The Multiresolution Fourier Transform:

A General Purpose Tool for Image Analysis

Andrew Calway B.Sc.

A thesis submitted to
The University of Warwick
for the degree of
Doctor of Philosophy

Computer Sci

September 1989

The Multiresolution Fourier Transform:
A General Purpose Tool for Image Analysis

Andrew Calway B.Sc.

A thesis submitted to
The University of Warwick
for the degree of
Doctor of Philosophy

September 1989

Summary

The extraction of meaningful features from an image forms an important area of image analysis. It enables the task of understanding visual information to be implemented in a coherent and well defined manner. However, although many of the traditional approaches to feature extraction have proved to be successful in specific areas, recent work has suggested that they do not provide sufficient generality when dealing with complex analysis problems such as those presented by natural images.

This thesis considers the problem of deriving an image description which could form the basis of a more general approach to feature extraction. It is argued that an essential property of such a description is that it should have locality in both the spatial domain and in some classification space over a range of scales. Using the 2-d Fourier domain as a classification space, a number of image transforms that might provide the required description are investigated. These include combined representations such as a 2-d version of the short-time Fourier transform (STFT), and multiscale or pyramid representations such as the wavelet transform. However, it is shown that these are limited in their ability to provide sufficient locality in both domains and as such do not fulfill the requirement for generality.

To overcome this limitation, an alternative approach is proposed in the form of the multiresolution Fourier transform (MFT). This has a hierarchical structure in which the outermost levels are the image and its discrete Fourier transform (DFT), whilst the intermediate levels are combined representations in space and spatial frequency. These levels are defined to be optimal in terms of locality and their resolution is such that within the transform as a whole there is a uniform variation in resolution between the spatial domain and the spatial frequency domain. This ensures that locality is provided in both domains over a range of scales. The MFT is also invertible and amenable to efficient computation via familiar signal processing techniques. Examples and experiments illustrating its properties are presented.

The problem of extracting local image features such as lines and edges is then considered. A multiresolution image model based on these features is defined and it is shown that the MFT provides an effective tool for estimating its parameters. The model is also suitable for representing curves and a curve extraction algorithm is described. The results presented for synthetic and natural images compare favourably with existing methods. Furthermore, when coupled with the previous work in this area, they demonstrate that the MFT has the potential to provide a basis for the solution of general image analysis problems.

Key Words

image analysis multiresolution Fourier methods feature extraction

Acknowledgements

This work was conducted while I was employed within the Department of Computer Science at the University of Warwick and was funded by the UK SERC. I would like to thank both institutions for their support during the past 3 years.

Many people within the department have contributed to this work and their help has been much appreciated. In particular, Jeff Smith provided essential software support and Liz Woolley helped in the typing of this thesis. I would also like to thank Michael Gould in the Photographic Department for his help with the photographs.

Thanks are also due to the past and present members of the Image and Signal Processing Group at Warwick - Abhir Bhalerao, Simon Clippingdale, Roddy McColl, Edward Pearson and Martin Todd. They have made numerous contributions and have provided a stimulating environment in which to work.

I am particularly indebted to my supervisor Dr. Roland Wilson. This work would not have been possible without his wide knowledge and profound understanding of the subject. Moreover, his enthusiasm and encouragement during many discussions has been a constant source of motivation.

Finally, I would like to express my gratitude to my parents for their support and encouragement, and to Helen for her endless patience and optimism.

Contents

List of Figures

CHAPTER ONE

INTRODUCTION

1.1. The Image Analysis Problem

For most people, describing what they see in an image does not present any great difficulty. Irrespective of the content of the image, they will be able to provide some sort of description which relates directly to the visual information and is a meaningful interpretation of the image. The process by which such descriptions are arrived at is an example of image analysis, and given its capacity for dealing with visual data, it is something that the human visual system (HVS) does particularly well.

In computer vision, which is concerned with building automated vision machines [5], the problem of image analysis is one of computing appropriate descriptions from image data. These data are derived from a visual scene using a suitable sensor, such as a camera, which projects the light energy reflected from objects in the scene onto a 2-d plane. In order that a computer can process this image, it is then converted into a discrete array of numbers (known as 'picture elements' or pixels) which represent the average luminance or colour in the vicinity of discrete grid points on the plane. This acquisition stage is analogous to the eye in the HVS and the resulting data to the so-called 'retinal image' [52]. Depending upon the application, the image data may also form part of a sequence to enable the machine to process motion, or may possibly be derived from several displaced sources in order to capture 3-d information.

Vision machines have applications in a wide range of areas, including medicine, commerce and scientific research [94][95]. Consequently, the requirements of image analysis vary widely and depend upon the underlying properties of the visual

information involved. For example, it may be sufficient to regard an image as essentially 2-d and produce a description which is based upon 2-d attributes, such as orientation in the image plane or relative spatial location. This type of analysis is often acceptable when dealing with data such as satellite imagery or some medical applications. On the other hand, it may be necessary to consider 3-d aspects of the scene, in which depth, surface orientation, etc, play an important role. This is clearly appropriate when dealing with more general vision problems. Other examples might be the analysis of motion or stereopsis. Furthermore, image analysis is often of use in areas related to computer vision, such as image coding [69], in which a perceptual component is often beneficial.

The diversity of the above requirements means that image analysis employs a multitude of techniques and methodologies. These range from identifying homogeneous regions of texture to deriving 3-d shape from shading in the 2-d plane. However, although these individual tasks are clearly important, it is also apparent that there is a common operation to be performed when addressing image analysis problems. To obtain a meaningful description from an image, it is necessary to transform the array of pixel values (which constitute the visual input) into a set of symbols with appropriate relationships, eg 'the object A is of class B and is at location C' [109]. In this sense, the problem of image analysis is well defined: the inference of a symbolic description from a given input signal. By recognising this property, it is possible to consider the individual tasks mentioned above within a single framework. In doing so it ensures that they are applied in a rigorous and coherent manner and not as an ad hoc collection of unrelated operations. It has been suggested that this approach is essential if solutions are to be found to the increasingly complex problems being presented by more general vision requirements [72][109][114].

A major requirement in deriving appropriate symbols to represent a visual scene is the extraction of meaningful features from the image data. These include lines, edges, and textural properties, and give rise to a feature description which provides a more suitable basis for determining the content and structure of the image. It is therefore vital that such features are chosen carefully and with due regard for the requirements of subsequent processing. Indeed, the derivation of effective feature descriptions is an important operation in all areas of image analysis [94][95]. This is considered in greater detail in the remainder of this chapter.

1.2. Feature Based Image Description

Feature description methods originate from their use in pattern recognition systems [44]. The basic paradigm for such systems is illustrated in fig 1.1, where a feature extraction process is implemented prior to a classification or decision stage. Broadly speaking, a feature is an entity which is obtained from the input data via an appropriate measurement and this then forms the input to the subsequent processing. If a set of features is chosen carefully, then the classification or decision stages can be made more effective and reduced in complexity. The basic idea is that the features should capture the important characteristics of the input data that underlie the particular patterns being represented. If this is achieved, then the discrimination of different patterns becomes a much easier task than simply considering the data in its original form.

An alternative view of the feature extraction operation in fig 1.1 is that it represents a transformation of the input data into what is known as the feature space [44]. The dimensions of this space are given by the chosen features in the set. Any further processing, such as classification or decision making, is then conducted within this feature space and its success is dependent upon the suitability of the space to solve the

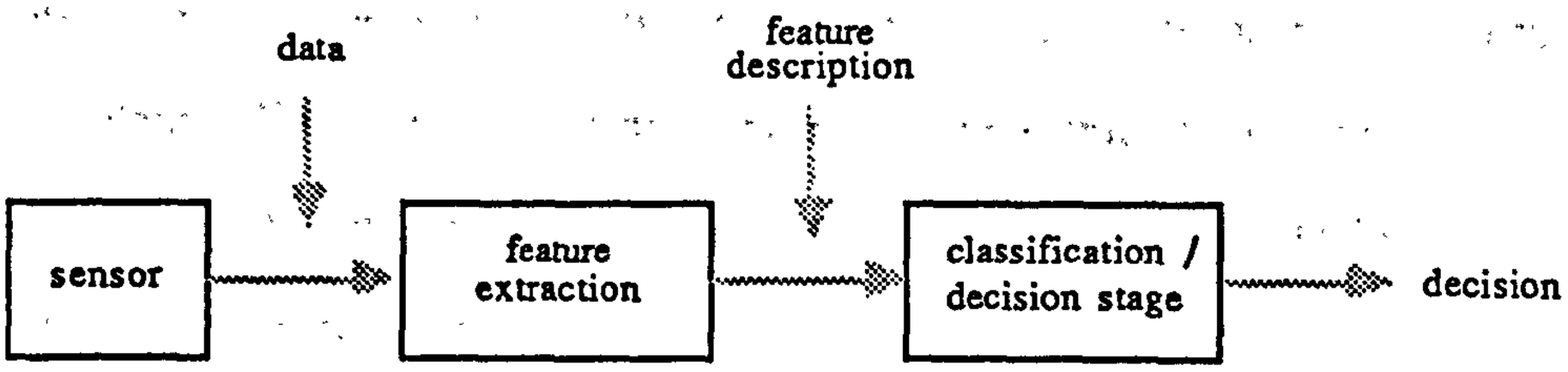


Figure 1.1 Basic pattern recognition paradigm.

problem, ie whether it provides sufficient class separability, noise immunity, etc. Regarding the use of feature descriptions in this way enables a more quantitative measure to be applied to the concept of a feature set characterising a particular problem [44].

The above ideas are readily extended to the problems of image analysis. In deriving appropriate symbols to represent a visual scene there is a need to understand and recognise the patterns being presented by the input array of pixels. These patterns are the result of the light reflected from entities in the scene impinging onto a 2-d plane. The patterns of interest are therefore those that convey information about the scene, eg shape, texture, boundaries of homogeneous regions, etc, and these then give clues about the content of the scene such as objects (their location and orientation) and the background environment. Thus although the pattern recognition paradigm of fig 1.1 is somewhat complicated by the diversity of the patterns being presented and the requirements of the problem, the motivation for extracting important features within the data is still applicable. Such features should be visually meaningful and capture the characteristics of the data that provide information about the visual scene. Having

obtained these features, the task of inferring the content of a scene becomes an easier and less complex operation in much the same way as in the original pattern recognition problems. Furthermore, the analogy of feature extraction representing a transformation of the data into a feature space enables these processes to be conducted in a more rigorous and well defined manner.

The use of feature descriptions has formed a part of image analysis since the idea of processing images was first conceived. The first serious attempt at a 'vision machine', proposed by Roberts in 1965 [92], was based upon the detection of edge features and since then the majority of image analysis systems have been based upon some sort of feature description. Amongst these, there are two areas which have received particular attention: the use of local features to identify lines and edges; and the use of textural features to represent regional properties. In the next section, the features used in these applications are considered and their important properties noted.

1.3. Applications of Feature Descriptions

1.3.1. Edge Detection

As mentioned above, edge detection methods were used in the earliest attempts at automated vision. The motivation is that transitions between luminance values in an image, which correspond to lines and edges, are perceptually important and provide a number of clues for interpreting the visual information. A basic assumption therefore is that the image consists of regions which are of constant or slowly varying luminance value and that the relevant local features are the boundaries between these regions, represented by relatively rapid transitions in luminance.

Methods of edge detection are based upon some local operation that has a maximum response in the vicinity of luminance discontinuities. By applying this operation to the whole image an estimate is obtained of the location of discontinuities and hence the local features. This apparently straightforward principle is complicated, however, by the fact that a discontinuity may also correspond to spurious features such as noise or random fluctuation. A considerable amount of attention is therefore paid to detecting only discontinuities that correspond to the features of interest.

The local operation in edge detection usually takes the form of convolving a 2-d kernel with the image. There are a vast number of possible kernels, ranging from simple masks [79] to more sophisticated designs based on signal processing principles [65]. As already noted, a significant performance criterion is noise immunity and there are a number of techniques used to reduce the effects of noise. These include spatial averaging [93] and frequency domain methods [65][96]. Other properties of lines and edges such as orientation and scale have also played an important part in successful detection [22][65]. These methods are considered in greater detail in chapter 4 of this thesis.

Local features can be used in a number of ways to enable a scene to be interpreted. Their main importance is that they convey structural information about the scene, eg the boundaries of objects are often characterised by an edge contour. This means that they are useful for both 2-d and 3-d interpretation tasks. These include the isolation and recognition of objects [5], inferring 3-d structure from motion [103] and stereopsis [72]. Lines and edges can also be significant in their own right to enable the identification of roads or waterways in satellite imaging for example [79].

Significant motivation for using local features has also come from the evidence provided by physiological investigations into the workings of the HVS. The major

advances in this area were initially based around the discovery of simple cells in mammalian vision that responded to such features [53]. Furthermore, the response is not simply to the features but specifically to those features in given orientations and with specific forms of motion (a comprehensive explanation of these findings is provided in [52]). This evidence prompted a number of workers to propose theoretical ideas for vision, the best known of these being probably that proposed by Marr [72], in which a wide range of vision tasks were explained using local features as a basis.

1.3.2. Texture Analysis and Segmentation

Another important application of feature descriptions has been in the area of characterising regional properties. It is often the case that some statistical or structural homogeneity amongst a group of pixels is perceptually important and can provide information about the content of a scene. These relationships can be analysed using textural features which capture the important properties of such groupings.

The classification of a texture which constitutes the whole image data has many connections with traditional pattern recognition problems. Referring to fig 1.1, textural features are derived from the data which are then input to an appropriate classifier. The features are chosen to reflect the various properties of texture including fineness and coarseness, directionality and regularity. There are a number of different approaches to the type of feature used and these are all based upon some global property of the pixels concerned. Examples are autocorrelation functions, spectral power density, and co-occurrence probabilities (see eg [48]). Thus in the case of the autocorrelation function, its fall off in a particular direction can characterise the coarseness of a texture and the directionality of texture can be determined from its Fourier spectrum.

The above classification of texture assumes that the image consists of a single textured region. However, an image will in general contain a number of contiguous regions and there is often a requirement in image analysis to segment the image into these regions. The textural features mentioned so far are derived from a global property of the image pixels and as such do not provide any absolute positional information, ie although they may provide information about the relative distribution of values over the image plane (eg an autocorrelation measure) they do not provide any direct position information. This means that individual regions can only be identified by treating the problem as a decision problem and deriving appropriate criteria, such as a maximum likelihood or Bayesian rule [44], to classify each pixel. Since this takes no account of spatial organisation, the estimation of region boundaries is often limited [113]. Several methods have been proposed to overcome this by providing the classification and decision stage with positional information. Examples include the use of features defined over a range of scales [98][113], iterative or relaxation techniques [17] and 'split-and-merge' methods [27]. In all these methods, however, there is inherently a trade-off between the resolution of position and class information and this must be addressed if the classification of regions is to be successful [113]. Indeed, this is a fundamental consideration when deriving feature descriptions, a point which will be returned to in later discussions (cf section 2.5.1).

The analysis of texture using feature descriptions has a number of applications. In the case of data which can be considered as 2-d, such as medical or satellite imagery, it has considerable use in segmenting the image into homogeneous regions [95]. These regions may correspond to significant areas, such as those corresponding to medical tumours, or maybe classified as belonging to certain types of crops, vegetation, etc. Textural features also have application when inferring 3-d aspects of a scene, where surface orientation can be determined which can then lead to information about 3-d shape [117].

1.4. The Need for a Unified Approach

As noted in the previous sections, the extraction of meaningful image features forms an important area of image analysis. However, recent work has suggested that many existing techniques are not sufficiently general to deal with complex analysis problems [113][114]. This section considers this in greater detail and proposes a basis for a possible solution.

A good example of the problem is apparent in the traditional approaches to the identification of lines and edges and the analysis of texture described in the section 1.3. The difficulty can be appreciated by taking a closer look at the features used in these areas and noting that there exists a fundamental difference in their properties. This relates to their relative locality in the spatial domain and can be summarised as follows:

(i) The features used to represent lines and edges are inherently local in the spatial domain; they are based upon local transitions in luminance value.

(ii) The features used to represent textural properties are more global in nature; they are based upon a population or subpopulation of pixels.

The implication of this is that the two types of feature refer to different aspects of the information being sought about a given image primitive. The features used in edge detection are providing essentially positional information, ie the location of luminance discontinuities. On the other hand, features used in the analysis of texture are providing information about the class of a region based upon some 'global' property of the pixels. Furthermore, the basic attributes of these features are apparently at odds - the need for locality to represent position and the need for a more global approach to

represent class.

This latter dichotomy underlies the lack of generality of the above methods. If there is a requirement for both local and global information, then neither of the two feature types in (i) and (ii) will provide a satisfactory solution. An example of this was noted in section 1.3.2 concerning the segmentation of textured regions: although texture may be classified by global features, the identification of boundaries in such an analysis is limited because it necessarily encompasses areas of the image in which there may exist several different regions [113]. Thus there is a conflict between the requirement for a global analysis to classify the regions and a more local approach in order to identify individual regions. A similar difficulty can be encountered when analysing an image using locally defined features. If the image contains homogeneous regions, then it is likely that many edge features will be detected within such regions. However, although in certain instances it may be possible to use this information for classification [40], in general it will not be sufficient to characterise the underlying region properties. Hence in this case the difficulty is reversed - the problem demands a more global approach instead of the local features used in the analysis.

The above discussion has wider implications for image analysis. In general problems, such as those presented by natural images, it is inevitable that a wide range of image properties will need to be considered, eg the classification of textured regions, the identification of boundaries between such regions, and at the same time the estimation of the location and orientation of lines and edges. To do this, however, the feature descriptions used must be generally applicable and not confined to a specific task. As already noted, it is often the case that existing approaches to feature description do not fulfill this requirement.

An immediate thought might be to make use of more than one approach to obtain the range of required features. However, such a solution presents numerous problems, not least in the fact that it would be difficult to combine the result of one method with that of another method. This would then inevitably lead to contradictory information. For example, a pixel or local region could be classified by an edge detection scheme as corresponding to an edge, while an analysis based on textural properties classifies it as belonging to a textured region - which method is more correct?

Clearly this does not represent a satisfactory solution. The overall conclusion must be that a more general approach is required, one in which the wide range of feature requirements can be met by a single method. A possible way of achieving this is to seek a description of the image (via a suitable transformation) that provides information about all perceptually important features. For example, such a description should have sufficient locality in the spatial domain to represent line and edge features and be sufficiently global to enable the definition of regional features. Analysis solutions based upon this description would then have access to image features that were suitable for a number of applications. In this sense it would represent a general purpose tool for image analysis.

There has recently been a growing amount of interest in the need for this type of approach and this has led to the use of methods which are both local in the spatial domain and in some class space [46][109][114]. These methods have been shown to provide advantages in a number of image analysis areas, including texture analysis and segmentation [66][99][113], line and edge extraction [65], and stereopsis [114]. Up to now, however, these techniques have remained essentially separate implementations of a general philosophy and have not taken the form of a unified image description. It is the aim of the work described in this thesis to derive such a description and apply it to a typical image analysis problem.

1.5. Thesis Outline

This thesis can be considered to consist of two parts: the derivation and definition of a unified image description; and the application of this description to the problem of extracting line and edge features from an image.

The requirements of a suitable image description are considered in chapter 2. It is shown that the description must have locality in both position and class space over a range of resolutions if it is to represent features that are generally applicable. A number of existing image transforms that may provide such a description are considered. However, it is argued that these fail to provide sufficient locality in both position and class space. This leads to the derivation of a new transform, known as the multiresolution Fourier transform (MFT).

This transform is formally defined in chapter 3. It has a hierarchical structure which can be viewed as providing local spectrum estimates over a range of spatial scales. The transform is invertible and can be efficiently implemented using familiar signal processing techniques. A general class of transforms is also introduced. Initial experiments and examples are presented for both synthetic and natural images.

Application of the MFT to the problem of extracting line and edge features is considered in chapters 4-6. A multiresolution image model based on such features is introduced in chapter 4. It is demonstrated that it has considerable generality and can be used to represent curves and boundaries.

Chapter 5 describes an estimation scheme for the model. This is based upon a local maximum likelihood estimator and a hierarchical decision process. A scheme for extracting curves is also defined within the same framework.

It is shown in chapter 6 that the MFT provides an effective and efficient tool for implementing the estimation scheme defined in chapter 5. Results are presented for both synthetic and natural images.

Finally, chapter 7 concludes the work with a synopsis and a discussion about the direction of future work.

1.6. Mathematical Preliminaries

It will be convenient in much of this thesis to make use of linear operator notation. This section presents the various conventions that are adopted.

Discrete signals will be represented by vectors which are indicated by boldface lower case letters. For example, the 1-d signal $v(i)$, $0 \le i < M$, is represented by

$$\mathbf{v} = [v_i] \qquad v_i = v(i) \qquad 0 \le i < M \tag{1.1}$$

where v_i is the ith component of the vector $\mathbf{v}$. Linear operators will be indicated by boldface upper case letters and for the 1-d case are assumed to be $(M \times M)$ matrices, ie

$$\mathbf{A} = [a_{kl}] \qquad 0 \le k,l < M \tag{1.2}$$

and if $\mathbf{v}$ in eqn (1.1) is assumed to be an $(M \times 1)$ column vector, then the vector $\mathbf{u} = \mathbf{A}\,\mathbf{v}$ is given by

$$u_k = \sum_{l=0}^{M-1} a_{kl}\, v_l \tag{1.3}$$

which is the result of operating on the vector **v** with the operator **A**.

The above notation is also adopted when the underlying problem has more than one dimension. In these cases, the components of vectors and operators are indexed by the appropriate number of indices, ie one for each dimension. For example, the 2-d signal $v(k,l)$, $0 \le k,l < M$, is represented by

$$\mathbf{v} = [\, v_{kl} \,] \qquad v_{kl} = v(k,l) \qquad 0 \le k,l < M \tag{1.4}$$

where v_{kl} is the (k,l)th component of the vector **v** and the vector $\mathbf{u} = \mathbf{Av}$, which is also defined on a 2-d support, is given by

$$u_{kl} = \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} a_{klmn}\ v_{mn} \tag{1.5}$$

where a_{klmn} are the components of the operator **A**. Note that this equation can also be written as a matrix operation by defining the vectors and operators in terms of column vectors and matrices [84].

It will be convenient to define sets of vectors and operators and these will be indexed either by subscript

$$[\ \mathbf{v}_0\ \mathbf{v}_1\ \mathbf{v}_2\ \ldots.\ \mathbf{v}_n\ ] \qquad [\ \mathbf{A}_0\ \mathbf{A}_1\ \mathbf{A}_2\ \ldots.\ \mathbf{A}_n\ ] \tag{1.6}$$

or within parentheses

$$[\ \mathbf{v}(0)\ \mathbf{v}(1)\ \mathbf{v}(2)\ \ldots.\ \mathbf{v}(n)\ ] \qquad [\ \mathbf{A}(0)\ \mathbf{A}(1)\ \mathbf{A}(2)\ \ldots.\ \mathbf{A}(n)\ ] \tag{1.7}$$

Several operators will be used frequently and it is worth defining them here to avoid unnecessary clutter in the text. This is done for the 1-d case, although they are readily extended to higher dimensions. To represent operations in the frequency domain, the discrete Fourier transform (DFT) operator will be used [91]. This is defined by

$$\mathbf{F} = [f_{kl}] \qquad f_{kl} = \frac{1}{\sqrt{M}} e^{-j\frac{2\pi}{M}kl} \tag{1.8}$$

and has an inverse operator $\mathbf{F}^+$

$$\mathbf{F}^+ \mathbf{F} = \mathbf{F} \mathbf{F}^+ = \mathbf{I} \tag{1.9}$$

where $^+$ indicates conjugate transpose and $\mathbf{I}$ is the identity operator.

Two other useful operators are those that represent shifts in frequency and position [111]. Thus the frequency shift operator $\mathbf{W}$ is defined by

$$w_{kl} = \delta(k-l)\ e^{j\frac{2\pi}{M}k} \tag{1.10}$$

where $\delta(k)$ is the Kronecker delta ($\delta(0) = 1$, $\delta(k) = 0$ for $k \neq 0$) and the circular shift operator $\mathbf{S}$ by

$$s_{kl} = \delta(k+1-l) \tag{1.11}$$

where $(k+1-l)$ is calculated modulo M.

These operators have the following properties

$$\mathbf{S^{-1}} = \mathbf{S^T} \qquad \mathbf{S} = \mathbf{F}\,\mathbf{W^+}\,\mathbf{F^+} = \mathbf{F^+}\,\mathbf{W}\,\mathbf{F} \tag{1.12}$$

$$\mathbf{W^{-1}} = \mathbf{W^+} \qquad \mathbf{W} = \mathbf{F}\,\mathbf{S}\,\mathbf{F^+} = \mathbf{F^+}\,\mathbf{S^T}\,\mathbf{F} \tag{1.13}$$

where ^{T} indicates the transpose operation.

Finally, it will also be useful to represent truncation and bandlimiting operations [111]. This can be done using the truncation operator $\mathbf{T}(\Gamma)$ defined by

$$t_{kl}(\Gamma) = \begin{cases} \delta_{kl} & 0 \le k < \Gamma \\ 0 & \text{else} \end{cases} \tag{1.14}$$

and the bandlimiting operator $\mathbf{B}(\Omega)$ defined by

$$\mathbf{B} = \mathbf{F^+}\,\mathbf{T}(\Omega)\,\mathbf{F} \tag{1.15}$$

where Γ and Ω are the truncation and bandlimiting intervals respectively.

CHAPTER TWO

TOWARDS A UNIFIED IMAGE DESCRIPTION

2.1. Introduction

The derivation of an appropriate feature description forms the first stage of most image analysis systems. It was shown in chapter 1 that many of the descriptions currently being utilised are not suitable when dealing with general analysis problems such as those presented by natural images. This led to the conclusion that a new approach was needed, in the form of an image description that provides information relevant to all perceptually important features. The purpose of this chapter is to discuss the requirements of such a description and then to define a transformation of the image data that will satisfy these requirements. The approach adopted is first to establish an underlying property of useful image features and then to base the required transformation upon this property.

2.2. A Common Property of Image Features

As noted in the previous chapter, image analysis involves making a transition from an array of pixel values to a symbolic description of the image. These symbols provide information about the content of an image, and as remarked by Marr [72], enable decisions to be made about "what is where". Such symbolic descriptions therefore have two important components [113]: a class component which indicates what something is; and a position component indicating where that something is located.

This property of symbols implies that the features used to derive them must have a common property, namely a degree of locality both in position and in some as yet unspecified class space. In other words, they should provide information about the spatial organisation of an image as well as information about the classification of separate entities. Note that this 'dual locality' is precisely what is missing from the earlier edge detection and texture analysis methods discussed in section 1.3. Each is based upon features which have locality in one domain but not in both.

A further important requirement of the features used to derive symbols is that over the range of possible features the locality in each domain should not be fixed. The features need to have various degrees of spatial locality and extend over different sized areas of the classification space. A good example of this is in the identification of homogeneous regions. These regions will have different spatial areas and will require different resolutions in the classification space. Hence the features used to represent such regions must have similar properties. This multiresolution requirement has been recognised by a number of workers [68][72].

Starting from two fundamental requirements of symbols, ie "what" and "where" information, has therefore led to a common property of the features needed to derive such symbols - a degree of locality in both position and class space which varies over the range of possible features. Consequently, if an image description is to enable such features to be derived then it must incorporate this important property. Although this is not the only property that may be considered, it is one that is generally applicable and is thus consistent with the need for generality sought in the present work. This can be contrasted with other approaches, such as those based on intensity changes in the image [72], in which it could be argued that generality has not been maintained [113].

It remains to decide upon a suitable classification space in which to base the required description. There are several possible candidates, but one that has a number of advantages is that provided by Fourier analysis. These methods have a well established theoretical base [14], and have found extensive use in identifying a wide range of image features [26][63][65][114]. In addition, there has been a considerable amount of work in physiology and psychology which suggests that it may be used in some form by the human visual system [21][24][50][104]. Although the latter is by no means conclusive, coupled with the theoretical and application arguments it does suggest that the Fourier domain is an appropriate choice.

2.3. Requirements for an Image Transform

In the previous section an essential property of the description sought in this work was derived. Based on this and other general properties, it is possible to formulate the requirements of an image transform that will be able to provide such a description. These are as follows:

(i) Locality - a property of meaningful features established in the previous section is that they should have a degree of locality in position and class space. If the classification space is assumed to be the Fourier domain, then this implies that the required description should provide information from both the spatial domain and the spatial frequency domain. In other words, the associated image transform should be one that represents a transformation into a space that is intermediate between the image and its Fourier transform. The resulting image description (or representation) is then known as a combined representation in space and spatial frequency (cf section 2.4).

(ii) Resolution - the previous section also established that the locality of features should exist over a range of scales. The implication of this for the type of transform described in (i) is that its resolution in space and spatial frequency should be adequate to represent the range of localisation exhibited by all the features of interest. For example, it should have sufficient spatial resolution to provide the required positional information and sufficient spatial frequency resolution to provide the required class information.

(iii) Invertibility - given that the above transform could be defined, then it will be advantageous if the transform were also invertible. This property will ensure that the information contained in the image is being preserved by the transformation, ie it provides an alternative representation of the image without loss of information. In addition, it will also provide a useful means of assessing the properties of models and further processing defined within the transform space.

(iv) Computational efficiency - as mentioned in chapter 1, image analysis is an ever expanding area of research with a wide range of applications. It is therefore desirable that the techniques employed are amenable to straightforward and efficient implementation. This is particularly true for techniques aimed at general use. The required transform should therefore have computational properties which are comparable to existing and established methods. A computationally inefficient and cumbersome transform will not only have limited use but also preclude general accessibility.

(v) Linearity - the final requirement for the transform is that it should be linear. This will ensure a predictable response in the transform space to the addition or weighting of features, and so will simplify analysis. Furthermore, it will enable familiar signal processing operations, such as filtering, to be considered within the transform framework.

In the following sections, existing techniques that are relevant to the above requirements are considered. An appropriate image transform based on these techniques is defined in section 2.6.

2.4. Combined Spatial and Spatial Frequency Representations

2.4.1. Introductory Remarks

In the previous section it was established that some form of representation which was intermediate between the image and the spatial frequency domain would possess properties essential to a unified image description. Representations that provide information about a signal from both its original domain and the corresponding frequency domain have received considerable attention in the literature, particularly for the 1-d case. The purpose of this section is to present and evaluate their characteristics. The best known of these fall broadly into two classes: linear forms based around the short-time Fourier transform (STFT); and bilinear forms, of which the Wigner distribution (WD) is the most widely used. For simplicity of notation and analysis, they will be considered in their 1-d form. Both classes are readily extended to the 2-d case.

2.4.2. Linear Forms

The linear forms can all be considered to be versions of the STFT. For a continuous 1-d signal $y(t)$, its STFT is defined by the following pair of equations [89]

$$Y(\tau,\omega) = \int_{-\infty}^{\infty} h(\tau-t)\, y(t)\, e^{-j\omega t}\, dt \qquad (2.1)$$

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(t-\tau)\ Y(\tau,\omega)\ e^{j\omega t}\ d\tau\ d\omega \qquad (2.2)$$

where the synthesis equation holds provided that

$$\int_{-\infty}^{\infty} h(t)\ f(-t)\ dt = 1 \qquad (2.3)$$

The functions $h(t)$ and $f(t)$ are known as the analysis and synthesis windows respectively.

The representation is clearly linear, ie if

$$y_2(t) = a\ y_1(t) + b\ y_0(t) \qquad (2.4)$$

where a and b are constants, then

$$Y_2(\tau,\omega) = a\ Y_1(\tau,\omega) + b\ Y_0(\tau,\omega) \qquad (2.5)$$

Note that for a given $\tau = \tau_0$, the function $Y(\tau_0,\omega)$ is the Fourier transform of the windowed signal $w(\tau_0,t)$

$$w(\tau_0,t) = h(\tau_0-t)\ y(t) \qquad (2.6)$$

and if $h(t)$ is chosen to be localised in both domains, then the STFT can be recognised

as a combined representation which is intermediate between the signal and its Fourier transform.

Two dimensional forms of eqns (2.1) and (2.2) are readily defined.

There are a number of related versions. The STFT of a discrete-time signal $x(n)$ is given by [89]

$$X(n,\omega) = \sum_{m=-\infty}^{\infty} h(n-m)\, x(m)\, e^{-j\omega m} \tag{2.7}$$

$$x(n) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_{r=-\infty}^{\infty} f(n-r)\, X(r,\omega)\, e^{j\omega n} d\omega \tag{2.8}$$

where the discrete window functions now satisfy

$$\sum_{m=-\infty}^{\infty} h(m)\, f(-m) = 1 \tag{2.9}$$

Eqn (2.7) represents the discrete signal $x(n)$ by a function of a continuous variable ω for each value of n and thus contains redundant information. This redundancy is minimised by defining a discrete STFT [89]

$$X(n,k) = \sum_{m=-\infty}^{\infty} h(nR-m)\, x(m)\, e^{-j\frac{2\pi}{M}km} \qquad 0 \le k < M \tag{2.10}$$

where R and M are the sampling intervals in each domain. Appropriate choice of

these parameters and synthesis window $f(n)$ means that the original signal can be reconstructed according to

$$x(n) = \frac{1}{M} \sum_{k=0}^{M-1} \sum_{m=-\infty}^{\infty} f(n-m)\, X(m,k)\, e^{j\frac{2\pi}{M}kn} \qquad (2.11)$$

provided that the two window functions satisfy

$$\sum_{s=-\infty}^{\infty} f(n-sR)\, h(sR-n+pM) = \delta(p) \qquad \text{for all } n \qquad (2.12)$$

Note the further constraint on suitable windows due to the sampling process.

The discrete STFT is the simplest example of a multirate filter bank [97][107]. These are defined by [97]

$$X(n,k) = \sum_{m=-\infty}^{\infty} h_k(nR-m)\, x(m) \qquad (2.13)$$

where the analysis windows $h_k(n)$ are no longer modulated versions of a baseband filter as in eqn (2.10), but are now independent functions of frequency. This has important implications in coding applications, where reconstruction need no longer rely upon satisfying the sampling theorem on a channel-by-channel basis as implied by eqns (2.11) and (2.12), but can make use of methods that are able to remove aliasing in the synthesis procedure [97]. The most popular of these methods is based upon quadrature mirror filter (QMF) techniques [42].

A relative of the STFT is the Gabor representation [45]. For a number of years this remained a separate entity and it was not until recently that a relationship was formally recognised [9]. This is probably due to the fact that it starts from a different viewpoint, namely the expansion of a continuous signal $y(t)$ in terms of a linear combination of elementary functions, ie

$$y(t) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} C_{kl}\ g_{kl}(t) \tag{2.14}$$

where

$$g_{kl}(t) = g(t-kT)\ e^{j(l\Omega t+\phi)} \tag{2.15}$$

$$g(t) = e^{-\alpha t^2} \tag{2.16}$$

and

$$T\Omega = \sqrt{\frac{\pi}{\alpha}} \tag{2.17}$$

As can be seen, the elementary functions $g_{kl}(t)$ are modulated and position shifted versions of a Gaussian window $g(t)$. This window is optimally concentrated into the intervals T and Ω, and as such satisfy a minimum uncertainty condition [45][83] (see later in this section). Consequently, the representation in eqn (2.14) expands the signal $y(t)$ in the signal/frequency plane defined by the indices (k,l). However, due to the fact that the functions $g_{kl}(t)$ are not orthogonal, an analytic solution for the

coefficients C_{kl} is not straightforward (Gabor suggested an iterative solution). Bastiaans [8][9] recently introduced the following solution based upon an auxiliary function $\gamma(t)$

$$C_{kl} = \int_{-\infty}^{\infty} y(t)\ \gamma^*(t-kT)\ e^{-jl\Omega t}\ dt \tag{2.18}$$

where * indicates complex conjugate and $\gamma(t)$ and $g(t)$ are related by a biorthonormality condition [9]

$$\int_{-\infty}^{\infty} \gamma(t)\ g_{kl}^*(t)\ dt = \delta(k)\ \delta(l) \tag{2.19}$$

This condition is related to that for the discrete STFT in eqn (2.12).

A 2-d version of the Gabor representation is also readily defined [37].

Efficient implementation of the various forms of the STFT is based upon fast Fourier transform (FFT) techniques [2][88]. This does not however apply to the Gabor representation, where calculation of the coefficients via eqn (2.18) is expensive in terms of computation [87]. To avoid this, Daugman [38] proposed an iterative scheme for his 2-d version.

The resolution of all these representations is determined by the windows and elementary functions. The degree of energy concentration in a given domain defines the resolution achievable in that domain. A central feature of the work of Gabor was the realisation that a signal cannot be simultaneously concentrated to an arbitrary extent

in both domains. This is known as the uncertainty principle in signal processing [45][83]. In fact, there exists a lower limit on the product of 'durations' of a signal in each domain and signals which satisfy this limit are regarded as having a minimum uncertainty condition. This applies to the elementary signals employed by Gabor and as such the representation in eqn (2.14) has optimal resolution. It is of course possible to employ windows with a similar property when implementing the STFT.

The STFT is the most well established combined representation due to its invertibility and computational advantages. It is used extensively in speech processing [90], linear time-varying filtering [89] and adaptive processing [2][61]. Multirate filter banks are also used widely, mainly in speech and image coding [97][106].

Application of the Gabor representation has been limited due to its computational difficulties. However, it is interesting to note in the context of this present work that it has received considerable interest from workers in vision research. Its relevance to the study of the visual system was first noted by Marcelja [71] in 1980, and this was supported by a number of other workers, eg [37][86]. More recently, the work of Daugman [38], Porat and Zeevi [87] and Friedlander and Porat [43] have implemented the representation for both 1-d and 2-d cases.

2.4.3. Bilinear Forms

Introduced to overcome the resolution restriction imposed on the linear forms by the window functions, bilinear combined representations provide a type of energy density measure of a signal at a given instant in each domain. The most notable of these are the Wigner distribution (WD) [29]-[31] and the ambiguity function (AF) [83]. Since these have a similar definition, being related by a 2-d Fourier transform [31], it suffices

here to consider only the more widely used WD.

The WD for a continuous signal $y(t)$ is given by [29]

$$W(t,\omega) = \int_{-\infty}^{\infty} y(t+\frac{\tau}{2})\ y^*(t-\frac{\tau}{2})\ e^{-j\omega\tau}\ d\tau \qquad (2.20)$$

and therefore at $t = t_0$, the WD is the Fourier transform of a function which corresponds to taking symmetrical correlation products about t_0. The term 'energy density measure' derives from its ability to preserve the energy distribution of the signal in both domains [15]. However, its use is often disputed due to the fact that the WD is not guaranteed to be nonnegative [15][29]. It is also worth noting that in general the WD is not invertible [29].

The use of the correlation term means that the WD is a bilinear function. In other words, the WD of the sum of two signals is not simply the sum of their respective WD's, ie if

$$y_2(t) = y_1(t) + y_0(t) \qquad (2.21)$$

then

$$W_2(t,\omega) = W_1(t,\omega) + W_0(t,\omega) + 2\,\mathrm{Re}\,W_{01}(t,\omega) \qquad (2.22)$$

where Re indicates the complex real part and $W_{01}(t,\omega)$ is the cross-WD

$$W_{01}(t,\omega) = \int_{-\infty}^{\infty} y_1(t+\frac{\tau}{2})\ y_0^*(t-\frac{\tau}{2})\ e^{-j\omega\tau}\ d\tau \qquad (2.23)$$

The WD of a discrete-time signal is not so well defined as in the continuous case and a number of versions exist. The following was adopted in [30]

$$W(n,\omega) = 2 \sum_{k=-\infty}^{\infty} x(n+k)\ x^*(n-k)\ e^{-j2k\omega} \qquad (2.24)$$

where n is the discrete variable and ω is a continuous frequency variable. This version suffers an aliasing problem which is common to a greater or lesser extent in others [32]. The problem derives from the fact that $W(n,\omega)$ is periodic with period π due to the factor 2 in the exponent of eqn (2.24). Since the spectra of discrete-time signals have period 2π, $W(n,\omega)$ will contain aliasing components. A number of techniques, including oversampling and prefiltering, have been proposed to overcome this difficulty [25][32][59].

The infinite summation in eqn (2.24) means that a variant of the WD needs to be employed in practice. This led to the pseudo-WD proposed in [29]

$$\tilde{W}(n,\omega) = 2 \sum_{k=-\infty}^{\infty} x(n+k)\ x^*(n-k)\ h(k)\ e^{-j2k\omega} \qquad (2.25)$$

where $h(n)$ is some finite window function. The effect is a smoothed WD in frequency, ie

$$\tilde{W}(n,\omega) = W(n,\omega)\ *_{\omega}\ H(\omega) \qquad (2.26)$$

where $H(\omega)$ is the DFT of $h(n)$ and $*_\omega$ denotes discrete convolution wrt the variable ω. Another example of this is the smoothed-WD introduced in [59]. Both of these forms of the WD can be implemented using FFT techniques.

In the continuous case, the bilinear forms do achieve better resolution than their linear counterparts. This is clear from eqn (2.20) where the lack of a window function means that maximal resolution is achieved [60]. However, in practice this advantage does not apply to either the pseudo or smoothed WD's. In these cases a window function is employed which means that their resolution is restricted by the uncertainty principle. In fact, the WD is further disadvantaged by its bilinearity, which often means that it is difficult to interpret for complicated signals, due to the presence of the cross-terms in eqn (2.22) [62]. Suggestions have been made for removing these cross-terms, although this has had limited success [59][62].

Despite this, the WD has been applied in a number of areas. These include optics [7], vision research [58] and speech analysis [105]. Of particular interest here is the work of Jacobson and Weschler [58]. These workers used a 2-d WD in order to define a model of the operation of simple and complex cells within the striate cortex.

2.5. Multiscale Representations

2.5.1. Fixed Window Size and Uncertainty

The resolution of the representations considered in the previous section is dependent upon a basis or window function. Such a function cannot be arbitrarily localised in both domains and therefore the representations are inherently restricted by uncertainty: defining the spatial resolution automatically sets a tight bound on the

maximum frequency resolution. What this implies in practice is that an arbitrary fixed window size must be employed, albeit based upon a window with optimal joint localisation.

This presents a considerable difficulty. As noted in section 2.2, image features that are generally applicable have varying degrees of locality in both space and spatial frequency. However, if one of the above representations is used, then it would mean adopting a trade-off between the locality of features in each domain. Indeed, choosing a given window size may provide sufficient frequency resolution to enable a feature to classify an object, but it may also mean that several objects are then present within the spatial extent of the window and thus prohibit identification.

The problem is clearly related to the uncertainty principle and is a fundamental problem when trying to simultaneously locate and classify image properties [109]. A good example of this is in the analysis of homogeneous regions mentioned at the beginning of this chapter. To obtain a reliable classification, it would be necessary to use features that are derived from a reasonably wide sample window, which would then preclude the accurate location of the boundary. In other words, the locality in position is sacrificed for locality in classification.

A possible solution to this dilemma is to make use of techniques that are finding increasing use in image processing. Multiscale methods recognise the limitations of an arbitrary fixed window size and seek an image representation in terms of windows that are scaled versions of each other. As it turns out, although these methods possess multiple resolution in the spatial domain, they fail to provide sufficient spatial frequency resolution. Nonetheless, their contribution in this area is important and, as will be shown later, they form a subset of the class of transforms sought in the present work.

2.5.2. Pyramid Representations and the Wavelet Transform

There is a wide range of methods that represent an image in terms of scaled and translated versions of a single window function. Although they often have a slightly different appearance in notation, their properties and underlying characteristics are essentially the same. The common feature is that they all consist of different resolutions of the image, which are then usually arranged in a pyramidal data structure. The generation of these different resolutions is typically achieved by some type of smoothing operation followed by a decimation process; the smoothing operator is then known as the 'generating kernel'. More recently, there has been interest in the use of kernels which have a degree of frequency and orientation selectivity, and to seek an orthogonal representation in terms of these kernels. The most notable contributions in this area are considered in this section.

The basic pyramid representation consists of a number of stacked 2-d arrays, each of which represents a different spatial resolution of the image. The bottom level is usually the image and subsequent arrays have dimensions (and resolutions) that decrease by a constant factor. For an image $v(k,l)$, $0 \le k,l < M$, where $M = 2^N$, a typical generation scheme is the following recursive smoothing and decimation operation

$$g_n(k,l) = \sum_{p=0}^{K-1} \sum_{q=0}^{K-1} w(p,q)\, g_{(n-1)}(2k+p, 2l+q) \qquad \begin{array}{l} 0 \le k,l < 2^{N-n} \\ 0 < n < N \end{array} \tag{2.27}$$

where $g_n(k,l)$ are the coefficients, or nodes, on level n of the representation, the generating kernel $w(p,q)$ is of finite size $K \times K$, and the image forms the initial level

$$g_0(k,l) = v(k,l) \qquad 0 \le k,l < M \tag{2.28}$$

The kernel $w(p,q)$ therefore defines the transformation function between the different resolutions.

The simplest example is the quadtree representation [100], where the kernel is the unweighted averaging of nodes in a 2×2 region, ie $K=2$ and $w(p,q)$ is given by

$$w(p,q) = \frac{1}{4} \qquad 0 \leq p,q < 2 \tag{2.29}$$

Each node on levels above the base level is therefore the average of its four 'child' nodes on the previous level. This representation is particularly easy to implement, since it is amenable to fast recursive computation.

Alternative kernels have been proposed which have a smoother spatial response and better joint localisation in the spatial and spatial frequency domain. Various examples and their associated properties have been considered in the literature [19][76][81]. An example is the Gaussian kernel [19], where w(p,q) approximates a Gaussian function defined on a limited support and with an appropriate variance.

The Laplacian pyramid representation is derived from the levels of a 'Gaussian' pyramid [16][18][35]. Each node is given by

$$d_n(k,l) = g_n(k,l) - g'_{(n+1)}(k,l) \tag{2.30}$$

where the nodes $g_n(k,l)$ are generated according to eqn (2.27), with $w(p,q)$ an approximation to a Gaussian function and the nodes $g'_{(n+1)}(k,l)$ are interpolated from the nodes $g_{(n+1)}(k,l)$ such that the dimensions of the arrays $g_n(k,l)$ and $g'_{n+1}(k,l)$ are the

same. From the above it can be seen that the levels of the Laplacian pyramid are the difference between successive levels of the Gaussian pyramid; it can be shown that this is equivalent to generating the levels by convolving the image with a kernel that approximates a Laplacian operator and then sampling the result [18]. The approximating kernel is known as a difference-of-Gaussians (DOG) [73]. Exact reconstruction is possible from this representation and its main use has been in image coding [18]. Its advantage over the simpler representations is that discontinuities in the image, such as lines and edges, are represented in the pyramid over a range of scales (a characteristic of the Laplacian operator) and these tend to be enhanced following reconstruction from a coding process. Since these features are important to the observer, the results have a more acceptable appearance than those based on methods less well matched to human visual perception.

A useful way to visualise the operations being performed in the above examples is to consider the frequency response of the different kernels. These are illustrated in figs 2.1a and 2.1b for the Gaussian and Laplacian pyramids respectively. The former is simply a lowpass filter whose bandwidth is a function of the level; the different levels are therefore just smoothed versions of each other. In the case of the Laplacian, the kernel is frequency selective; referring to a different circular band of the frequency domain for each level. Lines and edges, which correspond to an energy concentration in an orthogonal direction in frequency, are therefore represented on each level at a different scale (or frequency band). Note that the Laplacian pyramid has removed the redundancy apparent in the Gaussian based pyramid.

The above examples are all isotropic representations, eg lines and edges are represented equally, independent of their orientation. A natural extension is therefore to introduce anisotropy, or orientation selectivity, by dividing up the frequency domain as shown in the examples of figs 2.1c-2.1d. Such methods were recently

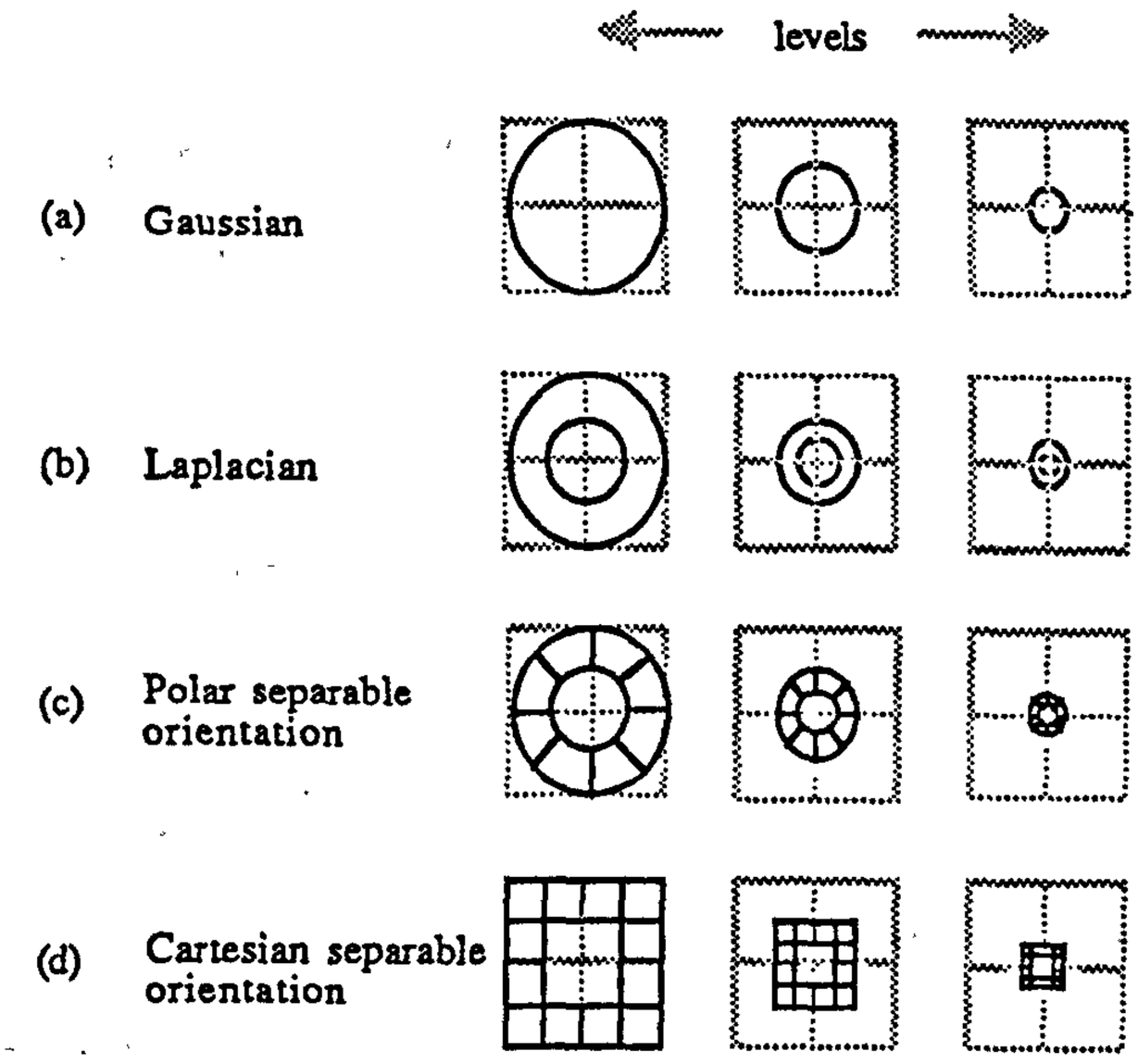


Figure 2.1. Frequency domain decomposition for various multiscale representations.

proposed by a number of workers [1][108][113]. Kernels are defined which have the appropriate frequency responses and then a filtering and decimation operation with the image produces the required data. From the responses in fig 2.1 it is clear that these recent methods are related to the combined representations considered in section 2.4. However, in these cases the frequency domain is represented on a logarithmic scale; the resolution varies (uniformly) over the domain. This correspondence was noted by Daugman [38] in his implementation of the Gabor representation, where he derived a version which was based upon oriented and logarithmically scaled elementary functions.

All the above versions have recently been placed into the theoretical framework of wavelet transforms [36][70]. The theory was originally defined by Meyer [77] and has been extended for both the 1-d and 2-d cases by a number of other workers [36]. It concerns the definition of functions which involve the projection of the signal onto a

space defined by dilations and translations of so called "wavelets" $\phi(t)$, ie for the 1-d continuous signal $x(t)$ [36]

$$U(\tau,\sigma) = \sigma^{-\frac{1}{2}} \int_{-\infty}^{\infty} \phi^*(\frac{t-\tau}{\sigma})\, x(t)\, dt \qquad (2.31)$$

where $U(\tau,\sigma)$ is known as the continuous wavelet transform. The redundancy in the 2-d 'scale-space' (τ,σ) [118], due to the overlapping of the wavelets, can be reduced by employing the discrete wavelet transform [36]

$$U(m,n) = \sigma_m^{-\frac{1}{2}} \int_{-\infty}^{\infty} \phi^*(\frac{t-n\tau_m}{\sigma_m})\, x(t)\, dt \qquad (2.32)$$

where the translation and dilation parameters now have discrete values

$$\sigma_m = \sigma_0^m \qquad \tau_m = \tau_0\, \sigma_0^m \qquad (2.33)$$

and τ_m depends upon σ_m so that as the scale parameter m increases, the translation steps move further apart to account for the widening of the functions $\phi(\cdot)$ in eqn (2.32). The pyramid representations discussed above are versions of this discrete form applied to 2-d discrete data.

In general wavelet transforms are not invertible, although by judicious choice of the wavelet function, inversion does become possible. An example is the Laplacian pyramid. Another example is when the wavelets are chosen to be orthonormal over the dilations and translations used in the transform [36]. The majority of recent work in these representations has concentrated on defining such orthogonal transforms,

where the wavelets or kernels are both frequency and orientation selective [1][70].

Multiscale methods have been applied to a wide range of image processing tasks. Perhaps the most popular is image data compression, where almost all the above versions have been applied. These include coders defined on the quadtree [110], the Laplacian pyramid [18] and the more recent orientation selective representations [1][108]. Other applications include segmentation [17][98][12], restoration [33] and edge detection [10][56].

However, although the above methods illustrate the advantages that can be gained by representing the image over a range of spatial resolutions, they do not provide the unified description sought in this work. This is because the limitations of a fixed window representation have only been addressed in terms of the spatial domain; the result in the frequency domain is still an arbitrary fixed resolution, albeit an orientation selective one. Furthermore, the frequency domain is no longer represented in a uniform way: separate subbands correspond to different spatial resolutions. These two properties are evident from the kernel frequency responses in fig 2.1.

To illustrate the effect of this upon an analysis problem, consider the available coefficients within one of the above representations that correspond to a finite spatial region. The coefficients are non-uniformly distributed across the frequency domain and they refer to different spatial resolutions of the region. The first problem, therefore, is how to combine this inhomogeneous set of coefficients to provide an effective classification space. Clearly this is not as straightforward as a simple Fourier representation. Secondly, the resolution in frequency is fixed by the choice of the scale parameter and generating kernel or wavelet, and hence the problem of classifying arbitrary features remains. For example, it would not be possible, given a division of the frequency domain as in fig 2.1c, say, to distinguish between two features which

have orientations closer than the (fixed) orientation bandwidth of the representation. In other words, these methods are limited by uncertainty in the frequency domain in a similar way to those discussed in section 2.4. It is shown in the next section that to overcome this problem it is necessary to adopt an approach which combines both of these methods into a single entity.

