RICE UNIVERSITY

Spatial and Temporal Image Prediction with Magnitude and Phase Representations

by

Gang Hua

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Approved, Thesis Committee:

M. T. Orchard

Michael T. Orchard, Chair, Professor
Electrical and Computer Engineering

Richard G. Baraniuk, Victor E. Cameron
Professor
Electrical and Computer Engineering

Wotao Yin, Assistant Professor
Computational and Applied Mathematics

Houston, Texas

December, 2010

ABSTRACT

Spatial and Temporal Image Prediction with Magnitude and Phase Representations

by

Gang Hua

In this dissertation, I develop the theory and techniques for spatial and temporal image prediction with the magnitude and phase representation of the Complex Wavelet Transform (CWT) or the over-complete DCT to solve the problems of image inpainting and motion compensated inter-picture prediction.

First, I develop the theory and algorithms of image reconstruction from the analytic magnitude or phase of the CWT. I prove the conditions under which a signal is uniquely specified by its analytic magnitude or phase, propose iterative algorithms for the reconstruction of a signal from its analytic CWT magnitude or phase, and analyze the convergence of the proposed algorithms. Image reconstruction from the magnitude and pseudo-phase of the over-complete DCT is also discussed and demonstrated.

Second, I propose simple geometrical models of the CWT magnitude and phase to describe edges and structured textures and develop a spatial image prediction (inpainting) algorithm based on those models and the iterative image reconstruction mentioned above. Piecewise smooth signals, structured textures and their mixtures can be predicted successfully with the proposed algorithm. Simulation results show that the proposed algorithm achieves appealing visual quality with low computational complexity.

Finally, I propose a novel temporal (inter-picture) image predictor for hybrid video coding. The proposed predictor enables successful predictive coding during fades, blended scenes, temporally decorrelated noise, and many other temporal evolutions that are beyond the capability of the traditional motion compensated prediction methods. The proposed predictor estimates the transform magnitude and phase of the desired motion compensated prediction by exploiting the temporal and spatial correlations of the transform coefficients. For the ease of implementation in standard hybrid video codecs, the over-complete DCT is chosen over the CWT. Better coding performance is achieved with the state-of-the-art H.264/AVC video encoder equipped with the proposed predictor. The proposed predictor is also successfully applied to image registration.

Acknowledgments

My sincere thanks go to my advisor, Dr. Michael Orchard, for his guidance, encouragement and generous sharing of his deep insights; to Dr. Richard Baraniuk and Dr. Wotao Yin for being on my thesis committee.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this dissertation, I solve the problems of inpainting and motion compensated inter-picture prediction with novel spatial and temporal image prediction techniques that benefit from the new perspective provided by the magnitude and phase representations of the Complex Wavelet Transform (CWT) and the over-complete DCT. Under those magnitude and phase representations, the formulation and modeling of complicated spatial and temporal image evolutions with the presence of edges, patterned textures, linear temporal distortion, and structured temporal interference are greatly simplified. First, I develop the theory and algorithms about image reconstruction from those magnitude and phase. Then, I propose simple models under those representations for describing the spatial and temporal image evolutions involved in the targeted inpainting and inter-picture prediction problems. Finally, I apply the developed image reconstruction algorithms to construct the desired prediction results.

1.1 Magnitude and Phase Representations for Image Prediction

The magnitude and phase of the CWT or the over-complete DCT are local measurements of signal energy amplitude and location respectively. For example, around an edge, the magnitude indicates the edge sharpness within a nearby region with non-zero signal energy; the phase is only significant in that region with non-zero magnitude and represents the information about the exact edge location. It follows that the magnitude is a smooth function along edges with uniform sharpness and the phase is close to linear around edges and within patterned textures under certain conditions. For spatial prediction, therefore, simple 2D geometrical models and 2D linear functions can closely approximate the CWT magnitude along edges and the CWT phase within patterned textures respectively. Those simple (linear) models while applied on the nonlinear magnitude and phase result in simple, fast, and nonlinear spatial predictors. For temporal (inter-picture) prediction, important information for rejecting temporal interference and correcting temporal distortion can be easily inferred from the involved spatial neighborhoods of the magnitude and phase. Equipped with the inferred information, a simple linear predictor in the CWT or over-complete DCT domain achieves successful nonlinear temporal prediction.

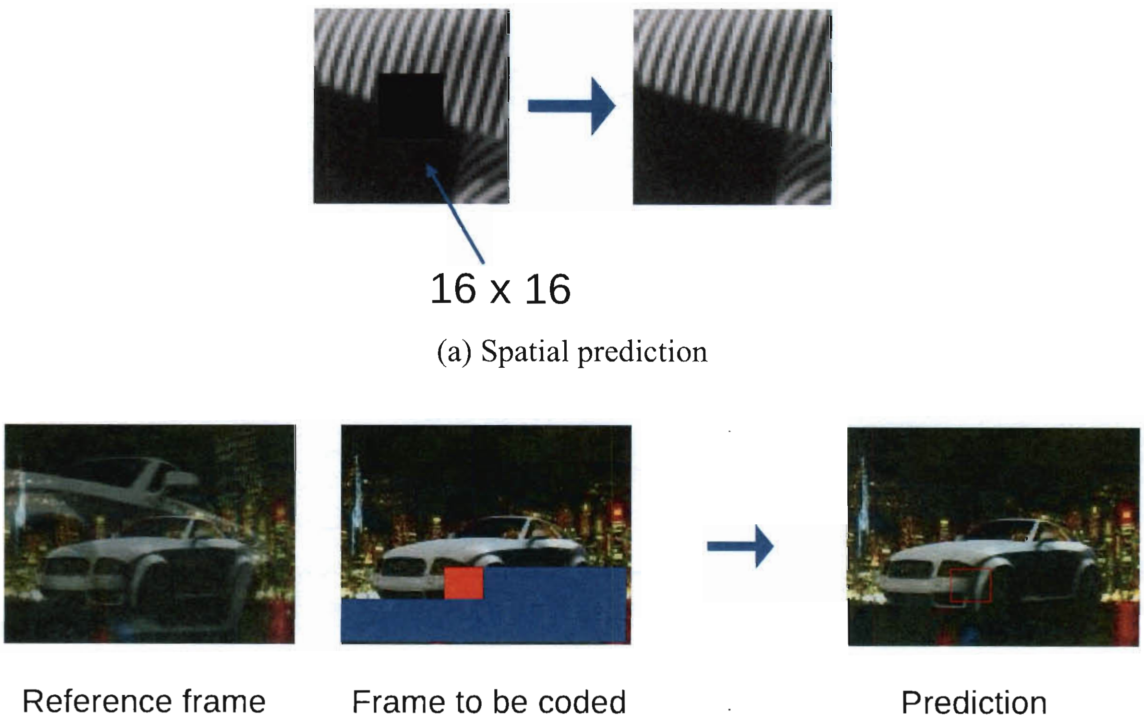
The magnitude and phase of the CWT or the over-complete DCT represent an image in "dual" ways in the sense that they both can represent an image alone, being complimentary measurements of local signal amplitude and location. It can be shown mathematically that an image is unique given its CWT magnitude or phase under certain conditions and the image can be reconstructed from the CWT magnitude or phase with some iterative algorithms. Similarly, the magnitude and pseudo-phase of the over-complete DCT can also reconstruct an image alone, although no mathematical proof is available. Therefore, an image can be predicted successfully if either its magnitude or phase is predicted correctly with the simple models mentioned above. It has to be noted though that, whenever possible, drawing information from both magnitude and phase is beneficial and desirable for the targeted image prediction problems.

Moreover, those magnitude and phase representations are known to match closely how the human visual system encodes visual information. There are psycho-physical and physiological experiments and findings showing that the human visual system is sensitive to localized 2D spatial phases and suggesting that the human visual system might be using localized 2D phases for encoding visual information internally ([1, 2, 3]). The CWT provides such a type of localized 2D spatial phase via the wavelet and filter design techniques and the CWT magnitude encodes image information in a "dual" way to the CWT phase. Therefore, performing prediction properly with those magnitude and phase may give results with high visual quality. The over-complete DCT and may be employed as an alternative when the CWT is not applicable (e.g., video encoding with block based coding and motion compensation scheme).

1.2 The Spatial and Temporal Image Prediction Problems

The spatial image prediction considered in this dissertation is the problem of predicting a certain (missing) region in an image from the neighboring known regions (Figure 1.1 (a)). The problem is also known as image inpainting [4, 5, 6] and it may occur under the situation of recovery of image and video from damages as well as spatial predictive image and video coding. As shown by the example in Figure 1.1, to predict the missing region successfully, smooth functions, edges and patterned textures all have to be interpolated correctly from the neighboring known regions.

Most existing inpainting works fall into the following 3 categories. First, diffusion based methods formulate inpainting as a variational problem in the pixel domain and solve it with partial differential equations [7, 8, 4, 9, 5, 10]. These methods propagate the pixel



(a) Spatial prediction

(b) Temporal prediction

Figure 1.1 : The considered spatial and temporal image prediction problems: (a) The spatial prediction recovers the missing 16x16 pixels from the available neighboring pixels (note that the missing block contains both edges and textures); (b) Temporal prediction estimates a 16x16 pixels block from the past frame image and the available causal blocks in the current frame (note that the past frame is a blended scene).

values in the surrounding regions into the missing region. Second, texture synthesis and examplar-based methods propagates the image information from know regions into the missing region at the patch level in the pixel domain [11, 12, 13]. Third, sparse coding based methods define inpainting as a non-linear minimization problem seeking a sparse solution under some fixed or learned dictionaries [14, 15, 16, 17]. Different from the existing methods, the proposed method in this dissertation interpolates the CWT magnitude and phase in the missing region from the magnitude and phase in the surrounding regions respectively.

The temporal prediction problem considered in this dissertation is aimed at the motion compensated inter-picture prediction for hybrid video coding (Figure 1.1 (b)). Traditional motion compensated prediction relies on translated blocks (may be low-pass filtered) from reference frames to directly match blocks in the frame to be coded. However, during fades, focus change, blended scenes, temporally decorrelated noise, and many other temporal evolutions, the traditional predictors will fail in finding a reasonably good temporal prediction from the reference frames, because the reference frames contain both a prediction relevant part and significant interference. The relevant part in a reference frame may be distorted (linearly scaled or filtered) from the frame to be coded and the interference may be noise, clutter, or another blending image. Therefore, to achieve successful inter-picture prediction, the temporal interference has to be rejected and the distortion in the relevant part has to be corrected. In this dissertation, I am interested in the temporal prediction problem under the above difficult conditions. This dissertation proposes a new inter-picture prediction method which estimates the over-complete DCT magnitude and pseudo-phase of the desired temporal prediction by exploiting their temporal and spatial correlations. This temporal prediction problem for motion compensated prediction is new and there is no directly related previous result reported.

1.3 The Proposed Work of the Dissertation

The spatial and temporal image prediction problems described above are very challenging interpolation or extrapolation problems, when the involved image signals exhibit some complicated patterns and evolutions in the spatial and temporal domain as shown in the

Figure 1.1. To successfully address these difficult problems, this dissertation exploits one simple idea: complicated spatial and temporal image patterns and evolutions become easy to model and predict under the magnitude and phase representations of the CWT or the over-complete DCT. Simple and straightforward modeling techniques while applied properly under those representations achieve high quality prediction results.

In this dissertation, I consider the spatial and temporal prediction problems by exploring the new perspective of magnitude and phase image representations.

First, to develop understanding and insights about this new perspective, I investigate the theory and algorithms about image reconstruction from the analytic magnitude and phase provided by the CWT. The conditions under which a signal is uniquely specified by its analytic magnitude or phase are presented. Iterative algorithms for reconstructing an image from its analytic magnitude or phase are proposed and the convergence of the proposed algorithms is analyzed. The extension of the results about the analytic magnitude and phase to the CWT and the over-complete DCT is also discussed and demonstrated.

Second, for the spatial prediction problem, I propose a simple inpainting method following the iterative algorithms for image reconstruction from the CWT magnitude and phase. The proposed method predicts the CWT magnitude and phase in the missing region band by band in the order of decreasing band energy with simple geometrical models (2D directional model for magnitude around edges and 2D linear model for phase within structured textures). Then, the proposed method constructs an estimate of the missing region through iterated image reconstruction from the predicted CWT magnitude and/or phase. Different from existing methods in the aforementioned 3 categories, the proposed method only estimate a few simple linear model parameters under the nonlinear magnitude and phase representation and does not need to solve complicated nonlinear optimization problems. Simulation results show that the proposed algorithm achieves appealing visual quality and competitive PSNR with low computational complexity.

Finally, for the temporal prediction problem, I propose a new inter-picture prediction technique that looks at the over-complete DCT magnitude and pseudo-phase of the reference frame and the frame to be coded, infers the information about the temporal interference and distortion from the spatial neighborhoods, and constructs a simple linear temporal filter in the over-complete DCT domain to reject the interference and correct the distortion. Better coding performance is achieved with the state-of-the-art H.264/AVC video coder equipped with the proposed predictor. The proposed temporal predictor is also successfully applied for image registration.

1.4 Organization of this Dissertation

This dissertation is organized as follows. Chapter 2 reviews the background information about a few related forms of magnitude and phase image representations for image prediction: the Fourier spectral representation, the analytic representation, the CWT, and the over-complete DCT. Chapter 3 develops the theory and algorithms for image reconstruction from the analytic magnitude or phase of the CWT. Chapter 4 and 5 describe the proposed spatial and temporal image prediction algorithms respectively. Chapter 6 concludes the dissertation.

Chapter 2

Image Representations with Magnitude and Phase

Appropriate representation and modeling of edge energy and location are critical to the successful prediction of the spatial and temporal image evolutions discussed in the previous chapter. For example, image inpainting around an edge or within patterned textures is essentially a task of predicting or modeling the spatial location of the edge or the regularly patterned group of edges in the missing image region. Temporal prediction consists of separating the energy of relevant and irrelevant image features (mostly edges), and learning and inverting of the temporal distortion (basically the sharpness and brightness changes of edges).

The magnitude and phase representation of the CWT is a suitable tool for describing and modeling edge energy and location. Edge energy and location are represented hierarchically by the CWT magnitude and phase respectively in an un-aliased, localized and multi-resolution fashion. The energy of edges and patterned textures is decomposed into different frequency and orientation bands almost alias-free. Within each band, the magnitude is a smooth function along edges with uniform sharpness and the phase is close to linear around edges and within patterned textures under certain conditions. Roughly speaking, the magnitude indicates the vicinity and sharpness of the edge, and the phase contains detailed information about edge location and edge profile. For the spatial and temporal image prediction problems, those magnitude and phase are suitable for local spatial interpolation and local modeling of temporal interference and distortion.

The mathematical basis of the magnitude and phase representation of the CWT is the analytic representation of real signals. Later in this dissertation, the theory and algorithms about image reconstruction from the analytic magnitude and phase of the CWT will be investigated and applied. The analytic representation is rooted in the Fourier spectral representation which is also a form of magnitude and phase representation of real signals, and image reconstruction from the Fourier magnitude and phase has already been researched extensively. This chapter first reviews the background information about these two basic types of magnitude and phase representations: the Fourier spectral representation and the analytic representation (including the CWT). Then, the over-complete DCT and other related representations are also briefly reviewed. The idea of modeling and estimating local image features with the magnitude and phase representations for the image prediction problems is also motivated and illustrated.

2.1 Introduction

A complex number c can be represented by its magnitude $|c|$ and phase $\angle c$, where $|c|$ measures the energy of c and $\angle c$ indicates the relative energy of the real and imaginary parts. Similarly, a real-valued signal can be represented by the magnitude and phase of a complex-valued function linearly derived from that real signal. Generally, the magnitude in those representations is some type of signal energy measurement and the phase contains detailed information about the signal structure.

The Fourier spectral representation and the analytic representation are probably the most widely used forms of magnitude and phase representations for real signals. The Fourier spectral magnitude and phase measure the strength and relative shift of global sinusoidal signal components respectively. It is well known that the Fourier phase carries important information about edge locations and the theory of image reconstruction from the Fourier magnitude or phase has been extensively researched. Those results on the Fourier representation will be reviewed briefly in Section 2.2, since they are deeply related to the analytic representation used in the CWT.

The analytic representation of a real signal is formed by discarding all the negative frequency components of the real signal (without loss of any information due to the spectral symmetry of real signals). The analytic magnitude indicates local signal energy amplitude and phase contains detailed information about signal structure. The CWT represents an image with a set of un-aliased, multi-resolution and localized analytic magnitude and phase that are suitable for modeling and predicting local image features within a small spatial and temporal neighborhood. The over-complete DCT represents an image with a similar set of DCT magnitude and pseudo-phase, which is a special case of magnitude and phase representation with the imaginary part of the complex-valued function being zero. The analytic representation and the CWT will be introduced in Section 2.3. The over-complete DCT and other related representations will be discussed in Section 2.4 and 2.5 .

2.2 Image Representation with the Fourier Magnitude and Phase

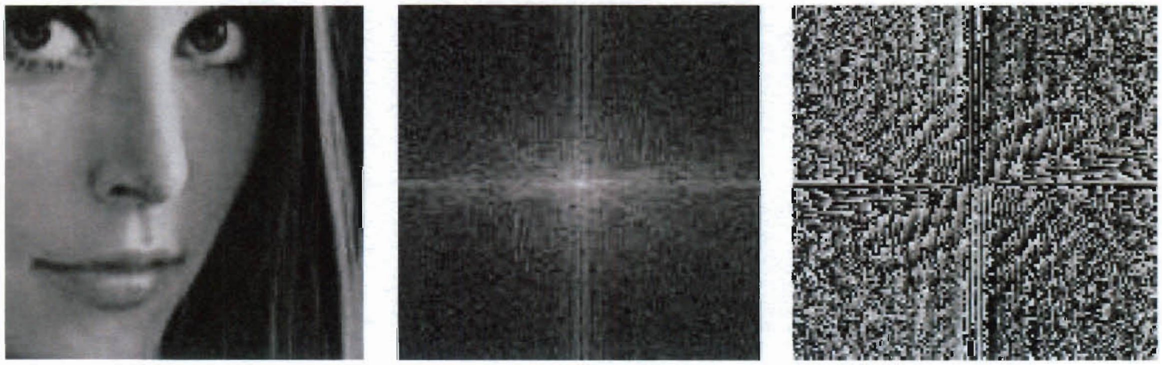
The Fourier transform provides a way of representing an image with its spectral magnitude and phase. In the discrete case, the Discrete Time Fourier Transform (DTFT) $X(\omega) \in \mathbb{C}$ of a signal $x \in \mathbb{R}^N$ is given below:

$$X(\omega) = \sum_{n=0}^{N-1} x(n) e^{-j\,\omega n}$$

The Fourier magnitude $\rho_x(\omega) = |X(\omega)|$ and phase $\theta_x(\omega) = \angle X(\omega)$ measure the strength and relative shift of the global sinusoidal signal components $e^{j\,\omega n} = \cos(\omega n) + \boldsymbol{j}\,\sin(\omega n)$ respectively. The extension of the Fourier transform to 2D (and higher dimensions) is straightforward.

$$X(u, v) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} x(n, m) e^{-j\,(un+vm)}$$

The Fourier magnitude and phase of a given image (a 128 by 128 block cropped from the standard Lena image) are shown in Figure 2.1. The importance of the Fourier phase for signal and image representation has been investigated in [18]. The authors demonstrated that most of the information about the edges is contained in the Fourier phase, but not in the Fourier magnitude by an experiment similar to Figure 2.2.



(a) an image

(b) the Fourier magnitude

(c) the Fourier phase

Figure 2.1 : The Fourier magnitude (in log scale and darker colors show smaller values) and phase (in linear scale from $-\pi$ to π and dark colors show smaller values) of an image

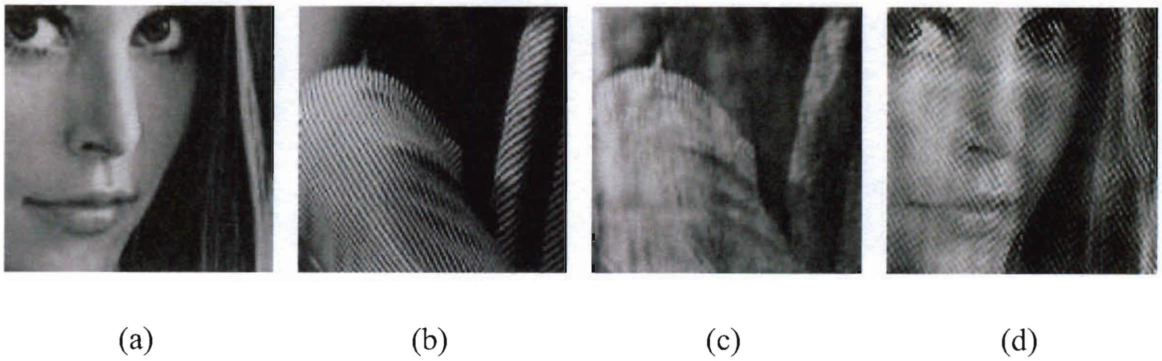


Figure 2.2 : The Fourier magnitude and phase experiment: (a) a block from Lena; (b) a block from Barbara; (c) inverse Fourier transform with Lena's magnitude and Barbara's phase (looks like Barbara); (d) inverse Fourier transform with Lena's phase and Barbara's magnitude (looks like Lena).

Furthermore, the authors of [18] proved that a sequence is unique given its Fourier phase under the following conditions and proposed iterative algorithms for image reconstruction from the Fourier phase.

Theorem 1 (The uniqueness given Fourier phase [18]). *A sequence which is known to be zero outside the interval* $0 \leq n \leq N-1$ *is uniquely specified to within a scale factor by* $(N-1)$ *distinct samples of its Fourier phase (or tangent of its phase) in the interval* $0 < \omega < \pi$, *if the sequence has a z-transform with no zeros on the unit circle or in conjugate reciprocal pairs.*

The uniqueness of a sequence with specified Fourier magnitude was considered in [19] and the following theorem was proposed.

Theorem 2 (The uniqueness given Fourier magnitude [19]). *A sequence which is known to be zero outside the interval* $0 \leq n \leq N-1$ *is uniquely specified to within a sign and/or a time shift by* $(N-1)$ *distinct samples of its Fourier magnitude in the interval* $0 < \omega < \pi$, *if it has a z-transform with all zeros outside (or inside) the unit circle.*

Given the above uniqueness conditions, a signal can be predicted successfully, as long as its Fourier magnitude or phase can be estimated correctly.

The above theorems about the Fourier magnitude and phase were extended to 2D and higher dimensions in [20, 21]. The image reconstruction results from the Fourier magnitude or phase are shown in Figure 2.3. It can be seen that the reconstruction result from only the magnitude is so poor that it cannot be used in real life applications like image and video coding. It has been shown in [21] that to obtain good reconstruction quality, the signs of all the Fourier phase variables may be required. The signal representation and reconstruction from the short time Fourier transform were also considered in [22, 23, 24].

