IMAGE RECONSTRUCTION FROM INCOMPLETE INFORMATION

by

ALAN STANLEY KWABWE

A Thesis submitted for the degree of Doctor of Philosophy in the Faculty of Engineering.

August 1984

Department of Electrical Engineering
Imperial College of Science and Technology
(University of London)
LONDON

To my Mother and Father

ABSTRACT

The general problem of image reconstruction from incomplete information or limited data is encountered in a number of diverse areas such as medical imaging, astronomy, geophysics, image processing etc.

This thesis considers two problems. The first concerns the reconstruction of images from the phase or magnitude of their Discrete Fourier Transform representation.

Phase only reconstruction is motivated by the intelligibility of the phase only image. Conditions under which an image may be uniquely reconstructed from its phase function are studied and both non-iterative and iterative algorithms for reconstruction are discussed. Phase only reconstruction when only noisy phase is available is studied and it is shown that the good choice of an initial magnitude estimate improves the reconstruction. Several methods of choosing a satisfactory initial magnitude are presented.

In the case of magnitude only reconstruction, theory indicates that irreducible finite support sequences are recoverable to within an equivalence class but satisfactory practical algorithms for reconstruction have not always been obtained. The amount of a priori information available will determine the rate of convergence and this thesis investigates the effect of specific information such as boundary information and its incorporation into the reconstruction.

Two important applications arising out of the above are

considered. A hybrid technique is developed for the restoration of linearly degraded images with additive noise. A magnitude only reconstruction algorithm is found to improve on conventional Wiener filtering.

The other is Fourier transform block coding, where techniques which seek to exploit the relative similarity of the magnitude functions in the blocks are investigated.

The second problem concerns incomplete image specification in the space domain. Specifically the case where an image is to be reconstructed from its intensities along the edge contours or their polygonal approximation.

A reconstruction method is developed which treats the problem as a constrained minimisation to find an image compatible with the given contour and intensity information.

ACKNOWLEDGEMENT

Its is my pleasure to acknowledge with gratitude the assistance of my supervisor Dr R.A. King whose patient guidance and direction has enabled me to develop my interest in this subject. His help and encouragement has been inspiring and invaluable at all stages right up to end.

Many of the ideas developed in this thesis have come out of enthusiastic discussions in Dr King's research group and I acknowledge the interest of the members of the group.

I would like to thank the staff and my colleagues at the Digital Communication Section who have helped to make my stay in the section a pleasant one. In particular, I was glad to work with Kumar Singarajah, Nasser Nasrabadi, Ansah Frimpong, Sanker Pal, Ahmet Kayran and all my other friends in the department.

I am grateful to my family and Tunde for their patience and understanding throughout this whole period when I have been so distant and preoccupied.

I would like to thank Anne O' Neill for her help and her careful typing of the script through very difficult circumstances.

While studying at Imperial College, I have benefited from association with many individuals whose names are too numerous to mention here.

CONTENTS

Table of Symbols

$x(n)$	1-D space domain sequence, taking on real values at integer points n.
$x(n,m)$	2-D space domain sequence
$x_m(n)$	1-D vector, representing the m^{th} row of the array $x(n,m)$ for all n also written as x_m.
$R(N)$	Region of support of the sequence $x(n)$. $x(n)$ is non-zero for $0<n<N$
x	matrix representing the array $x(n,m)$
$X(e^{j\omega})$	The Fourier transform of $x(n)$
$X(z)$	The z-transform of $x(n)$
$\|x\|$	The Euclidean norm of the vector x.
$T(\ )$	Mapping that represents the general iterative process.
x_p	the sequence at the pth iteration.
Δ	The Laplacian operator
x^o	Sequence to which an iterative process converges
F	Fourier transform operator
F^{-1}	Inverse Fourier transform operator
**	2-D convolution operator.
*	1-D convolution operator, or(superscript) complex conjugate.
log	Natural logarithm
$\log_{10}$	log to base 10
$\log_2$	log to base 2

CHAPTER ONE

INTRODUCTION

CHAPTER I

INTRODUCTION

1.1 Introduction

The problem of reconstruction from incomplete data can be said to be present in all forms of digital image processing because any digital image can be viewed as quantized samples of an analog image function.

For most purposes however, an image sampled above the Nyquist rate and digitized to a useful number of levels (say 256) can be considered to be a full or complete image and it is images reconstructed from much less information either in the time or frequency domain that are the subject of this thesis.

Applications abound throughout image processing and there are also many analogous problems in other engineering disciplines. A few of these applications are presented here.

Generally, the incomplete image information may be present in the time domain or in a transform domain and the reconstruction process may use either or both of these domains.

In the time domain we may have insufficient information due to decimation. Another application is sub-Nyquist sampling where the image function is sampled below the Nyquist rate, but at such a frequency that the aliasing components fall at predetermined frequencies and can be extracted.

In the transform domain, we have several application where it is necessary to reconstruct an image from incomplete information. The

Fourier phase problem, where only the Fourier transform magnitude is available for measurement and the phase must be reconstructed. In image transform coding, only a few samples of the transform co efficients are transmitted. These are chosen so that a reconstruction algorithm at the receiving end can give reasonable reconstruction of the image.

Another application in the transform domain, is image reconstruction from projections - especially when the imaging fast moving organs such as the beating heart. In this case we are interested both in reducing the total number of projections needed for reconstruction of static scenes as well as in producing images of reasonable quality from the few projections that can be taken within the very short 'static' viewing period of a fast moving object.

It is quite probable that the problem can be generalised to cover incomplete transform information of all kinds, as well as to consider image partial information that does not fall into either of the two categories discussed above - for example, a contour or polygonal approximation.

Many techniques, both theoretical and ad hoc, have been presented for solving the reconstruction problem in some applications.

A quick look at the history of image reconstruction shows that the basic approach to the problem has been two pronged. In the first instance, researchers have tried to establish, theoretically, the uniqueness of the partial information present, i.e. is the partial information uniquely linked to the original image function ?

The second aspect has been a consideration of different algorithms to perform the reconstruction. As we shall see in the review which follows, many of the successful algorithms are interactive algorithms and hence another approach to the problem has been one of algorithm acceleration and convergence study.

The importance of this subject may be illustrated by the breadth of applications where it occurs, they fall into 4 general categories:

Type 1: Those applications where only partial information is available and there is no other way of forming an image. This includes Tomographic reconstruction from limited projections, phase only and magnitude only reconstruction, High resolution Synthetic Aperture Radar (SAR).

Type 2: Where full information is present - but noisy or erroneous - Current techniques for image restoration and blind deconvolution would benefit from reconstruction methods.

Type 3: Where full information is present but some is selectively discarded for the purpose of data compression - e.g. Transform coding etc. and the received image from such a system may be treated as a limited data image, and reconstruction methods used to recover the original image.

This thesis addresses itself to several forms of this general problem. The general problem may be put in the following manner:

Limited samples of an image are obtained and an attempt is made to describe to the maximum extent possible, the image from which these samples were obtained, subject to

(i) Acceptance of all available data

(ii) Ensuring all extrapolated data are consistent with a priori knowledge.

(iii) Being neutral about data that are not measured and cannot be extrapolated.

Even though the above problem specification is applied to an image, it may often happen that the available incomplete information is not in the image domain itself but in another domain, uniquely related to the image function. Several significant problems in a number of diverse areas that have missing information in other domains include phase and magnitude only reconstruction, bandlimited spectrum extrapolation, tomographic reconstruction from projections etc.

In all these areas, the two basic problems to be addressed are

(i) Theoretical considerations of the uniqueness of the image derived from the partial information.

(ii) Practical algorithms to reconstruct images given the limited data. The main factors here are numerical complexity of the algorithms, rates and bounds of convergence.

1.2 Scope of the thesis:

This thesis addresses two subproblems of the image reconstruction problem.

The first concerns the reconstruction of images from the phase or magnitude of the fourier transform.

Interest in this area is motivated by several factors. The causative factor is the perceived different roles played by phase and

magnitude of the signal. A phase-only image is often intelligible while a magnitude-only image is not. Phase is considered more important than magnitude and this is utilised in phase only holograms and Fourier transform coding.

In the former, the magnitude is discarded and replaced by a constant, while in the latter, the phase is coded using more bits than the magnitude.

The relative importance of the phase has prompted suggestions that perhaps the full image may be reconstructed from phase only, with consequent reduction in the amount of information needed to represent an image.

A second factor is the presence of a number of applications especially in astronomy and optics, where only one or other of the Fourier spectral components is available for measurement.

Some form of reconstruction is then necessary to obtain an estimate of the full spectrum.

Phase only reconstruction with iterative algorithms has been found to converge most of the time in contrast to magnitude only reconstruction. There has however, not been a suitable analysis of the noise sensitivity of phase only image reconstruction. An analysis is presented in this thesis and consideration is also given to ways of improving phase only reconstruction both for the noisy and noise free cases.

Fast adaptive algorithms for both phase and magnitude reconstruction are developed.

Magnitude only reconstruction has had a longer history than phase only reconstruction, albeit a less successful one. Although

reconstruction is theoretically possible if the finite sequence is irreducible in practice the problem appears to have a very large number of local minima and convergence to the true global minimum is virtually impossible unless the initial starting point is very close or if stringent constraints are available. The amount and type of a priori information available as well as the constraints used will determine the rate of convergence of the iterative algorithms and this thesis investigates the effect of specific information such as extra phase information, boundary information etc.

Two important applications arising out of the above are considered. The first is the restoration of linearly degraded images with additive noise. Conventional restoration techniques use restoration filters (e.g. Wiener) which tend to restore the magnitude leaving the phase function unrestored. A hybrid technique is presented where phase restoration is carried out by using the restored magnitude and noisy phase as initial estimates in an iterative magnitude only reconstruction. The other application is in Fourier transform coding where techniques which seek to exploit the relative similarity of the magnitude functions among blocks of the image are investigated.

The second issue addressed in this thesis concerns incomplete image specification in the space domain specifically the case where an image is to be restored from a representation of its contours and texture. This is motivated by the intelligibility of contour images and the premise that the regions between contours have uniform textures.

It is shown that iterative methods can be used to achieve

reconstruction from this information and it is extended to the case where polygonal approximations are used for representation.

1.2 Outline of the thesis

Chapter 2 is a survey of some of the results which have appeared in the literature concerning image reconstruction from incomplete information. The general problem is presented and a brief historical background of some of the more popular application is given. A rather more extended survey is then made of the problems that have been considered in some detail in this thesis. These are the magnitude only and phase only reconstruction problems.

Due to the importance and wide use of iterative reconstruction methods in solving these problems, we include a brief review of these iterative approaches.

In chapter 3, the phase only reconstruction problem is examined in detail. The conditions under which an image can be uniquely represented by the phase of its DFT are considered and both iterative and non-iterative reconstruction algorithms are developed for the case where the correct phase function is available. The sensitivity of phase only reconstruction to both additive and quantisation noise is examined and it is shown that this sensitivity can be reduced if an estimate of the magnitude function is available.

Several techniques are examined for faster phase only reconstruction and in particular the effect of the initial amplitude on the rate of convergence of the iterative is exploited. Several ways of choosing an appropriate starting magnitude function are developed.

In chapter 4, a similar study to that carried out in chapter 3 is done for the case of reconstruction from magnitude only. This problem is generally less tractable and the uniqueness conditions are more general. A variety of reconstruction algorithms that have been used for this problem are discussed and compared and some reasons offered for their lack of success. Faster adaptive algorithms with extra constraints are studied. Due to the importance of a priori information on the success of the reconstruction methods, some special study is made of reconstruction from magnitude given some other information. In particular, it is shown that given the sign of the phase, the problem becomes a lot less ill conditioned. A general study is then made on the quality of reconstructions from other partitions of Fourier domain.

Chapter 5 is devoted to some frequency domain applications of image reconstruction. Some of the results developed in the previous chapters are used. Magnitude only reconstruction is used to estimate the phase of a degraded image that has been restored by a Wiener filter. The degraded image phase is taken as the first estimate in the iterative reconstruction. In the second part, a Fourier phase coding technique is developed. It uses the correct phase function and an estimate of the magnitude function as developed in chapter three.

Chapter 6 considers a related space domain reconstruction problem. It examines the question of whether a useful reconstruction of a gray scale image can be made from its contour representation. A similar constrained iteration approach is found to yield useful results. The reconstruction image is found as the solution to a

constrained optimisation problem, where the maximally smooth image consistent with the given information is sought.

Chapter 7 provides a brief summary of the results presented in the thesis. Some open questions and areas for future research are described.

CHAPTER TWO

BACKGROUND

CHAPTER 2

BACKGROUND

2.1 Introduction

This chapter provides a review of some of the more important application areas where the limited information problem is encountered. As well as reviewing the background and solutions to these problems, we also look at some generalisations of the iterative algorithms that have been applied to this problem.

2.2.1 Computerised tomography

Computerised Tomography (CT) is easily the one application of image reconstruction that has captured public imagination. The principles of computerised tomography apply widely. The basic radiographic problem involves the reconstruction from projection measurements of the linear attenuation co efficient integrated along the path of a collimated X-ray beam. The basic reconstruction techniques are based on Radon's rigorous solution to the problem. This has subsequently been further developed by others [2], [3] and reconstruction from projections has found many new applications in such diverse fields as geophysics, radio astronomy, structural biology, non-destructive testing, etc. In its medical applications, CT has been used in a lot of modalities other than X-rays e.g. Nuclear Magnetic Resonance (NMR), Positron Emission Tomography (PET), and in different ray geometries.

The original uniqueness results obtained by Radon are in general valid only when scans in all directions are available even though there are some exceptions [1] such as when the unknown function is radially symmetric.

The incomplete information appears in several forms in CT. For example the Fourier phase problem appears when CT systems that measure their data in the Fourier space are unable to measure the phase. In many practical applications, requirements for high temporal resolution of the presence of an X-ray opaque structure prevent the measurement of all the line integrals or it may be desirable to reduce dosage by exposing the patient for a shorter time. Attempts to use existing algorithms in the limited data problem result in images with severe streak artifacts [6]. Alternatively, the modality being used, such as in oceonagraphy or electrical prospecting, the area involved may be so large that it is impractical to collect data in 360°.

Some recent workers [7], [8] have considered utilising the Gerchberg Papoulis iterative frequency domain extrapolation algorithms [38, 39]. This however is only really applicable when the missing projections are in a continuous range and only for parallel beam geometry. The missing projections then correspond to a segment of the frequency spectrum and can be extrapolated. Considering the same problem, Sezan and Stark [9] use the method of projection onto convex sets [10] which can allow the incorporation of more a priori constraints.

2.2.2 Synthetic aperture radar

The limited data problem also occurs in Synthetic Aperture Radar (SAR)

In high resolution radar imaging, the fidelity of reconstructed images is limited by the high side lobes of the point spread functions (psf). These sidelobes are essentially artifacts resulting from spectral components missing in the measured spectral data and are higher when the measured spectrum is very discontinuous. One of the earliest approaches to this problem was the use of 2-D tapered windows to suppress the sidelobes which is effective when many independent data samples are available. However it may lead to severe degradation when only a few samples are available. Alternative methods that have been used include the Gerchberg Saxton algorithm.

2.2.3 Reconstruction from contours

The problem of reconstructing images from contours has not received much direct interest in the gray scale case. When the image is bilevel, this corresponds to the classical contour filling problem. There have been several approaches to this and similar problems [11] [20]. As with the phase, there have been approaches to image coding where it has been found beneficial to emphasize the contour information and treat it separately from the other image information. Two source modelling was first proposed by Schrieber [13] and since then many workers have used composite source models in transform coding.

In the main, this work has considered the spatial image as a composition of two images [edge image, difference image] or th transform co efficients as composite sources [low frequency, high frequency].

One recent approach to reconstruction from contours [17] uses

spline interpolation to estimate intermediate texture. The basis of this and other approaches is that one does not try to reproduce the image intensity like traditional coders, but only tries to find a reconstructed image that is compatible with the given information.

We consider the application of some of these reconstruction methods to the problems of recovering gray scale images from their contours and polygonal approximation.

Classical approaches have had the following basis. There has been interest in coding methods that code pictorial feature rather than pixel arrays. (Since human observers do not seem to perform quantitative analysis on each pixel point, but rather to search for distinguishing features such as edges or textural regions.) For example, methods based on contours, edges and texture regions have been studied.

If a quantized image is considered as a stack of slices or planes each of constant gray level, the image may be represented by the boundaries or contours of the gray levels in each slice.

These contour may then be efficiently encoded for image transmission. The basic premise of contour coding as a means of bandwidth compression is that an image will contain a much smaller number of contours compared to the total number of its pixels and that the contours can be more efficiently encoded.

While this may be true for some images and compression rates of 7:1 for black/white images and 1.5:1 for gray scale images are obtainable, the whole process is rather complicated and suitable only for rather low transmission rates [19] for gray scale images,

higher compression rates may be obtainable if some of the information is left out.

An approach related to contour coding as described above is edge coding. Here the image is convolved with an edge detector and the edge map obtained. This is of course composed of contours which may be efficiently coded as described previously. The image is also low pass filtered. Thus its data may be reduced by taking only the major low frequency co-efficients after applying a suitable transform e.g. Fourier.

The image may then be reconstructed using these two compressed representations. At rates of up to 0.25 bits/pixel have been reported [19] on images of little detail, but the method has a high level of system complexity requiring transform , edge detectors, edge linkers, followers linkers, etc.

Finally we look at texture coding which also falls within the context of our problem. The basis here is that if the image is segmented into disjoint regions of fairly constant texture, then the regions (say contours) and the measured texture. Compression is achievable because the regions boundaries may be efficiently coded and the texture simple represented - perhaps by a simply measured textural value.

Reconstruction of the image takes place when the regions are reconstructed and their internal texture synthesised. This approach is still of some research interest, with many workers looking both into efficient analysis and synthesis of texture.

2.2.4 Bandlimited spectrum extrapolation

The most popular reconstruction problem among research workers is band limited signal extrapolation. Here one attempts to extrapolate a finite segment of data, given that it is band-limited. Papoulis [38] considered this problem for continuous signals and proposed an algorithm for solving it which iterated between bandlimiting the estimated signal, and then replacing the known segment with its correct value. Convergence was proved by exploiting the properties of prolate spheroid wave functions.

Sabri and Stenart [25] proposed a single step closed form solution to the problem using an "extrapolation matrix". Cadzow [26] reconsidered the problem, and by discretising the problem arrived at a superior closed form solution.

Gerchberg [27] considered the same problem with the frequency and time domains reversed. He estimated the high frequencies of a finite length signal from the given low frequencies using a similar iterative algorithm.

A related problem which forms the subject matter of much of this thesis is the reconstruction of an image from samples of the phase or magnitude of its Fourier transform together with some extra information such as finite support etc.

2.2.5 Reconstruction from phase only or magnitude only.

The phase and magnitude of the Fourier transforms of images play different roles and it is well known [28,107] that many of the important features of a signal may be preserved if only the phase is retained.

There are a number of important application areas where only one or another of the components of the Fourier transform of a signal can

be measured directly. In such situations, the need arises for the reconstruction of the original signal from the one available component.

Reconstruction from magnitude is a fairly well established problem that has also been called the 'phase retrieval problem' and the 'Fourier phase problem'.

The problem is to reconstruct the original phase function of an image given the discrete magnitude spectrum. This is often necessary in some applications where only the magnitude is available or can be measured and the phase must be reconstructed. These include X-ray crystallography [40], radioastronomy [42], Electron microscopy [41] and image processing [44] in optical astronomy.

In X-ray crystallography, the molecular structure of crystals is to be inferred from the observed diffraction of pattern of X-rays. The diffraction pattern is related to the scattering density of the crystal by a Fourier transformation, but only the intensity (squared magnitude) can be measured.

To determine the crystal structure, knowledge of the phase is indispensable and phase retrieval must be carried out. In optical and electron microscopy, the index of refraction of a thin object or the height distribution of a surface may need to be determined from the intensity of the wave distribution in the image plane. To determine the structure phase information is needed.

Imaging through a turbulent atmosphere may reduce the resolution of objects well below the diffraction limits of the telescopes. However the development of speckle interferometric technique it is now possible to obtain diffraction limited information about the Fourier transform intensity of the object [43]. Because of these

and other important applications there has been a lot of research effort by workers in these areas on both the uniqueness formulations and reconstruction methods. This problem is also of extreme interest in electrical engineering because of the importance of the Fourier transform in the subject.

There is also the dual problem which arises when dealing with complex sequences in the space domain. Let $f(x)$ be the space domain signal in any of the applications above and $F(\omega)$ it Fourier transform.

Complex Valued functions.

$$f(x) = |f(x)| \exp [j\phi_f(x)] \tag{2.1}$$

$$F(\omega) = |F(\omega)| \exp \{j\phi_F(\omega)\} \tag{2.2}$$

are thus defined where $|f(x)|$ and $|F(\omega)|$ are the magnitudes of the signal and its Fourier transform, and $\phi_f(x)$ and $\phi_F(\omega)$ are their corresponding phase function

In one dimension (1-D) the two dual problems of whether $|F(\omega)|$ uniquely defines the phase of $F(\omega)$ given that $f(x)$ is time-limited and whether $|f(x)|$ uniquely defines the phase of $f(x)$ given that $F(\omega)$ is bandlimited were considered by Hofstetter [22] and Walther [45]. They showed that neither time nor band limitation is, in general, sufficient to ensure a unique solution to the phase retrieval problem.

This lack of uniqueness is attributed to the possibility of 'zero flipping'. The flipping of a zero about the $j\omega$-axis preserves the magnitude of the Fourier transform as well as the duration of the signal and hence allows 2^p (p = number of zeros) different signals having the same duration and Fourier transform magnitude as

$f(x)$. However a unique solution is obtained if all the zero are imaginary.

Because of this ambiguity, as a result of zero flipping, workers have searched for solutions based on the availability of additional information. Such an ara is electron microscopy. For example Greenway [46] has shown that the presence of any interval over which the field in the object plane is known to be zero, is sufficient to ensure a unique solution. Furthermore both the field in the exit pupil and in the image plane are known to be entire functions due to the finite extent of the field in the object plane and finite size of the aperture in the exit pupil. Hoenders [47] has shown that these constraints reduce the phase ambiguity to a single field $f(x)$ or $f^{*}(-x)$.

Another variation of this problem is the case where the field intensity in two planes known - e.g. electron microscopy where the field is measured both in the image plane as well as in the exit pupil plane giving both $|F(\omega)|$ and $|f(x)|$. Another example is when the field is measured in two slightly defocussed planes in an optical imaging system.

In the first case above, Huiser [24] showed that the solution is unique to within a constant phase factor for analytic functions and in the second case too, unique solutions have been obtained [47].

Further information has been obtained by adding a known reference signal to the unknown signal, prior to the observation or measurement of the magnitude [48] such a procedure is used in holography. Knowledge of the reference signal may allow the phase

information to be retrieved and the original complex function to be reconstructed.

With proper choice of reference signal, one may ensure that the observed signal is minimum phase and hence use the Hilbert transform to retrieve the phase. The use of Hilbert transform relations for phase retrieval has been considered by many workers e.g. [58].

Phase retrieval in the 1-D case has been widely studied as outlined above. Generally any given $|F(\omega)|$ corresponds to an enumerable (sometimes finite) set of $f(x)$ satisfying some conditions [22].

The reduced ambiguity of the two dimensional (2-D) phase retrieval problem was noted by several workers, but one of the earliest detailed studies of the 2-D case was by Bruck and Sodin [23]. They considered a case of reconstruction a discrete image function $f(x,y)$ from its discrete autocorrelation function. They Fourier transform is related to the corresponding z-transform which is a polynomial $P(z,w)$.

They defined an equivalence relationship between two polynomials

$$P(z,w) \Delta Ap(z,w)\ z^s\ w^t \qquad (2.2.1)$$

A is a constant

z^s represents a shift along the x axis

w^t represents a shift along the y axis

Knowing the spectrum modulus is equivalent to knowing $p(z,w)$, $p(z^{-1}, w^{-1})$ or its equivalent polynomial

$$Q(z,w) = p(z,w)\ p(z^{-1}, w^{-1})\ z^m w^n \qquad (2.2.2)$$

They show that there is a large number of multiple solutions in the 1-D case.Their basic postulate is that the uniqueness of a function is linked to the irreducibility of its z-transform.

In the 2D as opposed to the 1D case, there exist polynominals that cannot be factored. Since the probability of finding a non-factorisable polynominal is much higher than that of a factorisable polynominal, multiple solutions are not as common in the 2D case as the 1D case. Any polynomial of the form

$$p(z,w) = q(z) + w^k$$

is non-factorisable for k>1.

Bruck and Sodin further postulated that the factorisability of any polynominal geometrically implies latent or overt repeatability of the image elements. For example a structure may be superposed on other elements by shifts and constants. The same autocorrelation polynominal but with opposite shifts would lead to quite a different image function. So only images characterised by the above repeatability are not reconstructed uniquely. They did not provide a procedure for determining $p(z,w)$ from $Q(z,w)$ but conceded that it would be a quite complicated solution of a set of second order equations.

Recently, Hayes [30] extended the Bruck and Sodin postulate that the uniqueness of a 2D sequence with finite support is related to the irreducibility of its z-transforms. The polynominals are taken to be unique if they have at most one irreducible non-symmetric factor. Lawton [59] applies the Poisson summation formula in conjunction with certain properties of functions which are 2D generalisations of the prolate spheroid wave function to derive an

algorithm for determining $X(m,n)$ from $|X(u,v)|$ over parallelogram shaped regions.

Fienup [37] Quatieri [35] Hayes [30] used modified forms of Gerchberg and Saxton algorithm to solve the problem. Gonzalves [60] proposed an alternative algorithm which uses a polynomial expansion for the phase. The phase is modelled as a summation of Zernike polynomials and a cost vector.

Reconstruction from phase has only recently received the sort of attention that was given to the phase retrieval problem. This is rather suprising because of the observed importance of phase over magnitude in image intelligibility. The applications of phase-only images are not as numerous. Phase only images can be used for image alignment, taking advantage of the fact that the autocorrelation function for phase only images is an impulse [31].

In image transform coding, an important bit rate reduction may be obtainable if it were possible to code only the phase of the Fourier transform of an image. Many coding techniques allow more bits for coding the phase than the magnitude [61], [62].

Another area where image formation relies heavily on the intelligibility of phase only images is in Kinoforms or phase only holograms [21]. The quality of images reconstructed from Kinoforms would be greatly improved if the magnitude function could be recovered from the phase function.

In the area of blind deconvolution the signal of interest has been degraded by a blurring function about which detailed knowledge is not available. In some special cases, the distorting signal may be known to have a phase function that is approximately zero and

consequently the phase of the degraded signal is very similar to that of the original image. In these cases the problem becomes one of reconstruction from phase only. Examples of this situation occur in seismic signal processing [8] as well as in image processing [32].

Phase only reconstruction may also be useful in the estimation of the frequency response of a Linear Time Invariant (LTI), if the symmetry of an input can be controlled [28].

In the restoration of images degraded by additive noise, phase only reconstruction could be useful. Such systems with additive noise are sensor noise and quantization noise in low data transmission systems. Usually filters used for such restoration are of zero phase and consequently only estimate the magnitude function without modifying the phase function. Some significant improvements are made if both the phase and magnitude are estimated [28], [64].

An analytic solution to phase only reconstruction is possible through the Discrete Hilbert Transform (DHT) for minimum phase sequences [34] but this requires the unwrapped phase. Alternative closed form solutions have been considered [63], [65] but there are many restrictions on the size of the matrices.

An iterative approach has also been used [35]. This is similar to the ones used of phase retrieval and move alternately between the space and frequency domain, imposing known constraints in each domain.

2.3 Iterative reconstruction algorithms

For many incomplete information problems, explicit inversion formulas are not known or are difficult to use. Consequently iterative reconstruction algorithms have been widely applied. Some of these have already been outlined above.

The structure of these algorithms are quite similar; we simply alternate between forcing time domain and then frequency domain constraints on the signal. This simple idea of iterating between two domains has encouraged many others to try and apply the same concept to more complicated problems. For example, Malik [66] solves a multidimensional maximum entropy (MEM) spectral estimation problem by iterating between the correlation domain and the convolution inverse of the correlation domain, forcing constraints on the model power spectrum in both domains in an attempt to find the MEM power spectrum.

Finite impulse response (FIR) filter design algorithms, such as the Remez exchange have been deliberately designed to try to iteratively adjust the filter co efficients in the time domain in order to decrease the worst errors in the frequency domain. More extreme examples are iterative autoregressive moving average (ARMA) modelling algorithm of Eykoff [49] or the iterative inverse filtering algorithms of Konvalinka [50] which iterate between estimating residuals, poles and zeros in a manner that appears to solve the corresponding modelling problems.

Recognising the conceptual similarity of all these algorithms, as well their resemblance to certain iterative deconvolution algorithms, numerous authors have tried to unify the presentation

and convergence proofs of these algorithms. For example, using the notion of non-expansive and contraction mappings, Tom et al [67] have shown that when the solution to the reconstruction problem is unique, then convergence of the bandlimited and the phase-only reconstruction algorithms could be proved by showing that each iteration of the algorithm defined a strictly non-expansive mappings.

Fixed point theorems of Ortega [51] were then invoked to prove convergence. Schafer et al [52] take an identical approach to prove convergence of deconvolution and bandlimited extrapolation algorithms.

Youla [54] considered the Papoulis bandlimited extrapolation problems as only one example of a class of iterative projection algorithms involving two sets of constraints on projections of the unknown signal. By considering the more general problem in an abstract Hilbert space setting, he characterised the properties of the algorithm in terms of the 'angle' between the constraint spaces. Mosca [53] treated the same subject in depth analysing the various degeneracies possible in solving ill-behaved linear problems in infinite dimensional spaces.

Jain [55]interprets the bandlimited extrapolation problem as solving a least squares problem. They derive Papoulis iterative algorithm, discuss closed form solutions in terms of Discrete Polate Spheroid Function and they show that Cadzow's closed form solution is the minimum norm solution to the least squares problem. The least squares approach leads to a conjugate gradient iterative algorithm. Musicus [68] adopts a similar approach but uses Minimum Cross Entropy (MCEM) optimality criteria instead of simple leasts quares.

It appears that a particularly rewarding approach to the formulation of new algorithms is to define an objective function measuring the goodness of the estimate and then to optimise this function iteratively. When the objective function is convex, the resulting iterations are often contraction or non-expansion mappings and hence an algorithm is generated whose convergence can be easily verified.

CHAPTER THREE

PHASE ONLY RECONSTRUCTION

CHAPTER 3

PHASE-ONLY RECONSTRUCTION

3.1 Introduction

In the previous chapter, we have reviewed the background work in the general reconstruction problem. In this chapter, we re-open the discussion in the specific case of Phase Only Reconstruction (POR).

It is possible to develop a variety of conditions under which an image may be recoverable from its phase or magnitude or any other part of its Discrete Fourier Transform (DFT) representation.

In this chapter we examine in detail some of the conditions that have been developed for phase only reconstruction of signals with a view to application both in the later parts of this chapter and in the phase coding problem discussed in chapter five.

While a sequence generally cannot be uniquely defined in terms of only its DFT phase or magnitude, there are certain classes of sequences for which this unique specification may be possible. For example, there is a Hilbert transform relationship between phase and log-magnitude of the DFT of a minimum phase sequence. However as we discuss later, the minimum phase requirement is quite restrictive and hence it is necessary to study other conditions for uniqueness that may be satisfied by the images found in practical applications. Since the phase function may be subject to measurement error in some of the application areas, or to quantization noise in the coding application, it is necessary to study the effect of noise on phase only reconstruction.

This chapter is organised into two main parts. The first deals

with POR given the correct phase samples.

In section 3.2 we introduce some of the basic theory of sequence reconstruction from phase or magnitude, right up to the minimum phase condition. After this, we leave consideration of magnitude uniqueness to the next chapter and introduce several uniqueness conditions for the phase.

In section 3.3 we examine reconstruction algorithms that have been developed for POR and shown their suitability for reconstruction.

The second part of the chapter deals with some special extensions to the above theory and some consideration of the effects of noise. Section 3.4 is devoted to an experimental study of the effect of noise and to a consideration of the effect of noise on the iterative reconstruction.

Section 3.5 discusses the effect of the choice of the initial amplitude estimate on the convergence of the iteration.

3.2 Phase Uniqueness

Poisson's formulas or Hilbert transform relations can be developed formally from the properties of analytic functions. For example, given that the z-transforms are analytic, we have constraints such as the Cauchy-Riemann conditions relating the partial derivatives of the real and imaginary parts. Another is the Cauchy integral theorem, where the value of a complex function is specified everywhere inside a region of analyticity in terms of the values of the function on the boundaries of the region.

However, the approach developed here applies a basic causality

principle that allows a sequence to be recovered from its even part.

3.2.1 Uniqueness of real and Imaginary parts

Any arbitrary sequence $x(n)$ may be written as the sum of an even sequence $x_e(n)$ and an odd sequence $x_o(n)$

$$x(n) = x_e(n) + x_o(n) \tag{3.2.1a}$$

$$x_e(n) = \tfrac{1}{2}\,[x(n) + x(-n)] \tag{3.2.1b}$$

$$x_o(n) = \tfrac{1}{2}\,[x(n) - x(-n)] \tag{3.2.1c}$$

If $x(n)$ is causal, then it is possible to recover $x(n)$ from $x_e(n)$ and, except for $n = 0$, to recover $x(n)$ from $x_o(n)$. This is clear from the above equations since $x(-n) = 0$ for causal $x(n)$.

