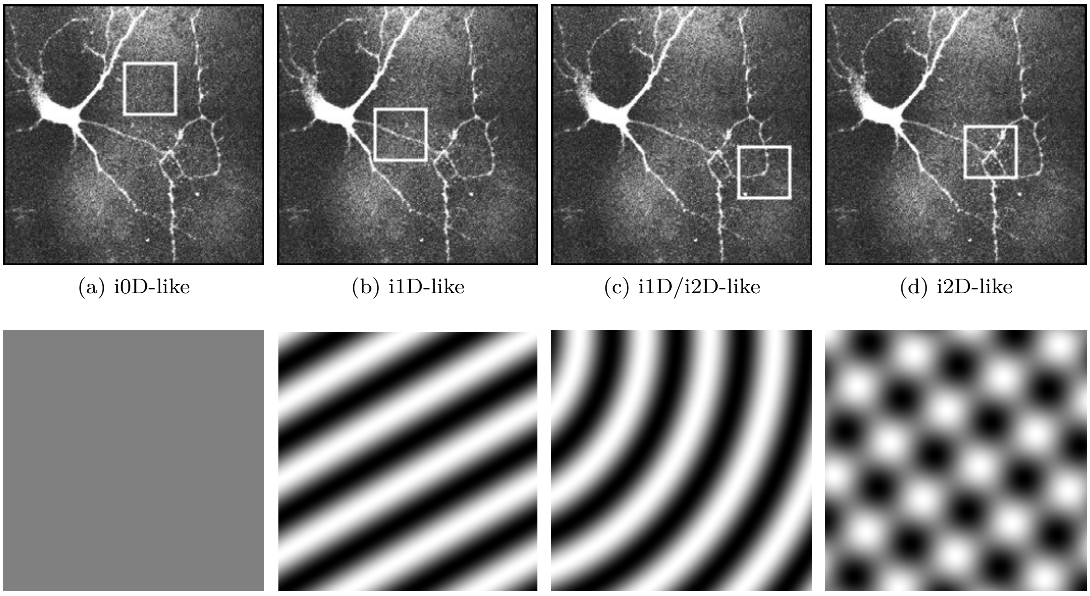
- i0D - The spectrum is predominately at the origin - $\hat{f}(\boldsymbol{\omega}) = c \cdot \delta(\boldsymbol{\omega})$.

- i1D - The spectrum is predominately centred upon a line through the origin - $\hat{f}(\boldsymbol{\omega}) = \hat{f}(\boldsymbol{\omega}) \cdot \delta(\mathbf{s}))$ where $\mathbf{s} = [\cos\theta, \sin\theta]$.

- i2D - Everything else.

Note, $\boldsymbol{\omega}$ denotes the coordinates of the Fourier domain representation of an image, which can be expressing polar coordinates $[\omega, \phi]$ where $\boldsymbol{\omega} = [\omega\cos\phi, \omega\sin\phi]$.

i0D features are flat areas in an image (Figure 1.4a), while i1D image features have only one direction of linear symmetry (Figure 1.4b). From the Fourier slice theorem it can be shown that the orientation vector $\mathbf{o}$ in the above spatial domain definition is equal to the orientation vector $\mathbf{s}$ in the above Fourier domain definition. Thus the orientation of linear symmetry of a signal can also be estimated from the symmetry of its local spectral representation [9] and therefore represented by a sinusoidal model (Figure 1.4c). i2D image features have multiple symmetries, for example, corners and junctions (Figure 1.4c). These features must be represented using more complex image models, such as multiple sinusoids (Figure 1.4h).



(a) i0D-like (b) i1D-like (c) i1D/i2D-like (d) i2D-like

(e) i0D Signal (f) i1D Signal (g) i1D/i2D Signal (h) i2D Signal

Figure 1.4: Examples of regions in an image with different intrinsic dimensions and the corresponding idealised signal models. Image taken from [118].

The local intrinsic dimension of an image thus can be used to classify image features into three types. Methods to obtain an estimate of intrinsic dimension include using 1st and 2nd order differential operators [5]. One recent approach is *continuous* intrinsic dimension, where local image structure is described by a point in a triangular space whose vertices corresponding to totally i0D, i1D or i2D features [28]. Noting that the intrinsic dimension can change with scale [64] we can consider local intrinsic dimension *itself* as an important image feature. Indeed the relative occurrence of i0D, i1D and i2D features in an image can be used to distinguish 'natural' images [61].

1.3.2 2D Quadrature Filters

Early methods developed for modelling local image structure using a sinusoidal model used 2D quadrature filters. Like the analytic signal, 2D quadrature filters are also constructed with the Hilbert transform. Consider an image, $f(\mathbf{z})$, that is locally i1D at a point of interest centred at $\mathbf{z} = \mathbf{0}$. The local structure can be represented as 1D function,

$$f(\mathbf{z}) = f(\langle \mathbf{z}, \mathbf{o} \rangle) \tag{1.11}$$

$$= f_{\mathrm{i1D}}(x), \tag{1.12}$$

where $\mathbf{o} = [\cos\theta, \sin\theta]$ with θ being the orientation of the local symmetry. According to the Fourier slice theorem, the Fourier transform of the image patch will have all non-zero coefficients concentrated along a line through the origin. Therefore, the local image structure can be exactly modelled as a sum of sinusoids,

$$f(\mathbf{z}) = \sum_k \alpha_k \cos(\omega_k \langle \mathbf{z}, \mathbf{o} \rangle + \phi_k), \tag{1.13}$$

and thus represented, at the point of interest, by a sinusoidal model,

$$f(\mathbf{z}) = A \cos\left(\langle \mathbf{z}, \mathbf{o} \rangle) + \phi\right), \tag{1.14}$$

which is equivalent to the analytic signal representation of $f_{\mathrm{i1D}}(x)$, with instantaneous amplitude, A, and phase, ϕ. The problem is how to obtain the amplitude, phase and now *orientation* values. Many different approaches have been proposed to achieve this goal.

For a 1D signal we use an even and odd (quadrature) pair of filters, with the odd filter given by the Hilbert transform of the even. The first class of methods developed for 2D images replicates this process. An even filter that is symmetric along the same axis as the local i1D symmetry is used,

$$h_e(\langle \boldsymbol{\omega}, \mathbf{o} \rangle) = h_e(-\langle \boldsymbol{\omega}, \mathbf{o} \rangle), \tag{1.15}$$

along with an odd filter given by its Hilbert transform along the same axis.

$$h_o(\langle \boldsymbol{\omega}, \mathbf{o} \rangle) = h_e(-\langle \boldsymbol{\omega}, \mathbf{o} \rangle). \tag{1.16}$$

The pair of filters are called *2D quadrature filters*. In the above equations and the rest of this thesis $h(\boldsymbol{\omega})$ will be used to denote the Fourier domain representation of a filter.

Using the output of the even and odd filters, the amplitude and phase of the sinusoidal model of the local image structure can be obtained. However, unless the orientation of the i1D feature is known beforehand, the problem remains as to what direction to apply the quadrature filter pair.

Not only that, the orientation of i1D features can *vary* over an image. One solution used in higher level methods such as phase congruency [59] is to use a *bank* of quadrature filters at multiple discrete orientations. For example, the filter bank used in [58] consists of eight 2D quadrature filters with orientations separated by 22.5 degrees.

Steerable Filters

The response to a discrete orientation 2D quadrature filter bank is not rotationally invariant. For example, an i1D feature oriented along the same axis as one of the filters will have a maximal response for that filter, whereas an i1D feature oriented along an axis in between two filters will have its energy split between them. For an i1D feature, Bigun and Granlund approached this problem by finding the eigenvectors of the *inertia matrix* of the local Fourier domain [9]. The components of the matrix are obtained from spatial domain filters, and localised by convolution with a window function. Early edge detection methods use the 1st order derivatives of an isotropic filter to estimate orientation [16, 109]. However, odd filters only respond to anti-symmetric features, and thus do not give a measure of phase.

Freeman and Adelson [37, 38] and Perona [98–100] introduced the concept of *steerable filters*, which also formed part of Simoncelli's work on *shiftable* parameters of wavelet transforms. Shiftability refers to the ability to modulate the components of a particular local image representation to synthesise the response to a filter or wavelet with respect to a particular parameter value. For steerable filters this parameter is orientation. Simoncelli's approach also looked at translation (a trivial case for linear filters) as well as scale.

A steerable filter is given by the linear combination of a finite set of *basis filters.* The basis filters are so called as their contribution can be modulated in order to synthesis the steerable filter kernel at any orientation, hence the term *steerable.* In [38], steerable filters are expressed in the spatial domain as

$$f^{\theta}(r,\phi) = \sum_{j=1}^{M} k_j(\theta) g_j(r,\phi), \tag{1.17}$$

where $f^{\theta}(r,\phi)$ is the filter in polar co-ordinates, $k_j(\theta)$ is an interpolation function, and $g_j(r,\phi)$ is a basis filter. Essentially, if a filter $f^{\theta}(r,\phi)$ can be represented by a finite Fourier series with maximum order N in the angular dimension, it can be steered by a set of $N+1$ basis functions whose vector of angular Fourier series components span the space of possible vectors of filter Fourier series components. These angular Fourier series components are known as *circular harmonics.*

The most common implementation of the basis filter bank for purely odd or even filters is actually $N+1$ copies of the original filter at equally spaced orientations over π radians. For filters with a mixture of odd and even symmetry it is $2N+1$ filters over 2π radians. Alternatively, one may instead describe a steerable filter by the sum of the $2N+1$ circular harmonics. This approach is simpler as the interpolation function is simply a complex exponential. That is,

$$f^{\theta}(r,\phi) = \sum_{n=-N}^{N} e^{\mathrm{i}n\theta} a_n(r) e^{\mathrm{i}n\phi}. \tag{1.18}$$

Having only non-zero even order circular harmonics results in an even filter, while having only non-zero odd orders results in an odd filter.

The beauty of steerable filters is that because of the linearity of the operators, the response to a filter at a particular orientation can be obtained from the same combination of the basis filter *responses* [38]. That is, we do not need to construct the steerable filter kernel itself. Furthermore, from the basis filter responses we can obtain the coefficients of a *trigonometric polynomial* that represents the angular response of the steerable filter as it is rotated. The orientation of the steerable filter that delivers the *maximum response* corresponds to the maximum of this polynomial. Owing to the Fourier slice theorem, steerable filters can also be constructed in the Fourier domain using the same principles.

Limitations

Returning the orientation estimation problem, we should be able to steer a quadrature filter pair (even steerable filter and its Hilbert transform) to find the optimal orientation at which to obtain the phase and amplitude of the sinusoidal model. However, there is a catch. In the frequency domain, the oriented Hilbert transform is a Heaviside step function, which in polar coordinates is given by

$$h^{\theta}(\omega,\phi) = \mathbf{1}_{[\theta-\pi/2,\theta+\pi/2)} - \mathbf{1}_{[-\theta-\pi/2,-\theta+\pi/2)}, \tag{1.19}$$

where θ is the orientation of the axis along which it is applied. Because of the sharp transition, the number of circular harmonic components needed to approximate this function is infinite. Therefore the approach used in [38, 99] and elsewhere is to *approximate* the Hilbert transformed part with a finite number of basis functions. Thus the pair of filters is *not* in perfect quadrature.

2D quadrature filters offer a local image representation in terms of a sinusoidal model consisting of amplitude, phase and orientation. However, if only one orientation is considered, then information will be lost where is the local image structure is i2D and contains more axes of symmetry. That is, the original signal cannot be recovered from the signal model parameters. If one instead uses a filter bank consisting of equally spaced orientations of the 2D steerable quadrature filter then both recovery of the original signal and steering is possible. However, orientation information is not explicit and the model parameters obtained from each of the component quadrature filters are no longer invariant to rotation of the image.

1.3.3 2D Analytic Signals

Whereas 2D steerable quadrature filters deal primarily with calculating local amplitude and phase parameters, another branch of research has focussed more on local image representation. These methods seek to extend the analytic signal to images and higher-dimensional signals by explicitly creating a higher-dimensional representation in which the original signal forms one dimension, and transforms of the signal form the other dimensions. In particular, the original signal can be recovered from any of the dimensions by inverting the transform and thus no information is lost. The parameters of the sinusoidal image model can then be obtained from the new signal.

Early attempts appear inspired by how the 1D analytic signal can be created using the Heaviside step (*signum*) function in the Fourier domain,

$$f_a(x) \quad \stackrel{\mathcal{F}}{\longleftrightarrow} \quad \hat{f}(\omega)\,(1 + \mathrm{i}\ \mathrm{sign}(\omega))\,, \tag{1.20}$$

meaning that is has no negative frequencies. We shall now review each of these attempts, summarised from the development by Bulow in [10, 12, 13].

Total Complex Signal

The total complex signal is the combination of an image and its *total Hilbert transform*, which is the Hilbert transform applied along both x and y axes [114]. In the Fourier domain this is,

$$f_a(\mathbf{z}) \quad \stackrel{\mathcal{F}}{\longleftrightarrow} \quad = \hat{f}(\boldsymbol{\omega})\,(1 - \mathrm{i}\,\mathrm{sign}(\omega_x)\mathrm{sign}(\omega_y))\,. \tag{1.21}$$

The amplitude and phase is calculated as the magnitude and argument of $f_a(\mathbf{z})$. However, this has the problem that the spectral signal is not causal, i.e. $\hat{f}_a(\mathbf{z}) \neq 0$ when $\mathbf{z} < 0$ [13]. Furthermore, the total Hilbert transform does not perform a $\pi/2$ phase shift [29] and is not rotationally invariant [140].

Partial Analytic Signal

The partial analytic signal is the combination of an image with its *partial Hilbert transform* along an orientation vector $\mathbf{o} = [\cos\theta, \sin\theta]$ [41]. In the Fourier domain this is,

$$f_a(\mathbf{z}) \quad \stackrel{\mathcal{F}}{\longleftrightarrow} \quad = \hat{f}(\boldsymbol{\omega})\,(1 + \mathrm{i}\,\mathrm{sign}(\langle\boldsymbol{\omega}, \mathbf{o}\rangle))\,. \tag{1.22}$$

The amplitude and phase are then calculated as the magnitude and argument of $f_a(\mathbf{z})$. However this only gives a measure of the symmetry along the orientation vector $\mathbf{o}$ and is therefore not rotationally-invariant.

Single Orthant Complex Signal

The signal orthant complex signal applies the partial Hilbert transform along both the x-axis and along the y-axis [44]. In this method all negative frequency components are zero. In the Fourier domain this is,

$$f_a(\mathbf{z}) \quad \xleftrightarrow{\mathcal{F}} \quad = \hat{f}(\boldsymbol{\omega})((1+\mathrm{sign}(\omega_x))(1+\mathrm{sign}(\omega_y)). \tag{1.23}$$

However the original signal is not recoverable; this requires an entire half plane of the Fourier spectrum instead of a single orthant. A work-around is to calculate the complex signal for an adjacent orthant as well, however this leads to two measurements of amplitude and phase which are difficult to interpret [13].

Quaternionic 2D Analytic Signal

The quaternionic analytic signal is a higher dimensional representation that uses the algebra of *quaternions* [45]. We recall that in a 1D signal any point can be described as symmetric or anti-symmetric, with the real part of the 1D analytic signal as the symmetric part, and the Hilbert transformed signal as the anti-symmetric imaginary part. The algebra of complex numbers is adequate for this representation, however for a 2D signal there are more symmetries to be considered. A 2D signal has a degree of symmetry or anti-symmetry along the x-axis, and can also have symmetry or anti-symmetry along the y-axis. The quaternionic Fourier transform [10, 23] encodes these symmetries onto the one real and three imaginary components of a quaternion, enabling the analysis of the phase for each symmetry.

The quaternionic 2D analytic signal is created by applying the same transform as used in the single orthant analytic signal applied to the quaternionic Fourier transform (QFT) spectrum,

$$\mathbf{z} = \hat{f}^q(\boldsymbol{\omega})((1+\mathrm{sign}(\omega_x))(1+\mathrm{sign}(\omega_y)), \tag{1.24}$$

where $\hat{f}^q(\boldsymbol{\omega})$ is the QFT of the signal. Converting back to the spatial domain the signal has four dimensions,

$$f_a(\mathbf{z}) = f_{ee}(\mathbf{z}) + \mathrm{i} f_{oe}(\mathbf{z}) + \mathrm{j} f_{eo}(\mathbf{z}) + \mathrm{k} f_{oo}(\mathbf{z}), \tag{1.25}$$

where $[i, j, k] \in \mathbb{H}$. An instantaneous amplitude, two i1D phases and a single i2D phase can then be calculated [10]. The two i1D phases represent the symmetry along the x and y axis, while the i2D phase is a measure of the overall symmetry of both. Compared to the previous methods, the instantaneous amplitude of the quaternionic 2D analytic signal is the best [13] however both amplitude and phase lack rotational invariance [29, 140].

Limitations

Each of the approaches reviewed so far meet Vakman's three basic properties of the analytic signal [126]. However, none are rotation invariant and therefore none will correctly determine the phase and amplitude of a single arbitrarily oriented i1D feature. The 2D complex signal and total analytic signal show that using properties of the 1D analytic signal - having the Hilbert transform for the imaginary component and zero negative frequencies - cannot be used to generate a rotationally invariant 2D analytic signal. The quaternionic 2D analytic signal hints that a higher dimensional representation is needed, however it is also rotationally variant.

1.4 2D Phase Using the Riesz Transform

1.4.1 Monogenic Signal

Felsberg and Sommer resolved the rotation variance problems of the early analytic signal attempts, introducing a rotationally invariant extension called the *monogenic signal*. Larkin, Bone and Oldfield separately introduced the *spiral quadrature* filter transform for the de-modulation of fringe patterns [62, 63], which is functionally equivalent to the monogenic signal. Their insight was the identification of the *Riesz transform* (RT), not the Hilbert transform, as the appropriate operator to extend the analytic signal to higher dimensions.

The monogenic signal is represented by a three dimensional vector consisting of an image and its RTs along two orthogonal axes,

$$\mathbf{f}_{\mathrm{M}}(\mathbf{z}) = [f(\mathbf{z}), \mathcal{R}_x[f](\mathbf{z}), \mathcal{R}_y[f](\mathbf{z})] \tag{1.26}$$

The above representation is a vector, but Felsberg's initial paper [29] uses quaternions to represent the monogenic signal, and subsequent work (e.g. [135]) uses a geometric algebra embedding. Geometric algebra contains sub-algebras isomorphic to complex numbers and quaternions [48], and thus provides a more generalised mathematical framework. For example, a phase shift for a 1D signal is represented by a rotation in 2D space. However, In this thesis we shall restrict ourselves to vector algebra for its simplicity and wider comprehension.

Riesz Transform

The Riesz transform, $\mathcal{R}_x$, of a multi-dimensional signal $f \in L_2(\mathbb{R}^N)$ along an axis, x, can be expressed as a convolution in the spatial domain, or a multiplication in the Fourier domain [29, 115],

$$\mathcal{R}_x f(\mathbf{z}) \quad \xleftrightarrow{\mathcal{F}} \quad \mathrm{i}\frac{\omega_x}{\|\boldsymbol{\omega}\|}\hat{f}(\boldsymbol{\omega}), \tag{1.27}$$

where $\hat{f}(\boldsymbol{\omega})$ is the Fourier transform of $f(\mathbf{z})$ with $\mathbf{z} = [x, y]$. It can be thought of as normalised derivative operator, similar to the Wirtinger operator [137], but without modification to the magnitude of the spectrum of $f(\mathbf{z})$. For example, the derivative along the x axis is given by

$$\mathcal{D}_x f(\mathbf{z}) \quad \stackrel{\mathcal{F}}{\longleftrightarrow} \quad \mathrm{i}\,\omega_x \, \hat{f}(\boldsymbol{\omega}). \tag{1.28}$$

For images and other 2D signals, $f \in L_2(\mathbb{R}^2)$, the Riesz transforms along the x and y axes can be combined into a single complex valued operator, the 2D complex Riesz transform (RT) [122, 135]

$$\mathcal{R}f(\mathbf{z}) \quad \stackrel{\mathcal{F}}{\longleftrightarrow} \quad \frac{\omega_x + \mathrm{i}\omega_y}{\|\boldsymbol{\omega}\|} \, \hat{f}(\boldsymbol{\omega}) = e^{\mathrm{i}\phi} \hat{f}(\omega, \phi), \tag{1.29}$$

where ω and ϕ are the radial and angular polar frequency domain coordinates given by $\boldsymbol{\omega} = [\omega \cos\phi, \omega \sin\phi]$. This complex embedding allows for easy rotation of the impulse response by multiplying by a complex exponential [124],

$$\mathcal{R}\{\delta\}(\mathbf{R}_\theta \mathbf{z}) = e^{-\mathrm{i}\theta} \mathcal{R}\{\delta\}(\mathbf{z}), \tag{1.30}$$

where $\mathbf{R}_\theta$ is a matrix that rotates the image axes by θ.

Given a sinusoidal image,

$$f_{\mathrm{S}}(\mathbf{z}) = A \cos(\omega_c \langle \mathbf{z}, \mathbf{o} \rangle + \phi), \tag{1.31}$$

at a point of interest at $\mathbf{z} = \mathbf{0}$ we may write

$$f_{\mathrm{S}} = A \cos(\phi). \tag{1.32}$$

It has been shown by means of the Radon transform [29] that the complex RT of a sinusoidal signal at the same location is given by

$$\mathcal{R}f_{\mathrm{S}} = A e^{\mathrm{i}\theta} \, \mathrm{i} \sin(\phi).$$

That is, the RT gives a *rotationally-invariant estimate* of the parameters of a sinusoidal image.

Image Model

Using this relationship, the monogenic signal also models local image structure as a single 2D sinusoid [29], as is done for 2D quadrature filters. Assuming a local coordinate system where the point of interest is located at $\mathbf{z} = \mathbf{0}$, the image model is

$$f_{\mathrm{S}}(\mathbf{z}) = A \cos(\langle \mathbf{z}, \mathbf{o} \rangle + \phi), \tag{1.33}$$

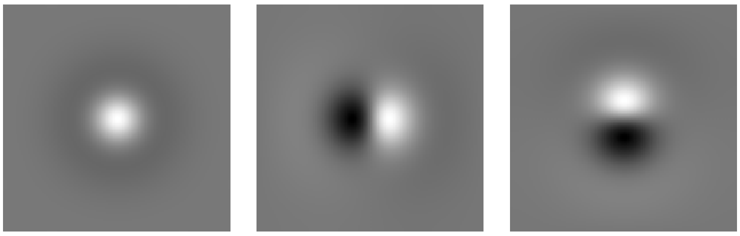
with sinusoid parameters amplitude, A, phase, ϕ, and orientation, θ. The parameters are obtained using the complex RT according to

$$A = |f + \mathrm{i}|\mathcal{R}f||\,, \tag{1.34}$$

$$\phi = \arg(f + \mathrm{i}|\mathcal{R}f|), \tag{1.35}$$

$$\theta = \arg(\mathcal{R}f), \tag{1.36}$$

where $A \in \mathbb{R}^+$, $\phi \in [-\pi/2, \pi/2)$ and $\theta \in [0, 2\pi)$. For a *local* representation the responses are filtered using a isotropic band pass filter (Figure 1.5).



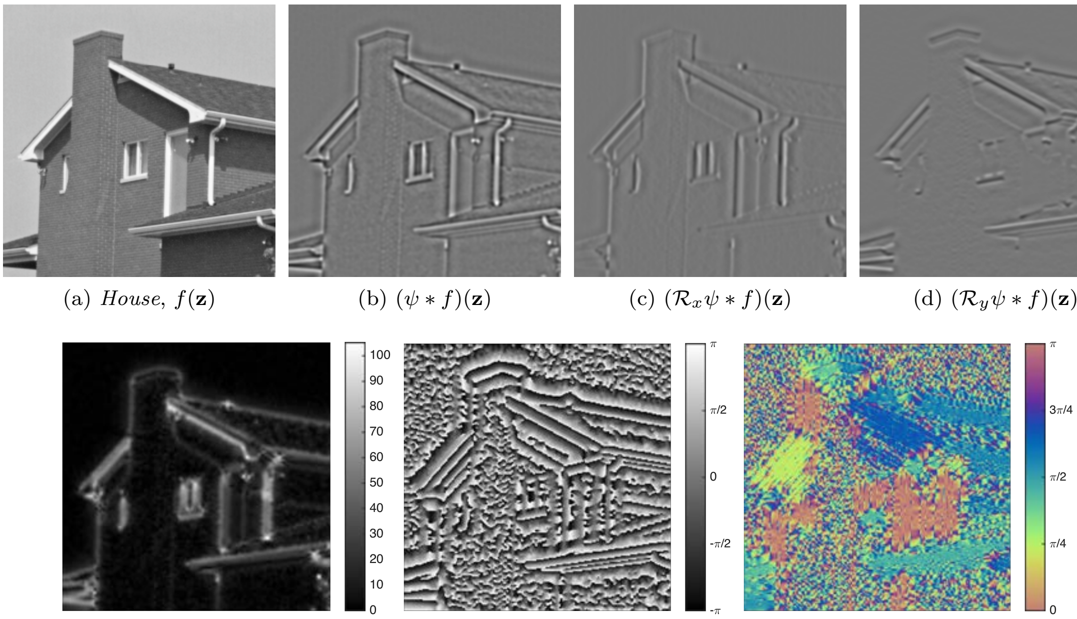
(a) $\psi(\mathbf{z})$ (b) $\mathcal{R}_x\psi(\mathbf{z})$ (c) $\mathcal{R}_y\psi(\mathbf{z})$

Figure 1.5: Isotropic band-pass filter (a) and its 1st order RTs (b,c) used to obtain the monogenic signal.

The monogenic-signal derived model achieves the desired split-of-identity of the original signal into invariant and equi-variant parameters [29]. Rotation of the input signal changes the orientation parameter, but not the amplitude or phase. Scalar addition or multiplication of the local image patch intensities changes the amplitude, but not the phase or orientation, and a phase shift of the Fourier spectrum only modifies the phase. Finally, the original signal can be recovered from the monogenic signal components.

The monogenic signal model was calculated for a the 256×256 pixel *House* image, localised using a Cauchy filter [11, 107] with peak wavelength 8 pixels and bandwidth factor $a = 2$ (Figure 1.6). As was observed for the analytic signal, the amplitude is high at the locations of strong image features [29]. The phase value is a measure of the local symmetry independent of amplitude; phase values near 0 or ϕ indicate an even structure, such as a line, and phase values near $\pm\pi/2$ indicate an odd structure, such as an edge. The orientation parameter describes the main axis of symmetry of the local structure [29].

The monogenic signal has been applied to diverse problems such as stereo image matching [25], image registration [79, 129, 138, 147], segmentation [3, 7], optical flow estimation [27, 34], face recognition [139], texture classification [146] and medical image analysis [52, 119]. Reconstruction of images from monogenic phase is more efficient than global phase based reconstruction [141].



(a) *House*, $f(\mathbf{z})$ (b) $(\psi * f)(\mathbf{z})$ (c) $(\mathcal{R}_x\psi * f)(\mathbf{z})$ (d) $(\mathcal{R}_y\psi * f)(\mathbf{z})$

(e) Amplitude (f) Phase (g) Orientation

Figure 1.6: Top row: *House* image and its monogenic signal components localised by the band-pass filter kernel $\psi(\mathbf{z})$. Bottom row: Sinusoidal model parameters of amplitude, phase and orientation derived from the responses.

Limitations

Care must be taken when interpreting the sinusoidal model derived from the monogenic signal. Firstly, the 0th order RT responds to both even and isometric structures, meaning that blobs, which would not be considered sinusoid-like, also give a large response. Secondly, the 0th order operator is isotropic, meaning that orientation is only calculated from the 1st-order RT, which only responds to odd structures. Thus the orientation estimate is poor near the centre of even features in the presence of noise [26]. This is noticeable in Figure 1.6d as line shaped discontinuities in the orientation estimate.

1.4.2 Higher-Order Signals

The poor orientation estimate around even structures is a drawback of using the monogenic signal in practical applications. Two solutions that have been proposed are to average the phase vector near even structures [26], or to include higher-order RT responses by using an expanded signal model. These methods include the structure multi-vector [26, 31], 2D analytic signal [136] and the signal multi-vector [135].

Higher-order Riesz transform

For 2D images, the higher-order responses are obtained using the n-th order complex RT, given by [122, 123, 135]

$$\mathcal{R}^n f(\mathbf{z}) \quad \stackrel{\mathcal{F}}{\longleftrightarrow} \quad e^{in\phi} \hat{f}(\omega, \phi), \tag{1.37}$$

where ω and ϕ are the radial and angular coordinates of the frequency spectrum, respectively. The 0th order RT, $\mathcal{R}^0$, is the identity operator. As for the 1st order RT, used in the monogenic signal, the impulse response is rotated by multiplying by a complex exponential to the same power as the RT order [124],

$$\mathcal{R}^n\{\delta\}(\mathbf{R}_\theta \mathbf{z}) = e^{-in\theta} \mathcal{R}\{\delta\}(\mathbf{z}), \tag{1.38}$$

where $\mathbf{R}_\theta$ is a matrix that rotates the image axes by θ. The odd and even order RTs of an image in the spatial domain are in conjugate according to

$$\mathcal{R}^n f(\mathbf{z}) = \overline{\mathcal{R}^{-n} f(\mathbf{z})} \qquad n \text{ even}, \tag{1.39}$$

$$\mathcal{R}^n f(\mathbf{z}) = -\overline{\mathcal{R}^{-n} f(\mathbf{z})} \qquad n \text{ odd}. \tag{1.40}$$

Like the Hilbert transform, the RT has an infinite impulse response and requires the entire image to compute. To construct a more localised operator, we can combine the RT with isotropic band-pass filter with enough vanishing moments [124], resulting in a *spherical quadrature filter* (SQF) [26, 29]. Figure 1.7 shows an example of some 0th to 3rd order SQFs constructed from the RTs of an isotropic filter.

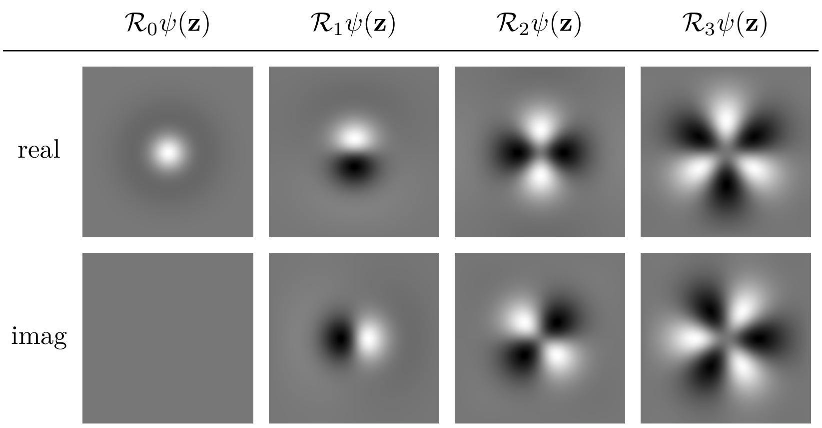


Figure 1.7: Real and imaginary components of the 0th to 3rd order SQFs constructed with the higher-order complex RT.

Similar to the Hilbert transform in 1D, the RT of a sinusoidal signal is equivalent to a phase shift. Consider an image given by the addition of K 2D sinusoidal signals,

$$f(\mathbf{z}) = \sum_{k=1}^{K} A_k \cos(w\langle \mathbf{z}, \mathbf{o}_k \rangle + \phi_k), \tag{1.41}$$

where $\mathbf{o} = [\cos\theta, \sin\theta], \theta \in [0, \pi)$ is the orientation vector, $A \in \mathbb{R}^+$ is the amplitude, and $\phi \in [0, 2\pi)$ is the phase of the sinusoid. It has been shown by means of the Radon transform [29, 135] that the n-th order RT of this signal at the point of interest, $\mathbf{z} = \mathbf{0}$, is

$$\mathcal{R}^n f = \begin{cases} \sum_k A_k e^{in\theta_k} \cos(\phi_k) & n \text{ is even,} \\ \sum_k A_k e^{in\theta_k} \, \mathrm{i} \sin(\phi_k) & n \text{ is odd.} \end{cases} \tag{1.42}$$

Since the higher-order RT kernels are orthogonal, the responses give separate estimates that can be used to solve for the sinusoid parameters. For example, for $K = 1$, amplitude can be found using any odd and even order. Phase can be estimated in the range $[0, \pi)$ using the 0th and any odd order, and in the range $[0, 2\pi)$ using the 0th and 1st orders. Orientation can be estimated in the range $[0, 2\pi/n)$ using the phase value and the n-th order response for $n \geq 1$. We note therefore, that to estimate orientation from even structures, even RTs orders of two or above must be used.

The previously mentioned approaches of the structure multi-vector [26, 31], 2D analytic signal [136] and signal multi-vector [135] all use the sinusoid estimates from (1.42) to obtain the parameters of a particular sinusoidal signal model. We shall review each along with some other local image structure representations using RTs.

Structure Multi-Vector

Felsberg and Sommer proposed the structure multi-vector [26], constructed from the 0th to 3rd order RT responses. They use a geometric algebra to describe the signal and encode the various symmetries that are present. The structure multi-vector model consists of two sinusoids at right angles,

$$f(\mathbf{z}) = \sum_{k=1}^{2} A_k \cos(\langle \mathbf{z}, \mathbf{o}_k \rangle + \phi_k), \tag{1.43}$$

where $\theta_2 = \theta_1 + \pi/2$. Including higher-order RTs leads to more orthogonal responses than model parameters (responses to RT orders above zero are complex, and therefore have two dimensions). The structure multi-vector deals with this by projecting the RT responses on to four complex basis functions which are then projected on to the five model parameters. Orientation is obtained from both even and odd RT orders, addressing the orientation problem of the monogenic signal.

2D Analytic Signal

Wietzke and Sommer proposed the 2D analytic signal [135], constructed from the 0th to 2nd order RT responses. The model consists of two sinusoids with the same phase,

$$f(\mathbf{z}) = \sum_{k=1}^{2} A_k \cos(\langle \mathbf{z}, \mathbf{o}_k \rangle + \phi). \tag{1.44}$$

Parameter estimation begins by solving for the apex angle, $\theta_1 - \theta_2$, from which the rest of the values are derived. However, there is an issue with the derivation method. When $|\mathcal{R}^2 f| > |\mathcal{R}^0 f|$ (60) in [135] gives a complex value for the apex angle, violating the model. This occurs for image structures, such as saddles, where the 2nd order RT response is larger than that of the other orders. Note, the name "2D analytic signal" is very broad, and shall actually be used to refer to the general class of all extensions of the analytic signal to 2D in this thesis.

Signal Multi-Vector

Wietzke and Sommer also proposed the signal multi-vector [135], constructed from the 0th to 3rd order RT responses. The authors use a quaternionic-valued matrix representation to describe the signal and encode the various symmetries that are present. The signal multi-vector model consists of two sinusoids without any restriction on the parameters,

$$f(\mathbf{z}) = \sum_{k=1}^{2} A_k \cos(\langle \mathbf{z}, \mathbf{o}_k \rangle + \phi_k). \tag{1.45}$$

As with the structure multi-vector, there are more orthogonal RT responses than parameters to estimate. The method deals with this by projecting the seven RT responses onto the six model parameters algebraically. However, again there are problems with the parameter estimation method. An image structure consisting of two equal amplitude sinusoids with even phases $\{0, \pi\}$ and opposite orientations $\{-\theta, \theta\}$ gives $\mathcal{R}^0 f = \mathcal{R}^1 f = \mathcal{R}^3 f = 0$ and $\mathcal{R}^2 f = 2A(e^{\mathrm{i}2\theta} + e^{-\mathrm{i}2\theta}) = 2A\cos(2\theta)$. Thus $\mathcal{R}^2 f$ is real-valued and the model parameters cannot be found due to having more unknowns than knowns. The addition of the 4th order RT response may therefore be required. Furthermore, the method of calculating orientation given by (130-131) in [135] uses only odd-order RT responses, and thus again the orientation estimate will be poor near even structures in the presence of noise.

Tensors

Other approaches do not explicitly use a sinusoidal signal model, and instead employ a tensor-based representation of local signal structure, where the RT is used in place of the traditional derivative operator. These include the boundary tensor [55, 56] proposed by Köthe and the monogenic curvature tensor [143] proposed by Zang and Sommer. The boundary tensor uses the 0th to 2nd order RTs to give estimates of phase-invariant *edge energy* (i1D features) and *junction energy* (i2D features). These values are combined to give the *boundary energy* which is an intrinsic-dimension invariant measure of feature strength. A mean orientation value is also obtained from the trace of the tensor. The monogenic curvature tensor [143] consists of an even and odd tensors formed using the 0th to 3rd order RTs. It gives amplitude, phase and orientation parameters along with a measure of the local curvature of the signal which can discern i1D and i2D features [142], and the angle of intersection between two i1D features [112].

Limitations

Apart from the problems with orientation estimation in some of these extensions, the main limitations are that either the image model is constrained or the number of RT orders used in its estimation is limited in order to have an analytic solution to the model parameters. Furthermore, the model is assumed to completely represent the local image structure, and thus to recover the signal from its amplitude, phase and orientation values the parameter estimation function should ideally be bijective, again restricting the number of RT orders.

1.5 2D Steerable Wavelets

The last approach to phase-based image representation that shall be reviewed is that of 2D steerable wavelets. The higher-order RT exists as complex exponential multiplier in the Fourier domain. This is equivalent to the circular harmonic functions (Fourier series angular component) of the 2D steerable filters reviewed earlier. However, typical steerable filters (e.g. [110]) often have a different radial frequency function depending on the order of the circular harmonic, due to either definition in the spatial domain, or the use of either wedge functions or derivatives to construct the higher-order components. However, if the basis filters are defined by circular harmonics of the frequency response of an isotropic filter *in the Fourier domain*, they are isomorphic with the set of RTs of the same filter.

Freeman and Adelson remarked that these kind of basis functions can used to construct *steerable wavelets*. Likewise, circular harmonics can be used as the basis functions for Simoncelli's steerable pyramid. More recently, after the realisation of the RT as the appropriate generalisation of the Hilbert transform to higher dimensions, much research has been performed on using the RT to generate *2D steerable wavelet frames* [94, 117, 123, 124].

In addition to steerability, the RT has properties of translation invariance, scale-invariance and inner-product preservation [29, 124]. Of particular relevance is that the RT is norm-preserving, $\|\mathcal{R}^n f\| = \|f\|$, and invertible, $\mathcal{R}^{-n}\{\mathcal{R}^n f\}(\mathbf{z}) = f(\mathbf{z})$, if $f(\mathbf{z})$ has zero mean. Thus given a higher dimensional representation of an image consisting of different order RTs, the original image can be reconstructed from any of the dimensions. This was one of the properties integral to the monogenic signal and subsequent 2D analytic signal approaches.

These properties also allow the generation of steerable wavelets from isotropic wavelet frames that satisfy three properties [124]. A steerable wavelet frame as it is commonly used in image processing is a steerable filter bank at multiple scales with the property that the original signal can be reconstructed using the responses to the filter bank across scales, known as the wavelet transform *coefficients*. There is no loss of information when applying the wavelet transform, in fact the coefficients are often highly redundant [124]. The coefficients are also usually sparse, meaning that a small number of values can be used to adequately reconstruct a much larger image [116]. Furthermore, with an appropriate choice of filters the image my be subsampled between scales [124].

Monogenic Wavelet Transform

Initial development surrounding wavelets and the RT was the *monogenic wavelet transform* proposed by a few different authors working along independent lines [47, 94, 122]. It consists of a isotropic wavelet frame and its 1st order RT. Held et al. [47] used the output of the transform to obtain a *multi-scale* representation of the monogenic signal from which sinusoidal model parameters could be obtained. Exploiting the reconstruction property of wavelets, they were able to perform image processing tasks such as brightness equalisation and de-screening by adjusting the model parameters before reconstructing the image. The process was to perform the transform, project onto the monogenic signal at each scale, adjust the model parameters making use of the invariant / equi-variant properties of amplitude, phase and orientation, then reconstruct the monogenic signal and reverse the wavelet transform. This approach was extended to colour images in [113].

Monogenic Curvelet Transform

The *monogenic curvelet transform* [117] developed by Storath solves the problem of having to approximate the Hilbert transformed part of a 2D quadrature filter. Instead of the Hilbert transform of an even directional wavelet, the RT is used. The pair of filters are then in quadrature in the sense of the monogenic signal. That is, given an even wavelet oriented along direction θ with frequency response defined by $h_e(\boldsymbol{\omega})$, the odd filter is given by $h_o(\boldsymbol{\omega}) = \mathcal{R}_\theta h_e(\boldsymbol{\omega})$ where $\mathcal{R}_\theta$ is the non-complex RT along the axis given by θ. The method is applied to curvelets, which have fixed angular support in the frequency domain and are therefore not steerable, but the principle could be applied to any even directional wavelet. The pair of the directional wavelet and its RT are known as *monogenic 2D quadrature filters*.

1.5.1 2D Steerable Wavelet Frames

Subsequent approaches in [123, 124] extended the monogenic wavelet concept by using higher-order RTs to create *2D steerable wavelet frames*. A framework for their design and use is laid out by Unser in [124]. This reference collates the developments from previous papers [121, 123, 125] as well as summarising various applications and is a good reference for the reader.

Some of the relevant findings shall now be presented. Restating proposition 4.1 from [124], if $h(\omega)$ is a radial frequency profile satisfying the following conditions:

$$h(\omega) = 0, \quad \forall\, \omega > \pi, \tag{1.46}$$

$$\sum_{i\in\mathbb{Z}} |h(2^i\omega)|^2 = 1, \tag{1.47}$$

$$\left.\frac{\mathrm{d}^n h(\omega)}{\mathrm{d}\omega^n}\right|_{\omega=0} = 0, \quad \text{for } n = 0, \ldots, N, \tag{1.48}$$

then the isotropic wavelet mother wavelet ψ with spectrum $\hat{\psi}(\boldsymbol{\omega}) = h(\|\boldsymbol{\omega}\|)$ generates a tight wavelet frame of $L_2(\mathbb{R}^2)$ whose basis functions, $\psi_{i,\mathbf{k}} = \psi_i(\mathbf{z} - 2^i\mathbf{k})$ with $\psi_i(\mathbf{z}) = 2^{-2i}\psi(\mathbf{z}/2^i)$, are isotropic with vanishing moments up to order N. In wavelet notation, i is the index of the scale of the wavelet, while $\mathbf{k}$ is the location in the image. The subscript notation is used for location as some wavelet frames are able to be *subsampled* between scales, changing the image domain.

Given a primary isotropic wavelet frame $\{\psi_{i,\mathbf{k}}\}_{i\in\mathbb{Z},\mathbf{k}\in\mathbb{Z}^2}$ that satisfies the above conditions, the higher-order RT can be used create a steerable wavelet frame $\{\psi_{i,\mathbf{k}}^{(m)}\}_{m\in\mathbb{N}_M^+, i\in\mathbb{Z},\mathbf{k}\in\mathbb{Z}^2}$ of $L_2(\mathbb{R}^2)$ by [124]

$$\psi_{i,\mathbf{k}}^{(m)} = \sum_{|n|\leq N} u_{m,n}\mathcal{R}^n\psi_{i,\mathbf{k}}, \tag{1.49}$$

where $\mathbf{U}$ is a complex valued shaping matrix of size $M \times (2N+1), M \geq 1$. The columns of $\mathbf{U}$ and denoted $\mathbf{u}_m$ and each describes the coefficients of a 2D steerable wavelet. The wavelet is the sum of the RTs of the primary isotropic wavelet, where the magnitude, $|u_{m,n}|$, modulates the n-th order RT and the argument, $\arg(u_{m,n})$, is its rotation. Exact reconstruction of the image from these wavelet coefficients is then possible according to

$$f(\mathbf{z}) = \sum_{i,\mathbf{k}}\sum_{m}\left\langle \psi_{i,\mathbf{k}}^{(m)}, f\right\rangle \psi_{i,\mathbf{k}}^{(m)}, \tag{1.50}$$

so long as $\mathbf{U}\mathbf{U}^H$ is a diagonal matrix whose elements sum to 1. The second condition results in an energy-preserving partition of the frequency spectrum, creating a Parseval-tight wavelet frame. A Parseval-tight wavelet frame means that we can invert the wavelet transform to obtain an exact reconstruction of the original image. Note that while scaling by 2^i allows for sub-sampling to create pyramidal decompositions, alternative partitions that are more narrowly spaced can be used, such as in [47]. For discrete images if the the second condition is relaxed to

$$\sum_{i\in\mathbb{Z}} |h_i(\omega)|^2 = 1, \tag{1.51}$$

then one also obtains a Parseval-tight frame, where i is the scale index of the wavelet. The third condition requires the primary wavelet to have at least N vanishing moments to account for the singularity of the RT at the origin, and for the wavelets to have sufficient spatial decay [124, 131].

If reconstruction or pyramidal decompositions are not of interest, the second condition can be abandoned, and an image analysed using an isotropic filter bank that preferably satisfies the third condition. That is, a filter bank consisting of the linear combination of spherical quadrature filters.

Circular Harmonic Frame

The basis of other 2D steerable wavelets are the *circular harmonic* (CH) wavelets. The CH frame, $\{\psi^n_{i,\mathbf{k}}\}_{|n|\in\mathbb{N}_N, i\in\mathbb{Z}, \mathbf{k}\in\mathbb{Z}^2}$, is given by the shaping matrix $\mathbf{U} = \mathbf{I}_{2N+1}$ [124]. It consists of wavelets given by the $-N$-th to N-th order RTs of the primary isotropic wavelet. These wavelets are also known as CH functions [51]. Each wavelet is given by

$$\psi^n_{i,\mathbf{k}} = \mathcal{R}^n \psi_{i,\mathbf{k}}. \tag{1.52}$$

The CH wavelets have equal norm with frame bounds of $2N + 1$. Reconstruction from a wavelet frame given by the shaping matrix $\mathbf{U}$ can be expressed as

$$f(\mathbf{z}) = \sum_{i,\mathbf{k}} \sum_{m} \left\langle \psi^{(m)}_{i,\mathbf{k}}, f \right\rangle \sum_{|n|\leq N} u_{m,n} \psi^n_{i,\mathbf{k}}. \tag{1.53}$$

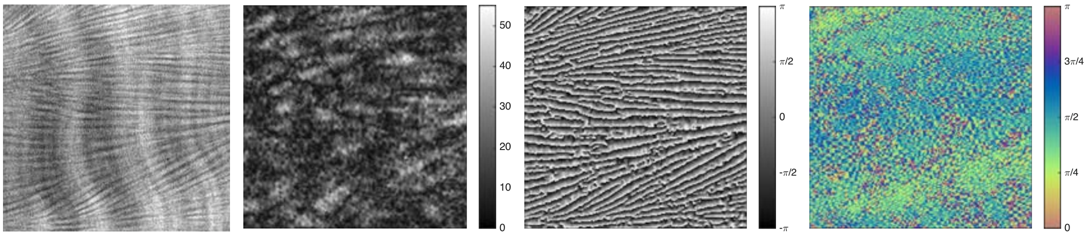
The CH wavelets are the wavelet equivalent of the SQFs previously mentioned (Figure 1.7). As such, the monogenic signal and other 2D analytic signals can be constructed from their responses.

1.6 Motivation

This thesis research began with a study of the monogenic signal. The sinusoidal model derived from the monogenic signal appeared to provide a useful representation of an image, particularly the orientation estimate. At the beginning of the research I was helping to fix the densitometer at the Australian Institute of Marine Science. It is used to measure the density of coral core slices to gather historical records of coral growth rate. The best axis along which to take the measurements is estimated by a human looking at an x-ray of the slice (Figure 1.8a). It was decided to apply the monogenic signal to estimate the best measurement axis in a more principled way. However, the aforementioned problem of the orientation estimation being from only odd orders gave a noisy and unsatisfactory response (Figure 1.8d)

Subsequently, the 2D analytic signal and signal multi-vector were also applied to see if they gave a better estimate. However the model of the 2D analytic signal cannot represent all parts of an image, and thus the code provided in [133] would result in divide-by-zero errors. These were masked by their graceful failure when implemented in the original C code, but MATLAB was not as forgiven in the ported version. Similarly, the signal multi-vector exhibited the same orientation estimation problem as the monogenic signal.

Improving the estimation of the sinusoidal model parameters to solve these problems was thus



(a) *Coral* (b) Amplitude (c) Phase (d) Orientation

Figure 1.8: Section of coral core image and the amplitude, phase and orientation of the sinusoidal model given by the monogenic signal.

the initial motivation behind this work. Since higher-order RTs give more estimates of the model parameters, the focus was on creating a method that could use these extra estimates. In particular, the following research questions were posed:

1. ***How can we derive the sinusoidal model of local image structure using higher-order RTs?***
2. ***How do we account for the extra terms without expanding the model like in previous approaches?***

The first attempt at addressing these questions used the 0th to 2nd order RTs to give an analytic solution to the sinusoidal model that estimated orientation from even structures as well as odd [72]. Subsequent work found the parameters by minimising the distance between a vector of RT responses at a point in an image, and the vector representing the model [73, 74].

Around the start of this research, Unser and others began publishing work on using the RT to generate steerable wavelet frames. Initially, this began with monogenic wavelets constructed using the first order RT, and later with 2D steerable wavelet frames constructed from higher-order RTs. Two methods in the wavelet literature that address the first research problem are the monogenic curvelet transform [117] and the even and odd harmonic wavelets briefly described in [124]. Recently, Puspoki et al. have been investigating the detection of M-fold (rotationally) symmetric junctions [104, 105], such as Y and X junctions, and although they do not investigate lines and edges, it is conceivable that the method could be also applied.

Since the wavelet literature provides a solid mathematical foundation for using higher-order RTs, it was decided to reformulate the research in this context. However, a number of observations were made that motivated a new approach:

- 2D steerable wavelet methods tend to be focussed on the tight wavelet frame property. The wavelets are chosen such that the transform is reversible. For example, the monogenic curvelet transform necessarily consists of $N+1$ copies of a steerable wavelet at equally spaced discrete orientations to have exact reconstruction, whereas to derive the sinusoidal model we need to find the single best orientation.
- The wavelet frames are generally designed or calculated for specific features or properties.

What if we want to use a variety of wavelets corresponding to a diverse set of features which *do not* form a frame?

- Recovery of the original signal from a higher dimensional representation is a feature of the monogenic signal, signal multi-vector and other 2D analytical signals. However, simply steering a wavelet to the best orientation of linear symmetry may deliver the signal model parameters, but lose information about i2D structures, and the original signal is not recoverable. A tiling of the frequency domain using copies of the wavelets at discrete scales and orientations will form a tight frame, however the wavelet coefficients will no longer be invariant / equi-variant with rotation.

Not wanting to limit the focus to a sinusoidal model, the scope of the thesis was expanded to investigate how to derive *general signal models* using higher-order RTs. This lead to the following research questions, which are the subject of this thesis:

3. ***How can we represent local signal structure using higher-order RTs in a principled way?***

4. ***How can we derive the parameters of a particular image model from this representation?***

5. ***How can we use wavelets that correspond to a particular structures of interest but do not form a frame, and yet still have exact reconstruction?***

6. ***What are the coefficients of wavelets that correspond to common image features?***

As a result, a general method of solving local image modelling problems using higher-order RTs has been developed. The main points of the method are as follows:

- Local image structure is represented as a *vector* of *circular harmonic* (CH) wavelet coefficients (RT responses) that is *weighted* to give phase invariance. The vector generalises existing signal vectors such as the monogenic signal.

- Wavelets are created to *match* particular local image structures of interest and are described by their weighted CH vector.

- The CH vector is split into *model* and *residual* components by correlation with the matched wavelet CH vector at the optimal orientation.

- The model component is used to describe image structures of interest, from which amplitude, orientation and phase parameters are obtained.

- The residual component describes the remainder of the local signal structure, and provides the rest of the information needed for exact reconstruction of the original signal.

The method and its applications are developed in the following chapters. The research is exploratory in nature, and is focussed on creating a new framework for local image analysis, not for a specific problem or application. The layout of the thesis is as follows:

- In chapter two the CH vector is introduced as a primary descriptor of local image structure. A generic local image model consisting of sets of feature components with different amplitudes and orientations is proposed. We show how to find the weighted CH vector matched to each model component. A suite of methods for solving for the amplitude and orientation parameters with respect to various model constraints is developed. Each method is demonstrated on a test image.

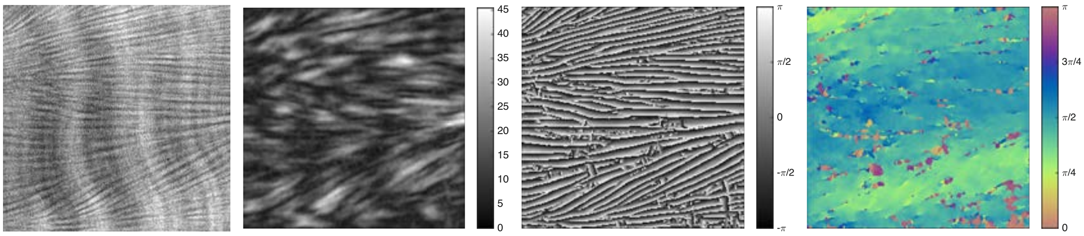
- In chapter three we apply the method to solving for the single sinusoidal model used by the monogenic signal and other quadrature filters. Choice of primary isotropic wavelet, number of RT orders and CH vector weighting is investigated. The usefulness of the sinusoidal model is demonstrated for orientation estimation. Furthermore, the residual component, a novel aspect of this work, is used along with the model component to create a new measure of intrinsic dimension that can be used for junction and corner detection.

- In chapter four a multi-sinusoidal model is proposed that can describe multiple i1D orientations. The model is applied to the problem of finding the amplitude and orientation of features consisting of additive i1D or occluded line components. A hybrid wavelet set for analysing occluded lines is introduced, as well as an order-dependent threshold for classifying by number of components. The method is compared to two recent approaches in the literature for estimating the orientation of lines. Finally, the multi-sinusoidal model is used to solve the initial problem that motivated the research, the estimation of growth direction in coral core X-ray images.

- In chapter five a half-sinusoidal model is introduced. The model can describe features consisting of line and edge *segments* radiating from a point, such as Y or X junctions. The model can also be used to design wavelets matched to specific features.

- In chapter six the same principles used to solve for model orientation are applied to finding the best *scale* at which to perform analysis. A Fourier series decomposition of the *radial* part of the frequency domain is used along with the RT (Fourier series in the angular frequency domain) to create filters shiftable in *both* orientation and scale. A quaternionic polynomial representation of the magnitude of the CH vector through scale is obtained. This gives a measure of local structure energy from which measures such as the maximum, mean, and variance of the scale response are found.

- Finally, the last chapter gives an overall summary of the work and outlines future research directions.



(a) *Coral* (b) Amplitude (c) Phase (d) Orientation

Figure 1.9: Section of coral core image and the amplitude, phase and orientation of the sinusoidal model using up to the 7th order RT.

1.7 Published Work

The following is a list of six conference and one journal papers that were published and the chapters to which they relate. All were co-authored with my supervisor, Dr. Paul Jackway, except one. The ideas within the papers were my own and I performed the writing, experiments and content generation of the material within. Dr. Jackway's contribution was to review the papers and offer suggestions for improvements.

1.7.1 Conference Papers

- R. Marchant and P. Jackway, "Generalised Hilbert Transforms for the Estimation of Growth Direction in Coral Cores," in *2011 International Conference on Digital Image Computing: Techniques and Applications*, IEEE, 2011, pp. 660–665

  Chapters 3 and 4

- R. Marchant and P. Jackway, "Feature Detection from the Maximal Response to a Spherical Quadrature Filter Set," in *2012 International Conference on Digital Image Computing Techniques and Applications (DICTA)*, Perth: IEEE, 2012, pp. 1–8

  Chapters 2 and 5

- R. Marchant and P. Jackway, "Local feature analysis using a sinusoidal signal model derived from higher-order Riesz transforms," in *2013 IEEE International Conference on Image Processing*, IEEE, 2013, pp. 3489–3493

  Chapter 3

- R. Marchant and P. Jackway, "Using Super-Resolution Methods to Solve a Novel Multi-Sinusoidal Signal Model," in *2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, IEEE, 2013, pp. 1–8

  Chapter 4

- R Marchant and P Jackway, "Modelling Line and Edge Features Using Higher-Order Riesz Transforms," in *Advanced Concepts for Intelligent Vision Systems*, ser. Lecture Notes in

Computer Science, vol. 8192, 2013, pp. 438–449
Chapter 5

- R. Marchant, "Scale Adaptive Filters," in *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, vol. 32, IEEE, 2015, pp. 1–8
Chapter 6

1.7.2 Journal Papers

- R. Marchant and P. Jackway, "A sinusoidal image model derived from the circular harmonic vector," *Journal of Mathematical Imaging and Vision*, 2016
Chapters 2 and 3

Chapter 2

Circular Harmonic Vector

In this chapter a general framework for local image analysis using higher-order RTs is presented. Three of the last four research questions proposed in the introduction are addressed: How do we represent local signal structure using higher-order RTs, how do we derive the parameters of a particular image model from this description, and how do we still maintain recovery of the original image?

To achieve these goals, an algebraic framework for both representing the higher-order RT responses and solving for the model parameters is required. The monogenic signal and signal multi-vector were formulated using both quaternion and geometric algebras in order to make explicit their geometrical interpretation. However, the algebras were specific to the number of RT orders used. Rather than delve into these representations we shall employ a simpler approach using vector and matrix algebra and leave it to the signal model to describe the geometry.

The proposed solution is to collect the higher-order RT responses into a *vector*, which forms a representation of local image structure at a particular frequency scale given by the band-pass filter used to localise the RT responses. This vector will then be used as the primary unit for deriving signal model parameters. The parameters are found by splitting the vector into model and residual components, and minimising the residual. The residual component provides the necessary information not captured by the model to allow recovery of the original image. In this way the representation is similar to previous 2D analytic signals, but allows for arbitrary models.

A general approach is presented which will be applied for specific image models in later chapters. The layout of the development is as follows:

- Introduction of circular harmonic vector as a primary description of the local image structure that generalises other signal vectors.
- Introduction of a generic local image model and its wavelet response.
- Methods of solving the signal model given various constraints.

2.1 Circular Harmonic Vector

Previous 2D analytic signals collected up to the 3rd order RT responses into either vector [29] or matrix [134, 143] forms. This section introduces the CH vector as a representation of local image structure from any number of higher-order RTs.

2.1.1 Circular Harmonic Frame

In the introduction we reviewed the CH wavelet frame, of which the CH wavelets are a basis for all other 2D steerable wavelets. The CH frame, $\{\psi^n_{i,\mathbf{k}}\}_{|n|\in\mathbb{N}_N, i\in\mathbb{Z}, \mathbf{k}\in\mathbb{Z}^2}$, is given by the shaping matrix $\mathbf{U} = \mathbf{I}_{2N+1}$ [124] and consists of wavelets given by the $-N$-th to N-th order RTs of the primary isotropic wavelet,

$$\psi^n_{i,\mathbf{k}} = \mathcal{R}^n \psi_{i,\mathbf{k}}. \tag{2.1}$$

The CH wavelets have equal norm with frame bounds of $2N + 1$.

The CH wavelets are the wavelet equivalents of the SQFs used by Felsberg , Wietzke , Zang and Sommer [29, 135, 143] to give localised representations of the monogenic single and other signal models. The set of wavelets in the CH frame includes all orders up to N, and wavelet theory gives us a rich mathematical basis to perform image reconstruction from the responses. That is, we can recover the original image from the wavelet coefficients. Formulating answers to the research questions using a wavelet context as opposed to filter banks is thus an attractive proposition.

However, to begin with we need a higher-dimensional representation of the image that includes these higher-order RT responses, such as is used in the monogenic signal. It is proposed to collect the $-N$-th to N-th order CH wavelet responses into a vector, called a *CH vector*. Applying the CH wavelet frame to an image $f \in L_2(\mathbb{R}^2)$, the CH vector of correlation coefficients at each scale and location is given by

$$\mathbf{f}_{i,\mathbf{k}} = \left[\left\langle f, \psi^{-N}_{i,\mathbf{k}}\right\rangle, ..., \left\langle f, \psi^{N}_{i,\mathbf{k}}\right\rangle\right]^T. \tag{2.2}$$

Thus the CH vector is a vector of higher-order RT responses up to order N, localised by the primary isotropic wavelet. The negative RT orders are included in the vector as the corresponding wavelets are required to synthesise other 2D steerable wavelets.

Because the CH frame bounds are $2N + 1$, to obtain exact reconstruction (frame bounds = 1) the CH vector coefficients must be *weighted*. Let $\mathbf{W}$ be a real-valued diagonal weighting matrix,

$$\mathbf{W} = \text{diag}(\mathbf{w}), \tag{2.3}$$

$$\mathbf{w} = [w_{-N}, ..., w_N], \tag{2.4}$$

where $w_{-n} = w_n$ and $\sum_{|n|\leq N} w_n^2 = 1$. The set of weighted CH wavelets $\{w_n\psi^n_{i,\mathbf{k}}\}_{|n|\in\mathbb{N}_N, i\in\mathbb{Z}, \mathbf{k}\in\mathbb{Z}^2}$

thus have a frame bound of 1. Exact reconstruction is then possible using the *weighted* CH vector, $\mathbf{Wf}$, according to

$$f(\mathbf{z}) = \sum_{i,\mathbf{k}} \sum_{|n| \leq N} (\mathbf{Wf}_{i,\mathbf{k}})_n w_n \psi^n_{i,\mathbf{k}}. \tag{2.5}$$

Thus like the monogenic signal and signal multi-vector, the CH vector is a higher-dimensional representation of an image from which the original image can be recovered.

2.1.2 CH Vector Local Descriptor

It is proposed that the CH vector is a *local descriptor* of image structure. Properties of the vector such as the individual CH wavelet responses and the magnitude of the vector describe different aspects of this. Each are discussed in the following sections. The i and $\mathbf{k}$ wavelet scale and location indices are dropped for clarity where appropriate, and one can assume that the CH vector at a point of interest is being referred to.

Channel Amplitude

The amplitude of the n-th component, $|f_n|$, represents the magnitude of the n-th order rotational symmetry of the local image structure.

This can be seen in Figure 2.1, which shows the 0th to 8th order CH wavelet responses for the *Board* image using a log-Gabor primary isotropic filter with wavelength 24 and $\sigma = 0.6$. The *Board* image was taken from [86]. See [11] for a reference on filters such as the log-Gabor.

It can be observed that:

- The 0th order is isotropic and responds to both isotropic and even structures (lines).
- The 1st order responds to odd structures (edges), as do subsequent odd-ordered wavelets.
- The 2nd order responds to even structures (lines), as do subsequent even-ordered wavelets. However it does not respond to T junctions, X junctions or corners.
- The 3rd order responds to edges, slightly off centre of corners and T junctions, but not to the X junctions.
- The 4th order responds to both the T and X junctions.
- Higher orders are more complex in their response, but also respond to features where the wavelet order is a multiple of the rotational symmetry of the feature.

Essentially, if a feature consists of either multiple line segments (or edge segments) radiating from a point, and the angle between two segments is a close to a multiple of $2\pi/n$, then the n-th order CH wavelet will respond to that feature. Thus corners, T junctions or X junctions whose

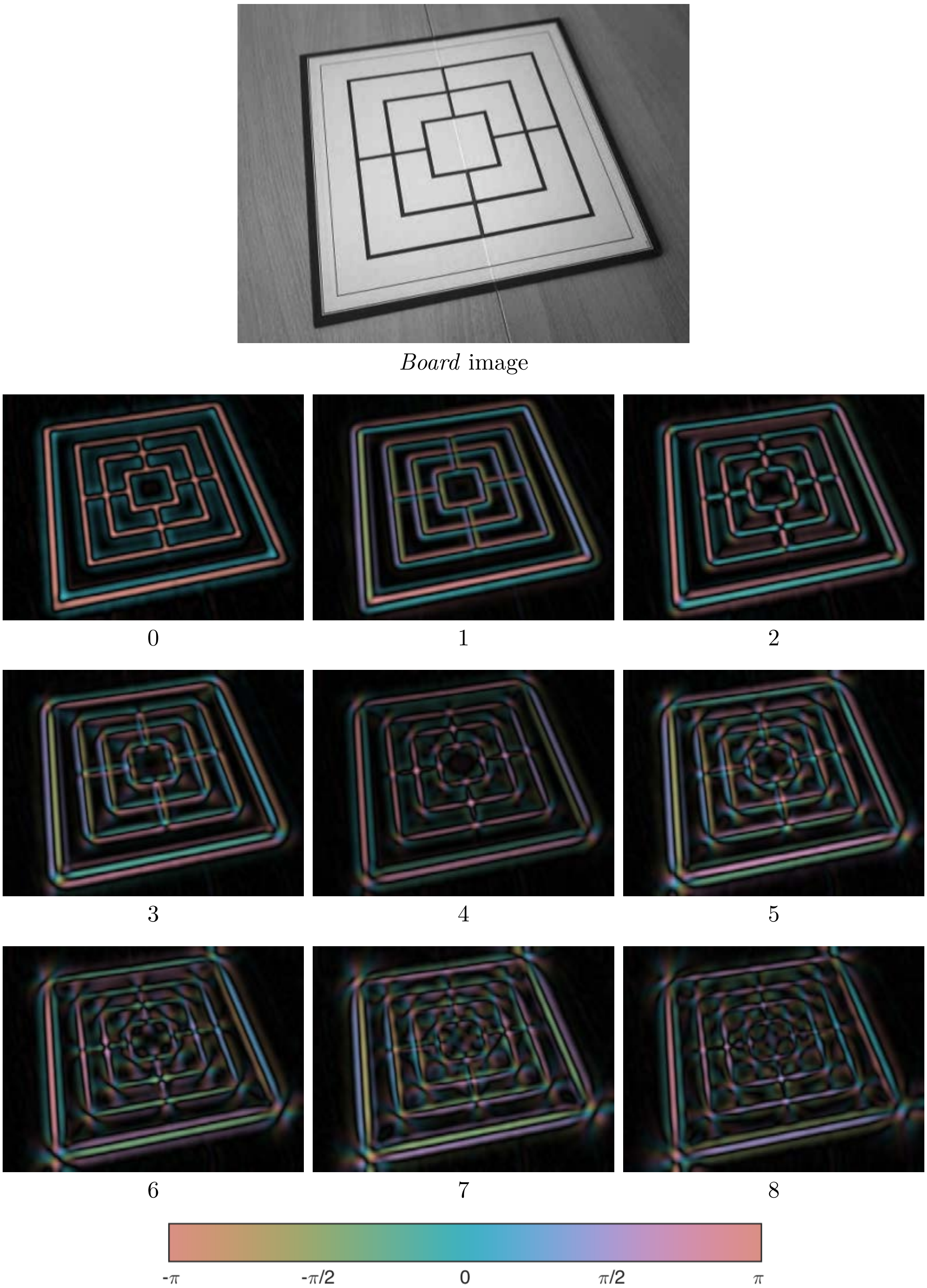


Figure 2.1: 0th to 8th order CH wavelet responses for the *Board* image using a log-Gabor filter with wavelength 24 and $\sigma = 0.6$. Brightness: magnitude of response, colour: angle of the response. For the 0th order, colour represents negative (red) or positive (blue) response.

line segments are close to $\pi/2$ radians apart, require the 4th order CH wavelet for analysis. This explains why the signal multi-vector model fails for two sinusoids at right angles with a phase of 0. The structure is that of an X junction and thus requires the 4th order to discriminate from a blob (0th order). However, only up to the 3rd order is used in the derivation of that model.

Higher-order RT kernels are orthogonal and have more axes of symmetry. Correlation of the CH wavelets with an image therefore gives independent measurements of the strength and orientation of the symmetries present in the local image structure. Signal models that involve extra symmetries therefore need high enough order RTs to be able to discriminate the features. This equates to a *larger* CH vector.

Channel Orientation

The argument of the n-th component, $\arg(f_n)$*, represents the orientation offset of the n-th order symmetry.*

This can be also be seen in Figure 2.1:

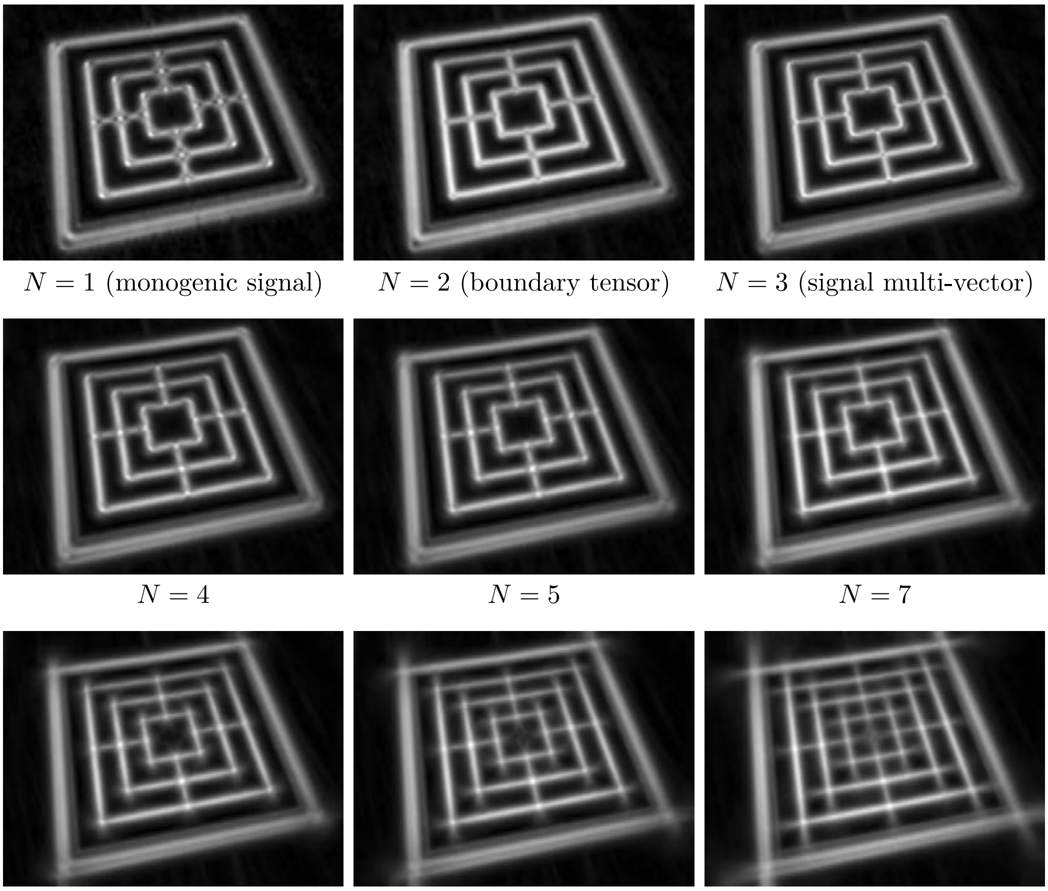
- The 0th order is isotropic and thus indicates the sign of the local structure compared to surrounding pixel values.
- The 1st order estimates the orientation of odd structures (edges) over $[0, 2\pi)$.
- The 2nd order estimates the orientation of even structures (lines) over $[0, \pi)$.
- Subsequent odd orders estimate the orientation of odd structures (edges) over $[0, 2\pi/n)$.
- Subsequent even orders estimate the orientation of even structures (lines) over $[0, 2\pi/n)$.
- Higher orders also give an orientation estimate for features where the wavelet order is a multiple of the rotational symmetry of the feature.

Thus higher-orders can be used to augment the orientation estimate of the lower orders. These extra estimates will be put to use in the next chapter, where they are applied to solving for the sinusoidal model.

Vector Norm

The norm of the CH vector, $\|\mathbf{Wf}\|$*, is a measure of local energy.*

The norm of the CH vector is high at the location of image features, as shown for the *Board* image in Figure 2.2 for different values of N. In the examples, the vector has been weighted so that all the odd orders have the same weight, all even orders have the same weight, and the 0th order weighting is $\sqrt{2}$ times higher than the other even orders, due to the inclusion of negative orders in



$N = 1$ (monogenic signal) $N = 2$ (boundary tensor) $N = 3$ (signal multi-vector)

$N = 4$ $N = 5$ $N = 7$

$N = 9$ $N = 13$ $N = 21$

Figure 2.2: CH vector magnitude of the *Board* image for different values of N using a log-Gabor filter with wavelength 24 and $\sigma = 0.6$. The CH vector was weighted to match that of the monogenic signal, boundary tensor and signal multi-vector.

the vector. Using this weighting scheme, the norm for $N = 1$ is the same as the monogenic signal amplitude, the norm for $N = 2$ is the same as the boundary tensor [56] magnitude, and the norm for $N = 3$ is the magnitude of the signal multi-vector.

For $N = 1$ there are locations near the T and X junctions where the response is quite low, as indicated by 'holes' in the energy image. These locations do not respond to either the 0th or 1st order CH wavelets. Adding the 2nd order (boundary tensor) improves the response, and by $N = 7$ the response to boundary objects (lines and edges) appears more uniform. However, as N increases further, the magnitude begins to smear along the direction of the lines. This is due to the increasing size of the CH wavelets, meaning a higher response is possible further from the feature. The change from holes in the magnitude to smearing suggests an ideal range of N when using the CH vector norm as a boundary measure.

Weighting

Weighting each order differently also allows us to control how each order contributes to the magnitude of the CH vector. In particular, the sets odd and even orders can be weighted equally so that the magnitude is phase invariant. This is covered in the next chapter.

Comparing CH Vectors

One may wish to compare different image structures using their corresponding CH vectors. A simple measure is the distance between the vectors,

$$\mathrm{d}(\mathbf{Wf}, \mathbf{Wg}) = \|\mathbf{Wf} - \mathbf{Wg}\|. \tag{2.6}$$

However, the distance varies with the magnitude of the vectors and thus the strength of the local structure. It is often desirable to compare structures by shape alone. The normalised weighted CH vector,

$$\frac{\mathbf{Wf}}{\|\mathbf{Wf}\|}, \tag{2.7}$$

gives an illumination-invariant description of the local image structure. That is, it describes its shape separately to its strength. Therefore to introduce illumination invariance we may normalise the vectors,

$$\mathrm{d}_{\mathrm{norm}}(\mathbf{Wf}, \mathbf{Wg}) = \left\| \frac{\mathbf{Wf}}{\|\mathbf{Wf}\|} - \frac{\mathbf{Wg}}{\|\mathbf{Wg}\|} \right\|, \tag{2.8}$$

or use the angle difference between them,

$$\gamma = \cos^{-1} \frac{\langle \mathbf{Wf}, \mathbf{Wg} \rangle}{\|\mathbf{Wf}\|\|\mathbf{Wg}\|}. \tag{2.9}$$

Rotation

The CH vector can be steered and its magnitude is invariant to rotations.

Rotation of the image causes a rotation of the CH vector components (1.30). For rotation of the image axes by θ, the CH vector is given by

$$\mathbf{f}(\mathbf{R}_\theta \mathbf{z}) = \mathbf{S}_{-\theta} \mathbf{f}(\mathbf{z}), \tag{2.10}$$

where $\mathbf{f}(\mathbf{z})$ is the image CH vector, and

$$\mathbf{S}_\theta = \mathrm{diag}\left[e^{-\mathrm{i}N\theta}, ..., e^{\mathrm{i}N\theta}\right] \tag{2.11}$$

is a diagonal rotation matrix. Since the rotation matrix components are unitary-magnitude complex exponentials we have

$$\|\mathbf{f}(\mathbf{R}_\theta \mathbf{z})\| = \|\mathbf{S}_{-\theta}\mathbf{f}(\mathbf{z})\|, \tag{2.12}$$

and thus the CH vector magnitude is *invariant to rotations*. This result is important in that it performs a split of identity of the original image. That is, we can represent the CH vector as a magnitude component (strength) times a normalised vector (shape) that can be separately rotated,

$$\mathbf{S}_\theta \mathbf{W}\mathbf{f} = \|\mathbf{W}\mathbf{f}\| \times \mathbf{S}_\theta \frac{\mathbf{W}\mathbf{f}}{\|\mathbf{W}\mathbf{f}\|}. \tag{2.13}$$

Scale

By using a wavelet frame, scale is implicit in the representation and the scale of the basis wavelet, i, indicates the size of the local structure under consideration.

Wavelets

Other 2D steerable wavelets are described by their CH vector. In a steerable wavelet frame, each wavelet CH vector corresponds to one of the columns of the shaping matrix $\mathbf{U}$. Therefore we may write the correlation of the image and a wavelet as the correlation of the wavelet CH vector, $\mathbf{u}_m$, and the image CH vector, $\mathbf{f}$. That is,

$$\left\langle \psi_{i,\mathbf{k}}^{(m)}, f \right\rangle = \mathbf{u}_m^H \mathbf{f}_{i,\mathbf{k}}, \tag{2.14}$$

and thus the vector of coefficients for a particular wavelet frame is given by $\mathbf{U}^H\mathbf{f}$. The response to a set of 2D steerable wavelets rotated to a particular orientation θ can therefore be obtained by rotating the CH wavelet responses according to $\mathbf{S}_\theta \mathbf{U}^H \mathbf{f}_{i,\mathbf{k}}$.

Thus wavelets in the 2D steerable wavelet frames in [124] can be represented by their CH vectors. However, in [124] wavelet frames are designed for specific applications and weighted to give exact reconstruction, whereas in the approach developed in this thesis, the CH wavelet frame is used so that the response to any steerable wavelet can be investigated. Thus weighting is applied to make the CH wavelet frame tight and can be adjusted, rather than fixed according to a specific set of wavelets.

Generalisation

The CH vector is isomorphic with previous 2D analytical signals

- The components of monogenic signal, structure multi-vector, 2D analytic signal, monogenic curvature tensor, boundary tensor and signal multi-vector can all be obtained from the CH vector with $N \leq 3$.

- The CH vector with $N = 1$ and weighting vector $\mathbf{w} = [1/2, 1/\sqrt{2}, 1/2]$ has a magnitude $\sqrt{2}$ times amplitude of the monogenic signal. The sinusoidal model parameters are obtained from the CH vector as

$$A = \sqrt{2}\|\mathbf{Wf}\|, \tag{2.15}$$

$$\phi = \arg(f_0 + \mathrm{i}\,|f_1|), \tag{2.16}$$

$$\theta = \arg(-\mathrm{i}\,f_1). \tag{2.17}$$

- The CH vector with $N = 2$ and weighting vector $\mathbf{w} = \sqrt{[1/2, 1, 1, 1, 1/2]\,/4}$ has a magnitude equal to the square root of the boundary tensor energy (see [55, 56]).

Since the CH vector can represent previous 2D analytic signals, many of the observations made for those signals can be similarly applied. For example, detection using the boundary tensor is possible using the CH vector with $N = 2$, and the multi-sinusoidal model of the signal multi-vector can be derived from the CH vector with $N = 3$.

Summary

The CH vector provides a description of the symmetries of the local image structure up to order N. Its magnitude describes strength separately to the shape, which is described by the normalised vector.

2.2 General Signal Model

The initial motivation behind this research was to use higher-order RTs to model local image structure as an oriented sinusoid,

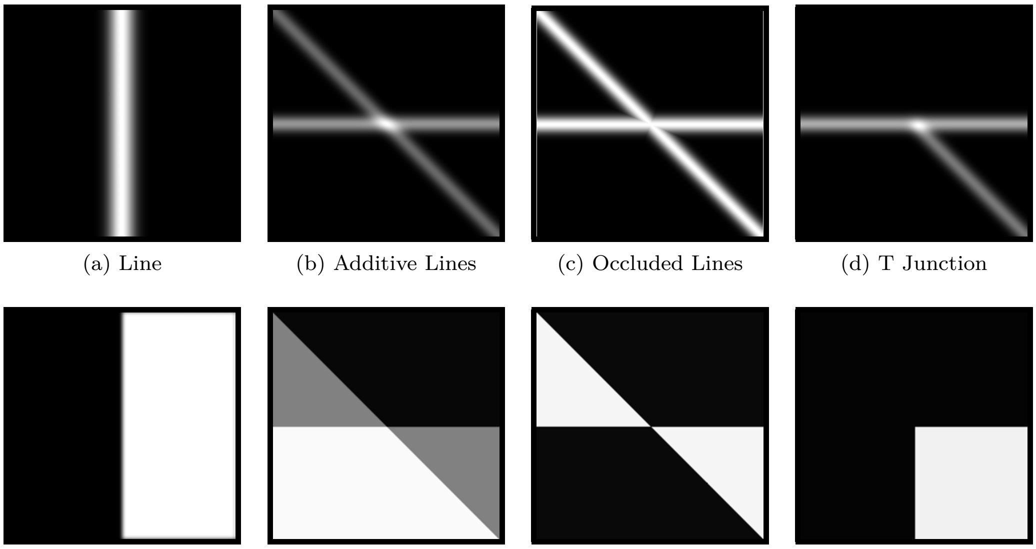
$$f(\mathbf{z}) = A\cos(\langle \mathbf{z}, \mathbf{o} \rangle + \phi), \tag{2.18}$$

as is performed when using 2D quadratic filters and the monogenic signal. However, the scope was expanded to develop a *general* signal model that could represent a wide range of common image features, such as corners and junctions. Examples of some common features are shown in Figure 2.3. Many can be described as the sum of individual components at a particular orientation. For example, a line is a single line component at one orientation (Figure 2.3a), an X junction is two line components at two orientations (Figure 2.3b) and a T junction is three line *segments* at three orientations (2.3d).

To investigate such features we need:

1. A general image model that can be used to represent these common types of structures.
2. A set of steerable wavelets that respond to the different model components.

3. A method of estimating the amplitude and orientation of each model component from the wavelet responses.



(a) Line (b) Additive Lines (c) Occluded Lines (d) T Junction

(e) Edge (f) Additive Edges (g) Chequer Edge (h) Corner

Figure 2.3: Examples of various image features that can be represented by lines and edges.

This approach is used in many existing methods of modelling local image structures, particularly for corners and junctions. For example, 2D steerable quadrature filters have been used for a phase-invariant estimate of multiple local orientations [37, 38, 99, 111] and one-sided quadrature filters have been used for junction parametrisation [50, 80]. Other methods use a pair of steerable wedge filters defined in the spatial domain, such as [110]. This approach has been expanded by Muhlich and others in [86, 88, 91], who steer *sets* of wedge filters with common orientations chosen to match the constraints of a particular structure. The method is called *multi-steerable matched filters* (MSMFs). For example, one can steer the combination of a single wedge filter and two wedge filters fixed 180 degrees apart to analyse a T junction. The idea being that if the model is known beforehand, the wedge filters can be fixed in number and angular relation.

The proposed method shares a similar sentiment to that of the MSMFs, in that different constraints are added to the model. However, unlike MSMFs structures of interest are not restricted to those which respond to a wedge, and instead 2D steerable wavelets are used as the measurement kernels. Furthermore, the following novel aspects are added to the approach:

4. The model parameters will be derived beginning with the CH vector representation of local image structure.

5. The CH vector will be split into model and residual vectors, for which the latter represents the part of the image that is *not* represented by the model, and solved by minimising the residual.

These latter points differentiate the method from previous approaches in the steerable filter and wavelet literature. Like 2D analytic signals, a general description of the local image structure is used, rather than beginning with a representation that is specific to one model. Specifically, the approach is to treat the vector of 2D steerable CH wavelet responses (CH vector) as the *primary descriptor* of local image structure. Then given a set of wavelets that match the components of particular signal model, the CH vector is split into model components that describe the structures of interest, and a residual component that describes the remainder. The model parameters are solved by minimising the residual CH vector magnitude.

This section introduces a general image model that describes common image structures, such as those in Figure 2.3. The subsequent sections detail how to solve for the parameters of the image model using the CH vector, given various constraints.

2.2.1 Proposed Model

All of the feature examples in Figure 2.3 can be described by the superposition of individual components rotated around a point located at the centre of the feature. A general image model is proposed for these kinds of features. It consists of the linear combination of K structures of interest, $\{u_k(\mathbf{z})\}_{k\in\mathbb{N}_K}$, with individual strengths (amplitudes), $\{\lambda_k\}$, and rotated to different orientations, $\{\theta_k\}$, plus a residual component, $f_\epsilon(\mathbf{z})$. At a point of interest, $\mathbf{z} = 0$, the model is given by

$$f(\mathbf{z}) = \sum_{k=1}^{K} \lambda_k u_k(\mathbf{R}_{\theta_k}\mathbf{z}) + f_\epsilon(\mathbf{z}). \tag{2.19}$$

For example, this general model for a T junction (Figure 2.3d), could consist of straight line and line-segment components and thus $K = 2$, or alternatively line-segment components and $K = 3$. T junctions with different angles or amplitudes could be represented by same model, except with different orientation or amplitude parameters.

A more expansive image model consists of the linear combination of K *sets* of one or more (M_k) image structures. The structures within each set have a common orientation, $\{\theta_k\}$, but different amplitudes $\{\lambda_{k,m}\}$. This model generalises the previous one, however both are introduced to aid the understanding of later derivations. The model is,

$$f(\mathbf{z}) = \sum_{k=1}^{K}\sum_{m=1}^{M_k} \lambda_{k,m} u_{k,m}(\mathbf{R}_{\theta_k}\mathbf{z}) + f_\epsilon(\mathbf{z}). \tag{2.20}$$

An example of this kind of model is the two-sinusoidal model of the signal multi-vector,

$$f(\mathbf{z}) = \sum_{k=1}^{2} A_k \cos(\langle \mathbf{z}, \mathbf{o}_k\rangle + \phi_k), \tag{2.21}$$

which can be written as multiple even and odd components

$$f(\mathbf{z}) = \sum_{k=1}^{2} \lambda_{k,1}\cos(\langle \mathbf{z}, \mathbf{o}_k\rangle) - \lambda_{k,2}\sin(\langle \mathbf{z}, \mathbf{o}_k\rangle), \tag{2.22}$$

where the scalars $\lambda_{k,1} = A_k \cos(\phi)$ and $\lambda_{k,2} = A_k \sin(\phi)$ are related to the sinusoid amplitude and phase. In fact, an amplitude and phase-vector representation can be given for a general set of features as

$$A_k = \left(\sum_m \lambda_{k,m}^2 \right)^{1/2}, \tag{2.23}$$

$$\boldsymbol{\phi}_k = \frac{[\lambda_{k,1}, ..., \lambda_{k,M_k}]}{A_k}. \tag{2.24}$$

The image structure in the first model, or the sets of image structures in the second, can be different or the same. These constraints will be explored later in the chapter.