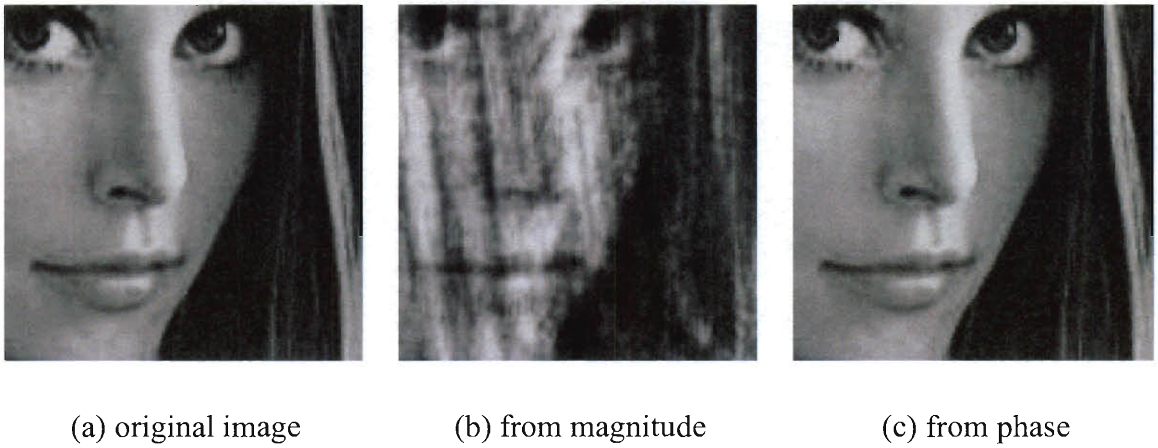
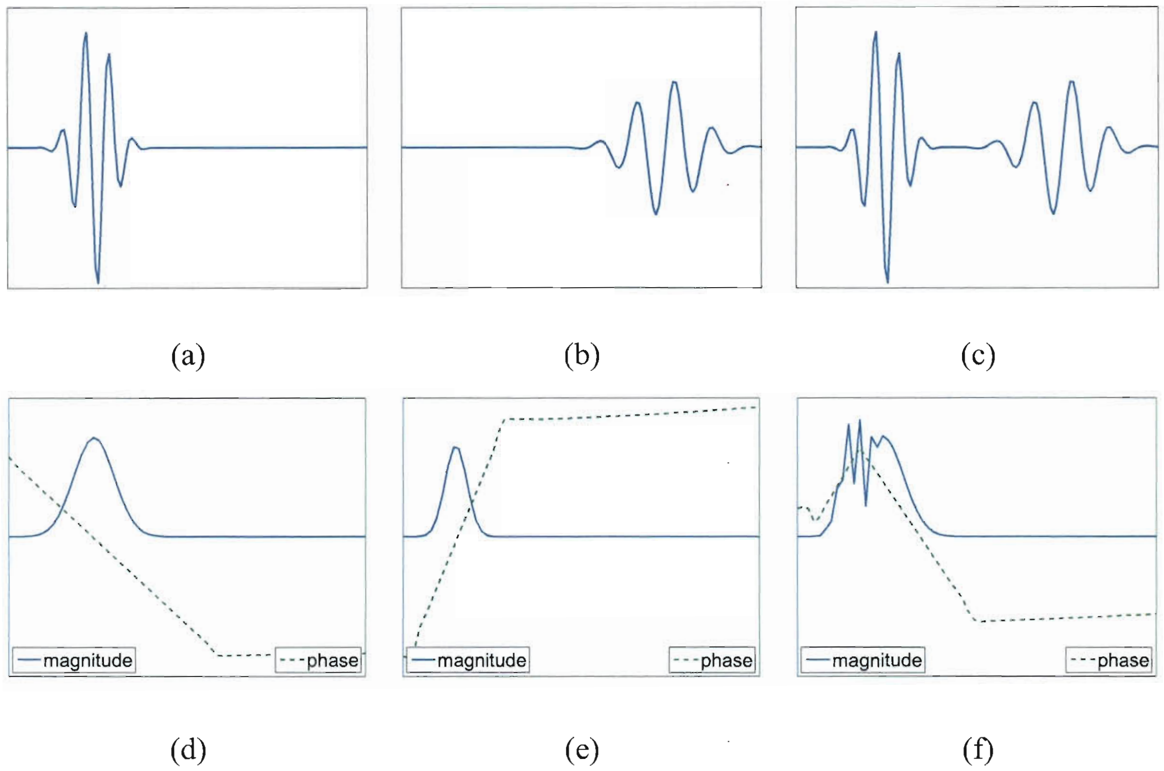

(a) original image (b) from magnitude (c) from phase

Figure 2.3 : The reconstruction results from the Fourier magnitude or phase: (a) the original image; (b) image reconstructed from magnitude (PSNR = 20.17dB); (c) image reconstructed from phase (perfect reconstruction).

For the inpainting and motion compensated inter-picture prediction problems discussed in the previous chapter, image prediction has to be done within a small local neighborhood. However, predicting local image features with the Fourier magnitude and phase is a very difficult task, because the Fourier magnitude and phase are global measurements of sinusoidal signal components. Any significant change in one local area of a signal re-


(a) (b) (c)

(d) (e) (f)

Figure 2.4 : The influence of local features on the Fourier magnitude and phase: (a) signal one; (b) signal two; (c) the sum of the two signals; (d, e, f) the Fourier magnitude and phase of (a, b, c), respectively.

sults in significant changes to almost all the Fourier magnitudes and phases. This effect is demonstrated in Figure 2.4. The Fourier magnitude and phase of two localized signals in (a) and (b) are shown in (d) and (e) respectively. The third signal (c) which is the sum of the two localized signals has Fourier magnitude and phase much more complicated than the two localized signal components. In reality, a signal or image is composed of a lot of local features and its Fourier magnitude and phase become very complex (e.g., Figure 2.1 (b) and (c)). Therefore, modeling and estimation of local image features with the Fourier magnitude and phase are difficult.

2.3 Image Representation with the Analytic Magnitude and Phase

The analytic magnitude and phase are suitable for representing oscillatory signals and have been widely used for characterizing band-pass signals and systems in communication theory [25]. The magnitude and phase indicate the envelop and oscillation of a signal respectively. The CWT uses the analytic magnitude and phase for representing the oscillatory real wavelet coefficients.

2.3.1 Analytic Magnitude and Phase

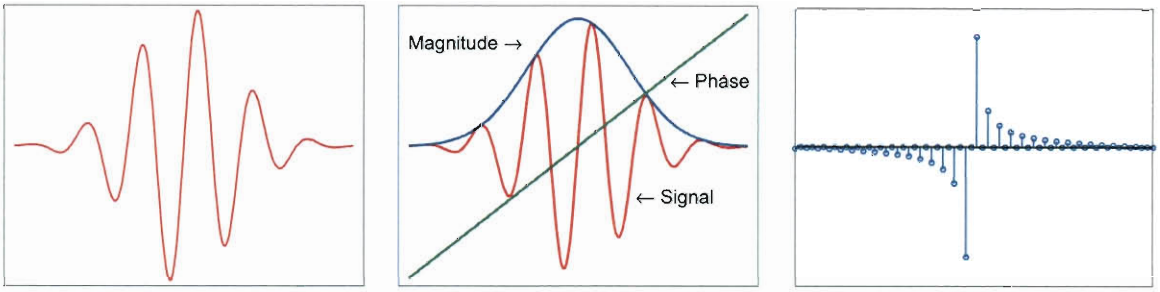
A real-valued signal $x(n)$ can be viewed as the real part of a complex-valued analytic signal $c(n) = x(n) + \boldsymbol{j}\, \widehat{x}(n)$, where $\widehat{x}(n)$ denotes the Hilbert transform of $x(n)$. The analytic magnitude and phase of $x(n)$ are $|c(n)|$ and $\angle c(n)$ respectively.

In the Fourier frequency domain,

$$C(\omega) = X(\omega)\; F_a(\omega) = \begin{cases} 2S(\omega), & \omega > 0 \\ S(0), & \omega = 0 \\ 0, & \omega < 0 \end{cases}$$

where $F_a(\omega)$ is the frequency response of an ideal analytic filter $f_a(n)$ which suppresses all the negative frequencies ($F_a(\omega) = 0$ if and only if $\omega < 0$). In the time domain, the impulse response of the ideal analytic filter is $f_a(n) = \delta(n) + \boldsymbol{j}\, h(n)$, where $h(n)$ is the impulse response of the Hilbert transform. Since the Fourier transform of real signals has the complex conjugate symmetry, removing the negative frequency components does not loose any information about $x(n)$.

The analytic magnitude and phase of an oscillatory signal and the impulse response of the Hilbert transform (approximately implemented with the Fast Fourier Transform (FFT)) are shown in Figure 2.5 [1]. As shown in the figure, the magnitude is the envelop of the signal and the phase indicates the oscillation cycle. In general, the analytic magnitude and phase reveal fundamental signal characteristics [26]: the magnitude defines the amplitude of local sinusoid and the phase relates to small local shifts.



(a) (b) (c)

Figure 2.5 : The analytic magnitude and phase representation approximately implemented with the FFT: (a) one real oscillatory signal; (b) its analytic magnitude and phase; (c) the impulse response of the Hilbert transform.

The analytic representation can be easily extended to 2D and higher dimensions by following [26, 27, 28]. For the 2D case, half of the 2D frequency plane can be discarded by analytic filters due to the symmetry of the 2D Fourier transform of real signals. In this dissertation, I use two complex signals $c_1(n, m)$ and $c_2(n, m)$ with non-zero response in the first and fourth quadrant of the 2D Fourier frequency plane respectively to represent a 2D real signal $x(n, m)$. The frequency response of the two corresponding 2D analytic

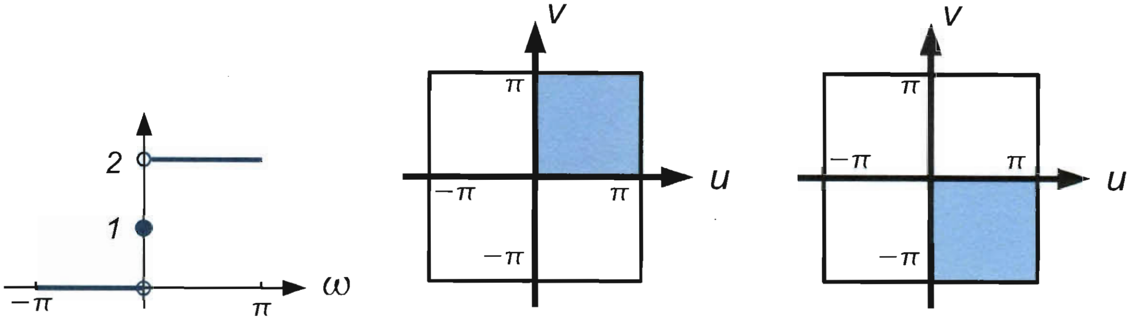
[1] In real applications, the analytic filter is typically designed to have compact support to achieve time and spatial localization.

filters ($F_1(u, v)$ and $F_2(u, v)$) have the following properties:

$$F_1(u, v) \neq 0, \quad \text{if and only if } u \geq 0 \text{ and } v \geq 0$$

$$F_2(u, v) \neq 0, \quad \text{if and only if } u \geq 0 \text{ and } v \leq 0$$

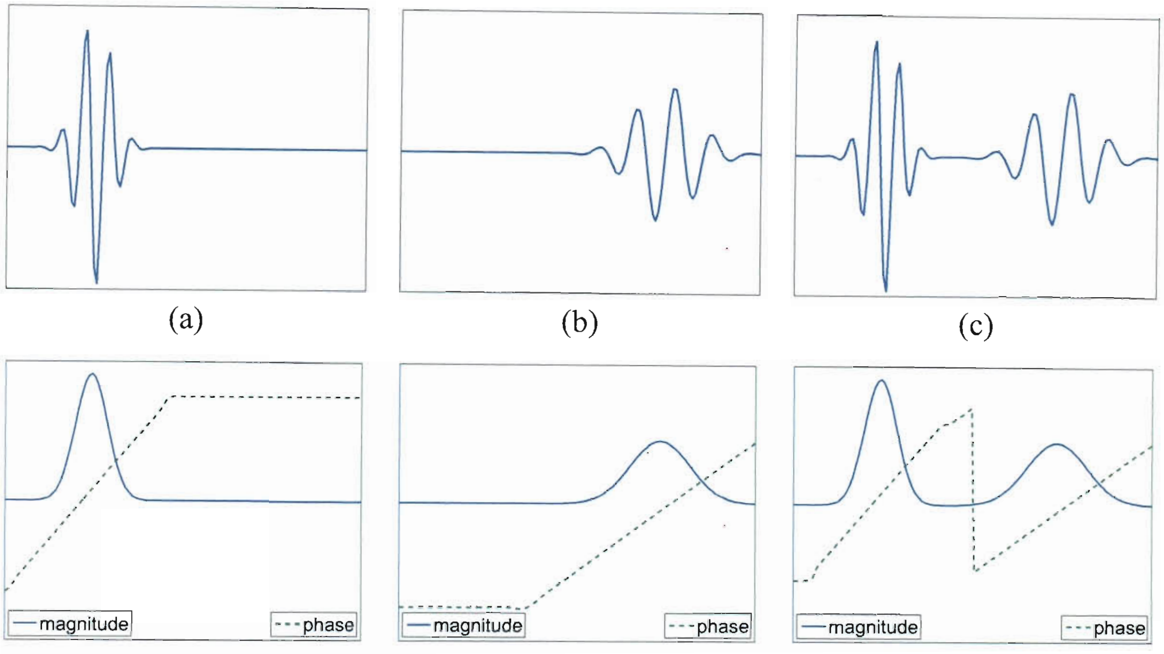
The ideal frequency responses of 1D and 2D analytic filters are depicted in Figure 2.6.



(a) $F_a(\omega)$ (b) $F_1(u, v)$ (c) $F_2(u, v)$

Figure 2.6 : The frequency response of 1D and 2D analytic filters: (a) ideal 1D analytic filter; (b, c) ideal 2D analytic filters.

The analytic magnitude and phase of the signals in Figure 2.4 are shown in Figure 2.7. The original signal structure can be clearly seen in the analytic magnitude and phase: the magnitude indicates where the oscillatory signal components are located and the phase shows the speed of signal oscillation. In contrast to the Fourier case, the analytic representation is so well localized that the two spatially separated signal components do not interfere with the analytic magnitude and phase of each other, which is important for modeling and predicting local image features.



(a) (b) (c)

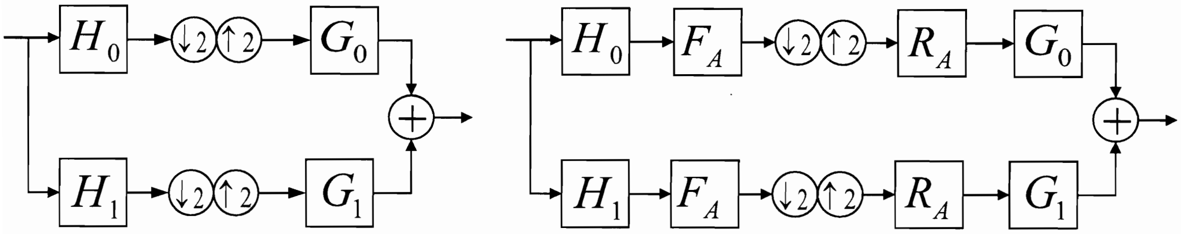
(d) (e) (f)

Figure 2.7 : The influence of local features on the analytic magnitude and phase: (a) signal one; (b) signal two; (c) the sum of the two signals; (d,e,f) the Fourier magnitude and phase of (a,b,c).

2.3.2 The Complex Wavelet Transform

The Complex Wavelet Transform (CWT) provides a multi-scale set of analytic magnitude and phase for representing signals. For example, the dual-tree complex wavelet transform [30, 31, 32] represents an image with a redundant collection of complex coefficients generated by a bank of bandpass, analytic filters. The filters are selected so that, with redundancy of two in each direction, each band offers an un-aliased representation of signal components in a particular frequency range. Following the quaternionic Fourier transform formulated in [26], other complex/quaternion wavelet transform variants have also been proposed in [27, 33] for image coding and in [34, 35] for edges geometry and image disparity estimation.

As shown in Figure 2.8, the CWT can be considered as the analytic representation of some real wavelet coefficients, where H_0 and H_1 are the wavelet analysis filters, G_0 and G_1 are the wavelet synthesis filters, and F_A and R_A are the analytic filter and its inverse respectively [36, 29].



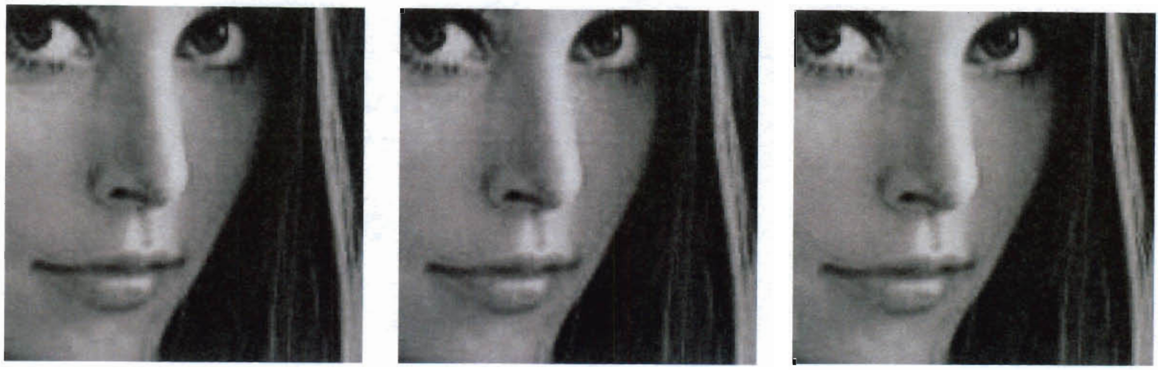
(a) The real wavelet

(b) The complex wavelet

Figure 2.8 : The filter bank structure of the real and complex wavelet.

The CWT inherits the advantages of the wavelet transform for image processing applications and introduces the new nonlinear magnitude and phase representation for images. The CWT magnitude has been exploited in some image denoising algorithms [37, 38, 39, 28] and the CWT phase has been used for edges geometry and image disparity estimation [34, 35, 40, 41]. However, this magnitude and phase representation still has not been thoroughly researched and its applications are very limited.

In my earlier work on inpainting [36], I observed the following interesting properties about the analytic magnitude and phase of the CWT. An image can usually be reconstructed perfectly or with very high visual quality from only its CWT magnitude or phase with POCS (Project Onto Convex Sets) type of iterative algorithms. Typically, the image reconstructed from the CWT phase is perfect. Sometimes, the image reconstructed from the CWT magnitude has very high visual quality, but has some very subtle position shift in



(a) original image (b) from magnitude (c) from phase

Figure 2.9 : The reconstruction results from the CWT magnitude or phase: (a) the original image; (b) image reconstructed from magnitude (PSNR = 44.10dB); (c) image reconstructed from phase (perfect reconstruction).

some local areas. In contrast to the Fourier spectral representation, successful reconstruction from the CWT magnitude without additional phase information is usually observed (please see Figure 2.9 and 2.3 for comparison).

In the next chapter, I will develop the conditions under which a signal is unique given its analytic magnitude or phase, and then extend the results to the CWT magnitude and phase. The algorithms for reconstructing an image from its CWT magnitude or phase will also be investigated. Those results will be applied to solve the spatial and temporal image prediction problems later in chapter 4 and 5.

2.4 The Over-complete DCT

The over-complete DCT is a translation invariant version of the block based DCT widely used in image and video coding. In 1D, a K-point over-complete DCT is a tight frame consisting of all the K possible K-point block based orthogonal DCT. Let $H_i \in \mathbb{R}^{N\times N}$ be the orthogonal transform of the i-th block based DCT ($0 \leq i < K$) of K points, then the K-

point over-complete DCT forward transform is $H = [H_0^T, H_1^T, \ldots, H_K^T]^T$. Therefore, the K-point 1D over-complete DCT is K times redundant. The inverse over-complete DCT is $G = \frac{1}{K}[H_0^T, H_1^T, \ldots, H_K^T]$, which is the average of all the K orthogonal block based IDCT. Equivalently, the over-complete DCT can also be considered as a K-channel un-decimated filter bank. The extension to 2D is straightforward and the $K \times K$ 2D over-complete DCT is K^2 times redundant.

Similar to the CWT, the over-complete DCT also offers an un-aliased representation of signal components in several (K for 1D and K^2 for 2D) frequency ranges. In addition, I observed that the magnitude and pseudo-phase (also called DCT-phase by some authors [42]) of the over-complete DCT can also be used to reconstruct an image. As shown in Figure 2.10, an image can also be reconstructed very well from the 4×4 over-complete DCT magnitude. The reconstruction from the pseudo-phase has low PSNR, but all the image details are recovered and only the local signal energy distribution is wrong. With some side information about the local signal energy, the image can also be correctly reconstructed from the over-complete DCT pseudo-phase.

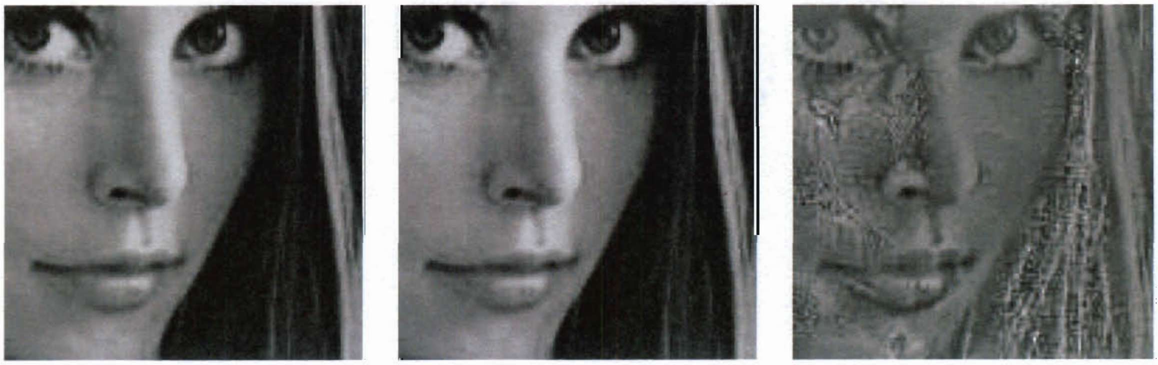


Figure 2.10 : The reconstruction results from the 4×4 over-complete DCT magnitude or pseudo-phase: (left) the original image; (middle) image reconstructed from magnitude (PSNR = 45.07dB); (right) image reconstructed from pseudo-phase (PSNR = 15.39dB).