The Fourier transform $X(e^{j\omega t})$ of $x(n)$, is generally complex and may be written as, $(t=1)$

$$\begin{aligned} X(e^{j\omega}) &= R_e\,[X(e^{j\omega})] + j\, I_m\,[X(e^{j\omega})] \\ &= X_r(e^{j\omega}) + j\, X_I\,(e^{j\omega}) \end{aligned} \tag{3.2.2}$$

Given that $x(n)$ is real, $X(e^{j\omega})$ is conjugate symmetric and its real part is even and the imaginary part odd.

Moreover it can be proved that $X_r(e^{j\omega})$ is the Fourier transform of $x_e(n)$ and that $X_I\,(e^{j\omega})$ is the Fourier transform of $x_o(n)$ - see appendix 1.

Therefore, if $X_r(e^{j\omega})$ is known, then provided $x(n)$ is real, causal and absolutely summable (stable), then $X_I\,(e^{j\omega})$ is also known. This is because $x_e(n)$ may be determined from $X_r(e^{j\omega})$ and hence $x(n)$ maybe determined using (3.2.1) from which $X_I\,(e^{j\omega})$ may be obtained.

Generally the z-transform for x(n)

$$X(z)\Big|_{z=re^{j\omega}} = X(re^{j\omega}) = \sum_{n=o}^{\infty} x(n)\ r^{-n}\ e^{-j\omega n} \qquad (3.2.3)$$

Since $x(n) = U_t(n)\ x_e(n)$, where $U_t(n) = \begin{cases} 2, & n>o \\ 1, & n=o \\ =, & n<o \end{cases}$

$$X(re^{j\omega}) = \sum([x_e(n)].[U_t(n)\ r^{-n}])\ \ e^{-j\omega n} \qquad (3.2.4)$$

i.e. $X(re^{j\omega})$ is the z-transform of the product $x_e(n)$ and $U_t(n)r^{-n}$. The z-transform is thus a Fourier transform that is the convolution of two Fourier transforms

$$X(re^{j\omega}) = X_r(e^{j\omega}) * \left(\frac{1 + r^{-1}e^{-j\omega}}{1 + r^{-1}e^{-j\omega}} \right) \qquad (3.2.5)$$

for r>1.

Using the complex convolution theorem,

$$X(z)\Big|_{z=re^{j\omega}} = \frac{1}{2\pi j} \oint_C \frac{X_R(v)\ (e^{j\omega} + r^{-1}v)}{(e^{j\omega} - r^{-1}v)}\, dv \qquad (3.2.6)$$

C is the unit circle

X(z) outside the unit circle is described in terms of the known real part on the unit circle.

If $r=e^{j\theta\omega}$, (3.2.6) is rewritten as a line integral

$$X(z)\Big|_{z = re^{j\omega}} = \frac{1}{2\pi}\int_{-\pi}^{\pi} X_R\,(e^{j\theta})\; P(\theta-\omega)\; d\theta + \frac{1}{2\pi}\int_{-\pi}^{\pi} X_R\,(e^{j\theta}) Q(\theta-\omega)\; d\theta \qquad (3.2.7)$$

$$P = Re\left[\frac{1 + r^{-1}\, e^{j\theta}}{1 - r^{-1}\, e^{j\theta}}\right] \quad ; \quad Q = Im\left[\frac{1 + r^{-1}\, e^{j\theta}}{1 - r^{-1}\, e^{j\theta}}\right]$$

Therefore

$$X\,(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X_R\,(e^{j\theta})\; P(\theta-\omega)\; d\theta \qquad (3.2.8)$$

and

$$X_I\,(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X_R\,(e^{j\theta})\; Q(\theta-\omega)\; d\theta \qquad (3.2.9)$$

The above has now been derived starting from a representation of the real, causal, stable sequence by its even component.

A similar analysis may be carried out starting with the sequence represented by its odd component.

i.e. $x(n) = x_o(n)\; U_t(n) + x(o)\; \delta(n)$.

This gives

$$X_R(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X_I(e^{j\theta})\, Q(\theta-\omega)\, d\theta + x(o) \qquad (3.2.10)$$

$$X_I(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X_R(e^{j\theta})\, P(\theta-\omega)\, d\theta + x(o) \qquad (3.2.11)$$

To obtain direct relations between X_I and X_R on the unit circle, take the limits in 3.2.8 and 3.2.10 as $r \longrightarrow 1$. This is alright if the integrals are performed first. Direct substitution is dangerous because of the singularity at Cot(O), but using care at the vicinity of the singularity and interpreting the integrals as Cauchy principle values, we have

$$X_I(e^{j\omega}) = \frac{1}{2\pi} P_c \int_{-\pi}^{\pi} X_R(e^{j\theta})\, \mathrm{Cot}\left(\frac{\theta-\omega}{2}\right) d\theta \qquad (3.2.12)$$

and

$$X_R(e^{j\omega}) = x(o) - \frac{1}{2\pi} P_c \int_{-\pi}^{\pi} X_I(e^{j\theta})\, \mathrm{Cot}\left(\frac{\theta-\omega}{2}\right) d\theta \qquad (3.2.13)$$

where P_c denotes Cauchy principle value.

So $H_I(e^{j\omega})$ is obtained by the periodic convolution of $\mathrm{Cot}(-w/2)$ with $H_R(e^{j\omega})$, taking special care in the vicinity of the singularity at $\theta = w$.

3.2.2 Recovery of the z-transform from the phase or magnitude on the unit circle

Consider X(z) the z-transform of the sequence x(n)

$$X(z) = |X(z)|e^{j\arg\{X(z)\}} \tag{3.2.14}$$

Its complex logarithm is

$$K(z) = \log\{X(z)\} = \mathrm{Log}\,|X(z)| + j\arg\{X(z)\} \tag{3.2.15}$$

If K(z) is also the z-transform of a sequence k(n), then as previously shown, $\log X(e^{j\omega})$ and $\arg X(e^{j\omega})$ will be Hilbert transforms of each other, provided k(n) is real, causal and stable.

In particular the above constraint of k(n) implies K(z) has a region of convergence including the unit circle and hence is analytic in a region including the unit circle, giving a convergent power series representation

$$K(z) = \sum_{n=0}^{\infty} k(n)\, z^{-n} \tag{3.2.16}$$

K(z) is infinite at both the poles and zeros of X(z), so there can be no poles or zeros of X(z) within the region of convergence associated with K(z).

The ambiguity of arg $\{X(z)\}$ is resolved by the constraint that $K(z)$ be analytic which in turn implies arg $\{X(\omega)\}$ must be a continuous function of ω and a further requirement is that for $x(n)$ real, $k(n)$ is also real.

If $k(n)$ is causal then $K(z)$ and consequently $X(z)$ can be recovered from $X_R(e^{j\omega}) = \text{Log } [X(e^{j\omega})]$ or $Xi(e^{j\omega}) = \arg X(e^{j\omega})\}$ and with the previous results,

$$\text{Log } |X(e^{j\omega})| = k(o) - \frac{1}{2\pi} P \int_{-\pi}^{\pi} \arg [X_I(e^{j\theta}] \text{ Cot}(\frac{\theta-\omega}{2}) \, d\theta \qquad (3.2.17)$$

and

$$\arg |X(e^{j\omega})| = \frac{1}{2\pi} P \int_{-\pi}^{\pi} \log [X_I(e^{j\theta}] \text{ Cot}(\frac{\theta-\omega}{2}) \, d\theta \qquad (3.2.18)$$

So $|X(e^{j\omega})|$ is only specified to a constant multiplier if $x(o)$ is not known.

This requirement that $\log |X(e^{j\omega})|$ and $\arg [X(e^{j\omega})]$ be a Hilbert transform pair is the minimum phase condition and corresponds to the requirement that $k(n)$ is causal and $K(z)$ is analytic everywhere outside the unit circle. Since $K(z) = \log X(z)$, there can be no poles or zeros of $X(z)$ outside the unit circle. This is an alternative expression of the minimum phase requirement.

While real and imaginary part sufficiency relations were developed via the z-transform, similar conditions can be developed for the DFT of finite length sequences with a suitable definition of causality. Specifically, it may be shown [33] that if
$X(k) = X_R(k) + jX_I(k)$ is the DFT of the finite length (or periodic sequence $x(n)$ then

$$jX_I(k) = \frac{1}{N} \sum_{m=0}^{N-1} X_R(m)\, V_N(k-m) \qquad (3.2.19)$$

$$\text{where } V_N(k) = \begin{cases} -j2\ \mathrm{CoT}\ (\pi k/n), & k \text{ odd} \\ 0 & K \text{ even} \end{cases}$$

and

$$X_R(k) = \frac{1}{N} \sum_{m=0}^{N-1} jX_I(m)\, V_N(k-m) + x(o) + x(N/2)\ (-1)^k \qquad (3.2.20)$$

Unlike the z-transform, it is not possible in general to develop similar sufficiency relations between log magnitude and phase of the DFT for minimum phase sequences. Previously, the log of the z-transform

$$\mathrm{Log}\ \{X(z)\} = \log |X(z)| + j \arg \{X(z)\}$$

was interpreted as a legitimate z-transform of a causal, stable sequence $k(n)$.

However because the logarithm of a transform $X(z)$ has singularities corresponding to both the poles and zeros of $X(z)$, its

inverse is generally of infinite duration. Consequently, the inverse transform of the logarithm of a transform can not in general be a discrete Fourier transform. Interpreting log {X(z)} as a DFT of a sequence of length N, corresponds to the aliased sequence

$$x_{p(n)} = \sum_{r=-\infty}^{\infty} k(n + rN) \qquad (3.2.21)$$

3.2.3 The Minimum phase condition

In the previous section, it was established that Hilbert transform relations exist between the log-magnitude and phase of the z-transform of minimum phase sequences. Some statements of the minimum phase condition were then made.

In order to be able to relate these results to practical images, it is necessary to develop equivalent statements of these conditions and evaluate their applicability.

(I) The z-transform of a minimum phase sequence has no poles or zeros outside the unit circle.

For a finite duration sequence, the z-transform has only zeros. To apply this condition, one would need to locate the positions of the zeros in the complex plane- a non-trivial matter when dealing with the 2-D sequences that represent images.

(II) Rewriting X(z) in its general form as a rational function

$$X(z) = Az^{n_o} \frac{\prod_{k=1}^{M_i} (1 - a_k z^{-1}) \prod_{k=1}^{M_o} (1 - b_k z)}{\prod_{k=1}^{P_i} (1 - c_k z^{-1}) \prod_{k=1}^{P_o} (1 - d_k z)} \qquad (3.2.22)$$

X(z) is minimum phase if it and its reciprocal are both analytic for $|z|>1|$. This excludes poles or zeros on or outside the unit circle, and with $n_o = 0$

$$X(z) = \frac{\prod_{k=1}^{M_i} (1 - a_k z^{-1})}{\prod_{k=1}^{P_i} (1 - c_k z^{-1})} \qquad (3.2.23)$$

Consequently the minimum phase sequence is causal and the unwrapped phase function has no linear phase component.

The third condition of course is that the unwrapped phase and log magnitude are related through the Hilbert transform. These are all fairly difficult to apply to arbitrary images.

3.2.4 Other uniqueness conditions

Uniqueness conditions distinctly different from the minimum phase condition are developed by considering that one way another sequence with the same phase may be formed is by convolving the given sequence

with a zero phase sequence.

In order to understand the effect of convolving arbitrary sequences with zero phase sequences, it is necessary to study the properties of zero phase sequences. In 1-D such a study may be done by considering the pole-zero plots of such sequences.

The Fourier transform $G(\omega)$ of a zero phase sequence $g(n)$ is real and non-negative, consequently $g(n)$ is conjugate symmetric for complex sequences, and even for real sequences.

$$g(n) = g^*(-n) \tag{3.2.25}$$

and

$$G(z) = G^*(1/z^*).$$

Therefore the singularities of the z-transform $G(z)$ occur in conjugate reciprocal pairs.

$$\tan [\phi_g(\omega)] = 0 \text{ for all } \omega \tag{3.2.26}$$

so $$\phi_g(\omega) = 0 \text{ or } \pi \text{ for all } \omega$$

Starting off with a finite length sequence $x(n)$ we can develop the constraints on its z-transform which will ensure that it cannot be written as a convolution of another finite length sequence and a zero phase sequence without betraying the conditions.

Consider the finite length sequence $x(n)$ which has no zero on the unit circle or in conjugate reciprocal pairs and the zero phase (even) sequence $g(n)$ which are convolved as follows:

$$y(n) = x(n) * g(n) \tag{3.2.27}$$

therefore

$$Y(z) = X(z)\, G(z).$$

1. Y(z) contains conjugate reciprocal zeros or poles
2. Y(z) contains zeros on the unit circle
3. The zeros of X(z) are replaced with poles in Y(Z) at conjugate reciprocal locations.

In order for y(n) to finite in length, Y(Z) must either have zeros on the unit circle or in conjugate reciprocal pairs. If we constrain y(n) to have no zeros on the unit circle or in conjugate reciprocal pairs, then it cannot be written as the convolution of x(n) and g(n).

Condition 1.

If x(n) and y(n) are real sequences of finite length, N, with z-transforms which have no zeros in conjugate reciprocal pairs or on the unit circle and $\phi_x(\omega) = \phi_y(\omega)$ at N-1 distinct frequencies. Then $x(n) = \beta y(n)$ for some positive real number β.

Proof:

Consider g(n) defined as

$$g(n) = x(n) * y(-n) \tag{3.2.28}$$

If x(n) and y(n) satisfy the requirements of the condition, then

$$\tan\{\phi_g(\omega_k)\} = 0 \text{ for } k = 1,2,.....N-1$$

$\omega_1..\omega_k$ are frequencies in the interval

G(ω) is real and g(n) is zero outside the interval [-N+1, N-1]

$$I_m[G(\omega_k)] = \sum_{n=-N+1}^{N-1} g(n) \sin(n\omega_k) = 0 \tag{3.2.29}$$

$$\Longrightarrow \quad \sum_{n=1}^{N-1} \left[g(n) - g(-n)\right] \mathrm{Sin}(n\omega_k) = 0$$

It follows that $[g(n) - g(-n)] = 0$ as the functions $\{\sin \omega, \sin 2\omega, \ldots\ldots \sin n\omega\}$ form a Chebyshev set, and consequently $g(n)$ is even, and

$$G(Z) = X(Z)\ Y(Z^{-1}) = G(Z^{-1}) = X(Z^{-1})\ Y(Z).$$

If $X(Z)$ has a k^{th} order zero at $z = z_o$ $0 < |z_o| < \infty$. $Y(1/z_o)$ must be finite since $y(n)$ if finite length.

However, $G(Z-1)$ must also have a k^{th} order zero at $z = z_o$. Since $X(Z)$ does not have zeros in conjugate reciprocal pairs $X(1/z_o)$ must be finite, consequently $Y(Z)$ must have a k^{th} order zero at $z = z_o$. Therefore, $X(Z)$ and $Y(Z)$ have the same zero set.

$$X(Z) = \beta\ Z^k\ Y(Z) \tag{3.2.30}$$

However if $\phi_x(\omega) = \phi_y(\omega)$, then k must be zero and thus the condition is proved. «»

In the 2-D case, the finite length constraint is replaced by finite support constraint on the sequence and the Z-transform is a polynomial in two complex variables z_1, z_2. The z-transform $X(z_1, z_2)$ of the 2D sequence $x(n_1,n_2)$ is defined as

$$X(z_1,z_2) = \sum_{n_1} \sum_{n_2} x(n_1,n_2)\ z_1^{-n_1}\ z_2^{-n_2} \tag{3.2.31}$$

In this and the following chapters, it is assumed that all the sequences considered have a rational z-transform with a region of convergence which includes the unit polydisk

$$|Z_k| = 1 \quad \text{for } k=1, 2$$

Then the Fourier transform exists and is given by

$$X(\omega_1,\omega_2) = X(z_1,z_2)\Big|_{z_1,z_2 = \exp\{j\omega_1\},\ \exp\{j\omega_2\}}$$

$$= \sum_{n_1} \sum_{n_2} x(n_1,n_2)\ e^{jn_1\omega_1}\ e^{jn_2\omega_2}$$

or written in terms of phase and magnitude

$$= |X(\omega_1,\omega_2)|\ \exp\ j\phi_x(\omega_1,\omega_2)$$

The region of support of $x(n_1,n_2)$ is $R(N_1,N_2)$. This is the region inside which the sequence is non-zero and outside which the sequence is always zero. We assume non-negative sequences

i.e. $x(n_1,n_2) = 0,\ 0>n_1>N_1,\ \ 0>n_2>N_2$

finally, $F(n_1,n_2)$ denotes the set of all 2-D sequences having for some N_1,N_2, a region of support $R(N_1,N_2)$.

The multidimensional equivalent of the poles and zeros of a 1-D z-transform are the zero contours of the irreducible factors of a multidimensional z-transform.

Consider the symmetric z-transform $X(Z)$, where

$$\tilde{X}(z_1,z_2) = \pm z_1^{-k_1} z_2^{-k_2} X(z_1^{-1}, z_2^{-1}) \qquad (3.2.32)$$

k - positive integers.

Since a 1-D sequence which has a z-transform with all of its zeros on the unit circle or in reciprocal pairs is symmetric, this notion represents an extension of the 'zero phase' properties to multidimensional sequences. Hence, the multidimensional uniqueness, a similar statement can be made [30].

Consider $x(n_1,n_2)$ and $y(n_1,n_2)$ which are 2-D sequences in the set $F(n_1,n_2)$ with support $R(N_1,N_2)$. Let $X(z_1,z_2)$ and $Y(z_1,z_2)$ be their respective z-transforms and let $M_1 > 2N_1-1$ and $M_2 > 2N_2-1$.

If $X(z_1,z_2)$ has no symmetric factors and

$$\phi_x(\omega_1,\omega_2)_M = \phi_y(\omega_1,\omega_2)_M \qquad \text{for } \omega_1 = 1,\ldots\ldots M_1$$
$$\omega_2 = 1,\ldots\ldots M_2$$
$$M = M_1, M_2$$

then $y(n_1,n_2) = \beta x(n_1,n_2)$ for some number β

3.3 Reconstruction algorithms.

The algorithms employed to actually reconstruct an image from its phase fall into two basic classes - non-iterative algorithms which rely on the solution of large systems of equations and iterative algorithms which successively generate better estimates of the image at each iteration by imposing well chosen constraints. Strategies combining these two have been suggested in the literature and it is feasible that the disadvantages of each method may be alleviated by a hybrid method. e.g. a non-iterative approach may be used to give a good first estimate which would enhance the convergence of the iterative algorithm.

We shall not however examine this approach in any more detail and we restrict ourselves to studying and comparing some alternative methods and analysing ways of improving their performance.

3.3.1 Non-iterative algorithms

The POR problem can be formulated in several ways that allow a non-iterative solution. This section examines some of the algorithms that can be developed from these formulations.

If the image is assumed to be minimum phase, a reconstruction

algorithm can be derived based on implementing the DHT to find the missing magnitude function. However computation of the Hilbert transform is quite involved and requires the unwrapped phase [33]. Phase unwrapping is itself quite difficult [34] and an active research problem. Alternative algorithms that avoid direct implementation of the DHT have been tried [35] but these involve the use of an iterative algorithm to implement the DHT of the given function and are mentioned again in the next section.

Another approach that is applicable to minimum phase sequences is through the use of Cepstral co-efficients.

The DFT of the finite support real sequence $x(n)$ is a polynominal in $e^{j\omega}$ and can be factored

$$X(\omega) = |X(\omega)| \exp [j(\phi(\omega)] = x(0) \prod_{r=1}^{M} (1 - z_r e^{-j\omega}) \qquad (3.3.1)$$

where

$|z_r| < 1$ and $|X(\omega)|$ and $\phi(\omega)$ are the magnitude and phase respectively.

Taking the complex logarithm gives

$$\log |X(\omega)| + j\phi(\omega) = \log(x(0)) + \sum_{r=1}^{M} \log (1 - z_r e^{-j\omega}) \qquad (3.3.2)$$

and using the identity $\log(1-t) = - \sum_{r=1}^{\infty} \frac{t^r}{r}$ for $|t| < 1$

$$\log |X(\omega)| + j\phi(\omega) = \log(x(0)) - \sum_{r=1}^{M} \sum_{n=1}^{\infty} \frac{z_r \, e^{-jn\omega}}{n}$$

$$= \log x(0) - \sum_{n=1}^{\infty} \frac{\sum_{n=1}^{M} z_r^n}{n} e^{-jn\omega}$$

$$= \log x(0) - \sum_{n=1}^{\infty} \frac{S_n}{n} e^{-jn\omega} \qquad (3.3.3)$$

where $S_n = \sum_{n=1}^{M} z_r^n$

The S_n/n terms are the cepstral co-efficients of $x(n)$ and equating real and imaginary parts

$$\log |X(\omega)| = \log \left| x(0) \right| - \sum_{n=1}^{\infty} \frac{S_n \cos (n\omega)}{n} \qquad (3.3.4a)$$

$$\phi(\omega) = \phi(0) + \sum_{n=1}^{\infty} \frac{S_n}{n} \sin(n\omega) \qquad (3.3.4b)$$

So with the knowledge of $\phi(\omega)$, the inverse sine transform can be used to compute the cepstral co-efficients which in turn can be used to obtain $\log|H(\omega)|$ from equation (3.34a) except for the term $\log|x(o)|$. Therefore from the knowledge of $\phi(\omega)$ alone, we can compute the sequence $x(n)$ up to a constant scale factor.

This has the same disadvantage of requiring the unwrapped phase which is usually not available. Furthermore the cepstral co- efficients involve the powers of the zeros of the sequence and the zeros whose magnitude is less than unity die out rapidly, leading to numerical problems of recovering them. Consequently this is not really a practical algorithm either.

In the next approach, a closed form relationship is developed between the given phase and the original sequence.

Consider the 2-D sequence $x(n_1, n_2)$ with support $R(N_1, N_2)$ whose DFT is

$$X(\omega_1, \omega_2) = \sum_{n=0}^{N_1-1} \sum_{n=0}^{N_2-1} x(n_1, n_2)\, e^{-jn_1\omega_1}\, e^{-jn_2\omega_2}$$

Suppose that the z-transform $X(z_1, z_2)$ does not have any symmetric factors then

$$\frac{X(\omega_1, \omega_2)}{X(-\omega_1, -\omega_2)} = \frac{\sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x(n_1, n_2)\, e^{-j(n_1\omega_1 + n_2\omega_2)}}{\sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x(n_1, n_2)\, e^{j(n_1\omega_1 + n_2\omega_2)}} \qquad (3.3.5)$$

$$= \frac{M_x(\omega_1,\omega_2)e^{j\phi_x(\omega_1,\omega_2)}}{M_x(\omega_1,\omega_2)e^{-j\phi_x(\omega_1,\omega_2)}}$$

where $M_x(\omega)$ and $\phi_x(\omega)$ are the 2-D magnitude and phase functions respectively.

$$M_x(\omega_1,\omega_2)e^{-j\phi_x(\omega_1,\omega_2)}\cdot\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1} x(n_1,n_2)\ e^{-j(n_1\omega_1+n_2\omega_2)}$$

$$= M_x(\omega_1,\omega_2)e^{j\phi_x(\omega_1,\omega_2)}\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1}x(n_1,n_2)e^{j(n_1\omega_1+n_2\omega_2)}$$

which reduces to

$$\sum_{n_1=0}^{N_1-1}\sum_{n_2=1}^{N_2-1} x(n_1,n_2)\ e^{-j\{n_1\omega_1+n_2\omega_2+\phi_x(\omega_1,\omega_2)\}}$$

$$= \sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1} x(n_1,n_2)e^{j\{n_1\omega_1+n_2\omega_2+\phi_x(\omega_1,\omega_2)\}}$$

This is equivalent to

$$\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1} x(n_1,n_2)\ \sin\{n_1\omega_1+n_2\omega_2+\phi_x(\omega_1,\omega_2)\} = -x(0,0)\ \sin\{\phi_x(\omega_1,\omega_2)\} \qquad (3.3.7)$$

So, given the phase $\phi(\omega)$ for (N-1 x N-1) distinct values of ω between 0 and π, then a system of linear equations in N-1 x N-1 unknowns is obtained if we assume $x(0,0) = 1$.

$$\sum_{n_1=1}^{N_1-1} \sum_{n_2=1}^{N_2-1} x(n_1,n_2) \sin\{n_1\omega_1+n_2\omega_{2+}\phi x(\omega_1,\omega_2)\} = -\sin\phi_x(\omega_1,\omega_2) \qquad (3.3.8)$$

This result has been derived before by other workers for 1-D sequences e.g. [36] using trigonometric considerations. The derivation outlined above for 2-D sequences appears to be better because it shows in a convincing way the importance of the absence of symmetric factors on phase only uniqueness. The system of equations can be solved exactly for N-1 unknowns using any standard technique.

If $x(0,0)$ is not known (3.3.8) only specifies $x(n_1,n_2)$ to within a scale factor. The potential problem of numerical instability caused by severe round off errors for large values of N is a limiting factor on the utility of this approach.

3.3.2 Iterative algorithms

Due to the difficulties outlined in the previous section, non-iterative solutions of the POR problem are not always practical. In fact for many other related problems such as the magnitude only reconstruction problem, such closed form relationships are very hard to come by and the more common approach is the one that utilises iterative algorithms.

The iterative algorithm discussed in this section is basically similar to the original algorithm used by Gerchberg and Saxton [39], Papoulis [38] and subsequently described in [37], [35] etc.

These algorithms alternately move between the space and frequency domains at each iteration imposing known constraints in each domain this family of algorithms is illustrated in fig. (3.3.1).

The basic algorithm used here is:

1° Make an initial guess of the unknown DFT magnitude $|X_o(\omega_1,\omega_2)|$ form the next estimate of the DFT by combining the magnitude guess with the known phase

$$X_1(\omega_1,\omega_2) = |X_o(\omega_1,\omega_2)| \exp\{j\phi_x(\omega_1,\omega_2)\}$$

Compute the inverse DFT

2° Apply the finite support and positivity constraints

$$x_p(n_1,n_2) = \begin{cases} x_p(n_1,n_2) & n_1 < N_1,\ n_1,n_2 \geq 0,\ n_2 < N_2 \\ x(0,0) & n_1 = n_2 = 0 \\ 0 & \text{Otherwise} \end{cases} \qquad (3.3.9)$$

$$\text{and } x_p(n_1,n_2) = \begin{cases} x_p(n_1,n_2) & \text{for } x_p(n_1,n_2) > 0 \\ 0 & \text{for } x_p(n_1,n_2) < 0 \end{cases} \qquad (3.3.10)$$

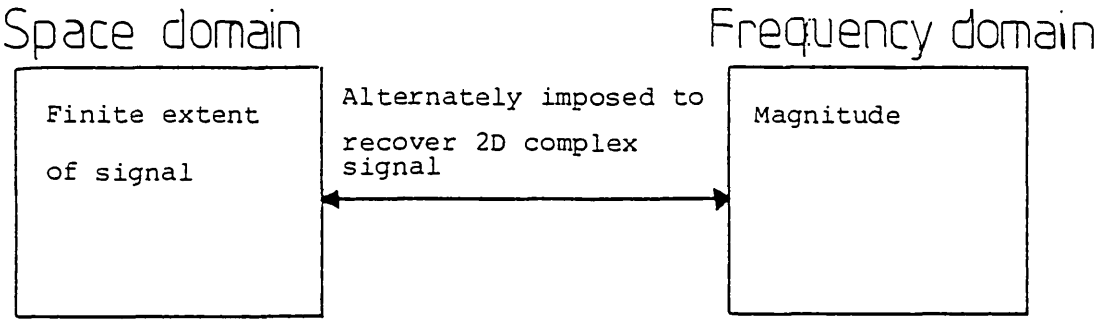
Compute the M-point DFT

3° The magnitude $|X_p(\omega_1,\omega_2)|$ of the DFT is used as the new estimate of $|X(\omega_1,\omega_2)|$ and the next estimate is formed as

$$X_{p+1}(\omega_1,\omega_2) = |X_p(\omega_1,\omega_2)| \exp[j\phi x(\omega_1,\omega_2)] \qquad (3.3.11)$$

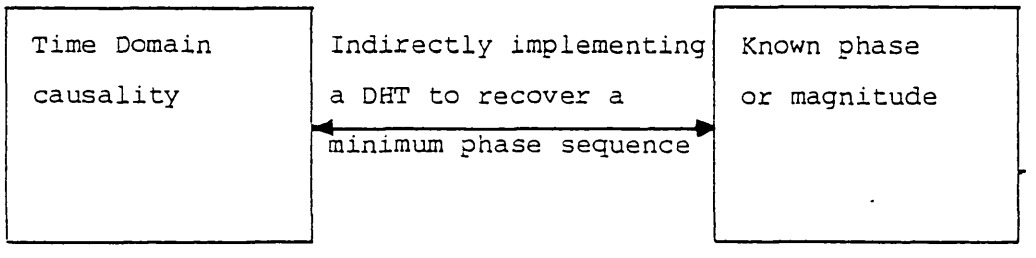
Compute the M-point IDFT

Go to 2°.



(a) Gerchberg and Saxton

(b) Fienup



(c) Quateri

Fig.3.1 Iterative reconstruction algorithms

The iteration has converged when a sequence with the correct support and phase function is obtained

This algorithm is illustrated in the block diagram of fig. 3.3.2

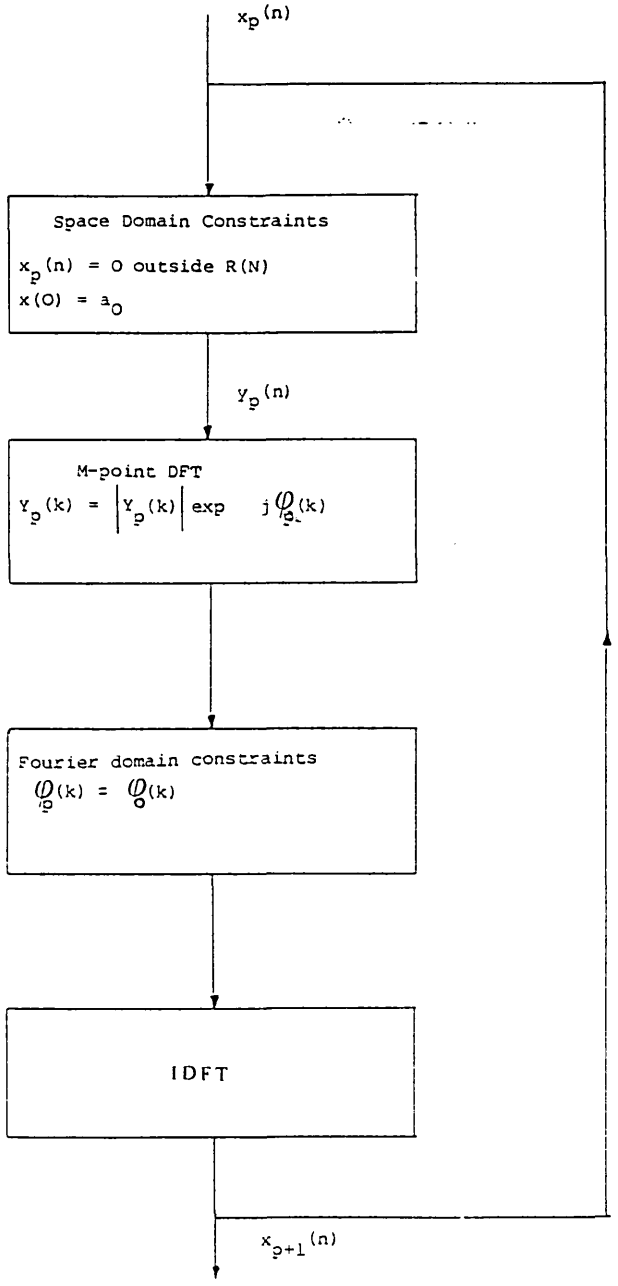


Fig. 3.3.2 Block diagram of iterative reconstruction algorithm

This algorithm was implemented on the computer and used to reconstruct the 128 x 128 pixel image shown in fig (3.3.3). (a) shows the original image and (b) shows the phase only image formed by taken the magnitude of the DFT to be unity. The reconstructed image after 30 iterations is shown in (c). In order to perform M-point DFT's for NxN images where M > 2N-1, it was found necessary to break up the input images into smaller blocks which were treated separately and later combined to form the image.

As a consequence of this approach, there sometimes appears a discernible 'blocking effect' in the reconstructed image, with the borders of the blocks imposing themselves on the image. This familiar problem with the blocking effect could be reduced by overlapping the blocks or filtering and serves to illustrate one of the problems involved in using this iterative algorithm, i.e. the necessity to perform 2N x 2N point DFT's. The other problem is the slow convergence of the algorithm.

Fig. (3.3.4) shows a plot of the mean square error Vs the number of iterations and it is clear that while the error decreases sharply at the beginning, the rate of convergence soon slows down.