Previous 2D analytic signals assume that the local image structure can be completely represented by the model. However, this limits both the number of RTs and the types of structures that can be represented. Adding the residual component removes this constraint, by representing the part of the structure that is not well-modelled. This allows us to use arbitrary features as the model components. Furthermore, we can compare the energy of the residual with that of the model to give an illumination invariant measure of how well the local structure is modelled. This will be demonstrated in later chapters.

2.2.2 General Solution

Solving the general model is possible using the CH vector. Typically this involves correlating the local image patch with wavelets designed to match the components of the model. However, in the proposed approach the CH vector describing the local image patch is split into vectors describing the model components plus a residual vector. The residual vector is then minimised to solve for the model parameters.

A 2D steerable wavelet, ψ_{u_m}, can be represented by the normalised weighted CH vector $\mathbf{W}\mathbf{u}_m$ where

$$\psi_{u_m} = \sum_{|n|\leq N} u_{m,n} \mathcal{R}\psi, \tag{2.25}$$

and

$$\mathbf{W}\mathbf{u}_m = \mathbf{W}\mathbf{f} / \|\mathbf{W}\mathbf{f}\|, \tag{2.26}$$

where $\mathbf{f}$ is the CH vector at the centre of the feature. For example, for a Y junction we would calculate the weighted CH vector at the position where the three line segments meet then normalise it.

Consider the set of K 2D steerable wavelets, $\{\psi^{(k)}\}_{k\in\mathbb{N}_K}$, that match the particular set of image structures we are interested in. These shall be referred to as *model* wavelets. Let $\{\mathbf{W}\mathbf{u}_k\}_{k\in\mathbb{N}_K}$ be the corresponding set of weighted and normalised CH vectors. For the first general model, the local image structure CH vector, $\mathbf{Wf}$, can be written as the sum of individually scaled and rotated versions of each model wavelet CH vector, plus the residual component, $\mathbf{W}\boldsymbol{\epsilon}$. The relationship for a particular scale i and location $\mathbf{k}$ is

$$f(\mathbf{z}) = \sum_{k=1}^{K} \lambda_k u_k(\mathbf{R}_{\theta_k}\mathbf{z}) + f_\epsilon(\mathbf{z}) \quad \xleftrightarrow{CH} \quad \mathbf{Wf} = \sum_k \lambda_k \mathbf{S}_{\theta_k}\mathbf{W}\mathbf{u}_k + \mathbf{W}\boldsymbol{\epsilon}. \tag{2.27}$$

Since we can reconstruct the image exactly from the CH vector, the image can be expressed as the sum of separate model and residual reconstructions,

$$f(\mathbf{z}) = f_\psi(\mathbf{z}) + f_\epsilon(\mathbf{z}), \tag{2.28}$$

where $f_\psi(\mathbf{z})$ is the part synthesised from the model wavelets,

$$f_\psi(\mathbf{z}) = \sum_{i,\mathbf{k}} \sum_{k=1}^{K} \sum_{|n|\leq N} \lambda_{k_{i,\mathbf{k}}} (\mathbf{S}_{\theta_{i,\mathbf{k}}}\mathbf{W}\mathbf{u}_k)_n, w_n \psi^n_{i,\mathbf{k}}, \tag{2.29}$$

and $f_\epsilon(\mathbf{z})$ is the residual image synthesised from the residual wavelets,

$$f_\epsilon(\mathbf{z}) = \sum_{i,\mathbf{k}} \sum_{k=1}^{K} \sum_{|n|\leq N} (\mathbf{W}\boldsymbol{\epsilon}_{i,\mathbf{k}})_n w_n \psi^n_{i,\mathbf{k}}. \tag{2.30}$$

The residual component is the missing part of the local structure that is not correlated with the wavelets but is needed for exact reconstruction of the image.

To solve the model, we shall choose values of λ_k and θ_k that minimise the residual component at each scale and location, so that the model wavelets explain as much of the image as possible. Since the CH wavelets $\{\psi^n\}_{|n|\in\mathbb{N}_N}$ are orthogonal, the L_2-norm of the residual wavelet $\psi_{\epsilon_{i,\mathbf{k}}}$ is proportional to the ℓ_2-norm of the residual vector $\mathbf{W}\boldsymbol{\epsilon}_{i,\mathbf{k}}$ [124]. Therefore

$$\|\psi_\epsilon\| \propto \|\mathbf{W}\boldsymbol{\epsilon}\| \tag{2.31}$$

$$\propto \left\| \mathbf{Wf} - \sum_{k=1}^{K} \lambda_k \mathbf{S}_{\theta_k}\mathbf{W}\mathbf{u}_k \right\|. \tag{2.32}$$

Letting $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_K]$ and $\boldsymbol{\theta} = [\theta_1, ..., \theta_K]$, the linear scale and rotation parameters that minimise the residual are given by

$$\boldsymbol{\lambda}, \boldsymbol{\theta} = \arg\min_{\boldsymbol{\lambda},\boldsymbol{\theta}} \left\| \mathbf{Wf} - \sum_{k=1}^{K} \lambda_k \mathbf{S}_{\theta_k}\mathbf{W}\mathbf{u}_k \right\|. \tag{2.33}$$

Once this equation has been solved we may compare the model and residual vectors to give a measure of how well the model represents the local image structure,

$$\gamma = \tan^{-1} \frac{\|\mathbf{W}\boldsymbol{\epsilon}\|}{\|\sum_k \lambda_k \mathbf{S}_{\theta_k} \mathbf{W}\mathbf{u}_k\|} \tag{2.34}$$

$$= \sin^{-1} \frac{\|\mathbf{W}\boldsymbol{\epsilon}\|}{\|\mathbf{W}\mathbf{f}\||}, \tag{2.35}$$

where $\gamma = 0$ means the structure CH vector is completely represented by the model CH vectors. The second general model consisting of sets of features will be considered later in the chapter.

2.3 Least-Squares With Global Orientation

Finding an exact solution for the residual minimisation equation is difficult due to the non-linearity from having different orientation parameters. By introducing some constraints, particularly by having a single orientation values, we can simplify the equations so that an exact solution can be found. In this section we shall explore the solutions for different constraints, starting with the general solution for wavelets with a single orientation and finishing with multiple sets. For simplicity and without loss of generality, we drop the weighting matrix $\mathbf{W}$, that is $\mathbf{Wf} \to \mathbf{f}$, so that the equations are easier to read.

2.3.1 Single Orientation Constraint

The first constraints we shall introduce are:

- The set of model wavelet CH vectors is linearly independent. This sets an upper level on the number of model wavelets that can be used to $2N + 1$.
- The wavelets are all rotated to the same orientation, that is $\theta_p = \theta_q \quad \forall p, q \in \mathbb{N}_M$.

This equates to a single set of M wavelets ($K = 1$) with a single orientation θ and different amplitudes. The model equation reduces to

$$f(\mathbf{z}) = \sum_{m=1}^{M} \lambda_m u_m(\mathbf{R}_\theta \mathbf{z}) + f_\epsilon(\mathbf{z}) \quad \overset{CH}{\longleftrightarrow} \quad \mathbf{Wf} = \sum_m \lambda_m \mathbf{S}_\theta \mathbf{W}\mathbf{u}_m + \mathbf{W}\boldsymbol{\epsilon}, \tag{2.36}$$

and thus the equation to solve is

$$\boldsymbol{\lambda}, \theta = \arg\min_{\boldsymbol{\lambda}, \theta} \left\| \mathbf{Wf} - \sum_m \lambda_m \mathbf{S}_\theta \mathbf{W}\mathbf{u}_m \right\|. \tag{2.37}$$

To solve this we begin by collecting all the normalised model CH vectors into the columns of a $2N + 1 \times M$ matrix $\mathbf{U}$, that is

$$\mathbf{U} = [\mathbf{u}_1, ..., \mathbf{u}_M], \tag{2.38}$$

and rewrite the model equation to solve as

$$\boldsymbol{\lambda}, \theta = \arg\min_{\boldsymbol{\lambda},\theta} \|\mathbf{f} - \mathbf{S}_\theta \mathbf{U}\boldsymbol{\lambda}\| \tag{2.39}$$

$$= \arg\min_{\boldsymbol{\lambda},\theta} \|\mathbf{S}_{-\theta}\mathbf{f} - \mathbf{U}\boldsymbol{\lambda}\|. \tag{2.40}$$

For a fixed value of θ, this becomes the classic linear least-squares problem. Using the properties $\|\mathbf{a}\|^2 = \langle \mathbf{a}, \mathbf{a}\rangle = \mathbf{a}^H\mathbf{a}$ and $(\mathbf{AB})^H = \mathbf{B}^H\mathbf{A}^H$, and since $\mathbf{S}_\theta\mathbf{S}_\theta^H = \mathbf{I}_N$ and $\mathbf{S}_{-\theta} = \mathbf{S}_\theta^H$, we have

$$\min_{\boldsymbol{\lambda},\theta} \|\mathbf{S}_{-\theta}\mathbf{f} - \mathbf{U}\boldsymbol{\lambda}\| = \min_{\boldsymbol{\lambda},\theta} \mathbf{f}^H\mathbf{S}_\theta\mathbf{S}_\theta^H\mathbf{f} - 2\boldsymbol{\lambda}^H\mathbf{U}^H\mathbf{S}_\theta^H\mathbf{f} + \boldsymbol{\lambda}^H\mathbf{U}^H\mathbf{U}\boldsymbol{\lambda} \tag{2.41}$$

$$= \min_{\boldsymbol{\lambda},\theta} \; -2\boldsymbol{\lambda}^H\mathbf{U}^H\mathbf{S}_\theta^H\mathbf{f} + \boldsymbol{\lambda}^H\mathbf{U}^H\mathbf{U}\boldsymbol{\lambda}. \tag{2.42}$$

At the minimum, the derivative with respect to $\boldsymbol{\lambda}$ will be 0, which gives

$$0 = -2\mathbf{U}^H\mathbf{S}_\theta^H\mathbf{f} + 2\mathbf{U}^H\mathbf{U}\boldsymbol{\lambda}, \tag{2.43}$$

$$\boldsymbol{\lambda} = (\mathbf{U}^H\mathbf{U})^{-1}\mathbf{U}^H\mathbf{S}_\theta^H\mathbf{f} \tag{2.44}$$

$$= \mathbf{U}^+\mathbf{S}_\theta^H\mathbf{f}, \tag{2.45}$$

where $\mathbf{U}^+$ is called the pseudo-inverse of $\mathbf{U}$. To solve for θ we substitute back into (2.42) to get

$$\min_\theta \|\mathbf{S}_{-\theta}\mathbf{f} - \mathbf{U}\boldsymbol{\lambda}_\theta\| = \min_\theta \; -2\boldsymbol{\lambda}_\theta^H\mathbf{U}^H\mathbf{S}_\theta^H\mathbf{f} + \boldsymbol{\lambda}_\theta^H\mathbf{U}^H\mathbf{U}\boldsymbol{\lambda}_\theta \tag{2.46}$$

$$= \min_\theta \; -2\mathbf{f}^H\mathbf{S}_\theta(\mathbf{U}^+)^H\mathbf{U}^H\mathbf{S}_\theta^H\mathbf{f} + \mathbf{f}^H\mathbf{S}_\theta(\mathbf{U}^+)^H\mathbf{U}^H\mathbf{U}\mathbf{U}^+\mathbf{S}_\theta^H\mathbf{f}. \tag{2.47}$$

Since $\mathbf{U}\mathbf{U}^+$ is Hermitian, $(\mathbf{U}^+)^H\mathbf{U}^H = (\mathbf{U}\mathbf{U}^+)^H = \mathbf{U}\mathbf{U}^+$ and $\mathbf{U}^+\mathbf{U} = \mathbf{I}_M$. The above equation simplifies as follows

$$\min_\theta \|\mathbf{S}_{-\theta}\mathbf{f} - \mathbf{U}\boldsymbol{\lambda}_\theta\| = \min_\theta -2\mathbf{f}^H\mathbf{S}_\theta\mathbf{U}\mathbf{U}^+\mathbf{S}_\theta^H\mathbf{f} + \mathbf{f}^H\mathbf{S}_\theta\mathbf{U}\mathbf{U}^+\mathbf{U}\mathbf{U}^+\mathbf{S}_\theta^H\mathbf{f} \tag{2.48}$$

$$= \min_\theta -2\mathbf{f}^H\mathbf{S}_\theta\mathbf{U}\mathbf{U}^+\mathbf{S}_\theta^H\mathbf{f} + \mathbf{f}^H\mathbf{S}_\theta\mathbf{U}\mathbf{U}^+\mathbf{S}_\theta^H\mathbf{f} \tag{2.49}$$

$$= \max_\theta \mathbf{f}^H\mathbf{S}_\theta\mathbf{U}\mathbf{U}^+\mathbf{S}_\theta^H\mathbf{f} \tag{2.50}$$

$$= \max_\theta \langle \mathbf{U}^H\mathbf{S}_\theta^H\mathbf{f}, \mathbf{U}^+\mathbf{S}_\theta^H\mathbf{f}\rangle. \tag{2.51}$$

Since each of the terms in the inner product are a function of θ we let

$$\boldsymbol{\lambda}(\theta) = \mathbf{U}^+\mathbf{S}_\theta^H\mathbf{f}, \tag{2.52}$$

$$\boldsymbol{\delta}(\theta) = \mathbf{U}^H\mathbf{S}_\theta^H\mathbf{f}, \tag{2.53}$$

which gives

$$\max_\theta \langle \mathbf{U}^H\mathbf{S}_\theta^H\mathbf{f}, \mathbf{U}^+\mathbf{S}_\theta^H\mathbf{f}\rangle = \max_\theta \sum_m \delta_m(\theta)\lambda_m(\theta). \tag{2.54}$$

Each of these functions are trigonometric polynomials with order $2N$, that is

$$\lambda_m(\theta) = \sum_{|n|\leq N} u^+_{m_n} w_n^2 f_n e^{-\mathrm{i}n\theta}, \tag{2.55}$$

$$\delta_m(\theta) = \sum_{|n|\leq N} \bar{u}_{m_n} w_n^2 f_n e^{-\mathrm{i}n\theta}. \tag{2.56}$$

We must thus find the maximum of the trigonometric polynomial

$$p(\theta) = \sum_m \delta_m(\theta)\lambda_m(\theta), \tag{2.57}$$

which has degree $4N$. One method to find the maximum is by finding locations where the root is zero. The derivative of the polynomial is

$$p'(\theta) = \frac{d}{d(e^{\mathrm{i}\theta})} p(\theta) \tag{2.58}$$

$$= \sum_{|n|\leq N} -\mathrm{i}np_n e^{-\mathrm{i}n\theta}, \tag{2.59}$$

where p_n are the coefficients of the polynomial. Using root finding we obtain up to $4N$ candidate values of θ,

$$\boldsymbol{\theta} = \{\theta \,:\, p'(\theta) = 0, \theta \in [0, 2\pi)\}. \tag{2.60}$$

Substituting each θ back into $p(\theta)$ we choose the value that gives the maximum

$$\theta_{\max} = \arg\max_{\theta\in\boldsymbol{\theta}} p(\theta), \tag{2.61}$$

from which $\lambda_m(\theta_{\max})$ and $\delta_m(\theta_{\max})$ can be calculated for each wavelet to give the amplitude of each model component.

Orthonormal Wavelets

If instead the wavelets are orthonormal, then $\mathbf{U}^H\mathbf{U} = \mathbf{I}_{2N+1}$ and thus the pseudo-inverse is simply $\mathbf{U}^+ = \mathbf{U}^H$. Therefore $\lambda_m(\theta) = \delta_m(\theta)$ and the trigonometric polynomial to find the maximum of becomes

$$p(\theta) = \sum_m \lambda_m^2(\theta). \tag{2.62}$$

Each of the image models explored in the later chapters are orthonormal, and this reduction will apply.

Range

Because of the multiplication of λ and δ in (2.57), it is possible to have both positive and negative values of λ. A negative value means that the wavelet, and thus the feature component, has been inverted. This can be useful in many situations. For example, consider a wavelet that is matched to a positive line feature. A large negative value for λ indicates a strong correlation with the inverse, a negative valued line feature.

However, in other situations only positive values of λ would be desired, such as if the line features were all positive. One would then solve

$$\theta = \arg\max_{\theta \in \boldsymbol{\theta}} p(\theta) \quad \text{s.t. } \lambda_m(\theta) >= 0. \tag{2.63}$$

For a single component to the model ($M = 1$) the positive constraint is equivalent to solving for the maximum of

$$p(\theta) = \lambda_1(\theta), \tag{2.64}$$

and the solution is easily obtained. However for more than one component, the solution is more complicated. The non-linearity of the condition means that the best solution does not necessarily correspond to one of the maxima of $p(\theta)$. Fortunately a model consisting of *multiple* components with the same polarity at a *single* orientation is not needed to analyse lines, edges, corners or junctions. Note, a model consisting of multiple components with the same polarity and each with *different* orientations is solved in a different manner later in the chapter. Note, that for a single component with either positive or negative amplitude one does not need to solve for the maximum of $\lambda_1(\theta)^2$. Instead one can find the roots of $\lambda_1(\theta)$ and pick the result that maximises $|\lambda_1(\theta)|$. This halves the degree of the polynomial and thus reduces computation time.

Example

The approach is demonstrated on the *Board* image (Figure 2.1). A set of four wavelets were chosen to model the image (Figure 2.4) by their additive combination at a *single common orientation.* The wavelets roughly correspond to line, corner, T junction and X junction features, which are present in the test image. Their construction is detailed in later chapters.

The shapes of the pseudo-inverse wavelets, corresponding to CH vectors of the columns of $\mathbf{U}^{+H}$ (Figure 2.4), are quite different to their corresponding feature wavelets. Since large values for both $\delta_m(\theta)$, corresponding to the model wavelets, and $\lambda_m(\theta)$, corresponding to the pseudo-inverse wavelets, will give a higher value for $p(\theta)$, we can infer that the pseudo-inverse wavelets help discriminate the model wavelets from each other. For example, the pseudo-inverse wavelet for the X junction ($m = 4$) consists of only the left segment, a component that all the other model wavelets lack.

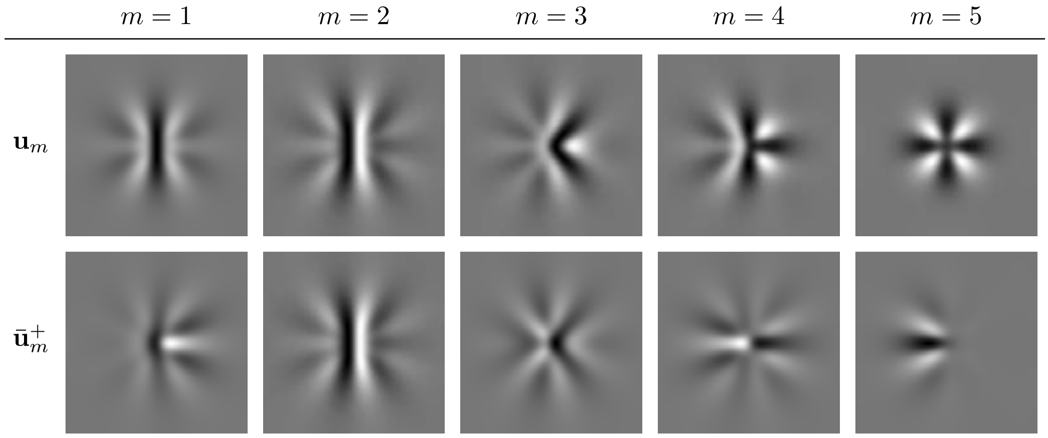


Figure 2.4: Set of model wavelets (top row) and the corresponding wavelets from the columns of the pseudo-inverse of $\mathbf{U}^{+H}$ (bottom row), for $N = 7$.

The resulting values for the amplitude of each model component, δ_m (Figure 2.5), show there is a high response to the line-like wavelet at lines, the corner-like wavelet at corners, and the X junction wavelet at the line intersections. Curiously, while the T junction wavelets respond to the lines, the response is not high at the actual T junction locations. Indeed, the pseudo-inverse wavelet for the T junction has a large *negative* response at the locations of T junctions, meaning that this feature is actually being subtracted in the model formulation. Note, the polarity of the edge response, δ_2, flips between positive and negative values in some locations because a positive edge at 0 degrees is the same as a negative edge at 180 degrees. Thus where the local signal structure is edge-like both solutions are possible.

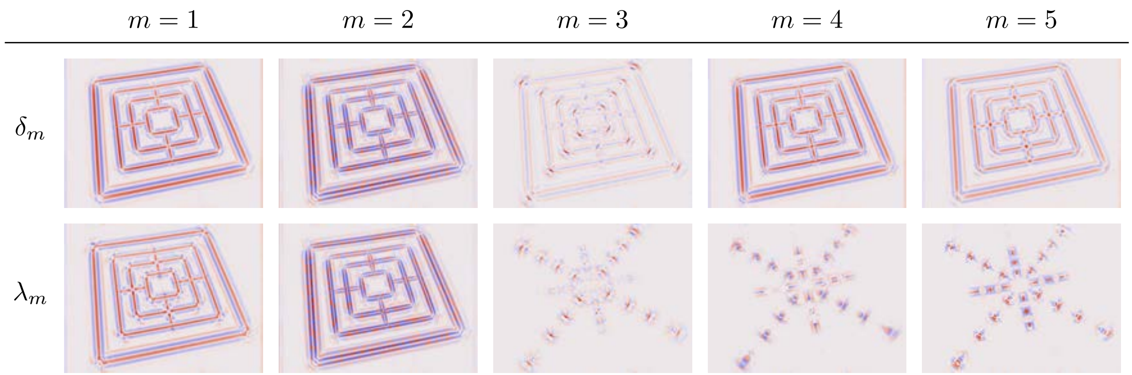
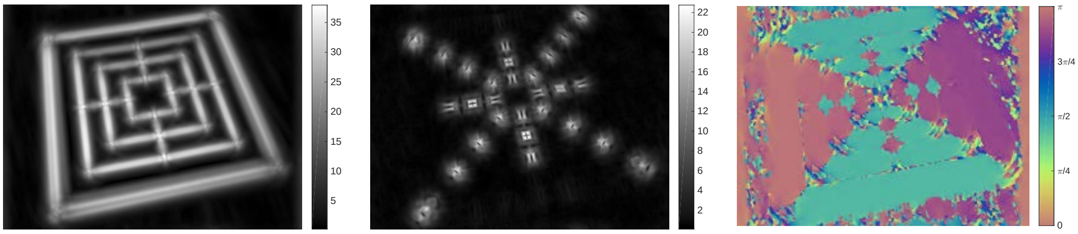


Figure 2.5: Test image response to the model wavelets from Figure 2.4 (top row) and their pseudo-inverse (bottom) from Figure 2.4. Red: positive response, blue: negative response.

The model norm is large at the location of features that are similar to the model wavelets (Figure 2.6a), while the residual norm is large where the features are not similar (Figure 2.6b). The model orientation gives an estimate of the feature orientation (Figure 2.6c); it is shown modulo π due to the aforementioned flipping of the edge orientation.



(a) Model norm (b) Residual norm (c) Orientation mod π

Figure 2.6: Model and residual vector magnitudes along with the global orientation estimate for the wavelet set in Figure 2.4.

Investigating the negative response to the T junction further, at the centre of the left-side T junction, the model CH vector is given by

$$34.1\mathbf{u}_1 + 1.6\mathbf{u}_2 + 1.1\mathbf{u}_3 - 24.1\mathbf{u}_4 + 26.0\mathbf{u}_5, \tag{2.65}$$

with orientation $\theta = 3$ radians clockwise. One would expect the T junction to be modelled by the T junction wavelet with a slight counter-clockwise rotation corresponding to $\theta = -0.15$ radians. If we construct the wavelet corresponding to the model CH vector it looks very much like a T junction. However, it is made up of a line, X junction and a flipped and inverted T junction, as shown in Figure 2.7:

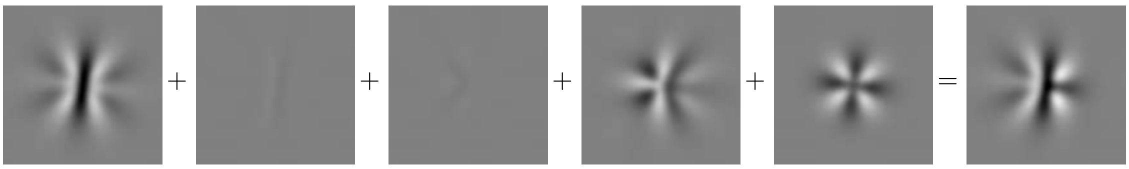


Figure 2.7: Model components of the left T junction solved using the wavelet set in Figure 2.4.

Therefore while the method may give the absolute minimum of the residual norm at this location, large λ values for the each wavelet suggest the model is being over-fit. Indeed from a image interpretation point-of-view, we would prefer a single large value for the T junction wavelet, rather multiple combinations of wavelets with a slightly better fit.

2.3.2 Regularisation

A common approach to prevent over-fitting is by introducing a penalty term on the magnitude of the coefficient vector, in our case $\boldsymbol{\lambda}$. The model becomes

$$\boldsymbol{\lambda}, \theta = \underset{\boldsymbol{\lambda}, \theta}{\arg\min} \left(\|\mathbf{f} - \mathbf{S}_\theta \mathbf{U} \boldsymbol{\lambda}\|_2 + \alpha \|\boldsymbol{\lambda}\|_p \right), \tag{2.66}$$

where p indicates the type of norm used. Three types are:

- $p = 0$: used in Akaike information criterion (AIC) [2]. It is equal to the number of non-zero elements of $\boldsymbol{\lambda}$.

- $p = 1$: used in methods such as LASSO [120]. It tends to give sparser values for $\boldsymbol{\lambda}$, meaning that there are fewer non-zero components.
- $p = 2$: known as Tikhonov regression or ridge regression. It tends to reduce the values of all parameters, rather than select for only a few, as a vector with many small coefficients tends to have a smaller ℓ_2 norm than one with a single large coefficient.

We shall investigate ℓ_2 and a ℓ_1-like regularisation for the model estimation problem.

ℓ_2 regularisation

Using ℓ_2 regularisation the problem becomes

$$\boldsymbol{\lambda}, \theta = \underset{\boldsymbol{\lambda},\theta}{\arg\min} \left(\|\mathbf{f} - \mathbf{S}_\theta \mathbf{U}\boldsymbol{\lambda}\|_2^2 + \alpha \|\boldsymbol{\lambda}\|_2^2 \right) \tag{2.67}$$

$$= \underset{\boldsymbol{\lambda},\theta}{\arg\min} \; -2\boldsymbol{\lambda}^H \mathbf{U}^H \mathbf{S}_\theta^H \mathbf{f} + \boldsymbol{\lambda}^H \mathbf{U}^H \mathbf{U}\boldsymbol{\lambda} + \alpha \boldsymbol{\lambda}^H \boldsymbol{\lambda}. \tag{2.68}$$

Following the same procedure as before, at the minimum the derivative with respect to $\boldsymbol{\lambda}$ will be 0, which gives

$$0 = -2\mathbf{U}^H \mathbf{S}_\theta^H \mathbf{f} + 2\mathbf{U}^H \mathbf{U}\boldsymbol{\lambda} + \alpha\boldsymbol{\lambda}, \tag{2.69}$$

$$\boldsymbol{\lambda} = (\mathbf{U}^H \mathbf{U} + \alpha \mathbf{I})^{-1} \mathbf{U}^H \mathbf{S}_\theta^H \mathbf{f}. \tag{2.70}$$

Letting $\mathbf{U}_\alpha^+ = (\mathbf{U}^H \mathbf{U} + \alpha \mathbf{I})^{-1} \mathbf{U}^H$, for a fixed orientation θ we have

$$\min_\theta \; -2\boldsymbol{\lambda}_\theta^H \mathbf{U}^H \mathbf{S}_\theta^H \mathbf{f} + \boldsymbol{\lambda}_\theta^H \mathbf{U}^H \mathbf{U}\boldsymbol{\lambda}_\theta + \alpha \boldsymbol{\lambda}_\theta^H \boldsymbol{\lambda}_\theta \tag{2.71}$$

$$= \min_\theta \; -2\mathbf{f}^H \mathbf{S}_\theta (\mathbf{U}_\alpha^+)^H \mathbf{U}^H \mathbf{S}_\theta^H \mathbf{f} + \mathbf{f}^H \mathbf{S}_\theta (\mathbf{U}_\alpha^+)^H \mathbf{U}^H \mathbf{U}\mathbf{U}^+ \mathbf{S}_\theta^H \mathbf{f} + \alpha \mathbf{f}^H \mathbf{S}_\theta (\mathbf{U}_\alpha^+)^H \mathbf{U}_\alpha^+ \mathbf{S}_\theta^H \mathbf{f}. \tag{2.72}$$

Since $\mathbf{U}\mathbf{U}^+$ is Hermitian, $(\mathbf{U}^+)^H \mathbf{U}^H = (\mathbf{U}\mathbf{U}^+)^H = \mathbf{U}\mathbf{U}^+$ and $\mathbf{U}^+ \mathbf{U} = \mathbf{I}_M$. The above equation simplifies to

$$\max_\theta \; -\mathbf{f}^H \mathbf{S}_\theta (\mathbf{U}_\alpha^+)^H \mathbf{U}^H \mathbf{S}_\theta^H \mathbf{f} - \alpha \mathbf{f}^H \mathbf{S}_\theta (\mathbf{U}_\alpha^+)^H \mathbf{U}_\alpha^+ \mathbf{S}_\theta^H \mathbf{f} \tag{2.73}$$

$$= \max_\theta \langle \mathbf{U}^H \mathbf{S}_\theta^H \mathbf{f}, \mathbf{U}_\alpha^+ \mathbf{S}_\theta^H \mathbf{f} \rangle - \langle \mathbf{U}_\alpha^+ \mathbf{S}_\theta^H \mathbf{f}, \mathbf{U}_\alpha^+ \mathbf{S}_\theta^H \mathbf{f} \rangle. \tag{2.74}$$

Since each of the terms in the inner product are a function of θ we let

$$\boldsymbol{\lambda}(\theta) = \mathbf{U}_\alpha^+ \mathbf{S}_\theta^H \mathbf{f}, \tag{2.75}$$

$$\boldsymbol{\delta}(\theta) = \mathbf{U}^H \mathbf{S}_\theta^H \mathbf{f}, \tag{2.76}$$

and as before, the polynomial to maximise to solve for θ is thus

$$p(\theta) = \max_{\theta} \sum_{m} \delta_m(\theta)\lambda_m(\theta). \tag{2.77}$$

The response to the *Board* image using the same set of wavelets as before, but this time using ℓ_2 regularisation with $\alpha = 0.1$ was calculated (Figure 2.8). In contrast to using no regularisation, the T junction wavelet ($m = 4$) now has a positive response at the locations of the T junctions.

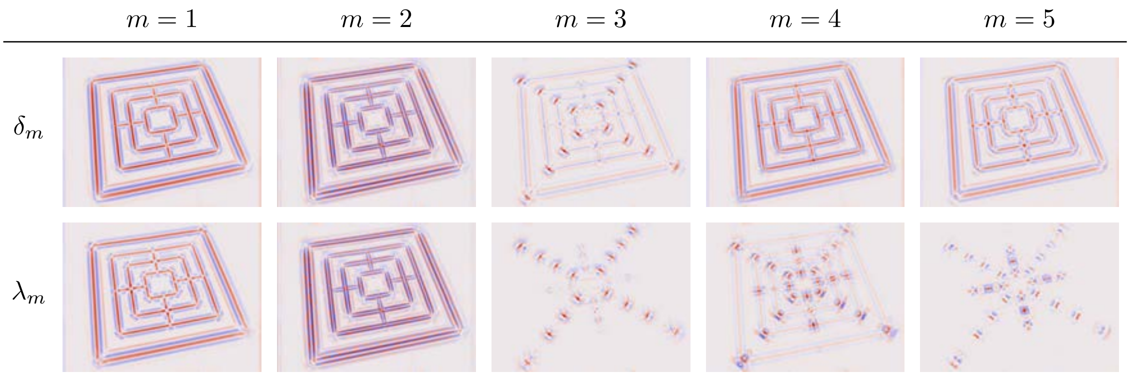
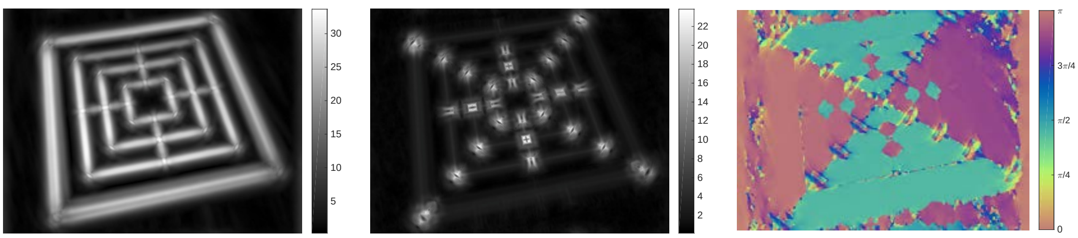


Figure 2.8: Test image response to the model wavelets from Figure 2.4 (top row) and the wavelets from the columns of the pseudo-inverse, $\mathbf{U}_{\alpha}^{+H}$ (bottom row), using ℓ_2 regularisation with $\alpha = 0.1$. Red: positive response, blue: negative response.

The model norm (Figure 2.9a), residual norm (Figure 2.9b) and orientation estimate (Figure 2.9c) appear unchanged.



(a) Model norm (b) Residual norm (c) Orientation mod π

Figure 2.9: Model and residual vector magnitudes along with the global orientation estimate for the wavelet set in Figure 2.4 using ℓ_2 regularisation with $\alpha = 0.1$.

Investigation the left-side T junction shows the model CH vector is given now by

$$13.52\mathbf{u}_1 - 1.4\mathbf{u}_2 + 0.8\mathbf{u}_3 + 18.9\mathbf{u}_4 + 3.8\mathbf{u}_5, \tag{2.78}$$

with orientation $\theta = 3$ radians. The T-junction response is now mainly comprised of the line wavelet and the T-junction wavelet (Figure 2.10). The earlier problem where the response was made up of line, X junction and a flipped and inverted T junction wavelets has been resolved by using regularisation.

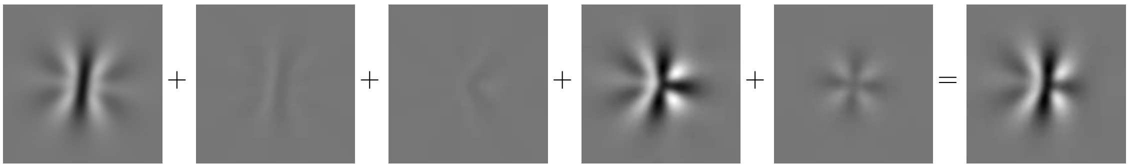


Figure 2.10: Model components of the left T junction solved using the wavelet set in Figure 2.4 at ℓ_2 regularisation with $\alpha = 0.1$.

ℓ_1 *penalty*

Using ℓ_1 regularisation the problem becomes

$$\boldsymbol{\lambda}, \theta = \underset{\boldsymbol{\lambda}, \theta}{\arg\min} \left(\|\mathbf{f} - \mathbf{S}_\theta \mathbf{U} \boldsymbol{\lambda}\|_2^2 + \alpha \|\boldsymbol{\lambda}\|_1 \right). \tag{2.79}$$

However, due to the ℓ_1 norm this is no longer a linear system of equations. One can use ℓ_1 solving methods such as *basis pursuit denoising* [18] to find the values of λ for a fixed orientation θ, however this does not extend to all orientations.

Instead we shall use the ℓ_1 norm to choose between multiple similar solutions to the non-regularised problem. The process is the same as solving the unregularised problem

$$\boldsymbol{\lambda}, \theta = \underset{\boldsymbol{\lambda}, \theta}{\arg\min} \|\mathbf{f} - \mathbf{S}_\theta \mathbf{U} \boldsymbol{\lambda}\|_2^2, \tag{2.80}$$

except when we obtain a set of candidate points for θ

$$\boldsymbol{\theta} = \{\theta \,:\, p'(\theta) = 0, \theta \in [0, 2\pi)\}, \tag{2.81}$$

we choose the point using the ℓ_1 norm as follows

$$\theta = \underset{\theta \in \boldsymbol{\theta}}{\arg\max} \left(p(\theta) - \alpha \|\boldsymbol{\lambda}(\theta)\|_1 \right). \tag{2.82}$$

This approach is not true ℓ_1 regularisation, it simply weights the solutions according to the sum of the components.

The response to the *Board* image using the same set of wavelets as before, but this time using the ℓ_1 penalty with $\alpha = 2$ (Figure 2.11) shows the ℓ_1 penalty also resolves the over-fitting problem, as the T junction wavelet ($m = 4$) has a positive response at the locations of the T junctions. The response to the corner, T, and X wavelets (Figure 2.11 bottom row) is lower around the line areas compared to the ℓ_2 norm, resulting in more a more localised response for the T junction in particular. Using the ℓ_1 penalty thus resolves the over-fitting problem while improving the semantic description of the junctions.

The model norm (Figure 2.12a) and residual norm (Figure 2.12b) appear unchanged from the ℓ_2 case. However, the orientation estimate (Figure 2.12c) has noisy regions in the low strength magnitude areas, due to the penalty term being larger than the polynomial value at that point.

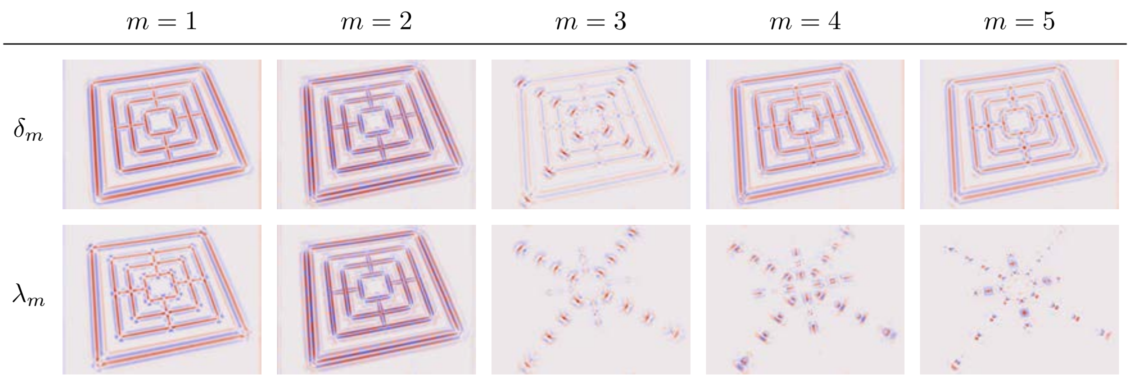
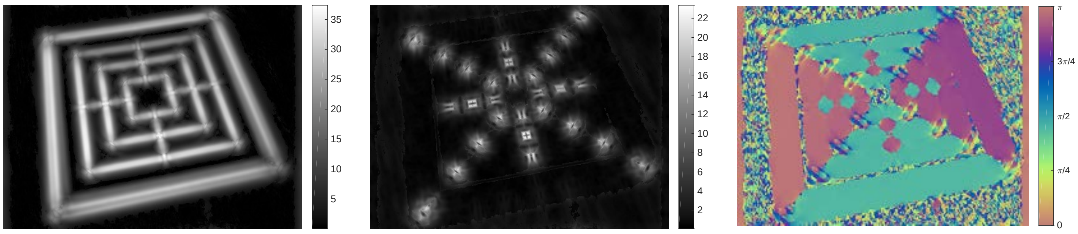


Figure 2.11: Test image response to the model wavelets from Figure 2.4 (top row) and the wavelets from the columns of the pseudo-inverse, $\mathbf{U}_\alpha^{+H}$ (bottom row), using an ℓ_1 penalty term with $\alpha = 2$.



(a) Model norm (b) Residual norm (c) Orientation mod π

Figure 2.12: Model and residual vector magnitudes along with the global orientation estimate for the wavelet set in Figure 2.4 using an ℓ_1 penalty $\alpha = 2$.

At the centre of the left-side T-junction, the model CH vector using the ℓ_1 penalty is given by

$$13.2\mathbf{u}_1 - 1.6\mathbf{u}_2 + 1.2\mathbf{u}_3 + 23.5\mathbf{u}_4 + 0.9\mathbf{u}_5, \tag{2.83}$$

with orientation $\theta = 3$ radians. The T junction response is mainly comprised of the line wavelet and the T-junction wavelet (Figure 2.13). The earlier problem where the response was made up of line, X junction and a flipped and inverted T junction wavelets has been also been resolved by using the ℓ_1 penalty.

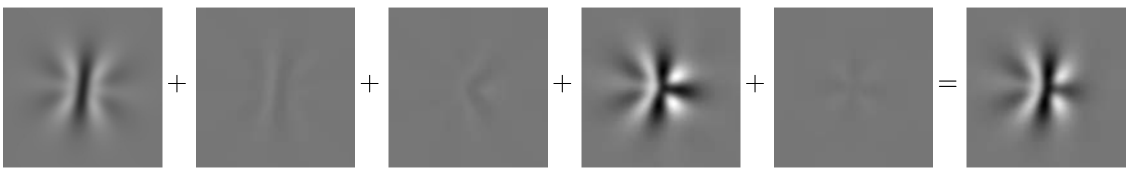


Figure 2.13: Model components of the left T junction solved using the wavelet set in Figure 2.4 at ℓ_2 regularisation with $\alpha = 2$.

2.3.3 Summary

This section has developed the tools needed to solve for the parameters a set of wavelets with common orientation, by choosing values that give the lowest residual. Using regularisation relaxes the problem so that solutions with a fewer number of high amplitude components can be found.

The examples given used a set of five wavelets, and the ratio of their amplitudes can be considered a multi-dimensional phase descriptor. In the following chapters, sets consisting of only two wavelets with common orientation will be used, where the phase value is a scalar given by their ratio.

2.4 Maximal Response

2.4.1 Single Wavelet, Single Orientation

In the last section we saw how penalising the large coefficient vectors gave feature strengths that were sparser. Taking this to the extreme, we can represent the local image structure by choosing from a set the single wavelet that has the maximum correlation with the structure. For example, if the local signal structure resembles a line, only the line-wavelet will have a non-zero coefficient. The optimisation problem becomes

$$\boldsymbol{\lambda}, \boldsymbol{\theta} = \underset{\boldsymbol{\lambda}, \boldsymbol{\theta}}{\arg\min} \|\mathbf{Wf} - \lambda_k \mathbf{S}_{\theta_k} \mathbf{W} \mathbf{u}_k\| \quad \text{such that } \|\boldsymbol{\lambda}\|_0 = 1. \tag{2.84}$$

As there will be only one non-zero coefficient for both $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$ we may instead write this as

$$\lambda, \theta, k = \underset{\lambda, \theta, k \in \mathbb{N}_K}{\arg\min} \|\mathbf{Wf} - \lambda \mathbf{S}_\theta \mathbf{W} \mathbf{u}_k\|, \tag{2.85}$$

where k is the wavelet type index. Solving this problem is easy, we find λ and θ for each wavelet individually and then choose the k for which λ is a maximum. Each wavelet from Figure 2.4 responds maximally to different features in the test image (Figure 2.14).

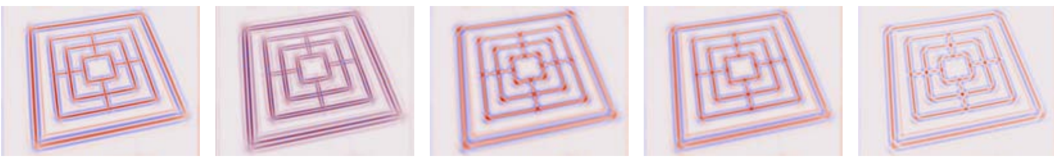
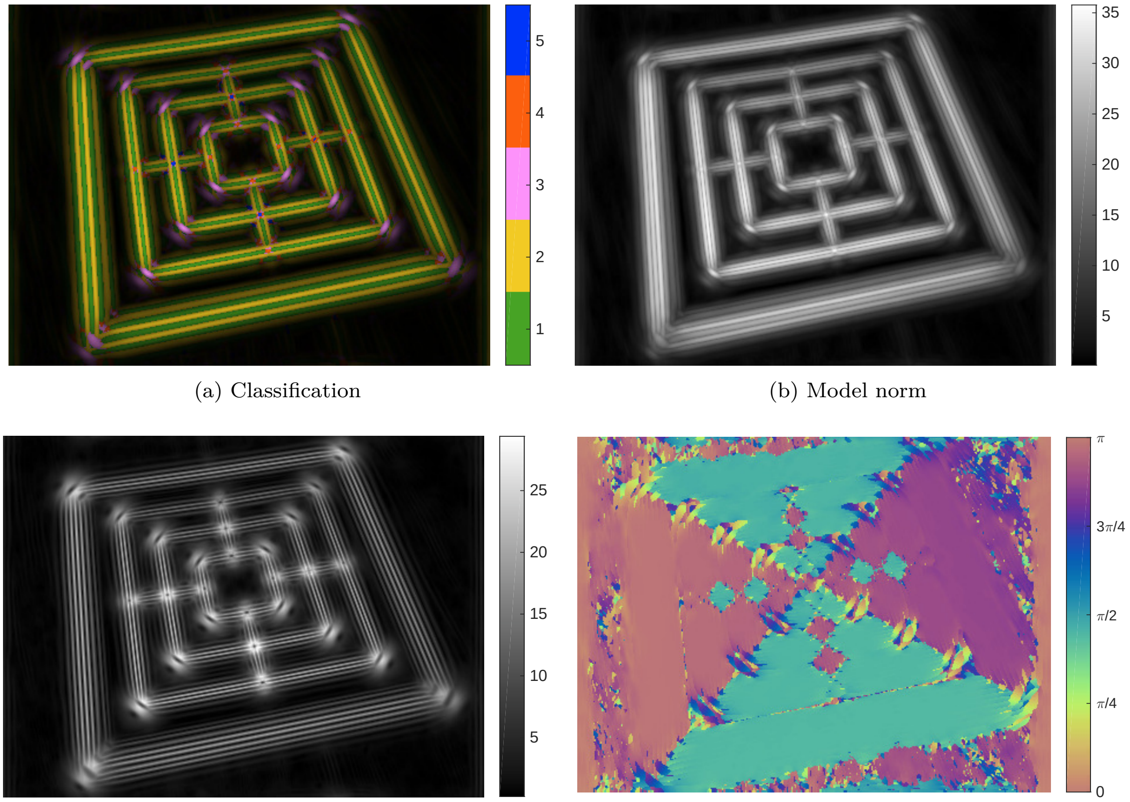
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
λ_k					

Figure 2.14: Test image response to the model wavelets from Figure 2.4 individually.

At each location of the test image the pixel is classified according to which wavelet gives the maximum response (Figure 2.15a). The class k clearly differentiates each of the feature types, and the absolute value of λ is high at the location of these features. Interpretation of the orientation must be made in the context of the feature class. The wavelets have been purposely aligned along their main axis of symmetry so that the orientation values can be compared. This is why there is a smooth transition from the line orientation to the edge orientation and so on. Design of a wavelet set should follow this principle in general.

The residual norm is large adjacent to the features as previously observed, but also at locations in between line and edge features. We can infer that locations that are half-line / half-edge are not well represented by either type and because only *one* wavelet is allowed in this model the residual will be high. In contrast, the approach of the previous section uses a model consisting of more than one wavelet and can represent combinations of structures, therefore the residual is reduced.



(a) Classification (b) Model norm

(c) Residual norm (d) Orientation mod π

Figure 2.15: Test image classification, model and residual vector norm along with the global orientation estimate for the maximal response to the set of CH wavelets in Figure 2.4.

2.4.2 Single Wavelet Set, Single Orientation

To reduce the residual we can instead use the maximal response to different *sets* of wavelets, where the combination of wavelets within a set represent feature components. For example, we may have one set consisting of the line and edge wavelets, and another consisting of the junction wavelets. Local signal structure would then be represented by a combination of wavelets from whichever set gives the lowest distance. This is represented by

$$\boldsymbol{\lambda}, \theta, k = \underset{\boldsymbol{\lambda},\theta,k\in\mathbb{N}_K}{\arg\min} \|\mathbf{Wf} - \mathbf{S}_\theta \mathbf{W}\mathbf{U}_k \boldsymbol{\lambda}_k\|, \tag{2.86}$$

where $\{\mathbf{U}_k\}_{k\in\mathbb{N}_K}$ is a set of matrices representing different sets of wavelets, with their CH vectors as columns. The solution is to solve the equation for each individual set of wavelets, and choose the k that corresponds to the minimum residual. Solving for each set uses the method from Section 2.3 and therefore can also use regularisation.

To demonstrate, two wavelet sets were created. The first consists of the line and edge wavelets from Figure 2.4 with $m = 1$ and $m = 2$ respectively, and the second consists of the corner, T

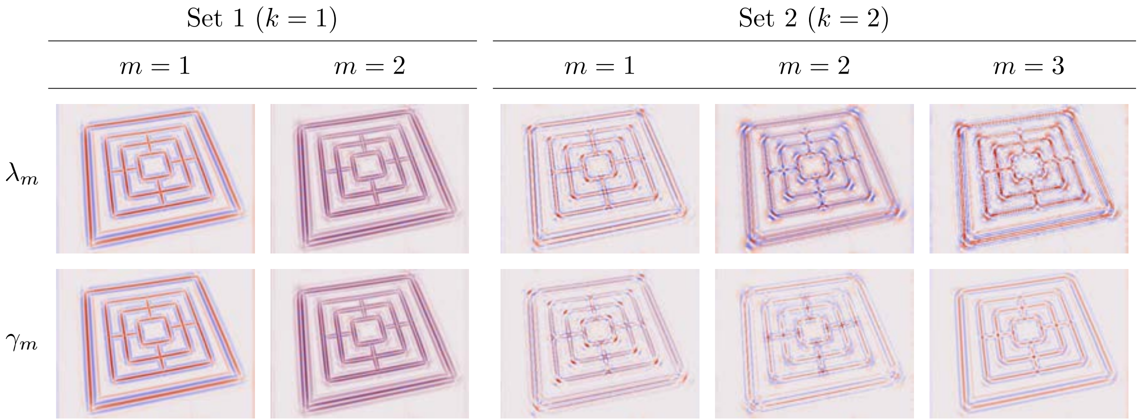


Figure 2.16: Test image response for each of the model wavelets from Figure 2.4 divided into a line and edge wavelet set (Set 1) and a corner, T and X junction set (Set 2).

junction, and X junction wavelets from Figure 2.4 with $m = 1$, $m = 2$ and $m = 3$ respectively. The responses for each wavelet (Figure 2.16) differ from previous examples in Section 2.3 in that the corner and junction set also have large responses at the locations of lines, as the line wavelet is no longer in the set.

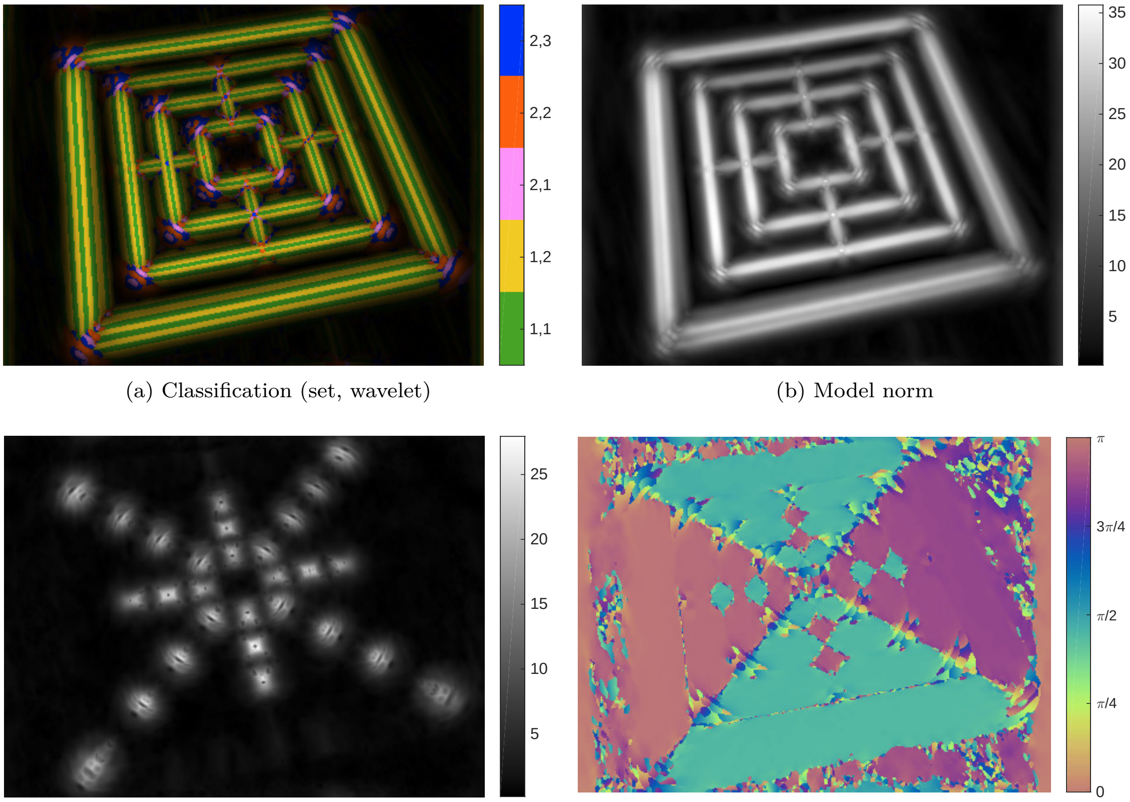
However, when compared to the previous example using maximal response to single wavelet, by using sets of wavelets the residual is no longer high around half-line and half-edge features (Figure 2.17c) and the model norm is smoother (Figure 2.17b), while classification is about the same (Figure 2.17a). This shows some image features are best represented by the combination of more than one wavelet.

The purpose of this example is to illustrate that neither collecting all the wavelets into one set with a common orientation (Figure 2.5) or treating all the wavelets individually (Figure 2.15) may be the best approach to local image modelling. Often one must combine complementary wavelets into the same set. In terms of models, a good example is the analysis of a honeycomb pattern. One could use a set of line and edge wavelets to describe the cell boundaries, and a set of Y junctions to describe the points where the boundaries of adjacent cells meet. The set of odd and even sinusoidal wavelets are one example that will be developed in Chapter 3 and the set of line-segment and edge-segment wavelets are another that will be developed in Chapter 5.

2.4.3 Multiple Wavelet Sets, Multiple Orientations

The previous example constrained the problem to representing the local image structure by a single wavelet or single wavelet set that gave the maximum response for a single orientation. This allowed for least-squares solutions featuring multiple different wavelets. Now we consider the case where the local signal structure is comprised of *multiple* wavelets at *multiple* orientations,

$$\mathbf{W}\mathbf{f} = \sum_{k=1}^{K} \lambda_k \mathbf{S}_{\theta_k} \mathbf{u}_{q_k} + \mathbf{W}\boldsymbol{\epsilon}, \tag{2.87}$$



(a) Classification (set, wavelet) (b) Model norm

(c) Residual norm (d) Orientation mod π

Figure 2.17: Classification model and residual vector magnitudes along with the global orientation estimate using the maximal response to two sets of wavelets. The first set contains a line and edge wavelet, the second contain a corner, T junction and X junction wavelet.

or multiple wavelet sets at multiple orientations

$$\mathbf{Wf} = \sum_{k=1}^{K} \mathbf{S}_{\theta_k} \mathbf{U}_{q_k} \boldsymbol{\lambda}_k + \mathbf{W}\boldsymbol{\epsilon}, \tag{2.88}$$

where q_k is the index of the wavelet set. The latter is the most general model. For example, consider a corner, T junction and an X junction. A corner could be modelled by two line-segments at 90 degrees, a T junction by a line and a line segment, and an X junction by two lines. Different model components, and thus different wavelets, at different orientations are therefore required. Furthermore, the wavelets do not have to be linearly independent and can even be multiple copies of the same wavelet at different orientations.

Iterative Solution

The solution to these types of problems is made possible by having the residual vector. The idea is to first model the structure using the best wavelet or wavelet set, then repeat the process using the remaining image structure as represented by the residual. Thus the method is iterative.

To begin with, we calculate the maximal response as in the previous section and choose the best wavelet set out of Q sets to choose from,

$$\boldsymbol{\lambda}_1, \theta_1, q_1 = \underset{\boldsymbol{\lambda}, \theta, q \in \mathbb{N}_Q}{\arg\min} \|\mathbf{W}\mathbf{f} - \mathbf{S}_\theta \mathbf{W}\mathbf{U}_q \boldsymbol{\lambda}\|, \tag{2.89}$$

then calculate the residual component from the result.

$$\mathbf{W}\boldsymbol{\epsilon}_1 = \mathbf{W}\mathbf{f} - \mathbf{S}_{\theta_1} \mathbf{W}\mathbf{U}_{q_1} \boldsymbol{\lambda}_1. \tag{2.90}$$

The process is repeated for the remaining wavelet sets up to K iterations, using the residual vector from the previous iteration instead of the original CH vector,

$$\boldsymbol{\lambda}_2, \theta_2, q_2 = \underset{\boldsymbol{\lambda}, \theta, q \in \mathbb{N}_Q}{\arg\min} \|\mathbf{W}\boldsymbol{\epsilon}_1 - \mathbf{S}_\theta \mathbf{W}\mathbf{U}_q \boldsymbol{\lambda}\|, \tag{2.91}$$

$$\cdots \tag{2.92}$$

$$\boldsymbol{\lambda}_K, \theta_K, q_K = \underset{\boldsymbol{\lambda}, \theta, q \in \mathbb{N}_Q}{\arg\min} \|\mathbf{W}\boldsymbol{\epsilon}_{K-1} - \mathbf{S}_\theta \mathbf{W}\mathbf{U}_q \boldsymbol{\lambda}\|. \tag{2.93}$$

Two set of wavelets, line and edge (set 1), and line segment and edge segment (set 2), shown in Figure 2.18, were used to model the *Board* image for $K = 2$ iterations. Each location in the image was classified for each iteration according to the set that gave the maximum response (Figure 2.19). Line and edge features had a maximal response to set 1 and a low residual after the first iteration, indicating that one component is enough to model them. Corner junctions had a high response to set 2 for two iterations, meaning that two line/edge-segment components are required to model them. T junctions had a high response to set 1 and then set 2, while X junctions had a high response to set 1 for two iterations. This shows that we can both describe features using different model components in the same model as well as differentiate them according to which model wavelets give high responses.

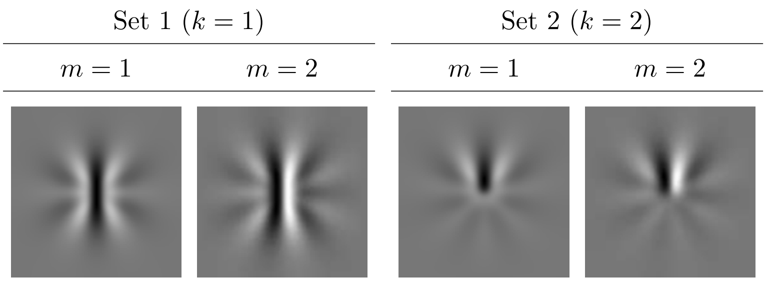


Figure 2.18: The sinusoidal (line and edge) wavelets developed in Chapter 3 (set 1), and the half-sinusoidal (line segment and edge segment) wavelets developed in Chapter 5 (set 2).

2.4.4 Identical Wavelet Sets, Multiple Orientations

If instead of choosing from Q different sets of wavelets, we can simplify the previous model to simply have K copies of the *same* set of wavelets. An example would be to model the T junction

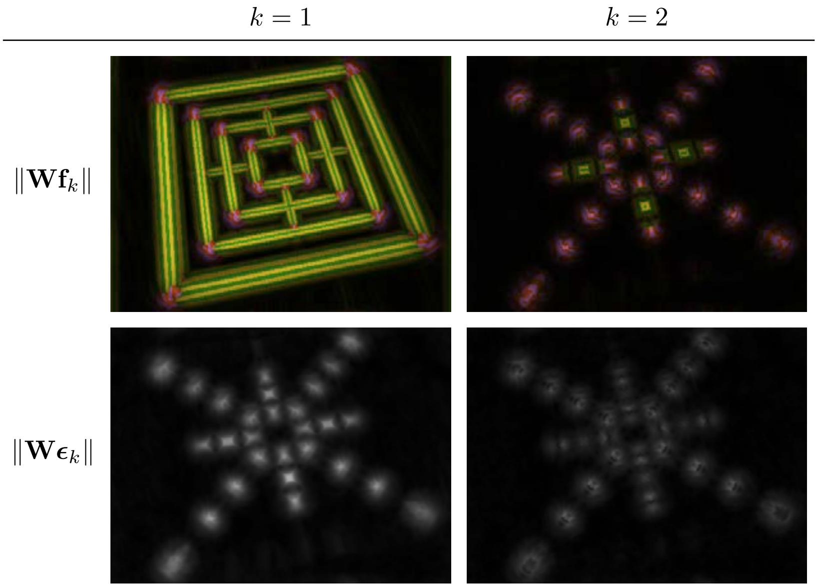


Figure 2.19: Result of modelling the test image with the two sets of wavelets in Figure 2.18. Top row: model norm for each iteration, coloured according to the individual wavelet with the largest response from the *set* with the largest response. Green: line wavelet, yellow: edge wavelet, pink: line segment wavelet, red: edge segment wavelet. Bottom row: residual norm.

using multiple copies of a line-segment at different strengths and orientations,

$$\mathbf{W}\mathbf{f} = \sum_k \mathbf{S}_{\theta_k}\mathbf{W}\mathbf{U}\boldsymbol{\lambda}_k + \mathbf{W}\boldsymbol{\epsilon}, \tag{2.94}$$

where $\mathbf{WU}$ is the matrix of wavelet CH vectors.

This type of model is what will be predominately used in the rest of this thesis. The general idea of analysis using the same wavelet set is used by, for example, Perona [98], Freeman [36], Michaelis and Sommer [80] and Simoncelli and Farid [110] to parametrise junctions. They steer a filter or filter pair and find peaks in the orientation response to determine the amplitude, orientation and number of components. Where the proposed approach differs is we have the residual component that describes the part of the signal that isn't well modelled, and thus an iterative method of solving for the both the amplitude and orientation parameters.

Iterative Solution

The iterative approach to solving this is the same as for different wavelet sets. First we find the parameters corresponding to the minimum distance,

$$\boldsymbol{\lambda}_1, \theta_1 = \arg\min_{\boldsymbol{\lambda},\theta} \|\mathbf{W}\mathbf{f} - \mathbf{S}_\theta\mathbf{W}\mathbf{U}\boldsymbol{\lambda}\|, \tag{2.95}$$

then repeat the process using the residual,

$$\boldsymbol{\lambda}_2, \theta_2 = \arg\min_{\boldsymbol{\lambda},\theta} \|\mathbf{W}\boldsymbol{\epsilon}_1 - \mathbf{S}_\theta \mathbf{W}\mathbf{U}\boldsymbol{\lambda}\|, \tag{2.96}$$

$$\cdots \tag{2.97}$$

$$\boldsymbol{\lambda}_K, \theta_K = \arg\min_{\boldsymbol{\lambda},\theta} \|\mathbf{W}\boldsymbol{\epsilon}_{K-1} - \mathbf{S}_\theta \mathbf{W}\mathbf{U}\boldsymbol{\lambda}\|.$$

Roots Solution

The traditional approach [36, 80, 98, 110] is to find peaks in the angular response of the wavelet set. In the context of the CH vector, recall that the polynomial we must find the maximum of is

$$p(\theta) = \sum_m \lambda_m(\theta)\delta_m(\theta). \tag{2.98}$$

The roots of the derivative polynomial $p'(\theta)$ give up to $4N$ possible candidates for the orientation, $\{\theta_i\}$. We choose the values that satisfy the following conditions:

1. $p'(\theta_i) = 0$: The derivative polynomial evaluates to 0.
2. $p''(\theta_i) < 0$: The second derivative polynomial evaluates to a negative value, meaning the point is at a local maximum.

From the candidate set we choose the θ_i that give the K largest values of $p(\theta)$, ordered from largest to smallest. Then λ and δ are calculated for each wavelet in the set. In contrast to the iterative process, this method only requires the polynomial roots to be calculated once. Finding the roots of a large polynomial is computationally expensive. In MATLAB, the in-built `roots` method, which uses the eigenvalues of the companion matrix, was the fastest implementation that could be found. Even so, the time taken to process an entire 512×512 image is in the order of seconds (Table 2.1). Others methods were implemented, such as the Durand-Kerner and Bairstow methods, however these were slower. Therefore the number of root finding operations is an important distinction between the iterative and roots methods.

Four components ($K = 4$) of the line-segment and edge-segment (half-sinusoidal) wavelet set (Figure 2.18) were used to describe the test image. The parameters we solved using both the iterative and roots methods (Figure 2.20). The third component ($k = 3$) of the model highlights interesting differences between the each method. Firstly, the roots method still has a response at the location of lines and edges whereas the iterative method does not. Secondly, the response at the centres of the junctions is lower for the roots method. This means model components given by the iterative method better describe these locations than the roots method.

Finally, one can also classify each pixel according to the number of components above a certain threshold (Figure 2.21). The iterative method appears to better classify the regions near corners.

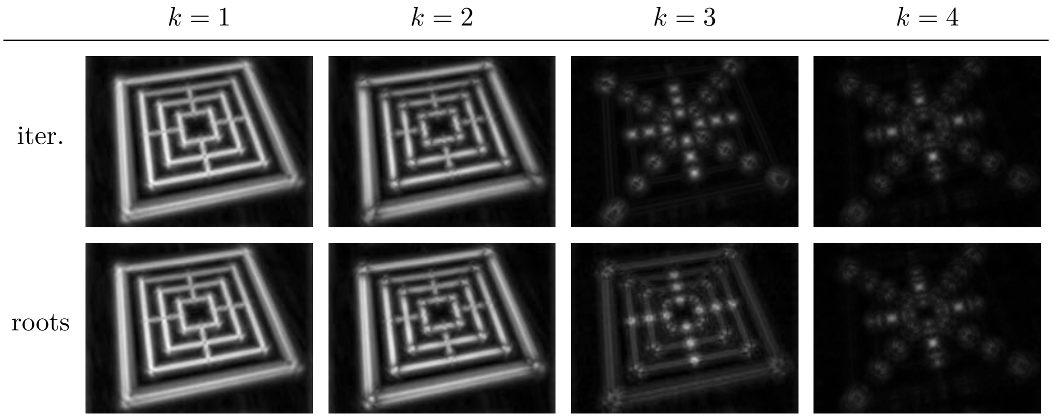
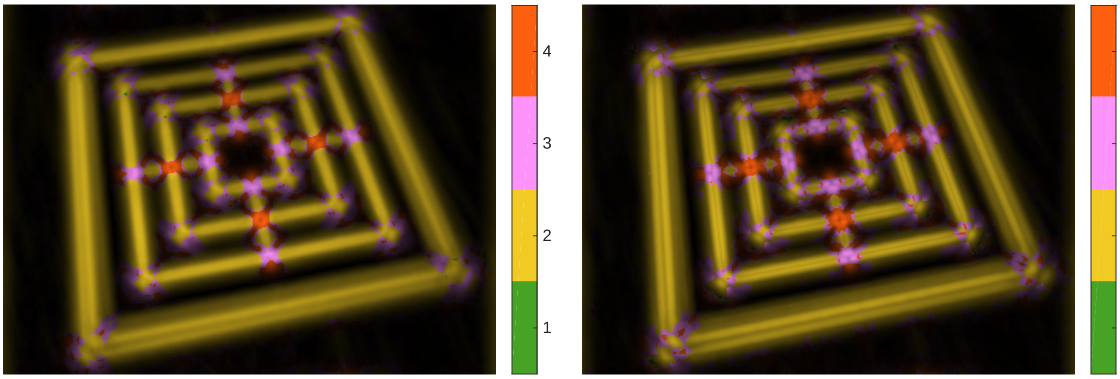


Figure 2.20: Norm of each model component of the test image found using the line-segment / edge-segment wavelet set (Figure 2.18) for four components, using either the iterative or roots methods.



(a) Iterative Method

(b) Roots Method

Figure 2.21: Classification of the test image using the line-segment / edge-segment wavelet set at up to four orientations. Points are classified by the number of model components with amplitude greater than 0.4 times the first model component amplitude. Brightest is the magnitude of the combined four model vectors.

2.4.5 Summary

This section has developed two methods of solving image models using multiple sets of wavelets at multiple orientations. The iterative method makes use of the residual component solve the model piece by piece. The roots method solves all the components at once, but is limited to models using multiple copies of the same component (wavelet set). These two methods will be referred to throughout the thesis. The test image response (Figures 2.21 and 2.20) suggests the iterative method gives slightly better parameters than the roots method for the second wavelet set in Figure 2.18.

It was necessary to introduce the line/edge (sinusoidal) and line-segment/edge-segment (half-sinusoidal) wavelet sets to provide good examples of the different types of maximal response models. The sinusoidal wavelets are developed properly in Chapters 3 and 4 and the half-sinusoidal wavelets in Chapter 5. The performance of the iterative and roots methods are tested for each wavelet set in these chapters.

2.5 Super-Resolution

Both the iterative method and root finding method allow us to obtain the parameters of a single wavelet set at K orientations. A recently developed method called *super-resolution* can also be applied when solving a model that uses a single wavelet or two orthogonal wavelets that meet certain criteria. Using super-resolution to solve for multiple wavelet orientations is a novel application of the method, and the connection does not appear to have been made before. One reason may be that it is a recent development. The other is that it also solves for model parameters by minimising the norm of residual vector, and thus lends itself to the CH vector approach.

2.5.1 Overview

Discrete basis pursuit denoising [18] is a method for finding sparse approximations of signals from a dictionary of elements. That is, finding a sparse solution for the least-squares problem,

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \leq \delta. \tag{2.99}$$

In the above, $\mathbf{y}$ can be thought of as a discrete signal, $\mathbf{A}$ as a matrix whose columns represent model components, and $\mathbf{x}$ as a vector of their amplitudes. The idea is to find the smallest number of model components such that norm of the residual signal, $\|\mathbf{y}-\mathbf{A}\mathbf{x}\|$, is below a certain threshold. A CH vector model where all the components are fixed in orientation could thus be solved using the same method, that is,

$$\min_{\boldsymbol{\lambda}} \|\boldsymbol{\lambda}\|_1 \text{ subject to } \|\mathbf{Wf} - \mathbf{WU}\boldsymbol{\lambda}\|_2^2 \leq \delta, \tag{2.100}$$

where $\mathbf{Wf}$ is the weighted image CH vector, $\mathbf{WU}$ is a matrix whose columns are the model wavelet CH vectors, and $\boldsymbol{\lambda}$ is a vector of their amplitudes. The restriction to a fixed orientation makes this approach limited in application regarding local feature analysis, as orientation is on of the parameters needed to represent features.

With some coaxing the method was applied to estimating the orientations of a model consisting of one component at different amplitudes and *discrete* orientations. The problem takes the form,

$$\min_{\boldsymbol{\lambda}} \|\boldsymbol{\lambda}\|_1 \text{ subject to } \|\mathbf{Wf} - \boldsymbol{\Theta}\boldsymbol{\lambda}\|_2^2 \leq \delta. \tag{2.101}$$

The matrix consists of columns of the model wavelet CH vector, $\mathbf{Wu}$, at discrete orientations over the range $[0, 2\pi)$. For example, if we wished to find the orientation of the components to within one degree, the matrix would have 360 columns consisting of the model wavelet CH vector rotated in one degree increments. Solving the problem thus gives a sparse solution for $\boldsymbol{\lambda}$ where the positions of the non-zero elements correspond to the orientations of the model wavelets, and their values to the amplitudes.

Solving the problem by making it discrete seemed inelegant. A search of the literature to find a continuous variation of the basis pursuit denoising algorithm was performed, and the recently developed method of the super-resolution of complex spike trains [14, 15] was discovered.

2.5.2 Super-Resolution

Super-resolution involves finding the location and amplitude of spikes in a complex spike train, from the low-frequency Fourier series components of that signal [15] which may be corrupted by noise [14]. Restating the problem from [14, 15], let $x(\theta)$ be a signal composed of the superposition of K complex-valued spikes with amplitudes $\{\alpha_k\}_{k\in\mathbb{N}_K} \in \mathbb{C}$, at locations $\{\theta_k\}_{k\in\mathbb{N}_K} \in [0, 2\pi)$. That is,

$$x(\theta) = \sum_{k=1}^{K} \alpha_k \delta(\theta - \theta_k). \tag{2.102}$$

Now let $\boldsymbol{\mathcal{F}}^N$ be the operator that maps a signal to a vector of its $-N$-th to N-th Fourier series coefficients. That is,

$$\begin{aligned} \{\boldsymbol{\mathcal{F}}^N x(\theta)\}_n &= \int_0^{2\pi} x(t) e^{-in\theta\phi}\, d\phi & (2.103) \\ &= \sum_{k=1}^{K} \alpha_k e^{-in\theta_k}. & (2.104) \end{aligned}$$

Letting $\mathbf{x} = \boldsymbol{\mathcal{F}}^N x(\theta)$ be this vector, the position and amplitude of the spikes can be estimated by solving

$$\min_{\boldsymbol{\alpha},\boldsymbol{\theta}} \|\tilde{x}(\theta)\|_{\mathrm{TV}} \quad \text{subject to} \quad \|\boldsymbol{\mathcal{F}}^N \tilde{x}(\theta) - \mathbf{x}\|_1 = 0, \tag{2.105}$$

where $\tilde{x}(\theta)$ is the *estimated* spike train, and $\|\tilde{x}(t)\|_{\mathrm{TV}}$ is the total-variation norm, which can be interpreted as the generalization of the ℓ_1 norm to the real line [14]. For this problem, it is equivalent to $\|\boldsymbol{\alpha}\|_1$ with the constraint that each element in the set of orientations $\boldsymbol{\theta}$ is unique.

If instead $y(\theta)$ is a complex-valued spike train corrupted by noise, that is, $y(\theta) = x(\theta) + \eta(\theta)$, with Fourier series coefficients given by $\mathbf{y} = \boldsymbol{\mathcal{F}}^N y(\theta)$, the super-resolution method estimates the values of the spike train by solving

$$\min_{\boldsymbol{\alpha},\boldsymbol{\theta}} \|\tilde{\mathbf{x}}(\theta)\|_{\mathrm{TV}} \quad \text{subject to} \quad \|\boldsymbol{\mathcal{F}}^N \tilde{\mathbf{x}}(\theta) - \mathbf{y}\|_1 < \delta, \tag{2.106}$$

where $\tilde{\mathbf{x}}(\theta)$ is the estimated spike train and δ is a parameter that relaxes the problem to account for the noise. For an unknown signal, $y(\theta)$, that we *assume* is a noisy complex-valued spike train, δ also relaxes the problem for situations where the signal cannot be properly modelled by a spike train. In this case, δ accounts for the non-spike part of the signal.

For an ideal spike train with no noise and $\delta = 0$, total variation minimisation will give a *unique* and therefore *exact* solution, to arbitrarily small precision, so long as the minimum separation between spikes is greater than $4\pi/N$ for complex valued spikes, or $3.74\pi/N$ for real valued spikes [15]. The minimum separation is the smallest distance between any set of spike locations and is given by [15]

$$\delta(\theta, \theta') = \inf_{(\theta,\theta') \in \Theta : \theta \neq \theta'} |\theta - \theta'|. \tag{2.107}$$

Note this is the wrap-around distance, that is, the distance between 0 and 2π would be 0.

Numerical simulations in [15] suggest that this separation constraint may be as low as $2\pi/N$ when there is a low number of spikes compared to N. Furthermore, for real spikes all of the *same* sign, the minimum distance is much smaller. The maximum number of spikes that can be resolved is N [15]. Importantly, the super-resolution problem does not need to be made discrete. It can be solved directly from the Fourier coefficients via a semi-definite program, of which the mathematical basis is developed in [15] along with links to example MATLAB code.

Recovery of the original spike train is possible even when spikes are close together, as shown in Figure 2.22. The example consists of four spikes recovered from five Fourier series components. The first two spikes are close together with only one corresponding local maxima in the low-pass response. If the position of the spikes were estimated simply by looking at the local extrema, only one spike would be found at this location, using super-resolution both spikes are found.

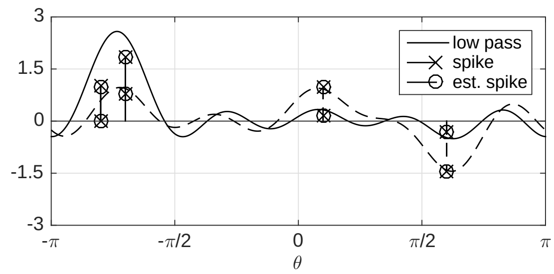


Figure 2.22: A complex-valued spike train shown with the signal reconstructed using just the first five Fourier series coefficients, and the spike train estimated from these coefficients.

2.5.3 Application to Model Orientation Estimation

Consider the low-pass response given by the sum of the first N Fourier series components of a unit-valued spike. If we think of this function as a model signal, then the super-resolution problem can be though of as finding the positions and amplitudes of the sum of these model components that make up the signal being analysed. Replace position with orientation, and we find that the super-resolution method can be applied to solve for CH vector models consisting of one or two wavelets at multiple orientations, with a few constraints.

To begin with, consider the model consisting of a single wavelet $\mathbf{Wu}$ at different orientations and amplitudes,

$$\mathbf{Wf} = \sum_{k=1}^{K} \lambda_k \mathbf{S}_{\theta_k} \mathbf{Wu} + \mathbf{W}\boldsymbol{\epsilon}. \tag{2.108}$$

An individual order of the image CH vector would therefore be given by

$$w_n f_n = \sum_{k=1}^{K} \lambda_k e^{-\mathrm{i}n\theta_k} w_n u_n + w_n \epsilon_n. \tag{2.109}$$

Letting $g_n = f_n/u_n$, this may be written as

$$g_n = \sum_{k=1}^{K} \lambda_k e^{-\mathrm{i}n\theta_k} + \epsilon_n/u_n. \tag{2.110}$$

This is equivalent to the n-th Fourier series coefficient of a complex-spike train,

$$x(\theta) = \sum_{k} \lambda_k \delta(\theta - \theta_q), \tag{2.111}$$

which has been corrupted by 'noise'. In this case, noise also means non-model components of the local image structure, that is, the residual. With the problem in this form we may then obtain the amplitudes and orientations by solving the super-resolution equation

$$\min_{\boldsymbol{\lambda},\boldsymbol{\theta}} \|\tilde{x}(\theta)\|_{\mathrm{TV}} \quad \text{subject to} \quad \|\boldsymbol{\mathcal{F}}^N \tilde{x}(\theta) - \mathbf{g}\|_1 < \delta,$$

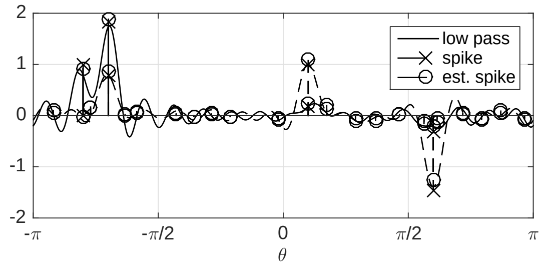
where $\mathbf{g}$ is the CH vector adjusted above.

There are a few caveats with this approach:

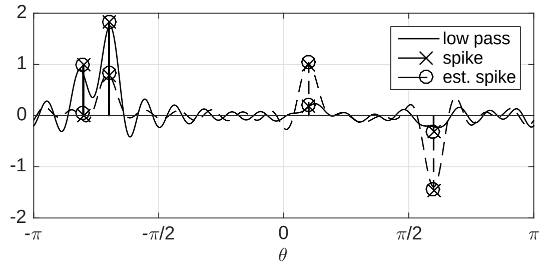
- It is assumed that the model wavelet has all non-zero components, u_n, otherwise g_n would be undefined. An exception is the sinusoidal wavelets covered in Chapter 4.
- It is also assumed the signal can be reasonably modelled by our choice of wavelet. If it cannot, there is no guarantee the signal will be angularly band-limited in the same way the wavelet is. Thus if for some model CH vector, $\mathbf{Wu}$, u_n is small, $g_n = f_n/u_n$ can be potentially quite large.
- Since we divide by the wavelet coefficients, any weighting is also divided out and thus there is no point considering different weightings.

2.5.4 Noise Component

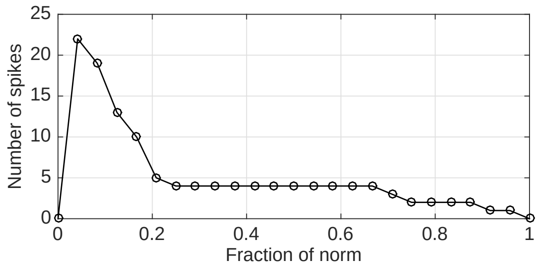
The choice of δ is important for solving the spike train parameters, as no solution is possible if δ is lower than the actual noise level [14]. When δ is set too low, no solution is returned. As δ is increased past some threshold, a high number of spikes are returned (Figures 2.23a and 2.23c), as the extra components are required to meet the norm constraint. Increasing δ further relaxes the problem and reduces the number of spurious spikes returned (Figure 2.23b), however as δ approaches $\|\mathbf{f}\|$ the number of spikes reduces to zero (Figure 2.23c).



(a) Solution for $\delta = 0.02 \times \|\mathbf{f}\|_2$



(b) Solution for $\delta = 0.5 \times \|\mathbf{f}\|_2$



(c) Number of spikes returned for different values of δ as a proportion of $\|\mathbf{f}\|_2$

Figure 2.23: Example solutions for a four-spike train with added noise and $N = 17$ for different values of δ. Solid line: real component, dashed line: imaginary component.

For a local image model, the error component consists of both actual noise and structures that are not well represented by the model. Therefore we must choose

$$\delta \geq \|\boldsymbol{\epsilon}'\|, \tag{2.112}$$

where $\epsilon'_n = \epsilon_n / u_n$. Therefore δ needs to be set higher at locations in the image where non-model structures are present, making estimating the actual value of δ difficult, even if the type and amount of noise is known. To account for this, δ is first initialised to a low value and then gradually increased until a solution is found. Typically, this solution will have a large number of spikes. To reduce the number of spikes we may

- Choose spikes with amplitude above a threshold.
- *Combine* nearby spikes with orientation separation below a threshold.

To reduce the number of spikes below a maximum, K, as dictated by the model we may

- Increase δ until K spikes remain.
- Choose the K largest amplitude spikes.
- Combine nearby spikes until K spikes are left.

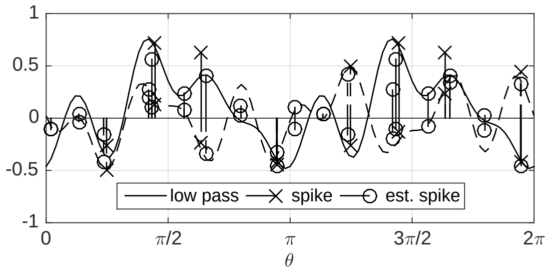
Combining spikes is preferred, as it seeks to incorporate information from the lower valued spikes, rather than discarding them as is the case with increasing δ or thresholding the amplitude. For example, In some cases two spikes are returned very close together when there is only one spike in the original signal. Increasing δ requires repeated application of the super-resolution algorithm and thus increases computational load. However, combining spikes requires few operations. The procedure is:

1. Solve for $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$ using a small value of δ such as $0.05 \times \|\mathbf{f}\|_2$, increasing if necessary.
2. Choose the element of $\boldsymbol{\lambda}$ with the largest magnitude, $|\lambda_k|$.
3. Find all elements of $\boldsymbol{\theta}$ which differ from θ_k by less than a separation threshold, inclusive of θ_k. Let $\mathbf{J}$ be the set of their indices.
4. Calculate a new orientation value using vector averaging,
$$\beta = \sum_{j \in \mathbf{J}} |\lambda_j| e^{\mathrm{i}\theta_j},$$
$$\theta_{\text{new}} = \arg(\beta).$$
5. Calculate a new complex amplitude by projecting the real and imaginary parts of the old components onto a new vector,
$$\Re(\lambda_k) = \frac{\left\langle \sum_{j \in J} \mathrm{Re}(\lambda_j) e^{\mathrm{i}\theta_j}, \beta \right\rangle}{|\beta|},$$
$$\Im(\lambda_k) = \frac{\left\langle \sum_{j \in J} \mathrm{Im}(\lambda_j) e^{\mathrm{i}\theta_j}, \beta \right\rangle}{|\beta|}.$$
6. Remove the remaining components with indices in the set $\mathbf{J}$; $\theta_{j \in \mathbf{J} \setminus k}$ and $\lambda_{j \in \mathbf{J} \setminus k}$ from $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$, respectively.
7. Repeat until no more components can be removed.

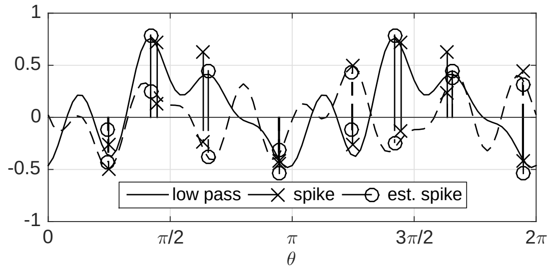
Combining spikes gives a similar result to increasing δ for the example eight-spike train signal in Figure 2.24. The solution using a low δ results in multiple smaller spikes corresponding to the actual spikes. Once δ is increased to $0.6 \times \|\mathbf{f}\|$ a close approximation with the same number of spikes is reached (Figure 2.24b). Keeping δ low but using spike combining also achieves similar results albeit with two small extra components (Figure 2.24c). The procedure of combining components that are nearby in orientation can be applied to the normal multiple orientation CH vector models as well.

2.5.5 Orientation Separation

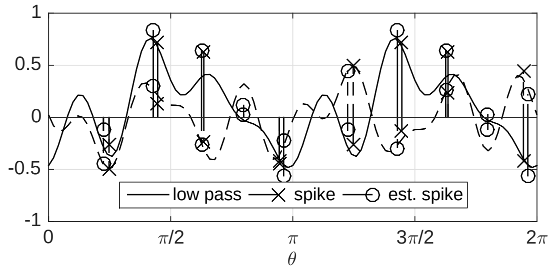
As mentioned previously, theoretical results in [15] give a minimum separation between complex-valued spikes of $4\pi/N$, and real-valued spikes of $3.74\pi/N$, while the minimum separation is almost



(a) Solution for $\delta = 0.05\|\mathbf{f}\|_2$



(b) Solution for $\delta = 0.6\|\mathbf{f}\|_2$



(c) Solution for $\delta = 0.05\|\mathbf{f}\|_2$, spikes combined for separation $< \pi/N$

Figure 2.24: Comparison between increasing δ (b) and spike combining (c) for a noisy spike train (a) with $N = 9$.

zero if the spikes are the same sign. This equates to being about to resolve up to $N/4$ spikes, thus when applied to local image modelling, one would need up to the 8th order RT to calculate a model with two wavelets ($K = 2$). However, numerical simulations in [15] suggest that the minimum separation may be as low as $2\pi/N$, meaning that up to $N/2$ spikes can be resolved. The $N/2$ limit also applies to real-valued spikes of the same sign, as there need to be at least two unknowns per spike [15]. In later chapters the orientation separation will be investigated for specific models that can be solved for using the super-resolution method.