The over-complete DCT is preferred over the CWT for solving the temporal prediction problem for video coding, because most standard video encoders employ the block based coding and motion compensated prediction scheme. Therefore, in this dissertation, the over-complete DCT will be employed for the temporal image prediction problem in Chapter 5.

2.5 Other Related Representations

Image representations with Gabor-like localized phase was proposed and analyzed in [43, 3, 44]. These schemes are partly motivated by the research on the biological representation of visual information at the level of the visual cortex. In [33], the author attempted in several ways to construct image representations separating local signal energy and local signal structure (similar to the localized magnitude and phase) for image coding. The spherical coder proposed in [33] may be considered as treating the absolute values of wavelet coefficients as magnitude and the signs as phase. Then, the absolute values of wavelet coefficients in each band are again encoded by the total energy of the band and a set of phase variables indicating the spatial energy distribution. The uniqueness of a signal with specified magnitude of "general" real or complex frame coefficients are considered in [45]. However, the "general" frames cannot be localized in time, frequency or scale in any way.

2.6 Summary

This chapter reviewed the background information about several types of magnitude and phase representations: the Fourier spectral representation, the analytic representation, the

CWT, and the over-complete DCT. In general, those magnitude and phase contain important information about edge energy and location and can be used for modeling spatial and temporal image evolutions discussed in the previous chapter. In addition, an image can be reconstructed accurately given only those magnitude or phase under certain conditions.

In the next chapter, I develop the theory of image reconstruction from the analytic magnitude and phase of the CWT. I relate the analytic magnitude and phase to the Fourier magnitude and phase, develop the conditions under which a signal is uniquely specified by its analytic magnitude and phase, and discuss the extension to multi-resolution and higher dimensions to match the situation of the CWT. The over-complete DCT is also discussed and it may be employed when the application of the CWT is not allowed (e.g., for the temporal prediction problem for video coding).

Chapter 3

Image Reconstruction from the CWT Magnitude or Phase

In this chapter, I consider image reconstruction from the analytic CWT magnitude or phase. As discussed in the previous chapter, the theory of image reconstruction from the Fourier spectral magnitude or phase has been extensively researched. It is well known that under certain conditions, a signal is uniquely specified by its Fourier magnitude or phase and may be reconstructed with some iterative algorithms [18, 19, 20, 21, 46]. However, image reconstruction from the analytic CWT magnitude or phase used in this dissertation is still not well understood. This chapter develops the following fundamental results: (1) the conditions under which a signal is unique given its CWT magnitude or phase; (2) the algorithms for reconstructing an image from its CWT magnitude or phase.

This chapter is organized as follows. Section 3.1 develops the conditions under which a 1D signal is uniquely specified by its analytic magnitude or phase. Section 3.2 proposes iterative algorithms for reconstructing a signal from its analytic magnitude or phase and shows the results about the convergence of the proposed algorithms. Section 3.3 discusses the extension of the uniqueness conditions and reconstruction algorithms to the multi-resolution 2D CWT. The section also provides some insights about the difference between magnitude and phase, the quality of the reconstructed image, and the geometrical structure of the magnitude or phase representation. Section 3.4 discusses image reconstruction from the magnitude or pseudo-phase of the over-complete DCT. To solve the temporal image prediction problem for video coding, the DCT is preferred over the CWT because most standard video encoders employ the block based DCT and motion compensated prediction scheme. Section 3.5 is a short summary of this chapter. In Section 3.6, some basic properties of the analytic representation and the CWT are listed.

3.1 The Uniqueness in Terms of Analytic Magnitude or Phase

In this dissertation, for a signal $x \in \mathbb{C}^N$ with magnitude $\rho = |x| \in \mathrm{I\!R}^N$ and phase $\theta = \angle x \in \mathrm{I\!R}^N$, I use a simple notation of $\rho e^{j\,\theta}$ $(= x)$ to indicate the magnitude and phase form of x even if $N > 1$. It is clear from the context that x, ρ, and θ are all vectors.

3.1.1 Uniqueness Given 1D Analytic Magnitude or Phase

In this subsection, I consider the uniqueness of a 1D discrete signal $x \in \mathrm{I\!R}^N$ given its analytic magnitude or phase. Suppose a discrete analytic filter $F \in \mathbb{C}^{N \times N}$ is used to construct the analytic magnitude $\rho(x) = |Fx|$ and phase $\theta(x) = \angle(Fx)$. If circular boundary extension is used, F is circulant and diagonalizable by the DFT matrix W:

$$\rho(x)e^{j\,\theta(x)} = Fx = W^{-1}\Lambda_F W x = W^{-1}(\Lambda_F W x)$$

where the diagonal matrix Λ_F contains the frequency response of the analytic filter. Therefore, the analytic magnitude and phase are the Fourier magnitude and phase of the filtered spectrum sequence $\Lambda_F W x$ (please ignore the difference between W and W^{-1}). Note that there is a very important difference from the normal Fourier magnitude and phase of real time domain signal: the filtered spectrum $\Lambda_F W x$ is always single sided by the definition of the analytic filter.

After recognizing the connection to the Fourier transform, we know the theory for Fourier magnitude and phase [18, 19, 20, 21] applies to the sequence $\Lambda_F W x$ instead of x for analytic magnitude and phase. However, the global uniqueness conditions for Fourier magnitude differ significantly from the observation in the previous chapter: reconstruction from the CWT magnitude is much more common than the reconstruction from Fourier magnitude. Therefore, in this section, I use a new method to develop the local uniqueness conditions for the analytic magnitude and phase. With the new method, I show that the analytic magnitude and phase have about the same uniqueness conditions. That is, if a signal is unique given is analytic phase, it is also unique (locally) given its analytic magnitude.

In order to be general for both the redundant transforms and critically down-sampled transforms, I assume that x is a bandpass signal living in a known frequency band of $[K_1, K_2]$ $(0 \leq K_1 \leq K_2 < \frac{N}{2})$. Again for generality, I only assume that the analytic filter $F = A + \boldsymbol{j}\, B$ $(A, B \in \mathbb{R}^{N \times N})$ is circulant and suppresses all the negative frequency components.

The local uniqueness of magnitude $\rho(x) = |Fx|$ and analytic phase $\theta(x) = \angle(Fx)$ can be determined by the Jacobians of $\theta(x)$ and $\rho(x)$ with respect to x, namely $J_{\rho(x)}$ and $J_{\theta(x)}$. Following the geometrical definition of the Jacobians, we have $\Delta\rho = J_{\rho(x)}\Delta x$, $\Delta\theta = J_{\theta(x)}\Delta x$, and $\Delta\rho + \boldsymbol{j}\, D_{\rho(x)}\Delta\theta = D_{e^{-\boldsymbol{j}\,\theta(x)}} F \Delta x$ for all $\Delta x \in \mathbb{R}^N$, where D_x denotes the diagonal matrix with vector x on its main diagonal. Let $J_{c(x)} = J_{\rho(x)} + \boldsymbol{j}\, D_{\rho(x)} J_{\theta(x)}$ and cancel the Δx terms on both sides, we have the following proposition which gives a simple relationship between $J_{c(x)}$ and the analytic filter matrix F.

Proposition 3.1 (The Jacobians of the magnitude and phase). *Let $F \in \mathbb{C}^{N \times N}$ be an an-*

alytic filter and the magnitude $\rho(x)$ and phase $\theta(x)$ of a signal $x \in \mathrm{I\!R}^N$ are defined by $\rho(x)e^{\boldsymbol{j}\,\theta(x)} = Fx$ (where $\rho(x) \in \mathrm{I\!R}^N_+, \theta(x) \in \mathrm{I\!R}^N$), then

$$
\begin{aligned}
J_{c(x)} &= J_{\rho(x)} + \boldsymbol{j}\; D_{\rho(x)} J_{\theta(x)} \\
&= D_{e^{-j\,\theta}}\, F \\
J_{\rho(x)} &= \Re\{D_{e^{-j\,\theta}}\, F\} \\
J_{\theta(x)} &= D^{-1}_{\rho(x)} \Im\{D_{e^{-j\,\theta}}\, F\}
\end{aligned}
$$

where $\Re\{v\}$ and $\Im\{v\}$ denote the real and imaginary parts of a complex vector v respectively, and D_v denotes the diagonal matrix with v on its main diagonal.

Note that, when an element of $\rho(x)$ is zero, the corresponding element of $\theta(x)$ can be arbitrarily defined. Since we are mostly interested in $D_{\rho(x)}\Delta\theta(x)$ rather than $\Delta\theta(x)$, zeros elements in $\rho(x)$ do not cause any real definition problem.

Alternatively, $J_{\rho(x)}$ and $J_{\theta(x)}$ can be derived as in the following proposition which is another format of the proposition above.

Proposition 3.2 (The Jacobians of the magnitude and phase). *Let $F = A + \boldsymbol{j}\; B$ be an analytic filter with $A, B \in \mathrm{I\!R}^{N\times N}$. The the magnitude $\rho(x)$ and phase $\theta(x)$ of x are defined by $\rho(x)e^{\boldsymbol{j}\,\theta(x)} = Fx$ (where $\rho(x) \in \mathrm{I\!R}^N_+, \theta(x) \in \mathrm{I\!R}$). The Jacobians of $\theta(x)$ and $\rho(x)$ are*

$$
\begin{aligned}
J_{\rho(x)} &= D^{-1}_{\rho(x)}\; (D_{Ax}A + D_{Bx}B) \\
J_{\theta(x)} &= D^{-2}_{\rho(x)}\; (D_{Ax}B - D_{Bx}A)
\end{aligned}
$$

if and only if $\rho(x)$ has no zero elements, where D_v denotes the square matrix with vector v on its diagonal.

Proof: The analytic phase $\theta(x)$ and magnitude $\rho(x)$ of the transform coefficients of a signal $x \in \mathrm{I\!R}^N$ are given below

$$\theta_k(x) = \arctan\left(B_k x, A_k x\right)$$
$$\rho_k(x) = \sqrt{(A_k x)^2 + (B_k x)^2}$$

where

$$\arctan(y, x) = \begin{cases} \arctan(\frac{y}{x}), & \text{if } x > 0 \\ \arctan(\frac{y}{x}) + \pi, & \text{if } x < 0 \\ \frac{\pi}{2}\,\mathrm{sgn}(y), & \text{if } x = 0 \end{cases}$$

Note that the following partial derivatives exist and are continuous except at $x = y = 0$.

$$\frac{\partial \mathrm{arctan}(y, x)}{\partial x} = \frac{-y}{x^2 + y^2}$$
$$\frac{\partial \mathrm{arctan}(y, x)}{\partial y} = \frac{x}{x^2 + y^2}$$

Then the partial derivatives of $\rho(x)$ and $\theta(x)$ are

$$\frac{\partial \theta_k(x)}{\partial x_l} = \frac{-B_k x}{(A_k x)^2 + (B_k x)^2} A_{k,l} + \frac{A_k x}{(A_k x)^2 + (B_k x)^2} B_{k,l}$$
$$\frac{\partial \rho_k(x)}{\partial x_l} = \frac{A_k x}{\sqrt{(A_k x)^2 + (B_k x)^2}} A_{k,l} + \frac{B_k x}{\sqrt{(A_k x)^2 + (B_k x)^2}} B_{k,l}$$

In matrix form, we have

$$J_{\rho(x)} = D^{-1}_{\rho(x)} \left(D_{Ax} A + D_{Bx} B\right)$$

$$J_{\theta(x)} = D^{-2}_{\rho(x)} \left(D_{Ax} B - D_{Bx} A\right)$$

From the above, we have that the Jacobians exist, if and only if $\rho(x)$ has no zero element (*i.e.*, $\rho(x) > 0$), However, following Proposition 3.1, this non-zero magnitude requirement is not really necessary. $\square$

The uniqueness given the magnitude or phase can be determined by examining the rank of the Jacobians ($J_{\theta(x)}$ and $J_{\rho(x)}$), since they specify the tangent plane of $\theta(x)$ and $\rho(x)$ at x. Given the frequency band $[K_1, K_2]$ and $\theta(x)$ or $\rho(x)$, if for all non-zero $v \in \mathbb{R}^N$ in band $[K_1, K_2]$, $J_{\rho(x)}v \neq 0$ or $J_{\theta(x)}v \neq 0$, then x is locally unique within a small neighborhood of x. For phase, local uniqueness implies global uniqueness, because if $\theta(x) = \theta(y)$, then $\theta(ax + by) = \theta(x)$ for all $a, b \geq 0$ and $a + b = 1$.

Theorem 3 (Uniqueness given 1D analytic magnitude or phase). *Suppose $F \in \mathbb{C}^{N\times N}$ is a circulant analytic filter which removes all the negative frequencies and $W \in \mathbb{C}^{N\times N}$ is the DFT transform matrix. Let $S_x(z)$ be the z transform of WFx and $S_x(z)$ is know to be zero outside the range between z^{-K_1} and z^{-K_2} ($0 \leq K_1 \leq K_2 < \frac{N}{2}$).*

$$S_x(z) = \sum_{k=K_1}^{K_2} a_k z^{-k}$$

If and only if at least one of a_{K_1} and a_{K_2} is non-zero and $S_x(z)$ has no zeros on the unit circle or in complex conjugate reciprocal pairs, within frequency band $[K_1, K_2]$, (1) given the analytic phase $\theta(x)$, x is globally unique up to a scale factor ; (2) given the analytic magnitude $\rho(x)$, x is locally unique when $a_0 \neq 0$ or locally unique up to a phase shift when $a_0 = 0$.

Proof: First, following Proposition 3.1, we have $\Delta\rho = J_{\rho(x)}\Delta x$, $\Delta\theta = J_{\theta(x)}\Delta x$, and

$\Delta\rho + \boldsymbol{j}\, D_{\rho(x)}\Delta\theta = D_{e^{-j\,\theta(x)}}F\Delta x$ for all $\Delta x \in \mathbb{R}^N$. Let $v = \Delta x$, we have

$$\begin{aligned}(Fx)^* \odot (Fv) &= (D_{\rho(x)}D_{e^{-j\,\theta(x)}})(Fv)\\ &= D_{\rho(x)}(\Delta\rho + \boldsymbol{j}\, D_{\rho(x)}\Delta\theta)\\ &= D_{\rho(x)}(J_{\rho(x)} + \boldsymbol{j}\, D_{\rho(x)}J_{\theta(x)})v \qquad (3.1)\end{aligned}$$

where $\odot$ and $*$ denote element-wise multiplication and complex conjugation respectively [1]. Therefore, $J_{\theta(x)}v = 0$ or $J_{\rho(x)}v = 0$ is equivalent to $(Fx)^* \odot (Fv)$ being pure real or pure imaginary respectively.

Second, let the z transform of WFv be $S_v(z)$ (in the same way as $S_x(z)$), then the z transform of $W((Fx)^* \odot (Fv))$ is the polynomial $S(z) = S_x^*(1/z^*)S_v(z)$. Therefore, $(Fx)^* \odot (Fv)$ being pure real or pure imaginary is equivalent to $S(z)$ has complex conjugate symmetry or anti-symmetry about z^0 respectively. According to the following lemma, the zeros of $S_x^*(1/z^*)S_v(z)$ are on the unit circle or in complex conjugate reciprocal pairs.

Lemma: Generalized Complex Conjugate Symmetry. *A FIR sequence $S(z)$ has generalized complex conjugate symmetry (*i.e., $S^*(1/z^*) = e^{j\,\alpha}z^M S(z)$ *for some $\alpha \in \mathbb{R}$ and $M \in \mathbb{Z}$), if and only if the zeros of $S(z)$ are on the unit circle or in complex conjugate reciprocal pairs.*

Finally, if and only if $S_x(z)$ have no zeros on the unit circle or in complex conjugate reciprocal pairs, the symmetry of $S(z)$ about z^0 requires that $S_v(z) = re^{j\,\phi}S_x(z)$ ($r, \phi \in \mathbb{R}$). Combined with the fact $x, v \in \mathbb{R}^N$, we conclude that, for $v \neq 0$ in band $[K_1, K_2]$, (1) $J_{\theta(x)}v = 0$ if and only if $v = rx$; (2) $J_{\rho(x)}v = 0$ if and only if $a_0 = 0$ and $v = r\widehat{x}$

[1] Note that the $D_\rho^{-1}(x)$ terms on in $J_{\rho(x)}$ and $J_{\theta(x)}$ are all canceled out in the last line.

(where $\widehat{x} = \Im\{Fx\}$ is the Hilbert transform of x). It is easy to show that if $a_0 = 0$ and $y = x\cos\alpha + \widehat{x}\sin\alpha$ (i.e., $S_y(z) = e^{j\,\alpha}S_x(z)$) for any $\alpha \in \mathrm{I\!R}$, $\rho(y) = \rho(x)$ and $\theta(y) = \theta(x) + \alpha$.

Therefore, under the conditions stated in the theorem, x is globally unique up to a scale factor given $\theta(x)$ and x is locally unique up to at most a phase shift given $\rho(x)$.

□

In plain language, a signal is locally unique up to an analytic phase shift given its analytic magnitude and is globally unique up to a linear scaling factor given its analytic phase, if the analytically filtered spectrum ($\Lambda_F Wx$) is not symmetric and the information about the frequency band $[K_1, K_2]$ of the signal is known. Note that the uniqueness conditions are about the same for magnitude and phase.

3.1.2 Signal Space Geometry Specified by the Analytic Magnitude or Phase

The uniqueness theorem reveals that the signal space geometry specified by the analytic magnitude or phase is very similar to 2D concentric circles (see Figure 3.1). For any zero mean $x \in \mathrm{I\!R}^N$ satisfying the uniqueness conditions, any $y \in \{rx : r > 0\}$ has the same phase as x; any $y \in \{x\cos\alpha + \widehat{x}\sin\alpha : \alpha \in \mathrm{I\!R}\}$ has the same magnitude as x ($\widehat{x} = \Im\{Fx\}$ is the Hilbert transform of x). Starting from x, by continuously changing the phases α, we can keep the magnitude unchanged and travel all the way through $\widehat{x}$, $-x$ and $\widehat{x}$, and then back to x again (just like traveling on a circle); or, we may continuously change the scaling factor r to reach bigger or smaller circles.

In summary, the signals with the same analytic phase lies in a 1D linear subspace and

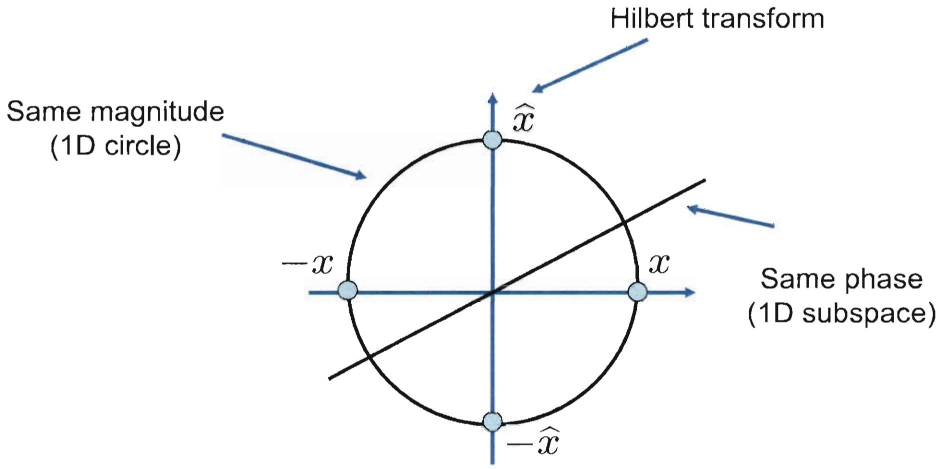


Figure 3.1 : The signal space geometry specified by the analytic magnitude and phase.

the signals with the same analytic magnitude lies on a 1D circle in a local neighborhood. This result seems suggesting that the magnitude encodes visual in a dual way to the phase, since a signal is unique given either its magnitude or phase. The above geometry gives some intuition about the reconstruction of a signal from its analytic magnitude or phase with Projection Onto Convex Sets (POCS [47]) type of iterative algorithms. Since the 1D subspace is a convex set, POCS algorithm for reconstruction from phase is guaranteed to converge to the 1D subspace. A circle at any small local neighborhood is very close to a convex affine space. So, POCS type of algorithm for reconstruction from magnitude is likely to converge to a nearby point on the circle, if the starting point is close enough to the circle.

3.1.3 Frequency Band Information $[K_1, K_2]$

In the uniqueness theorem, the frequency band $[K_1, K_2]$ is assumed to be known exactly (at least one of a_{K_1} and a_{K_2} is non-zero). If the signal x is allowed to go beyond the band

$[K_1, K_2]$ to $[K_1 - 1, K_2 + 1]$, there are two other interesting situations.

First, a signal y not limited in band $[K_1, K_2]$ (but in $[K_1 - 1, K_2 + 1]$) may have the same analytic magnitude or phase as x. As in the uniqueness theorem above, $S_v(z)$ may have one more pair of symmetric zeros than $S_x(z)$ and still keep $S_x^*(1/z^*)S_v(z)$ generalized complex conjugate symmetric about z^0. Theoretically, those symmetric factors in the filtered spectrum can be eliminated if the signal is known to satisfy the uniqueness condition. In real applications, however, that information may be unavailable.

Second, shift in frequency of $S_x(z)$ (equivalently, modulation in time) keeps the magnitude unchanged. For example, $S_x(z) = z^{-K_1}$ and $S_y(z) = z^{-K_1 - 1}$ both have the same constant analytic magnitude. The shift of $S_x(z)$ does not show up in Theorem 3 about magnitude, because it is not a local variation of signal x for discrete signals.

With unknown frequency band, the uniqueness given analytic magnitude or phase may break. Fortunately, when the analytic magnitude and phase of the CWT is concerned in practice, information about the signal components around band boundary can typically be obtained from the previous band (lower frequency band). The cross band information helps to fix the signal frequency band boundary as well as to resolve the ambiguity of same magnitude circle in Figure 3.1 (see more discussion in Section 3.3).

3.1.4 Singular Values of the Jacobians

Previously, we have looked at the rank of the Jacobians to determine the uniqueness of the analytic magnitude and phase representation. The Singular Values Decompositions (SVD) of the Jacobians, $J_{\rho(x)}$ and $D_{\rho(x)}J_{\theta(x)}$, provide more detailed information about the analytic magnitude and phase representation.

The magnitude $\rho(x)$ and phase $\theta(x)$ together form a locally orthogonal coordinate. So, if we let $J_{c(x)} = J_{\rho(x)} + \boldsymbol{j}\ D_{\rho(x)} J_{\theta(x)}$, then, we have the following proposition about the SVD of $J_{c(x)}$.