3.3.2.1 Adaptive acceleration

1. The Gerchberg iteration may be considered as a functional relationship between consecutive estimates

i.e. $x_{p+1}(n) = T(x_p(n))$, where $x(n)$ is a vector representing the 2-D image, this may be rewritten as

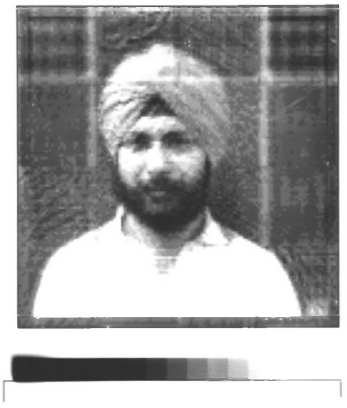
$$x_{p+1}(n) = (1 - \lambda_p)x_p(n) + \lambda_p T(x_p(n)) \qquad (3.3.12)$$

λ_p is a relaxation parameter

and $T()$ represents the general iterative process.

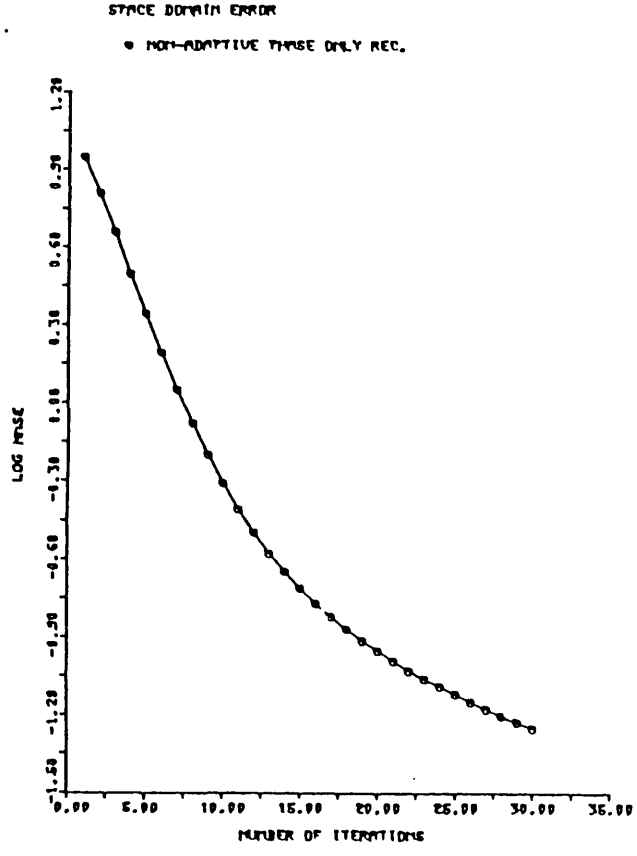
(a) Original 'Man' image

(b) phase only image

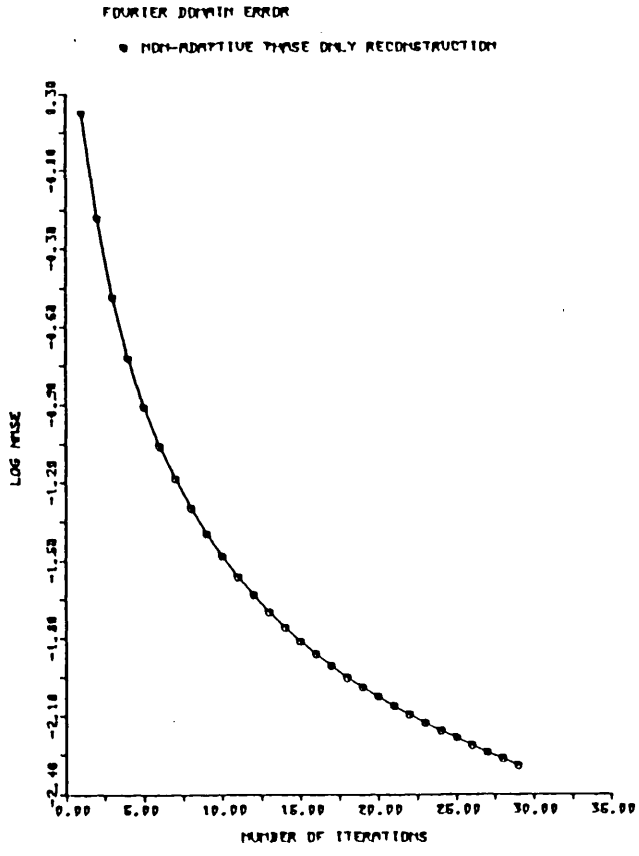


(c) Reconstruction after 30 iteration of POR algorithm

Fig. 3.3.3 Phase Only Reconstruction



(a) Space domain error



(b) Fourier domain error

Fig. 3.3.4 MMSE V's error behaviour of the non-adaptive algorithm

$x_{p+1}(n) = \lambda_p r_p(n)$ where, $r_p(n) = T(x_p(n)) - x_p(n)$

If $\lambda_p = 0$ trivial

$0 < \lambda_p < 1$ under relaxed.

λ_p may be chosen to minimise a measure of the estimates that fall outside region of support.

It is a non-trivial problem to determine what the optimum choice of λ_p should be, particularly as it implies an assumption about the relative importance of the a priori constraints. In the phase only iteration it is fairly clear that the finite support constraint is important and λ_p was chosen to minimise the Euclidean norm of the vector of non-zero points outside the region of support.

An alternative choice of λ_p would emphasize the non-negativity (or positivity) constraint and could be chosen to minimise the euclidean norm of the vector of non-negative points within the region of support. Suppose that the elements of $x(n)$ within $R(M)$, $(M = 2N)$ are partitioned into a vector V_x such that,

$$V_x = \begin{bmatrix} V_x^{(1)} \\ \\ V_x^{(2)} \end{bmatrix}$$

Where $V_x^{(1)}$ is the NxN vector of non-negative elements and $V_x^{(2)}$ is the vector of negative elements in the image.

Consequently, rewrite

$$V_{p+1} = V_p + \lambda_p r_p$$

and

$$\begin{bmatrix} V_{p+1}^{(1)} \\ \\ V_{p+1}^{(2)} \end{bmatrix} = \begin{bmatrix} V_p^{(1)} \\ \\ V_p^{(2)} \end{bmatrix} + \lambda_p \begin{bmatrix} r_p^{(1)} \\ \\ r_p^{(2)} \end{bmatrix} \qquad (3.3.13)$$

Convergence is obtained when $V_{p+1}^{(2)} = 0$, so choose $\lambda_p = \hat{\lambda}_p$ to minimise the euclidean norm of $V_{p+1}^{(2)}$

i.e $$\left[\frac{d}{d\lambda_p} \left\| V_{p+1}^{(2)} \right\| \right]_{\lambda_p = \hat{\lambda}_p}$$

The euclidean norm is defined as [106]

$$\|x\| = (x,x)^{1/2} \qquad \text{where } (x,x) \text{ is the inner product}$$

hence:

$$\frac{d}{d\lambda_p} \left\| V_{p+1}^{(2)} \right\|^2 = \frac{d}{d\lambda_p} \left[V_{p+1}^{(2)},\ V_{p+1}^{(2)} \right] = \frac{d}{d\lambda_p} \left[V_p^{(2)} + \lambda_p r_p^{(2)} ,\ V_p^{(2)} + \lambda_p r_p^{(2)} \right]$$

using the identity

$$(x + y,\ z) = (x,z) + (y,z)$$

$$\frac{d}{d\lambda_p} \left\| V_{p+1}^{(2)} \right\| = \frac{d}{d\lambda_p} (V_p^{(2)}, V_p^{(2)}) + \frac{d}{d\lambda_p} (V_p^{(2)}, \lambda_p r_p^{(2)}) + \frac{d}{d\lambda_p} (\lambda_p r_p^{(2)}, V_p^{(2)})$$

$$+ \frac{d}{d\lambda_p} (\lambda_p r_p^{(2)}, \lambda_p r_p^{(2)}) = 0$$

Since $(\alpha x, y) = \alpha (x,y)$, we have

$$\frac{d}{d\lambda_p} \left\| V_{p+1}^{(2)} \right\| = \frac{d}{d\lambda_p} (V_p^{(2)}, V_p^{(2)}) + \frac{d}{d\lambda_p} \lambda_p (V_p^{(2)}, r_p^{(2)}) + \frac{d}{d\lambda_p} \lambda_p (V_p^{(2)}, r_p^{(2)})$$

$$+ \frac{d}{d\lambda_p} \lambda_p^2 (r_p^{(2)}, r_p^{(2)}) = 0$$

$$= 0 + (r_p^{(2)}, V_p^{(2)}) + (r_p^{(2)}, V_p^{(2)}) + 2\lambda_p (r_p^{(2)}, r_p^{(2)}) = 0$$

$$= 2(r_p^{(2)}, V_p^{(2)}) + 2\lambda_p (r_p^{(2)}, r_p^{(2)}) = 0$$

hence $$\lambda_p = -\frac{(r_p^{(2)}, v_p^{(2)})}{(r_p^{(2)}, r_p^{(2)})}$$

$$= -\frac{(r_p^{(2)}, v_p^{(2)})}{\|r_p^{(2)}\|} \qquad (3.3.14)$$

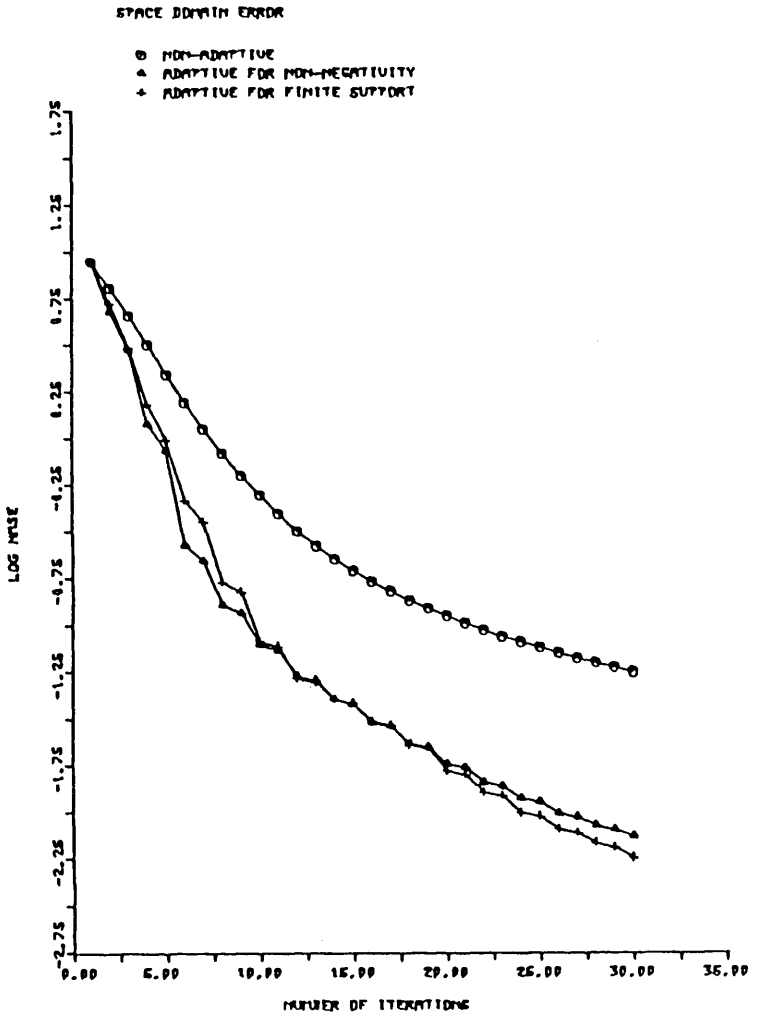
The adaptive algorithm described above was implemented for the phase only iteration and the graph in fig 3.3.5 shows the relative performance of the three algorithms implemented so far.

3.4 Effect of noise on the phase only reconstruction

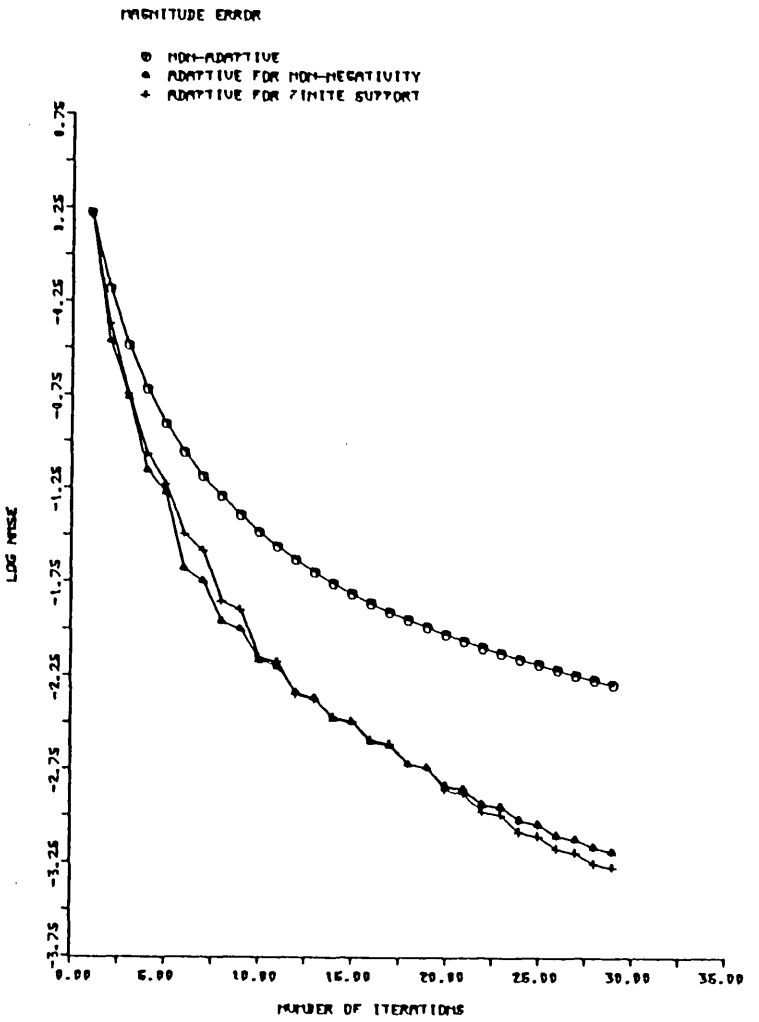
In any application setting, there will be limits to the accuracy to which the phase can be measured or computed and thus the available phase may have been degraded by measurement noise, quantisation noise etc.

Consequently it is important to understand the sensitivity of the reconstruction methods to errors in the phase samples. Some experimental results have been reported for the effect [63] of noisy phase on the reconstruction of a 1-D sequence with the non-iterative algorithm (3.3.8).

This section provides experiments with images for iterative reconstruction algorithms and attempts to provide a theoretical analysis of the errors introduced by noisy phase for images and its effect on the iteration. Since the convergence of the phase only iteration has been proved, it is possible to show that provided the image sequence corresponding to the noisy phase sequence is close to the correct sequence the iteration converges, even though the point to which it converges may change. This is proved in appendix III.



(a) Space domain error



(b) Fourier domain error

Fig. 3.3.5 Adaptive POR reconstruction algorithms

Two kinds of noise effects were studied.

The first was gaussian noise added directly to the undegraded phase and reconstruction attempted. The second investigated the effect of various levels of quantisation noise by varying the number of levels used to quantise the phase. Various methods have been proposed for quantising the phase, e.g. Andrews [69], Pohlig [61]. The method used here is broadly the same as that proposed by Pohlig and makes no assumption about the distribution of the phase. This method is discussed in some detail in chapter five, but it can be broadly described as allocating more levels to code the lower frequencies. It is interesting to note that the bit allocation study based on the phase gives broadly the same result as one based on the magnitude. That is that the number of bits needed is inversely proportional to frequency - this probably explains the success of coding techniques that used bit allocation for the phase based on magnitude considerations.

The normalised mean square error (NMSE) given by

$$\frac{\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1}[x(n_1,n_2) - \hat{x}(n_1,n_2)]^2}{\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1}x(n_1,n_2)^2} \qquad (3.4.1)$$

was used to compare the effect of different noise levels. The iteration was allowed to run for 30 iterations in each case and the NMSE computed. A graph of noise level Vs NMSE was plotted for various error levels and these are shown in fig. 3.4.1. It is of interest to observe the error in the reconstructed sequence as a function of the number of iteration. Although the error in the reconstructed sequence will be the same in the limit, as that obtained in the non-iterative algorithm, it may be possible that the error after a finite number of iterations is less than the error of the convergent solution.

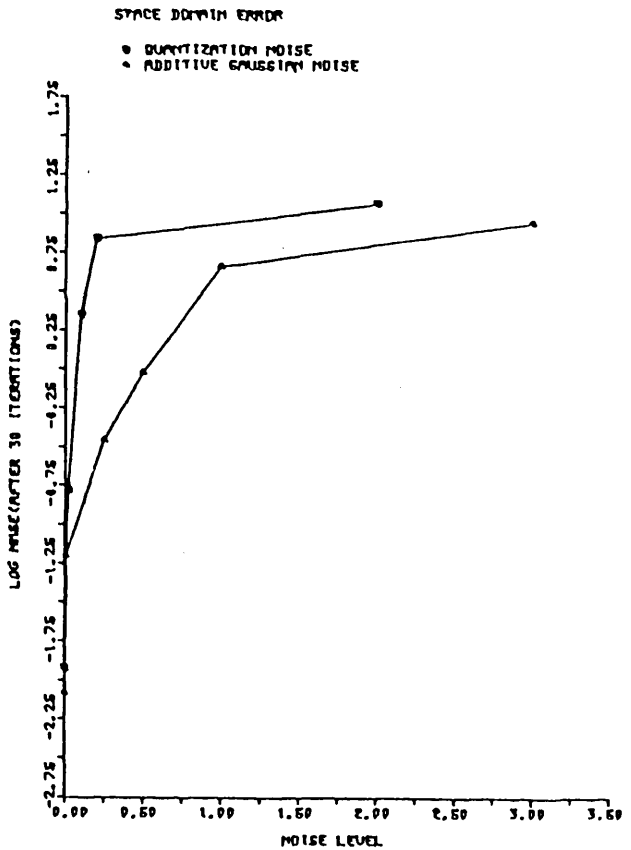
One possibility investigated here is the attempt to mitigate the effects of phase noise by combining it with an estimate of the correct magnitude. The magnitude may be estimated in any one of the ways discussed in some detail in the next section, or may be a measure of the noisy magnitude. The graph in fig. 3.4.2 shows that convergence is faster with the magnitude estimate and the intelligibility of the reconstructed image shows that phase coding with a magnitude estimate is feasible.

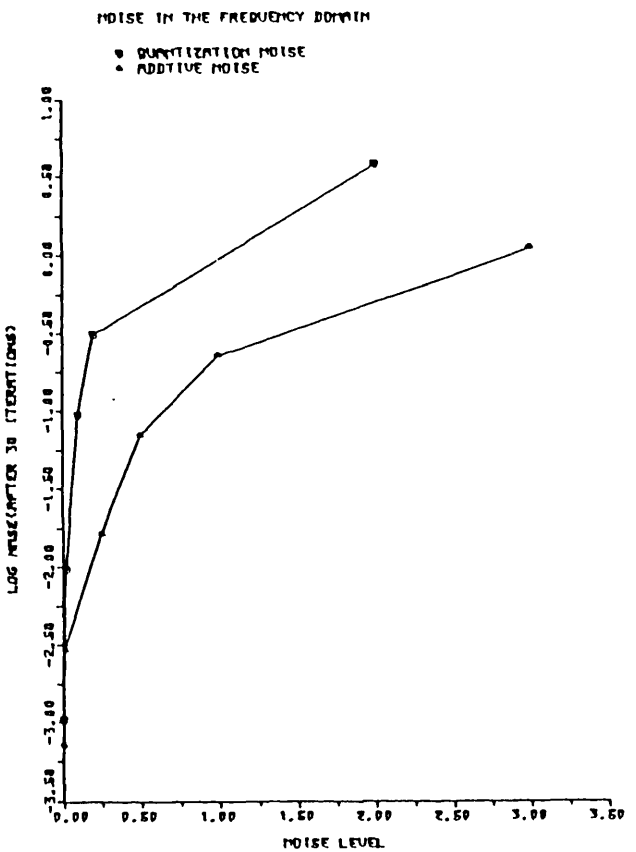
3.5 Effect of initial magnitude on phase reconstruction

Conventionally, the phase-only image is defined as the image formed by performing the inverse DFT on a function having the same phase as the original image and unity or constant magnitude

i.e. $F_p\ (n_1,n_2) = F^{-1}\ [1 \times \exp\ [\phi(\omega_1,\omega_2)]]$

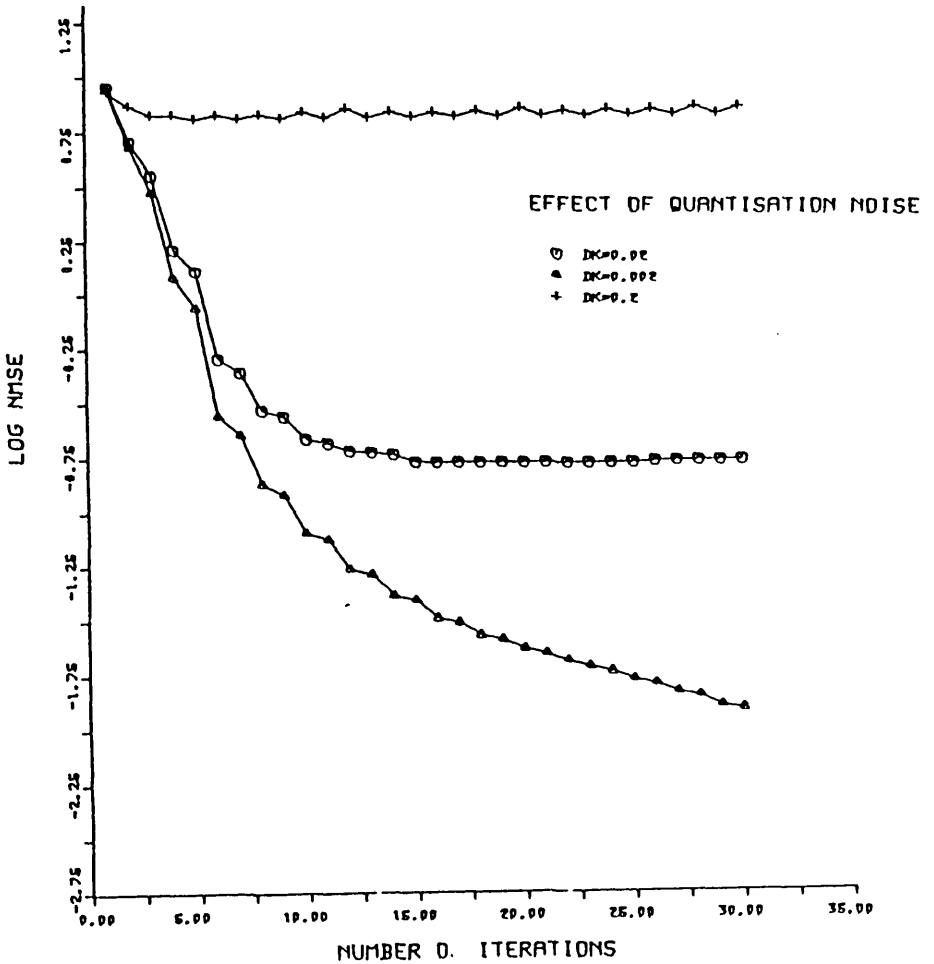
In the phase only iteration, whatever magnitude guess is used, serves only as an initial estimate and is later allowed to change so as to



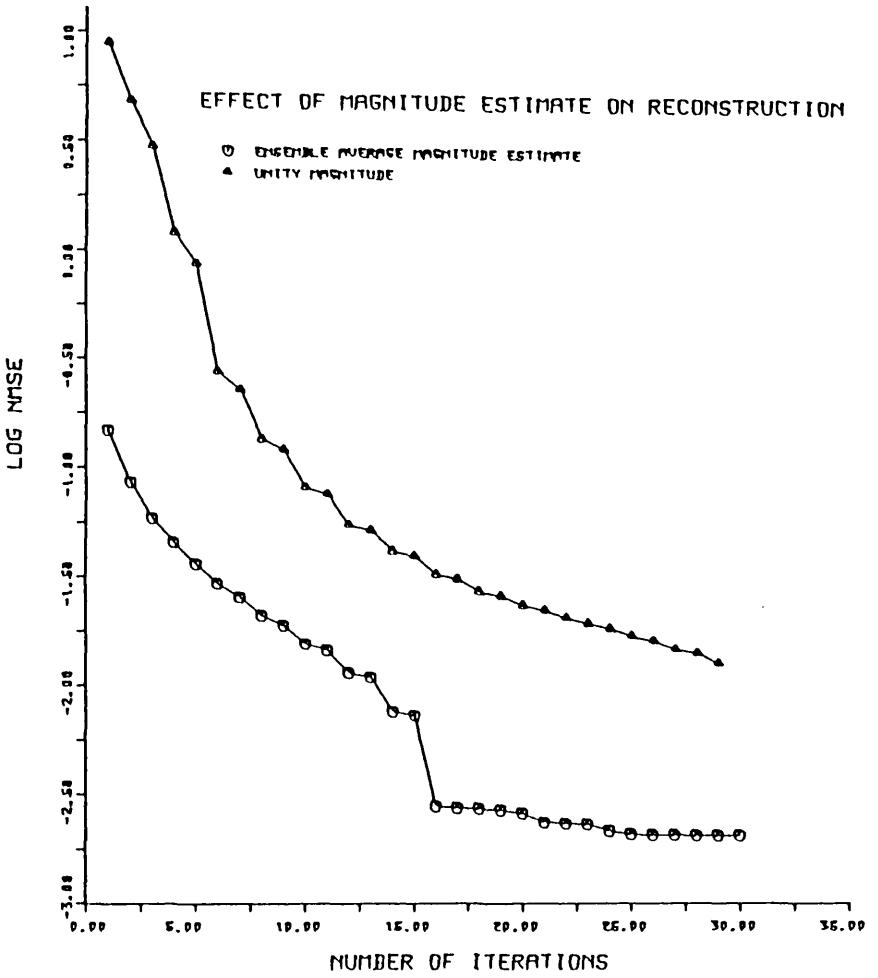


ISE

Fig. 3.4.1 Reconstruction from noisy phase



(a) Quantization noise V's Number of iterations



(b) Magnitude estimation for DK = 0.002

Fig. 3.4.2 Effect of magnitude estimation on reconstruction

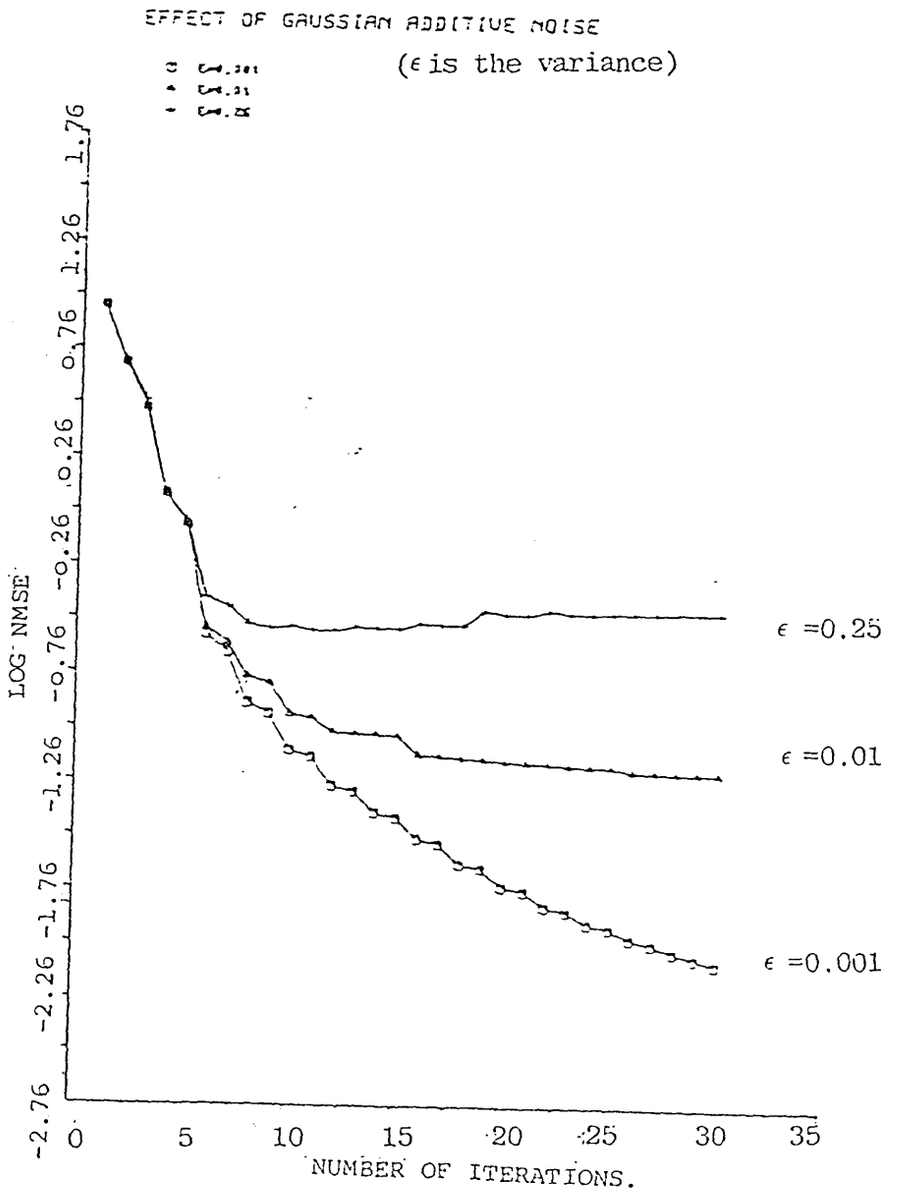


Fig. 3.4.2 Effect of additive gaussian noise on convergence of the phase only iteration.

satisfy the given space and frequency domain constraints.

We have already observed that phase only images as described above bear a close resemblance to the original image and it has also been observed that even when the magnitude is a random function some intelligibility is maintained.

Obviously there are limits to the sort of values that the magnitude function can be allowed to take so as to give an intelligible first estimate. For example zero would be illogical and extremely large values would also be out of the question.
It has already been established that the iteration converges much faster if a first estimate close to the final image is obtained so it can be argued that clever choice of the initial magnitude guess can substantially speed up the iteration.

Clearly, it would be to some advantage if further a priori information could be incorporated into the choice of the magnitude function and while this may not be possible in any specific manner, there is some information known about the magnitude function.

A case in point is the knowledge that the original sequence is a real positive space limited sequence. The fourier transform is conjugate symmetric.

i.e. $$X(e^{j\omega}) = X^{*}(e^{-j\omega}) \qquad (3.5.2)$$

hence

$$Re[X(e^{j\omega})] = -Re[X(e^{-j\omega})] \qquad (3.5.3)$$

and $$Im[X(e^{j\omega})] = -Im[X(e^{-j\omega})]$$

hence the real part of the Fourier transform is an even function and the imaginary part is an odd function.

Similarly, expressed in polar co-ordinates

$X(e^{j\omega}) = |X(e^{j\omega})|e^{j\ arg[X]}$; $iX(e^{-j\omega}) = |Xe^{(-j\omega)}|e-jarg[X]$

consequently

for a real sequence $x(n)$, the magnitude of the Fourier transform is an even function of ω and the phase $arg[X(e^{j\omega})]$ is an odd function of ω.

The above argument suggests the projection of the space domain constraint that a signal is real into a Fourier domain constraint on the phase of magnitude functions. i.e. This is a constraint on the magnitude and phase that has hitherto not been used explicitly in reconstruction, but it means that some bad guesses such as an odd function for magnitude can be ruled out.

A commonly used method of obtaining a reasonable first estimate for the magnitude function is to take an ensemble average of the magnitudes of several images of the same class as the image we are trying to reconstruct. This has the disadvantage of requiring that we know the image class and of course also raises the question of how to define the image classes.

A different method which does not require this type of knowledge simply uses the mathematical function that appears to best approximate the magnitude function. It is well known that the ordered magnitude function is 'low pass' in nature, having the larger values at the low frequencies and smaller values at the higher frequencies. This is true for most images and is used a lot in transform coding schemes where the low frequency co-efficients are quantized using more bits than the high frequency co-efficients. Consequently a simple low pass function will probably give better

magnitude representation than either unity, constant or random magnitude. Some experimentation may still be required to select the parameters of the functions. To illustrate this, some experiments were performed with some magnitude functions.