2.6 Computation Speed

2.6.1 Root Finding

Apart from the super-resolution method, minimising the residual vector ℓ_2 norm is common to all the methods presented. The main problem to solve in each is finding the value of θ corresponding to the maximum of the trigonometric polynomial,

$$p(\theta) = \sum_m \delta_m(\theta)\lambda_m(\theta), \tag{2.113}$$

where $\delta_m(\theta)$ and $\lambda_m(\theta)$ are the angular response polynomials of the m-th wavelet in a set and its pseudo-inverse, respectively. There are two parts to solving this

1. **Polynomial multiplication:** Multiplying $\delta_m(\theta)$ and $\lambda_m(\theta)$ results in another trigonometric polynomial with twice the number of coefficients. The multiplication can be implemented by

convolving two discrete signals made up of the coefficients of each polynomial.

2. **Root finding:** The common method of finding the location of the maximum is to find the roots of the derivative of the polynomial.

The convolution and root finding operations make up most of the computational load when implementing the method (Table 2.2). The time is dependent on the size of the CH vectors (N) and the type of model:

- Models with a single wavelet can be solved for without squaring the polynomial. The solution is given by finding the maximum and minimum values of

$$p(\theta) = \lambda(\theta) \tag{2.114}$$

  which requires no convolution operation and root finding on a smaller 2N degree polynomial.

- Models with a single wavelet where the degree of the polynomial can be reduced to $p(n\theta)$ require less operations.

- Models with multiple wavelets in a set require one convolution for each wavelet, and root finding is performed on a $4N$ degree polynomial.

- Models with multiple wavelets in a set where the degree of the polynomial can be reduced to $p(n\theta)$ require less operations. An example is the sinusoidal model in the next chapter, which can be expressed as $p(2\theta)$.

- Models with K copies of the same wavelet or wavelet set require K root finding operations using the iterative method, but only one when using the roots method.

Root finding is computationally expensive. For example, a two-wavelet set with $N = 7$ requires finding the roots of a degree 28 polynomial which takes approximately 21.2 seconds using MATLAB `roots` on a single core of a 2.5Ghz Intel Core i7 processor for a 256 by 256 pixel image (Table 2.2). As a consequence, a variety of different root finding algorithms were investigated for implementation in MATLAB. However, none were faster than the inbuilt MATLAB `roots` function. Instead a different approach was investigated. Each order of the polynomial is an estimate of the location of the maximum. An iterative process was developed to make a rough approximation of the location of the maximum using each estimate, and is described in the next section.

2.6.2 Quick Approximation

A quick approximation method for finding the maximum of a trigonometric polynomial was developed. Consider a trigonometric polynomial of degree $2N$ given by

$$p(\theta) = \sum_{|n| \leq N} c_n e^{in\theta}. \tag{2.115}$$

The argument of n-th coefficient c_n gives an estimate of θ in the range $[0, 2\pi/n)$, and therefore n possible estimates for θ over the entire range $[0, 2\pi)$, while the magnitude of c_n is the strength of each estimate. We can think of these as representing n equiangular complex vectors for each positive order $|n| \in \mathbb{N}_N$,

$$b_{n_m} = \delta(n)|c_n| \exp\left(\mathrm{i}\frac{2\pi m + \arg(c_n)}{n}\right), \tag{2.116}$$

where $m \in \mathbb{N}_{n-1}$ is the index of the estimate, and $\delta(n)$ is a weighting function reflecting that higher orders are more sensitive to orientation changes and are therefore better estimates.

Next we choose one vector from each order and sum them to give a combined vector,

$$v_{\mathbf{m}} = \sum_{n=1}^{N} b_{n_{m_n}}, \tag{2.117}$$

where $\mathbf{m} \in \mathbf{M}^{(N)}$ and

$$\mathbf{M}^{(N)} = \{M_n^{(N)} \in \mathbb{N}_{n-1} \mid n = [1, ..., N]\} \tag{2.118}$$

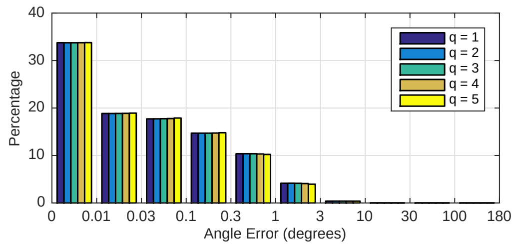
is the set of all possible indices of the estimates up to order N. There are thus $N!$ combinations we can create. Out of these $N!$ we choose the $v_{\mathbf{m}_N}$ with the greatest magnitude. Its argument is the final estimate for θ.

$$\theta = \arg v_{\mathbf{m}} \qquad \text{where } \mathbf{m} = \max_{\mathbf{m} \in \mathbf{M}_N} |v_{\mathbf{m}}|. \tag{2.119}$$

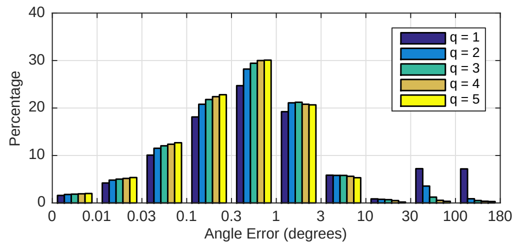
However, $N!$ possible combinations to search through becomes very large for large N, and defeats the purpose of a quick algorithm.

Instead, the following hybrid scheme is proposed. Firstly, all combinations are calculated for each of the first q orders, giving $q!$ vectors, from which the q vectors with the largest magnitude are chosen. For each of these q vectors, the vectors for the next order are added and the combined vector with the largest magnitude is kept. The number of vectors remains at q. This is repeated for the remaining orders, giving q final vectors, from which θ is chosen as the argument of the vector with largest magnitude. Using $\delta(n) = n^2$ for the order weighting function was found to give good results.

The performance of the approximation method depends on the type of model and the distribution of the features in the image. For example, if a structure in an image is represented by two components of similar magnitude, the polynomial will have two peaks with similar amplitude and thus the approximation method is more likely to pick the incorrect maximum. To get an idea of the error, orientation was calculated using the sinusoidal model (Chapter 3) for the first scale of the *Pentagon* image using $N = 7$, equal weighting, and using either MATLAB `roots` or the approximation method (Figure 2.25). The error between the approximation and the roots method was calculated, and split into two classes depending on whether the local structure had a larger model component or larger residual component. Over half the errors in the model class were below 0.01 degrees, and increasing q gave only a marginal improvement in error, while for the residual class the errors were higher, but still mostly less than 3 degrees. Increasing q reduced the number of large errors (above 30 degrees). These errors were due to structures that could be modelled with two components of similar magnitude.



(a) Model norm > residual norm



(b) Resid norm > model norm

Figure 2.25: Histogram of sinusoidal model orientation errors using the quick approximation method for the *Pentagon* image with $N = 7$ and equal weighting. Errors are divided into two classes: errors where the model norm was larger than the residual norm (a), and the opposite (b).

Computation time for the approximation method versus MATLAB `roots` was calculated for a 256 × 256 pixel image. The model parameters were solved using each method on a single core of a 2.5Ghz Intel Core i7 processor. Two different models were experimented for: a single model component resulting in a $2N$ degree polynomial (Table 2.1), and a two-wavelet set resulting in a $4N$ degree polynomial (Table 2.2). For both model types, the approximation method was over 100 times faster for $q = 1$; however, jumping from $q = 4$ to $q = 5$ resulted in a significant slow down. The time complexity for the initial q estimates and the subsequent search is $\mathcal{O}(q! + q\sum_{q+1}^{N} k)$ which makes $q > 5$ longer than the MATLAB roots method and therefore impractical. Choosing $q = 3$ is a good trade off between speed and accuracy.

N	poly. deg.	conv.	`roots`	proposed method				
				$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
2	4	0.0	4.5	0.0	0.0	0.0	0.0	0.0
3	6	0.0	4.9	0.0	0.0	0.1	0.1	0.1
5	10	0.0	6.2	0.0	0.1	0.2	0.7	3.4
7	14	0.0	8.3	0.1	0.1	0.3	0.9	5.1
9	18	0.0	10.6	0.1	0.1	0.3	1.3	6.7
11	22	0.0	14.2	0.1	0.2	0.5	1.6	7.9
13	26	0.0	17.2	0.1	0.2	0.4	1.7	8.8

Table 2.1: Single wavelet model convolution and orientation solving time (seconds) for the MATLAB roots method versus the proposed approximation method for $q \in [1, ..., 5]$, for the 256 × 256 *Pentagon* image.

A K-component model requires K polynomial maximum finding operations using the iterative approach but only one with the roots approach. However, the quick approximation technique can be used with the iterative method as it only needs to find a single maximum, where as the roots

N	poly. deg.	conv.	`roots`	proposed method				
				$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
2	8	0.1	5.4	0.0	0.1	0.1	0.1	0.1
3	12	0.1	7.2	0.1	0.1	0.2	0.2	0.2
5	20	0.2	12.9	0.1	0.1	0.4	1.6	7.6
7	28	0.2	21.2	0.1	0.2	0.6	2.1	10.8
9	36	0.4	38.7	0.1	0.2	0.6	2.6	13.6
11	44	0.4	65.0	0.2	0.3	0.8	3.5	17.1
13	52	0.6	87.6	0.2	0.4	0.9	4.4	20.4

Table 2.2: Two wavelet model convolution and orientation solving time (seconds) for the MATLAB roots method versus the proposed approximation method for $q \in [1, ..., 5]$, for the 256 $\times$ 256 *Pentagon* image.

method needs to find K local maxima and thus the approximation cannot be used. This makes the combination of the quick approximation and the iterative method a faster implementation, and has been used in most of the examples in the rest of this thesis. A quick approximation method that can find multiple local extrema is the subject of future work.

Quartic Solvers

The location of the maximum of a degree 4 polynomial can be found analytically using a quartic solver. The small number of operations involved means the operation takes less than 0.1 seconds for a 256 $\times$ 256 pixel image. For larger N, and thus larger degree polynomials, the quartic solver can still be used by only considering the first two orders of the polynomial. Note, this is not the same as limiting the number of RT orders when a multiple wavelet set is used, as since the polynomial to solve for is the sum of square of the two angular response polynomials for each wavelet, all RT orders are used to calculate the first three coefficients that are input into the quartic solver.

2.6.3 Super-Resolution

The MATLAB implementation of the super-resolution method in [15] is extremely slow but consistent. The time to calculate the parameters of a four-spike signal ranged from 2.7 seconds for $N = 4$ to 2.8 seconds for $N = 13$. This suggests there is a large computational load in setting up the semi-definite program as the actual order only increases the time slightly. Note this is for solving a *single* model, that is, one pixel in the image. Therefore, using the super-resolution method on a whole image would be impractical. Instead it is better to apply it at particular points of interest. Speeding up the semi-definite program for small N would be a worthwhile endeavour but is outside the scope of this thesis.

2.6.4 Implementation

In digital signal processing the RT is implemented using the discrete Fourier transform (DFT). In the continuous frequency domain we have the negative and positive order RT responses related by

$$\mathcal{R}^n f(\mathbf{z}) = \overline{\mathcal{R}^{-n} f(\mathbf{z})} \qquad \text{for even } n, \tag{2.120}$$

$$\mathcal{R}^n f(\mathbf{z}) = -\overline{\mathcal{R}^{-n} f(\mathbf{z})} \qquad \text{for odd } n. \tag{2.121}$$

However, this does not hold when the RT is implemented using the DFT of discrete images with even width or height. For example, let $F[i, j] = \mathrm{DFT}\{f[\mathbf{z}]\}$ where $f[\mathbf{z}]$ is a discrete image of size $M \times N$ pixels. When M and N are odd, all the values of $F[i, j]$ (except for $F[0, 0]$) are in a conjugate pair with $F[M-i, N-j]$. When either M or N are even there is an extra row or column of coefficients corresponding to wavelength 2 pixels with no conjugate pair. This asymmetry means the conjugate symmetry relationship above no longer holds, and the trigonometric polynomials used to solve the model parameters (2.56) are not real valued.

Two solutions are proposed:

- The image can be resized to have odd dimensions. However, this is only possible if no wavelet sub-sampling is required.
- The extra row or column in the DFT matrix with no conjugate pair is set to 0 before calculating the RT. The row or column is added back in during reconstruction.

2.7 Summary

This chapter developed the the CH vector, a new higher-dimensional representation of local image structure. It builds on previous 2D analytic signals such as the monogenic signal and signal multi-vector by adding extra higher-order RT components. The magnitude of the CH vector is a measure of the strength of the local signal structure that is invariant to rotation, while the normalised CH vector is a description of the shape of the structure that is invariant to illumination changes. The CH vector thus achieves the same split of identity common to the previous 2D analytic signal representations.

In previous approaches, the local image structure was completely represented by the image model, restricting the type of model and the number of orders of RT that could be used. The CH vector allows one to use a generic local image model consisting of sets of one or more components with different amplitudes and orientations. A method to solve for the model parameters was developed, whereby the CH vector is split into model and residual vectors and the residual component is minimised. Adding the residual component allows the use of arbitrary model CH vectors and any number of RT orders while still maintaining the ability to recover the original signal. The method is formulated in the context of wavelets; correlation of the image CH vector with the model CH vector is equivalent to correlation of image with the model wavelet. In this sense the method bridges the two approaches to local image analysis - 2D analytic signals and 2D steerable matched wavelets.

The residual component is a useful representation of the part of the local image structure that is not well modelled. The ratio of the residual vector magnitude to the model vector magnitude is therefore an illumination invariant measure of how well the model explains the structure. Having the residual vector also led to development of an iterative method of solving for multiple model components. The method enables the analysis of local image structure with different sets of model wavelets, even if they are not linearly independent. Another novel aspect is the application of the super-resolution method to solving for copies of one or two model components at different orientations, however it is computationally expensive. Even solving for orientation using polynomial root finding takes a significant amount of time. To address this, a quick approximation method was developed so that using the CH vector to solve models is fast enough to be employed in practical applications. This is necessary if the method is going to be widely adopted.

The CH vector representation is more useful than the tensor or geometric algebra representations of previous 2D analytic signals, as it admits any number of RT orders. These previous approaches have shown that the energy of phase-invariant descriptors is a good measure for the presence of image features, and this chapter has demonstrated that the CH vector magnitude is also high at the location of features. In Chapter 6 the shape of the local image structure, as represented by the normalised CH vector, will also be investigated for general feature detection.

The specific image models used in some of the examples given in this chapter are developed in the next three chapters. The matched CH wavelets are derived for each model, and the relationship between number of RT orders, N, and the ability to accurately resolve model parameters is explored.

Chapter 3

Single Sinusoidal Model

In this chapter we turn our attention to the initial problem of obtaining a phase-based representation of local image structure using a sinusoidal model derived from higher-order RT responses. The sinusoidal model can be used to parametrise lines and edges (Figure 3.1) and distinguish them from junctions and corners. The framework introduced in the previous chapter details the mathematical basis from which the sinusoidal model parameters can be found. The process is

- Create a model of the local image structures of interest.
- Derive the set or sets of matched wavelets.
- Calculate the parameters of the model using the appropriate method.

We shall also explore different weighting schemes, particularly those that make the magnitude of the CH vector phase-invariant. The advantage of having a description of the non-model part of the local image structure will be made clear in the latter part of the chapter, where we use the residual to develop a new representation of *intrinsic dimension*.

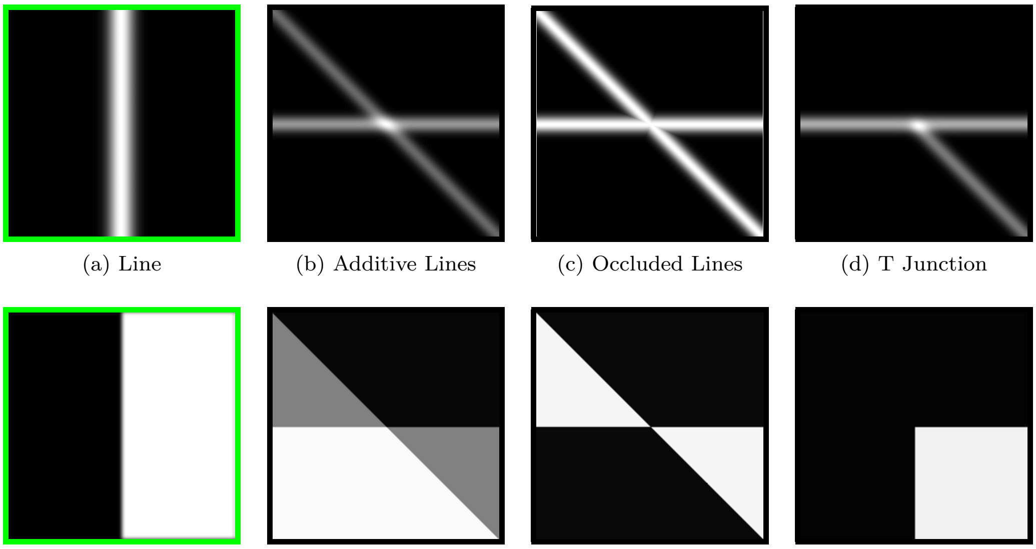
This chapter deals with a *single sinusoidal model*. Preliminary attempts at deriving the model can be found in [72, 74] and much of this chapter is based on work that has been published in [76].

3.1 Model

3.1.1 Multi-Sinusoidal Model

The monogenic signal models local image structure as a single sinusoid,

$$\begin{aligned} f(\mathbf{z}) &= f_{\mathrm{S}}(\mathbf{z}) && (3.1) \\ &= A_k \cos(\langle \mathbf{z}, \mathbf{o}_k \rangle + \phi_k), && (3.2) \end{aligned}$$



(a) Line (b) Additive Lines (c) Occluded Lines (d) T Junction

(e) Edge (f) Additive Edges (g) Chequer Edge (h) Corner

Figure 3.1: Examples of various image features. Lines (a) and edges (e) are well-modelled by the single sinusoidal model. Non-sinusoidal features (b-d,f-h) give a high residual component.

with amplitude, A, phase, ϕ, and orientation vector, $\mathbf{o} = [\cos\theta, \sin\theta]$. The structure multi-vector, 2D analytic signal and signal multi-vector use a model consisting of two sinusoids,

$$f(\mathbf{z}) = \sum_{k=1}^{2} \underbrace{A_k \cos(\langle \mathbf{z}, \mathbf{o}_k \rangle + \phi_k)}_{f_{\mathrm{S}}(\mathbf{z})}, \tag{3.3}$$

with various constraints on the parameters. To generalise this model, an expanded multi-sinusoidal signal model consisting of K oriented sinusoids with differing amplitude, phase and orientation plus a residual component is proposed. The local image structure at a point of interest, $\mathbf{z} = 0$, is modelled by

$$f(\mathbf{z}) = \sum_{k=1}^{K} \underbrace{A_k \cos(\langle \mathbf{z}, \mathbf{o}_k \rangle + \phi_k)}_{f_{\mathrm{S}}(\mathbf{z})} + f_\epsilon(\mathbf{z}), \tag{3.4}$$

where $f_{\mathrm{S}}(\mathbf{z})$ is a single sinusoidal model component and $f_\epsilon(\mathbf{z})$ is the residual.

In this chapter, the model shall be restricted to a single sinusoid, that is, $K = 1$, and multiple sinusoids will be discussed in the next chapter. The single sinusoidal model is

$$f(\mathbf{z}) = \underbrace{A \cos(\langle \mathbf{z}, \mathbf{o} \rangle + \phi)}_{f_{\mathrm{S}}(\mathbf{z})} + f_\epsilon(\mathbf{z}). \tag{3.5}$$

Adding the residual component allows for the inclusion of higher-order RT responses up to any order, as it then becomes possible to choose model parameters that satisfy the RT sinusoidal response equations,

$$\mathcal{R}^n f = \begin{cases} Ae^{in\theta}\cos(\phi) + \mathcal{R}^n f_\epsilon, & n \text{ is even,} \\ Ae^{in\theta}\,\mathrm{i}\sin(\phi) + \mathcal{R}^n f_\epsilon, & n \text{ is odd,} \end{cases} \tag{3.6}$$

for arbitrary images. In contrast, the monogenic signal [29], 2D analytic signal [136] and signal multi-vector [135] require the image to be completely represented by the model, limiting the number of RT orders that can be used. The task now is to choose appropriate values of amplitude, phase and orientation using the methods described in the previous chapter.

3.1.2 Matched Wavelets

We must first create a set of 2D steerable wavelets that are matched to the particular structure of interest. In this case, the proposed model is represented by a purely sinusoidal image given by

$$f_{\mathrm{S}}(\mathbf{z}) = A\cos(\omega_0\langle \mathbf{z}, \mathbf{o}\rangle + \phi), \tag{3.7}$$

where $\mathbf{o} = [\cos\theta, \sin\theta]$ and ω_0 is the sinusoid frequency.

Thus we must find the wavelet that matches the sinusoid for a particular amplitude, A, phase, ϕ and orientation, θ. Let $\{\psi_n\}_{|n|\leq N}$ be the set of CH wavelets up to order N, generated from an isotropic wavelet ψ. For simplicity, let the sinusoid frequency be located at the centre of the wavelet passband $h(\omega)$ such that $|h(\omega_0)| = 1$. The value of the sinusoid CH vector $\mathbf{f}_{\mathrm{S}}$ at the origin is therefore given by

$$f_{\mathrm{S_n}} = \langle f_{\mathrm{S}}, \psi^n\rangle \tag{3.8}$$

$$= \mathcal{R}^{-n}(f_S * \psi)(\mathbf{0}) \tag{3.9}$$

$$= \begin{cases} Ae^{-in\theta}\cos(\phi), & n \text{ is even,} \\ Ae^{-in\theta}\,\mathrm{i}\sin(\phi), & n \text{ is odd.} \end{cases} \tag{3.10}$$

The sinusoidal image CH vector can be written as a function of amplitude, phase and orientation,

$$\mathbf{f}_{\mathrm{S}}(A, \phi, \theta) = A\mathbf{S}_\theta \cos\phi\, \mathbf{s}_e + A\mathbf{S}_\theta \sin\phi\, \mathbf{s}_o, \tag{3.11}$$

where $\mathbf{s}_e$ and $\mathbf{s}_o$ are orthogonal CH vectors given by

$$s_{e_n} = 1 \qquad \text{if } n \text{ even}, 0 \text{ otherwise,} \tag{3.12}$$

$$s_{o_n} = -\mathrm{i} \qquad \text{if } n \text{ odd}, 0 \text{ otherwise.} \tag{3.13}$$

The vector $\mathbf{s}_e$ has only even orders and therefore represents an even wavelet, while $\mathbf{s}_o$ has only odd orders and therefore represents an odd wavelet. The sinusoidal image CH vector is thus given by the linear combination of these two wavelet vectors, rotated to the same orientation.

Thus to describe a sinusoid we require a wavelet set consisting of two matched wavelets that correlate with the even and odd parts of the signal. Setting orientation to 0, the two matched wavelets for a given weighting matrix $\mathbf{W}$ are

$$\mathbf{W}\mathbf{f}_e = \mathbf{s}_e\mathbf{W}/\sqrt{W_e}, \tag{3.14}$$

$$\mathbf{W}\mathbf{f}_o = \mathbf{s}_o\mathbf{W}/\sqrt{W_o}, \tag{3.15}$$

where W_e and W_o are the sum of the even and odd weights respectively,

$$W_e = \sum_{n \text{ even}, |n| \in \mathbb{N}_N} w_n^2, \tag{3.16}$$

$$W_o = \sum_{n \text{ odd}, |n| \in \mathbb{N}_N} w_n^2. \tag{3.17}$$

The model sinusoidal CH vector can be expressed as the scaled and rotated sum of the model wavelet CH vectors

$$\mathbf{W}\mathbf{f}_{\mathrm{S}}(A, \phi, \theta) = \lambda_e \mathbf{S}_\theta \mathbf{W}\mathbf{f}_e + \lambda_o \mathbf{S}_\theta \mathbf{W}\mathbf{f}_o, \tag{3.18}$$

where

$$\lambda_e = \sqrt{W_e}\, A \cos\phi, \tag{3.19}$$

$$\lambda_o = \sqrt{W_o}\, A \sin\phi. \tag{3.20}$$

An example of the two types of model wavelets is shown in Figure 3.4, generated from a Simoncelli-type primary wavelet for different values of N. The wavelets for $N = 1$ are the monogenic signal wavelets. As N increases the wavelets elongate along the axis perpendicular to the sinusoid orientation.

3.1.3 Solution

The proposed sinusoidal model of an arbitrary image, f, with the point of interest located at $\mathbf{z} = \mathbf{0}$, and localised by an isotropic wavelet, ψ_i, is

$$(f * \psi_i)(\mathbf{z}) = \underbrace{A \cos(\langle \mathbf{z}, \mathbf{o} \rangle + \phi)}_{f_{\mathrm{S}}} + f_\epsilon(\mathbf{z}). \tag{3.21}$$

Using the sinusoidal model wavelets previously derived, we may now write the image CH vector as the sum of model and residual components,

$$\mathbf{Wf} = \mathbf{W}\mathbf{f}_{\mathrm{S}}(A, \phi, \theta) + \mathbf{W}\boldsymbol{\epsilon} \tag{3.22}$$

$$= \lambda_e \mathbf{S}_\theta \mathbf{W}\mathbf{f}_e + \lambda_o \mathbf{S}_\theta \mathbf{W}\mathbf{f}_o + \mathbf{W}\boldsymbol{\epsilon}. \tag{3.23}$$

The model parameters are solved for by minimising the residual, $\|\mathbf{W}\boldsymbol{\epsilon}\|$. Since the model vectors are orthonormal and have the same orientation the solution is given by the model for solving a wavelet set with a single orientation from Section 2.4.2. Orientation is thus given by the maximum of the polynomial

$$\theta = \max_{\theta} \lambda_e(\theta)^2 + \lambda_o(\theta)^2, \tag{3.24}$$

where

$$\lambda_e(\theta) = \mathbf{W}\mathbf{f}_e^H \mathbf{S}_\theta^H \mathbf{W}\mathbf{f} \tag{3.25}$$

$$= \langle \mathbf{Wf}, \mathbf{S}_\theta \mathbf{W}\mathbf{f}_e \rangle, \tag{3.26}$$

$$\lambda_o(\theta) = \langle \mathbf{Wf}, \mathbf{S}_\theta \mathbf{W}\mathbf{f}_o \rangle. \tag{3.27}$$

The polynomial $p(\theta) = \lambda_e(\theta)^2 + \lambda_o(\theta)^2$ has degree $4N$. However, since $\mathbf{f}_e$ only has non-zero even orders and $\mathbf{f}_o$ only has non-zero odd orders, $p(\theta)$ will only have non-zero even coefficients. Therefore it can be written as a degree $2N$ trigonometric polynomial in 2θ,

$$p(2\theta) = \lambda_e(\theta)^2 + \lambda_o(\theta)^2. \tag{3.28}$$

Solving for θ therefore gives estimates in the range $[0, \pi)$. The solution method employed depends on the maximum order, N. When $N = 1$, the only possible values for W where $W_e = W_o$ are $w_0 = 1/\sqrt{2}$ and $w_1 = 1/2$. The resulting wavelets are the monogenic wavelets [124] and the sinusoidal model parameters can be derived analytically without root finding, as follows:

$$A = \sqrt{2}\|\mathbf{Wf}\|, \tag{3.29}$$

$$\phi = \arg(f_0 + \mathrm{i}\,|f_1|), \tag{3.30}$$

$$\theta = \arg(-\mathrm{i}\, f_1). \tag{3.31}$$

where $\phi \in [0, \pi)$ and $\theta \in [-\pi, \pi)$. Note that for a sinusoidal model, a rotation of π radians is equivalent to a sign change of the phase. For example, a sinusoid with $\{\phi = \pi/2, \theta = 0\}$ is equivalent to one with $\{\phi = -\pi/2, \theta = \pi\}$. Therefore two ranges for phase and orientation can be used interchangeably for the model: one can either restrict orientation to the half circle, or restrict phase to the half circle. To change between the two representations we have

$$\phi_{[0,\pi)} = |\phi_{[-\pi,\pi)}|, \tag{3.32}$$

$$\theta_{[-\pi,\pi)} = \begin{cases} \theta_{[0,\pi)} & \text{if } \phi_{[-\pi,\pi)} > 0, \\ \theta_{[0,\pi)} - \pi & \text{if } \phi_{[-\pi,\pi)} \leq 0, \end{cases} \tag{3.33}$$

and

$$\theta_{[0,\pi)} = \theta_{[-\pi,\pi)} \pmod{\pi}, \tag{3.34}$$

$$\phi_{[-\pi,\pi)} = \begin{cases} \phi_{[0,\pi)} & \text{if } \theta_{[-\pi,\pi)} \geq 0, \\ -\phi_{[0,\pi)} & \text{if } \theta_{[-\pi,\pi)} < 0. \end{cases} \tag{3.35}$$

When comparing model parameters, the interaction between phase and orientation should be taken into account. For example, an even sinusoid with $\phi = 0$, and $\theta = 0$ is the same as one with $\theta = \pi$. In these cases a double angle representation, such as $2\phi_{[0,\pi)}$ or $2\theta_{[0,\pi)}$, is useful.

For larger N, finding the maximum typically involves finding the roots of the derivative of $p(2\theta)$. The roots are candidate values for the orientation which corresponds to the maximum. For $N = 2$, the polynomial has degree 4 and the roots can be solved for analytically using a quartic solver. For larger orders a numerical solution is required in most cases. The resulting value is within the range $[0, \pi)$. The quick approximation method from Section 2.6.2 can also be used.

Once θ has been found, we have

$$A \cos\phi = \frac{\lambda_e(\theta)}{\sqrt{W_e}}, \tag{3.36}$$

$$A \sin\phi = \frac{\lambda_o(\theta)}{\sqrt{W_o}}, \tag{3.37}$$

and thus amplitude and phase are given by

$$A = \sqrt{\frac{\lambda_e(\theta)^2}{W_e} + \frac{\lambda_o(\theta)^2}{W_o}}, \tag{3.38}$$

$$\phi = \arg\left(\frac{\lambda_e(\theta)}{\sqrt{W_e}} + \mathrm{i}\frac{\lambda_o(\theta)}{\sqrt{W_o}}\right), \tag{3.39}$$

where $A \in \mathbb{R}^+$ and $\phi \in [-\pi, \pi)$. Finally, the residual vector is given by

$$\mathbf{W}\boldsymbol{\epsilon} = \mathbf{W}\mathbf{f} - \mathbf{W}\mathbf{f}_{\mathrm{S}}(A, \phi, \theta). \tag{3.40}$$

One of the advantages of using the CH wavelets, as opposed to other steerable basis functions, is that we may synthesise the image from the CH vector responses. Furthermore, we may split the image into two parts by synthesising separately from the model and residual coefficients, according to

$$f(\mathbf{z}) = f_{\mathrm{S}}(\mathbf{z}) + f_{\epsilon}(\mathbf{z}), \tag{3.41}$$

where

$$f_{\mathrm{S}}(\mathbf{z}) = \sum_{i,\mathbf{k}} \sum_{|n|\leq N} (\mathbf{W}\mathbf{f}_{\mathrm{S}_{\mathrm{i},\mathbf{k}}})_n w_n \psi^n_{i,\mathbf{k}}, \tag{3.42}$$

$$f_{\epsilon}(\mathbf{z}) = \sum_{i,\mathbf{k}} \sum_{|n|\leq N} (\mathbf{W}\boldsymbol{\epsilon}_{i,\mathbf{k}})_n w_n \psi^n_{i,\mathbf{k}}. \tag{3.43}$$

3.1.4 Example Solution

The sinusoidal model was calculated for the first four scales of the pyramidal decomposition of the *Pentagon* image (Figure 3.2). The CH wavelet frame used to obtain the CH vector was generated using a Simoncelli-type isotropic wavelet [101] and the 0th to 7th order RTs. Each scale was subsampled by two, and the odd and even orders were each equally weighted such that $W_e = W_o$ (3.60). The amplitude at each scale is high at the location of linear features, indicating these are well described by the sinusoidal model. The phase value describes the symmetry of the image structure, independently of the amplitude. The orientation shows the main axis of symmetry regardless of the phase or amplitude values. Thus the *split of identity* property [29] of phase-based image representations is preserved.

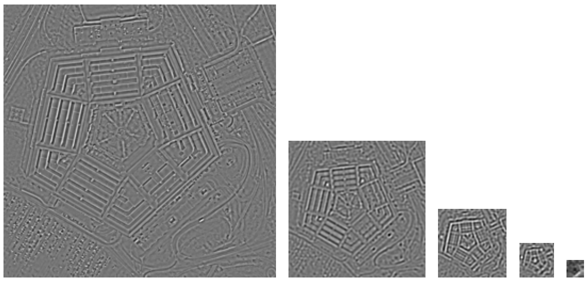
Reconstruction was performed using the model component for each scale (Figure 3.2g), as well as separate reconstruction from all the model (Figure 3.2h) and residual components (Figure 3.2i). Reconstruction from the model appears to act like a wide-band rotation-invariant line and edge filter. In contrast, the residual reconstruction contains features which have multiple axes of symmetry, such as corners and junctions, therefore these are not well represented by a sinusoidal model.

3.2 Method Parameters

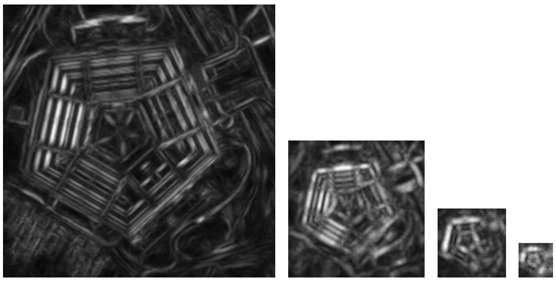
Three choices must be made when applying the model:

- The primary isotropic basis filter, $\psi(\omega)$, from which to construct the CH vector.
- The number of RT orders, N, in the CH vector.
- The values for the weights, $\mathbf{W}$, of the CH vector..

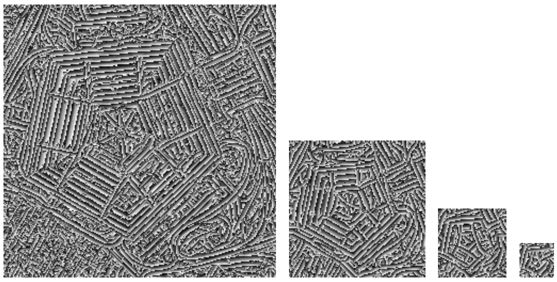
(a) Original Image



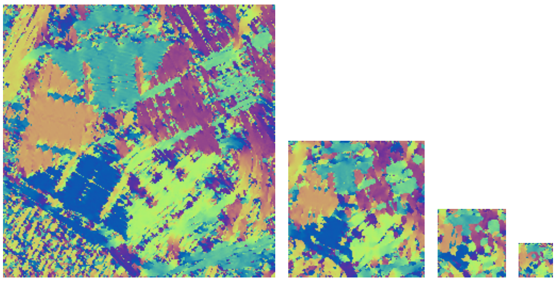
(b) Isometric Wavelet Response (4 scales + low-pass)



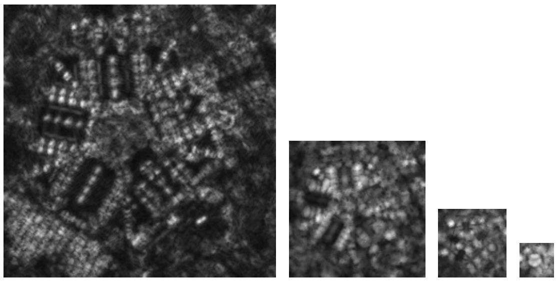
(c) Amplitude



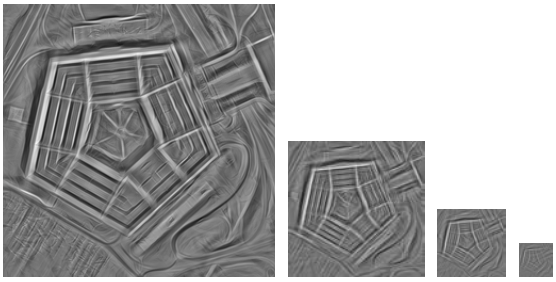
(d) Phase



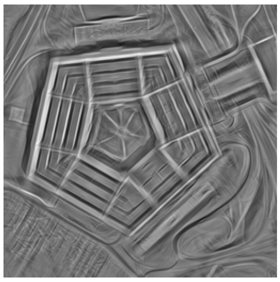
(e) Orientation



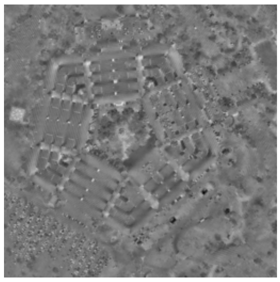
(f) Norm of residual



(g) Model Reconstruction for each scale



(h) Model Reconstruction



(i) Residual Reconstruction

Figure 3.2: Decomposition of a 256 $\times$ 256 pixel version of the *Pentagon* image into amplitude, phase, orientation and residual components over four scales using a pyramidal Meyer wavelet scheme and $N = 7$.

3.2.1 Choice of Basis Filter

To construct CH wavelet frames, the primary isotropic basis filter must have at least N vanishing moments [124]. Wavelets such as the Simoncelli [101], Papadakis [106], Meyer [21] and variance optimised wavelets (VOW) in 1D [95] and 2D [96] satisfy these conditions; however, they contain discontinuities in the frequency domain that lead to a slow decay in the spatial domain [131]. In contrast, a smooth wavelet such as the second Meyer wavelet in [21] may be a better choice.

For filter banks, the basis filter should also have the minimum number of vanishing moments and a smooth frequency profile to ensure fast decay. The log-Gabor filter [32] is often used for quadrature filters, as it is possible to construct large bandwidth frequency response with zero mean. In [11] it is shown that the difference-of-Gaussian (DoG) and Cauchy ($h(\omega) = n_c\,\omega^a e^{-\sigma\omega}$) quadrature (Hilbert transform) filters are better for edge detection [11]. However, the DoG filter has only one vanishing moment and a large minimum bandwidth, while the number of vanishing moments of the Cauchy filter is dependent on its bandwidth. The log-Gabor filter has infinite vanishing moments and thus remains an suitable choice for RT derived filters. An extension on the log-Gabor filter is

$$h(\omega) = \exp\left(-\left|\frac{\log^a(\omega/\omega_0)}{a\log^a(\sigma)}\right|\right), \tag{3.44}$$

where increasing a gives a more compact frequency response and shorter tail. The normal log-Gabor filter is given by $a = 2$. The -3dB bandwidth in octaves is

$$\beta = 2\log(\sigma)\sqrt[a]{a\log(2)}/\sqrt{\log 2}. \tag{3.45}$$

Figure 3.3 shows the formulas, frequency response and sinusoidal wavelet constructed for different basis filters. It can be observed that the filters with sharp transitions have more oscillations further from the centre. The log-Gabor filter is the smoothest due to its long tail.

3.2.2 Effect of N

Higher-order CH wavelets have a higher order of rotational symmetry. Therefore, increasing N increases the complexity of the local signal structure that the CH vector can represent. However, higher-order CH wavelets also have a larger spatial extent, increasing the size of the local image patch under consideration. This is because the magnitude of the radial frequency response remains constant due to the RT.

The model wavelets thus also increase in size with increasing N. An example of the two types of sinusoidal model wavelets for different values of N is shown in Figure 3.4. The even monogenic wavelet ($N = 1$) has no directionality, hence the problem with resolving orientation near even structures. As N increases, the wavelets become elongated along the axis perpendicular to their orientation, becoming more orientation selective due to a narrower angular profile but only

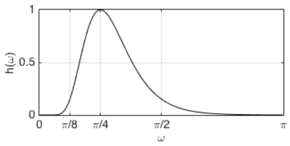
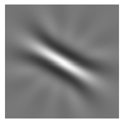
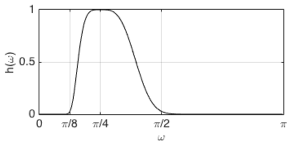
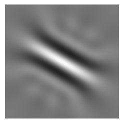
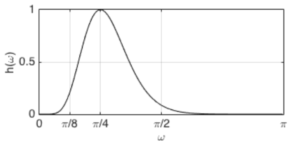
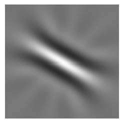
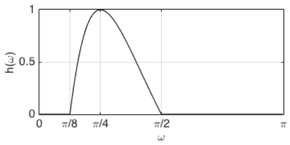
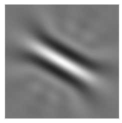
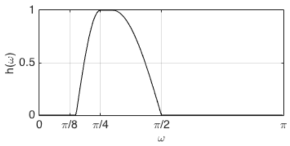
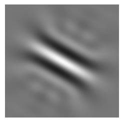
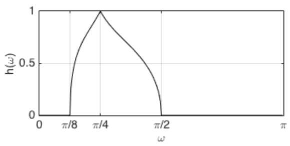
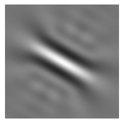
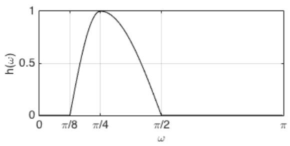
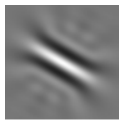
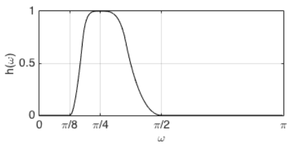
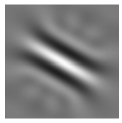
Filter	$h(\omega)$	Spectrum	Wavelet
log-Gabor (original)	$\exp\left(-\frac{\log^2(\omega/\omega_0)}{2\log^2(\sigma)}\right)$		
log-Gabor (extended)	$\exp\left(-\left\lvert\frac{\log^a(\omega/\omega_0)}{a\log^a(\sigma)}\right\rvert\right)$		
Cauchy	$n_c\,\omega^a e^{-a\frac{\omega}{\omega_0}}$ where $n_c = \frac{1}{\omega_0^a e^{-a}}$		
Simoncelli	$\begin{cases} \cos(\frac{\pi}{2}\log_2(\frac{\omega}{\omega_0})) & \frac{\omega_0}{2} \le \omega < 2\omega_0 \\ 0 & \text{otherwise} \end{cases}$		
Papadakis	$\begin{cases} \sqrt{\frac{1-\cos(10\omega-3\pi/2)}{2}} & \frac{3}{5}\omega_0 \le \omega < \omega_0 \\ 1 & \omega_0 \le \omega < \frac{6}{5}\omega_0 \\ \sqrt{\frac{1+\cos(5\omega-3\pi/2)}{2}} & \frac{6}{5}\omega_0 \le \omega < 2\omega_0 \\ 0 & \text{otherwise} \end{cases}$		
VOW	$\begin{cases} \left(\frac{1}{2} + \frac{\tan(\frac{3}{4}+\frac{2}{3}\log_2(\frac{\omega}{\omega_0}))}{2\tan(\frac{3}{4})}\right)^{\frac{1}{2}} & \frac{\omega_0}{2} \le \omega < \omega_0 \\ \left(\frac{1}{2} - \frac{\tan(\frac{3}{4}+\frac{2}{3}\log_2(\frac{\omega}{2\omega_0}))}{2\tan(\frac{3}{4})}\right)^{\frac{1}{2}} & \omega_0 \le \omega < 2\omega_0 \\ 0 & \text{otherwise} \end{cases}$		
Meyer (v1)	$\begin{cases} \sin(\pi\frac{\omega}{\omega_0} - \frac{\pi}{2}) & \frac{\omega_0}{2} \le \omega < \omega_0 \\ \cos(\frac{\pi}{2}\frac{\omega}{\omega_0} - \frac{\pi}{2}) & \omega_0 \le \omega < 2\omega_0 \\ 0 & \text{otherwise} \end{cases}$		
Meyer (v2)	$\begin{cases} 2\sqrt{2}\log_2^2(\frac{2\omega}{\omega_0}) & \frac{\omega_0}{2} \le \omega < \frac{\omega_0}{\sqrt{2}} \\ \sqrt{1-8\log_2^4(\frac{\omega}{\omega_0})} & \frac{\omega_0}{\sqrt{2}} \le \omega < \sqrt{2}\omega_0 \\ 2\sqrt{2}(2-\log_2(\frac{2\omega}{\omega_0}))^2 & \sqrt{2}\omega_0 \le \omega < 2\omega_0 \\ 0 & \text{otherwise} \end{cases}$		

Figure 3.3: Different filters (top section) and wavelets (bottom section) along with their formulas. A plot of the radial frequency response is shown for wavelength 8 pixels ($\omega_0 = \pi/4$) along with the even sinusoidal wavelet for $N = 13$. Parameters used were $\sigma = 0.7$ for the log-Gabor filter, $\sigma = 0.7$ and $a = 4$ for the log-Gabor (extended) filter, $a = 8$ for the Cauchy filter.

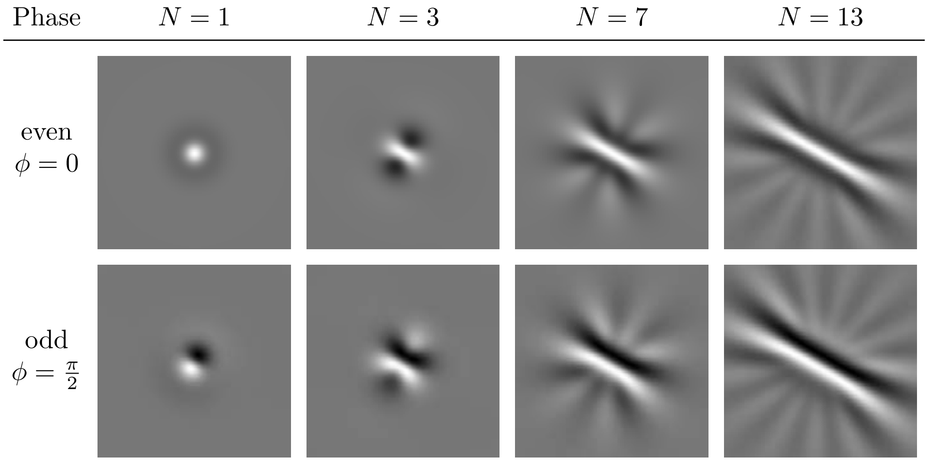


Figure 3.4: Sinusoidal matched wavelets for different N, phase, ϕ, and $\theta = \pi/3$. The top row shows the even wavelets, and the bottom row shows the odd wavelets. Different phases are obtained by a linear combination of both. For $N = 1$ the even wavelet has no directionality.

responding to longer linear features.

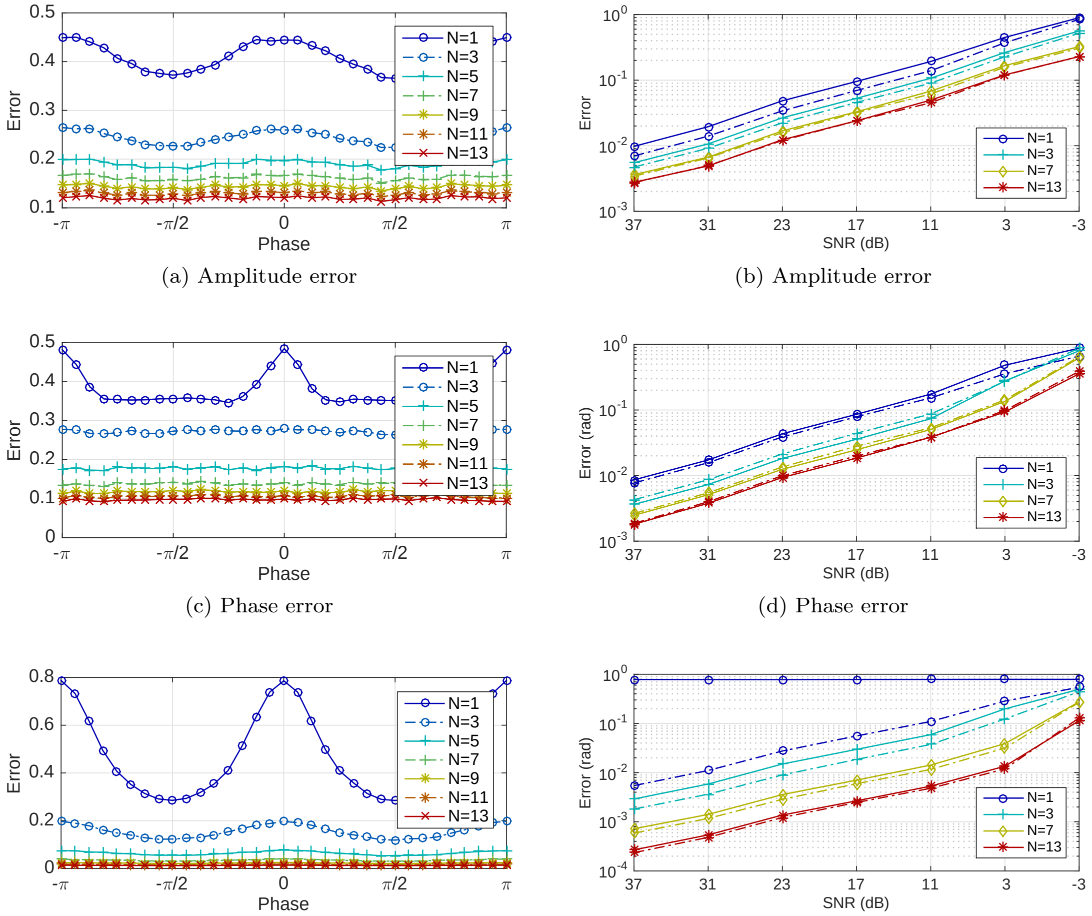
Noise

To quantify the effect of increasing N on sinusoidal model accuracy, the amplitude, phase and orientation were calculated for a zero-mean, 512×512 pixel sinusoidal image with different levels of additive white Gaussian noise. An all-pass basis filter was used and the CH vector was weighted using the phase-invariant equal weighting scheme (3.60). The mean error in estimated model parameters was compared to the phase of the sinusoid for 3dB signal-to-noise ratio (SNR) (Figure 3.5). Increasing N decreased the average error for all parameters. However, both the amplitude and phase errors appear to reach a plateau around $N = 13$, after which increasing N gives little improvement. The orientation estimate though was particularly improved, with a ten times reduction in error between $N = 3$ and $N = 13$. As expected, the orientation error is high for the monogenic signal ($N = 1$) at even locations, regardless of noise. For a SNR greater than 3dB, the errors vary proportionally with the noise standard deviation.

Qualitative Image Results

The effect of increasing N for was compared for the *Pentagon* image (Figure 3.6) with the sinusoidal model calculated using a log-Gabor primary filter with wavelength 8 pixels, $\sigma = 0.65$, and $N \in \{1, 3, 7, 13, 21\}$. For the monogenic signal ($N = 1$), the amplitude is large and the phase is the same for both isometric features (blobs) and lines, and thus they cannot be differentiated from the model parameters alone. As N increases, the model becomes more selective for longer linear features due to the increasing elongation of the model wavelets. This is particularly noticeable going from $N = 7$ to $N = 21$, as the roof edges are no longer broken up. Blobs also have a reduced amplitude response; those in the lower right quadrant have almost disappeared by $N = 13$ and instead the amplitude is large only at the location of linear features. Likewise, the residual norm

Error vs Phase	Error vs Noise



(a) Amplitude error

(b) Amplitude error

(c) Phase error

(d) Phase error

(e) Orientation error

(f) Orientation error

Figure 3.5: Average model error for a sinusoidal image with added Gaussian noise (SNR: 3dB) for different N compared to the original image phase (first column). Average model error for $\phi = 0$ (solid lines) and $\phi = \pi/2$ (dashed lines) for different amounts of noise (second column). The phase-invariant equal weighting scheme was used.

is large at the location of corners and junctions which have multiple linear symmetries.

The orientation estimate also becomes smoother with increased N, and eventually larger features begin to dominate. In contrast, the orientations of curved lines appear to become more disjoint. The increased size of the wavelet, and thus local image patch, means curves are less well modelled by a sinusoid at larger N. The residual norm image confirms this, showing a larger magnitude for curved structures as N increases. Overall, less of the image is well described by the sinusoidal model with larger N, and the residual norm increases for most locations. The effect can be seen in the separate reconstructions from the model and residual components, with more of the structure identifiable in the residual reconstruction. A qualitative assessment suggests that for this image, $N = 7$ provides a good balance between resolution of linear features and too much energy in the residual component.

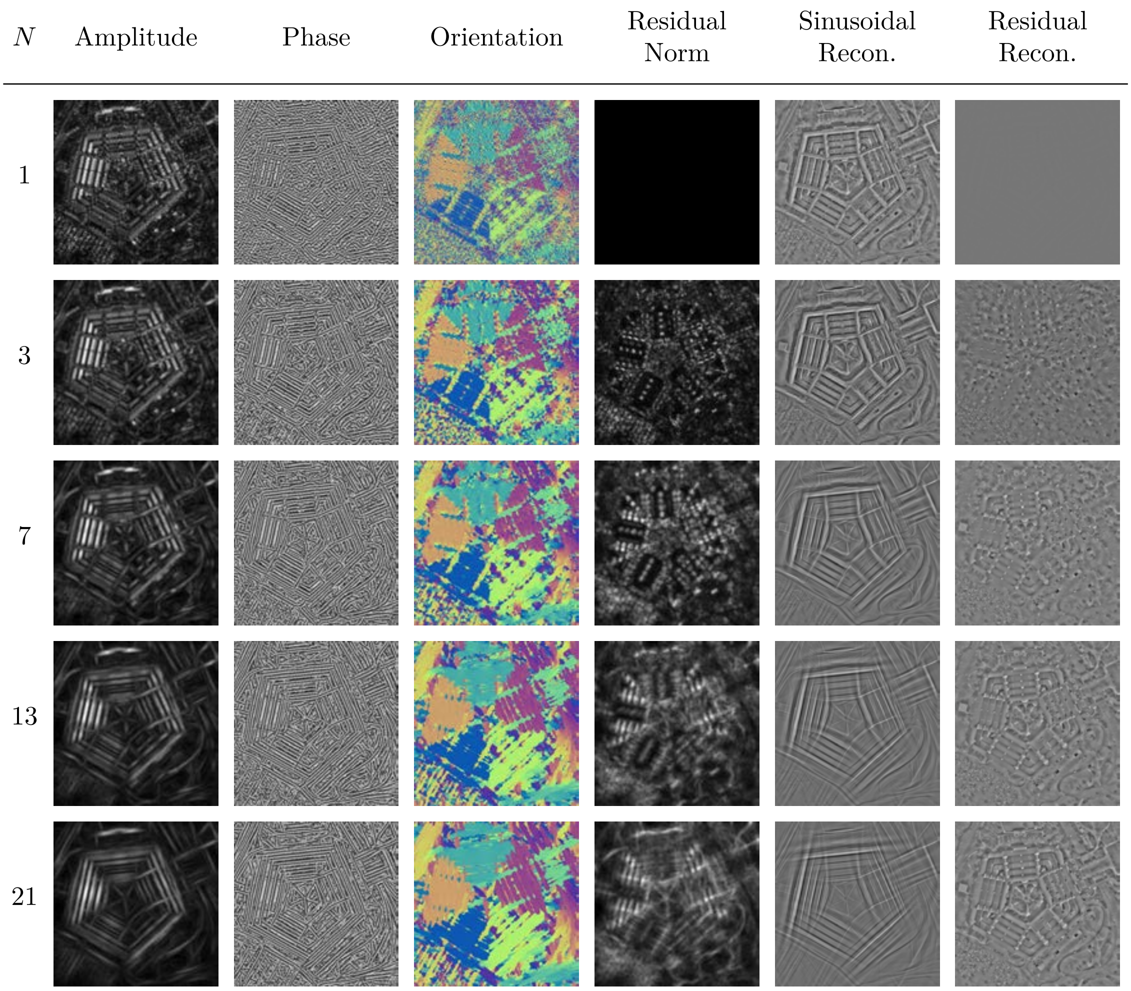


Figure 3.6: Sinusoidal model parameters for the second scale of the *Pentagon* image for different values of N. Also shown is reconstruction from the sinusoidal model, and reconstruction from the residual component, using four scales and not including the low-pass response.

Computation Time

Since the orientation is obtained using a polynomial, $p(2\theta)$, with degree 4N, the computation times given in Table 2.2 apply to the sinusoidal model.

3.2.3 Choosing Weights

The weighting matrix scales each RT order in the image CH vector, and therefore different weightings affect the CH vector magnitude and values of the model parameters for a given image structure. The choice of weights determines the angular profile of the sinusoidal model wavelets and the phase-invariance of the CH vector magnitude.

The magnitude of the CH vector of a purely sinusoidal image is given by

$$\|\mathbf{W}\mathbf{f}_{\mathrm{S}}(A,\phi,\theta)\| = A\|\cos\phi\mathbf{W}\mathbf{S}_\theta\,\mathbf{s}_e + \sin\phi\mathbf{W}\mathbf{S}_\theta\,\mathbf{s}_o\| \tag{3.46}$$

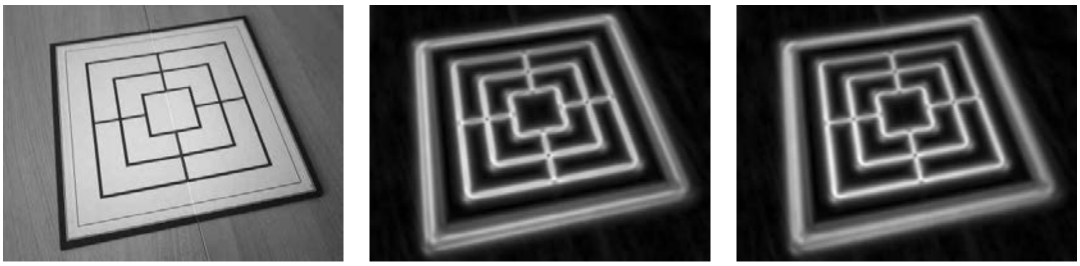
$$= A\sqrt{\cos^2\phi\,W_e + \sin^2\phi\,W_o} \tag{3.47}$$

$$= \begin{cases} A\sqrt{W_e}, & \phi = 0,\pi \\ A\sqrt{W_o}, & \phi = \pm\pi/2. \end{cases} \tag{3.48}$$

If $W_e \neq W_o$ the magnitude of the CH vector is affected by the phase. If we choose $W_e = W_o = 1/2$, the odd and even components are each weighted equally and the magnitude is invariant to phase. That is,

$$\|\mathbf{W}\mathbf{f}_{\mathrm{S}}(A,\phi,\theta)\| = A/\sqrt{2}. \tag{3.49}$$

This is desirable as it preserves the invariance properties of previous approaches. For example, the monogenic signal vector magnitude is invariant to phase. If instead each individual order is weighted equally, then $W_e \neq W_o$, and the effect is subtle but noticeable as an increase in magnitude at the location of odd features for odd N, and even features for even N, although the effect diminishes with increasing N (Figure 3.7).



(a) *Board* image (b) Each order weighted equally (c) Phase-invariant weighting, $W_e = W_o$

Figure 3.7: CH vector magnitude of the second scale of the *Board* image (a) using all orders equally weighted (b) and odd and even orders equally weighted, for $N = 3$.

The weighting also determines the angular response of the model wavelets in the frequency domain, which is given by the trigonometric polynomial

$$h_u(\theta) = \sum_{|n|\leq N} w_n u_n e^{-\mathrm{i}n\theta}, \tag{3.50}$$

where $\mathbf{Wu}$ is the weighted model wavelet CH vector. For the sinusoidal model wavelets to be useful to estimate the orientation of even structures in $\theta = [0,\pi)$ then at least w_0 and w_2 should be non-zero. Likewise, to estimate orientation in odd structures over $\theta = [0, 2\pi)$ then at least w_1 should be non zero.

To pick the coefficients of $\mathbf{W}$ it is proposed to maximise the energy of the angular response of the sinusoidal wavelets inside of a window $h(\theta)$ by adapting the method described in [103, 105, 124] for designing *prolate spheroidal wavelets*. Each sinusoidal wavelet is 2nd order rotationally symmetric or anti-symmetric and therefore has its angular response concentrated at two points π radians apart. The same energy window can be used for each. Let $v(2\theta)$ be a positive window function, symmetric at both $\theta = 0$ and $\theta = \pi$. That is,

$$v(\theta) = v(-\theta), \tag{3.51}$$

$$v(\theta) = v(\theta - \pi), \tag{3.52}$$

and let $u(\theta)$ describe the angular response of a wavelet, $\mathbf{u}$, that is

$$u(\theta) = \sum_{n=-N}^{N} u_n u_n e^{-\mathrm{i}n\theta}. \tag{3.53}$$

Then the energy within the window is given by [103, 105, 124]

$$\begin{aligned} E[w] &= \int_{-\pi}^{\pi} u(\theta)^2 v(\theta)\, d\theta & (3.54)\\ &= \sum_{n'=-N}^{N} \sum_{n=-N}^{N} \bar{u}_{n'} u_n \int_{-\pi}^{\pi} v(\theta)\, d\theta & (3.55)\\ &= \mathbf{u}^H \mathbf{V} \mathbf{u}, & (3.56) \end{aligned}$$

where $V_{n,n'} = \int_{-\pi}^{\pi} e^{\mathrm{i}(n-n')\theta} v(\theta)\, d\theta$ since $\mathbf{u}^H\mathbf{u} = 1$.

Two types of orthogonal symmetric functions that fit this window are an even function with extrema of the same sign at 0 and π, and an odd function with extrema of opposite signs at 0 and π, where $u(\theta) = -u(\theta - \pi)$. The eigenvectors corresponding to the largest two eigenvalues of $\mathbf{V}$ thus describe the even and odd wavelets in the sinusoidal model. Let $\mathbf{u}_1$ and $\mathbf{u}_2$ be these eigenvectors. Each either has only odd orders, or only even orders. The final weighting is given by the absolute value of each order scaled by $\sqrt{2}$, since both are of unit norm. That is,

$$w_n = \frac{|u_{1_n} + u_{2_n}|}{\sqrt{2}}, \tag{3.57}$$

and thus $\|\mathbf{w}\| = 1$ and the weighting matrix is thus $\mathbf{W} = \mathrm{diag}(\mathbf{w})$.

A simple window function consists of two rectangular functions with angular width B separated by π,

$$v(\theta) = \mathrm{rect}\left(\frac{\theta}{B}\right) + \mathrm{rect}\left(\frac{\theta + \pi}{B}\right). \tag{3.58}$$

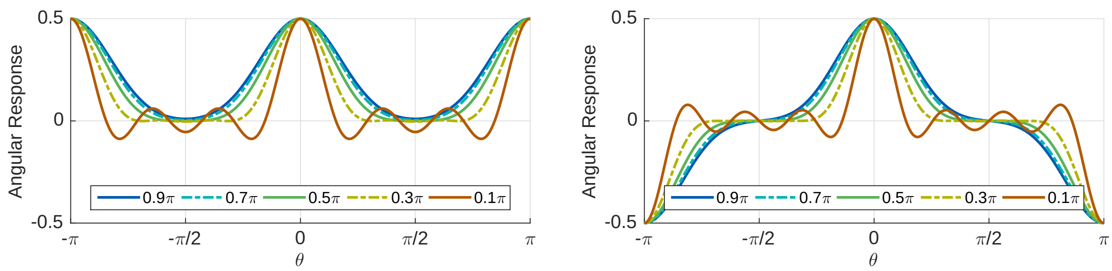
The values of $\mathbf{V}$ for this function are

$$V_{n,n'} = \begin{cases} 2B, & n - n' = 0, \\ \frac{2\sin(B(n-n'))}{n-n'}, & n - n' \text{ is even}, \\ 0, & n - n' \text{ is odd}, \end{cases} \tag{3.59}$$

where B is the width of the rectangle in radians. When B approaches 0 the set of even components become each equally weighted, as do the odd, so that $W_e = W_o$. This equal weighting scheme is given by

$$w_n = \begin{cases} \frac{1}{\sqrt{2(N+1)}} & \text{if } (N-n) \text{ is even} \\ \frac{1}{\sqrt{2N}} & \text{if } (N-n) \text{ is odd}. \end{cases} \tag{3.60}$$

Figure 3.8 shows an example of even and odd angular profiles for different values of B and $N = 7$. For smaller B, the angular response has a narrower peak but larger oscillations. For larger B, the response is smoother and wider, but less orientation selective. To quantify the amount of oscillation, an experiment was performed to measure the ratio of the energy under the side lobes to the total energy, for different values of N and B. The ratio was less than 0.1% (indicating small oscillations) when $B > 5.64/N - 6.57/N^2$. For $N = 7$ this equals approximately 0.21π. The quantity $5.64/N - 6.57/N^2$ will be referred to as $B_{0.1}$ later in this thesis, and the window width can be adjusted as a multiple of this value in order normalise the design of the wavelets across different values of N.



(a) Even wavelet ($\phi = 0$) (b) Odd wavelet ($\phi = \pi/2$)

Figure 3.8: Angular response of odd and even sinusoidal model wavelets ($N = 7$) for different window widths, B (shown in legend).

An example of the even and odd sinusoidal wavelets with B as a multiple of $B_{0.1}$ is shown in Figure 3.9. As the width is increased, the angular variations in the wavelet shape also decrease, however, the wavelets are also less elongated and therefore less orientation selective.

Repeating the noisy sinusoidal image experiment with different values for B it was found that the equal weighting scheme ($B \approxeq 0$) gave the best results (Figure 3.10). The error increase is likely due the narrowness of the main lobe combined with the even distribution of noise energy in the side lobes negating their effect on the response. However, in natural images the local

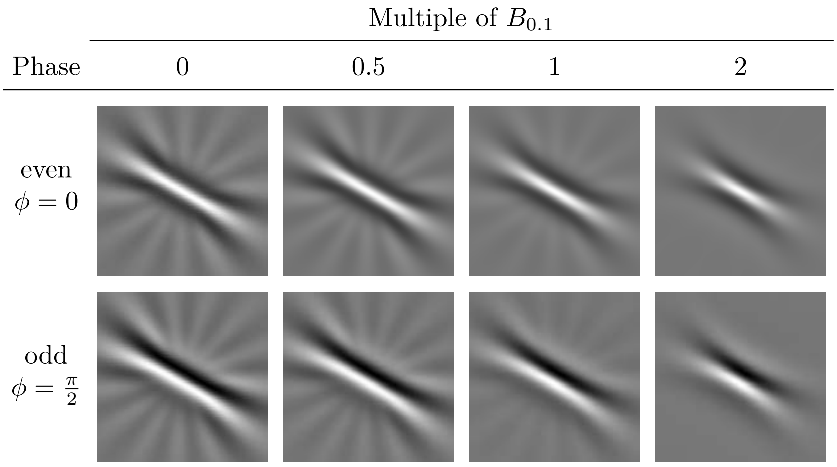
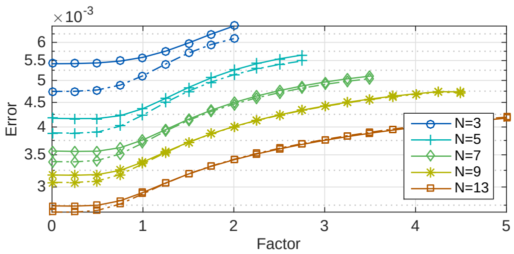
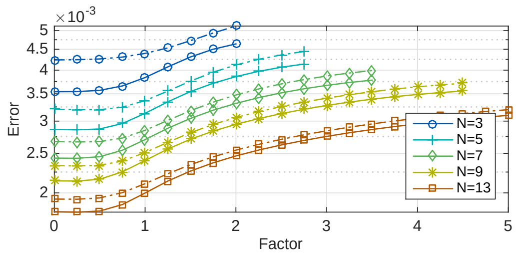


Figure 3.9: Effect of weights chosen using the windowed energy minimisation method on the odd and even sinusoidal wavelets for $N = 13$. The window width is shown as a multiple of $B_{0.1}$.

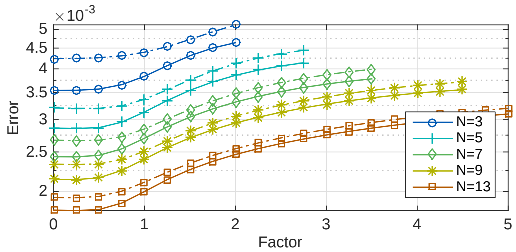
image structure can have multiple discrete elements. In that case, it may be advantageous to have smaller oscillations (higher value of B) so that the extra parts interfere less with the main response. Furthermore, if one wished to model using multiple sinusoids, a larger B may be helpful to reduce the correlation between the model wavelets at different orientations.



(a) Amplitude error



(b) Phase error



(c) Orientation error

Figure 3.10: Average model error for a sinusoidal image with a small amount of added Gaussian noise (SNR: 37dB) for different N and different weightings obtained using the windowed energy minimisation method with B as a multiple of $B_{0.1}$.

3.3 Intrinsic Dimension

The advantage of a sinusoidal model is the parametrisation of the local image structure into amplitude, phase and orientation values, which can be analysed separately. Deriving the model starting with the CH vector is different to other quadrature filter type methods in that we are left with a residual vector that describes the non-model part of the local image structure. From the split into model and residual components a representation of the local *intrinsic dimension* can be developed.

Intrinsic dimension describes the linear symmetry of the local image structure. Flat areas with constant intensity are intrinsically 0D (i0D) as they can be described by a single value. Linear features, such as lines and edges, are intrinsically 1D (i1D), as they vary along a single axis and can be represented by a 1D function. More complex structures, such as corners and junctions, have multiple symmetries and are intrinsically 2D (i2D) [29, 61, 142, 145]. The strict definition is,

$$f(\mathbf{x}) \in \begin{cases} \{\text{i0D}\} & \text{if } f(\mathbf{z}) = \text{constant}, \\ \{\text{i1D}\} & \text{if } f(\mathbf{z}) = f(\langle \mathbf{z}, \mathbf{o} \rangle), \mathbf{o} = [\cos\theta, \sin\theta], \\ \{\text{i2D}\} & \text{otherwise}. \end{cases} \tag{3.61}$$

However, typical images do not have perfectly linear structures and are often corrupted by noise, meaning that most structures would be classified as i2D. Instead, a continuous representation of intrinsic dimensionality is necessary, which in turn requires a local structure descriptor that is able to discriminate classes. A classic example is the structure tensor [9, 35] whose eigenvalues describe the strength of the local symmetry along two main axes. If only one eigenvalue is large, the local structure is i1D, if both are large it is i2D, otherwise it is i0D. This the basis behind the popular Harris corner detector [46]. In [28] a continuous representation was introduced that plots the eigenvalues onto a bounded triangle, whose barycentric coordinates give a probability of belonging to each class. The representation can be applied to the output of other descriptors that discriminate between i1D and i2D structures.

A drawback of the structure tensor is that it is only computed from 1st-order derivatives and therefore can have a double response for roof edges (thick lines) [56]. A improvement proposed was the boundary tensor [56], which gives a boundary energy value consisting of line / edge (i1D) and junction (i2D) energies, using the 0th to 2nd-order RTs. In fact, the boundary energy is equal to the square of the CH vector norm for $N = 2$ with weighting $\mathbf{w} = \sqrt{[1/2, 1, 1, 1, 1/2]/4}$. Likewise, the boundary tensor i1D energy measure roughly corresponds to the energy of the sinusoidal model vector, and the boundary tensor junction energy to the energy of the residual vector for the same weighting and N. The monogenic curvature tensor [112] uses the same basis functions as the boundary tensor but adds phase and curvature measurements to give a richer description of the local structure.

The sinusoidal model appeared more descriptive of i1D structures and the residual component was higher around i2D locations for larger N for the *Pentagon* image (Figure 3.6). This suggests the model and residual components can be used to discriminate between these classes. Therefore we shall use the proposed sinusoidal model calculated using higher-order RTs to develop a representation of intrinsic dimension.

3.3.1 Model Response

The first step is to determine the sinusoidal model response to an i1D structure. Consider an image, f, that is locally i1D at a point of interest centred at $\mathbf{z} = \mathbf{0}$ when filtered by an isotropic wavelet, ψ. The local structure can be represented as 1D function,

$$(f * \psi)(\mathbf{z}) = f_{\mathrm{i1D}}(x), \tag{3.62}$$

where $x = \langle \mathbf{z}, \mathbf{o} \rangle$ and $\mathbf{o} = [\cos\theta, \sin\theta]$ with θ being the orientation of symmetry. According to the Fourier slice theorem, the Fourier transform of the image patch will have all non-zero coefficients concentrated along a line through the origin. Therefore, the local image structure can be exactly modelled as a sum of sinusoids,

$$(f * \psi)(\mathbf{z}) = \sum_k \alpha_k \cos(\omega_k x + \phi_k), \tag{3.63}$$

which can also be expressed as a single sinusoid,

$$(f * \psi) = A(x) \cos\left(\phi(x)\right), \tag{3.64}$$

which is equivalent to the analytic signal representation of $f_{\mathrm{i1D}}(x)$, with local amplitude, A, and phase, ϕ. It follows that the overall RT responses are given by the sum of the RT responses of the individual sinusoids, and thus

$$\mathcal{R}^n(f * \psi)(\mathbf{z}) = \begin{cases} A(x)e^{\mathrm{i}n\theta}\cos(\phi(x)) & n \text{ is even,} \\ A(x)e^{\mathrm{i}n\theta}\,\mathrm{i}\sin(\phi(x)) & n \text{ is odd.} \end{cases} \tag{3.65}$$

The magnitudes of all the even order responses are equal, as are all the odd order responses. Let $\mathbf{f}_{\mathrm{i1D}}$ be the CH vector generated for this image structure at $\mathbf{z} = \mathbf{0}$, using ψ as the primary isotropic wavelet. The structure is completely described by the sinusoidal model, that is,

$$\mathbf{f}_{\mathrm{i1D}} = \mathbf{f}_{\mathrm{S}}(A, \phi, \theta) \tag{3.66}$$

$$= A\mathbf{S}_\theta(\cos\phi\,\mathbf{s}_e + \sin\phi\,\mathbf{s}_o). \tag{3.67}$$

It follows that the magnitude of the residual component is zero

$$\min_{A,\phi,\theta} \|\boldsymbol{\epsilon}\| = \min_{A,\phi,\theta} \|\mathbf{f}_{\mathrm{i1D}} - \mathbf{f}_{\mathrm{S}}\| \tag{3.68}$$

$$= 0, \tag{3.69}$$

which means that an i1D signal can be completely reconstructed from the sinusoidal model wavelets rotated to the same orientation.

3.3.2 Complex Exponential Representation

Since the sinusoidal model can completely represent an i1D signal, it follows that the residual component represents the other parts of the local structure with a different shape or orientation. The ratio of residual vector norm to the model vector norm is therefore a measure of where the overall local structure lies on a i1D to i2D scale. Furthermore, it is invariant to the magnitude of the local structure (CH vector norm). We can represent this relationship in the form of a complex exponential, $\mathbf{d}_0$, given by

$$\mathbf{d}_0 = \|\mathbf{W}\mathbf{f}_\mathrm{S}\| + \mathrm{i}\|\mathbf{W}\boldsymbol{\epsilon}\| \tag{3.70}$$

$$= \|\mathbf{W}\mathbf{f}\| e^{\mathrm{i}\gamma_0}, \tag{3.71}$$

where $\gamma_0 = \tan^{-1} \frac{\|\mathbf{W}\boldsymbol{\epsilon}\|}{\|\mathbf{W}\mathbf{f}_\mathrm{S}\|}$ is the angle representing the ratio between $\|\mathbf{W}\boldsymbol{\epsilon}\|$ and $\|\mathbf{W}\mathbf{f}_\mathrm{S}\|$.

While it is possible to have a zero residual response, it is not possible to have a zero sinusoidal model response, as the model wavelets will always positively correlate with some part of the local image structure. This means that the upper bound of possible values of γ_0 will always be less than $\pi/2$, and can change according to the number of orders and weighting scheme. For a purely sinusoidal signal, we have $\mathbf{W}\mathbf{f} = \mathbf{W}\mathbf{f}_\mathrm{S}$ and $\|\mathbf{W}\boldsymbol{\epsilon}\| = 0$, and therefore $\gamma_0 = 0$ is the lower bound.

By finding this upper bound, we can adjust γ_0 so that the range is always $[0, \pi/2)$ regardless of weighting, and by extension, N. Consider an image CH vector, $\mathbf{f}$, that is zero for every order except for n, and without loss of generality let $f_{|n|} = 1$. Calculating the sinusoidal signal model for a weighting scheme, $\mathbf{W}$, we obtain

$$\|\mathbf{W}\mathbf{f}_\mathrm{S}\| = \begin{cases} 2w_n^2/\sqrt{W_e} & n \neq 0, \text{even} \\ 2w_n^2/\sqrt{W_o} & n \neq 0, \text{odd} \\ w_n^2/\sqrt{W_e} & n = 0. \end{cases} \tag{3.72}$$

Since $\|\mathbf{W}\boldsymbol{\epsilon}\| = \sqrt{\|\mathbf{W}\mathbf{f}\|^2 - \|\mathbf{W}\mathbf{f}_\mathrm{S}\|^2}$,

$$\gamma_0 = \begin{cases} \tan^{-1}\sqrt{\frac{W_e}{2w_n^2} - 1} & n \neq 0, \text{even} \\ \tan^{-1}\sqrt{\frac{W_o}{2w_n^2} - 1} & n \neq 0, \text{odd} \\ \tan^{-1}\sqrt{\frac{W_e}{w_n^2} - 1} & n = 0. \end{cases} \tag{3.73}$$

The upper bound, $\gamma_{\max}$, is the maximum value of γ_0 in the above equation for all n. Note that as N increases, w_n tends to get smaller, and therefore $\gamma_{\max}$ gets closer to $\pi/2$. The rescaled intrinsic dimension representation is thus

$$\mathbf{d}_1 = \|\mathbf{W}\mathbf{f}\| e^{\mathrm{i}\gamma_1} \tag{3.74}$$

$$\text{where } \gamma_1 = \frac{\gamma_0}{\gamma_{\max}} \frac{\pi}{2}. \tag{3.75}$$

3.3.3 i2D Detection

The norm of the residual component is large around corners and junctions (Figure 3.6). Thus the next step is to use the i2D part of the intrinsic dimension representation as a corner and junction detector, in the same way the junction energy is used for the boundary tensor [56]. The proposed detection measure is thus the imaginary part of the intrinsic dimension:

$$\begin{aligned}\det_{\mathrm{i2D}}(\mathbf{d}) &= \Im\{\mathbf{d}\} &(3.76)\\ &= \|\mathbf{Wf}\|\sin(\gamma). &(3.77)\end{aligned}$$

However, common i2D features can have a large i1D component. For example, a T junction will give a large sinusoid amplitude representing the top bar that a 'Y' junction will not. The angle, γ_1, will therefore be different at the centre of these two junctions. To compensate we shall rescale γ_1 using a sigmoidal function so that the angle values are closer together. The new representation is

$$\mathbf{d}_2 = \|\mathbf{Wf}\|e^{\mathrm{i}\gamma_2}, \tag{3.78}$$

where $\gamma_2 = \pi/2\eta(2\gamma/\pi, h, s)$ which is a sigmoidal function given by the regularised incomplete gamma function as follows

$$\eta(x, h, s) = \begin{cases} I_\gamma(x, s, s/h - s)) & h \leq 0.5 \\ 1 - \eta(1 - x, 1 - h, s) & h > 0.5 \end{cases} \tag{3.79}$$

with $x \in [0, 1]$ and $h \in [0, 1]$. The gamma function was chosen because the slope and position of the curve can be easily manipulated and the output values cover the complete range from 0 to 1. In the above equation, h roughly corresponds to the halfway point, that is, for $x = h$, $g(x, h, s) = 0.5$. Increasing the value of s increases the steepness of the slope of the function at this point (Figure 3.11).

Rescaling γ_1 improves the detection location for a slanted chequer feature (Figure 3.12). Because the centre of the feature is less i2D than the surrounding areas it results in an off-centre detection. After scaling γ_1, the energy is more centred and thus the detected location is brought closer to its true position.

3.3.4 Corner and Junction Response

The intrinsic dimension representation, $\mathbf{d}_2$, was calculated for 15 corner and junction features using $N \in \{2, 3, 7, 13\}$, and sigmoidal function parameters $h = 1/3$ and $s = 2.4$, and is shown in Figure 3.13. A depiction of intrinsic dimension using the structure tensor is given for comparison, using the largest eigenvalue as the i1D component, the smallest as the i2D component, and $\gamma_{\max} = \pi/4$.

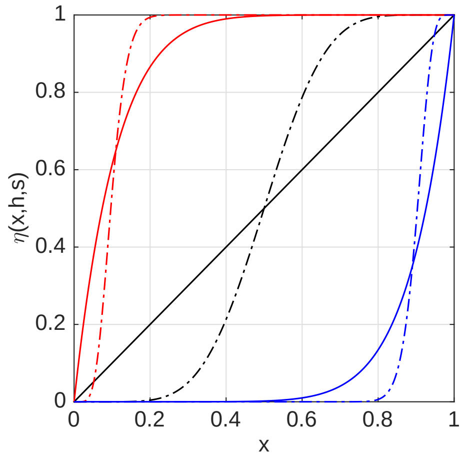


Figure 3.11: Sigmoidal function created from the incomplete gamma function for halfway values (h) of 0.1 (red), 0.5 (black) and 0.9 (blue) and slope values (s) of 1 (solid line) and 8 (dotted line).

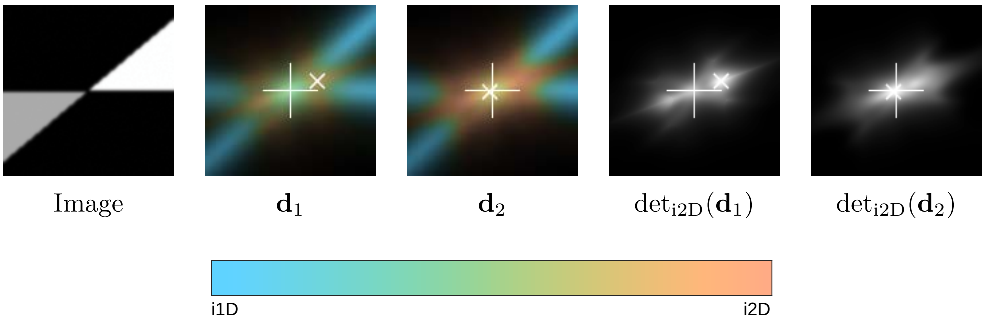


Figure 3.12: Intrinsic dimension representation and i2D detection measure before and after rescaling γ_1 for $N = 7$. In the 2nd and 3rd images, brightness corresponds to $\|\mathbf{d}\|$ and colour to γ. The actual feature centre is indicated by a plus; the maximum of the detection measure is indicated by a cross.

The images are 128 x 128 pixels in size, and were constructed from either line or wedge segments radiating from the centre pixel, followed by some Gaussian blurring. A log-Gabor filter with wavelength 32 pixels and $\sigma = 0.6$ was used to localise the model response, and a Gaussian filter with $\sigma = 6$ was used for the structure tensor. Brightness is equal to $\|\mathbf{d}_2\|$ which is the same as the norm of the image CH vector. Colour describes γ_2, with blue indicating i1D and red indicating i2D. An iso-luminant colour map from [60] was used to ensure correct perception.

The magnitude of the responses appear phase-invariant, as both line and edge features have similar patterns of magnitude and intrinsic dimension angle. It can be observed that for smaller N both the magnitude and the extent of the i2D region is concentrated more towards the centre of the features, due to the smaller size of the wavelets. However, for smaller N there are also regions of low magnitude near the centre of features with more than two segments. This is particularly visible in the fifth and sixth images for $N = 2$. As N increases, the magnitude becomes more uniform, showing that a larger N is required for boundary estimation of complex features. However, this also causes a smearing of the magnitude response along the direction of line and edge segments, particularly noticeable in the seventh image as well as in Figure 3.2. Again this is due to the increased size of the wavelets.