Proposition 3.3 (SVD of $\boldsymbol{J_{\rho(x)} + j\ D_{\rho(x)} J_{\theta(x)}}$). *Suppose the analytic filter F is circulant ($\Lambda_F = WFW^{-1}$ is diagonal). For all $x \in \mathrm{I\!R}$, the singular value decomposition of $J_{c(x)} = J_{\rho(x)} + \boldsymbol{j}\ D_{\rho(x)} J_{\theta(x)}$ is given below:*

$$J_c(x) = U \Sigma V^H$$

where $\Sigma = |\Lambda_F|$, $U = D_\theta^{-1} W^{-1}$ and $V = W S_p^*$ (S_p is given via $\Lambda_F = \Sigma S_p$).

Proof: Similar to the uniqueness theorem, we have

$$\begin{aligned} J_{c(x)} &= J_{\rho(x)} + \boldsymbol{j}\ D_{\rho(x)} J_{\theta(x)} \\ &= D_\theta^{-1}\ F \\ &= D_\theta^{-1}\ W^{-1}\ \Lambda_F\ W \\ &= (D_\theta^{-1} W^{-1})\ \Sigma\ (S_p W) \end{aligned}$$

Let $U = D_\theta^{-1} W^{-1}$ and $V = W^{-1} S_p^*$, then we know that they are both unitary. Since Σ is a non-negative diagonal matrix, we have the SVD: $J_{c(x)} = U \Sigma V^H$. □

By definition, for a standard analytic filter $F \in \mathbb{C}^{N \times N}$, we should have, for all $x \in \mathrm{I\!R}^N$,

$$\begin{aligned} \|Fx\|_2 = \|x + \boldsymbol{j}\ \widehat{x}\|_2 &= \sqrt{\|x\|_2^2 + \|\widehat{x}\|_2^2} \\ &= \sqrt{2\|x\|_2^2} = \sqrt{2}\ \|x\|_2 \end{aligned}$$

In practice, F is also designed by some filter design technique to satisfy this norm preserving property. Then, we have the following property about the maximum singular values of the Jacobians.

Proposition 3.4 (Maximum Singular Values of $\boldsymbol{J_{\rho(x)}}$ and $\boldsymbol{D_{\rho(x)}J_{\theta(x)}}$). *Let the maximum singular values of $J_{\rho(x)}$ and $D_{\rho(x)}J_{\theta(x)}$ be $\sigma_{\rho(x)}$ and $\sigma_{\theta(x)}$ respectively. If the analytic filter F satisfies that $\|F\ x\|_2 = \sqrt{2}\ \|x\|_2$, then*

$$\sigma^{\max}_{\rho(x)} \leq \sqrt{2}$$

$$\sigma^{\max}_{\theta(x)} \leq \sqrt{2}$$

Proof: Let v be the right singular vector of $J_{\rho(x)}$ associated with the maximum singular value $\sigma^{\max}_{\rho(x)}$.

$$\begin{aligned}\sqrt{2}\,\|v\|_2 &= \|Fv\|_2 \\ &= \|D_{e^{-j\,\theta}}Fv\|_2 \\ &= \|\left(J_{\rho(x)} + \boldsymbol{j}\ D_{\rho(x)}J_{\theta(x)}\right)\ v\|_2 \\ &\geq \|J_{\rho(x)}v\|_2 \\ &= \sigma^{\max}_{\rho(x)}\|v\|_2\end{aligned}$$

Therefore, we have $\sigma^{\max}_{\rho(x)} \leq \sqrt{2}$.

The other inequality, $\sigma^{\max}_{\theta(x)} \leq \sqrt{2}$, can be derived similarly. □

Combining Proposition 3.3 and the uniqueness theorem (Theorem 3), we have the following conclusions about the SVD of $D_{\rho(x)}J_{\theta(x)}$ and $J_{\rho(x)}$. First, the largest singular values of $D_{\rho(x)}J_{\theta(x)}$ and $J_{\rho(x)}$ are both $\sqrt{2}$ and the associated right singular vectors are $\widehat{x}$

and x respectively ($||D_{\rho(x)}\Delta\theta(x)|| \leq \sqrt{2}||\Delta x||$ and $||\Delta\rho(x)|| \leq \sqrt{2}||\Delta x||$). That is, the largest changes in $D_{\rho(x)}\theta(x)$ and $\rho(x)$ come from the changes of x in the direction of $\widehat{x}$ and x respectively. Therefore, the phase is most effective in encoding the local shift (α in $\{x\cos\alpha + \widehat{x}\sin\alpha : \alpha \in \mathrm{I\!R}\}$) and magnitude is most effective in encoding the local signal energy (r in $y \in \{rx : r > 0\}$).

Second, the smallest singular value of $D_{\rho(x)}J_{\theta(x)}$ is 0 and the corresponding singular vector is x (i.e., $||x||$ is arbitrary). The smallest singular value of $J_{\rho(x)}$ is 0 or very close to 0 and the corresponding singular vector is $\widehat{x}$, unless x has very big DC components. That is, $x + r\widehat{x}$ has exactly or almost the same magnitude as x for small $r > 0$ (i.e., in a very small neighborhood, a small piece of a circle is very close to the tangent of the circle there). Therefore, the reconstruction from magnitude without any information of phase is may be less accurate than the reconstruction from phase for very detailed structures.

3.2 Iterative Reconstruction from Analytic Magnitude or Phase

In this section, I present iterative algorithms for signal reconstruction from the specified analytic magnitude or phase. Then, I show the theoretical results on the convergence of the iterative algorithms.

3.2.1 An Algorithm for Reconstruction from Analytic Magnitude

The following simple iterative algorithm can be used to reconstruct a signal $x \in \mathrm{I\!R}^N$ from its analytic magnitude $\rho \in \mathrm{I\!R}^N$. In the reconstruction algorithm below, F is the analytic filter as defined previously; G is the "inverse" of F in ℓ_2 sense, *i.e.*, $y = \Re\{G\rho e^{j\,\theta}\} \in \mathrm{I\!R}^N$ minimizes the norm $||Fy - \rho e^{j\,\theta}||_2$.

```
Iterative Reconstruction from the analytic magnitude
Given initial signal estimate $x_0$ and the magnitude $\rho$
(1) Let $k = 1$
(2) Compute the magnitude and phase: $\rho_k e^{j\,\theta_k} = F x_{k-1}$
(3) Update signal estimate: $x_k = \Re\{G\,\rho\, e^{j\,\theta_k}\}$
(4) Let $k = k + 1$ and go to (2).
```

In the above algorithm, the frequency band information ($[K_1, K_2]$) required by Theorem 3 is not considered at all. In real applications, the analytical magnitude and phase of the CWT will be actually employed and the filter bank structure of the CWT will take care of the frequency band information automatically (discussed later in Section 3.3). The influence of this lack of band information is illustrated in Figure 3.2 later on.

Proposition 3.5 (The monotonic decreasing of magnitude error). *The reconstruction algorithm described above monotonically decreases the error in magnitude before it reaches a fixed point*, i.e.,

$$\|\rho_{k+1} - \rho\|_2 \leq \|\rho_k - \rho\|_2$$

where equality holds if and only if $x_k = x_{k-1}$.

Proof: It can be shown that $|r_1 - r_2| \leq |r_1 - r_2 e^{j\,\alpha}|$ holds true for $r_1, r_2 \geq 0$ and $\alpha \in \mathbb{R}$. Then, in vector form, we have

$$\|\rho_{k+1} - \rho\|_2 \leq \|\rho_{k+1} - \rho e^{j\,(\theta_k - \theta_{k+1})}\|_2 = \|\rho_{k+1} e^{j\,\theta_{k+1}} - \rho e^{j\,\theta_k}\|_2$$

By the definition of G, $\rho_{k+1} e^{j\,\theta_{k+1}} = F x_k$ is the unique nearest point to $\rho e^{j\,\theta_k}$ in the valid signal space, *i.e.*,

$$\|\rho_{k+1} e^{j\,\theta_{k+1}} - \rho e^{j\,\theta_k}\|_2 \le \|\rho_k e^{j\,\theta_k} - \rho e^{j\,\theta_k}\|_2 = \|\rho_k - \rho\|_2$$

where equality holds if and only if $x_k = x_{k-1}$.

Combine the above two,

$$\|\rho_{k+1} - \rho\|_2 \le \|\rho_k - \rho\|_2$$

where equality holds if and only if $x_k = x_{k-1}$. □

In the algorithm and proposition above, the given magnitude $\rho \in \mathbb{R}^N$ is not required to be the magnitude of any signal $x \in \mathbb{R}^N$

Theorem 4 (The convergence of the algorithm for reconstruction from magnitude). *The reconstruction algorithm above gives a bounded sequence $\{x_k\}_{k=0}^{\infty}$. The limit x of any convergent subsequence (which must exist) of $\{x_k\}_{k=0}^{\infty}$ is a fixed point of the algorithm.*

Proof: For the convergence, we apply the Global Convergence Theorem in [48] which is cited below:

Let A be an algorithm on X, and suppose that, given x_0, the sequence $\{x_k\}_{k=0}^{\infty}$ is generated satisfying $x_{k+1} \in A(x_k)$.

Let a solution set $\Gamma \subset X$ be given, and suppose

1. all points x_k are contained in a compact set $S \subset X$.

2. there is a continuous function Z on X such that

   (a) if $x \notin \Gamma$, then $Z(y) < Z(x)$ for all $y \in A(x)$

(b) if $x \in \Gamma$, then $Z(y) \leq Z(x)$ for all $y \in A(x)$

3. the mapping A is closed at points outside Γ

Then the limit of any convergent subsequence of $\{x_k\}$ is in Γ.

Follow the notation of the global convergence theorem, we define the solution set as the fixed point set $\Gamma = \{x \in \mathbb{R}^N : A(x) = x\}$. All the signals with the specified magnitude, if they exist, are in Γ, since they satisfy $A(x) = x$.

First, all x_k are bounded in R^N with the specified magnitude ρ. So, they are in a compact set.

Second, we define function $Z(x) = \|\rho(x) - \rho\|_2$ as the error measure in magnitude as in the previous Proposition 3.5. Then, for $y = A(x)$, we have (a) $Z(x) = Z(y)$ for all $x \in \Gamma$; (b) $Z(y) < Z(x)$ for all $x \notin \Gamma$.

Third, the algorithm is a continuous map on a compact set of R^N. So, it is a closed map.

Therefore, the limit of any convergent subsequence of $\{x_k\}_{k \in \mathbf{Z}}$ is a fixed point of A.

□

With Proposition 3.5 and Theorem 4, I established that the reconstruction algorithm generates a convergent sequence or subsequence with a limit of a fixed point of the algorithm and the error in magnitude always drops down before convergence.

In practice, the proposed iterative reconstruction algorithm usually is able to converge to the desired signal when a good initial estimate is supplied and additional frequency band information is enforced (will be discussed in more details in Section 3.3.3). Some simulation examples are given in the following section.

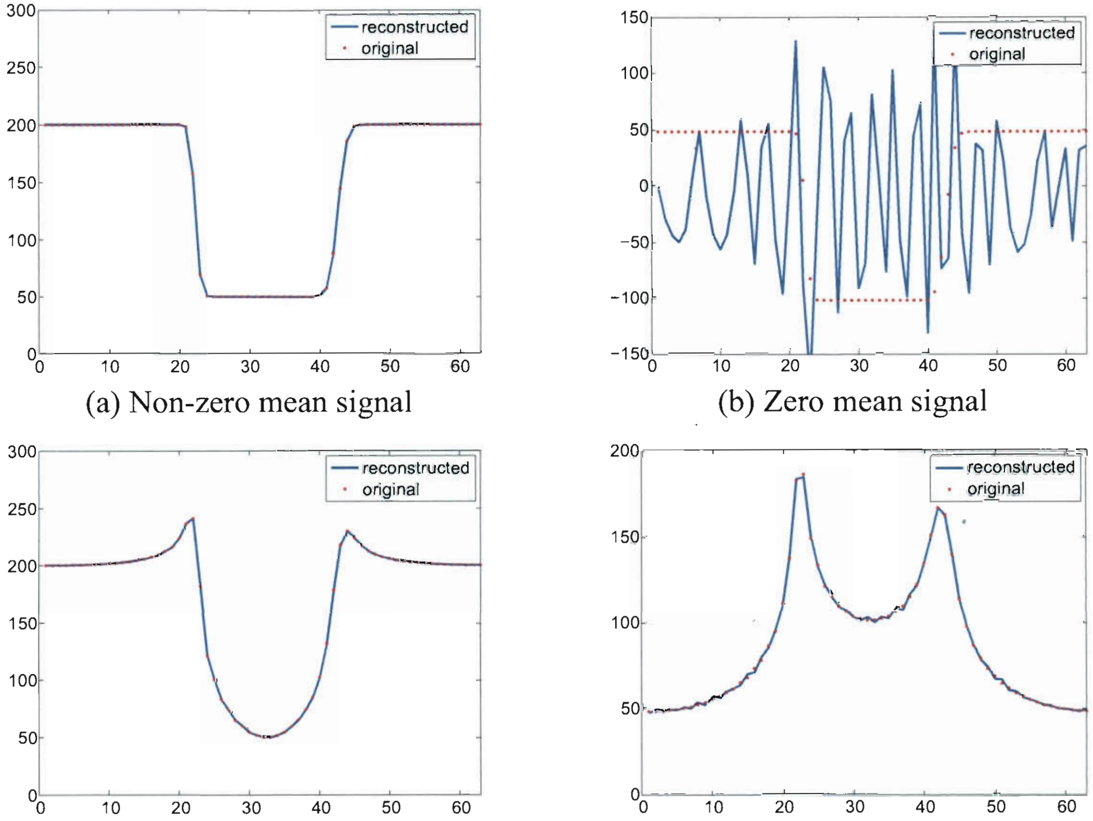
3.2.2 Simulation Examples for Iterative Reconstruction from Analytic Magnitude

As discussed previously, when the uniqueness theorem (Theorem 3) holds, the signals with the same magnitude are isolated points or isolated 1D circles in the given frequency band $[K_1, K_2]$.

First, if a signal x has non-zero positive mean (e.g., the lowpass band of the CWT of an image), the iterative algorithm is observed to converge to x starting from a positive constant signal, or converge to $-x$ starting form a negative constant signal. Figure 3.2 (a) and (c) show such an example of perfect reconstruction from the analytic magnitude. In this case, the non-zero mean condition implies the local uniqueness, the frequency band knowledge of $K_1 = 0$ and $a_0 \neq 0$, and a good starting point of a constant signal (DC signal) which falls into the desired frequency band.

Second, if $a_0 = 0$ (e.g., for the highpass bands of the CWT, $a_0 \approx 0$), the iterative reconstruction algorithm is observed to converge to some signal with very close magnitude if starting from some white Gaussian random signal x_0. Figure 3.2 (b) and (d) are such a simulation example. The signal x is the same signal in Figure 3.2 (a) with mean removed. The starting signal x_0 is a white Gaussian random vector. In this case, the reconstruction algorithm converges to a wide band signal y with a very close analytic magnitude (SNR=37.1dB in Figure 3.2 (d)). As discussed previously, the signal y is a fixed point of the iterative algorithm and the reconstruction y is wide band since the starting point x_0 is wide band. This wrong convergence problem can be avoided by choosing better initial estimate x_0 as shown in Figure 3.3.

Figure 3.3 shows simulation examples with more reasonable initial estimate for the



(a) Non-zero mean signal

(b) Zero mean signal

(c) Magnitude of (a)

(d) Magnitude of (b)

Figure 3.2 : The signals reconstructed from analytic magnitude: (a) reconstruction of a non-zero mean signal (perfect reconstruction); (b) reconstruction of a zero mean signal (signal in (a) with mean removed); (c) the analytic magnitude of (a); (d) the analytic magnitude of (b) (the relative magnitude difference is 37.1dB).

iterative reconstruction algorithm. In (a) and (c), the initial estimate is chosen to be a lowpass filtered version of the original signal with additive white Gaussian noise. The reconstructed signal has exactly the same analytic magnitude as the original signal and a constant phase shift. In (b) and (d), the initial estimate is a sinusoid and the result is the same as in (a) and (c). These simulation results match exactly the uniqueness theorem (Theorem 3).

As shown by the above examples and the uniqueness theorem, some extra information is required to resolve the ambiguity of the phase shift. In real applications, reconstruc-

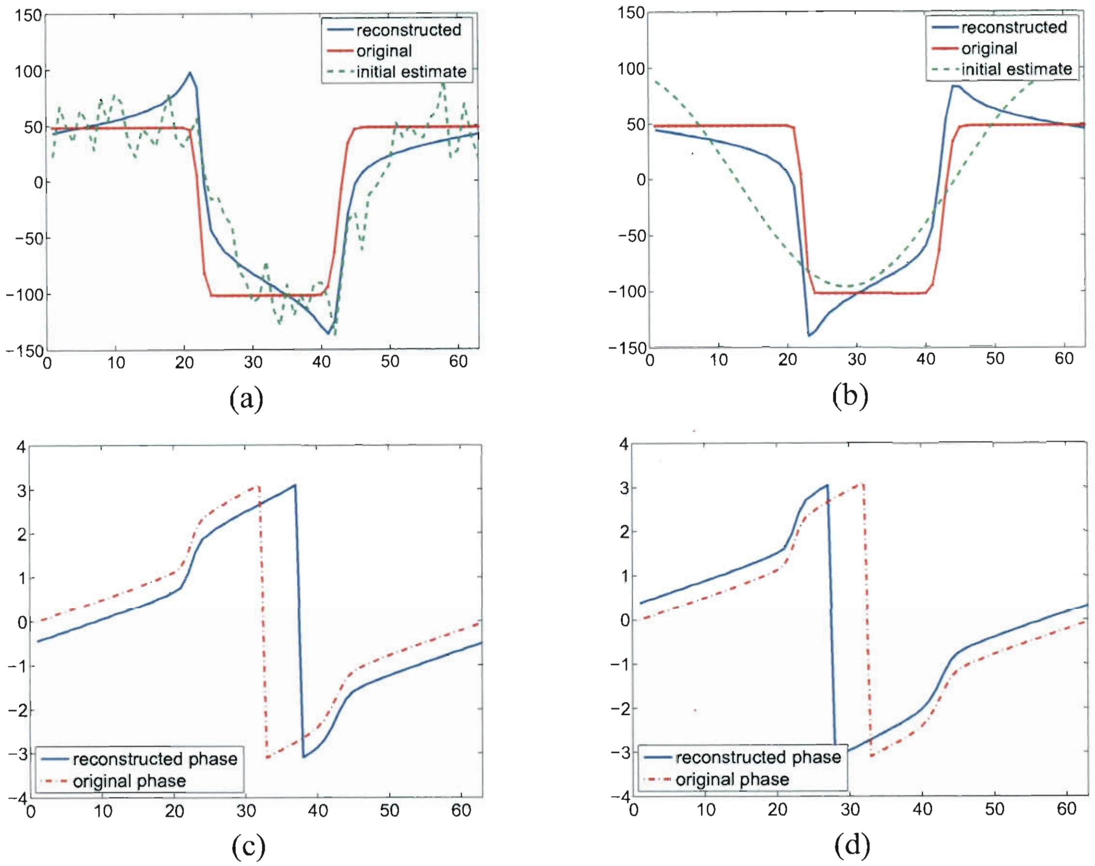


Figure 3.3 : The signals reconstructed from the analytic magnitude (same signal as in Figure 3.2 (b)): (a) reconstruction starting with an initial estimate of lowpass filtered original signal plus additive white Gaussian noise (reconstructed signal has exactly the same magnitude as the original); (b) reconstruction starting with an initial estimate of a sinusoid (reconstructed signal has exactly the same magnitude as the original); (c) the analytic phase of the original and reconstructed signal in (a) (note that the phase difference is constant); (d) the analytic phase of the original and reconstructed signal in (b) (note that the phase difference is constant)

tion from analytic magnitude of the CWT can typically make use of the lower pass band information to decide the unknown phase shift (will be discussed later in Section 3.3).

3.2.3 An Algorithm for Reconstruction from Analytic Phase

The following simple iterative algorithm can be used to reconstruct a signal $x \in \mathbb{R}^N$ from its analytic phase $\theta \in \mathbb{R}^N$. In the reconstruction algorithm below, F and G are defined as previously in Section 3.2.1.

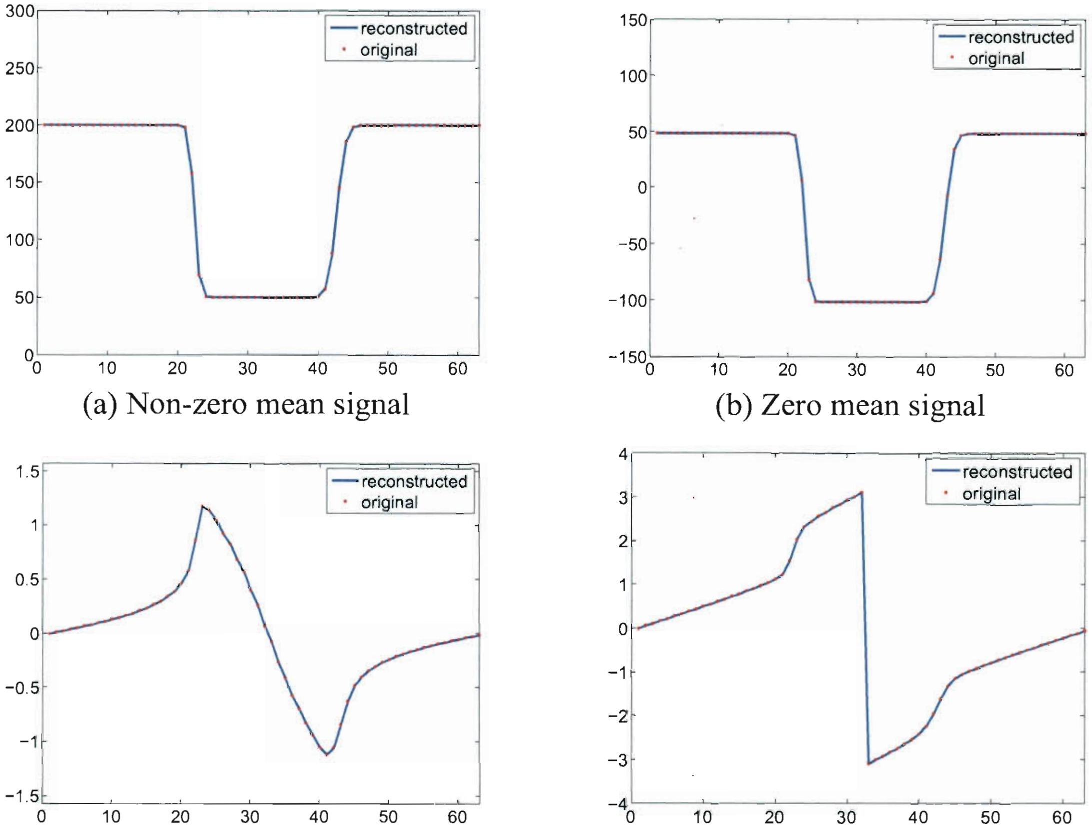
```
Iterative Reconstruction from the analytic phase
Given initial signal estimate $x_0$ and the phase $\theta$
(1) Let $k = 1$
(2) Compute the magnitude and phase:  $\rho_k e^{j\,\theta_k} = F\,x_{k-1}$
(3) Update signal estimate:  $x_k = \Re\{G\,\rho_k\,e^{j\,\theta}\}$
(4) Let $k = k + 1$ and go to (2).
```

The algorithm above is exactly the same as the iterative reconstruction algorithm from the Fourier phase proposed in [19], except being applied on the analytic phase. Its convergence has also been proved [19, 49] as a form of iterative non-expensive signal reconstruction. The main result is cited as in the below theorem.