The figures illustrate the quality of the initial estimates using the amplitude functions discussed previously. This is shown for three test images and considers

(i) unity amplitude . (See Fig. 3.3.3a)

(ii) ensemble average of the amplitudes from three images. (Fig.3.5.1b)

(iii) exponential amplitude function - formed by defining an exponential low pass filter function . (See Fig. 3.5.1d)

(iv) Butterworth amplitude function- formed as before defining a low pass function as:

$$H(\omega_1,\omega_2) = \frac{1}{1 + [D_0 / D(\omega_1,\omega_2)]^2} \qquad (3.5.4)$$

(See Fig. 3.5.1c)

The better quality of the images produced using these mathematical functions raises the whole question of representation of the magnitude function. What functions are acceptable and how do they relate to the nature of the image itself ?

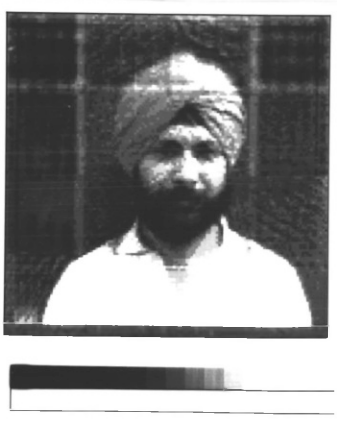
If standard functions can be defined to model the amplitude spectrum for some images classes, then the computational burden may be further reduced. Notwithstanding the difficulty in relating image 'classes' to reconstruction behaviour, results seem to indicate that

(a) Exponential low pas H_o = (10000, D_o = 1.0

(b) Ensemble average

(c) Butterworth lowpass $D_o = 1.0$

(d) Exponential lowpass $D_o = 1.0$



(e) Reconstruction after 30 iterations with adaptive algotithms

Fig. 3.5.1 Phase only images with various starting functions.

only low detail images may be intelligible without phase information - medium-high detail images may be quite intelligible without magnitude information and even better with an exponential Magnitude frequency. The figures show the image after 30 iterations

CHAPTER FOUR

MAGNITUDE ONLY RECONSTRUCTION

CHAPTER 4

MAGNITUDE ONLY RECONSTRUCTION

4.1 Introduction

As outlined in chapter 2, Magnitude Only Reconstruction (MOR) has been studied in the past in order to solve certain problems that arise as result of inability to obtain measurements of the phase function. As a result of the importance of some of these application areas, there has been a concerted research effort to provide solutions to the problem and consequently a large body of knowledge is available.

In general, without any extra information it is not possible to uniquely obtain the phase function from the magnitude function[57]. In many problems however, extra information such as known region of support [37] or known intensity in the space domain [39] is available and in these cases reconstructions are possible.

At the same time, as a result of the wide use of the Fourier transform in electrical engineering and image processing, there is interest in studying the slightly artificial problem of reconstruction from partial Fourier transform information. i.e. other partial representations such as the magnitude and the sign of the phase.

The main motivation of the work outlined in this chapter has been to study the MOR problem with a view to later applying it to estimate the phase in an image restoration problem described in chapter five. This problem is slightly different because the

available magnitude function is only a MMSE estimate of the real intensity and there is a noisy phase function available. This chapter is organised as follows:

Section 4.2 considers the extent to which a 2D function may be defined by its Fourier transform magnitude.

Section 4.3 looks at non-iterative reconstruction methods and the effect of known boundary conditions on these reconstructions.

Section 4.4 presents a detailed study of Gerchberg-Saxon type iterative reconstruction algorithms

Section 4.5 considers the problem when certain portions of the phase or magnitude information are available.

4.2. Magnitude uniqueness

One way to form another sequence $y(n)$ which has the same Fourier magnitude as $x(n)$ is to convolve $x(n)$ with an all-pass sequence $g(n)$

$$y(n) = x(n) * g(n) \tag{4.2.1}$$

where the Fourier magnitude $|G(\omega)| = 1$

Note that

$$|G(\omega)|^2 = G(\omega)\ G^*(\omega) = 1$$

Consequently the autocorrelation $r_g(n)$ of $g(n)$, equals the unit sample function

$$r_g(n) = g(n) * g(-n) = \delta(n) \tag{4.2.2}$$

and its z-transform

$$R_g(z) = G(z)\ G^*(1/z^*) = 1 \tag{4.2.3}$$

from (4.2.3) it follows

$$G^{-1}(z) = G^*(1/z^*)$$

therefore G(z) consists of conjugate reciprocal pole/zero pairs. In addition, G(z) may have no singularities on the unit circle..

If x(n) is a sequence whose z-transform X(z) has no conjugate reciprocal pole/zero pairs and g(n) is an arbitrary all-pass sequence other than a delayed unit sample function, Their convolution y(n) is :

$$y(n) = x(n) * g(n)$$

and

$$Y(z) = X(z)\ G(z)$$

The following statement can be made about X(z) and Y(z)

Y(z) contains conjugate reciprocal pole/zero pairs or the poles or zeros of X(z) are reflected about the unit circle. This gives one way of defining conditions for x(n) and y(n) to be specified by their Fourier transform magnitude.

Theorem

Let x(n) and y(n) be real sequences with z-transforms which have no conjugate reciprocal pole/zero pairs and, in addition,

(a) all the poles or zeros of X(z) and Y(z) (except at z =0 or $z^{-1} = 0$) are either inside or outside the unit circle.

If

$|X(\omega)| = |Y(\omega)|$, then $x(n) = \pm y(n+k)$ for some interger k.

This theorem is satisfied by both minimum phase as well as maximum phase sequences. Since minimum (maximum) phase sequence have additionally no singularities at $z^{-1} = 0 (z=0)$, the magnitude of the Fourier transform uniquely specifies a minimum phase or maximum phase sequence to within a multiplicative sign factor.

Proof

Let x(n) and y(n) satisfy the conditions of the theorem. If the Fourier transform magnitudes of x(n) and y(n) are equal, then their autocorrelations are equal.

Equivalently,

$$X(z)\ X(z^{-1}) = Y(z)\ Y(z^{-1}) \qquad (4.2.4)$$

When all the zeros of X(z) and Y(z) are inside the unit circle, suppose X(z) has a k^{th} order zero at $z = z_o$ where $0 < z_o < 1$. Since X(z) has no conjugate reciprocal pole/zero pairs, then X(z) does not have a pole at $z = 1/z_o$ and

$$R_y(z) = Y(z)\ Y(z^{-1})$$

must have a k^{th} order zero at $z=z_o$. However since $|z_o|^{-1} > 1$ and since Y(z) has no zeros outside the unit circle, then $Y(1/z_o) \neq 0$ and Y(z) must have at least k zeros at $z = z_o$.

Finally since Y(z) has no conjugate reciprocal pole/zero pairs, then Y(z) must have exactly k zero at $z = z_o$. Reversing the roles of X(z) and Y(z), it follows that X(z) and Y(z) have the same zero set for $0 < |z| < \infty$. By a similar argument, the same result holds for the case in which zeros of X(z) and Y(z) are outside the unit circle.

Repeating the argument for poles, it follows that the poles of X(z) and Y(z) are identical for $0 < |z| < \infty$. Thus

$$Y(z) = \beta\ z^{k}\ X(z) \qquad ;\ \beta \text{ is complex, k is integer}$$

Since $|Y(\omega)| = |X(\omega)|$, it follows $|\beta| = 1$, which implies $\beta = \pm 1$ as x(n) and y(n) are real. Therefore $y(n) = \pm x(n + k)$ as desired.

There are other classes of sequences which are uniquely defined by the magnitude of their Fourier transforms. When X(n) and Y(n) are even sequences, with $|X(\omega)| = |Y(\omega)|$,

$$X(z)\ X(z^{-1}) = Y(z)\ Y(z^{-1})$$

Since $x(n)$ and $y(n)$ are even, $X(z) = X(z^{-1})$ and $Y(z)=Y(z^{-1})$
It follows that

$$X^2(z) = Y^2(z) \qquad (4.2.5)$$

therefore $X(z) = \pm Y(z)$ and consequently, $x(n) = \pm y(n)$. Therefore an even sequence is defined to within a sign by the magnitude of its Fourier transform. As another example, suppose $x(n)$ is a real finite length sequence which is zero outside $[0, N-1]$ with $x(0) \neq 0$.

In this case, since $X(z)$ is a polynomial in z^{-1} over the real numbers, it can be shown that if $X(z)$ is irreducible and if $Y(n)$ is any finite length sequence with $|Y(\omega)| = |X(\omega)|$, then either $y(n) = \pm x(n)$ or $y(n) = \pm x(-n)$. However, due to the fundamental theorem of Algebra, no polynomial of degree greater than two is irreducible over the real numbers. This constrains $x(n)$ to be of length three or less. These constraints do not encompass a very large or useful class of 1-D sequences.

In fact in the case of the finite length signal on $[0,N]$, there are up to 2^{N-1} sequences which have the same magnitude function. The z-transform is simply a polynominal in z^{-1} and $H(z)$ has poles only at $z = 0$. So as many as 2^{N-1} different phase curves can be formed simply by reflecting zeros about the unit circle.

In 1-D the z-transform is represented as a product of prime factors;

$$X(z) = \alpha z^{n_0} \prod_{j=1}^{N} (z - z_j) \qquad (4.2.6)$$

where α is real and n_0 non-negative integers

The autocorrelation polynomial

$$Q(z) = X(z)\ X(z^{-1})$$

$$= \prod_{j=1}^{N} (z - z_i)\ (z - z_j^{-1}) \qquad (4.2.7)$$

If all $|z_i| \neq 1$, there are up to 2^N solutions.

If all $|z_i| \neq 1$, the solution is unique because among the roots of $p(z)$ there must be z_i and $z_j^{-1} = z_j^*$. If there are r roots with $|z_i| = 1$ the problem has up to 2^{N-1} solutions. In 2-D there exist polynomials that cannot be factored (prime polynomials). Most 2-D image functions can usually, but not invariably be represented as prime polynomials[24].

The first treatment of the 2-D uniqueness question appears to have been by Bruck and Sodin [23]. They postulated that the uniqueness of a 2-D sequence will finite support is related to the irreducibility of its z-transform. A slightly more general result has been obtained by Hayes [30].

If the z-transform of an image corresponds to a prime polynomial then given the autocorrelation function, one can only construct two solutions differing by a 180° rotation.

The autocorrelation polynomial is defined as before (4.2.7)

$$Q(z_1,z_2) = z^{-1}\ \alpha^2 \prod_{k=1}^{P} X_k(z_1,z_2)\ X_k(z_1^{-1},\ z_2^{-1}) \qquad (4.2.8)$$

$X_k(z_1,z_2)$ are non trivial irreducible polynomials in z^{-1}

$Q(z_1,z_2)$ and $|X(\omega_1,\omega_2)|$ contain the same information about $x(n_1,n_2)$ as they are uniquely derivable from each other, and so ability to recover $x(n_1,n_2)$ from $|X(\omega_1,\omega_2)|$ is equivalent to recovery of $X(z_1,z_2)$ from $Q(z_1,z_2)$. Clearly such recovery cannot be achieved unambiguously. The sign of α and the linear phase terms cannot be determined and it is impossible to establish whether $X_k(z_1,z_2)$ is a factor of $X(z_1,z_2)$.

This ambiguity is an extension of the fact that another finite duration sequence with the same Fourier magnitude can be generated simply by reflecting a zero of $X(z_1,z_2)$ about the unit polydisk.

Certain information is irretrievable when the phase is absent. The closest equivalence that can be defined in the absence of phase is $y(n_1,n_2) = \pm\, x(n_1 \pm k_1,\ n_2 \pm k_2)$. This provides an equivalence class that is related to within a delay, a sign and time reversal of the original sequence.

Bates [56] introduces the concept of the 'form' of an image which is similar to the above equivalence class. Sequences in this equivalence class share the same Fourier magnitude, but there exist sequences outside the class that also have the same magnitude.

Using the knowledge that the only way to generate another sequence with the same Fourier magnitude is to convolve it with an all pass sequence, conditions have been obtained [30] under which only one 'equivalence class' exists for a given Fourier magnitude.

To generate another sequence outside the given equivalence class, with the same Fourier magnitude, it is necessary to replace one or more non-trivial factors $X_k(z_1,z_2)$ of $X(z_1,z_2)$ with $X_k(z_1^{-1},\ z_2^{-1})$. If the factor is symmetric however, this only changes the sign.

Consequently, it follows that the number of equivalence classes with magnitude $|X(\omega_1,\omega_2)|$ is most $2^{(p-1)}$ where p is the number of non-symmetric irreducible factors in $X(z_1,z_2)$.

Therefore to leave only one equivalence class with the given Fourier magnitude, $X(z_1,z_2)$ must have at most one irreducible non-symmetric factor.

i.e.

$$X(z_1,z_2) = P(z_1,z_2) \prod_{k=1}^{P} X_k(z_1, z_2) \qquad (4.2.9)$$

$P(z_1,z_2)$ is irreducible and non symmetric

$X_k(z_1, z_2)$ is irreducible and symmetric

Note that this uniqueness is only for an equivalence class and not for a particular sequence.

4.3 Reconstruction algorithms

A considerable number of papers on the phase problems have proposed schemes for phase recovery from magnitude based on the analytical properties fields. One possible method is to use the Hilbert transform relations and the locations of the complex zeros of the magnitude [58]. Other approaches include apodisation[103] etc. algorithm. These methods however, have not proved very practical for complicated two dimensional images.

One of the reasons why the magnitude only reconstruction problem is complicated, is the non-linear relationship between the image sequence and its Fourier transform magnitude. Unlike the phase, we can not obtain a linear closed form relationship such as (3.3.8).

The Fourier transform magnitude of an image sequence $x(n,m)$ may always be used to obtain the autocorrelation $r(n,m)$,

$$r(n,m) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left| F(\omega_1, \omega_2) \right|^2 e^{-j(\omega_1 n + \omega_2 m)} \qquad (4.3.1)$$

One possible solution to the phase retrieval problem could thus consists of solving these non-linear equations for $x(n,m)$. However such a solution would be very complicated for large numbers of equations and unknowns.

4.3.1.2 Effect of known boundary values:

If boundary values of the sequence are known, the non-linear equations (4.3.2) can be replaced by linear equations. These will be easier to solve than the non-linear system.

Consider the DFT of $X(\omega_1, \omega_2)$ of the pixel array $x(n,m)$

$$X(\omega_1, \omega_2) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x(n,m) \exp \{-j(\frac{\omega_1 n}{N}, \frac{\omega_2 m}{M})\}$$

The autocorrelation function $r(n,m)$

$$r(n,m) = x(n,m) \; ** \; x(-n,-m)$$

** is the 2-D convolution operator

$$r(n,m) = \sum_{k=1-N}^{N-1} \sum_{\ell=1-M}^{M-1} x(n,m)\, x(n+k,\, m+\ell) \qquad (4.3.2)$$

$$n = 0, 1, \ldots\ldots, 2N-1$$

$$m = 0, 1, \ldots\ldots, 2M-1$$

When $x(n,m)$ is a finite duration sequence with support $R(M,N)$, the autocorrelation sequence has support $R(2M-1, 2N-1)$ and is symmetric. The boundary of the autocorrelation function is formed by convolution of the boundary sequences forming $x(n,m)$. This can be seen be considering the autocorrelation function borders.

$$r(n,M-1) = r(-n,1-M) = x(n\ 0) * x(-n,M-1)$$

$$r(N-1,m) = r(1-N,-m) = x(0,m) * x(1-N,-m)$$

Let the m^{th} row of the array $x(n,m)$ be denoted by the vector $x_m(n)$ for $n=1, 2, \ldots\ldots N$. Equivalently written as x_m.

Let the m^{th} row of the array $r(n,m)$ be denoted by $r_m(n)$ for $n=1, 2,\ldots,N$ and equivalently written as r_m.

Observe from (4.3.2) that the M-2 row is formed as

$$x_o(-n) * x_{M-2}(n) + x_{M-1}(n) * x_1(-n) = r_{M-2}(n) \qquad (4.3.3)$$

for $n=1,..,N$

If the vectors of boundary values $\{x_o(-n),\ x_{M-1}(n)\}$ are known, (4.3.3) represents a set of 2N-1 linear equations in the variables $x_1(n)$ and $x_{M-2}(n)$.

Suppose now, that the first (k-1) rows and the last (k-1) rows are known, we write as before

$$r_{M-k}(n) = x_o(-n) * x_{M-k}(n) + \sum_{\ell=1}^{k-z} x_\ell(-n) * x_{M-k+\ell}(n) \qquad (4.3.4)$$

which may be written as

$$x_o(-n) * x_{M-k}(n) + x_{M-1}(n) * x_{k-1}(-n) = r_{M-k}(n) \qquad (4.3.5)$$

where

$$r_{M-k}(n) = r_{M-k}(n) - \sum_{\ell=1}^{k=2} x_{\ell}(-n) * x_{M-k+\ell}(n)$$

(4.3.3) shows that the M-2 row of r(n,m) can be calculated from the boundary values in the vectors x and $x_o(n)$ and $x_{M-1}(n)$, $x_m(N-1)$ and $x_m(o)$, and it can be solved for $x_1(n)$ and $x_{M-2}(n)$.

The rows x_{M-k} and x_{k-1} can be calculated recursively from the values of x_{M-1} and x_{ℓ} for $\ell = 1,2,.....k-2$.

4.4 Iterative algorithms

In this section we look at the basis of the iterative reconstruction methods that are applicable to this problem. The Gerchberg-Saxton (GS) algorithm discussed in the previous chapter for the phase only reconstruction problems was first proposed to solve the two intensity phase retrieval problem [39].

The basic GS algorithm for magnitude only reconstruction is as follows:

1° Make an initial guess $\phi_o(\omega_1,\omega_2)$ of the unknown function $\phi(\omega_1,\omega_2)$. Form the next estimate of the DFT by combining the phase guess with the known magnitude function

$$X_1(\omega_1,\omega_2) = |X(\omega_1,\omega_2)| \exp\{j\phi_o(\omega_1,\omega_2)\} \qquad (4.4.1)$$

compute the inverse DFT

2° Apply known space domain constraints to form the current space domain estimate $x_p(n_1,n_2)$

$$x_p(n_1,n_2) = \begin{cases} x_p(n_1,n_2) & \text{within } R(N_1,N_2) \\ \\ 0 & \text{elsewhere} \end{cases} \qquad (4.4.2)$$

and the non-negative constraint,

$$x_p(n_1,n_2) = \begin{cases} x_p(n_1,n_2) & \text{for } x_p(n_1,n_2) > 0 \\ 0 & \text{for } x_p < 0 \end{cases} \qquad (4.4.3)$$

compute the DFT

3° Use the phase of this DFT as the next estimate $\phi_p(\omega_1,\omega_2)$ of the phase function and form the new DFT estimate as

$$X_{p+1}(\omega_1,\omega_2) = |X(\omega_1,\omega_2)| \exp\{j\phi_p(\omega_1,\omega_2)\} \qquad (4.4.4)$$

compute the DFT, Go to 2.

$x_p(n_1,n_2)$ is the p^{th} iterate of the image being sought. It is expected to have support $R(N_1,N_2)$ - i.e. non-zero on an $N_1 x N_2$ grid. The DFT's calculated are $(2N_1 x 2N_2)$ DFT's.

While this basic algorithm was found to give reasonable but slow reconstruction in the phase only reconstruction it is far less successful in the case of magnitude only reconstruction. The error behaviour of this algorithm for our magnitude only reconstruction is shown fig. 4.1 and is in agreement with the results reported by many other workers [37, 39].

The algorithm's convergence characteristics was observed using

(1) only the non-negativity constraint

(2) only the finite support constraint

(3) both constraints

and the error curves corresponding to these three cases are also shown.

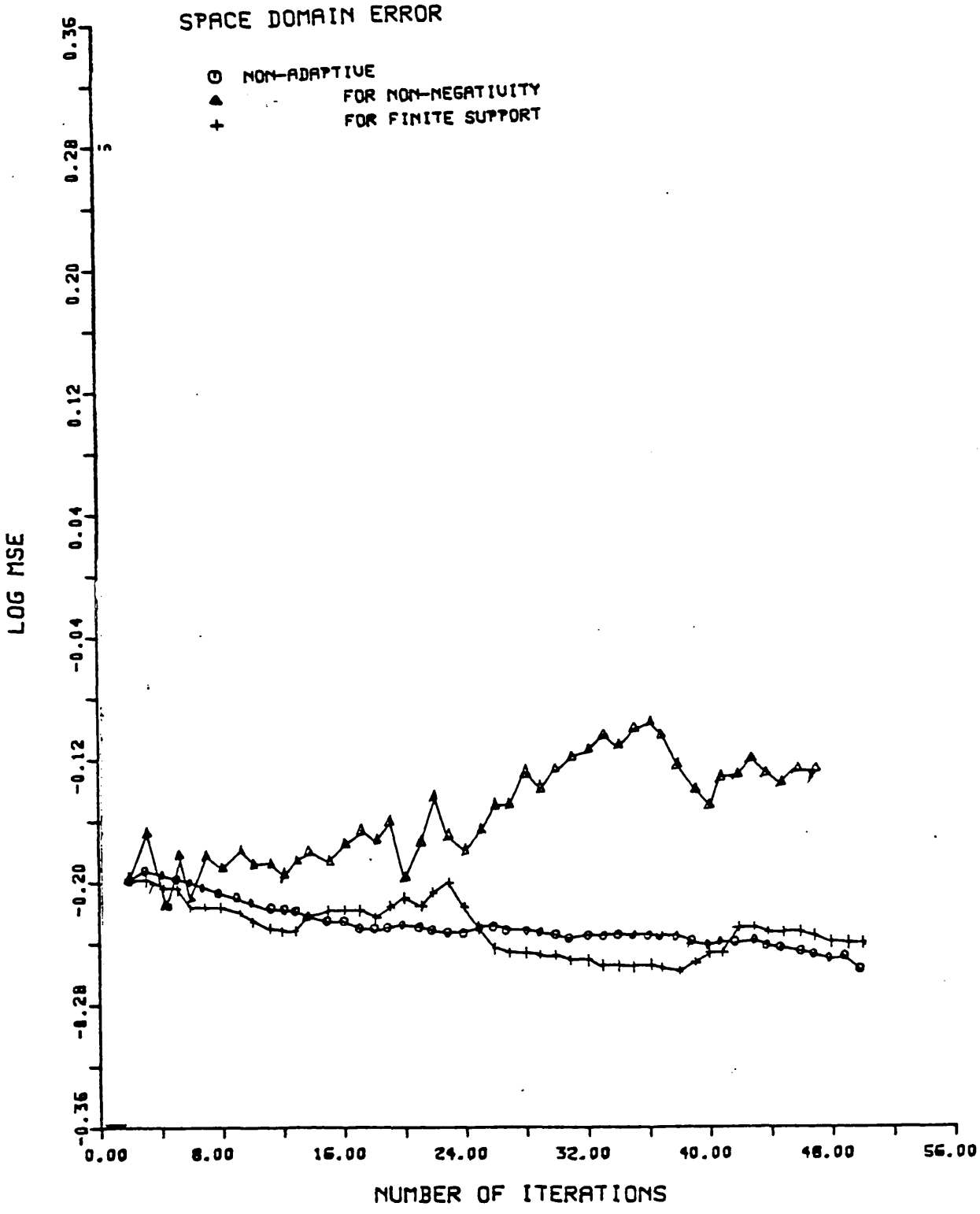


Fig. 4.1 Non adaptive magnitude only reconstruction algorithm

4.3.2.1 Adaptive acceleration

As in the previous chapter, we can model the iteration as a mapping T, and form the next estimate of the space domain signal as

$$x_{p+1}(n_1,n_2) = x_p(n_1,n_2) + \lambda_p[T(x_p(n_1,n_2)) - x_p(n_1,n_2)] \qquad (4.4.5)$$

As before the question arises on the optimum choice of the relaxation parameter λ_p.

Furthermore, (4.4.5) represents a whole class of algorithms that may be formed both by changing the way λ_p is formed and by changing the way that the next estimate is formed.

The difficulty of choosing λ_p to give maximum convergence is a key barrier to the development of fast adaptive algorithms. Choosing λ_p to minimise the euclidean norm of the vector of points falling outside R(n) has proved to be a good criterion. However it is suboptimum because while it takes into account the known R(N), it fails to take into account any other information that may be available a priori. e.g. rough location of known or guessed objects in the image.

If some information is available about the rough location of say edges, an adaptive parameter can be chosen so as to minimise the possible introduction of non-existent objects. Such an application where this information is available is in a Wiener restored image, where the main edges are present though they may not be clearly defined.

The adaptively accelerated algorithm was applied to magnitude only reconstruction for three choices of λ_p

(i) using non-negativity

Here λ_p is chosen to mininise the euclidean norm of non-zero points outside the region of support as in 3.3.

(ii) using finite support

λ_p is chosen to minimise negative points as above

(iii) using an edge criterion

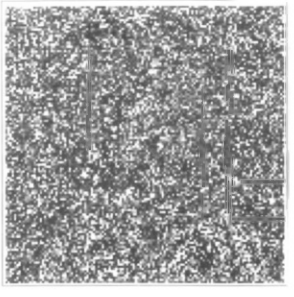
λ_p is chosen to minimise gradients exceeding a set threshold in an area designated as being non-edge.

and the convergence characteristics are shown in fig. 4.2.

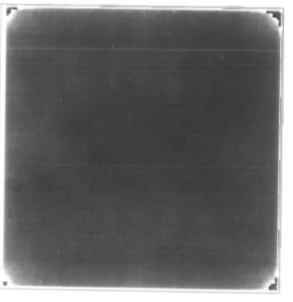
While it is difficult to compare the relative importance of these constraints on the basis of few iterations taken, they seem to bear out the view put by Fienup [105] that this basic algorithm eventually converges after thousands of iterations if the non-negativity constraint is used. They also indicate why experiments which are based on the finite support constraint are not generally successful even after 1000 iterations. The magnitude of reconstruction problem is very ill-conditioned and there are many local minima so that unless a very good starting function is obtained, it is quite easy for the algorithm to converge to a local minimum.

The shape of the error Vs iteration graph shows that the error decreases at first, then slows down and remains relatively unchanged for many iterations before falling. This suggests two ways in which the basic algorithm may be made to converge faster.

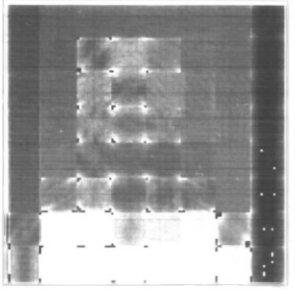
The first method involves overestimation of the region of support. The problems being considered in this thesis differ from many real problems in that the region of support is known exactly.



(a) Finite support



(b) Non-negativity.



(c) Spatial constraints.

Fig. 4.2 Reconstructed images after 100 iterations of adaptive MOR algorithms with various constraints.

It is observed that if the region of support of the sequence is assumed to be greater than $R(N_1,N_2)$ and the finite support constraint is applied

$$X_p(n_1,n_2) = \begin{cases} x_p(n_1,n_2) & 0<n_1<M_1 \quad 0<n_2<M_2 \\ 0 & \text{otherwise} \end{cases} \qquad (4.4.6)$$

where $M_1>N_1$, $M_2>N_2$

there is a very rapid decrease in the error components falling outside $R(M_1,M_2)$, while very little change takes place within $R(M_1,M_2)$.

Instead of applying the usual region of support constraint to $R(N)$, we apply it to $R(M)$ where $M>N$. As more iterations are performed, $R(M)$ is progressively brought closer to $R(N)$. This gives faster overall convergence to zero when an adaptive relaxation parameter is chosen on the basis of minimising the components outside the region of support.

The second method involves re-initialisation of the iteration. Since the convergence of the iteration decreases afer a few iterations, due to local minima that force it to stagnate, faster convergence may be obtained if an adaptive parameter is calculated over several iterations.

So far, we have considered the sequence before each iteration as the latest estimate of the sequence we seek - consequently we have made it satisfy space domain constraints.

The formulation, in the previous section, of the iteration as a mapping leads to a selection of options on how the next estimate is to be formed. For example is it best to replace the points outside $R(N)$ with zero or with some other value ?

The input - output concept is a generalisation of these types of algorithms. The block diagram is shown in the fig. 4.3 below

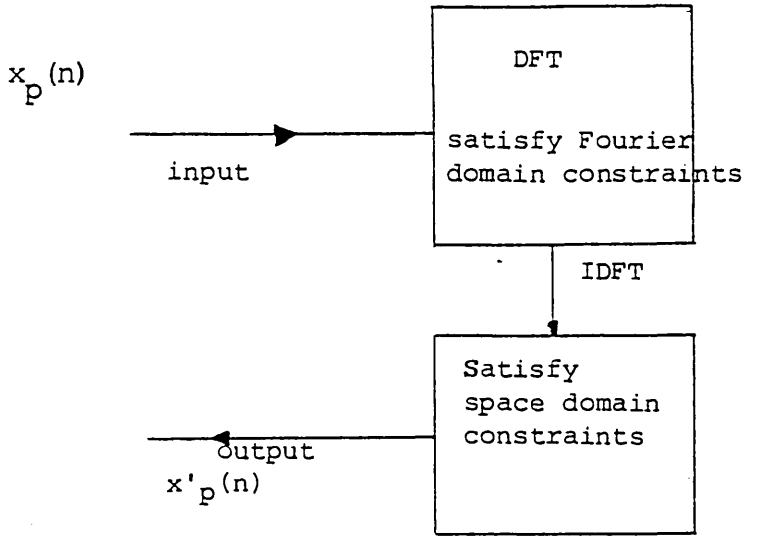


Fig. 4.3 input - output concept.

The input is seen as a driving force for the output and not necessarily as the previous estimate. λ_p is chosen to act on $x_p(n)$ so as to move $x_{p+1}(n)$ closer to satisfying the constraints. $x_p(n)$ does not have to satisfy space domain constraints, which allows greater flexibility in choosing λ_p.

The input of the iteration can be formed in a number of ways, one of which is found to be identical to the basic GS algorithm(4.4.8) for $\beta = 1$

Algorithm 1

$$x_{p+1}(n) = x_p(n) + \beta\delta_p(n) \qquad (4.4.7)$$

$$= \begin{cases} x_p(n) & \text{in } R(N) \\ x_p(n) - \beta x'_p(n) & \text{outside } R(N) \end{cases}$$

where

$x_{p+1}(n)$ is the next input to be used

$x'_p(n) = T(x_p(n)), \qquad r_p(n) = x'_p(n) - x_p(n)$

$x_p(n)$ is the previous input

$$\delta_p(n) = \begin{cases} 0 & \text{in } R \\ -x'_p(n) & \text{outside } R \end{cases}$$

This is the basic input output algorithm. An input $x_p(n)$ is chosen in some way and put through the iteration to produce an output $x'_p(n)$.

Generally $x'_p(n)$ will not satisfy the finite support constraints, so it is now necessary to go back and form a new input $x_{p+1}(n)$ from the old input, but adjusting it to ensure that the next output will have a smaller magnitude of components outside R(N). The starting input may be chosen completely freely and need not satisfy the space domain constraints.

The adaptive parameter β is chosen on the basis that a small change in the input results in a change of the output in the same general direction. So since a change is required outside the region of support to move the components to zero, such a change must be made outside the region of support of the input sequence. The non-negativity constraint is applied at the same time to all the algorithms. The proofs of convergence of these algorithms are given in appendix [III].

Algorithm 2

$$x_{p+1}(n) = x'_p(n) + \beta\delta_p(n) \qquad (4.4.8)$$

$$= \begin{cases} x'_p(n) & \text{in } R(N) \\ x'_p(n) - \beta x'_p(n) & \text{outside } R(N) \end{cases}$$

The next input is formed from the present output in the regions of support and the output is modified outside the region of support.

Algorithm 3

Alternatively the bottom lines of (4.4.7) and (4.4.8) may be combined to give yet another algorithm.

$$x_{p+1}(n) = \begin{cases} x'_p(n) & \text{in } R(N) \\ x_p(n) - \beta x'_p(n) & \text{outside } R(N) \end{cases} \qquad (4.4.9)$$

Algorithm 4

$$x_{p+1}(n) = \begin{cases} x_p(n) + \lambda r_p(n) & \text{in } R(N) \\ x'_p(n) - \beta x'_p(n) & \text{outside } R(N) \end{cases} \qquad (4.4.10)$$

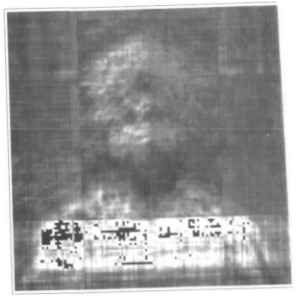
this algorithm is suggested by the adaptive acceleration discussed earlier and the three previous algorithms.

λ is chosen as before, and is used to modify the components within the region of support while β is also chosen so as to push the components outside the region of support to zero. The convergence characteristics of these algorithms are shown in fig. 4.4 their convergence is better than that of the basic algorithm.

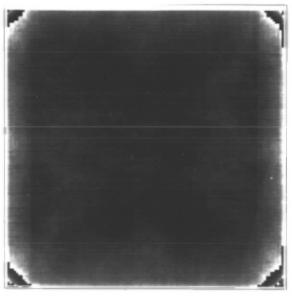
4.5 Reconstruction given some phase information

In this section the problem is widened to one of reconstruction from partial Fourier domain information. It is observed that the magnitude only reconstruction problem is more tractable when some extra information about the phase is provided.