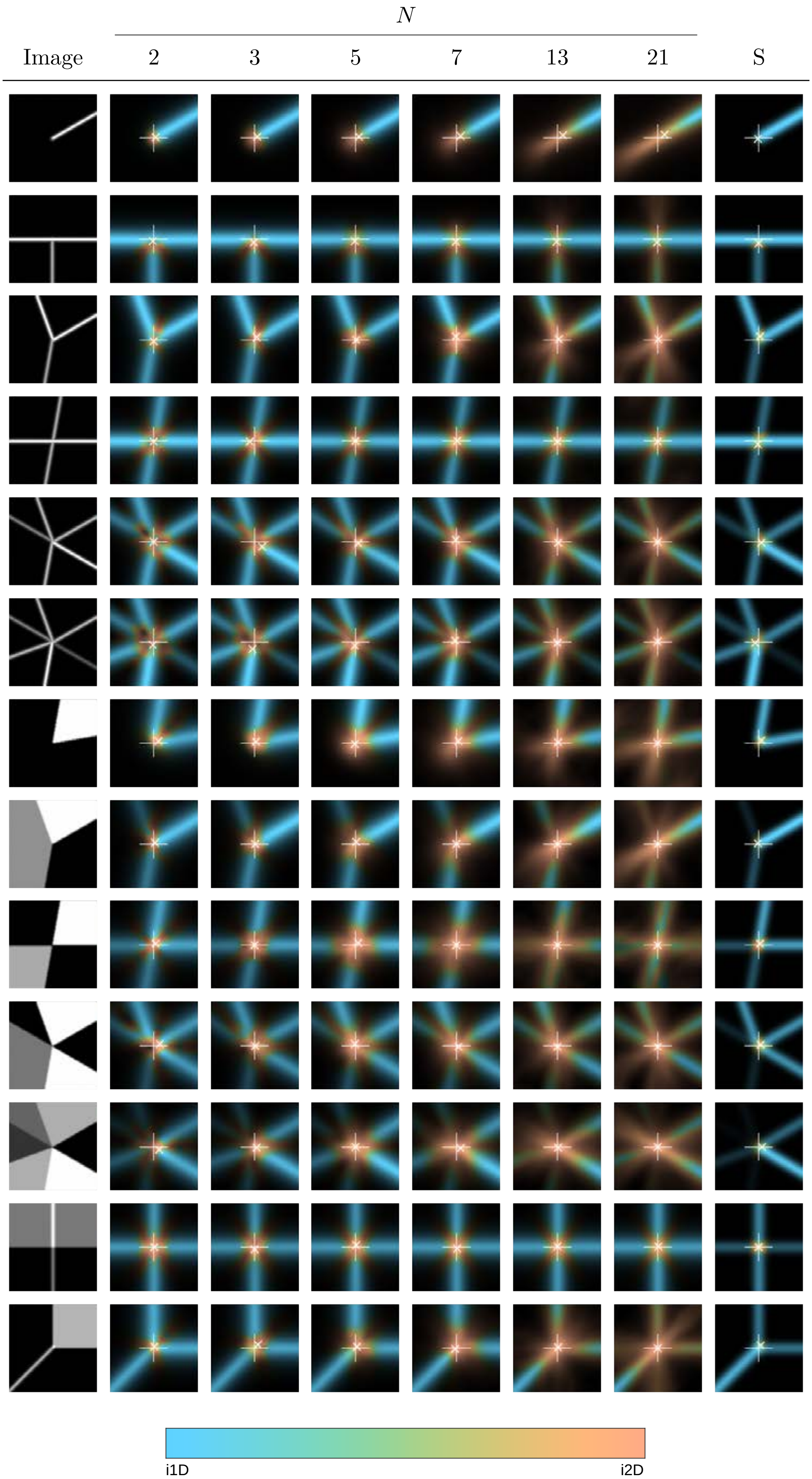


Figure 3.13: Examples of image features along with their intrinsic dimension representation from the CH vector for different N along with that of the structure tensor (S). Brightness represents magnitude, $\|\mathbf{d}_2\|$, colour represents angle, γ_2. The centre of the feature is indicated by the plus symbol, an the location of the detection point is indicated by the cross symbol.

Apart for the single line segment image, detection position improves with increasing N (Figure 3.13). Figure 3.14 shows the range of detection position errors for the same set of features for different N, compared to those using the boundary tensor junction energy and Harris corner detector from the structure tensor. The filter sizes were the same as for Figure 3.13. As N increases, the position error decreases, and appears to plateau after about $N = 11$. Detecting the actual centre of an i2D feature is important for methods that use steerable filters to find the orientation of component line or edge segments, such as in [86] and [70]. If the position is off-centre the calculated orientations can be affected. Therefore, applying the proposed detection measure using larger N should be useful for these methods.

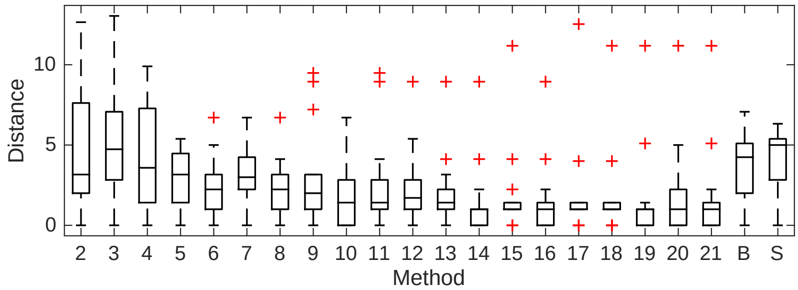


Figure 3.14: Range of detection location errors (pixels) for the feature set in Figure 3.13 using the sinusoidal model method for different N (numbered), compared to the boundary tensor (B) and the Harris corner detector from the structure tensor (S). The large outlier for higher values of N is due to the single line segment feature.

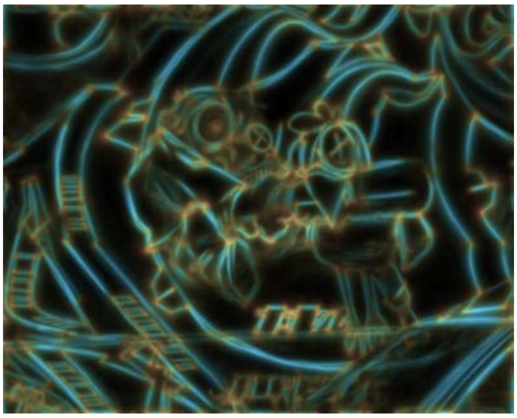
3.3.5 Test Set Evaluation

To gauge the repeatability of the detector under different image transformations, a grey-scale version of the test set from [81] was used. The test set has seen popular use for the evaluation of local descriptors and interest point detectors, such as in [4, 82, 108]. It consists of eight subsets of images, each with an original image and five or six transformed images. The transforms are viewpoint change (subset: *graffiti*, *bricks*), scale and rotation (subset: *boat*, *bark*), blur (subset: *bikes*, *trees*), illumination (subset: *cars*) and JPEG compression (subset: *ubc*).

The sinusoidal model was calculated over four scales using a log-Gabor filter with $\sigma = 0.6$ and wavelength $\{4, 8, 16, 32\}$ pixels. The i2D detection measure, without angle rescaling, for each scale was added together to give a final detection score. Adding the scores was found to give better results than choosing the maximum from each scale. Candidate detection points were chosen as the locations of the local maxima in a 3 pixel radius area. Computational load was almost wholly taken up with calculating the sinusoidal model, and thus the time to calculate the detections was approximately four times longer than that given in Table 2.2 for each N. An example of a pair of images from the viewpoint *graffiti* set along with their intrinsic dimension representation for the first scale and the top 100 detections is shown in Figure 3.15.

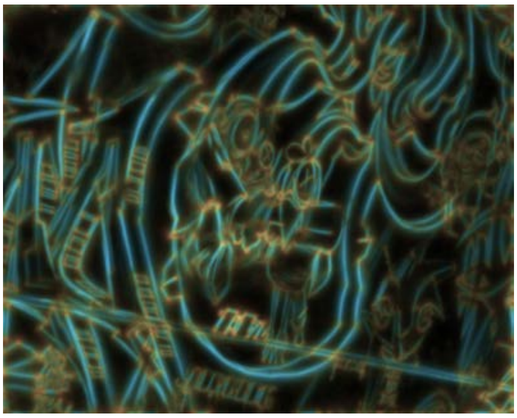
The detections from the original image in each subset were compared to each of the transformed images. Any points that were not within the common area to each image were discarded.

(a) First image in *Graffiti* test set



(b) Intrinsic dimension representation of (a)

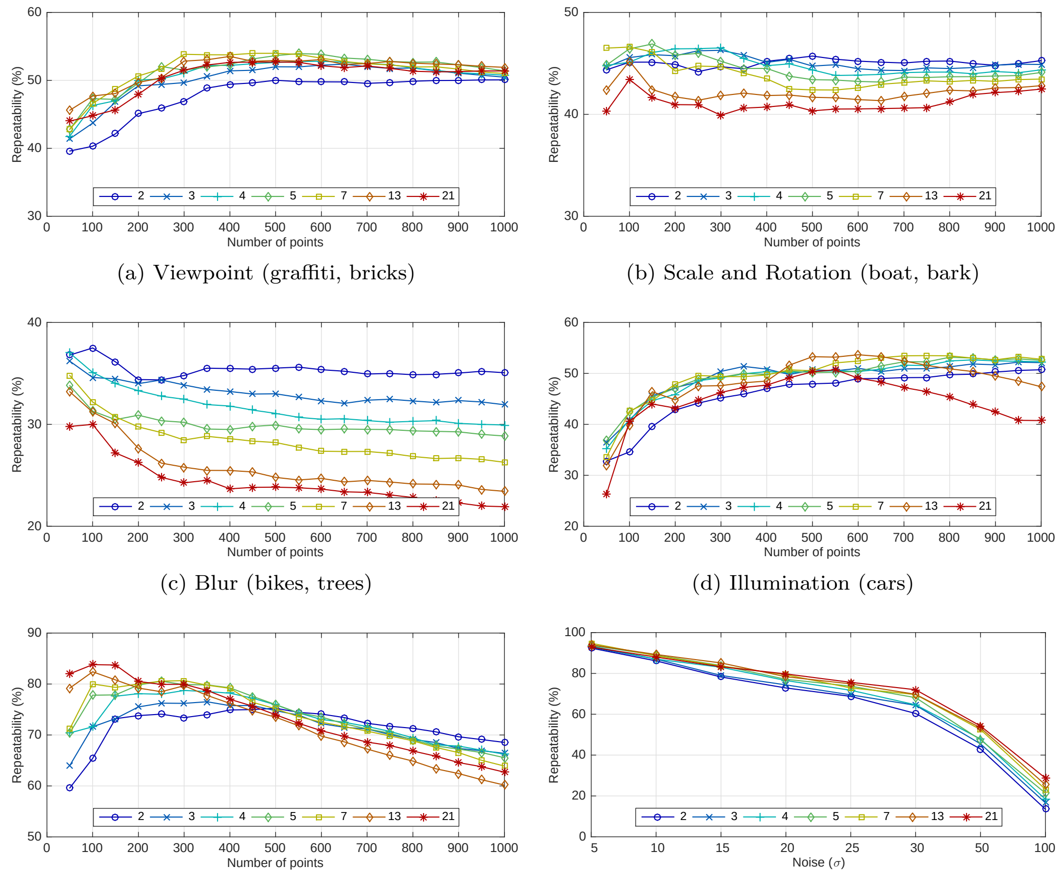
(c) Second image in *Graffiti* test set



(d) Intrinsic dimension representation of (b)

(e) Matched detections (green) and unmatched (cyan) between the first image (circle) and the second image (cross).

Figure 3.15: Example of the top 100 detections matched between the first and second images of the *Graffiti* test set from from [81].



(a) Viewpoint (graffiti, bricks)
(b) Scale and Rotation (boat, bark)
(c) Blur (bikes, trees)
(d) Illumination (cars)
(e) JPEG Compression (ubc)
(f) Noise (100 points, all subsets)

Figure 3.16: Average i2D detection repeatability for different values of N (shown in legend) and number of detection points, evaluated for various image transformations (a-e) and additive Gaussian noise (f) using the test set from [81].

Each detection point was considered matched if there was a corresponding detection point in the transformed image within a distance of three pixels. However multiple correspondences were not allowed. Repeatability was calculated as the average of the fraction of matched points in the first image and the fraction of matched points in the transformed image, as in [4]. These values were then averaged for each transformation type to give an overall score. The results are shown in Figures 3.16a - 3.16e. Different levels of Gaussian noise were also added to the original image from each subset and the repeatability calculated. The results for the top 100 points averaged across all subsets is shown in Figure 3.16f.

The repeatability varied with increasing N, the type of transformation and number of points. For the viewpoint, scale and rotation, and illumination subsets, the middle range value of $N = 7$ had the highest repeatability for smaller numbers of points, whereas low range N values had the highest repeatability for larger numbers of points. For the blur subsets, any increase in N reduced the scores dramatically, and $N = 2$ gave the best results. In contrast, increasing N increased repeatability consistently for the JPEG subset, for less than 300 points. This is due to the increased size of the wavelets averaging out the block-like compression artefacts. Repeatability increased with N for the noise experiment for the same reason.

Using a value of $N = 7$ appears to give the best all-round score. Qualitative analysis of the detection images revealed that for larger values of N, detection of curved lines started to increase, due to the lower correlation with model wavelets because of their increased size. Using $N = 7$ appears to be a good compromise between having enough RT orders to discriminate more complex junctions and corners from i1D features, yet having compact enough model wavelets to follow the curved lines in the image sets. It was also found that if only one scale (wavelength = 8 pixels) is used, repeatability remains generally the same except for a slight decrease in the blur subset. This shows that performing detection at a single scale is sufficient to capture most of the interest points in the test set, and has the benefit of reducing computation time.

Contour-based corner detectors along with the Harris-Laplace [82] and Laplacian-of-Gaussian [64] detectors were tested in [4] on the same set of images. The results for $N = 7$ and 200 points were compared to that of the best detector in each subset in [4]. For our detection method, repeatability was approximately 20% better for the viewpoint subset, 15% better for the scale and rotation subset, same for the blur and JPEG subsets, and 20% worse for the illumination subset.

3.4 Reconstruction

The other advantage of the proposed approach is its wavelet embedding. We can reconstruct the image from either its model or residual components. Reconstructing an image from only the sinusoidal model components acts like a wide-band linear (i1D) filter, while reconstructing from only the residual components filters for i2D areas. This interesting effect was shown in Figure 3.6 where the *Pentagon* image was reconstructed from four wavelet scales for different values of N. Increasing N resulted in the model reconstruction becoming more selective towards larger linear features, due to the increased size of the wavelets. Taking this concept further, reconstruction can be performed after manipulation of the model parameters to achieve various image processing tasks.

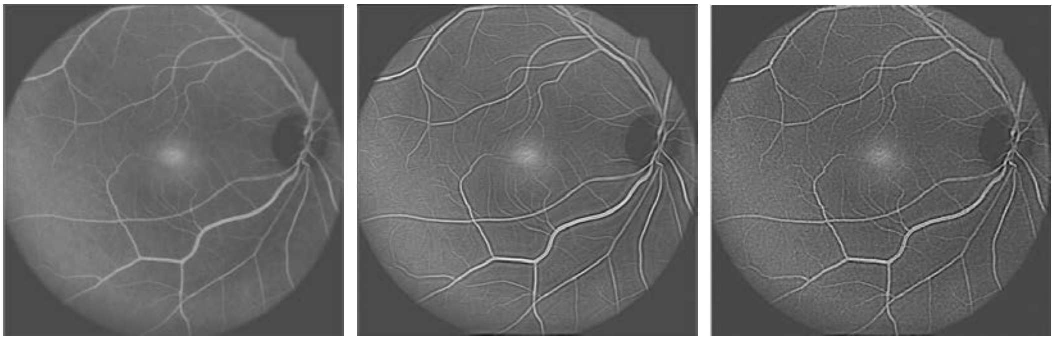
The process is:

1. Calculate the sinusoidal model components at each scale,
2. Adjust the amplitude, orientation or phase parameters.
3. Recreate the model CH vector using the adjusted parameters.
4. Reconstruct the image.

Examples using amplitude and orientation are presented in this section.

3.4.1 Amplitude

Modulating the model amplitude can also be used to enhance the image. For example, Figure 3.17a shows an image of a retina, with linear features that correspond to blood vessels. These features were enhanced by multiplying the sinusoidal model amplitude for the first two (high frequency) scales by four and reconstructing (Figure 3.17b). If instead both the model and residual parts for the first two scales are amplified (Figure 3.17c), the non-linear parts of the image are also enhanced, which leads to a noisier reconstruction in this case. The residual components and low-pass channel were included in the reconstruction.

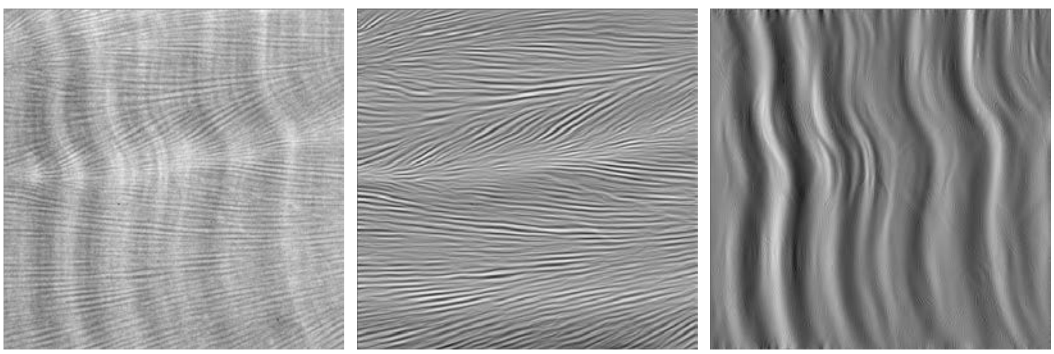


(a) Retina image (b) Model enhanced (c) Both enhanced

Figure 3.17: Retina image (a) with linear features enhanced (b) and all features enhanced (c) for the first two scales.

3.4.2 Orientation

Figure 3.18a shows an x-ray image of a slice of *Porites* coral. The horizontal features are the hollow tubes along which the coral organisms grow, while the vertical lines are yearly growth rings [72]. The task is to split the image into horizontal and vertical components. To do this we can reconstruct from only the model components with $\theta \in [\pi/4, 3\pi/4]$ to isolate the tubes (Figure 3.18b), and reconstruct from the remaining model components to isolate the growth rings. Fine features in the growth rings are maintained which would have been otherwise lost if using filters at different scales. Images like this may benefit from analysis with a two-sinusoidal signal model and this is explored further in the next chapter. The low-pass component was not included in the reconstruction.



(a) Coral x-ray (b) Horizontal recon. (c) Vertical recon.

Figure 3.18: Coral x-ray image (a) reconstructed separately from horizontal (b) and vertical (c) model components.

3.5 Summary

In this chapter we have introduced a novel sinusoidal signal model constructed from higher-order RTs. Previous 2D analytic signal approaches to modelling image structure as a sinusoid [29, 135] only used up to the 3rd order RT. The model consists of a sinusoidal and a residual component. The sinusoidal component describes the strength (amplitude), symmetry (phase) and orientation of i1D features. The residual component describes the remaining local image structure, and therefore has a high magnitude at the location of i2D features. Increasing the number of RT orders improves the estimation of model parameters, especially orientation.

A method of weighting the CH vector to ensure phase-invariance of the vector magnitude was also proposed. Phase invariance is important when using the CH vector magnitude as a boundary energy measure. The benefit of having a residual component in the model led to the development of a novel measure of intrinsic dimension. The complex intrinsic dimension measure can be used for boundary detection (amplitude), i1D detection (real part) and i2D detection (imaginary part). Tests on a common feature detection image set showed i2D detection performance on par with other contemporary methods.

Many methods that use the monogenic signal for phase and orientation estimation should be improved using this approach. In particular, simply adding the 2nd order RT gives an estimate of the orientation from even structures and is quick to compute as the orientation parameters can be found analytically using quartic solvers. Even with the 2nd order RT an intrinsic dimension representation is possible.

Chapter 4

Multi-Sinusoidal Signal Model

The previous chapter demonstrated that lines and edges are i1D structures that are well represented by a single sinusoidal model. The large residual component of the model at junctions and corners indicates they are i2D structures with multiple orientations of symmetry. Therefore, alternative image models are required to represent these more complex features.

This chapter develops the *multi-sinusoidal* image model that was introduced at the start of the previous chapter. It consists of the addition of multiple sinusoids at different amplitudes, phases and orientations plus a residual component. The model can represent features consisting of the addition of i1D components, such as crossed-line junctions, as well as give phase-invariant estimates of the multiple orientations of symmetry that may be present in the local image structure.

The chapter layout is as follows:

- The matched wavelets for the model components are developed.
- The iterative, roots and super-resolution methods of solving for the model parameters are investigated to see the effect of number of RT orders on model error. Super-resolution is particular effective in resolving sinusoidal components close in orientation.
- A procedure for junction classification is introduced. In particular, the difference between additive lines and occluded lines is investigated, as occluded lines are common in many images but do not match an additive model.
- The model is compared to two contemporary methods for orientation estimation of two crossed lines.
- The usefulness of having extra sinusoids in the model is demonstrated on the original motivating problem of estimating orientation in coral core x-ray images. The model combined with the multi-scale CH wavelet method is ideally suited to this problem as the images consist of multiple i1D features at different orientations and scales.

Preliminary work on solving a multi-sinusoidal model using RTs and analysing coral x-ray images has been published in [75] and [72], respectively.

4.1 Multi-sinusoidal signal model

We now consider a *multi*-sinusoidal signal model consisting of K oriented sinusoids with differing amplitude, phase and orientation plus a residual component. This model was briefly introduced in Section 3.1.1, it models local image structure at a point of interest, $\mathbf{z} = 0$, as

$$f(\mathbf{z}) = \sum_{k=1}^{K} \underbrace{A_k \cos(\langle \mathbf{z}, \mathbf{o}_k \rangle + \phi_k)}_{f_{\mathrm{S}}(\mathbf{z})} + f_\epsilon(\mathbf{z}), \tag{4.1}$$

where $f_{\mathrm{S}}(\mathbf{z})$ is a single sinusoidal model component with amplitude A, phase ϕ, and orientation θ where $\mathbf{o} = [\cos\theta, \sin\theta]$, and $f_\epsilon(\mathbf{z})$ is the residual component. Higher-order RTs give more estimates of the sinusoid parameters. Since the individual sinusoidal components in the model are linearly combined, we may write

$$\mathcal{R}^n f = \begin{cases} \sum_k A_k e^{\mathrm{i}n\theta_k} \cos(\phi_k) + \mathcal{R}^n f_\epsilon, & n \text{ is even,} \\ \sum_k A_k e^{\mathrm{i}n\theta_k}\, \mathrm{i}\sin(\phi_k) + \mathcal{R}^n f_\epsilon, & n \text{ is odd.} \end{cases} \tag{4.2}$$

For this model we have to find *multiple* values of amplitude, phase and orientation. To do this we find the CH vectors *matched* to the sinusoidal model component and apply either the iterative or roots method to solve for their values.

4.1.1 Matched Wavelets

As derived in the last chapter, the single sinusoidal CH vector can be written as a function of amplitude, phase and orientation,

$$\mathbf{f}_{\mathrm{S}}(A, \phi, \theta) = A\mathbf{S}_\theta \cos\phi\, \mathbf{s}_e + A\mathbf{S}_\theta \sin\phi\, \mathbf{s}_o, \tag{4.3}$$

where $\mathbf{s}_e$ and $\mathbf{s}_o$ are orthogonal CH vectors given by

$$s_{e_n} = 1 \qquad \text{if } n \text{ even}, 0 \text{ otherwise}, \tag{4.4}$$

$$s_{o_n} = -\mathrm{i} \qquad \text{if } n \text{ odd}, 0 \text{ otherwise}. \tag{4.5}$$

Setting orientation to 0, the two matched wavelets for a given weighting matrix $\mathbf{W}$ are

$$\mathbf{W}\mathbf{f}_e = \mathbf{s}_e \mathbf{W}/\sqrt{W_e}, \tag{4.6}$$

$$\mathbf{W}\mathbf{f}_o = \mathbf{s}_o \mathbf{W}/\sqrt{W_o}, \tag{4.7}$$

where W_e and W_o are the sum of the even and odd weightings respectively,

$$W_e = \sum_{n \text{ even}, |n| \in \mathbb{N}_N} w_n^2, \tag{4.8}$$

$$W_o = \sum_{n \text{ odd}, |n| \in \mathbb{N}_N} w_n^2. \tag{4.9}$$

The model single sinusoidal CH vector can thus be expressed as the scaled and rotated sum of the model wavelets,

$$\mathbf{Wf}_{\mathrm{S}}(A, \phi, \theta) = \lambda_e \mathbf{S}_\theta \mathbf{Wf}_e + \lambda_o \mathbf{S}_\theta \mathbf{Wf}_o, \tag{4.10}$$

where

$$\lambda_e = \sqrt{W_e}\, A \cos\phi, \tag{4.11}$$

$$\lambda_o = \sqrt{W_o}\, A \sin\phi. \tag{4.12}$$

Since the multi-sinusoidal signal model is a *linear* combination of single sinusoids, we may write them as the sum of rotated single sinusoidal CH vectors,

$$\mathbf{Wf} = \sum_{k=1}^{K} \mathbf{Wf}_{\mathrm{S}}(A_k, \phi_k, \theta_k) + \mathbf{W}\boldsymbol{\epsilon} \tag{4.13}$$

$$= \sum_{k=1}^{K} \mathbf{S}_{\theta_k} \left(\lambda_{e_k} \mathbf{Wf}_e + \lambda_{o_k} \mathbf{Wf}_o \right) + \mathbf{W}\boldsymbol{\epsilon}. \tag{4.14}$$

4.2 Solution

The model consists of a single wavelet set at multiple orientations and thus can be solved for using the iterative or roots method for the *single wavelet set, multiple orientation* model in Section 2.4.2. Furthermore, the model can be solved by the super-resolution method by adding the second wavelet to the first as an imaginary component.

4.2.1 Iterative Method

The iterative method begins by solving for one sinusoid using the image CH vector, as is done for the single sinusoidal model.

$$[A_1, \phi_1, \theta_1] = \arg\min_{A,\phi,\theta} \| \mathbf{Wf} - \mathbf{Wf}_{\mathrm{S}_1} \|. \tag{4.15}$$

The second sinusoid is solved by using the *residual* component. Let $\mathbf{W}\boldsymbol{\epsilon}_1 = \mathbf{W}\mathbf{f} - \mathbf{W}\mathbf{f}_{\mathrm{S}_1}$ be the first residual component. The second sinusoid is given by

$$[A_2, \phi_2, \theta_2] = \underset{A,\phi,\theta}{\arg\min} \|\mathbf{W}\boldsymbol{\epsilon}_1 - \mathbf{W}\mathbf{f}_{\mathrm{S}_2}\|. \tag{4.16}$$

From this we obtain the second residual $\mathbf{W}\boldsymbol{\epsilon}_2 = \mathbf{W}\mathbf{f} - \mathbf{W}\mathbf{f}_{\mathrm{S}_2}$, and the process is repeated for each of the K sinusoids in the model, that is

$$[A_k, \phi_k, \theta_k] = \underset{A,\phi,\theta}{\arg\min} \|\mathbf{W}\boldsymbol{\epsilon}_{k-1} - \mathbf{W}\mathbf{f}_{\mathrm{S}_k}\|. \tag{4.17}$$

Note that iterative solving does not guarantee an overall optimum as the individual CH vectors are not orthogonal. However, it does guarantee that the residual component decreases in magnitude with each iteration since

$$\|\mathbf{W}\boldsymbol{\epsilon}_k\|^2 = \|\mathbf{W}\mathbf{f}_{\mathrm{S}_k}\|^2 + \|\mathbf{W}\boldsymbol{\epsilon}_{k+1}\|^2. \tag{4.18}$$

4.2.2 Roots Method

The roots method solves for all of the sinusoids at one time. Recall that the polynomial to solve for the sinusoidal model is

$$p(2\theta) = \lambda_e(\theta)^2 + \lambda_o(\theta)^2. \tag{4.19}$$

For a single sinusoid, we find the root of the derivative of $p(2\theta)$ that gives the maximum value of the polynomial. For multiple sinusoids we find the *set* of roots where the second derivative of $p(2\theta)$ is negative, meaning that they correspond to peaks in the angular response. Given the set of roots, $\{\theta_k\}$, the amplitude and phase values are then found by evaluating $\lambda_e(\theta_k)$ and $\lambda_o(\theta_k)$. The final step is to order the sinusoids in descending order of amplitude, as is inherently done in the iterative method, as this enables a consistent interpretation of the model.

There are some differences between output of the roots and iterative methods.

- There is no guarantee that the residual component using K roots will decrease if K is increased.
- The quick approximation method (Section 2.6.2)of estimating the location of the maximum of a trigonometric polynomial cannot be used, and normal root finding procedures must be employed.
- The number of large components may be different.

An example of the last difference is shown in Figure 4.1 for the angular response polynomial of a feature consisting of two closely oriented sinusoids. The iterative process returns two large sinusoidal components. However, the roots method can only return components where there are peaks, and therefore returns one. This property can be taken advantage of in both ways. The iterative procedure could be used resolve components that are closer together in orientation, while the roots process guarantees one only gets components at maxima of the response.

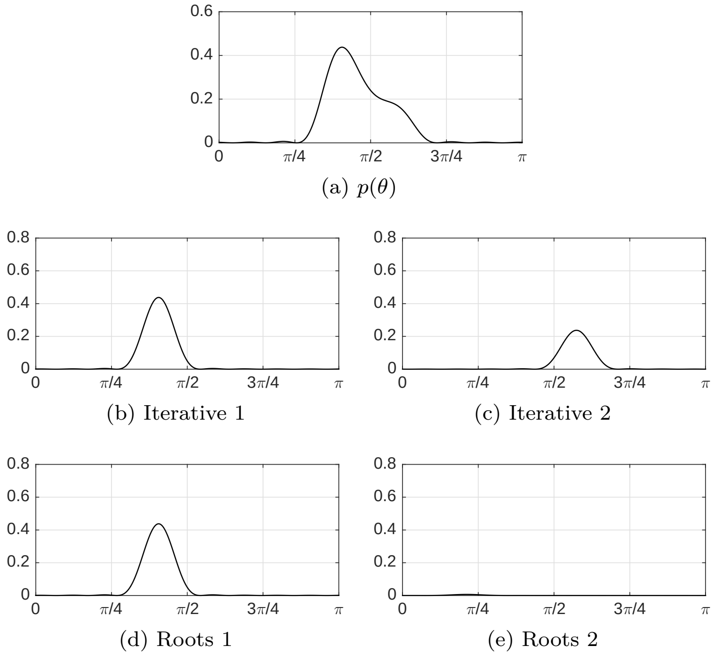


Figure 4.1: Sinusoidal angular response polynomial (a) and the angle response of the first two model components returned using the iterative method (b-c) and the roots method (d-e) for $N = 9$

4.2.3 Super-Resolution

In Section 2.5 the *super-resolution* method was introduced to solve models consisting of a single wavelet at multiple orientations. The method can be extended to solve for the multi-sinusoidal model as well. Weighting has no effect with the super-resolution method, therefore we express the model CH vector as

$$\mathbf{f} = \sum_{k=1}^{K} \mathbf{f}_{\mathrm{S}}(A_k, \phi_k, \theta_k) + \boldsymbol{\epsilon} \tag{4.20}$$

$$= \sum_{k=1}^{K} \lambda_{e_k} \mathbf{S}_{\theta_k} \mathbf{s}_e + \lambda_{o_k} \mathbf{S}_{\theta_k} \mathbf{s}_o + \boldsymbol{\epsilon} \tag{4.21}$$

$$= \sum_{k=1}^{K} A_k \cos \phi_k \mathbf{S}_{\theta_k} \mathbf{s}_e + A_k \sin \phi_k \mathbf{S}_{\theta_k} \mathbf{s}_o + \boldsymbol{\epsilon}. \tag{4.22}$$

Let $s_\phi = \cos \phi + \mathrm{i} \sin \phi$. Since $\mathbf{s}_e + \mathbf{s}_o$ is a vector of all ones, substituting this in gives

$$\mathbf{f} = \sum_{k=1}^{K} A s_\phi \mathbf{S}_{\theta_k} (\mathbf{s}_e + \mathbf{s}_o) + A \bar{s}_\phi \mathbf{S}_{\theta_k} (\mathbf{s}_e + \mathbf{s}_o) + \boldsymbol{\epsilon} \tag{4.23}$$

$$= \sum_{k=1}^{K} \frac{A}{2} s_\phi \mathbf{S}_{\theta_k} + \frac{A}{2} \bar{s}_\phi \mathbf{S}_{\theta_k} + \boldsymbol{\epsilon}. \tag{4.24}$$

An individual order of the CH vector is thus

$$f_n = \sum_{k=1}^{K} \frac{A_k}{2} s_{\phi_k} e^{-in\theta} + \frac{A_k}{2} \bar{s}_{\phi_k} e^{-in\theta} + \boldsymbol{\epsilon}. \tag{4.25}$$

Therefore the model CH vector can be rewritten as the Fourier series of a spike train

$$\mathbf{f} = \boldsymbol{\mathcal{F}}^N x(\theta) + \boldsymbol{\epsilon}, \tag{4.26}$$

where

$$x(\theta) = \sum_{k=1}^{K} \left(\frac{A_k}{2} s_{\phi_k} \delta(\theta - \theta_k) + \frac{A_k}{2} \bar{s}_{\phi_k} \delta(\theta - \pi - \theta_k) \right). \tag{4.27}$$

Each sinusoid corresponds to a *conjugate* pair of spikes. The complex amplitudes of the spikes, $A_k/2s_{\phi_k}$ and $A_k/2\bar{s}_{\phi_k}$, are given by the amplitude and phase of the sinusoid components. The spikes are located at θ_k and $\pi + \theta_k$ respectively, and these locations correspond to the orientations of the sinusoids. Crucially, the double spike representation means that all the orders are non-zero and thus the super-resolution method can be used.

Let $\boldsymbol{\alpha} = \{A_k/2s_{\phi_k}\}_{k \in \mathbb{N}_K}$ be the set of all the complex amplitudes. Then we can find $\{A_k\}$ and $\{\theta_k\}$ using the super-resolution method by minimising the residual component

$$\min_{\boldsymbol{\alpha},\boldsymbol{\theta}} \|\tilde{x}(\theta)\|_{\mathrm{TV}} \quad \text{subject to} \quad \|\boldsymbol{\mathcal{F}}^N \tilde{x}(\theta) - \mathbf{f}\|_1 \leq \delta \tag{4.28}$$

and it follows that $\|\boldsymbol{\epsilon}\|_1 \leq \delta$. The resulting spikes are in conjugate pairs at θ and $\theta + \pi$, so we only need to consider the spikes with a location smaller than π. If α_k is the complex amplitude of one of these spikes, the corresponding amplitude and phase of the sinusoid components are given by

$$A_k = 2|\alpha_k|, \tag{4.29}$$

$$\phi_k = \arg(\alpha_k). \tag{4.30}$$

4.2.4 Example Image

The multi-sinusoidal model for $K = 2$ was calculated for the *Pentagon* image, obtained using the first scale of a Meyer wavelet decomposition with $N = 13$ using the iterative method, and is shown in Figure 4.2. The values of the amplitude and residual norm images are all on the same scale and can be directly compared. The amplitude of the second sinusoid is large at the locations of the cross-like junctions on the roof of the pentagon, indicating at least two strong components of linear symmetry. It can be observed that at locations where the first sinusoid favours the features *perpendicular* to the main pentagon shape, the second sinusoid will represent the main shape. For example, the blue dots in the large orange area in the lop left quadrant of the first orientation image are orange in the second image. The example shows that extra sinusoidal components are needed to represent multiple linear symmetries.

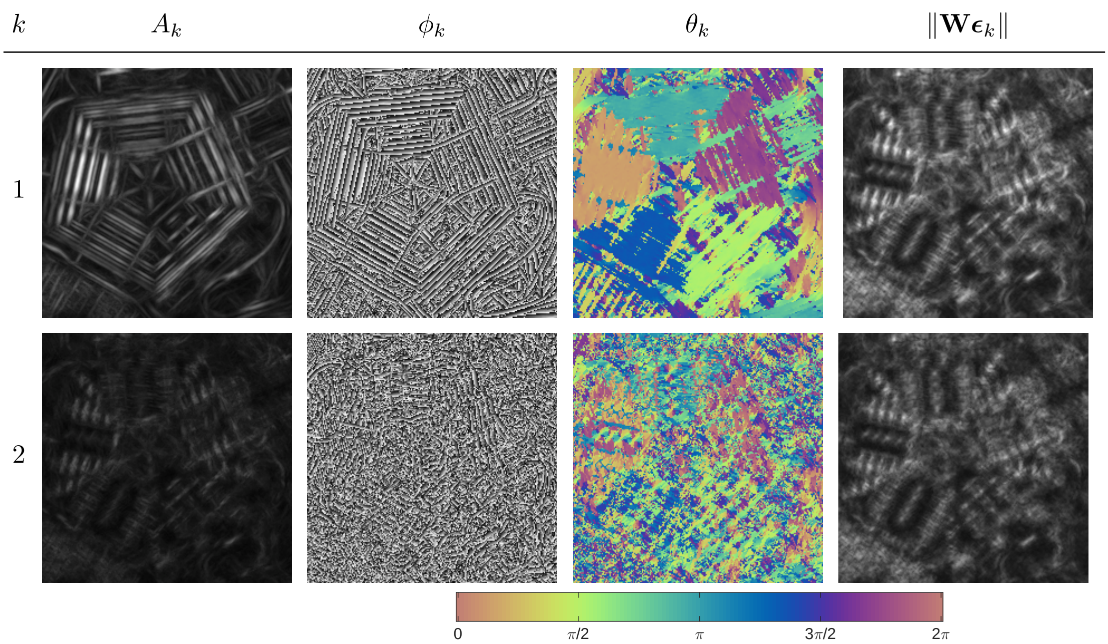


Figure 4.2: Decomposition of the first scale of a 256 × 256 pixel version of the *Pentagon* image into two sinusoidal model components using the second Meyer wavelet, $N = 13$ and phase-invariant equal weighting.

4.3 Model Accuracy

Given more than one linear symmetry component, we wish to know how many RT orders are needed to resolve their parameters accurately. In this section we shall investigate the differences in model accuracy between each solution method for different values of N and weighting. The iterative and roots methods are treated separately to the super-resolution method, owing to the latter not using weights.

4.3.1 Iterative and Roots Method

Accuracy of the model estimation using the iterative and roots methods was compared for $K = 2$ sinusoids for different values of N and weighting. In particular, the orientation difference between the two sinusoids was varied to see the effect on the model estimation, and thus no noise was added.

A set of 100 test CH vectors was constructed from the addition of two sinusoid CH vectors with amplitude ratio randomly varying between 0.5 and 1.5 and uniformly distributed random phase. The orientation difference was then varied from 2.5 to 90 degrees in 2.5 degree increments, for a total of 3600 test vectors. That is,

$$\mathbf{Wf} = \mathbf{Wf}_{\mathrm{S}_1}(A_1, \phi_1, \theta_1) + \mathbf{Wf}_{\mathrm{S}_2}(A_2, \phi_2, \theta_2). \tag{4.31}$$

The model parameters were solved for using either the iterative or roots method to give two estimated single sinusoidal model CH vectors $\mathbf{f}'_{S_1}$ and $\mathbf{f}'_{S_2}$. Since the order of the estimated model components does not necessarily match that of the original model, the estimated components were paired with the original model components such that there was a *minimum* distance between CH vectors. The amplitude, orientation and phase error were calculated for each pair, and averaged over the 100 test vectors for each orientation using the following formulas:

$$A_{\text{error}} = \sum_k |A_k - A'_k| / K, \tag{4.32}$$

$$\phi_{\text{error}} = \sum_k |\arg(\cos(\phi_k - \phi'_k) + \mathrm{i}\sin(\phi_k - \phi'_k))| / K, \tag{4.33}$$

$$\theta_{\text{error}} = \sum_k |\arg(\cos(2\theta_k - 2\theta'_k) + \mathrm{i}\sin(2\theta_k - 2\theta'_k))| / 2K. \tag{4.34}$$

The double angle is used in the calculation for θ since the methods return ϕ over the entire circle, but θ only over the half-circle.

Effect of N

The effect of N was investigated using the phase-invariant equal weighting scheme ($B \approxeq 0$) (Figure 4.3). For both methods, the error is greatest when the orientation separation is close to 0. This is due the orientation selectivity of the wavelets as when the sinusoids are close together they interfere and are instead perceived as a single sinusoid. The iterative method is better at resolving closely oriented sinusoids, due to the ability to discern two components from a single wider peak as demonstrated in Figure 4.1.

Increasing N increases the orientation selectivity, and thus a higher N can be seen to result in smaller orientation differences being resolved with less error. When applying the model, the expected minimum orientation difference of feature can therefore be used to guide the choice of N. As a rough guide, when the orientation difference is above π/N for the iterative method, or $1.5\pi/N$ for the roots method, the errors appear to plateau. Furthermore, increasing N reduces the error *floor* at larger orientation differences, and the error floor for the iterative method is lower than the roots method. This is due to the iterative method removing components from the model, and therefore removing the effect of the angular oscillations of that component on the estimates for subsequent components.

A common approach to multiple component estimation with steerable filters is to find the maxima of the response polynomial, that is, the roots method. These results show that the iterative approach, made possible by having the residual vector, gives better parameter estimates for a sinusoidal model.

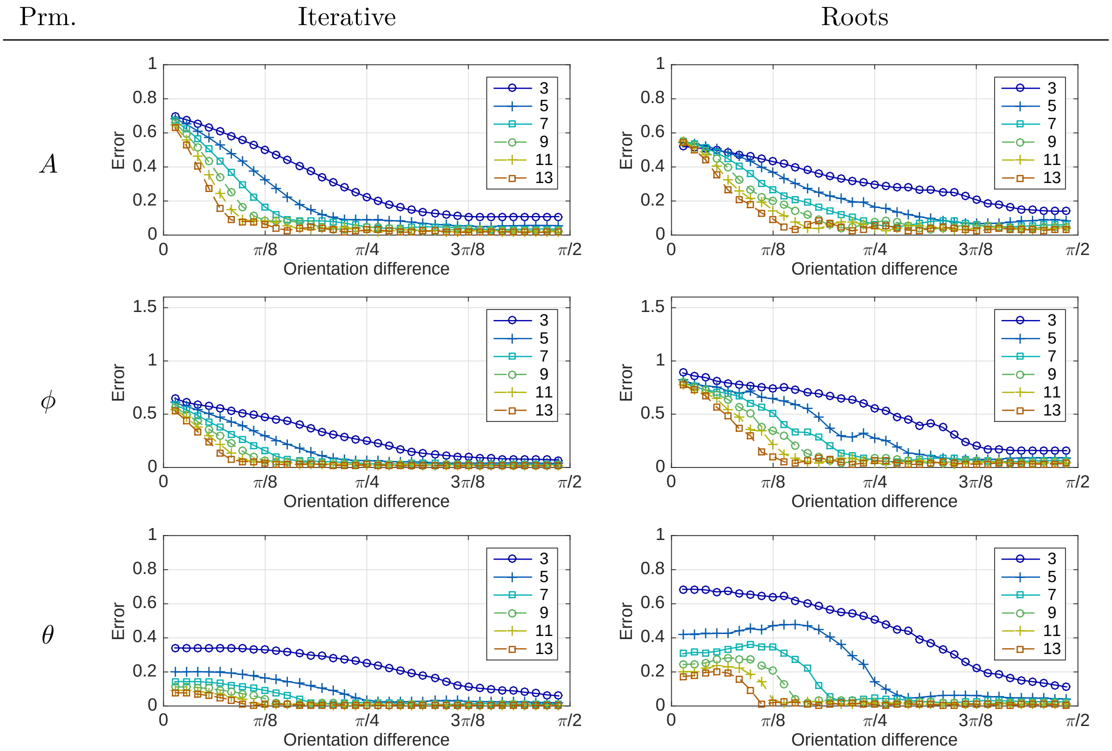


Figure 4.3: Average error of the estimated model parameters for two sinusoids with varying amplitude, random phase, and varying orientation separation for different values of N (shown in legend), using the iterative and root solvers.

Effect of Weighting

In the previous chapter it was shown that adjusting the weights could be used to reduce the oscillations in the angular response of the wavelets, at the expense of orientation selectivity (main lobe width). It was remarked that reducing the oscillations would reduce the correlation between wavelets in a multi-sinusoidal model. To test this theory, the same test was performed as above, except with a fixed value of $N = 13$ and the weighting varied instead (Figure 4.4). The weights were calculated using the energy maximisation method with the window width B chosen as a multiple of the constant, $B_{0.1} = 5.64/N - 6.57/N^2$. A factor of 0 is equivalent to the equal weighting used in the previous experiment.

As expected, increasing B increases the minimum orientation difference before the errors begin to plateau. That is, the ability to accurately resolve sinusoids close in orientation is reduced due to the wider angular response profile. However, past this point, increasing B *decreased* the error floor by orders of magnitude thanks to the reduced oscillations.

The same experiment was conducted with different values of added Gaussian noise. As the noise was increased the error floor also increased. After the noise CH vector magnitude was greater than approximately 0.15 times the image CH vector magnitude the noise floor for all weightings was approximately the same. Therefore in higher noise images using $B = 0$ remains a good choice for the weighting.

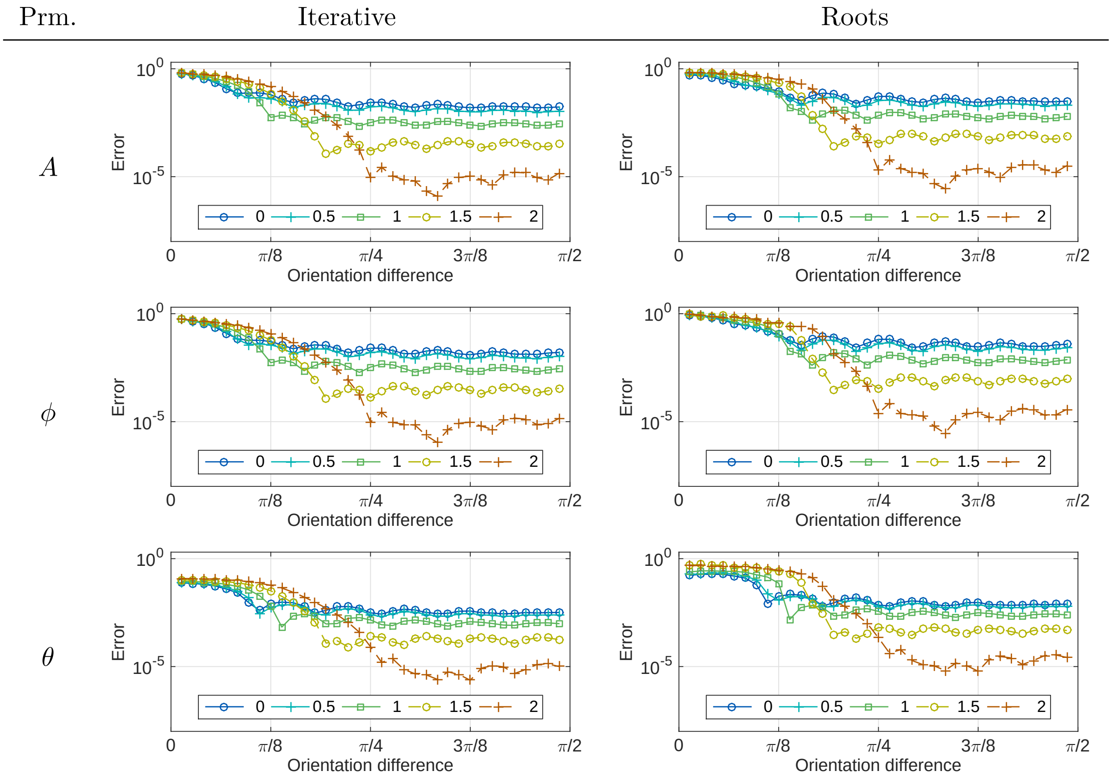


Figure 4.4: Average error of the estimated model parameters for two sinusoids with varying amplitude, random phase, and varying orientation separation for different values of B as a multiple of $B_{0.1}$ (shown in legend) and $N = 13$, using the iterative and root solvers.

4.3.2 Super-Resolution Method

A similar experiment testing the accuracy of the model was performed using the super-resolution method, albeit with a reduced number of vectors due to the increased computation time of the method. Two sets of CH vectors were created, one set consisting of two equal amplitude sinusoids with the same phase of 0 and one set with two equal amplitude sinusoids with opposite phases of 0 and π. Essentially this corresponds to having spikes with the same or different sign. The orientation difference was then varied from 2.5 to 90 degrees in 2.5 degree increments, for a total of 72 test vectors. Since the super-resolution method does not use a weighting scheme, only the accuracy of the model estimation versus N was investigated.

Numerical experiments involving spike trains in [15] suggest a orientation separation as small as $2\pi/N$ may be enough to guarantee finding an exact result. However, those results used single, uncorrelated spikes, where as the sinusoids in the model are represented as a *pair* of spikes. Results of the test (Figure 4.5) show that when both phases are 0, the separation required is approximately $\pi/2N$, while for opposite phases it is approximately π/N. These values are one quarter and half of the experiment in [15]. After these thresholds, the error is practically zero.

In contrast, the previous two methods have a much larger separation threshold after which the average error stays relatively constant. Therefore the super-resolution method can be used to resolve sinusoidal model components that are close in orientation using a smaller value of N.

Unfortunately, the large computation time of the MATLAB code from [15] (Section 2.6.3) currently limits the practical application of the super-resolution method as it is infeasible to model an entire image.

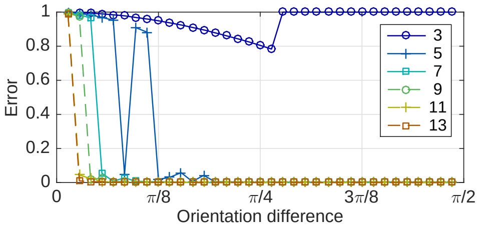
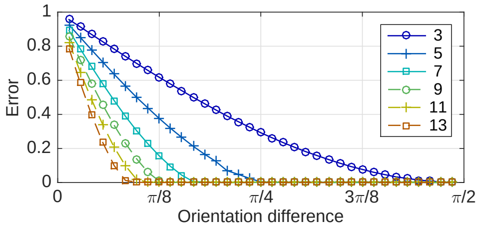
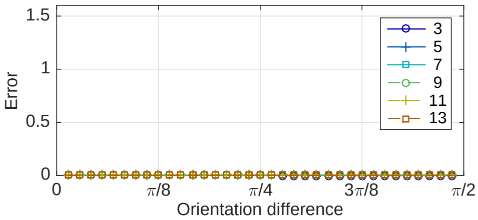
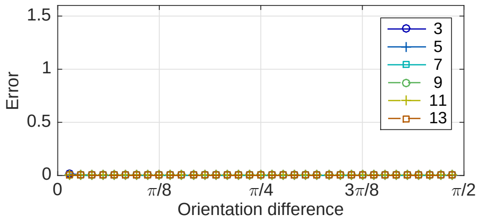
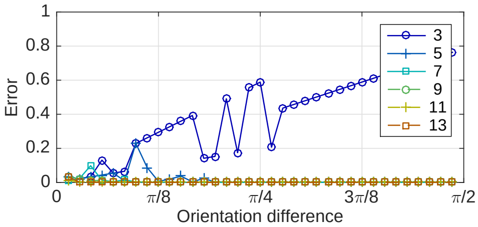
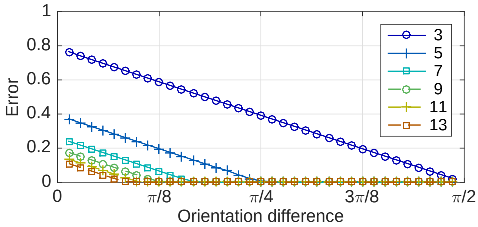
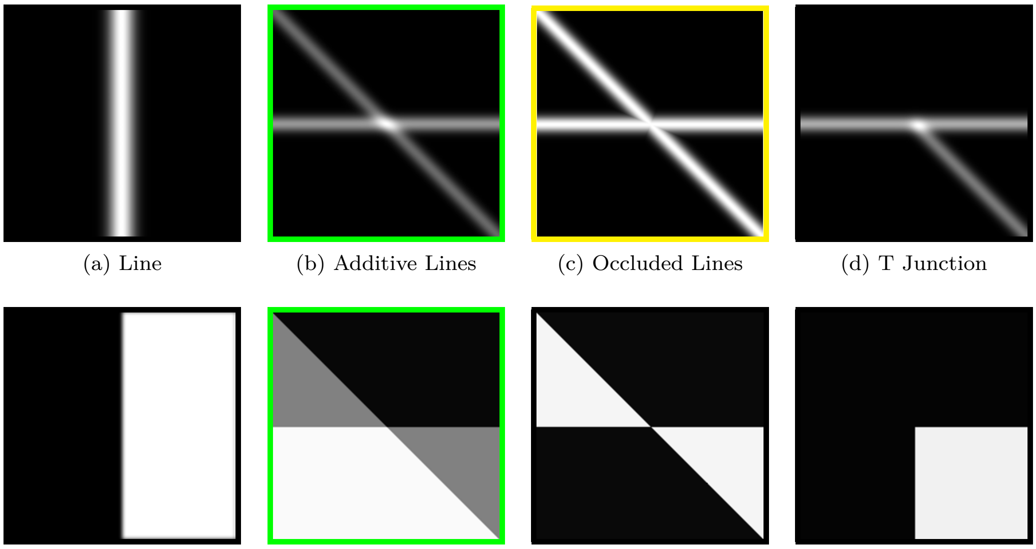
Prm.	$\phi = [0, 0]$	$\phi = [0, \pi/2]$
A		
ϕ		
θ		

Figure 4.5: Average error of the estimated model parameters for two sinusoids with unit amplitude, phase of $\{0, 0\}$ or $\{0, \pi/2\}$, and varying orientation separation for different values of N (shown in legend), using the super-resolution method.

4.4 Junction Analysis

In this section the multi-sinusoidal model is applied to the task of analysing multiple-line and multiple-edge features. Additive line (Figure 4.6b) and additive edge features (Figure 4.6f) can be modelled using multiple sinusoids as they can be expressed as a sum of i1D features. Occluded line (Figure 4.6c) features require the CH wavelets to be of a certain size. Other features (Figures 4.6d, 4.6g and 4.6h) are better modelled by line and wedge segments, which is the focus of the next chapter.



(a) Line (b) Additive Lines (c) Occluded Lines (d) T Junction

(e) Edge (f) Additive Edges (g) Chequer Edge (h) Corner

Figure 4.6: Examples of various image features. Lines and edges are well-modelled by a single sinusoid (a, e). Additive lines and edges are well modelled by a superposition of sinusoids (green - b, f), occluded lines have an extra component in the centre (yellow - c), and others are better modelled using line-segment or wedge-segments (e, f, g).

4.4.1 Feature Response

Additive i1D Features

To begin with we shall find the model response to multiple additive i1D features. A line or an edge is a i1D structure which varies along one axis of symmetry. Locally it can be represented as a sum of sinusoids with the same orientation,

$$
\begin{aligned}
f_{\mathrm{i1D}}(\mathbf{z}) &= f_{\mathrm{i1D}}(\langle \mathbf{z}, \mathbf{o} \rangle) \\
&= \sum_k A_k \cos(w_k \langle \mathbf{z}, \mathbf{o} \rangle + \phi_k),
\end{aligned}
$$

where $\mathbf{o} = [\cos\theta, \sin\theta]$, and can be modelled completely by a single sinusoidal model. That is,

$$
f_{\mathrm{i1D}}(\mathbf{z}) = A(\mathbf{z}) \cos(\langle \mathbf{z}, \mathbf{o} \rangle + \phi(\mathbf{z})).
$$

A feature consisting of the *addition* of a line, edge or any combination i1D features can be locally represented as the sum of i1D signals,

$$f(\mathbf{z}) = \sum_k f_{\mathrm{i1D}}(\langle \mathbf{z}, \mathbf{o}_k \rangle),$$

and therefore by the multi-sinusoidal model as

$$f(\mathbf{z}) = \sum_k A_k(\mathbf{z}) \cos(\langle \mathbf{z}, \mathbf{o}_k \rangle + \phi(\mathbf{z})).$$

The RT responses to this type of feature are therefore

$$\mathcal{R}^n f(\mathbf{z}) = \begin{cases} \sum_{k=1}^{K} A_k(\mathbf{z}) e^{\mathrm{i} n \theta_k} \cos(\phi_k(\mathbf{z})) & n \text{ is even,} \\ \sum_{k=1}^{K} A_k(\mathbf{z}) e^{\mathrm{i} n \theta_k} \, \mathrm{i} \sin(\phi_k(\mathbf{z})) & n \text{ is odd.} \end{cases}$$

from which the CH vector of the feature can be obtained. Note that within the local area of this type of feature, the amplitude and phase of the model components vary between locations, but the *orientation remains the same.* This is important as it means that the model can be evaluated at locations other than the centre of the feature and still return the correct orientation.

Occluded Lines

While lines may be i1D features, often crossed lines in an image are not additive but instead occluded. For example, the intersection of the black lines in Figure 4.7 is same shade of black, not a darker shade. Additive lines are more like crossed lines drawn by a highlighter, where multiple strokes result in a darker colour.

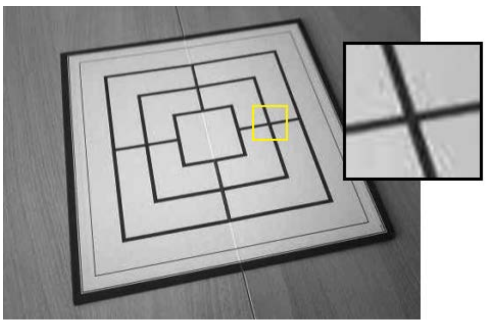


Figure 4.7: Example of an occluded crossed line feature.

Therefore we must investigate the CH vector response of occluded line features, which shall be defined as the superposition of lines of the same sign. Instead of modelling these feature as the addition of i1D components, we model them as the maxima (or minima) of the their intensity. For positive valued lines (white on black) the model is

$$f(\mathbf{z}) = \max_{k \in K} \{ f_{\mathrm{line}_k}(\mathbf{z}) \}, \tag{4.35}$$

and the sinusoidal model for this structure becomes

$$f(\mathbf{z}) = \max_{k \in K}\{A_k(\mathbf{z}) \cos(\langle \mathbf{z}, \mathbf{o}_k \rangle + \phi(\mathbf{z}))\}. \tag{4.36}$$

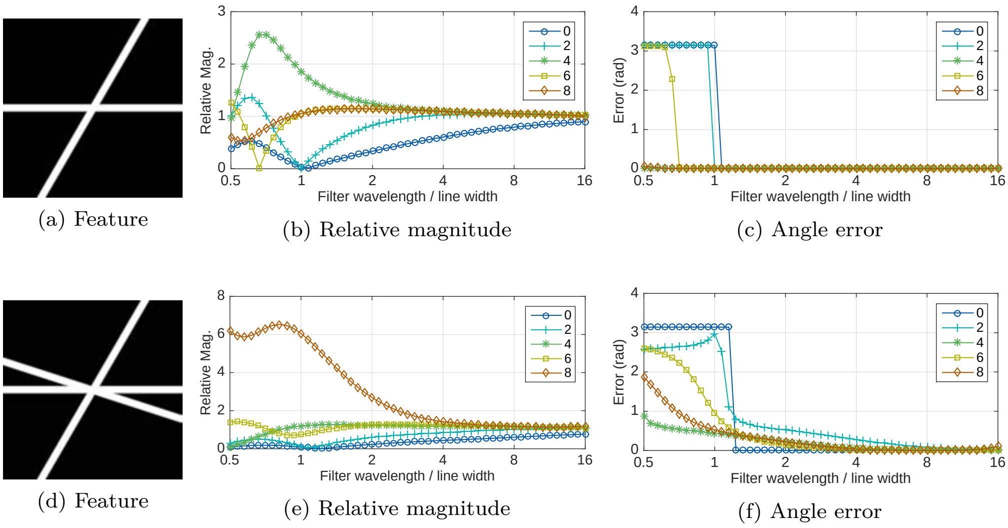
The maximum operator will always result in a value for $f(\mathbf{z})$ that is *less* than the equivalent additive model, that is

$$\max_{k \in K}\{A_k(\mathbf{z}) \cos(\langle \mathbf{z}, \mathbf{o}_k \rangle + \phi(\mathbf{z}))\} \leq \sum_{k=1}^{K} A_k(\mathbf{z}) \cos(\langle \mathbf{z}, \mathbf{o}_k \rangle + \phi(\mathbf{z})), \tag{4.37}$$

and thus we may write the occluded model as the additive model minus a non-negatively valued occlusion function $f_{occ}(\mathbf{z})$

$$f(\mathbf{z}) = \sum_{k=1}^{K} A_k(\mathbf{z}) \cos(\langle \mathbf{z}, \mathbf{o}_k \rangle + \phi(\mathbf{z})) - f_{occ}(\mathbf{z}). \tag{4.38}$$

The *normalised* CH vector at the centre of two occluded line features (Figure 4.8) was investigated using a log-Gabor filter with $\sigma = 0.65$. The differences between the magnitude and angle of each CH vector component (RT order) and the ideal additive version of the feature were calculated for different filter wavelengths. The magnitude is expressed in relative terms in Figure 4.8b and 4.8e. For example, a value of 2 means the occluded vector component has twice the magnitude of the ideal component, whereas a value of 1 is equal. The amplitude errors are larger for both smaller wavelets (smaller wavelength) and lower orders. In particular, the 0th order CH wavelet response is always lower in magnitude than the ideal version, and the angle error indicates that it has the wrong sign for the smaller wavelengths.



(a) Feature (b) Relative magnitude (c) Angle error

(d) Feature (e) Relative magnitude (f) Angle error

Figure 4.8: Differences in the magnitude (b,e) and angle (c,f) of each CH vector order (shown in legend) between an ideal additive multiple line feature and a non-zero width occluded line feature (a,d).

The occluded error component, $f_{occ}(\mathbf{z})$, in the previous equation is fixed in size. Therefore increasing the wavelength increases the size of the wavelet and thus reduces the magnitude of the response to the error component. Likewise, higher-order CH wavelets have a larger spatial extent with more energy concentrated further from the centre. These observations lead to the following proposed strategies to deal with occluded lines:

- **Increase the number of orders N.** Including higher orders increases the size of the wavelet and therefore reduces the occluded error response.
- **Increase the wavelength of the primary filter.** The examples in Figure 4.8 suggest a minimum wavelength of two to four times the line width is necessary.
- **Remove lower order components.** The lower order CH wavelets are the smallest in spatial extent and most affected by the occluded error. By setting their weighting to 0, we reduce the influence of the error. The remaining weights must be adjusted such that $W_e = W_o$ to ensure the amplitude is remains phase-invariant. Note this means the image CH vector is *also* weighted the same way.

Removing Lower Orders

An example of wavelets constructed by removing the lower orders is shown in Figure 4.9 for $N = 13$, numbered according to the minimum non-zero order. For example, 2 indicates the 0th and 1st order components were weighted to 0. Removing lower orders attenuates the centre of the wavelet but also reduces its linear appearance and increases the amplitude of oscillations outside the main lobe. It appears that lower orders are *not* required to estimate the model over the full orientation range, as if there are enough higher orders the interference effects still produce a wavelet with a single directional axis.

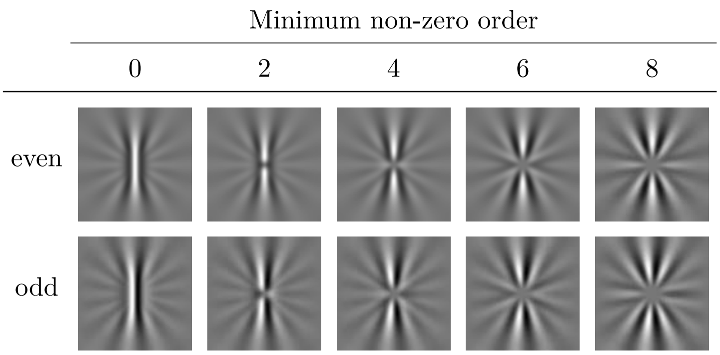
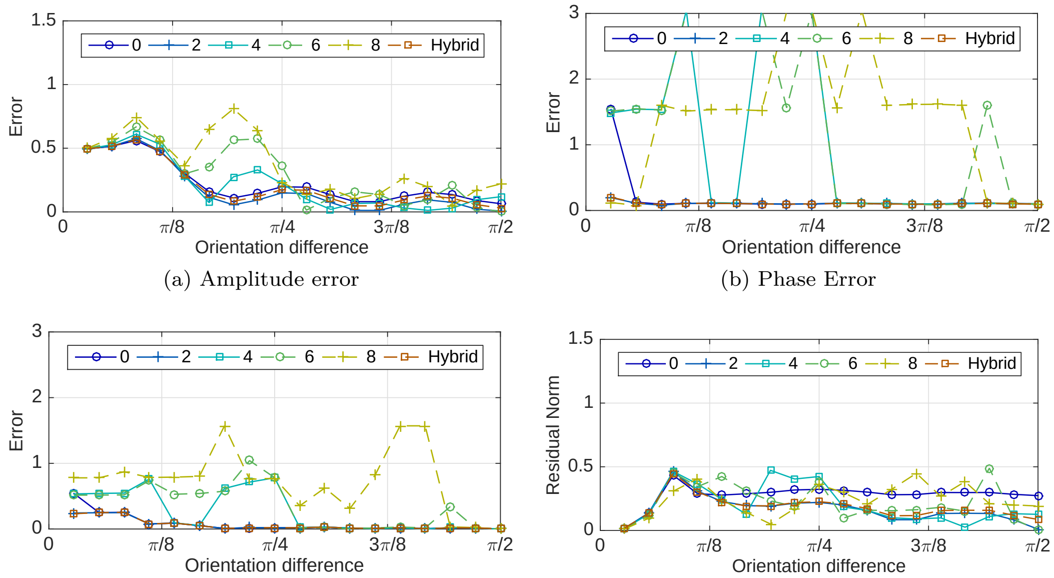


Figure 4.9: Even and odd order sinusoidal wavelets with the lower RT orders set to zero for $N = 13$ and phase-invariant equal weighting. The number indicates the minimum non-zero order.

The effect of removing lower orders was investigated by finding the model parameters for an occluded two line feature (such as Figure 4.8a) with orientation difference ranging from $\pi/36$ to π radians. The model was calculated for $N = 13$, phase-invariant equal weighting, the iterative method and using a log-Gabor filter with $\sigma = 0.65$ and wavelength twice the width of the line.

The response to a *single* line feature with the same line width was used to normalise the results (Figure 4.10). Note that for a single occluded line feature, the residual is zero.



(a) Amplitude error (b) Phase Error

(c) Orientation Error (d) Residual magnitude

Figure 4.10: Amplitude, phase and orientation errors long with the residual norm for an occluded two line feature solved using the iterative method and wavelets with the lower orders removed. The minimum non-zero order is shown in the legend. The CH vector was calculated using a log-Gabor primary filter with wavelength twice the line width and $\sigma = 0.65$, and $N = 13$.

The amplitude, phase and orientation errors are lowest for the both normal sinusoidal wavelets with no orders removed, and the wavelets with the 0th and 1st order removed (using 2nd order and above). In particular, amplitude errors for the two line occluded feature were larger than for two additive features (Figure 4.3), but the phase and orientation errors were almost the same. However, if the filter wavelength is reduced to be the same as the line width (not shown), phase and orientation errors are introduced. This suggests that phase and orientation of occluded features can be found reasonably accurately for filter wavelengths *twice the line width and above*.

Setting any more than the second order to zero gives phase and orientation errors at certain orientation differences. This can be explained by the increase in amplitude of the off-axis oscillation of the angular response giving a large off-axis response. Thus rather than isolating a single line, the maximum response actually occurs when the crossed lines of the feature are at the same angle difference as two side lobes. Removing the 0th and 1st orders is of benefit in reducing the residual norm slightly, shown as a proportion of the image CH vector norm in Figure 4.10. Where this is of benefit is:

- The model vector explains more of the image CH vector.
- The effect of the occluded part of the feature is reduced for subsequent iterations.
- The residual is closer to the ideal residual of zero, meaning that thresholding based on the residual may be more accurate.

While removing lower orders has some advantages, the CH vector is intended to be a primary descriptor of local image structure and removing the 0th and 1st order removes information about isotropic local image structures (blobs). Therefore, a hybrid system is proposed where the CH vector is weighted normally and the first sinusoidal component found using the normal sinusoidal wavelets with no lower orders removed. Then subsequent sinusoids are found using the sinusoidal wavelets with the 0th and 1st orders removed. The results for the hybrid scheme (Figure 4.10) show the approach works in that the residual component remains lower while the parameter errors remain the same. A hybrid approach is only possible using the iterative solving method, and thus by having the residual component.

Other Types

Some features are not well described by a sinusoidal model. The chequer pattern in Figure 4.6f is defined by edges; however, it is an even structure that does not respond to odd order RTs. As a result, the corresponding model has multiple components that do not match the orientation of the edges, and thus analysis is not straight-forward. The corner feature (Figure 4.6g) has a similar problem. In contrast, the T junction (Figure 4.6e) generates sinusoid components with the same orientation as the lines; however, there are extra components to compensate for the half-line segment not continuing through the centre. Features such as these are better modelled by multiple lines or edges radiating from a point, which is the subject of the next chapter.

4.4.2 Classification

When calculating the multi-sinusoidal model one of the choices will be how many sinusoidal components to include in the model. This is particularly relevant to the problem of classifying junctions according to the number of significant i1D components they contain. In fact, the problem is common to any model featuring multiple components and it is worth while exploring several strategies.

Maximum Component

The first approach is to only include model components where the amplitude is greater than a certain fraction, $\delta \in [0, 1]$, of the maximum component's amplitude. That is

$$\mathrm{class}(\mathbf{f}) = \mathrm{card}\{k \,|\, \lambda_k > \delta \times \lambda_{\max}, k \in \mathbb{N}_K\}, \tag{4.39}$$

where $\lambda_{\max}$ is the amplitude of the maximum component. Since models can have multiple parts to a single component, such as the even and odd parts of the sinusoid component, consider λ_k to be the magnitude of the CH vector corresponding to the k-th component of the model. This approach is invariant to the strength of the local image structure, however is does not take into account the amplitudes of the other components when setting the threshold.

K-adjustable Threshold

It was thought that an improved approach might be to set the threshold as a proportion of the the CH vector magnitude. However, this approach is not invariant to the number of components. For example, consider a feature constructed of K components with unit norm and thus $\lambda_k = 1$. We have

$$\sum_{k=1}^{K} \lambda_k^2 \geq \|\mathbf{Wf}\|^2 \tag{4.40}$$

and thus

$$\sqrt{K} \geq \|\mathbf{Wf}\| \tag{4.41}$$

for $k \in \mathbb{N}_K$. Therefore, the threshold needs to depend on $\frac{1}{\sqrt{K}}$. A classification method taking this into account is

$$\text{class}(\mathbf{f}) = \max_{k \in \mathbb{N}_K} \lambda_k > \delta \times \|\mathbf{Wf}\|/\sqrt{k} \qquad \text{where } k \in \mathbb{N}_k, \lambda_k > \lambda_{k+1}. \tag{4.42}$$

In detail, if the k-th component has an amplitude greater than $\delta/\sqrt{k}$ times the CH vector norm, then the first k sinusoids will also be above the threshold. We choose the largest k for which this holds true.

4.4.3 Applying The Model

The procedure to apply the multi-sinusoidal model for junction analysis is:

1. **Choose a method:** The iterative and root methods are fast enough to be applied to an entire image. The iterative method gives better results but takes K times longer when using MATLAB `roots` to find the maximum. However, the iterative method can make use of the quick approximation method from Section 2.6.2 which is faster. This is the recommended approach. The super-resolution method is computationally expensive but gives the best results, it is only feasible when applied at selected points in the image. In this case, either the iterative or roots methods can be used to location possible junctions.

2. **Choose N:** Larger N gives a smaller separation constraint allowing for lines or edges with closer orientation to be resolved, as well as better noise performance. The disadvantages are increased computation time and increased kernel size. The experiments in Section 4.3 give a guide as to what value of N to choose in order to resolve features with a certain minimum orientation separation between components. If this is not known, as a general guide choose $N \geq 6K$ for the iterative method, $N \geq 8K$ and roots method, and $N \geq 4K$ for the super-resolution method, for useful results.

3. **Choose W:** The equal weighting scheme ($B \approxeq 0$) gives the best model estimate when i1D features are close in orientation. However, using the energy maximisation method with a larger window width, (as a multiple of $B_{0.1}$) reduces oscillations in the angular response of the wavelet and improves estimates where i1D features have a larger orientation difference. A larger value of B is best employed with larger values of N to maintain the orientation selectivity of the wavelets.

4. **Calculate Model:** Obtain the CH vector at the desired locations and scales in the image and evaluate the model. The hybrid scheme may be used to reduce the residual vector magnitude.

5. **Detect junctions:** Detection of candidate junction locations can be performed using the intrinsic dimension measure as described in Section 3.3.3. When using the super-resolution method, restricting the analysis to candidate locations will speed up computation dramatically.

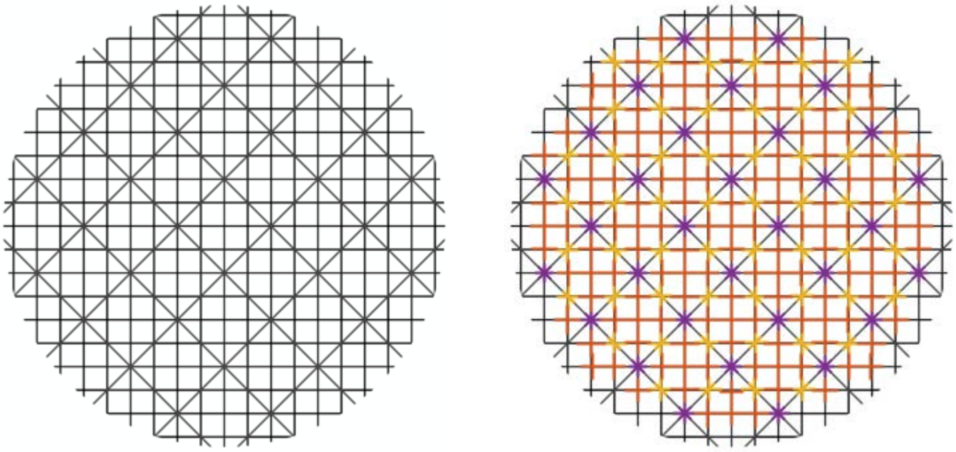
6. **Analysis:** Classify the local image structure according to the model parameters. Phase indicates either even (line) or odd (edge) structures along the corresponding orientation, while the number of large amplitude components corresponds to the number of lines or edges present.

4.4.4 Example Image

An example of junction classification for a test image is shown in Figure 4.11. The original image (340×340 pixels) consists of occluded crossed lines forming junctions of two, three or four lines. The analysis procedure described in the previous section was applied. The multi-sinusoidal model was found at a single scale using a log-Gabor filter with $\omega = 8$ and $\sigma = 0.65$, $N = 13$, equal weighting, and using the iterative process with hybrid wavelets. Locations of the junctions were detected using the local maximum of the i2D junction measure in a 3 pixel radius. Outer edge junctions were excluded from the analysis.

At each detected location the multi-sinusoidal model was used to classify the junction using the K-adjusted method with a threshold of $\delta = 0.5$. The model components are shown as lines in Figure 4.11b where the length and orientation of the lines corresponds to the amplitude and orientation of the model components respectively. The lines are coloured according to the classification of the junction. All the junctions are classified correctly with a mean orientation error of 0.008 radians.

For some junctions, the detection location is not at the crossing point of the lines yet the model is still accurate. This highlights the advantage of the phase-based representation given by sinusoidal models. The symmetry measure (phase) is split from the amplitude and orientation, and thus it does not matter if the detection point is at the centre of the line (even symmetry) or at the edge of the line (odd symmetry). Therefore, if one is mainly interested in the orientation of component structures then the detection point does not have to be at the exact centre of the

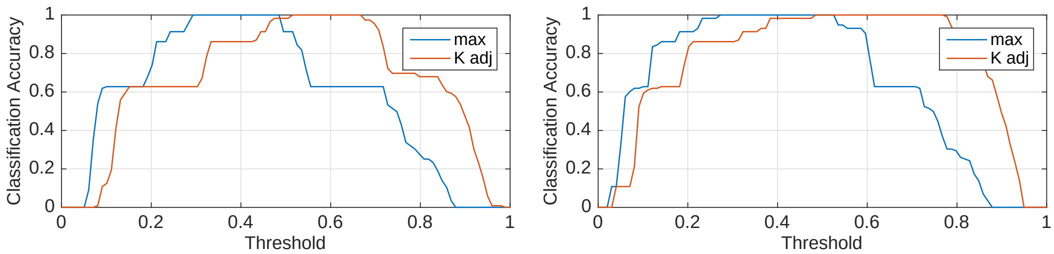


(a) Test Image (b) Classification

Figure 4.11: Test image and resulting junction classification using the multi-sinusoidal model with $N = 13$. Red: two lines, orange: 3 lines, purple: four lines.

structure, as is required for other approaches such as multi-steerable matched filters [86].

The classification accuracy of the K-adjusted method compared to the maximum method was found for varying with threshold values, using both normal (Figure 4.12a) and hybrid wavelets (Figure 4.12b). For both wavelet types, the K adjusted method has a wider area of 100% correct classification, and a larger area under the curve. The probability of choosing a good threshold is thus higher with this method. The hybrid wavelets also result in a higher classification rate over a wider range of threshold values. This example suggests that hybrid wavelets are a better approach to classifying occluded crossed line junctions.



(a) Non-hybrid (b) Hybrid

Figure 4.12: Classification accuracy for different thresholds using the maximum and K-adjusted thresholding methods for the image in Figure 4.11, and either normal sinusoidal wavelets (a) or the hybrid-wavelets (b).

4.5 Experiment: Orientation Estimation

A common problem in image analysis is the estimation of the angles of the individual line or edge features present in a corner or junction feature. For example, these angles can be used for camera calibration [90] or source separation, shown in the latter part of this chapter. The multi-sinusoidal model is suited to the parametrisation of crossed lines. In this section, the model using the iterative, roots or super-resolution methods is compared with two recent approaches, *multi-steerable matched filters* [86] and *mixed orientation parameters* [85], for the analysis of crossed line orientation. An overview of different approaches to multiple orientation estimation is described below.

Multi-Steerable Matched Filters

Steerable wedge-shaped filters are asymmetric kernels that can be used to detect junctions and corners [110, 132]. Wedge filters are constructed from polar basis functions in the spatial domain that have been windowed to restrict their spatial extent. The filter is steered through 2π radians to find peaks in the angular response [110]. Whereas a symmetric steerable filter has one peak for a line, a wedge filter will have two. Wedge filters can be designed to respond to lines or edges or both, and are also sensitive to initial location.

Multi-steerable matched filters (MSMFs) extend the wedge filter paradigm by considering multiple wedge filters matched to the structure of interest [86, 88, 91]. For example, a T junction would require three wedge filters, crossed lines would require two pairs of wedge filters fixed π radians apart, and corners a single variable angle wedge. Initial orientations of feature components are estimated using a crude sampling of the angular response. These orientations are then used to solve for the model parameters using a damped least-squares algorithm (Levenberg–Marquardt) [86]. Different models can be tested and the most representative chosen.

Mixed Orientation Parameters

Mixed orientation parameters (MOP) is a method to analyse additive i1D structures with two orientations [85], additive and occluded i1D structures with two or more orientations [1, 89], and corners and junctions [87]. Rather than estimate orientation at a single location, MOP incorporates the estimates at each point in the neighbourhood of the feature of interest. Higher-order tensors are calculated at each point, with the order depending on the maximum desired orientations in the model. For example, if there were up to four orientations, the fourth order tensor would be used. Singular value decomposition is performed on a matrix of all the tensors (expressed as a row) in the local neighbourhood. The result corresponding to the smallest singular value is the tensor with the minimum difference to the set of tensors in a least-squares sense; that is, the most representative tensor. The orientations are then obtained from eigenvectors of this tensor (in a diagonal matrix).

Steerable Filters

Estimating multiple orientations with symmetric steerable filters involves rotating an appropriated designed filter and looking for peaks in the response [38, 100]. The filter can be designed to match a specific structure of interest [50], or 2D quadrature filters can be used to give a response to both lines or edges. The multi-sinusoidal model from the CH vector falls into the family of 2D steerable filter and wavelet methods. However, the iterative approach to solving for the model parameters differs from other steerable filter methods in that the CH vector is used as a starting point and each component is found by minimising the residual vector iteratively, rather than simply choosing the peaks in the angular response (roots method). The method results in better model estimation, due to the removal of the effects of wavelet side lobes after every iteration and the ability to resolve more closely oriented components.

Applying the super-resolution method to solve for the model parameters is a new approach. Although it may remain a novelty thanks to its enormous processing time, it solves the entire model at once, similar to MSMFs. However, unlike MSMFs it does not require any initial orientation estimates and thus is not sensitive to local maxima. Often steerable filters are created using higher-order derivatives and can also be expressed as the sum of complex basis functions, where the derivatives have been combined into a single complex operator called the Wirtinger operator [137]. The Wirtinger operator is the derivative form of the RT, and thus many methods that use derivatives can be modified to use RTs. In particular, RT tensors [54] could be used for the MOP method.

One important concept to note is that the magnitude of the response of the RT to a sinusoid or i1D signal is the same among all even higher-order RTs, and among all odd higher-order RTs. With derivative-based steerable filters, the magnitude changes with order. The super-resolution model also requires the magnitude of the higher-order responses to be the same for i1D components. The assumption of a spike train hinges of the ability to project i1D signals, which exist as lines in the 2D spectrum, on to 1D spikes. If the response was not the same across orders then the projection would need to be weighted according to the response, requiring the radial part of the spectrum of the i1D feature to be known. Thus the use of the RT enables these approaches.

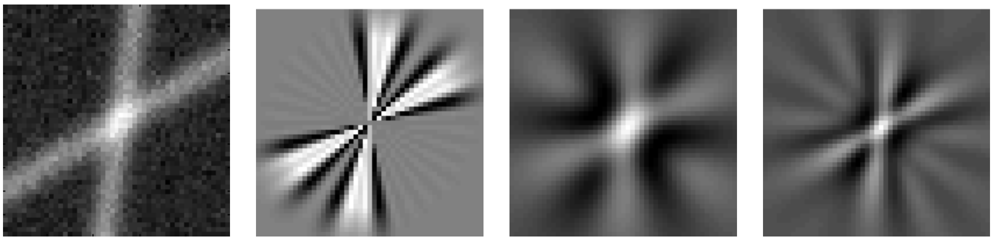
Experimental Results

The performance of the iterative, roots and super-resolution methods was compared to the MSMF and MOP methods for multiple orientation estimation. A synthetic image of two crossed lines was created with the orientation separation ranging from 5° to 90° in 5° increments, and with different amounts of additive Gaussian noise. The orientations of the lines were estimated using the following methods:

- Multi-sinusoidal model calculated for $N \in \{7, 13\}$, $K = 2$, using a log-Gabor filter with $\omega \in [22, 12]$ respectively and $\sigma = 0.65$. The wavelength of the filter was chosen so that the spatial extent of the filter kernel would be approximately the same as that for the MSMF method. The model was evaluated using the iterative, root and super-resolution methods. Spikes within 5 degrees were combined for the super-resolution method (Section 2.5).
- MOP using a second-order tensor evaluated over 48×48 and 12×12 pixel windows centred on the image feature.
- MSMF created using 28 orders, wedge angle 20° and radius 24 pixels. The Levenburg-

Marquardt solver was used with initial parameters set to the maxima of the angular response sampled at 10° spacings. A crossed line model was assumed, and thus two pairs of wedges were used whose components were spaced 180° radians apart.

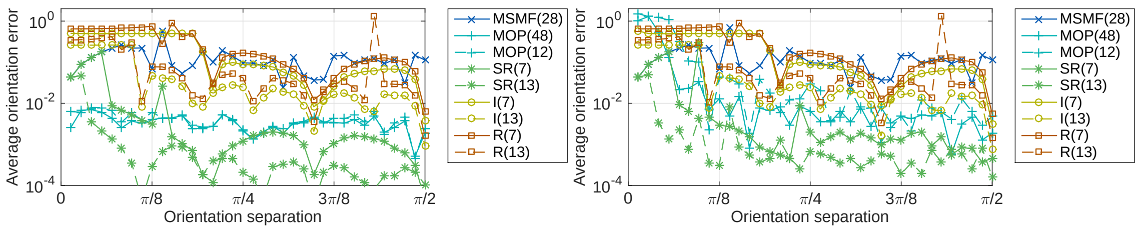
An example of a noisy additive crossed line feature and the MSMF and sinusoidal model wavelets corresponding to the ideal response are shown in Figure 4.13.



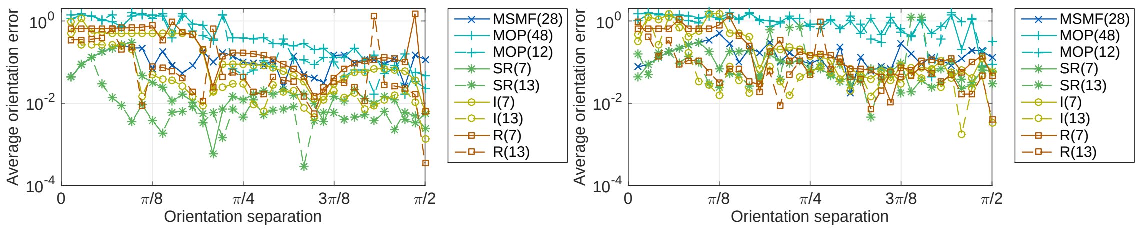
(a) Line feature (b) MSMF (N=28) (c) CH ($N = 7$) (d) CH ($N = 13$)

Figure 4.13: Noisy line feature and the MSMF and sinusoidal model wavelets corresponding to the ideal response at the centre

The estimated orientation error was calculated for different amount of Gaussian noise with $\sigma \in [0, 0.01, 0.1, 1]$ and is shown in Figure 4.14. The super-resolution method is more accurate than all the other methods, except for MOP with no noise and very small separation. Using $N = 13$ orders instead of $N = 7$ improves the error for all the CH vector methods, and both the iterative and the roots method with $N = 13$ have better performance than the 28-order MSMF except for a small separation. As expected, the iterative method outperforms the roots method by a small margin. The MSMF method assumes a double line model and for small separations returns two model components even if only a single orientation was detected, which explains the better small-separation performance.



(a) No noise (b) Gaussian noise, $\sigma = 0.01$



(c) Gaussian noise, $\sigma = 0.1$ (d) Gaussian noise, $\sigma = 1$

Figure 4.14: Orientation estimation errors for a two-line feature, using the half-sinusoidal model solved with iterative (I) and roots (R) methods and line-segment model solved using super-resolution (SR) method for $N \in \{7, 13\}$, compared to MSMF with $N = 28$ and MOP.

MOP performs very well at low noise levels and small separations. The ability to resolve orientations with small separations is due to estimates being made at many pixel locations, especially away from centre of the junction where the lines are more spatially distinct. However, errors for the MOP method rapidly increase with increasing noise. One explanation is that pixels were being included in the window that were not part of the line feature, but reducing the window to 12×12 pixels did not improve the results. Pre-filtering the image around the scale of the features to reduce noise should improve the MOP performance.

The results show that solving for crossed line orientation using the CH vector and either the iterative or roots methods is on par or better than MSMFs and MOP, two of the most recent methods from the literature. Using super-resolution gives superior performance compared to all methods, especially for small orientations. It is remarked in [28] that it is perhaps better to average the output of smaller, simpler filter kernels at many pixel locations, like as is performed with MOP, rather than use the the output of a single large, complex filter, like as is performed with steerable filters. Indeed, construction of the structure tensor [35] uses local averaging of the derivative to give an estimate of local orientation. A combined approach where the CH vector is averaged over an area before solving for the model parameters may therefore improve orientation estimation, although one would need to take the phase of the local symmetry into account, as CH vectors of linear structures with opposite phases would cancel out. This is the subject of future research.