Theorem 5 (The convergence of the iterative reconstruction from phase [19, 49]). *In the above algorithm, the squared error $||x_k - x||^2$ is non-increasing with each iteration. If the signal is unique (up to a scale factor) with the specified phase, the algorithm will converge to the desired signal.*

Alternatively, the step 3 in the iterative algorithm can be modified slightly from replacing the phase θ_k with θ to projection to the linear subspace with phase θ. Then, the iterative algorithm becomes a POCS algorithm when viewed from the transform domain and the convergence can also be derived as a direct result of POCS [47]. Practically, both replacing phase and projection work for reconstruction, although replacing phase is much simpler computationally. Conceptually, the perspective of POCS is a better interpretation when estimation and other constraints are involved in real applications (e.g., when some models are employed to describe the phase or the signal as in Chapter 4).

3.2.4 Simulation Examples for Iterative Reconstruction from Analytic Phase



(a) Non-zero mean signal

(b) Zero mean signal

(c) Phase of (a)

(d) Phase of (b)

Figure 3.4 : The signals reconstructed from the analytic phase: (a) perfect reconstruction of a non-zero mean signal; (b) perfect reconstruction of a zero mean signal; (c) the analytic phase of (a); (d) the analytic phase of (b).

In Figure 3.4, the same signals in Figure 3.2 are used to show the results of the iterative reconstruction from the analytic phase. For both zero mean and non-zero mean signals, the iterative algorithm can perfectly reconstruct the desired signal.

3.3 Uniqueness and Reconstruction from the CWT Magnitude or Phase

3.3.1 The Analytic Magnitude and Phase of the CWT

The CWT magnitude and phase of a signal can be considered as the analytic magnitude and phase of some real wavelet transform coefficients of that signal. Therefore, the extension of the uniqueness theorem (Theorem 3) to the multi-resolution CWT seems straightforward.

First, a real wavelet filter bank is applied to decompose a signal $s(n)$ into real wavelet coefficients $\{x(n;k) : k \in \Phi\}$ and then the analytic filter F is applied to transform $x(n;k)$ to the analytic magnitude and phase representation $\{\rho(n;k), \theta(n;k)\}$. If the wavelet coefficients $x(n;k)$ in each band satisfies the uniqueness conditions, $x(n;k)$ can be uniquely specified by $\rho(n;k)$ or $\theta(n;k)$. Since the signal x can be determined by the wavelet coefficients in all the bands $\{x(n;k) : k \in \Phi\}$, x must be uniquely specified by $\{\rho(n;k) : k \in \Phi\}$ or $\{\theta(n;k) : k \in \Phi\}$.

However, the above extension formulation holds strictly true only for the DC band ($x(n;k)$, $k = 0$) of the wavelet coefficients (assuming that the image pixels take non-negative values). Two details required by the uniqueness theorem have to be considered carefully for the AC bands ($x(n;k)$, $k \neq 0$).

First, an AC band typically has large portions of zeros in the smooth regions of an image, which leads to a lot of zeros in the analytic magnitude $\rho(n;k)$. The uniqueness theorem assumes an ideal analytic filter which has an infinite filter length (2.5 (c)). When a large consecutive set of the magnitude is zero, the Jacobians become ill conditioned (very small singular values). In real implementation, the analytic filter will typically be designed to have compact support and those very small singular values become zeros. Effectively, the separated parts become independent (the magnitude and phase of each part has no information about other parts). In the following, I will discuss how to extend the uniqueness theorem in incorporate information about the zero elements in the magnitude.

Second, an AC band $x(n;k)$ usually has zero mean, because the wavelet AC band filters are high pass filters which suppress the DC component. Therefore, an AC band typically is only locally unique up to a phase shift given the analytic magnitude according to the uniqueness theorem. It is not clear at this point how the set of multi-resolution magnitude $\rho(n;k)$ together can resolve the unknown phase shift in every AC band. Later in this section, I will also explain by some examples how the multi-resolution reconstruction algorithm find out the local neighborhood of uniqueness and determine the unknown phase shift.

A. Localized signal with zero elements in the magnitude

For real life signals (wavelet with compact support), the magnitude is typically zero within smooth areas in an image and is non-zero around edges. The uniqueness theorem may be extended in the following way to incorporate the information about zero elements. Suppose the locations of zero magnitude are known and only signals with the same set of zero magnitude are interested (similar to the known frequency band in Theorem 3). The following proposition gives the condition on the existence of zero magnitude.

Proposition 3.6 (Zero magnitude condition). *The k-th element of the magnitude $\rho(x) = |Fx|$ is zero (*i.e.*, $\rho_k(x) = 0$), if and only if $S_x(z)$ has a zero at $e^{-j\frac{2k\pi}{N}}$, where $S_x(z)$ is the z transform of WFx (same as in Theorem 3).*

Proof: Since $\rho_k(x) = 0$ is equivalent to the k-th element of Fx being 0, the IDFT of the sequence $S_x(z)$ has a zero at the k-th location. Therefore, $S_x(z)$ has a zero at $e^{-j\frac{2k\pi}{N}}$. □

The k-th element of the magnitude is zero ($\rho_k(x) = 0$) if and only if $S_x(z)$ has a zero at $e^{-j\frac{2k\pi}{N}}$. Then, $S_v(z)$ in the proof of Theorem 3 has to have the same zero to maintain $\rho_k(v) = 0$, since we are only interested in v with the same zero at its k-th element.

In this situation, given the analytic phase or magnitude, each segment of the signal with non-zero magnitude is unique up to a scale factor or phase shift respectively. The information about the unknown scale factor or phase shift for each segment can typically be derived from the lowpass bands because of the filter bank structure of the CWT. With given magnitude, the information about the locations of the zero magnitude elements is available and used in the reconstruction algorithm. With given phase, that information may also be inferred from the lower pass bands of the CWT.

B. Uniqueness given multi-resolution magnitude or phase for signals with zero mean AC bands

For a signal $s(n)$ with zero mean AC bands $x(n; k)$, the uniqueness theorem states that $x(n; k)$ is only locally unique up to a phase shift given the magnitude $\rho(n; k)$ and that $x(n : k)$ is globally unique up to scaling factor. So, for magnitude, a good initial estimate is needed and the unknown phase shift has to be decided. For phase, the unknown scaling factor has to be obtained.

The multi-resolution decomposition greatly helps the reconstruction from magnitude or phase in this situation. Conceptually, a lowpass band typically can be recovered first which gives a good initial estimate for the highpass band through the inter-band dependency and redundancy of the CWT. The overlapping of adjacent bands in the CWT gives very accurate information about the frequency components around the band boundary. With this information, the ambiguity of the unknown phase shift and the unknown scaling factor in highpass band can both be resolved. When all the CWT bands are reconstructed in parallel, the forward and inverse CWT transform will impose the cross-band relationship automatically. Therefore, the ambiguous phase shift of AC band can be fixed automatically (will be demonstrated by simulation examples in Section 3.3.3).

3.3.2 Extension to Higher Dimensions

The extension of the uniqueness theorem to higher dimensions follows from the extension of analytic filter to higher dimensions [27]. For example, consider a 2D signal $x \in \mathrm{I\!R}^N$ with $\sqrt{N} \times \sqrt{N}$ pixels. The 2D magnitude $\rho(m, n)$ and phase $\theta(m, n)$ in each band are constructed by filtering $x(m, n)$ with a 2D analytic filter F with single quadrant frequency response (Figure 2.6). We can construct polynomials $S_x(z_1, z_2)$ and $S_v(z_2, z_2)$ in a similar way as for 1D and let $S(z_1, z_2) = S_x^*(1/z_1^*, 1/z_2^*)S_v(z_1, z_2)$. Since Equation 3.1 in the uniqueness theorem holds for higher dimensions, we conclude that $S(z_1, z_2)$ should have no non-trivial symmetric factors ($f(z_1, z_2) = e^{j\,\alpha} z_1^{M_1} z_2^{M_2} f^*(1/z_1^*, 1/z_2^*)$) for the phase and magnitude to be unique.

3.3.3 Simulation Examples for the Analytic CWT Magnitude and Phase

In this section, I present some simulation examples of the iterative reconstruction from the analytic CWT magnitude or phase. The signals in Figure 3.5 are used in the simulation. The wavelet lowpass band in (b) is a signal with non-zero mean and the wavelet highpass band in (c) is a signal with zero mean.

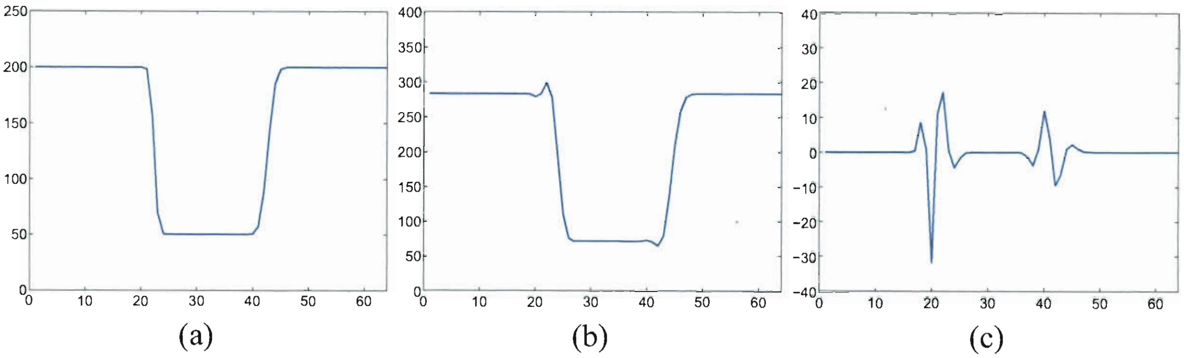
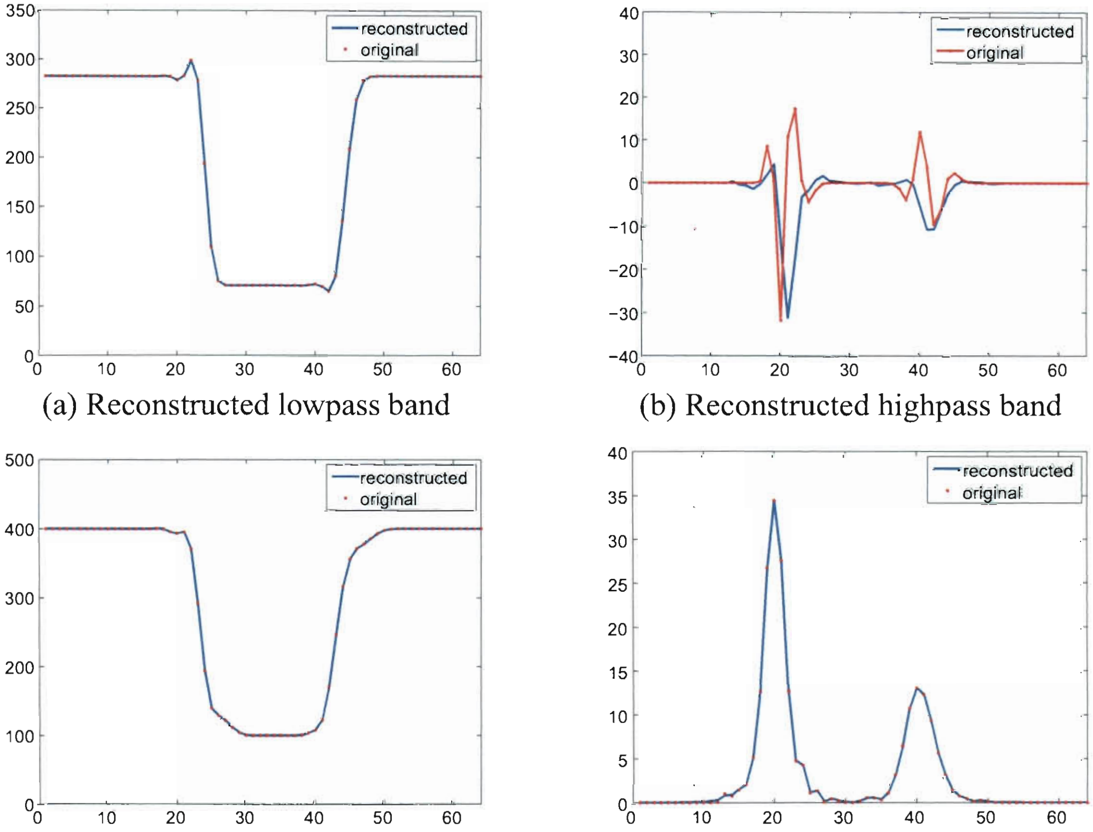


Figure 3.5 : The signals used in the simulation for reconstruction from the analytic CWT magnitude and phase: (a) a signal with two edges; (b) the wavelet lowpass band of (a); (c) the wavelet highpass band of (a).



(a) Reconstructed lowpass band

(b) Reconstructed highpass band

(c) Magnitude of (a)

(d) Magnitude of (b)

Figure 3.6 : The signals reconstructed from the analytic magnitude: (a) reconstructed wavelet lowpass band (perfect reconstruction); (b) reconstructed wavelet highpass band; (c) the analytic magnitude of (a) (perfect reconstruction); (d) the analytic magnitude of (b) (relative error in magnitude $\frac{\|\rho-\widehat{\rho}\|_2}{\|\rho\|_2} = 0.6\%$).

With the iterative algorithm in the previous section, the wavelet lowpass band and highpass band are reconstructed from their analytic magnitude respectively. The results are shown in Figure 3.2. In (a), the lowpass band is perfectly reconstructed from an initial estimate of a positive constant signal. In (b), the highpass band is reconstructed from an initial estimate of a zero signal. Although, the reconstruction in (b) is not good, the magnitude in (d) is very close.

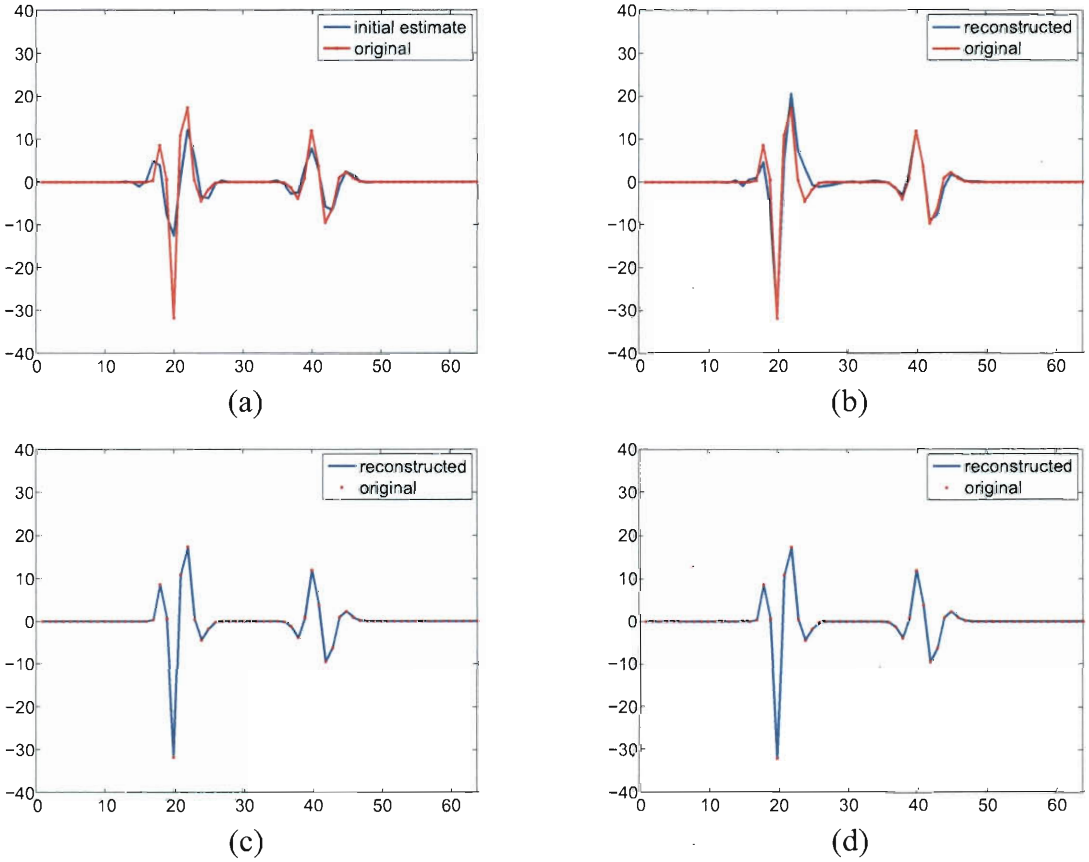
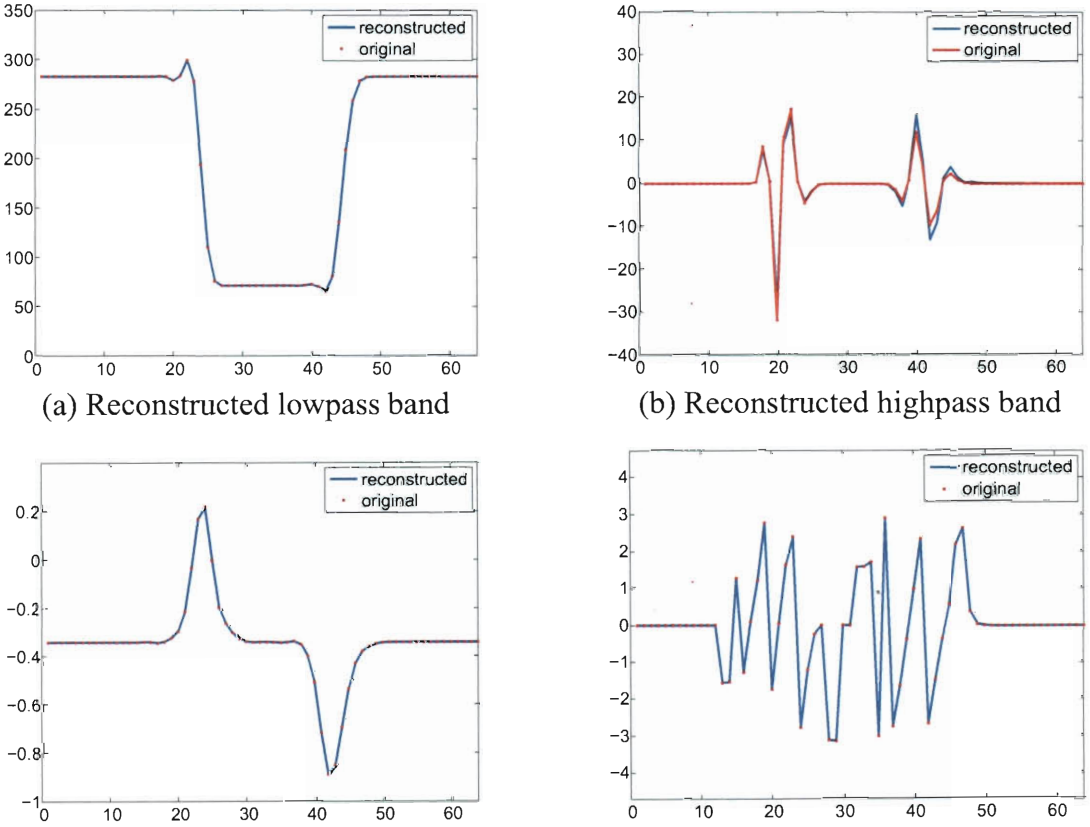


Figure 3.7 : The reconstruction from the CWT magnitude: (a) initial estimate given by projecting the lowpass band to highpass band; (b) reconstruction result with initial estimate in (a) ; (c) reconstruction result by repeatedly projecting the lowpass band to the highpass band in every reconstruction iteration (perfect reconstruction); (d) reconstruction results with estimating the two CWT in parallel (perfect reconstruction).

In Figure 3.7, the influence of the CWT on the reconstruction results is demonstrated.

In (a), the projection of the lowpass band to the highpass band is shown to be a good initial estimate for reconstructing the highpass band. In (b), the reconstruction result improves significantly by using the initial estimate in (a). In (c) and (d), perfect reconstruction of the highpass band from the CWT magnitude is achieved by repeatedly projecting the lowpass band to the highpass band and reconstruct both the lowpass and highpass band in parallel respectively.