2.6. A Multiresolution Approach

2.6.1. Summary of Existing Methods

The joint localisation property of image features suggested the use of a combined spatial and spatial frequency representation to provide a unified description. The two classes of such representations, linear and bilinear, were considered in section 2.4. Of these, a version of the linear STFT can be defined to have a number of advantages: optimal localisation; invertibility; and computational efficiency. However, an inherent limitation is that a trade-off must be adopted between the resolution obtained in each domain, leading to a compromise which is inevitably inadequate for some features.

Multiscale methods provide a partial solution to this limitation by seeking to represent the image over a range of spatial resolutions. However, as was shown in the last section, this leads to representations in which the frequency domain has an arbitrarily fixed resolution and is non-uniformly represented. It is therefore desirable to seek a more general description, one in which both domains are represented in a uniform manner and over a multiplicity of resolutions.

2.6.2. A Unified Description

In order to provide the required description, a new transformation is adopted in this work which is based upon the STFT and is a generalisation of the multiscale methods. The basic idea is to combine a set of STFT's into a single hierarchical transform. The individual levels are then defined so that within the limits imposed by uncertainty, the resolution in each domain varies uniformly over the transform, ranging from the original signal to its DFT. Specifically for the 2-d case, the bottom level is the original image and the top level is the DFT, while the intermediate levels are STFT's with increasing spatial frequency resolution and decreasing spatial resolution, where the change of resolution is by a factor of two in each domain. Hence the transform contains the 'full range' of resolutions in both domains. This new description is known as the *multiresolution Fourier transform* (MFT).

A concrete example will help to make the above description clear. Consider the 1-d signal $v(i)$ represented by the vector $\mathbf{v}$, where $v_i = v(i)$, $0 \le i < M$ and $M = 2^N$, then its MFT vector $\mathbf{u}(n)$ is given by

$$u_{ik}(n) = \left[\mathbf{W}^{k\Omega_n} \mathbf{S}^{-i\Gamma_n} \mathbf{g}(n) \right]^+ \mathbf{v} \qquad (2.34)$$

$$0 \le i < \Omega_n \qquad 0 \le k < \Gamma_n \qquad \Omega_n = 2^n \qquad \Omega_n \Gamma_n = M \qquad 0 \le n \le N$$

where the operators $\mathbf{W}$ and $\mathbf{S}$ are the frequency and position shift operators defined in section 1.6 and $\mathbf{g}(n)$ is a family of analysis vectors which have optimal localisation in a signal domain interval of size Γ_n and a frequency domain interval of size Ω_n. A comparison with eqn (2.10) shows that the MFT, for a given value of n, is a discrete STFT with an analysis window given by the vectors $\mathbf{g}(n)$ and resolutions in each domain defined by the parameters Γ_n and Ω_n. There are N levels of the transform and

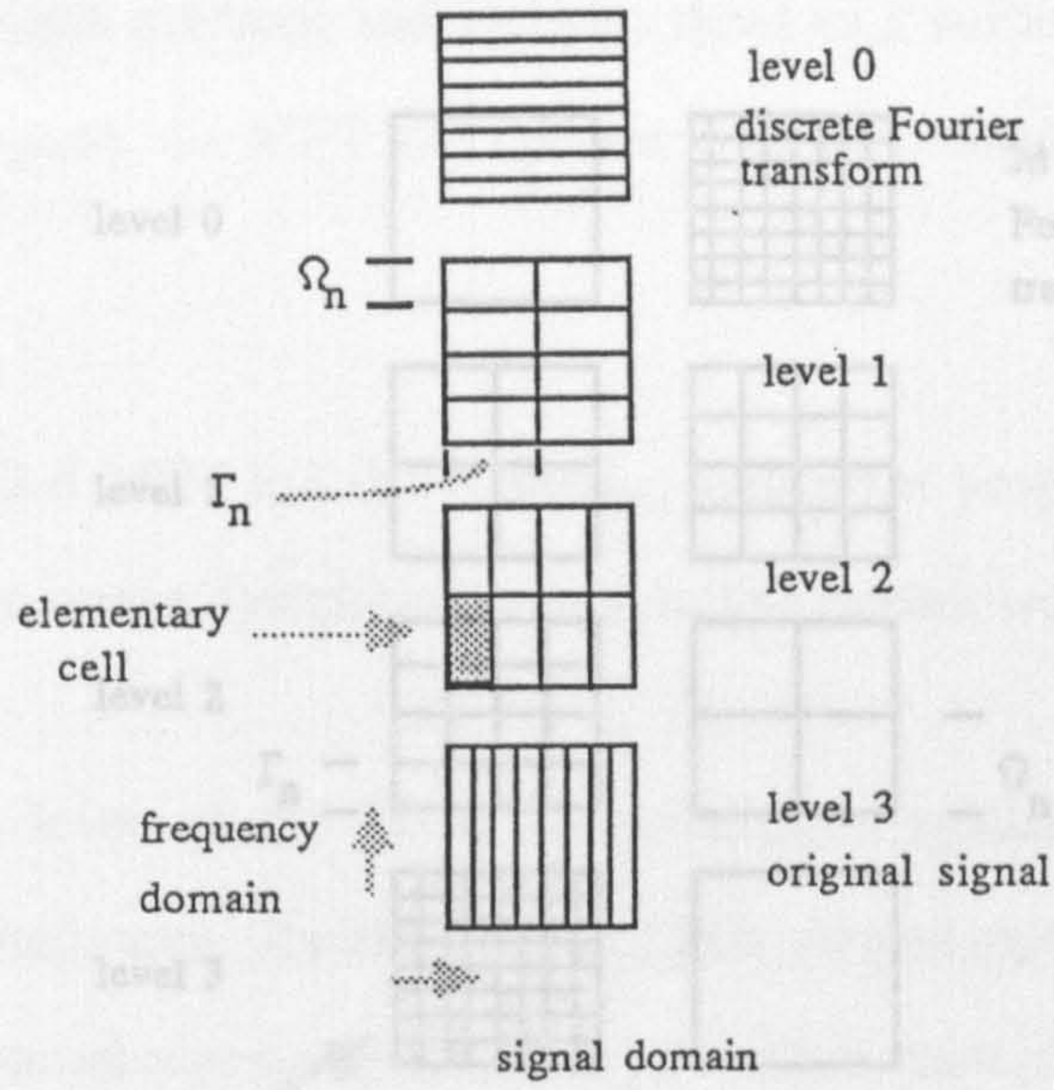


Figure 2.2. Signal/frequency diagram for 1-d MFT.

the level index n defines the ratio of resolution between the two domains as illustrated in fig 2.2.

The rectangular boxes or 'elementary cells' in fig 2.2 correspond to the position and frequency shifted versions of the analysis vectors in eqn (2.34), ie $\mathbf{W}^{k\Omega_n}\,\mathbf{S}^{-i\Gamma_n}\,\mathrm{g}(n)$. Since each coefficient of the MFT is derived from the inner product of these versions with the original signal, each cell can be considered to represent a single degree of freedom within the transform. The dimensions of the cells, Γ_n and Ω_n, are the effective durations of the analysis vectors in each domain. In other words, the vectors are defined such that their 'duration product' satisfies

$$\Gamma_n\ \Omega_n\ =\ M \qquad \Omega_n\ =\ 2^n \qquad 0 \le n \le N \tag{2.35}$$

It is clear that this relationship ensures that both domains are completely represented

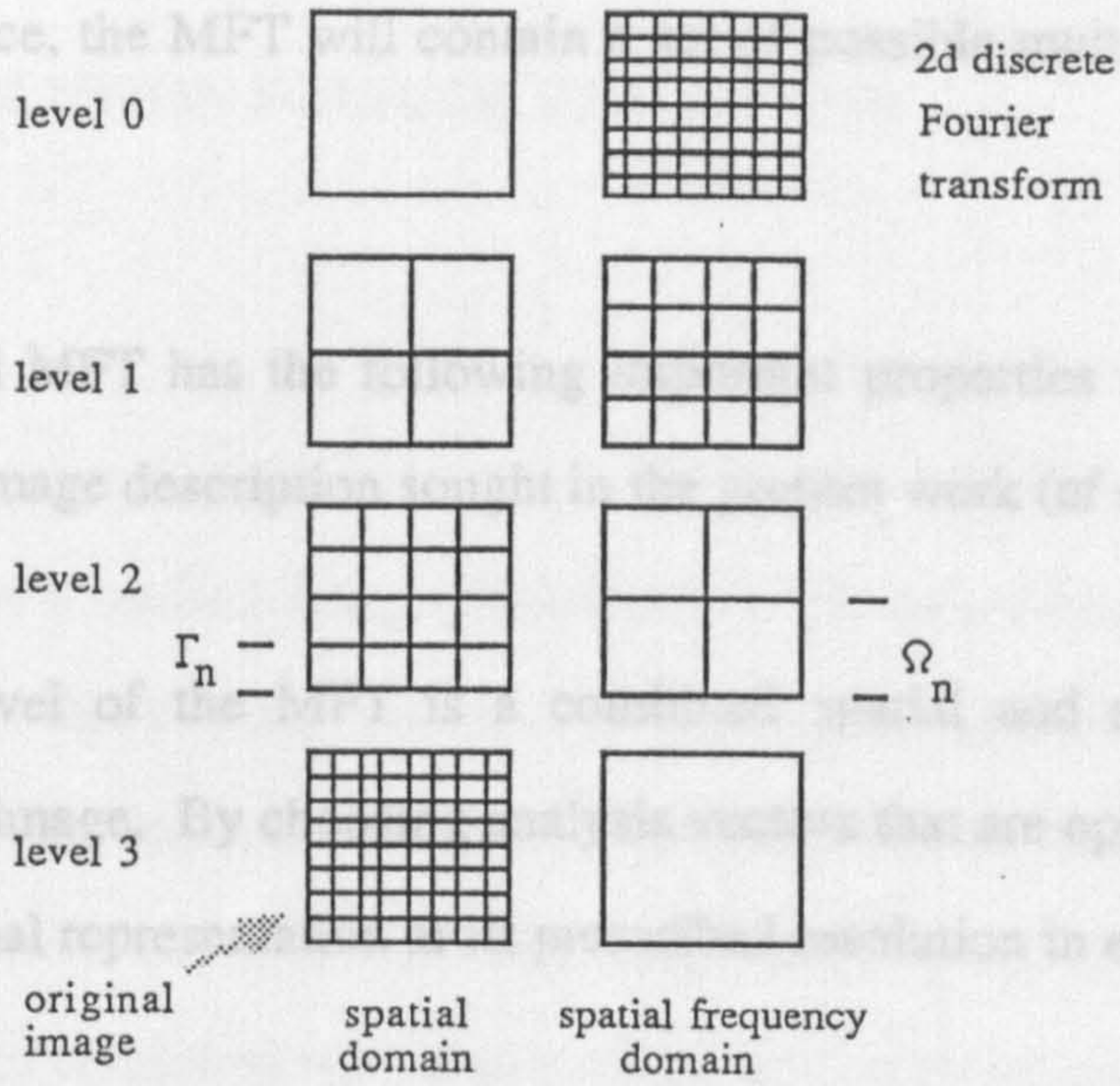


Figure 2.3. Spatial/spatial frequency diagram for 2-d MFT.

by the MFT at each level.

An extension of the above example to the 2-d case is straightforward. The resulting tessellation of each domain for a cartesian separable implementation is illustrated in fig 2.3. In this case there are five dimensions in the transform; two spatial, two spatial frequency and a level or resolution index. The elementary cells of a 2-d MFT corresponding to a single degree of freedom are therefore 4-d 'hypercubes', which in fig 2.3 are symmetrical in both the spatial and spatial frequency dimensions.

It is important to note from fig 2.3 (and fig 2.2) that for a given level (and thus resolution) both the spatial and spatial frequency domain are uniformly represented within the MFT. Recall that this was not the case for the multiscale methods discussed in section 2.5. Furthermore, it is clear from figs 2.1 and 2.3 that the MFT is in fact a

generalisation of these methods and includes them as a subset. In other words, for a given coordinate space, the MFT will contain a set of possible multiscale representations.

To conclude, the 2-d MFT has the following important properties which satisfy the requirements of the image description sought in the present work (cf section 2.3):

(i) Locality - a level of the MFT is a combined spatial and spatial frequency representation of the image. By choosing analysis vectors that are optimally localised, each level is an optimal representation at its prescribed resolution in each domain.

(ii) Resolution - the MFT contains a multiplicity of resolutions in both domains. These range from the image to its DFT. The different resolutions consist of coefficients that are uniformly distributed across the whole domain. Hence there will exist a set of coefficients in the MFT that can represent an arbitrary degree of locality in either domain exhibited by a given feature.

(iii) Invertibility - since the STFT can be defined to be invertible by judicious choice of window function, the MFT has a similar property.

(iv) Computational efficiency - each level of the MFT can be efficiently computed in a similar manner to a STFT using familiar FFT techniques.

(v) Linearity - the STFT is linear by definition and this also applies to the MFT.

In the remainder of this thesis, the MFT is considered in greater detail and its application to a typical image analysis problem is presented.

CHAPTER THREE

THE MULTIRESOLUTION FOURIER TRANSFORM

3.1. Introduction

The requirements for a unified image description were considered in the previous chapter. It was shown that existing methods fail to provide a suitable solution and that a more general approach needs to be taken. This led to the introduction of the multiresolution Fourier transform (MFT), which combines the approach of multiscale techniques with that of combined spatial and spatial frequency representation. In this chapter, the transform is considered in more detail. The forward and inverse transforms are defined using linear operator notation and the important properties of the transform are noted. It is then shown that there exists a general class of such transforms and the details of this class are considered. The chapter concludes by describing an efficient implementation scheme and presenting various examples of the transform.

3.2. Forward Transform Definition

The MFT was introduced in section 2.6. It has a hierarchical structure in which each level resembles a STFT with minimum uncertainty window functions. The resolution of these levels, determined by the parameters Γ_n and Ω_n for the signal and frequency domains respectively, varies uniformly over the transform from the original signal to its DFT according to a *scale parameter n*. For a signal vector **v**, the transform vector **u** is given by

$$\mathbf{u} = \mathbf{G}^{+}\,\mathbf{v} \tag{3.1}$$

where the operator **G** is partitioned into *level operators* **G**(n)

$$\mathbf{G} = \left[\, \mathbf{G}(0)\;\; \mathbf{G}(1)\;\; \ldots\ldots\;\; \mathbf{G}(n)\;\; \ldots\ldots\;\; \mathbf{G}(N) \,\right] \tag{3.2}$$

and the vector **u** is partitioned into *level vectors* **u**(n)

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}(0) \\ \mathbf{u}(1) \\ \cdot \\ \mathbf{u}(n) \\ \cdot \\ \mathbf{u}(N) \end{bmatrix} \tag{3.3}$$

where the resolution of **u**(n) in the signal and frequency domains is determined by the parameters Γ_n and Ω_n. The operators **G**(n) are then defined by the *analysis vectors* $\mathbf{g}_{ik}(n)$

$$\mathbf{G}(n) = \left[\, \mathbf{g}_{00}(n) \ldots\ldots \mathbf{g}_{0(\Gamma_n-1)}(n) \ldots\ldots \mathbf{g}_{ik}(n) \ldots\ldots \mathbf{g}_{(\Omega_n-1)(\Gamma_n-1)}(n) \,\right] \tag{3.4}$$

where

$$\mathbf{g}_{ik}(n) = \mathbf{W}^{k\Omega_n}\;\mathbf{S}^{-i\Gamma_n}\;\mathbf{g}(n) \tag{3.5}$$

$$0 \le i < \Omega_n \qquad 0 \le k < \Gamma_n \qquad \Omega_n = 2^n \qquad \Omega_n\Gamma_n = M \qquad 0 \le n \le N$$

and **W** and **S** are the frequency and position shift operators defined in eqns (1.10) and (1.11). A given vector $g_{ik}(n)$ is therefore a frequency and position shifted version of a basic analysis vector $g(n)$.

The definition of the set of analysis vectors $g(n)$ follows directly from the discussion in section 2.6. The requirement is that their energy should be concentrated into intervals of size Γ_n and Ω_n in the signal and frequency domains respectively. These intervals can be represented by the truncation and bandlimiting operators $\mathbf{T}(\Gamma_n)$ and $\mathbf{B}(\Omega_n)$ defined in section 1.6. Using these operators, it is possible to define a set of functions which satisfy the required energy concentration criteria and provide an invertible transform which can be efficiently implemented (cf sections 3.2 and 3.6). These are from the class of finite prolate spheroidal sequences (FPSS) [111] and they are defined by the following eigenvalue problem

$$\mathbf{B}(\Omega_n)\ \mathbf{T}(\Gamma_n)\ g(n) \;=\; \lambda_0\ g(n) \tag{3.6}$$

where λ_0 is the largest eigenvalue of the operator $\mathbf{B}(\Omega_n)\ \mathbf{T}(\Gamma_n)$. The vectors $g(n)$ are therefore bandlimited and it can be shown that among bandlimited vectors they have maximum energy concentration in the interval defined by $\mathbf{T}(\Gamma_n)$ [111]. Note that from the above equations, these vectors ensure that the signal and frequency domains are partitioned as illustrated in fig 2.2.

In a similar manner to the STFT [89], it is possible to consider two separate interpretations of a level of the MFT for a given value of n. This can be done by noting that from eqns (3.1)-(3.5), the coefficients $u_{ik}(n)$ are given by

$$u_{ik}(n) \;=\; g^{+}(n)\ \mathbf{S}^{i\Gamma_n}\ \mathbf{W}^{-k\Omega_n}\ \mathbf{v} \tag{3.7}$$

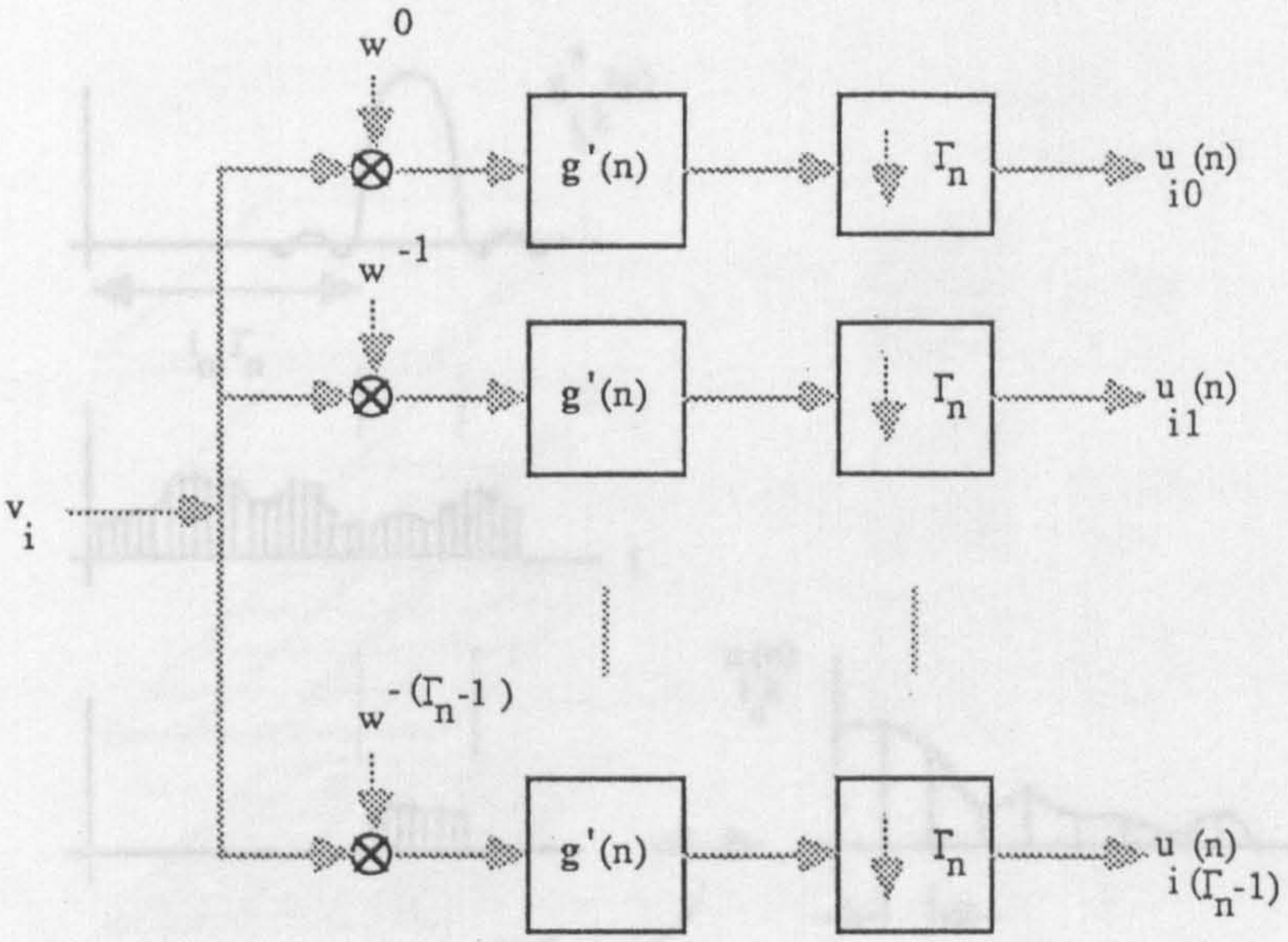


Figure 3.1. Filter bank interpretation of 1-d MFT.

and that this equation can be interpreted in one of two ways. First, for a given value of $k = k_0$, the coefficients are the sampled output of a filter with impulse response $g_l'(n) = g_{l'}^*(n)$, $l' = M - l \pmod M$, and input given by the frequency shifted signal $\mathbf{W}^{-k_0\Omega_n}\,\mathbf{v}$, where the sampling factor is Γ_n. In this case, the vector $\mathbf{u}(n)$ can be represented by the filter bank arrangement shown in fig 3.1. Secondly, for a given value of $i = i_0$, the coefficients form an estimate of a sampled Fourier spectrum of a local region in the signal domain centred at $i_0\Gamma_n + \Gamma_n/2$, with sampling interval Ω_n. This region corresponds to weighting the signal by a function which is given by a shifted version of the vector $\mathbf{g}^*(n)$ and the interpretation is illustrated in fig 3.2.

As noted above, a given level of the MFT resembles a discrete STFT. What distinguishes the MFT is the incorporation of the scale parameter n, which gives the transform a multiresolution structure. Its significance is that it determines the resolution in each domain of the vectors $\mathbf{u}(n)$. For $n = 0$, the vector $\mathbf{u}(0)$ is the DFT of the

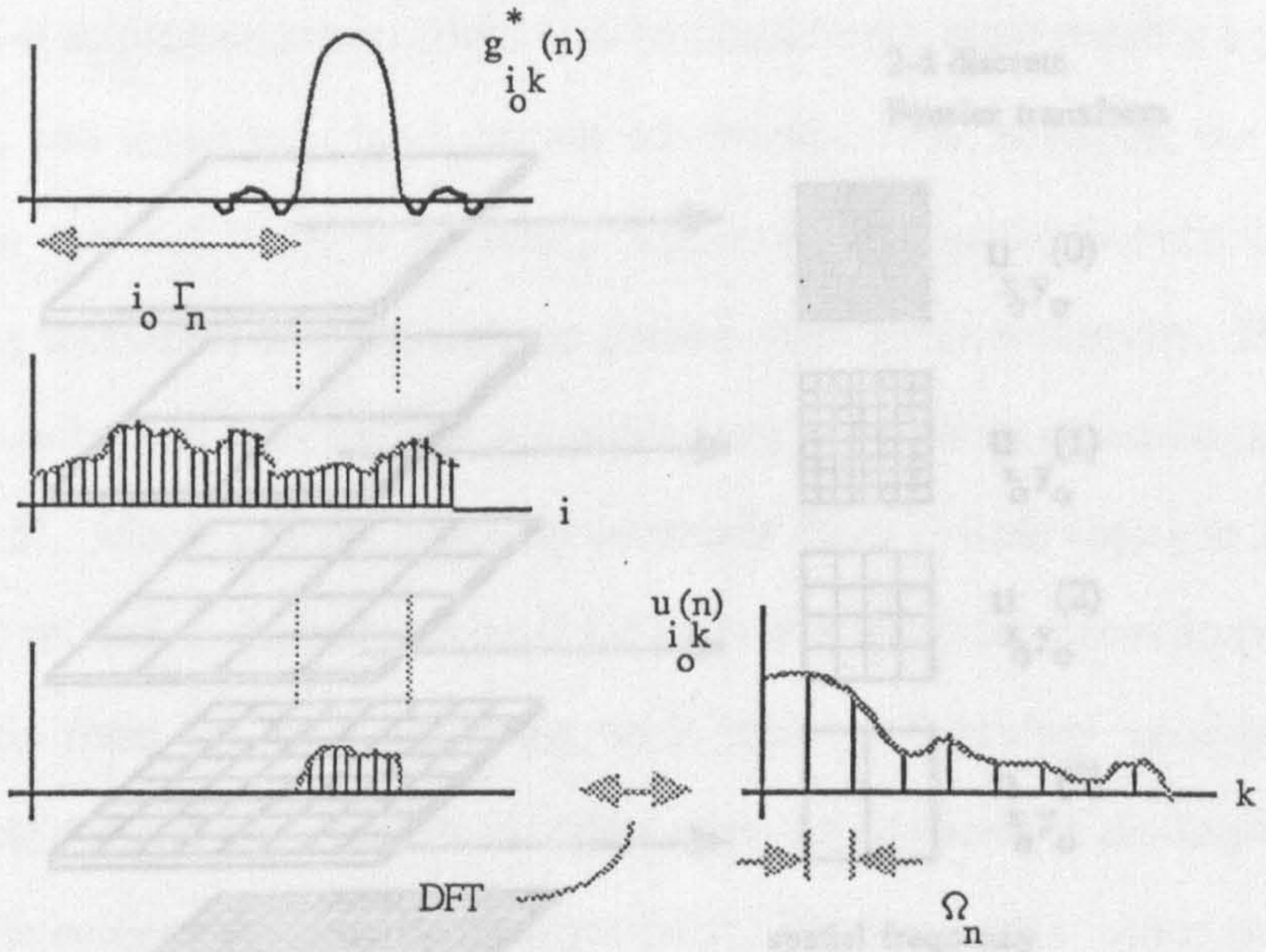


Figure 3.2. Local spectrum interpretation of 1-d MFT.

signal

$$\mathbf{u}(0) = \mathbf{G}^{+}(0)\,\mathbf{v} = \mathbf{F}\,\mathbf{v} \qquad \mathbf{G}^{+}(0) = \mathbf{F} \tag{3.8}$$

As n increases, the resolution in the signal domain increases and the resolution in the frequency domain decreases, culminating in the original signal

$$\mathbf{u}(N) = \mathbf{G}^{+}(N)\,\mathbf{v} = \mathbf{v} \qquad \mathbf{G}^{+}(N) = \mathbf{I} \tag{3.9}$$

Generalisation of the transform to 2-d is straightforward, particularly for a cartesian separable implementation. In this case, the 2-d transform operator $\mathbf{G}_2$ is simply the Kronecker product of its 1-d counterparts $\mathbf{G}_x$ and $\mathbf{G}_y$ [84]

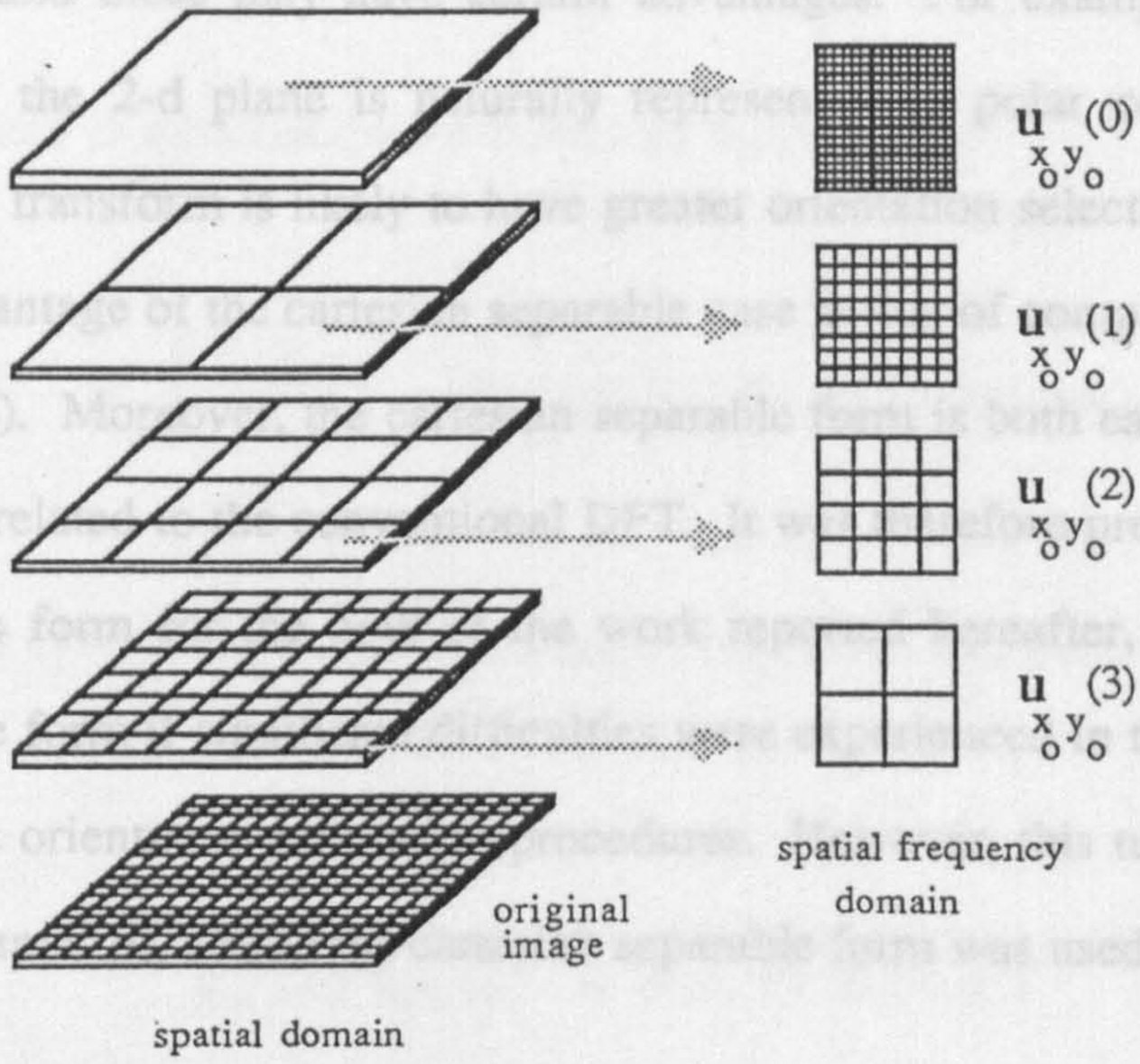


Figure 3.3. Local spectrum interpretation of 2-d MFT.

$$\mathbf{u} = \mathrm{G}_2\ \mathbf{v} = (\ \mathrm{G}_x \otimes \mathrm{G}_y\ )\ \mathbf{v} \tag{3.10}$$

where the 2-d image $v(i,j)$, $0 \le i,j < M$, is represented by stacking its rows into the $(M^2 \times 1)$ image vector **v**, ie

$$v_{(iM+j)} = v(i,j) \qquad 0 \le i,j < M \tag{3.11}$$

This 2-d version can be interpreted in a similar way to the 1-d case. For the remainder of this thesis it will be convenient to adopt the 'local spectrum' interpretation illustrated in fig 3.3. The level vectors $\mathbf{u}(n)$ are then considered to consist of a set of local 2-d spectrum estimates, each referring to a square region of the image and denoted by the vectors $\mathbf{u}_{xy}(n)$, $0 \le x,y < \Omega_n$.

Alternative 2-d implementations could also be considered, most notably a polar separable version, and these may have certain advantages. For example, the concept of orientation in the 2-d plane is naturally represented in polar coordinates and the corresponding transform is likely to have greater orientation selectivity. However, an important advantage of the cartesian separable case is that of computational efficiency (cf section 3.6). Moreover, the cartesian separable form is both easier to analyse and more directly related to the conventional DFT. It was therefore provisionally decided to employ this form for the bulk of the work reported hereafter, only adopting the polar separable form if significant difficulties were experienced in the implementation of for example orientation estimation procedures. However, this turned out not to be the case (cf chapter 6), and so the cartesian separable form was used exclusively.

3.3. The Inverse Transform

3.3.1. Exact Inversion

A signal may be reconstructed from each level vector $\mathbf{u}(n)$ of its MFT by the application of an inverse operator $\mathbf{H}(n)$

$$\mathbf{v} = \mathbf{H}(n)\ \mathbf{u}(n) \tag{3.12}$$

which is related to the forward operator by

$$\mathbf{v} = \mathbf{H}(n)\ \mathbf{G}^{+}(n)\ \mathbf{v} \tag{3.13}$$

and hence

$$\mathbf{G}^{+}(n) \;=\; \mathbf{H}^{-1}(n) \tag{3.14}$$

If $\mathbf{H}(n)$ is defined in terms of *synthesis vectors* $\mathbf{h}_{ik}(n)$

$$\mathbf{H}(n) \;=\; \left[\; \mathbf{h}_{00}(n) \ldots \mathbf{h}_{0(\Gamma_n-1)}(n) \ldots \mathbf{h}_{ik}(n) \ldots \mathbf{h}_{(\Omega_n-1)(\Gamma_n-1)}(n) \right] \tag{3.15}$$

then from eqns (3.4) and (3.14), the analysis vectors and synthesis vectors are related by

$$\mathbf{g}_{ik}{}^{+}(n)\; \mathbf{h}_{i'k'}(n) \;=\; \delta(i-i')\; \delta(k-k') \tag{3.16}$$

For a unitary transform, in which the analysis vectors are orthonormal, the synthesis vectors are simply equal to their analysis counterparts [84]. This is not the case for the MFT. Although the vectors in eqn (3.5) are orthogonal across frequency bands [111]

$$\mathbf{g}_{ik}^{+}(n)\; \mathbf{g}_{i'k'}(n) \;=\; 0 \qquad k \neq k' \tag{3.17}$$

the same does not apply with respect to position in the signal domain. However, there does exist an alternative set of vectors which satisfy eqn (3.16) and enable an exact reconstruction.

These are defined in a similar manner to the analysis set, ie

$$\mathbf{h}_{ik}(n) \;=\; \mathbf{W}^{k\Omega_n}\, \mathbf{S}^{-i\Gamma_n}\; \mathbf{h}(n) \tag{3.18}$$

where the frequency domain vectors

$$\hat{\mathbf{g}}(n) = \mathbf{F}\ \mathbf{g}(n) \qquad \hat{\mathbf{h}}(n) = \mathbf{F}\ \mathbf{h}(n) \tag{3.19}$$

are related by

$$\hat{g}_i^*(n)\ \hat{h}_i(n) = \begin{cases} 1/\Omega_n & 0 \le i < \Omega_n \\ 0 & \text{else} \end{cases} \tag{3.20}$$

and the component magnitudes of $\hat{\mathbf{g}}(n)$ are non-zero in the frequency band of interest (see appendix I)

$$|\ \hat{g}_i(n)\ | \begin{cases} \neq 0 & 0 \le i < \Omega_n \\ = 0 & \text{else} \end{cases} \tag{3.21}$$

In other words, the vectors $\mathbf{h}(n)$ are defined to have an inverse frequency response to that of the vectors $\mathbf{g}(n)$.

To see that such a choice of $\mathbf{h}(n)$ is correct, note that from eqns (3.5) and (3.18), the lhs of eqn (3.16) can be written as

$$\mathbf{g}_{ik}{}^{+}(n)\ \mathbf{h}_{i'k'}(n) = \mathbf{g}^{+}(n)\ \mathbf{S}^{i\Gamma_n}\ \mathbf{W}^{-k\Omega_n}\ \mathbf{W}^{k'\Omega_n}\ \mathbf{S}^{-i'\Gamma_n}\ \mathbf{h}(n) \tag{3.22}$$

Noting that for $\Gamma_n\ \Omega_n = M$

$$\mathbf{S}^{i\Gamma_n}\ \mathbf{W}^{k\Omega_n} = \mathbf{W}^{k\Omega_n}\ \mathbf{S}^{i\Gamma_n} \tag{3.23}$$

and using the properties of the operators S and W defined in eqns (1.12) and (1.13), eqn (3.22) becomes

$$\begin{aligned} \mathbf{g}_{ik}{}^{+}(n)\ \mathbf{h}_{i'k'}(n) &= \mathbf{g}^{+}(n)\ \mathbf{S}^{(i-i')\Gamma_n}\ \mathbf{W}^{(k'-k)\Omega_n}\ \mathbf{h}(n) \\ &= \mathbf{g}^{+}(n)\,\mathbf{F}^{+}\,\mathbf{F}\,\mathbf{S}^{(i-i')\Gamma_n}\,\mathbf{W}^{(k'-k)\Omega_n}\,\mathbf{F}^{+}\,\mathbf{F}\,\mathbf{h}(n) \\ &= \mathbf{g}^{+}(n)\,\mathbf{F}^{+}\,\mathbf{F}\,\mathbf{S}^{(i-i')\Gamma_n}\,\mathbf{F}^{+}\,\mathbf{S}^{(k-k')\Omega_n}\,\mathbf{F}\,\mathbf{F}^{+}\,\mathbf{F}\,\mathbf{h}(n) \\ &= \hat{\mathbf{g}}^{+}(n)\ \mathbf{W}^{(i-i')\Gamma_n}\ \mathbf{S}^{(k-k')\Omega_n}\ \hat{\mathbf{h}}(n) \end{aligned} \tag{3.24}$$

where $\hat{\mathbf{g}}(n)$ and $\hat{\mathbf{h}}(n)$ are defined as in eqn (3.19). Now, given the bandlimiting properties of $\mathbf{g}(n)$ and $\mathbf{h}(n)$

$$\mathbf{T}(\Omega_n)\ \hat{\mathbf{g}}(n) = \hat{\mathbf{g}}(n) \qquad \mathbf{T}(\Omega_n)\ \hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n) \tag{3.25}$$

and the definition of $\mathbf{h}(n)$ in eqn (3.20), eqn (3.24) becomes for $k \neq k'$

$$\mathbf{g}_{ik}{}^{+}(n)\ \mathbf{h}_{i'k'}(n) = 0 \qquad k \neq k' \tag{3.26}$$

while for $k = k'$

$$\begin{aligned} \mathbf{g}_{ik}{}^{+}(n)\ \mathbf{h}_{i'k'}(n) &= \sum_{l=0}^{\Omega_n-1} \hat{g}_l^{*}(n)\ \hat{h}_l(n)\, e^{j\frac{2\pi}{M}\Gamma_n(i-i')l} \qquad k = k' \\ &= \frac{1}{\Omega_n} \sum_{l=0}^{\Omega_n-1} e^{j\frac{2\pi}{M}\Gamma_n(i-i')l} \qquad k = k' \\ &= \begin{cases} 1 & i = i' \\ 0 & i \neq i' \end{cases} \qquad k = k' \end{aligned} \tag{3.27}$$

and hence

$$\mathbf{g}_{ik}{}^{+}(n)\ \mathbf{h}_{i'k'}(n) \;=\; \delta(i-i')\ \delta(k-k') \tag{3.28}$$

as required by eqn (3.16).

The above inversion is directly related to the sampling theorem and is an intuitive result when one considers the synthesis filter bank in fig 3.4, where the input to the kth channel are the coefficients $u_{ik}(n)$ for $0 \le i < \Omega_n$. The synthesised signal is then generated by interpolating (expanding and filtering) each of the Γ_n channels using a filter with impulse response $h_i(n)$, frequency shifting each output by Ω_n, and summing over all the channels. It can be seen from the analysis filter bank in fig 3.1 and the relationship between the analysis and synthesis filters in eqn (3.20), that each channel of the analysis filter bank is reconstructed exactly and that the original signal results following summation.

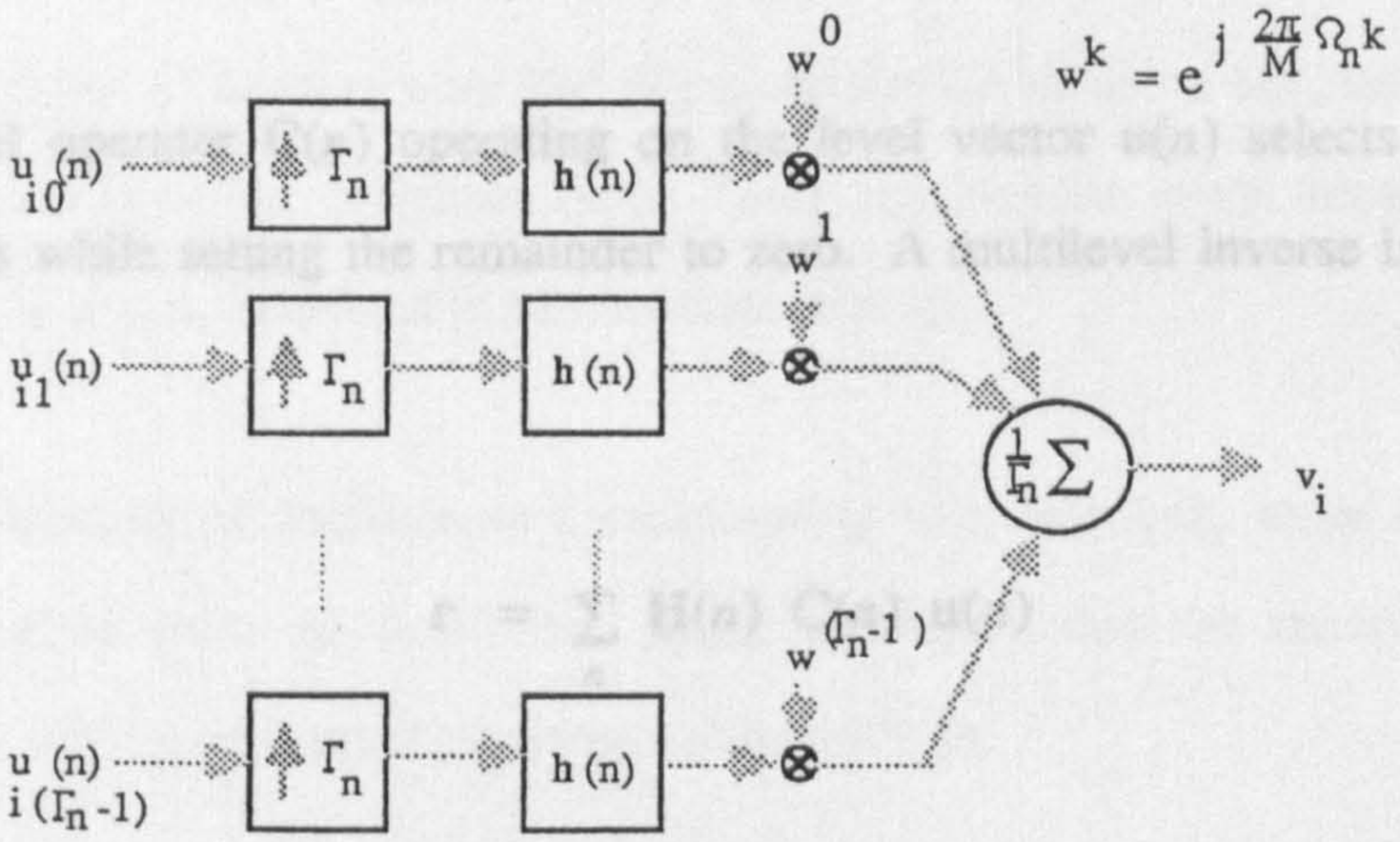


Figure 3.4. Synthesis filter bank interpretation of 1-d inverse MFT.

3.3.2. A Multilevel Inverse

It was noted in chapter 2 that useful image features have a degree of locality in both the spatial and spatial frequency domain which varies over a number of different resolutions. This implies that they will be optimally represented within the structure of a 2-d MFT by a set of coefficients from several different levels. Given that it would be possible to define a selection process to identify the relevant coefficients, the question then arises as to whether it is possible to invert a 'sparse' transform, ie one that contains coefficients from different levels with no single level having a complete set.

This can be investigated by defining an operator to represent the selection of coefficients from different levels of the MFT. Denoting this 'selection operator' by **C** and considering the 1-d case, it is given by

$$c_{iki'k'}(n) = \begin{cases} \delta(i-i')\ \delta(k-k') & \text{iff selection criterion true at } u_{ik}(n) \\ 0 & \text{else} \end{cases} \tag{3.29}$$

ie the level operator $\mathbf{C}(n)$ operating on the level vector $\mathbf{u}(n)$ selects a number of coefficients while setting the remainder to zero. A multilevel inverse is then defined as

$$\mathbf{r} = \sum_n \mathbf{H}(n)\ \mathbf{C}(n)\ \mathbf{u}(n) \tag{3.30}$$

where $\mathbf{r}$ is the reconstructed signal and $\mathbf{H}(n)$ is the inverse operator defined in the previous section.

Although trivial cases can be identified in which it is possible to exactly reconstruct an image from coefficients selected from different levels of a 2-d MFT, in general such a reconstruction will yield errors. The scale of these errors are determined by several factors:

(i) the degree of localisation in the spatial domain of the selected coefficients.

(ii) the form of the synthesis vector $\mathbf{h}(n)$.

(iii) the spectral distribution of the original image $\mathbf{v}$.

The above factors can be appreciated by considering the filter bank in fig 3.4 and noting that an arbitrary selection operator $\mathbf{C}(n)$, operating on the vector $\mathbf{u}(n)$, corresponds to a non-uniform sampling process. Since each level is a complete representation in terms of the sampling theorem, this additional sampling will result in aliasing errors in the reconstruction. These errors can then be determined from the synthesis formula in eqn (3.13). It suffices here, however, to note that the factors listed above will determine the degree of aliasing and that these are similar to those encountered in other sampling/reconstruction problems [95]. Their significance when reconstructing an image from a sparse transform is summarised below:

(i) The selection of coefficients corresponding to a relatively large spatial region (dependent upon the level) will in general mean that the central area of the region will contain small errors on reconstruction.

(ii) A synthesis vector with a smoothly varying spatial response will result in more 'acceptable' errors, reducing the edge ripples normally associated with aliasing effects.

(iii) A reconstruction for an image that possesses relatively large low frequency components will result in unpleasant high frequency aliasing as the spectrum 'folds back'. This is particularly true for natural images.

3.4. Properties of the Transform

3.4.1. Linearity and Shift Invariance

The MFT is a linear transform, ie for the vector $\mathbf{w} = a\,\mathbf{x} + b\,\mathbf{y}$

$$\mathbf{G}^{+}\,\mathbf{w} \;=\; a\,\mathbf{G}^{+}\,\mathbf{x} \;+\; b\,\mathbf{G}^{+}\,\mathbf{y} \tag{3.31}$$

where a and b are constants. This follows directly from eqns (3.1)-(3.5). It is worth noting that although not considered in the present work, this linearity property would have important advantages if one were to consider the definition of linear filtering operations within the context of the transform (see Portnoff [89]).

A level of the transform represents shifts in position up to a factor Γ_n

$$\mathbf{G}^{+}(n)\;\mathbf{S}^{r\Gamma_n}\;\mathbf{v} \;=\; \mathbf{S}^{r\Gamma_n}\;\mathbf{G}^{+}(n)\;\mathbf{v} \qquad\qquad 0 \le r < \Omega_n \tag{3.32}$$

and shifts in frequency up to a factor Ω_n

$$\mathbf{G}^{+}(n)\;\mathbf{W}^{s\Omega_n}\;\mathbf{v} \;=\; \mathbf{W}^{s\Omega_n}\;\mathbf{G}^{+}(n)\;\mathbf{v} \qquad\qquad 0 \le s < \Gamma_n \tag{3.33}$$

This shift invariant property is readily shown by considering a level of the MFT for the position and frequency shifted signal $\mathbf{W}^{s\Omega_n}\ \mathbf{S}^{r\Gamma_n}\ \mathbf{v}$ and noting that from eqns (3.4) and (3.5)

$$\begin{aligned} \mathbf{u}(n) &= \mathbf{G}^{+}(n)\ \mathbf{W}^{s\Omega_n}\ \mathbf{S}^{r\Gamma_n}\ \mathbf{v} \\ &= \left[\ \mathbf{S}^{-r\Gamma_n}\ \mathbf{W}^{-s\Omega_n}\ \mathbf{G}(n)\right]^{+}\ \mathbf{v} \\ &= \mathbf{S}^{r\Gamma_n}\ \mathbf{W}^{s\Omega_n}\ \mathbf{G}^{+}(n)\ \mathbf{v} \end{aligned} \tag{3.34}$$

3.4.2. Local Spectrum Estimation

The 2-d MFT can be interpreted as a set of 'local' spectrum estimates spaced uniformly over the image plane (section 3.2). If the transform is to be utilised in a way which assumes this interpretation, then the properties of these spectra need to be considered.

This is best achieved by again considering the 1-d case and noting that from eqns (3.7) and (3.23), the coefficients $u_{ik}(n)$ for a given $i = i_0$ are related to the shifted signal $\mathbf{S}^{i_0\Gamma_n}\ \mathbf{v}$ by

$$u_{i_0k}(n) = \mathbf{g}^{+}(n)\ \mathbf{W}^{-k\Omega_n}\ \mathbf{S}^{i_0\Gamma_n}\ \mathbf{v} \tag{3.35}$$

From eqn (1.13) this becomes

$$u_{i_0k}(n) = \left[\ \mathbf{F}\ \mathbf{g}(n)\right]^{+}\ \mathbf{S}^{k\Omega_n}\ \mathbf{F}\ \mathbf{S}^{i_0\Gamma_n}\ \mathbf{v} \tag{3.36}$$

In other words, the spectrum estimate $u_{i_0 k}(n)$, $0 \le k < \Gamma_n$, is a sampled version of the convolution between the (shifted) signal spectrum and the spectrum of the analysis vector $\mathrm{g}(n)$. The estimate is therefore a biased estimate and the nature of this bias is illustrated in fig 3.5. Note that for a given frequency $\omega_0 = 2\pi\Omega_n k_0 / M$, the estimate is the sum of frequency coefficients in its vicinity weighted by the components of the vector $\mathbf{S}^{-k_0\Omega_n}\,\mathbf{F}\,\mathrm{g}(n)$.

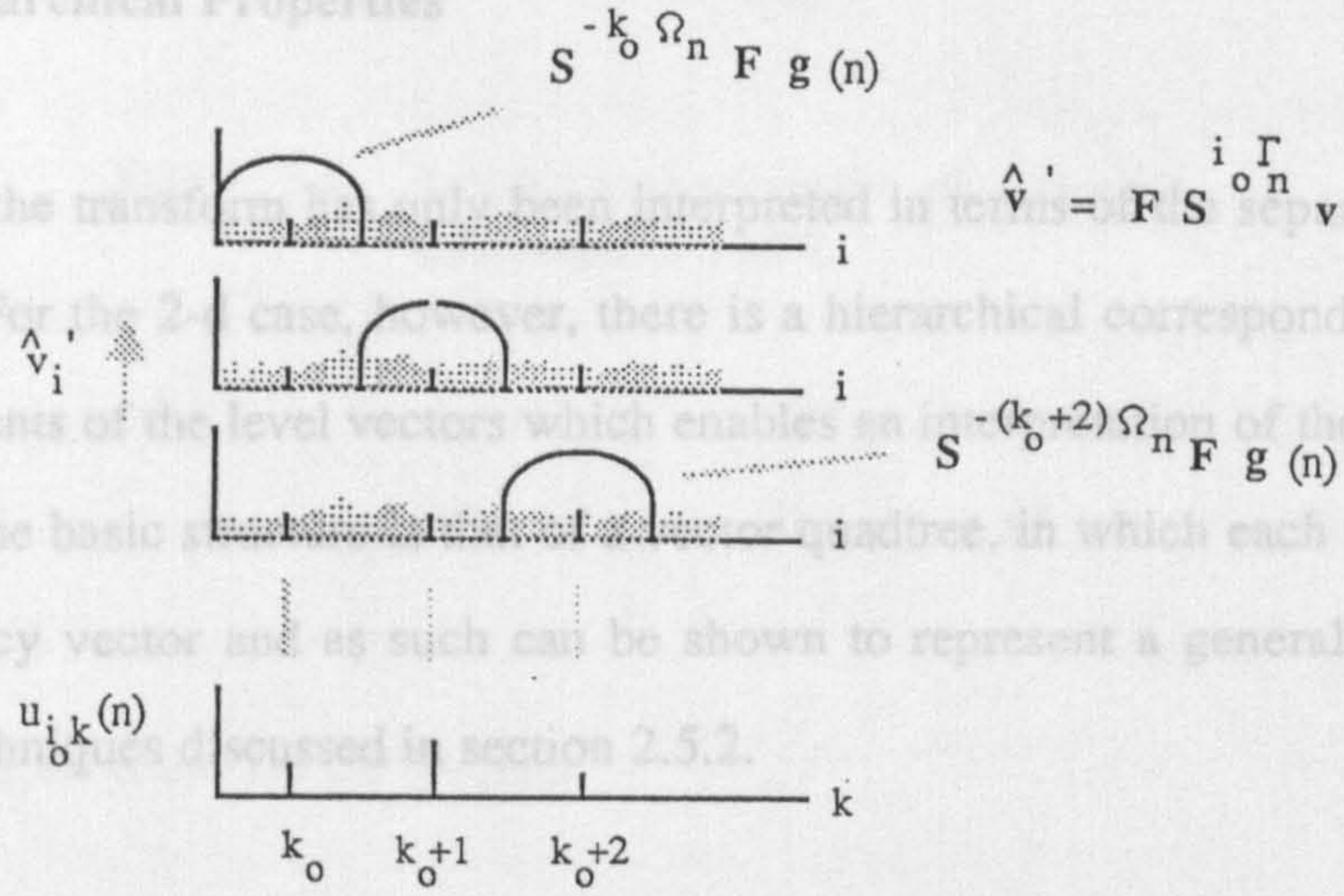


Figure 3.5. Spectrum estimate $u_{i_0 k}(n)$ interpreted as a sampled convolution.

The above is a familiar problem in spectrum estimation when the estimate is of the form defined by eqn (3.35) [49][64]. As can be seen from fig 3.5, the bias is reduced by using an analysis vector which has a small effective bandwidth, although this in turn implies that a trade-off must be made with the locality of the estimate (principle of uncertainty). Hence the optimal vectors in this respect are the finite prolate spheroidal sequences. This not unsurprising conclusion was also arrived at by Harris [49] in a comprehensive survey of available analysis windows. Furthermore,

Thomson [101] has made use of this optimality for general spectrum estimation. The local spectra of the MFT therefore constitute optimal estimates in terms of locality given their predefined resolution in each domain. It is also worth noting that the above bias can also be reduced by applying a prewhitening operation to the input signal [64][101] (cf section 3.6.4).