4.6 Application: Measuring Coral Growth Orientation

We now have the tools to tackle the original problem that motivated this research, the biological image analysis task of estimating growth orientation in coral core x-ray images. Initial work on this problem was previously published in [72]. Subsequently, the method has been updated, making use of the CH vector with higher-order RTs instead of only the 0th to 2nd order RT, estimating orientation over multiple scales instead of a single scale, and splitting the image into polyp and growth ring images using wavelet reconstruction.

4.6.1 Overview

The Great Barrier Reef is the largest reef system in the world, comprised of over 2900 individual reefs and stretching along 2600 km of the coast of Queensland [49]. It is an important cultural icon, and supports extensive fishing and tourism industries. As such the health of the reef ecosystem is important, and is the subject of much research in the biological sciences. One proxy for reef health is coral growth rate, a function of the rate and density of extension to its calcareous skeleton. Results from a long running *Porites* coral (Figure 4.15) core sampling program conducted by the Australian Institute of Marine Science [19] have shown a decrease in Great Barrier Reef coral growth rate over fifteen years until 2009. [22].

To measure the growth rate, a cylindrical core is removed from the coral and sliced into planar

Figure 4.15: *Porites* coral colony with brightly coloured Christmas-tree worms on the surface.

sections approximately 7mm thick. The density along the length of the coral is then measured every 0.25mm using a gamma ray densitometer [17]. Alternating high and low density growth bands (similar to tree rings) delineate the measurements into years enabling the estimation of yearly calcification rate [67]. However a limitation of the densitometer instrument is that it is restricted to measuring along a *straight* line. As a result, the coral section is x-rayed to determine the *major growth axis* along which the best measurements would be taken. Two drawbacks to this method are that the major growth axis is determined by eye, and it is often not parallel with the varying direction of coral growth along the sample. These methodological limitations may lead to inaccuracies in the estimation of growth rate.

Two dimensional coral x-ray images consist primarily of i1D features at different scales and orientations (Figure 4.16). There are small lines corresponding to the hollow coral polyp tubes, and larger perpendicular lines corresponding to the varying density of the yearly growth bands. This makes analysis scale dependent and difficult using standard methods. In this section we shall estimate the local orientation of line patterns in the x-ray images using the multi-sinusoidal signal model and CH vector method. The model is calculated across multiple scales, from which an orientation estimate is obtained by combining that from the small polyp lines and the larger growth rings.

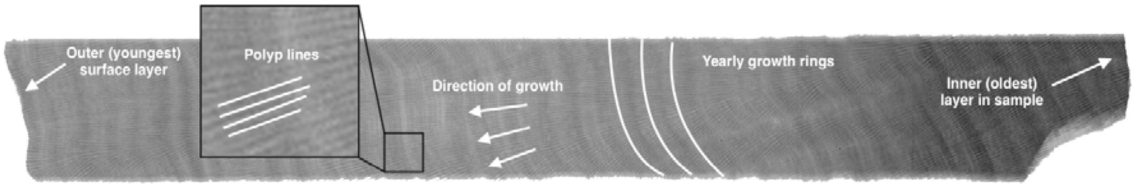


Figure 4.16: Coral x-ray image COB71A_1 with smaller horizontal and larger vertical i1D components.

4.6.2 Method Overview

The previous work on this problem [72] used a single sinusoidal model obtained from the 0th to 2nd order RTs applied at a single scale corresponding to the polyp lines. These smaller features were used to obtain the orientation estimate. Reviewing this work, some issues and their solutions have been identified:

1. Adding the 2nd order RT gave a much improved orientation estimate compared to the monogenic signal. However, it was still quite noisy, especially in low intensity areas of the image, and thus required median filtering. Adding higher-order RTs may therefore improve the estimate.

2. At some locations in the image the response to the vertical growth rings was stronger than for the horizontal polyp lines. This resulted in the sinusoidal model being oriented along the vertical lines, and thus the orientation estimate being off by approximately 90 degrees. Using the multi-sinusoidal model with two sinusoids should model these locations better, as one sinusoid will correspond to the polyp lines and one will correspond the growth rings (Figure 4.16).

3. The growth rings are perpendicular to the horizontal polyp lines, however only the polyp lines were used for the orientation estimate. Combining orientation estimates from multiple scales will allow us to use the growth rings as well.

The proposed improved method to obtain the orientation estimate is

1. Segment the coral from the background.

2. Calculate the multi-sinusoidal model at multiple scales using the CH vector method with iterative solving.

3. Identify the two main global orientations (polyps and growth rings) and classify each sinusoid component accordingly.

4. Create two multi-scale orientation estimates, one from polyps and one from the growth rings, and combine them.

4.6.3 Segmentation

The coral images appear on a uniform light background. We wish to segment the coral from the background for two reasons. The first is to identify coral pixels so only these are used in the orientation estimate. The second is to change the intensity of the background pixels to more closely match that of the coral. A light background creates a strong edge feature at the boundaries of the coral, which dominates the orientation estimate in the area. By making the background a similar colour and smoothing the edge, this effect can be reduced.

Each pixel in the image, $f(\mathbf{z})$, is classified as coral if its intensity differs from the background intensity by more than a threshold. The background intensity is calculated as the mode of the pixel values on the border of the image.

$$m(\mathbf{z}) = \begin{cases} 1 & |f(\mathbf{z}) - \text{mode(background)}| > \text{threshold}, \\ 0 & \text{otherwise}. \end{cases} \tag{4.43}$$

The resulting binary mask is morphologically closed using a radius 3 disk-shaped structuring element to fill any internal holes, then morphologically eroded by a radius 9 disk to bring the edges of the mask away from the edges of the coral.

A new background image, $b(\mathbf{z})$, that is the same size as the original image is created by replicating the mean value of each unmasked image column then smoothing with a Gaussian filter with $\sigma = 3$. The background of the coral image (masked area) is replaced with this new background and the edge of the transition is also smoothed using a Gaussian filter with $\sigma = 3$ to create a new image $\mathbf{f}'(\mathbf{z})$,

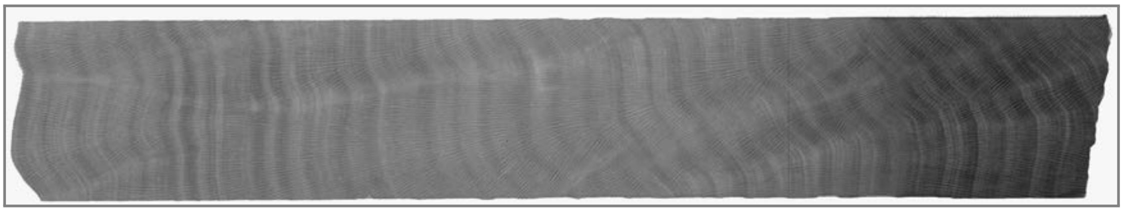
$$f'(\mathbf{z}) = f(\mathbf{z}) \cdot m_s(\mathbf{z}) + b(\mathbf{z}) \times (1 - m_s(\mathbf{z})), \tag{4.44}$$

$$\tag{4.45}$$

where $m_s(\mathbf{z}) = m(\mathbf{z}) * g(\mathbf{z}; 3)$ and $g(\mathbf{z}; \sigma)$ is a Gaussian filter with variance σ^2. The adjusted image (Figure 4.17c) has a much smoother transition from the edge to the background, and therefore the edges of the coral sample will have less effect on the orientation estimate of the i1D features within.

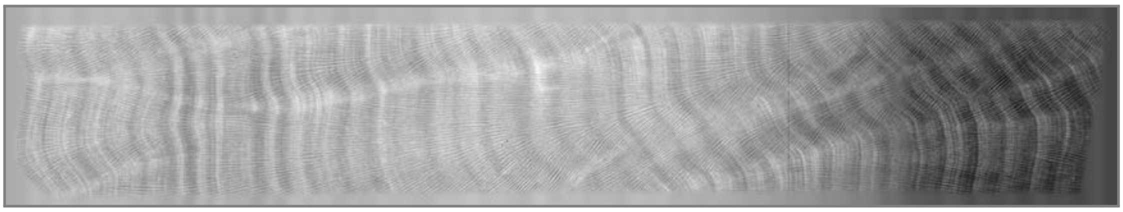
4.6.4 Multi-Sinusoidal Model

There are two dominant i1D features that make up most of the local image structure, therefore we shall use a multi-sinusoidal model with $K = 2$ sinusoids. This means at least $N = 4$ RT orders are required for the CH vector. For this application $N = 7$ was chosen as a trade off between orientation selectivity (high N) and accuracy of reconstruction using model components (low N). Reconstruction accuracy for a given N is explored in Section 5.8 in the next chapter. For the primary isotropic filter we later wish to reconstruct the image so a wavelet frame is necessary.



(a) *COB71A_3* x-ray image

(b) Blurred coral area mask



(c) Adjusted image ready for processing

Figure 4.17: Original coral x-ray image (a), the smoothed mask (b) and the background filled image to reduce edge effects (c) which will be used for processing.

As such, the second Meyer type wavelet is used (Figure 3.3) due it its smoother transition in the frequency domain and thus faster spatial decay. Finally, the model is solved using the iterative method.

The model was calculated for five scales (Figure 4.18). Most of the energy of the horizontal polyp lines appears in the first sinusoid of the first and second scales, while most of the energy of the vertical growth rings appears in the first sinusoid in the second to fifth scales. However, we notice that there are still high amplitude vertical patches in the second sinusoid for the first and second scales. Furthermore, in the forth and fifth scales the second sinusoid also has high amplitude in certain areas. In locations where there are horizontal ridge lines, the first sinusoid in the fifth scale represents these features instead of the vertical growth lines, which are instead picked up by the second sinusoid. These results validate the choice of a multi-sinusoidal model, as the model is representing both the fine details of the vertical growth rings, and the broad details of the horizontal ridges, that otherwise would be relegated to the residual in a single sinusoidal model. Most of the energy for the polyp tubes is in scales one to three, and most of the energy for the growth rings is in scales ones to five.

4.6.5 Orientation Classification

The next step to classify each sinusoid in the model as belonging either the polyp tubes class (horizontal), the growth rings class (vertical) or neither. The average orientation of the polyp

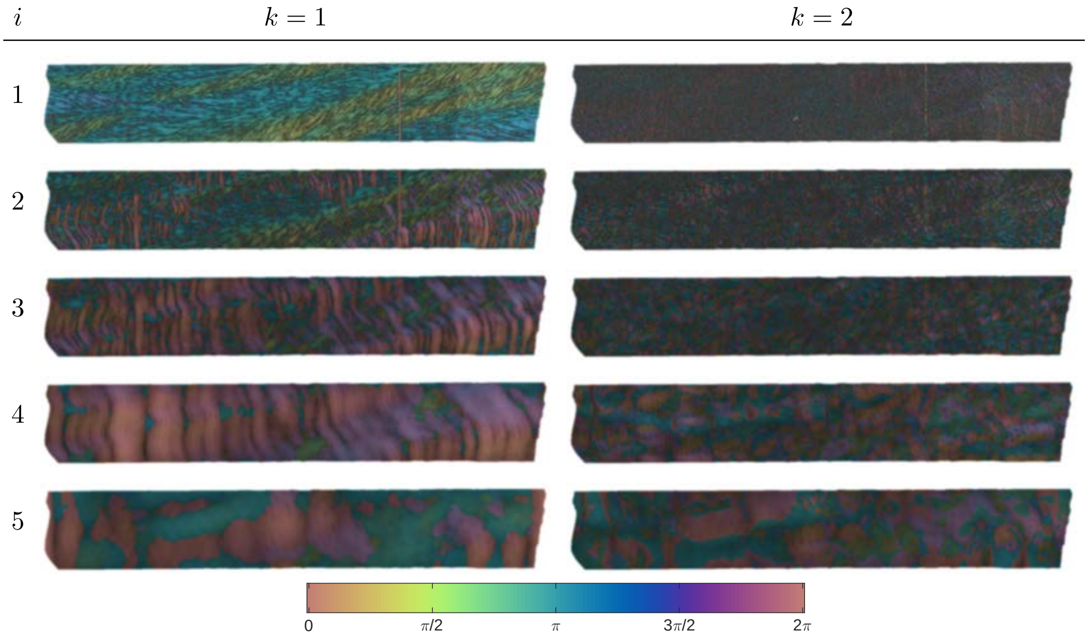


Figure 4.18: Multi-sinusoidal signal representation of the first five scales of the coral image. Brightness represents amplitude A, as $\log(A+2)$; colour represents orientation.

tubes appears to change over the length of the sample and must be adjusted for. The image is split into nine sections along the horizontal axis. The orientation estimates for both sinusoids, all scales, and locations within both the coral segment and the section, are collected for each set,

$$\boldsymbol{\theta}_s = \{(\theta_i(\mathbf{z}))_k \mid k \in [1,2], i \in [1, ...5], m(\mathbf{z}) = 1, \mathbf{z} \in \text{section}(s)\}, \tag{4.46}$$

where i is the scale index, k is the sinusoid component index, s is the section index and $m(\mathbf{z})$ is the mask. k-means clustering is performed to estimate the two predominant orientations for each set. The orientation closest to $\pi/2$ (horizontal) corresponds to the polyp tubes, while the other corresponds to the growth rings. Figure 4.19 shows the combined histogram of the orientations from all the sets along with their class for the example image in Figure 4.17. There is some overlap between the classes due to the changing average orientation as we move horizontally through the image.

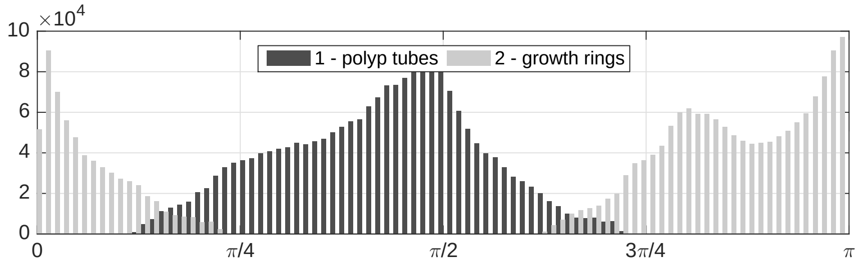


Figure 4.19: Histogram of orientation for the polyp tube and growth rings classes.

The multi-sinusoidal model is now split into two *single* sinusoidal models corresponding to each class according to the orientation parameter. Generally, each sinusoid will belong to a different class. However, if the both have same class, only the largest amplitude sinusoid will be used, and the amplitude of the other sinusoid is set to zero. That is, at a particular scale and location the model parameters are given by

$$\{A, \phi, \theta\}_{\text{class}} = \begin{cases} \{A, \phi, \theta\}_1 & \text{if } \theta_1 \in \boldsymbol{\theta}_{\text{class}}, \\ \{A, \phi, \theta\}_2 & \text{if } \theta_1 \notin \boldsymbol{\theta}_{\text{class}} \text{ and } \theta_2 \in \boldsymbol{\theta}_{\text{class}}, \\ \{0, 0, 0\} & \text{otherwise}, \end{cases} \tag{4.47}$$

where class $\in$ {polyp, growth ring}.

An example of the process is shown in Figure 4.20. At the first scale, the first sinusoid (largest amplitude) mainly belongs to the polyp class, while the second sinusoid mainly belongs to the growth ring class. For scales three and four this is reversed. At the second scale the classes are more evenly split. Again this shows the advantage of a multi-sinusoidal model, as without the second sinusoid all the orientation information in the θ_2 rows would be lost.

4.6.6 Combined Orientation

The orientations for each scale are averaged over I scales to give a final measurement for each class. To do this the amplitude and orientation are expressed as a complex exponential using the double angle representation. The double angle is used because orientation is defined over the half-circle, and thus 0 and π actually represent the same angle. The average representation is

$$m_{\text{class}} = \frac{1}{I} \sum_{i=1}^{I} (A_i)_{\text{class}} e^{\mathrm{i}2(\theta_i)_{\text{class}}} \tag{4.48}$$

where i is the scale index. The overall orientation estimate for each class is then given by the argument

$$\theta_{\text{class}} = \arg(m_{\text{class}})/2. \tag{4.49}$$

The addition of the first I scales is different for each class (Figure 4.21). For the example image, the combined estimate for the polyp class does not appear to change significantly until the addition of the fifth scale, as this scale includes the large horizontal ridges present in the image. The effect can be seen as a change in orientation in the light-green section on the right side of the image between the fourth and fifth scales in Figure 4.21. Since the polyps are small features we shall only use the first three scales to construct the estimate. All five scales appear to improve the estimate for the growth ring class and so all will be used.

The estimates for each class are combined, again by taking the average (Figure 4.21). Because the growth ring orientation is perpendicular to the polyp orientation, it is rotated by 90 degrees

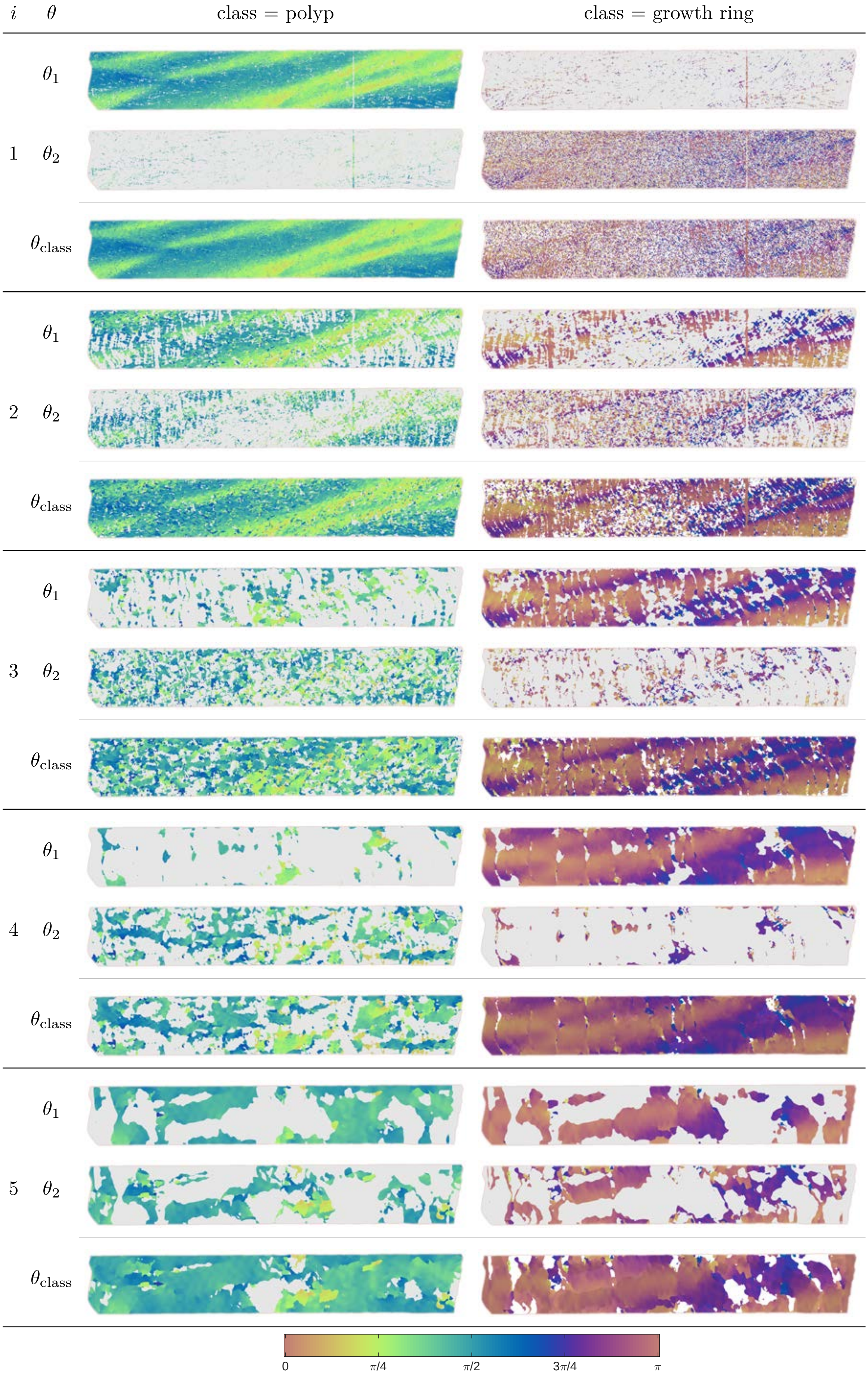


Figure 4.20: Assignment of first (θ_1) or second (θ_2) sinusoid to either the polyp or growth ring classes, for the first five scales.

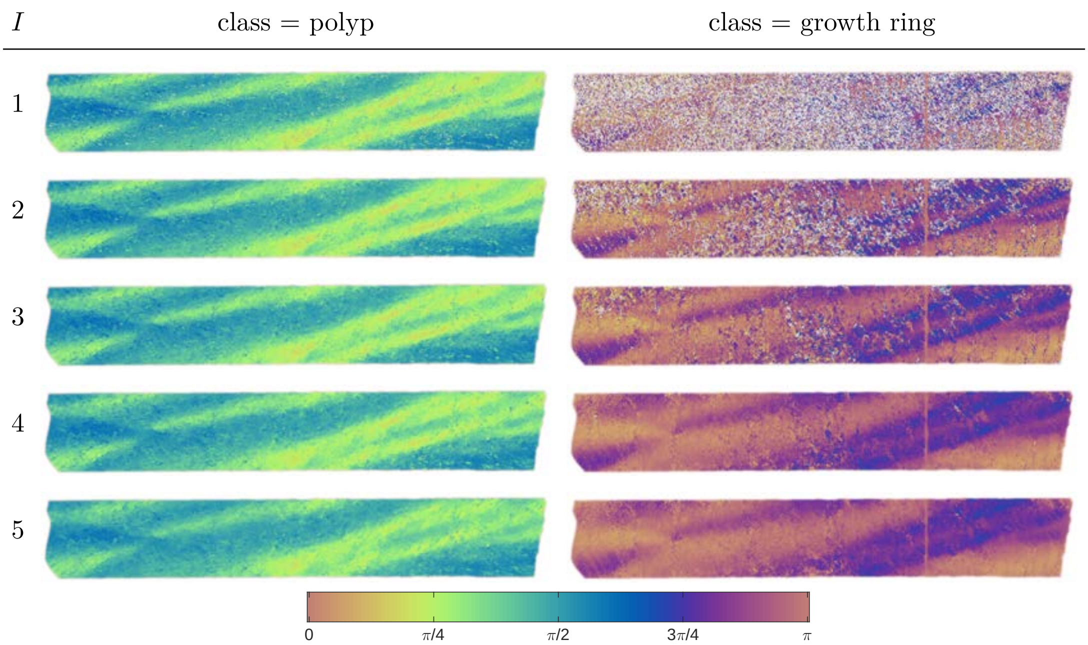


Figure 4.21: Average orientation estimate for each class, calculated using the first to I-th scale. As more scales are used the orientation estimate improves.

to match. That is,

$$m_{\text{combined}}(\mathbf{z}) = m_{\text{polyp}}(\mathbf{z}) + (m_{\text{ring}}(\mathbf{z}) \cdot e^{\mathrm{i}\pi}), \tag{4.50}$$

$$\theta_{\text{combined}}(\mathbf{z}) = \arg(m_{\text{combined}}(\mathbf{z}))/2, \tag{4.51}$$

and finally the estimate is smoothed using a Gaussian filter with $\sigma = 3$

$$m_{\text{smooth}}(\mathbf{z}) = m_{\text{combined}}(\mathbf{z}) * g(\mathbf{z}; 3), \tag{4.52}$$

$$\theta_{\text{smooth}}(\mathbf{z}) = \arg(m_{\text{smooth}}(\mathbf{z}))/2. \tag{4.53}$$

The combination of each estimate and the smoothed result is shown in Figure 4.22. The vertical line in the ring estimate is an artefact of the image. Figure 4.23 shows the estimate super-imposed on the original coral image. A quiver plot is drawn over the image in order to make a qualitative assessment of the accuracy of the estimation as there is no ground truth.

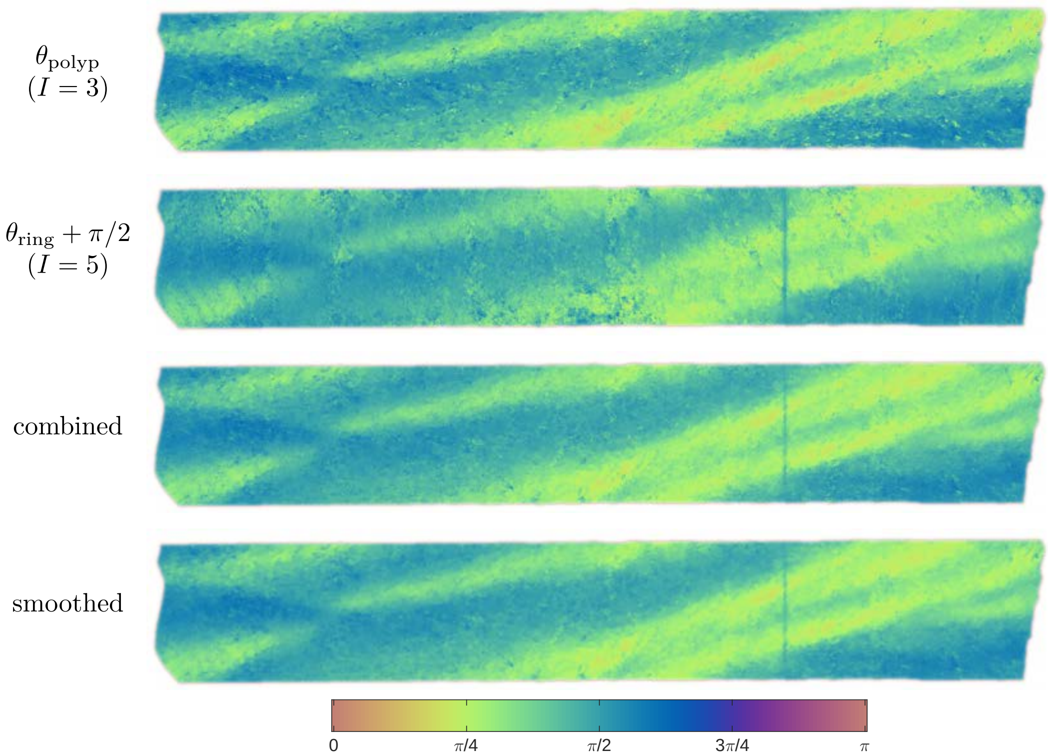


Figure 4.22: Average polyp and growth ring orientations, combined then smoothed.

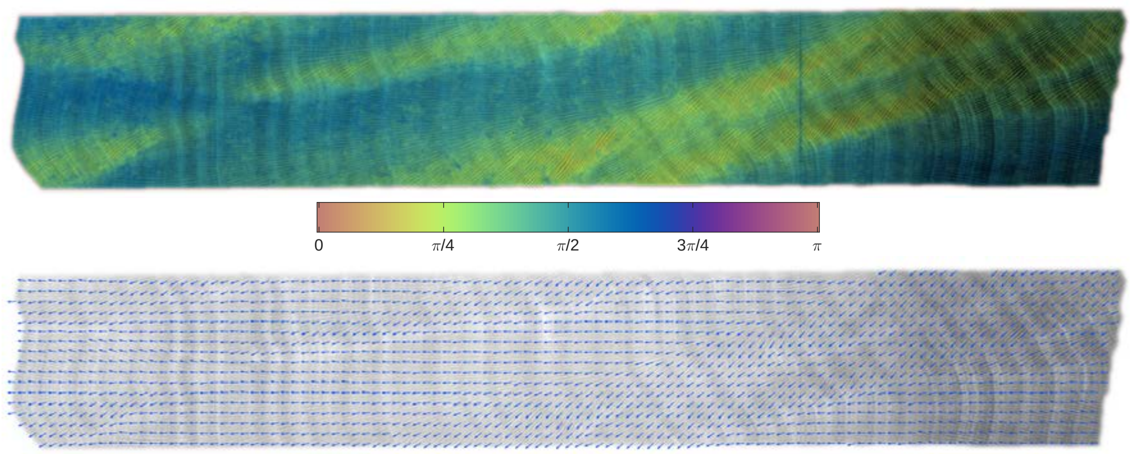


Figure 4.23: Final orientation estimate super-imposed over the original coral image as using colour (top) and a quiver plot (bottom). The quiver plot image was used to make a qualitative assessment of the estimation accuracy by comparing the arrows to the polyp tube lines.

4.6.7 Reconstruction

One of the advantages to using wavelets is that we can reconstruct images from the model components. Adjustment of the model parameters before reconstruction can be used as a method of image processing and modification, as was demonstrated in the last chapter. We shall reconstruct from the polyp and growth ring models to create two separate images of the coral that isolates each feature type. The model CH vector used for reconstruction for each class, at a particular scale and location, is given by

$$(\mathbf{Wf}_\mathrm{S})_\mathrm{class} = \mathbf{Wf}_\mathrm{S}(A, \phi, \theta)_\mathrm{class} \tag{4.54}$$

where

$$\mathbf{Wf}_\mathrm{S}(A, \phi, \theta) = \frac{A}{\sqrt{2}}\mathbf{S}_\theta \mathbf{W}\left(\cos\phi\, \mathbf{f}_e + \sin\phi\, \mathbf{f}_o\right). \tag{4.55}$$

Reconstruction of both classes was performed for the example image using the first five wavelet scales. The original image has been clearly decomposed into polyp and growth ring features (Figure 4.24). In particular, the fine details of the growth rings are preserved where they are present in the image. This would not be possible if a simple smoothing filter had been used to removed the fine polyp features.

However, there are some noisy elements in the reconstruction, particularly for the growth ring class (Figure 4.25b). This effect is due to the reconstruction from all components in the class, even those sinusoidal CH vectors with orientations quite different to the orientation estimate. To compensate we can smooth perpendicular to the global orientation estimate axis by modulating the sinusoidal components according to their deviation from the growth orientation at the point. Various metrics could be used for the modulation. We shall use the power of the cosine of the angle difference between the model component orientation and the smoothed orientation estimate. That is,

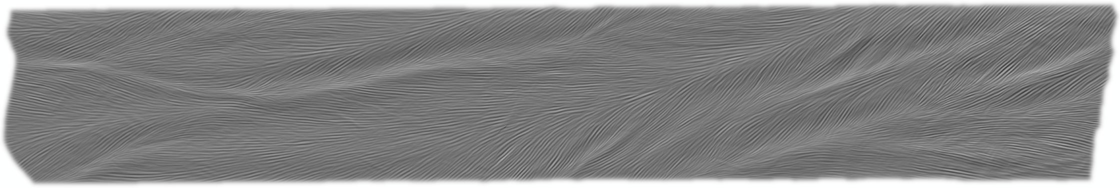
$$\mathbf{Wf}'_\mathrm{S}(A, \phi, \theta) = \mathbf{Wf}_\mathrm{S}(A, \phi, \theta) \times |\cos(\theta - \theta_\mathrm{est.})|^\alpha. \tag{4.56}$$

Increasing the power increases the amount of smoothing (Figure 4.25). Another method that also works well is to use the angular response of the wavelets themselves to determine the modulation,

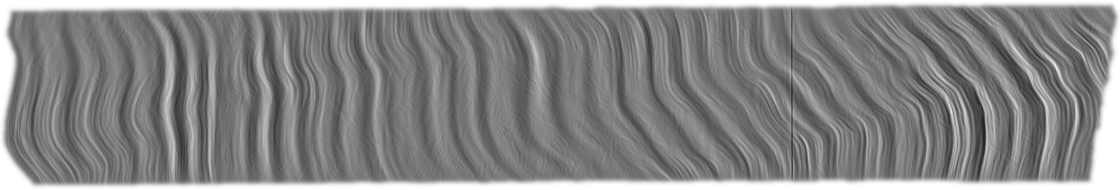
$$\mathbf{Wf}'_\mathrm{S}(A, \phi, \theta) = \mathbf{Wf}_\mathrm{S}(A, \phi, \theta) \times \langle \mathbf{Wf}_\mathrm{S}(A, \phi, \theta), \mathbf{Wf}_\mathrm{S}(A, \phi, \theta_\mathrm{est.})\rangle. \tag{4.57}$$

4.6.8 Optimal Sampling Axis

Armed with the orientation estimate we can now determine the optimal axis along which to measure the coral density. Using the density versus distance measurements, coral calcification rate is

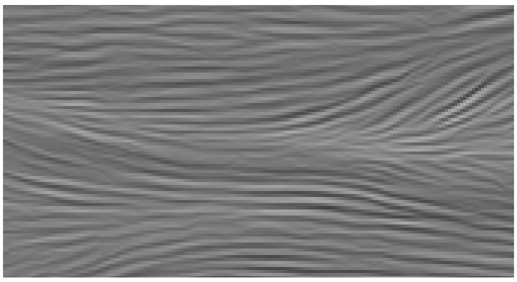


(a) Polyp class

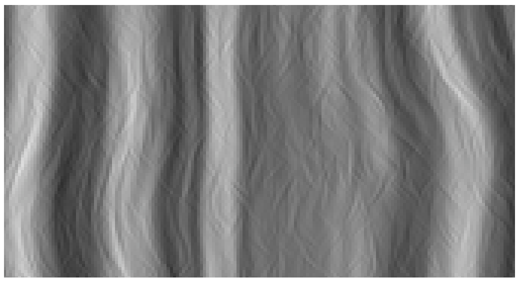


(b) Ring class

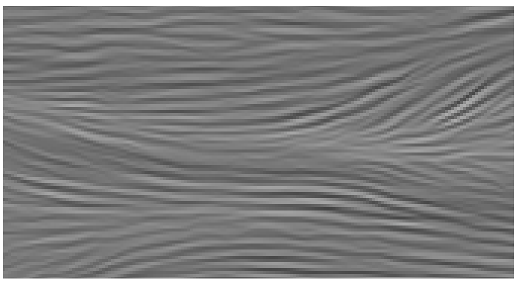
Figure 4.24: Reconstruction of the coral x-ray image from the polyp and growth ring classes of model components, using the first five scales.



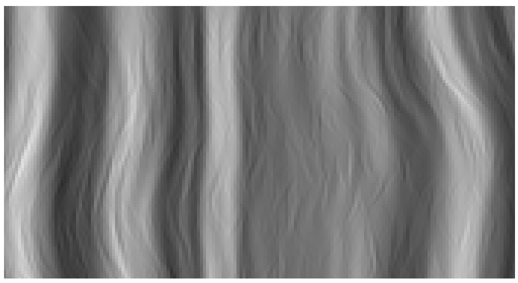
(a) Polyp



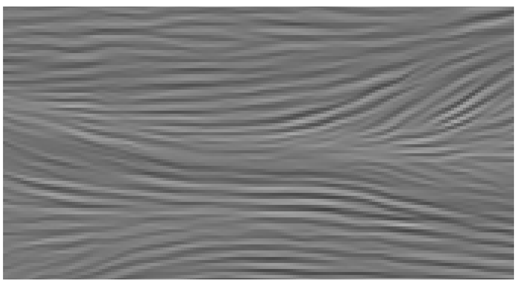
(b) Growth ring



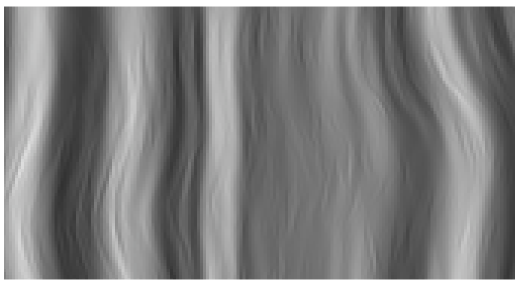
(c) Polyp smoothed ($\alpha = 8$)



(d) Growth ring smoothed ($\alpha = 8$)



(e) Polyp smoothed ($\alpha = 32$)



(f) Growth ring smoothed ($\alpha = 32$)

Figure 4.25: Effect of orientation smoothing on the reconstruction using the cosine of the angle difference between the component orientation and the overall orientation estimate raised to the power of α.

calculated as

$$\text{calcification rate} = \text{density}(x) \cdot \text{extension rate}(x), \tag{4.58}$$

where x is the distance along the axis in the direction of growth. Extension rate is given by

$$\text{extension rate} = \frac{\Delta x}{\Delta t}, \tag{4.59}$$

where t is time and is determined by locating the positions of the yearly growth rings.

These measurements must be taken along the direction of growth, which varies along the length of the coral sample. However, due to the construction of the densitometer, measurements are restricted to being taken along a straight line, and thus the local growth orientation will differ from the orientation along which the measurements are taken. This means the distance travelling along the measurement axis will be greater than the actual growth length, resulting in over-estimation of the extension rate.

For an angle of α radians between the measurement axis and the local orientation axis, the extension rate will be

$$\text{extension rate}(\alpha) = \frac{\Delta x \sec(\alpha)}{\Delta t}. \tag{4.60}$$

The error is thus

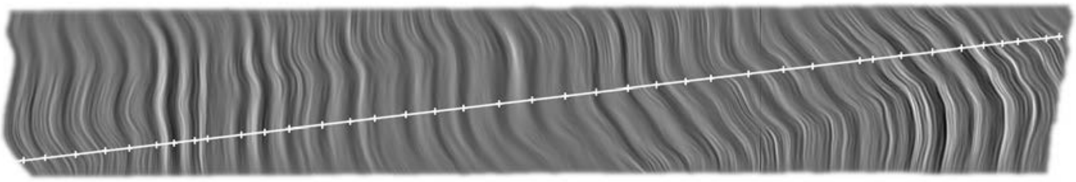
$$\text{error}(\alpha) = \frac{\text{extension rate}(\alpha) - \text{extension rate}(0)}{\text{extension rate}(0)} \tag{4.61}$$

$$= \sec(\alpha) - 1. \tag{4.62}$$

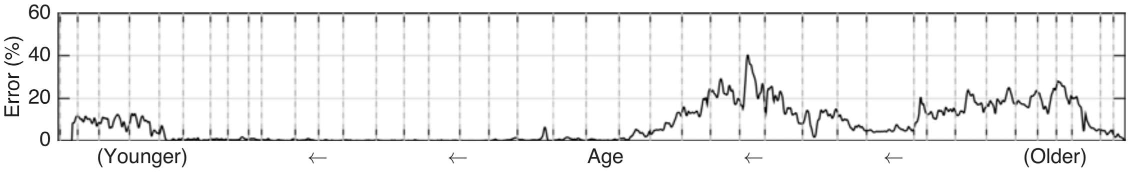
For small angle differences the error is negligible. However if the measurement axis goes through areas of large deviation the error will be high. It is important to note that the error is always *positive*, and therefore does not average out. Instead, the extension rate, and thus the calcification rate, can be corrected by multiplying by $\cos(\alpha)$ to compensate.

The best measurement axis is defined as the straight line path through the image that minimises the square of this error. A rough search was performed using 36 different start positions and 36 different end positions for a total of 1296 paths, and the path with the minimum average error was chosen as the major growth axis (Figure 4.26a). Local minima (high density) of the values of the reconstructed growth ring image along the major growth axis are used to delineate the series into years. The error approaches 40% in some locations of the example image (Figure 4.26b).

The major growth axis and corresponding estimation error were calculated for 12 coral core x-ray samples consisting of three slices from four cores (Figure 4.27). Errors were up to 40% in some samples, and the majority of errors occurred in the growth bands of the older years, which are further from the surface of the coral.



(a) Ring image showing best path and local minima



(b) Extension rate error

Figure 4.26: Image reconstructed from the growth ring model components (a) showing the estimated major growth axis (white line) and the positions of greatest density (white crosses), and the error calculated from the orientation estimate along this axis (b). Vertical dotted lines delineate the estimated yearly growth rings.

4.6.9 Interpretation

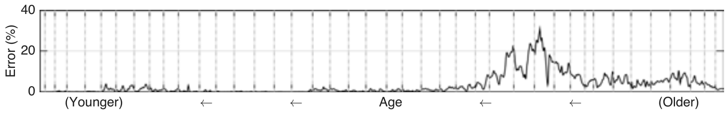
Application of the multi-sinusoidal model provides an estimate of local coral growth orientation from both polyp and growth ring features. It gives a more principled way of finding the major growth axis compared to subjective human estimation by eye. The objective is to measure density along a line as close as possible to the actual orientation. However, where this is not possible, measurements of extension and calcification rate can be compensated using the error estimate given by the cosine of the local orientation angle and the major growth axis angle.

The error is always positive, that is, it leads to an over-estimation of the calcification rate (Figure 4.27). It is interesting to note that higher errors are tend to be in the older growth bands. This may be due to the method of obtaining the cores. The angle of drilling into the coral is a judgement made by the diver using the surface appearance of the coral. However, coral do not grow out in perfect straight lines, and thus the deeper into the coral the more likely the growth axis is different from the drilling axis. This is particularly noticeable in the COB71A-3 sample used in the previous examples.

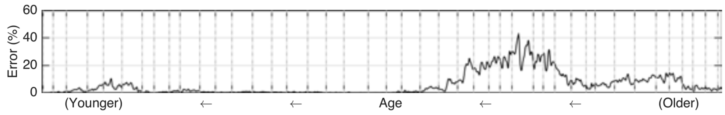
Calcification rates reported in [22] range from 1.5 to 1.75 g/cm^2, and extension rates from 1.25 to 1.42 cm/yr, over the core record from 1900 to 2009. Thus the estimation errors shown in Figure 4.27 are around the same magnitude as the reported changes in calcification rate. Therefore application of the growth estimation technique we have presented may lead to more accurate findings. The errors are also only calculated along the plane of the core slice. It may be that the actual growth axis has a component in the orthogonal plane as well, further exacerbating the errors. A further improvement would be to estimate the orientation along the orthogonal plane, perhaps using the length of the polyp tube lines as a proxy.

Sample	Estimation Error

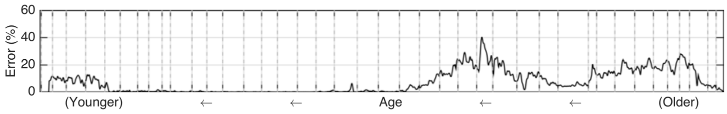
COB71A-1



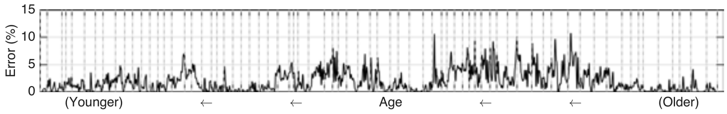
COB71A-2



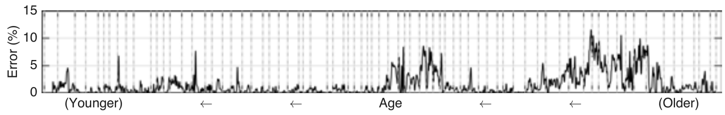
COB71A-3



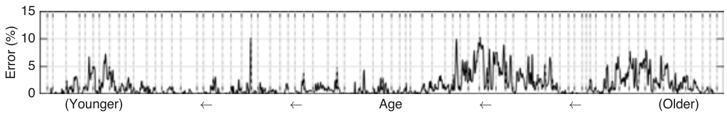
MYR032A-1



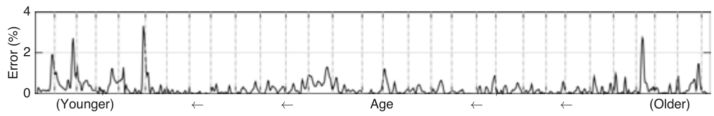
MYR032A-2



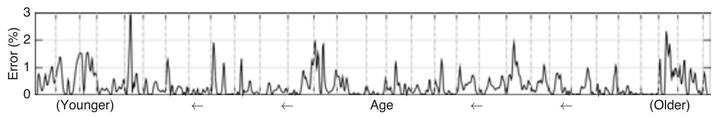
MYR032A-3



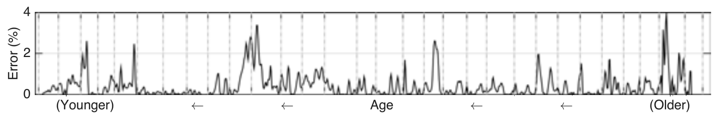
TNT05B1-1



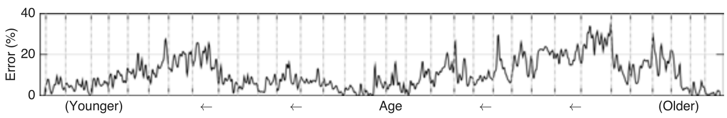
TNT05B1-2



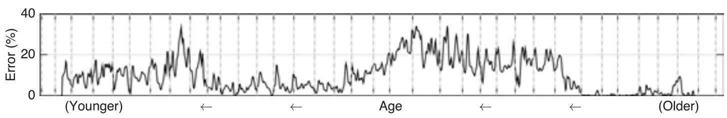
TNT05B1-3



TNT05B4-1



TNT05B4-2



TNT05B4-3

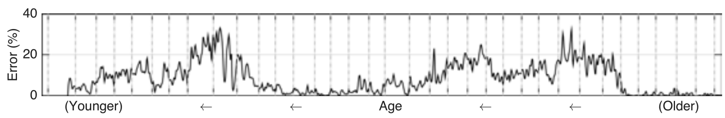


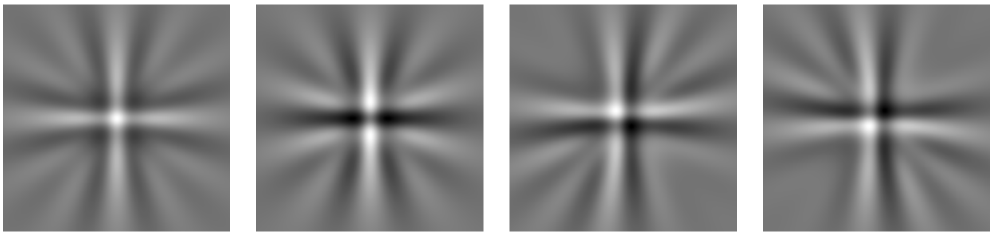
Figure 4.27: Error estimates for 12 coral core samples. COB71A-3 was used in the previous examples in this section. Vertical dotted lines delineate the estimated yearly growth rings.

4.7 Summary

This chapter developed the multi-sinusoidal image model, consisting of multiple sinusoids with different amplitude, phase and orientation plus a residual component. The model can be solved using either the iterative, roots or super-resolution methods. The roots method is the typical approach to analysing local orientation or features with steerable filters. It simply looks for peaks in the angular response polynomial. The iterative method is made possible by starting with the CH vector and having the residual component. It can resolve sinusoidal components closer together in orientation. The super-resolution method gives the best estimation of the model parameters for a given N, but is very slow using the current implementation and therefore impractical for analysing an entire image.

The multi-sinusoidal model can be used for multiple orientation estimation, and is on par or out-performs the contemporary MSMF and MOP methods for the estimation of the orientation of two crossed lines. A two sinusoidal model was applied to the problem of estimating the growth direction in coral core x-rays. The general method can be adapted to multi-scale orientation estimation for any image.

Previous 2D analytic image models [135] used up to two sinusoidal components. This approach extends the model to have any number of components by adding more RT orders. One aspect that has not been explored is to constrain the model. For example, we might constrain a two-sinusoidal model to have the components are right-angles, as is performed for the structure multi-vector in [26]. In this case the wavelet set is expanded to incorporate the different phase permutations (Figure 4.28).



(a) $\phi = [0, 0]$ (b) $\phi = [0, \pi]$ (c) $\phi = [\pi/2, \pi/2]$ (d) $\phi = [\pi/2, -\pi/2]$

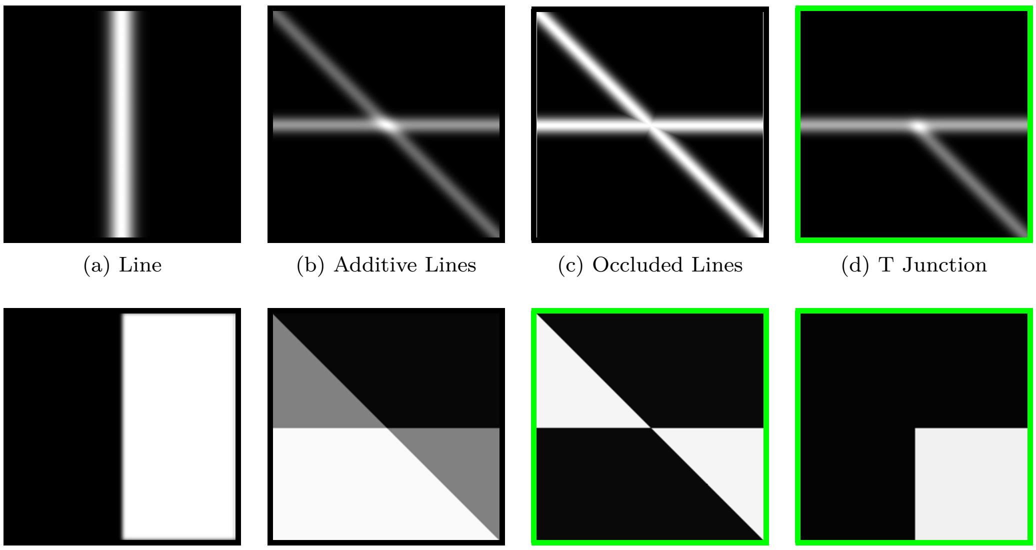
Figure 4.28: Set of wavelets for finding the parameters of a two-sinusoidal model with each component at right-angles, for $N = 13$.

Chapter 5

Half-Sinusoidal Model

5.1 Introduction

Junctions and corners are i2D features that are not well-represented by a single sinusoidal model. The multi-sinusoidal model can be used to model i2D features that are composed of the rotation and addition of i1D components, such as crossed lines. However, other junction and corner structures that look similar are not well-modelled by the multi-sinusoidal model either. For example, the chequer pattern in Figure 5.1g is comprised of edge-like features, but it is actually an even structure that does not respond to odd order RTs. As a result, the corresponding multi-sinusoidal model has multiple components that do not match the orientation of the edges. The corner feature (Figure 5.1h) and T junction (Figure 5.1d) only deliver sinusoids with the correct orientation close to the centre point, after which the residual remains large, resulting in extra high amplitude model components to compensate for the one-sided segments.



(a) Line (b) Additive Lines (c) Occluded Lines (d) T Junction

(e) Edge (f) Additive Edges (g) Chequer Edge (h) Corner

Figure 5.1: Examples of various image features. Junctions and corner features (green) are better modelled using line-segment or wedge-segments rather than a combination of sinusoids.

The use of steerable filters to analyse junctions and corners is an established technique. Both

Perona [98] and Freeman and Adelson [38] demonstrated using 2D steerable quadrature filters for multiple orientation analysis. However, like the multi-sinusoidal model, these give a double response to line or edge segment features and therefore cannot distinguish them from crossed lines. Improving on quadrature filters, Michaelis and Sommer [80] constructed a pair of elongated one-sided odd and even filters using Gaussian derivatives to properly analyse both line and edge-segment junctions. Simoncelli and Farid [110] applied steerable wedge filters constructed using windowed circular harmonics in the spatial domain for the same purpose. However, their filters are not frequency selective and a wedge shape is not the best operator to analyse lines as it becomes larger with distance.

Jacob and Unser [50] also designed a steerable filter using higher-order Gaussian derivatives to analyse single wedge-shaped features, such as Figure 5.1h. The corner angle is calculated from the basis filter coefficients; however, the method does not analyse multiple wedge-like features such as Figure 5.1g. Mota et al. [85] proposed the MOP technique, discussed in Section 4.5, for analysing corners and junctions. However, the orientation estimate is obtained using derivatives in the local area, and thus is only given over the half-circle range, $[0, \pi)$. This means one cannot distinguish between features consisting of either a line or a line-segment at the same orientation, such as Figures 5.1b, 5.1d, 5.1f and 5.1g. Subsequent work on the same approach by Muhlich and Aach [87] also suffers from this restriction. Muhlich and Aach proposed a better alternative, the MSMFs discussed in Section 4.5. MSMFs are steerable wedge filters constructed in the same manner as Simoncelli and Farid's [110]. However, if the number of feature segments is known beforehand, then *multiple* wedge filters are steered *together* so that interference effects caused by closely orientated segments are taken into account.

Some methods have been created specifically using the RT. Zang and Sommer [144] proposed the phase of the monogenic curvature signal as a measure of how line-like or edge-like an intersection is, as well for calculating its angle [112]. Puspoki and Unser [104] designed 2D steerable wavelets to detect the location and overall orientation of segment-type junctions. However, the application is constrained to only junctions where the orientation separation between components is equal, and the junction components are either all line segments or all edge segments. They have recently introduced other wavelets designed to detect specific features [103], such as T junctions, similar in concept to Marchant and Jackway in [73].

In this chapter, a model for parametrising both junction and corner features is proposed. These features shall be described by multiple line segments or wedges radiating from a point. Wedge features can be considered as consisting of two edge segments, with multiple wedges defined by their adjoining edges. For example, Figure 5.1f could be described by four edge segments radiating from the centre, while Figure 5.1d appears as three line segments radiating from the centre.

As a result, a combined line and edge segment model is developed. In this model, wavelets matched to archetypal line and edge segments are solved for as a set with a single orientation parameter (see Section 2.4.2). The amplitude and orientation of the set is invariant to the feature type - line or edge segment - which is instead described by a phase parameter. The model shall be called the *half-sinusoidal* model, for which the multi-sinusoidal model of the previous chapter is a special case corresponding to two half-sinusoidal components constrained to be mirrored pairs. The approach is similar to that proposed by Michaelis and Sommer [80] except that higher-order RT responses within the presented CH vector framework are used.

The chapter layout is as follows:

- The half-sinusoidal model is introduced and the wavelets matched to line-segment and edge-segment archetypes are derived.
- Experiments on the effect of RT orders (N) and weighting in resolving feature segments are performed.
- Use of the model for junction analysis is discussed with particular attention to occluded line-segment junctions.
- The validity of the model as a general image descriptor is compared to the sinusoidal model, with application to orientation estimation.
- The half-sinusoidal wavelets are used to construct junction specific wavelets.

Preliminary work on solving a half-sinusoidal model using RTs has been published in [70, 73].

5.2 Line-Segment Feature

5.2.1 Model

We shall define a line-segment feature as an image structure that can be represented by K *line segments* of different strengths $\{A_k \in \mathbb{R}\}_{k\in\mathbb{N}_K}$ and orientations $\{\theta_k \in [0, 2\pi)\}_{k\in\mathbb{N}_K}$ radiating from a point. Let a single line segment oriented at 0 radians and radiating from the point $\mathbf{z} = 0$ be represented by the function $f_{\mathrm{L}}(\mathbf{z})$. The local image structure can be modelled as line-segment feature consisting of multiple additive line segments plus a residual component,

$$f(\mathbf{z}) = \sum_{k=1}^{K} A_k f_{\mathrm{L}}(\mathbf{R}_{\theta_k}\mathbf{z}) + f_\epsilon(\mathbf{z}), \tag{5.1}$$

where $\mathbf{R}_\theta$ is an matrix that rotates the image axes by θ. The image CH vector can therefore be written according to the model as

$$\mathbf{Wf} = \sum_{k=1}^{K} A_k \mathbf{S}_{\theta_k}\mathbf{W}\mathbf{u}_{\mathrm{L}} + \mathbf{W}\boldsymbol{\epsilon}, \tag{5.2}$$

where $\mathbf{W}\mathbf{u}_{\mathrm{L}}$ is the weighted normalised CH vector of the wavelet matched to an ideal line segment.

5.2.2 Matched Wavelet

The line-segment matched wavelet will now be found from the RT integral equations. Consider an idealised line segment consisting of a zero-width line. In polar coordinates, $\mathbf{z} = [r\cos\theta, r\sin\theta]$, an ideal line segment with amplitude, A_k, and orientation, θ_k, can be written as

$$f_{\mathrm{L}(A_k,\theta_k)}(r,\theta) = A_k \frac{\delta(\theta-\theta_k)}{r} \tag{5.3}$$

and an ideal line-segment *feature* is thus given by

$$\sum_{k=1}^{K} A_k \frac{\delta(\theta-\theta_k)}{r}. \tag{5.4}$$

Let ψ be a suitable primary isotropic wavelet with N or more vanishing moments and frequency response given by $h(k)$, where k is the radial part of the polar coordinates of the spectrum, $\boldsymbol{\omega} = [k\cos\phi, k\sin\phi]$. The spatial domain expression of the n-th order CH wavelet $\mathcal{R}^n\psi$ is

$$\mathcal{R}^n\psi = \frac{1}{2\pi}\int_{\boldsymbol{\omega}\in\mathbb{R}^2\setminus B_\epsilon(0)} \left(\frac{\omega_x + \mathrm{i}\omega_y}{\|\boldsymbol{\omega}\|}\right)^n h(k)\; e^{\mathrm{i}\langle\boldsymbol{\omega},\mathbf{z}\rangle}\, d\boldsymbol{\omega},$$

or in polar coordinates,

$$\mathcal{R}^n\psi = \frac{1}{2\pi}\int_0^\infty h(k)\left\{\int_0^{2\pi} e^{\mathrm{i}n\phi} e^{\mathrm{i}kr\cos(\phi-\theta)} d\phi\right\} k\,dk, \tag{5.5}$$

where $\boldsymbol{\omega} = [k\cos\phi, r\sin\phi]$. Substituting $\phi' = \phi - \theta$ we have

$$\mathcal{R}^n\psi = \frac{1}{2\pi} e^{\mathrm{i}n\theta}\int_0^\infty h(k)\left\{\int_0^{2\pi} e^{\mathrm{i}n\phi'} e^{\mathrm{i}kr\cos(\phi')} d\phi'\right\} k\,dk.$$

The Bessel function is defined as

$$J_n(k) = \frac{1}{2\pi}\int_0^{2\pi} e^{\mathrm{i}k\sin\phi} e^{-\mathrm{i}n\phi}\, d\phi \tag{5.6}$$

or with a change of variable as

$$J_n(k) = \frac{(-\mathrm{i})^n}{2\pi}\int_0^{2\pi} e^{\mathrm{i}k\cos\phi} e^{-\mathrm{i}n\phi}\, d\phi. \tag{5.7}$$

Substituting in the Bessel function we arrive at

$$\mathcal{R}^n\psi = \frac{1}{(-\mathrm{i})^n} e^{\mathrm{i}n\theta}\int_0^\infty h(k) J_{-n}(kr)\, k\,dk.$$

And finally, since $J_{-n}(k) = (-1)^n J_n(k)$ for $n \in \mathbb{N}^+$

$$\mathcal{R}^n\psi = (-\mathrm{i})^n e^{\mathrm{i}n\theta}\int_0^\infty h(k) J_n(kr)\, k\,dk.$$

As expected, the CH wavelet spatial domain expression is polar separable with an angular part depending on θ, and a radial part depending on both the RT order, n, and the primary wavelet frequency response, $h(k)$. The n-th order CH wavelet response to a ideal line segment f_{L} with amplitude A and orientation θ' is given by their correlation,

$$\langle f_{\mathrm{L}(A,\theta')}, \mathcal{R}^n\psi\rangle = (f_{\mathrm{L}(A,\theta')} * \mathcal{R}^{-n}\phi)(\mathbf{0}) \tag{5.8}$$

$$= \int_0^{2\pi}\int_0^{\infty} A\frac{\delta(\theta-\theta')}{r}\left\{(-\mathrm{i})^n e^{-\mathrm{i}n\theta}\int_0^{\infty} h(k)J_n(kr)\,k\,dk\right\} r\,dr\,d\theta \tag{5.9}$$

$$= A(-\mathrm{i})^n e^{-\mathrm{i}n\theta'}\int_0^{\infty}\int_0^{\infty} h(k)J_n(kr)\,k\,dk\,dr. \tag{5.10}$$

Since $\int_0^\infty J_n(kr)dr = 1/k$ for $k \geq 0$ we have

$$\langle f_{\mathrm{L}(A,\theta')}, \mathcal{R}^n\psi\rangle = A(-\mathrm{i})^n e^{-\mathrm{i}n\theta'}\int_0^{\infty} h(k)\,dk. \tag{5.11}$$

So long as $\int_0^\infty h(k)\,dk$ is finite, the magnitude of the response for each order is equal. Using the conjugate relation between the negative and positive CH orders (1.40), the CH vector corresponding to the ideal line segment is thus

$$\mathbf{f}_{\mathrm{L}}(A,\theta) = cA\left[(-\mathrm{i})^{|-N|}e^{-\mathrm{i}N\theta},\dots,(-\mathrm{i})^{|N|}e^{\mathrm{i}N\theta'}\right]^T \tag{5.12}$$

$$= cA\mathbf{S}_\theta\left[(-\mathrm{i})^{|-N|},\dots,(-\mathrm{i})^{|N|}\right]^T, \tag{5.13}$$

where $c = \int_0^\infty h(k)\,dk$ is a constant depending on the isotropic basis wavelet. It follows that $\mathbf{W}\mathbf{u}_{\mathrm{L}}$, the normalised CH vector for an ideal line segment orientated 0 degrees, is for a weighting matrix $\mathbf{W}$ given by

$$\mathbf{W}\mathbf{u}_{\mathrm{L}} = \frac{\mathbf{W}\mathbf{f}_{\mathrm{L}}}{\|\mathbf{W}\mathbf{f}_{\mathrm{L}}\|} \tag{5.14}$$

$$= \left[w_{-N}\,(-\mathrm{i})^{|-N|},\dots,w_N\,(-\mathrm{i})^{|N|}\right]^T. \tag{5.15}$$

This line segment CH vector can now be used to solve the multiple line-segment image model in (5.2).

5.3 Wedge-Segment Features

We shall define a wedge segment feature as an images structure that can be represented by K *wedge segments* of different strengths $\{A_k \in \mathbb{R}^+\}_{k\in\mathbb{N}_K}$, each having two edges with orientations $\{\theta_{k,1},\theta_{k,2} \in [0,2\pi)\}_{k\in\mathbb{N}_K}$ radiating from a point. Let the wedge segment be represented by the function $f_{\mathrm{W}(\theta_1,\theta_2)}(\mathbf{z})$ with the segment edges radiating from the point $\mathbf{z} = 0$. The local image structure can be modelled by a wedge-segment feature consisting multiple wedge segments plus a residual component,

$$f(\mathbf{z}) = \sum_k A_k f_{\mathrm{W}(\theta_{k,1},\theta_{k,2})}(\mathbf{z}) + f_\epsilon(\mathbf{z}), \tag{5.16}$$

where the wedge feature is given by

$$f_{\mathrm{W}(\theta_1,\theta_2)}(r,\theta) = \begin{cases} 1 & \text{where } \theta \in [\theta_1, \theta_2), \\ 0 & \text{otherwise.} \end{cases} \tag{5.17}$$

The image CH vector can therefore be written as

$$\mathbf{Wf} = \sum_{k=1}^{K} A_k \mathbf{Wf}_{\mathrm{W}(\theta_{k,1},\theta_{k,2})} + \mathbf{W}\boldsymbol{\epsilon}, \tag{5.18}$$

where $\mathbf{Wu}_{\mathrm{W}}$ is the weighted normalised CH vector matched to the wavelet matched to the ideal wedge segment.

5.3.1 Matched Wavelet

The wedge-segment matched wavelet will now be found from the RT integral equations. Using the same primary isotropic wavelet, ψ, as before, the correlation of the wedge segment with the n-th order CH wavelet is given by

$$\langle A f_{\mathrm{W}(\theta_1,\theta_2)}, \mathcal{R}^n\psi\rangle = \int_0^{2\pi}\int_0^{\infty} A f_{\mathrm{W}(\theta_1,\theta_2)} \left\{(-\mathrm{i})^n e^{-\mathrm{i}n\theta} \int_0^{\infty} h(k) J_n(kr)\, k\, dk\right\} r\, dr\, d\theta \tag{5.19}$$

$$= A(-\mathrm{i})^n \int_{\theta_1}^{\theta_2} e^{-\mathrm{i}n\theta}\, d\theta \int_0^{\infty}\int_0^{\infty} h(k) J_n(kr)\, k\, dk\, r\, dr. \tag{5.20}$$

The integrals involving k and r do not yield readily. Integrating with respect to r first does not converge [40] while integrating with respect to k first requires a function for $h(k)$ with an analytical expression for its integral. For example, if $h(k)$ represents the Cauchy kernel (Figure 3.3), then the double integral can be evaluated for certain N [40].

Instead consider the Bessel function recurrence relation for $n \in \mathbb{N}$,

$$\frac{2n}{x} J_n(x) = J_{n-1}(x) + J_{n+1}(x). \tag{5.21}$$

Substituting $x = kr$ and rearranging we get

$$J_n(kr) = \frac{kr}{2n} J_{n-1}(kr) + \frac{kr}{2n} J_{n+1}(kr). \tag{5.22}$$

Now consider the integral

$$\int_0^{\infty}\int_0^{\infty} h(k) J_n(kr)\, dk\, dr. \tag{5.23}$$

It is similar to the integral we need to solve, but without the extra k and r terms that make the integral difficult. It simplifies to

$$\int_0^\infty \int_0^\infty h(k) J_n(kr)\, dk\, dr = \int_0^\infty \frac{h(k)}{k}\, dk \tag{5.24}$$

which importantly does not depend on the order, n. Using the recurrence relation we can also write this integral as

$$\int_0^\infty \frac{h(k)}{k}\, dk = \tag{5.25}$$

$$\frac{1}{2n} \int_0^\infty \int_0^\infty h(k) J_{n-1}(kr)\, k\, dk\, r\, dr + \frac{1}{2n} \int_0^\infty \int_0^\infty h(k) J_{n+1}(kr)\, k\, dk\, r\, dr. \tag{5.26}$$

For simplicity define

$$D = \int_0^\infty \frac{h(k)}{2k}\, dk \tag{5.27}$$

$$h_n = \int_0^\infty \int_0^\infty h(k) J_n(kr)\, k\, dk\, r\, dr. \tag{5.28}$$

Substituting into the previous recurrence relationship we simplify to

$$2D = \frac{1}{n} h_{n-1} + \frac{1}{n} h_{n+1}. \tag{5.29}$$

Since $h(k)$ represents a primary isotropic filter kernel with zero mean, the 0th order response to a wedge function is also 0 (5.20), and therefore $h_0 = 0$. Thus we may write

$$2D = h_0 + h_2 \tag{5.30}$$

$$= h_2, \tag{5.31}$$

which relates the even orders to the odd orders,

$$2D = h_1/2 + h_3/2, \tag{5.32}$$

$$h_2 = h_1/2 + h_3/2. \tag{5.33}$$

Now we shall relate the odd orders to each other. A wedge with angles $\theta_1 = 0$ and $\theta_2 = \pi$ is an edge oriented at $\theta = \pi/2$. From the sinusoidal model we know that the magnitude of the response of odd order CH wavelets to an edge is the same. That is,

$$\langle f_{\mathrm{W}}(-\pi/2, \pi/2), \mathcal{R}^n \psi \rangle = A e^{-\mathrm{i}n\pi/2} \mathrm{i} \sin(-\pi/2) \tag{5.34}$$

$$= A(-\mathrm{i})^{n+1}. \tag{5.35}$$

Then from (5.20) and (5.28) we may write, for odd n,

$$A(-\mathrm{i})^n h_n \int_{-\pi/2}^{\pi/2} e^{-\mathrm{i}\theta}\, d\theta = A(-\mathrm{i})^{n+1}, \tag{5.36}$$

and thus

$$\begin{aligned}
\mathrm{i}A\, h_1 \int_0^{\pi} e^{-\mathrm{i}(\theta-\pi/2)}\, d\theta &= \mathrm{i}A\, h_n \int_0^{\pi} e^{-\mathrm{i}n(\theta-\pi/2)}\, d\theta, &(5.37)\\
h_1 \int_0^{\pi} e^{\mathrm{i}\theta}\, d\theta &= h_n \int_0^{\pi} e^{\mathrm{i}n\theta}\, d\theta, &(5.38)\\
-2\mathrm{i}h_1 &= \frac{-2\mathrm{i}}{n} h_n, &(5.39)\\
nh_1 &= h_n, &(5.40)
\end{aligned}$$

for odd values of n. Substituting in to (5.33) we find for all n that

$$\begin{aligned}
h_n &= nh_1 &(5.41)\\
&= nD &(5.42)
\end{aligned}$$

which relates the responses for each order. For an ideal wedge segment the correlation with a particular order CH wavelet is thus

$$\begin{aligned}
\langle Af_{W(\theta_1,\theta_2)}, \mathcal{R}^n\psi\rangle &= A(-\mathrm{i})^n h_n \int_{\theta_1}^{\theta_2} e^{-\mathrm{i}n\theta}\, d\theta &(5.43)\\
&= A(-\mathrm{i})^n \left(\frac{-\mathrm{i}(e^{-\mathrm{i}n\theta_1} - e^{-\mathrm{i}n\theta_2})}{n} \right) nD &(5.44)\\
&= A(-\mathrm{i})^{n+1} D \left(e^{\mathrm{i}n\theta_1} - e^{\mathrm{i}n\theta_2} \right). &(5.45)
\end{aligned}$$

Using the conjugate relation between the negative and positive CH orders (1.40), the CH vector corresponding to the wedge is thus

$$AD\,(\mathbf{S}_{\theta_2} - \mathbf{S}_{\theta_1}) \left[(-\mathrm{i})^{|-N+1|}, ..., (-\mathrm{i})^{|N+1|} \right]^T. \tag{5.46}$$

It follows that $\mathbf{W}\mathbf{u}_{\mathrm{W}}$, the normalised CH vector for an ideal line segment orientated 0 degrees, is for a weighting matrix $\mathbf{W}$ given by

$$\begin{aligned}
\mathbf{W}\mathbf{u}_{\mathrm{W}} &= \frac{\mathbf{W}\mathbf{f}_{\mathrm{W}}}{\|\mathbf{W}\mathbf{f}_{\mathrm{W}}\|} &(5.47)\\
&= \frac{(\mathbf{S}_{\theta_2} - \mathbf{S}_{\theta_1}) \left[w_{-N}(-\mathrm{i})^{|-N+1|}, ..., w_N(-\mathrm{i})^{|N+1|} \right]^T}{\left\| (\mathbf{S}_{\theta_2} - \mathbf{S}_{\theta_1}) \left[w_{-N}(-\mathrm{i})^{|-N+1|}, ..., w_N(-\mathrm{i})^{|N+1|} \right]^T \right\|}. &(5.48)
\end{aligned}$$

The wedge segment CH vector can be used to solve the multiple wedge-segment image model in (5.18) if the wedges are constrained to have a fixed orientation.

5.3.2 Edge Segment

In the previous equation (5.48), the wedge segment CH vector is expressed as the difference between a vector rotated to θ_1 and the same vector rotated to θ_2, which are the orientations of the wedge edges. We shall call this vector oriented at 0 degrees the *edge segment* vector $\mathbf{u}_{\mathrm{E}}$. An edge segment vector at orientation θ is thus given by

$$\mathbf{f}_{\mathrm{E}(A,\theta)} = A\mathbf{S}_\theta \left[(-\mathrm{i})^{|-N+1|}, ..., (-\mathrm{i})^{|N+1|}\right]^T. \tag{5.49}$$

In the above equation, f_{E_0}, the 0-th order response to the to edge segment, is non-zero and set to $-\mathrm{i}$. The actual value could be anything, since for a wedge feature the edge segment vector are of equal and opposite amplitudes and thus the 0-th order components cancel out.

However, the value does have an effect on the corresponding edge wavelet and the CH vector that represents it. If the 0-th order is set to 1 or $-\mathrm{i}$ for example, then the normalised matched CH vector is simply given by

$$\mathbf{W}\mathbf{u}_{\mathrm{E}} = \frac{\mathbf{W}\mathbf{f}_{\mathrm{E}}}{\|\mathbf{W}\mathbf{f}_{\mathrm{E}}\|} \tag{5.50}$$

$$= \mathbf{W}\mathbf{f}_{\mathrm{E}}. \tag{5.51}$$

Interestingly, for this edge wavelet the CH vector weights are $-\mathrm{i}$ times that of the line wavelet. However, it also means that the 0-th CH wavelet is imaginary, and thus the response to the edge wavelet has an imaginary part which is undesirable. If the 0-th order is set to 0, then the normalised matched CH vector is given by

$$\mathbf{W}\mathbf{u}_{\mathrm{E}} = \frac{\mathbf{W}\mathbf{f}_{\mathrm{E}}}{\|\mathbf{W}\mathbf{f}_{\mathrm{E}}\|} \tag{5.52}$$

$$= \mathbf{W}\mathbf{f}_{\mathrm{E}}/\sqrt{W_{\mathrm{E}}}, \tag{5.53}$$

where

$$W_{\mathrm{E}} = 1 - w_o^2. \tag{5.54}$$

For this edge wavelet the CH vector values are $-\mathrm{i}$ times that of the line wavelet only if the line wavelet 0-th order is also set to 0. When solving using iterative and root finding methods one should set the 0th order of the edge wavelet to 0. When solving using super-resolution it must be set to $-\mathrm{i}$.

An alternative, more general expression for the wedge CH vector model (5.18) is to represent it by its edges. That is, the sum of K rotated and scaled *edge* segment CH vectors

$$\mathbf{W}\mathbf{f} = \sum_{k=1}^{K} A_k \mathbf{S}_{\theta_k} \mathbf{W}\mathbf{u}_{\mathrm{E}} + \mathbf{W}\boldsymbol{\epsilon}, \tag{5.55}$$

where $A_k \in \mathbb{R}$ can be both positive and negatively valued and is constrained by

$$\sum_{k=1}^{K} A_k = 0. \tag{5.56}$$

Thus using the edge-segment CH vector we are able to solve wedge image models with *arbitrary* orientations.

5.4 Half-Sinusoidal Signal Model

We shall combine both the line-segment and edge-segment into a single representation that models the local image structure as copies of the additive combination of a line-segment and edge-segment pair at multiple amplitudes and orientations. That is

$$\mathbf{Wf} = \sum_{k \leq K} \lambda_{\mathrm{L}_k} \mathbf{S}_{\theta_k} \mathbf{W}\mathbf{u}_{\mathrm{L}} + \lambda_{\mathrm{E}_k} \mathbf{S}_{\theta_k} \mathbf{W}\mathbf{u}_{\mathrm{E}}. \tag{5.57}$$

We can express this in amplitude, phase and orientation parameters according to

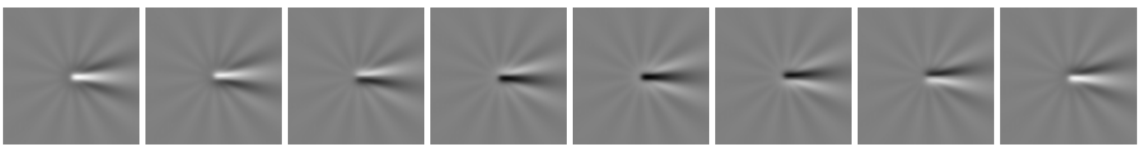
$$\mathbf{Wf} = \sum_{k=1}^{K} A_k \cos\phi_k \mathbf{S}_{\theta_k} \mathbf{W}\mathbf{u}_{\mathrm{L}} + A_k \sin\phi_k \mathbf{S}_{\theta_k} \mathbf{S}_{\theta_k}, \mathbf{W}\mathbf{u}_{\mathrm{E}} \tag{5.58}$$

where

$$A_k = |\lambda_{\mathrm{L}_k} + \mathrm{i}\lambda_{\mathrm{E}_k}|, \tag{5.59}$$

$$\phi_k = \arg(\lambda_{\mathrm{L}_k} + \mathrm{i}\lambda_{\mathrm{E}_k}). \tag{5.60}$$

The amplitude parameter, A_k, thus represents feature strength separately to the *phase* parameter, ϕ_k, that represents feature type, that is, line segment or edge segment. An example of the wavelets corresponding to different values of ϕ is shown in Figure 5.2. The model shall be called the *half-sinusoidal* model, as the even sinusoidal wavelet may be constructed from two equal amplitude line-segments orientated 180 degree apart, corresponding to $K = 2$, $\phi_1 = \phi_2 = 0$ and $\theta_1 = 0$, $\theta_2 = \pi$, and the odd sinusoidal wavelet from two opposite amplitude edge-segments oriented 180 degrees apart, corresponding to $K = 2$, $\phi_1 = \pi/2$, $\phi_2 = -\pi/2$ and $\theta_1 = 0$, $\theta_2 = \pi$. The half-sinusoidal model can therefore represent i1D features, edge-segment and line-segment corners, and edge-segment and line-segment junctions. That is, all of the example features in Figure 5.1.



(a) 0 (b) $\pi/4$ (c) $\pi/2$ (d) $3\pi/4$ (e) π (f) $-3\pi/4$ (g) $-\pi/2$ (h) $-\pi/4$

Figure 5.2: Example of half-sinusoidal kernels for different phase values and $N = 12$

5.5 Model Solution

Each of the three models, line-segment, edge-segment and half-sinusoidal, may be solved using the iterative or roots methods, while the only the line-segment and edge segment models can be solved using the super-resolution method. This is because unlike the odd and even sinusoidal wavelets, the line-segment and edge-segment wavelets are not orthogonal with rotation.

Iterative and Roots Methods

The procedure for solving the half-sinusoidal model using the iterative or roots methods is similar to that performed for the multi-sinusoidal model. The orientation is given by the maximum (iterative) or local maxima (roots) of the polynomial

$$p(\theta) = \lambda_L(\theta)^2 + \lambda_E(\theta)^2, \tag{5.61}$$

where $\lambda_L(\theta)$ and $\lambda_L(\theta)$ are the angular responses of a line segment or edge segment respectively.

Recall that the odd and even sinusoidal model CH vectors have either non-zero odd or non-zero even orders, and thus the polynomial to solve is only degree $2N$. In contrast, the half-sinusoidal model components have all non-zero orders, and therefore the polynomial to solve will be degree $4N$. Thus for the same size CH vector, solving the half-sinusoidal model is more computationally intensive. Note that the 0th order of the edge wavelet *must be* real for the above polynomials to be real valued and thus calculation using root finding to work properly.

If using only the line-segment or edge-segment model, we can instead solve the simpler $2N$ degree polynomials $\lambda_L(\theta)$ or $\lambda_E(\theta)$, respectively (Section 2.6.1). Furthermore, if one wishes to limit the line-segment model to having either only positively or negatively valued components, then the solution may be found by only considering possible orientations at which the polynomial values match the desired sign (Section 2.3.1).

Super-resolution

The super-resolution method can only be applied to either the line-segment or edge-segment models, and not the half-sinusoidal model. The image CH vector for the line-segment model is given by

$$\mathbf{Wf} = \sum_{k=1}^{K} A_k \mathbf{S}_{\theta_k} \mathbf{W} \mathbf{u}_{\mathrm{L}} + \mathbf{W}\boldsymbol{\epsilon}. \tag{5.62}$$

Because all the components are non-zero, the super-resolution method is valid. As the method divides out any weighting, we may write this as

$$\mathbf{f} = \sum_{k \leq K} A_k \mathbf{S}_{\theta_k} \mathbf{u}_{\mathrm{L}} + \boldsymbol{\epsilon} \tag{5.63}$$

and thus

$$f_n = A_k e^{-\mathrm{i}n\theta}(-\mathrm{i})^{|n|} + \boldsymbol{\epsilon}. \tag{5.64}$$

Letting $g_n = f_n/(-\mathrm{i})^{|n|}$, the coefficients g_n are equal to the Fourier series coefficients of a spike train given by

$$x(\theta) = \sum_k A_k \delta(\theta - \theta_k) \tag{5.65}$$

which can then be solved for using the super-resolution method.

Likewise the CH vector for the edge-segment model is given by

$$\mathbf{Wf} = \sum_{k=1}^{K} A_k \mathbf{S}_{\theta_k} \mathbf{W} \mathbf{u}_{\mathrm{E}} + \mathbf{W}\boldsymbol{\epsilon}. \tag{5.66}$$

The super-resolution method can be only applied if the components are all non-zero, therefore we use the version of the edge wavelet where the 0th order is set to $-\mathrm{i}$. As the method divides out any weighting, we may write this as

$$\mathbf{f} = \sum_{k \leq K} A_k \mathbf{S}_{\theta_k} \mathbf{u}_{\mathrm{E}} + \boldsymbol{\epsilon} \tag{5.67}$$

and thus

$$f_n = A_k e^{-\mathrm{i}n\theta}(-\mathrm{i})^{|n+1|} + \boldsymbol{\epsilon}. \tag{5.68}$$

Letting $g_n = f_n/(-\mathrm{i})^{|n+1|}$, the coefficients g_n are equal to the Fourier series coefficients of a spike train given by

$$x(\theta) = \sum_k A_k \delta(\theta - \theta_k) \tag{5.69}$$

which again can be solved for using the super-resolution method.

Unfortunately, the super-resolution method cannot be applied directly to the half-sinusoidal model as the line and edge wavelets are not orthogonal for all orientations, unlike the even and odd sinusoidal wavelets. This means that a line segment wavelet will respond to an edge segment feature at a certain orientation offset, resulting in spurious components in the spike train representation.

5.5.1 Example

The half-sinusoidal model was calculated for the first scale of the *Pentagon* image, for $K = 4$ components and $N = 13$ RT orders, using the iterative method and the second Meyer wavelets (Figure 5.3). The sinusoidal model is shown for comparison on the bottom row. We observe that line and edge features require two line or edge segment components to represent, as indicated by the strong amplitudes of both the first and second components at the location of i1D features.

Because an i1D feature presents two peaks in the angular response of a half-sinusoidal component, the orientation and phase estimates flip by π radians between some locations, depending on the relative strength of each model component. The normal sinusoidal model appears to provide a better description at these locations. However, two half-sinusoidal components appear better able to represent sharp curves than a single sinusoidal model component. This is shown by locations along the curve in the lower right quadrant of the image having a high residual for the sinusoidal model, but not for two components of the half-sinusoidal model.

5.6 Model Accuracy

This section follows a similar development to Section 4.3 in the previous chapter. Each solution method shall be investigated for the ability to resolve two model components with varying orientation, for different values of N and CH vector weighting. The weighting shall be based on the sinusoidal model scheme with $W_e = W_o$, to ensure the CH vector norm is still phase-invariant. Finding the error with respect to N gives us an idea of how many RT orders to choose when performing analysis.

5.6.1 Iterative and Roots Method

Accuracy of the half-sinusoidal model estimation for the iterative and roots methods was tested using a CH vector consisting of $K = 2$ half-sinusoidal components. The orientation difference between each was varied to see the effect on errors in the parameter estimation, and thus no noise was added.

Two sets of CH vectors were constructed:

- Addition of two line-segment vectors with amplitude ratio randomly varying between 0.5 and 1.5 and orientation difference varied from $\pi/36$ to π in 36 increments. This CH vector represents the response to a line-segment corner with a variable angle. It is equivalent to two half-sinusoidal components with $A_1/A_2 \in [0.5, 1.5]$ and $\phi_1 = \phi_2 = 0$.
- Addition of two edge-segment vectors with opposite amplitudes, amplitude ratio randomly varying between -0.5 and -1.5, and difference in orientation varied from $\pi/36$ to π in 36 increments. This CH vector represents the response to a wedge-type corner with variable angle. It is equivalent to two half-sinusoidal components with $A_1/A_2 \in [0.5, 1.5]$ and $\phi = [\pi/2, -\pi/2]$.

The vectors are represented by two half-sinusoidal components, $\mathbf{Wf}_{\mathrm{H}}$ as follows:

$$\mathbf{Wf} = \mathbf{Wf}_{\mathrm{H}_1}(A_1, \phi_1, \theta_1) + \mathbf{Wf}_{\mathrm{H}_2}(A_2, \phi_2, \theta_2). \tag{5.70}$$

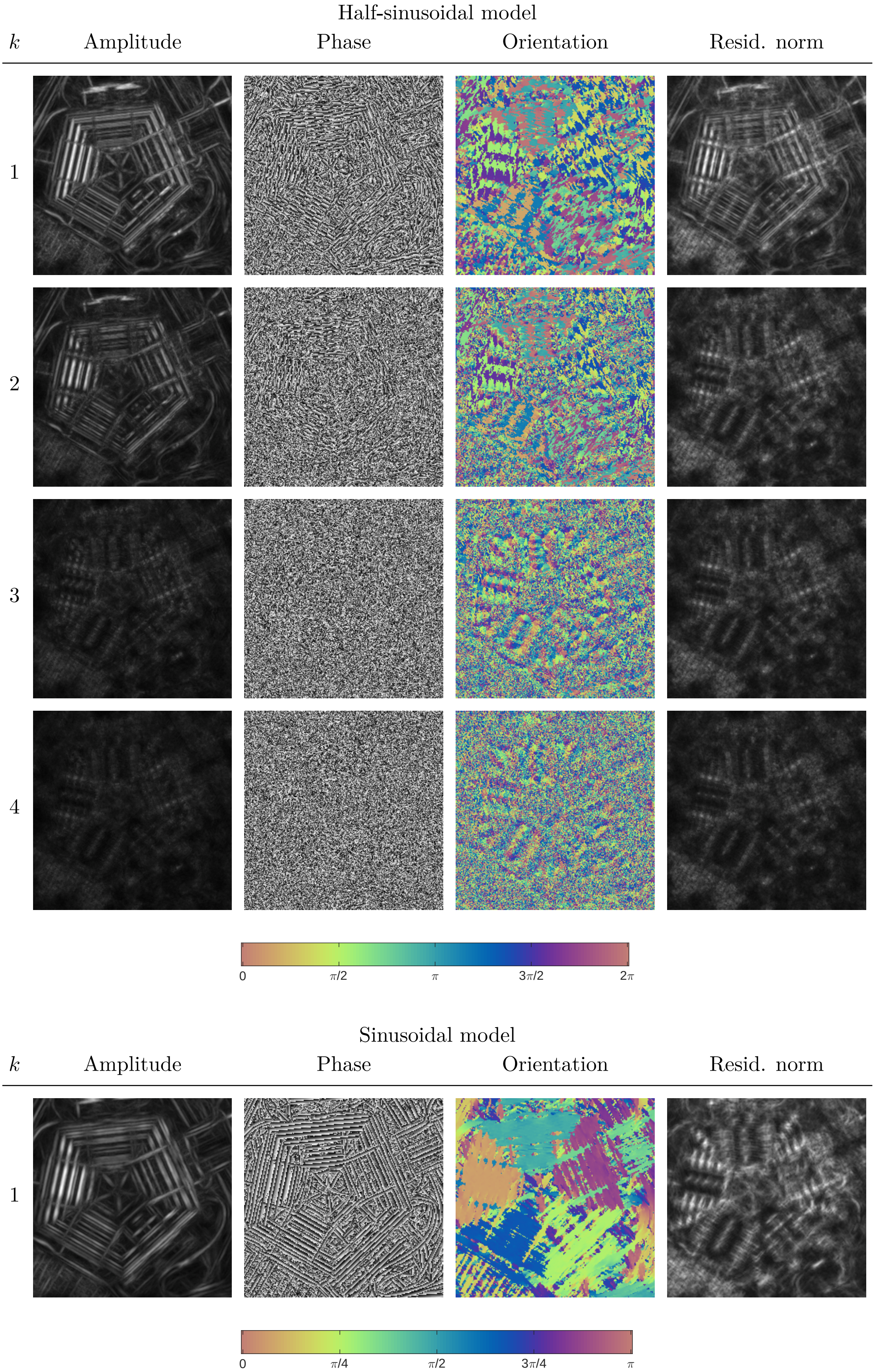


Figure 5.3: First four components of the half-sinusoidal model of the first scale of the *Pentagon* image using $N = 13$, compared to the sinusoidal model (bottom row).