For example an image formed from the magnitude function and a phase that is either 0 or π, with the correct sign of the phase is

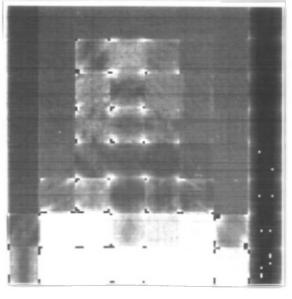


(a) Algorithm 1



(b) Algorithm 2

(c) Algorithm 3



(d) Algorithm 4

Fig. 4.4 Reconstructed images after 100 iterations of various Magnitude only reconstruction algorithms.

intelligible. It is much closer to the final image and the iteration converges much faster giving a very good image after a few iterations. Two cases are examined. In one the correct magnitude function and the sign of the phase are available while in the other only the sign of the phase is available.

(1) Sign and Magnitude

In the Fourier domain, the DFT of $x(n_1,n_2)$ is formed as

$$X(\omega_1,\omega_2) = \begin{cases} |X(\omega_1,\omega_2)| \exp[j\,\theta(\omega_1,\omega_2)] & \text{for } S(\omega_1,\omega_2) = +1 \\ |X(\omega_1,\omega_2)| \exp[j\,2\alpha - \theta(\omega_1,\omega_2)] & \text{for } S(\omega_1,\omega_2) = -1 \end{cases} \tag{4.5.1}$$

where $S(\omega_1,\omega_2) = \begin{cases} +1 & \text{for } \alpha - \pi \leq \phi(\omega_1,\omega_2) \leq \alpha \\ -1 & \text{otherwise} \end{cases}$

$\phi(\omega_1,\omega_2)$ is the correct phase of the DFT and takes on values from $-\pi$ to π.

α is a parameter that is used to divide up the phase function so that the 'sign' can be set for any chosen α threshold. For example when $\alpha = \pi/2$, then $S(\omega_1,\omega_2)$ corresponds to the algebraic sign of the real part of $X(\omega_1,\omega_2)$. α takes on values between 0 and π.

$\theta(\omega_1,\omega_2)$ is the phase estimate that is used in the absence of the correct phase $\phi(\omega_1,\omega_2)$. The first estimate is formed as

$$X_o(\omega_1,\omega_2) = |X(\omega_1,\omega_2)| \exp j\,\theta_o(\omega_1,\omega_2) \tag{4.5.2}$$

where $\theta_o(\omega_1,\omega_2) = \begin{cases} 0 & \text{for } S(\omega_1,\omega_2) = +1 \\ \pi & \text{for } S(\omega_1,\omega_2) = -1 \end{cases}$

This image is quite intelligible and an iterative reconstruction converges quite quickly to give a good reconstruction

(II) Sign of the real part

In the previous case, we had both the magnitude and sign functions. However, it is evident that a sign only image is also intelligible.

$$S(\omega_1,\omega_2) = \begin{cases} +1 & -\pi/2 < \phi(\omega_1,\omega_2) < \pi/2 \\ -1 & \text{otherwise} \end{cases} \qquad (4.5.3)$$

An initial estimate of the DFT is formed as

$$X(\omega_1,\omega_2) = 1.\ \exp\{j\theta(\omega_1,\omega_2)\} \qquad (4.5.3)$$

where

$$\theta_o(\omega_1,\omega_2) = \begin{cases} 0 & S(\omega_1,\omega_2) = +1 \\ \pi & S(\omega_1,\omega_2) = -1 \end{cases}$$

The image formed by taking the inverse DFT of $X(\omega_1,\omega_2)$ is shown in fig. 4.5 after reconstruction.

4.6 Summary

Reconstruction of an image from the magnitude of its Fourier spectrum is generally not possible unless extra information is available. The available information depends very much on the application area, but ofter non-negativity and finite support can be assumed. If the support is known and non-negativity assumed, an iterative algorithm that was both these constraints can be used, but convergence is slow. Modifications of the basic algorithm, using various starting images which do not necessarily satisfy space domain constraints, can be faster but the rate is still much lower than the phase only iteration.

Fig. 4.5 Image reconstructed from sign information.

When some extra phase information is present, the iteration converges very quickly, giving good reconstruction.

CHAPTER FIVE

SOME FREQUENCY DOMAIN APPLICATIONS

CHAPTER 5

SOME FREQUENCY DOMAIN APPLICATIONS.

5.1 General Introduction

This chapter considers some type 2 applications of image reconstruction. It is divided into two parts. In part I results developed in chapter four are used to improve the restoration of noisy linearly degraded images by estimating the phase as well as the magnitude function.

In part II, an investigation is made into the viability of phase only coding to reduce the bit rate of the more usual magnitude/phase coding.

PART I : RESTORATION OF NOISY LINEARLY DEGRADED IMAGES.

5.2 Introduction

Both blind deconvolution and image restoration may be improved by using image reconstruction techniques to estimate the phase of the image, together with the magnitude estimation that is conventionally performed. In blind deconvolution, the signal of interest has been degraded by a blurring function about which complete knowledge is not available.

In some special cases, the distorting signal may be known to have a phase function that is approximately zero and consequently the phase of the degraded image is very similar to that of the original

image. In these cases the problem becomes one of reconstruction from phase only e.g. wavelet reconstruction in seismic signal processing.

A problem that may be approached in a similar way, is the restoration of images blurred by linear convolutional systems with additive noise. Conventional restoration techniques include inverse, Wiener, Power Spectral Estimation (PSE) and Spectral Subtraction (SS) filters [71],[72].

When viewed in the frequency domain, it is observed as shown in section 5.2.1 that whilst the Fourier Magnitude Spectrum (FMS) is a modified or restored version of the degraded image's FMS, the Fourier Phase Spectrum (FPS) of the filtered image is essentially unrestored and effectively identical to the FPS of the degraded image. In this sense, these filters may be said to be 'phaseless'.

In this part of the chapter a method is presented to reconstruct the FPS as well as the FMS of an image and so enhance the visual quality of the restored image. In the next section it is shown that a combined FMP-FPS restoration can be realised by cascading a conventional statistical filter which serves to estimate the FMS, with a generalised deterministic Gerchberg-Saxton iterative algorithm which serves to estimate the FPS. The experimental cascade system is described in section 5.2.3 and results discussed in section 5.2.4.

5.2.1 Statistical Image Restoration Filters

Linear convolutional blurring systems with additive noise can be described by the discrete formulation

$$[g] = [h]**[f] + [n] \qquad (5.2.1)$$

Where [g], [f] and [n] are NxN matrices representing the sampled output degraded images, input image and noise fields respectively. [h] is a dimensional matrix representation of the 2-D blurring system. In the Fourier domain,

$$[G] = [H] \times [F] + [N] \tag{5.2.2}$$

where [G[, [H] and [N] are the corresponding DFT'S of [g], [f] and [n]. [H] is a matrix arising from the "diagnoalisation" of [h] in the Fourier domain. The X and / operators define elementwise multiplication and division of two matrices (or vectors) of identical dimensions. The image and noise processes are modelled as zero mean stationary random fields with power spectra $[P_f]$ and $[P_N]$.

Then the inverse and Wiener restorations are defined as:

$$\text{inverse:} \quad [F] = [G]/[H] = [L_i] \times [G] \tag{5.2.3}$$

$$\text{Wiener:} \quad [F] = \frac{[P_f] \times [H]^*}{[H] \times [H]^* \times [P_f] + [P_N]} [G] \tag{5.2.4}$$

$$= [L_W] \times [G]$$

Similarly, estimator equations can be derived for PSE and SS filters respectively.

Inverse filtering performs poorly in the presence of noise. It is aimed solely at removing the effects of blur and essentially ignores the presence of additive noise. There are many problems with the inverse filter. For example the image blur and formation process may not be invertable and consequently L_i above may not exist. To overcome this problem some methods use pseudo inverses. The other

problems are that the frequency response H usually falls off at high frequencies and if high frequncy noise is present, this may lead to severe noise amplifications.

Finally, the inverse transfer function blows up at the zeros of H which would cause severe difficulties. Improved restoration quality is obtained using Wiener filtering techniques.

Wiener filters explicitly take the presence of noise into account and incorporate a priori statistics of the noise.

In the discrete Wiener filter formulation, it is desired to cause the estimate f as the minimum mean square error (MMSE) estimate.

$$\min_{f} E[(f-\hat{f})^T (f-\hat{f})] \qquad (5.2.5)$$

The transfer function of the resulting filter to give the optimal estimate has been found [100] to be as shown in (5.2.4); and in the limit with no noise it reduces to the inverse filter.

Several variation of the Wiener filter [72] have been proposed. The limitations of the Wiener filter are that it is not particularly well suited to the way the human visual system works, largely because of its reliance on the MMSE criterion. It is overly concerned with noise suppression, and the stationary assumptions that must be made to make the filter computationally feasible make it insensitive to abrupt charges. This is tends to smooth edges and reduce contrast.

Thus the Wiener filter scarifices too much resolution in favour of noise suppression. As mentioned above, a variety of other filters have been proposed which alleviate some of the short comings of hte Wiener filter, [99], [72].

Because of the importance of phase in images, some filters have proposed to take the phase into account for several specific types of PSF [32] but these seem to require large amounts of a priori information.

We shall restrict our analysis to the Wiener filter but the other filters share the same deficiency of not estimating the phase so the same approach would be equally applicable. Another aspect is that many of the restoration approaches are based on homogenous random field models for the images and noise, and on least squares error criteria - both of which are nt optimum or even fair assumptons. Consequently, an image model based on segmentation of the image would be desirable, and breaking the image into phase and magnitude is a form of segmentation.

It is easily shown from the defining equations that the FPS of the degraded image and the FPS of the restored image , are apart from phase shifts introduced by the blurring system, identical. Noisy induced phase degradations are not compensated for in the restoration process.

This does not appear surprising since the restoration filters are derived from statistical estimation criteria defined in terms of energy related quantities such as MMSE, Power Spectra and second order moments. Hence they estimate the 'spectral energy' which is a characterisation related to the FMS.

Except in cases of extreme degradation, degraded images may generally be adequately restored by well designed restoration filters. One is usually able to recognise the main global or gross features and structures of an image so restored.

Restoration quality is necessarily imperfect in the sense that local features such as edges may be obscured or smoothed out or that noise levels are too objectional in non-edge regions. High noise suppression image restoration is in practice unattainable by linear filtering techniques which are subject to the well known resolution-vs-noise dilemma. In the absence of a valid statistical phase estimation technique, a deterministic and iterative algorithm is developed for phase reconstruction.

The spatial masking effect of the human visual system (HVS) [73] is a property where the eye is able to tolerate relatively high noise levels in edge regions of an image but not in flat or smoothly varying regions of an image. In contrast to the rather non-subjective restoration criteria implicit in filtering techniques, such a non-linear restoration process would be more adaptive and responsive to the HVS criteria of image intelligibility.

5.2.2 The Hybrid reconstruction technique

In section 5.2.1 it was seen that conventional filtering provides an optimal estimate of the FMS of the image. Given such an estimate, one could then attempt to estimate, the FPS via a magnitude only GS iterative reconstruction algorithm. This combined FMS-FPS restoration could in principle be realised by cascading a conventional filter which estimates the FMS with an iterative reconstruction algorithm which estimates the FPS. However as noted earlier this case may fail to converge if a good first estimate of the phase function is not available. Convergence may be enforced if further constraints more restrictive in scope than finite support and positivity were imposed.

One way of specifying further constraints is suggested by the interrelationship between edge structure and the FPS of an image. If information pertaining to the spatial edge activity of the image is known, then successive spatial domain estimates, generated by the GS algorithm, could be constrained to conform in edge structure to a priori known edge information, and so indirectly constrain the FPS. Such information can be obtained by an 'edge region detection' on the output of the statistical restoration filter. Information on the rough location of edges is often available from the restored image.

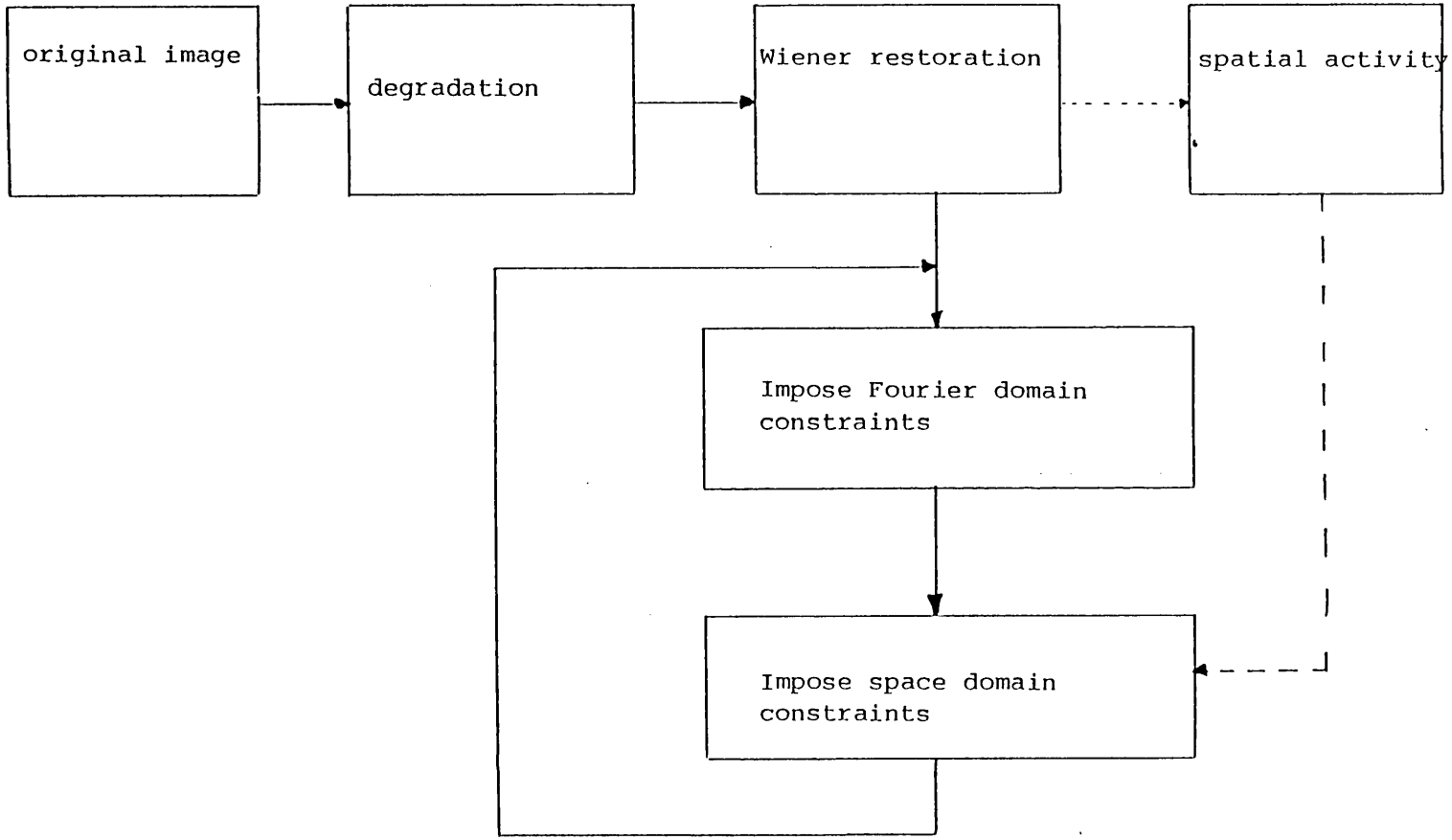
5.2.3 Cascade Algorithms

The block diagram illustrates the algorithm (see fig. 5.2.1). The degraded image is first restored by a Wiener filter, this filtered image's FMS is taken as a good estimate of the true image's FMS. The FPS is then estimated by a modified GS reconstruction algorithm using the FMS estimate as its driving input.

In the Fourier domain, the FMS is constrained to that obtained by the restoration filter above, while the FPS is unchanged. In the space domain, positivity, finite support constraints are applied in addition to a 'spatial activity' constraint.

The spatial activity constraint is determined by edge detecting the filtered image and identifying approximately its 'edge' and 'non-edge' regions. The adaptive spatial activity constraint set then consists of bounding the current iterate's pixel amplitude to lie within an adaptive amplitude window centered on the amplitude of the previous iterate's corresponding pixel. For image pixels known

Fig. 5.2.1 Block diagram of the hybrid reconstruction method



to be located in the non-edge regions the amplitude width is narrow,thus minimising intensity variations in this area. But for pixels located in edge regions, the width is extended to permit the amplitude variability of edge pixels and so preserve edge structure. Noise propagation is controlled as the adaptive spatial constraint tends to restrict noise amplification to mainly the edge regions, where by virtue of the spatial masking effect it is not too subjectively noticeable.

To study the effectiveness of this technique, it was applied to a degraded image and various algorithms were tested.

The original 128x128 'man' image is shown in fig. 5.2.2. In fig. 5.2.3, the degraded 'man' image obtained by blurring with a gaussian point spread function and adding zero mean noise, SNR of 21 dB is shown.

The Wiener restored image is shown in fig. 5.2.4. Note that even though noise levels in the picture are well suppressed, this is at the expense of somewhat poor resolution and edge quality.

The final FMS-FPS restoration obtained by the hybrid cascade system is shown in fig. 5.2.5. The 'blocking effect' present in the restored image arises from the fact that the modified as phase reconstruction algorithm is implemented by initially partitioning the image into smaller blocks so as to minimize the memory requirements.

5.2.4 Further discussion

Three versions of the GS algorithm are implemented.

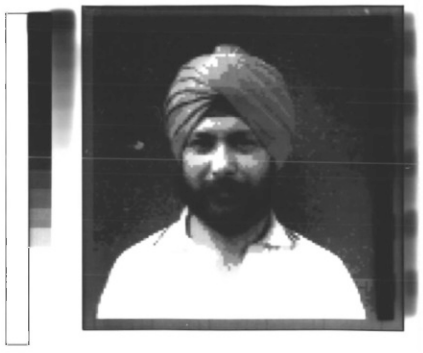
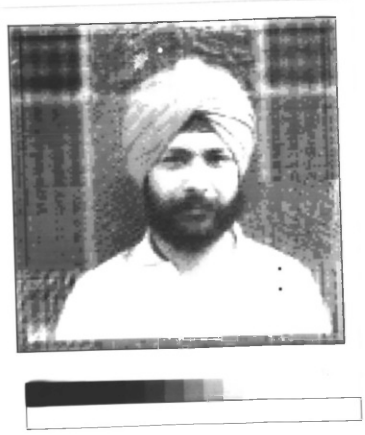


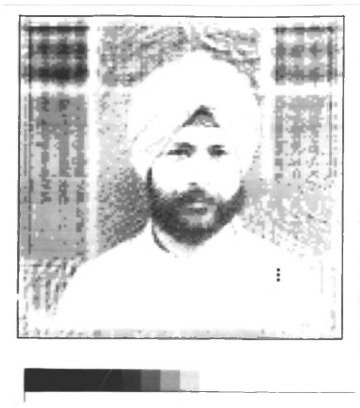
Fig. 5.2.2 Original 'Man' image

Fig. 5.2.3 Degraded image, SNR = 21 dB

Fig. 5.2.4 Wiener restored image



(a) Adaptive GS with both non-negativity and finite support



(b) Spatial activity constraint

Fig. 5.2.5 Reconstructed image after 20 iterations

A. GS with positivity and finite support constraints.

This is the conventional GS algorithm without any acceleration and imposing positivity and finite support constraints.

B. GS with adaptive acceleration.

The next iterate is multiplied by an adaptive factor chosen to minimize the euclidean norm of the vector of image points outside the region of support.

C. GS with spatial activity and reinitialisation.

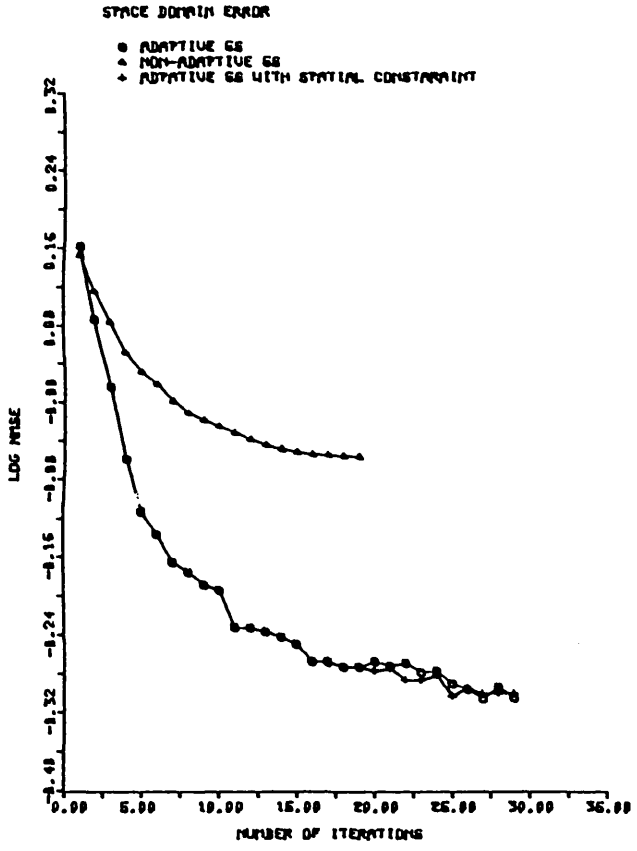
The spatial activity constraint is applied as follows:

$$V_{p+1} = \begin{cases} V_{p+1} & \text{in non-edge region } (s=0) \\ V_p & \text{in non edge region } (s=1) \\ V_{p+1} & \text{in designated edge region} \end{cases} \qquad (5.2.6)$$

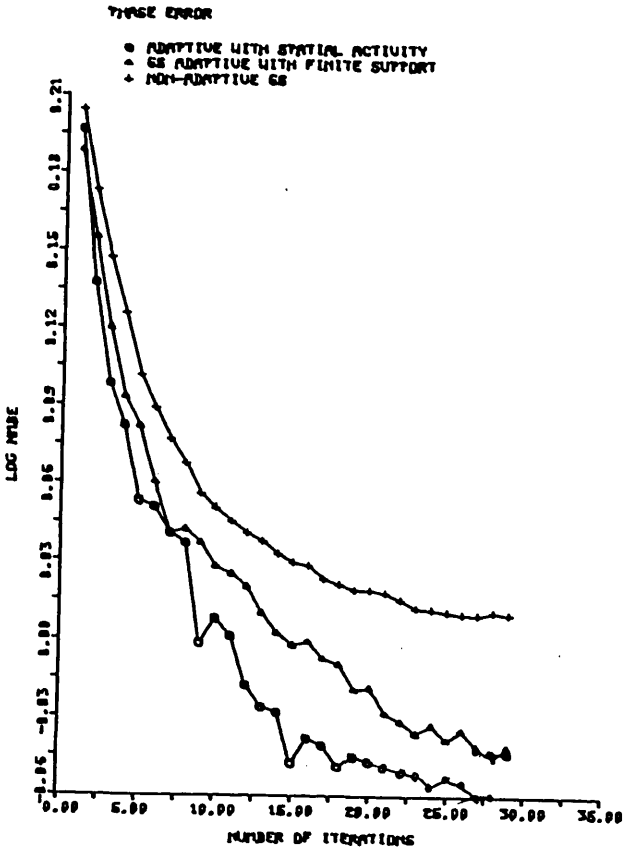
where V_{p+1} is the vector of points at the $(p+1)^{th}$ iteration,
S is the spatial activity and is set to 1 if the next iteration appears to introduce wide variation in the non-edge region.

The convergence of these versions are compared by evaluating the normalised error between the true FPS and the current FPS estimate. The graphs showing these convergence characteristics are given in fig. 5.2.6.

The graphs show that both the phase domain (5.2.6a) and the space domain (5.2.6b) NMSE decreases with the iterations. As before, adaptivity using the finite support constraint leads to a considerable improvement in convergence rate over the non-adaptive algorithm. The spatial activity constraint is seen to reduce the error further , but it is unsteady and may increase the error if it is not chosen carefully. However the visual quality of the reconstructed image using the spatial activity constraint is better.



(a) Space domain



(b) Phase domain

Fig. 5.2.6 Error behaviour of reconstruction algorithms.

PART II: FOURIER TRANSFORM PHASE CODING OF IMAGES

5.3.1 Introduction

In chapter 3 we considered reconstruction from phase only both for the case where the full phase function is present , and for the case where only noisy samples of the phase are available. We also considered several techniques for selecting good starting magnitude functions.

We now consider an application of these results to Fourier transform image coding. Usually both the phase and magnitude are coded and transmitted and at the reciever an image is reconstructed from them. Monochrome images have beencoded at bit rates of 1.0 to 1.5 bits/pixel [19] with mean square error distortion less that 0.5%. Successful coding schemes developed for Fourier transform coding have been found to rely on the assignment of considerably more bits to the phase than the magnitude [61].

For example Pearlman and Gray [62], using rate-distortion theory in the source coding of the DFT, derive performance bounds and encoding guidelines for the direct fixed-rate MMSE data compression of the DFT of a stationary real sequence.

Their technique yields a theoretical measure of the relative importance of phase over magnitude in compression, with the result that phase must be coded with 1.37 bits more than the magnitude for the ergodic stationary source assumed.

Since an image may be reconstructed from the phase of its Fourier transform, we consider coding only the phase and then using phase only image reconstruction to reconstruct the image from the coded phase. The other application of interest here of course is the

Kinoform - a phase only hologram. The intelligibility of its reconstructed image could be greatly enhanced by phase only reconstruction. To avoid the time and computational problems posed by iterative reconstruction, interest is focussed on obtaining an excellent first estimate which may preclude the need for iteration.

In this part of the chapter, we investigate a Fourier transform coding technique where the phase is coded accurately for each block of an image, but not the magnitude. The aim is to improve the compression ratio of adaptive Fourier transform coding methods by sending the phase of all the blocks but only the magnitude of a few and replacing the other blocks with magnitude functions as discussed in chapter three.

In the absence of any specific information we may say that the distribution of magnitude co-efficients follows a general 'low-pass' form with larger values of the low frequencies. This coding method produces acceptable and even comparable images to conventional adaptive Fourier transform coding at lower bit rate. In effect it may be argued that the phase only reconstruction simply suggests another adaptivity criterion.

This part of the chapter is organised as follows. In the next section we look generally at unitary transforms and transform coding. Next we look at the Fourier transform and the need for its efficient coding and investigate the theoretical feasibility of phase only coding.

In section 5.3.4 we develop a method of coding the phase for our application and we describe the coding technique in section 5.3.5

5.3.1 Unitary transforms and transform coding

The forward unitary transform of and $N_1 x N_2$ image array $f(n_1,n_2)$ is an $N_1 x N_2$ transformed array defined by

$$F(m_1,m_2) = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} f(n_1,n_2)\ A(n_1,\ n_2;\ m_1,m_2) \qquad (5.3.1)$$

where $A(n_1,\ n_2;\ m_1,m_2)$ is the forward transform kernel. The inverse transformation from the transform domain to the space domain is defined by

$$f(n_1,n_2) = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} F(m_1,m_2)\ B(n_1,\ n_2;\ m_1,m_2) \qquad (5.3.2)$$

where B is the inverse transform kernel. A and B must satisfy orthonomality conditions [101].

Unitary transforms have found wide application in image processing particularly image coding. They provide a spectral decomposition of an image into co efficients that tend to isolate certain features of an image. For example the first spectral component is proportional to average image brightness and the higher sequency components are measures of the image edge content. The block diagram of a transform coding system is illustrated on the next page. The number of bits used to code a coefficient could be given by the equation below which uses the maximun variance zone.

$$N_b(u,v) = \frac{N_B}{N^2} + 2\log_{10} V_f(u,v) - \frac{2}{N^2} \sum_{a=1}^{N} \sum_{b=1}^{N} \log_{10} V_f(a,b) \qquad (5.3.3)$$

where

$N_b(u,v)$ bits allocated to the coefficient at (u,v),

N_B total number of bits used to code the NxN image.

$V_f(a,b)$ the variance of a transform coeffient.

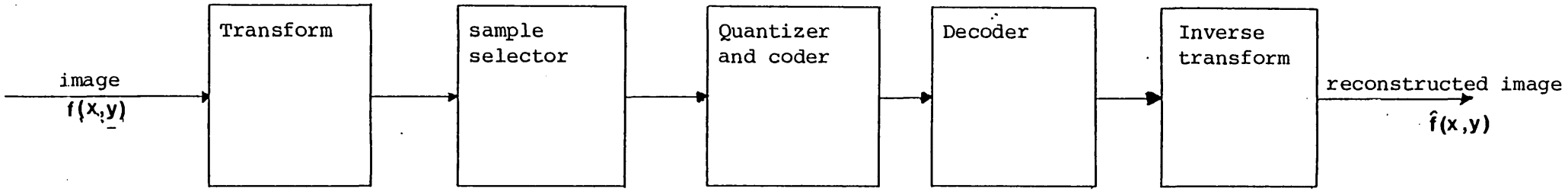


Fig. 5.3.1 Transform coding system.

A transform is performed on the image producing a set of transform co efficients which are quantized and coded for transmission.

Fourier transforms were the first to be used [74,75]. Later Pratt [76] utilised the Hadamard transform with a considerable decrease in computational requirements. Other investigation have since utilised Karhunen-Loeve (77], Haar[78], Sine [102] and Cosine transforms to name some of the most popular ones. The Karhunen-Loeve transform provides minimum mean square error performance but requires statistical knowledge of the source and does not posses a fast computational algorithm. The Haar has a fast algorithm but results in relatively large coding error.

The sine and cosine transforms have fast algorithms and have been shown to approach the efficiency of the Karhunen-Loeve transform for Markov process image data [102,103]. Various comparisons have been made of the performance of these transforms under certain coding criteria.

The basic premise of transform coding is that the transform co efficients have an energy distribution that is more suitable for coding than the image pixels. The co efficients are more decorrelated and the energy in the transform domain tends to be clustered into a relatively small number of samples at the lower sequency.

The samples to be used are selected either using a zonal sampling method or a threshold criterion. The co efficients that are not selected are replaced by zero's.

The most commonly used zonal sampling technique is a bit allocation technique, where the co efficients are allocated a

to (5.3.3). The phase is allowed to take twice the number of levels as the magnitude. This and similar adaptive techniques give better MSE and subjective performance than non-adaptive techniques.

5.3.2 Feasibility of phase only coding

Assuming that an image may be satisfactorily reconstructed from its phase only, we investigate the bit rate requirement in quantising the phase to give the same sort of distortion as coding techniques that code both the phase and magnitude co-efficients. For uniform quantisation the quantisation interval δV is given by

$$\delta V = \frac{\phi_{MAX}}{L} = \frac{\phi_{Max}}{2^B}$$

ϕ_{Max} - expected maximum value of the phase. We expect the phase to vary uniformly between $-\Pi$ and Π

B - no bits/pixel

L - the number of levels used to quantise the phase.

Assuming that the decision thresholds are set half way between the levels (MMSE criterion) the maximum quantization noise will be

$$e = \frac{\delta V}{2}$$

$$e = \frac{\delta V}{2} = \frac{\phi_{MAX}}{2 \times 2^B} = \frac{2\Pi}{2 \times 2^B}$$

$$e = \frac{\Pi}{2^B}$$

$$B = \log_2 (\Pi/e) \text{ bits/pixel} \qquad (5.3.4)$$

To achieve NMSE of 1%, e should be less than $\Pi \times 10^{-2}$

and consequently from (5.3.4) this corresponds to a bit rate requirement of B = 6.6 bits/pixel. Consequently on an NMSE basis, both a low distortion rate and a low bit rate cannot be achieved by trying to code only the phase and reconstructing from the coded phase by using the phase only reconstruction algorithm.

However, coding is feasible if a good first magnitude estimate is obtained. The phase quantization strategy which is adopted here makes no assumption about the nature of the probability distribution of the phase and follows an approach by Pohlig [61].

The number of quantization levels used to code the phase of a given co-efficient is inversely proportional to the frequency associated with that co-efficient. To justify this statement, consider the (NxN) image of f(m,n) with DFT F(i,k). We define a continuous periodic function p(x,y) as

$$p(x,y) = \sum_{i=0}^{N-1} \sum_{i=0}^{N-1} F(i,k) \exp \{ix + ky\} 2\Pi j/N \qquad (5.2.5)$$

The image f(m,n) is thus composed of samples of the periodic function p(x,y) at x = 0,.....N-1, y= 0,.....N-1. This trivially has a Fourier series since p(x,y) above is a weighted sum of harmonic sinusoids.

The phase is quantized with quantization interval $\Delta\phi(i,k)$ since the position of the $(i+k)^{th}$ sinusoid is determined by the phase $\phi(i,k)$ of F(i,k), there will be a corresponding quantization interval in the position of the $(i+k)^{th}$ sinusoid. If we take the approach

that the phase quantization interval be chosen so that the corresponding quantization in the position of the $(i+k)^{th}$ sinusoid is constant with respect to frequency $(i+k)^{th}$, then

$$\frac{\Delta(x,y)}{N/(i+k)} = \frac{\Delta\phi}{2\pi} \quad , \qquad i,k \neq 0.$$

where $N/(i+k)$ is the period of the $(i+k)^{th}$ sinusoid.