3.4.3. Hierarchical Properties

Up to now the transform has only been interpreted in terms of the separate level vectors **u**(n). For the 2-d case, however, there is a hierarchical correspondence between the coefficients of the level vectors which enables an interpretation of the transform as a whole. The basic structure is that of a vector quadtree, in which each node is a spatial frequency vector and as such can be shown to represent a generalisation of the pyramid techniques discussed in section 2.5.2.

The quadtree is such that each node is defined as the local spectral coefficients

$$\mathbf{u}_{xy}(n) \qquad 0 \le n \le N \qquad 0 \le x,y < \Omega_n \tag{3.37}$$

which has associated child nodes

$$\mathbf{u}_{(2x+r)(2y+s)}(n+1) \qquad 0 \le x,y < \Omega_n \quad 0 \le r,s < 2 \tag{3.38}$$

The leaf nodes are then the image **v**

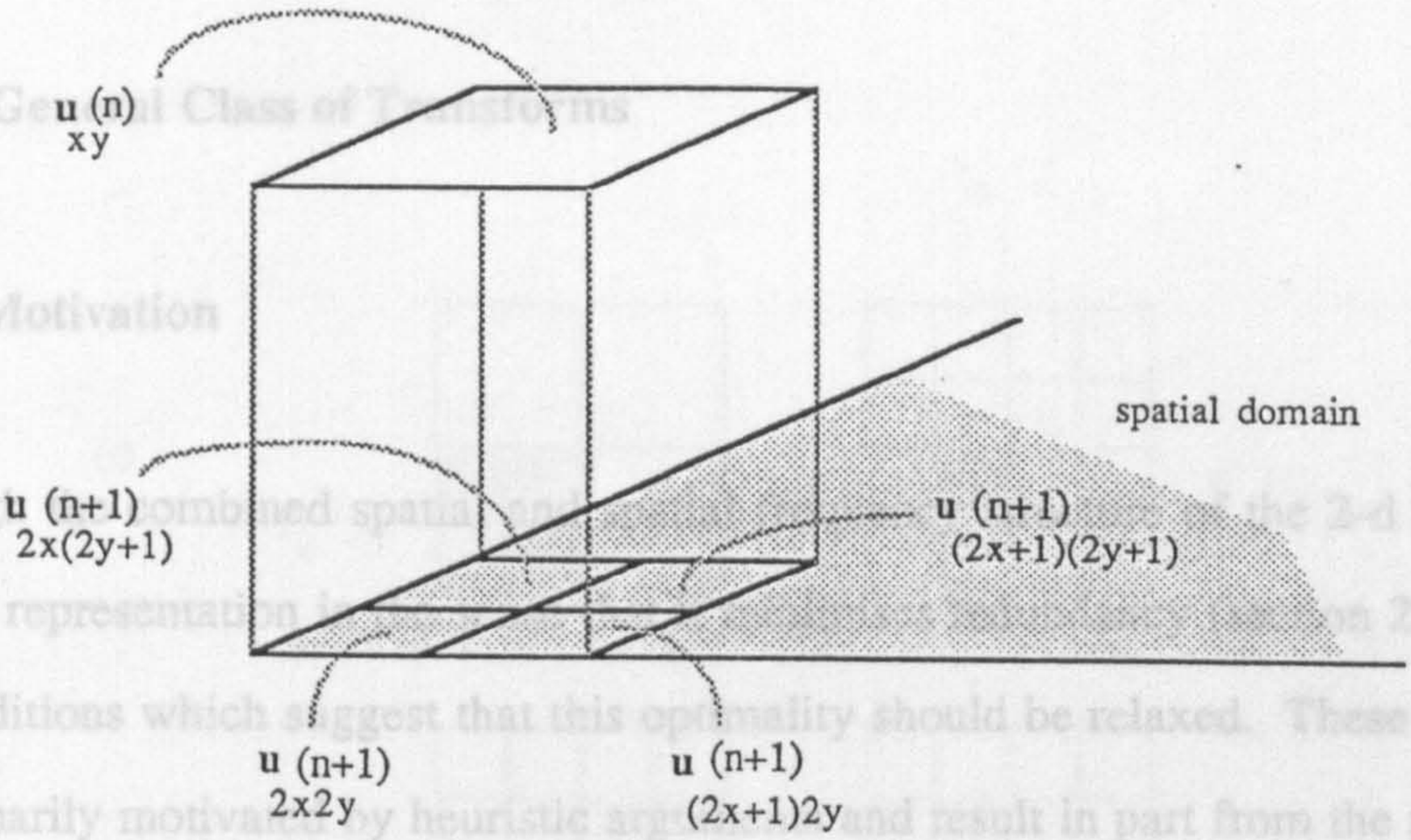


Figure 3.6. Spatial correspondence between parent and child node in 2-d MFT.

$$u_{xy\,00}(N) \;=\; v_{xy} \qquad 0 \le x,y < M \tag{3.39}$$

and the root node is the 2-d DFT vector

$$\mathbf{u}_{00}(0) \;=\; \mathbf{F}\ \mathbf{v} \tag{3.40}$$

The relationship between parent and child nodes is one of spatial correspondence as illustrated in fig 3.6. The frequency vector at the parent node refers to the (square) local region whose four quadrants are the regions referred to by each of the child vectors (section 3.2).

3.5. A General Class of Transforms

3.5.1. Motivation

Although the combined spatial and spatial frequency structure of the 2-d MFT is an optimal representation in the sense that it minimises redundancy (section 2.6.2), there are conditions which suggest that this optimality should be relaxed. These conditions are primarily motivated by heuristic arguments and result in part from the application of the MFT to image analysis problems. The outcome is a general class of the transforms which are related to the original version defined in section 3.2.

The goal is to achieve greater locality of energy in the spatial dimension of the spatial/spatial frequency diagram of fig 2.3. In other words, each analysis vector in the MFT should have increased energy concentration in its respective spatial region. The resulting transform would then have improved spatial locality, which is clearly of interest when analysing local image regions. However, from the principle of uncertainty, this necessitates an increase in the bandwidth of the signals, ie the duration product of eqn (2.35) becomes

$$\Omega_n \cdot \Gamma_n = \sigma M \qquad \sigma > 0 \tag{3.41}$$

where σ is known as the *relaxation parameter*. A level of the new spatial/spatial frequency diagram is illustrated in fig 3.7a for $\sigma = 2$, where the dotted lines indicate the new frequency region of the cell A. Note that there is now an overlapping in frequency of the cells.

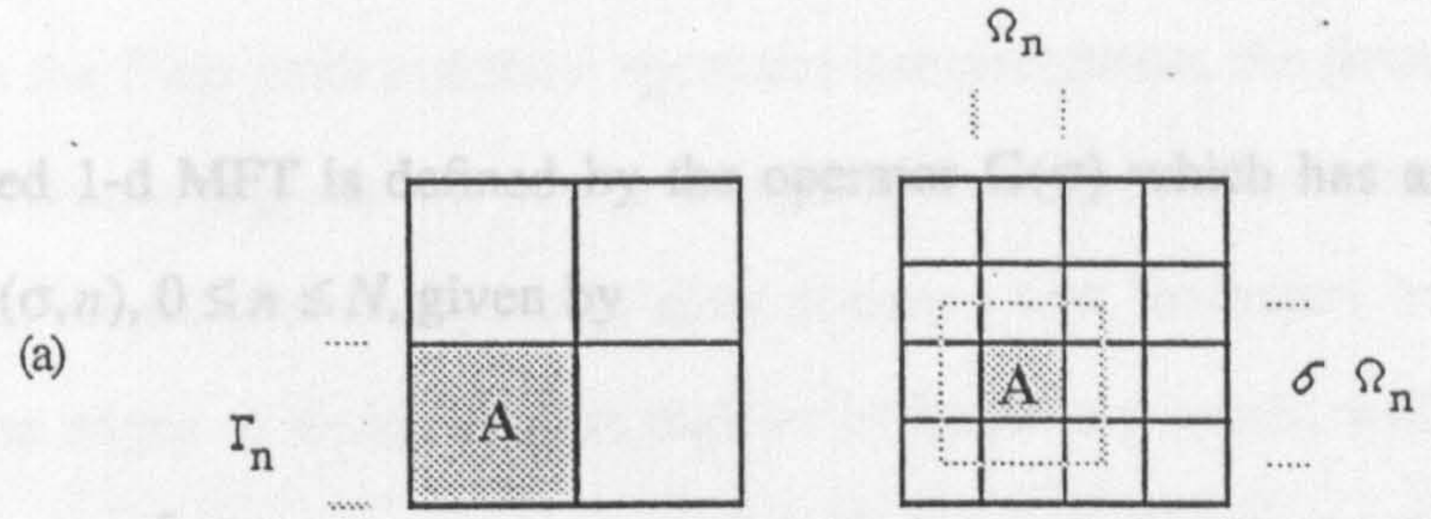


Figure 3.7. Relaxed spatial/spatial frequency diagram for generalised transform.

A consequence of this bandwidth expansion is that the diagram of fig 3.7a is no longer an optimal description in terms of the sampling theorem (section 2.6.2). To achieve an optimal description, the spatial separation of the cells needs to be reduced such that

$$s = \Gamma_n/\sigma \qquad \sigma > 0 \tag{3.42}$$

The resulting diagram is shown in fig 3.7b. Note that now there is also an overlapping of cells in the spatial dimension and that the degree of overlapping in space and frequency is determined by the relaxation parameter σ, which can be set as appropriate. However, from fig 3.7b and in terms of computational efficiency (cf section 3.7), a suitable value is $\sigma = 2$. The implication of the above is that a general class of MFT's can be defined, the structure of which are considered in the following two sections.

3.5.2. Filter Relaxation and Spatial Oversampling

A generalised 1-d MFT is defined by the operator $G(\sigma)$ which has associated level operators $G(\sigma,n)$, $0 \le n \le N$, given by

$$G(\sigma,n) = \left[g_{00}(\sigma,n) \ldots g_{0(\Gamma_n-1)}(\sigma,n) \ldots g_{ik}(\sigma,n) \ldots g_{(\sigma\Omega_n-1)(\Gamma_n-1)}(\sigma,n) \right] \quad (3.43)$$

where

$$g_{ik}(\sigma,n) = W^{k\Omega_n} S^{-i\Gamma_n/\sigma} g(\sigma,n) \quad (3.44)$$

$$0 \le i < \sigma\Omega_n \qquad \sigma > 0 \qquad 0 \le k < \Gamma_n \qquad \Omega_n = 2^n \qquad \Gamma_n \Omega_n = M$$

and the analysis vectors $g(\sigma,n)$ are adaptations of the FPSS $g(n)$

$$g(\sigma,n) = B(\sigma\Omega_n)\ T(\Gamma_n)\ g(n) \quad (3.45)$$

These generalised analysis vectors are therefore relaxed in terms of bandwidth with respect to the vectors $g(n)$ (from Ω_n to $\sigma\Omega_n$) and will consequently have greater energy concentration in the interval defined by $T(\Gamma_n)$.

Comparison of eqns (3.5) and (3.44) indicates that in the signal domain the coefficients of the generalised transform are oversampled in comparison with the original transform. This ensures that the representation is complete in terms of the sampling theorem (ie it accounts for the increase in bandwidth of the analysis vectors) and retains the invertibility of the transform (cf section 3.5.3). However, this also means that the number of coefficients is increased by a factor σ.

Interpretation of this generalised transform is similar to that for the original transform. However, in the filter bank and local spectrum interpretations, the frequency intervals and signal domain regions are now overlapping (as opposed to being contiguous). This provides an additional advantage since it means that 'boundary events', ie those falling on the edges of signal domain regions or frequency bands, will be less disadvantaged than in the non-overlapping original version. These events will now be within the energy concentration regions of the analysis vectors and will have greater energy value in the transform. This useful property resembles the use of overlapping data samples in spectrum estimation problems [49][91]. The properties outlined in section 3.4 also apply to the generalised version, although the local spectra on each level will have more bias (dependent on the value of σ), since the bandwidth of the analysis vectors has been increased.

3.5.3. General Inverse Transform

The inversion is best considered using the synthesis filter bank in fig 3.4. There are two possible inverse procedures, one is similar to that of the original inverse, while the other makes use of the overlapping in frequency apparent in the generalised version.

From figs 3.1 and 3.4, it is clear that the original signal can be reconstructed from the coefficients by defining the synthesis vector as before, ie

$$\hat{g}_i^*(\sigma,n)\ \hat{h}_i(\sigma,n) = \begin{cases} 1/\Omega_n & 0 \le i < \Omega_n \\ 0 & \text{else} \end{cases} \qquad (3.46)$$

where

$$\hat{g}(\sigma,n) = \mathbf{F}\ g(\sigma,n) \qquad \hat{h}(\sigma,n) = \mathbf{F}\ h(\sigma,n) \tag{3.47}$$

Note that the frequency interval remains at Ω_n and is not $\sigma\,\Omega_n$, the bandwidth of the new analysis vectors. The inversion follows from the discussion in section 3.3.1.

An alternative inverse can also be defined which makes use of the overlapping in the frequency responses of the analysis vectors. Referring to fig 3.4, instead of requiring that each bandpass signal is to be reconstructed exactly, it is sufficient to require that the summation of the analysis-synthesis frequency response products result in a 'flat' response overall. Obviously this applies in the original inverse definition. However, this no longer implies that the synthesis vector must have an inverse frequency response to that of the analysis vector. Provided the overall response is unity then the original signal is reconstructed. This enables the adoption of synthesis vectors without frequency domain discontinuities (as in the original inverse) and enables the analysis vector to be reapplied for synthesis without noticeable error (cf section 3.6.3).

3.6. Implementation

3.6.1. The Forward and Inverse Transform

Implementation of the MFT can be considered using the analysis and synthesis filter banks in figs 3.1 and 3.4. Referring to fig 3.1, the transform coefficients are generated by uniform sampling of the output of a set of filters, each with impulse response $g'(n)$ and inputs which are frequency shifted versions of the input signal. The signal is

reconstructed in the opposite manner (fig 3.4) - expanding, application of synthesis filter, frequency shifting and summation of the resulting bandpass signals. These two operations can be efficiently implemented when the computation is performed in the frequency domain and use is made of the fast Fourier transform (FFT) [91].

It is simplest to consider the original 1-d MFT and note that from eqn (3.7) the coefficients $u_{ik}(n)$ for a given $k = k_0$ are given by the sampled convolution relationship

$$u_{ik_0}(n) \;=\; \mathbf{g}^{+}(n)\; \mathbf{S}^{i\Gamma_n}\; \mathbf{W}^{-k_0\Omega_n}\; \mathbf{v} \tag{3.48}$$

Using the sampling property of the DFT [14], the Ω_nth order DFT with respect to the index i of these coefficients is given by

$$\hat{u}_{\omega k_0}(n) \;=\; \frac{1}{\Gamma_n} \sum_{r=0}^{\Gamma_n-1} \hat{g}^{*}_{(\omega+r\Omega_n)}(n)\; \hat{v}_{(\omega+k_0\Omega_n+r\Omega_n)} \qquad 0 \le \omega < \Omega_n \tag{3.49}$$

where $\hat{\mathbf{g}}(n) = \mathbf{F}\,\mathbf{g}(n)$, $\hat{\mathbf{v}} = \mathbf{F}\,\mathbf{v}$, and the subscripts are calculated modulo M. Since $\mathbf{g}(n)$ is bandlimited (eqn 3.25), ie $\hat{g}_{\omega}(n) = 0$ for $\omega \ge \Omega_n$, the above equation reduces to

$$\hat{u}_{\omega k_0}(n) \;=\; \frac{1}{\Gamma_n}\; \hat{g}^{*}_{\omega}(n)\; \hat{v}_{(\omega+k_0\Omega_n)} \qquad 0 \le \omega < \Omega_n \tag{3.50}$$

which is simply the product of the analysis vector frequency response and a shifted version of the signal spectrum. Hence, the coefficients $u_{ik_0}(n)$ can be generated by applying eqn (3.50) to the DFT of the input signal and forming the Ω_nth order inverse DFT of the result.

Reconstruction of the original signal is the reverse procedure. From eqns (3.20) and (3.50) the vector $\hat{\mathbf{v}}$ is given by

$$\hat{v}_{(\omega+k\Omega_n)} = M\ \hat{h}_\omega(n)\ \hat{u}_{\omega k}(n) \qquad 0 \le \omega < \Omega_n \quad 0 \le k < \Gamma_n \qquad (3.51)$$

where $\mathbf{h}(n) = \mathbf{F}\ \mathbf{h}(n)$. The original signal is therefore reconstructed by forming the Ω_nth order DFT wrt the index i of the coefficients $u_{ik}(n)$ for each value of k, implementing eqn (3.51) and deriving the Mth order inverse DFT of the result. Both this and the forward transform procedure can be efficiently implemented using the FFT (cf section 3.6.2). Since the generalised transform is also based on bandlimited analysis vectors, its implementation takes a similar form.

For a 2-d cartesian separable implementation it is convenient to introduce an appropriate frequency shift in the MFT so that the spatial frequency vectors possess a symmetry property. This is particularly useful when extracting orientation information (cf chapter 6). In the present work, this is achieved by frequency shifting the input signal such that

$$\mathbf{v}' = \mathbf{W}^{\frac{(M-1)}{2}}\ \mathbf{v} \qquad (3.52)$$

It can then be shown that the components of the spatial frequency vectors $\mathbf{u}_{xy}(n)$ of the resulting MFT are related by

$$u_{xykl}(n) = u^*_{xyk'l'}(n)\, e^{-j\frac{2\pi}{\Omega_n}(x+y)} \qquad (3.53)$$

$$0 \le k,l < \Gamma_n \quad k' = \Gamma_n - 1 - k \quad l' = \Gamma_n - 1 - l$$

where the phase term $2\pi(x+y)/\Omega_n$ results from the definition of the analysis vectors in eqn (3.6), ie the truncation interval defined by the operator $\mathbf{T}(\Gamma_n)$. The magnitude values of the components of $\mathbf{u}_{xy}(n)$ therefore possess a 2-fold rotational symmetry about the notional centre of the 2-d lattice (k,l). This symmetry property enables the extraction of orientation information within such a lattice to be easier from both a conceptual and practical point of view.

3.6.2. Computational Requirements

As demonstrated in the previous section, implementation of the transform can be efficiently based upon the FFT. This enables a 2-d DFT of order M to be calculated using $M^2\log_2 M$ complex multiplications [91]. In this section the number of calculations required to generate a complete 2-d MFT (and inverse) is considered in terms of this basic operation.

Consider first the generation of an intermediate level $\mathbf{u}(n)$, $1 \le n \le N-1$. With reference to the previous section, by extending eqn (3.50) to the 2-d case and assuming the image DFT to be available, the number of required operations $\mathcal{N}_n$ is given by

$$\begin{aligned} \mathcal{N}_n &= \Gamma_n^2 \left(\Omega_n^2 + \Omega_n^2 \log_2 \Omega_n \right) \qquad \Omega_n = 2^n \\ &= M^2 \left(1 + n \right) \end{aligned} \tag{3.54}$$

where the product in eqn (3.50) is of dimension Ω_n^2 for the 2-d case and each Ωth order inverse DFT requires $\Omega_n^2 \log_2 \Omega_n$ operations.

Thus for the complete transform consisting of N levels (including the initial DFT of the image which also forms level 0)

$$\mathcal{N} = \sum_{n=1}^{N-1} \mathcal{N}_n + M^2 \log_2 M \tag{3.55}$$

which becomes from eqn (3.54)

$$\mathcal{N} = \frac{M^2}{2} \left[(\log_2 M)^2 + 3 \log_2 M - 2 \right] \tag{3.56}$$

and is less than a factor of $\log_2 M$ times that for the FFT. Obviously the same number of operations is required for the inverse transform. In the case of a generalised transform, computation increases according to the relaxation parameter σ

$$\mathcal{N} = \sigma^2 \frac{M^2}{2} \left[(\log_2 M)^2 + \log_2 M - 2 \right] + M^2 \log_2 M \tag{3.57}$$

3.6.3. Finite Prolate Spheroidal Sequences

The analysis and synthesis vectors of the forward and inverse transform are derived from FPSS's according to eqns (3.6) and (3.20). In the case of a generalised transform, a single vector can be utilised for both analysis and synthesis without noticeable error.

For a given order of transform, the vectors need only be generated once and can then remain in situ. In the case of the original transform defined in section (3.2) this

requires the solution of the eigenvalue problem in eqn (3.6) and then the analysis vectors for the 2-d cartesian separable transform are given by

$$g_2(n) = g_x(n) \otimes g_y(n) \tag{3.58}$$

where $g_x(n)$ and $g_y(n)$ are the analysis vectors of the corresponding 1-d transforms. Software routines are readily available to solve eigenvalue problems and since computational time is not an issue, these can be used accordingly. However, if these are not available, then the successive approximation method for the FPSS's suggested by Wilson [111] provides a suitable alternative. This latter technique was used in the work described here. The magnitude response in the spatial and spatial frequency domain of the 2-d analysis vectors $g_2(n)$ for $n = 3$ and $n = 4$ are shown in fig 3.8.

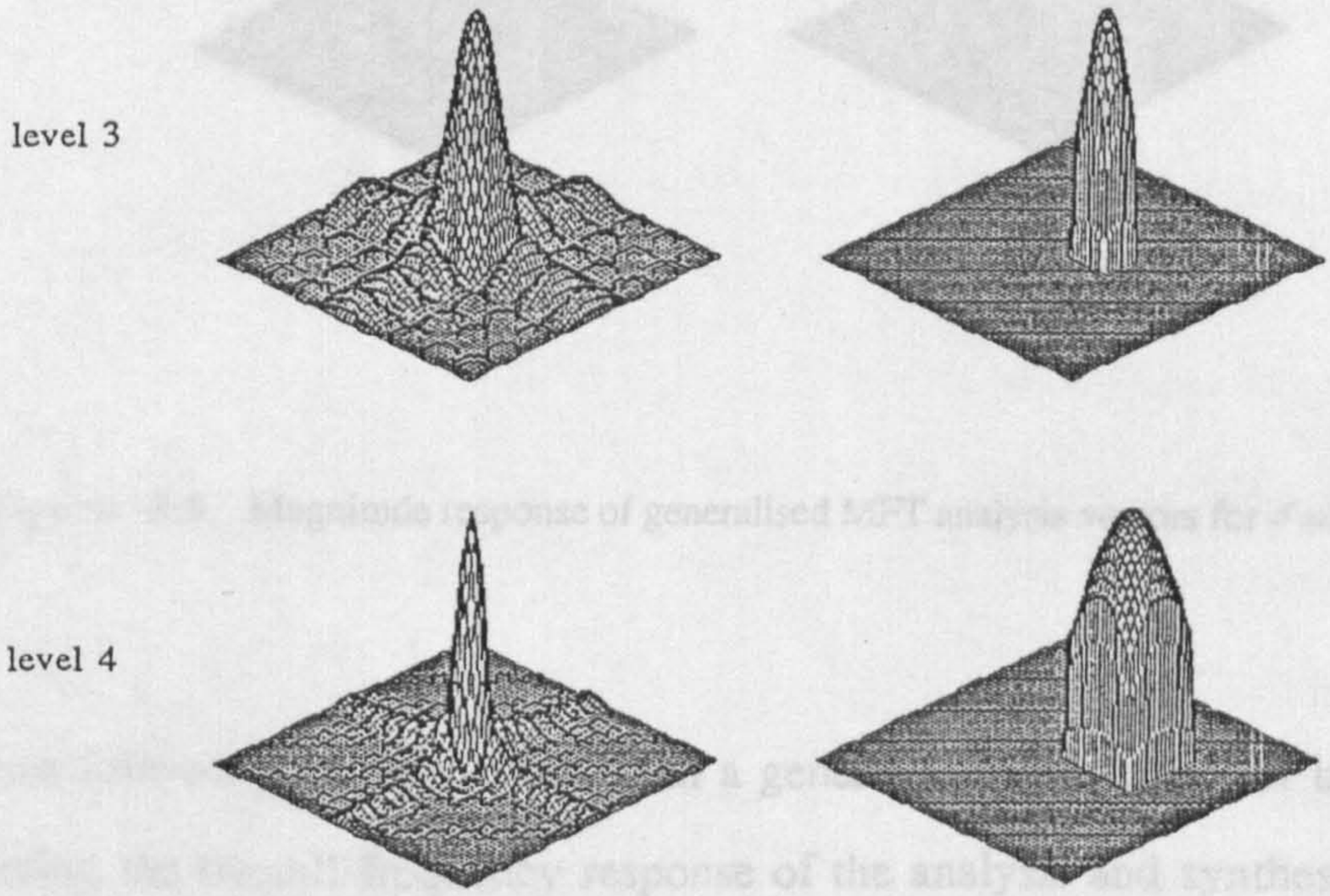


Figure 3.8. Magnitude response of MFT analysis vectors for M=64.

For a generalised version of the MFT, both forward and inverse transforms can be implemented using the same set of vectors $g(\sigma, n)$ given by eqn (3.45). With a relaxation parameter $\sigma = 2$, these vectors are shown in fig 3.9 for $n = 3$ and $n = 4$, where it should be noted that in the spatial domain the vectors have greater energy concentration and less sideband energy than those in fig 3.8.

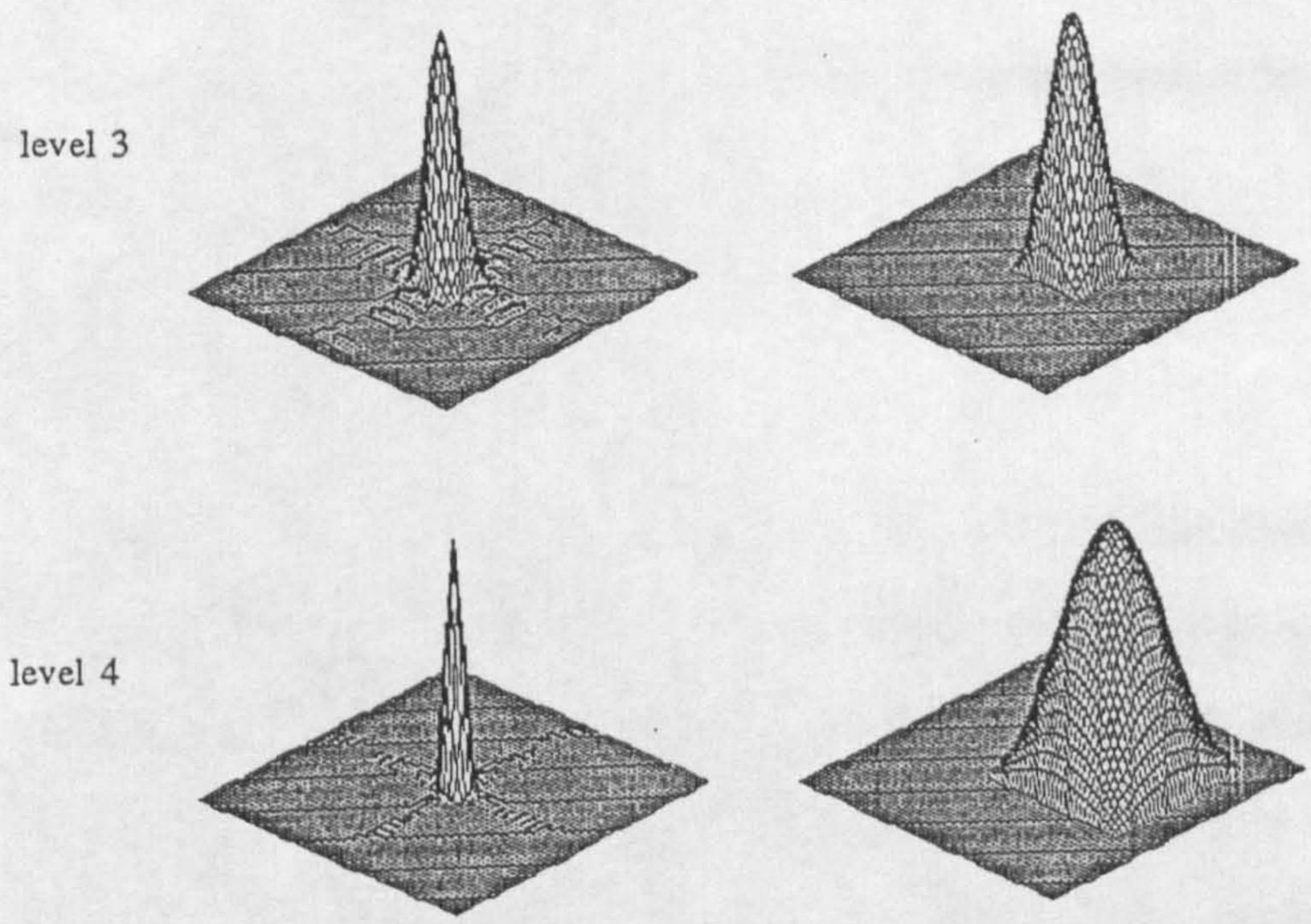
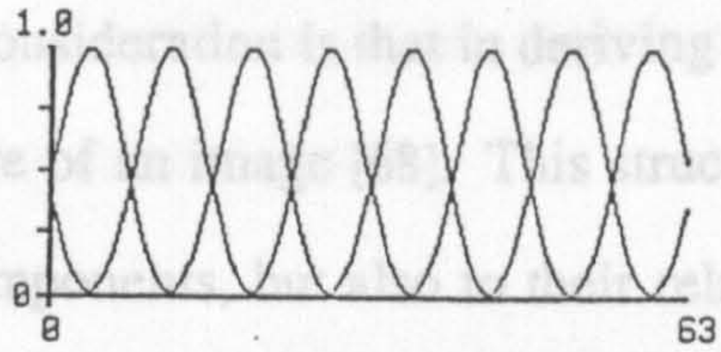


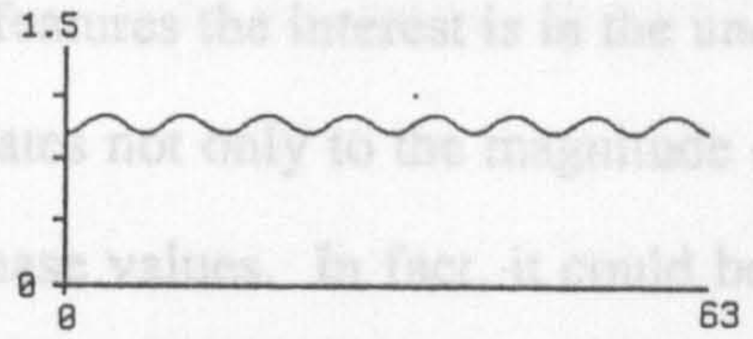
Figure 3.9. Magnitude response of generalised MFT analysis vectors for σ=2 and M=64.

The error following reconstruction from a generalised transform can be assessed by considering the overall frequency response of the analysis and synthesis filter banks (section 3.5.3). Fig 3.10a shows the set of analysis-synthesis frequency response products corresponding to each channel of the filter banks for level 3 of a 1-d generalised MFT with $\sigma = 2$. The summation of these is shown in fig 3.10b, where the maximum deviation from unity is 0.08 and results in negligible error following reconstruction.



(a)

Analysis-synthesis frequency response products for generalised MFT with σ =2 and M=64.



(b)

Summation of responses in (a).

Figure 3.10. Combined analysis-synthesis frequency response for generalised MFT.

3.6.4. Prewhitening

As noted in section 3.4.2, the local spectra of the MFT constitute biased estimates of the 'true' spectrum. Prewhitening is a technique which is often employed in spectrum estimation methods to reduce this bias [64][101] and can therefore be used to further improve the MFT estimates. However, before defining the type of prewhitening used in this work, it is worth noting some additional reasons for considering its use.

The first thing to note is that in general natural images contain significant low frequency energy. An examination of the spectral envelope of such an image will reveal that the majority of the energy is centred about the dc value. However, the interesting image features, such as lines, edges and textural properties, correspond to the higher frequency ranges of the spectrum. Thus, in order to reduce any bias that the predominant low frequency energy might cause in the analysis of these features, it would be beneficial to "even-out" the magnitude values of the frequency components, ie to employ some form of prewhitening.

A further consideration is that in deriving image features the interest is in the underlying structure of an image [68]. This structure relates not only to the magnitude of the spectral components, but also to their relative phase values. In fact, it could be suggested that the latter has greater importance, a view supported by Oppenheim and Lim in their work on the importance of phase in images [82]. The work demonstrated that the magnitude distribution of an image spectrum is not crucial to the structure of an image, whereas the relative phase values appear to be highly significant. This suggests that any analysis should place greater importance on the relationship between phase values (cf chapters 4 and 5) and should not be unduly biased by the magnitude distribution. However, this should not mean that magnitude values are totally ignored, since small random fluctuations (noise) in the spectrum will have correspondingly random phase values.

Prewhitening of spectra is widely adopted in spectrum estimation with varying degrees of complexity [64][101]. A simple approach is adopted here. It takes the form of weighting the image spectrum by a symmetric pre-emphasis function

$$w(\rho) = \begin{cases} 1 - \cos^{a}\left(\frac{\rho\pi}{2A}\right) & \rho < A \\ 1 & \text{else} \end{cases} \tag{3.59}$$

where ρ is the radial frequency in radians and A and a define the degree of pre-emphasis. The operation is therefore equivalent to highpass filtering the image using a filter which has a raised cosine roll-off and is consequently a very simple attempt to prewhiten the spectrum. However, it is straightforward to implement and has been found to be effective for the work described later in this thesis (cf chapter 6). The selection of parameter values will obviously depend upon the particular image

involved, whether any noise energy is known a priori to be present and the resolution of any subsequent processing.

3.7. Examples and Preliminary Experiments

3.7.1. Synthetic Images

Examples of the 2-d MFT for two synthetic images are presented. Each is designed to illustrate in a simple way the properties and potential of the transform. The images are monochrome and of dimension 512×512 pixels with an 8-bit grey level at each pixel. Complex magnitude values of the MFT coefficients for a given level are displayed on a discrete lattice of size $\Omega_n \times \Omega_n$, where at each point the frequency coefficients $u_{xykl}(n)$, $0 \le k,l < \Gamma_n$, are shown as a 2-d block. A grid is placed on the display to indicate the individual blocks. This display technique corresponds closely to the local spectrum interpretation of the 2-d MFT (cf section 3.2).

The first example illustrates the ability of the transform to decorrelate image data in the spatial/spatial frequency plane. The synthetic image is shown in fig 3.11a and consists of three regions: an oriented texture; a 2-d sinusoid; and a random texture. The oriented texture is an impulse noise field filtered by an oriented Gaussian filter

$$v(k,l) = g(k,l) * n(k,l) \tag{3.60}$$

where

$$g(k,l) = e^{-(ak^2+bkl+cl^2)} \tag{3.61}$$

and

$$n(k,l) = \begin{cases} 1 & \xi(k,l) \geq 0.5 \\ 0 & \xi(k,l) < 0.5 \end{cases} \tag{3.62}$$

and $\xi(k,l)$ is a normally distributed random variable with variance 1.

The DFT of the image is shown in fig 3.11b and levels 4 and 5 of its MFT are shown in figs 3.11c and 3.11d. It can be seen that the MFT representations are intermediate between the extremes of the image and its DFT. In the original image, no information is apparent concerning the properties of the different regions; whereas in the DFT, this property information is available (eg note the spatial frequency impulses corresponding to the sinusoid), although no indication is given about the location of the regions. The MFT levels, however, provide both property and positional information at two different resolutions, level 4 having greater spatial frequency resolution and level 5 greater spatial resolution.

The second example demonstrates the multiresolution property of the transform. Fig 3.12a shows an image which consists of different sized circular discs of fixed luminance value centred nonuniformly over the image plane. This image was prewhitened using the function described in section (3.6.4) with $a = 2$ and $A = \pi/16$, and then levels 3, 4, and 6 of its MFT were calculated. These are shown in the remaining plates of fig 3.12. Note that at levels with high spatial resolution, the boundaries of the circles are represented by energy concentrated in an orthogonal orientation within the local spectra of the MFT. In contrast, on levels with high frequency resolution, a given circle is represented within a single spectrum by symmetrically distributed energy.

3.7.2. Natural Images

The MFT for the 'girl' image shown in fig 3.13a is presented. The image was prewhitened using the function in eqn (3.59) with $a = 2$ and $A = \pi/16$. Levels 3-5 of the generalised MFT with $\sigma = 2$ for a 256×256 pixel version of this image are shown in figs 3.13b-d. A 512×512 pixel version is used in the following chapters of this thesis, however the number of MFT coefficients is then 1024×1024 which exceeds the display capacity of the equipment used and therefore cannot be shown here.

In addition to the properties already illustrated for the synthetic images of the previous section, several factors should be noted concerning the results in fig 3.13.

(i) The change in spatial and spatial frequency resolution between the levels of the transform.

(ii) The distribution of energy in the local spectra: oriented and concentrated in the vicinity of line and edge features; uniformly distributed in more complex regions.

(iii) Within certain regions features are adequately represented by local spectra at high spatial frequency resolution (eg the mirror edge), whilst other regions require greater spatial resolution (eg the eyes).

3.7.3. Threshold Coding

A simple experiment to assess the validity of the MFT representation is worth recording. For a subset of transform levels the coefficient magnitudes are thresholded such

that those below the threshold are set to zero while those above remain as before. The transform is then inverted using the multilevel inverse of section 3.3.2 and the result assessed.

This is conveniently represented by a threshold selection operator **C**, similar to those discussed in section 3.3.2, which is defined as (for the 1-d case)

$$c_{iki'k'}(n) = \begin{cases} \delta(i-i')\ \delta(k-k') & |\ u_{ik}(n)\ | \geq t_n \\ 0 & \text{else} \end{cases} \tag{3.63}$$

where t_n is the threshold and is chosen to give a uniform distribution of coefficients over the levels. The multilevel inverse **r** is then given by

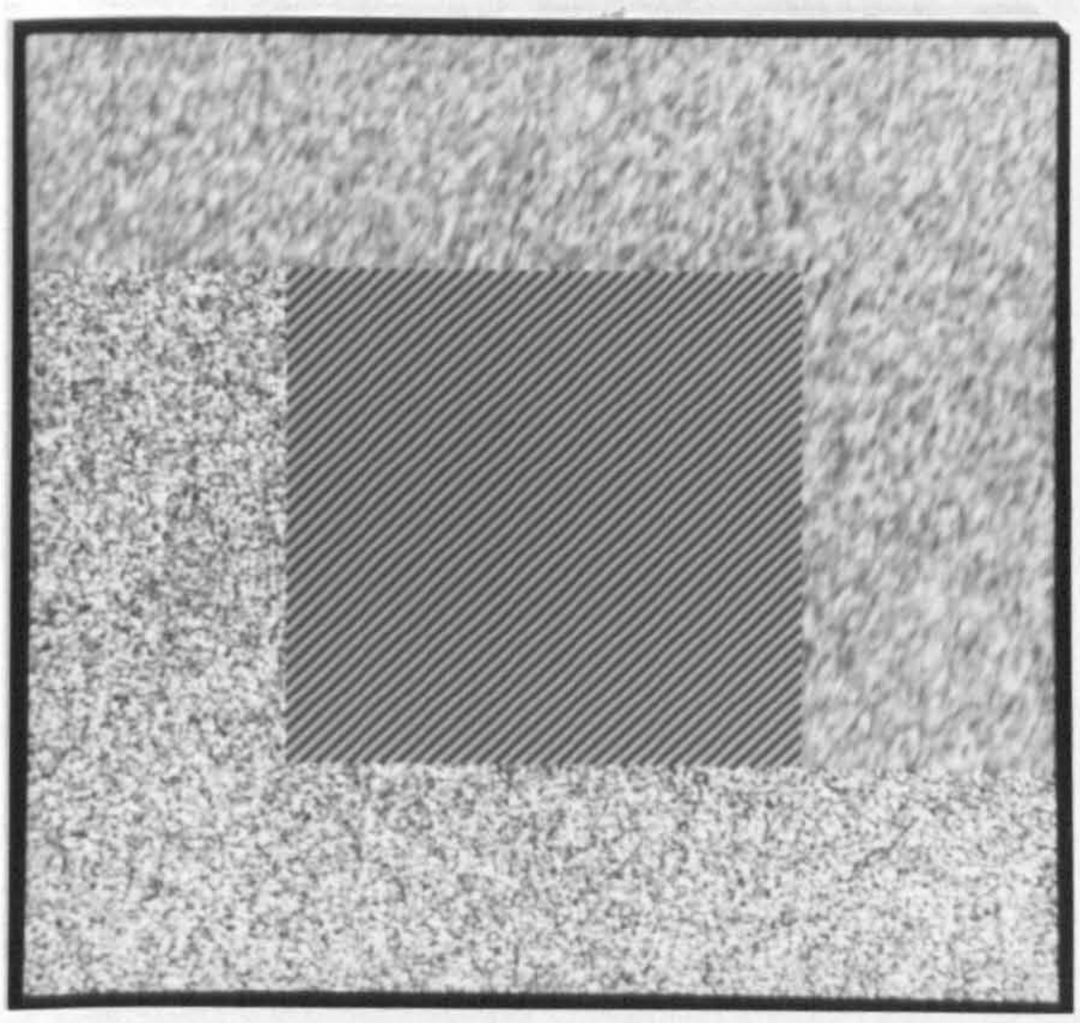
$$\mathbf{r} = \sum_n \mathbf{H}(n)\ \mathbf{C}(n)\ \mathbf{u}(n) \tag{3.64}$$

where $\mathbf{H}(n)$ is the inverse operator.

The above was implemented for the MFT ($\sigma = 1$) of a 512×512 pixel version of the 'girl' image. Levels 4-6 were used in the scheme and the total number of coefficients prior to thresholding was $3 \times 512 \times 512$. Figure 3.14 shows the result of two reconstructions in which the number of coefficients after thresholding has been reduced to 7% and 4% respectively.

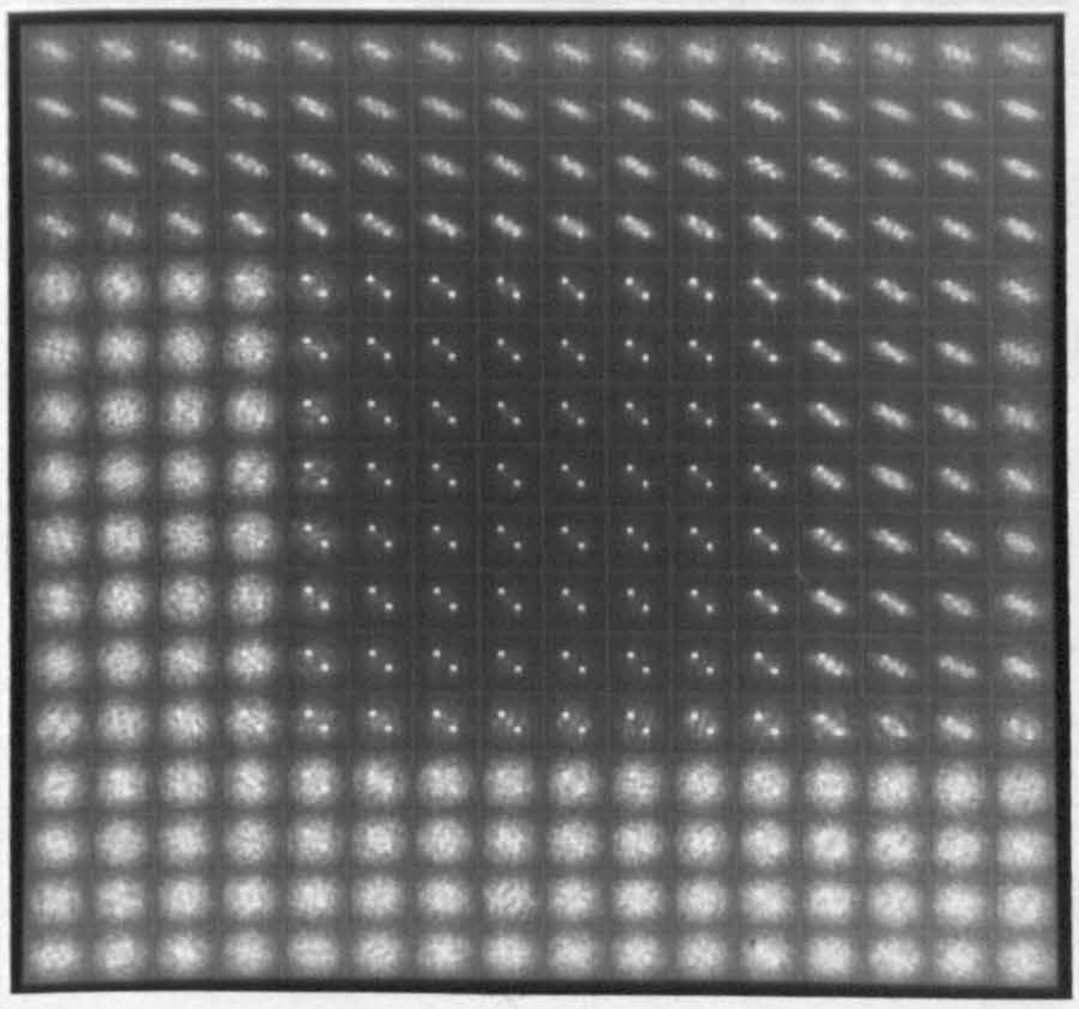
There are two points to note about these results. The first is that even using small numbers of coefficients the reconstructions are of acceptable quality. The main error is aliasing caused by the inversion using a subset of coefficients (cf section 3.3.2) and it is particularly visible in fig 3.14b. This takes the form of ripples in the vicinity of

edge features and if studied carefully can be seen to exist at various frequencies depending upon the level from which it was derived from. The second point to note is that the reconstructed images retain the important features of the original. In particular, notice that the perceptually significant edge features are retained, ensuring that the images have a sharpness quality comparable to that of the original. Thus, although this experiment is very simple and the choice of coefficients is based solely on their magnitude value, it does illustrate that the MFT is capable of emphasising the significant features in an image. This ability underlies the use of the multiscale methods in coding applications [1][70][108].

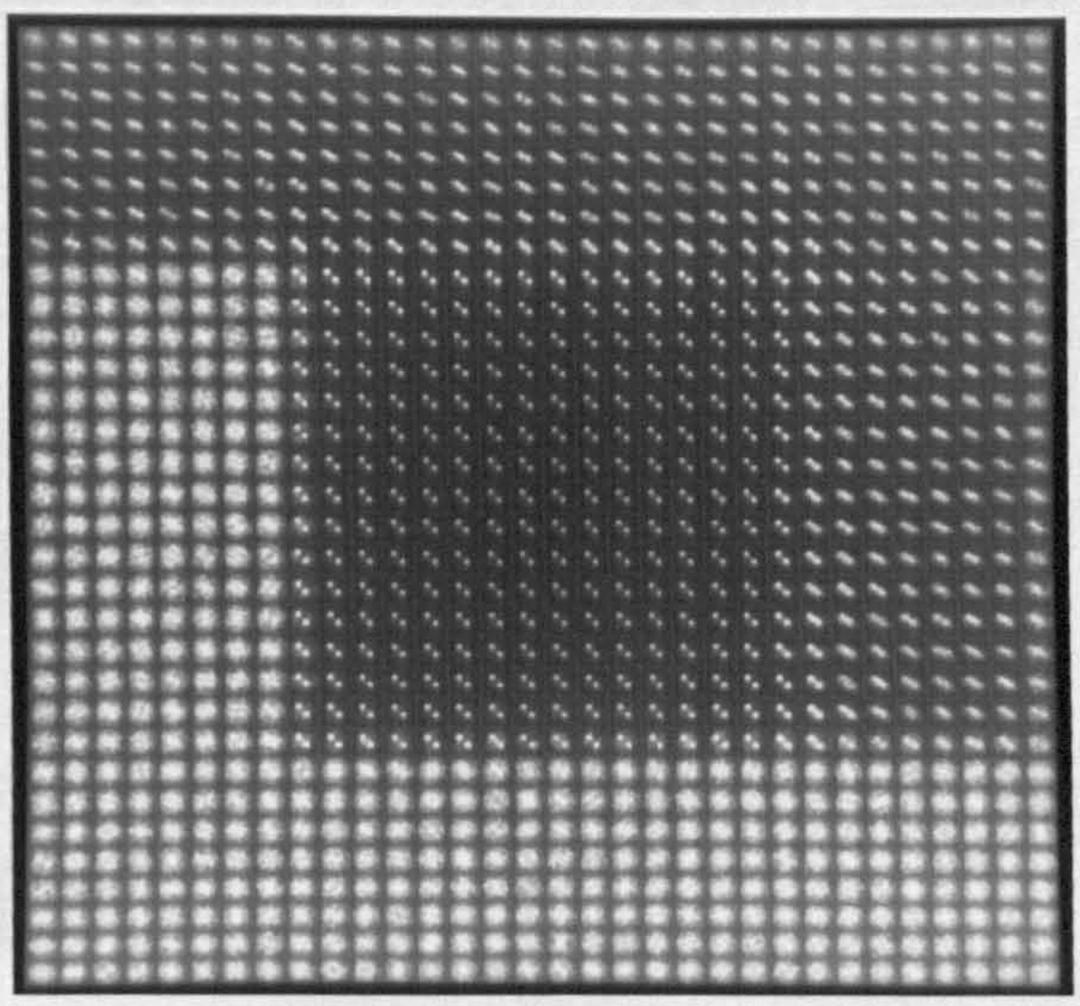


(a) Original image.

(b) Discrete Fourier transform.

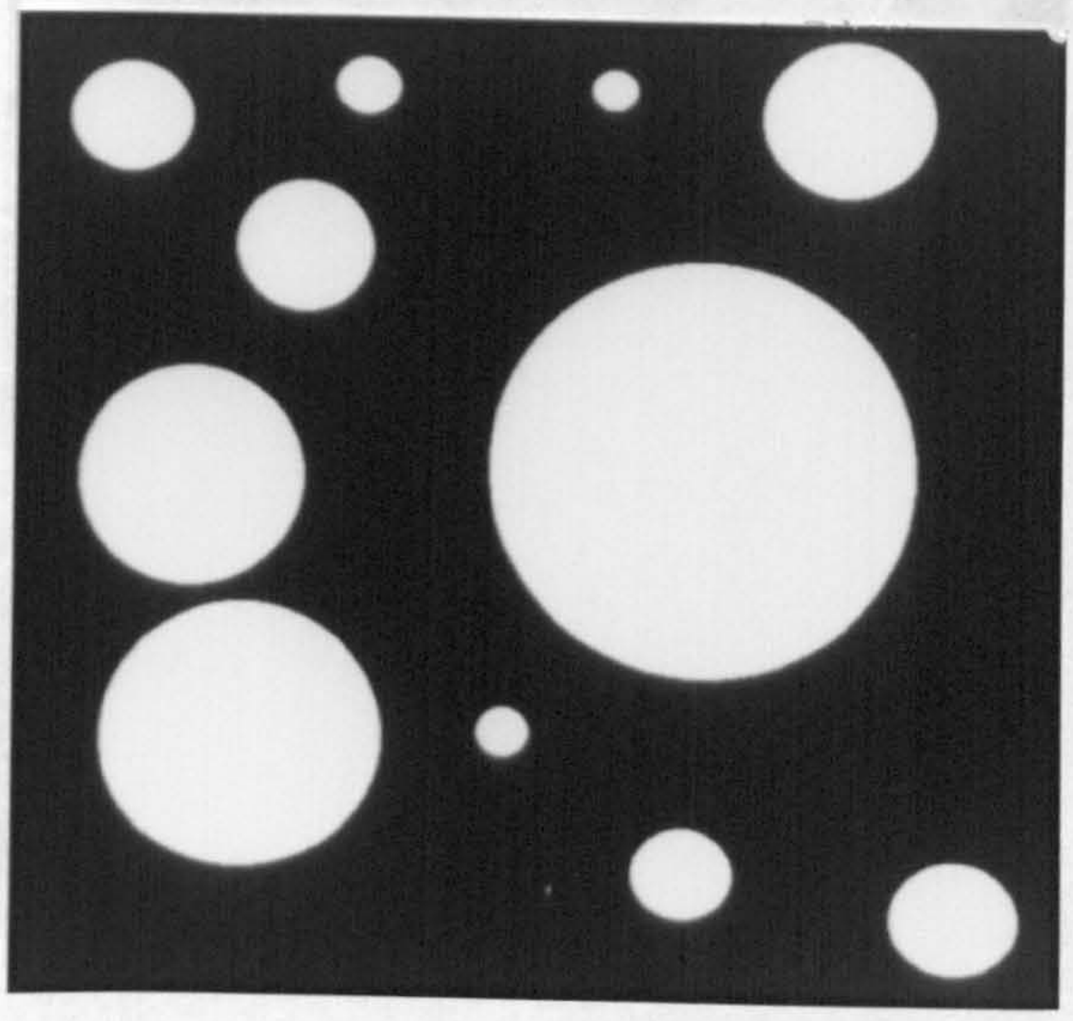


(c) MFT level 4.

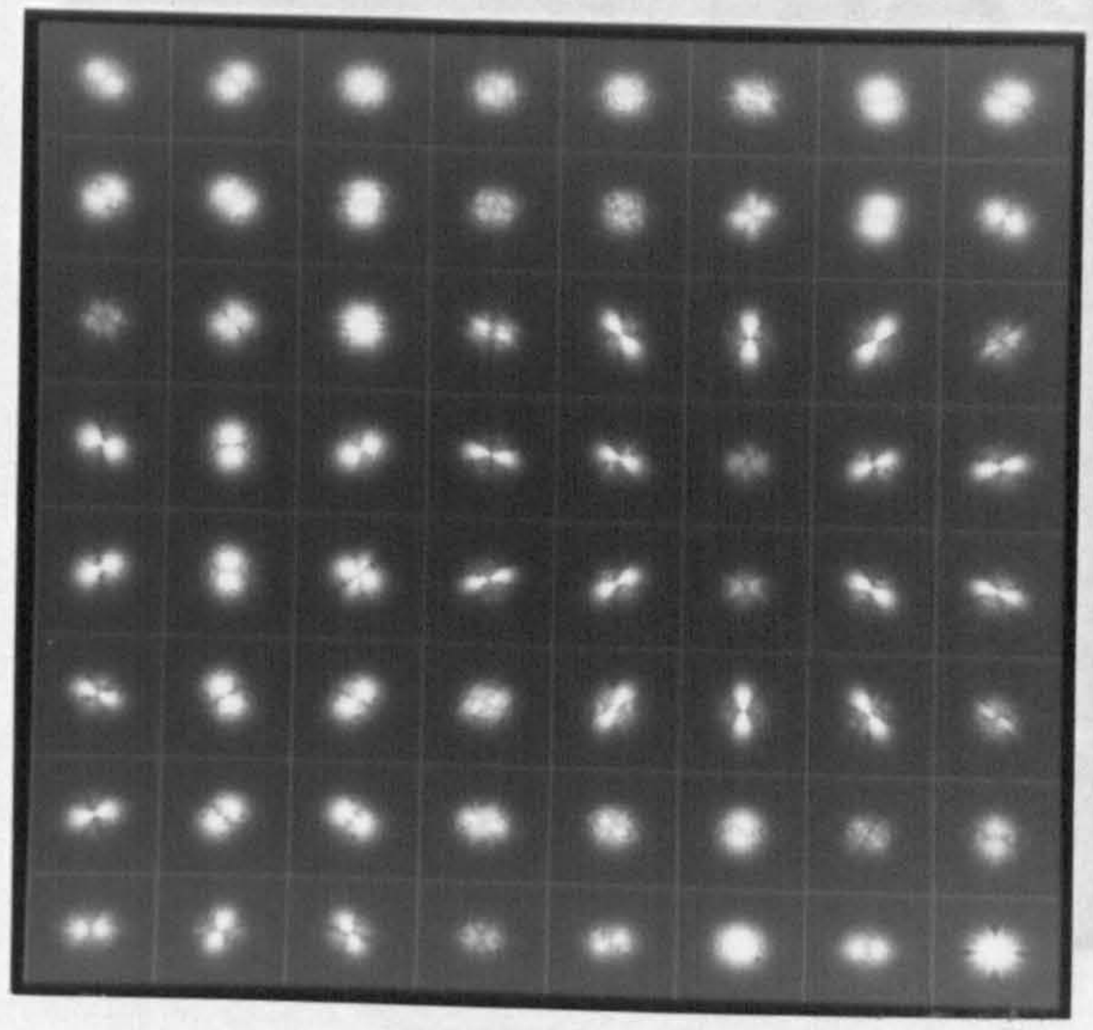


(d) MFT level 5.