The model parameters were solved for using either the iterative or roots method to give two estimated half-sinusoidal model CH vectors $\mathbf{f}'_{H_1}$ and $\mathbf{f}'_{H_2}$. Since the order of the estimated model components does not necessarily match that of the original model, the estimated components were paired with the original model components so that there was *minimum* distance between their CH vectors. The amplitude, orientation and phase error were calculated for each pair, and the errors calculated according to the mean difference (amplitude) and mean angular difference (phase and orientation) (Figure 5.6), as follows

$$A_{\text{error}} = \sum_k |A_k - A'_k| / K, \tag{5.71}$$

$$\phi_{\text{error}} = \sum_k |\arg(\cos(\phi_k - \phi'_k) + \mathrm{i}\sin(\phi_k - \phi'_k))| / K, \tag{5.72}$$

$$\theta_{\text{error}} = \sum_k |\arg(\cos(\theta_k - \theta'_k) + \mathrm{i}\sin(\theta_k - \theta'_k))| / K. \tag{5.73}$$

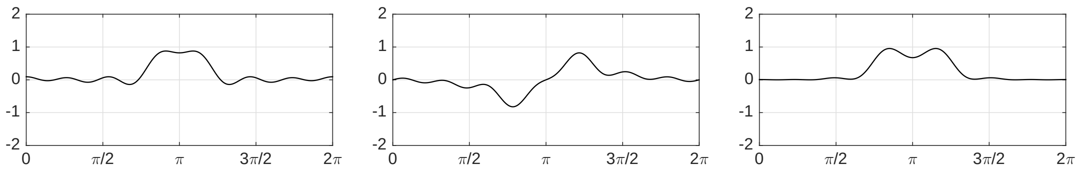
The first test set was used to evaluate the line-segment only model and the second test set was used to evaluate the edge-segment only model as well (Figures 5.7 and 5.8). Since there is no phase parameter due to only one wavelet in each model, only the amplitude and orientation errors were calculated.

Effect of N

The effect of N was investigated using the phase-invariant equal weighting scheme ($B \approxeq 0$) for the half-sinusoidal model (Figure 5.6). For both methods, the error is greatest when the orientation difference is close to 0. Both methods are able to accurately estimate the parameters of line-segment features at a smaller orientation separation than for edge-segment features. This is because a small-angle wedge segment correlates with the line-segment wavelet as well, leading to a single peak in the response (Figures 5.4 and 5.5).

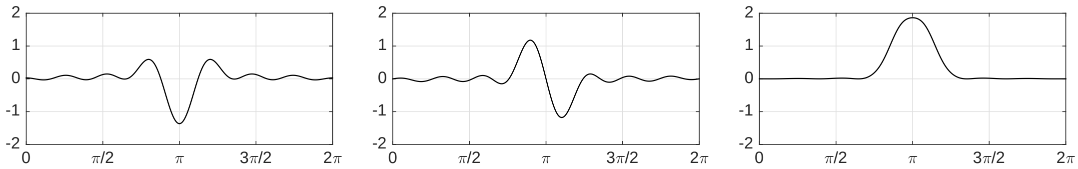
Increasing N increases the orientation selectivity of the wavelets, allowing for smaller orientation differences to be resolved with less error. Furthermore, increasing N reduces the error *floor* at larger orientation differences. The line-segment corner was able to be resolved at smaller orientation differences than the edge segment (wedge) corner. As a rough guide, one can resolve orientation differences larger than $2\pi/(N+4)$ for line segment features, and $4\pi/(N+4)$ for edge segment features using the half-sinusoidal model.

The effect of N was also investigated for solving the line-segment feature using on the line-segment model (Figure 5.7) and the edge-segment feature using the edge-segment model (Figure 5.8). For the line-segment feature, using only line-segments in the model instead of the full half-sinusoidal model improved the amplitude and orientation estimates slightly. For the edge-segment feature, using only edge-segments in the model instead of the full half-sinusoidal model improved the amplitude and orientation estimates considerably. These results suggest that one should use the edge-segment only model to more accurately parametrise edge-segment features such as corners



(a) Line-segment $\lambda_L(\theta)$ (b) Edge-segment $\lambda_E(\theta)$ (c) Combined $p(\theta)$

Figure 5.4: Line-segment wavelet (a), edge-segment wavelet (b) and combined (c) angular response polynomials for a two line-segment feature with components spaced 0.2π radians apart.



(a) Line-segment $\lambda_L(\theta)$ (b) Edge-segment $\lambda_E(\theta)$ (c) Combined $p(\theta)$

Figure 5.5: Line-segment wavelet (a), edge-segment wavelet (b) and combined (c) angular response polynomials for a two edge-segment feature (wedge) with components spaced 0.2π radians apart.

and junctions when using smaller N.

Effect of Weighting

To test the effect of weighting for the half-sinusoidal model, the same features were modelled as for the previous half-sinusoidal experiment except with a fixed value of $N = 13$ and the weighting varied instead (Figure 5.9). The weights were calculated using the energy maximisation method for the sinusoidal wavelet with the window width B chosen as a multiple of the constant, $B_{0.1} = 5.64/N - 6.57/N^2$. The reason this weighting was used rather than one designed specifically for the half-sinusoidal wavelet is for a consistent approach to weighting when using either model. A factor of 0 is equivalent to the equal weighting used in the previous experiment.

As expected, increasing B increases the minimum orientation difference before the errors begin to plateau. That is, the ability to accurately resolve model components close in orientation is reduced, due to the wider angular response profile. In contrast to the multi-sinusoidal model, past this point increasing B does *not* decrease the error floor until close to π radians. The same experiment was conducted with different values of added Gaussian noise. As the noise was increased the error floor also increased. After the noise CH vector magnitude was greater than approximately 0.15 of the image CH vector magnitude the noise floor for all weightings was approximately the same. Therefore in higher noise images using $B = 0$ remains a good choice for the weighting.

5.6.2 Super-Resolution Method

The effect of N on solving the line-segment and edge-segment models was also tested with the super-resolution method. The effect of weighting was not tested as weights are not used with this method, and there is no phase error comparison because the restriction to only line segment or edge segment types means no phase value is obtained. Numerical experiments involving spike trains

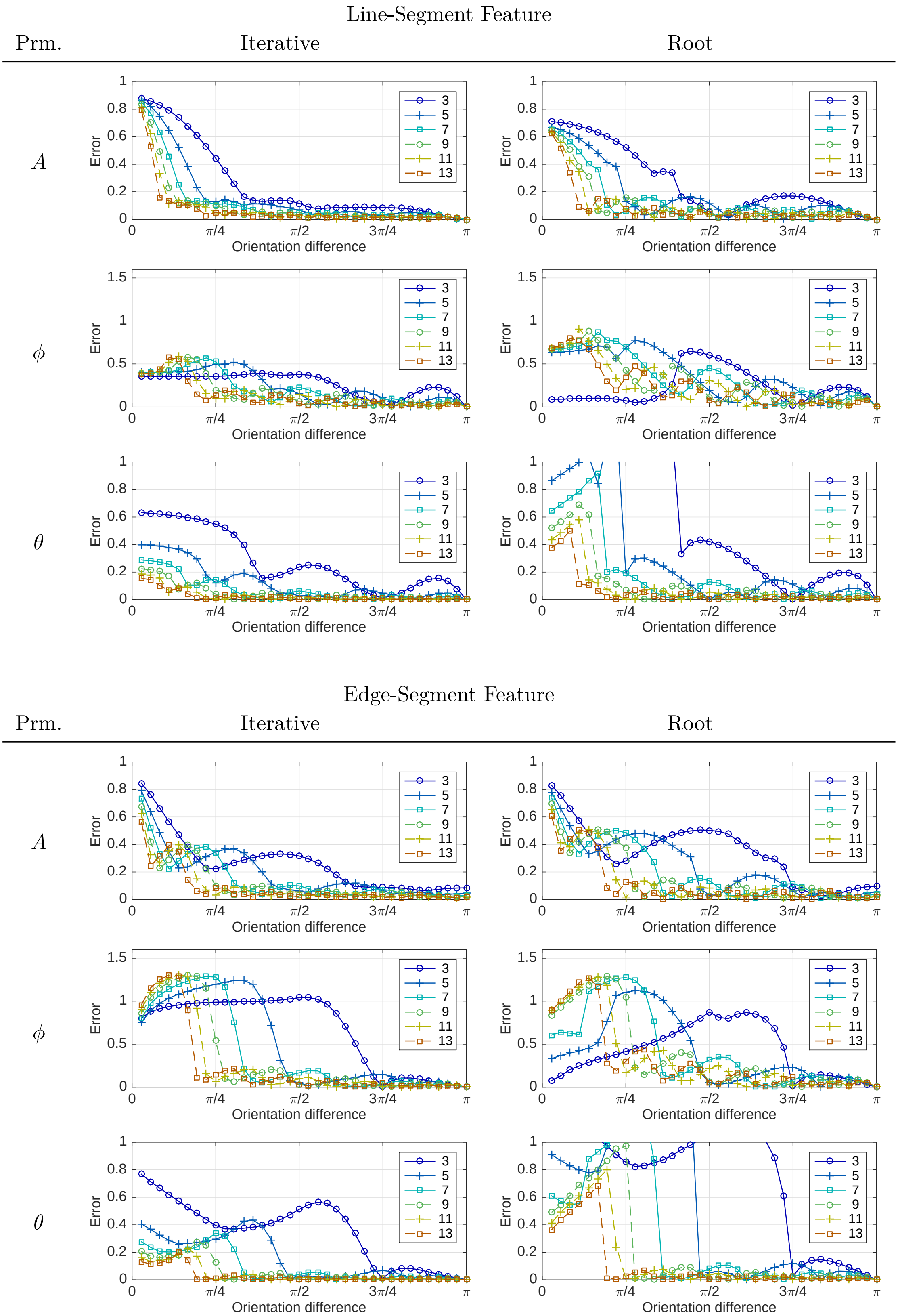


Figure 5.6: Average error of the estimated half-sinusoidal model parameters for an ideal two line-segment or two edge-segment feature with unit amplitude versus orientation difference for different values of N (shown in legend), using the iterative and root solvers.

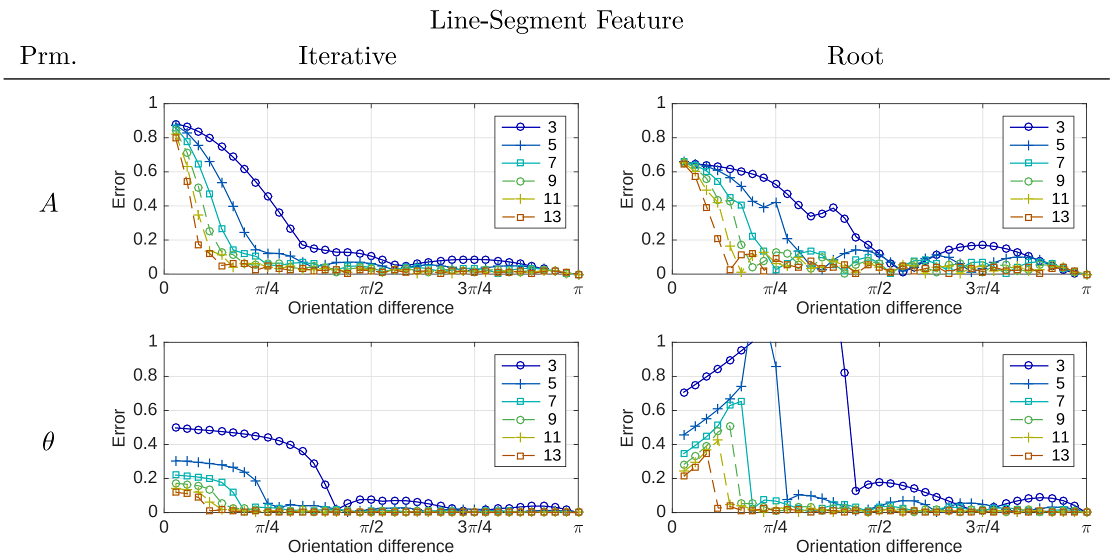


Figure 5.7: Average error of the estimated line-segment model parameters for an ideal two line-segment feature versus orientation difference for different values of N (shown in legend), using the iterative and root solvers.

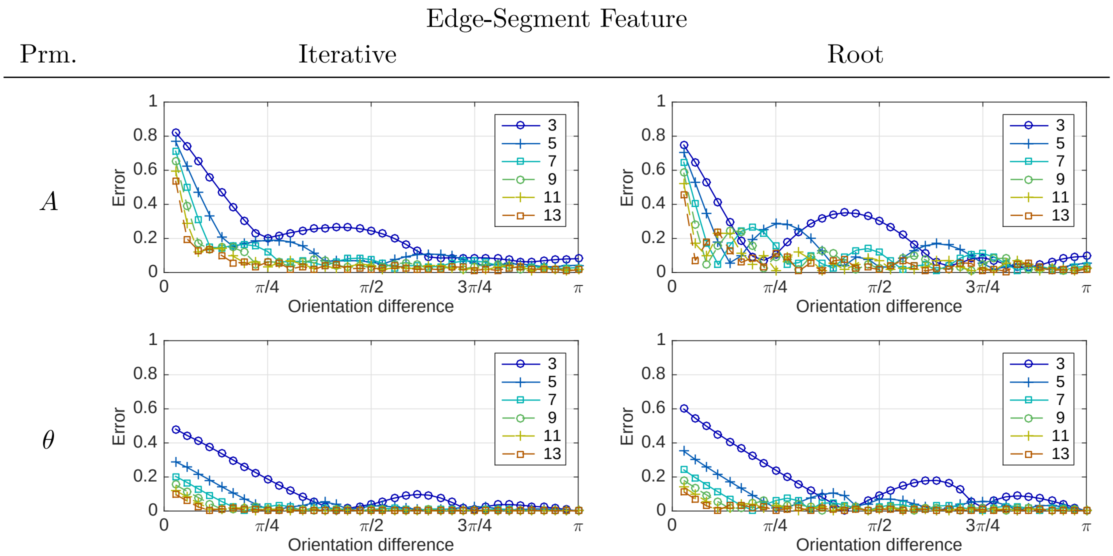


Figure 5.8: Average error of the estimated edge-segment model parameters for an ideal two edge-segment feature versus orientation difference for different values of N (shown in legend), using the iterative and root solvers.

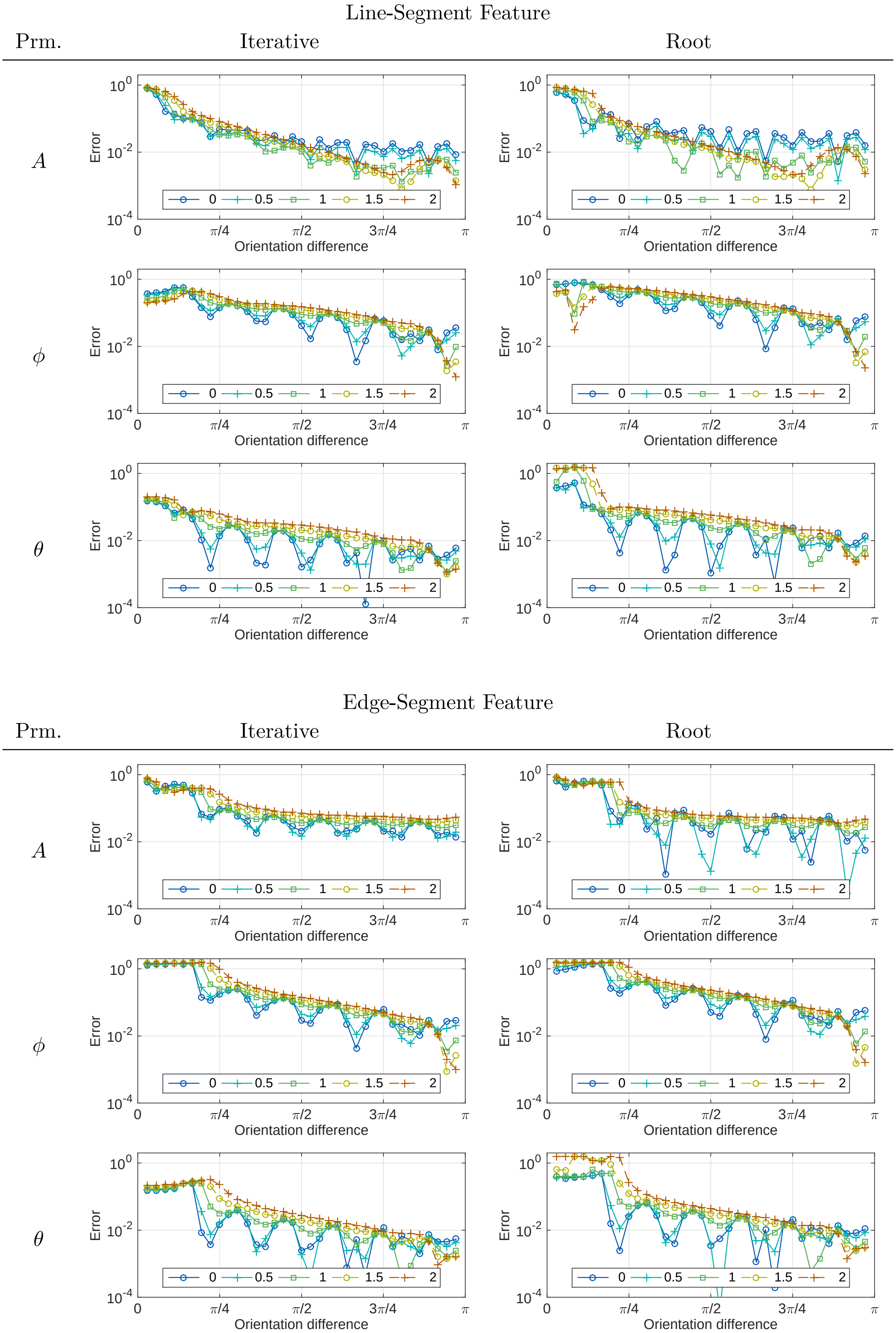


Figure 5.9: Average error of the estimated half-sinusoidal model parameters for an ideal two line-segment or two edge-segment feature with unit amplitude versus orientation difference for different weighting parameters B as a factor of $B_{0.1}$ (shown in legend), for $N = 13$ using the iterative and root solvers.

in [15] suggest a orientation separation as small as $2\pi/N$ may be enough to guarantee finding an exact result. The results in Figure 5.10 show that for the line segment feature (equivalent to spikes of the same sign), the minimum orientation separation required is less than $\pi/(N+1)$, while for the edge segment feature (spikes of opposite signs) a separation of approximately $\pi/(N-1)$ is needed. Above these values the error is practically zero. In contrast, the iterative and roots methods require a much larger separation threshold until the error floor is reached. Therefore the super-resolution method can be used to resolve line or edge-segment model components that are close in orientation with a smaller value of N. However, as for the multi-sinusoidal model, the large computation time makes the super-resolution method impractical to apply to an entire image.

Prm.	Line Segment	Edge Segment

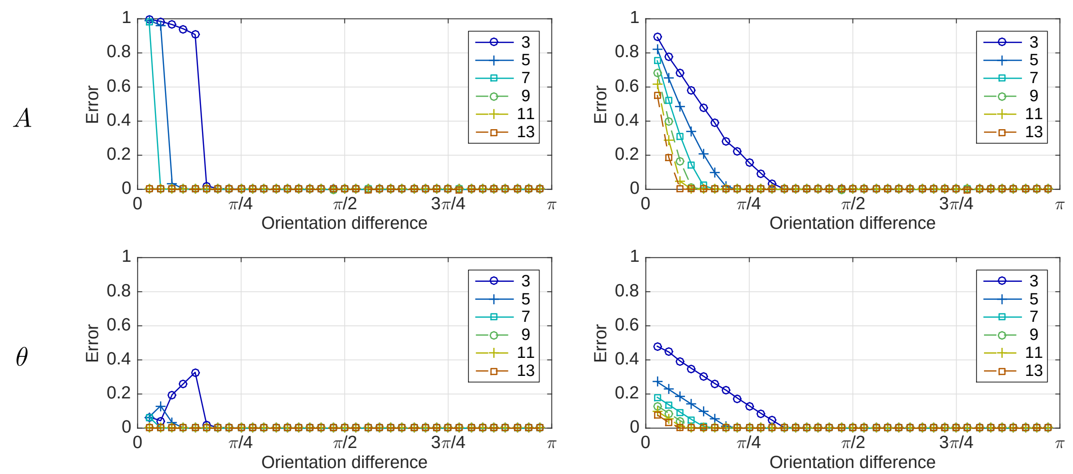


Figure 5.10: Average error of the model parameters for either an ideal two line-segment or two edge-segment feature with unit amplitude and varying orientation difference for different values of N (shown in legend), using the super-resolution method.

5.7 Junction Analysis

As the half-sinusoidal model encompasses both line-segments and edge-segments, we shall focus on application of that particular model to the task of analysing multiple-line-segment and multiple-edge-segment features.

5.7.1 Feature Response

Additive Segment Features

The response to both additive ideal line-segment features and edge-segment features has been derived earlier in the chapter. However, one of the assumptions of the line-segment model was of an ideal line with zero width. The difference between the ideal response and the response to a non-zero width line was tested for a line segment with a rounded end and a line segment with a square end. The *normalised* CH vector at the centre of each feature was found using a log-Gabor filter with $\sigma = 0.65$. The differences between the magnitude and angle of each vector component and the ideal additive version of the feature were calculated for different filter wavelengths, and the magnitude is expressed in relative terms (Figure 5.11). For example, a value of 2 means the occluded vector component has twice the magnitude of the ideal component, whereas a value of 1 is equal.

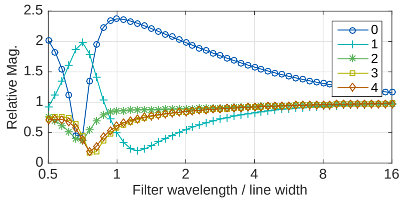
The amplitude errors are larger for both smaller wavelets (smaller wavelength) and lower orders. In particular, the 0th order CH wavelet has a large error for the rounded line segment, and does not reach the correct value until the filter wavelength is approximately 16 times the line width. All the other orders reach approximately the correct value when the filter wavelength is four times the line width or more. In contrast, a normal fixed width line feature (two line segments 180 degrees apart) is completely modelled by the sinusoidal model. These results show that the filter wavelength should be at least twice the line width for the actual line-segment vector to be similar in angular component to the ideal line-segment model vector.

Occluded Line Segment

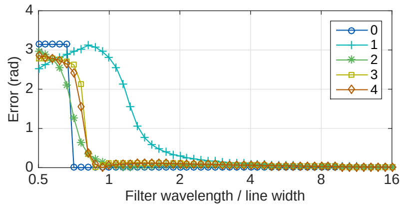
As for the crossed line features in the previous chapter, often line segment features in an image are not additive but instead occluded (Figure 5.12). Instead of modelling occluded features as the addition of line-segment components, we shall model them as the maxima (or minima) of the their intensity. For positive valued lines (white on black) we assume

$$f(\mathbf{x}) = \max_{k \in \mathbb{N}_K} \{ f_{\mathrm{L}(A_k, \theta_k)}(\mathbf{z}) \}. \tag{5.74}$$

(a) Round end

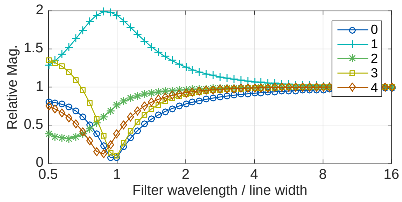


(b) Relative magnitude

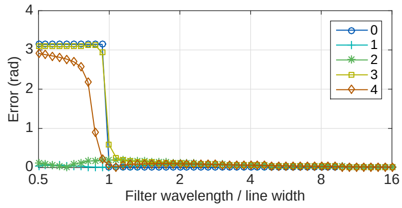


(c) Angle error

(d) Square end



(e) Relative magnitude



(f) Angle error

Figure 5.11: Differences in the magnitude (b,e) and angle (c,f) of each CH vector order (shown in legend) between an ideal line-segment and either a rounded (a) or square (d) non-zero width line-segment.

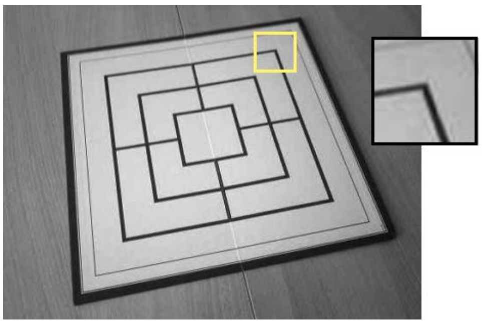


Figure 5.12: Example occluded line-segment feature.

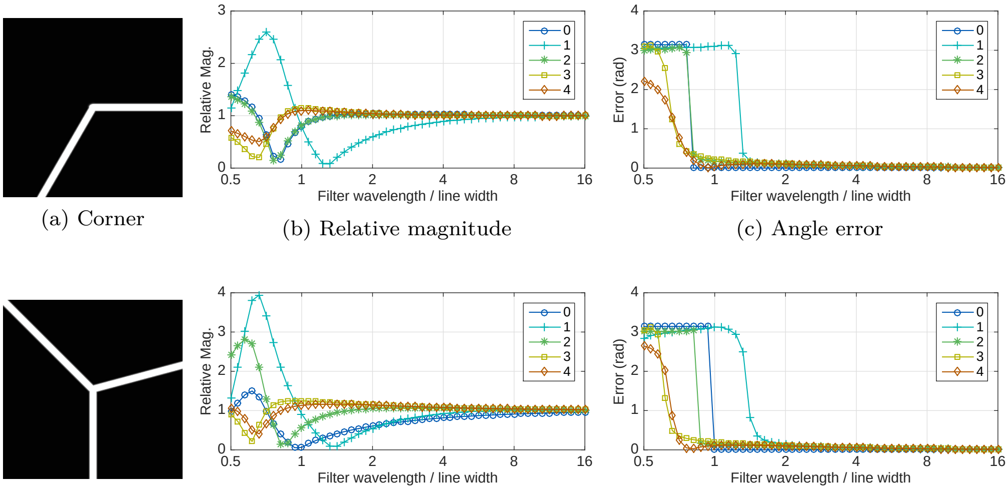
The maximum operator will always result in a value that is less than the equivalent additive model, that is ,

$$\max_{k \in K} \{ f_{\mathrm{L}(A_k,\theta_k)}(\mathbf{z}) \} \leq \sum_{k=1}^{K} f_{\mathrm{L}(A_k,\theta_k)}(\mathbf{z}), \tag{5.75}$$

and thus we may write the occluded model as the additive model minus a non-negatively valued occlusion function $f_{occ}(\mathbf{z})$

$$f(\mathbf{x}) = \sum_{k=1}^{K} f_{\mathrm{L}(A_k,\theta_k)}(\mathbf{z}) - f_{occ}(\mathbf{z}). \tag{5.76}$$

The different between the response of two occluded line-segment features consisting of two or three line segments and the ideal model for the same amplitude and orientation parameters was investigated (Figure 5.13). As before, the amplitude errors are larger for both smaller wavelets (smaller wavelength) and lower orders. In particular, both the 0th and 1st order CH wavelets differ greatly in magnitude and angle from the response of the ideal feature when the filter wavelength is below twice the line width. The *angle* of each RT order response was correct when the wavelength of the filter was greater than twice the line width. This gives a rough minimum bound on the filter size to use for analysing occluded line junctions to ensure reasonably accurate parameter estimation.



(a) Corner (b) Relative magnitude (c) Angle error

(d) Y junction (e) Relative magnitude (f) Angle error

Figure 5.13: Differences in the magnitude (b,e) and angle (c,f) of each CH vector order (shown in legend) between an ideal multiple line-segment feature and a non-zero width line-segment feature (a,d).

The occluded error component, $f_{occ}(\mathbf{z})$, changes size depending on the angles between the line segments. Potential strategies to deal with occluded line segments are the same as those for the multi-sinusoidal model:

- **Increase the number of orders** N : Including higher orders increases the size of the wavelet and therefore reduces the influence of the larger occluded errors in the lower orders.
- **Increase the size (wavelength) of the primary filter:** The examples in Figure 5.13 suggest a minimum wavelength of at least twice the line width is necessary.
- **Remove lower order components:** The lower order CH wavelets are the smallest in spatial extent and most affected by the occluded error. Therefore we reduce the overall error by setting their weights to 0.
- **Combine estimates over multiple scales:** Good features tend to have the same shape over multiple scales. This will be investigated in the next chapter.

Removing Lower Orders

An example of wavelets constructed by removing the lower orders is shown in Figure 5.14 for $N = 13$, numbered according to the minimum non-zero order. For example, 2 indicates the 0th and 1st order components were weighted to 0. Removing lower orders attenuates the centre of the wavelet but also reduces its linear appearance and increases the amplitude of oscillations outside the main lobe.

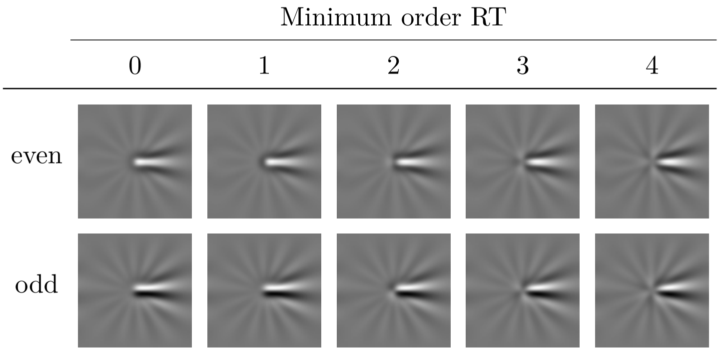
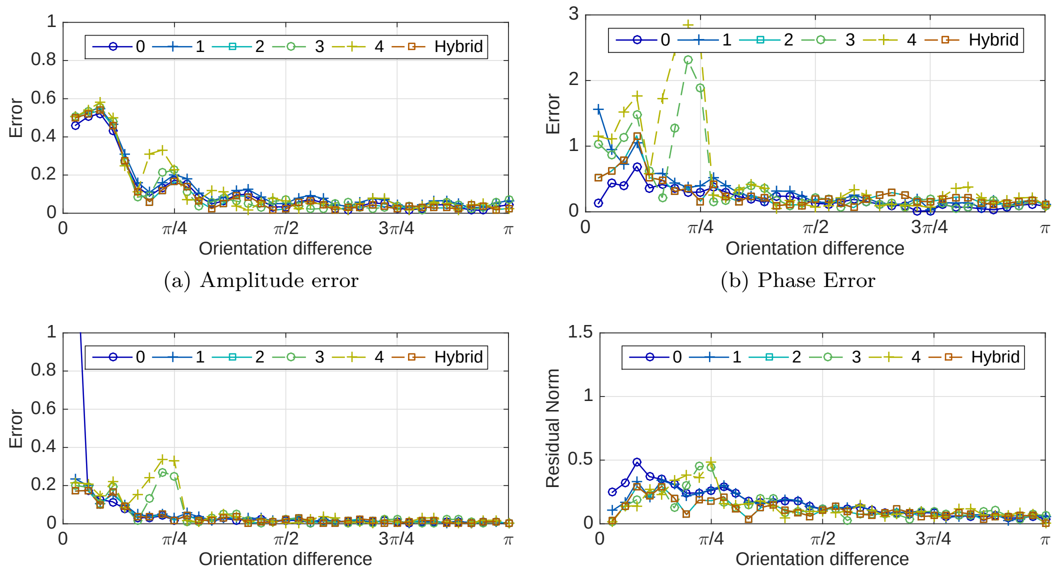


Figure 5.14: Line-segment and edge-segment wavelets with the lower RT orders set to zero for $N = 13$ and phase-invariant equal weighting. The number indicates the minimum non-zero order.

The effect of removing lower orders was investigated by finding the model parameters for an occluded two line-segment feature (such as Figure 5.13a) with orientation difference ranging from $\pi/36$ to π radians. The model was calculated for $N = 13$, phase-invariant equal weighting, the iterative method and using a log-Gabor filter with $\sigma = 0.65$ and wavelength twice the width of the line (Figure 5.15). The response to a *single* line-segment feature with the same line width was used to normalise the results. Note that in contrast to the sinusoidal model for a line feature, the residual component for a single non-zero width *line-segment* feature is *not* zero.



(a) Amplitude error
(b) Phase Error
(c) Orientation Error
(d) Residual magnitude

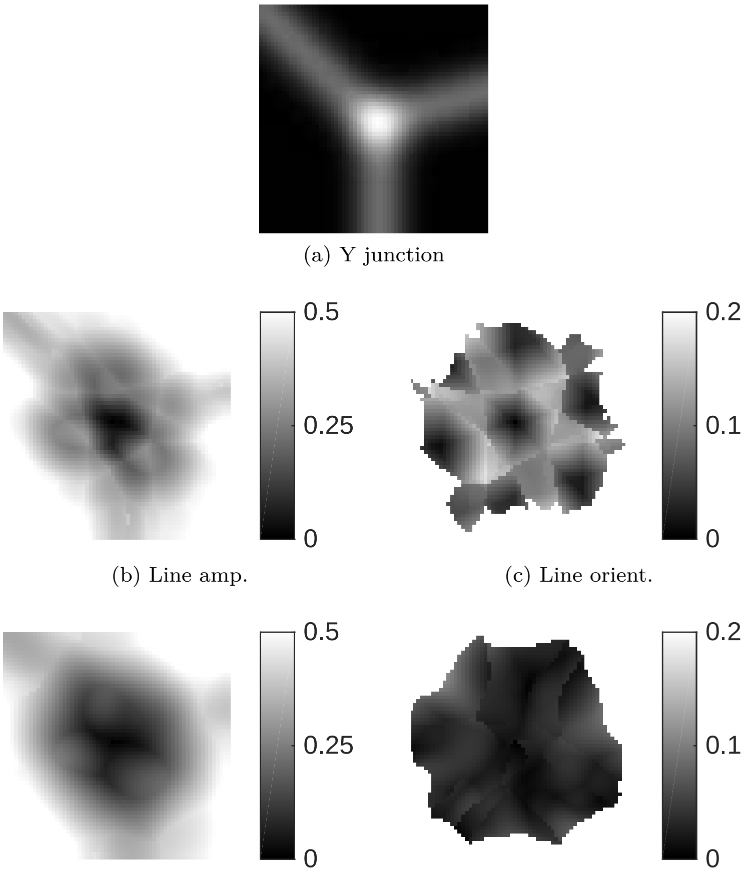
Figure 5.15: Amplitude, phase and orientation errors along with the residual norm for an occluded two line-segment feature solved using the iterative method and wavelets with the lower orders removed. The minimum non-zero order is shown in the legend. The CH vector was calculated using a log-Gabor primary filter with wavelength twice the line width and $\sigma = 0.65$, and $N = 13$.

The amplitude, phase and orientation errors are lowest for the both the normal half-sinusoidal wavelets with no orders removed, and the wavelets with the 0th and 1st order removed (using 2nd order and above) (Figure 5.15). Ideally, the residual should be zero, meaning that the model is completely representing the feature. Zeroing the 0th and 1st orders showed some benefit in the slight reduction of the residual norm (shown as a proportion of the image CH vector norm). A hybrid system was also trialled, where the normal half-sinusoidal wavelets were used for the first iteration and the wavelets with the 0th and 1st orders removed for the second. This scheme gave a small improvement, reducing the residual magnitude where the orientation difference was smaller and occluded component larger. Note, a hybrid approach is only possible using the iterative solving method.

5.7.2 Off-Centre Response

One may ask if there is any benefit in using the half-sinusoidal model if one is only interested in the analysis of purely line-segment or purely edge-segment features, especially since the experiments on model error versus orientation separation showed that using a line or edge-segment model matched to the corresponding line or edge-segment feature gave lower errors (Figures 5.7 and 5.8) compared to using the half-sinusoidal model (Figure 5.6). The advantage of having both line and edge-segments in the model is for *off-centre parametrisation*. The segment model proposed in this chapter and other approaches such as MSMFs assume the point of analysis is at the centre of the junction. However, detection examples in Section 3.3.4 showed that the proposed i2D detector along with the Harris corner detector and boundary tensor do not necessarily detect the centre of a junction.

Performing analysis at an off-centre point leads to errors in the model parameters. Figure 5.16 shows an example for a Y junction feature. The average difference between the amplitude and orientation of the Y junction segments and the calculated model parameters was found for each point in the image, using either the line segment model (Figures 5.16b and 5.16c) or the half-sinusoidal model (Figures 5.16d and 5.16e). The errors in the local area are much reduced for the half-sinusoidal model, especially for the orientation estimate. The inclusion of the edge-segment thus improves parameter estimation if the feature is modelled off centre, making analysis less reliant on perfectly detecting the centre location.



(a) Y junction

(b) Line amp.

(c) Line orient.

(d) Half amp.

(e) Half orient.

Figure 5.16: Error in parameter estimation of a Y junction (a) according to analysis location when using the line-segment model (b,c) compared to the half-sinusoidal model (d,e).

5.7.3 Classification Procedure

To demonstrate the usefulness of the half-sinusoidal model, junction classification was performed on the *Lab* and *Flintstones* test images (Figures 5.18a and 5.18b). The *Lab* image has many corner features that should be well modelled by the edge-segment part of the model, while the cartoon appearance of the *Flintstones* image has different regions of fairly constant colour bordered by dark lines, and thus the junctions between regions have both line and edge characteristics.

The procedure followed was the same for each image:

1. **Primary filter:** A log-Gabor filter with wavelength 8 pixels and a wide bandwidth given by $\sigma = 0.5$ was selected. The log-Gabor filter has a smooth frequency profile and thus a smooth spatial decay, and we are not interested in reconstruction.

2. **CH vector:** A value of $N = 13$ was used to construct the CH vector. According to the results in Figure 5.6 this should allow features with components differing in orientation by more than $\pi/8$ radians (22.5 degrees) to be resolved. The phase-invariant equal weighting scheme was used.

3. **Intrinsic dimension:** The sinusoidal model was obtained from the CH vector, and the

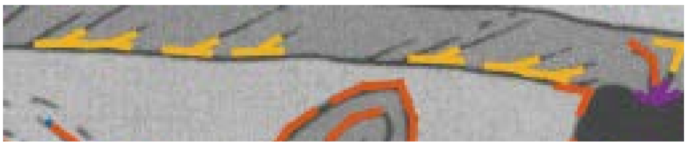
model norm and residual norm were used to construct the intrinsic dimension representation from Chapter 3. The intrinsic dimension angle was adjusted with the sinusoidal function from Section 3.3.3 with parameters $h = 0.5$ and $s = 2$ (Figures 5.18c and 5.18d).

4. **Detection:** Possible corners and junctions were detected as local maxima of the i2D detection score in an 8 pixel radius. Points with a score less than 0.2 times the maximum score were removed (Figures 5.18e and 5.18f).

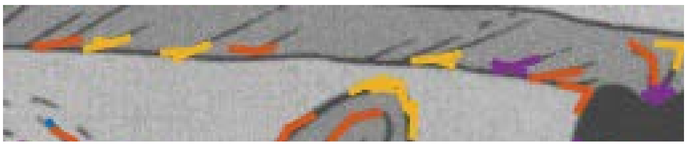
5. **Parametrisation:** The half sinusoidal model was calculated using up to $K = 4$ components. Each detection point was then classified according to the invariant classification thresholding method introduced in Section 4.4.2 using a threshold of 0.5. Figures 5.19a and 5.19b show the classifications. The length of each line is proportional to the corresponding amplitude of the component. The colours of the lines indicate the classification.

6. **Post processing:** Some of the junctions have multiple close together line components, for example the three circular 'pacman' shapes in the foreground of the *Lab* image in Figure 5.19a. Since when using $N = 13$ the errors increase below an orientation separation of $\pi/8$ radians, any components with differences lower than this threshold were combined using the spike combining method from Section 2.5.4. The post-processed classification results in Figures 5.19c and 5.19d show a qualitative improvement in parameter estimation.

For a final comparison the same process was followed for each image, except only using the edge-segment wavelet for the *Lab* image (Figure 5.19e) and only the line-segment wavelet for the *Flintstones* image (Figure 5.19f). Some corners appear to have better classification in the *Lab* image, but the angle of the components deviates slightly from that found with the half-sinusoidal model. For the *Flintstones* image, junctions that have both line and edge characteristics are handled correctly by the half-sinusoidal model, but poorly parametrised if only the line-segment wavelet is used. This is particularly noticeable in the upper section of the *Flintstones* image, repeated below in Figure 5.17. We can observe that the parametrisation using the line-segment only model does not match the line segment components in the image.

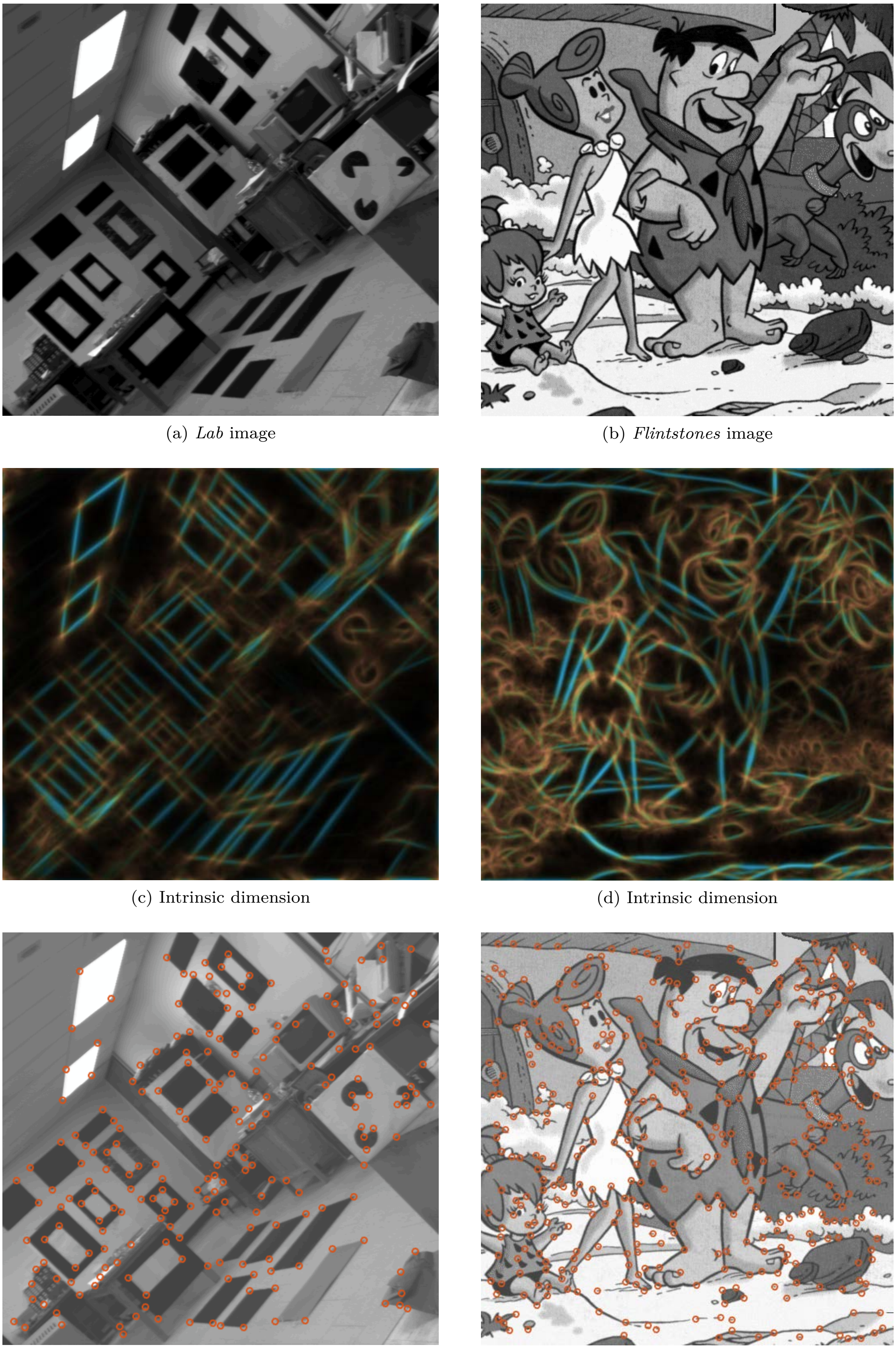


(a) Parametrisation using half-sinusoidal model



(b) Parametrisation using line-segment model

Figure 5.17: Difference in parametrisation of features that have both line and edge characteristics using the half-sinusoidal model (a) and the line-segment model (b). Orange: two segments, yellow: three segments, purple: four segments.



(a) *Lab* image

(b) *Flintstones* image

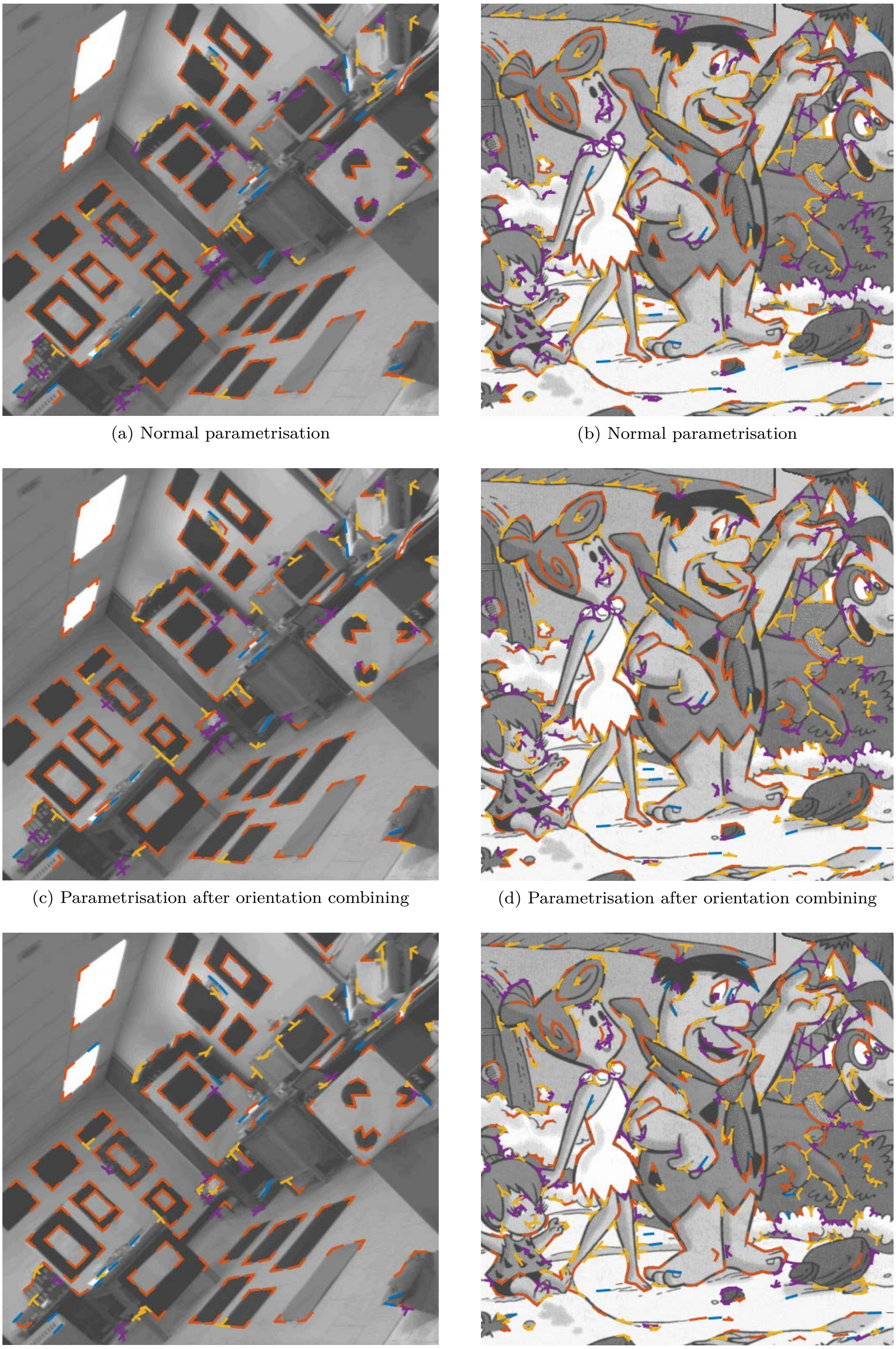
(c) Intrinsic dimension

(d) Intrinsic dimension

(e) i2D detections

(f) i2D detections

Figure 5.18: i2D junction and corner detection for the *Lab* and *Flintstones* images using a log-Gabor filter (wavelength $= 8$, $\sigma = 0.5$) and $N = 13$.



(a) Normal parametrisation

(b) Normal parametrisation

(c) Parametrisation after orientation combining

(d) Parametrisation after orientation combining

(e) Parametrisation using only the edge-segment

(f) Parametrisation using only the line-segment

Figure 5.19: Classification and parametrisation of the detection junction and corner locations using the half-sinusoidal model (a,b), the half-sinusoidal model with orientation combining (c,d) or either the edge-segment (e) or line-segment (f) models. Orange: two segments, yellow: three segments, purple: four segments.

5.7.4 Experiment: Orientation Estimation

Accurate estimation of the *orientation* of the individual line or edge segments that are present in a corner or junction feature is used in applications such as camera calibration [90]. The half-sinusoidal model is suited to the parametrisation of corners and junctions. To evaluate its performance, a similar experiment to Section 4.5 of the previous chapter was performed for a corner junction consisting of two line segments instead of a crossed line feature.

A synthetic image consisting of two line-segments was created with the orientation separation ranging from 5° to 180° in 5° increments, and with different amounts of additive Gaussian noise. The junction parameters were solved for using the half-sinusoidal model and iterative and roots solvers, or the line-segment model using the super-resolution method, and compared to the results of the MSMF method. The MOP method was not applicable as it only estimates orientation over the half circle, and therefore cannot distinguish between acute angle corners and obtuse angle ones.

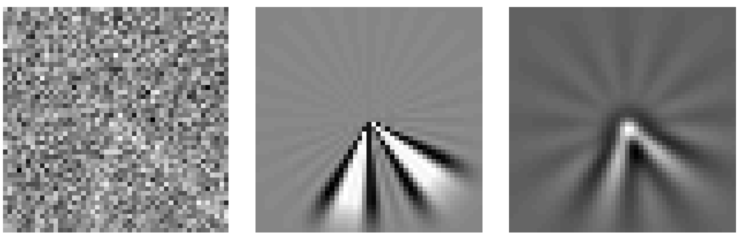
The specifics of each method used are as follows:

- Half-sinusoidal model calculated for $N \in [7, 13]$, $K = 2$, using a log-Gabor filter with $\omega \in [22, 12]$ respectively and $\sigma = 0.65$. The wavelength of the filter was chosen so that the spatial extent of the filter kernel would be approximately the same as that for the MSMF method. The model was evaluated using the iterative and roots methods.

- Line-segment model using the same parameters above but evaluated using the super-resolution method.

- MSMF created using 28 orders, wedge angle 20° and radius 24 pixels. Levenburg-Marquardt solver was used with initial parameters set to the maxima of the angular response sampled at 10° spacing. A two line-segment model was assumed and the method constrained to two wedges.

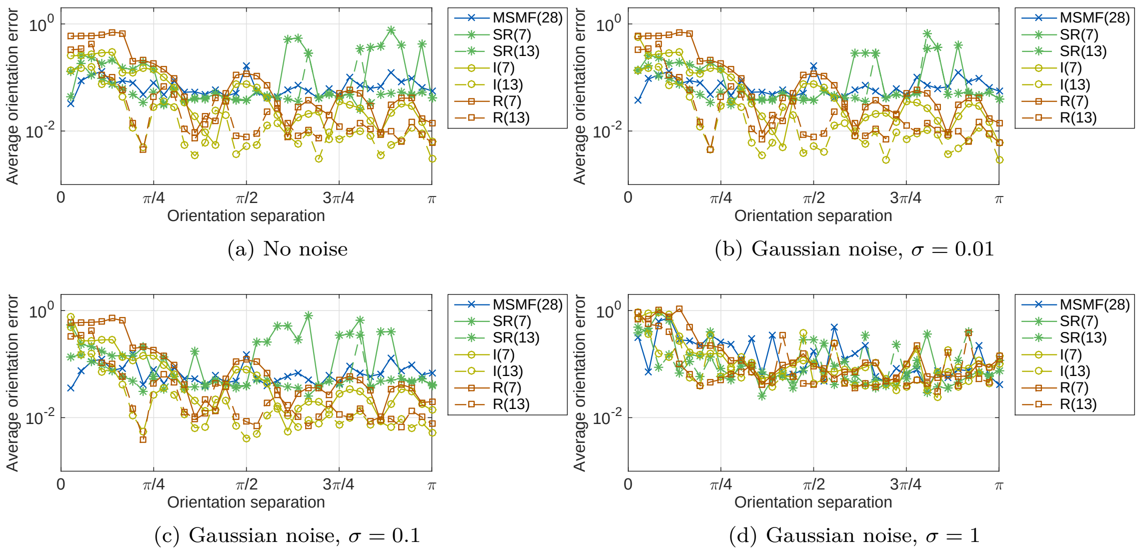
An example of a noisy additive crossed line feature and the MSMF and sinusoidal model wavelets corresponding to the ideal response are shown in Figure 5.20.



(a) Line feature (b) MSMF ($N = 28$) (c) CH ($N = 13$)

Figure 5.20: Noisy line feature and the MSMF and half-sinusoidal model wavelets corresponding to the ideal response at the centre.

The estimated orientation error was calculated for different amount of Gaussian noise with $\sigma \in [0, 0.01, 0.1, 1]$ and is shown in Figure 5.21. Unlike for the crossed line experiment in the previous chapter, the super-resolution method and MSMF method perform equally well and it is the iterative method followed by the roots methods that give lower errors overall. This is explained by the sensitivity of the MSMF method and the super-resolution method to centre position, where even a single pixel shift in the centre location will change the estimates by a couple of degrees. In contrast, the iterative and roots methods are used with the half-sinusoidal model, which is less sensitive to off-centre analysis (Figure 5.16). The MSMF authors propose finding the centre to sub-pixel accuracy and interpolating the orientation estimates in a 3×3 pixel region, presumably to improve the accuracy [86]. Using the half-sinusoidal model this is likely not required.



(a) No noise

(b) Gaussian noise, $\sigma = 0.01$

(c) Gaussian noise, $\sigma = 0.1$

(d) Gaussian noise, $\sigma = 1$

Figure 5.21: Orientation estimation errors for a line-segment corner, using the half-sinusoidal model solved with iterative (I) and roots (R) methods and line-segment model solved using super-resolution (SR) method for $N \in \{7, 13\}$, compared to MSMF with $N = 28$.

Michaelis and Sommer [80] mention that it would be beneficial to be able to use both line-segments and lines in the one model. This concept is the novel aspect of MSMFs compared to a normal wedge-filter approach, as with MSMFs one can add constraints to the wedge orientations to better estimate the orientations of the segments of a T junction, for example. We note that the same idea is easily achieved using the CH vector by solving using two sets of wavelets, that of the sinusoidal model and that of the half-sinusoidal model, using the iterative method. In fact, one could use combinations of any sets of wavelets, for example a line-segment and corner, if there was a practical application.

5.8 Image Representation

Both the multi-sinusoidal model and half-sinusoidal model are useful for representing various types of junctions and corners, but what about an image as a whole? If the CH wavelets are constructed from a suitable isotropic wavelet frame then we can reconstruct the image exactly from the model and residual components. Therefore to see how well either model represents the image as a whole, it is proposed to reconstruct an image from only the first K *model* components of the first four scales plus the low pass component, and compare the reconstructed result to the original image. The *peak signal-to-noise ratio* (PSNR $= 20\log_{10}(255/\text{MSE})$, where MSE is the mean square error) will be used as the comparison measure, as it is commonly used in the wavelet denoising literature (for example [102]) to compare the performance of different algorithms.

If we solve using the iterative method we are guaranteed that increasing the number of model components decreases the magnitude of the residual component, that is $\|\mathbf{W}\boldsymbol{\epsilon}_{k+1}\| < \|\mathbf{W}\boldsymbol{\epsilon}_k\|$. This in turn should increase the accuracy of the reconstruction from the model components. A visual example is given for the multi-sinusoidal model in Figure 5.22 and half-sinusoidal model in Figure 5.23. It shows reconstruction of the *Pentagon* image for different numbers of model components and N using four scales of a Simoncelli wavelet decomposition with subsampling.

We can make the following qualitative observations:

- Lower numbers of RT orders (N) gives a better reconstruction.
- Increasing the number of model components gives a better reconstruction.
- There is less smearing of the image along linear components, such as the pentagon walls, when using the half-sinusoidal model.

The first observation makes sense when we consider that a Parseval-tight wavelet frame can be constructed by a set of $N+1$ odd or even sinusoidal wavelets, or $2N+1$ line-segment or edge-segment wavelets, at equally spaced orientations around the half-circle or circle, respectively [124]. The second is explained by the residual decreasing with each extra model component. The third observation is due to the ability of the half-sinusoidal model to better represent a wider range of features. For example, a corner with two segments will be described by the multi-sinusoidal model has having two sinusoids along the same orientation, however there will still be a large residual as the segments are not i1D signals.

An experiment was performed to measure the average PSNR of the reconstruction for both models on grey-scale versions of four image sets:

1. Common test images such as *Lena*, *Barbara* and *Lab*. (15 images)
2. Aerial images of roads from the USC-SIPI dataset (`http://sipi.usc.edu/database/`). (15 images)
3. Texture images also from the USC-SIPI dataset. (21 images)
4. A selection of natural images from the BSD500 data set [77]. (10 images)

Example images are shown in Figure 5.24.

The PSNR results (Figure 5.25) confirm the previous observations that both decreasing orders and increasing the number of components improves the reconstruction. The reconstruction was

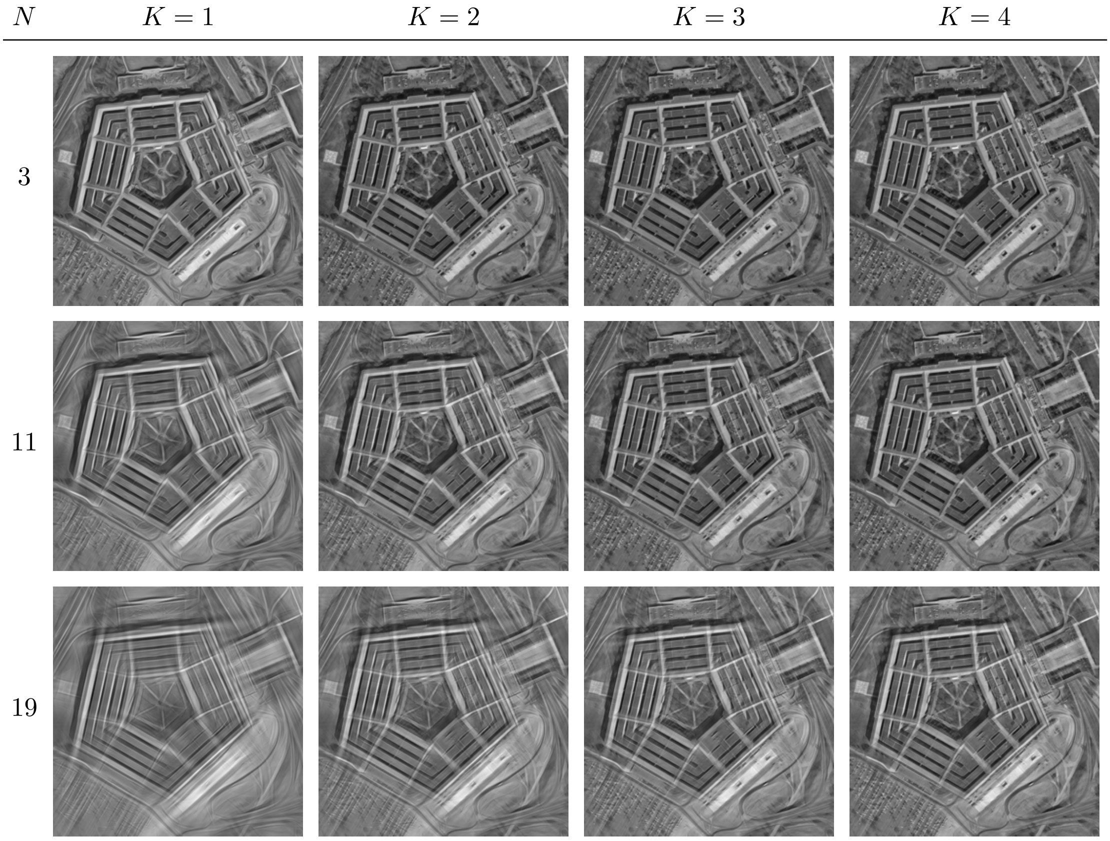


Figure 5.22: *Pentagon* image reconstruction from the multi-sinusoidal model with K components, using four scales of the Simoncelli wavelet different numbers of and RT orders, N.

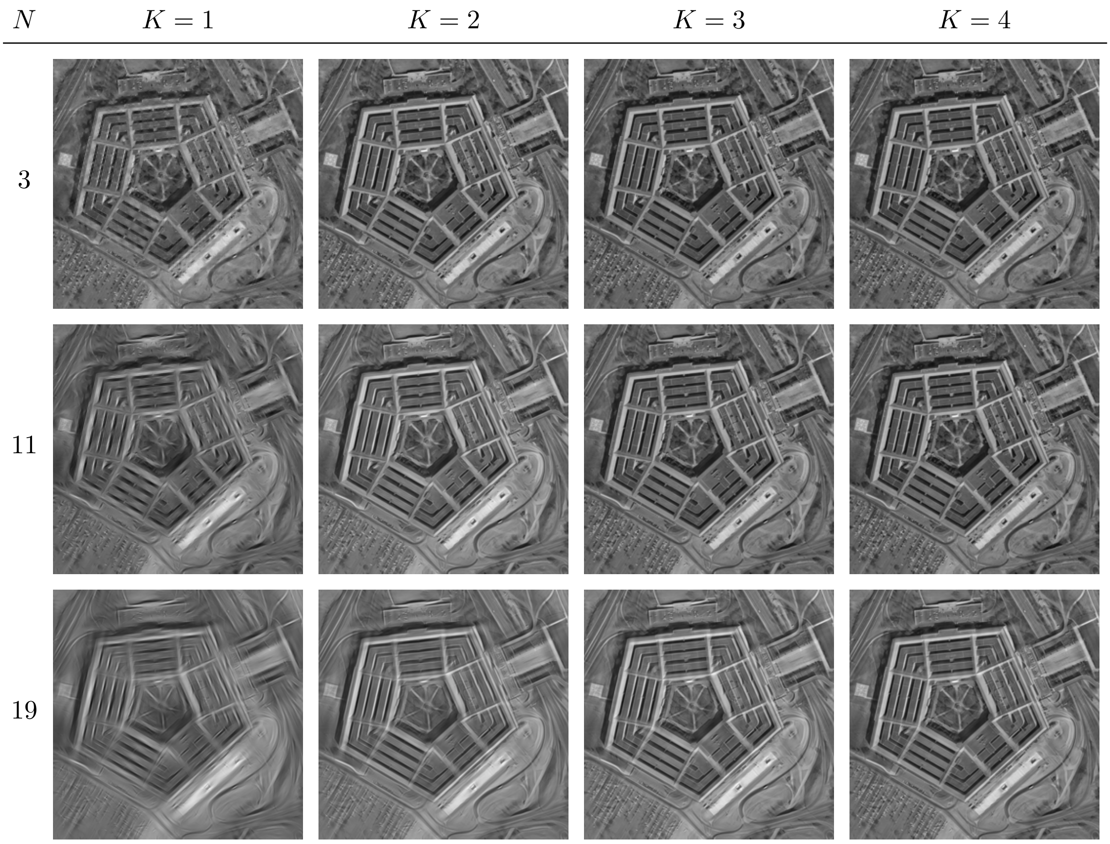


Figure 5.23: *Pentagon* image reconstruction from the multi-sinusoidal model with K components, using four scales of the Simoncelli wavelet different numbers of and RT orders, N.

best for the aerial image set, followed by the common and natural images, then the texture set. We can infer that the aerial photographs contain many i1D line and edge features which are well described by the two models, while the texture images contain more i2D components that respond to neither. Textures may be better described using the entire CH vector as a feature descriptor rather than projecting onto a particular model. The results also give a general guide as to what can be achieved if performing wavelet de-noising using the model components. For example, when using one sinusoid component and $N = 7$ the PSNR lies between 24dB and 27dB, thus any further improvement would require inclusion of the residual component.

For a single component, the sinusoidal model gives a better reconstruction. while for two or more components, the half-sinusoidal model is better. The difference is due to the model shape. To represent an i1D structure we only need *one* sinusoid, whereas for the half-sinusoidal model we need *two*. This can be seen back in Figure 5.3 where there is still a high residual component at the location of lines and edges when using just one half-sinusoidal component. However, when using *two* half-sinusoidal components both i1D features *and* corners are well modelled, and thus the reconstruction is better than for two sinusoids. Therefore, when using the half-sinusoidal model, at least $K = 2$ components should be derived.

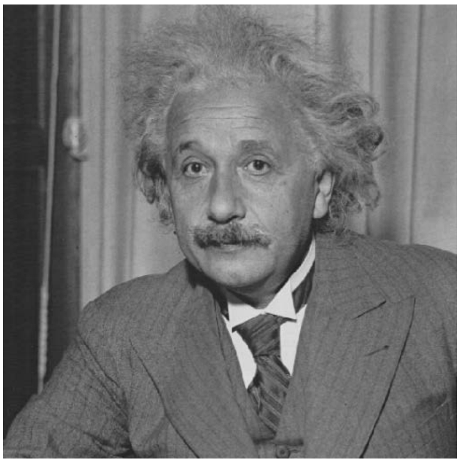
5.8.1 Orientation Estimation

One of the advantages of the sinusoidal model is that it gives a phase-invariant estimate of the main linear symmetry of the local image structure. For example, the sinusoidal model orientation estimated using a log-Gabor filter (wavelength 8, $\sigma = 0.5$) and $N = 7$ (Figure 5.26b) for the *Tree Rings* image (Figure 5.26a) is smooth with few discontinuities and appears to match the orientation of the curved rings. In the absence of a ground truth, having a similar smoothness to the features in the image is a good indication of accuracy. The first and second component of the half-sinusoidal model also gives an estimate of the local orientation of i1D features. Since their estimate is over $[0, 2\pi)$ instead of $[0, \pi)$ for the sinusoidal model, they are expressed modulo π (Figures 5.26c and 5.26d) for i1D orientation estimation. However, the estimate is less smooth with discontinuities not present in the original features. At every i1D-like location there are *two* possible orientations along the rings, clockwise or anti-clockwise, that the first component could match most strongly to. The discontinuities occur where the estimate flips between each direction and because the curvature of the local feature means the change in angle is less than π radians.

However, by averaging the two estimates we can obtain a combined estimate of the orientation over $[0, \pi)$,

$$\theta_{\mathrm{av}} = \arg\left(A_1 e^{\mathrm{i}\theta_1} + A_2 e^{\mathrm{i}(\theta_2+\pi)}\right) \bmod \pi, \tag{5.77}$$

where A and θ are the individual amplitude and orientation parameters of the two-component half-sinusoidal model. The average appears to be much smoother that the individual estimates



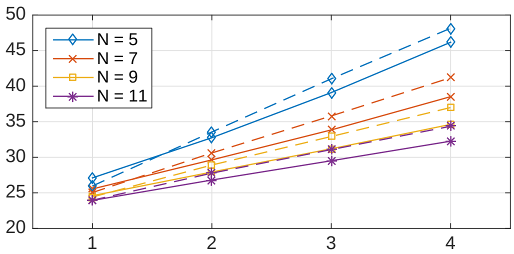
(a) Common

(b) Aerial

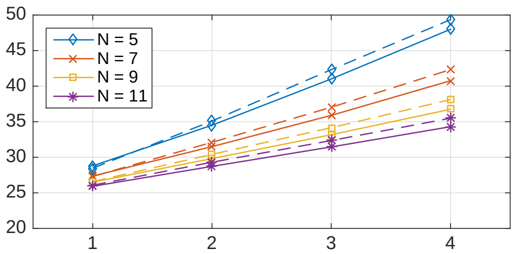
(c) Texture

(d) Natural

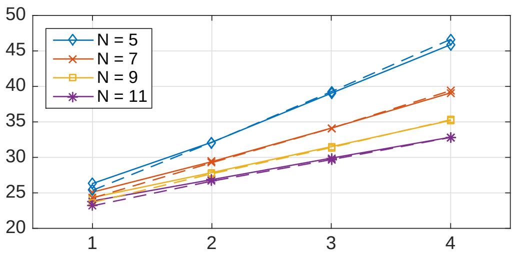
Figure 5.24: Example images from each test set.



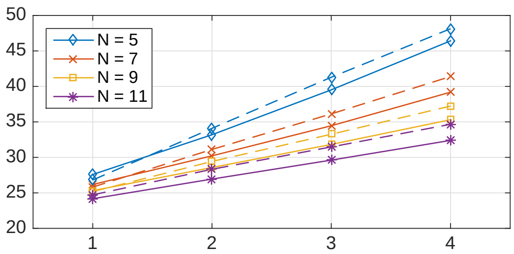
(a) Common images



(b) Aerial images



(c) Texture images



(d) Natural images

Figure 5.25: Average PNSR for the images in each test set reconstructed using K model components for the sinusoidal model (solid line) versus the half-sinusoidal model (dashed line) and different values of N.

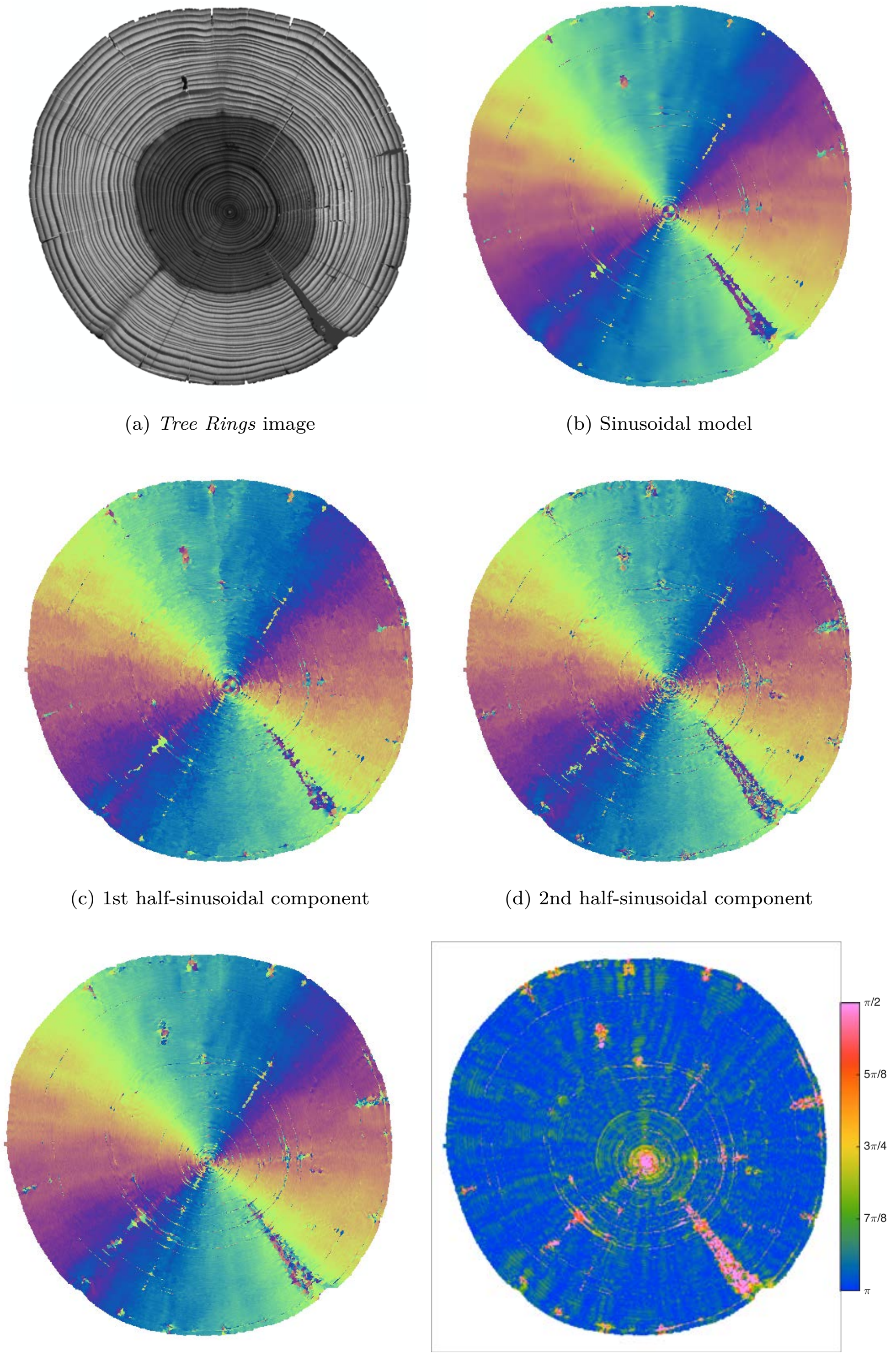
(Figure 5.26e), although not as good as the sinusoidal model. Neither model gives a good estimate at the very centre of the rings though, likely because the wavelets are too large. At a larger scale, the models see curved features as i1D features or line/edge segments. When the radius of curvature of the feature is around the same magnitude of the wavelet, the structure deviates too much from this assumption for a good estimation.

We can also calculate a measure of the *difference* between the two orientations,

$$\theta_{\text{diff}} = \arg(\cos(\theta_1 - \theta_2) + \mathrm{i}\sin(\theta_1 - \theta_2)). \tag{5.78}$$

The difference for the *Tree Rings* image is shown in Figure 5.26f as θ_{diff}. Values below $\pi/2$ have been truncated, such as in the pink region at the centre. Thus the two-component half-sinusoidal model gives an orientation estimate of both i1D features and the curved features, as well as an indication of the curvature with the angle difference. However, the angle difference appears to vary with number of RT orders used, and is therefore not a consistent measure.

One could consider that applying the sinusoidal model, or other orientation estimation methods such as MOPs and the structure tensor, as also performing an averaging of the curve orientation in the local area. The half-sinusoidal approach is novel in that it explicitly averages two segment components. The model is therefore useful for performing both orientation estimation and junction analysis in the one operation. However, as a general junction or corner detector the residual of the sinusoidal model is a more straight forward measure than using the half-sinusoidal model, because the residual of a two-component half-sinusoidal model is low for both i1D features *and* i2D corner features and thus requires further analysis of the component orientations.



(a) *Tree Rings* image

(b) Sinusoidal model

(c) 1st half-sinusoidal component

(d) 2nd half-sinusoidal component

(e) Average half-sinusoidal

(f) Difference between half-sinusoidal component orientation

Figure 5.26: Sinusoidal model orientation (b) compared to the first (c), second (d) and averaged (e) half-sinusoidal model components. The difference between the orientations of each half-sinusoidal component are shown in (f).

5.9 Feature Sets

An advantage of the line and edge segment wavelets is that they can be used to construct more complex wavelets matched to specific junctions. For example, adding three line segment wavelets spaced $2\pi/3$ radians apart in orientation gives a wavelet matched to a Y junction. Thus rather than deriving matched wavelets for a particular feature or angular response profile, we can find the half-sinusoidal model parameters for the feature and used the resulting model CH vector to construct the wavelet. This is equivalent to adding constraints to the model.

Table 5.1 shows the CH vectors for various features created from the addition of either the line or edge segment wavelet for $N = 4$. The shape of the wavelets depends on the number of RT orders used. Some examples of common features and the corresponding wavelets for different values of N are shown in Figure 5.27. The following observations can be made:

- As N increases, the overall size of the wavelets increase, however the individual segments have the same narrow angular profile. This increases the orientation selectivity of the wavelets.
- Features with n-th order rotational symmetry require at least the N-th order CH wavelet to discriminate the feature from an isotropic blob. This can be seen for the Y and X junctions in particular. It gives a visual example of why the signal multi-vector, which only uses up to the 3rd order RT, cannot be used to determine the angle of two perpendicularly crossed lines despite having two sinusoidal components in the model.
- Features with higher-order symmetry often require the addition of more than one RT order to improve their shape. For example, the shape of the wavelet corresponding to the chequer feature is the same for $N \in \{2, 3, 4\}$, and changes every $2 + 4k$ orders, where k is a positive integer. This must be taken into account when performing experiments as N would need to be varied enough to change the wavelet.

5.9.1 Example

In [73] Marchant and Jackway used line, edge, Y, T, X and blob matched wavelets with $N = 4$ to analyse the junctions of a bee-wing image. A similar approach was applied to the dragonfly wing shown in Figure 5.28a. Wavelets corresponding to each of these features with equal amplitude line or edge segments were used. The procedure followed to analyse the image was

1. **CH vector:** The CH vector was calculated over three scales of a log-Gabor filter (wavelength $\in \{8, 16, 32\}$, $\sigma = 0.65$). Three scales were used in order to capture the different junction sizes and because according to the principle of phase congruency [58], which shall be discussed in the next chapter, good features tend to have the same *shape* over multiple scales.
2. **Feature modelling:** The amplitude and orientation response for each feature wavelet was obtained for each scale. An illumination invariant detection score was calculated by dividing

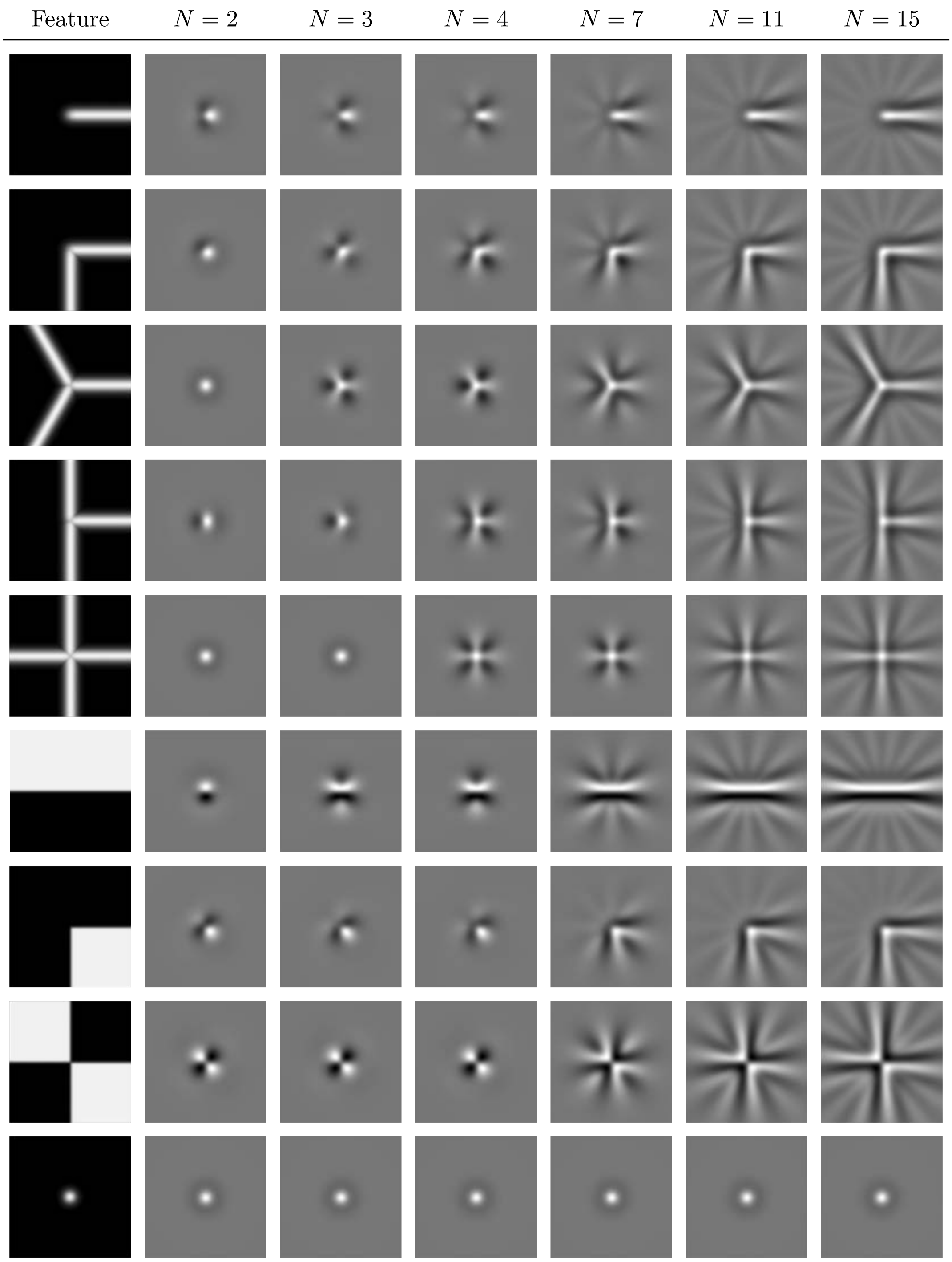


Figure 5.27: Matched wavelets constructed using the half-sinusoidal wavelets for for common image features and different values of N.

Feature	f_{-4}	f_{-3}	f_{-2}	f_{-1}	f_0	f_1	f_2	f_3	f_4	$\|\mathbf{f}\|$
Line Segment:	1	i	-1	-i	1	-i	-1	i	1	$\sqrt{9}$
Line: $\theta = [0, \pi]$	2	0	-2	0	2	0	-2	0	2	$\sqrt{20}$
'Y': $\theta = [\frac{\pi}{6}, \frac{5\pi}{6}, -\frac{\pi}{2}]$	0	3	0	0	3	0	0	3	0	$\sqrt{27}$
'T': $\theta = 0, \pi, -\pi/2]$	3	1	-1	1	3	-1	-1	-1	3	$\sqrt{33}$
'X': $\theta = [\frac{\pi}{4}, \frac{3\pi}{4}, -\frac{\pi}{4}, -\frac{3\pi}{4}]$	-4	0	0	0	4	0	0	0	-4	$\sqrt{48}$
Edge Segment:	i	-1	-i	1	0	-1	i	1	-i	$\sqrt{9}$
Edge: $\theta = [0, \pi]$	0	-2	0	2	0	-2	0	2	0	$\sqrt{16}$
Corner: $\theta = [0, \frac{\pi}{2}]$	0	-1+i	-2i	1+i	0	-1+i	2i	1+i	0	$\sqrt{16}$
Chequer: $\theta = [0, \frac{\pi}{2}, \pi, -\frac{\pi}{2}]$	0	0	-4i	0	0	0	4i	0	0	$\sqrt{32}$
Blob:	0	0	0	0	1	0	0	0	0	1

Table 5.1: CH vectors for different feature archetypes for $N = 4$ with no weighting.

the the model CH vector norm by the image CH vector norm for each feature type and scale,

$$\mathrm{det}_{\mathrm{m},i} = \frac{\|\mathbf{W}\mathbf{f}_{\mathrm{m},i}\|}{\|\mathbf{W}\mathbf{f}_i\|}, \tag{5.79}$$

where m is the feature type index, and i is the scale index.

3. **Detection score:** A final detection score was calculated by summing the individual scale scores for each type. Again, summing over multiple scales instead choosing the maximum scale refines the detection points to good features whose energy is spread across the spectrum. The measure is

$$\mathrm{det}_{\mathrm{m}} = \sum_{i=1}^{3} \mathrm{det}_{\mathrm{m},i}. \tag{5.80}$$

4. **Classification:** Each point was then classified according to the type which gave the maximum detection (Figure 5.28b),

$$\mathrm{score} = \max_m \mathrm{det}_{\mathrm{m}}, \tag{5.81}$$

$$\mathrm{class} = \arg\max_m \mathrm{det}_{\mathrm{m}}. \tag{5.82}$$

5. **Detection locations:** Candidate junction locations were chosen as the largest score in a local 5 pixel radius area from *only* the pixels that were classified as a Y, T, or X junction feature. This can be done simply by zeroing the score at the location of the other classes and using the local maxima. It is a crucial aspect of this method. The other features are necessary to discriminate each type of structure, however we cannot use the local maxima without accounting for class otherwise we would miss detections of the junctions due to adjacent higher scores corresponding to the uninteresting features (line, edge, blob).

6. **Post-processing:** Detection locations where the sum of the model vector norm,

$$\sum_{i=1}^{3} \|\mathbf{W}\mathbf{f}_{\mathrm{m},i}\|, \tag{5.83}$$

was below a noise threshold were removed (Figure 5.28c).

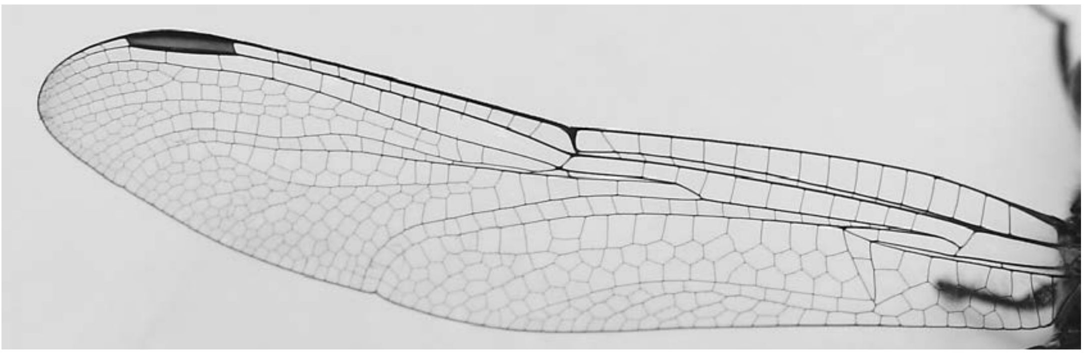
One problem with the wavelet set used was that many of the T junctions in the image consist of a heavy line for the bar and a light line for the stem. This resulted in these locations being classified as *lines* instead of T junctions. To compensate, the T junction was adjusted to give a higher score for these features by reducing the amplitude of the stem component by half. The detection was particularly improved along the edges of the wings (Figure 5.28d).