The number of levels is thus:

$$L = \frac{2\pi}{\Delta\phi} = \frac{N}{(i+k)\Delta(x,y)}$$ which is inversely proportional to frequency

the number of bits/pixel is

$$b \cong \frac{1}{N} \sum_{k=1}^{N} \sum_{i=1}^{N} \log_2 \left\{ N/\left[(i+k)\ \Delta(x,y)\right] \right\} \qquad (5.3.6)$$

5.2.3 An adaptive Fourier phase coding technique

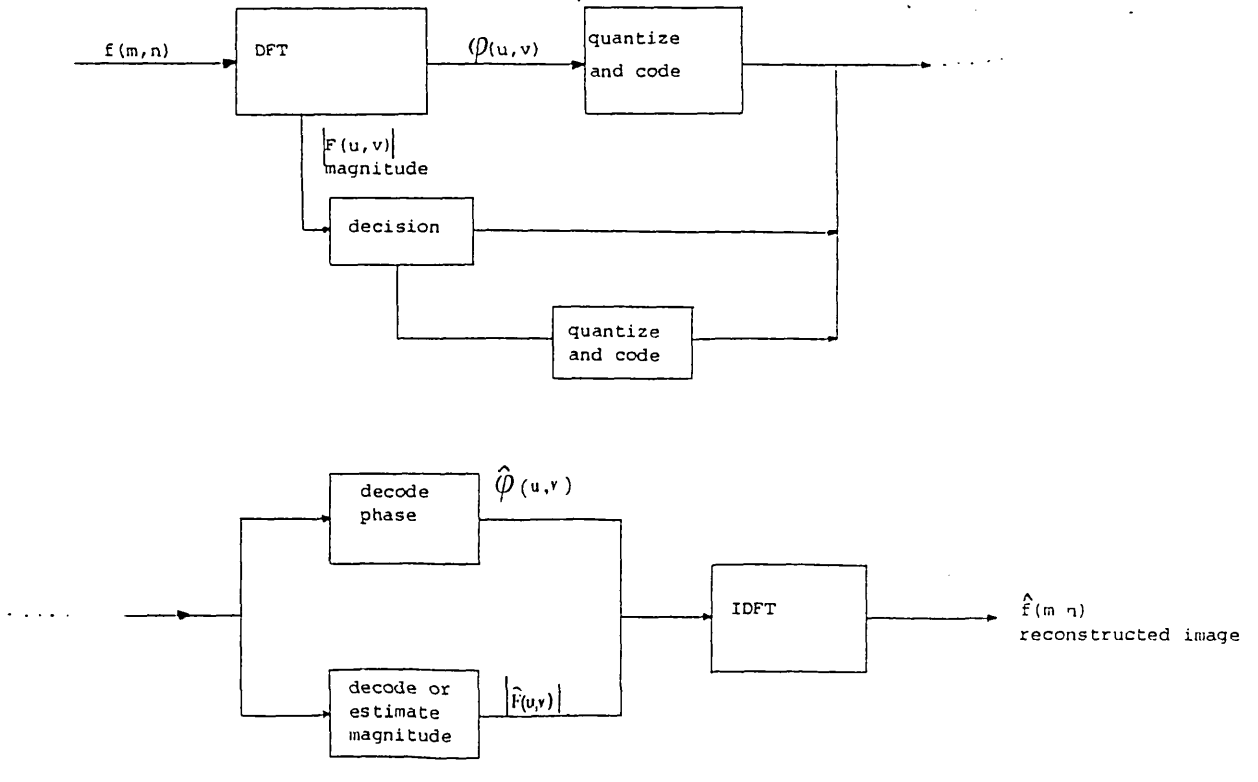


Fig. 5.3.2 The adaptive phase coding scheme.

The NxN image pixel array is subdivided into nxn arrays and the DFT of each is performed independently.

The phase of each block is quantised and coded as explained in the previous section. The magnitude is handled by an adaptive encoder which makes a decision on whether or not a magnitude or array should be sent.

Several techniques are available for selecting the magnitude functions to be used for reconstruction, for example

(i) The magnitude of four neighbouring blocks could be combined to produce an ensemble average to be used for all four. This need only be sent once every four blocks.

(ii) A texture segmentation technique may be used to divide the imageinto several regions. The blocks in each region are used to provide an ensemble average magnitude function used for all blocks in that region. consequently the block ensemble average for each sector is sent, together with a description of the location regions.

(iii An exponential function may generated for the magnitude as described in chapter 3.

At the receiver, the image is reconstructed using the quantized phase and an estimate of the magnitude function. In our experiment the starting amplitude function was selected as an ensemble average and so iterations were performed to recover the magnitude function, using the given phase information as discussed in chapter three. The recovered images are shown for various quantization rates in Fig. 5.3.3.

(a) phase coded at 1.0 bit/pixel, with a magnitude contributions of 0.2 bit/pixel

(b) pnase coded at 0.8 bit/pixel magnitude contribution of 0.2 bit/pixel

(c) Reconstruction after 30 iterations phase at 0.8 bit/pixel, no magnitude sent (ensemble average used)

Fig. 5.3.3 Phase coding.

The results show that acceptable images can be obtained by coding the phase only but to minimise bit-rate it is essential to have a good magnitude first estimate. The method has the major drawback of requiring several iterations to reconstruct the image at the receiver, but the number of iterations may be reduced if a good first estimet can be formed at the receiver.

Consequently coding only the phase does not appear to lead to practical advantages over methods that code both the phase and magnitude. Nevertheless this approach is promising because it seeks to take full advantage of the 'segmentation' properties of the DFT.

CHAPTER SIX
IMAGE RECONSTRUCTION FROM POLYGONAL APPROXIMATIONS

CHAPTER 6

IMAGE RECONSTRUCTION FROM POLYGONAL APPROXIMATIONS

6.1 Introduction

In the previous chapters we have been concerned with reconstructing images given partial specifications in the frequency domain.

In many ways the problem of reconstructing images from partial specification in the space domain is similar. For example phase only images are visually related to edge detected or contour images. Consequently the problem of phase only image reconstruction is related to the problem of reconstruction from contours.

In this chapter we investigate the problem of reconstruction from edges or contours and extend it to reconstruction from polygonal approximations. Data compaction may be achieved by representing an image by its contours for many classes of images [19], and regenerating the images using contours and a texture generation rule [17]. Further compaction may be achieved if such contours can themselves be uniquely represented by polygonal approximations and later reconstructed.

Relatively little work appears on the recovery of contours from polygonal approximations (PA) mainly because most applications see PA's as a last stage in scene recognition rather than an intermediate stage. Some obvious improvements may be obtainable here. For example if a computer vision system is being used to recognise a

scene by obtaining polygonal approximations, better scene descriptors may be obtained if an image synthesised from the polygonal approximations is fed back and compared with the input image.

This chapter considers a simple polygonal approximation algorithm that is particularly suitable for reconstruction of a contour, which is then used in a second stage of texture synthesis to reconstruct a gray scale image.

In general, it is not possible to uniquely specify an image in terms of regions and their properties (e.g. texture) and then recover it without loss of information. However, it is shown here that some recovery within acceptable distortion can be achieved.

The relationship between an image and its contours has been treated by a number of authors especially in the bilevel case.

Pavel[15] defined a framework for looking at skeletons as deformed images and hence the synthesis of images is seen as a restoration or reconstruction process. Grenander [16] formalised a concept of pure patterns where grammars were constructed to generate images.

Schemes have been reported for recovering gray scale images from their contours and texture information e.g. Lemay [17], Delp[14] etc.

While the general problem is quite ambiguous and many results are scene dependent, many useful results have been obtained. The main thrust of this chapter is to demonstrate the feasibility of contour reconstruction, extend it to polygonal approximations and show that it is a possible approach to the problem of image coding at high compression ratios.

The chapter is organised as follows. The next section briefly reviews the background to texture analysis and modelling appropriate to the reconstruction problem. Section 6.3 proposes a suitable texture contour parametrisation scheme and reconstruction algorithm. In section 6.4 polygonal approximations and their relationships to contours are presented and in section 6.5 a simple algorithm for contour recovery is discussed. The final section presents some results of reconstruction from contours and polygonal approximations.

6.2 Image texture analysis

The problem of obtaining reasonable spatial models for images has been considered by many recent workers and has led to some good results in image enhancement and coding. As an example, if the image is a sample of a Markov field then cosine transform coding can be shown to be nearly optimal in the mean square sense [81].

A Gaussian-Markov model of the image is also used by various studies using distortion rate theory to examine coding performance [80].

In this section we look briefly at the different methods available for the representation and analysis of the texture in a region. The premise that texture can be parametrized around contours and that consequently the regions texture can be represented by samples of the texture at the contour is examined.

Several methods have been proposed for texture analysis and measurement. Qualitatively, a region of uniform texture has a characteristic repetitiveness. Hawkins [85] specifies three ingredients upon which the notion of texture appears to depend. These are:

(i) Some local order is repeated over a region which is large in comparison to the orders size,

(ii) The order consists in the non-random arrangement of elementary parts,

(iii) The parts are roughly uniform entities having approximately the same dimensions everywhere within the textured region.

6.2.1 Quantitative texture measures

(a) Fourier Spectrum Textures Analysis

The degree of texture coarseness is proportional to spatial period, hence a region of coarse texture should have its Fourier spectral energy concentrated at low spatial frequencies. At the same time, regions of fine texture should exhibit a concentration of spectral energy at high spatial frequencies. A correspondence does exist to some degree but difficulties often arise because of spatial changes in the period and phase of pattern repetitions.

For example Fourier spectral analysis helps in the detection and classification of coal miners black lung disease [19] which appear as visual textural deviations from the norm, but the same method may fail in analysing aerial photographs due to the considerable spectral overlap of different natural textures such as urban and rural regions.

A 2-D transform characterises the image as a weighted sum of brightness pattern (basis functions). The coefficients of the transform may thus be regarded as an indication of the correlation of a particular basis function with an image field. If the basis pattern is of the same form as a feature in the image when that feature can be known simply by monitoring the appropriate transform co-efficient.

Unfortunately such simple correspondence does not happen much in practice, but many studies have been done to link or examine parts of the Fourier spectra to the image features, e.g. Lendris and Stanley [86].

(b) The Spatial Autocorrelation Function

This is defined as

$$A(\varepsilon,\eta: j,k) = \frac{\sum_{m=j-W}^{j+W}\sum_{n=k-W}^{k+W} f(m,n)\, f(m-\varepsilon, n-\eta)}{\sum_{m=j-W}^{j+W}\sum_{n=k-W}^{k+W} [f(m,n)]^2} \qquad (6.2.1)$$

$$\varepsilon,\eta = 0 \pm 1, \pm 2, \ldots, \pm T$$

A is the autocorrelation function calculated over a $(2W+1)^2$ window at each point of the image $f(j,k)$ for the offset values ε,η.

The expectation here is that coarsely textured regions will exhibit a higher correlation for a fixed shift (ε,η) than finely textured ones so that the spread of the autocorrelation function will reflect the texture coarseness. Consequently a suitable measure of the spread such as the second moment (6.2.2) may be used as a measure of the texture.

$$T(j,k) = \sum_{\varepsilon=-T}^{T}\sum_{\eta=-T}^{T} \varepsilon^2,\eta^2\, A(\varepsilon,\eta : j,k) \qquad (6.2.2)$$

(c) Edge Activity

The number of edge points in a neighbourhood about a point can be

used as a texture measure. This is formed as:

$$T(j,k) = \frac{1}{(2\omega+1)^2} \sum_{m=j-W}^{j+W} \sum_{n=k-W}^{k+W} \varepsilon(m,n) \qquad (6.2.3)$$

over a $(2W+1)\times(2W+1)$ window

where $\varepsilon(m,n)$ is an edge image formed by applying some form of edge detection on the image. W is the neighbourhood window size.

(d) Joint Occurence Matrices

A number of texture measures have been proposed based on the joint amplitude histogram of pairs of geometrically related image points [82].

If the pair of pixels $F(j,k)$ and $F(M,n)$ with intensisties $0 < a, b < L-1$ are separated by r radial units at an angle θ with respect to the horizontal axis, let $P(a,b,i,j,k,r,\theta)$ represent the 2D histogram measurement of the image field over some $(2W+1)\times(2W+1)$ window. The two dimensional histogram can be considered as an estimate of the joint probability distribution.

For each pair of the set $\{j,k,r,\theta\}$ the 2D histogram may be regarded as an LxL array of numbers relating the statistical dependence of pairs of pixels. Such arrays are joint occurrence matrices (also called gray level dependence matrices) and are useful texture measures.

6.2.2 Texture field models

A number of texture field models have been used. These include the

Julesz model [86], linear programming model [88], autoregressive model [87], etc.

As an example, consider the autoregressive model

$$y(i,j) = \sum_{m=0}^{a} \sum_{n=0}^{b} \theta_{mn}\, y(i-m,\ j-n) + u(i,j)$$

θ_{mn} are the regression coefficients.

The image $y(i, j)$ is modelled as a 2-D discrete homogeneous Gaussian - Markov field, where

$E[u(i,j).y((i-m,\ j-n)] = 0$; $0<m<a$, $0<n<b$, $m=n\neq o$ (pixel uncorrelated with pixels in recursion region)

$E[u(i,j)\ u(k,l)] = \sigma^2\ \delta_{ik}\ \delta_{jl}$

$E[u(i,j)] = 0$; $E[.]$ is the expectation operator (zero mean)

$\delta_{ik} = 1$; if $i = k$ and

$= 0$; if $i \neq k$.

For the special case of $a = b = 1$,

$$y(i,j) = \theta_1\, y(i-1,j) + \theta_2\, y(i-1,j-1) + \theta_3\, y(i,j-1) + u(i,j)$$

by row concanteration, (6.2.5)

$y(k) = \theta^T\, Z(k-1) + u(k)$ Where, (6.2.7)

$Z^T(k-1) = y(k,N), y(k-N-1), y(k-1)$ the past history of the process.

and $\theta^T = (\theta_1, \theta_2, \theta_3)$

The mean square error is

$$J_N(\theta) = \sum_{k=1}^{N^2} [y(k) - \theta^T\, z(k-1)]^2 \qquad (6.2.8)$$

$k \neq \{\text{initial condition set}\}$

To form an estimate of the texture, select the regression co-efficients θ and σ^2 (variance of the zero mean gaussian noise μ) so as to minimise the mean square error.

The problem of determining perceptually sufficient texture descriptors is as yet unsolved. In general, the mean, variance and autocorrelation function are not sufficient texture descriptors even though the 'Julesz conjecture' third order and higher order density differences between texture field paires are not discernable by human vision [87].

A texture model may be used for texture analysis by estimating the parameters of the model that would result in a a synthesised texture that matches the pertinent statistics of the texture to be analysed.

Direct estimation of the parameters of a model is difficult. The Julesz model is non-linear, while the linear programming model is complex. The autoregressive model requires the measurement of the first order and second order moments of the texture and to estimate the probability density of the model driving process.

One approach described by Faugeras[89] is ullustrated below in fig. 6.2.1. The texture field is first decorrelated by a whitening operator to produce a field $\omega(j,k)$ that forms an estimate of the independent, identically distributed driving process of the autoregressive model. The histrogram of over some window is measured to estimate the probability density $p(\omega)$ of the driving process.

The whitening operator is derived from the measured testure autocorrelation function (6.2.1(b)). The first four moments of the histogram - mean, standard deviation and kurtosis are used to represent the histogram. The autocorrelation function is represented in a similar way by its histogram

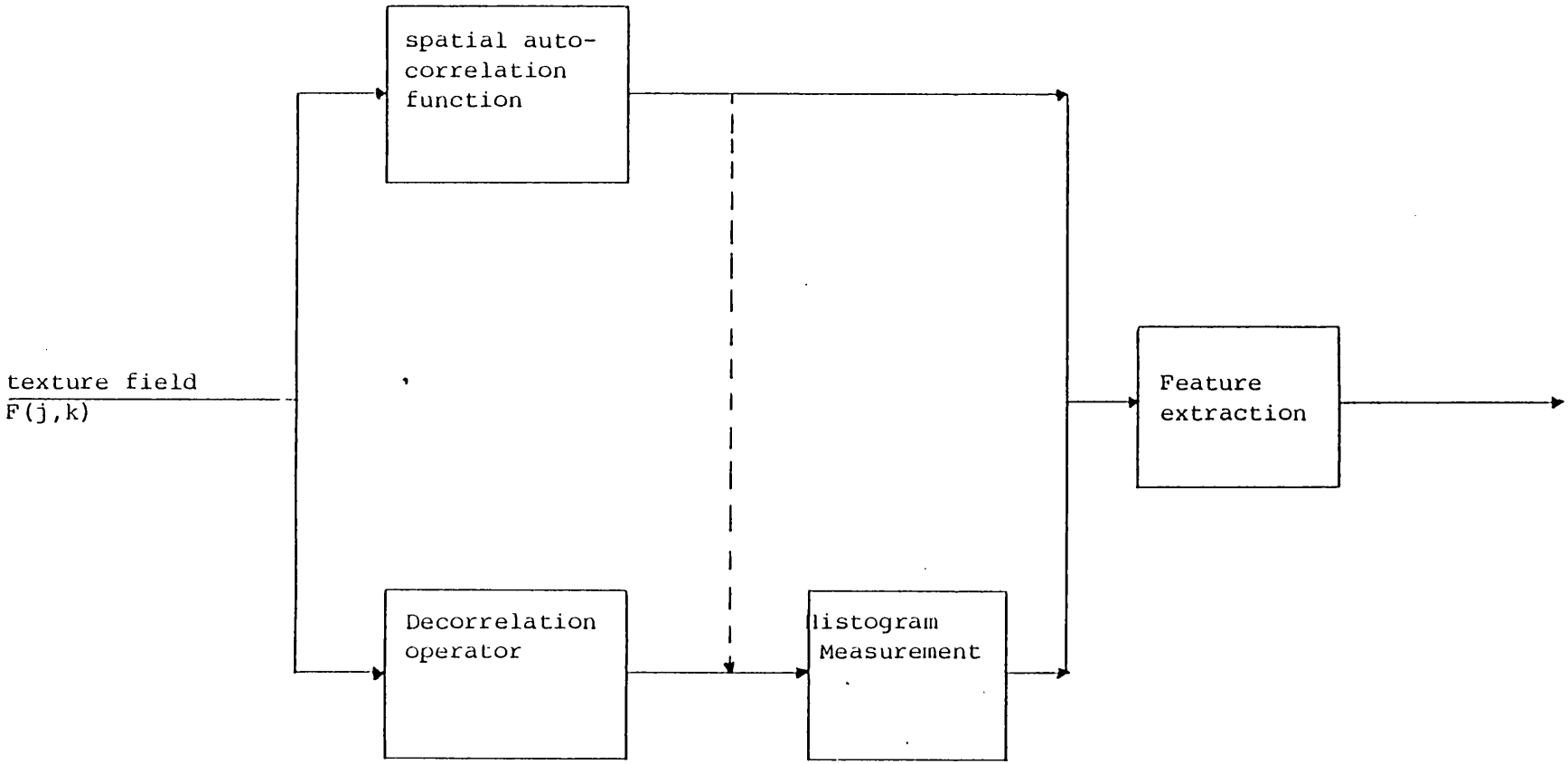


Fig. 6.2.1 Decorrelation method of texture feature extraction

The result is an eight dimensional feature vector that gives good classification but is limited by computational bounds on the whitening step and the autocorrelation function measurement.

As a computational simplification, the whitening operation can be replaced by a simple approximate decorrelation operator such as the Laplacian or Sobel and the characterisation of the autocorrelation function may be totally eliminated.

The decorrelation operator produces edges of the texture field and the sufficiency of measurements on this edge field in characterising a natural texture field [89] suggests that testure is uniform where there are no edges and is discontinuous at the edges.

Edges contain a lot of visual cues such as surface orientation, 3D shape etc. and consequently recovery of this information from edges has received a lot of attentuation [90].

6.2.3 Edge and contour extraction

The extraction of edges from an image is quite difficult and unreliable. While many edge detection operators and segmentation algorithms have been proposed and used, they do not always give all the essential edge information that correctly characterises the

image. Most useful techniques such as Roberts gradient technique [91] give a fairly good approximation to the ideal with a few extraneous lines. A missing edge might occur for example when a physical edge does not result in a visible brightness discontinuity.

In order to use contours for reconstruction, they must be extracted as nearly perfectly as possible and spurious edges must be eliminated.

Simple linear edge detection methods involve performing a discrete spatial differentiation. Ofter this is achieved by convolving the image function with a gradient mask.

A commonly used mask is the Laplacian edge detector below

$$H = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \qquad (6.2.9)$$

The new image intensity $g(x,y)$ at the point (x,y) is formed from the image $f(x,y)$ by using the mask above to give

$$g(x,y) = f(x-1,y) + f(x+1,y) + f(x,y+1) - 4f(x,y) + f(x,y-1) \qquad (6.2.10)$$

Several other masks may be used, and an 'edge map' may be formed by thresholding the resultant edge image.

Prewitt [96] considers several discrete differentiation masks. These compass gradient masks have maximum response to certain direction, producing maximum output for luminance charges in a preferred direction.

Non-linear edge detection systems utilise non-linear combinations of pixels over a limited windows area. Examples include Roberts [91] where the edge image is formed by the cross operations given as

$$g(x,y) = ([f(x,y) - f(x+1,y+1)]^2 + [f(x,y+1) - f(x+1,y]^2)^{\frac{1}{2}} \quad (6.2.11)$$

Another operator attributed to Sobel [92] uses the 3x3 window below to describe the pixel numbering convention.

a	b	c
d	e	f
g	h	i

and defines

$$S_x = (C + 2f + i) - (a + 2d + g)$$

$$S_y = (g + 2h + i) - (a + 2b + 1)$$

the edge detected image gives the gradient at point e as

$$g = \sqrt{S_x^2 + S_y^2} \quad (6.2.12)$$

Kirsh [94] describes another 3x3 operator. Many other techniques have been proposed and compared - For example Rosenfeld [95] Marr [93], Davis[97].

The approach used in this thesis starts off with an edge detection algorithm followed by a thinning and thresholding operation. The Roberts edge detector is used to produce the first edge image. This method has been found to produce edge images that are well suited to this application, giving both good luminance and textures edges.

For data compression it is now essential to operate on the edge image and produce an edge map that contains all the essential edge information but which contains edgges that are one pixel wide rather than several pixels wide.

Contours are thus formed by reducing the dimensionality of the edges. This is necessary So that polygonal approximations can be used to act on the contours so obtained for compression. To obtain the best contour from the edges one must seek the maximum of the edge detection operator in each window, as this is often the point of steepest change.

One way is to take two derivative operations. The maximum of the first derivative operator corresponds to the zero crossings of the second. This is the principle of the Marr-Hildreth operator [93] which detects the zero crossings of hte output after the application of the Laplician (6.2.9). The whole question of obtaining the correct contours is fraught with difficulty because it is very easy to miss out certain parts of the contour when you use a second derivative operation.

The Marr-Hildreth operator is directional and will give correct contours for horizontal and vertical directions. The horizontal and vertical gradient operators work well for most orientations but leave gaps when the edge deviates slightly from the horizontal or vertical. An alternative is a thresholding operation where a decision is made about the edge intensity threshold and all the edge pixels with intensities less than this are removed. The disadvantage with this approach is that it assumes a global intensity over the image or window area. This does not work well for all the contours,

but works well in those cases where the chosen thresholds correspond to a particularly dominant edge. The best approach is a combination of an edge follower and adaptive thresholder. Once one correct edge point is obtained the others on the same edge are grown from it and the threshold is chosen locally. This allows different thresholds for different areas.

6.3 Contour and texture parametrisation of images.

Given the contours and the texture information a suitable data structure can be used to represent the image for purposes of compression. The overal block diagram is shown in fig. 6.3.1 .

Where F(j,k) is the image pixel array and p(c,t) is the contour and texture representation. The block diagram of the reconstruction system used is shown in fig. 6.3.2

Having obtained the correct contours, several options are available for the parametrisation of the image. One option would be to take measures of the texture in the regions on either side of the contours, and later regenerate the texture using a suitable model and texture synthesis algorithm. This requires a fair amount of computation good closed regions are not always formed. Computing

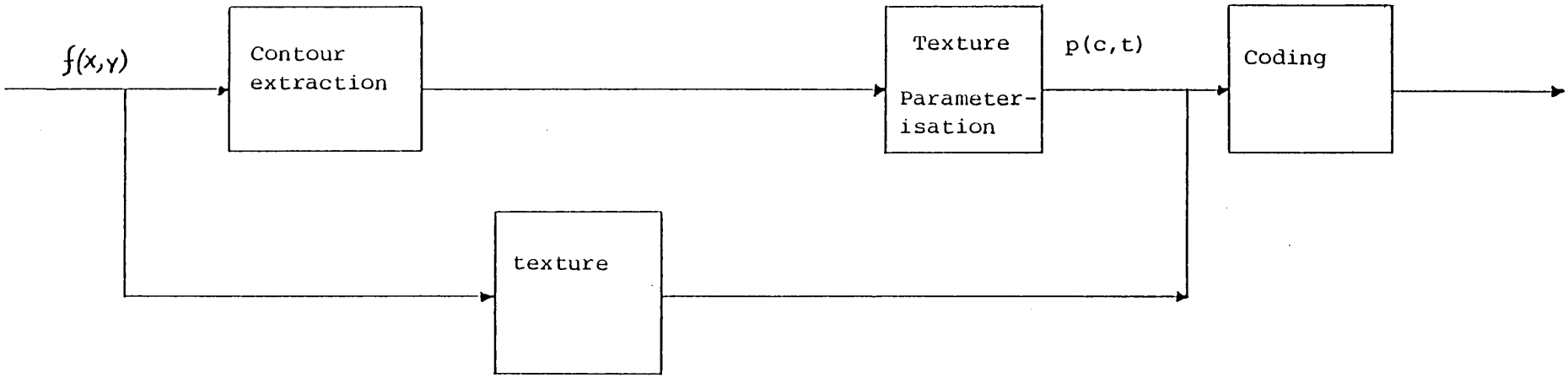


Fig. 6.3.1 Contour and texture parameterisation.

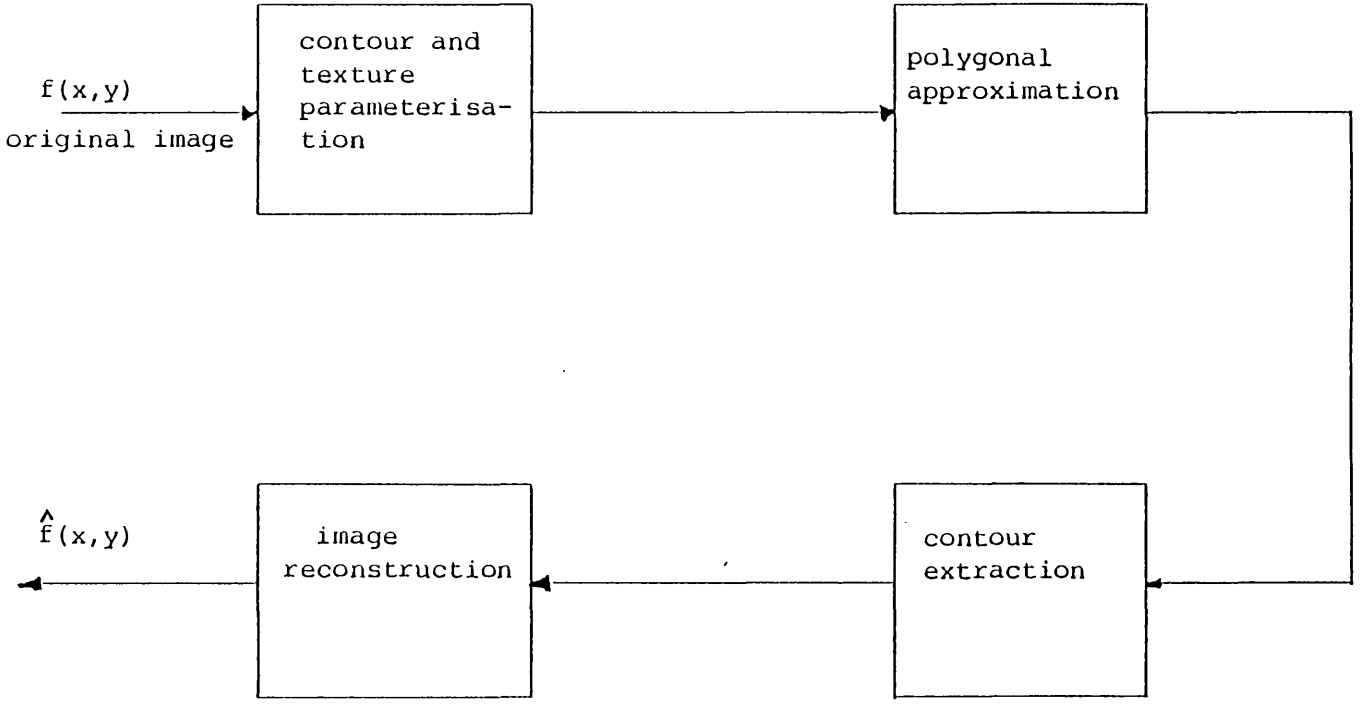


Fig. 6.3.2 Block diagram of reconstruction system.

texture measures on an irregular grid and later regenerating this texture over the whole image requires an inordinate amount of processing and storage.

The appracch used here and by others[17] is to represent the texture on either side of the contour by the pixel intensities and to recover the image by inserting these intensities and then estimating the intensities in the non-edge region.

The given contour information does not uniquely specify the image and there are several images that may be compatible with the contour information.

All these images should not contain any more contours in the areas between the given contours. This suggests that the image intensity between the contours should be maximally smooth and a reconstruction constraint should minimise the probability of introducing any new contours.

When the image being sought is bilevel, this corresponds to the classical contour filling problem.

In the gray level case several approaches are considered.

(i) Spline interpolation.

If the contour information and the texture on either side of the contours is given, this approach has been found to give a reconstruction image compatible with the given information. The intermediate texture is obtained by spline interpolation using the given texture information.

(ii) Iterative Constraint

The missing texture may be estimated using an iterative reconstruction algorithm, subject to the constraint that no new

contours may be introduced. The reconstruction problem is formulated as a constrained optimisation problem as follows: Given a local measure of image intensity variation at pixel (x,y) involving the derivatives of intensity with respect to the co-ordinates; $F(f_x, f_y, f_{xy}.....)$, the reconstructed image f(x,y) is obtained as the minimisation of this measure over the images.

$$\min_{f(x,y)} \iint F(f_x, f_y, f_{xy}.....)\, d\,x\, dy \qquad (6.3.1)$$

subject to known f(x,y) along the contours.

f_x is the derivative with respect to the x-co-ordinate at (x,y)

f_y is the derivative with respect to the y-co-ordinate at (x,y)

$F(f_x, f_y, f_{xy}.....,)$ is a measure of image intensity variation at (x,y) and is a function of the derivatives.

If the squared magnitude, F, is used as a measure of this variation,

$$F = f_x^2 + f_y^2$$

Recasting the above problem in a discrete form, with the derivatives replaced by differences, we need to minimize

$$C = \sum_x^M \sum_y^N [f(x,y) - f(x,y-1)]^2 + [f(x,y) - f(x-1,y)]^2 \qquad (6.3.2)$$

where f(x) is the pixel intensity at (x,y)

The minimisation is carried out at all points except at the

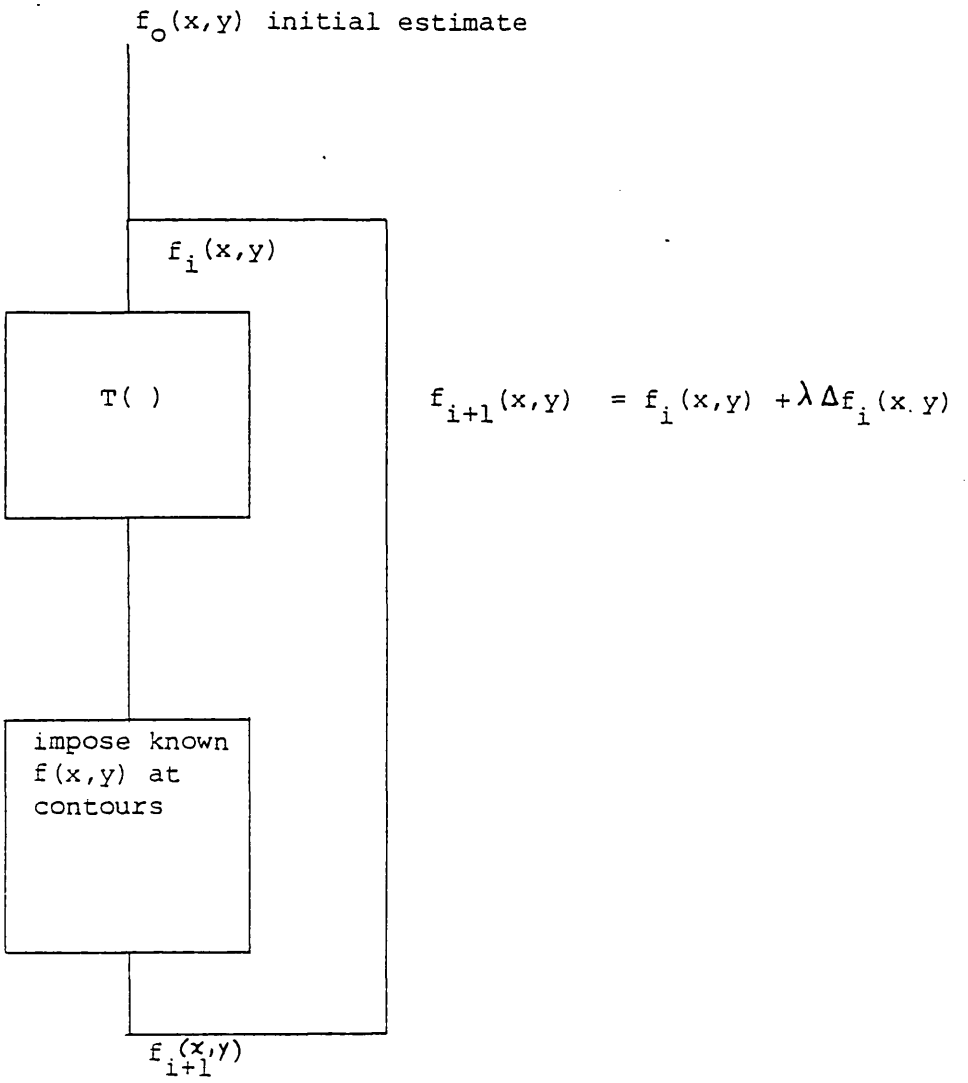


Fig. 6.3.3 Iterative algorithm for reconstruction.

contour where the given values of intensity are used. To minimise, a sufficient and necessary condition is

$$\frac{\partial C}{\partial f(x,y)} = 0 \qquad (6.3.3)$$

defining the Laplacian

$$\Delta f(x,y) = f(x-1,y) + f(x, y-1) + f(x+1,y) + f(x,y+1) - 4f(x,y) \qquad (6.3.4)$$

which implies that if (6.3.3) is satisfied, $\Delta f(x,y) = 0$ at all points except the contour points.