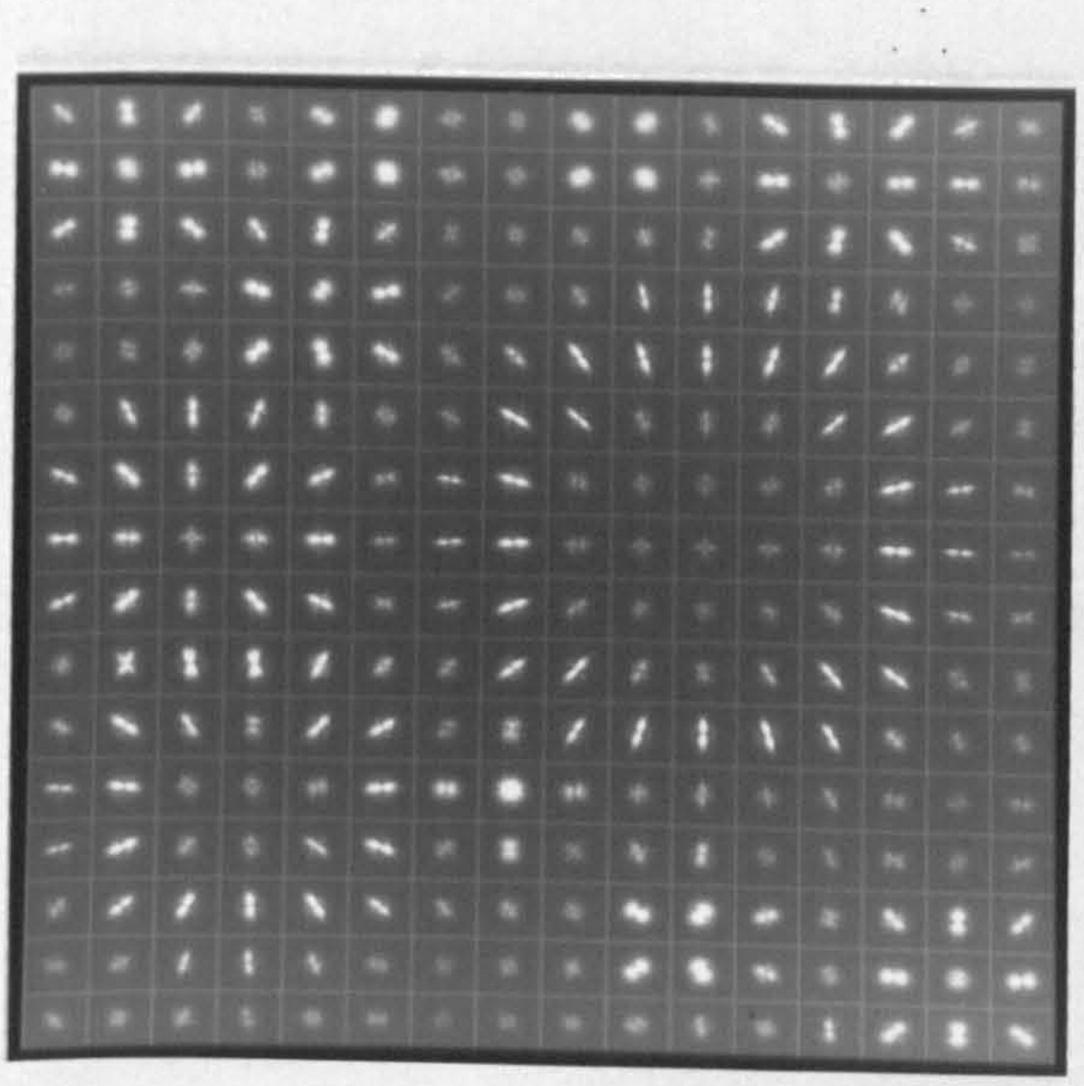
Figure 3.11. MFT examples for 'textures' image.



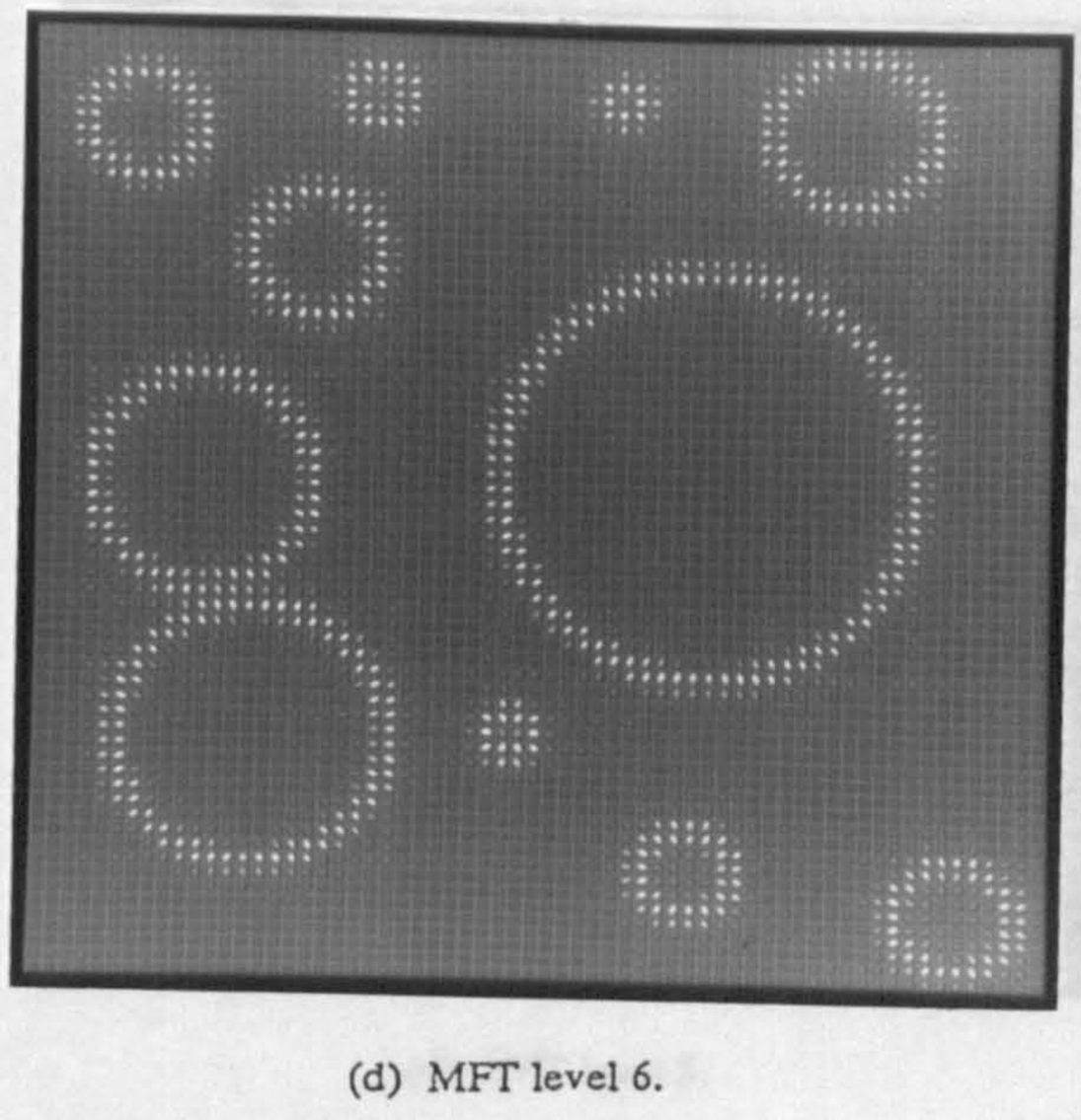
(a) Original image.



(b) MFT level 3.



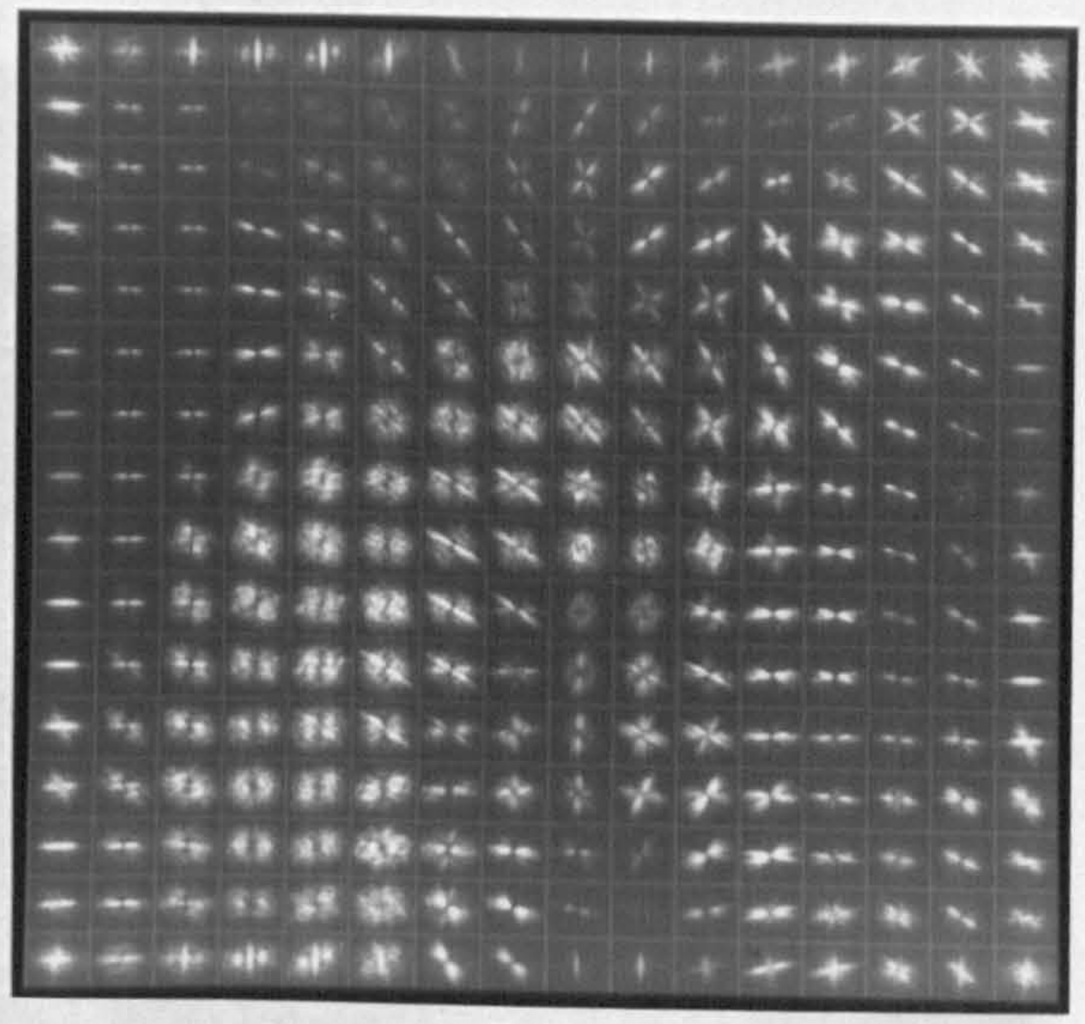
(c) MFT level 4.



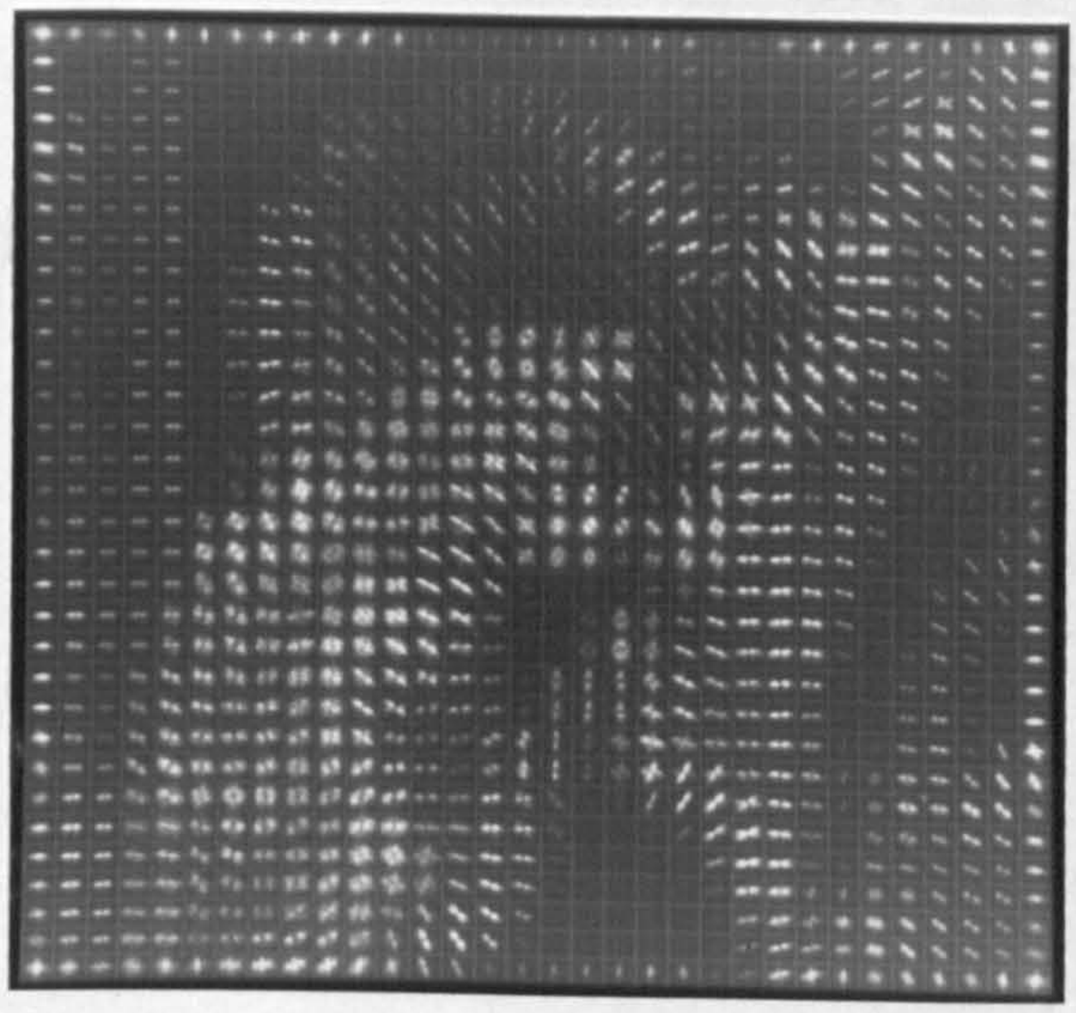
(d) MFT level 6.

Figure 3.12. MFT examples for 'discs' image.

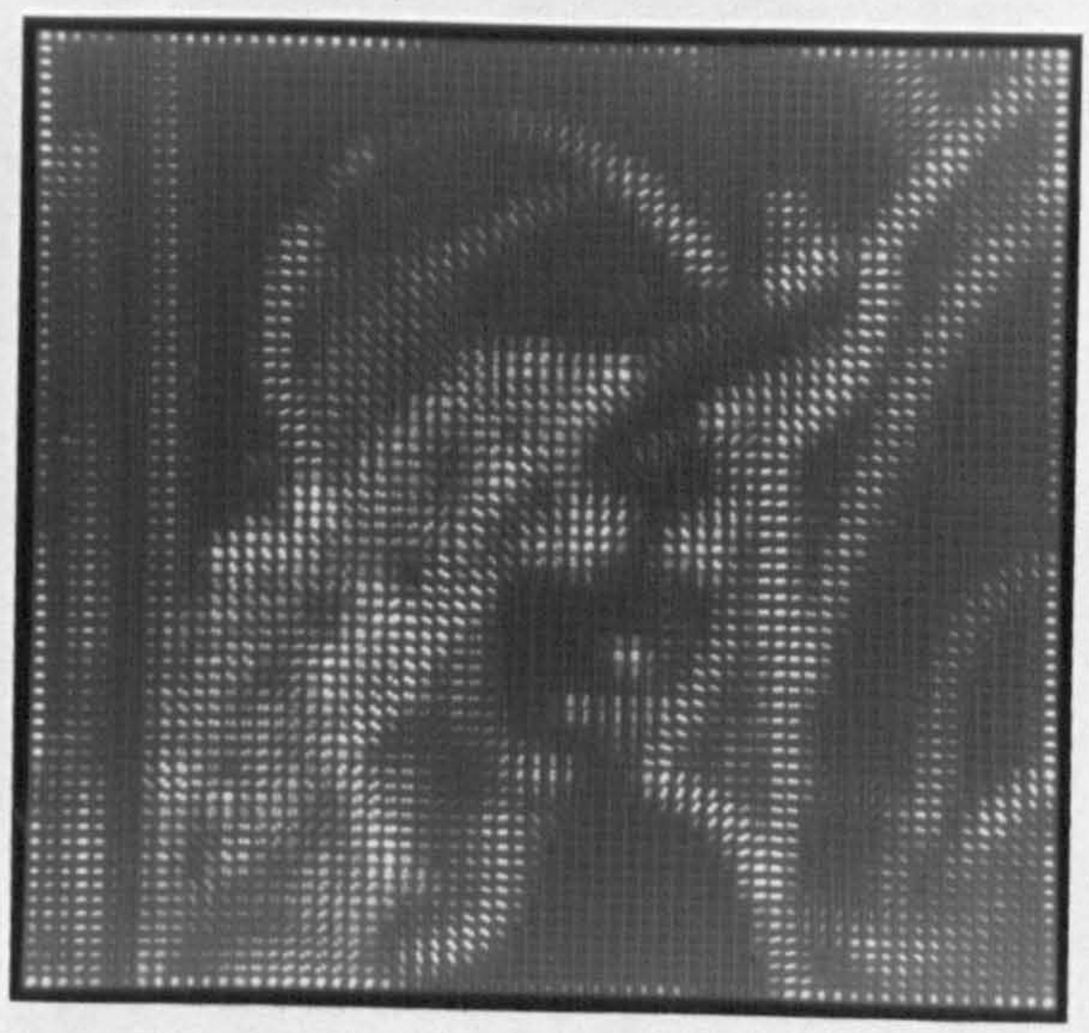
(a) Original image.



(b) MFT level 3.



(c) MFT level 4.



(d) MFT level 5.

Figure 3.13. MFT examples for 'girl' image.

(a) Reconstruction from 7% of coefficients.

(b) Reconstruction from 4% of coefficients.

Figure 3.14. Threshold coding examples for 'girl' image.

CHAPTER FOUR

REPRESENTING LOCAL IMAGE FEATURES

4.1. Introduction

The aim of the remaining chapters in this thesis is to illustrate how the MFT can be used to detect and estimate image features. The approach taken is to consider a specific example, which provides a general framework and methodology for using the transform. As noted in chapter 1, feature descriptions have been mainly applied to the representation of lines and edges and texture. These areas are accepted as being important in a wide range of computer vision and image processing tasks. It is therefore appropriate to use one of these as the example to be considered in this work.

Of the above possibilities, it was felt that it would be more instructive to consider the analysis of local image features such as lines and edges. The reason for this choice was twofold. First, the existing methods in this area are most often based within some form of single resolution or simple pyramid framework - less consideration has been given to taking a more complex multiscale or multiresolution approach. The work will therefore provide a useful insight into a predominantly untried area of feature extraction. On the other hand, there has been considerable work done in the extraction of textural features within a multiscale framework. The work of Spann and Wilson [99][112][113] in this area has shown clearly the advantages that can be gained when applied to segmentation problems. Since the MFT is a generalisation of multiscale techniques (cf section 2.6), it would be possible to extend their ideas and yield similar results. Hence, the advantages for texture analysis have already been shown, it remains to show that similar advantages can be gained when dealing with spatially localised features.

The purpose of this chapter is to consider the various aspects of local feature extraction and to review the existing approaches to the problem. Once this has been done, the problem will then be defined in a multiresolution context, leading to an image model which forms the basis of a detection and estimation scheme described in chapter 5. Finally, the model is shown to be generally applicable and readily extended to more complex features.

4.2. A Review of Existing Methods

4.2.1. General Properties

An examination of a typical image will reveal the existence of regions which are one dimensional, have an associated orientation and contain a boundary between constant luminance levels. Indeed, an image can be considered to consist entirely of such regions, each contiguous to one another and of different size and orientation. These regions are typically known as lines and edges, although a more general term is local image features.

Such features are an important part of image structure and often form part of more global attributes such as boundaries and curves [72]. Their existence can derive from the boundary of an object, indicating its shape and orientation, or they can characterise lighting and reflectance in an image. Indeed, the cartoon drawing, which is based upon emphasising these features and often provides an instantly recognizable image, is an example which illustrates this structural importance. Support for this is also provided by evidence from studies into the workings of the human visual system. Physiologists, prompted by the pioneering work of Hubel and Wiesel [53], have demonstrated the existence of so called 'simple cells' in the striate cortex that respond to

lines and edges at specific orientations [52].

4.2.2. Detection Methods

The importance of local features has been recognised both in computer vision and image processing [5][94][95]. Detection and estimation schemes have received considerable attention, leading to a wide range of approaches which are based upon different underlying models. There are several reviews available in the literature [39][85][94][95].

The aim of these methods is to detect the luminance discontinuity, and in some cases the orientation, associated with a local feature. The various forms include: gradient methods [22][92][93]; the detection of zero crossings in the second derivative [47][73]; the use of edge masks or templates [79]; frequency domain methods [65][96]; various parametric approaches [54][55]; and those based upon statistical models [51].

There are several issues that effect the choice of an appropriate method. These include noise sensitivity, spatial resolution and computational efficiency. The appropriate method is also often dictated by the application. For instance, if the images involved are relatively noiseless and contain well defined edges, then a simple edge mask is likely to be acceptable. However, if the images are more complex or of imperfect quality, then a less noise sensitive and variable resolution method would be more appropriate.

A recent development in this area has been the recognition that the orientation associated with a local feature is an important characteristic. Some of the above mentioned schemes are based only upon detecting luminance discontinuity and this ignores the

two dimensionality of the problem: for a local feature to exist it must have an associated orientation. Indeed, it is this anisotropy that distinguishes the local feature from a random fluctuation in the image. This is supported by the evidence mentioned earlier that simple cells in the striate cortex are not just sensitive to any line or edge, but to these features at specific orientations.

The importance of orientation has been incorporated into the recent advances in this area. Intended for general application, such as the processing of natural images, these new developments have been based upon the use of local operators which are tuned to a finite number of orientations. The results of applying these operators are then combined to give an estimate of the feature strength and its orientation. There are several versions available of this general technique [22][65]. These methods either employ prefiltering [22] or are defined in the frequency domain [65] to minimise noise sensitivity. They are also applied over a range of spatial resolutions to detect both the fine detail and broad edges in an image.

4.2.3. Curves and Boundaries

Once local features have been detected using an appropriate scheme, the question then arises as to whether it is possible to combine a suitable subset of them into a single entity, such as a curve or object boundary. Indeed, isolated features by themselves have limited use in an analysis. This problem has been considered in the literature and several reviews are available [5][95]. Methods vary according to the amount of prior knowledge that is assumed about the boundaries in the image. This knowledge may consist of local constraints such as a maximum curvature value or more global constraints such as specific shape identification. However, the most important issue in all these methods is that they are dependent to a greater or lesser extent upon the

initial local feature detection.

The methods fall broadly into two classes. The first is edge following, in which features that are in close proximity are joined up according to some boundary acceptance function and so form a list of features that follow the boundary. The acceptance function can take various forms, ranging from a simple shortest distance measure to a more complex curvature measure. These methods necessarily involve some type of search process and this has been done using heuristic methods [3][23][74] and dynamic programming [28][78]. The advantage of these methods is that they are generally applicable since only local boundary constraints are imposed. They can also be reasonably efficient in terms of computation. However, they are critically dependent upon the feature detection input, since if an edge is to be followed then all the relevant sections must be present. Boundary gaps or complicated fine detail areas will cause problems.

The second class consists of methods which are based around the Hough transform [5][57]. The basic idea is to determine some parametric representation of the boundary, create a discrete accumulator array with the parameters as coordinates and then for each local feature increment the appropriate components in the accumulator array that correspond to boundaries that 'contain' the feature. Local maxima in the accumulator array then indicate the boundaries present in the image. These methods have been considered extensively by many workers, and the variations range from simple straight line detectors [41] to more general shape detection [4]. The advantage of these methods is that they are less dependent upon the detected features since the process is global and not effected by local errors such as gaps. When a specific shape is required to be detected in an image these methods are appropriate. However, there are a number of computational problems and also a requirement for a sufficiently large number of detected features [5].

4.2.4. Comments on Existing Methods

The above review of existing methods is not intended to provide an exhaustive reference, but is an attempt to illustrate the various aspects involved in representing and subsequently detecting local image features. The main conclusion that can be drawn is that the area has received considerable attention over a period spanning almost 25 years and as yet no unified approach has emerged. The various schemes and even recent developments, although achieving considerable success, have remained essentially ad hoc with no regard for any subsequent analysis. A good example of this is illustrated in the case of curve and boundary identification considered in the previous section. Although methods of feature detection have improved, curve and boundary identification has remained a separate process, implemented within an entirely different framework. This approach completely ignores the inherent relationship between the two feature types. In the next section it is shown that this need not be the case and that a unified approach to both feature detection and curve identification can be adopted by making use of multiresolution techniques.

4.3. Adopting a Multiresolution Approach

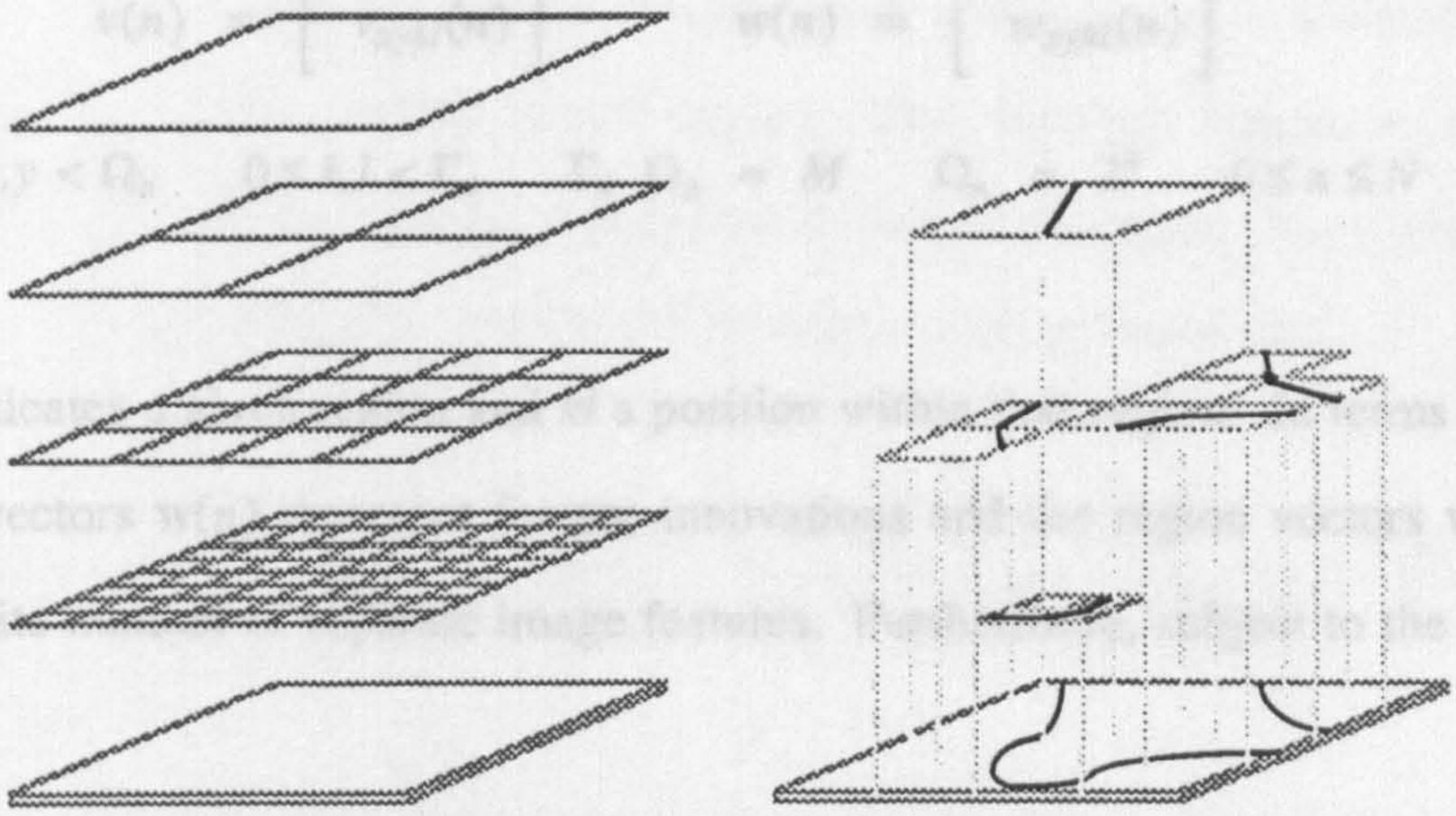
4.3.1. An Image Model

The important property of local features is that they are characterised by regions which can be assumed to be essentially one dimensional, each containing a luminance boundary and having an associated orientation. It is this property that underlies the recent developments in detection methods [22][65]. A multiresolution approach to the problem can also be based upon this property, and provide a number of advantages.

The use of any detection methods implies the existence of some underlying image model. In the models mentioned above, the image is assumed to consist of the 1-d regions that characterise local features. A general model would allow such regions to be of any shape and size, although in practice a regular structure is imposed upon the model. The local operators used in the detection process define the size and shape of the regions within the model. A certain amount of variation is introduced by using operators at different spatial sizes, although this has yet to be incorporated into any coherent framework.

It is possible however to take a more general approach by basing the detection process upon a multiresolution image model. This has a hierarchical structure and represents an image by local features defined at different spatial resolutions. A simple example, and the one that is adopted in this work, is to assume that the image consists of contiguous square regions, each corresponding to a local feature with a particular orientation. The size of these regions can range from a single pixel to the complete image, as shown in fig 4.1a, and a typical example of the model is illustrated in fig 4.1b for a simple line drawing. Note that the fine detail in the image is represented by short line segments corresponding to small regions, whereas large isolated lines are represented by larger regions at a decreased spatial resolution.

Although the above model does not enable any variation in region shape, it does incorporate a well defined range of spatial scalings of the prototype square shape and this will be shown to have considerable generality for representing features. It also has the advantage that it is a simple structure which is easy to understand and this has important implications when considering an appropriate detection process. However, before it can be utilised in such a scheme, it must first be considered in a mathematical form. This is the subject of the next section.



(a) Spatial regions of model (b) Typical example

Figure 4.1. Multiresolution image model.

4.3.2. Linear Recursive Form of the Model

The model described in the previous section is a simple example of a general class of multiresolution image models [116]. These have a linear structure and are defined by the following recursive operation.

$$\mathbf{v}(n) \;=\; \mathbf{A}(n)\;\mathbf{v}(n-1) \;+\; \mathbf{B}(n)\;\mathbf{w}(n) \qquad 0 < n \leq N \tag{4.1}$$

where the vectors $\mathbf{v}(n)$ and $\mathbf{w}(n)$ represent 2-d arrays of size $M \times M$, $M = 2^N$. These arrays are divided into $\Omega_n \times \Omega_n$ contiguous regions, each of size $\Gamma_n \times \Gamma_n$, and arranged as in fig 4.1a. The number of regions is defined by the scale parameter n and

the vectors are indexed by four indices to indicate the region structure

$$\mathbf{v}(n) = \left[v_{xykl}(n) \right] \qquad \mathbf{w}(n) = \left[w_{xykl}(n) \right] \tag{4.2}$$

$$0 \le x,y < \Omega_n \qquad 0 \le k,l < \Gamma_n \qquad \Gamma_n \, \Omega_n = M \qquad \Omega_n = 2^n \qquad 0 \le n \le N$$

where xy indicates a given region and kl a position within that region. In terms of the model, the vectors $\mathbf{w}(n)$ represent feature innovations and the region vectors $\mathbf{w}_{xy}(n)$ contain a finite number of separate image features. Furthermore, subject to the initial condition

$$\mathbf{v}(0) = \mathbf{B}(0) \, \mathbf{w}(0) \tag{4.3}$$

the resulting image $v\,(i,j)$ is given by

$$v\,(i,j) = v_{ij00}(N) \tag{4.4}$$

The purpose of the operators $\mathbf{A}(n)$ and $\mathbf{B}(n)$ in eqn (4.1) is to 'construct' the image by selecting appropriate regions from both the previous levels via $\mathbf{v}(n-1)$ and from the innovation levels $\mathbf{w}(n)$. This general model therefore enables the image to be represented by features that are defined over the complete range of resolutions, from per pixel to global definition.

To obtain the simple model adopted in this present work it is necessary to impose two additional constraints upon this general model. First, the innovation vectors $\mathbf{w}_{xy}(n)$ are restricted to a class that contains single local features

$$w_{xykl}(n) \;=\; h_{nxy}(k,l) \;+\; g\,(\,x\Gamma_n + k,\; y\Gamma_n + l\,) \qquad (4.5)$$

where $h_{nxy}(k,l)$ is some locally defined real function which has an associated orientation θ_{nxy} and centroid position vector $\eta_{xy}(n)$. This function represents the local feature and is confined within the region boundaries, ie the vector $\eta_{xy}(n)$ is defined with respect to the centre of the region and its components are such that

$$\eta_{xy}(n) \;=\; [\;\eta_{xy0}(n)\,,\,\eta_{xy1}(n)\;] \qquad -\frac{\Gamma_n+1}{2} \le \eta_{xy0}(n)\,,\,\eta_{xy1}(n) \le \frac{\Gamma_n+1}{2} \qquad (4.6)$$

as illustrated in fig 4.2. The function $g(x,y)$ in eqn (4.5) is a smooth real function which is globally defined and represents a lowpass version of the image.

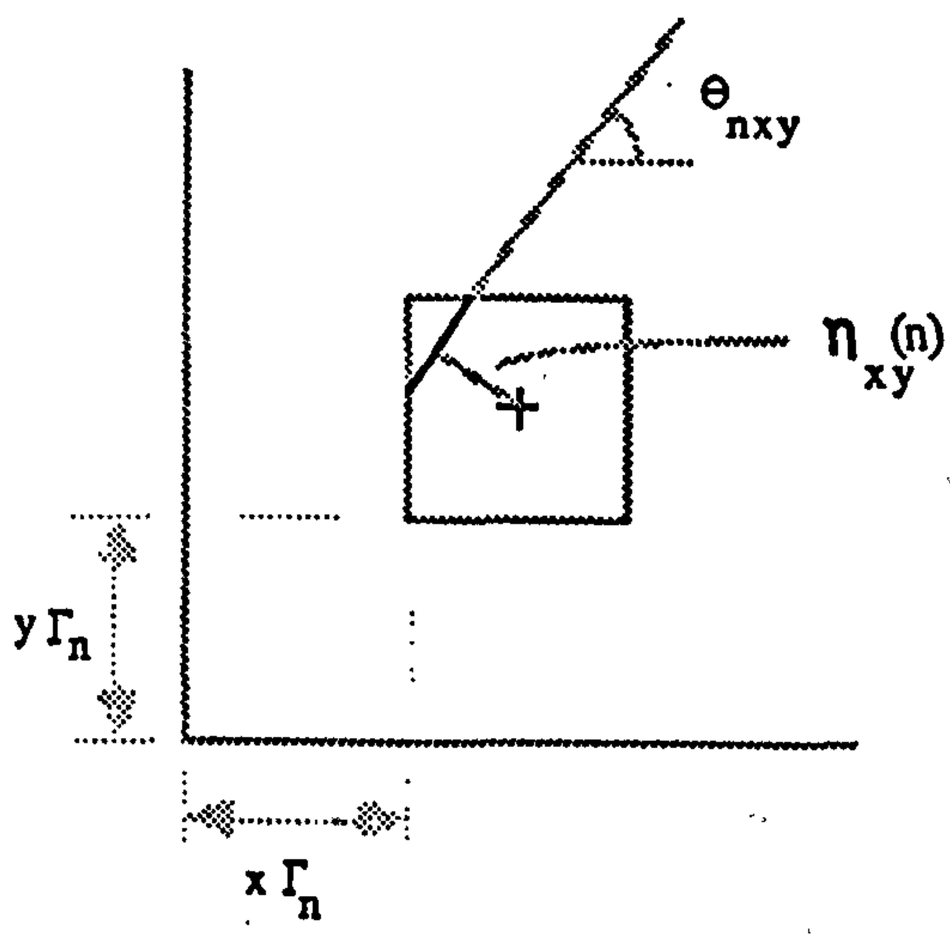


Figure 4.2. Parameters of function $h_{nxy}(k,l)$.

Secondly, the hierarchical structure of the general model is limited by ensuring that a given image region is represented exclusively by a single innovation vector at an appropriate scale. In other words, from eqn (4.1), the operators $\mathbf{A}(n)$ and $\mathbf{B}(n)$ are

defined such that a vector $\mathbf{v}_{xy}(n)$ is either equal to a new feature vector

$$\mathbf{v}_{xy}(n) = \mathbf{w}_{xy}(n) \tag{4.7}$$

or is equal to a quadrant of the relevant vector on the previous level

$$v_{xykl}(n) = v_{rspq}(n-1) \qquad 0 \le k,l < \Gamma_n \tag{4.8}$$

$$r = \lceil \frac{x}{2} \rceil \quad s = \lceil \frac{y}{2} \rceil \quad p = (x-2r)\Gamma_n + k \quad q = (y-2s)\Gamma_n + l$$

where the notation $\lceil x/2 \rceil$ indicates that $x/2$ is truncated to the nearest integer. The above criterion ensures that the model structure resembles the example in fig 4.1b.

4.3.3. Curve Representation

The image model considered in the previous two sections provides a straightforward way of representing curves in an image. It is a general approach and avoids a number of problems associated with traditional methods.

Curves are represented in a piecewise manner using the local features in the model. Each feature is assumed to represent a section of the curve and the change in orientation between adjacent sections provides a local curvature measure. An example is illustrated in fig 4.3. Note that the regions associated with the local features, and thus the curve sections, vary in size according to the curvature value: small regions corresponding to high curvature and vice versa. Features are formed into a curve representation providing they satisfy a maximum curvature and 'best fit' criterion.

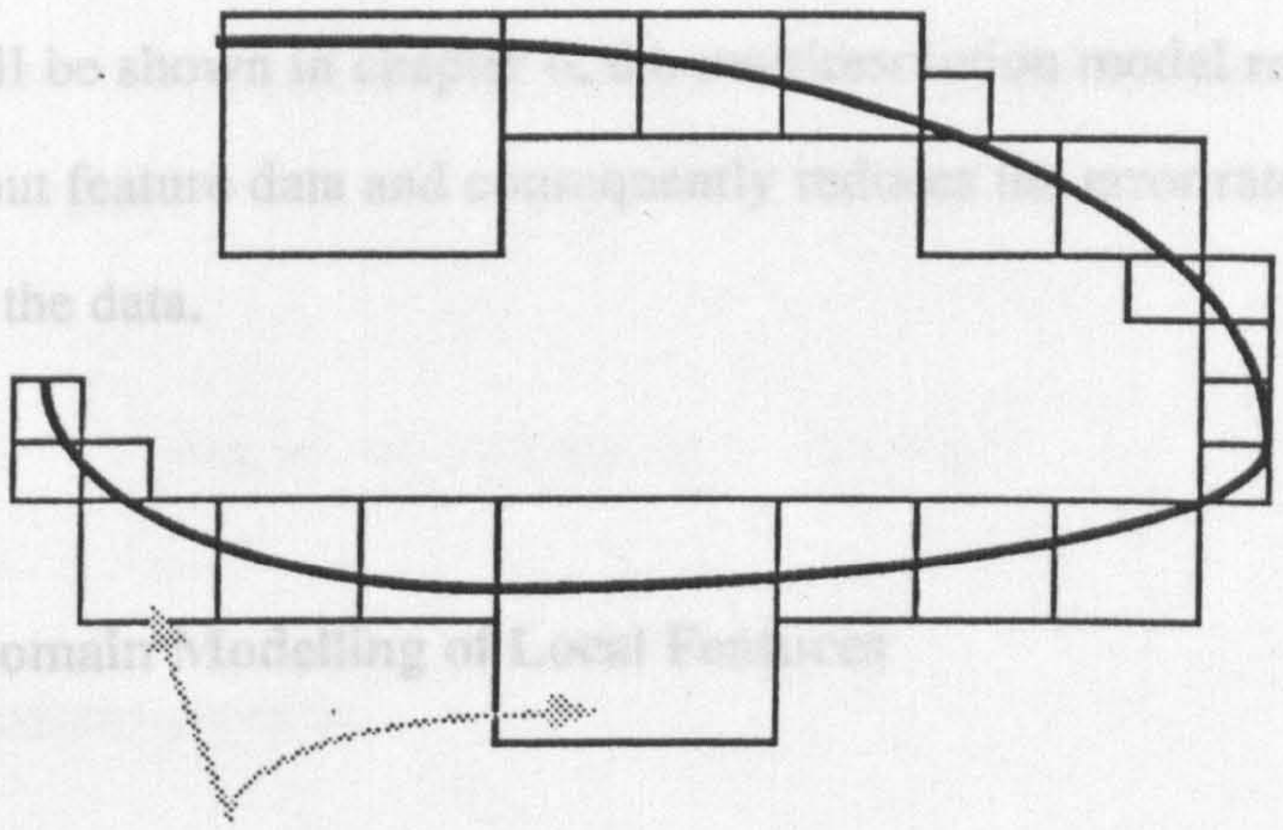


Figure 4.3. Curve representation in the multiresolution image model.

This is discussed in section 5.5, where a hierarchical curve extraction scheme is described.

The above method of representing curves has a number of advantages. It requires only a local curvature criterion to be met, but at the same time it leads to an extraction process that avoids several problems associated with other general methods such as edge following. The first is that the input features are defined over a local region and not on a per pixel basis, reducing the chance of local errors 'breaking' the curve. A further advantage, and a more important one, is that the hierarchical structure of the model in which the curve is represented means that the extraction can be implemented in a hierarchical manner, regarding sections of the curve in a fine-to-coarse analysis as opposed to 'following' them sequentially. This not only reduces the searching involved in the extraction process, but also provides a capability to fill in gaps on a local basis, an ability which has traditionally only been associated with parametric

methods such as the Hough transform.

In short, and as will be shown in chapter 6, the multiresolution model removes redundancy from the input feature data and consequently reduces the error rate when curves are extracted from the data.

4.4. Frequency Domain Modelling of Local Features

4.4.1. Motivation

The model described in the previous section is based upon regions which contain local features. These are represented in eqn (4.5) by a locally defined function $h_{nxy}(k,l)$. The purpose of this section is to describe a suitable model for this function.

As discussed in section 4.2.2, a number of techniques exist for modelling local features, including both spatial and frequency domain methods. In the present work, a model based upon the latter is adopted. These methods have a well defined representation of local orientation and can be defined to minimise noise sensitivity [65][96].

For simplicity, the model is introduced in the context of a continuous spatial domain with ideal oriented local features and ignoring any constraint imposed by the limited size of the regions. This general approach is then extended to account for the locality within the image model and to enable the existence of local feature segments within the regions.

4.4.2. Continuous Case

An ideal local feature in the continuous 2-d spatial domain can be represented by the following oriented region

$$v(x,y) = v(x\cos\theta + y\sin\theta) \tag{4.9}$$

and its Fourier transform given by

$$V(\omega_x,\omega_y) = \delta(\omega_x\sin\theta - \omega_y\cos\theta)\, V(\omega_x\cos\theta + \omega_y\sin\theta) \tag{4.10}$$

where $V(\omega)$ is some complex function in one dimension and $V(\omega_x,\omega_y)$ is confined to a line which is perpendicular to the orientation of the feature in the spatial domain. Simple examples of such regions are line and edge features, where the $v(x)$ in eqn (4.9) are ideal rectangular pulses and step functions respectively.

However, these are not the only features which can be defined in such a way. For example, an oriented texture also falls into the same class and has an equally concentrated spectrum. To differentiate between these and the local features such as lines and edges, it is necessary to consider the nature of the complex function $V(\omega)$ in eqn (4.10).

Local features give rise to an essentially linear component in the phase of this function. Specifically, if η is defined as the centroid of $v(x)$, then the argument of $V(\omega)$ is given by (see Papoulis [83])

$$\text{Arg}[\,V(\omega)\,] = \omega\,\eta + \varepsilon \tag{4.11}$$

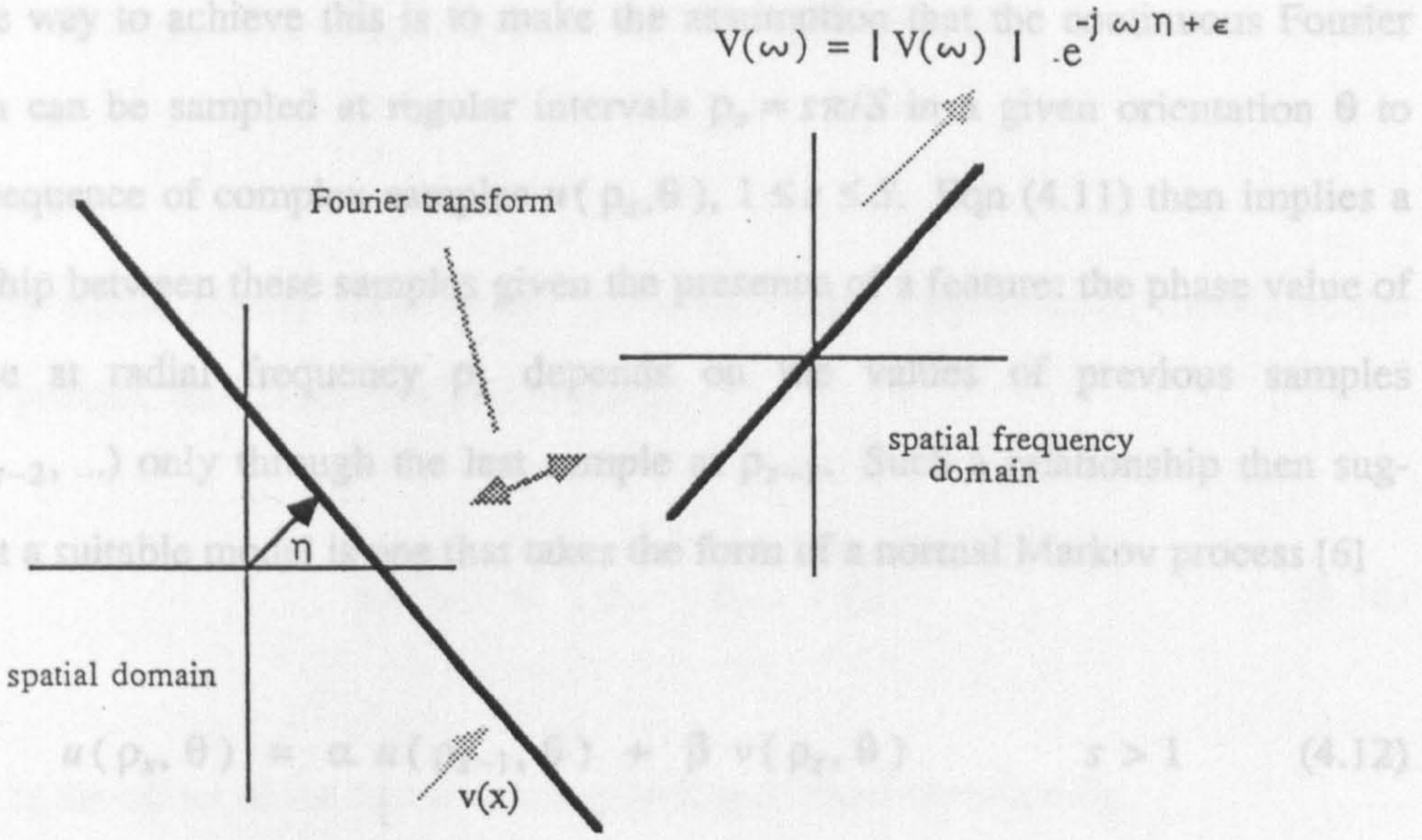


Figure 4.4. Linear phase property of ideal local feature.

where ε is a phase constant. In other words, the offset of these features is directly proportional to the phase variation of the spectrum in an orthogonal orientation as illustrated in fig 4.4. In direct contrast, fields such as an oriented texture do not possess this property, the phase variation in these cases tending to be of a random nature.

The importance of this linear phase property underlies the work of Oppenheim and Lim [82]. These workers showed that the randomisation of phase in an image led to a critical breakdown in its structure, ie the visually significant lines and edges were severely distorted. It is interesting to note that a similar experiment with the magnitude of the spectrum, although causing noticeable error, did not lead to unrecognizable results, thus confirming the crucial part played by the phase. The logical outcome of this is to incorporate such information into the feature model.

A simple way to achieve this is to make the assumption that the continuous Fourier spectrum can be sampled at regular intervals $\rho_s = s\pi/S$ in a given orientation θ to yield a sequence of complex samples $u(\rho_s,\theta)$, $1 \le s \le S$. Eqn (4.11) then implies a relationship between these samples given the presence of a feature: the phase value of a sample at radial frequency ρ_s depends on the values of previous samples $(\rho_{s-1}, \rho_{s-2}, ...)$ only through the last sample at ρ_{s-1}. Such a relationship then suggests that a suitable model is one that takes the form of a normal Markov process [6]

$$u(\rho_s, \theta) = \alpha\, u(\rho_{s-1}, \theta) + \beta\, v(\rho_s, \theta) \qquad s > 1 \qquad (4.12)$$

$$u(\rho_1, \theta) = v(\rho_1, \theta) \qquad (4.13)$$

where

$$\alpha = |\alpha|\, e^{-j\phi(\theta)} \qquad 0 \le |\alpha| < 1 \qquad (4.14)$$

and $|\alpha|^2 = 1 - \beta^2$. The complex innovations $v(\rho_s, \theta)$ are normally distributed with zero mean and unit variance

$$E\, v(\rho_s, \theta) = 0 + j0 \qquad (4.15)$$

$$E\, v(\rho_s, \theta_1)\, v^*(\rho_r, \theta_2) = \delta(s - r)\, \delta(\theta_1 - \theta_2) \qquad (4.16)$$

It is then easily shown that

$$E\ u(\rho_s, \theta_1)\, u^*(\rho_r, \theta_2) = \alpha^{(s-r)}\ \delta(\theta_1 - \theta_2) \qquad s \geq r \qquad (4.17)$$

The above model clearly incorporates the linear phase requirement, indeed if the feature is present, then from eqns (4.11) and (4.12)

$$\phi(\theta_k) = (\rho_s - \rho_{s-1})\ \eta_k \qquad (4.18)$$

where η_k is the offset of the feature corresponding to the orientation θ_k.

Note that the model of eqn (4.12) tends to the continuous case as $S \rightarrow \infty$. Furthermore, if it is assumed that there are a total of $0 \leq k < K$ local features with spatial frequency orientations θ_k, then the Fourier spectrum $u(\rho_s, \theta)$ can be written as

$$u(\rho_s, \theta) = \sum_{k=0}^{K-1} u(\rho_s, \theta_k)\ \delta(\theta - \theta_k) \qquad (4.19)$$

where $u(\rho_s, \theta_k)$ is defined according to eqn (4.12).

4.4.3. Adaptation to Local Analysis

The feature model needs to be modified to account for the limited region size prescribed by the image model and the consequent requirement for some type of local analysis. To represent such an analysis at a scale defined by the parameter n, define a circular region of radius r_n and assume that there are K features with spatial frequency orientations θ_k and spatial offsets η_k uniformly distributed between $(0, \pi)$ and $(-r_n, r_n)$ respectively.