(a) Reconstructed lowpass band

(b) Reconstructed highpass band

(c) Phase of (a)

(d) Phase of (b)

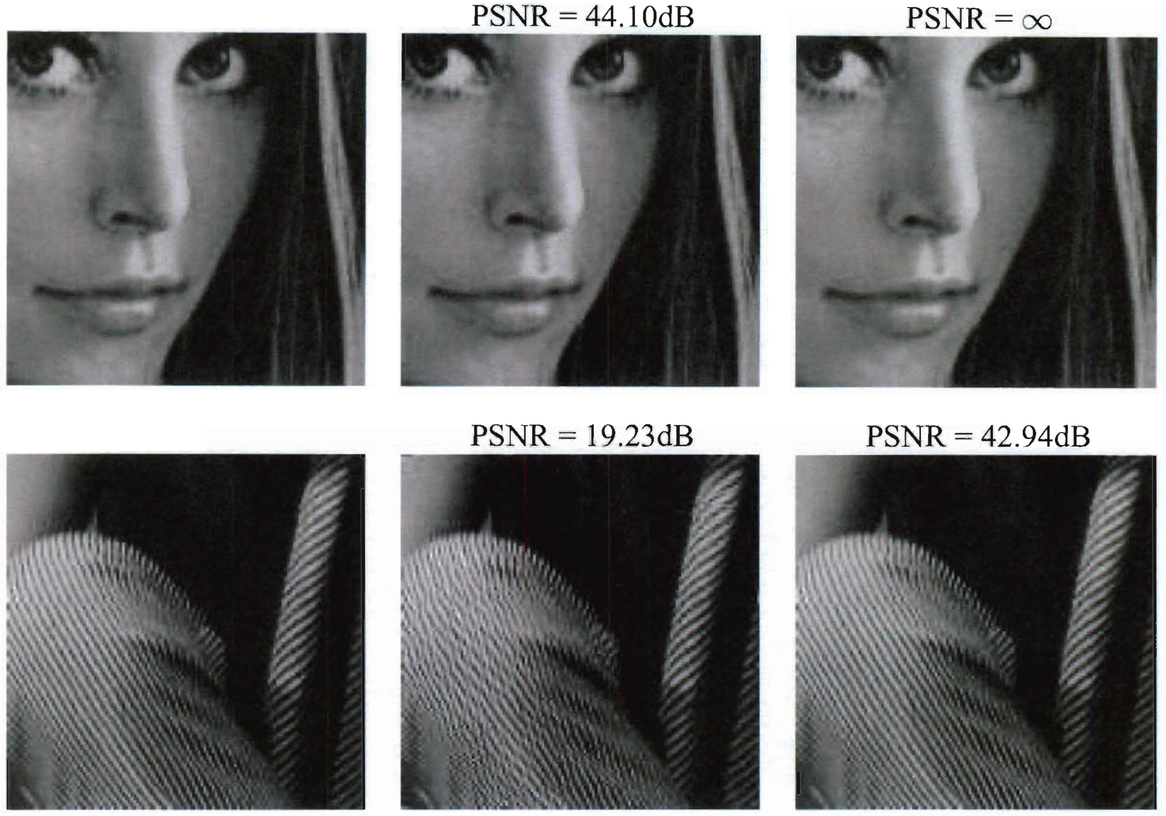
Figure 3.8 : The signals reconstructed from the CWT phase: (a) reconstructed wavelet lowpass band (perfect reconstruction); (b) reconstructed wavelet highpass band; (c) the analytic phase of (a) (perfect reconstruction) ; (d) the analytic phase of (b) (the maximum phase deviation is 0.001π).

For the reconstruction from the CWT phase, the problem is simpler, because the lowpass band and highpass band can be reconstructed (up to a scaling factor) independently as shown in Figure 3.8. The result in (b) shows a little problem in deciding the right energy of the two pulses: the reconstructed left pulse is a little smaller and the right one is a little bigger than the original signal. This is exactly the zeros in magnitude problem just discussed in Section 3.3.1. When the two bands are reconstructed in parallel, the lowpass band can help decide the right scaling factor for the two pulses. Since the two bands have the relationship just discussed above, the scaling factor of one band can be typically derived from that of the other band, if the signal components around the band boundary are not zero. So, a signal can even be reconstructed given only the CWT phase of both bands without further information about the signal energy in each band.

Figure 3.9 shows some simulation examples for reconstruction from the 2D CWT magnitude and phase with 300 iterations starting from white Gaussian noise. For a crop from Lena image, the reconstruction from both magnitude and phase are very good. The magnitude reconstruction has problem on locating some of the hair strips. In the areas with problems, typically, some of the phase variables in a CWT band of the reconstructed image have an extra offset close to π, i.e., the signal has the wrong polarity (positive or negative) there in that CWT band.

For the crop from the Barbara image, the reconstruction from magnitude has low PSNR, because the magnitude has trouble determine the direction of the texture (the signal is not unique given the magnitude in this situation). The reconstruction from the phase is not perfect (but with very good visual quality), because the convergence is very slow on some of the texture area in determining the local magnitude.

The iterative reconstruction processes of the two crops above are illustrated in Figure 3.10. The initial estimates are white Gaussian random noise with positive mean. So, the



original from magnitude from phase

Figure 3.9 : The reconstruction results from the CWT magnitude or phase after 300 iterations starting from white Gaussian noise: (left) the original image; (middle) image reconstructed from magnitude; (right) image reconstructed from phase.

convergence at the beginning is slightly slow. After the first iteration (the first column), the image contents are recognizable. After iteration 10 or 20 (the second or third column), the most of the image details are recovered, although reconstructions from magnitude show blurry or shifted edges at some places and reconstructions from phase are noisy in smooth areas. The convergence is typically much faster in practice, if a better initial estimate or some other information is available as in many real applications.

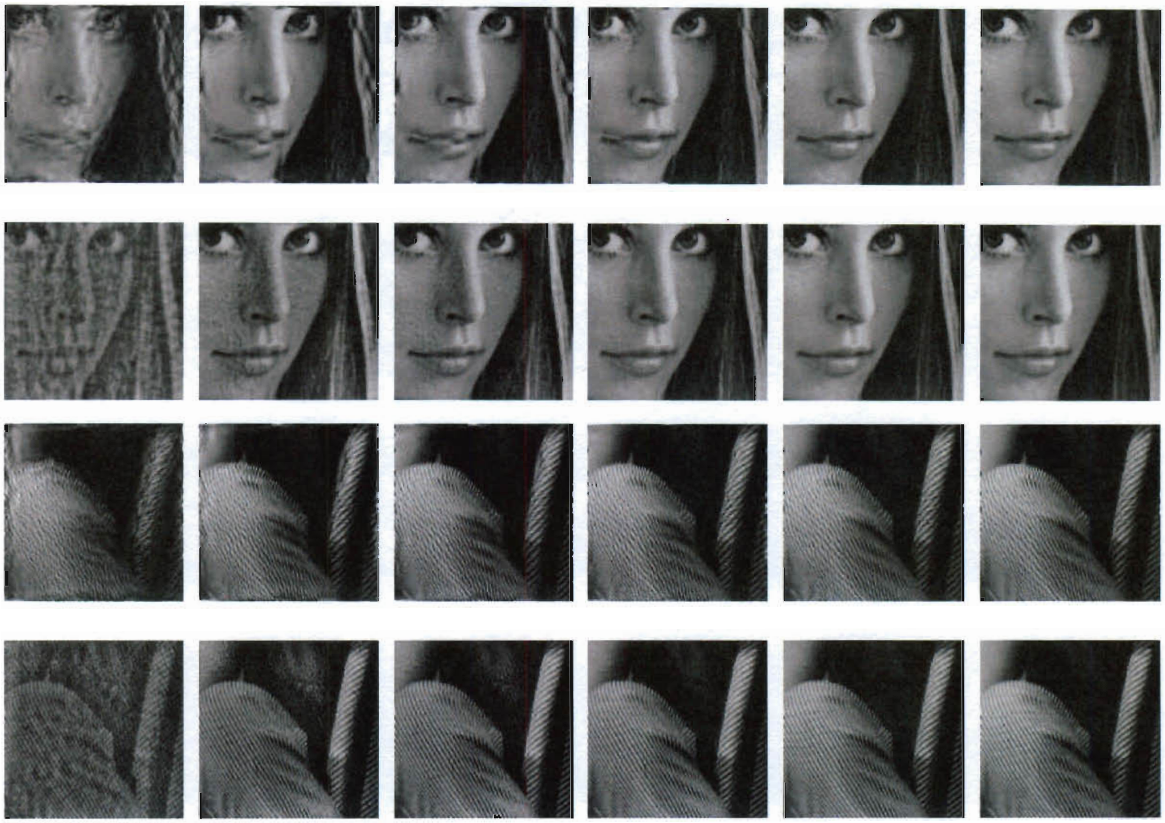


Figure 3.10 : The process of reconstruction from the CWT magnitude or phase: (first row) Lena crop from magnitude; (second row) Lena crop from phase; (third row) Barbara crop from magnitude; (fourth row) Barbara crop from phase; (from left to right) reconstruction after iteration 1, 10, 20, 40, 80, and 160 respectively.

3.4 The Magnitude and Pseudo-phase of the Over-complete DCT

In this chapter, I have developed the theory and algorithms for image reconstruction from the analytic CWT magnitude and phase. However, for some applications, the over-complete DCT is preferred over the CWT (e.g., the temporal image prediction problem discussed in this dissertation). In those applications, the over-complete DCT can be employed as an alternative to the CWT. The DCT pseudo-phase (sometimes called DCT-phase by other authors) has been exploited in the fields of image registration and block matching motion estimation [50, 42, 51]. The DCT magnitude rarely appears as a topic in image processing research. In this section, I discuss image reconstruction from the magnitude and pseudo-phase of the over-complete DCT.

The iterative reconstruction algorithms for analytic magnitude and phase discussed in Section 3.2 can be applied to the over-complete DCT magnitude and pseudo-phase with only a little modification. The convergence properties for the analytic magnitude (Proposition 3.5 and Theorem 4) still hold true for the over-complete DCT. The reconstruction from the DCT pseudo-phase can also be considered as projection onto convex sets (POCS) just like the reconstruction from the CWT phase.

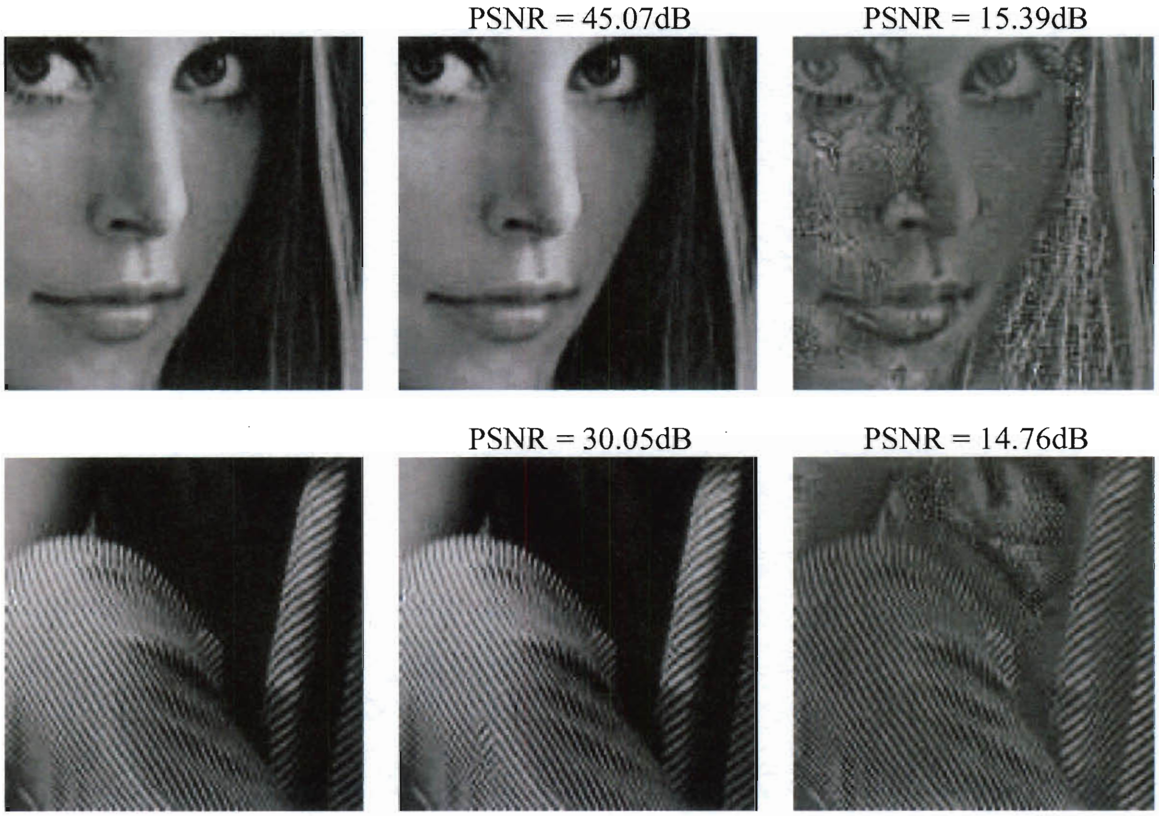


Figure 3.11 : The reconstruction results from the over-complete DCT magnitude or pseudo-phase: (left) the original image; (middle) image reconstructed from magnitude; (right) image reconstructed from pseudo-phase.

I observe that images can be reconstructed quite well from the magnitude and pseudo-phase of the over-complete DCT, although I cannot give theoretical results about the uniqueness. As shown in Figure 3.11, the reconstruction from the magnitude of the over-complete DCT is very similar to the reconstruction from the CWT magnitude (Figure 3.9): the magnitude can reconstruct good looking images, but has trouble in determining some details and orientation of some textures. The DCT pseudo-phase can reconstruct almost all the image details, but has trouble in determining the local signal energy. The reconstruction result from the DCT pseudo-phase in Figure 3.11 is very impressive, considering the fact that a pseudo-phase variable carries only one bit of information. For real applications, some information about the local signal energy may also be available so that the reconstruction may have much better local energy distribution.

3.5 Summary

This chapter considers the image reconstruction from the analytic magnitude and phase of the CWT. The conditions under which a signal is unique given its analytic magnitude or phase are presented. Iterative algorithms for reconstructing a signal from its analytic magnitude or phase are proposed and the convergence of proposed algorithms are analyzed. The above results are extended to the 2D CWT and illustrated with simulation examples. Image reconstruction from the over-complete DCT magnitude or pseudo-phase is also discussed for problems to which the CWT is not applicable.

In the next chapter, I apply simple models on both the CWT magnitude and phase to construct a new spatial image prediction algorithm, following the image reconstruction theory and algorithms developed in this chapter. In Chapter 5, I model the magnitude and pseudo-phase of the over-complete DCT under complicated spatial and temporal image evolutions involved in motion compensated video prediction for video coding.

3.6 Appendix: Some Properties of the Analytic Magnitude and Phase

This section presents some basic properties of the analytic magnitude and phase representation. Since the CWT can be considered as a set of analytic bandpass filters, I look at the properties of the analytic bandpass filters and the analytic magnitude and phase of the filtered coefficients.

First, if an ideal bandpass analytic filter has linear phase and zero delay (zero-phase), then it is complex conjugate symmetric.

Proposition 3.7 (Zero-phase analytic bandpass filter). *A zero-phase analytic filter* $h(t) = h_r(t) + \boldsymbol{j}\, h_i(t)$ *has even real part* $h_r(t)$ *and odd imaginary part* $h_i(t)$, i.e., $h^*(-t) = h(t)$.

Proof: By definition, the Fourier transform of the filter response $h(t)$ should be a real function ($\mathscr{F}\{h(t)\} = H(\omega) \in \mathbb{R}$). Therefore, we have $h(t) = h^*(-t)$. □

A bandpass analytic filter $h(t)$ can be simply designed by modulating a zero-phase lowpass filter (a even real function $h_0(t) = h_0(-t) \in \mathbb{R}$) to the pass-band center frequency ω_c, then we have $h(t) = h_0(t)e^{\boldsymbol{j}\,\omega_c t}$ and $h^*(-t) = h_0(-t)e^{\boldsymbol{j}\,\omega_c t} = h(t)$.

Note that in the special case of discrete analytical filter with center frequency $\omega_c = \frac{\pi}{2}$ (the pass-band is $[0, \pi]$), we have $h_r(n) = h_0(n)\cos(\frac{\pi}{2}n)$ and $h_i(n) = h_0(n)\sin(\frac{\pi}{2}n)$.

Proposition 3.8 (Phase at edge center). *The phase of the analytic bandpass filtered coefficient at the center of one anti-symmetric edge is* $\frac{\pi}{2}$ *or* $-\frac{\pi}{2}$.

Proof: Suppose the edge is $x(t) = s(t - t_0)$ where $s(t)$ is an odd function of t. Let the analytic bandpass filter be $h(t) = h_r(t) + \boldsymbol{j}\, h_i(t)$, then $h_r(t)$ and $h_i(t)$ are real even and real odd functions respectively. The filtered coefficients are given below:

$$\begin{aligned} c(t) &= x(t) * h(t) = s(t - t_0) * h(t) \\ &= s(t - t_0) * (h_r(t) + \boldsymbol{j}\, h_i(t)) \\ &= s(t - t_0) * h_r(t) + \boldsymbol{j}\, (s(t - t_0) * h_i(t)) \end{aligned}$$

The coefficient at the edge center is:

$$\begin{aligned} c(t_0) &= \int_\tau (x(t_0 - \tau)h_r(\tau) + \boldsymbol{j}\, x(t_0 - \tau)h_i(\tau))\, \mathrm{d}\tau \\ &= \int_\tau (s(-\tau)h_r(\tau) + \boldsymbol{j}\, s(-\tau)h_i(\tau))\, \mathrm{d}\tau \\ &= 0 - \boldsymbol{j} \int_\tau s(\tau)h_i(\tau)\, \mathrm{d}\tau \\ &= \boldsymbol{j}\, K \qquad (K \neq 0) \end{aligned}$$

Therefore, $c(t_0)$ is pure imaginary, *i.e.*, its phase is $\frac{\pi}{2}$ or $-\frac{\pi}{2}$.

Note that the correlation of $s(\tau)$ and $h_i(\tau)$ can be computed in the frequency domain

$$-\boldsymbol{j} \int_\tau s(\tau)h_i(\tau)\, \mathrm{d}\tau = \boldsymbol{j}\, \langle \boldsymbol{j}\, s(\tau)\, ,\, \boldsymbol{j}\, h_i(\tau) \rangle$$

Since $\boldsymbol{j}\, s(\tau)$ for edges from black to white and $\boldsymbol{j}\, h_i(\tau)$ typically both have positive response at positive frequency and negative response at negative frequency, the inner product should be positive. Therefore, black to white edges have center phase $\frac{\pi}{2}$ and white to black edges $-\frac{\pi}{2}$. □

Proposition 3.9 (Linear phase analytic signal). *An analytic bandpass filtered signal has linear phase* $c(t) = m(t)e^{\boldsymbol{j}\,(\omega_c t + \theta)}$ *(where* $m(t) \in \mathbb{R}$ *is the analytic magnitude), if and only* *if the Fourier transform of $c(t)$ is generalized complex conjugate symmetric about ω_c.*

$$C(\omega_c + \omega) = C^*(\omega_c - \omega)e^{j\,2\theta}$$

Proof: Here, we relax the condition of $m(t) \geq 0$ to $m(t) \in \mathbb{R}$.

If $c(t) = m(t)e^{j\,(\omega_c t + \theta)}$, let $\mathscr{F}\,\{m(t)\} = M(\omega) = M^*(-\omega)$, then,

$$C(\omega) = M(\omega - \omega_c)e^{j\,\theta}$$

$$C(\omega_c + \omega) = M(\omega)e^{j\,\theta}$$

$$C(\omega_c - \omega) = M(-\omega)e^{j\,\theta}$$

$$C(\omega_c + \omega) = C^*(\omega_c - \omega)e^{j\,2\theta}$$

If $C(\omega_c + \omega) = C^*(\omega_c - \omega)e^{j\,2\theta}$, let $m(t) = \mathscr{F}^{-1}\left\{e^{-j\,\theta}C(\omega_c + \omega)\right\}$, then we have $m(t) \in \mathbb{R}$ and $c(t) = m(t)e^{j\,(\omega_c t + \theta)}$. □

I observed that the CWT phase are approximately linear around edges and some structured textures. These features we met in images are approximately symmetric (odd or even), hence they have linear Fourier phase. The following proposition gives the necessary and sufficient condition of the coefficients of these symmetric features to have linear phase.

Proposition 3.10 (Linear phase analytic signal). *Suppose a signal $x(t)$ and the analytic filter $h(t)$ both have linear Fourier phase,* i.e., $X(\omega) = X_m(\omega)e^{-j\,(\omega t_x + \theta_x)}$ *and* $H(\omega) = H_m(\omega)e^{-j\,(\omega t_h + \theta_h)}$, *where $X_m(\omega), H_m(\omega) \in \mathbb{R}$. The filtered coefficients $c(t) = x(t) * h(t)$ has linear phase with slope ω_c, if and only if the magnitude $X_m(\omega)H_m(\omega)$ is symmetric about frequency ω_c.*

Proof: Let $X_m(\omega)H_m(\omega) = M(\omega - \omega_c)$, then we have

$$\begin{aligned} C(\omega) &= X_m(\omega)e^{-j\,(\omega t_x + \theta_x)}H_m(\omega)e^{-j\,(\omega t_h + \theta_h)} \\ &= X_m(\omega)H_m(\omega)e^{-j\,(\omega(t_x+t_h)+\theta_x+\theta_h)} \\ &= M(\omega - \omega_c)e^{-j\,(\omega-\omega_c)(t_x+t_h)}e^{-j\,(\omega_c(t_x+t_h)+\theta_x+\theta_h)} \end{aligned}$$

Therefore, $C(\omega)$ is generalized complex conjugate symmetric about ω_c, if and only if $X_m(\omega)H_m(\omega)$ is symmetric about ω_c (*i.e.*, $M(\omega) \in \mathbb{R}$ is even). According to the previous Proposition, $c(t)$ has linear phase, if and only if $X_m(\omega)H_m(\omega)$ is symmetric about ω_c. Then, we have

$$\begin{aligned} c(t) &= m(t - t_x - t_c)e^{j\,\omega_c t}e^{-j\,(\omega_c(t_x+t_h)+\theta_x+\theta_h)} \\ &= m(t - t_x - t_c)e^{j\,(\omega_c(t-t_x-t_h)-\theta_x-\theta_h)} \end{aligned}$$

where $m(t) = \mathscr{F}^{-1}\{M(\omega)\} \in \mathbb{R}$. □

From the about two propositions, it becomes clear that analytic filter output $c(t)$ has linear phase, if and only if $c(t)$ has symmetric Fourier magnitude. Since many narrow band signals are approximately symmetric about its band center frequency, they have approximately linear phase. For example, the CWT coefficients of many structured textures are narrow band analytic signals. Therefore, these signals have approximately linear phase.

Chapter 4

Spatial Image Prediction Based on the Geometrical Modeling of the CWT Magnitude and Phase

4.1 Introduction

Spatial image prediction has many important applications such as the predictive coding of images and videos, the recovery of damaged image blocks due to errors in transmission or storage, and the removal of scratches in old paintings [52, 8, 4, 5, 14, 15, 16, 17]. This prediction problem has also been known as "inpainting" among museum restoration artists. Inpainting algorithms predict or interpolate a missing or unknown region of an image from information provided by surrounding known regions based on some assumed model for images. This chapter proposes a novel inpainting algorithm that interpolates smooth regions, edges, and patterned textures in images based on simple geometrical models placed on the CWT magnitude and phase of the unknown image.