As before, we can consider the mapping from the given $f_i(x,y)$ to an estimate $f_{i+1}(x,y)$, so that to solve the above iteratively. $f_{i+1}(x,y) = T\{f_i(x,y)\}$ where $T(.)$ is a transformation from the ith to the (i+1)th iteration. Linearized this becomes

$$f_{i+1}(x,y) = f_i(x,y) + \lambda \Delta f_i(x,y) \qquad (6.3.5)$$

where λ is a relaxation parameter.

$f_o(x,y)$, the initial estimate is obtained by using the given values of texture at the contour points and 0 elsewhere. The Laplacian is the gradient of C with respect to f so the iterative algorithm above can be considered as a steepest descent algorithm for the minimisation of C with respect to f.

On a variation of this algorithm, the unknown texture is initially generated using a texture model such as the autoregressive model discussed previously.

$$\Delta f(m,n) = (f(m-1,n) + f(m,n-1) + f(m+1,n) + f(m,n+1) - 4f(m,n)$$
$$i \qquad i+1 \qquad i+1 \qquad i \qquad i \qquad i$$

It is interesting to consider the convergence behaviour of this algorithm.

The error between the latest estimate $f_i(x,y)$ and original image is

$$e_i(x,y) = f_i(x,y) - f(x,y)$$

Where $f(x,y)$ is the smooth reconstruction being sought.

Using this linear relationship, we obtain a similar equation to 6.3.5 for the error at the $(i+1)^{th}$ iteration as

$$e_{i+1}(x,y) = e_i(x,y) + \lambda \Delta e_i(x,y)$$

Edges represent high frequency information and the iteration tries to prevent the introduction of edges (high frequencies) in the non-edge region. The iteration can thus be modelled as a repeated low pass filtering operation in the non edge region . Consequently, we may take

$$E_{i+1}(\omega_1,\omega_2) = H(\omega_1,\omega_2)\ E_i(\omega_1,\omega_2) \qquad (6.3.6)$$

where $E_{i+1}(\omega_1,\omega_2)$ is the DFT of $e_{i+1}(x,y)$

and $H(\omega_1,\omega_2)$ is the transfer function of the low pass filter.

Since the points at the contour not affected, the filter is not shift invariant so this analysis is not entirely rigid, but we can evaluate

$$H(\omega_1,\omega_2) = \frac{E_{i+1}(\omega_1,\omega_2)}{E_i(\omega_1,\omega_2)} = \frac{F\{e_i(x,y) + \lambda \Delta e_i(x,y)\}}{F\{e_i(x,y)\}}$$

This reduces to (see appendix 4)

$$= \frac{[1-4\lambda] + \lambda[\exp(j\omega_1) + \exp(j\omega_2)]}{1 - \lambda[\exp(-j\omega_1) + \exp(-j\omega_2)]} \qquad (6.3.7]$$

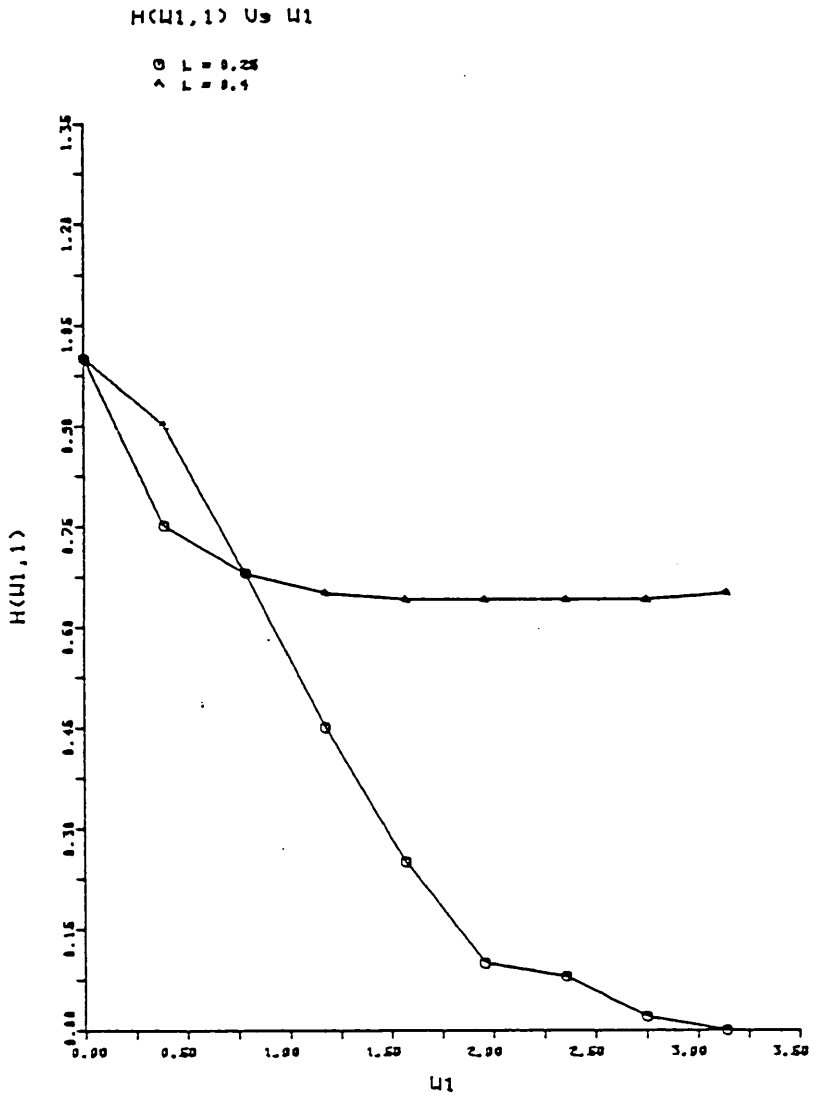


Fig. 6.3.4 Graph of $H(w_1,1)$ V's

6.4 Polygonal approximations

In the general polygonal approximation problem we want to find a polygon that closely approximates a given curve while having a much smaller number of vertices. This problem is stated as follows:

Given a curve C

$$C = \{x_i, y_i) \mid \mid i=1,.....,N\}$$

determine the minimum set of m (usually) straight line segments $(c_1.....,c_m)$ that will approximate the curve subject to a closeness criterion.

So the requirements are

(i) find the minimum number of segments (m) for a given error E, and/or

(ii) find the minimum total length of polygonal sides for a given number of segments m and error E.

While segments are usually chosen such that the data points along them are on straight lines, it is often useful to describe other forms e.g. arcs.

The solution to this problem is generally non-unique and optimal methods of solution such as dynamic programming [18] have a prohibitively high computational cost. As a result, there have been many proposals in the literature of algorithms that produce a sub-optimal solution but with much lower computational cost [12] . The converse problem is the reconstruction of the contour from the polygonal approximation. This may be stated as one of finding the best curve C to fit the given set of straight line segments and some constraints.

Apart from the different approximation methods, there are differing estimation methods for the error between the polygonal approximation and the curve.

We propose an algorithm that is suboptimal because it is constrained to have points of the original image in the polygon. But this will be useful in reconstructing the contour. It makes local decisions about whether to extend the segment and the advantage of doing this is that they may be done in reverse, hence giving a

contour that would have satisified them. i.e. the same decisions can be made when recovering the contour about whether the segment can be broken any further

The general situation is illustrated in the figure below

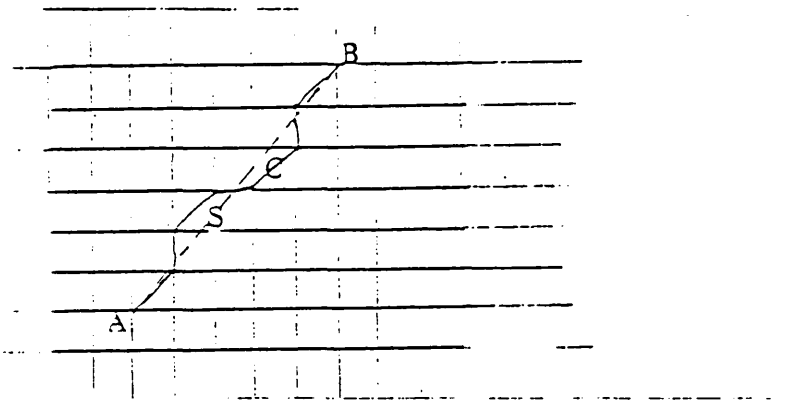


Fig. 6.4.1 Geometrical illustration of polygonal approximation

The algorithm

input: Co-ordinates of the curve C(I), I = 1,....., N

output: Selected co-ordinates corresponding to j^{th} segment

$S_j(k)$, k = 1,....., M

definition: f(.) is a function which determines how close the segment with the present extension is to the segment without the extension.

E is the allowed error for f(.) above.

j is an index of the number of segments j=1,.....m.

k is an index of the points belonging to j^{th} segment.

T is a temporary segment.

1. initialise j=0
2. perform procedure SEGMENT until end of contour
3. If end-of-segment = time

$j = j + 1$

$k = 0$

Else

$k = k + 1$

Go to 2

Prodcedure SEGMENT

n = next point along contour

$T = S_j + n$.....extended segment

IF $f(T) - f(S_j) > E$

THEN end-of-segment = true.

Else

$S_j(k) = n$

End

At each local point, E is defined as the maximum allowable angular divergence which was chosen as 22° in our experiments. f(.) was taken simply as the new angle. To assist in the detection of gradual changes, an expected short segment length of about 5 points on the grid was used, so that a comparison was also made every neighbouring sets of 5 points in the contour and these would be merged if they don't differ to much, otherwise they would be split. Fig. 6.6.3 shows a polygonal approximation obtained using this algorithm.

6.5 Contour recovery

Since the polygonal approximation is so close to the contour, it may well be that if grid points are selected at random close to the approximating polygon side, we may obtain a reasonable contour.

A better reconstruction may be obtained by applying the algorithm outlined above, in reverse. (See fig. 6.5.1). Given the contour points in the polygonal approximation, we want to find a path along the grid which satisfies the rules for constructing the polygon. Even though the path joining two vertices in a connected graph is not unique in the general case, knowing the rules for constructing the polygon, considerably reduces the ambiguity and the contour may be grouwn between the vertices. At each point the next point is determined so as to satify these rules.

This time the algorithm starts from point B (fig. 6.4.1) and works its way towards point A, at the same time satisfying the constraints on angular divergence that were used to construct the polygon segment. There is only a finite number of curves that can satisfy these constraints and in most cases are not really going to affect the acceptability of the reconstructed contour.

If it becomes impossible to satisfy the rules the algorithm back tracks to revise its decisions. A buffer store is kept of the previous points. If a contour fails to meet the end vertex after a large number of points (far beyond the expected number), the algorithm backtracks. In our work this was loosely defined as 2(X+Y) where x = direction, Y is length in Y direction of polygon. The solution can be made almost unique if the number of contour points represented by each polygonal segment is kept as well. This

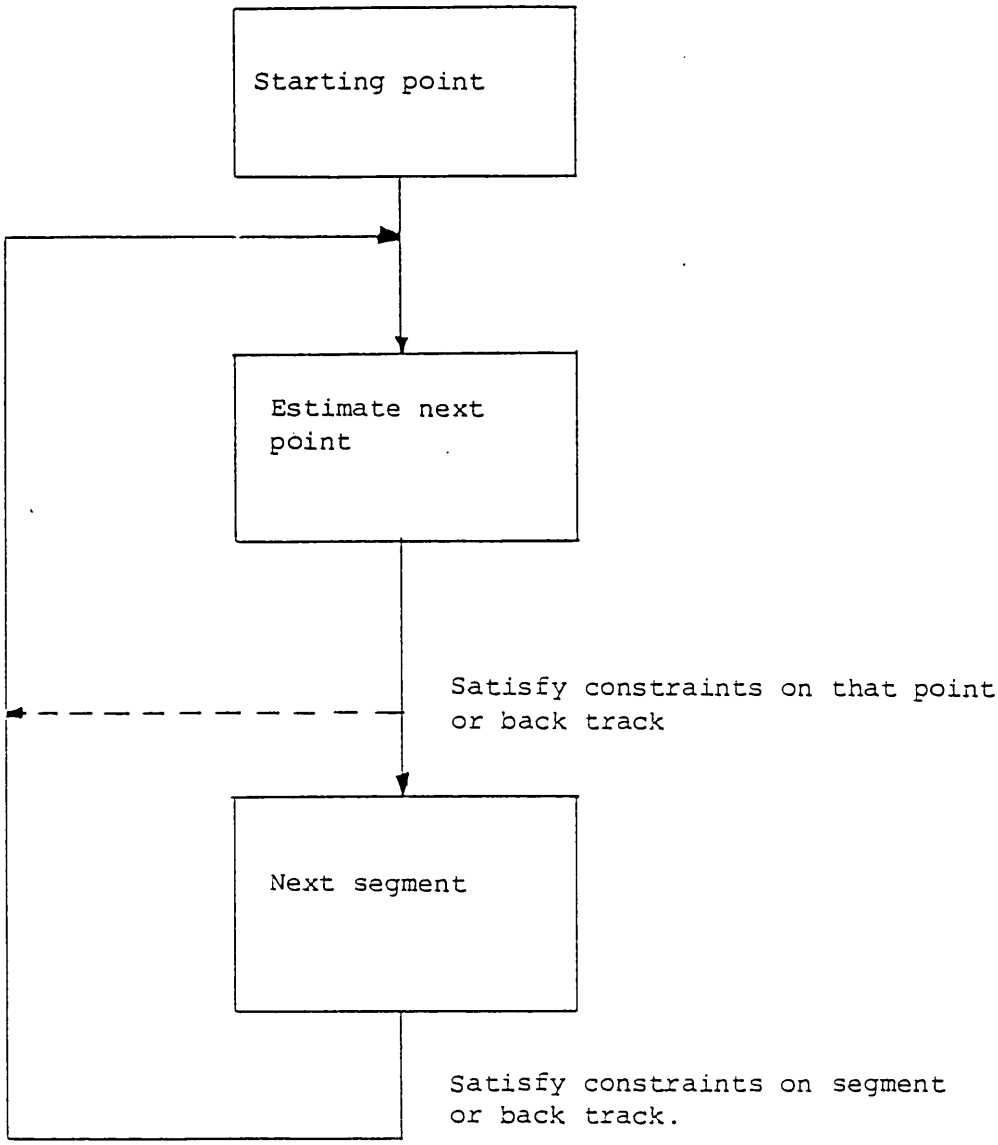


Fig. 6.5.1 Contour recovery.

has the disadvantage of increasing the number of points. But if a parametric representation of the contour and texture has been obtained, this information would be present anyway from the number of texture points

6.6 Experimental results and Conclusions

The ideas outlined in the previous sections were applied to the image of fig. 3.3.3 Fig. 6.6.2 shows the contour image of the test image and fig.6.6.3 shows the polygonal approximations obtained using the algorithms outlined in section 6.4 Fig. 6.6.4 shows the results of reconstructing the contours from the polygonal approximations. Fig. 6.6.5 shows the reconstructed images using spline interpolation (a) and the iterative approach outlined in section 6.3 (6.6.5(b)).

The results indicate that subjectively acceptable reconstructions from contour and texture parametrised information are possible.The compression obtainable depends somewhat on the nature of the image. For example the 'script' images is represented just 3 polygonal approximation curves, or a total of 39 segments. The parametric representation then consists of 3 1-D vectors of 42 points, encoding the x-co-ordinates, y-co-ordinates and intensity.

These have a lot of inherent redundancy and may be further encoded among themselves. For example the first intensity along the contour may be fully encoded and the others would only have a residual coded so that they may be predicted from it. Using similar methods to encode the co-ordinates, the bit rate required is obtained by adding up all the bits needed to represent each vector and dividing by the total needed for the original image.

Fig. 6.6.1 Edge image

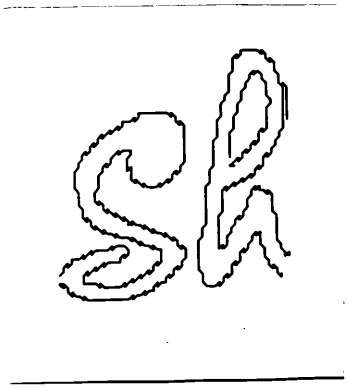


Fig. 6.6.2 Contour image

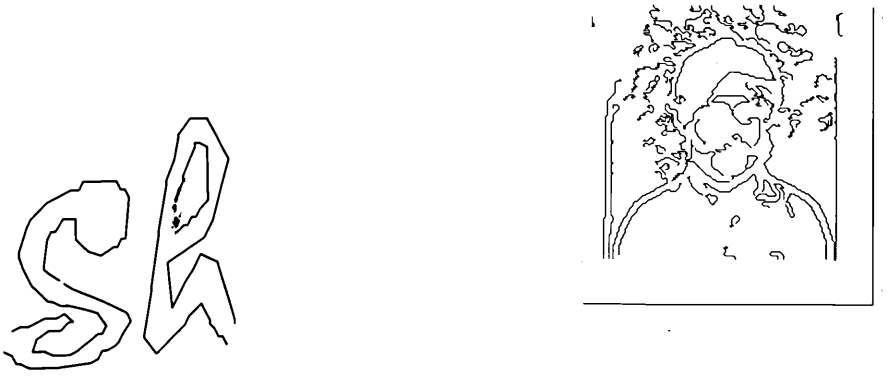


Fig. 6.6.3 Polygonal approximations

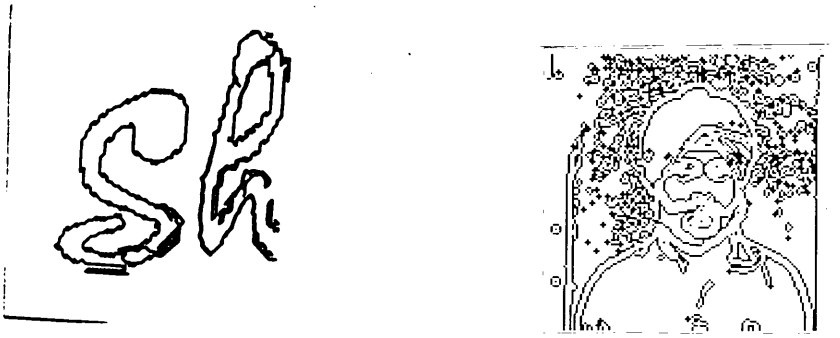


Fig. 6.6.4 Contour reconstruction

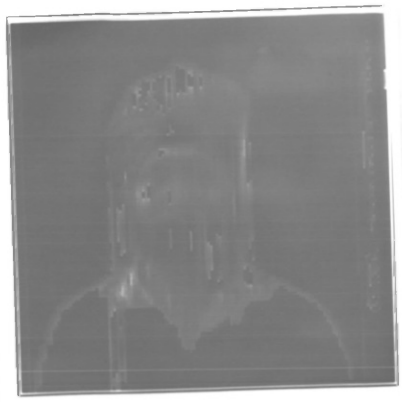


Fig. 6.6.5 Spline interpolation

Fig. 6.6.6 Constrained iteration

The original required 64x64x7 bits and the parametric required 42x2x4+42x5 bits

This allows up to 4 bits for encoding the differences in the co-ordinates and 5 bits for the intensity differences. This is about 0.13 bits/pixel and may be reduced further. The 'man' image shown however, uses about 0.4 bits/pixel.

The criteria of maximum smoothness gives and acceptable image and 30 iterations seems to be about the required number for reconstruction. The main factor on which success depends is the correct choice of edge image and extraction of the contours. Under some conditions, a good starting image may be available - e.g. in coding image sequences, the previous sequence is available and it would give faster convergence. The quality of the reconstructed images is worse than the quality of transform coded images at the higher bit rates, but is better at the very low bit rates.

CHAPTER SEVEN

CONCLUSIONS

CHAPTER 7

CONCLUSIONS

This chapter presents the main conclusions of this thesis and some suggestions for research areas where future research could be useful.

Issues related to three problems of image reconstruction were considered. These were image reconstruction from the phase or magnitude of its Fourier transform and from information about its contours.

The conditions under which a sequence is uniquely defined by its magnitude or phase function have been examined. In particular, it was shown that a finite support constraint is in most cases, sufficient for a 2-D sequence to be uniquely defined to within a scale factor by its phase. The finite support constraint is sufficient for a class of sequences which are related to within a sign, linear shift and time reversal to be uniquely defined by the magnitude of the Fourier transform.

Practical algorithms for reconstructing the image from its phase were presented. Both iterative and non-iterative algorithms were discussed. The effect of noise on the phase only reconstruction was investigated.

The convergence of the phase only iteration is reduced when the available phase is noisy even though the fixed point to which it converges will not change if the iterative mapping is a contraction over the domain of the noisy sequence.

It is seen that the selection of a good starting magnitude function can speed up the phase only reconstruction from noisy phase and several methods of selecting the starting magnitude are examined.

While the error in the final reconstructed image will be the same for both the non-iterative and iterative reconstruction algorithms, it is possible that the image at some point in the iteration will be better than the one to which it converges. Consequently it is important to study the iteration theoretically, using error measures that can relate to the subjective quality of the image. More work is also required in developing relationships between the phase and the space domain image. Such relationships could be useful in developing constraints for estimating the phase. Further theoretical work is needed on uniqueness in the case where the available phase is noisy or quantised, and on reconstruction methods to reduce the effect of the error.

An issue that is related to the above problem is Fourier transform coding, which is studied in the second part of chapter five.

A Fourier transform coding technique is examined where only the phase is coded and transmitted and the magnitude is reconstructed from the phase information of the receiver. The method is shown to provide a good reconstructed image, but bit-rate and quality of the image depend on the ability to form a good estimate of the magnitude function at the receiver. Our results indicate that fewer iterations would be required if such an estimate was available.

To develop practical coding algorithms using only the phase more work is required on selecting magnitude functions, and on coding the phase function.

Adaptive techniques that take full advantage of the apparent link between phase and image edge structure as well as between magnitude and average image brightness are required. One such technique could classify an image into several regions where the same starting magnitude function could be used.

The question of how best to code the phase is still an open problem. When both magnitude and phase have been coded, it is usually on the basis that more bits should be used to represent the larger co-efficients at the lower frequency. The same approach is used for other transforms. However, this form of lowpass filtering will tend to destroy edge structure at high compression ratios and it may be better to choose co-efficients by some other technique. As shown, a slightly different analysis based on the phase, gives broadly the same result that lower frequencies need more bits.

Further studies of coding methods that use only the phase are needed because of applications such as the kinoform. There is also need for the development of solutions to the non-iterative phase reconstruction algorithm for reasonable sizes of image arrays.

In Chapter four, the importance of a priori information for the success of iterative algorithms for magnitude only reconstruction was studied. Reconstruction is generally not possible without a priori information. The type of information available depends on the application area, but the two most commonly applied constraints are the non-negativity and finite support constraints. However, it is seen that simply applying these constraints does not generally lead to quick convergence, even with an accelerated algorithm.

A more powerful concept is to consider the design of starting

images which will produce an image that satisfies the constraints after passing through the iteration.

The convergence of the iteration improves when some information about the phase is provided, but this kind of information may not be available in practice.

There is considerable scope for further work on reconstruction algorithms. Further theoretical work is needed on the relationship between constraints in the frequency and space domains so that adaptivity parameters can be designed not only to speed up the iterations but also to satisfy the constraints in both domains.

The uniqueness of the magnitude is still an open problem especially in the case of complex valued sequences and on reconstruction from a noisy magnitude function.

One application investigated in this thesis showed that the GS iterative algorithm will reconstruct from a noisy (Wiener estimate) magnitude when a noisy estimate of the phase is available.

In chapter five, a method is presented to estimate the phase as well as the magnitude of the spectrum of an image that has been degraded by linear blur and additive noise. The method consists of a conventional statistical restoration filter which serves to restore the magnitude function, followed by a Gerchberg Saxton type iterative algorithm to estimate the phase function.

Using a finite support constraint is found to reduce the phase domain error, and the spatial activity constraint leads to visually better images. This method provides a practical way to obtain better resolution for images that are restored using a zero phase filter. More work is needed on techniques to estimate the correct

phase from the noisy measured phase function.

In Chapter six, the reconstruction of images from polygonal approximations is shown to give reasonable images at high compression ratios. The image is represented by the polygonal approximation of its contours and the pixel intensity along the contours. The contours are obtained by a derivative and thresholding operation on the edge image and their correct extraction is essential to the recovery of an image that resembles the original.

The image pixels are recovered by an iterative algorithm that minimizes the square of the derivatives in the non-edge region subject to the constraint that the edge pixels maintain their known intensities.

This approach is found to give useful images and could possibly be developed further. More work is needed on algorithms that extract the correct contours as this makes a lot of difference to the quality of the final image, and on developing good first estimates. Since these methods are mostly concerned with reproducing the correct relationships between the pixels, rather than the actual intensity values, the final image quality may be improved by some form of image enhancement. For example histogram equalisation could be used to redistribute the intensities on the known range. The fact that a number of iterations must be performed is a disadvantage as is slows down the response time, but the calculations are neighbourhood operations and could be performed very quickly in parallel.

Extracting the correct edges from an image is still an

interesting research area with many algorithms being proposed for various classes of images. Image reconstruction from these edges may be used to improve the performance of an edge detecting system, if the image reconstructed from the contours is compared with the original. This may be important in some applications where a much compressed version of the image scene needs to be stored with very limited resources or must be transmitted along a very narrow bandwidth channel.

Many of the reconstruction algorithms discussed in this thesis and in the literature are ad hoc and based on heuristic or intuitive arguments. In part this is because of the difficulty in properly quantifying the concept of image information. For example it is difficult to provide a simple measure that would explain why a phase only image is intelligible and a magnitude image is not.

There is therefore a need to develop mathematical or probabilistically based reconstruction methods and more work is needed on developing models that are appropriate to image reconstruction from incomplete information.

Summary

Three reconstruction problems have been considered in some detail:

1. Reconstruction from the phase of the Fourier transform representation of an image
2. Reconstruction from the magnitude of the Fourier transform representation of an image.
3. Reconstruction from polygonal approximations of the contours of an image.

It is shown that real image functions are uniquely defined to within a scale factor by their phase. There is a very clear visual link between the original image and the 'phase-only' image. Estimates of the image are obtained from the phase only image both by solving linear closed form equations between the original image and the phase and by using iterative algorithms, applying known constraints in each domain as we alternately move between space and frequency domains. The non-iterative method is limited by the size of the matrix as it requires inversion of the matrix. A subimage size of 8x8 is reasonable. The iterative algorithm is also limited by size as it requires 2N-1 x 2N-1 size DFT's for an NxN image. It is also limited by time as a number of iterations (around 30) are needed for reconstruction. Furthermore, the convergence of the iteration is reduced when the phase function is noisy or quantized. In this case, good performance is obtained by selecting good starting magnitude functions and by using adaptive relaxation techniques. Consequently, the success of iterative phase only reconstruction as a compression method depends on the ability to form excellent starting magnitude functions and to perform many iterations quickly. This is possible with array processors and may be even faster if hybrid digital/optical processing is used to perform the DFT's.

Magnitude only reconstruction is generally not successful without a priori information. Magnitude only images are not as intelligible as phase only images and the relationship between magnitude only and original images is non-linear and consequently very difficult to solve non-iteratively. The basic iterative algorithm converges extremely slowly, but reconstruction improves when further a priori information about phase or its sign is available. Unfortunately such information is

unobtainable in most practical situations. Frequently only the modulus is measurable so that while the algorithms are effective, they are not readily applicable to real life problems. The methods however, do provide a practical way to obtain better resolution for images that have been restored using a zero phase filter.

Reconstruction from polygonal approximations of contours, gives intelligible images at very high compression ratios. The compression obtainable is greater for images with significant contour content such as script images and compares very favourably with other coding methods.

Future Work

The following problems, among others, remain to be solved;

1. Developing mathematical or probabilistically based reconstruction methods.
2. Theoretical examination of relationships between the image and its phase.
3. Uniqueness of the phase only representation when the available phase is noisy or quantized.
4. Uniqueness of the magnitude for complex valued sequences.
5. Criteria for coding the phase.
6. Fast algorithms for performing the iterations.

Solutions to these problems would go a long way to improve understanding of image reconstruction and contribute to the development of image processing science.

APPENDIX I

Proof

$$X(e^{j\omega}) = X_r(e^{j\omega}) + jX_I(e^{j\omega})$$

The Fourier Transform of $x_e(n) = \sum_{-\infty}^{\infty} x_e(n)e^{-j\omega n}$

$$= \tfrac{1}{2}\sum_{-\infty}^{\infty}[x(n) + x(-n)]e^{-j\omega n}$$

$$= \tfrac{1}{2}\sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n} + \tfrac{1}{2}\sum_{n=-\infty}^{\infty} x(-n)e^{-j\omega n}$$

$$= \tfrac{1}{2}\left[X(e^{j\omega}) + x(n)\ e^{j\omega}\right] = \tfrac{1}{2}\left[X(e^{j\omega}) + X^{x}(e^{-j\omega})\right]$$

$$= \tfrac{1}{2}\left[X_r(e^{j\omega}) + jX_I(e^{j\omega}) + X_r(e^{j\omega}) - jX_I(e^{j\omega})\right] = X_r(e^{j\omega})$$

APPENDIX II

Proof of phase uniqueness

Theorem:

Let $x(n_1,n_2)$, $y(n_1,n_2) \in F((n_1,n_2)$ with support $R(N_1,N_2)$.

If $X(z_1,z_2)$ has no symmetric factors and ϕ

$$\phi_x(\omega_1, \omega_2))_M = \phi_y(\omega_1, \omega_2)_M \quad \text{for} \quad \begin{matrix} M = M_1, M_2 \\ \omega_1 = 1,.....M_1 \\ \omega_2 = 1,.....M_2 \end{matrix}$$

then $y(n_1,n_2) = \beta x(n_1,n_2)$ for some positive number β.

If $\tan [\phi_x(\omega_1, \omega_2))_M] = \tan \phi_y[(\omega_1, \omega_2))_M]$

$y(n_1,n_2) = \beta\ x(n_1,n_2)$ for some number β

Proof

Let $x(n_1,n_2)$, $y(n_1,n_2)$ satisfy the theorem.