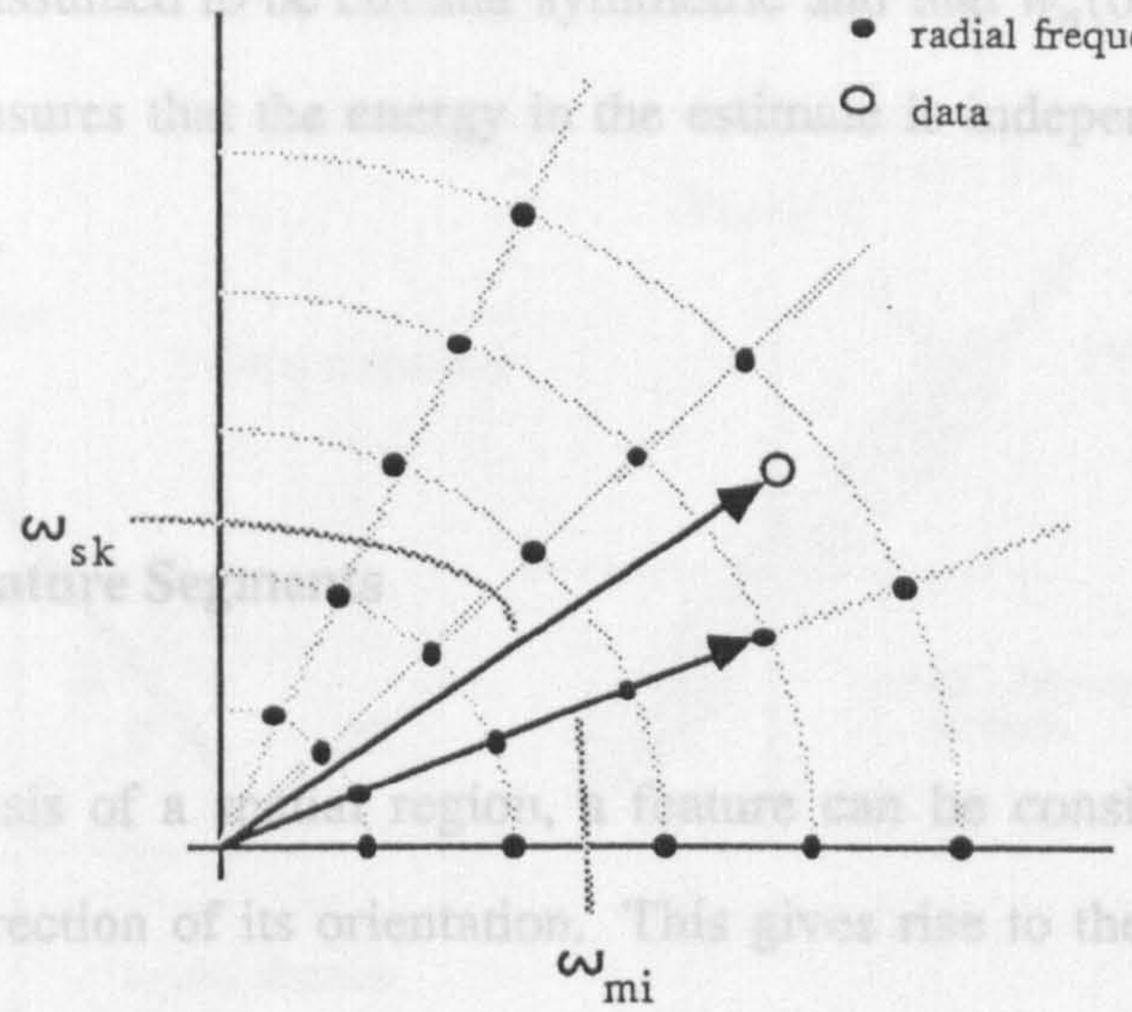


Figure 4.5. Position vectors ω_{mi} and ω_{sk}.

Since the region has a finite extent, an appropriate spectral estimate will have finite resolution (cf section 3.4.2). Ignoring the errors due to the sampling of the continuous spectrum, ie for large S, such an estimate can be approximated by

$$u(m,i) \;=\; \frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} \sum_{s=1}^{S} u(\rho_s,\theta_k)\; w_n(\|\omega_{mi} - \omega_{sk}\|) \qquad (4.20)$$

where the individual components have radial frequency $\rho_m = m\pi/2r_n$, $1 \le m \le M$, and orientations $\psi_i = i\pi/L$, $0 \le i < L$. The vectors $\boldsymbol{\omega}_{mi}$ and ω_{sk} correspond to the position in the spatial frequency domain of the components and the data, ie

$$\omega_{mi} \;=\; [\,\rho_m \cos\psi_i\,,\,\rho_m \sin\psi_i\,] \qquad \omega_{sk} \;=\; [\,\rho_s \cos\theta_k\,,\,\rho_s \sin\theta_k\,] \qquad (4.21)$$

as shown in fig 4.5. The transformed data window $w_n(\omega)$ in eqn (4.20) therefore represents the 'smearing' or bias introduced by the finite resolution estimate. For

simplicity it is assumed to be circular symmetric and that $w_n(0) = 1$. The factor $1/\sqrt{K}$ in eqn (4.20) ensures that the energy in the estimate is independent of the number of features present.

4.4.4. Local Feature Segments

In a local analysis of a spatial region, a feature can be considered to have a finite length in the direction of its orientation. This gives rise to the concept of a *feature segment*, which has its centroid within the region boundaries. Indeed, such segments are more consistent with the locally defined functions $h_{nxy}(k,l)$ in the image model, where the centroid is given by the position vector $\eta_{xy}(n)$. By adopting a similar approach to the previous sections, it is possible to incorporate these segments into the local feature model.

It is best to consider a segment as being generated by the multiplication of the 'infinite' length feature by a smooth and locally defined window function. This will be centred at the centroid of the segment and have a small effective area in comparison with the scale window discussed in the previous section. In the frequency domain, this corresponds to convolving the spectrum with the transformed window, the effect being to smooth out the energy concentration as illustrated in fig 4.6. Note the finite concentration of energy in the form of an oriented and elongated region.

The phase variation across a perpendicular section of this region will be linear and directly proportional to the position of the segment. To see this, consider the ideal case of a continuous oriented segment

$$v(x,y) = v(x \cos\theta + y \sin\theta) v'(x \sin\theta - y \cos\theta) \qquad (4.22)$$

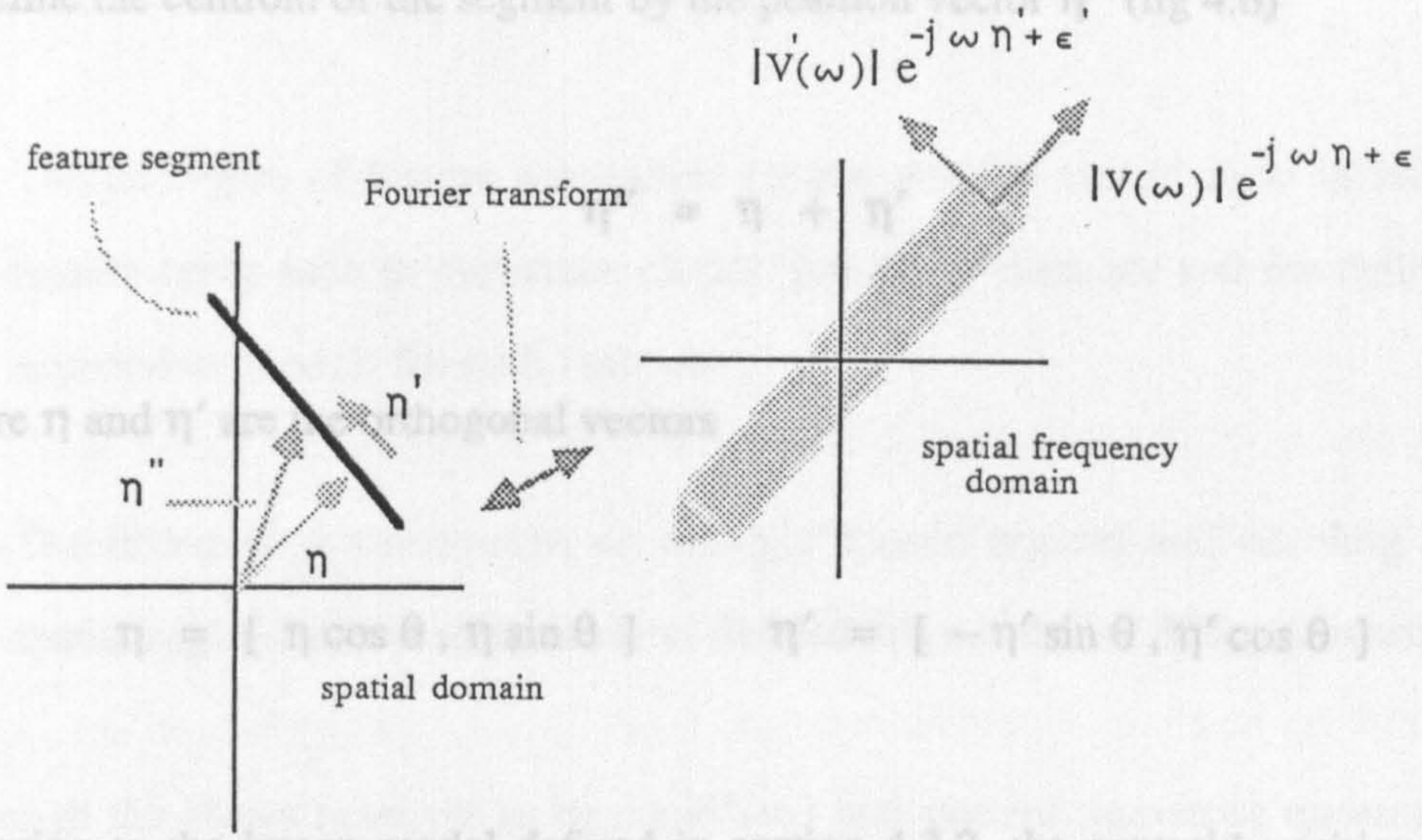


Figure 4.6. Linear phase property of local feature segment.

where $v'(x)$ defines a 1-d variation perpendicular to the orientation θ. The Fourier transform of this segment is

$$V(\,\omega_x\,,\,\omega_y\,) \;=\; V(\,\omega_x \cos\theta + \omega_y \sin\theta\,)\; V'(\,\omega_x \sin\theta - \omega_y \cos\theta\,) \qquad (4.23)$$

where $V'(\,\omega\,)$ represents the smoothing out of the spectrum due to the finite length of the segment as illustrated in fig 4.6.

Using a similar argument to before, this function will have a linear phase component which is proportional to the centroid of $v'(x)$, ie

$$\mathrm{Arg}\,[\,V'(\omega)\,] \;=\; \omega\;\eta' \;+\; \varepsilon' \qquad (4.24)$$

where ε' is a phase constant. By combining the two centroids η and η', it is possible to define the centroid of the segment by the position vector η'' (fig 4.6)

$$\eta'' = \eta + \eta' \tag{4.25}$$

where η and η' are the orthogonal vectors

$$\eta = [\, \eta\cos\theta\,,\, \eta\sin\theta\,] \qquad \eta' = [\, -\eta'\sin\theta\,,\, \eta'\cos\theta\,] \tag{4.26}$$

Referring to the image model defined in section 4.3.2, the centroid position vectors $\eta_{xy}(n)$ corresponding to the local features $h_{nxy}(k,l)$ in eqn (4.5) are equivalent to the vectors η'' defined above.

4.5. A General Class of Models

This chapter has considered a particular form of multiresolution image model. Although it provides an adequate framework for representing local features, it is a simple example of a more general class of models which are capable of representing more complex features.

Various forms of the model have already been applied in a number of image processing areas. The work of Spann and Wilson [113] in texture analysis has already been mentioned, but other applications include image restoration [33][34] and coding [102]. All these have illustrated the advantages that can be gained from a multiresolution approach.

The model can be extended to represent more complex features. There are two aspects to this:

(i) The definition of feature innovation vectors $\mathbf{w}(n)$ in eqn (4.1) to include other feature types such as curvature, circles, junction points, etc and the definition of appropriate models for such features.

(ii) The lifting of the restriction on multiple feature regions and enabling a given spatial region to have many features defined at a number of different resolutions.

Although the above have yet to be considered and present numerous questions concerning implementation and suitable models etc, the important thing to note is that such extensions are considered within the same framework and not as a separate process. In other words, the image analysis problem is still being considered in a unified manner.

An additional point to note, and one that will be illustrated in chapter 6, is that the MFT provides a suitable framework for estimating the parameters of these models. This further enhances the idea of a unified approach, since not only are the image models defined as a similar structure but also the means of estimation can be based within the same representation space.

CHAPTER FIVE

A DETECTION AND ESTIMATION ALGORITHM

5.1. Introduction

The image model introduced in chapter 4 has a hierarchical structure and is based upon local features defined at different spatial resolutions. There are three parameters associated with each feature: the scale n at which it is defined; its orientation θ_{nxy}; and its position vector $\eta_{xy}(n)$. A frequency domain model was defined for these local features.

The purpose of this chapter is to describe a detection and estimation algorithm which assumes the above model. There are three phases to the algorithm. First, for contiguous spatial regions of a given size, parameter estimates based on the frequency domain model are obtained for features in uniformly distributed orientations. Secondly, these estimates are used to detect the single feature regions prescribed by the image model. Finally, the detected local features are used to extract curves in the image according to the representation described in section 4.3.3. The algorithm assumes the availability of local spectrum estimates over the full range of scales and corresponding to the spatial regions illustrated in fig 4.1a. It is shown in chapter 6 that these estimates are provided by the MFT of the image.

5.2. Estimation of Local Features

The frequency domain model of local features defined in section 4.4 is based upon a normal Markov process. It will be shown that the parameters of this model can be

estimated using a maximum likelihood (ML) estimation scheme. The continuous model is considered first and is then extended to include the effects of local analysis and the estimation of feature segments.

Recall that for the continuous case, a local feature is modelled in the frequency domain as a normal Markov process

$$u(\rho_s, \theta) = \alpha\, u(\rho_{s-1}, \theta) + \beta\, v(\rho_s, \theta) \qquad s > 1 \qquad (5.1)$$

where

$$\alpha = |\alpha|\, e^{-j\phi(\theta)} \qquad (5.2)$$

and the spectral coefficients $u(\rho_s, \theta)$ lie in an orientation θ which is perpendicular to the spatial orientation of the feature and $v(\rho_s, \theta)$ is a normally distributed complex random variable with zero mean. It is readily shown, eg [6], that given this model a ML estimate for α can be obtained from the correlation statistic

$$R(\theta) = \frac{1}{S-1} \sum_{s=1}^{S-1} u^*(\rho_s, \theta)\, u(\rho_{s+1}, \theta) \qquad (5.3)$$

where the resulting estimates are unbiased, since from eqns (4.16) and (4.17)

$$E[R(\theta) \mid |\alpha|=0] = E\, v^*(\rho_s, \theta)\, v(\rho_{s+1}, \theta) = 0 \qquad (5.4)$$

and

$$\mathrm{E}\,[\,R(\theta)\mid\,|\alpha|>0\,]\;=\;\alpha\;=\;|\alpha|e^{-j\phi(\theta)} \tag{5.5}$$

Furthermore, the magnitude value $|R(\theta)|$ provides a certainty measure for the existence of the feature which is based upon the energy of the projection of the signal onto the subspace spanned by the model of eqn (5.1), and not upon a simple energy calculation in the given orientation. This means that a distinction is made between the local features of interest and other oriented features (cf section 4.4.2). The offset of the feature follows from eqn (4.18)

$$\tilde{\eta}\;=\;\frac{\tilde{\phi}(\theta)}{(\rho_{s+1}\;-\;\rho_s)} \tag{5.6}$$

where

$$\tilde{\phi}(\theta)\;=\;\mathrm{Arg}\,[\,R(\theta)\,] \tag{5.7}$$

The above estimation is based upon a continuous model that takes no account of the local analysis required by the image model. As discussed in section 4.4.3, such an analysis must be performed using a finite resolution spectrum. An approximation for such a spectrum is defined in eqn (4.20). This assumes that there are K uniformly distributed features with orientations θ_k and offsets η_k at a scale defined by the parameter n. The individual spectral coefficients $u\,(m,i)$ have radial frequency $\rho_m = m\pi/2r_n$, $1 \le m \le M$, and orientation $\psi_i = i\pi/L$, $0 \le i < L$.

From this local spectrum it is possible to derive a set of correlation statistics $R\,(i)$ corresponding to each spectral orientation ψ_i

$$R(i) = \frac{1}{M-1} \sum_{m=1}^{M-1} u^*(m,i)\ u(m+1,i) \qquad (5.8)$$

Clearly the estimates derived from these statistics will be biased, and this can be investigated by considering their expected values given the presence of K features, ie from eqns (4.17) and (4.20)

$$\mathrm{E}[\,R(i)|K\,] = \frac{1}{K(M-1)} \sum_{m=1}^{M-1} \sum_{k=0}^{K-1} \sum_{s=1}^{S} \sum_{r=1}^{S} w_n(\|\omega_{mi} - \omega_{sk}\|)\ w_n(\|\omega_{(m+1)i} - \omega_{rk}\|)\ \alpha_k^{|r-s|} \qquad (5.9)$$

where $\alpha_k = |\alpha_k| e^{-j\phi_k}$ and $\phi_k = \pi\,\eta_k / 2r_n$ is the phase increment representing the offset of the kth feature.

As was explained in section 4.4.3, the data window $w_n(\omega)$ in eqn (5.9) represents the smearing caused by the finite resolution of the estimated spectrum. If this smearing is assumed to be localised by the appropriate choice of window and sampling intervals, then the complicated sum in eqn (5.9) can be approximated by

$$\mathrm{E}[\,R(i)|K\,] = \frac{1}{K(M-1)} \sum_{m=1}^{M-1} \sum_{k \in \Lambda_i} w_n(\|\omega_{mi} - \omega_{mk}\|)\, w_n(\|\omega_{(m+1)i} - \omega_{(m+1)k}\|)\ \alpha_k \qquad (5.10)$$

where $k \in \Lambda_i$ iff $|\theta_k - i\pi/L| \le \pi/L$. This equation simplifies still further if it is assumed that only one feature is present, ie $K = 1$

$$\mathrm{E}[\,R(i)|1\,] = \begin{cases} \dfrac{\alpha_1}{M-1} \displaystyle\sum_{m=1}^{M-1} w_n(\|\omega_{mi} - \omega_{m1}\|)\, w_n(\|\omega_{(m+1)i} - \omega_{(m+1)1}\|) & |\theta_1 - \dfrac{i\pi}{L}| \le \dfrac{\pi}{L} \\ 0 & |\theta_1 - \dfrac{i\pi}{L}| > \dfrac{\pi}{L} \end{cases} \qquad (5.11)$$

By comparing this with the continuous case in eqn (5.5), it can be seen that the average magnitude of the statistic is reduced due to the effect of the data window, ie

$$\sum_{m=1}^{M-1} w_n(\|\omega_{mi}-\omega_{m1}\|)\; w_n(\|\omega_{(m+1)i}-\omega_{(m+1)1}\|) \;\leq\; M-1 \qquad (5.12)$$

However, if it is assumed that the addition of overlapping data windows leads to a 'flat response', and that M and L are large, then the summation of adjacent statistics gives

$$\mathrm{E}[\,R(i)+R(i+1)\mid 1\,] \begin{cases} \approx \alpha_1 & \frac{i\pi}{L} \leq \theta_1 \leq \frac{(i+1)\pi}{L} \\ = 0 & \text{else} \end{cases} \qquad (5.13)$$

Therefore, although the statistic is reduced by the misalignment of the sampling orientations ψ_i with that of the feature θ_1, there is significant contribution in the orientations which straddle θ_1.

From eqn (5.10), the existence of more than one feature, $K > 1$, will result in the corruption of the statistic due to interference from features with similar orientations. In addition, other orientations will in general contain significant feature energy.

In section 4.4.4 the concept of a local feature segment was introduced and it was shown that its centroid is given by the position vector

$$\eta'' = \eta + \eta' \qquad (5.14)$$

where η and η' are defined in eqn (4.26) and illustrated in fig (4.6).

From eqns (4.24) and (4.26), an estimate for the vector η' can be obtained from the linear phase component within the spectral coefficients in an orientation perpendicular to that corresponding to the feature. As in the case of $\tilde{\eta}$ above, an estimate of the offset in this orientation can be obtained via a correlation statistic $C(\theta)$, ie for the continuous spectrum $u(\rho_s, \theta)$

$$C(\theta) = \frac{1}{S(S-1)} \sum_{s=1}^{S} \sum_{r=1-S/2}^{S/2-1} u^*(\rho_{sr}, \theta + \gamma_{sr})\ u(\rho_{s(r+1)}, \theta + \gamma_{s(r+1)}) \qquad (5.15)$$

where

$$\rho_{sr}^2 = \rho_s^2 + \rho_r^2 \qquad \gamma_{sr} = \tan^{-1}(r/s) \qquad (5.16)$$

as illustrated in fig 5.1.

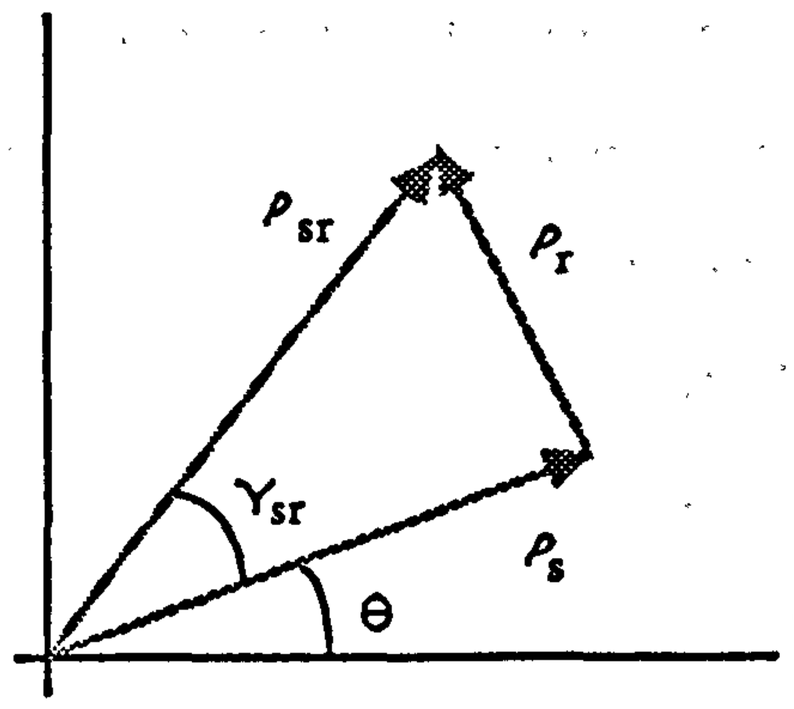


Figure 5.1. Illustration of parameters ρ_{sr} and γ_{sr}.

Hence the statistic $C(\theta)$ is calculated in a perpendicular direction to that of $R(\theta)$, using data symmetrically about the orientation θ and averaging in the radial direction.

Assuming the presence of a single feature, then the required estimate is

$$\tilde{\eta}' = \frac{\text{Arg}\,[\,C(\theta)\,]}{(\rho_{r+1} - \rho_r)} \tag{5.17}$$

However, unlike the estimation of the offset η, the existence of more than one feature in the continuous case will result in a corruption of the statistic $C(\theta)$. This is because it is calculated from coefficients which are at different orientations about the feature orientation θ. To reduce this corruption, the correlation in eqn (5.15) is modified to give the following weighted correlation

$$C(\theta) = \frac{1}{S(S-1)} \sum_{s=1}^{S} \sum_{r=1-S/2}^{S/2-1} \upsilon(\rho_r)\; u^*(\rho_{sr}, \theta + \gamma_{sr})\; u(\rho_{s(r+1)}, \theta + \gamma_{s(r+1)}) \tag{5.18}$$

where the real function $\upsilon(\rho)$ is localised and symmetric about $\rho = 0$. This modification ensures that the statistic is calculated within an elongated region central about the orientation θ similar to that in fig 4.6, hence limiting corruption from other features to those with similar orientations. In the case of a local analysis, it is possible to derive a set of such correlation statistics $C(i)$ in uniformly distributed orientations ψ_i. As in the case of the statistics $R(i)$, errors will be introduced by the finite resolution spectral estimate and the misalignment of features with the discrete orientations. Once again these errors are minimised if a single feature is present.

To conclude, assuming the availability of a finite resolution spectrum estimate $u(m,i)$ for a given local region, it is possible to obtain ML estimates $\tilde{\alpha}_i$ for features in uniformly distributed orientations $\psi_i = i\pi/L$. If a single feature is present in the local region, then there will be significant contribution in those estimates corresponding to the orientations which straddle the orientation corresponding to that of the feature.

The magnitudes of these estimates are reduced in the presence of more than one feature, particularly when features are in similar orientations. Estimates of the position of features in the orientation ψ_i can be obtained via $\tilde{\phi}(i) = \mathrm{Arg}[\, R\,(i)\,]$ and the position of segments by including the estimate in the orthogonal orientation, $\tilde{\phi}'(i) = \mathrm{Arg}[\, C\,(i)\,]$.

5.3. Single Feature Regions

5.3.1. Motivation

The image model defined in chapter 4 is based upon regions which contain a single local feature. In the previous section, however, a region-based estimation scheme was described in which parameter estimates were obtained for features in a finite number of uniformly distributed orientations. To provide a decision scheme for the image model, it is therefore necessary to devise a method which identifies those regions that satisfy the single feature hypothesis.

An obvious way to proceed is to consider the distribution of estimate magnitudes (or certainty measures) over all orientations. A local maximum in this distribution would indicate the presence of a feature in a given orientation and a single, highly concentrated maximum would suggest that the underlying region is likely to contain just the one feature. What is required therefore is a measure of the anisotropy or 'orientedness' of the distribution. One such measure can be obtained by adopting the analogy of point masses rotating about a fixed axis, a system which is often encountered in mechanics. Using the theory of the moments of inertia [13], it is possible to determine the principal axis, or in this case the principal orientation, of such a system and to obtain an estimate of the degree of mass concentration. This is considered in the next

section.

An additional criterion for testing the single feature hypothesis is that of scale consistency: if a region is considered to contain a single feature at a given scale, then this should be confirmed by any information obtained at a smaller scale, the assumption being that no contradictory evidence should be obtained within any subregions if only one feature is present. Such a consistency test is considered in section 5.3.3.

5.3.2. Principal Orientation

A system of point masses rotating about a fixed axis can be analysed by considering the moments of inertia. The technique is used extensively in various areas of mechanics [13]. A related problem is an analysis of the dispersion of a set of (weighted) points with respect to the centroid of the set, and the determination of the principle axes. The latter has found use in image processing to represent orientation in both 2-d and 3-d space [5][67].

Define the set of points by the position column vectors $\mathbf{x}_i$, where the origin is assumed to be the centroid of the points. The 'inertia tensor' is then given by [13]

$$\mathbf{P} = \sum_i m_i \mathbf{P}_i \tag{5.19}$$

where

$$\mathbf{P}_i = \mathbf{x}_i \mathbf{x}_i^T \tag{5.20}$$

and m_i is the weight associated with the point whose position vector is $\mathbf{x}_i$. Since $\mathbf{P}$ is real symmetric, its eigenvectors are orthogonal and it can be shown (eg [5]) that the eigenvector corresponding to the smallest eigenvalue defines the direction of minimum dispersion. In other words, the principal axis is given by the eigenvector corresponding to the largest eigenvalue. The difference between the eigenvalues is then a measure of the dispersion, eg in the 2-d case one large eigenvalue and one small one indicates that the points are concentrated about the orientation given by the angle of the eigenvector corresponding to the largest eigenvalue.

The above theory can be applied in the present 2-d case. From section 5.2, the set of 'point masses' are the estimate magnitudes $|\tilde{\alpha}_i| = |R(i)|$ with position vectors

$$\mathbf{x}_i = [\ \sqrt{2}\cos\psi_i\ ,\ \sqrt{2}\sin\psi_i\ ] \tag{5.21}$$

corresponding to the orientations $\psi_i = i\pi/L$, $0 \le i < L$. The following tensor is then defined

$$\mathbf{P} = \sum_i |\tilde{\alpha}_i| \mathbf{P}_i = \begin{bmatrix} p_1 & p_2 \\ p_2 & -p_1 \end{bmatrix} \tag{5.22}$$

where

$$\mathbf{P}_i = \mathbf{x}_i\ \mathbf{x}_i^{\mathrm{T}} - \mathbf{I} = \begin{bmatrix} \cos 2\psi_i & \sin 2\psi_i \\ \sin 2\psi_i & -\cos 2\psi_i \end{bmatrix} \tag{5.23}$$

is the traceless tensor corresponding to the orientation ψ_i and having identical eigenvectors and difference of eigenvalues to $\mathbf{x}_i\ \mathbf{x}_i^{\mathrm{T}}$. The eigenvalues of $\mathbf{P}$ are then related

by

$$\lambda_1 = -\lambda_2 = (p_1^2 + p_2^2)^{\frac{1}{2}} \tag{5.24}$$

and thus a suitable measure of the dispersion of the estimates $|\tilde{\alpha}_i|$ is the statistic

$$\zeta = p_1^2 + p_2^2 \tag{5.25}$$

which represents the difference in the eigenvalues. The principal orientation is then given by

$$\theta_0 = \tan^{-1}\left[\frac{e_1(1)}{e_0(1)}\right] \tag{5.26}$$

where e(1) is the eigenvector corresponding to the eigenvalue λ_1.

Recall that the requirement is to determine when the estimates $|\tilde{\alpha}_i|$ correspond to a single feature region. To assess whether the statistic ζ can be used for this task, consider its expected value when a single feature is present with an orientation θ. Ignoring any errors due to sampling when calculating the correlation statistics $R(i)$, ie L and M are assumed to be large in eqn (5.11), then from eqns (5.22) and (5.25) this is given by

$$\begin{aligned} E[\zeta \mid 1] &= E[|R(\theta)|^2] (\cos^2 2\theta + \sin^2 2\theta) \\ &= E[|R(\theta)|^2] \end{aligned} \tag{5.27}$$

where $R(\theta)$ is the correlation statistic calculated in the orientation θ. Now, if an additional feature is present in an orientation uniformly distributed between $(\theta, \theta + \pi/2)$, then this expected value becomes

$$E[\zeta|2] = \frac{1}{2\pi} E[|R(\theta)|^2] \int_{\theta}^{\theta+\frac{\pi}{2}} (\cos 2\theta + \cos 2\theta_1)^2 + (\sin 2\theta + \sin 2\theta_1)^2 \, d\theta_1 \quad (5.28)$$

where it is assumed that the region energy is independent of the number of features and that the magnitude of the correlation statistic is the same for both features. The above equation simplifies to

$$E[\zeta|2] = \frac{1}{2} E[|R(\theta)|^2] = \frac{1}{2} E[\zeta|1] \quad (5.29)$$

and the statistic ζ is reduced by half if the region contains two features. It is clear that further reduction will occur if more features are present. Therefore, in this case, ζ provides a suitable statistic to test the single feature hypothesis. However, its effectiveness is obviously determined by the ability of the correlation statistics to represent separate orientations, which is further dependent upon the availability of a local spectral representation with sufficient resolution.

5.3.3. Scale Consistency

An additional criterion for testing the single feature hypothesis for a given region is that of scale consistency. This is based upon the premise that if a region contains only one feature, then an analysis of subregions should not indicate anything to the

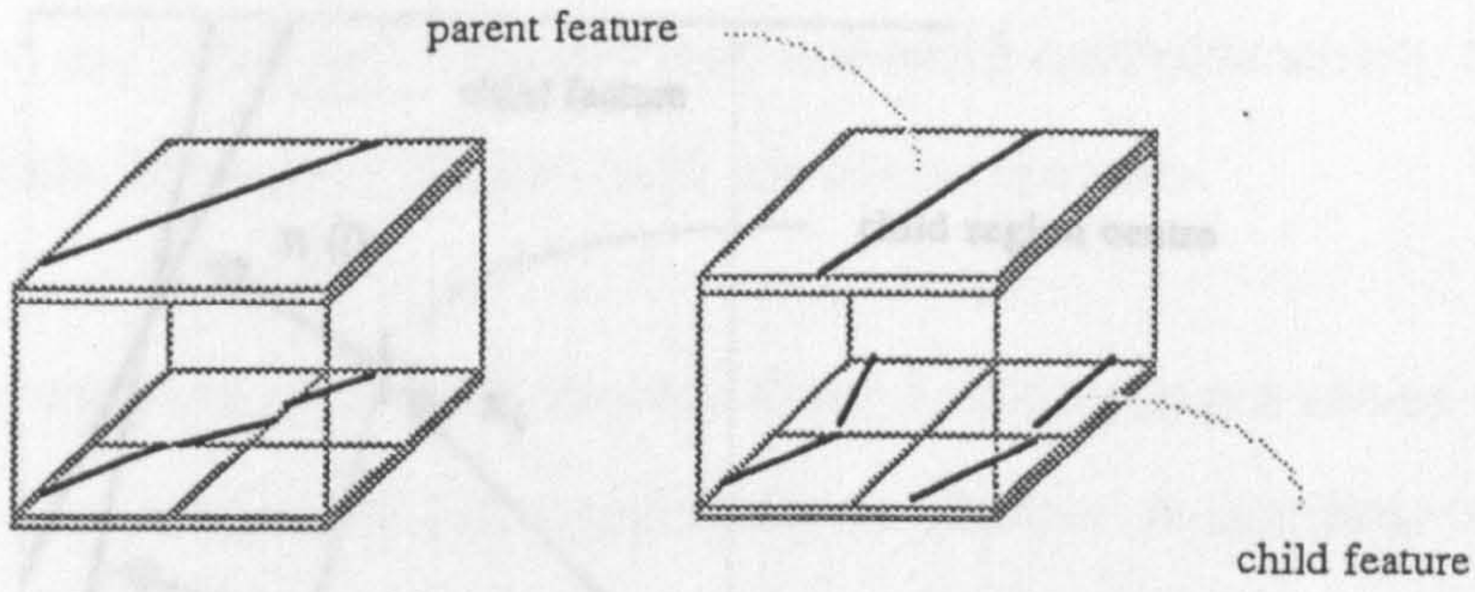
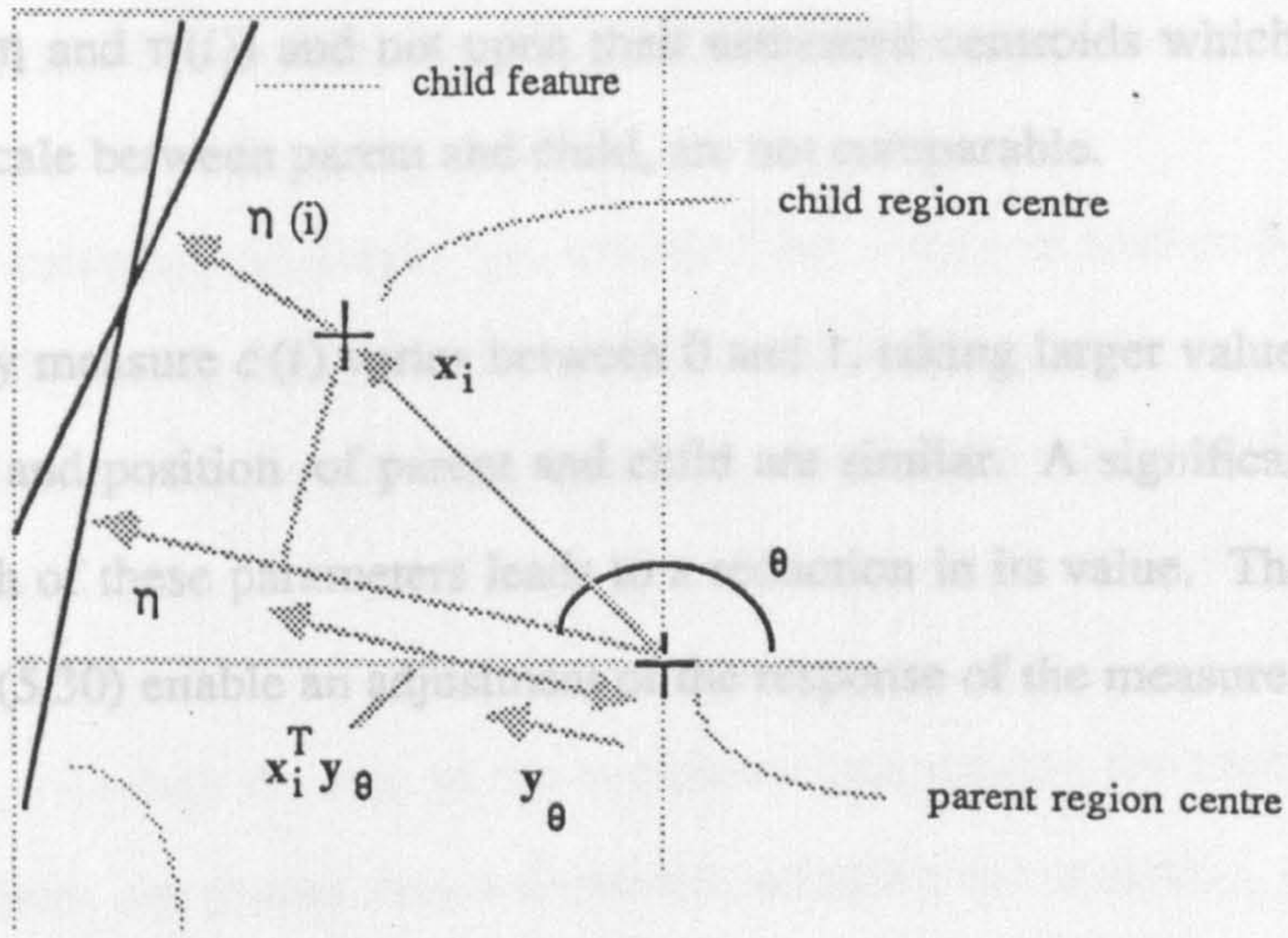


Figure 5.2. Consistent and inconsistent feature information.

contrary, eg the existence of a feature in a subregion which does not correspond to the feature already identified.

In this work, the consistency test is based upon the four quadrants of a region which correspond to the adjacent level in the multiresolution model framework. This is illustrated in fig 5.2, where the single feature identified in the parent region is compared with those (if any) identified in its four child regions. If the features within the child regions agree with that at the parent level as in fig 5.2a, then the single feature hypothesis is accepted within the parent region. However, if as in fig 5.2b, one or more children disagree, then the hypothesis is rejected. The test also requires that either each child region should satisfy the single feature criterion described in the previous section or has no significant contribution in its feature estimates; otherwise the information at the child level is regarded as indeterminate and the hypothesis at the parent level is rejected.

There are two parameters in the above agreement criterion: the orientation and position of the features. Denoting the orientation and position vector of the parent feature by θ and $\eta'' = \eta + \eta'$ respectively, a suitable consistency measure with respect to each child is



$$\Delta\eta_i = \|\eta\| - x_i^T y_\theta - \|\eta(i)\|$$

Figure 5.3. Calculation of position difference $\Delta\eta_i$.

$$c(i) = \begin{cases} \cos^a\left(\frac{\Delta\eta_i\,\pi}{2A}\right)\,\cos^a(\theta-\theta_i) & |\Delta\eta_i|<A \quad |\theta-\theta_i|<\pi/2 \\ 0 & \text{else} \end{cases} \qquad A,a>0 \qquad (5.30)$$

where θ_i is the orientation of the feature associated with ith child and

$$\Delta\eta_i = \|\eta\| - x_i^T y_\theta - \|\eta(i)\| \qquad (5.31)$$

represents the difference in position indicated by the parent and the ith child as illustrated in fig 5.3. In this case, $\eta''(i) = \eta(i) + \eta'(i)$ is the position vector corresponding to the ith child, x_i indicates the position of the child region centre with respect to that of the parent region, and y_θ is a unit vector in the orientation θ. Note that this difference measure is based upon the feature offsets in the direction orthogonal to their

orientation (ie η and $\eta(i)$) and not upon their estimated centroids which, because of the change in scale between parent and child, are not comparable.

The consistency measure $c(i)$ varies between 0 and 1, taking larger values when both the orientation and position of parent and child are similar. A significant difference in either or both of these parameters leads to a reduction in its value. The parameters a and A in eqn (5.30) enable an adjustment of the response of the measure.

5.4. A Hierarchical Detection Scheme

The detection criteria described in the previous section can be combined into a single hierarchical scheme. The aim is to identify the separate and contiguous regions of the multiresolution image model defined in section 4.3 and illustrated in fig 4.1.

The scheme is based upon the availability of local spectrum estimates for the full range of region sizes. Local feature estimates in uniformly distributed orientations corresponding to each region are obtained according to section 5.2. It is useful to visualise these estimates as forming the nodes of a quadtree, where the root node refers to the largest region, ie the image, and subsequent child nodes refer to quadrants as the image is recursively split into squares.

The detection scheme is a recursive process that starts at the root node and then proceeds as follows:

(i) Calculate the sum of the correlation magnitudes $|R(i)|$ over all orientations ψ_i. These may be normalised values according some hierarchical normalisation process (cf section 6.2.3). If the sum is less than a threshold, the region is classified

as 'lowpass' and the process stops by truncating the tree at the current node.

(ii) Otherwise, calculate the anisotropy statistic ζ according to section 5.3.2. If it is less than a threshold, the region is classified as containing more than one feature. The region is then split and the process is repeated for each child node (quadrant).

(iii) Otherwise, calculate the sum of the correlation magnitudes for each child node. If one or more are greater than a threshold, calculate the statistic ζ and the consistency measure $c(i)$ for the relevant children. If either of these is less than a threshold, an inconsistency is noted and the region is split, repeating the process for each child node.

(iv) Otherwise, the region is classified as containing a single feature and the process stops by truncating the tree at the current node.

Once the above scheme is completed, the image will be represented by a truncated quadtree that divides it into different sized spatial regions which are classified as either containing a single local feature or are essentially lowpass regions, ie they are defined by the function $g(x,y)$ in eqn (4.5). The resulting structure resembles that in fig 4.1b.

The parameters of the multiresolution image model have therefore been determined: a finite set of features each defined at a scale n and having orientations θ_{nxy} and position vectors $\eta''_{xy}(n)$. As discussed in chapter 4, these local features can be used to extract curves in an image and this is considered in the next section.

5.5. Curve Extraction

5.5.1. Recursive Curve Forming

The representation of curves within the multiresolution image model was described in section 4.3.3. It is a piecewise representation, with local features defined at different spatial resolutions representing sections of the curve. The local curvature is given by the change in orientation between adjacent features and a given curve satisfies a maximum curvature criterion. Using the local features that result from the detection scheme described in the previous section, it is possible to base a curve extraction scheme upon this representation.

The scheme is recursive and operates on the truncated quadtree that results from the local feature detection. Recall that the nodes in this tree correspond to square regions of the image, and that the leaf nodes are either classified as single feature regions or lowpass regions. The idea is to use an upward directed process within this tree to form a curve by combining features at successively lower spatial resolutions, until at some node the curve is completely identified. This involves applying the following steps at each node in the tree, starting at the root node:

(i) For each child node that is not a leaf node apply steps (i)-(iii).

(ii) Form all possible curve sections by combining appropriate features or curve sections defined at each child node. This is the local curve forming operation and is described below.

(iii) Assign these sections to the current node.

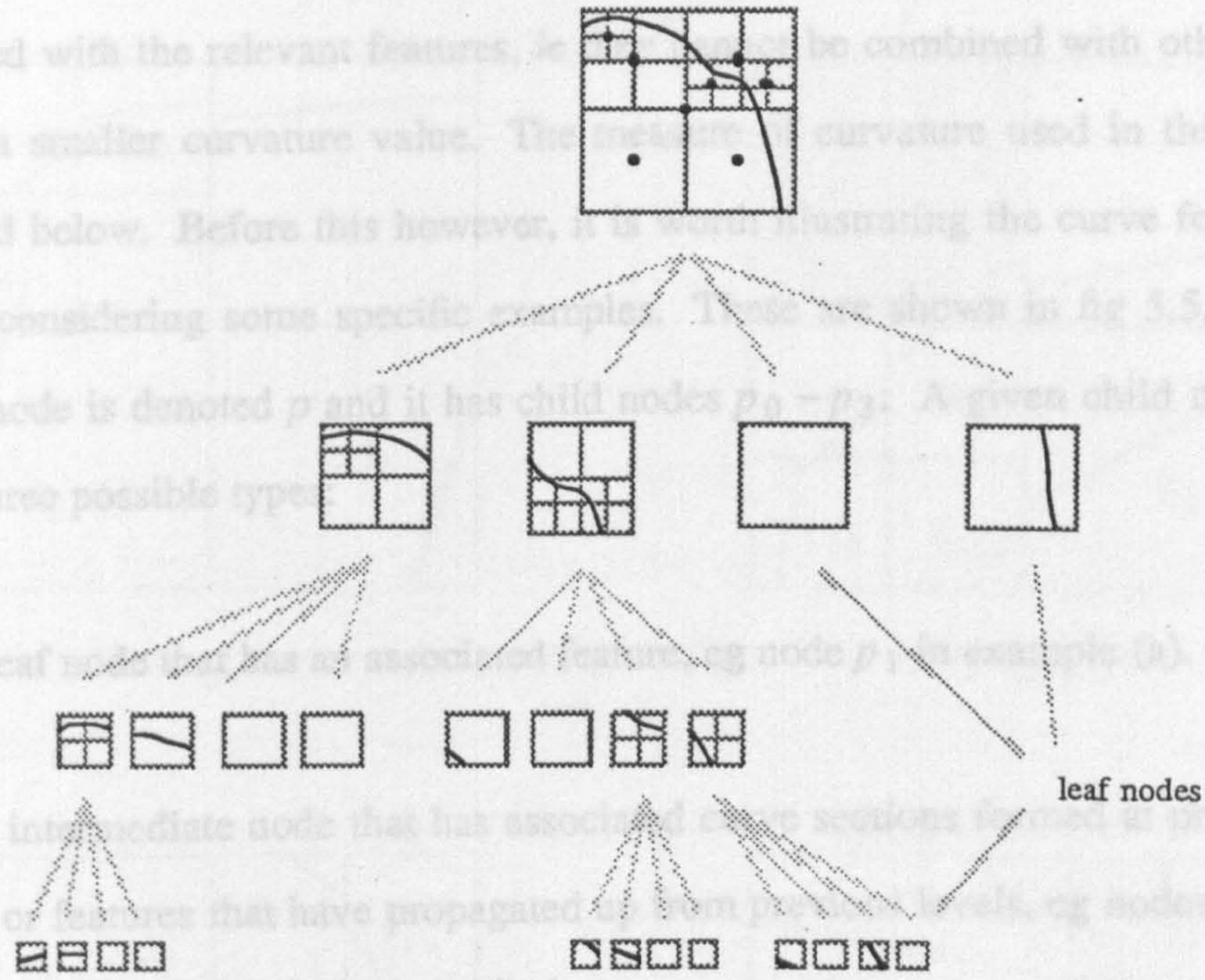


Figure 5.4. Hierarchical curve extraction.

The scheme is therefore a fine-to-coarse analysis that begins by descending to the nodes whose child nodes are all leaf nodes. Curve sections are then formed and these are propagated up through the tree to enable them to be combined with other similarly formed sections or features defined at lower spatial resolutions. An example of a curve formed in this way is illustrated in fig 5.4, where the complete curve is extracted after three levels of the tree have been processed.

The local curve forming operation used in the above scheme is based upon a heuristic search using the features or curve sections defined at the child nodes of the current node. The requirement is to combine appropriate features or sections into new sections according to a maximum curvature and 'best fit' criteria. In other words, a curve section is formed provided that the curvature represented by adjacent features is less than a threshold and that these curvature values are the minimum that can be

associated with the relevant features, ie they cannot be combined with other features to give a smaller curvature value. The measure of curvature used in this search is described below. Before this however, it is worth illustrating the curve forming process by considering some specific examples. These are shown in fig 5.5, where the current node is denoted p and it has child nodes $p_0 - p_3$. A given child node can be one of three possible types:

(i) A leaf node that has an associated feature, eg node p_1 in example (a).

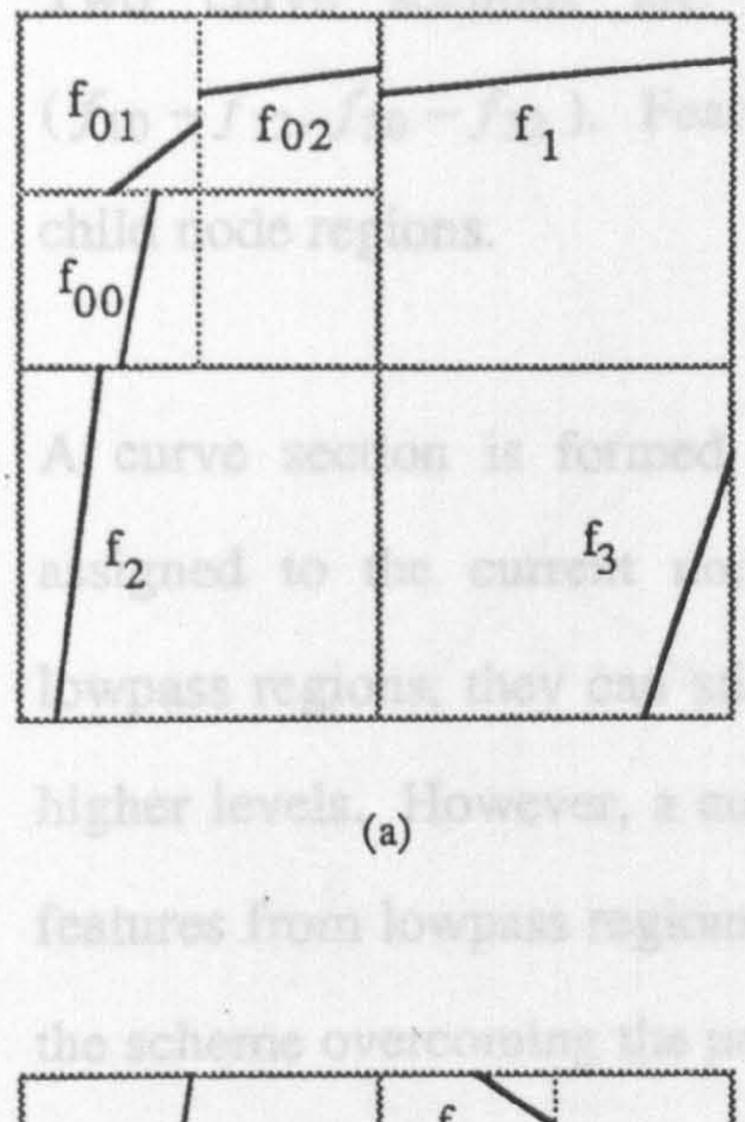
(ii) An intermediate node that has associated curve sections formed at previous levels or features that have propagated up from previous levels, eg nodes p_2 and p_1 in examples (b) and (c) respectively.

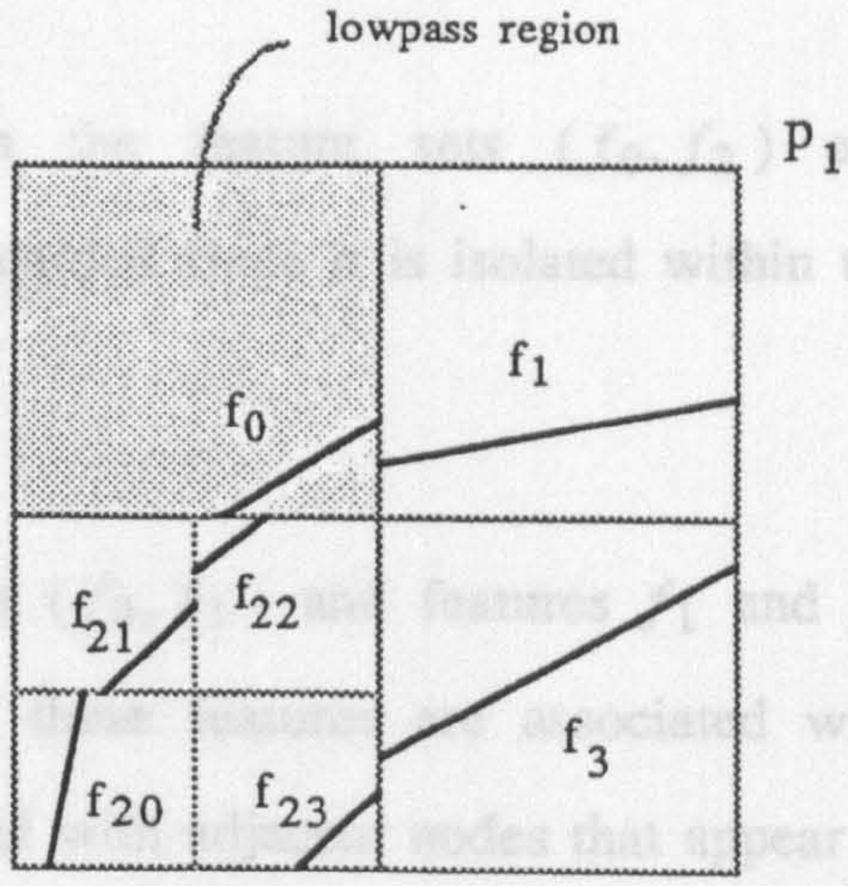
(iii) A leaf node that has been classified as representing a lowpass region, although has an associated feature estimate, eg node p_0 in example (b).

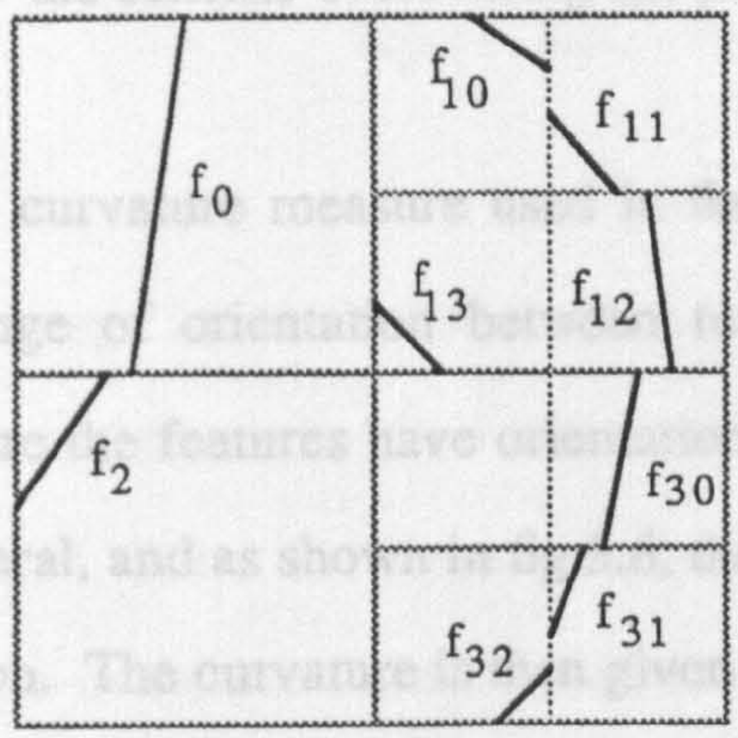
The outcome of the search process in the examples (a)-(d) is summarised below (see also the table in fig 5.5)

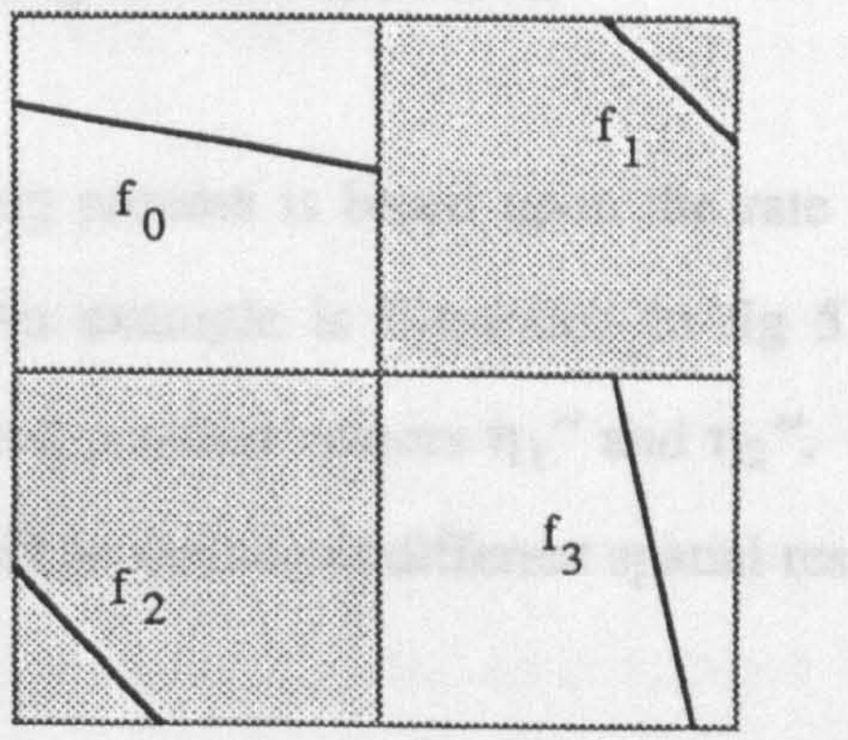
(a) A new curve section is formed by combining features ($f_2, f_{00} - f_{02}, f_1$). Feature f_3 is also assigned to the current node since it is defined at the boundary of the child regions and could therefore combine with adjacent nodes at a higher level.