It is clear that two types of image information need to be interpolated by any reasonable image model for inpainting. Within smooth regions, gray levels of the missing region should be smoothly interpolated based on surrounding gray levels. Many linear methods (polynomial interpolation, band-limited interpolation, etc.) perform this processing well. But when surrounding pixel values indicate that some spatial structure (piece-wise smooth structure like an edge or edges, or a patterned texture) passes through the missing region, a second type of interpolation is needed. In such cases, it is perhaps clearer to view the structure itself as being interpolated, rather than the pixel values. For example, inpainting a region containing a sharp edge involves first smoothly interpolating the contour defined by the edge, and then smoothly interpolating the pixel values on either side of the edge. Similarly, inpainting a region surrounded by patterned texture involves replicating the surrounding structure smoothly through the missing region. Because this second type of interpolation involves estimating the locations of structure features, nonlinear processing approaches are necessary.

Most existing inpainting works fall into 3 categories ([17, 13]). First, diffusion based methods formulate inpainting as a variational problem and compute the missing region as the solution to a set of nonlinear PDEs used to propagate information from the surrounding areas at pixel level. [7, 8, 4, 9, 5, 10]. These variational approaches work well on piece-wise smooth image structures but poorly on textures. Second, sparse coding based approaches [14, 15, 16, 17] define the missing region as the solution to an optimization problem seeking to maximize the sparsity of the image's linear expansion with respect to some sparse representation dictionary of image basis vectors. These approaches are quite complex computationally, and their performance depends heavily on the choice and design of the dictionaries (using learned dictionaries [53] improves performance with further increased complexity). Good performance on both piece-wise smooth (cartoon-like) structure and texture image component can be achieved simultaneously by choosing a dictionary as the combination of one sub-dictionary designed for edges (e.g., the wavelet) and the other for textures (e.g., the over-complete DCT) [16]. Third, texture synthesis and examplar-based methods propagate the image information from know regions into the missing region at the patch level. Classic texture synthesis method [11] works for regions with only textures. After decomposing an image into a texture layer and a piece-wise smooth structure layer, texture synthesis and some diffusion base method can work on the two layers respectively and their results can be combined together finally [5]. Examplar-based methods propagates known patches into the missing region with composite piece-wise smooth structures and textures by defining some patch priority terms to encourage proper fill-in of patches on piece-wise smooth structures [12, 13].

This chapter proposes a much more direct and unified approach to interpolating both gray levels and spatial structures (including both piece-wise smooth structures and textures), using the magnitude and phase representation of the CWT. The proposed approach follows the image reconstruction work in the previous chapter, and also showcases the advantages of the magnitude and phase representation of the CWT. The central idea of the proposed approach comes from one observation: the missing region of an image is correctly interpolated from surrounding regions if the CWT magnitude and/or phase corresponding to that region can be correctly interpolated from surrounding regions. To interpolate the CWT coefficients in each band, I separately interpolate their magnitudes and their phases. That is, using the CWT, I translate the inpainting problem into many simpler interpolation problems of each band's magnitude and phase fields. The inpainting problem is solved if both the magnitude and phase are correctly predicted. If only the magnitude or phase is correctly interpolated, the iterative reconstruction algorithm discussed in the previous chapter may be applied to recover the correct image. The CWT magnitudes represent the local band energy, and are typically very smooth in the highest energy bands associated with edges or patterned textures. Thus, although any approach can be used to interpolate the magnitude fields, I find that very simple directional smoothing of these fields gives very good results.

The CWT phases are only significant for coefficients with large magnitudes, and, for such coefficients, the phases represent the location of the band's energy. Using linear-phase CWT filters, I find the unwrapped phase fields associated with edges and patterned texture are approximated well by linear interpolation models. In summary, piece-wise smooth structures and textures can be inpainted simultaneously by correctly interpolating the CWT magnitude and phase respectively. It should be noted that, since the very simple linear interpolation models are applied to parameters (magnitudes and phases) that are nonlinearly related to the image pixel values, they do not correspond to linear modeling assumptions on the image itself.

The proposed algorithm has the following advantages. First, the proposed method has low computational complexity, because the CWT magnitude and phase of the missing region are estimated with simple and explicit models and there are no complex nonlinear optimization problems to solve. Second, piece-wise smooth structures and textures are inpainted simultaneously by correctly interpolating the CWT magnitude and phase. Therefore, the decomposition of the image into two layers is not necessary. Third, the proposed method gives very natural looking inpainting results for edges and patterned textures. The good visual quality of the proposed simple models on edges and textures may be due to the fact that the CWT magnitude and phase match the way human visual system encoding those visual information.

The chapter is organized as follows. In section 4.2, I present the implementation and notations of the 2D CWT used in this chapter and motivate the idea of using the simple models mentioned above for predicting the CWT magnitude and phase. In section 4.3, I propose a new iterative inpainting algorithm based on the proposed models. Section 4.4

gives some simulation results and Section 4.5 concludes this chapter.

4.2 The CWT Magnitude and Phase for Images

The 2D CWT is a multi-resolution representation of images. In its magnitude and phase form, the CWT decomposes an image $f(x, y)$ into a set of magnitudes $\rho(x, y; k)$ and phases $\theta(x, y; k)$, where k is in an index set Φ of scales and orientations. The CWT magnitude represents a smoothed measurement of the local signal energy for the designated frequency band, and the CWT phase indicates the location of that energy relative to the position of each coefficient.

$$\mathrm{CWT}(f) \Rightarrow \left\{\rho(x, y; k)e^{j\ \theta(x,y;k)} : k \in \Phi\right\}$$

The frequency response of the CWT filter bank used in the implementation of this chapter is shown in Figure 4.1 [27, 33, 28]. With carefully designed filters and redundancy, the CWT is nearly alias-free. In higher dimensions, the CWT is approximately shift and rotational invariant.

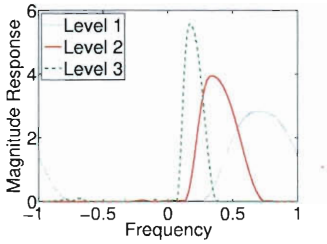
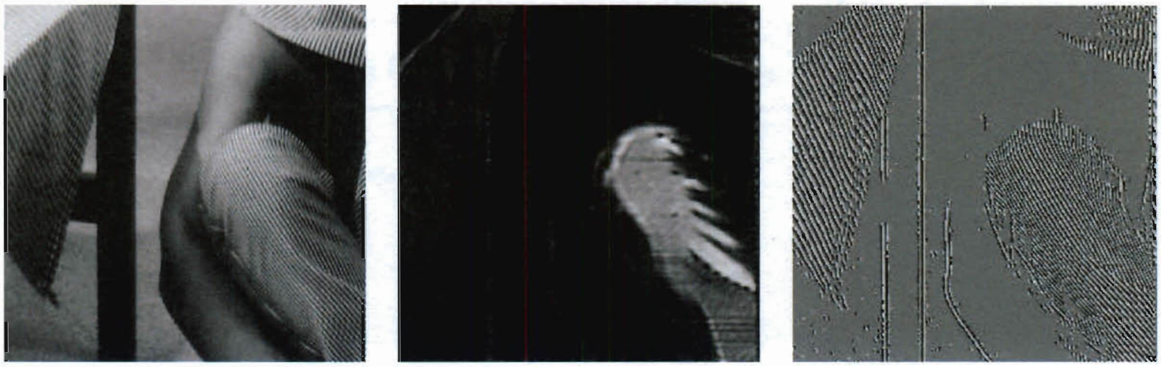


Figure 4.1 : The frequency response of the CWT on the first three scales

The CWT magnitude and phase exhibit strong geometrical properties around edges and within patterned texture areas (e.g., Proposition 3.10 gives a condition for linear CWT phase). As illustrated in Figure 4.2, the CWT magnitude is a smooth ridge-like function around edges (e.g., the arm and the chair leg). If an edge is symmetric, the CWT phase may be approximately linear under certain conditions. Patterned textures (e.g., the pants and the table cloth) can usually be decomposed into local directional narrow band 2D components by the CWT and each component has nearly constant or patterned magnitude and approximately linear unwrapped phase. Based on the above observations, I propose to employ simple 2D directional model and 2D linear model to estimate the missing magnitude and phase respectively for inpainting.



(a) Barbara (b) CWT magnitude (c) CWT phase

Figure 4.2 : The geometrical regularity of the CWT magnitude and phase

4.3 The Proposed Inpainting Algorithm

4.3.1 The Problem Formulation

Suppose in an image f, the region f_a is known and the region f_b is missing (or need to be predicted for the purpose of predictive coding). The missing region f_b has to be estimated from the information available in f_a with some assumed image model.

$$f = \begin{bmatrix} f_a \\ f_b \end{bmatrix}$$

In the CWT domain, there are roughly corresponding missing regions of the magnitude (ρ) and phase (θ) in each band. Therefore, the original inpainting problem of estimating f_b is translated into the problem of estimating (ρ_b, θ_b) in each band.

$$\rho = \begin{bmatrix} \rho_a \\ \rho_b \end{bmatrix}, \qquad \theta = \begin{bmatrix} \theta_a \\ \theta_b \end{bmatrix}$$

As discussed above, ρ and θ have strong geometrical properties around edges and within patterned textures. In section 4.3.2, simple 2D direction model and 2D linear model will be presented to describe those geometrical properties and estimate the missing magnitude and phase respectively.

$$\widehat{\rho} = \begin{bmatrix} \rho_a \\ \widehat{\rho}_b \end{bmatrix}, \qquad \widehat{\theta} = \begin{bmatrix} \theta_a \\ \widehat{\theta}_b \end{bmatrix}$$

Here, I assume that the models hold true around and within the missing region and the image with a set of magnitude and phase satisfying the models is a reasonable estimate of the original image.

If both the magnitude and phase estimation are perfect (i.e., $\widehat{\rho} = \rho$ and $\widehat{\theta} = \theta$), the image f can be recovered trivially with the inverse CWT ($\widehat{f} = \mathrm{ICWT}(\widehat{\rho}, \widehat{\theta})$). For edges, typically, the magnitude can be estimated accurately and the phase may not fit the linear model very well when getting away from the center of the edge. And for many patterned textures, the phase may fit the linear model very closely and the magnitude may not be smooth (patterned instead). In those cases, only the magnitude or phase can be estimated every well by the proposed models and the iterative reconstruction algorithm in the previous chapter can be employed to recover the missing image block. Inspired by image reconstruction from its CWT magnitude and phase, in section 4.3.3, I propose an iterative estimation algorithm to enforce the geometrical models on the magnitude and/or phase from high energy bands to low energy bands. In each iteration, the proposed algorithm looks at each CWT band, estimates and verifies the geometrical model parameters and then interpolate the magnitude and phase.

4.3.2 Geometrical Models for the CWT Magnitude and Phase

For the CWT magnitude, a simple 2D directional model works well when the missing block size is small. Suppose a block of m by m magnitudes is missing in one CWT band and denote the i-th column of the missing magnitudes as ρ_i. The columns ρ_i are modeled as different shifts of a common magnitude profile function p:

$$\rho_i = D(p, \tau_i)$$

where $D(p, \tau)$ is the shift of p with amount τ. I assume that the variable τ_i changes smoothly with column number i (it can be linear, quadratic or more complex functions of i). A linear model of τ_i is adequate for inpainting the missing blocks with size 16 by 16 in real images [1], because within such a small region image structures are close to be straight. Therefore, in this chapter, I simply choose τ_i to be linear in i when inpainting 16

[1] The block size for image and video coding is usually 16 by 16.

by 16 missing image blocks:

$$\tau_i = w_\rho i + \tau_\rho$$

where w_ρ and τ_ρ are linear model parameters.

The unwrapped phases in the missing region are simply modeled as a 2D linear function:

$$\theta_{i,j} = w_c i + w_r j + \phi$$

where i and j are the column and row numbers respectively; w_c, w_r and ϕ are 2D linear model parameters.

The magnitude and phase estimation with the proposed models is illustrated by an example shown in Figure 4.3. To interpolate the magnitude in Figure 4.3 (a) from (b), let ρ_l and (ρ_r) denote the columns of the magnitudes just to the left and right of the missing region and determine their relative shift $\widehat{\tau}$ by minimizing the difference between $D(\rho_l, \tau)$ and ρ_r. Assume that the columns of magnitudes in the missing region $(\rho_i, l < i < r)$ are shifts of ρ_l and ρ_r and the shift τ_i changes linearly with the column number i. Therefore, the shift operator D can be used to estimate all the columns of magnitudes in the missing region as below:

$$\begin{aligned}\widehat{\tau_i} &= \frac{i-l}{r-l}\widehat{\tau} \\ \widehat{\rho_i} &= \frac{r-i}{r-l}\mathrm{D}(\rho_l, \widehat{\tau_i}) + \frac{i-l}{r-l}\mathrm{D}(\rho_r, -\widehat{\tau} + \widehat{\tau_i})\end{aligned}$$

For the estimation of the missing phases, the unwrapped CWT phase is fit to a linear 2D plane with the current estimate of the magnitude as weights. If the linear model fits the phases in the surrounding areas very well, it will be used to predict the phases in the

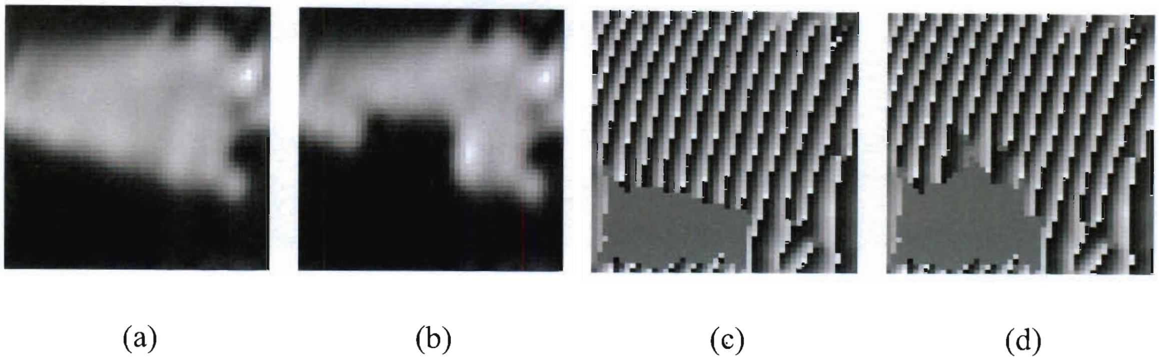


Figure 4.3 : The CWT magnitude and phase of the true image ((a) and (c)) and of the image with a missing block ((b) and (d)) (the phases associated with very small magnitudes are set to 0 for better visualization).

missing region. Otherwise, the phase is not estimated. For the case of Figure 4.3, the phase in the missing region can be recovered accurately by a linear plane.

4.3.3 The Proposed Iterative Inpainting Algorithm

To address the inpainting problem, I propose an iterative algorithm to construct an image estimate $\widehat{f}$ with the simple CWT magnitude and phase models in the previous section. In the n-th iteration, a new image estimate $\widehat{f_n}$ is obtained by applying the proposed models on the CWT magnitude and phase of the previous image estimate $\widehat{f}_{n-1}$ to predict the missing image region.

The basic idea is to recover the CWT bands with high signal energy (typically parent bands) in earlier iterations than bands with low signal energy. As discussed in the previous chapter, recovering a high energy parent band first will help reconstructing the low energy child bands from magnitude or phase (e.g., Figure 3.7). If a parent band has significant error (e.g., the edge direction is wrong in the current estimation), its child bands may have the same error and cause trouble in estimation.

Let B_k denote the CWT band with the k-th high band signal energy in the missing region. At the n-th iteration, the CWT magnitude and phase in $B_1, B_2, \ldots, B_n$ are estimated. A band B_k ($k \leq n$) is recovered trivially if both its CWT magnitude and phase can be estimated accurately with the models. If only the magnitude or the phase of B_k can be estimated accurately, B_k can still be reconstructed by the iterative reconstruction algorithm discussed in the previous chapter. The overall structure of this proposed inpainting algorithm is very similar to the iterative reconstruction algorithm in that the magnitude or phase in B_k will be repeatedly estimated and enforced in the rest of the iterations. So, B_k gets estimated for the first time at the k-th iteration and can be reconstructed after a few more iterations. After processing all the bands, the final estimate $\widehat{f}$ is obtained.

Two details have to be taken care of for the basic idea above to work. First, the actual CWT band energy in the missing region is unknown. Second, the initial estimate $\widehat{f_0}$ may contain some strong spurious edges which may cause trouble in applying the proposed magnitude and phase models and lead to poor estimate of the missing block.

To solve these two problems at the same time, I pick a threshold T_n for the n-th iteration to determine all the bands to process ($B_1, B_2, \ldots, B_n$). At the n-th iteration, only the bands with maximum CWT magnitude in the missing region of the previous iteration image estimate $\widehat{f}_{n-1}$ no less than T_n are interpolated with the proposed models. The CWT magnitude of all the bands (interpolated or not processed) will be hard thresholded by T_n before computing the new estimate $\widehat{f_n}$ to remove the influence of the spurious edges. At the beginning, set the initial threshold T_1 to the maximum of the CWT magnitude in the missing region of $\widehat{f_0}$ in all CWT bands (except the DC band). In the first iteration, T_1 is big enough such that only one band will be estimated and all the spurious edges in $\widehat{f_0}$ will be removed after thresholding with this big threshold. At the end of the n-th iteration, the maximum CWT magnitude in the missing region of $\widehat{f}_n$ are computed for all bands. The new threshold T_{n+1} is set to the maximum band magnitude just below T_n. Therefore, ideally, one new band will get into consideration in the next iteration. In this way, the maximum number of iterations N can be set to the total number AC bands of the CWT used.

The proposed algorithm is outline as follows. A final noise floor threshold T_f is specified such that the algorithm will not waste time on processing bands with only noise.

```
The Proposed Iterative Inpainting Algorithm
Given total iteration number $N$ and final threshold $T_f$
(1) Compute the initial estimate $\widehat{f}_0$ and $(\widetilde{\rho}_0, \widetilde{\theta}_0) = \mathrm{CWT}(\widehat{f}_0)$.
(2) Set $n = 1$ and compute threshold $T_1$ from $\widetilde{\rho}_0$.
(3) Determine the bands to interpolate given $T_n$.
(4) Interpolate $(\widetilde{\rho}_{n-1}$, $\widetilde{\theta}_{n-1})$ to get $(\widetilde{\rho}_n$, $\widehat{\theta}_n)$
(5) Hard threshold $\widetilde{\rho}_n$ with $T_n$ to get $\widehat{\rho}_n$.
(6) Compute the inverse CWT: $[\widetilde{f}_a^T, \widetilde{f}_b^T]^T = \mathrm{ICWT}(\widehat{\rho}_n, \widehat{\theta}_n)$.
(7) Compute new estimate $\widehat{f}_n = [f_a^T, \widetilde{f}_b^T]^T$.
(8) Compute the CWT: $(\widetilde{\rho}_n, \widetilde{\theta}_n) = \mathrm{CWT}(\widehat{f}_n)$
(9) Calculate a new threshold $T_{n+1}$ from $\widetilde{\rho}_n$.
(10) Set $n = n + 1$ and go to (3) while $T_n > T_f$ and $n < N$.
(11) Output the final result.
```

Figure 4.4 shows some examples of the iterative inpainting process. The thresholds $\{T_0, T_1, T_2, T_3\}$ are fixed to $\{85, 75, 65, 55\}$ for all the three rows for better illustration. The first column shows the input images with the missing blocks. The second column shows the initial estimates of the missing blocks. There are some spurious edges because of the initial estimate is obtained by a simple method of averaging the neighboring pixels. The rest of the columns show the inpainting results after the first four iterations of the proposed algorithm. After the first iteration with $T_0 = 85$, all the spurious edges are removed, since T_0 is large enough to clean all the AC bands. With just the first four iterations, the inpainting results look very natural.

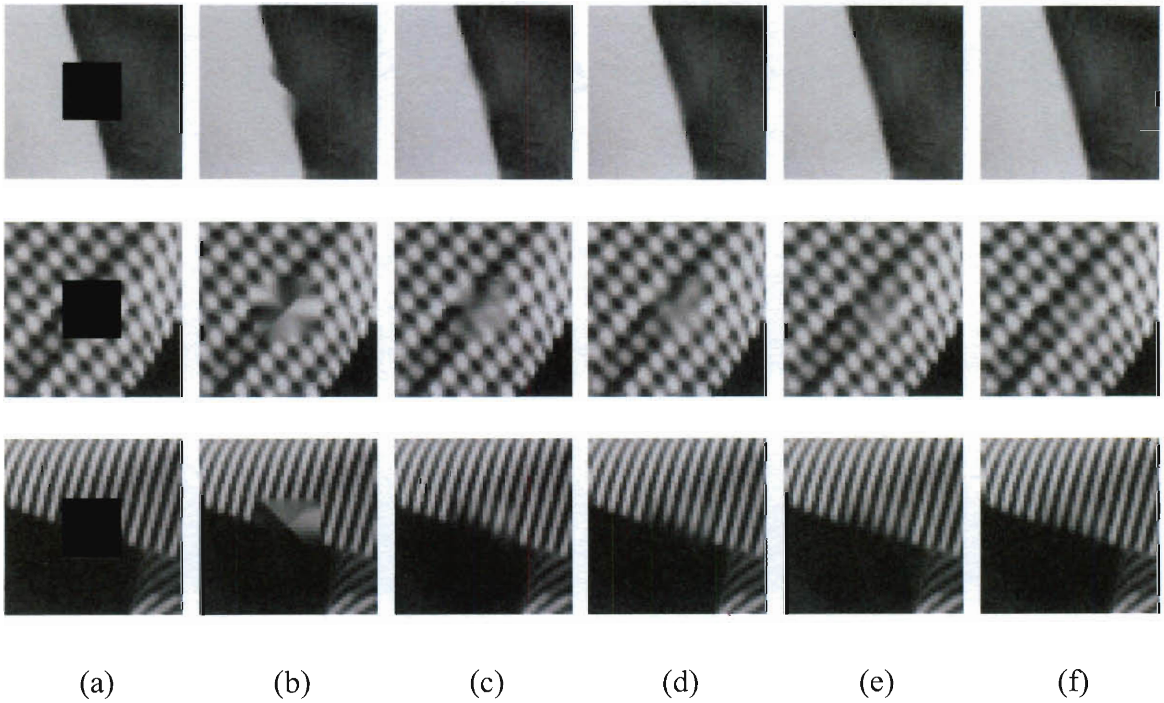


Figure 4.4 : The process of the proposed inpainting algorithm: (a) input images with missing blocks; (b) initial estimates; (c) after the first iteration with $T_0 = 85$; (d) after the second iteration with $T_1 = 75$; (e) after the third iteration with $T_2 = 65$; (f) after the fourth iteration with $T_3 = 55$.