5.9.2 Wavelet Selectivity

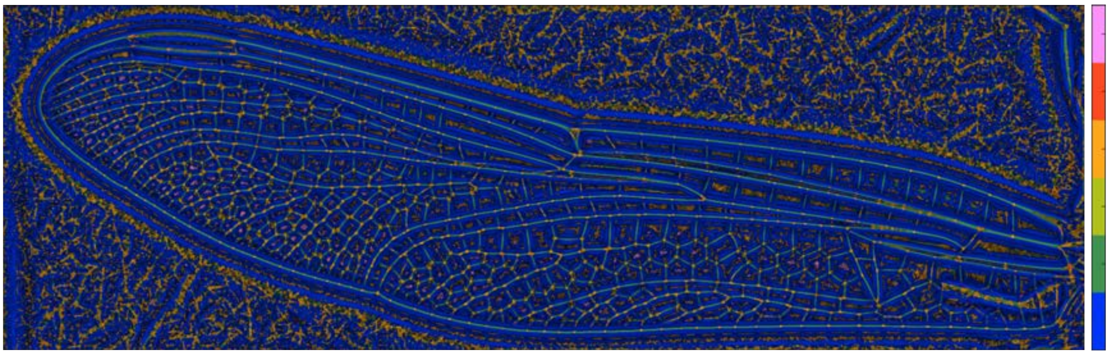
Adding extra RT orders does not necessarily improve the detection performance using feature-matched steerable wavelets (Figure 5.27). Increasing N causes the wavelets to become more orientation selective due to their increased spatial extent. This is beneficial for obtaining more precise orientation estimates, but in practical applications it can lead to worse detection performance. In the *Dragonfly* example, increasing N from 4 to 13 resulted in much fewer junctions being correctly classified. This is because the junctions in the image do not always exactly match the archetypal features. For example, the stem of a T junction may not be perpendicular to the bar, and thus will not correlate with the bar segment of the archetypal T junction wavelet with components at right angles.

Recent approaches to detecting rotationally symmetric junctions, such as Y and X junctions, and junctions with specific angular profiles, such as T junctions, using CH wavelets have been proposed by Puspoki et al. in [105] and [103], respectively. They also add extra CH orders to increase the orientation selectivity of the wavelets and consequently the detection performance under noise conditions. Their approach is very similar to that proposed above, and thus will also suffer from the aforementioned problem of detecting semantically similar but warped junctions. Indeed, most of the experimental results in [103] are constrained to be perfectly rotationally symmetric.

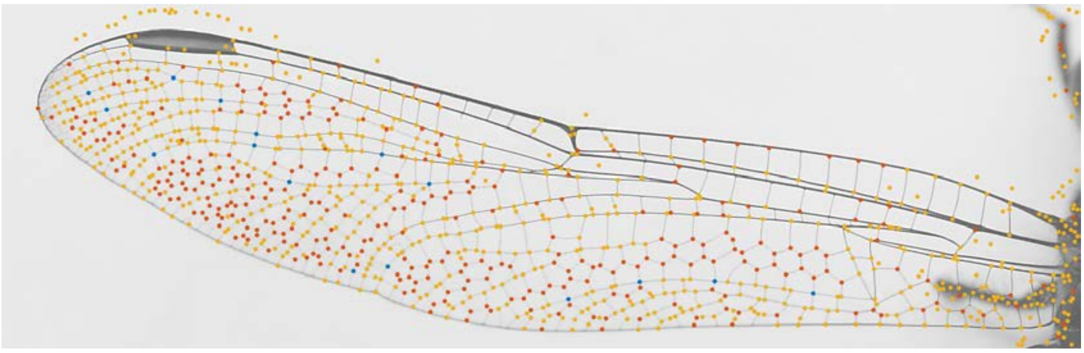
Paradoxically, it seems that instead of increasing N, using just enough RT orders to be discriminative and combining detection scores over multiple scales is a better approach to detecting junctions with variable orientation components. Another approach would be to use wedge filters as their angular profile can be adjusted so that features with a wider range of segment orientations will still have a high correlation. A third alternative is go back to using the half-sinusoidal model. It allows for more orientation selective wavelets by using higher-order RTs, but since they can be individually rotated they will still match junctions with varying angles. A plausible scheme is therefore to check if the half-sinusoidal model components fall within certain orientation ranges that match the features of interest. This is the subject of future work.



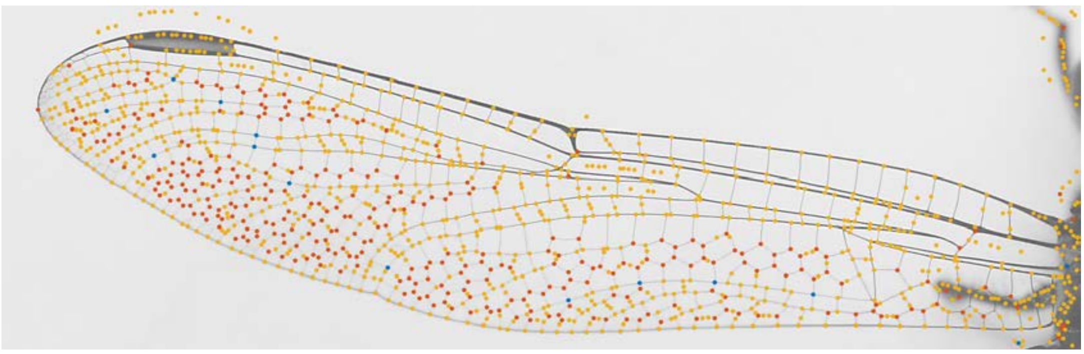
(a) *Dragonfly* image



(b) Classification (colour) and detection score (brightness) for each feature type.



(c) Junctions classified as either Y (red), T (orange) or X (blue).



(d) Junctions classified as either Y (red), T (orange) or X (blue) using an alternative T junction wavelet.

Figure 5.28: *Dragonfly* image junctions classified using a feature wavelet set consisting of line, edge, Y junction, T junction, X junction and blob wavelets for $N = 4$.

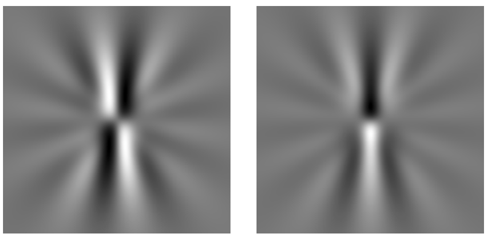
5.10 Summary

The main theoretical contributions in this chapter is the derivation of the RT response to a line-segment and edge-segment feature. Line and edge segments do not have straight-forward frequency domain representations like for the sinusoidal model, so the derivation was performed in the spatial domain. It is interesting to note that the integral along a line radiating from the centre of a CH wavelet is the same regardless of order.

The wavelets corresponding to both feature types were combined into a half-sinusoidal model. The advantage of this model is that it provides a phase-based representation of junctions and corners, where amplitude describes the strength of the segments, orientation describes their direction, and phase describes whether they are a line or an edge segment. A further advantage is that off-centre estimation of junctions is improved compared to using a single line-segment or edge-segment wavelet type. This is a problem for other techniques such as MSMFs.

The half-sinusoidal model can be solved for using either the iterative or roots methods. However, using the super-resolution method restricts the model to exclusively either line-segment or edge-segment components. The advantage is that one may resolve model components closer in orientation than when using the iterative or roots methods. An interesting new application is the use of the two-component half-sinusoidal model as an alternative to the sinusoid model for the orientation estimation of i1D features. The estimate is given by the average of the two model orientation parameters, while their difference is an indication of how curved the local image structure is.

Using these line and edge segment archetypes, other more complex features can be constructed. However, the improved orientation selectivity with increasing N may actually give worse results if the actual features do not match the exact orientations of the archetypes. In these situations calculating the half-sinusoidal model and classifying based on the model parameters may be a better approach. Another possible application of the half-sinusoidal model is analysis of the chequer edge pattern in Figure 5.1 using a pair of *odd* quadrature wavelets, shown in Figure 5.29.



(a) Odd-odd wavelet (b) Even-odd wavelet

Figure 5.29: Odd quadrature wavelet pair for the analysis of chequer patterns, for $N = 11$.

Chapter 6

Scale Harmonic Wavelets

6.1 Introduction

One of the early decisions in the image processing pipeline is the scale at which image analysis should be performed. For example, when applying Canny's famous edge detector [16] one must first choose the variance of the Gaussian filter with which to smooth the image. Likewise, in the previous chapter, we saw that the wavelength of a log-Gabor filter should be at least twice the width of a line-segment for junction analysis with the CH vector. However, features can occur at many different sizes in an image, or over a continuous range of scales, such as in the *Dune* image in Figure 6.1. The dunes in the distance appear at both coarse and fine scales, while the line features corresponding to ripples in the sand become continuously smaller towards the centre of the image.

Figure 6.1: *Dune* image consisting of features with continuously varying scale.

A multi-scale approach to the analysis of an image is therefore useful. One may wish to evaluate a range of scales and chose the best one at which to perform analysis, or combine the response from multiple scales, to achieve a degree of scale invariance. For example, when applying the i2D feature detection method in Chapter 3, the sum of the response over four scales was used, while in Chapter 4, five scales were used to obtain the orientation estimates for coral core analysis.

A common approach to scale selection is to choose the best scale using the set of responses to a discrete filter bank. The best scale being the one that gives the maximum response according to a particular measure. Lindeberg proposed a method of automatic selection using the determinant of the Hessian matrix of normalised derivatives [64], and investigated other differential entities for scale selection and image matching in [65] and [66], respectively. Fdez-Valdivia et al. [24] used the response to 2D Gabor filters as a measure, and Kadir and Brady [53] used a saliency score given by the entropy of the local image patch. Selection of the appropriate scale is a crucial aspect one of the most popular feature descriptors, the *scale-invariant feature transform* (SIFT) [69]. It locates features by finding the maximum response to the difference of Gaussian operator over many discrete scales, usually three or four per octave.

One problem with using discrete filter banks is that they are centred at specific frequencies (or scales). Thus if the main frequency component of a structure of interest lies between two filter channels, its energy will be split across them. For image processing operations that rely on the magnitude of the filter response this may cause issues, such as the energy of a desired feature falling below a detection threshold. The obvious approach is to use more finely spaced scales, however this increases the number of filtering operations that must be performed.

Along with filters that are steerable in orientation, Simoncelli [111] and Perona [99] also introduced filters that are *shiftable* in scale. The same steerable principle applies, a set of complex exponential functions in the radial frequency domain deliver a set of basis filters, which can be linearly combined to approximate a filter dilated to a particular scale. Ng and Bharath [93] applied this idea to estimate the coefficients of the scale basis filters from a set of discrete responses. In this way they were able to obtain a *continuous* trigonometric polynomial representation of the energy of a the response to a steerable quadrature filter across scale, and thus select the best scale as that which corresponds to the maximum of the polynomial.

In this chapter a method of scale selection and scale adaptive analysis using the CH vector is presented. Instead of derivative based measures such as the determinant of Hessian, we shall use the energy of the CH vector as the measure for selecting the optimal scale. The development of the method is as follows:

1. A wavelet frame that can be used to approximate a filter shifted through scale, using the principles in [99], is constructed.
2. The frame is augmented with higher-order RTs to create a steerable wavelet frame. The responses from the wavelets are collected into a quaternionic matrix that represents the *continuous* CH vector response shifted through scale.
3. The polynomial representing the CH vector energy is used for scale selection and the generation of other statistics that give information about the scale response. Other measures for picking the best scale are proposed.
4. The CH vector scale response is used to develop a continuous version of phase congruency.

The work in this chapter is preliminary in nature but has been included in the thesis due to the novel approach. Earlier work on some sections has been published in [71]; however, it neglected the connection with Perona's deformable kernels [99].

6.2 Scale Harmonic Wavelet Frame

The goal of this development is to create a continuous representation of the CH vector through scale. To do this we need a suitable primary isotropic wavelet frame from which the higher-order RTs can be added to give a steerable wavelet frame. In this section, we shall generate a primary isotropic wavelet frame from the Fourier series basis functions of an interval in the logarithmic frequency domain, similar to the filters proposed in [99]. The advantage of these wavelets is that they can approximate other isotropic wavelets, which can then be shifted through scale by steering the basis filter responses, much the same as how 2D steerable wavelets can be steered through orientation using the CH wavelet responses.

6.2.1 Concept

To begin with, consider a simple isotropic filter bank consisting of a 2D low-pass filter with frequency response, $l(\omega)$, and a high-pass filter, $h(\omega)$, where $\omega = |\boldsymbol{\omega}|$ and $\boldsymbol{\omega}$ is the Fourier domain coordinates. The filter bank is a simple Parseval-tight wavelet frame with infinite vanishing moments, and is compactly supported such that $h(\omega)$ is 0 below a certain frequency. That is,

$$l(\omega)^2 + h(\omega)^2 = 1, \tag{6.1}$$

$$\left.\frac{\mathrm{d}^n h(\omega)}{\mathrm{d}\omega^n}\right|_{\omega=0} = 0, \text{for } n \in \mathbb{N}, \tag{6.2}$$

$$l(\omega) = 0 \quad \forall\, \omega > \omega_c, \tag{6.3}$$

$$h(\omega) = 0 \quad \forall\, \omega < \omega_c/2, \tag{6.4}$$

where ω_c is the cut-off frequency. Below that, the high pass filter spectrum tapers down to 0 at $\omega_c/2$. An example of this type of filter bank based on Simoncelli's wavelets is [101],

$$h(\omega) = \begin{cases} \cos\left(\frac{\pi}{2}\log_2\left(\frac{\omega}{\omega_c}\right)\right), & \omega_c/2 < \omega \leq \omega_c, \\ 1, & \omega > \omega_c, \\ 0, & \omega \leq \omega_c/2, \end{cases} \tag{6.5}$$

$$l(\omega) = \sqrt{1 - h(\omega)^2}, \tag{6.6}$$

which we will use for the rest of the chapter.

Next, a set of $2M + 1$ complex-valued basis wavelets is constructed by multiplying the high pass filter response by a complex exponential in the *logarithmic frequency domain*,

$$h^m(\omega) = h(\omega)\, e^{\mathrm{j}mr}, \tag{6.7}$$

where $m \in [-M, ..., M]$. Note, that j is used instead of i for the imaginary number, as i will be used for the complex RT later. The parameter r is a scaled and shifted logarithmic frequency value given by.

$$r = \pi \frac{\log_2(\omega) - r_c}{r_w}, \tag{6.8}$$

$$r_c = \log_2(\omega_c/2), \tag{6.9}$$

$$r_w = \log_2(\sqrt{2}\pi) - r_c, \tag{6.10}$$

such that $r = 0$ corresponds to the low frequency cut-off, $\omega = \omega_c/2$, and $r = \pi$ corresponds to highest valid frequency in the discrete image spectrum, $\omega = \sqrt{2}\pi$. An octave bandwidth in the normal frequency domain corresponds to a bandwidth of π/r_w in this scaled and shifted logarithmic domain. To convert from the logarithmic frequency, r, to wavelength, the following formula is used,

$$\lambda = 2^{-\left(\frac{r}{\pi} r_w + r_c\right)}. \tag{6.11}$$

The functions that generate these basis wavelets are thus the first M Fourier series functions of the subspace $L_2(r), r \in [0, 2\pi)$. The logarithmic frequency domain is used because image features tend to be distributed logarithmically, and thus it is common to partition the spectrum as such. For a discrete 2D image, the minimum wavelength in the Fourier domain is $\sqrt{2}$, so the above complex filters will only perform $n/2$ cycles within the image spectrum. The spectrum of the first six basis wavelets are shown in Figure 6.2 for a cut-off frequency of $\omega_c = 2\pi/32$ (wavelength 32).

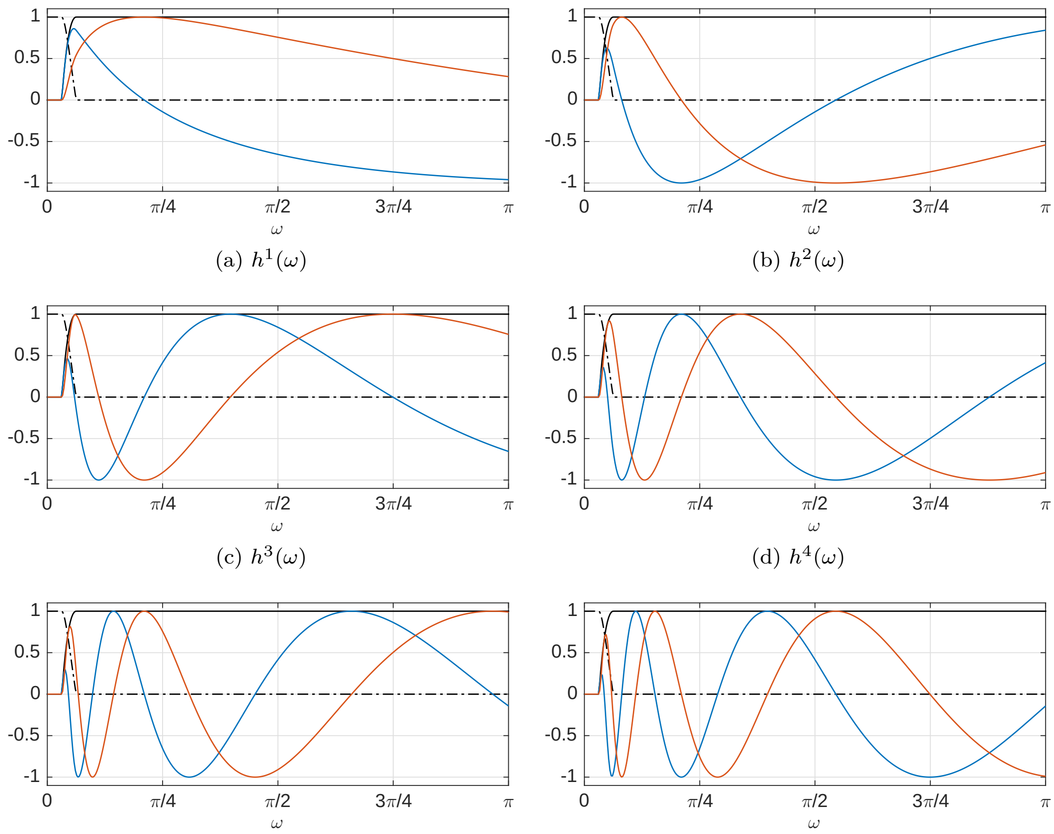
Given that

$$l(\omega)^2 + \sum_{|m| \leq M} |h^m(\omega)|^2 = 1, \tag{6.12}$$

the set of wavelets also form a tight wavelet frame for a discrete image or image with no frequencies above $\sqrt{2}\pi$. Such an image shall be assumed for the remainder of the development. The frame bounds are $2M + 1$, and there is $2M + 1$ redundancy; however, they cannot be sub-sampled.

These basis wavelets shall be called *scale harmonic* (SH) wavelets. We can collect the response of an image to the set of SH wavelets into a *scale harmonic* (SH) vector, similar to the CH vector. It is expressed by a convolution as

$$\mathbf{g}(\mathbf{z}) = \left[(f * \psi^M)(\mathbf{z}), ..., (f * \psi^{-M})(\mathbf{z})\right], \tag{6.13}$$



(a) $h^1(\omega)$ (b) $h^2(\omega)$ (c) $h^3(\omega)$ (d) $h^4(\omega)$ (e) $h^5(\omega)$ (f) $h^6(\omega)$

Figure 6.2: Real (blue) and imaginary (red) radial frequency spectrum of the first to sixth order basis filters, $h^m(\omega)$, for $M = 11$ and $\omega_c = 2\pi/32$. Dotted black line: Primary low pass frequency response, solid black line: primary high pass frequency response.

or in the wavelet notation with $\mathbf{k}$ as the location index,

$$\mathbf{g}_\mathbf{k} = \left[\langle f_\mathbf{k}, \psi_\mathbf{k}^{-M}\rangle, ..., \langle f_\mathbf{k}, \psi_\mathbf{k}^{-M}\rangle\right], \tag{6.14}$$

where the wavelets are defined as

$$\psi^m(\mathbf{z}) \quad \stackrel{\mathcal{F}}{\longleftrightarrow} \quad h^m(\boldsymbol{\omega}). \tag{6.15}$$

For exact reconstruction the frame must be Parseval tight. Let $\mathbf{V}$ be a weighting matrix,

$$\mathbf{V} = \mathrm{diag}(\mathbf{v}) \tag{6.16}$$

$$= [v_{-M}, ..., v_M], \tag{6.17}$$

where $\|\mathbf{v}\| = 1$. Reconstruction of the original image is then possible from the *weighted* SH vector plus the low-pass response,

$$f(\mathbf{z}) = (f * \psi_{\text{low}})(\mathbf{z}) + \sum_\mathbf{k} \sum_{|m|\leq M} (\mathbf{V}\mathbf{g}_\mathbf{k})_m v_m \psi_\mathbf{k}^m. \tag{6.18}$$

Note, that unlike normal wavelet frames, there is *no scale parameter.*

6.2.2 Approximating Isotropic Wavelets

Just as one can use the Fourier series to approximate other functions, we can linearly combine these SH basis wavelets to approximate other isotropic wavelets. In the Fourier domain this is expressed as

$$h_{\mathbf{c}}(\omega) = \sum_{n=-M}^{M} c_m h^m(\omega) \tag{6.19}$$

where $\mathbf{c} \in \mathbb{C}^{2M+1}, c_n = \bar{c}_{-m}$ is a complex coefficient vector which defines the new wavelet frequency response $h_{\mathbf{c}}(\omega)$. Two methods are proposed to choose the coefficients:

- Finding the Fourier series coefficients of the desired filter function in the logarithmic frequency domain.
- Maximising the energy inside the envelope of a desired filter function in the logarithmic frequency domain.

Fourier Series

Let $v(\omega)$ be an ideal filter function with centre frequency $\omega_0 \in [\omega_c, \sqrt{2}\pi)$. One set of SH wavelet coefficients for the filter approximation are the Fourier series coefficients of $v(r)$,

$$c_m = \int_0^{2\pi} v(r) e^{-\mathrm{j}mr}\, dr, \tag{6.20}$$

where r is the logarithmic frequency. A simple ideal filter is a rectangular function with width α and centre τ on the logarithmic frequency scale. Thus for

$$v(r) = \begin{cases} 1 \text{ where } r \in [\tau - \alpha, \tau + \alpha], \\ 0 \text{ otherwise,} \end{cases} \tag{6.21}$$

we have

$$c_m = \int_{\tau-\alpha}^{\tau+\alpha} e^{-\mathrm{j}mr}\, dr \tag{6.22}$$

$$= e^{-\mathrm{j}m\tau} \frac{2\sin(\alpha m)}{m}. \tag{6.23}$$

Alternatively, this filter can be expressed as a filter centred at $r = 0$ and shifted up in frequency by the scale parameter, τ. In this case, we may write

$$v_m = \frac{2\sin(\alpha m)}{m} \tag{6.24}$$

and multiply by $e^{-\mathrm{j}m\tau}$ to shift the filter through scale.

Energy Maximisation

Using the same filter function we can instead maximise the energy within the window, applying the maximisation technique used in [105] for rotationally symmetric wavelets. Let $v(r)$ be the ideal filter function, and $h_{\mathbf{c}}(r)$ be the frequency spectrum of the filter approximation generated using the coefficients $\mathbf{c}$. The energy functional in the logarithmic frequency domain is

$$E[v] = \int_0^{2\pi} h_{\mathbf{c}}(r)^2 v(r)\, dr \tag{6.25}$$

$$= \sum_{|m'|\leq M} \sum_{|m|\leq M} \bar{\mathbf{c}}_{m'} \mathbf{c}_m \int_0^{2\pi} v(r)\, dr \tag{6.26}$$

$$= \mathbf{c}^H \mathbf{V} \mathbf{c} \tag{6.27}$$

where $V_{m,m'} = \int_0^{2\pi} e^{\mathrm{i}(m-m')r} v(r)\, dr$. Since $\mathbf{c}^H\mathbf{c} = 1$, choosing $\mathbf{c}$ as the eigenvector of $\mathbf{V}$ with the largest eigenvalue gives the coefficients with maximum energy within the window. As for the Fourier series approximation, we can instead find the coefficients of $\mathbf{c}$ when the filter is centred at $\tau = 0$ and multiply by $e^{-\mathrm{j}m\tau}$ to shift through scale. Increasing the number, M, of basis filters gives a better approximation of the desired filter function (Figure 6.3). In particular, enough basis functions are needed to reduce oscillations outside the window; 15 are needed for the one octave window, and seven for a two octave window, when $\omega_c = 2\pi/64$. Note that using a window width of two octaves gives a filter of approximately one octave bandwidth using the energy minimisation method and $M = 11$. This type of filter shall be used in the remainder of this chapter.

The basis filter generating functions have only $m/2$ periods within the spectrum, as due to their periodicity the filter spectrum approximation will wrap-around the interval and thus a buffer is needed. For example, a filter centred near the start of the interval wraps around to the end of the interval, as shown in Figure 6.4. However, because of the buffer, the wrap-around part of the response lies in the region from $r = \pi$ to $r = 2\pi$ which is outside of the valid frequency domain of a discrete image and therefore does not affect the result. For approximations which are practically zero outside of a limited support, such as the examples in Figure 6.3, the number of periods within the frequency spectrum could be increased by changing the scaling of r in (6.10). This would mean a lower number of basis filters would be needed to achieve a similar approximation. If too few basis filters are used, the wrap-around part of the spectrum will infringe into the image domain, resulting in high frequency components being included with a low frequency filter, and vice-versa.

6.2.3 Isotropic Scale Response

All the operations to approximate a filter in the frequency domain using SH wavelets are linear. Therefore we can approximate a filter, and therefore filtering of an image, using the same linear combination of the SH wavelet *responses*. This is an important distinction between this method and

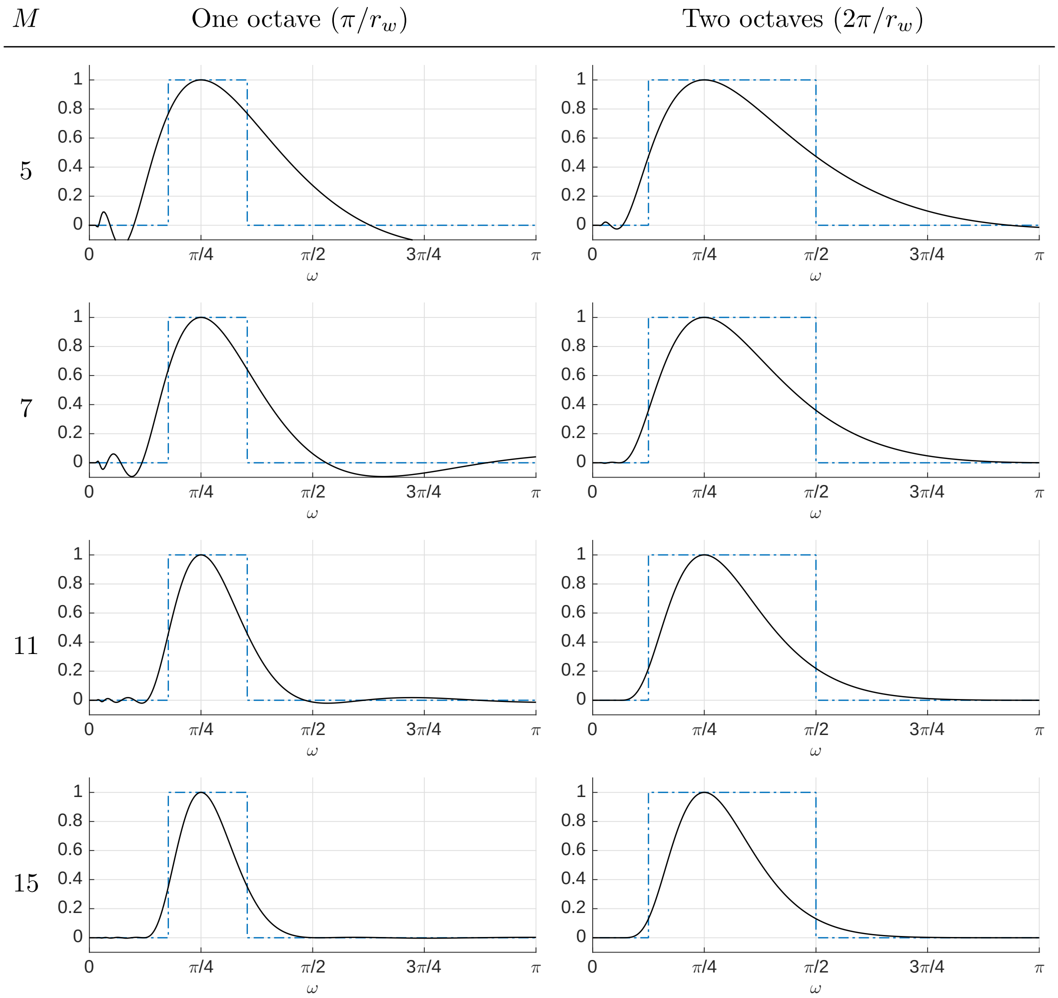


Figure 6.3: Spectrum of a filter (black line) with centre frequency $\omega_c = 2\pi/8$ constructed using the energy minimisation method different values of M and a window width of one or two octaves (dotted blue line). A two octave window results in a filter with approximately one octave bandwidth.

a discrete filter bank. Normally, to calculate the response to M discrete filters, M inverse Fourier transform operations are required. With the SH wavelet frame, any number of filter responses can be approximated from M SH wavelets, and still only $M+1$ inverse Fourier transform operations are required. (The negative orders are simply the conjugate of the positive).

From the SH wavelet responses we may obtain a *polynomial* representation of the response to a filter through scale, for a particular location in the image. Let $\mathbf{v}$ be the coefficients corresponding to the band pass filter profile centred at $\tau = 0$. The coefficients of the filter shifted to scale τ are therefore given by $\mathbf{S}_\tau \mathbf{v}$, where $\mathbf{S}_\tau$ is the same rotation matrix used in previous chapters,

$$\mathbf{S}_\tau = \mathrm{diag}\left[e^{-\mathrm{j}M\tau}, ..., e^{\mathrm{j}M\tau}\right]. \tag{6.28}$$

Normalising such that $\|\mathbf{v}\| = 1$ we see that these coefficients are a good choice for the scale weighting matrix $\mathbf{V}$. Due to the linearity of the operations, the filter response at scale τ is easily

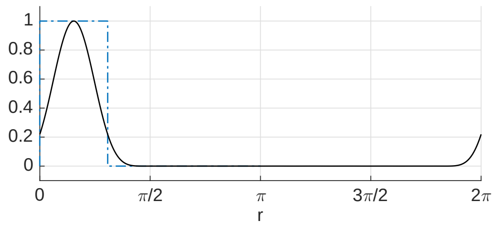


Figure 6.4: Spectrum of a filter constructed for $M = 11$ with window width $2\pi/r_w$ shown on the scaled and shifted logarithmic frequency domain. The filter response wraps around but does not extend into the valid part of the image spectrum, $\omega \in [\omega_c/2, \sqrt{2}\pi]$, which is the interval $r \in [0, \pi]$.

calculated from the SH wavelet responses, $\mathbf{g}$, as

$$\langle f, \psi_{\mathbf{v}}(\tau)\rangle = \sum_{|m|\leq M} (\mathbf{S}_\tau \mathbf{V}\mathbf{g})_m. \tag{6.29}$$

This can be written as a trigonometric polynomial in τ,

$$p(\tau) = \sum_{|m|\leq M} v_m g_m e^{\mathrm{j}m\tau}, \tag{6.30}$$

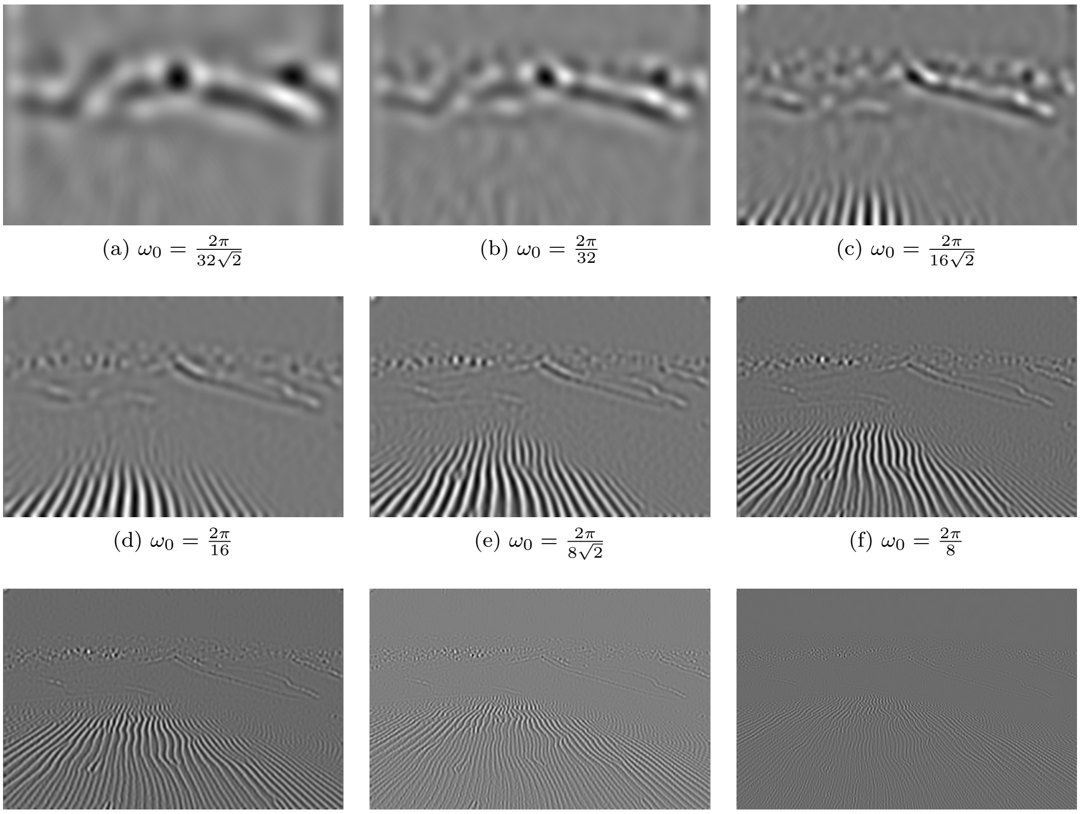
which represents the *continuous* response to the isotropic filter as it is shifted through scale. The filter is therefore *shiftable* in the scale parameter τ.

An example of filter responses for the *Dunes* image using $M = 11$ and $\omega_c = 2\pi/64$ is shown in Figure 6.5. Different sized features are isolated at different scales. The dunes in the background of the image appear in the larger scales, whereas the ripples in the sand appear in the smaller scales depending on their size.

6.3 Scale Circular Harmonic Wavelet Frame

The magnitude of the scale response given by the SH wavelets gives an idea of the scale at which the local image structure is strongest. However, the SH basis wavelets are isotropic and so any filter approximated by their linear combination is also isotropic. Therefore the scale response only represents the magnitude of even, symmetric image structures, such as lines, and not odd, anti-symmetric structures such as edges, and thus is not phase invariant.

To perform scale selection we need a measure that also responds to anti-symmetric features, and thus odd wavelets are needed to augment the isotropic response. The CH vector gives a representation of local image structure at a particular scale, and with appropriate weighting the CH vector magnitude is a measure of local energy that is invariant to phase. Higher-order RTs of a suitable primary isotropic wavelet frame also form a 2D steerable wavelet frame if they match the constraints in [124]. Since the SH wavelets meet these conditions, a new set of 2D steerable wavelets can be created from their RTs. This gives us a representation of the CH vector through scale, from which the CH vector energy is a phase-invariant measure that can be used for scale



(a) $\omega_0 = \frac{2\pi}{32\sqrt{2}}$ (b) $\omega_0 = \frac{2\pi}{32}$ (c) $\omega_0 = \frac{2\pi}{16\sqrt{2}}$

(d) $\omega_0 = \frac{2\pi}{16}$ (e) $\omega_0 = \frac{2\pi}{8\sqrt{2}}$ (f) $\omega_0 = \frac{2\pi}{8}$

(g) $\omega_0 = \frac{2\pi}{4\sqrt{2}}$ (h) $\omega_0 = \frac{2\pi}{4}$ (i) $\omega_0 = \frac{2\pi}{2\sqrt{2}}$

Figure 6.5: Response to an isotropic filter evaluated at discrete scales corresponding to the centre frequency ω_0, using the SH wavelets for $M = 11$, $\omega_c = 2\pi/64$ and window width $2\pi/r_w$.

selection.

6.3.1 Adding Higher-Order Riesz Transforms

The CH vector for each SH basis wavelet is given by the higher-order RTs of the wavelet. Since both the SH and CH wavelets are complex valued, an alternative representation is required to differentiate the complex part of the SH wavelet and the complex part of the CH wavelet. A *quaternionic* representation shall be used. Quaternions are an extension of the complex number system consisting of three imaginary components, i, j and k, where

$$\mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = \mathrm{ijk} = -1 \tag{6.31}$$

We shall let the complex part of the RT be represented by i, and the complex part of the SH wavelets be represented by j. Then the CH vector using the SH wavelet indexed by m as its primary wavelet is

$$\mathbf{Wf}^m = \left[\left\langle f, w_{-N} v_m \psi^{-N,m,}\right\rangle, ..., \left\langle f, w_N v_m \psi^{N,m}\right\rangle\right]^T, \tag{6.32}$$

where $\psi^{n.m}$ is a quaternion-valued wavelet given by the RT of the SH wavelet,

$$\psi^{n,m} = \mathcal{R}^n \psi^{0,m} \quad \overset{\mathcal{F}}{\longleftrightarrow} \quad e^{\mathrm{i}n\theta} h^m(\boldsymbol{\omega}) \tag{6.33}$$

$$\overset{\mathcal{F}}{\longleftrightarrow} \quad e^{\mathrm{i}n\theta} e^{\mathrm{j}mr} h(r). \tag{6.34}$$

Collecting the CH vectors for each SH basis wavelet into the columns of an $N \times M$ matrix we have

$$\begin{bmatrix} \langle f, w_{-N} v_{-M} \psi^{-N,-M} \rangle & \cdots & \langle f, w_{-N} v_M \psi^{-N,M} \rangle \\ \vdots & \ddots & \vdots \\ \langle f, w_N v_{-M} \psi^{N,-M} \rangle & \cdots & \langle f, w_N v_M \psi^{N,M} \rangle \end{bmatrix} \tag{6.35}$$

which can be written as

$$\mathbf{WFV}, \tag{6.36}$$

where $\mathbf{W}$ is the weighting matrix of the CH vector, $\mathbf{V}$ is the weighting matrix of the scale basis wavelets and defines the shiftable filter, and $\mathbf{F}$ shall be called the *scale circular harmonic* (SCH) matrix,

$$\mathbf{F} = \begin{bmatrix} \langle f, \psi^{-N,-M} \rangle & \cdots & \langle f, \psi^{-N,M} \rangle \\ \vdots & \ddots & \vdots \\ \langle f, \psi^{N,-M} \rangle & \cdots & \langle f, \psi^{N,M} \rangle \end{bmatrix}. \tag{6.37}$$

Since the weighted SH wavelets form a primary isotropic wavelet frame, the wavelets of the SCH matrix also form a steerable wavelet frame according to the conditions in [124], with $(2M+1) \times (2N+1) + 1$ redundancy. Due to the wavelets being in conjugate, only $(M+1) \times (N+1)$ filtering operations need to be made when implementing the method.

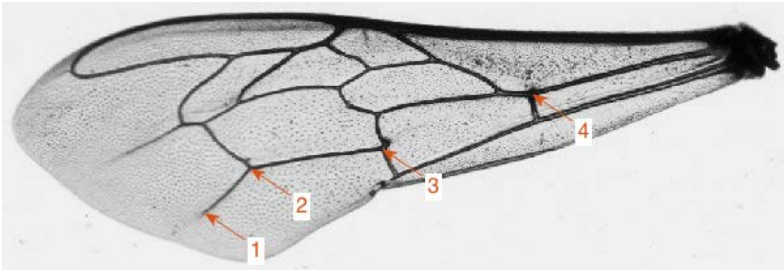
Using this representation, the response to a 2D steerable wavelet, $\mathbf{Wu}$, rotated by θ and shifted in scale by τ is given by

$$p(\theta, \tau) = \sum_n \sum_m \left(\mathbf{S}_\theta \mathbf{W} \mathbf{u}^H \, \mathbf{WFV} \, \mathbf{S} \tau \right)_{n,m} \tag{6.38}$$

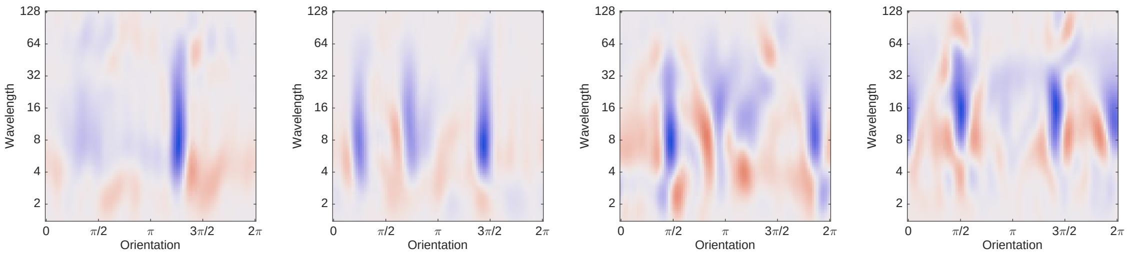
$$= \sum_n \sum_m w_n \bar{u}_n w_n F_{n,m} v_m e^{\mathrm{i}n\theta} e^{\mathrm{j}mr}. \tag{6.39}$$

This is a real-valued trigonometric polynomial in two variables. The range of orientation values is $\theta \in [0, 2\pi)$ and the range of scale values that fall within the image spectrum is $\tau \in [0, \pi)$.

An example of this bivariate polynomial for the line-segment wavelet from the previous chapter is shown for the labelled points in Figure 6.6, using $M = 11$ and $N = 13$, and the phase-invariant equal weighting scheme with window width of $B_{0.1}$. The scale parameter τ has been converted to wavelength using (6.11). Both the shape, scale and intensity of the features is clearly observed in the polynomial images. For example, three peaks are observed in the response for the Y junction at point 2, and its maximum is at a wavelength of approximately 8 pixels. On the other hand, the T junction at point 4 has its peak around a wavelength of 16 pixels.



(a) *Beewing* image



(b) Point 1 (c) Point 2 (d) Point 3 (e) Point 4

Figure 6.6: Visual representation of the bivariate polynomial given by SCH matrix response for a line-segment feature, at four points on the *Bee-wing* image. Red represents a positive line-segment response, blue is negative. The SCH matrix was calculated using the same parameters as the previous figure.

6.3.2 Energy Response

In Section 2.1.2 we saw how the magnitude of the CH vector is a measure of the strength of the local image structure which is rotationally invariant. If it is weighted such that $W_e = W_o$ the measure is also *phase-invariant*. Therefore the magnitude of the CH vector *shifted through scale* gives a phase-invariant measure of strength of the local image structure through scale. Thus the CH vector magnitude is a measure that can be used for scale selection.

Each column of the SCH matrix represents the CH vector obtained from the m-th order SH wavelet. Each row represents the coefficients of complex-valued (in i) trigonometric polynomial (in τ) describing the response to n-th order CH wavelet as it is shifted through scale. We can write the latter polynomial as

$$
\begin{aligned}
f_n(\tau) &= \sum_m (\mathbf{WFVS}_\tau)_{n,m} && (6.40)\\
&= \sum_m w_n F_{n,m} v_m e^{\mathrm{j}m\tau}. && (6.41)
\end{aligned}
$$

The polynomial that represents the *energy* (square of magnitude) of the CH vector, which shall be denoted $A^2(\tau)$, is therefore given by the sum of squares of the magnitude of the individual CH orders,

$$A^2(\tau) = \sum_{|n|\leq N} |f_n(\tau)|_{\mathrm{i}}^2, \tag{6.42}$$

where $|.|_{\mathrm{i}}$ means the absolute value with respect to the complex number i, for example $|ae^{\mathrm{i}b}ce^{\mathrm{j}d}|_{\mathrm{i}} = ace^{\mathrm{j}d}$. Rewriting the CH wavelet scale response we get

$$|f_n(\tau)|_{\mathrm{i}}^2 = \left|\sum_m w_n \Re_{\mathrm{i}}\{F_{n,m}\} v_m e^{\mathrm{j}mr} + \mathrm{i}\sum_m w_n \Im_{\mathrm{i}}\{F_{n,m}\} v_m e^{\mathrm{j}mr}\right|^2 \tag{6.43}$$

$$= \left(\sum_m w_n \Re_{\mathrm{i}}\{F_{n,m}\} v_m e^{\mathrm{j}mr}\right)^2 + \left(\sum_m w_n \Im_{\mathrm{i}}\{F_{n,m}\} v_m e^{\mathrm{j}mr}\right)^2 \tag{6.44}$$

which is the addition of two real-valued degree $4M$ trigonometric polynomials in $e^{\mathrm{j}\tau}$. The functions $\Re_{\mathrm{i}}$ and $\Im_{\mathrm{i}}$ give the real and imaginary parts with respect to i, for example, $\Re_{\mathrm{i}}\{e^{\mathrm{i}a}e^{\mathrm{j}b}\} = \cos a\, e^{\mathrm{j}b}$. When implementing the algorithm in MATLAB for example, it is necessary to split the CH vector into real and imaginary parts as only complex number arithmetic is supported.

Adding each of the $2N + 1$ polynomials from each CH wavelet order gives the CH vector energy polynomial, $A^2(\tau)$, which is thus another degree $4M$ polynomial. Unfortunately, to obtain a polynomial representation for the CH vector *magnitude* would require the square root of $A^2(\tau)$ to be obtained algebraically. This has no solution in most cases, so we shall work with the CH vector energy instead.

Image Example

The CH vector magnitude was calculated at different frequencies and for different numbers of RT orders (N), using $M = 11$ scale basis filters with cut-off frequency $\omega_c = 2\pi/64$ and filter window width $2\pi/r_w$ (Figure 6.7). It is observed that the sand dune features in the distance have a high magnitude for long wavelengths, while the *edges* of the sand dunes have a large magnitude across a wide range of wavelengths. The ripples in the foreground respond to different scale values in proportion to their size. The magnitude thus provides a clue as to which scale the local structure should be analysed, although at some locations it appears that more than one scale would be appropriate.

Increasing N appears to smooth the response, and reduces the chequer-pattern that can be seen for $N = 1$. This pattern is particularly noticeable around the i2D features on the left horizon of the image for $\omega_0 = 2\pi/16$. Their regular nature does not appear in the original image, suggesting that it is an artefact of having too few RT orders. The energy for $N = 1$ is that of the monogenic signal vector, and the same effect can also be seen in the monogenic signal examples in previous chapters. As seen for the magnitude examples in Section 2.1.2 and the i2D detection in Section 3.3.3, adding extra RT orders gives a better local energy estimate. The CH vector energy therefore provides an alternative measure for selecting the best scale at which to apply analysis.

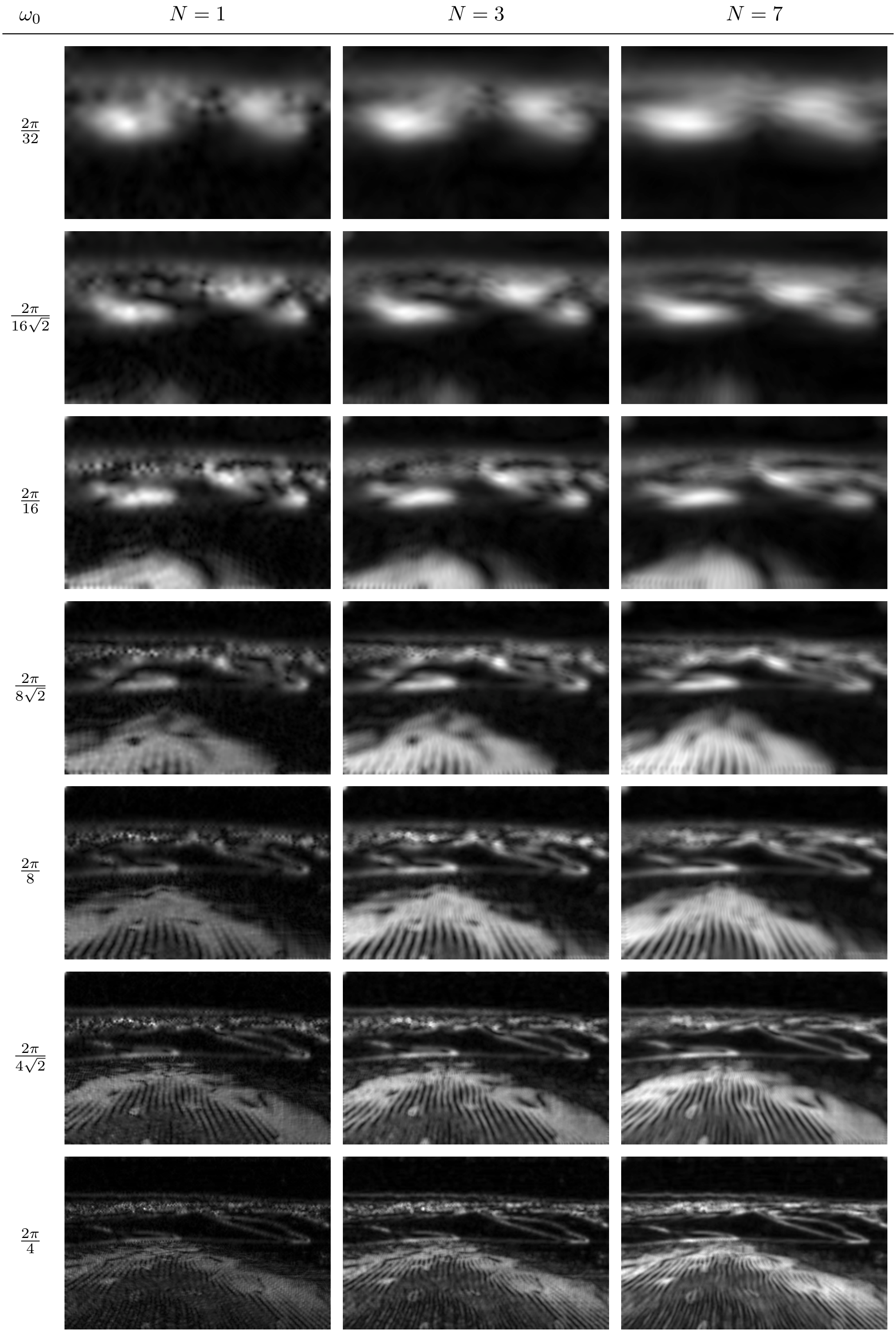


Figure 6.7: CH vector magnitude evaluated at discrete scales corresponding to the centre frequency ω_0, compared to N, using the SCH matrix with $M = 11$, $\omega_c = 2\pi/64$ and window width $2\pi/r_w$.

6.4 Scale Selection

6.4.1 Local Energy Statistics

The SCH matrix formed by the response of the SH wavelets and their RTs allows us a phase-invariant *continuous* measure of the CH vector energy, and thus local structure strength, over scale. An obvious statistic to calculate for scale selection is the scale at which the energy response polynomial is at a maximum. This is the same procedure that was used with the polynomial representation of quadrature filter energy in [93]. It is given by

$$\tau_{\mathrm{max}} = \underset{\tau\in[0,\pi)}{\arg\max} \quad A^2(\tau) \tag{6.45}$$

The maximum can be calculated by finding the roots of the derivative of this polynomial, then choosing the root that gives to the maximum.

However, we may also calculate some other statistics of the distribution *analytically.* These also give useful descriptions of the scale response that go beyond a simple "best scale". For example, we may wish to know if the energy is spread across a wide range of scales, if there are actually two scales at which the responses are high, or if the response is skewed towards one end of the frequency spectrum. These measures are given by the various moments of the energy polynomial, as described below.

Total Energy

The total energy is integral of the response,

$$A_{\mathrm{energy}} = \int_a^b A^2(\tau)\, d\tau, \tag{6.46}$$

which can be evaluated analytically. The integration limits a and b give the range of scale over which to calculate the statistic. An integration range of $[0, \pi]$ corresponds to the valid domain of the discrete image frequency spectrum.

Mean Scale

The mean scale (first moment) of the energy distribution is given by

$$\tau_{\mathrm{mean}} = \int_a^b \tau \frac{A^2(\tau)}{A_{\mathrm{energy}}}\, d\tau. \tag{6.47}$$

Unlike the scale maximum, the mean scale does not require root finding, and can be evaluated analytically.

Variance

Using the mean scale we can calculate the energy distribution across scale. The variance (second moment) is given by

$$\sigma_\tau^2 = \int_a^b (\tau - \tau_{\text{mean}})^2 \frac{A^2(\tau)}{A_{\text{energy}}}\, d\tau. \tag{6.48}$$

Lindeberg notes that strong features have their energy spread across scale [65], and this is one of the requirements behind phase-congruency based feature detection [58]. Therefore we expect variance to be higher at the locations of image features, although this assumes a uni-modal distribution.

Skewness and Kurtosis

Normalised skewness (how lopsided the distribution is) and kurtosis (a measure of its peakiness) are given by the third and fourth moments normalised by the variance according to

$$(\sigma_\tau^{(k)})^2 = \int_a^b (\tau - \tau_{\text{mean}})^k \frac{A^2(\tau)}{A_{\text{energy}}} \frac{1}{\sigma^k}\, d\tau, \tag{6.49}$$

with $k = 3$ and $k = 4$ respectively.

Bi-modality

Sarle's coefficient is a measure of how bi-modal a distribution is. It is given by the skewness and kurtosis as

$$\beta = \frac{(\sigma_\tau^{(3)})^4 + 1}{(\sigma_\tau^{(4)})^2}, \tag{6.50}$$

where a value of 0 indicates a purely uni-modal response, while a value of 1 indicates a purely bi-modal response.

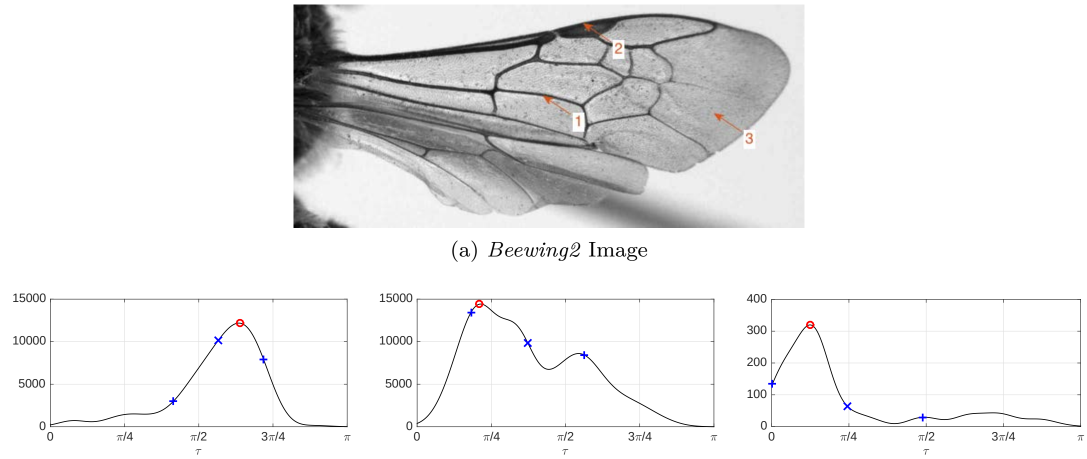
Integration Limits

The range of scales included when calculating the statistics is determined by the integration limits a and b. Choosing $a = 0$ and $b = \pi$ includes all the scales from the lower frequency cutoff, $\omega_c/2$, to the maximum frequency, $\sqrt{2}\pi$. Increasing a excludes longer wavelengths, as does increasing the cutoff frequency, ω_c. Likewise, reducing b removes high frequencies from the calculation, and thus decreases the influence of small features. Using $a = \pi/r_w$ starts the integration from $\omega = \omega_c$, where the high pass magnitude first reaches 1, and choosing $b = -\pi \log_2(\omega_c)/r_w$ sets the upper limit of integration to $\omega = 0.5$ which is wholly contained within the spectrum. Thus choosing these values excludes the part of the scale response which tapers, if so desired. However, using the full integration range of $[0, \pi]$ when calculating the mean scale appears to give better results for high frequency areas.

Example

The statistics were calculated for the *Beewing2* image (Figure 6.9a) to give an idea of the scale distribution of image features such as lines and edges (Figure 6.9b). The SCH matrix was obtained using $\omega_c = 2\pi/64$, $M = 11$, filter width $2\pi/r_w$, $N \in \{1, 3, 7\}$, and weighted using the phase-invariant equal weighing scheme.

The average energy is high around both strong odd and strong even image features, such as lines and edges, demonstrating phase invariance. The vein patterns on the bee wing are correctly identified as being at a fine scale, as are the small blob like textures. Both the variance and the bi-modality coefficient are very high in the textured areas of the bee wing cells. This is due to there being both a high frequency response to the texture, and another large low frequency response to the cell area (Figure 6.8d). The veins also have a reasonable scale variance and they are the only image feature with a large negative skewness. This is due to a large high-frequency response with a long tail down into the low frequencies (Figure 6.8b). In contrast, the edge features have a more uniform distribution of energy (Figure 6.8c). These results suggest that the scale response may be useful in discriminating line, edge and blob feature types.



(a) *Beewing2* Image

(b) Line feature

(c) Edge feature

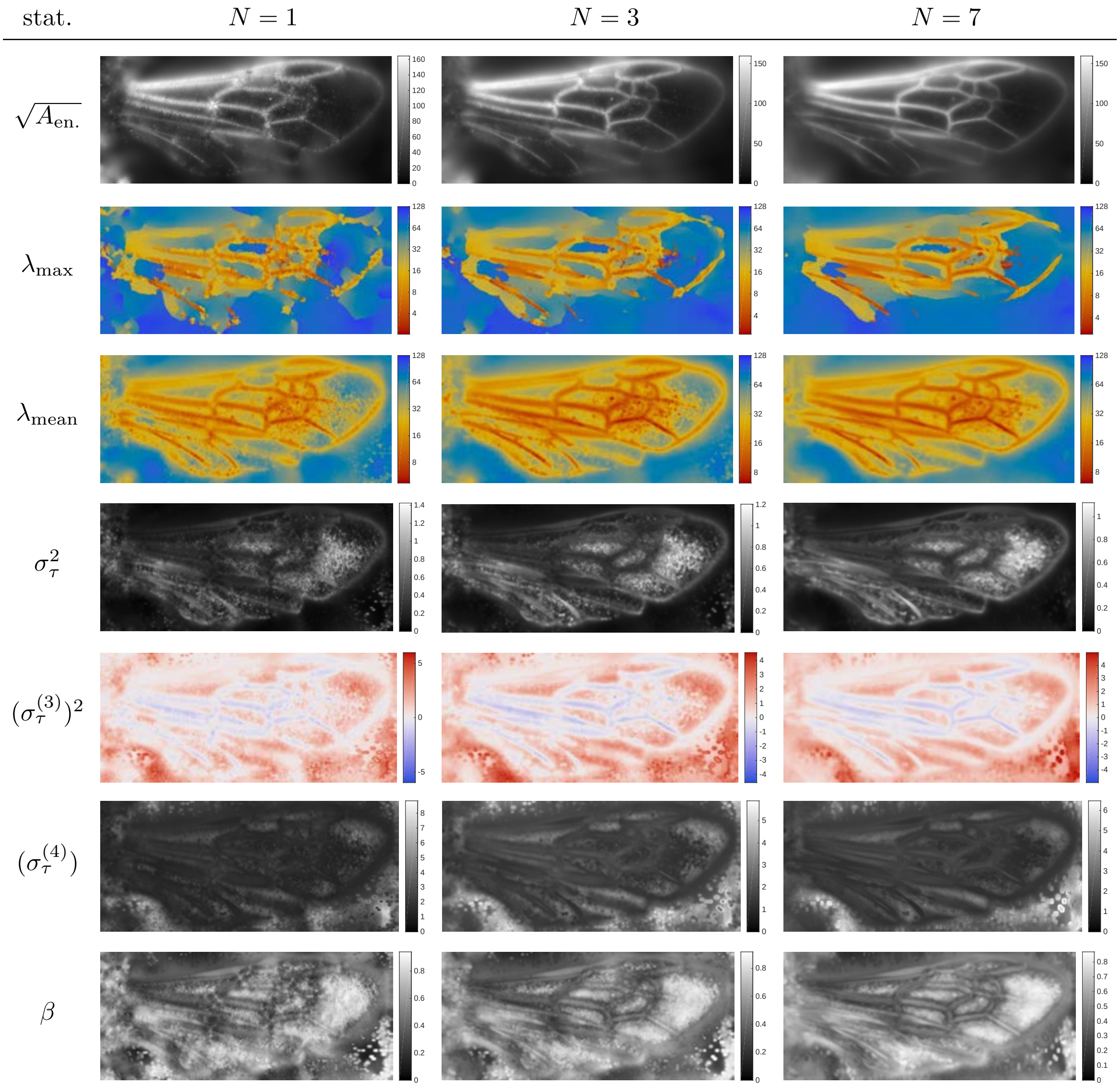
(d) Blob feature

Figure 6.8: CH vector energy scale response for three points on the *Beewing2* image. The SCH matrix was calculated using $M = 11$, $N = 7$, $\omega_c = 2\pi/64$ and window width $2\pi/r_w$. Scale maximum (red circle), mean scale (blue cross) and one standard deviation either side of the mean (blue plus) are also shown.

6.4.2 Scale Selection

When analysing an image using a particular image model we may also wish to identify the best scale. As when solving for orientation, the scale that gives the maximum response is an obvious choice. However, there is an added complexity in finding the solution, as the maximum response now depends on both the model wavelet *orientation* as well as the scale. Therefore we must search over the entire bivariate trigonometric polynomial, as represented by the SCH matrix, in both θ

(a) *Beewing2* image



(b) *Beewing2* statistics

Figure 6.9: Image statistics calculated for the *Beewing2* image from the CH vector magnitude scale response for different values of N. From top to bottom row: Square root of scale energy, scale maximum, mean scale, variance, skewness, square root of kurtosis, bi-modality coefficient. The SCH matrix was calculated using $M = 11$, $\omega_c = 2\pi/64$ and window width $2\pi/r_w$.

and τ. That is, the best scale and orientation are given by

$$\theta, \tau = \underset{\theta,\tau}{\arg\max} \sum_n \sum_m \left(\mathbf{S}_\theta \mathbf{W}\mathbf{u}^H \, \mathbf{W}\mathbf{F}\mathbf{V}\,\mathbf{S}\tau\right)_{n,m}. \tag{6.51}$$

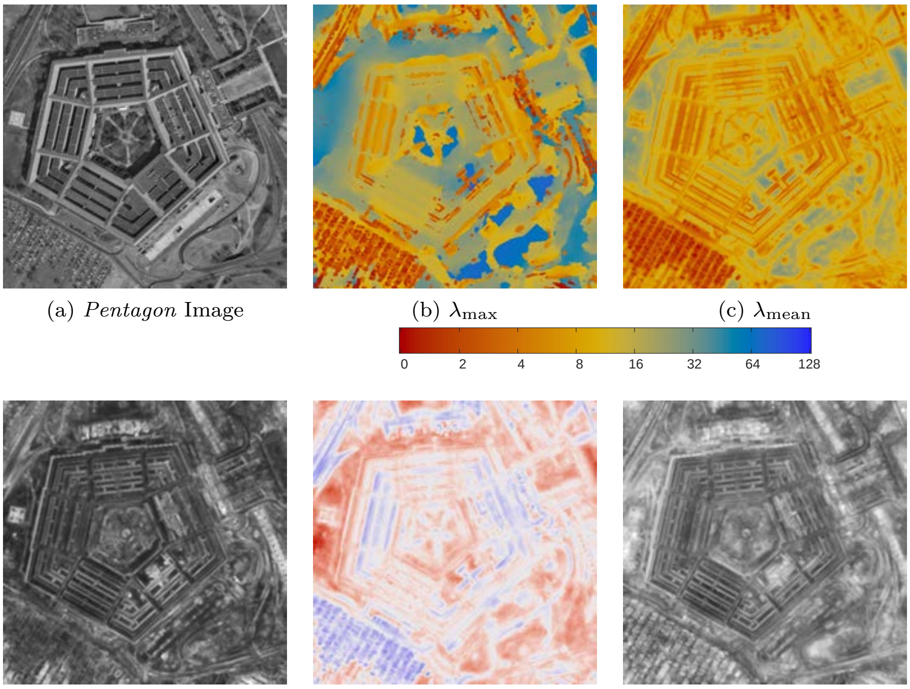
Experiments in Section 2.6.2 showed that solving for the maximum of a normal univariate trigonometric polynomial is time consuming due to the root finding required. For bivariate polynomials more advanced methods must be employed, for example [97]. It is likely that these methods would be even more involved, although it is difficult to determine the increase in time required due to the lack of available MATLAB implementations. Furthermore, in the case of models with multiple components, the scale that gives the maximum response to a single model component may not necessarily be the best choice. For example, if the response at another scale could be modelled as *multiple components*, but with individually smaller amplitudes, that may be more desirable as the model would be more descriptive and have a lower residual.

CH Vector Energy

Instead of searching for maximum in both scale and orientation, we shall instead use the statistics of CH vector energy scale response, $A^2(\tau)$, to select the best scale. The typical approach is to use the scale at which the measure is a maximum, which is $\tau_{\max}$ when using the CH vector energy. However, the scale maximum, and thus the energy, can be discontinuous in many locations, such as for the *Pentagon* image in Figure 6.10b. This is due to the bi-modality of the scale response (Figure 6.10f) combined with the non-linearity of picking a maximum.

Mean Scale

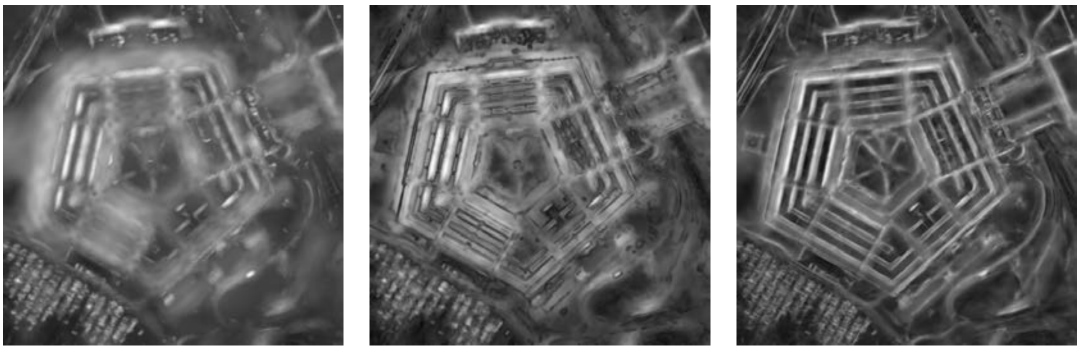
An alternative therefore is the mean scale, τ_{mean}. Since the operations to derive the mean scale are all linear, it gives a continuous estimate and is faster to compute compared to root finding for the scale maximum. However, a comparison of the CH vector energy calculated at the mean scale (Figure 6.11b) and the scale maximum (Figure 6.11a) for the *Pentagon* image, shows that at locations on the edge of the pentagon the energy is very low when using the mean scale, whereas one would expect a high magnitude due to the edge feature present. These locations have a high bi-modality coefficient (Figure 6.10f), and thus the low response is explained by the mean scale actually lying between two or more peaks of the scale response.



(a) *Pentagon* Image (b) λ_{max} (c) λ_{mean}

(d) σ_τ^2 (e) $(\sigma_\tau^{(3)})^2$ (f) β

Figure 6.10: Image statistics calculated for the *Pentagon* image using the CH vector magnitude scale response and $N = 7$. The SCH matrix was calculated using $M = 11$, $\omega_c = 2\pi/64$ and window width $2\pi/r_w$. λ_{max} is the wavelength corresponding to τ_{max}, λ_{mean} is the wavelength corresponding to τ_{mean}.



(a) $A(\tau_{\text{max}})$ (b) $A(\tau_{\text{mean}})$ (c) $\int_{\tau_{\text{mean}}-\sigma}^{\tau_{\text{mean}}+\sigma} A_n(\tau)/2\sigma$

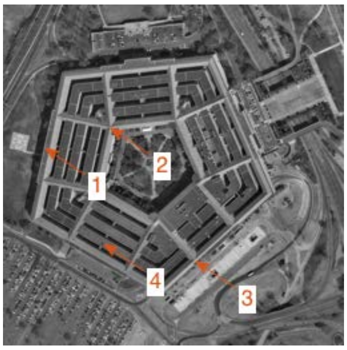
Figure 6.11: CH vector magnitude evaluated at the scale maximum (a), the mean scale (b), and using mean scale averaging (c), for $N = 7$. The SCH matrix was calculated using $M = 11$, $\omega_c = 2\pi/64$ and window width $2\pi/r_w$.

Mean Scale Averaging

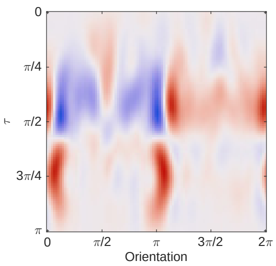
The existence of a single best scale is implicit in the idea of scale selection. However, this is not necessarily the case. To give an example, consider the SCH matrix polynomial of the response to a line-segment wavelet, along with the CH vector magnitude response, obtained at the centres of three junction features in the *Pentagon* image shown in Figure 6.12. Analysis of each response shows that the interpretation is scale dependent:

- The first point is located at the edge of the pentagon (Figure 6.12a) where the mean scale gives a low magnitude response (Figure 6.11). We see that the distribution of energy is bi-modal (Figure 6.12c), and the mean scale does indeed lie between two peaks where the magnitude is lower. The scale maximum is at the larger scale peak, where $\tau \approx 0.5\pi$. The two blue peaks in the SCH polynomial (Figure 6.12b) at this scale indicate the CH vector evaluated at this point would be modelled as two negative line-segments. However, there is another peak at $\tau = 0.75\pi$. Evaluated at this point we would end up with two *positive* line-segments in the model.
- At point 2 (Figure 6.12a) the scale maximum and mean scale coincide. The feature in the image appears as four positively valued line-segments radiating from a point. The SCH polynomial (Figure 6.12d) shows that the line-segment response remains coherent over a wide range of scales at the orientations of the feature components. However, at the scale maximum (and mean scale) there is a *negative* response at an orientation of approximately $\pi/8$ which is larger in magnitude than the positive response at $\pi/2$. In this case, a single scale does not take into account the larger spread of energy of the feature components that indicates that positive line-segments are more semantically correct.
- At point 3 (Figure 6.12a) the scale maximum is at $\tau = 0.75\pi$; however, the majority of the energy is below this scale (Figure 6.12g). The corresponding junction in Figure 6.12a resembles a T junction consisting of two thin line segments and one thick segment. The SCH polynomial shows that there are three peaks in the line segment response corresponding to each of these components, however one peaks at approximately $\tau = 0.5\pi$ while the other two are at the scale maximum. Evaluating the CH vector are either of these scales would likely deliver the wrong model.
- At point 4 (Figure 6.12a) the majority of the energy is concentrated at the scale maximum is at $\tau = 0.55\pi$ (Figure 6.12i). Of all the example features this point is best represented by the response at a single scale.

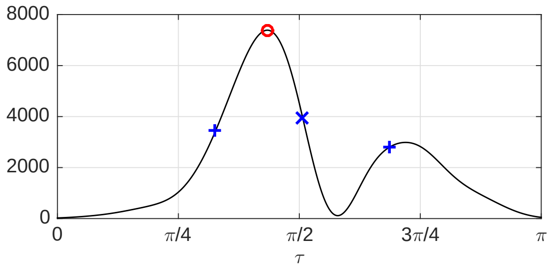
It is clear that a single scale does not always adequately represent the underlying features present, and therefore may not deliver the expected parametrisation when feature components have different sizes. Furthermore, a single scale does not take into account the coherence of energy



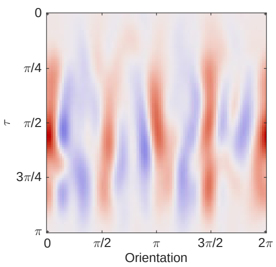
(a) *Pentagon* Image



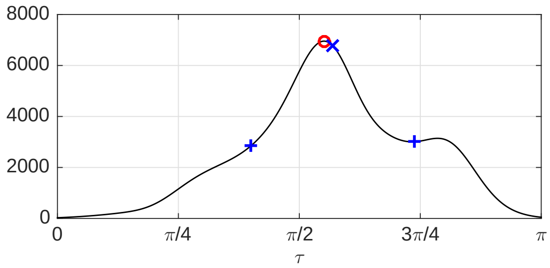
(b) SCH matrix 1



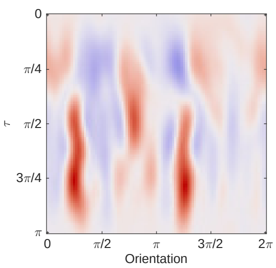
(c) CH energy 1



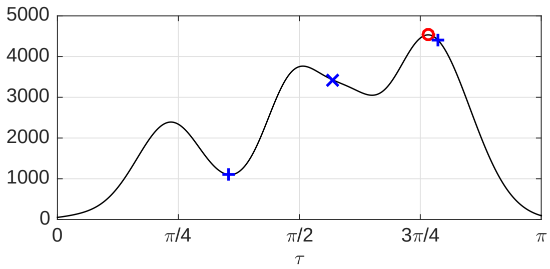
(d) SCH matrix 2



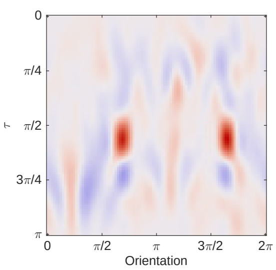
(e) CH energy 2



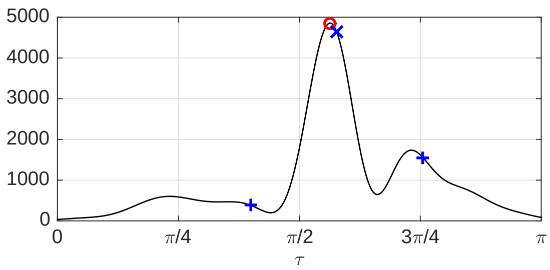
(f) SCH matrix 3



(g) CH energy 3



(h) SCH matrix 3



(i) CH energy 3

Figure 6.12: Visual representation of the bivariate polynomial given by SCH matrix response to a line-segment feature at points 1 (b), 2 (d), 3 (f) and 4 (h) in the *Pentagon*, along with the CH vector energy scale response at points 1 (c), 2 (e), 3 (g) and 4 (i). In the polynomial representation, red represents a positive line-segment response and blue is negative response. In the CH vector energy polynomial, a red circle is scale maximum, a blue cross is mean scale, and the blue plus indicates one standard deviation either side of the mean scale.

with orientation through scale that we see in both the *Pentagon* (Figure 6.12) and *Beewing* (Figure 6.6) SCH polynomials. This similarity over scale indicates a good feature.

As an alternative to evaluating the CH vector at the scale maximum or mean scale, it is proposed to instead *average* the CH vector over a *range* of scales. The statistics given by the CH magnitude can be used to choose the averaging interval. An analytically derived approach is to centre the interval at the mean scale and set its width based on the variance. A large variance indicates a more spread-out response, and thus that the averaging interval should be wide.

This process shall be referred to as *mean scale averaging*, where each RT order of the new average CH vector is given by the integral of its scale response over a certain interval,

$$\bar{f}_n = \int_{\tau_{\text{mean}}-c\sigma}^{\tau_{\text{mean}}+c\sigma} \frac{f_n(\tau)}{2c\sigma} \tag{6.52}$$

where σ is the standard deviation from the energy statistics and c is a constant. Using this approach with $c = 1$, the magnitude of the CH vector is now high at the edges of the *Pentagon* image (Figure 6.11c) compared to magnitude at the mean scale (Figure 6.11b). Qualitatively, the energy also appears to larger at the locations of image features than the energy given by the scale maximum (Figure 6.11a). However, some parts of the *Pentagon* are reduced in magnitude. Investigation of these locations finds that the scale response is narrow (Figure 6.12i) or that the CH vector changes shape.

6.4.3 Model Estimation

After the CH vector has been constructed from the response at either the scale maximum, mean scale, or using mean scale averaging, it can be used for image analysis as usual. An example of the single sinusoidal model of the *Pentagon* image calculated using each method is shown in Figure 6.13. Again, we notice that the amplitude and residual norm are discontinuous for the scale maximum case, while the mean scale has areas of low amplitude due to bi-modal distributions, also resulting in poor estimates of orientation in these locations. The scale maximum appears to have the smoothest orientation estimate, followed by the mean scale averaging. In particular, we notice that mean scale averaging picks up the cross-bars on the pentagon roof along with their orientation, and that the residual norm for this approach appears to correspond best to the locations of i2D features such as the corners of the pentagon and the cross-bar junctions on the roof.

These approaches show that the scale that gives the maximum response is not necessarily the best scale for modelling image features. In the case of the sinusoidal model, mean scale averaging appears to give model parameters that are semantically more similar to the features as they appear in the image. Another advantage is that mean scale averaging can be calculated analytically.

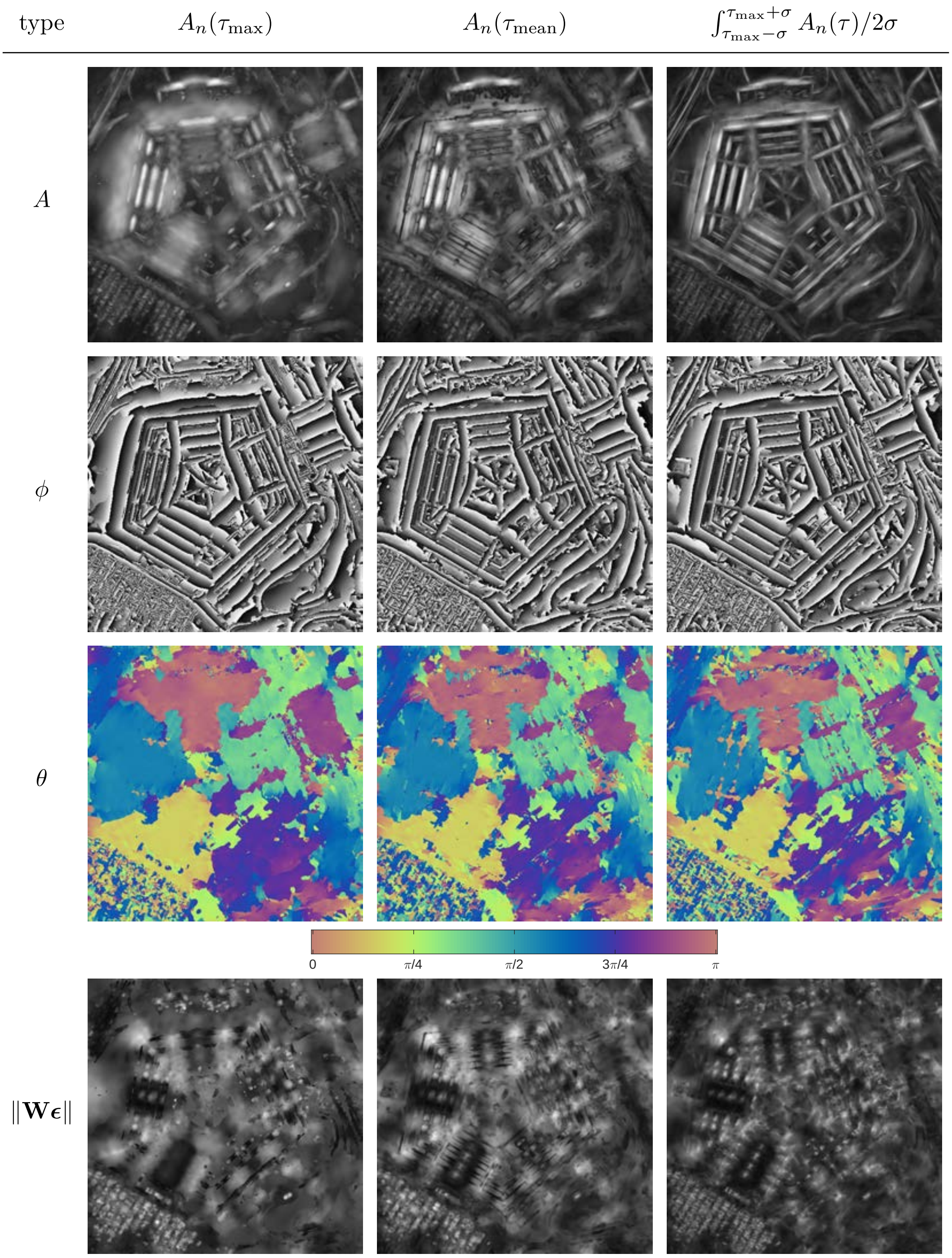


Figure 6.13: Amplitude, A, phase, ϕ, orientation, θ, and residual norm, $\|\mathbf{W}\boldsymbol{\epsilon}\|$, of the sinusoidal model of the *Pentagon* image, calculated at the scale maximum, mean scale, and using mean scale averaging, for $N = 7$. The SCH matrix was calculated using $M = 11$, $\omega_c = 2\pi/64$ and window width $2\pi/r_w$.

6.5 Orientation Statistics

Development of the concepts regarding statistics of the scale response led to a further idea not covered in the previous chapters: Can we derive useful statistics of the *angular* response of a wavelet, as given by the correlation with the CH vector. For example, the mean orientation could be a useful measure in place of the orientation maximum that is used to solve for model parameters. In this section, the circular moments are investigated for this purpose.

6.5.1 Circular Moments

Consider an image model consisting of a set of wavelets at a single orientation. This is the general model used where there is only one orientation parameter, such as for the single sinusoidal model. It is given by

$$f(\mathbf{z}) = \sum_m \lambda_m u_m(\mathbf{R}_\theta \mathbf{z}) + f_\epsilon(\mathbf{z}) \quad \stackrel{CH}{\longleftrightarrow} \quad \mathbf{Wf} = \sum_m \lambda_m \mathbf{S}_\theta \mathbf{W} \mathbf{u}_m + \mathbf{W}\boldsymbol{\epsilon}, \tag{6.53}$$

where $u_m(\mathbf{z})$ are the individual model components and $\mathbf{Wu}_m$ are their CH vectors. The angular response is given by the degree $4N$ polynomial (Section 2.3)

$$p(\theta) = \sum_m \delta_m(\theta)\lambda_m(\theta) \tag{6.54}$$

from which we normally choose the θ that gives the maximum of $p(\theta)$.

The polynomial $p(\theta)$ represents the energy of the response to the wavelet set as it is rotated through 2π radians. The orientation maximum is one statistic that we have used throughout this thesis, however the statistics used for the scale distribution could similarly be applied to orientation. The orientation energy, analogous to A_{energy} for the scale response, is given by the zeroth moment,

$$\theta_{\text{av}} = m_0 \tag{6.55}$$

$$= \int_0^{2\pi} p(\theta)\, d\theta \tag{6.56}$$

$$= 2\pi c_0, \tag{6.57}$$

where c_n is the n-th coefficient of $p(\theta)$. Note, there are $2N + 1$ coefficients $\{c_{-n}, ..., c_n\}$. However, calculating the mean orientation using the standard formula,

$$\theta_{\text{mean}} = \int_0^{2\pi} \theta \, \frac{p(\theta)}{\theta_{\text{av}}}\, d\theta \tag{6.58}$$

for an angular response centred at 0 radians would give $\theta_{\text{mean}} = \pi$ which is not the desired result. Instead, circular statistics [33] must be employed. The circular mean of the angular response polynomial is given by the first trigonometric moment,

$$m_1 = \int_0^{2\pi} p(\theta) e^{-\mathrm{i}\theta} \, d\theta \tag{6.59}$$

$$= c_1, \tag{6.60}$$

from which $\theta_{\text{mean}} = \arg(c_1)$. The normalised circular variance is then obtained as the ratio of the amplitudes of the zeroth moment and the first moment,

$$\sigma_\theta^2 = 1 - \frac{|m_1|}{m_0}. \tag{6.61}$$

Thus the circular moments give information about the angular response distribution that can be found analytically.

6.5.2 Application to the Sinusoidal Model

The single sinusoidal model is a good candidate for the using the mean orientation, as it consists of a one component at a single orientation. The first two moments of the angular response polynomial (in 2θ) of the sinusoidal wavelet pair were compared to the sinusoidal model parameters for the second scale of the *Pentagon* image (Figure 6.14). We notice that:

- The square root of the zeroth circular moment (Figure 6.14a) is approximately proportional to the CH vector magnitude (Figure 6.14d).
- The square root of the absolute value of the first circular moment (Figure 6.14b) is approximately proportional to the sinusoidal model amplitude (Figure 6.14e).
- The circular variance is low at the location of i1D features (Figures 6.14c) similar to the normalised residual magnitude (Figure 6.14e).
- The mean orientation (Figure 6.14g) is approximately equal to the sinusoidal model orientation (Figure 6.14h) at the location of i1D features.

The similarity in orientation between the moments and model makes sense when we consider that i1D features have a single main orientation with a unimodal angular distribution, therefore the location of the orientation maximum will be similar to the centre of the distribution (the mean orientation). The circular variance confirms this observation, as the difference between the two orientation estimates (Figure 6.14i) is only large where the variance is large. The circular variance may be a good measure for the half-sinusoidal model as well, as it will be large where there are multiple component responses.

We now have some proxy measures for the sinusoidal model amplitude and orientation. If we are mainly interested in i1D locations, these measurements give reasonable approximations that are much faster to compute, since the operations are linear and thus root finding is not required.