(b) Two curve sections are formed from the feature sets ($f_{20} - f_{22}, f_0, f_1$) and (f_{23}, f_3). Note that the lowpass node p_0 contributes its feature f_0 since it completes a curve. This is an example of the ability of the extraction scheme to fill in curve gaps.









example	curve sections	features
(a)	(f_2 , f_{00} - f_{02} , f_1)	f_3
(b)	(f_{20} - f_{22} , f_0 , f_1) (f_{23} , f_3)	
(c)	(f_{10} - f_{12} , f_{30} - f_{32}) (f_0 , f_2)	
(d)	(f_0 , f_3)	f_2 , f_3

Figure 5.5. Local curve forming examples.

(c) Two curve sections are formed from the feature sets (f_0, f_2) and ($f_{10} - f_{12}, f_{30} - f_{32}$). Feature f_{13} is discarded since it is isolated within the child node regions.

(d) A curve section is formed from features (f_0, f_3) and features f_1 and f_2 assigned to the current node. Although these features are associated with lowpass regions, they can still be combined with adjacent nodes that appear at higher levels. However, a curve section may not be formed by combining only features from lowpass regions. Once again the curve (f_0, f_3) is an example of the scheme overcoming the problem of missing sections in a curve.

The curvature measure used in the curve forming process is based upon the rate of change of orientation between two features. An example is illustrated in fig 5.6, where the features have orientations θ_1 and θ_2 and position vectors η_1'' and η_2''. In general, and as shown in fig 5.6, these features will be defined at different spatial resolution. The curvature is then given by

$$c = 1 - \frac{1}{4}\left[\,\|\mathbf{y}_{2\theta_1} + \mathbf{d}'\| + \|\mathbf{y}_{2\theta_2} + \mathbf{d}'\|\right] \tag{5.32}$$

where $\mathbf{y}_{2\theta_1}$ and $\mathbf{y}_{2\theta_2}$ are unit vectors in the orientations $2\theta_1$ and $2\theta_2$, $\mathbf{d}'$ is the unit vector in the same direction as the difference vector between the feature positions

$$\mathbf{d}' = \frac{\mathbf{d}}{\|\mathbf{d}\|} \qquad \mathbf{d} = \mathbf{x} + \eta_2'' - \eta_1'' \tag{5.33}$$

and $\mathbf{x}$ is the centre offset vector between the two regions associated with the features as shown in fig 5.6.

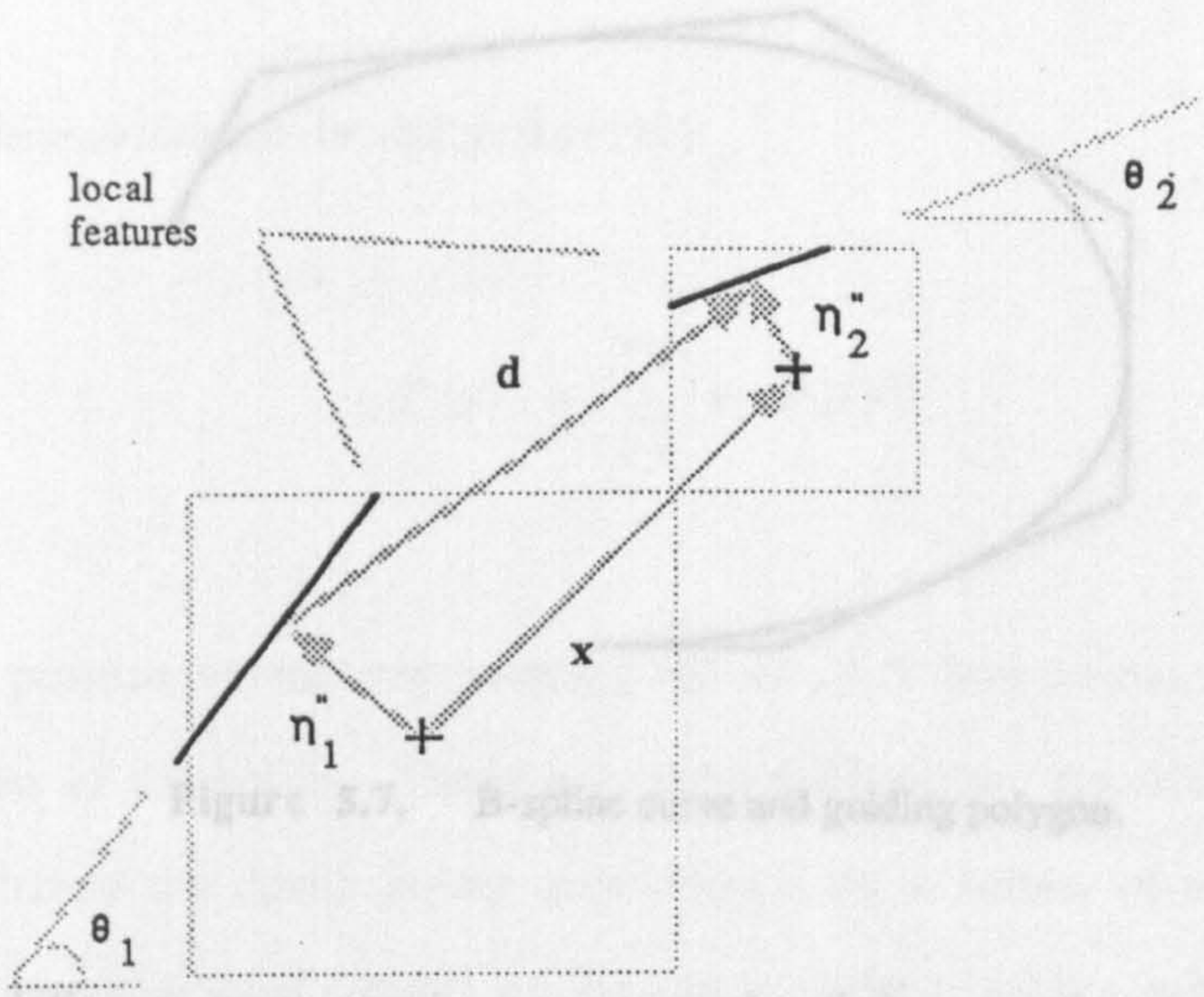


Figure 5.6. Parameters of local curvature measure.

The change in orientation represented by the two features is accounted for in the curvature measure c by the addition of the unit vectors $\mathbf{y}_{2\theta_1}$ and $\mathbf{y}_{2\theta_2}$ with the difference vector $\mathbf{d}'$. Note that these vectors are defined using a double angle representation in order to overcome the ambiguity associated with feature orientations: an orientation θ is equally represented by vectors whose orientations are θ and $\theta + \pi$. The use of the double angle overcomes this by ensuring that each orientation has a unique vector [65]. The resulting curvature measure in eqn (5.32) varies between 0 and 1, attaining a maximum value when the features are perpendicular and a minimum value when they have the same orientation.

5.5.2. B-Spline Polynomial Fitting

The result of the curve extraction scheme is that a given curve is represented by local features which have an associated position and orientation. The representation is therefore piecewise and each section is given by a straight line. It is possible,

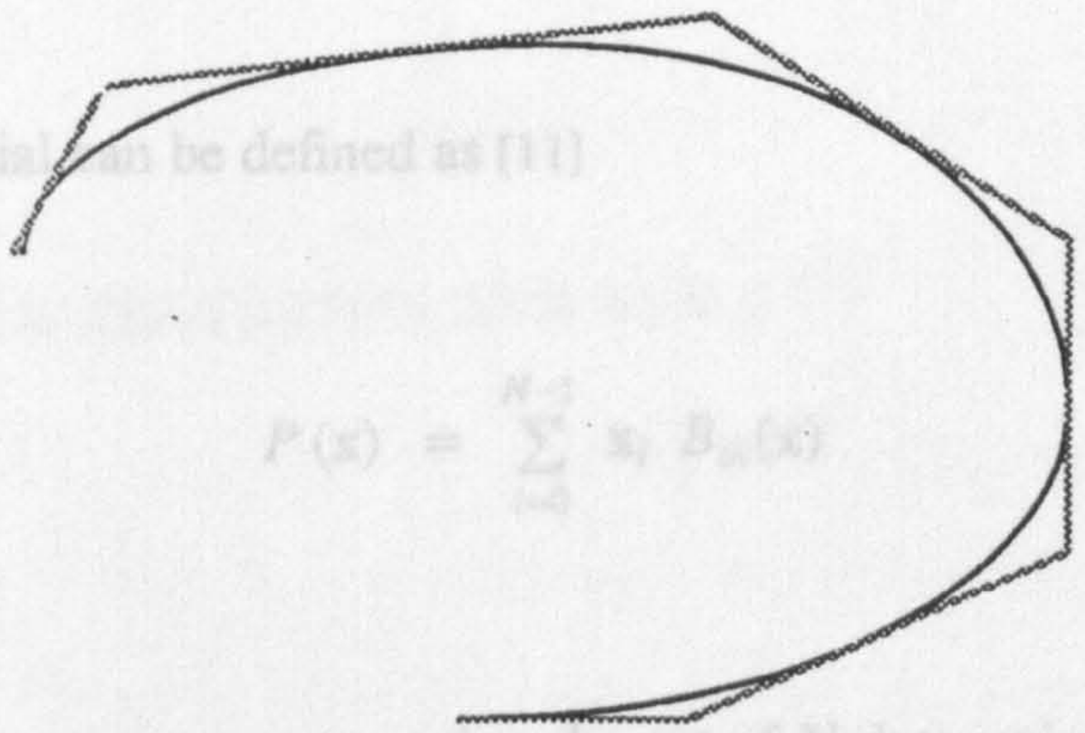


Figure 5.7. B-spline curve and guiding polygon.

however, to define a more continuous curve by fitting a polynomial to the set of feature positions. Such a curve is clearly more suited to those found in natural images.

There are a number of techniques available for fitting polynomials to a set of data points. A review of the various methods is provided in [5][11] and they include algebraic planar curves, parametric and rational polynomials and B-spline methods. The latter is frequently used since it is particularly easy to implement and has several advantages which are relevant when dealing with natural curves.

The B-spline is a piecewise polynomial curve which is controlled by a guiding polygon, where the vertices of the polygon are the set of data points. An example is illustrated in fig 5.7. Cubic polynomials are normally used for the splines since they are the lowest order in which the curvature can change sign. The main advantage of B-splines, apart from the fact that they can be computed efficiently, is that the resulting curves appear to be a natural choice given the particular guiding polygons. In other words, they closely approximate curves that might have been fitted with the aid of the human eye. This makes the technique particularly suitable for image analysis.

A B-spline polynomial can be defined as [11]

$$P(\mathbf{x}) = \sum_{i=0}^{N-1} \mathbf{x}_i \, B_{ik}(\mathbf{x}) \tag{5.34}$$

where $\mathbf{x}_i$ are position vectors representing the set of N data points and $B_{ik}(\mathbf{x})$ are the basis functions of the spline. These nonnegative functions are defined on a limited support and hence the curve points only depend on a subset of neighbouring data points. This provides local control on the shape of the resulting polynomial approximation. The parameter k in eqn (5.34) controls the continuity of the curve. There are a number of different basis functions that can be used [11] and these vary in complexity. In the present work those defined in [80] were adopted, primarily because they can be generated using a simple recursive procedure.

A curve resulting from the extraction scheme can be considered to have associated features f_i, $0 \leq i < L$, each defined at positions represented by the 2-d vectors $\boldsymbol{\eta}_i''$. Using eqn (5.34), the B-spline approximation for this curve is then given by

$$P(\mathbf{x}) = \sum_{i=0}^{L-1} \boldsymbol{\eta}_i'' \, B_{ik}(\mathbf{x}) \tag{5.35}$$

where $\mathbf{x}$ is a 2-d vector representing a point on the spline and $B_{ik}(\mathbf{x})$ is defined as in [80] with $k = 3$. Although this is a simple example of fitting a spline to the set of feature positions (eg it might have been possible to make use of the orientation information associated with each feature), it was found that the results obtained were acceptable (cf chapter 6) and that limited return would be gained from using a more complicated method.

CHAPTER SIX

ESTIMATOR IMPLEMENTATION AND RESULTS

6.1. Introduction

In the previous two chapters an image model based on local features has been introduced and a detection and estimation scheme described. The purpose of this chapter is to show how the MFT can be used to implement the estimation scheme and to consider two important issues in this implementation, namely the problem of phase ambiguity when estimating the parameters of the linear phase model and the normalisation of the calculated statistics used in the estimation process. Note that once an appropriate implementation method has been defined, then the detection of single feature regions and the extraction of curves follows directly from the discussion in the previous chapter. This chapter also considers the problem of reconstructing an image from the estimated parameters of the model and shows that if the MFT is used as an estimation tool, then this reconstruction can be achieved in a consistent and well defined way. Finally, the chapter concludes by presenting feature estimation, single feature detection and curve extraction results for synthetic and natural images.

6.2. An MFT Based Estimator

6.2.1. The Estimator

The MFT provides local spectrum estimates over the full range of scales. Specifically, if it is considered as a vector quadtree (cf section 3.4.3), then its spatial frequency vectors $\mathbf{u}_{xy}(n)$ refer to the individual regions of the multiresolution image model shown in

fig 4.1a. Hence the MFT has a structure which is the same as that which underlies the model and it therefore serves as a basis for implementing the estimation scheme.

Before considering the implementation in detail, it is worth describing the idea behind the scheme. From chapter 4, the estimation of local features is based upon the existence of a linear phase component within the spatial frequency coefficients of a local spectrum estimate in an orientation perpendicular to that of the feature. The multiresolution image model assumes that the image consists of such features defined within different sized spatial regions (a typical example is shown in fig 4.1b). Therefore, since the MFT provides local spectrum estimates for these regions it can be used as an estimation tool. Specifically, if a region contains a local feature as illustrated in fig 6.1, then this will give rise to a linear phase component in the MFT coefficients which is directly related to the centroid position of the feature. Thus by operating on the MFT coefficients to detect this property the existence or nonexistence of a local feature can be estimated.

The above does assume, however, that the local spectrum estimates provided by the MFT are not significantly biased. It was shown in section 3.4.2 that in terms of locality the estimates represent optimal estimates since the vectors used in the generation have minimum uncertainty. However, the local feature estimation described in section 5.2 was based on a number of assumptions about the type of window functions used and also upon a polar separable implementation. The MFT used in this work is cartesian separable and therefore any estimation of the local features must necessarily be an approximation. Despite this, experiments have shown that such an approximation does not lead to any significant bias and that acceptable estimates can be obtained. The details of how this is achieved is considered in the remainder of this section and the results are presented in section 6.4.1.

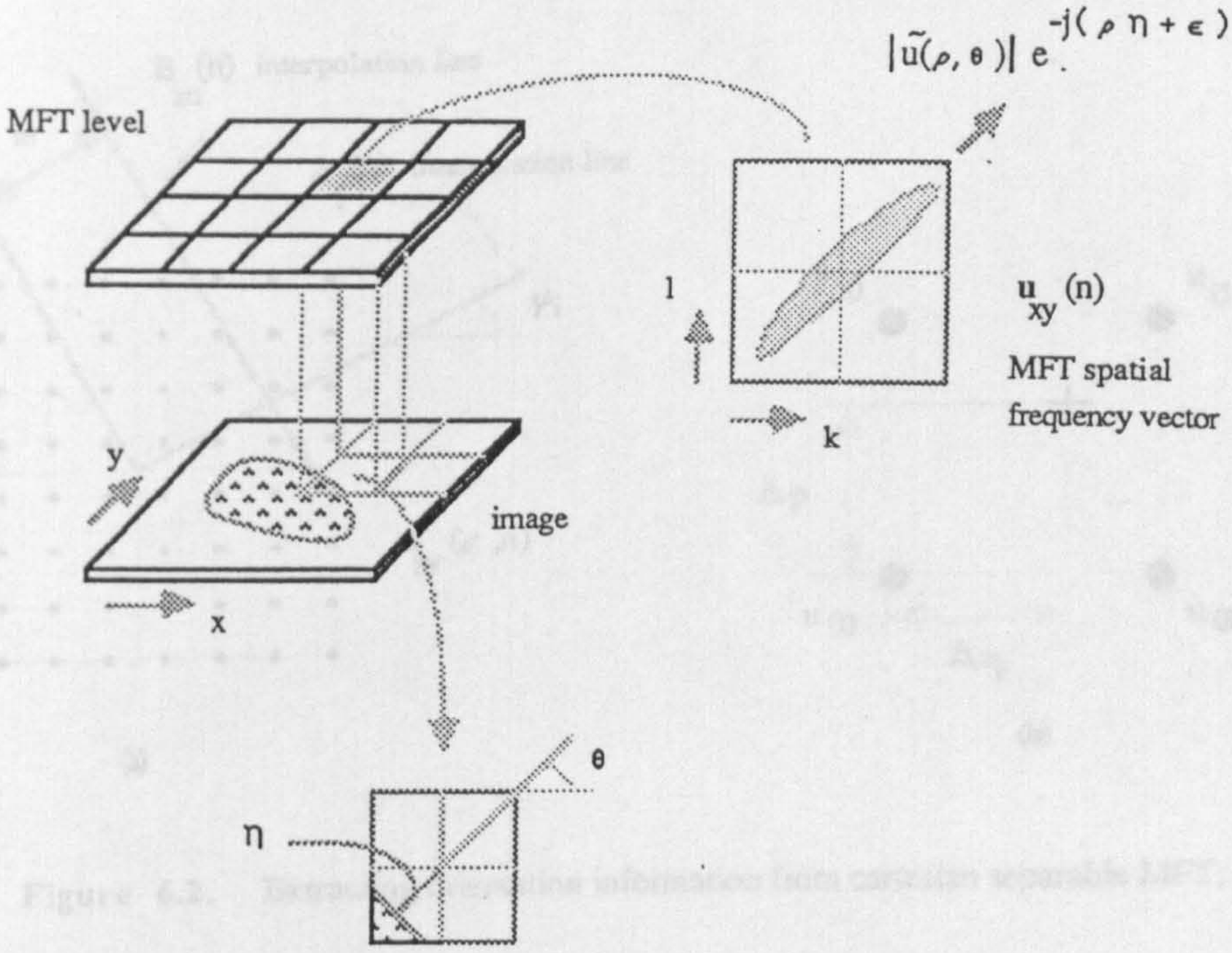


Figure 6.1. Estimation of local feature offset using the MFT.

Recall from the previous chapter that the estimation scheme is based upon calculating the correlation statistics $R(i)$ and $C(i)$ in a discrete number of uniformly distributed spatial frequency orientations ψ_i. As discussed above, these can be calculated by using the local spectrum estimates on each level of the MFT. Adopting the notation of chapter 3 for a generalised 2-d MFT, there are $\sigma\,\Omega_n \times \sigma\,\Omega_n$ spatial frequency vectors $\mathbf{u}_{xy}(\sigma, n)$ with resolution $\Gamma_n \times \Gamma_n$. In the present work these are defined on a cartesian lattice and it is therefore necessary to obtain the correlation statistics by interpolation. Towards this, define the following set of orientation vectors of dimension $\Gamma_n/2 \times 1$ for each spatial frequency vector

$$\mathbf{z}_{xyi}(n) \;=\; \mathbf{A}_i(n)\ \mathbf{u}_{xy}(\sigma, n) \qquad 0 \le i < \Gamma_n/2 \tag{6.1}$$

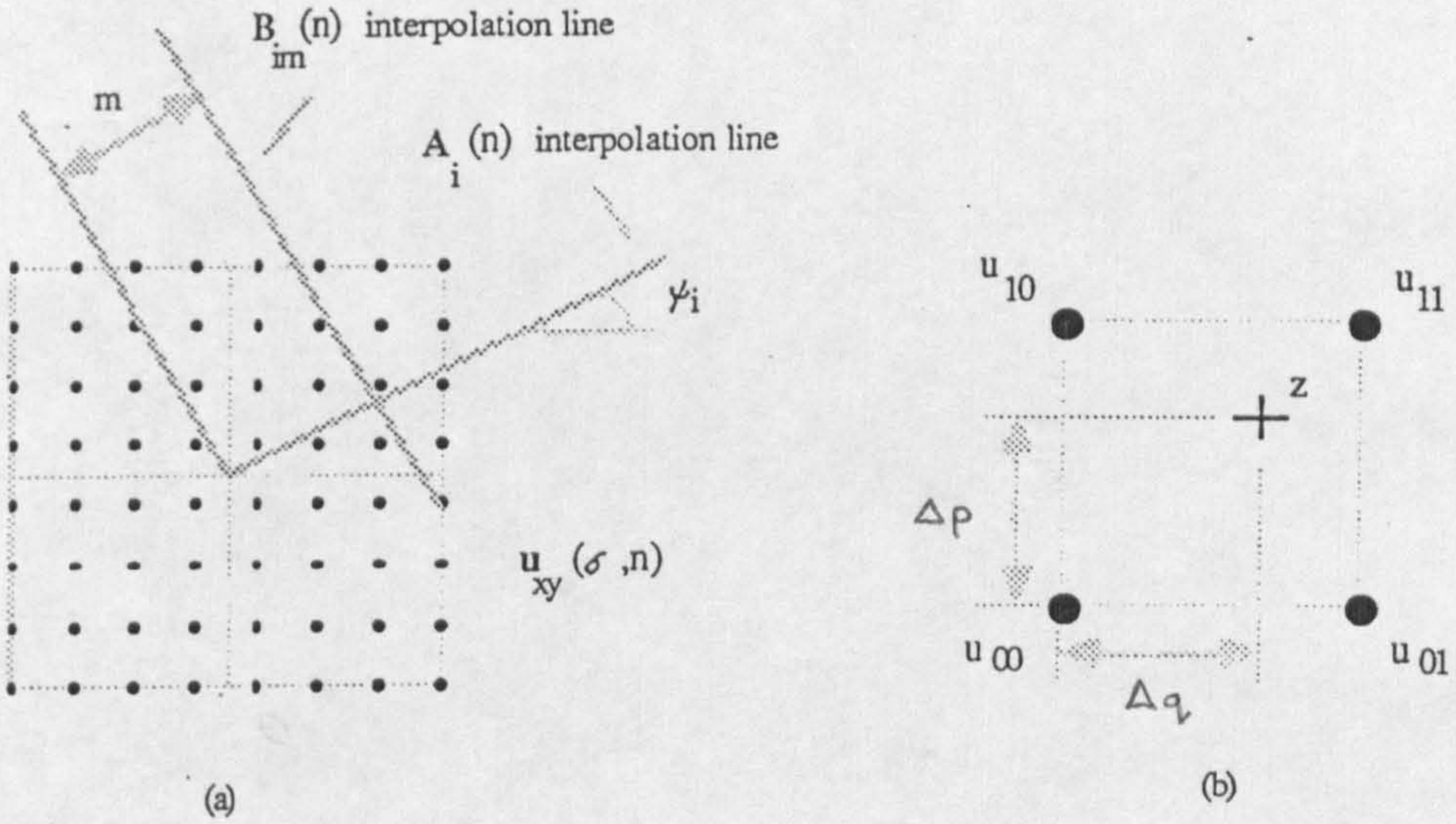


Figure 6.2. Extracting orientation information from cartesian separable MFT.

$$\mathbf{w}_{xyim}(n) \;=\; \mathbf{B}_{im}(n)\;\mathbf{u}_{xy}(\sigma,n) \qquad 0 \le i,m < \Gamma_n/2 \tag{6.2}$$

where the operator $\mathbf{A}_i(n)$ defines the interpolation in an orientation $\psi_i = i\pi L/\Gamma_n$ within the 2-d lattice of $\mathbf{u}_{xy}(n)$. The resulting vector $\mathbf{z}_{xy}(n)$ is then a polar representation of this vector. In a similar manner, $\mathbf{B}_{im}(n)$ defines an interpolation in the perpendicular direction with an offset m as illustrated in fig 6.2a.

For each orientation, the components of $\mathbf{z}_{xyi}(n)$ and $\mathbf{w}_{xyim}(n)$ are uniformly spaced at unit intervals. In the present work they are obtained using a bilinear interpolation formula, which has the advantage that it is simple to implement and computationally efficient. Higher order methods could be used, although the results presented in section 6.4 suggest that the extra complexity and computation would have limited return. In general, a required component will lie in a square region whose corner values are given by four available coefficients as shown in fig 6.2b. The interpolated value is

then given by

$$z = (1-\Delta p)(1-\Delta q)\,u_{00} + \Delta p\,(1-\Delta q)\,u_{10} + \Delta q\,(1-\Delta p)\,u_{01} + \Delta p\,\Delta q\,u_{11} \qquad (6.3)$$

where $u_{00} - u_{11}$ are the available coefficients and Δp, Δq is the position of the required value z with respect to the coefficient u_{00}.

The correlation statistics can be defined in terms of the above orientation vectors. From eqns (5.8) and (5.18), they are defined for each MFT vector $\mathbf{u}_{xy}(n)$ as

$$R_{xyi}(n) = \mathbf{z}^{+}_{xyi}(n)\ \mathbf{X}(n)\ \mathbf{z}_{xyi}(n) \qquad (6.4)$$

$$C_{xyi}(n) = \frac{2}{\Gamma_n} \sum_{m=0}^{\Gamma_n/2-1} \mathbf{w}^{+}_{xyim}(n)\ \mathbf{X}(n)\ \mathbf{Y}(n)\ \mathbf{w}_{xyim}(n) \qquad (6.5)$$

where $\mathbf{X}(n)$ is the non-circular shift operator

$$x_{kl}(n) = \delta(k+1-l) \qquad 0 \le k,l < \Gamma_n/2 \qquad (6.6)$$

and $\mathbf{Y}(n)$ represents the weighting function in eqn (5.18). In the current implementation this is defined to be

$$\upsilon_{kl}(n) = \begin{cases} \delta(k-l) & \frac{\Gamma_n}{4} - a_n \le k \le \frac{\Gamma_n}{4} + a_n \\ 0 & \text{else} \end{cases} \qquad (6.7)$$

where a_n is given by

$$a_n = \begin{cases} 1 & n \geq n_1 \\ 2^{n_1-n} & n_2 \leq n < n_1 \qquad n_1 > n_2 \\ 2^{n_1-n_2} & n < n_2 \end{cases} \tag{6.8}$$

and sets the range of coefficients for calculating $C_{xyi}(n)$ according to the spatial frequency (and thus orientation) resolution of $\mathbf{u}_{xy}(n)$, ie it increases the number of coefficients as the resolution increases (cf section 5.2).

In terms of the MFT coefficients, eqns (6.4) and (6.5) become

$$R_{xyi}(n) = \mathbf{u}^{+}_{xy}(\sigma,n)\ \mathbf{A}^{+}_{i}(n)\ \mathbf{X}(n)\ \mathbf{A}_{i}(n)\ \mathbf{u}_{xy}(\sigma,n) \tag{6.9}$$

$$C_{xyi}(n) = \frac{2}{\Gamma_n} \sum_{m=0}^{\Gamma_n/2-1} \mathbf{u}^{+}_{xy}(\sigma,n)\ \mathbf{B}^{+}_{im}(n)\ \mathbf{X}(n)\ \mathbf{Y}(n)\ \mathbf{B}_{im}(n)\ \mathbf{u}_{xy}(\sigma,n) \tag{6.10}$$

For an orientation ψ_i, a certainty measure for a feature is indicated by the magnitude of $R_{xyi}(n)$. The position of this feature is then derived from the arguments of the statistics $R_{xyi}(n)$ and $C_{xyim}(n)$ according to eqns (5.6) and (5.17), ie from fig 4.6

$$\eta_{xyi}''(n) = \eta_{xyi}(n) + \eta_{xyi}'(n) \tag{6.11}$$

where

$$\eta_{xyi}(n) = [\, \eta_{xyi}(n) \cos\psi_i \,,\, \eta_{xyi}(n) \sin\psi_i \,] \tag{6.12}$$

$$\eta'_{xyi}(n) = [\, -\eta'_{xyi}(n) \sin\psi_i \,,\, \eta'_{xyi}(n) \cos\psi_i \,] \tag{6.13}$$

and

$$\eta_{xyi}(n) = \frac{\Gamma_n}{2\pi} \operatorname{Arg}[\, R_{xyi}(n) \,] \qquad \eta'_{xyi}(n) = \frac{\Gamma_n}{2\pi} \operatorname{Arg}[\, C_{xyi}(n) \,] \tag{6.14}$$

A measure of the distribution of features and an estimate of the principal orientation within the region can then be obtained from the inertia tensor **P** described in section 5.3.2. Having obtained this orientation from the eigenvector **e**, say, then an estimate of the 'principal feature' position can be derived from the additional statistics

$$R_{xy\theta}(n) = \mathbf{z}^{+}_{xy\theta}(n)\ \mathbf{X}(n)\ \mathbf{z}_{xy\theta}(n) \tag{6.15}$$

$$C_{xy\theta}(n) = \frac{2}{\Gamma_n} \sum_{m=0}^{\Gamma_n/2-1} \mathbf{w}^{+}_{xy\theta m}(n)\ \mathbf{X}(n)\ \mathbf{Y}(n)\ \mathbf{w}_{xy\theta m}(n) \tag{6.16}$$

where

$$\theta = \tan^{-1}\left[\frac{e_1}{e_0}\right] \qquad 0 \le \theta < \pi \tag{6.17}$$

and $\mathbf{z}_{xy\theta}(n)$, $\mathbf{w}_{xy\theta m}(n)$ are the vectors in the principal orientation θ and $\theta + \pi/2$ respectively.

6.2.2. Phase Ambiguity

The correlation statistics determine the average phase increment between adjacent components in a given orientation. This increment is then directly proportional to the position of the feature. However, due to the inherent periodicity of phase values, there exists a problem of ambiguity in determining this position. In fact, that derived from the principal value of the phase may equally refer to a number of other possibilities. To see this, note from eqn (6.14) that the phase increments, $\Delta\phi$ say, are given by

$$\Delta\phi = \frac{2\pi}{\Gamma_n}(\eta + m\,\Gamma_n) \qquad (6.18)$$

where η is the required position value. Therefore, as shown in fig 6.3a for the 1-d case, a given $\Delta\phi$ corresponds to the set of values $\eta + m\Gamma_n$.

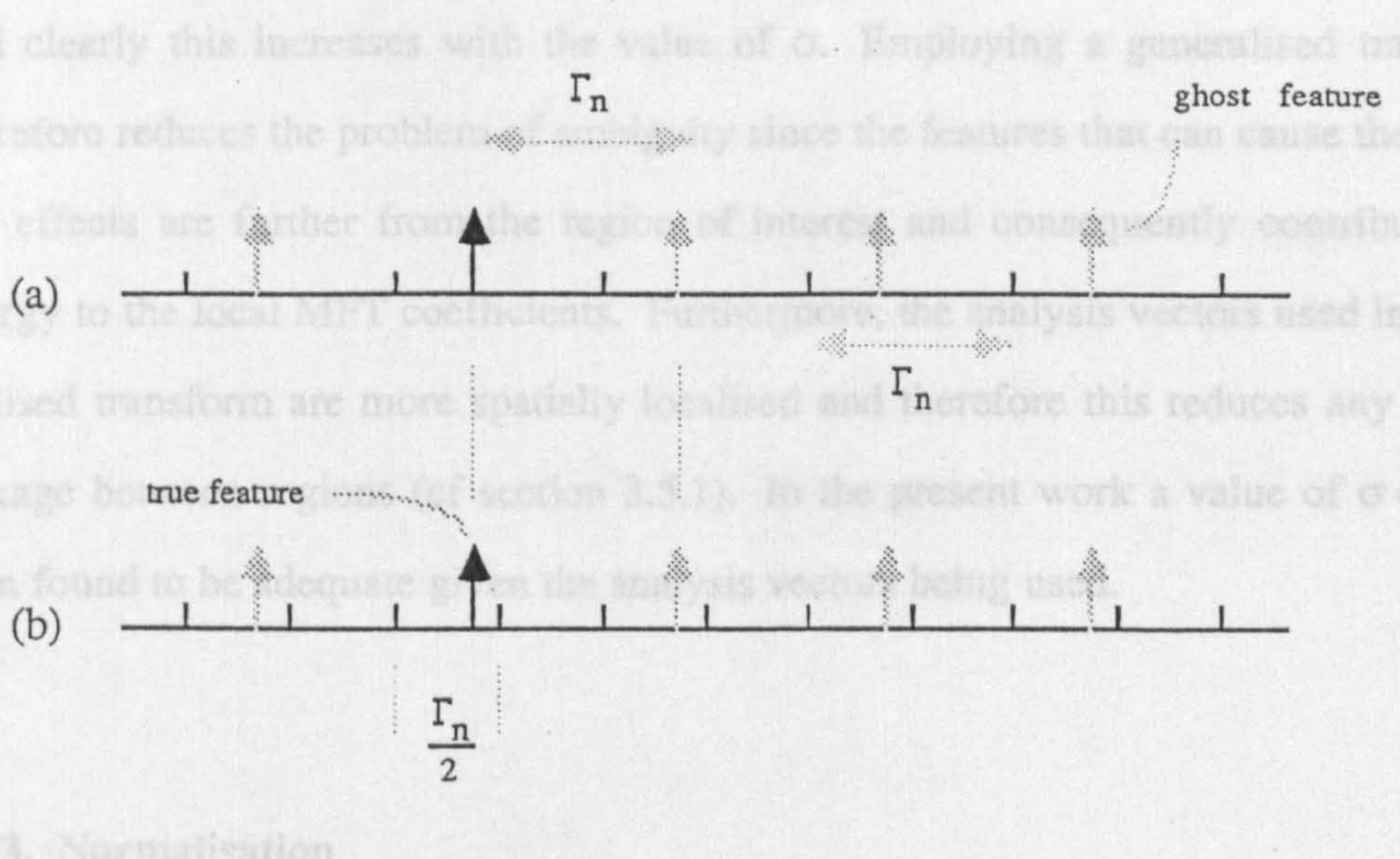


Figure 6.3. Feature offset repetition due to phase ambiguity.

Of course, since the MFT analysis vectors are localised, significant contribution in the correlation statistics will be due to those features in the vicinity, ie either within the

region itself or in an adjacent one. However, this still leaves the possibility of ghosting in regions that are next to a particularly strong feature. A way of reducing this is to make use of the generalised MFT. To see this, compare the example of fig 6.3a with that of fig 6.3b which shows the offset repetition for a generalised MFT with $\sigma = 2$. The region size in these two examples is Γ_n and $\Gamma_n/2$ respectively. However, since the frequency resolution is the same, the offset repetition remains at intervals of Γ_n in both cases. Now, if it is assumed that an offset value for a region is discarded if it falls outside of the boundary, then what is the minimum distance between a feature that gives rise to a 'ghost feature' and the centre of the region? Denoting this by Δ, it is given by

$$\Delta = \Gamma_n \left(1 - \frac{1}{\sqrt{2}\sigma} \right) \tag{6.19}$$

and clearly this increases with the value of σ. Employing a generalised transform therefore reduces the problem of ambiguity since the features that can cause the ghosting effects are further from the region of interest and consequently contribute less energy to the local MFT coefficients. Furthermore, the analysis vectors used in a generalised transform are more spatially localised and therefore this reduces any energy leakage between regions (cf section 3.5.1). In the present work a value of $\sigma = 2$ has been found to be adequate given the analysis vectors being used.

6.2.3. Normalisation

A feature which is less localised will have a correspondingly lower contribution in the correlation statistics. However, is such a feature less important than a well defined neighbour? This is a difficult question and involves issues such as the use of

thresholding and the importance of context. Nonetheless, a reasonable way to proceed is to employ some type of local normalisation.

In its simplest form this would consist of scaling all the measurements within a given region with respect to the maximum in that region. However, when regions have a small area then this would disregard the existence of larger regions containing a single feature. A compromise is to employ a *hierarchical normalisation* process which can be readily defined on the structure of the MFT. The basic idea is that the normalisation factor for a given region is derived not only from the measurements of that region but also from those at higher scales. In other words, within the MFT the measurements at child nodes are normalised to a certain extent by their ancestor nodes (cf section 3.4.3).

One such scheme can be defined in terms of a recursive process. Define the normalisation factor for a region xy at scale index n as

$$\bar{R}_{xy}(n) = (1-\beta)\,\bar{R}_{x'y'}(n-1) + \beta \sum_i |R_{xyi}(n)| \qquad n>0 \quad 0\le\beta\le 1 \qquad (6.20)$$

$$\bar{R}_{xy}(0) = \sum_i |R_{xyi}(0)| \qquad (6.21)$$

where

$$x' = \lceil \frac{x}{2} \rceil \qquad y' = \lceil \frac{y}{2} \rceil \qquad (6.22)$$

The correlation statistics are then modified such that

$$R'_{xyi}(n) = \frac{R_{xyi}(n)}{\bar{R}_{xy}(n)} \qquad (6.23)$$

The value of β in eqn (6.20) determines the 'size' of the local normalisation. For example, a small value will mean that the magnitude of the statistics are influenced by features in a wider surrounding area. As β increases this area reduces until when $\beta = 1$ it is simply a region by region normalisation.

6.2.4. Computational Requirements

Assuming the availability of an appropriate MFT, the additional calculations required to determine the correlation statistics are considered in this section.

For each spatial frequency vector the set of orientation vectors $\mathbf{z}_{xyi}(n)$ and $\mathbf{w}_{xyim}(n)$ need to be calculated. Assuming that these are calculated in Γ_n/L orientations and that each component is derived from a bilinear interpolation which requires the equivalent of 2 complex multiplications, the number of complex multiplications per spatial frequency vector is then (cf eqns (6.1),(6.2) and (6.5))

$$\begin{aligned} \mathcal{N}_I &= \frac{2\Gamma_n}{L}\left[\frac{\Gamma_n}{2} + \frac{\Gamma_n}{2}(2a_n + 1)\right] \\ &= \frac{2\Gamma_n^2}{L}(a_n + 1) \end{aligned} \tag{6.24}$$

where a_n is given by eqn (6.8).

From eqns (6.4) and (6.5), the number required to calculate the correlation statistics $R_{xyi}(n)$ and $C_{xyi}(n)$ is

$$\mathcal{N}_R = \frac{\Gamma_n}{L}\left(\frac{\Gamma_n}{2} - 1 + 2a_n\frac{\Gamma_n}{2}\right)$$

$$= \frac{\Gamma_n^2}{L}\left(a_n + \frac{1}{2} - \frac{1}{\Gamma_n}\right) \qquad (6.25)$$

Hence for a given spatial frequency vector

$$\mathcal{N}_{xy}(n) = \mathcal{N}_I + \mathcal{N}_R \qquad (6.26)$$

and for the $\sigma^2\,\Omega_n^2$ on each level of a generalised transform

$$\mathcal{N}(n) = \sigma^2\,\Omega_n^2\;\mathcal{N}_{xy}(n)$$

$$= \sigma^2\,\Omega_n^2\,(\,\mathcal{N}_I + \mathcal{N}_R\,)$$

$$= \sigma^2\,\Omega_n^2\left[\frac{2\Gamma_n^2}{L}(a_n+1) + \frac{\Gamma_n^2}{L}\left(a_n+\frac{1}{2}-\frac{1}{\Gamma_n}\right)\right]$$

$$= \frac{\sigma^2 M^2}{L}\left(3a_n + \frac{5}{2} - \frac{1}{\Gamma_n}\right)$$

$$\approx \frac{\sigma^2 M^2}{2L}\left(6a_n + 5\right) \qquad (6.27)$$

For example, using a generalised MFT with $\sigma = 2$ and estimating $\Gamma_n/2$ orientations per region (L=2), this gives

$$\mathcal{N}_0(n) \approx 6\,a_n + 5 \qquad (6.28)$$

complex multiplications per pixel. In the present work a_n has been limited to a maximum value of 2, ie $\mathcal{N}_0(n) \approx 17$, and has led to satisfactory results.

6.3. Image Reconstruction

An appropriate method of establishing the validity of a signal model is to reconstruct the signal from the parameter estimates. In the present case, this is particularly straightforward given the invertibility of the MFT.

The idea is to reconstruct the relevant MFT coefficients and then use the inverse procedure to provide a reconstructed image. In the multiresolution image model, the parameters obtained refer to local features defined at different spatial resolutions and these are determined from the spatial frequency vectors on different levels of the MFT. Hence to produce a reconstructed image, these vectors are first reconstructed and the resulting pseudo-MFT levels inverted using the multilevel inverse procedure described in section 3.3.2.

For each local feature in a given orientation, the correlation statistics used in the estimation scheme provide an estimate of the linear phase increment in an orthogonal orientation within the relevant spatial frequency vector. The phase values of the coefficients can therefore be reconstructed by using the linear model defined in section 4.4.2 with the phase increment given by the estimated value. The linear model also assumes a constant phase component, ie ε in eqn (4.11), and this can be calculated at the same time as the correlation statistics and then added in at the reconstruction stage.

However, the correlation statistics give no indication of the magnitude distribution along each orientation, although they do provide a measure of the energy variation over all orientations given the linear phase model. It is therefore necessary to model the magnitude distribution using some other method. This could consist of deriving a separate model for the distribution and estimating its parameters from the transform

coefficients, eg approximating it using an *n*th order polynomial. In the present work, however, a much simpler approach is adopted. The radial magnitude is modelled using a simple prototype function which is weighted according to the distribution of magnitudes within the correlation statistics. Despite its simplicity, this approach has been shown to give acceptable results.

The multiresolution image model is based upon regions which contain single local features. Hence the spatial frequency vectors must be reconstructed from information referring to a single orientation. This introduces the problem of modelling the response in the other orientations, ie how should the phase values in these orientations relate to those in the orientation corresponding to that of the feature and how should the magnitude values vary? Once again there are obviously methods that could be employed to estimate these parameters from the data having identified a single feature region. However, experiments have shown that a simple approach can also be taken in this case and yield satisfactory results. This involves assuming that the phase response varies in only the orientation of interest and that the magnitude response resembles the elongated region associated with a feature segment illustrated in fig 4.6. The implication of this approach is that the resulting feature after reconstruction will be centred about an orientation orthogonal to itself, ie there will be no offset along its orientation, and its length will be determined by the function which is used to produce the width of the elongated region. The results presented in the next section show that these effects do not produce gross errors in the reconstructed image.

A spatial frequency vector of the MFT is therefore reconstructed in the following manner. For a feature corresponding to an orientation θ detected at scale index n and within spatial region *xy*, the associated vector is reconstructed using the correlation statistics $R_{xy\theta}(n)$ such that

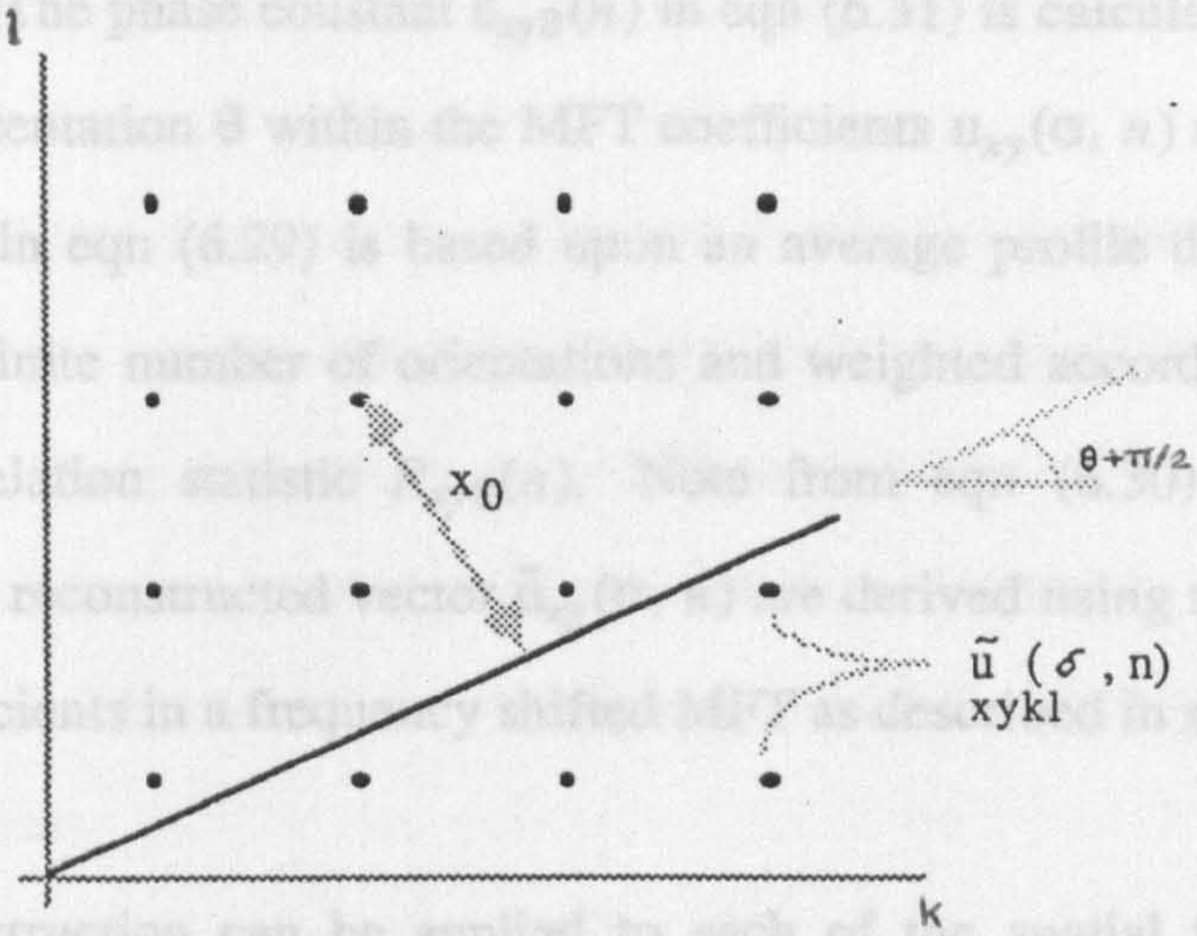


Figure 6.4. Reconstruction of MFT coefficients.

$$\tilde{u}_{xykl}(\sigma, n) = m_\theta(k,l)\, e^{-j\phi_{xy\theta}} \qquad 0 \le k < \Gamma_n \quad \frac{\Gamma_n}{2} \le l < \Gamma_n \tag{6.29}$$

and

$$\tilde{u}_{xykl}(\sigma, n) = \tilde{u}^{*}_{xyk'l'}(\sigma, n)\, e^{-j\frac{2\pi}{\Omega_n}(x+y)} \tag{6.30}$$

$$0 \le k < \Gamma_n \quad 0 \le l < \frac{\Gamma_n}{2} \quad k' = \Gamma_n - 1 - k \quad l' = \Gamma_n - 1 - l$$

where

$$\phi_{xy\theta} = \mathrm{Arg}\,[\,R_{xy\theta}(n)\,]\; x_0 + \tilde{\varepsilon}_{xy\theta}(n) \tag{6.31}$$

is the reconstructed phase value according to the linear model of section 4.4.2 and x_0 is the perpendicular distance between the coefficient and the orientation $\theta + \pi/2$ as

shown in fig 6.4. The phase constant $\tilde{\varepsilon}_{xy\theta}(n)$ in eqn (6.31) is calculated as the average constant in the orientation θ within the MFT coefficients $\mathbf{u}_{xy}(\sigma, n)$ and the magnitude function $m_\theta(k,l)$ in eqn (6.29) is based upon an average profile derived from these coefficients in a finite number of orientations and weighted according to the magnitude of the correlation statistic $R_{xy\theta}(n)$. Note from eqn (6.30) that half of the coefficients in the reconstructed vector $\tilde{\mathbf{u}}_{xy}(\sigma, n)$ are derived using the symmetry property of the coefficients in a frequency shifted MFT as described in section 3.6.1.

The above reconstruction can be applied to each of the spatial frequency vectors corresponding to the single feature regions detected by the scheme described in chapter 5. This will then result in a set of pseudo-MFT levels $\mathbf{s}(\sigma, n)$ with spatial frequency vectors given by

$$\mathbf{s}_{xy}(\sigma, n) = \begin{cases} \tilde{\mathbf{u}}_{xy}(\sigma, n) & \text{iff} \quad x,y \in \Lambda_n \\ \mathbf{0} & \text{else} \end{cases} \tag{6.32}$$

where $\mathbf{0}$ is the null vector and Λ_n denotes the set of spatial regions on level n of the multiresolution image model that contain single features. The reconstructed image is then given by

$$\mathbf{r} = \sum_{n \in \Lambda} \mathbf{H}(\sigma, n)\ \mathbf{s}(\sigma, n) \tag{6.33}$$

where Λ denotes the selected levels and $\mathbf{H}(\sigma, n)$ is the inverse operator defined in sections 3.2.1 and 3.5.3.

6.4. Experimental Results

The estimation and detection schemes described in this and the previous chapter were implemented for three images: the 'discs' in fig 3.12a; the 'girl' in fig 3.13a; and the 'boats' in fig 6.7a. These images are 512×512 pixels with an 8-bit grey level at each pixel. All three were prewhitened using the function defined in section 3.6.4 with $a = 2$ and $A = \pi/16$. The results are shown in figs 6.5-6.8 and they are considered in the following sections.

6.4.1. Local Feature Estimation

The estimation of local features using the MFT described in this chapter was implemented for the three images. Recall from section 6.2 that the scheme involves obtaining estimates in a number of uniformly distributed orientations within the spatial frequency vectors on each level of the transform. These estimates consist of the position of the feature, ie its centroid $\eta''_{xy}(n)$, and a certainty measure given by the magnitude of the correlation statistic $R_{xyi}(n)$.

The results for several MFT levels are shown in figs 6.5a-d, 6.6a-d, and 6.7b-e. Apart from fig 6.5d, these results were obtained using generalised MFT's with $\sigma = 2$. The estimated features on a given level are displayed by constructing an image in which each feature is represented by a straight line within the spatial region referred to the spatial frequency vector from which the estimate was obtained. This line is at the appropriate orientation and position and extends to the boundaries of the region. If the position estimate is such that it would render the line outside of the region, the feature is assumed to be in an adjacent region and the feature is not displayed. The luminance value of each line is then set to the magnitude of the correlation statistic $R_{xyi}(n)$. In

the case of the natural images, these values were first normalised using the hierarchical process described in section 6.2.3.

For all the images it can be seen that the scheme successfully identifies features at a given resolution provided that they satisfy the line or edge segment model. At higher levels of the MFT where the spatial resolution is greater (smaller region sizes), this condition is met by short segments and consequently fine detail is well represented. This detail is lost at lower levels, where the spatial resolution is reduced and the model is only appropriate for larger features. A good example of this can be seen in the 'discs' image. On level 3 (fig 6.5a), where the local region size is 32×32, only the larger discs are represented, whereas on level 5 (fig 6.5c) all the discs are well represented. This is also apparent in both of the natural images. The eyes and feathers are visible at the higher levels of the 'girl' image and only the larger features such as the mirror, hat, etc, at the lower levels. Similar points can be made concerning the results for 'boats' image.

To illustrate the problem of phase ambiguity and the advantages that can be gained when using a generalised MFT (cf section 6.2.2), fig 6.5d shows the feature estimates obtained from level 5 of an MFT with $\sigma = 1$ for the 'discs' image. The ghosting that results from phase ambiguity is clearly visible in this result; regions adjacent to edge features contain ghost features parallel to the true edge and at a distance which is a multiple of the region size Γ_n. Note that this is not apparent in fig 6.5b, which is the comparable level in terms of region size of an MFT with $\sigma = 2$.

The background noise in the estimates for the natural images results mainly from the normalisation process which was used. In order to ensure that interesting features with relatively low values are not missed, the normalisation process is employed to increase their value. However, the consequence of this is that spurious features are

also increased in value and therefore there exists a trade-off between identifying all the relevant features and being able to ignore the spurious ones. Although on each level this would appear to be a problem, the later results will show that the identification of the features corresponding to the overall image model is not adversely effected by this noise due to the requirement for scale consistency (cf section 5.3.3).

The results presented compare favourably with the recent developments in detecting such features, eg [22][65]. In particular, the method of detecting features at different resolutions may be compared with scale space methods [10]. It should also be pointed out that although based on different properties, the use of frequency domain characteristics has previously been shown to be successful, eg in the work of Knutsson [65]. However, it should also be emphasised that the overall approach in this work is that these features represent an intermediate stage and that the next task is to select those features that optimally represent the image according to the multiresolution image model of section 4.3. The results of implementing this process are considered in the next section.

6.4.2. Hierarchical Feature Detection

The image model defined in section 4.3 is based upon regions which contain single local features. In sections 5.3 and 5.4, a scheme was described which is designed to identify such regions from the results of the feature estimation presented in the previous section. The scheme is based upon principal orientation and scale consistency criteria, and takes the form of a recursive hierarchical process as defined in section 5.4. Once this process is complete, the result is a truncated quadtree in which the leaf nodes refer to contiguous regions of different sizes and these regions are classified as

either containing a single feature or as being a lowpass region.

The detection scheme was implemented for the three images using the feature estimates already presented. The results are shown in figs 6.5e, 6.6e, and 6.7f, where the single feature regions are indicated by a line within the relevant region at an appropriate orientation and position, and the lowpass regions are indicated by a fixed luminance value. Figures 6.6f and 6.7g show the single feature regions superimposed upon the original natural images.

The results illustrate the ability of the scheme to select regions containing a single feature and to classify lowpass regions. In the case of the 'discs' image all the features have been selected at an appropriate resolution and the lowpass regions have been successfully identified. For example, the larger disc is primarily represented by features from level 3 (cf fig 6.5a) and the smaller discs are represented by features from levels 4 and 5 (cf figs 6.5b-c). In particular, note that where a larger region contains more than one feature the scheme has successfully split the region until only single feature regions remain, eg in areas where two discs are close together. This is a good example of the scheme being able to select the resolution appropriate to a given area of the image.

Similar remarks can be made about the results obtained for the natural images. Thus the boundary of the hat in the 'girl' image is represented by features from level 3 (cf fig 6.6a), whilst the detail of the feathers and eyes are represented at level 6 (cf fig 6.6d). The splitting of regions is also apparent and this has been particularly successful at the junctions of curves. For example, the junction between the shoulder and the hair has been split until it is represented by four single feature regions. Note also that although in the feature estimates there was significant noise present in several areas, this has not led to a large number of single feature regions being detected in those

areas. This is due in part to the imposition of the scale consistency requirement, which in the case of spurious and random features is not generally satisfied.