Consider the sequence

$$g(n_1,n_2) = x(n_1,n_2), * y(-n_1,-n_2) \tag{A2.1}$$

$x(n_1,n_2)$ and $y(n_1,n_2)$ have support $R(N_1,N_2)$ hence their convolution $g(n_1,n_2)$ is zero outside $-N_1<n_1< N_1$, $-N_2<n_2< N_2$. Therefore if $M>2N-1$, then the M-point DFT of $g(n_1,n_2)$ is the product of the M-point DFT's of $x(n_1,n_2)$ and $y(n_1,n_2)$

$$G(k_1,k_2)_M = X(k_1,k_2)_M \cdot Y(-k_1,-k_2)_M \text{ where } M = M_1, M_2 \tag{A2.2}$$

Thus if $\phi_x(k_1,k_2)_M = \phi_y(k_1,k_2)_M$ or $\tan[\phi_x(k_1,k_2)_M] = \tan[\phi_y(k_1,k_2)_M]$

Then $G(k_1,k_2)_M$ must be real and since $g(n_1,n_2)$ is non-zero in the region $-N_1<n_1< N_1$, $-N_2<n_2< N_2$, it follows that $g(n_1,n_2)$ must be even, and its Fourier transform real.

$$G(z_1,z_2) = G(z_1^{-1},z_2^{-1}) \qquad \text{(A2.3)}$$

therefore from A2.2 and A2.4

$$X(z_1,z_2) = Y(z_1^{-1},z_2^{-1}) = X(z_1^{-1},z_2^{-1})\ Y(z_1,z_2)$$

Multiplying by $z_1^{-N_1}\ z_2^{-N_2}$, results in

$$X(z_1,z_2)\quad \tilde{Y}(z_1,z_2)\ z_1^{-m_1}\ z_2^{m_2} = \tilde{X}(z_1,z_2)\quad Y(z_1,z_2)\ z_1^{-n_1}\ z_2^{-n_2}, \qquad \text{(A2.4)}$$

where $m_1,m_2 > 0$ and $n_1,n_2 > 0$ are integres

and $\tilde{X}(z_1,z_2) = z_1^{-N_1}\ z_2^{-N_2}\ X(z_1^{-1},z_2^{-1})$

Consider a non trivial irreducible factor $X_k(z_1,z_2)$. If a polynomial has two differenct factorisations, the factors in each can always be ordered in such a way that the factors are associated - i.e. the factors in both factorization will be equal to within a factor of zero degree. Therefore $X_k(z_1,z_2)$ must be associated either with a factor of $\tilde{X}(z_1,z_2)$ of a factor of $Y(z_1,z_2)$.

But, if $X_k(z_1,z_2)$ is associated with a factor of $\tilde{X}_k(z_1,z_2)$ then $X_k(z_1,z_2) = \alpha\tilde{X}_i(z_1,z_2)$ for some i. If i = k this implies

$X_k(z_1,z_2) = \alpha^2\ \tilde{X}_k(z_1,z_2)$ and X_k is symmetric. (If $\alpha=\pm1$)

If $i \neq k$

$$X_k(z_1,z_2)\ X(z_1,z_2) = \alpha\ \tilde{X}_i(z_1,z_2)\ X_i(z_1,z_2) = \alpha A(z_1,z_2)$$

and $A(z_1,z_2)$ is a symmetric factor of $X(z_1,z_2)$.

This is excluded by the theorem, so each non trivial irreducible factor $X_k(z_1,z_2)$ of $X(z_1,z_2)$ must be associated with a factor of $Y(z_1,z_2)$.

In a similar manner, each non-trivial irreducible factor $\tilde{Y}_k(z_1,z_2)$ of $Y(z_1,z_2)$ be associated with a factor of $X(z_1,z_2)$. If $Y(z_1,z_2)$ has no non-trivial symmetric factors, each irreducible factor of $Y(z_1,z_2)$ must be associated with $X(z_1,z_2)$. Therefore $X(z_1,z_2)$ and $Y(z_1,z_2)$ have an associated set of non-trivial irreducible factors and may only differ at most by a trivial factor.

i.e. $Y(z_1,z_2) = \beta\ z_1^{k_1}\ z_2^{k_2}\ X(z_1,z_2)$

If no constraints are made on the factors of $Y(z_1,z_2)$ note that each non-trivial irreducible factor of $X(z_1,z_2)$ must be associated with a factor of $Y(z_1,z_2)$.

Consequently $X(z_1,z_2)$ and $Y(z_1,z_2)$ are related by

$$Y(z_1,z_2) = z_1^{m_1}\ z_2^{m_2}\ P(z_1,z_2)\ X(z_1,z_2)$$

where

$m = m_1,m_2$ - integers

$P(z_1,z_2)$ is a polynomial in Z^{-1}.

Since $Y(z_1,z_2)$ and $X(z_1,z_2)$ are both polynomials in Z^{-1} and $X(z_1,z_2)$ contains no trivial factors $Q(z_1,z_2) = z_1^{m_1}\ z_2^{m_2}\ P(z_1,z_2)$ must also be a polynomial in Z^{-1}.

This represents the Z-transform of a sequence $q(n_1,n_2)$ which must be even in order for the phase or target phase of $x(n_1,n_2)$ and $y(n_1,n_2)$ to be equal. Therefore $Q(z_1,z_2) = \beta$ and the theorem follows.

APPENDIX III

Convergence of iterative algorithms

We state and prove convergence theorems for iterations of the form

$$x_{P+1} = x_P + \lambda(r_P)$$

where $r_P = T(x_P) - x_P$

$T(X_P)$ is the mapping between consecutive iterates of the basic (unrelaxed) iteration ie $X_{P+1} = T(X_P)$.

Definitions

(i) A mapping T with domain D in R^n and range in R^m will be denoted by $T: D \subset R^n \rightarrow R^m$

(ii) A mapping $T: D \subseteq R^n \rightarrow R^n$ is non expansive on a set D. D if

$$T_x - T_y < x-y \quad \text{for all } x,y \in D_o$$

(iii) A fixed point x^o of the iteration T is defined as any solution of the equation

$$x - Tx = 0.$$

Theorem A3.1

Suppose that under the Euclidean norm T: $D \subset R^n \to R^n$ is non expansive on the closed convex set $D_o \subset D$. Assume, further, that $TD_o \subset D_o$ and that D_o contains a fixed point of T. Then for any $\lambda \in (0,1)$ and $x_o \in D_o$, the iteration

$$x_{p+1} = \lambda x_p + (1-\lambda)\, Tx_p, \qquad p = 0,1,\ldots, \tag{A3.1}$$

converges to a fixed point of T in D_o

Proof

The convexity of D_o ensures that the sequence (A3.1) is well defined and remains in D_o. If x^o is a fixed point of G in D_o then in the Euclidean norm

$$\left\| x_{p+1} - x^o \right\|^2 = \lambda^2 \left\| x_p - x^o \right\|^2 + (1-\lambda)^2 \left\| Tx_p - x^o \right\|^2 + 2\omega(1-\omega)\, (Tx_p - x^o)^T (x_p - x^o) \tag{A3.2}$$

and

$$\left\| x_p - Tx_p \right\|^2 = \left\| (x_p - x^o) \right\|^2 + \left\| (Tx_p - x^o) \right\|^2 - 2(Tx_p - x^o)^T (xp - x^o) \tag{A3.3}$$

after multiplying (A3.3) by $\lambda(1-\lambda)$ and adding to (A3.2) we obtain:

$$\left\| x_{p+1} - x^o \right\|^2 + \lambda(1-\lambda) \left\| x_p - Tx_p \right\|^2 \tag{A3.4}$$

$$= \lambda \left\| x_p - x^o \right\|^2 + (1-\lambda) \left\| Tx_p - Tx^o \right\|^2 < \left\| x_p - x^o \right\|^2$$

Therefore for any m>o

$$\lambda(1-\lambda) \sum_{p=0}^{m} \left\| x_p - Tx_p \right\|^2 < \sum_{p=0}^{m} \left[\left\| x_p - x^o \right\|^2 - \left\| x_{p+1} - x^o \right\|^2 \right]$$

$$= \left\| x_o - x^o \right\|^2 - \left\| x_{m+1} - x^o \right\|^2$$

$$< \left\| x_o - x^o \right\|^2$$

Which proves that for m , the series on the left converges, and , in particular that

$$\lim_{p \to \infty} \left\| x_p - Tx_p \right\| = 0$$

Since

$$\left\| x_{p+1} - x^o \right\| = \left\| \lambda (x_p - x^o) \right\| + (1-\lambda) \left\| Tx_p - Tx^o \right\|$$

$$< \left\| (x_p - x^o) \right\| \leqslant \left\| x_j - x^o \right\| \leqslant \left\| x_o - x^o \right\| \qquad \text{(A3.5)}$$

for all $p > 0$, $j < p$

It follows that the sequence $\{x_p\}$ is bounded and hence has a convergent subsequence $\{x_{pi}\}$ which, by the closedness of D_o, must have its limit point y^o in D_o.
Then (A3.1) shows that

$$\lim_{i \to \infty} (x_{pi+1} - y^o) = \lim_{i \to \infty} (x_{pi} - y^o) + (1-\lambda) \lim_{i \to \infty} (Tx_{pi} - x_{pi}) = 0$$

or by the continuity of T, that $y^o = Ty^o$, therefore (A3.5) holds with y^o instead of x^o, and accordingly the whole sequence X_p must converge to the fixed point y^o.

Note that the adaptive acceleration has changed the point to which the iteration converges.

Appendix 3b

Lemma Contraction mapping theorem

Suppose that $T:D \subset R^n \rightarrow R^n$ maps a closed set $D_o \subset D$ into itself and that

$$\| T_x - T_y \| < \alpha \| x-y \| \quad \text{for all } x,y \in D_o \qquad \text{(A3b.1)}$$

for some $\alpha < 1$.

Then for any starting point $x_o \in D_o$, the sequence

$$x_{p+1} = Tx_p \qquad p = 0,1,\ldots, \qquad \text{(A3b.2)}$$

Converges to the unique fixed point X^o of T in D_o

and

$$\| x_p - x^o \| \leq [\alpha/(1-\alpha)] . \| x_p - x_{p-1} \| . \quad = 1,2\ldots, \qquad \text{(A3b.3)}$$

Note that (A3b.3 provides a computable error estimate so that if α is known, the actual error after p iterations can be bounded in terms of the last step $x_p - x_{p-1}$.

(b) The convergence of approximate or Noisy sequences.

Given that the exact sequence x converges under an iteration such as the phase only iteration, it can be shown that the approximate sequence such as may be formed by quantization of x also converges to the same fixed point if it is contained in a domain under which the iterative mapping is a contraction.

Theorem

Let $T:D \subset R^n \rightarrow R^n$ be a contraction on $D_1 \subset D$ (with constant α) and $D_o \subset D_1$ a closed set such that $TD_o \subset D_o$.

[By the contraction mapping theorem, the sequence in (A3b.2) starting from $x_o \subset D_o$ converges to the unique fixed point x^o of T in D_o]

Let $y_p \subset D_1$ be any sequence and define

$$\varepsilon_p = \left\| Ty_p - y_{p+1} \right\| , \; p = 0, 1, \ldots\ldots \qquad \text{(A3b.4)}$$

then

$$\left\| y_{p+1} - x^o \right\| \leq [1/(1-\alpha)] \left[\alpha \left\| y_{p+1} - y_p \right\| + \varepsilon \right] \qquad \text{(A3b.5)}$$

$$\left\| y_{p+1} - x^o \right\| \leq \left\| x_p - x^o \right\| + \sum_{j=0}^{p} \alpha^{p-j} \varepsilon_j + \alpha^{p+j} \left\| x_o - y_o \right\| , p = 0,1.. \qquad \text{(A3b.6)}$$

and

$$\lim_{p \to \infty} y_p = x^o \quad \text{IFF} \quad \lim_{p \to \infty} \varepsilon_p = 0$$

Proof

the estimate (A3b.5) follows from

$$\left\| y_{p+1} - x^o \right\| \leq \left\| y_{p+1} - Ty_p \right\| + \left\| Ty_p - Ty_{p+1} \right\| + \left\| Ty_{p+1} - Tx^o \right\|$$

$$< \varepsilon_p + \alpha \left\| y_p - y_{p+1} \right\| + \alpha \left\| y_{p+1} - x^o \right\|$$

and (A3b.6) is obtained from

$$\left\| x_{p+1} - y_{p+1} \right\| \leq \left\| Tx_p - Ty_p \right\| + \left\| Ty_p - y_{p+1} \right\|$$

$$\leq \alpha \left\| x_p - y_p \right\| + \varepsilon_p \leq \ldots \leq \sum_{j=0}^{p} \alpha_{p-j} \varepsilon_j + \alpha_{p-1} \left\| x_o - y_o \right\|$$

together with

$$\left\| y_{p+1} - x^{o} \right\| \leq \left\| y_{p+1} - x_{p+1} \right\| + \left\| x_{p+1} - x^{o} \right\|$$

Suppose that $\lim_{p \to \infty} \varepsilon_p = 0$ and, more specifically, that for

given $\varepsilon > 0$, $\varepsilon_p \leq \varepsilon_{p_o}$ for $p \geq p_o$. Then with

$\gamma_p = \sum_{j=0}^{p} \alpha_{p-j}\, \varepsilon_p$, we have

$$\gamma_p \leq \alpha_{p-po}\, \gamma_{po} + \sum_{j=p_o+1}^{p} \alpha_{p-j}\, \varepsilon_j \leq \alpha_{p-po}\, \gamma_{po} + \varepsilon\, [\alpha_{po+1}/(1-\alpha)]$$

which shows that $\lim_{p \to \infty} \gamma_p = 0$

thus by (A3b.6), $\lim_{p \to \infty} y_p = x^{o}$

Conversely, if $\lim_{p \to \infty} y_p = x^{o}$, then

$$0 \leq \varepsilon_p = \left\| Ty_p - yp+1 \right\| \leq \left\| Ty_p - Tx^{o} \right\| + \left\| x^{o} - y_{p+1} \right\|$$
$$\leq \alpha \left\| y_p - x^{o} \right\| + \left\| x^{o} - y_{p+1} \right\|,$$

which implies that $\lim_{p \to \infty} \varepsilon_p = 0$

No assumptions are made about the noisy sequence, except that it is contained in a domain $D_1 \subset D$ in which T is a contraction. It need not lie in D_o.

The relationship between the exact sequence and the approximate sequence is given by (A3b.6).

APPENDIX 4

Proof of equation 6.3.7.

The error between the latest estimate $f_i(x,y)$ and the original image $f(x,y)$ at the i^{th} iteration is related to the error at the $i+1^{th}$ iteration by

$$e_{i+1}(x,y) = e_i(x,y) + \lambda \Delta e_i(x,y) \qquad (A4.1)$$

where

$$\Delta e_i(x,y) = e_{i+1}(x-1,y) + e_{i+1}(x,y-1) + e_i(x,y+1) + e_i(x+1,y) - 4\, e_i(x,y)$$

The transfer function of the lowpass operator modelling the iteration is thus

$$H(\omega_1,\omega_2) = \frac{E_{i+1}(\omega_1,\omega_2)}{E_i(\omega_1,\omega_2)} = \frac{F\left\{e_i(x,y) + \lambda \Delta e_i(x,y)\right\}}{F\left\{e_i(x,y)\right\}} \qquad (A4.2)$$

$$= \frac{(1-4\lambda)\, F\left\{e_i(x,y)\right\} + \lambda\left[F\left\{e_{i+1}(x-1,y)\right\} + F\left\{e_{i+1}(x,y-1)\right\} + F\left\{e_i(x,y+1)\right\} + F\left\{e_i(x+1,y)\right\}\right]}{F\left\{e_i(x,y)\right\}}$$

Using the relationships .. $F\left\{e_{i+1}(x+1,y)\right\} = \exp(j\omega_1).F\left\{e_{i+1}(x,y)\right\}$

and $$H(\omega_1,\omega_2) = \frac{E_{i+1}(\omega_1,\omega_2)}{E_i(\omega_1,\omega_2)}$$

we get

$$H(\omega_1,\omega_2) = (1-4\lambda) + \lambda\left[H(\omega_1,\omega_2)\exp(-j\omega_1) + \exp(-j\omega_2) + \exp(j\omega_1) + \exp(j\omega_2)\right]$$

hence

$$H(\omega_1,\omega_2) = \frac{(1-4\lambda) + \lambda\left[\exp(j\omega_1) + \exp(j\omega_2)\right]}{1 - \lambda\left[\exp(-j\omega_1) + \exp(-j\omega_2)\right]} \qquad (A4.3)$$

REFERENCES

REFERENCES

[1] R.M. Mersereau, A.V. Oppenheim, "Digital reconstruction of multidimensional signals from their projections" Proc. IEEE, Vol. 62, No. 10, pp. 1319 - 1338, October 1974.

[2] D. Ludwig, "The Radon transform on Euclidean spaces", Comm Pure and applied Math, Vol XIX, pp. 49 - 81, 1966,.

[3] R.N. Bracewell and A.C. Riddle, "Inversion of fan bean scans in Radio astronomy", Astrophys J., No 150, pp. 427 - 434, 1966.

[5] B.P. Medoff, W.R. Brody and A. Macovski, "The use of a priori information in image reconstruction from limited data", ICASSP83, pp. 131 - 134, Boston 1983.

[6] G.S. Harrel et al, "Stop action cardiac computed tomography", Radiology, Vol. 123, pp. 515, 1977

[7] K.C. Tan and V. Perez-Mendez, "Tomographical imaging with limited-angle input", J. Opt. Soc. Am., Vol. 71, No. 5, May 1981

[8] T. Sato, et al., "Tomographic image reconstruction from limited projections using iterative revisions in image and transform spaces", Appl. Opt., Vol. 20, No. 3, Pp. 395, Feb. 1981.

[9] M.T. Sezan and H. STark, "Image restoration in CT by the method of projection onto convex sets", ICASSP 83 Boston, 1983.

[10] D.C. Youla, "Generalised image restoration by the method of alternating projections", IEEE Trans. Circuits and Systems, Vol. CAS - 25, PP. 694 - 702, 1978.

[11] T. Pavlidis, Structural pattern recognition, Springer-Verlag, New York 1977

[12] T. Pavlidis,"Algorthims for shape analysis of contours and waveforms", IEEE Trans. pattern Analy. Mach. Intell., PAMI-2, pp. 301 - 312, 1980.

[13] W.F. Screiber, C.F. Knapp, "TV Bandwidth reduction by digital coding", IRE National convention record, Vol. 6, Pt. 4, pp. 88 - 89, 1958.

[14] E.J. Delp et al, "Image data compression using autoregressive time series models", Pattern recognition, Vol. 11, pp. 313 - 323, 1979.

[15] M. Pavel, "Skeletal Categories", Pattern Recognition, Vol. 11, pp. 325 - 327, 1979.

[16] U. Grenander, "Foundations of pattern analysis", Quart. Appl. Maths, Vol. 27, pp. 1 - 55, 1969.

[17] G.L. Lemay and J.D. Dessimoz, "Recovery of gray scaled images from contour processed representations", ICASSP83, pp. 116 - 117.

[18] R. Bellman, "On the approximation of curves by line segments using dynamic programming", Comm. ACM4, 1961.

[19] W.K. Pratt, "Digital image processing", John Wiley, 1978.

[20] M. Aoki, "Restoration of binary image from boundary information", Proc. 6th intd. conf. on Pattern Recog., Munich 1982.

[21] D. Kermish, "Image reconstruction from phase information only", J. Opt. Soc. Am., Vol. 60, pp. 15 - 17, Jan. 1970.

[22] E.M. Hofstetter, "Construction of time-limited functions with specified autocorrelation function", IEEE Trans. Info. Theory, Vol. IT-18, pp. 119 ff, 1964.

[23] Yu M. Bruck and L.G. Sodin, "On the ambiguity of the image reconstruction problem", Optics Comm., Vol. 30, No. 3, Sept. 1979.

[24] A.M.J. Huiser and P. Van Toorn, "Ambiguity of the phase reconstruction problem", Optics letters, Vol. 5, pp. 499 - 501, Nov. 1980.

[25] M. Shaker Sabri and William Stenaart, "An approach to Bandlimited signal extrapolation", IEEE Trans. circl. syst., Vol. CAS-22(9), pp. 4 - 12, Sept. 1975.

[26] J.A. Cadzow, "An extrapolation procedure for bandlimited signals", IEEE Trans. Acoust. Speech Sig. Proc., Vol. ASSP-27, pp. 4, Feb. 1979.

[27] R.W. Gerchberg, "Supersolution through error energy reduction", Optica acta 21, pp. 709 - 720, 1974.

[28] A.V. Oppenheim and J.S. Lim, "The importance of phase in signals", Proc. IEEE, Vol. 69, pp. 529 - 541, May 1981.

[29] J.R. Fienup, "Phase retrieval algorithms: a comparison", Applied Optics, Vol. 21, pp. 2758 - 2769, August 1982.

[30] M. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform", IEEE Trans. Acoust, Speech and signal processing, Vol. ASSP-30, pp. 140 - 154, April 1982.

[31] C.D. Juglin, D.C. Hines, "The phase correlation image alignment method", Proc. 1975 int. Conf. cybernetics and society, pp. 163 - 165, Sept. 23 - 25, 1975.

[32] T.G. Stockman et al, "Blind deconvolution through digital signal processing", Proc. IEEE, vol. 63, pp. 679ff, 1975.

[33] B. Gold, A.V. Oppenheim and C.M. Radar, "Theory and implementation of the discrete Hilbert Transform", in Proc. Symp. Comput Process. Comm., Vol. 19, New York, Polytechnic Press 1970.

[34] J.M. Tribolet, "A new phase unwrapping algorithm", IEEE Trans. Acoust. Speech, signal processing, Vol. ASSP-25, April 1977.

[35] T.F. Quatieri and A.V. Oppenheim, " Iterative techniques for minimum phase signal reconstruction from phase or magnitude" IEEE Trans. Acoust. Speech, signal processing, Vol. ASSP-29, No. 6, pp. 1187ff, Dec. 1981.

[36] M.H. Hayes, J.S. Lim and A.V. Oppenheim, "Signal reconstruction from phase or magnitude", IEEE Trans. ASSP, Vol. ASSP-28, pp. 672 - 680, Dec. 1980.

[37] J. R. Fienup, "Reconstruction of an object from the modulus of its Fourier transform", Optics letters, Vol. 3, pp. 27 - 29, July 1978.

[38] A. Papoulis, "A New algorithm in Spectral analysis and Bandlimited extrapolation", IEEE Trans. Cir. and Syst., pp. 735 - 742, Sept. 1975.

[39] R.W. Gerchberg and W.O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures", Optik 35, pp. 237 - 246, 1972.

[40] G.N. Ramachandrar and R. Srinivasan, "Fourier Methods in crystallography, Wiley interscience, New York, 1970.

[41] A.J.J. Drenth , A.M.J. Huiser and H.A. Ferweda, "The problem of phase retrieval in light and electron microscopy of strong objects, Optica Acta 22, pp. 615ff, 1975.

[42] M. Ryle, "Radio telescopes of large resolving power", Rev. Mod. Phys., Vol. 47, pp. 557 - 566, July 1975.

[43] J. R. Fienup, "Space object imaging through the turbulent atmosphere", Optical engineering, Vol 18, pp. 529ff, Sept. 1979.

[44] R.N. Bracewell, "Computer image processing", Annu. Rev. Aston Astrophysics, Vol. 17, pp. 113 - 114, 1979.

[45] A. Walther, "The question of phase retrieval in optics", Optica Acta 10, pp. 41 - 49, 1963.

[46] A.H. Greenway, "Proposal for phase recovery from a single intensity distribution", Optics letters, Vol. 1, pp. 10 - 12, 1977.

[47] B.J. Hoenders, "On the solution to the phase retrieval problem", J. Maths Phys, Vol. 16, pp. 1719, Sept. 1975.

[48] R.H.T. Bates, "On phase problems I", Opti, Vol. 51, pp. 161 - 179, 1978.

[49] Pieter Eyekoff, "System identification - Parameter and state estimation", John Wiley and Sons, New York, 1974.

[50] I.S. Konvalinka, M.R. Matransek, "Simultaneous estimation of poles and zeros in speech analysis and ITIE - iterative inverse filtering algorithm", IEEE Trans. Acoust. Speech, Signal Proc., Vol. ASSP-27, No. 5, Oct. 1979.

[51] J.M. Ortega and W. Rheinbolt, Iterative solution of nonlinear equations in several variables", Academic press, New York, 1970.

[52] R.S. Schafer, R.M. Mersereau, M.A. Richards, "Constrained iterative reconstruction algorithms", Proc. IEEE, Vol. 69, No. 4, pp. 432 - 450, April 1980.

[53] E. Mosca, "On a class of ill-posed estimation problems and a related gradient iteration", IEEE Trans. Auto. Contr., Vol. AC-17, No. 4, pp. 459 - 465, Aug. 1972.

[54] D. Youla, Generalised image restoration by the method of alternating orthogonal projections", IEEE Trans, Circuits, Systs, Bol. CAS-25, No. 9, pp. 694 - 702, Sept. 1978.

[55] A.K. Jain and S. Ranganath, "Extrapolation algorithms for discrete signals with applications in spectral estimation", IEEE Trans. Acoust. Speech, Signal Proc, Vol. ASSP-29, No. 4, pp. 830 - 845, August 1981.

[56] R.H.T. Bates, "Uniquness of solutions to two-dimensional Fourier phase problems for localised and positive images", Computer vision, graphics and image processing 25, pp. 205 - 217, 1984.

[57] R.E. Burge, et al, "The phase problem", Proc. Roy. Soc., London A350, 191, 1976.

[58] R.E. Burge et al, "The application of dispersion relationships to phase retrieval" J. Phys. D. Appli. phy., Vol. pp. L61 - L68, 1974.

[59] W. Lawton, "A numerical algorithm for 2-D wavefront reconstruction from intensity measurements in a single phase", SPIE Vol. 231, 1980 Int. Optic Comput. Conf., pp. 94ff.

[60] R.A. Gonzalves, "Phase retrieval and diversity in adaptive optics", Optical Engineering, Vol. 21, No. 5, pp. 829 - 832, 1982.

[61] S.C. Pohlig, "Fourier transform phase coding of images", IEEE Trans. Acoust. Speech, and Sig. Proc., Vol. ASSP-28, pp. 339 - 341, June 1980.

[62] W.A. Pearlman, R.M. Gray, "Source coding of the discrete Fourier transform", IEEE Trans. information theory, vol. IT-24, pp. 683 - 692, Nov. 1968.

[63] C. Espy and J.S. Lim, "Effects of noise on signal reconstruction from Fourier transform phase", ICASSP82, pp. 1833 - 1836.

[64] R.A. King A.J. Singarajah, S.A. Kwabwe, "A hybrid technique to restore the Fourier phase and magnitude spectra of noisy linearly degraded images", ICASSP83, Boston.

[65] B. Yegnarayana, A. Dhayalan, "Non iterative techniques for minimum phase signal reconstruction from phase or magnitude", ICASSP83, Boston, pp. 639 - 641.

[66] Naveed Malik, "1 and 2-D maximum entropy spectral estimation ", M.I.T. Sc. D. thesis Nov. 1981.

[67] V.T. Tom et al, "Convergence of iterative non-expansive signal reconstruction algorithms, IEEE Trans Acoust. Speech and Signal Proc., Vol. ASSP-29, pp. 1052 - 1058.

[68] B.R. Musicus, "Iterative algorithms for optimal signal reconstruction and parameter identification given noisy and incomplete data", Ph.D. thesis, M.I.T. 1982.

[69] H.C. Andrews and A.G. Tescher, "The role of adaptive phase coding in two and three dimensional Fourier and Walsh Image Compression", Proceedings of the Walsh function symposium, Washington DC, March 1974.

[70] M. Tasto and P.A. Wintz, "Image coding by adaptive block Quantization", IEEE Trans. Comm. Tech., Vol. Com-19 pp. 956 - 972, Dec. 1971.

[71] H.C. Andrews, B.R. Hunt, "Digital image restoration", Prentice Hall inc., Englewood Cliffs, New Jersey, 1977.

[72] J.S. Lim, "Image restoration by short space spectral spectral subtraction", IEEE Trans. Acoust. Speech and Signal Proc., Vol. ASSP-28, No. 2, April 1980.

[73] D.J. Granath, "The role of Human Visual models in image processing", Proc. IEEE, Vol. 16, pp. 552 - 562, May 1981.

[74] H.C. Andrews, W.K. Pratt, "Fourier transform coding of images", Hawaii international conference on system science, pp. 677 - 679, Jan. 1968.

[75] G.B. Anderson and T.S. Huang, " Fourier transformation for picture bandwidth compression", IEEE Trans. Comm. Com-20, pp. 488 - 491, June 1972.

[76] H.C. Andrews and W.K. Pratt, "Transform image coding" Proceeding computer processing in communications, Polytechnic, pp. 63 - 84, 1969.

[77] A. Habibi and P.A. Wintz, "Image coding by linear transformation and block quantization", IEEE Trans. Comm Tech, Vol. COM-19, No. 1, pp. 50 - 63, Feb. 1971.

[78] K.R. Rao, M.A. Narasimhan, K. Revuluri, "Image data processing by Hadamard-Haar transforms, IEEE Trans. Comput., Vol. C-23, No. 9, pp. 888 - 896, Sept. 1975.

[79] L.S. Taylor, "The phase retrieval problem", IEEE Trans Ant and Prop., Vol. AP-29, pp. 386 - 391, March 1981.

[80] M. Tasto and P.A. Wintz, "A bound on the rate-distortion function and application of images", IEEE Trans. Inf. Theory, IT-18, pp. 100 - 159, Jan. 1972.

[81] J.F. Hayes, A. Habibi and P.A. Wintz, "Rate distortion function for a Gaussian source model of images", IEEE Trans. Inf. Theory, IT-16, pp. 567 - 608, July 1970.

[82] R.M. Haralick, K. Shanmugan and I. Dinstein, "Texture features for image classification", IEEE Trans. of Syst. Man and Cyber., SMC-3, pp. 610 - 621, Nov. 1973.

[83] H. Freeman, "On the encoding of arbitrary geometric configurations", IRE Trans. on Electronic Computers, Vol. EC-10, No.2, pp. 260 - 268, 1961.

[84] P.E. Danielsson, "An improved segmentation and coding algorithm for binary and non-binary images, IBM journal, Vol. 26, No. 6, pp. 698 - 707, 1982.

[85] J.K. Hawkins, "Textural properties for pattern recognition" in picture processing and psychopictorics, B.S. Lipkin and A. Rosenfeld eds, Academic press, pp. 347 - 370, N.Y. 1970.

[86] G.J. Lendaris and G.L. Stanley, "Diffraction pattern sampling for automatic pattern recognition", Proc. IEEE,58, pp. 198 - 216, 2 February 1970.

[87] W.K. Pratt, O.D. Faugeras and A. Gagolowicz, "Visual discrimation of stochastic texture fields", IEEE Trans. Syst. Man, Cybern, Nov. 1978.

[88] A. Gagalowicz, "Stochastic texture fields synthesis from a priori given second order statistics", Proc. IEEE Computer Society Conf. Patter Recognition and Image Proc., Chicago, pp. 376 - 381, Aug. 6 - 8 1979.

[89] O.D. Faugeras and W.K. Pratt, " Decorrelation method of texture feature extraction", IEEE Trans. Patt. Anal. Mach, inte., Vol. PAMI-2, pp. 323 - 332, July 1980.

[90] K.A. Stevens, "Surface perception from local analysis of texture and contours", MIT, AI Labs. AI-TR 512, 1980.

[91] L.G. Roberts, "Machine perception of three dimensional solids", in Optical and Electrooptical information processing, pp. 159 - 197, (J.T. Tippett et al eds.)

[92] R.O. Duda and P.E. Hart, Pattern Classification and Scene analysis, John Wiley, 1973.

[93] D. Marr and E. Hildreth, "Theory of edge detection", Proc. Roy. Soc. B, Vol. 207, pp. 187 - 217, 1980.

[94] R. Kirsh, "Computer determination of constituent structure of Biological images", Computer and Biomed. Res., Vol. 4, No. 3, pp. 315 - 328, 1971.

[95] A. Rosenfeld, "Edge and curve detection for visual scene analysis", IEEE Trans. Computers, C-20, No. 5, pp. 562 - 569, May 1971.

[96] J.M.S. Prewitt, "Object enhancement and extraction", in Picture processing and Pyschopictorics, B.S. Lipkin and A. Rosenfeld, eds, Academic Press, New York,pp. 111ff, 1970.

[97] L.S. Davis and A. Mitchie, "Edge detection in textures", Computer graphics and Image Processing, Vol-12, pp. 25 - 39, 1980.

[98] A.S. Willisky, "Digital signal processing and control and estimation theory", The MIR Press, 1979.

[99] W.K. Pratt, "Generalized Wiener filtering computation techniques", IEEE Trans. Comp. C-21, pp. 636 - 641, 1972.

[100] C.W. Helstrom, "Image restoration by the method of least squares", J. Opt. Soc. Am., Vol. 57, No. 3, pp. 297 - 303, March 1967.

[101] N. Ahmed and K.R. Rao, "Orthogonal transfomrms for digital signal processing", Springer-Verlag, 1975.

[102] A.K. Jain, "A Sinusoidal family of unitary transforms", IEEE Trans. Patter anal. Mach. Intel., Vol. PAMI-1, pp. 356 - 365, Oct. 1979.

[103] N. Ahmed, T. Natarajan and K.R. Rao, "Discrete Cosine transform", IEEE Trans. Comp., Vol. C-23, pp. 90 - 93, Jan. 1974.

[104] G. Ross et al., "A Solution to the phase problem based on the theory of entire functions", OPTIK, Vol. 49, pp. 71 - 80, 1977.

[105] J.R. Fienup, "Comments" on the reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform, IEEE Trans. on Acoust., Speech and Signal Processing, Vol ASSP-31, No. 3, June 1983.

[106] B. Julesz, "Visual pattern discrimation", IRE Trans. Inform. Theory, Vol. IT-8, pp. 84 - 92, February 1962.

[107] T. S. Huang, "The importance of phase in image processing filters", IEEE Trans. on Acoust., Speech and Signal Proc., vol. ASSP-23, No. 5, Dec. 1975.