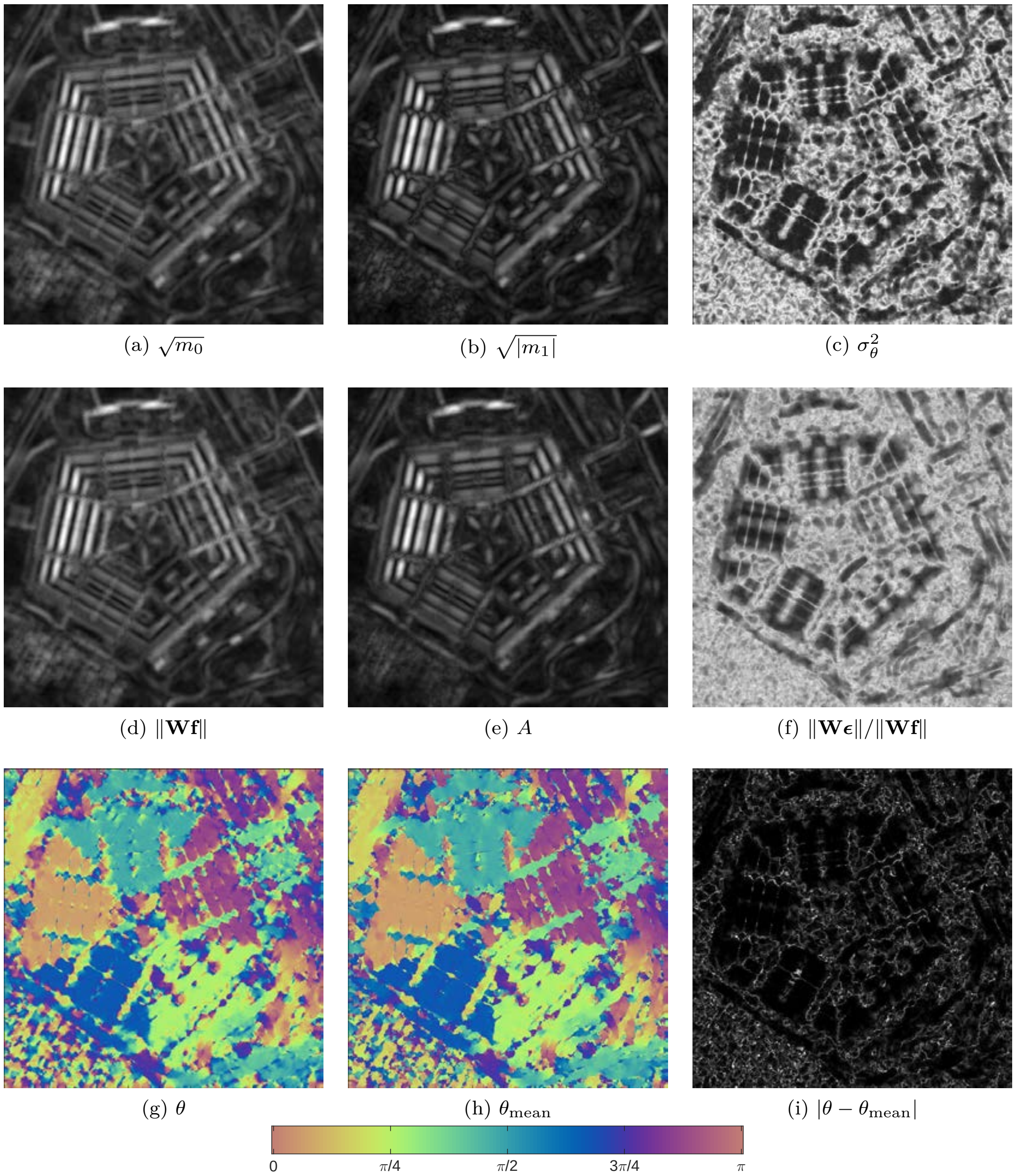


Figure 6.14: Circular moments of the second scale of the *Pentagon* image compared to the parameters of the sinusoidal model for $N = 7$. From top left: square root of 0th circular moment (a), square root of absolute value of 1st circular moment (b), circular variance (c), CH vector norm (d), sinusoidal model amplitude (e), normalised residual norm (f), sinusoidal orientation (g), mean orientation (h), and different between each orientation estimate (i). The SCH matrix was calculated using $M = 11$, $\omega_c = 2\pi/64$ and window width $2\pi/r_w$.

However, using the iterative approach to solve the multi-sinusoidal model, or any model with multiple components at different orientations, relies on accurately locating each component in turn. Therefore using the moment approximations is not feasible for models with components at multiple orientations. The circular variance is similar to the normalised residual magnitude (Figure 6.14f) and thus can be used to discriminate i1D and i2D locations. It provides an alternative to calculating the sinusoidal model parameters for a measure of intrinsic dimension.

Another interesting observation is that the mean orientation is continuous, owing to its linear construction. This can be seen along the lines joining each of the five roof sections of the pentagon (Figure 6.14h). In contrast, as we move across this boundary the change in the sinusoidal model orientation (Figure 6.14g) is abrupt where the adjacent roof becomes the stronger feature. The smoothness of the mean orientation may be of benefit if one is using measures such as the orientation gradient and requires a bound on the energy of the measure.

Note, that the moments for the sinusoidal model are calculated on $p(2\theta)$, that is, using the double angle polynomial, because the normal polynomial has only non-zero even orders (Section 3.1.3). If for example, a 3rd order rotationally symmetric Y junction model was used then it would be calculated on $p(3\theta)$, or $p(4\theta)$ for a symmetric X junction. If this is not performed, the first moment will be zero. Circular moments give a fast approximation of single component model parameters.

6.5.3 Scale Response

The linear construction of the circular moments means that we may also derive their scale response from the CH vector scale response. This provides two new measures that can be used for scale selection. The zeroth moment can be used in place of the CH vector energy, and the first moment can be used in place of a single model component amplitude, which does not have a polynomial scale response expression.

The scale response for the zeroth moment is given by multiplying the individual order CH polynomials as

$$p_0(\tau) = \sum_m \sum_{|n|\leq N} (w_n u_{m,n} w_n f_n(\tau)) \times (w_{-n} u_{m,-n} w_{-n} f_{-n}(\tau)) \tag{6.62}$$

which results in a degree 4N polynomial. A complex-valued polynomial representing the q-th moment is given by

$$p_q(\tau) = \sum_m \sum_{n=-N+q}^{N} (w_n u_{m,n} w_n f_n(\tau)) \times (w_{N-q} u_{m,N-q} w_{N-q} f_{N-q}(\tau)) \tag{6.63}$$

for which the squared magnitude, which is a real-valued polynomial with degree 8N, is given by

$$|p_q(\tau)|_{\mathrm{i}}^2 = \Re_i(p_q(\tau))^2 + \Im_i(p_q(\tau))^2. \tag{6.64}$$

Again, because the sinusoidal model angular response is only over the half circle, to calculate the first circular moment for its response one would use $q = 2$ in the above equations. Likewise, for an n-th order rotationally symmetric model one would use $q = n$.

In regards to the sinusoidal model, the absolute value of the first circular moment is large at the location of i1D features. Thus we can use its scale response to select the best scale for the analysis of lines and edges. The mean scale and scale energy were calculated using the traditional CH vector magnitude, the zeroth moment, and the magnitude of the first moment squared (Figure 6.15). The scale energy and the mean scale are different for the first moment due to the selectivity for i1D features, while there is little difference for the zeroth moment and the CH vector magnitude.

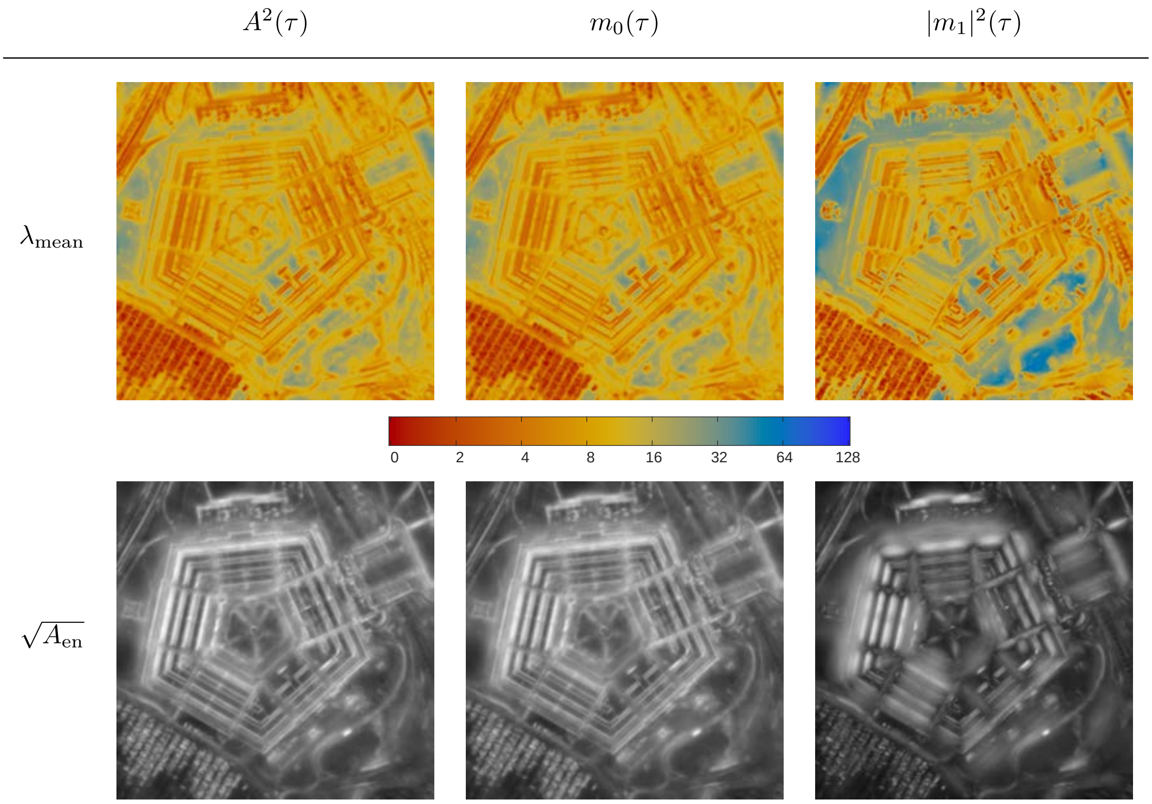


Figure 6.15: Mean scale (top row) and square root of the scale energy, (bottom row) calculated using the CH vector energy, $A^2(\tau)$, zeroth circular moment $m_0(\tau)$, and absolute value of the first circular moment squared, $|m_1|^2(\tau)$, scale response polynomials. The SCH matrix was calculated using $M = 11$, $\omega_c = 2\pi/64$ and window width $2\pi/r_w$.

6.6 Vector Congruency

The SCH matrix polynomials shown for various junctions in Figures 6.6 and 6.12 each demonstrated a similar response to a line-segment wavelet over multiple scales. For example, the polynomial at a Y junction in the *Beewing* image (point 2, Figure 6.6), repeated below in Figure 6.16, shows a similar response from a wavelength of 4 pixels to a wavelength of 64 pixels. These examples suggest the *shape* of the local image structure, as described by the CH vector, is similar at multiple scales at the location of good features.

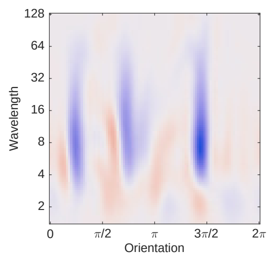


Figure 6.16: Visual representation of bivariate polynomial for SCH matrix response to a line-segment wavelet, at point 2 on the *Beewing* image in Figure 6.6.

Phase congruency is a method of feature detection that relies on a similar concept. It detects features by their similarity in phase through scale, as phase is an illumination invariant measure of symmetry, and thus shape. This section develops a new measure similar to phase congruency that uses the CH vector as the shape descriptor.

6.6.1 Phase Congruency

The analytic signal describes local 1D signal structure using a sinusoidal model, where the amplitude (energy) is high at the presence of features [83, 84, 100] and the phase can be used to classify their type [128]. Morrone and Owens proposed that the local energy is high at the locations of features because the Fourier components are all in phase, and thus constructively interfere [84]. The notion of the phase of multiple sinusoids being coherent is called *phase congruency*. For example, the Fourier series basis sinusoids at the edges of a square wave or the peaks of a triangle wave all have the same phase (Figure 6.17). These locations therefore have high phase congruency.

However, detection using local amplitude varies with illumination and contrast. Kovesi developed an alternative method of detecting features using phase congruency from quadrature filters [58]. Instead of using local energy or Fourier series components, it was found that at the location of features the phase of a sinusoidal model remained relatively constant over multiple scales. Based on this observation, a phase congruency measure was developed.

Given the analytic signal complex exponential representation, $f_a(x) = A(x)e^{\mathrm{i}\phi(x)}$, calculated using a set of quadrature filters at different scales, the *illumination and contrast invariant* measure

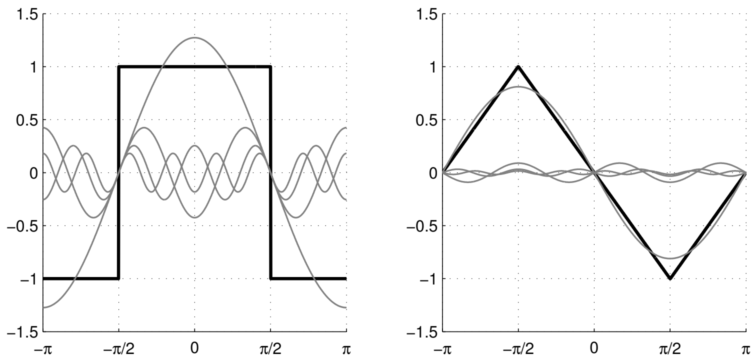


Figure 6.17: First four Fourier series components (grey lines) of a square and triangle wave. Based on Figure 1 in [58].

of phase congruency is given by [58]

$$PC(x) = \frac{\| \sum_i f_{a_i}(x) \|}{\sum_i \|f_{a_i}(x)\|} \tag{6.65}$$

$$= \frac{\| \sum_i A_i(x) e^{\mathrm{i}\phi_i(x)} \|}{\sum_i A_i(x)}, \tag{6.66}$$

where i is the scale index. The measure is the ratio of the amplitude of the sum of analytic signal exponentials divided by the sum of the amplitudes. If the phase is similar for all scales then these values will be equal and the phase congruency measure will be 1. A value of 0 is the lowest possible value and indicates all the exponentials completely cancel out. An example is shown in Figure 6.18. If the individual responses (solid lines) have the same phase (angles) the sum of their individual lengths would match the length of their sum and give a phase congruency score of 1. If instead the phases were different, the sum of their individual lengths would be more than the length of their sum, and the phase congruency score would be less than 1.

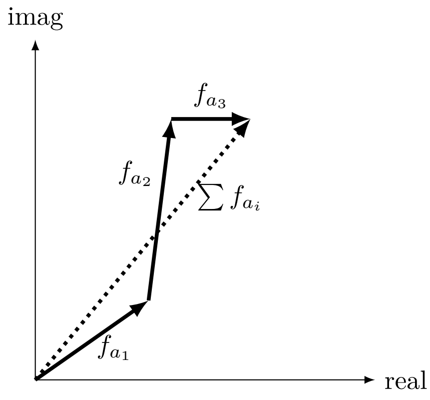


Figure 6.18: Three analytic signal vectors and their sum. The phase of each representation is the angle of the arrow and the amplitude is the length. When all the phases are equal, the length of the individual vectors and the length of their sum are the same.

The phase congruency concept was extended to 2D by using quadrature filters at multiple orientations [58]. In the 2D case, there are multiple phase values each corresponding to a particular orientation. So instead of a complex exponential, a *vector* consisting of each component of the quadrature filter pair responses was used. That is, for eight discrete orientations the vector would have 16 components [57]. Letting $\mathbf{\Phi}(x)$ represent this vector, the phase congruency measure is similar to the 1D case,

$$PC(x) = \frac{\| \sum_i \mathbf{\Phi}_i(x) \|}{\sum_i \| \mathbf{\Phi}_i(x) \|}, \tag{6.67}$$

and the final version of the measure proposed in [58] is

$$PC(x) = \frac{W(\{\mathbf{\Phi}_i(x)\}) \lfloor \| \sum_i \mathbf{\Phi}_i(x) \| - T \rfloor}{\sum_i \| \mathbf{\Phi}_i(x) \| + \epsilon}, \tag{6.68}$$

where $\{\Phi_i\}$ is a set of phase vectors at different scales, W penalises the measure if the amplitude of the phase vectors is not spread across scales, T is a noise threshold and ϵ is a small value to prevent numerical calculation errors.

2D phase congruency using quadrature filters at discrete scales and orientations has been shown to detect image features in an illumination and phase invariant manner [58]. It can be made specific to detecting corners and junctions using the second circular moments of the discrete orientation response [59]. A subsequent approach used the 2D Hilbert transform of the local image patch [130] to generate the phase vectors. Felsberg [30] used the phase of the monogenic signal to detect lines and edges and Zang [142] used the phase values given by the monogenic curvature signal to detect i2D features. In this section a new measure of phase congruency is developed that uses the CH vector as the phase vector in (6.67).

6.6.2 Discrete CH Vector Phase Congruency

The overall concept behind phase congruency is that the *shape* of a feature remains relatively constant over a wide range of scales. The CH vector is a representation of local image structure (shape) and its components are orthogonal. Therefore, it is proposed to substitute the CH vector for the phase vector (6.67) to obtain a new congruency measure of how similar the local image structure is through scale, and thus detect features.

The basic congruency measure using the CH vector is thus

$$PC = \frac{P_{\text{total}}}{P_{\text{sum}}} \tag{6.69}$$

where

$$P_{\text{total}} = \left\| \sum_i \mathbf{W}\mathbf{f}_i \right\| \tag{6.70}$$

$$P_{\text{sum}} = \sum_i \| \mathbf{W}\mathbf{f}_i \|. \tag{6.71}$$

Phase congruency is typically calculated from the response to one or more quadrature filters at discrete scales. Phase congruency using the CH vector can similarly be obtained using the RTs of a discrete filter bank or a wavelet decomposition. However, we may also use the SCH matrix by simply evaluating the CH vector at a set of discrete scales given by $\{\tau_i\}$. The amplitude of the sum of CH vectors is then

$$P_{\text{total}} = \left\| \sum_i \mathbf{Wf}(\tau_i) \right\| \tag{6.72}$$

$$= \left(\sum_{|n|\leq N} \left| \sum_i f_n(\tau_i) \right|^2 \right)^{1/2}, \tag{6.73}$$

and the sum of amplitudes is

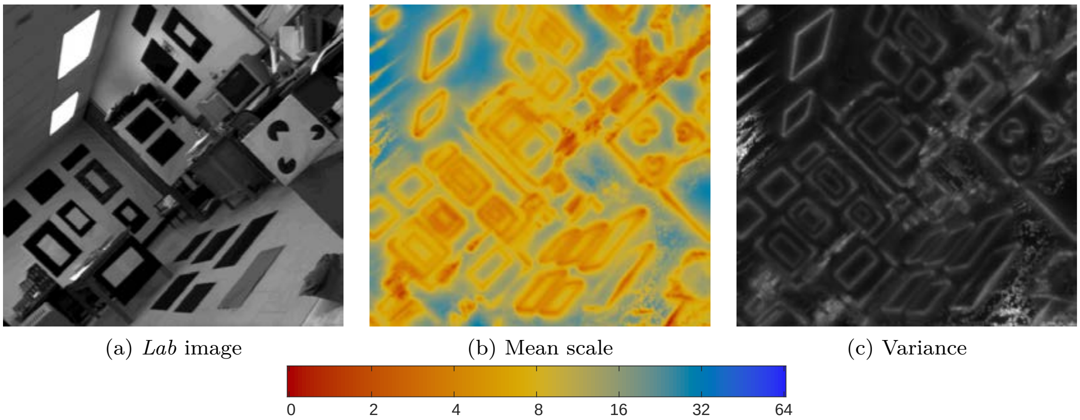
$$P_{\text{sum}} = \sum_i \|\mathbf{Wf}(\tau_i)\| \tag{6.74}$$

$$= \sum_i \left(\sum_{|n|\leq N} |f_n(\tau_i)|^2 \right)^{1/2}, \tag{6.75}$$

where $\mathbf{Wf}(\tau)$ is obtained from the SCH matrix using (6.41). An advantage of using the SCH matrix is that the CH vectors at any number of discrete scales can be calculated without additional filtering operations.

Discrete phase congruency using the SCH matrix was calculated for the *Lab* image (Figure 6.19a) for $M = 11$, $\omega_c = 2\pi/32$ and a filter window width of $2\pi/r_w$. The results are shown in Figure 6.20 for either 5, 10 or 20 evenly spaced discrete scales from the frequency interval $[w_c, \pi]$ and a maximum RT order (N) of either 1,3 or 7. In the result images, the brightness of a pixel is proportional to its congruency score, while the colour indicates the value of P_{total}, which is the sum of CH vector magnitudes and thus a rough measure of the strength of the underlying image structure. This representation will be used for the congruency images in the rest of this section. The *Lab* image was chosen as test image because it contains many features of different strengths and orientations and varying illumination across the image. We can observe that the mean scale is high (small feature size) and the variance is high (energy spread across scales) at the locations of features, in Figures 6.19b and 6.19c respectively.

Using the CH vector gives high phase congruency scores at the location of image features such as lines and edges. Furthermore, the score is high for both strong (yellow) and weak (blue) image structures. This illustrates the illumination and contrast invariant nature of the measure. Increasing N improves the results in some locations. For example, for $N = 1$ the score at each of the centre corner features in the three "pacman" shapes (Figure 6.20) is low, while for $N = 3$ it is high. Increasing N appears to smear the response due to the larger size of the wavelets, but also reduces the amount of noise, as does increasing the number of scales. In this image, most of the noise is due to quantisation of the pixel values from compression.



(a) *Lab* image (b) Mean scale (c) Variance

Figure 6.19: Mean scale (b) and variance (c) of the CH vector energy scale response for the *Lab* test image (a). The SCH matrix was calculated using $M = 11$, $N = 3$, $\omega_c = 2\pi/32$ and window width $2\pi/r_w$.

6.6.3 Continuous CH Vector Congruency

The SCH matrix provides a *continuous* representation of the CH vector response over scale. Therefore, rather than evaluating phase congruency at many discrete points, we can *integrate* the CH vector polynomial over a scale interval. In this case, the equations are

$$P_{\text{total}} = \left\| \int_a^b \mathbf{Wf}(\tau) \right\| \tag{6.76}$$

$$= \left(\sum_{|n| \le N} \left| \int_a^b w_n f_n(\tau) \right|^2 \right)^{1/2} \tag{6.77}$$

and

$$P_{\text{sum}} = \int_a^b \|\mathbf{Wf}(\tau)\| \tag{6.78}$$

$$= \int_a^b \sqrt{A^2(\tau)} \tag{6.79}$$

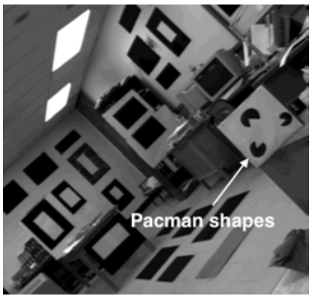
where a and b give the range of scale to integrate. Choosing $a = 0$ and $b = \pi$ corresponds to the frequency spectrum from $\omega_c/2$ to $\sqrt{2}\pi$.

The value of P_{total} can be calculated analytically, as it is possible to find an expression for the integral of a trigonometric polynomial. For example,

$$\int_a^b p(\theta) = \int_a^b \sum_{|n| \le N} c_n e^{in\theta} \tag{6.80}$$

$$= \left[c_0\theta + \sum_{|n| \le N \setminus 0} \frac{c_n}{in} e^{in\theta} \right]_a^b \tag{6.81}$$

$$= c_0(b-a) \sum_{|n| \le N \setminus 0} \frac{c_n}{in} (e^{inb} - e^{ina}). \tag{6.82}$$



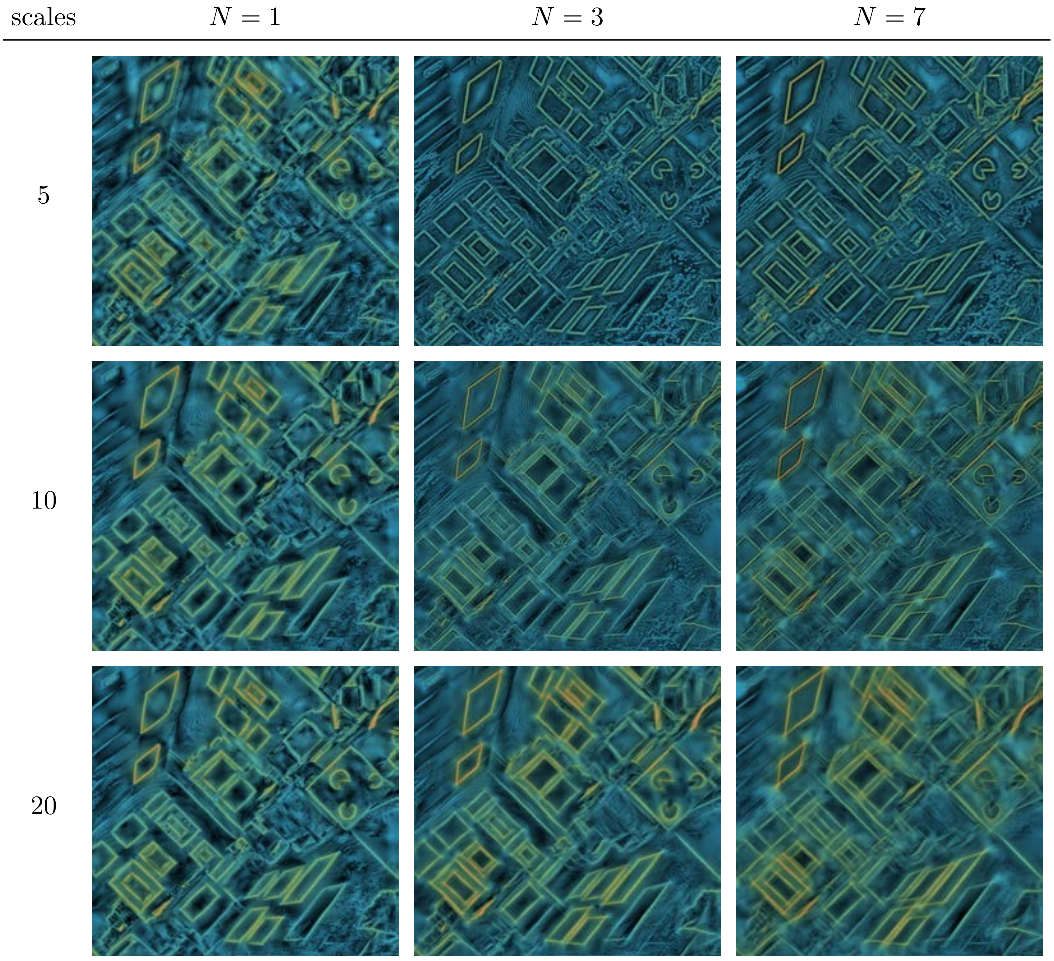


Figure 6.20: Discrete phase congruency calculated from the CH vector scale response for different numbers of scales and N. The scales were equally space to cover the interval $[0, \pi]$ and the SCH matrix was calculated using $M = 11$, $\omega_c = 2\pi/32$ and window width $2\pi/r_w$.

Note that if the integration limits are simply 0 and 2π the exponential terms cancel out and the equation simplifies to

$$\int_0^{2\pi} p(\theta) = 2\pi c_0. \tag{6.83}$$

However, finding the value of P_{sum} involves calculating the square root of $A^2(\tau)$, and thus cannot be found analytically unless $A^2(\tau)$ can be expressed as the square of a polynomial, which is unlikely as it consists of the sum of other polynomials. A new measure of *continuous* phase congruency is therefore proposed, which is able to be computed analytically from the SCH matrix. The new measure shall be referred to as *continuous vector congruency* and is given by

$$VC = \frac{V_{\text{total}}}{V_{\text{sum}}}. \tag{6.84}$$

The expression for V_{total} is similar to that of P_{total} (6.77), except that the integral is *divided by the integration range* before squaring,

$$V_{\text{total}} = \left\| \frac{1}{b-a} \int_a^b \mathbf{Wf}(\tau) \right\| \tag{6.85}$$

$$= \left(\sum_{|n| \le N} \left| \frac{1}{b-a} \int_a^b w_n f_n(\tau) \right|^2 \right)^{1/2}. \tag{6.86}$$

Similarly, the expression for V_{sum} differs from P_{sum} (6.79) in that the square of the CH vector magnitude polynomial is used, as this has an analytic expression,

$$V_{\text{sum}} = \left(\frac{1}{b-a} \int_a^b \|\mathbf{Wf}(\tau)\|^2 \right)^{1/2} \tag{6.87}$$

$$= \left(\frac{1}{b-a} \int_a^b A^2(\tau) \right)^{1/2}. \tag{6.88}$$

Thus vector congruency can be calculated *analytically* from the SCH matrix. A discrete version of is also possible and is given by

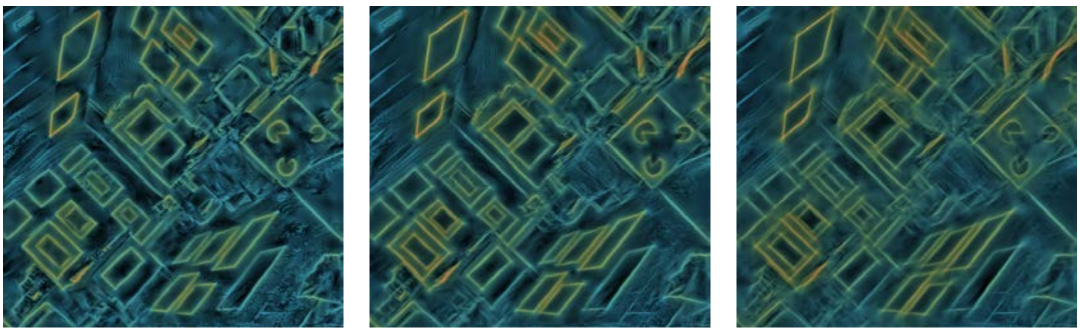
$$V_{\text{total}} = \left(\sum_{|n| \le N} \left| \frac{1}{I} \sum_{i=1}^{I} w_n f_n(\tau_i) \right|^2 \right)^{1/2} \tag{6.89}$$

$$V_{\text{sum}} = \left(\frac{1}{I} \sum_{i=1}^{I} A^2(\tau_i) \right)^{1/2}. \tag{6.90}$$

Vector congruency differs subtly from phase congruency. For a CH vector scale response where the normalised vector $\mathbf{Wf}(\tau)/\|\mathbf{Wf}(\tau)\|$ is constant and only the magnitude, $\|\mathbf{Wf}(\tau)\|$, changes, phase congruency will give the maximum value of 1. However, vector congruency will only give a result of 1 if both the vector *and* its magnitude are constant. To illustrate, consider a simple 1D descriptor evaluated at three scales that gives the values $\{1,1,0\}$. Phase congruency gives $P_{\text{total}} = 2$, $P_{\text{sum}} = 2$ and thus $PC = 1$. Whereas vector congruency gives $V_{\text{total}} = 2/3$, $V_{\text{sum}} = \sqrt{2/3}$ and thus a lower value of $VC = \sqrt{2/3} \approx 0.82$.

This difference is not necessarily bad. The simple formula for phase congruency in (6.69) does not account for the spread of energy across the scale spectrum. Indeed if the amplitudes are zero for all except one of the discrete scales, the phase congruency measure will still be 1. In practical terms this means a single large response at one scale is enough to give a high phase congruency score. This was the reason Kovesi [58] included a spread weighting, W, in (6.68), to penalise the score if the energy was not spread across scale. A similar penalty is thus built in to the vector congruency measure by way of its construction, while still remaining illumination invariant.

Continuous vector congruency using the SCH matrix was calculated for the *Lab* image for $M = 11$, $\omega_c = 2\pi/32$, a filter window width of $2\pi/r_w$, and over the interval $\tau \in [0,\pi]$. The results are shown in Figure 6.21 for a maximum RT order (N) of either 1,3 or 7. As with phase congruency, vector congruency gives high scores at the location of image features, such as lines and edges, regardless of the strength of the local image structure. Thus it has the same illumination and contrast invariant nature as phase congruency. Likewise, increasing N appears to smear the response but also reduces the amount of noise, however the effect is less pronounced. The main difference between the measures is that vector congruency has a lower score in the flat regions of the image. These regions correspond to areas of low variance in Figure 6.19, meaning that the energy is not spread across scale. This highlights the aforementioned difference where vector congruency inherently penalises a lack of spread of energy.



(a) $N = 1$ (b) $N = 3$ (c) $N = 7$

Figure 6.21: Continuous vector congruency calculated from the CH vector scale response for different values of N. The integration limits were $[0,\pi]$ and the SCH matrix was calculated using $M = 11$, $\omega_c = 2\pi/32$ and window width $2\pi/r_w$.

6.6.4 Interval, Noise and Weighting

When implementing Kovesi's final version of the phase congruency measure [58], shown in (6.68), four aspects must be decided upon,

- The phase vector that is used to describe the local image structure.

- The scales at which the phase vector is calculated.
- A noise threshold that attenuates the score in regions where the magnitude of the phase vector is low.
- A weighting function to penalise the score if there is insufficient energy spread across scales.

In the proposed CH vector congruency, the phase vector is the CH vector given by the SCH matrix, and we must decide on the number of SH basis wavelets (M), RT orders (N), cut-off frequency (ω_c), and filter window width. Instead of the number of discrete scales, the integration limits must be chosen. Furthermore, we can expand the basic vector congruency equation to include terms for noise and energy spread weighting to

$$VC = \frac{W(\mathbf{Wf}(\tau))\lfloor V_{\text{total}} - T \rfloor}{V_{\text{sum}} + \epsilon} \tag{6.91}$$

where $\lfloor \cdot \rfloor$ is a function that sets negative values to zero.

Regarding the SCH matrix, the first choice to make is the filter width. Using the energy approximation method a filter of approximately one octave is given by a window width of $2\pi/r_w$ (Figure 6.3). The cut-off frequency sets the minimum frequency (maximum wavelength) that is used in the calculations, and determines the value of r_w (6.10). It should be set low enough to include the desired range of feature scales. Values of $2\pi/32$ and $2\pi/64$ have been used in the examples, for images ranging from 256×256 to 512×512 pixels. The number of SH orders (M) should be set high enough to reduce the oscillations outside the filter window to an acceptable level (Figure 6.3). The number of RT orders must be large enough to adequately discriminate the interesting features in the image but not too large to cause smearing of the response. A value between $N = 3$ and $N = 7$ is recommended. Using $N = 1$ (the monogenic signal) showed artefacts at the location of i2D features (Figure 6.7).

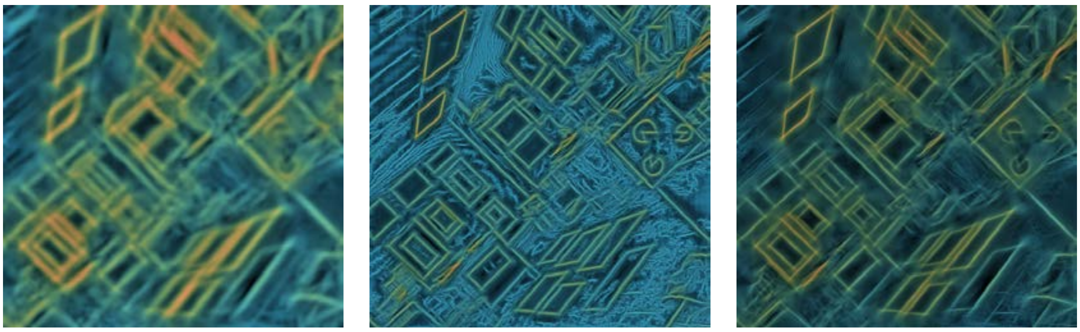
In the remainder of this section the effects of integration limits, noise thresholding and energy spread weighting will be demonstrated.

Integration Limits

The range of scales included in the vector congruency measure is determined by the integration limits a and b. Choosing $a = 0$ and $b = \pi$ includes all the scales from the lower frequency cutoff, $\omega_c/2$, to the maximum frequency, $\sqrt{2}\pi$. For a maximum vector congruency of 1 the vector must be the same across all the scales. Therefore if lower frequencies (longer wavelengths) are included, the local area for which the congruency is calculated increases, and the response will tend to towards larger features. Increasing a excludes longer wavelengths, as does increasing the cutoff frequency, ω_c. Likewise, reducing b removes high frequencies from the calculation, and thus decreases the response of small features. Using $a = \pi/r_w$ starts the integration from $\omega = \omega_c$, where the high pass magnitude first reaches 1, and choosing $b = -\pi \log_2(\omega_c)/r_w$ sets the upper limit of integration to $\omega = \pi$ which is wholly contained within the spectrum. Choosing these values excludes the part of the scale response which tapers.

Removing high frequencies ($[a, b] = [0, \pi/2]$) suppresses noise but results in a less localised congruency response (Figure 6.22a). Removing low frequencies ($[a, b] = [\pi/2, \pi]$) increases the score for smaller features and noisy locations but is more localised (Figure 6.22b). Changing the limits from the valid frequency range, $[a, b] = [0, \pi]$, to the entire range, $[a, b] = [0, 2\pi]$, appears to give little change in the result (Figure 6.22c). The integral of the squared polynomial over this interval is simply 2π times the 0th order. Since finding the entire square of a trigonometric polynomial requires $(n + 1)^2$ multiplications, but finding only the 0th order requires only $2n + 1$, using integration limits of $[a, b] = [0, 2\pi]$ will speed up the congruency calculation.

The previous example (Figure 6.21) showed that increasing N reduced noise but caused a smearing of the response. In this example (Figure 6.22), also with a larger N, we see that by restricting the integration range to the higher frequencies the smear is much reduced and the score is more consistent at the expense of increased noise. Thus one may restrict vector congruency to the higher frequencies in order to account for the increased size of the wavelets when using larger N.



(a) Lower frequencies, $[0, \pi/2]$ (b) Higher frequencies, $[\pi/2, \pi]$ (c) Entire range, $[0, 2\pi]$

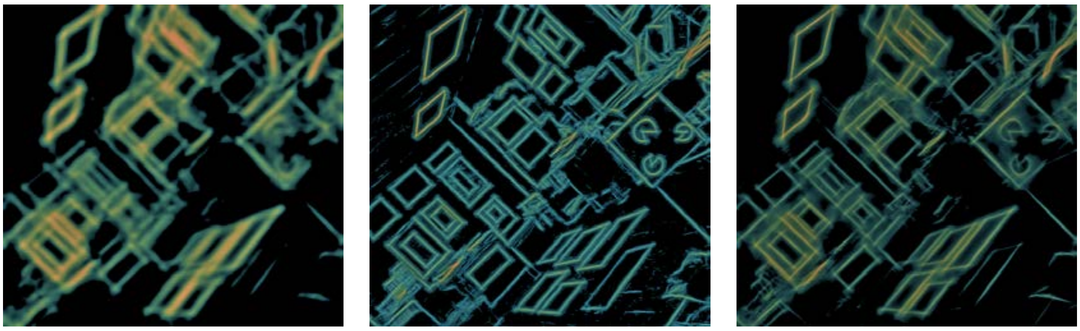
Figure 6.22: Continuous vector congruency calculated from the CH vector scale response for $N = 7$ and different integration limits (shown in brackets). The SCH matrix was calculated using $M = 11$, $\omega_c = 2\pi/32$ and window width $2\pi/r_w$.

Noise Thresholding

The noise seen for the high frequency interval (Figure 6.22) is due to low strength image structures (blue colour). Therefore it can be removed by increasing the noise threshold, T, in (6.91) above zero. Setting the threshold to the median value of V_{total} gives good results for the high frequency range (Figure 6.23).

Spread Weighting

Even though the vector congruency has an inherent penalty for non-uniform spread, we may wish to further penalise features with a small energy spread. Kovesi [58] used a spread measure given



(a) Lower frequencies, $[0, \pi/2]$ (b) Higher frequencies, $[\pi/2, \pi]$ (c) Valid range, $[0, \pi]$

Figure 6.23: Continuous vector congruency calculated from the CH vector scale response for $N = 7$, different integration limits (shown in brackets), and noise threshold set to the median of V_{total}. The SCH matrix was calculated using $M = 11$, $\omega_c = 2\pi/32$ and window width $2\pi/r_w$.

by the average amplitude of each scale divided by the maximum amplitude. A sigmoidal function was used to create the weighting function, W in (6.68), which is then multiplied with the phase congruency score.

A similar approach is possible with vector congruency. We can divide the total energy, A_{energy}, of the CH vector magnitude scale response by the value at the scale maximum and the integration range,

$$s_{\text{en}} = \frac{\frac{1}{b-a}\int_a^b A^2(\tau)\, d\tau}{\|\mathbf{Wf}(\tau_{\max})\|^2} \tag{6.92}$$

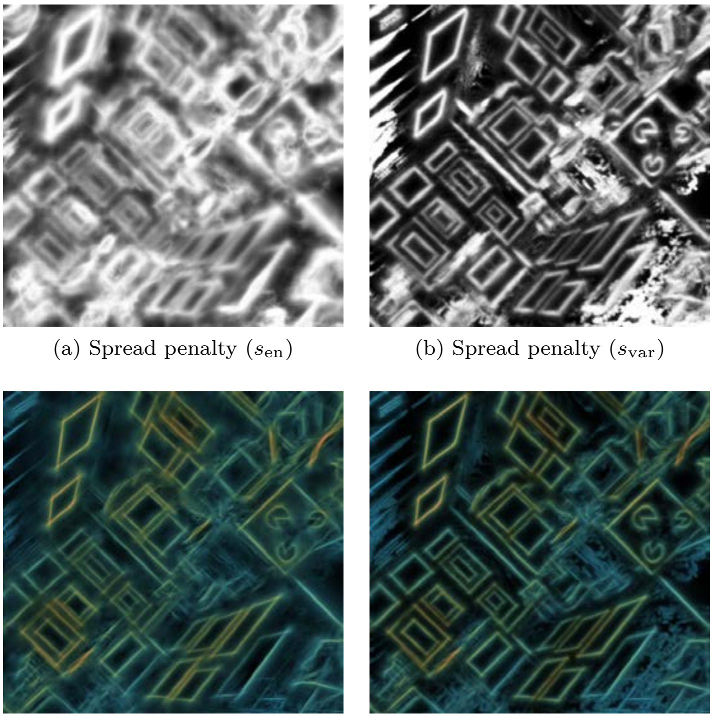
and use this as the input into a sigmoidal function to create the weighing term. Unlike the sigmoidal function in [58], the function from Section 3.3.3 shall be used as it delivers values over the full range from 0 to 1. Alternatively, the variance, skewness and kurtosis measures give information the shape of the energy distribution that could be used. For example, by dividing the variance by its maximum possible value we obtain a value between 0 and 1 that can also be fed into the sigmoidal function to create the weighting term.

$$s_{\text{var}} = \frac{8\sigma_\tau^2}{(b-a)^2} \tag{6.93}$$

Figure 6.24 shows an example of the weighting along with its congruency output using the same SCH parameters as the previous example, and either the energy spread measure using the maximum, s_{en}, or using the variance, s_{var}. The sigmoidal function parameters used were cutoff = 0.45, slope = 4, and cutoff = 0.2, slope = 2, respectively. For each method the response from noise in the flat areas is reduced.

6.6.5 Models

Neither phase congruency nor vector congruency discriminate between types or features. Lines, edges, corners and junctions all give large congruency scores. An extension proposed by Kovesi [59]



(a) Spread penalty (s_{en}) (b) Spread penalty (s_{var})

(c) Vector congruency (s_{en}) (d) Vector congruency (s_{var})

Figure 6.24: Spread weighting calculated from the scale maximum (a) and the normalised variance (b) of the CH vector energy response for $N = 7$. Sigmoidal function cutoff and slope were $[0.45, 4]$ and $[0.2, 2]$, respectively. Continuous vector congruency calculated using integration limits $[0, \pi]$ and multiplied by the spread weighting from the scale maximum (c) and the normalised variance (d). The SCH matrix was calculated using $M = 11$, $\omega_c = 2\pi/32$ and window width $2\pi/r_w$.

used circular moments of the discrete angular response given by a set of orientated quadrature filters to detect i2D features. Another proposed by Zang [142] used the phase of the monogenic curvature signal as the shape descriptor to also detect i2D features. In this section, vector congruency is combined with modelling from the previous chapter, and a new method of detecting i1D and i2D features is developed.

To begin with, consider the numerator of the vector congruency equation,

$$V_{\mathrm{total}} = \left\| \frac{1}{b-a} \int_a^b \mathbf{Wf}(\tau) \right\|, \tag{6.94}$$

$$\tag{6.95}$$

and let

$$\overline{\mathbf{Wf}} = \frac{1}{b-a} \int_a^b \mathbf{Wf}(\tau) \tag{6.96}$$

be the average CH vector over the integration range. This vector can be split in model and residual components, where the model part represents the particular structure of interest. That is,

$$\overline{\mathbf{Wf}} = \overline{\mathbf{Wf}}_{\mathrm{model}} + \overline{\mathbf{W}\epsilon}. \tag{6.97}$$

Vector congruency can then be calculated using either the model component or residual component individually to give a congruency score related to those features.

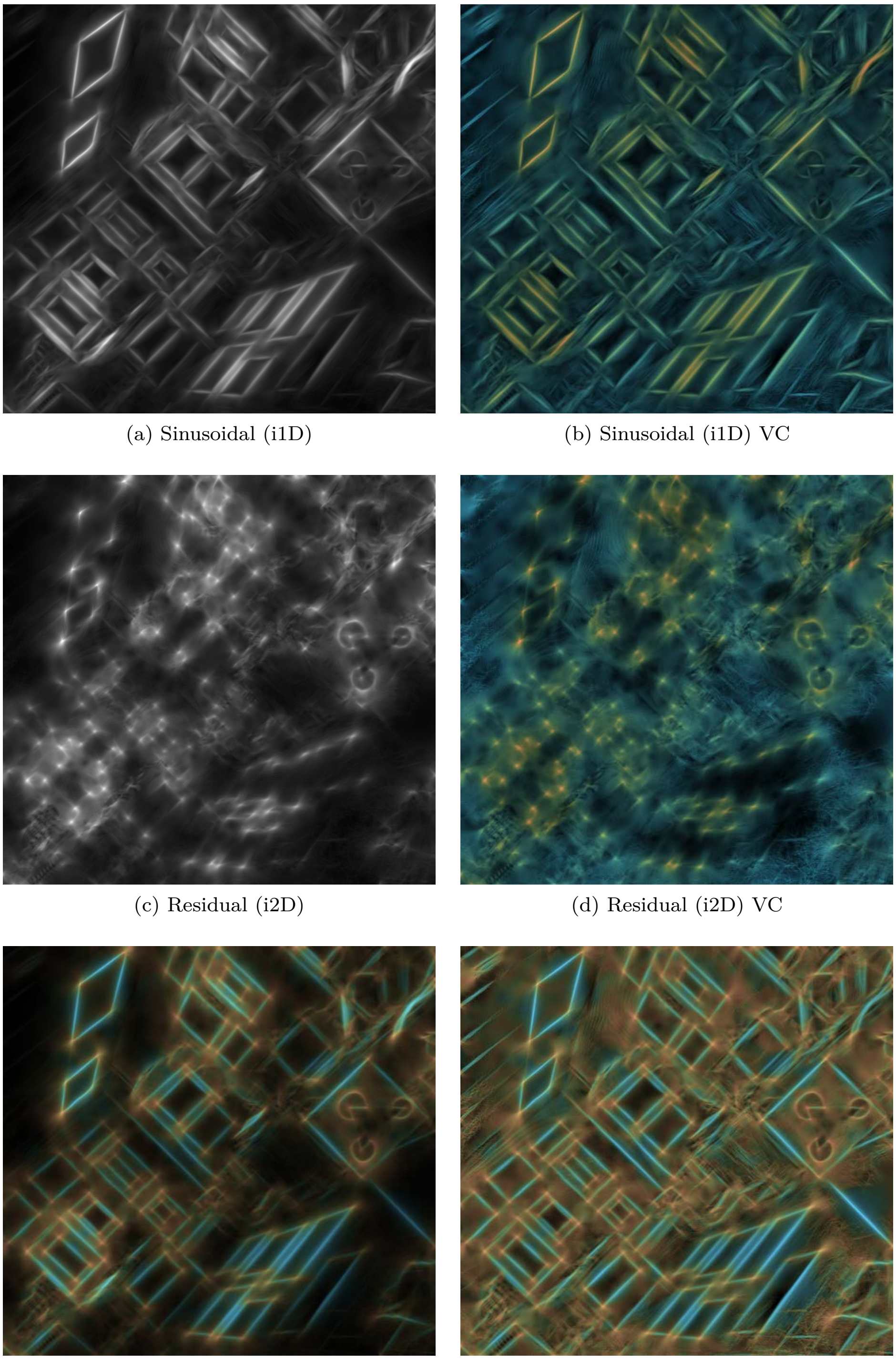
$$VC_{\text{model}} = \|\overline{\mathbf{W}\mathbf{f}}_{\text{model}}\|/V_{\text{sum}}, \tag{6.98}$$

$$VC_{\text{resid}} = \|\overline{\mathbf{W}\epsilon}\|/V_{\text{sum}}. \tag{6.99}$$

Sinusoidal Model

The sinusoidal model and residual vector magnitudes were calculated using the average CH vector, $\overline{\mathbf{W}\mathbf{f}}$, and compared to vector congruency (6.99) using the same vectors (Figure 6.25). The model magnitude obtained from the average CH vector is high at the locations of i1D features, however the value varies depending on the strength of the feature. In contrast, the model vector congruency score is high at the location of i1D features regardless of their amplitude. This is shown in Figure 6.25b where bright blue lines indicate high congruency but low strength. A similar observation can be made for the residual component, where i2D features with weak strength give a similar vector congruency score to those with strong strength (Figure 6.25d).

Combining the model and residual components using either the average CH vector (Figure 6.25e) or vector congruency (Figure 6.25f) gives a measure of intrinsic dimension. As before, the strength of the normal intrinsic dimension measure varies with the strength of the local image structure, whereas the vector congruency version is normalised. This may improve i2D feature detection where illumination varies over an image.



(a) Sinusoidal (i1D) (b) Sinusoidal (i1D) VC

(c) Residual (i2D) (d) Residual (i2D) VC

(e) Intrinsic Dimension (f) Intrinsic Dimension VC

Figure 6.25: Sinusoidal model magnitude (a), residual magnitude (c) and intrinsic dimension (d) calculate using the average CH vector (left column), compared to the same measurements (b,d,f respectively) calculated using vector congruency (right column) with integration limits $[0, \pi]$. The SCH matrix was calculated using $M = 11$, $\omega_c = 2\pi/32$ and window width $2\pi/r_w$.

6.7 Summary

This chapter introduced a new local image descriptor called the SCH matrix, which describes the CH vector over a continuous range of scale as well as orientation. The matrix is given by the responses to a wavelet frame. The frame is constructed from an isotropic scale shiftable wavelet frame, based on the principles of Perona's deformable kernels [99], augmented with its higher-order RTs to create a 2D steerable wavelet frame, according to Unser's conditions in [124].

We are able to derive a polynomial expression for the CH vector energy scale response through linear combination of the matrix components. One may select the scale that gives the maximum energy using this novel measure. Other statistics can be derived analytically, such as the mean scale and variance, to give further insight into the scale distribution of features. From the matrix we are also able to describe the response to a wavelet in both orientation and scale as a bivariate polynomial. Inspired by the mean scale statistics, the 0th and 1st circular moments of the wavelet angular response were proposed as new measures for scale selection. The argument of the first circular moment is a measure of mean orientation, and is a faster alternative to solving for the orientation maximum using root finding. One may also average the CH vector over a variable scale interval, such as either side of the mean scale, as an alternative to limiting analysis to a single scale.

Finally, a new continuous version of phase congruency called vector congruency was introduced. It is constructed by the integration of the SCH matrix polynomial over a particular scale interval. Calculating the numerator of the vector congruency equation is similar to mean scale averaging of the CH vector magnitude, except that the scale interval is fixed. Indeed if we replace the CH vector *magnitude* in the numerator with the response to a wavelet, vector congruency gives an illumination invariant detection measure. This was demonstrated for the sinusoidal model and its residual to give an i1D and i2D detector, but could be modified for any feature. It is important to remember that the vector congruency results do not depend on the strength of the local image structure, only on the coherence of the CH vector through scale. This makes the measure illumination invariant.

Since the CH vector scale polynomials can be added, subtracted, multiplied, integrated and differentiated algebraically, any algorithms that use these operations will also result in a continuous polynomial function. Division (polynomial inverse) and square root operations are not possible except in special cases, so algorithms that use these will need to be evaluated at discrete scales. A continuous representation of filter responses is still of benefit, however, as it allows for evaluation at any number of arbitrary scales without the need for more filtering. By using integration limits of $[0, 2\pi]$, many of the integral operations can be sped up, as integration of a trigonometric polynomial over this range is simply equal to $2\pi c_0$, where c_0 is the 0th order coefficient. Note that this does not mean that only the 0th order SH wavelet is used. The scale response polynomial, $A^2(\tau)$, is given by the *square* of the scale response polynomial for each order RT, and therefore all SH orders contribute to the value of the coefficient.

One idea that is implicit in the previous chapters of this thesis and other wavelet based detection and orientation estimation methods, for example [55, 86, 104, 142], is that the *components* of junction features are of a similar scale, and thus there is a best scale at which to analyse them. However, feature energy is spread over a range of scales and feature components can have their maximum response at different scales, as many of the examples in this chapter showed. This implies that a single scale may not be sufficient to capture all the information necessary for analysis. The SCH matrix therefore is a useful descriptor for analysing these types of features, and applying it to these problems is the subject of future work.

Chapter 7

Conclusion

7.1 Contributions

This thesis has developed a complete framework for local image analysis using higher-order RTs. Many different tasks that would normally be performed using different methods, such as local image representation, interest point detection, feature parametrisation and phase-independent energy measurement, can be achieved starting with the CH vector.

There are several novel concepts which underpin this approach:

- Collecting higher-order RTs responses into the CH vector are a crucial aspect of the method. The CH vector generalises previous 2D analytic signals and gives information about the symmetries of the local image structure up to a certain order. The normalised CH vector is an illumination invariant description of shape, while the magnitude is a measure of the local energy. Rather than start with a set of wavelets specific to a feature or problem, the CH vector is a general descriptor that provides information on a wide range of common image structures.
- Models are solved by splitting the CH vector into model and residual components and choosing the parameters that minimise the residual magnitude. This allows us to use arbitrary wavelets that are suited to the image analysis task at hand, rather than designed for the tight frame property. It also allows for weighting to be treated separately to wavelet shape.
- The inclusion of a residual component enables the novel iterative method of solving for model parameters. In previous approaches the orientations of model components were found from the local maxima in the angular response (the roots method in previous chapters). The iterative method accounts for the side lobe oscillations of the wavelets, and thus allows more closely oriented feature component to be resolved. The iterative method also allows for the wavelets to be modified between iterations. For example, one can use both sinusoidal and half-sinusoidal components in the one representation.

Using the CH vector method, three models were developed that can describe common image features.

- The first model was the sinusoidal model used by the monogenic signal and other 2D quadrature approaches. Whereas previous 2D analytic signal were limited in the number of RT orders used, the CH vector approach meant that higher-order RTs could be used to both get estimates from even structures and to tune the wavelets towards larger i1D structures. The benefit of the residual component was again demonstrated with a new method of junction and corner detection that is a general extension of the previous approaches of the boundary tensor and monogenic curvature tensor.

- The sinusoidal model was extended to a multi-sinusoidal model. This model allows for multiple orientation estimation up to any order by adding more higher-order RTs. One particular application is the analysis of images with multiple additive i1D patterns, such as coral core x-rays. The wavelet basis is useful for this model, as one can separate an image into different orientation classes by reconstructing from different model components.

- A half-sinusoidal model was also developed to analyse line-segment and edge-segment junctions in the one model. This model is more general than the previous two as it is able to represent both i1D and i2D features equally well. Furthermore, wavelets matched to specific features can be created by considering the CH vector given by the half-sinusoidal model with set parameters.

Thus we have a primary descriptor of image structure, a method of detecting points of interest, and three models that describe common image features, all derived from the 2D analytic signal that is the CH vector. The models have the illumination and rotation invariances of the monogenic signal and previous approaches, and with appropriate weighting the CH vector magnitude is also phase invariant. It was found that increasing the number of RT orders increased the size of the wavelets and thus their orientation selectivity. However, it also increased the size of the image patch under consideration. It was also found that one needs at least the 4th order RT to distinguish two crossed lines from an isotropic blob.

One of the most useful results is that one can calculate the sinusoidal model analytically from both odd and even structures using the 0th to 2nd order RTs and a quartic solver. Many applications currently use the monogenic signal as a quick method of calculating local orientation. Simply adding the 2nd order RT should improve these methods will little increase in processing time.

In the last chapter the CH vector was extended to the SCH matrix, which describes the CH vector over scale. The scale of image analysis is an important parameter that is often picked through trial and error or using some previous knowledge given by a human operator. From the SCH matrix a continuous representation of the CH vector magnitude and other measures such as circular harmonics can be derived. This allows for automatic scale selection in the same CH vector framework that will be used for detection and modelling. In contrast, other methods will perform scale selection using measures, such as the determinant of Hessian, that may not be related to the operators used for subsequent analysis.

The framework allows the description of local image structure using models with parameters that are invariant or equi-variant with illumination, rotation and scale.

7.2 Future Work

The sinusoidal model and multi-sinusoidal model are specific to line and edge features, while the half-sinusoidal model is a more general representation. Further work would be to investigate models and their corresponding wavelets which are more specific to certain feature types, such those suggested for chequer patterns (Figure 5.29) or sinusoids at fixed angles (Figure 4.28). Thanks to the residual vector and the iterative method, images could then be analysed using a dictionary of different models, or combination thereof.

A second research direction is the use of the CH vector as a texture descriptor. Other descriptors that encode low order rotational symmetry have proved successful for texture classification [20], character recognition [92] and object recognition [42]. Preliminary work has shown that k-means clustering of the normalised CH vector, rotated using the sinusoidal model orientation, automatically classifies regions in an image by shape. The challenge is to learn the feature CH vectors that are most discriminative.

One of the more interesting observations from this research is that features occur over a range of scales, and that feature *components* can have different scales. It suggests that a single scale may not provide enough information to adequately describe a feature. To compensate one may include information from multiple scales, such as in the last chapter where we averaged the CH vector around the mean scale. Another approach could be to use an image model that includes a scale parameter for the model components, such as

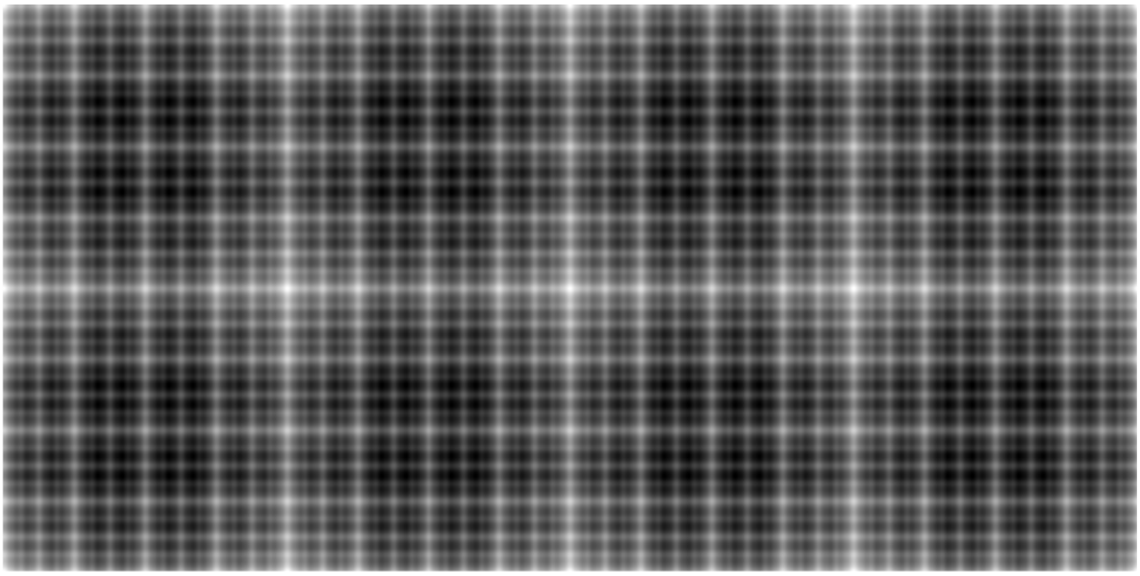
$$f(\mathbf{z}) = \sum_{k=1}^{K} \lambda_k u_k(\mathbf{R}_\theta(2^{i_k}\mathbf{z})) + f_\epsilon(\mathbf{z}), \tag{7.1}$$

where i_k is the dilation (scaling) factor for the k-th model component $u_k(\mathbf{z})$. Estimating the model components then involves estimating both the orientation and scale parameters from the SCH matrix. To estimate the components, a quick approximation method like from Section 2.6.2, would need to be developed for a bivariate polynomial, as would an iterative method that takes into account the spread of energy of a feature component.

7.3 Code Repository

It is hoped the CH vector framework finds wide use within the image analysis community. To help this process a MATLAB repository of functions and examples is available at:

`https://github.com/geometrikal/ch-vector`



Bibliography

[1] T. Aach, C. Mota, I. Stuke, M. Muhlich, and E. Barth, "Analysis of Superimposed Oriented Patterns," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3690–3700, 2006.

[2] H Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[3] R. Ali, M. Gooding, M. Christlieb, and M. Brady, "Advanced phase-based segmentation of multiple cells from brightfield microscopy images," in *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, IEEE, 2008, pp. 181–184.

[4] M. Awrangjeb, G. Lu, and C. S. Fraser, "Performance Comparisons of Contour-Based Corner Detectors," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 4167–4179, 2012.

[5] E. Barth, "Image Encoding, Labeling, and Reconstruction from Differential Geometry," *Graphical Models and Image Processing*, vol. 55, no. 6, pp. 428–446, 1993.

[6] E. Bedrosian, "The Analytic Signal Representation of Modulated Waveforms," *Proceedings of the IRE*, vol. 50, no. 10, pp. 2071–2076, 1962.

[7] A. Belaid, D. Boukerroui, Y Maingourd, and J.-F. Lerallut, "Phase-Based Level Set Segmentation of Ultrasound Images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 1, pp. 138–147, 2011.

[8] R. Bennett, "The intrinsic dimensionality of signal collections," *IEEE Transactions on Information Theory*, vol. 15, no. 5, pp. 517–525, 1969.

[9] J. Bigun and G. Granlund, "Optimal orientation detection of linear symmetry," in *Proceedings of the IEEE First International Conference on Computer Vision: London, Great Britain*, vol. 54, 1987, p. 258.

[10] T. Billow and G. Sommer, "Multi-dimensional signal processing using an algebraically extended signal representation," in *Algebraic Frames for the Perception-Action Cycle*, ser. Lecture Notes in Computer Science, G. Sommer and J. Koenderink, Eds., vol. 1315, Springer Berlin / Heidelberg, 1997, pp. 148–163.

[11] D. Boukerroui, J. A. Noble, and M. Brady, "On the Choice of Band-Pass Quadrature Filters," *Journal of Mathematical Imaging and Vision*, vol. 21, no. 1, pp. 53–80, 2004.

[12] T. Bulow and G. Sommer, "Hypercomplex signals-a novel extension of the analytic signal to the multidimensional case," *IEEE Transactions on Signal Processing*, vol. 49, no. 11, pp. 2844–2852, 2001.

[13] T. Bülow and G. Sommer, "A Novel Approach to the 2D Analytic Signal," in *Computer Analysis of Images and Patterns*, Springer, 1999, pp. 25–32.

[14] E. J. Candès and C. Fernandez-Granda, "Super-Resolution from Noisy Data," *Journal of Fourier Analysis and Applications*, vol. 19, no. 6, pp. 1229–1254, 2013. arXiv: `arXiv:1211.0290v2`.

[15] ——, "Towards a Mathematical Theory of Super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014. arXiv: `arXiv:1203.5871v3`.

[16] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 1986.

[17] B. E. Chalker and D. J. Barnes, "Gamma densitometry for the measurement of skeletal density," *Coral Reefs*, vol. 9, no. 1, pp. 11–23, 1990.

[18] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[19] T. F. Cooper, G. De'Ath, K. E. Fabricius, and J. M. Lough, "Declining coral calcification in massive Porites in two nearshore regions of the northern Great Barrier Reef," *Global Change Biology*, vol. 14, no. 3, pp. 529–538, 2008.

[20] M. Crosier and L. D. Griffin, "Using Basic Image Features for Texture Classification," *International Journal of Computer Vision*, vol. 88, no. 3, pp. 447–460, 2010.

[21] I. Daubechies, A. Grossmann, and Y. Meyer, "Painless nonorthogonal expansions," *Journal of Mathematical Physics*, vol. 27, no. 5, p. 1271, 1986.

[22] G. De'ath, J. M. Lough, and K. E. Fabricius, "Declining Coral Calcification on the Great Barrier Reef," *Science*, vol. 323, no. 5910, pp. 116–119, 2009.

[23] T. Ell, "Quaternion-Fourier transforms for analysis of two-dimensional linear time-invariant partial differential systems," in *Proceedings of 32nd IEEE Conference on Decision and Control*, IEEE, 1993, pp. 1830–1841.

[24] J. Fdez-Valdivia, J. Garcia, J. Martinez-Baena, and X. Fdez-Vidal, "The selection of natural scales in 2D images using adaptive Gabor filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 5, pp. 458–469, 1998.

[25] M. Felsberg, "Disparity from Monogenic Phase," in *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, 2002, pp. 248–256.

[26] ——, "Low-Level Image Processing with the Structure Multi-vector," PhD thesis, 2002.

[27] ——, "Optical Flow Estimation from Monogenic Phase," in *Complex Motion*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 1–13.

[28] M. Felsberg, S. Kalkan, and N. Krüger, "Continuous dimensionality characterization of image structures," *Image and Vision Computing*, vol. 27, no. 6, pp. 628–636, 2009.

[29] M. Felsberg and G Sommer, "The monogenic signal," *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 3136–3144, 2001.

[30] M. Felsberg and G. Sommer, "A New Extension of Linear Signal Processing for Estimating Local Properties and Detecting Features," in *22. DAGM Symposium Mustererkennung*, 357, vol. 357, Citeseer, 2000, pp. 195–202.

[31] ——, "Image Features Based on a New Approach to 2D Rotation Invariant Quadrature Filters," in *Computer Vision — ECCV 2002*, vol. 2350, Springer Berlin Heidelberg, 2002, pp. 369–383.

[32] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America A*, vol. 4, no. 12, p. 2379, 1987.

[33] N. Fisher, T. Lewis, and B. Embelton, *Statistical analysis of circular data*. Cambridge: Cambridge University Press, 1996.

[34] P.-E. Forssén and H. Spies, "Multiple Motion Estimation Using Channel Matrices," in *Complex Motion*, 1, vol. 2007, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 54–65.

[35] W. Forstner and E. Gulch, "A fast operator for detection and precise location of distinct points, corners and centres of circular features," in *ISPRS Intercommission Workshop*, 1987.

[36] W. Freeman, "Steerable filters and local analysis of image structure," PhD thesis, 1992.

[37] W. Freeman and E. Adelson, "Steerable filters for early vision, image analysis, and wavelet decomposition," in *[1990] Proceedings Third International Conference on Computer Vision*, IEEE Comput. Soc. Press, 1990, pp. 406–415.

[38] ——, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991.

[39] D. Gabor, "Theory of communication," *Electrical Engineers - Part III: Radio and Communication Engineering, Journal of the Institution of*, vol. 93, no. 26, pp. 429 –457, 1946.

[40] I. Gradshteĭn and I. Ryzhik, *Table of Integrals, Series, and Products*. 2007.

[41] G. Granlund and H. Knutsson, *Signal Processing for Computer Vision*. Kluwer Academic Publishers, 1995.

[42] L. D. Griffin, M. Lillholm, M. Crosier, and J. van Sande, "Basic Image Features (BIFs) Arising from Approximate Symmetry Type," in *Scale Space and Variational Methods in Computer Vision*, ser. Lecture Notes in Computer Science X, vol. 5567, Springer Berlin / Heidelberg, 2009, pp. 343–355.

[43] S Hahn, "The instantaneous complex frequency concept and its application to the analysis of building up oscillations in oscillators," *Proc. Vibration Problems*, vol. 1, pp. 29–46, 1959.

[44] S. Hahn, “Multidimensional complex signals with single-orthant spectra,” *Proceedings of the IEEE*, vol. 80, no. 8, pp. 1287–1300, 1992.

[45] W. Hamilton, “On quaternions, or on a new system of imaginaries in algebra,” *Philosophical Magazine*, vol. 25, no. 3, pp. 489–495, 1844.

[46] C. Harris and M. Stephens, “A Combined Corner and Edge Detector,” in *Proceedings of the Alvey Vision Conference 1988*, vol. 15, Alvey Vision Club, 1988.

[47] S Held, M Storath, P Massopust, and B Forster, “Steerable Wavelet Frames Based on the Riesz Transform,” *IEEE Transactions on Image Processing*, vol. 19, no. 3, pp. 653–667, 2010.

[48] D. Hestenes and G. Sobczyk, *Clifford Algebra to Geometric Calculus*. Reidel, 1987.

[49] D. Hopley, K. E. Parnell, and P. J. Isdale, “The Great Barrier Reef Marine Park: Dimensions and Regional Patterns,” *Australian Geographical Studies*, vol. 27, no. 1, pp. 47–66, 1989.

[50] M. Jacob and M Unser, “Design of steerable filters for feature detection using canny-like criteria,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1007–1019, 2004.

[51] G. Jacovitti and A. Neri, “Multiresolution circular harmonic decomposition,” *IEEE Transactions on Signal Processing*, vol. 48, no. 11, pp. 3242–3247, 2000.

[52] N. Joshi, S. Bond, and M. Brady, “The segmentation of colorectal MRI images,” *Medical Image Analysis*, vol. 14, no. 4, pp. 494–509, 2010.

[53] T. Kadir and M. Brady, “Saliency, Scale and Image Description,” *International Journal of Computer Vision*, vol. 45, no. 2, pp. 83–105, 2001.

[54] H. Knutsson, C.-F. Westin, and M. Andersson, “Structure Tensor Estimation: Introducing Monomial Quadrature Filter Sets,” in *New Developments in the Visualization and Processing of Tensor Fields*, 2012, pp. 3–28.

[55] U. Köthe, “Edge and junction detection with an improved structure tensor,” in *Pattern Recognition*, Springer, 2003, pp. 25–32.

[56] U. Köthe, “Low-level Feature Detection Using the Boundary Tensor,” in *Visualization and Processing of Tensor Fields*, Springer, 2006, pp. 63–79.

[57] P. Kovesi, “Invariant measures of image features from phase information,” PhD thesis, University of Western Australia, 1996.

[58] ——, “Image features from phase congruency,” *Videre: Journal of Computer Vision Research*, vol. 1, no. 3, pp. 1–26, 1999.

[59] ——, “Phase congruency detects corners and edges,” in *The Australian Pattern Recognition Society Conference*, 2003, pp. 309–318.

[60] ——, “Good Colour Maps: How to Design Them,” pp. 1–42, 2015. arXiv: `1509.03700`.

[61] G Krieger and C Zetzsche, “Nonlinear image operators for the evaluation of local intrinsic dimensionality,” *IEEE Transactions on Image Processing*, vol. 5, no. 6, pp. 1026–1042, 1996.

[62] K. G. Larkin, “Natural demodulation of two-dimensional fringe patterns II Stationary phase analysis of the spiral phase quadrature transform,” *Journal of the Optical Society of America A*, vol. 18, no. 8, p. 1871, 2001.

[63] K. G. Larkin, D. J. Bone, and M. A. Oldfield, “Natural demodulation of two-dimensional fringe patterns I General background of the spiral phase quadrature transform,” *Journal of the Optical Society of America A*, vol. 18, no. 8, p. 1862, 2001.

[64] T. Lindeberg, “Feature Detection with Automatic Scale Selection,” *International Journal of Computer Vision*, vol. 30, no. 2, pp. 79–116, 1998.

[65] ——, “Scale Selection Properties of Generalized Scale-Space Interest Point Detectors,” *Journal of Mathematical Imaging and Vision*, vol. 46, no. 2, pp. 177–210, 2013.

[66] ——, “Image Matching Using Generalized Scale-Space Interest Points,” *Journal of Mathematical Imaging and Vision*, vol. 52, no. 1, pp. 3–36, 2015.

[67] J. Lough and D. Barnes, “Intra-annual timing of density band formation of Porites coral from the central Great Barrier Reef,” *Journal of Experimental Marine Biology and Ecology*, vol. 135, no. 1, pp. 35–57, 1990.

[68] P. Loughlin and B. Tacer, “Instantaneous frequency and the conditional mean frequency of a signal,” *Signal Processing*, vol. 60, no. 2, pp. 153–162, 1997.

[69] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[70] R Marchant and P Jackway, “Modelling Line and Edge Features Using Higher-Order Riesz Transforms,” in *Advanced Concepts for Intelligent Vision Systems*, ser. Lecture Notes in Computer Science, vol. 8192, 2013, pp. 438–449.

[71] R. Marchant, “Scale Adaptive Filters,” in *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, vol. 32, IEEE, 2015, pp. 1–8.

[72] R. Marchant and P. Jackway, “Generalised Hilbert Transforms for the Estimation of Growth Direction in Coral Cores,” in *2011 International Conference on Digital Image Computing: Techniques and Applications*, IEEE, 2011, pp. 660–665.

[73] ——, “Feature Detection from the Maximal Response to a Spherical Quadrature Filter Set,” in *2012 International Conference on Digital Image Computing Techniques and Applications (DICTA)*, Perth: IEEE, 2012, pp. 1–8.

[74] ——, “Local feature analysis using a sinusoidal signal model derived from higher-order Riesz transforms,” in *2013 IEEE International Conference on Image Processing*, IEEE, 2013, pp. 3489–3493.

[75] ——, "Using Super-Resolution Methods to Solve a Novel Multi-Sinusoidal Signal Model," in *2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, IEEE, 2013, pp. 1–8.

[76] ——, "A sinusoidal image model derived from the circular harmonic vector," *Journal of Mathematical Imaging and Vision*, 2016.

[77] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 2, IEEE Comput. Soc, 2001, pp. 416–423.

[78] W. McLean and D. Elliott, "On the p-norm of the truncated Hilbert transform," *Bulletin of the Australian Mathematical Society*, vol. 38, no. 03, p. 413, 1988.

[79] M. Mellor and M. Brady, "Phase mutual information as a similarity measure for registration," *Medical Image Analysis*, vol. 9, no. 4, pp. 330–343, 2005.

[80] M. Michaelis and G. Sommer, "Junction classification by multiple orientation detection," in *Computer Vision—ECCV'94*, 1994, pp. 101–108.

[81] K. Mikolajczyk and C. Schmid, "Scale & Affine Invariant Interest Point Detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.

[82] ——, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[83] M. C. Morrone and D. C. Burr, "Feature Detection in Human Vision: A Phase-Dependent Energy Model," *Proceedings of the Royal Society B: Biological Sciences*, vol. 235, no. 1280, pp. 221–245, 1988.

[84] M. Morrone and R. Owens, "Feature detection from local energy," *Pattern Recognition Letters*, vol. 6, no. 5, pp. 303–313, 1987.

[85] C. Mota, I. Stuke, T. Aach, and E. Barth, "Estimation of Multiple Orientations at Corners and Junctions," in *Pattern Recognition*, 2004, pp. 163–170.

[86] M. Muehlich, D Friedrich, and T Aach, "Design and Implementation of Multisteerable Matched Filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 279–291, 2012.

[87] M. Muhlich and T Aach, "Analysis of Multiple Orientations," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1424–1437, 2009.

[88] M Mühlich, T. Dahmen, and T Aach, "Doubly-rotated matched filtering," *15th European Signal Processing Conference (EUSIPCO-2007)*, pp. 2169–2173, 2007.

[89] M. Mühlich and T. Aach, "A Theory of Multiple Orientation Estimation," in *Computer Vision–ECCV 2006*, 2006, pp. 69–82.

[90] ——, "High Accuracy Feature Detection for Camera Calibration: A Multi-steerable Approach," in *Pattern Recognition*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 284–293.

[91] M. Muhlich, T. Dahmen, and T. Aach, "Design of Multi-Steerable Filters and their Application for the Detection of Corners and Junctions," in *2007 IEEE International Conference on Image Processing*, IEEE, 2007.

[92] A. J. Newell and L. D. Griffin, "Natural Image Character Recognition Using Oriented Basic Image Features," in *2011 International Conference on Digital Image Computing: Techniques and Applications*, IEEE, 2011, pp. 191–196.

[93] J. Ng and A. A. Bharath, "Steering in Scale Space to Optimally Detect Image Structures," in *Lecture Notes in Computer Science*, ser. Lecture Notes in Computer Science, vol. 3021, Springer, 2004, pp. 482–494.

[94] S. Olhede and G. Metikas, "The Monogenic Wavelet Transform," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3426–3441, 2009.

[95] P. Pad, V. Uhlmann, and M. Unser, "VOW: Variance-optimal wavelets for the steerable pyramid," in *2014 IEEE International Conference on Image Processing (ICIP)*, vol. 0, IEEE, 2014, pp. 2973–2977.

[96] ——, "Maximally Localized Radial Profiles for Tight Steerable Wavelet Frames," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2275–2287, 2016.

[97] N. M. Patrikalakis and T. Maekawa, *Shape Interrogation for Computer Aided Design and Manufacturing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 1–408.

[98] P. Perona, "Steerable-scalable kernels for edge detection and junction analysis," *Image and Vision Computing*, vol. 10, no. 10, pp. 663–672, 1992.

[99] ——, "Deformable kernels for early vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, pp. 488–499, 1995.

[100] P. Perona and J. Malik, "Detecting and localizing edges composed of steps, peaks and roofs," in *[1990] Proceedings Third International Conference on Computer Vision*, IEEE Comput. Soc. Press, 1990, pp. 52–57.

[101] J Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 49–71, 2000.

[102] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain.," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–51, 2003.

[103] Z. Puspoki, V. Uhlmann, C. Vonesch, and M. Unser, "Design of Steerable Wavelets to Detect Multifold Junctions," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 643–657, 2016.

[104] Z. Puspoki and M. Unser, "Template-Free Wavelet-Based Detection of Local Symmetries," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 3009–3018, 2015.

[105] Z. Puspoki, C. Vonesch, and M. Unser, "Detection of symmetric junctions in biological images using 2-D steerable wavelet transforms," in *2013 IEEE 10th International Symposium on Biomedical Imaging*, IEEE, 2013, pp. 1496–1499.

[106] J. R. Romero, S. K. Alexander, S. Baid, S. Jain, and M. Papadakis, "The geometry and the analytic properties of isotropic multiresolution analysis," *Advances in Computational Mathematics*, vol. 31, no. 1-3, pp. 283–328, 2009.

[107] C. Ronse, "On idempotence and related requirements in edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 5, pp. 484–491, 1993.

[108] E. Rosten, R. Porter, and T. Drummond, "Faster and Better: A Machine Learning Approach to Corner Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 105–119, 2010.

[109] S. Sarkar and K. Boyer, "On optimal infinite impulse response edge detection filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 11, pp. 1154–1171, 1991.

[110] E. Simoncelli and H. Farid, "Steerable wedge filters for local orientation analysis," *IEEE Transactions on Image Processing*, vol. 5, no. 9, pp. 1377–1382, 1996.

[111] E. Simoncelli, W. Freeman, E. Adelson, and D. Heeger, "Shiftable multiscale transforms," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 587–607, 1992.

[112] G. Sommer, L. Wietzke, and D. Zang, "Monogenic Curvature Tensor as Image Model," in *Visualization and Processing of Tensor Fields*, 357, vol. 2, Springer, 2009, pp. 281–301.

[113] R. Soulard, P. Carre, and C. Fernandez-Maloigne, "Vector Extension of Monogenic Wavelets for Geometric Representation of Color Images," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 1070–1083, 2013.

[114] H. Stark, "An extension of the Hilbert transform product theorem," *Proceedings of the IEEE*, vol. 59, no. 9, pp. 1359–1360, 1971.

[115] E. Stein, *Singular Integrals and Differentiability Properties of Functions*. Princeton University Press, 1971.

[116] M. Storath, "The monogenic curvelet transform," in *2010 IEEE International Conference on Image Processing*, IEEE, 2010, pp. 353–356.

[117] ——, "Directional Multiscale Amplitude and Phase Decomposition by the Monogenic Curvelet Transform," *SIAM Journal on Imaging Sciences*, vol. 4, no. 1, pp. 57–78, 2011.

[118] C. Sun and P. Vallotton, "Fast linear feature detection using multiple directional non-maximum suppression," *Journal of Microscopy*, vol. 234, no. 2, pp. 147–157, 2009.

[119] T. Szilagyi, A. K. Harvey, S. M. Brady, L. Cochlin, and N. Joshi, “Quantitative analysis of tendon ECM damage using MRI,” in *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, IEEE, 2010, pp. 1365–1368.

[120] R. Tibshirani, “Regression Shrinkage and Selection via the Lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[121] M Unser, N. Chenouard, and D. Van De Ville, “Steerable Pyramids and Tight Wavelet Frames in <formula formulatype=”inline”><tex Notation=”TeX”>$L_{2}({}^{d})$</tex> </formula>,” *IEEE Transactions on Image Processing*, vol. 20, no. 10, pp. 2705–2721, 2011.

[122] M Unser, D. Sage, and D. Van De Ville, “Multiresolution Monogenic Signal Analysis Using the Riesz-Laplace Wavelet Transform,” *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2402–2418, 2009.

[123] M Unser and D Van De Ville, “Wavelet Steerability and the Higher-Order Riesz Transform,” *IEEE Transactions on Image Processing*, vol. 19, no. 3, pp. 636–652, 2010.

[124] M. Unser and N. Chenouard, “A Unifying Parametric Framework for 2D Steerable Wavelet Transforms,” *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 102–135, 2013.

[125] M. Unser and D. Van De Ville, “Higher-order riesz transforms and steerablewavelet frames,” in *2009 16th IEEE International Conference on Image Processing (ICIP)*, IEEE, 2009, pp. 3801–3804.

[126] D. Vakman, “On the analytic signal, the Teager-Kaiser energy algorithm, and other methods for defining amplitude and frequency,” *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 791–797, 1996.

[127] ——, “Analytic waves,” *International Journal of Theoretical Physics*, vol. 36, no. 1, pp. 227–247, 1997.

[128] S. Venkatesh and R. Owens, “On the classification of image features,” *Pattern Recognition Letters*, vol. 11, no. 5, pp. 339–349, 1990.

[129] C. Wachinger and N. Navab, “Alignment of Viewing-Angle Dependent Ultrasound Images,” in *Medical Image Computing and Computer-Assisted Intervention {MICCAI} 2009*, vol. 12, 2009, pp. 779–786.

[130] K. Wang, P. Xiao, X. Feng, and G. Wu, “Image feature detection from phase congruency based on two-dimensional Hilbert transform,” *Pattern Recognition Letters*, vol. 32, no. 15, pp. 2015–2024, 2011.

[131] J. P. Ward, K. N. Chaudhury, and M. Unser, “Decay Properties of Riesz Transforms and Steerable Wavelets,” *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 984–998, 2013. arXiv: `arXiv:1301.2525v1`.

[132] Weichuan Yu, K Daniilidia, and G. Sommer, “Rotated wedge averaging method for junction characterization,” in *Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Comput. Soc, 1998, pp. 390–395.

[133] L. Wietzke, “Algebraic Representation and Geometric Interpretation of Hilbert Transformed Signals,” PhD thesis, 2011.

[134] ——, “Algebraic Representation and Geometric Interpretation of Hilbert Transformed Signals (OLD),” PhD thesis, 2011.

[135] L. Wietzke and G. Sommer, “The Signal Multi-Vector,” *Journal of Mathematical Imaging and Vision*, vol. 37, no. 2, pp. 132–150, 2010.

[136] L. Wietzke, G. Sommer, and O. Fleischmann, “The geometry of 2D image signals,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2009, pp. 1690–1697.

[137] W Wirtinger, “Zur formalen Theorie der Funktionen von mehr komplexen Veränderlichen,” *Mathematische Annalen*, vol. 97, pp. 357–376, 1927.

[138] A. Wong and J. Orchard, “Robust Multimodal Registration Using Local Phase-Coherence Representations,” *Journal of Signal Processing Systems*, vol. 54, no. 1-3, pp. 89–100, 2009.

[139] M. Yang, L. Zhang, L. Zhang, and D. Zhang, “Monogenic Binary Pattern (MBP): A Novel Feature Extraction and Representation Model for Face Recognition,” in *2010 20th International Conference on Pattern Recognition*, IEEE, 2010, pp. 2680–2683.

[140] D. Zang, “Signal modeling for two-dimensional image structures and scale-space based image analysis,” PhD thesis, 2007.

[141] D. Zang and G. Sommer, “Phase Based Image Reconstruction in the Monogenic Scale Space,” in *Pattern Recognition*, 357, Springer, 2004, pp. 171–178.

[142] ——, “Detecting Intrinsically Two-Dimensional Image Structures Using Local Phase,” in *Pattern Recognition*, 357, vol. 357, Springer, 2006, pp. 222–231.

[143] ——, “The Monogenic Curvature Scale-Space,” in *Combinatorial Image Analysis*, 357, vol. 357, Springer Berlin / Heidelberg, 2006, pp. 320–332.

[144] ——, “Signal modeling for two-dimensional image structures,” *Journal of Visual Communication and Image Representation*, vol. 18, no. 1, pp. 81–99, 2007.

[145] C. Zetzsche and E. Barth, “Fundamental limits of linear filters in the visual processing of two-dimensional signals,” *Vision Research*, vol. 30, no. 7, pp. 1111–1117, 1990.

[146] L. Zhang, L. Zhang, Z. Guo, and D. Zhang, “Monogenic-LBP: A new approach for rotation invariant texture classification,” in *2010 IEEE International Conference on Image Processing*, IEEE, 2010, pp. 2677–2680.

[147] W. Zhang, J. M. Brady, H. Becher, and J. A. Noble, “Spatio-temporal (2D+T) non-rigid registration of real-time 3D echocardiography and cardiovascular MR image sequences,” *Physics in Medicine and Biology*, vol. 56, no. 5, pp. 1341–1360, 2011.