However, although the majority of important local features have been suitably selected in the 'girl' image, there are a few anomalies and these are also apparent to a greater extent in the more complex 'boats' image. There are two main sources of error:

(i) Lowpass classification - although in general the classification of lowpass regions has been acceptable (eg the shoulder and background in the 'girl' image) there are cases where a lowpass region appears to correspond to a single feature. Examples of this are at the top of the hat and in parts of the column on the left hand side. However, examination of these areas in the original image will reveal that the discontinuities corresponding to the local features are not significant and thus detection is made more difficult. This can be verified by reference to the feature estimates in figs 6.6a-d and is a typical problem encountered when detecting local features. As mentioned in the previous section, the utilisation of a normalisation process reduces these problems to a certain extent, although there is necessarily a trade-off between detecting these relevant features and those that are spurious (the scale consistency requirement is not sufficient to reject all spurious features). Note, however, that although in this case the regions have been classified as lowpass, the information concerning the features in those regions is still available and can be used in further processing (cf section 5.5.1 and 6.4.3).

(ii) Complex regions - in a number of cases, particularly in the 'boats' image, the scheme has classified a region as containing a single feature when it appears to contain several features. There are two main reasons for this: features are either too close in terms of position and/or orientation with respect to the resolution of the respective

level of the MFT; or there exists a relatively strong feature amongst weaker ones. In these cases the simplicity of the model and the single feature detection criteria are not sufficient to separate the individual features. This problem will be discussed further in chapter 7.

Despite the above difficulties, the results of the feature detection have illustrated that the overall approach and the use of the MFT as an estimation tool is capable of providing multiresolution feature descriptions of an image. This has only been shown to a limited extent by other methods, eg the multiscale representations [1][70][108], and confirms the view that such descriptions can only be achieved if a sufficiently general image representation such as the MFT is used (cf section 7.3). In the next section, results are presented for the curve extraction scheme described in section 5.5 and these further illustrate the benefits that can be gained from this approach.

6.4.3. Curve Extraction

As described in section 4.3, the multiresolution image model enables the representation of curves in an image. Using the single feature detection results presented in the previous section, curves can be extracted from the image based upon this representation and using the extraction scheme defined in section 5.5.1. This is a hierarchical process that constructs curves in a fine-to-coarse analysis over the truncated quadtree resulting from the single feature detection (fig 5.4). Once the curves have been extracted they are represented in a piecewise manner by a set of local feature segments. These can then be used to derive a B-spline representation of the curves as described in section 5.5.2.

The results of the curve extraction are shown in figs 6.5f, 6.6g, and 6.7h. In these results separate B-spline curves are indicated by unbroken lines of fixed luminance value. Thus in the 'discs' image there are 11 curves each corresponding to a complete circle. In the 'girl' image the hat, mirror, shoulder, etc, are all represented by single isolated curves and similar examples can be found in the 'boats' image. The scheme has therefore been successful in extracting the major curves in these images. Note in particular that in a number of cases the scheme has filled in the gaps (lowpass regions) that are apparent in the detection results of the previous section. As noted in that section, although some regions may be classified as being lowpass, the information concerning the features in those regions is still available and can be used to complete a given curve as these results have illustrated. A good example of this is the top of the hat in the 'girl' image.

In the case of small and detailed curves within complex areas of the image, the scheme has been less successful. This is apparent in a number of cases for the 'boats' image. One reason for this is the reliance on the single feature detection results, which as noted in the previous section are often unreliable in complex areas. An additional problem associated with the curve extraction scheme is that the construction of a curve is based upon orientation and spatial position information. At higher levels of the MFT, which correspond to sufficiently short feature segments to represent detailed curves, both orientation and spatial information are necessarily at low resolution. Hence curve extraction is limited and possibly not an appropriate operation to be addressing in such areas. This is again related to the principle of uncertainty and will be discussed further in section 7.2.

As noted in section 4.2, there have been a number of methods proposed for identifying curves and boundaries. Of these, methods which are based upon some type of edge following are comparable to the techniques used in this work, eg [3][23][74][78]. The

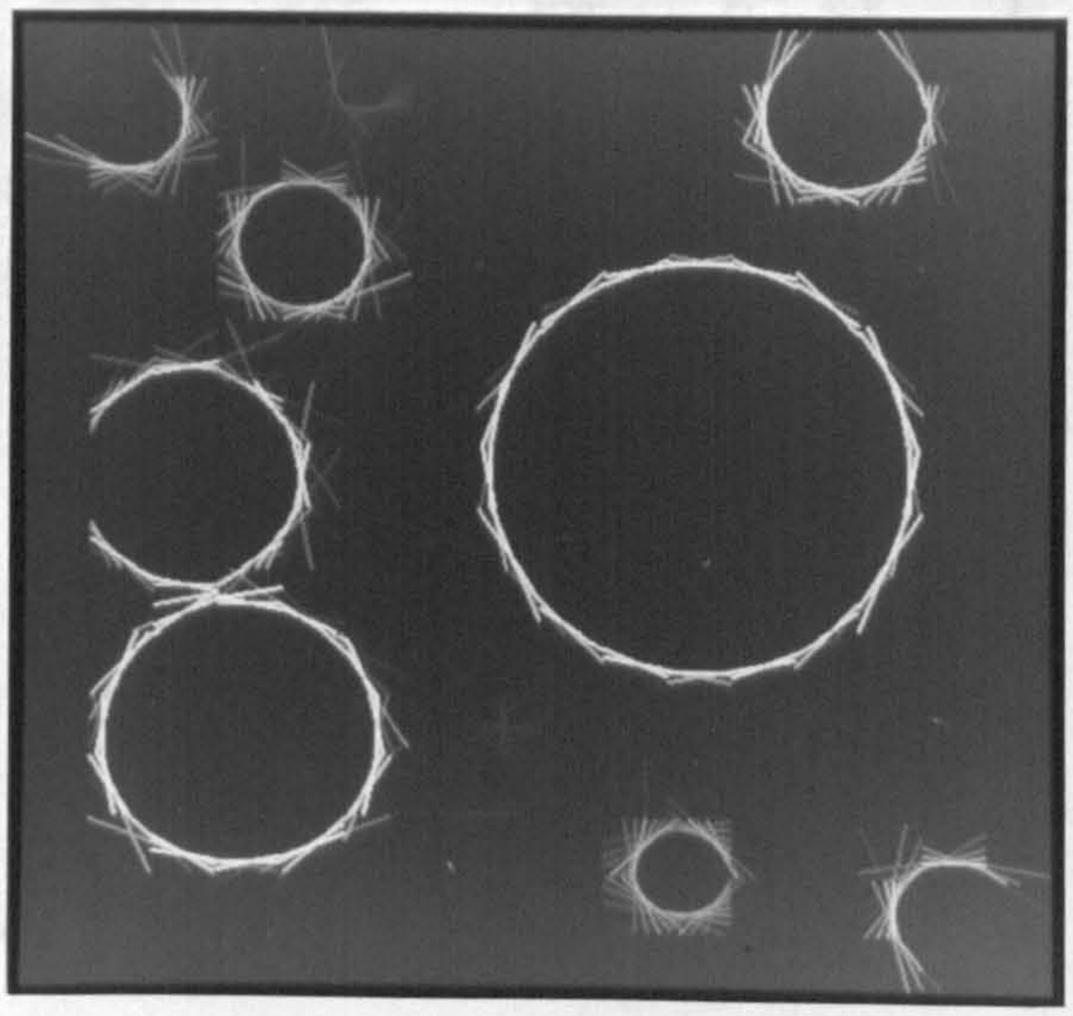
majority of work in this area has been done using either medical or satellite imagery [3][79] or 'object inspection' type images [74], although some work in image coding has used natural scenes, eg results are presented in [23] for the 'girl' image. The results presented here compare favourably with the latter.

6.4.4. Image Reconstruction

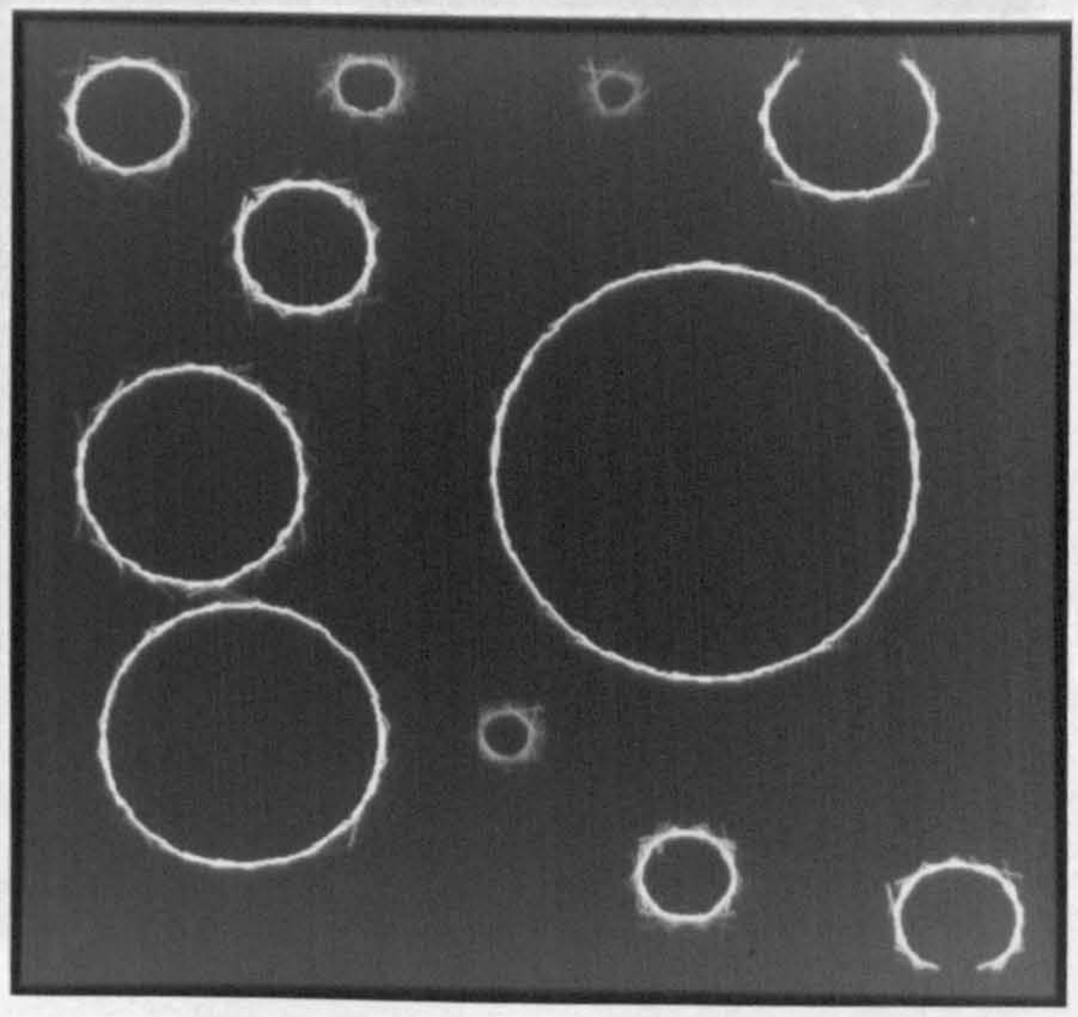
Section 6.3 described a method of reconstructing an image using local features defined at different spatial resolutions. The idea is to reconstruct the relevant levels of the MFT according to the local feature model using the parameters derived from the single feature detection scheme. The resulting pseudo-MFT is then inverted using the multilevel inverse procedure described in section 3.3.2. The scheme was implemented using the estimates for the 'girl' image presented in section 6.4.2. and the results are shown in figs 6.8a-f.

Figures 6.8a-c show separate reconstructions from the features on levels 4-6 and fig 6.8d shows a combined reconstruction from levels 3-6. Since the features are derived from prewhitened images which are effectively high pass filtered versions of the original (section 3.6.4), these reconstructed images are all high pass images. Despite this, it can be seen that the reconstructions clearly correspond to line and edge features at different resolutions and that this indicates the suitability of the model to represent local image features. To demonstrate this further, fig 6.8f shows the result of adding the high pass feature image in fig 6.8d to the lowpass image in fig 6.8e which approximates the lowpass version removed by the prewhitening process. Comparison of figs 6.8e and 6.8f reveals that the addition of the reconstructed edge features increases the sharpness of the former.

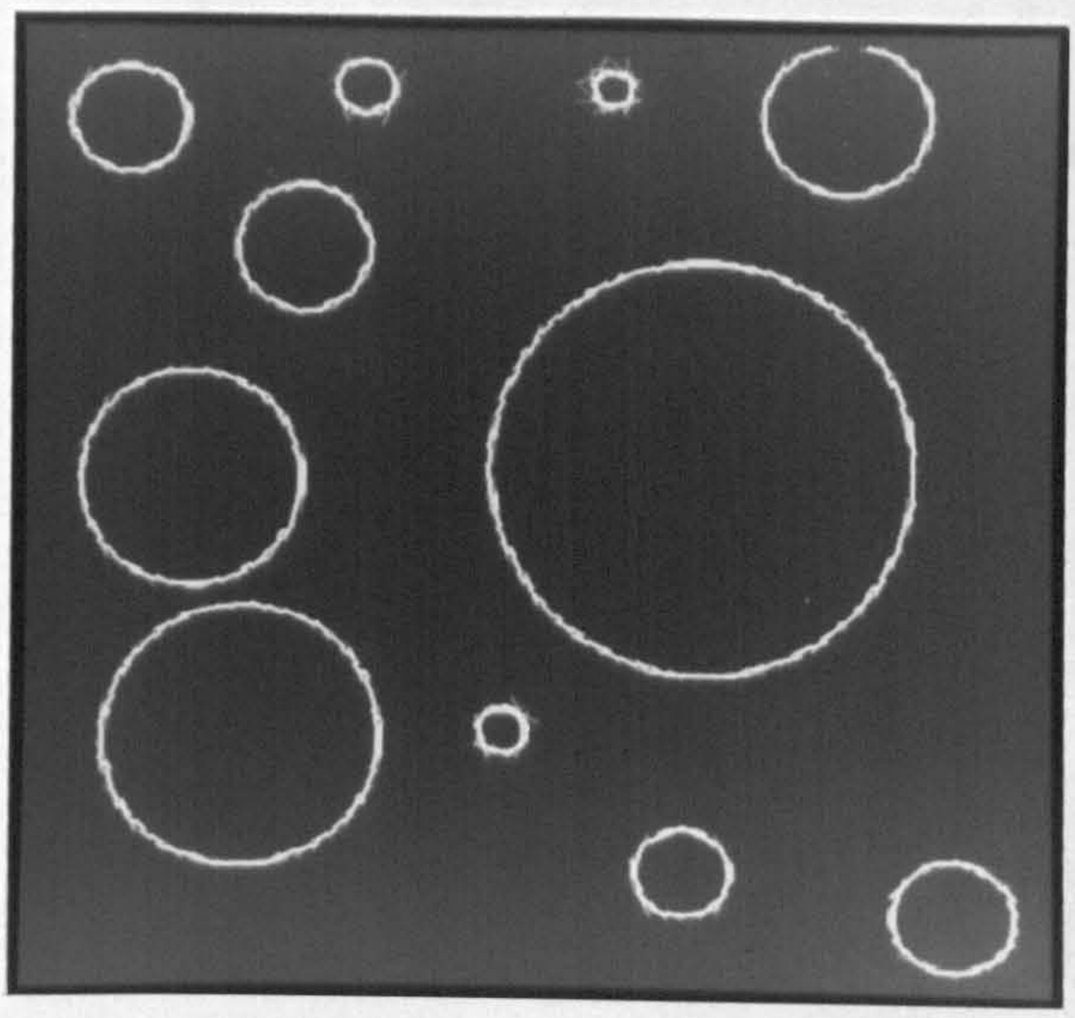
These reconstruction results are particularly important since they illustrate the usefulness of the MFT as an estimation tool. In particular, if the MFT is used to estimate the parameters of a given model, then the process of synthesising images based on those parameters can be achieved in a relatively straightforward manner and without any adverse effects caused by the structure of the transform.



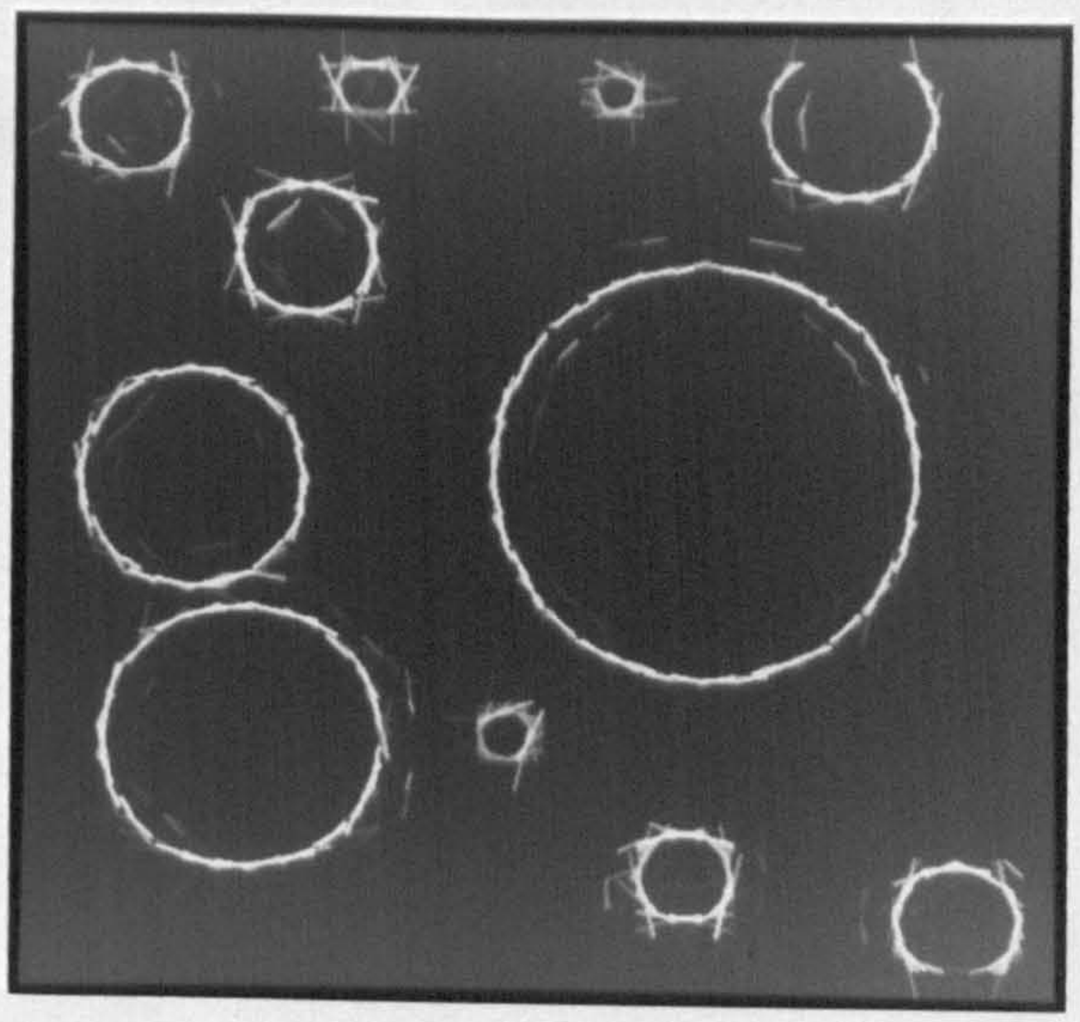
(a) Local feature estimates level 3.



(b) Local feature estimates level 4.

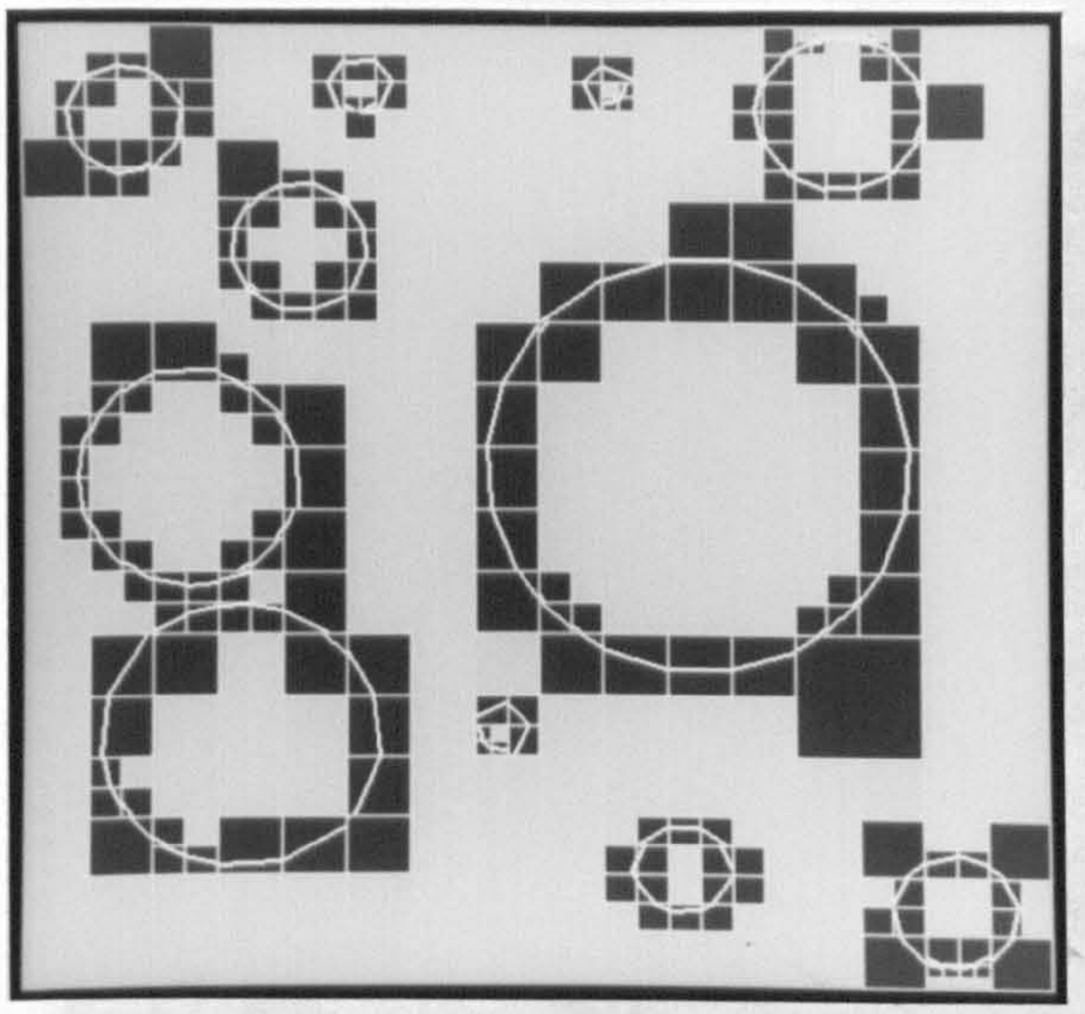


(c) Local feature estimates level 5.

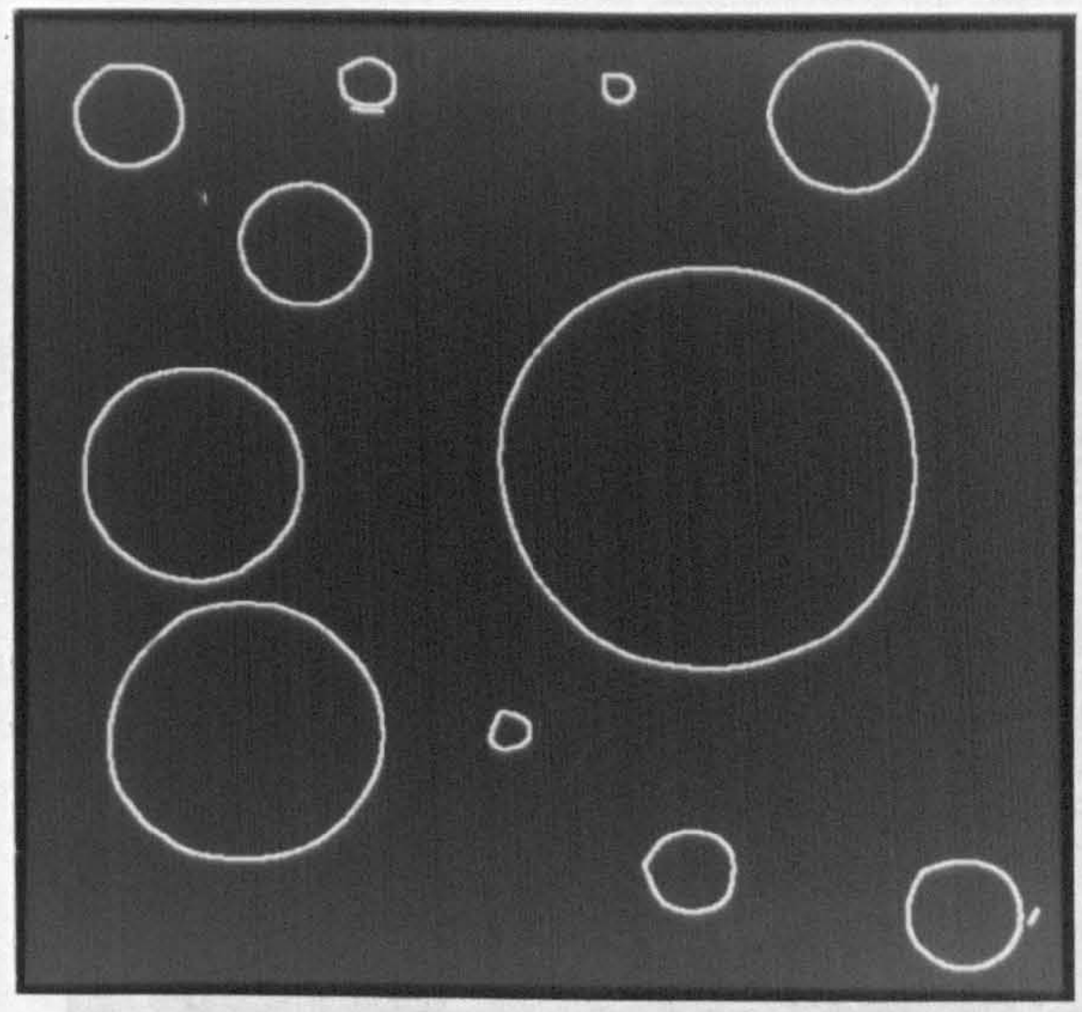


(d) Local feature estimates level 5 ($\sigma_{\cdot} = 1$).

Figure 6.5. Results for 'discs' image.



(e) Hierarchical feature detection.



(f) Curve extraction.

Figure 6.5. (cont) Results for 'discs' image.

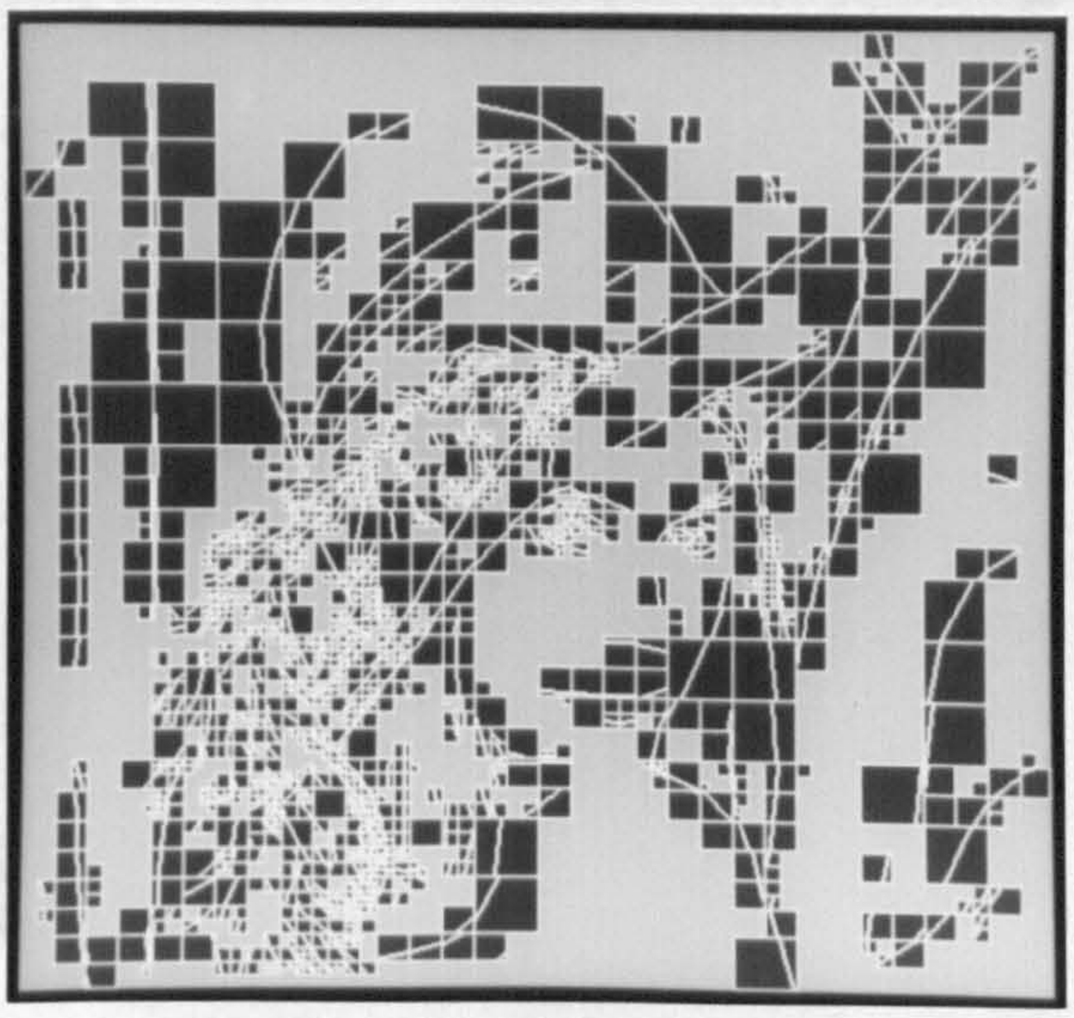
(a) Local feature estimates level 3.

(b) Local feature estimates level 4.

(c) Local feature estimates level 5.

(d) Local feature estimates level 6.

Figure 6.6 Results for 'girl' image.



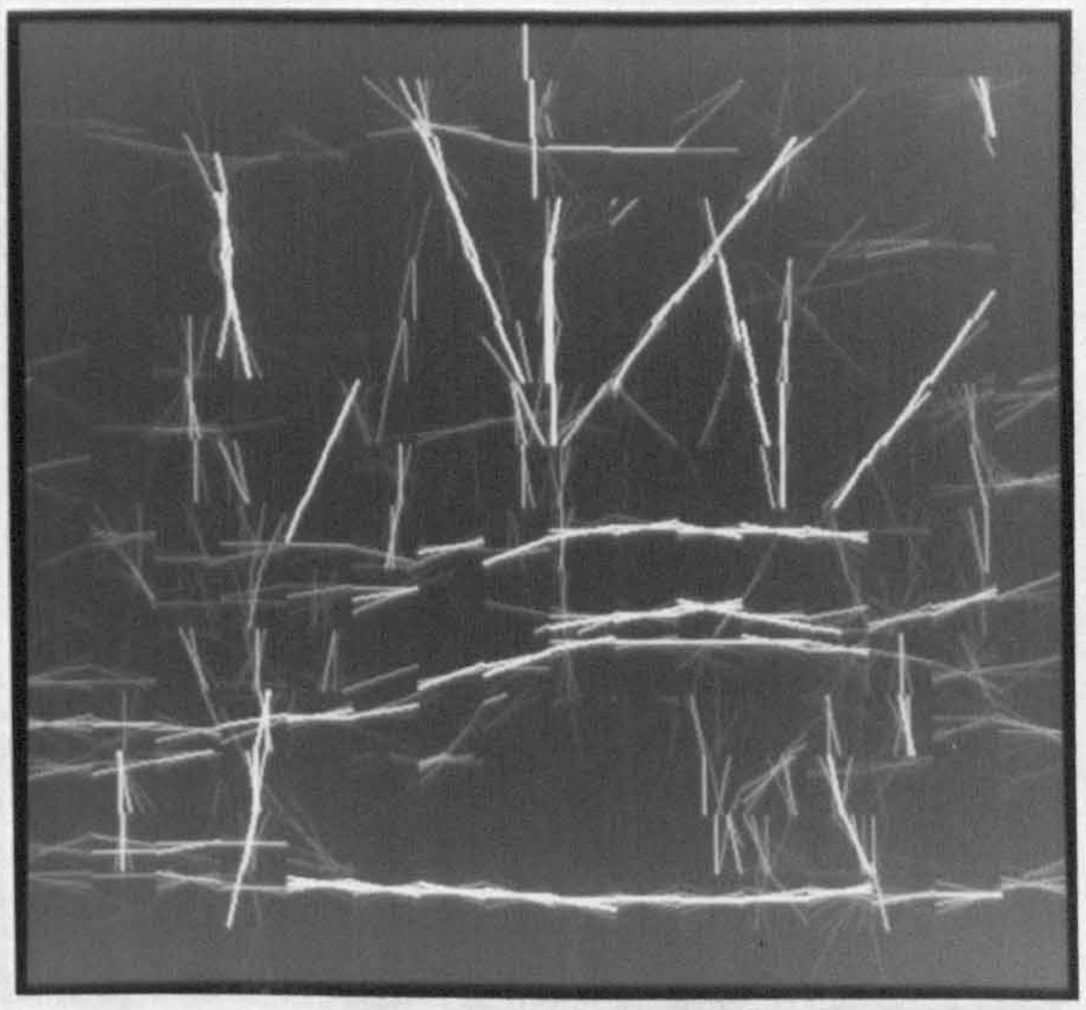
(e) Hierarchical feature detection.

(f) Hierarchical feature detection (overlay).

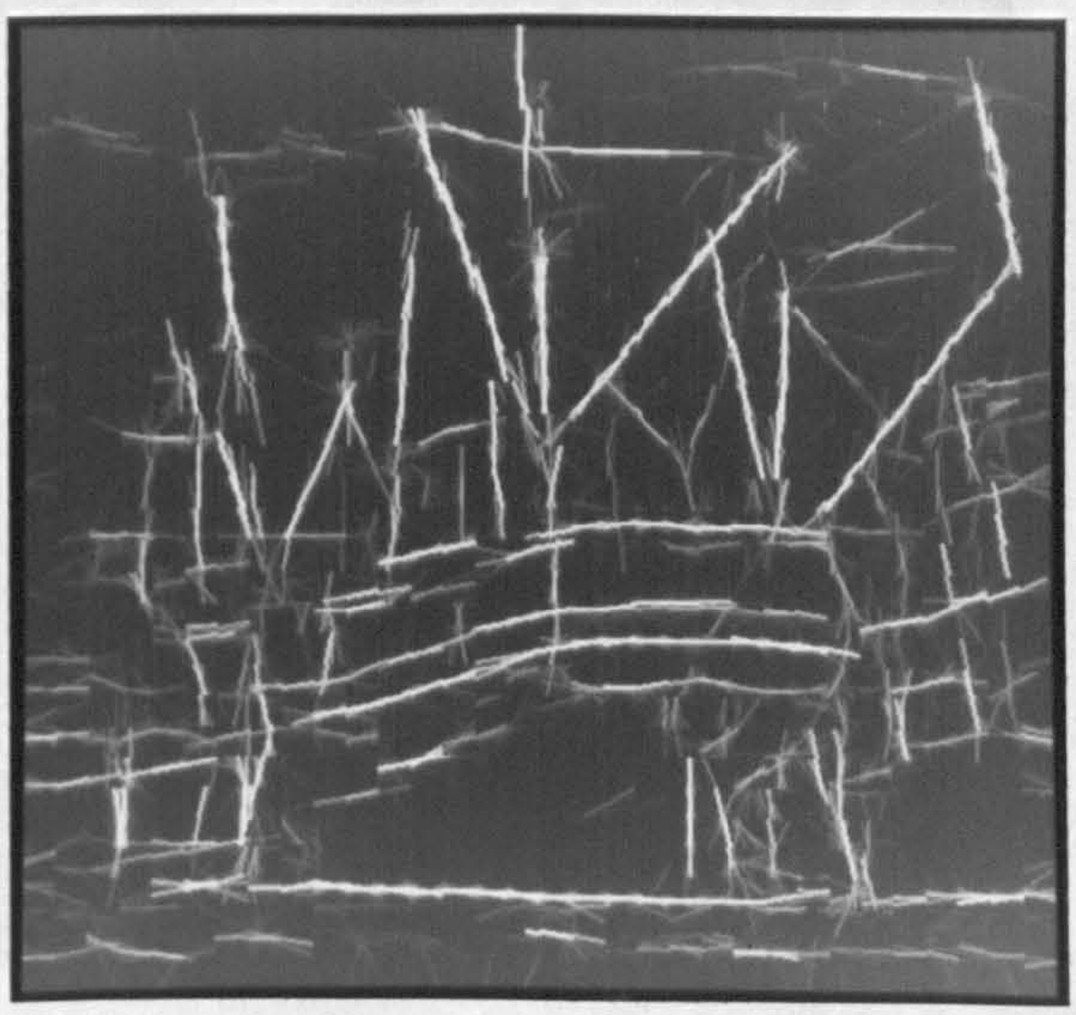
(g) Curve extraction.

Figure 6.6. (cont) Results for 'girl' image.

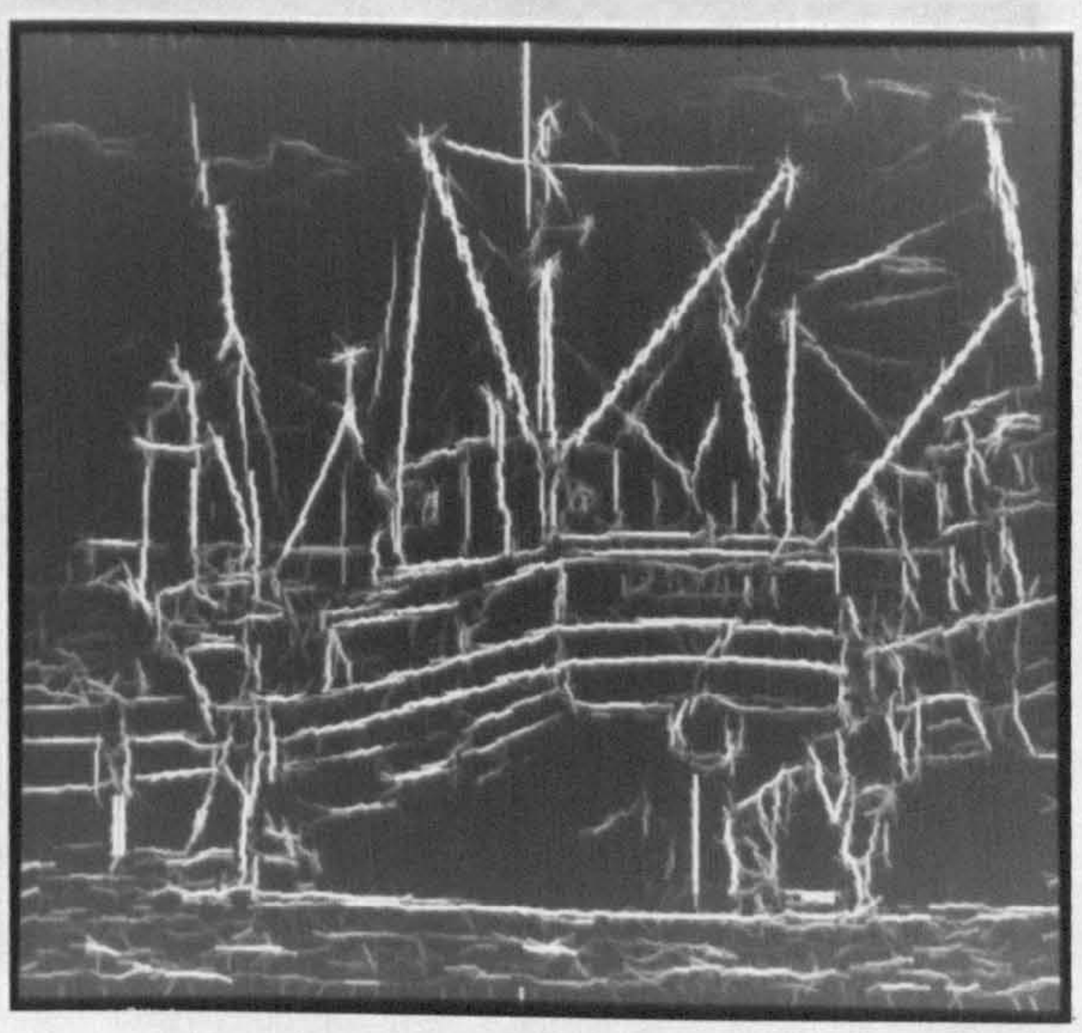
(a) Original image.



(b) Local feature estimates level 3.

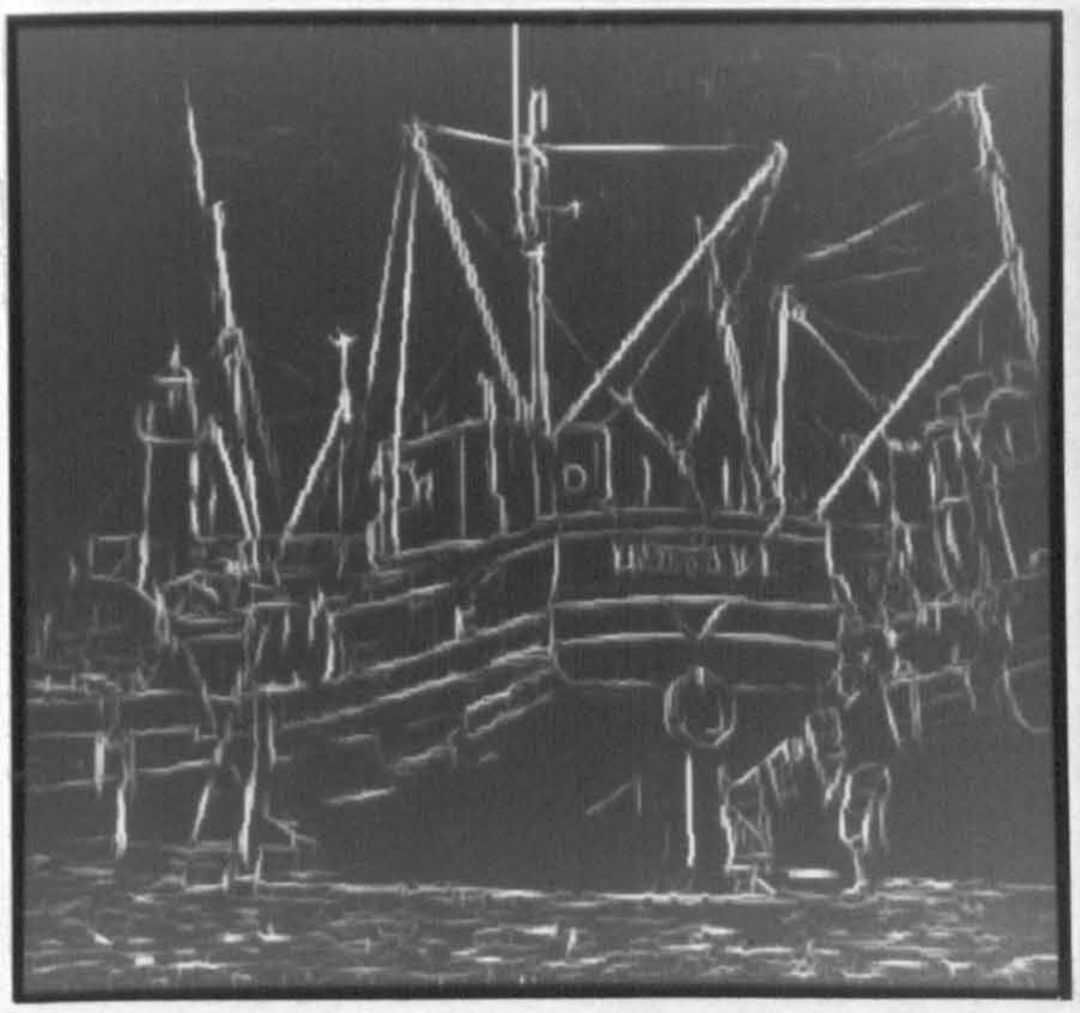


(c) Local feature estimates level 4.

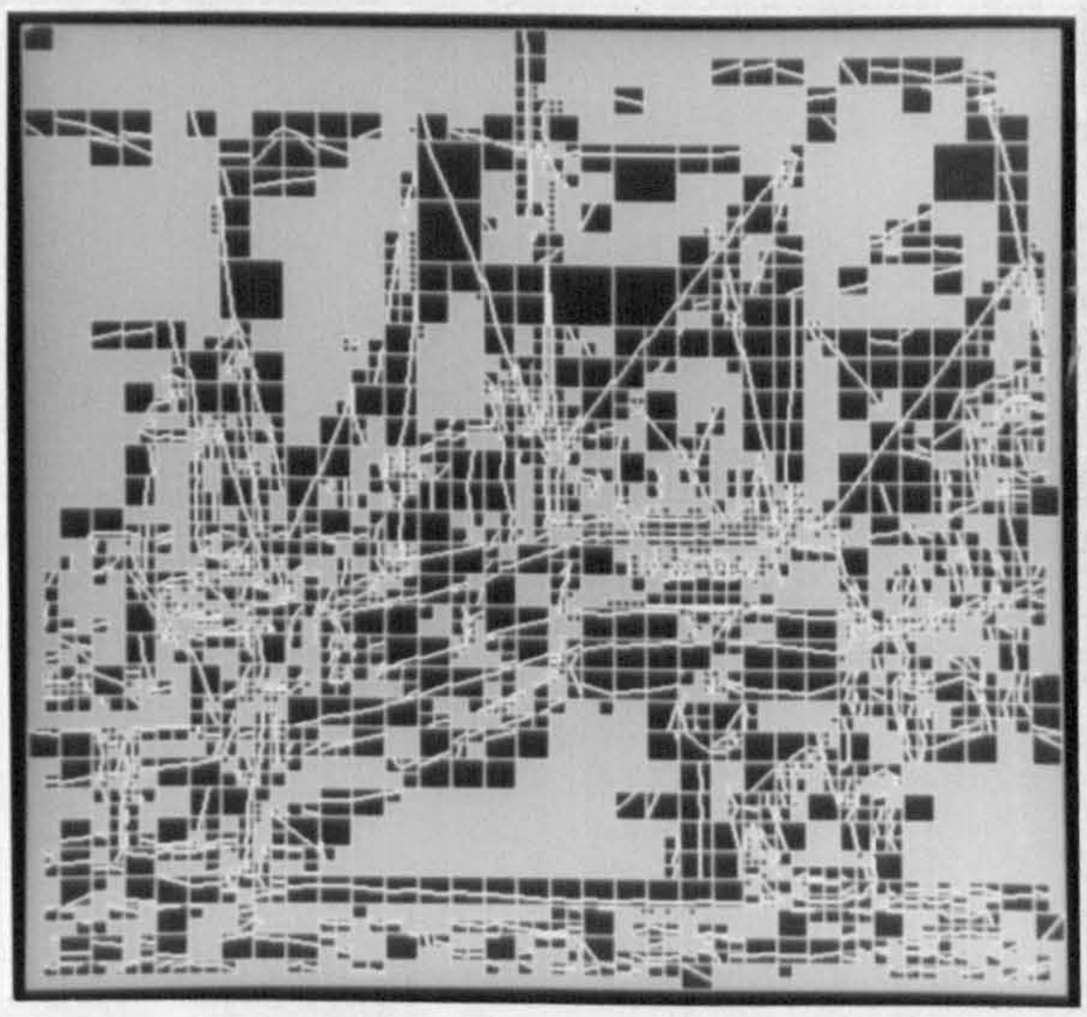


(d) Local feature estimates level 5.

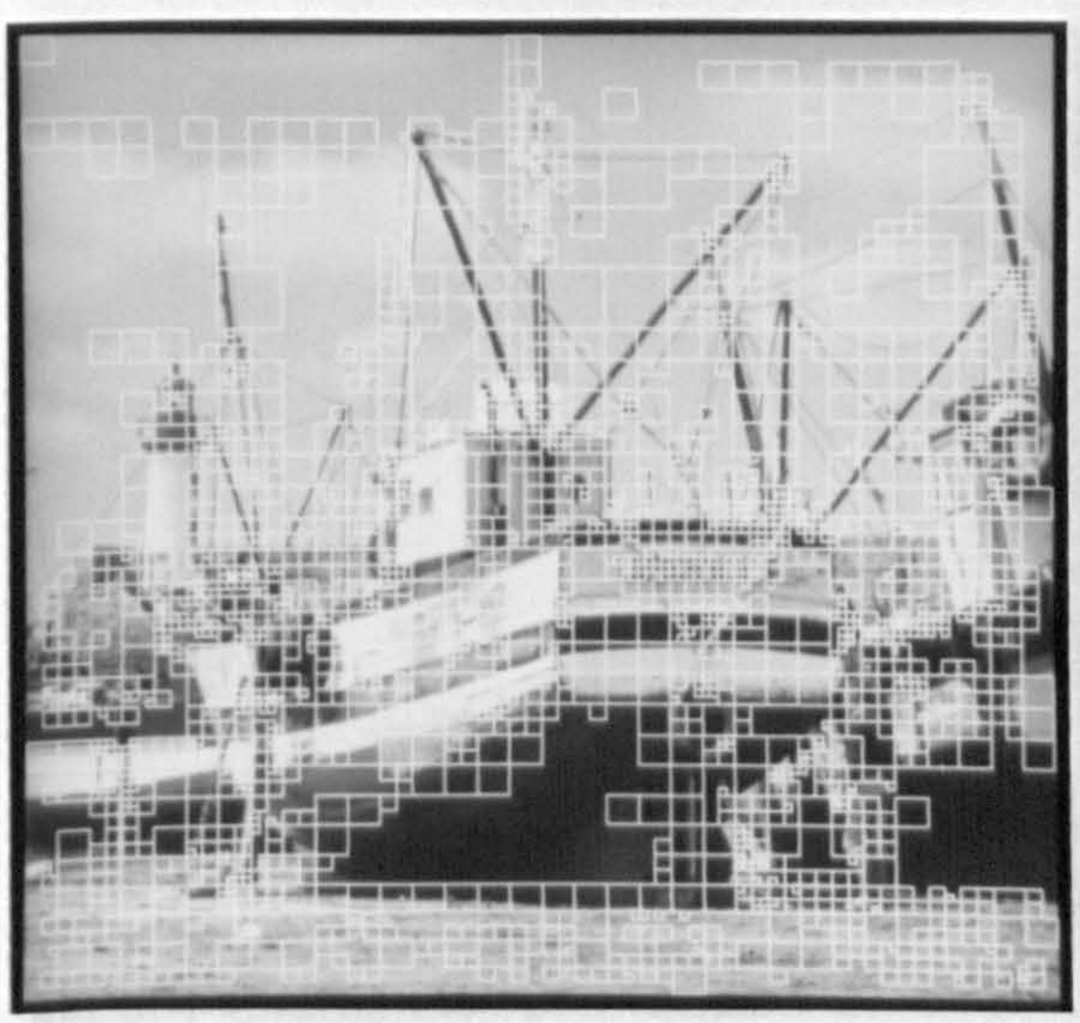
Figure 6.7. Results for 'boats' image.



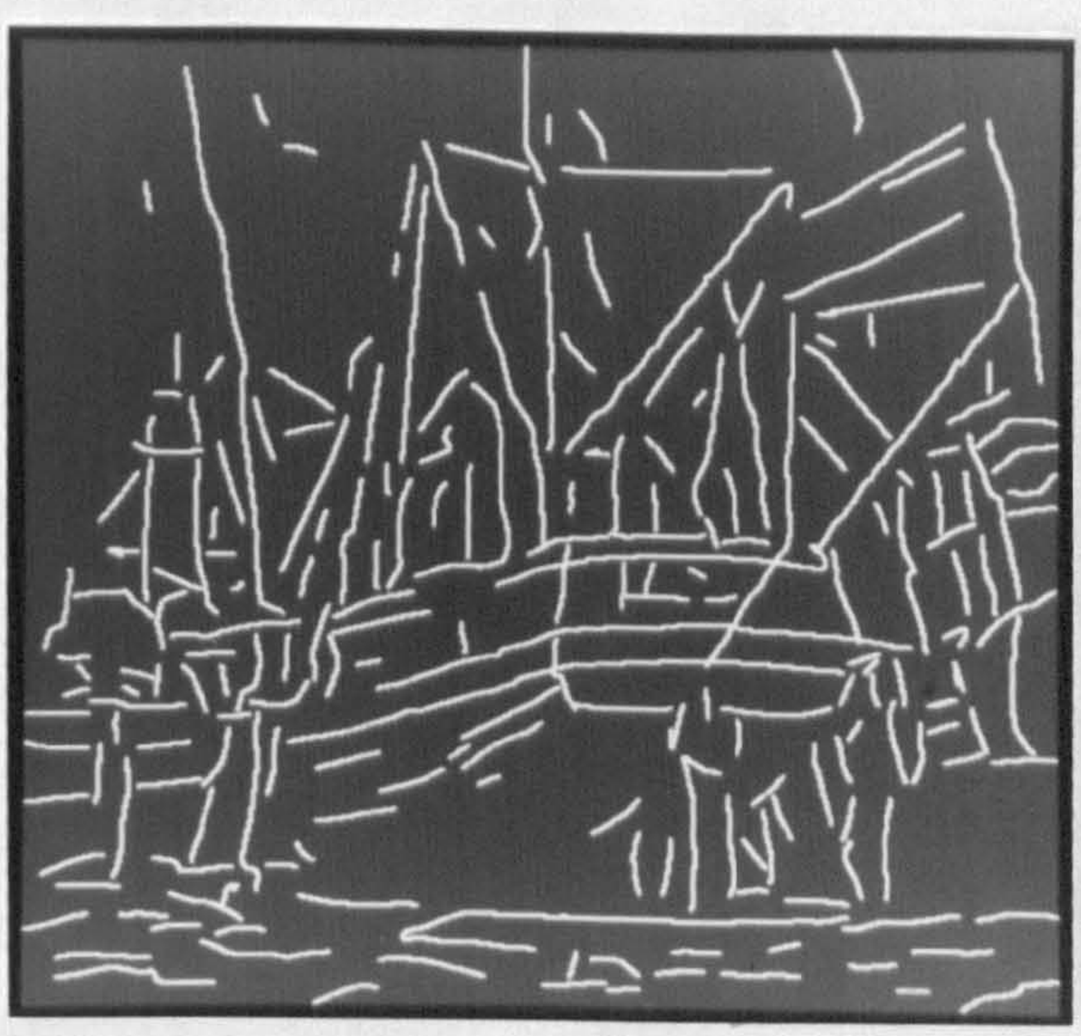
(e) Local feature estimates level 6.



(f) Hierarchical feature detection.

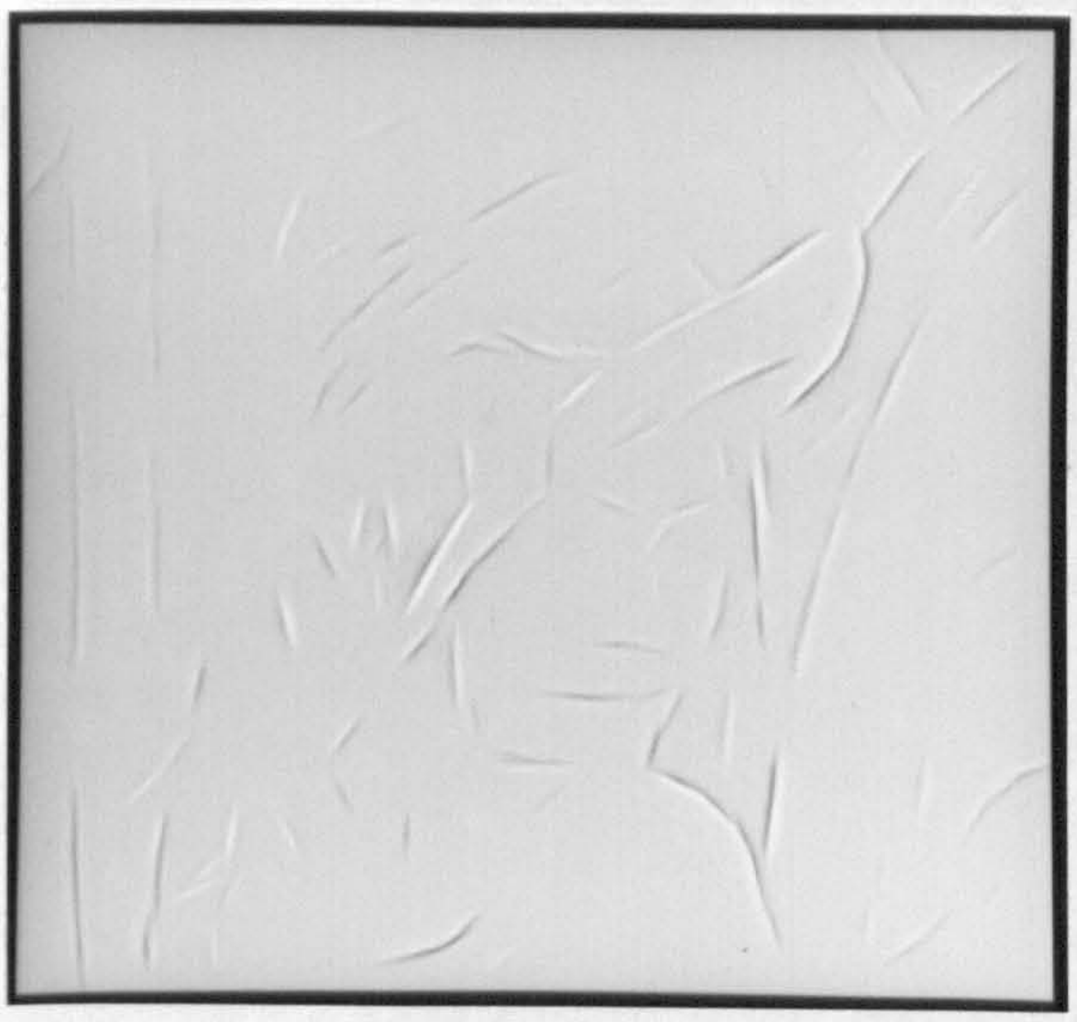


(g) Hierarchical feature detection (overlay).

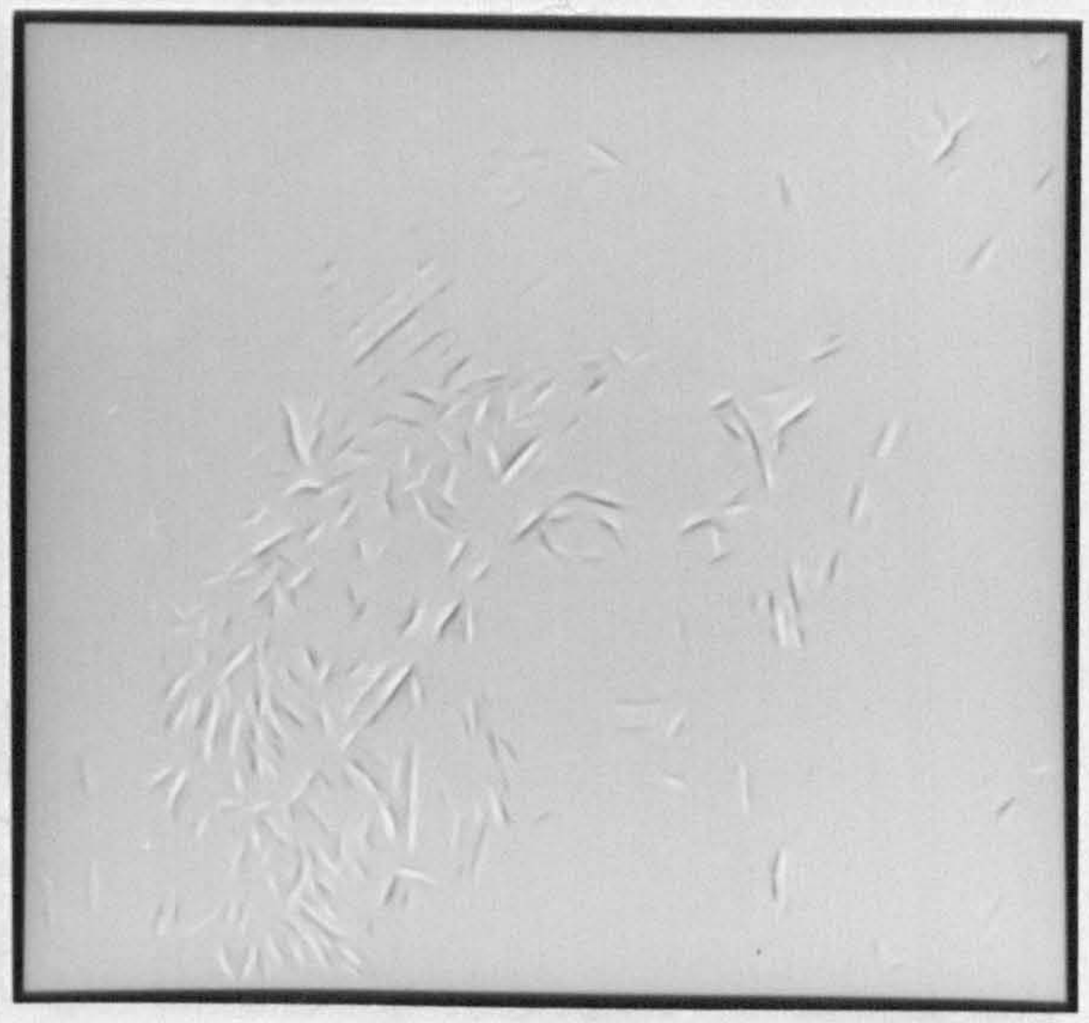


(h) Curve extraction.

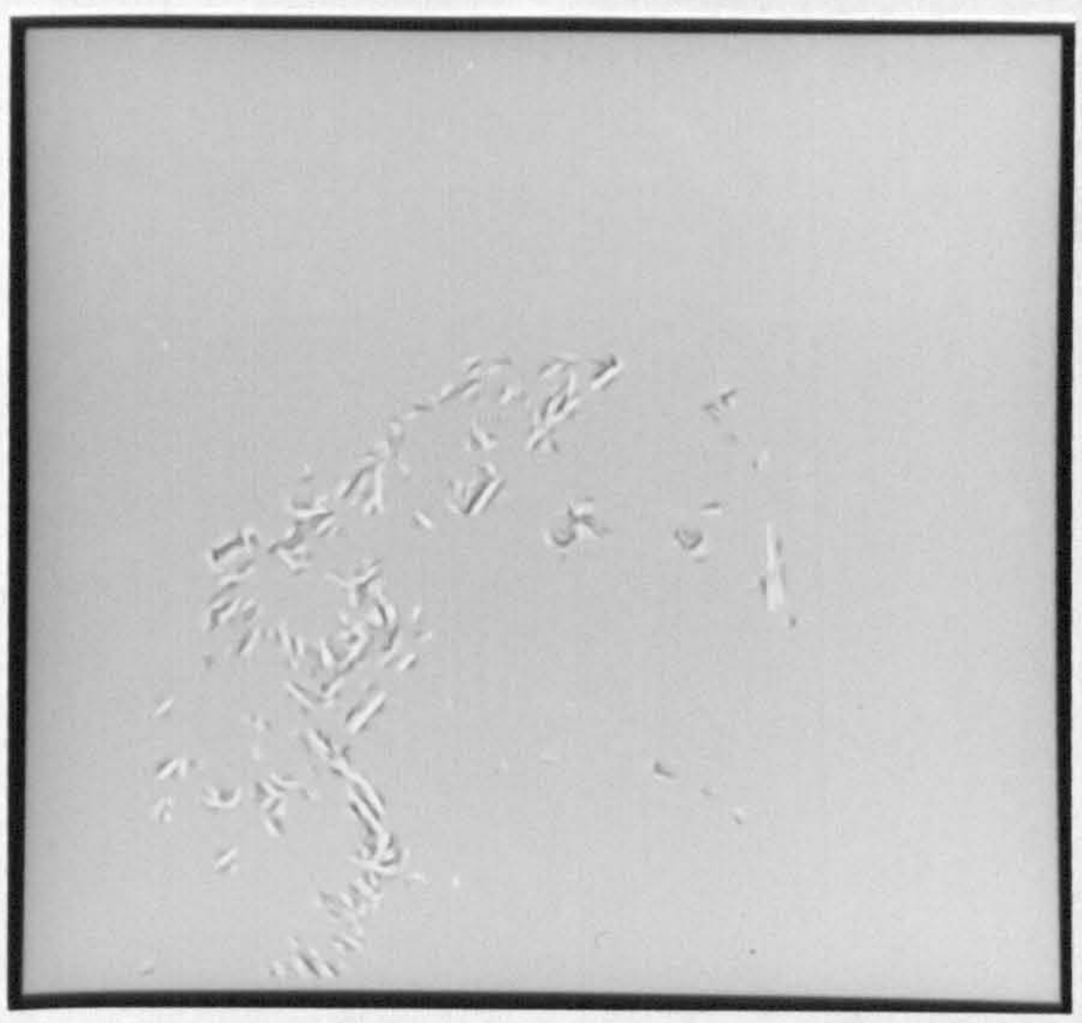
Figure 6.7. (cont) Results for 'boats' image.



(a) Reconstruction from level 4 features.



(b) Reconstruction from level 5 features.



(c) Reconstruction from level 6 features.

(d) Reconstruction from levels 3-6 features.

Figure 6.8. Image reconstruction results.

(e) Lowpass filtered image.

(f) Summation of lowpass image and reconstructed image in fig 6.8d.

Figure 6.8. (cont) Image reconstruction results.

CHAPTER SEVEN

CONCLUSIONS AND FURTHER WORK

The work described in this thesis has addressed the problem of defining an image description that could form the basis for a unified approach to feature extraction. It can be considered to consist of two parts: the derivation and definition of a suitable description in the form of the multiresolution Fourier transform (MFT); and the application of the transform to a typical image analysis problem. To conclude the thesis, it is appropriate to recall the principal arguments and results of these two areas and discuss their implications in a wider sense and with regard to the direction of further work.

7.1. A Unified Image Description

The importance of extracting meaningful features from an image in order to derive appropriate symbolic descriptions was discussed in chapter 1. Using the principles of pattern recognition methods, it was shown that the identification of such features enables the inference of structure and content from an image. Two specific and important examples of this were considered in greater detail: the use of local features such as lines and edges; and the use of regional features such as those associated with textural properties. Both of these are used extensively in image analysis, with applications ranging from 2-d segmentation to the analysis of motion and stereopsis.

It was noted, however, that the traditional methods used to extract these features exhibit a limitation when addressing general analysis problems. This relates to their relative locality in the spatial domain - the extraction of local features has been primarily

based upon local operations, whilst regional features have been extracted using essentially global methods. The limitation of this is that the two approaches represent a dichotomy, neither is suitable for both types of feature extraction. However, when dealing with general analysis problems, such as those presented by natural images, there is a need to extract a wide range of features and hence if an analysis is based upon these traditional methods it would mean using several different approaches. The result of this would be a collection of unrelated operations combined into a single ad hoc solution. Although in certain circumstances this might be acceptable, it cannot represent a general solution of the feature extraction problem.

The conclusion of the discussion in chapter 1 was that an approach was needed in which a range of feature extraction problems could be addressed. It was suggested that one way of achieving this was to seek a description of the image (via a suitable transformation) that characterises the properties of all the features of interest. For example, such a description should have sufficient spatial locality to be able to represent lines and edges and be sufficiently global to represent regional properties. As was pointed out, these ideas are not new and there has been considerable interest in such an approach. In particular, the work of Granlund [46] and Wilson et al [109][112][114] was noted. These workers have made a number of advances in this area and have demonstrated the benefits that can be gained. However, although the basic ideas were laid out in [112], these have not been incorporated into a single image description. Thus, motivated by the previous work, the aim of this work has been to arrive at a contender for such a description.

The requirements of a suitable description and the associated transformation were investigated in chapter 2. It was shown that an essential property of meaningful features is that they provide information about position and class, ie "what" something is and "where" that something is located. This characteristic has been noted before

[72][113][114] and relates directly to the inherent nonstationarity of image properties. Furthermore, it implies that this nonstationarity must be incorporated into the required description if it is to be generally applicable, ie it must possess locality in both position and in some class space [113]. In other words, if useful features possess position and class information, then the description from which they are derived must provide this information.

A further characteristic of useful features noted in chapter 2 is that their degree of locality in position and class space varies from one feature to the next. This relates to the idea of features existing at different resolutions and has been generally recognised [68][72]. The implication for a unified description is that not only should it have locality in both position and class space but that this locality should exist over a wide range of scales. This important requirement has formed the basis for the work described in this thesis and although it is not the only one that could be postulated, it was felt that it represented a fundamental property of features and as such is consistent with the need for generality.

The above observations assume the existence of a suitable class space in which to base the required description. Feature extraction methods have utilised a number of possibilities, including both statistical and Fourier methods. It was noted in chapter 2, however, that the latter has a number of advantages (in particular its tractability) and was therefore adopted. As a result, the requirements of a suitable description were formalised as follows:

(i) It should have locality in both the spatial domain and in the spatial frequency domain.

(ii) The locality in requirement (i) should exist over a wide range of scales in both domains.

A transformation of the image data that would provide such a description was therefore identified as one which represents a transformation into a space that provides both spatial and spatial frequency information. Furthermore, the resolution in each domain within that space should be sufficient to represent the features of interest. It was also noted that if such a transformation was to be useful then it must possess other important properties, including amenability to efficient computation, invertibility and linearity.

Signal transformations that provide information from the original signal domain and the corresponding frequency domain (so-called combined representations) were reviewed in chapter 2. These were classified into two groups: linear forms and bilinear forms. It was noted that of these, a 2-d version of the linear short-time Fourier transform (STFT) possessed the most advantageous properties. These included invertibility, efficient computational properties, and a superior resolution in practical cases.

However, in common with the other representations, the STFT was shown to be limited in its ability to provide arbitrary resolution simultaneously in both domains. This results because it is necessarily based upon a windowing operation and the uncertainty principle [45] precludes any window function from having an arbitrarily high concentration in both the spatial and spatial frequency domain. The consequence of this is that a resolution trade-off must be adopted and this further implies that the requirement (ii) above is not satisfied. The relationship between this locality restriction and the difference between traditional feature extraction methods can therefore be noted: in both cases there is an absence of simultaneous locality in position and class. Indeed, this has prompted the suggestion that the uncertainty principle lies at the heart

of the vision problem [109].

A partial solution to the problem of uncertainty is to make use of multiscale methods and these were also reviewed in chapter 2. They have found extensive use in image processing in response to the need to process images over a range of spatial scales. The general form is that of a pyramid structure [100] in which the image is represented at different spatial resolutions on each level. However, although recently there has been a move towards some form of frequency selectivity [1][70][108], these methods fail to provide sufficient locality in the frequency domain and in fact are still limited in this respect by uncertainty.

After reviewing the above transformations, the overall conclusion of chapter 2 was that it was necessary to adopt a more general approach. This led to the introduction of the MFT, which is essentially a generalisation of the multiscale and combined representations. It has a hierarchical structure in which each level resembles a STFT with minimum uncertainty window functions. The outermost levels of this structure are the original signal and its discrete Fourier transform (DFT), while the intermediate levels are such that there is a uniform variation in resolution in each domain between these two extremes. Such a structure therefore has locality in both domains and over a full range of resolutions - something which existing methods fail to provide independently.

The MFT was formally defined in chapter 3. It was expressed in the form of a linear operator which is partitioned into level operators and is defined in terms of a set of analysis vectors. These represent the windowing operation performed by each level of the transform and are position and frequency shifted versions of a set of basic vectors. These vectors are defined to have minimum uncertainty according to the resolution of the different levels and are derived from the family of finite prolate spheroidal

sequences (FPSS) [111]. Each level of the MFT can be interpreted in one of two ways: either as the output of a filter bank with bandpass filters having contiguous frequency responses; or as a set of local spectrum estimates referring to contiguous regions of the signal domain. The transform was defined in its 1-d form, although it was shown that it is readily extended to the 2-d case. A cartesian separable implementation is particularly straightforward (it is simply the Kronecker product of its 1-d counterparts) and it was adopted throughout this work.

The MFT is an invertible transform. This results from the bandlimiting properties of the analysis vectors and an inverse operator was defined in chapter 3. The inversion procedure can also be considered in terms of a synthesis filter bank. The concept of a multilevel inverse was also introduced, in which a sparse transform containing a subset of coefficients on several levels is inverted. Although in general such an inversion is not exact, it was noted that by selecting an optimal set of coefficients it could be hoped to minimise errors. This method of inversion was used in later work.

Several properties of the transform were noted in chapter 3. These were: linearity; position and frequency shift invariance up to a given factor; local spectral properties; and hierarchical properties. It was shown that the local spectra of the MFT represent optimal estimates in terms of locality due to the minimum uncertainty properties of the analysis vectors. It was also noted that the hierarchical structure of the transform can be considered as a vector quadtree in which the local spectra correspond to the nodes of the tree.

Chapter 3 also introduced a general class of MFT. This was motivated by the need for increased locality in the spatial domain of a 2-d MFT. A generalised transform is defined by a relaxation parameter which determines the increase in locality. However, this locality can only be achieved at the expense of an increase in the number of

coefficients to ensure that the transform remains complete and invertible. It has been found in this work that a generalised 2-d MFT which has four times the number of coefficients of the original form provides a sufficient increase in spatial locality for the analysis problems considered.

The basic structure of the MFT means that it can be efficiently implemented by making use of methods based upon the fast Fourier transform (FFT). It was shown that the implementation of each level of the transform corresponds to a series of filtering operations which can be conveniently implemented in the spatial frequency domain. As a result, the MFT requires only an order of magnitude increase in computation over that for a conventional radix-2 FFT (approximately by a factor of $\log M$ for an image of dimension $M \times M$ pixels).

Several examples and experiments illustrating the properties of the transform were presented in chapter 3. These demonstrated the ability of the transform to decorrelate the image data in the spatial/spatial frequency plane and to do so over a number of different resolutions. Examples were presented for both synthetic and natural images. A simple threshold coding experiment was also recorded in which it was shown that important image features were emphasised within the transform coefficients.

7.2. Local Feature Estimation

Chapters 4-6 of this thesis considered the application of the MFT to a typical image analysis problem. It was decided to address the problem of extracting local image features such as lines and edges. This decision was based on the observation that the use of multiresolution methods in this area has received very little attention, although they would appear to offer a number of advantages. Furthermore, it was noted that

such an approach has been applied to the alternative area of texture analysis [99][112][113] and that the present work would therefore be complementary to previous investigations.

A review of existing approaches to local feature extraction was presented in chapter 4. These range from simple spatial templates to more sophisticated frequency domain methods. However, as noted earlier, the majority of methods are designed specifically for edge detection and do not admit generalisation to other feature types. In this sense, the area of local feature extraction epitomises the problem being addressed in this work. This was further emphasised in the review of existing approaches to curve detection based on such features also presented in chapter 4. This task is often performed within a separate framework from the original feature extraction (which is surprising given the inherent relationship between the two feature types) and is a good example of a series of ad hoc solutions being applied in order to arrive at the required result.

To demonstrate that multiresolution methods, and in particular the MFT, can form a basis for a more coherent approach, a multiresolution image model was introduced. It has a hierarchical structure in which the image is assumed to consist of different sized square regions each containing a single local feature. The model is a simple example of a more general class of models which have a linear recursive form and are designed to incorporate image properties over a full range of resolutions [116]. In the present work, these properties are restricted to single oriented local features defined within contiguous regions. The structure of the model is also amenable to the representation of curves and boundaries. This takes the form of a piecewise representation in which the curve or boundary is represented by local feature segments. It was shown that due to the multiresolution structure of the model, this type of representation provides a number of advantages including the reduction of redundancy and enabling the

definition of computationally efficient curve extraction schemes.

The local features within the image model are represented by a frequency domain model. It was noted that the use of frequency domain properties has previously been shown to be advantageous when extracting local features [65][96]. The model used in this work is based upon the observation that an oriented local feature such as an edge or line segment will give rise to an essentially linear phase relationship between spatial frequency coefficients in that orientation corresponding to the feature. The linear component of this relationship is then directly related to the position of the feature. It was noted that this property of such features underlies the importance of phase information in images [82]. Furthermore, by incorporating the relationship into a suitable model it enables these features to be distinguished from other oriented features such as those associated with some forms of texture. A local feature model was therefore introduced in the form of a normal Markov process in the relevant orientation of the spatial frequency domain. Within this model, the linear phase component is represented by the phase value of the complex recursion coefficient associated with the process. By assuming the existence of a finite number of features, this also led to the definition of a general model for the (continuous) frequency domain.

This continuous model was then adapted to take account of the locality requirement in the image model. Features are assumed to exist within local regions of the image and therefore some type of local analysis is required which necessarily means that the derived frequency information will be of finite resolution [64][83]. It was shown that this could be approximated by the introduction of a window function into the model to represent the smearing of the frequency domain due to the local analysis window. This approach was also extended to include the concept of a feature segment in which an "infinite" feature is replaced by one having finite length in the direction of its orientation. It is worth pointing out here that these adaptations of the continuous model

attempt to take account of the discrete nature of practical problems and in so doing make a number of simplifications and approximations. However, as the later results illustrated, they do not appear to introduce any gross errors into the model.

An estimation and detection scheme for the multiresolution image model was described in chapter 5. The scheme is based upon the assumption that local spectrum estimates are available corresponding to the admissible regions of the model. It was shown that a ML estimation scheme could be defined for the local feature model and that this yielded an unbiased estimate in a given orientation assuming a continuous spectrum. In the case of a local analysis, only a finite number of orientations can be considered and this will necessarily be biased due to the smearing introduced by the window in the frequency domain. However, it was shown that if a single feature was assumed to be present, then this bias is limited and acceptable estimates can be obtained in orientations that straddle the orientation corresponding to the feature. A similar scheme was defined to estimate the parameters of feature segments.

The multiresolution image model is based upon contiguous regions containing single local features and chapter 5 described a hierarchical detection scheme to identify such regions given feature estimates in a finite number of orientations. This is a recursive process which employs a principal orientation measurement for each potential single feature region and a scale consistency criterion between a region and its subregions. These criteria were designed to ensure that the regions selected not only possessed properties associated with them containing a single feature (ie significant contributions in estimates centred about one orientation) but that this information is confirmed by estimates obtained within their subregions. The result of the process is the selection of a number of contiguous spatial regions and these can be considered to form the leaf nodes of a truncated quadtree in which nodes refer to different sized spatial regions.

Chapter 5 also described a curve extraction scheme based upon the representation of curves within the model. This is a recursive curve forming process performed on the truncated quadtree resulting from the single feature detection. The basic idea is that curves are constructed in a piecewise manner by a fine-to-coarse analysis as the process ascends the tree. At a given node, local segments of the curve are formed using a local curvature measure and a heuristic search amongst the local features and curve segments defined at its four child nodes. It was shown that the scheme has the potential to perform 'gap-filling' due to the multiresolution structure of the image model and that it is amenable to efficient implementation.

It was shown in chapter 6 that the MFT can be used as an effective estimation tool for the local feature estimation scheme described in chapter 5. Specifically, the local spectra of the transform are those that are assumed to be available in the design of the estimation scheme. The relevant statistics can therefore be calculated from these local spectra and this leads to estimates for the parameters of the local feature model, ie position and certainty measures for features in a number of discrete orientations. This was shown not to require excessive computation.

An important consideration when implementing the estimation scheme is the problem of phase ambiguity and this was investigated in chapter 6. The periodicity of phase values means that the estimation of the linear phase component prescribed by the feature model leads to ambiguity as to the position of a feature and gives rise to the detection of ghost features. It was shown that this can be reduced by employing a generalised transform which has the dual effect of reducing the ambiguity and increasing the spatial concentration of the local spectrum estimates (and hence reducing leakage between regions). Another important issue in estimating local features is that of normalisation, ie assessing the relative importance of two features given a difference in their certainty measures. The approaches to this problem can vary between relying

entirely upon the estimated values or normalising the values on a region by region basis. A compromise has been adopted in this work so that some kind of simple context dependency can be incorporated into a normalisation scheme. This is related to the probability of finding a feature in a given region and is a hierarchical process in which a feature certainty measure is normalised according to the existence of features in its vicinity, where the size of the vicinity can be defined as appropriate.

Results of implementing the local feature estimation, single feature detection and curve extraction schemes on synthetic and natural images were presented in chapter 6. In the case of the feature estimation, the results demonstrated that feature parameters could be successfully estimated at different resolutions on the levels of the MFT. These range from detailed features at higher levels to more coarse features at lower levels. However, although in general these results have shown the validity of the scheme, there are a number of issues that need to be considered further:

(i) Noise sensitivity - the performance of the scheme in the presence of noise was not investigated. There are a range of applications in which image quality may be degraded and if the scheme is to be effective in these cases then its noise sensitivity will need to be assessed. The fact that the local feature model is based upon frequency domain properties does mean that it should be possible to minimise the effects of noise [65][96] and this would represent an important line of further research.

(ii) Normalisation - it was noted above that the normalisation process used in the scheme introduces a degree of context dependency. However, this has not been quantified in any way and the selection of normalisation parameters was based purely on subjective criteria, ie whether it produced a reasonable trade-off between feature and noise visibility. It would therefore be advantageous and theoretically more acceptable if this operation could be defined in a rigorous manner, although such an

exercise is complicated by the inherent nonlinearity of the operation.

(iii) Prewhitening - the local feature estimation results were derived from images that had been prewhitened using the simple method described in chapter 3. The justification for this preprocessing was to reduce any bias that may be caused by the predominantly lowpass frequency envelope of natural images. Since interesting features exist within highpass intervals of the frequency domain (eg lines, edges and texture) any such bias would be undesirable in a feature analysis operation. However, the approach adopted in this work is very simple (it is effectively a high pass filtering operation) and it is likely that a number of advantages could be gained by employing a more sophisticated technique [64][101]. Such an approach may also be more amenable to direct reconstruction of the image from the feature estimates which in this work has been based upon the addition of an approximation to the lowpass image removed by the prewhitening function.

The single feature detection and curve extraction results based on the above feature estimates were also shown to be generally successful. In the case of the former, features were identified at appropriate resolutions and it was demonstrated that the scheme is capable of splitting regions of the image until single feature regions are obtained. The resulting feature estimates then represent a considerable reduction in redundancy over conventional edge representations of images. However, as noted in chapter 6, there are cases in which difficulties are encountered and these are particularly apparent in complex areas of natural images. The problems relate to an inability of the scheme to separate features within such regions and derives from the inappropriateness and simplicity of the single feature criteria employed, ie the principal orientation and scale consistency measures. A further development of these criteria would therefore be beneficial, particularly in the area of establishing more sophisticated models to represent the presence of several features or specific combinations of

features, eg vertices, shapes, etc. This is relevant to the discussion in chapter 4 concerning the definition of a general class of feature based multiresolution image models.

The above difficulty with complex regions also gives problems for the curve extraction algorithm. Although the significant and isolated curves were identified for both synthetic and natural images, it was apparent that it was not suitable in detailed and complex regions of the image. This is not surprising given the simplicity of the method and it could be argued that in such regions a piecewise linear representation of a curve is not appropriate. Once again this suggests that more complex models in such regions would be beneficial. An example of this might be to extract vertices from an image in order to identify curve endings [75].

Chapter 6 also presented results of reconstructing images from the single feature estimates. This represents an important attribute when using the MFT in image analysis since it provides a coherent way of assessing the appropriateness of models defined within the transform structure (note that this facility is not generally possible in other approaches to image analysis). The results presented illustrate the applicability of the overall image model and the local feature model. However, the method of reconstruction used in this work does involve a number of simplifications and it is anticipated that the results could be improved by further work in this area. In particular, the method of modelling the magnitude response of the MFT local spectra by using a weighted average magnitude profile could be replaced by a more sophisticated scheme. As mentioned earlier, an alternative prewhitening method could also improve the quality of the reconstruction. It should be emphasised that these issues are important since the invertibility property of the MFT is one of its advantages and that the whole question of obtaining reconstructions for a given image model will therefore represent a significant area of further research.

7.3. Concluding Remarks

As mentioned at the beginning of this chapter, the work described in this thesis has investigated the possibility of defining an image description which could form the basis of a unified approach to feature extraction. This led to the introduction of the MFT and its application to the problem of extracting local features from an image. The transform was shown to have properties that are generally applicable and the application results were comparable to those of existing methods. However, does this imply that the original specification for a 'unified approach' has been met?

Of course, such a question can only be fully answered by illustrating the use of the MFT in a wide range of different tasks. This will clearly involve a considerable amount of work and in any case the intention is that the transform should form a basis for further research into new feature types as well as existing requirements. However, the properties of the MFT are based upon a substantial amount of previous work. This has considered the use of a basic structure for feature extraction [46][113][114] and has noted the important role played by uncertainty [109][113] and the use of multiresolution techniques [113][116]. Applications have included texture analysis and segmentation [66][99][112][113], line and edge extraction [65], image coding and restoration [33][34][102] and stereopsis [114]. It is envisaged that such operations could also be incorporated into the framework of the MFT. The implication therefore is that although this work has only considered one particular aspect of feature extraction, previous work does suggest that similar results could be obtained in other areas. Thus, although it has not been completely demonstrated that the MFT can provide a unified approach to image analysis, the potential for achieving this goal is apparent.

It is also worth recalling another issue that was discussed earlier in chapter 2. There has been a considerable amount of interest in recent years concerning the use of

multiscale methods in image analysis (particularly in recent publications on the Wavelet transform [36][70]) and it is reasonable to compare the MFT with these methods. As noted in chapter 2, the MFT can be regarded as a generalisation of these methods and consequently could be expected to be able to perform the work reported in the multiscale literature. Up to now this work has been mainly concerned with applications in image coding, although there has been some work reported on edge detection [56][70]. However, this and previous work has demonstrated that a multiresolution approach can be used in a wider range of applications and achieve acceptable results. Although this is by no means conclusive, it does suggest that the generalisation provided by the MFT is a more appropriate form on which to base solutions to image analysis problems.

To conclude, the MFT provides an image description that possesses properties which are applicable to general feature extraction. Although in specific applications it may be possible to define a nearly optimal or efficient solution, the MFT has the potential to provide a useful and effective solution for a number of different problems. Furthermore, it is a transform which has a well defined mathematical basis and can be computed in an efficient manner using familiar signal processing techniques. In this sense it fulfills the requirements of a general purpose image analysis tool. However, there is still a large amount of work to be done before its potential can be brought to fruition and it is hoped that this thesis will provide a suitable basis for this to be achieved.

APPENDIX I

Non-Zero Frequency Response of Analysis Vectors

It is shown that the analysis vectors g(n) of the 1-d MFT have non-zero magnitude within their respective frequency band, ie

$$| \hat{g}_i(n) | \begin{cases} \neq 0 & 0 \leq i < \Omega_n \\ = 0 & \text{else} \end{cases} \qquad \text{(A1)}$$

where

$$\hat{\mathbf{g}}(n) \;=\; \mathbf{F}\ \mathrm{g}(n) \qquad \text{(A2)}$$

and

$$\mathbf{B}(\Omega_n)\ \mathbf{T}(\Gamma_n)\ \mathrm{g}(n) \;=\; \lambda_0\ \mathrm{g}(n) \qquad \text{(A3)}$$

Equation (A3) can be expressed in terms of the vectors $\hat{\mathbf{g}}(n)$. This can be achieved by applying the operator **F** to both sides to give

$$\mathbf{T}(\Omega_n)\ \mathbf{F}\ \mathbf{T}(\Gamma_n)\ \mathrm{g}(n) \;=\; \lambda_0\ \mathbf{F}\ \mathrm{g}(n) \qquad \text{(A4)}$$

and then from eqn (A2)

$$\mathbf{T}(\Omega_n)\ \mathbf{B}(\Gamma_n)\ \hat{\mathbf{g}}(n) \;=\; \lambda_0\ \hat{\mathbf{g}}(n) \qquad \text{(A5)}$$

which can also be written as

$$\mathbf{T}(\Omega_n)\ \mathbf{B}(\Gamma_n)\ \mathbf{T}(\Omega_n)\ \hat{\mathbf{g}}(n) = \lambda_0\ \hat{\mathbf{g}}(n) \qquad \text{(A6)}$$

where the introduction of the operator $\mathbf{T}(\Omega_n)$ has no effect since the vector $\hat{\mathbf{g}}(n)$ is by definition truncated to the interval defined by $\mathbf{T}(\Omega_n)$ [111].

The above equation then enables a reduced problem to be defined, ie

$$\mathbf{A}(n)\ \mathbf{x}(n) = \lambda_0\ \mathbf{x}(n) \qquad \text{(A7)}$$

where $\mathbf{A}(n)$ is a $\Omega_n \times \Omega_n$ operator and $\mathbf{x}(n)$ is a $\Omega_n \times 1$ vector st

$$a_{ik}(n) = b_{ik}(\Gamma_n) \qquad 0 \le i,k < \Omega_n \qquad \text{(A8)}$$

and

$$x_i(n) = \hat{g}_i(n) \qquad 0 \le i < \Omega_n \qquad \text{(A9)}$$

From eqn (A1), the problem is now to show that the component magnitudes of the vector $\mathbf{x}(n)$ are non-zero.

This can be done by rewriting eqn (A7) as

$$\mathbf{W}^{\alpha}\ \mathbf{A}(n)\ \mathbf{W}^{-\alpha}\ \mathbf{W}^{\alpha}\ \mathbf{x}(n) = \lambda_0\ \mathbf{W}^{\alpha}\ \mathbf{x}(n) \qquad \text{(A10)}$$

where $\mathbf{W}^{\alpha}$ is a $\Omega_n \times \Omega_n$ frequency shift operator and α is a positive integer. The proof then follows from the theorem of Perron and Frobenius which states that the eigenvectors of an irreducible nonnegative matrix have real and positive components [84]. This can be applied to eqn (A10), where after some straightforward although extensive manipulation, it can be shown that the operator $\mathbf{W}^{\alpha}\,\mathbf{A}(n)\,\mathbf{W}^{-\alpha}$ is both irreducible and nonnegative for some value of α. Hence the vector $\mathbf{W}^{\alpha}\,\mathbf{x}(n)$ will have nonnegative components. Since the operator $\mathbf{W}^{\alpha}$ implies the multiplication of each component by a complex exponential with unit magnitude (eqn 1.8), the component magnitudes of the vector $\mathbf{x}(n)$ are also nonnegative. Combining this with eqns (A1) and (A9) completes the proof.

Any pages, tables, figures or photographs, missing from this digital copy, have been excluded at the request of the university.

REFERENCES

[1] E.H.Adelson, E.Simoncelli, R.Hingorani, *Orthogonal Pyramid Transforms for Image Coding*, Proc. SPIE Visual Commun. and Image Proc. II, Cambridge MA, 1987.

[2] J.B.Allen, L.R.Rabiner, *A Unified Approach to Short-Time Fourier Analysis and Synthesis*, Proc. IEEE vol. 65, 1558-1564, 1977.

[3] G.P.Ashkar, J.W.Modestino, *The Contour Extraction Problem with Biomedical Applications*, Comput. Graph. Image Proc. 7, 331-355, 1978.

[4] D.Ballard, *Generalising the Hough Transform to Detect Arbitrary Shapes*, Pattern Recognition 13, 111-122, 1981.

[5] D.Ballard, C.M.Brown, *Computer Vision*, Englewood Cliffs NJ, Prentice-Hall, 1982.

[6] M.S.Bartlett, *An Introduction to Stochastic Processes*, 3rd ed., Cambridge, Cambridge University Press, 1978.

[7] M.J.Bastiaans, *The Wigner Distribution Function and its Application to First Order Optics*, J. Opt. Soc. Am. 69, 1710-1716, 1979.

[8] M.J.Bastiaans, *Gabor's Expansion of a Signal into Gaussian Elementary Signals*, Proc. IEEE vol. 68, 538-539, 1980.

[9] M.J.Bastiaans, *A Sampling Theorem for the Complex Spectrogram and Gabor Expansion of a Signal into Gaussian Elementary Signals*, Opt. Eng. vol. 20, 594-598, 1981.

[10] F.Bergholm, *Edge Focusing*, IEEE Trans. vol. PAMI-9, 726-741, 1987.

[11] P.J.Besl, *Geometric Modelling and Computer Vision*, Proc. IEEE vol. 76, 936-958, 1988.

[12] D.Blostein, N.Ahuja, *A Multiscale Region Detector*, Comput. Vision Graph. Image Proc. 45, 22-41, 1989.

[13] A.I.Borisenko, I.E.Tarapov, *Vector and Tensor Analysis with Applications*, New York, Dover, 1979.

[14] R.N.Bracewell, *The Fourier Transform and its Application*, Int. ed., Singapore, McGraw-Hill, 1986.

[15] N.G. de Bruijn, *Uncertainty Principles in Fourier Analysis*, Inequalities, O.Shisha (ed.), New York, Academic Press, 57-71, 1967.

[16] P.J.Burt, *Fast Filter Transforms for Image Processing*, Comput. Graph. Image Proc. 16, 20-51, 1981.

[17] P.J.Burt, T.H.Hong, A.Rosenfeld, *Segmentation and Estimation of Image Region Properties Through Cooperative Hierarchical Computation*, IEEE Trans vol. SMC-11, 802-809, 1981.

[18] P.J.Burt, E.H.Adelson, *The Laplacian Pyramid as a Compact Image Code*, IEEE Trans vol. COM-31, 532-540, 1983.

[19] P.J.Burt, *Fast Algorithm for Estimating Local Image Properties*, Comput. Vision Graph. Image Proc. 21, 368-382, 1983.

[20] A.Calway, R.Wilson, *A Unified Approach to Feature Extraction Based on an Invertible Image Transform*, Proc. 3rd IEE Int. Conf. Image Processing, 651-655, Warwick, 1989.

[21] F.W.Campbell, J.G.Robson, *Application of Fourier Analysis to the Visibility of Gratings*, J. Physiol. vol. 197, 551-566, 1968.

[22] J.Canny, *A Computational Approach to Edge Detection*, IEEE Trans. vol. PAMI-8, 679-698, 1986.

[23] S.Carlsson, *Sketch Based Coding of Grey Level Images*, Signal Processing 15, 57-83, 1988.

[24] P.Cavanagh, *Size and Rotation Invariance in the Visual System*, Perception vol. 7, 167-177, 1978.

[25] D.S.K.Chan, *A Non-Aliased Discrete-Time Wigner Distribution for Time-Frequency Signal Analysis*, Proc. ICASSP, 1333-1336, Paris, 1982.

[26] C.H.Chen, *A Study of Texture Classification Using Spectral Features*, Proc. 6th Int. Conf. Patt. Rec., 1074-1077, Munich 1982.

[27] P.C.Chen, T.Pavlidis, *Image Segmentation as an Estimation Problem*, Comput. Graph. Image Proc. 12, 153-172, 1980.

[28] Y.P.Chien, K.S.Fu, *A Decision Function Method for Boundary Detection*, Comput. Graph. Image Proc. 3, 125-140, 1974.

[29] T.A.C.M. Claasen, W.F.G.Mecklenbräuker, *The Wigner Distribution - A Tool for Time-Frequency Signal Analysis Part I: Continuous-time Signals*, Philips J. Res. 35, 217-250, 1980.

[30] T.A.C.M. Claasen, W.F.G.Mecklenbräuker, *The Wigner Distribution - A Tool for Time-Frequency Signal Analysis Part II: Discrete-time Signals*, Philips J. Res. 35, 276-300, 1980.

[31] T.A.C.M. Claasen, W.F.G.Mecklenbräuker, *The Wigner Distribution - A Tool for Time-Frequency Signal Analysis Part III: Relations with other Time-Frequency Signal Transformations*, Philips J. Res. 35, 372-389, 1980.

[32] T.A.C.M. Claasen, W.F.G.Mecklenbräuker, *The Aliasing Problems in the Discrete-Time Wigner Distributions*, IEEE Trans vol. ASSP-31, 1067-1072, 1983.

[33] S.C.Clippingdale, R.G.Wilson, *Least-Squares Image Estimation on a Multiresolution Pyramid*, Proc. ICASSP, 1409-1412, Glasgow, 1989.

[34] S.C.Clippingdale, *Multiresolution Image Modelling and Estimation*, University of Warwick PhD Thesis, 1988.

[35] J.L.Crowley, R.M.Stern, *Fast Computations for the Difference of Lowpass Transform*, IEEE Trans. vol. PAMI-6, 212-221, 1984.

[36] I.Daubechies, *Orthogonal Bases of Compactly Supported Wavelets*, Comm. on Pure and Appl. Math. vol. XLI, 909-996, 1988.

[37] J.G.Daugman, *Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimised by Two-Dimensional Visual Cortical Filters*, J. Opt. Soc. Am. vol. 2, 1160-1169, 1985.

[38] J.G.Daugman, *Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression*, IEEE Trans. vol. ASSP-36, 1169-1179, 1988.

[39] L.S.Davis, *A Survey of Edge Detection Techniques*, Comput. Graph. Image Proc. 4, 248-270, 1975.

[40] L.S.Davis, S.A.Johns, J.K.Aggarwal, *Texture Analysis Using Generalised Cooccurrence Matrices*, IEEE Trans. PAMI-1, 251-259, 1979.

[41] R.O.Duda, P.E.Hart, *Use of the Hough Transformation to Detect Lines and Curves in Pictures*, Commun. ACM 15, 11-15, 1972.

[42] D.Esteban, C.Galand, *Application of Quadrature Mirror Filters to Split-Band Coding*, Proc. ICASSP, Hartford, 1977.

[43] B.Friedlander, B.Porat, *Detection of Transient Signals by the Gabor Representation*, IEEE Trans. vol. ASSP-37, 169-180, 1989.

[44] K.Fukunaga, *Introduction to Statistical Pattern Recognition*, New York, Academic Press, 1972.

[45] D.Gabor, *Theory of Communication*, Proc. IEE 93, 429-441, 1946.

[46] G.H.Granlund, *In Search of a General Picture Processing Operator*, Comput. Graph. Image Proc. 8, 155-173, 1978.

[47] R.M.Haralick, *Zero-Crossings of Second Directional Derivative Edge Operator*, IEEE Trans. vol. PAMI-6, 58-68, 1984.

[48] R.M.Haralick, *Statistical Image Texture Analysis*, in Handbook of Patt. Rec. and Image Proc. ed. T.Y.Hong, K.S.Fu, 247-279, 1986.

[49] F.J.Harris, *On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform*, Proc. IEEE 66, 51-83, 1978.

[50] L.O.Harvey, M.J.Gervais, *Visual Texture Perception and Fourier Analysis*, Percept. and Psychophys. vol. 24, 534-542, 1978.

[51] J.S.Huang, D.H.Tseng, *Statistical Theory of Edge Detection*, Comput. Vision Graph. and Image Process. 43, 337-346, 1988.

[52] D.H.Hubel, *Eye, Brain, and Vision*, New York, Sci. Am. Lib., 1988.

[53] D.H.Hubel, T.N.Wiesel, *Receptive Fields of Single Cells in the Cat's Striate Cortex*, J. Physiol. 148, 574-591, 1959.

[54] M.F.Hueckel, *An Operator which Locates Edges in Digitised Pictures*, J. ACM 18, 113-125, 1971.

[55] R.A.Hummel, *Feature Detection Using Basis Functions*, Comput. Graph. Image Proc. 9, 40-55, 1979.

[56] R.A.Hummel, *Representations Based on Zero-Crossings in Scale Space*, Proc. IEEE Comput. Vision and Patt. Rec., Miami Beach FL, 204-209, 1986.

[57] J.Illingworth, J.Kittler, *A Survey of the Hough Transform*, Comput. Vision Graph. and Image Process. 44, 87-116, 1988.

[58] L.Jacobson, H.Weschler, *A New Paradigm for Computational Vision Based on the Wigner Distribution*, University of Minnesota, Tech. Rep., 1982.

[59] L.Jacobson, H.Weschler, *The Composite Pseudo Wigner Distribution (CPWD): A Computable and Versatile Approximation to the Wigner Distribution (WD)*, IEEE ICASSP, 254-256, Boston MA, 1983.

[60] A.J.E.M.Janssen, *Gabor Representation and Wigner Distribution of Signals*, Proc. ICASSP, San Diego, 1984.

[61] D.L.Jones, T.W.Parks, *A High Resolution Data-Adaptive Time-Frequency Representation*, Proc. ICASSP, 1987.

[62] D.L.Jones, T.W.Parks, *A Resolution Comparison of Several Time-Frequency Representations*, Proc. ICASSP, Glasgow, 1989.

[63] M.Kass, A.Witkin, *Analyzing Oriented Patterns*, Comput. Vision Graph. Image Proc. 37, 362-385, 1987.

[64] S.M.Kay, S.L.Marple, *Spectrum Analysis - A Modern Perspective*, Proc. IEEE vol. 69, 1380-1419, 1981.

[65] H.Knutsson, *Filtering and Reconstruction in Image Processing*, Linköping University, PhD Thesis, 1982.

[66] H.Knutsson, G.H.Granlund, *Texture Analysis Using Two-Dimensional Quadrature Filters*, CAPAIDM Workshop, Pasadena CA, 1983.

[67] H.Knutsson, *A Tensor Representation of 3-D Structure*, IEEE ASSP Workshop on Multidimensional Signal Processing, Noordwijkerhout, Holland, 1987.

[68] J.J.Koenderink, *The Structure of Images*, Biol. Cybern 50, 363-370, 1984.

[69] M.Kunt, A.Ikonomopoulos, M.Kocher, *Second-Generation Image Coding Techniques*, Proc. IEEE 73, 549-573, 1985.

[70] S.G.Mallat, *A Theory for Multiresolution Signal Decomposition: The Wavelet Representation*, IEEE Trans. vol. PAMI-11, 674-693, 1989.

[71] S.Marcelja, *Mathematical Description of the Responses of Simple Cortical Cells*, J. Opt. Soc. Am. vol. 70, 1297-1300, 1980.

[72] D.Marr, *Vision*, San Francisco, Freeman, 1982.

[73] D.Marr, E.Hildreth, *Theory of Edge Detection*, Proc. R.Soc. London, 187-217, 1980.

[74] A.Martelli, *An Application of Heuristic Search Methods to Edge and Contour Detection*, Commun. ACM 19, 73-83, 1976.

[75] G.Medioni, Y.Yasumoto, *Corner Detection and Curve Representation Using Cubic B-Splines*, Comput. Vision Graph. Image Proc. 39, 267-278, 1987.

[76] P.Meer, E.Baugher, A.Rosenfeld, *Frequency Domain Analysis and Synthesis of Image Pyramid Generating Kernels*, IEEE Trans. vol. PAMI-9, 512-522, 1987.

[77] Y.Meyer, *Principe d'Incertitude, Bases Hilbertiennes et Algebres d'Operateurs*, Bourbaki Seminar 662, 1985-86.

[78] U.Montanari, *On the Optimal Detection of Curves in Noisy Pictures*, Commun. ACM 14, 335-345, 1971.

[79] R.Nevatia, K.Ramesh Babu, *Linear Feature Extraction and Description*, Comput. Graph. Image Proc., 13, 257-269, 1980.

[80] W.M.Newman, R.F.Sproull, *Principles of Interactive Computer Graphics*, 2nd ed, New York, McGraw-Hill, 1979.

[81] L.O'Gorman, A.C.Sanderson, *A Comparison of Methods and Computation for Multi-Resolution Low- and Band-Pass Transforms for Image Processing*, Comput. Vision Graph. Image Proc. 37, 386-401, 1987.

[82] A.V.Oppenheim, J.S.Lim, *The Importance of Phase in Signals*, Proc. IEEE vol. 69, 529-541, 1981.

[83] A.Papoulis, *Signal Analysis*, Int. ed., Singapore, McGraw-Hill, 1986.

[84] M.C.Pease, *Methods of Matrix Algebra*, New York, Academic Press, 1965.

[85] T.Peli, D.Malah, *A Study of Edge Detection Algorithms*, Comput. Graph. Image Proc. 20, 1-21, 1982.

[86] D.A.Pollen, S.F.Ronner, *Visual Cortical Neurons as Localised Spatial Frequency Filters*, IEEE Trans. SMC-13, 907-916, 1983.

[87] M.Porat, Y.Y.Zeevi, *The Generalised Gabor Scheme of Image Representation in Biological and Machine Vision*, IEEE Trans. vol. PAMI-10, 452-468, 1988.

[88] M.R.Portnoff, *Implementation of the Digital Phase Vocoder using the Fast Fourier Transform*, IEEE Trans. vol. ASSP-24, 243-248, 1976.

[89] M.R.Portnoff, *Time-Frequency Representation of Digital Signals and Systems Based on Short-Time Fourier Analysis*, IEEE Trans. vol. ASSP-28, 55-69, 1980.

[90] M.R.Portnoff, *Short-Time Fourier Analysis of Sampled Speech*, IEEE Trans. vol. ASSP-29, 364-373, 1981.

[91] L.R.Rabiner, B.Gold, *Theory and Application of Digital Signal Processing*, Englewood Cliffs NJ, Prentice-Hall, 1975.

[92] L.Roberts, *Machine Perception of 3-Dimensional Solids*, in Optical and Electrooptical Information Processing, ed. J.Tippett, Cambridge MA, MIT Press, 1965.

[93] A.Rosenfeld, M.Thurston, *Edge and Curve Detection for Visual Scene Analysis*, IEEE Trans. vol. C-20, 562-569, 1971.

[94] A.Rosenfeld, *Image Analysis: Problems, Progress and Prospects*, Pattern Recognition 17, 3-12, 1984.

[95] R.J.Schalkoff, *Digital Image Processing and Computer Vision*, New York, Wiley, 1989.

[96] K.Shanmugam, F.M.Dickey, J.A.Green, *An Optimal Frequency Domain Filter for Edge Detection in Digital Pictures*, IEEE Trans vol. PAMI-1, 37-69, 1979.

[97] M.J.T.Smith, T.P.Barnwell, *A New Filter Bank Theory for Time Frequency Representation*, IEEE Trans. vol. ASSP-35, 314-326, 1987.

[98] M.Spann, R.Wilson, *A Quad-Tree Approach to Image Segmentation which Combines Statistical and Spatial Information*, Patt. Rec. vol. 18, 257-269, 1985.

[99] M.Spann, *Texture Description and Segmentation in Image Processing*, University of Aston, PhD Thesis, 1985.

[100] S.L.Tanimoto, T.Pavlidis, *A Hierarchical Data Structure for Picture Processing*, Comput. Graph. Image Proc. 4, 104-119, 1975.

[101] D.J.Thomson, *Spectrum Estimation and Harmonic Analysis*, Proc. IEEE 70, 1055-1096, 1982.

[102] M.Todd, R.Wilson, *An Anisotropic Multiresolution Image Data Compression Algorithm*, Proc. ICASSP, 1969-1972, Glasgow, 1989.

[103] S.Ullman, *Analysis of Visual Motion by Biological and Computer Systems*, IEEE Computer 14, 57-69, 1981.

[104] A.J.Van Doorn, J.J.Koenderink, *The Structure of the Human Motion Detection System*, IEEE Trans. vol. SMC-13, 916-922, 1983.

[105] E.Velez, R.Absher, *Transient Analysis of Speech Signals Using the Wigner Time-Frequency Representation*, Proc. ICASSP, 2242-2245, 1989.

[106] M.Vetterli, *Multidimensional Sub-Band Coding: Some Theory and Algorithms*, Signal Processing vol. 6, 97-112, 1984.

[107] M.Vetterli, *A Theory of Multirate Filter Banks*, IEEE Trans. vol. ASSP-35, 1987.

[108] A.B.Watson, *The Cortex Transform: Rapid Computation of Simulated Neural Images*, Comput. Vision Graph. Image Proc. 39, 311-327, 1987.

[109] R.Wilson, G.H.Granlund, *The Uncertainty Principle in Image Processing*, IEEE Trans. vol. PAMI-6, 758-767, 1984.

[110] R.Wilson, *Quadtree Predictive Coding*, Proc. ICASSP, San Diego, 1984.

[111] R.Wilson, *Finite Prolate Spheroidal Sequences and Their Applications I: Generation and Properties*, IEEE Trans. vol. PAMI-9, 787-795, 1987.

[112] R.Wilson, M.Spann, *Finite Prolate Spheroidal Sequences and Their Applications II: Image Feature Description and Segmentation*, Proc. IEEE Trans. vol. PAMI-10, 193-203, 1987.

[113] R.Wilson, M.Spann, *Image Segmentation and Uncertainty*, Letchworth, Research Studies Press, 1988.

[114] R.Wilson, H.Knutsson, *Uncertainty and Inference in the Visual System*, IEEE Trans. vol. SMC-18, 305-312, 1988.

[115] R.Wilson, A.D.Calway, *A General Multiresolution Signal Descriptor and Its Application to Image Analysis*, Proc. EUSIPCO, Grenoble, 1988.

[116] R.Wilson, S.C.Clippingdale, *A Class of Nonstationary Image Models and Their Applications*, Proc. IMA Conf. Math. in Sig. Process., Warwick, 1988.

[117] A.P.Witkin, *Recovering Surface Shape and Orientation from Texture*, Artificial Intelligence 17, 17-45, 1981.

[118] A.P.Witkin, *Scale-Space Filtering*, Proc. IJCAI, Karlsruhe, 1983.