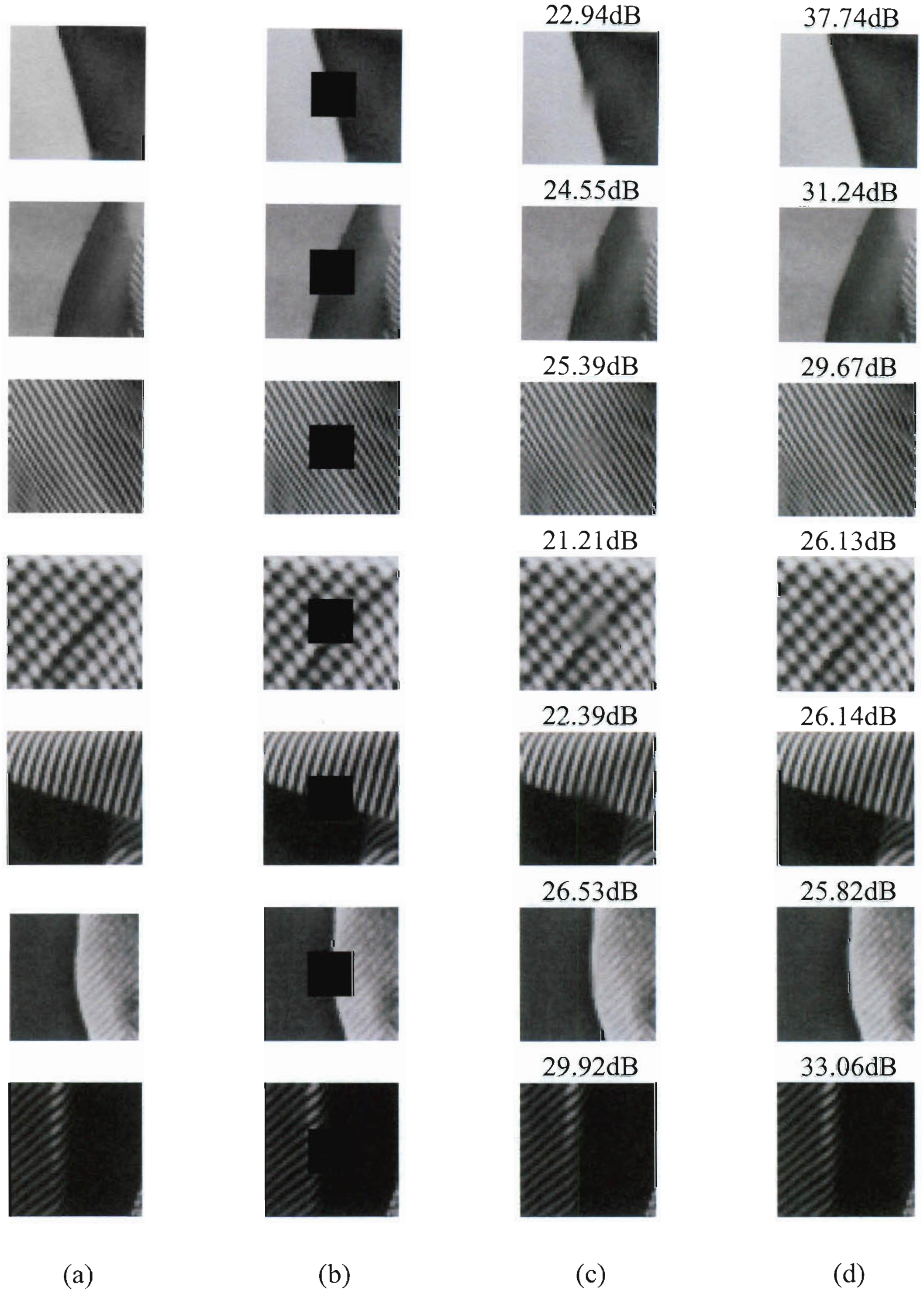


Figure 4.5 : Inpainting simulation results: (a) clean images, (b) missing blocks, (c) results of the iterated denoising algorithm, and (d) results of the proposed method (the dB numbers in (c) and (d) are the PSNR of the missing blocks)

4.4 Simulation Results

Some simulation results of the proposed algorithm are shown in Figure 4.5. A 3-level CWT is used in the simulation because it is enough for the missing block size of 16 by 16. The results of the iterated denoising method in [14, 15] with 16 by 16 DCT are also shown for comparison. To finish all the 7 shown examples, it takes less than 10 seconds with the proposed algorithm in Matlab, while the C code of the iterated denoising method takes about 260 seconds on the same computer. The image blocks are all from Lena and Barbara with 16 by 16 missing blocks. The proposed method takes only about 10 iterations to finish each of the 7 examples, since only a few number of the CWT bands needs to be interpolated. The proposed algorithm generates inpainting results with very good visual quality and mostly good PSNR. It has to be noted that PSNR is not a very good performance measure for the proposed method, because small error in the estimation of edge and patterned texture locations may results in big drop in PSNR although the visual quality still keeps about same. For example, on the forth row of Figure 4.5 (the table cloth in the standard Barbara image), the prediction result looks very natural despite the 26.1dB PSNR. On the sixth row, the original edge has a small curvature and the proposed algorithm straighten the edge because of the simple directional model on magnitude, which leads to a low PSNR of 25.8dB. A few more inpainting simulation examples are shown in Figure 4.6.

It may appear to the readers that the proposed linear models are too simple for real life images. However, they are very effective when applied on the CWT magnitude and phase and combined with the proposed iterative procedure. The performance is determined mostly by if the magnitude interpolation direction can be estimated correctly in a few CWT bands with high energy. If the estimated direction is right, edges can be recovered correctly and thus the inpainting results look natural. Sometimes, especially when complex image structures exist around the missing block, the proposed algorithm may have problem in estimating model parameters and result in edges or textures with wrong directions.

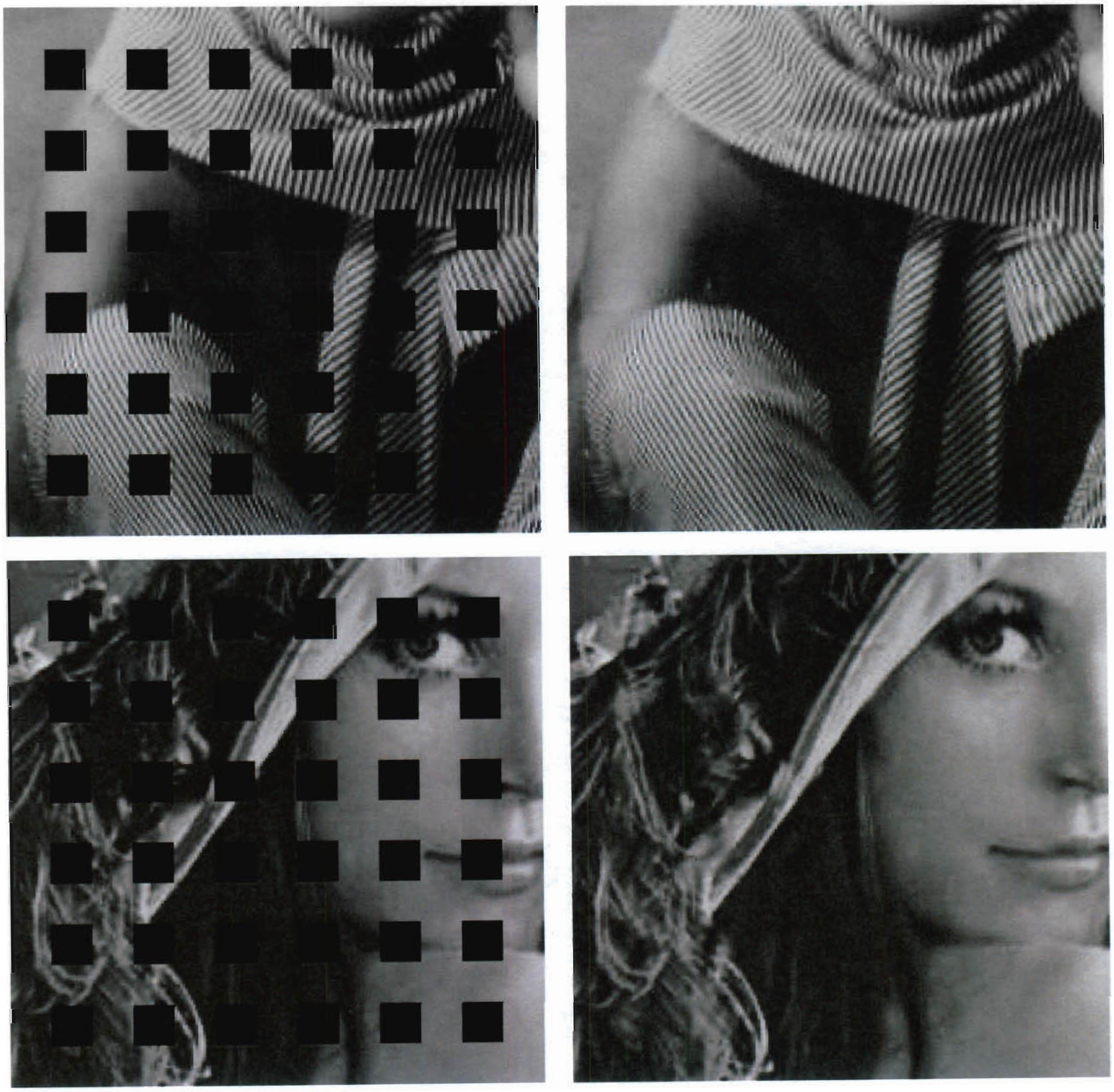


Figure 4.6 : Inpainting simulation results: (left) missing blocks, (right) results of the proposed method

4.5 Conclusions

Under the magnitude and phase representation of the CWT, edges and patterned textures can be closely approximated by employing simple linear models on the edge location and the 2D phase fields respectively. In this chapter, I constructed a new iterative inpainting algorithm by applying the above simple models and following the image reconstruction theory and algorithms developed for the CWT in the previous chapter. The proposed algorithm is very simple (linear edge location and linear phase models) and fast (about 10 iterations to finish each simulation example). It gives inpainting results with appealing visual quality for piecewise smooth signals, patterned textures and their mixtures.

It has to be noted that more sophisticated sparse coding based algorithms [16, 53, 17] may give better inpainting results through powerful nonlinear optimization techniques. The proposed algorithm, however, is seeking to solve the problem from a very different new perspective: the magnitude and phase representation of the CWT. The simplicity and effectiveness of the proposed method demonstrates the advantages of the magnitude and phase representation of CWT in dealing with important image features like edges, patterned textures, and their mixtures.

Chapter 5

Temporal Image Prediction for Hybrid Video Coding

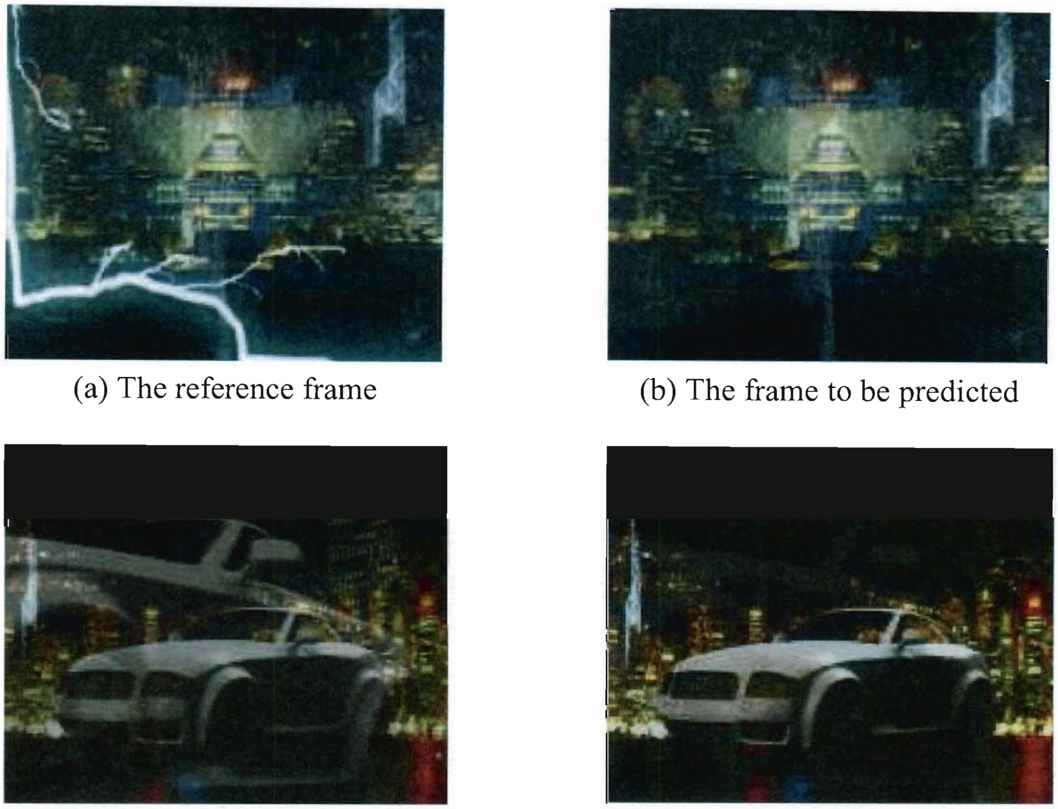
In this chapter, I propose a novel temporal (inter-picture) image prediction technique for the motion compensated prediction employed in hybrid video coding. The proposed prediction technique enables successful inter-picture predictive encoding during fades, blended scenes, intensity modulations, linear distortions (e.g., focus variations), structure clutter, temporally decorrelated noise, and many other evolutions under which motion compensated predictors used in traditional hybrid video coders fail in generating reasonable image prediction.

Under the aforementioned video evolutions, the reference frame blocks to be used in motion compensated prediction is modeled as consisting of two superimposed parts: one part that is relevant for prediction and the other part that is irrelevant (it could be noise, interference, or both). By performing prediction within a small spatial and temporal neighborhood under sparse over-complete representations (e.g., the CWT or the over-complete DCT) of the video frames, the proposed technique allows completely automated and blind learning of the evolutions of the video frames and the separation of the prediction-relevant part from the irrelevant part. This separated relevant part is then used to enable better prediction than what would be possible with the traditional block matching based prediction methods.

Experimental results on images and video frames show that the proposed method provides successful predictions under a variety of complex transitions, distortions and interferences. The proposed prediction method is also implemented to operate inside a state-of-the-art video compression codec and results show significant improvements on scenes that are hard to encode using traditional prediction techniques. The proposed algorithm can also be employed in other applications like image registration with little modification.

5.1 Introduction

As video compression matures, temporal (inter-picture) prediction techniques that try to yield significant performance improvements must concentrate on providing gains over ever more sophisticated evolutions in video. Traditional inter-picture prediction techniques rely on translated blocks from reference frames to directly match blocks in the frame to be coded. However, translated reference frame blocks may contain prediction-irrelevant interference. In simple cases the interfering signal can be as unstructured as white noise, whereas in more difficult cases, the interference can be as structured as the class of interested signals. In many types of evolutions in video, such as fades from one scene to the other, blended scenes, spatial modulations, linear distortions, temporally decorrelated noise, the prediction-irrelevant part can become severe and significantly hurt prediction accuracy, resulting in the encoding of expensive "INTRA" macroblocks. Figure 5.1 shows two examples of the reference frame consisting of both the prediction-relevant and prediction-irrelevant parts. For example, the lightning bolt in Figure 5.1 (a) will adversely affect the temporal prediction of the frame in (b) in portions of that frame; the scene fading out in Figure 5.1 (c) may render straightforward block matching prediction completely useless.



(a) The reference frame

(b) The frame to be predicted

(c) The reference frame

(d) The frame to be predicted

Figure 5.1 : Two-frame transitions from a commercial video sequence: (a) and (b) show a transition with prediction-irrelevant lightning bolt and temporally decorrelated rain drops; (c) and (d) illustrate a fade-in and fade-out transition with motion.

In this chapter, I consider the temporal evolutions over which motion compensated prediction employed by state-of-the-art codecs [54, 55] results in unusable predictors and thus non-differentially encoded (INTRA) frames and macroblocks. Figure 5.2 uses example video frames composed of standard test images to illustrate a simple subset of the temporal evolutions that can be effectively handled by using the proposed prediction technique.

The basic idea of the proposed work of this chapter is as follows. The CWT or the overcomplete DCT decomposes an image into spatially sparse and smooth magnitude and phase

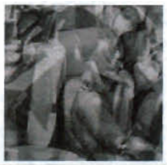
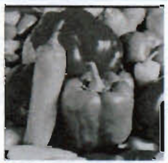
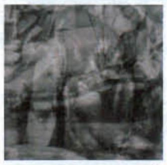
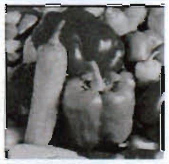
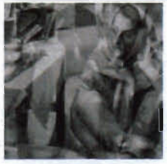
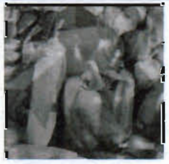
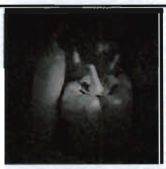
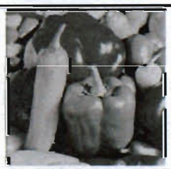
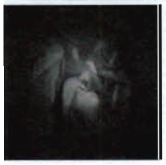
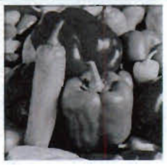
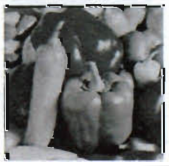
Temporal Evolution	Reference frame	Frame to be Coded	Required processing for each predicted block	Prediction	Prediction Accuracy (PSNR)
Scene transition from a blend of two scenes (peppers & barbara). (a)			Denoise, *find* peppers out of the blend of peppers & barbara, amplify peppers.		28.954 dB
Scene transition from a blend of three scenes (peppers, barbara & boat). (b)			Denoise, find peppers out of the blend of peppers, barbara & boat, amplify peppers.		26.874 dB
Scene transition with a cross-fade. One scene (barbara) fades out, the other (peppers) fades in. (c)			Denoise, find barbara, reduce barbara, find peppers, amplify peppers.		34.952 dB
Brightness change due to a lightmap. (d)			Denoise, find lightmap, invert lightmap.		31.697 dB
Scene transition from a blend with a brightness change. (e)			Denoise, find lightmap, invert lightmap, find peppers out of the blend of peppers & barbara, amplify peppers.		27.274 dB

Figure 5.2 : Example temporal evolutions that can be targeted using the prediction algorithm proposed in this chapter. The frame to be coded is predicted using the reference frame in a hybrid video compression setting. Both frames have additive Gaussian noise of standard deviation $\sigma_w = 5$. The fourth column provides a summary of the required processing for successful prediction (the proposed algorithm accomplishes these results using simple low-level prediction). Traditional motion compensated prediction results in significant prediction errors and ends up with non-differential encoding. The prediction accuracy is shown in the last column. Note that the prediction is successful even under complicated scenarios that involve brightness changes and sophisticated fades. The algorithm manages to "fish-out" scenes, recombine them, correct lighting, etc., to form these predictors.

fields. In the reference frame, the prediction relevant and irrelevant parts can typically be separated by the CWT or the over-complete DCT into different bands, and a coefficient c_i and its spatial neighborhood of $Q(c_i)$ have strong correlations. In the frame to be coded, there is a corresponding coefficient d_i and neighborhood $Q(d_i)$ at the same spatial location. Therefore, by looking at $Q(d_i)$ and $Q(c_i)$ jointly, d_i (or its magnitude and/or phase) may be temporally predicted from c_i with some assumed model trained over $Q(d_i)$ and $Q(c_i)$ ($\widehat{d_i} = P(c_i)$). In this chapter, a simple linear temporal model is used ($\widehat{d_i} = P(c_i)$), which corresponds to a linear scaling of the magnitude and a constant phase shift. Applying spatial magnitude and phase models similar to the previous chapter is conceptually possible, but is computationally prohibitive, since it has to be done on a coefficient by coefficient basis.

This chapter proposes a predictor using the over-complete DCT under which the frame to be coded and the reference frames to be used for prediction are all assumed to be sparse and smooth in space. By using the spatially causal information, the proposed predictor estimates temporal correlations, constructs predictions of the transform coefficients of the frame to be coded, and finally performs an inverse transform to obtain the equivalent pixel domain prediction. The proposed method first transforms signals to the over-complete DCT domain. Then, in this transform domain, I show that very simple predictors can be designed to isolate the prediction-relevant part of the reference blocks, learn the spatial and temporal video evolution, and perform efficient prediction. The parameters of these predictors are derived from causal information enabling completely automated and blind operation. The utilized over-complete representation allows multiple predictions for each sample in signal domain, which are averaged and combined into a single prediction. Conceptually, the proposed technique processes reference frame blocks that are about to be used in prediction so that they become much better predictors of the frame to be coded. Figure 5.3 shows the positioning of the proposed technique inside a basic hybrid video encoder.

In contrast to the proposed method, established work in inter-picture prediction is for-

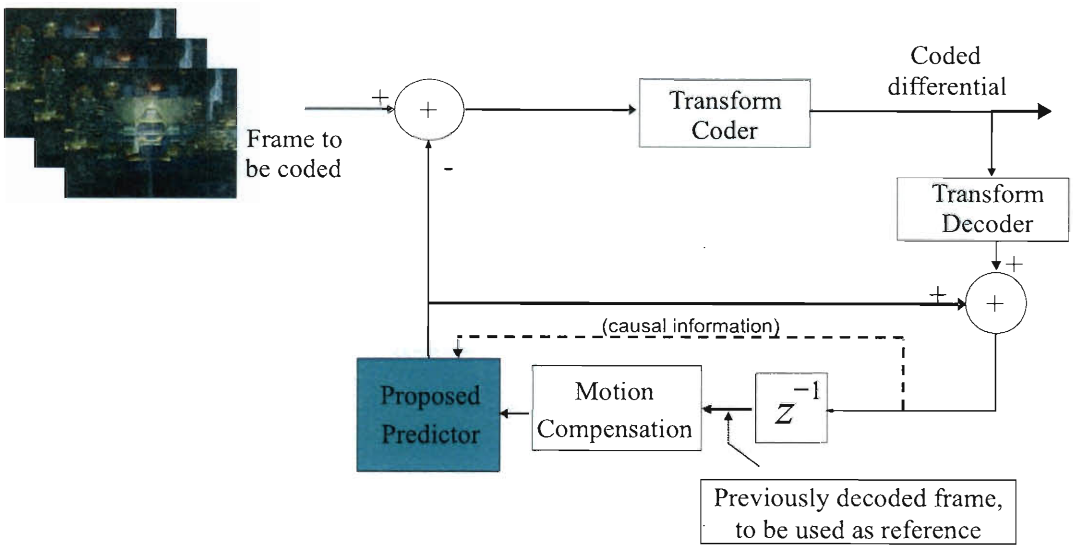


Figure 5.3 : The proposed predictor inside a basic hybrid video encoder: the proposed predictor is incorporated as a module that processes motion compensated prediction estimates so that they become better predictors of the frame to be coded.

mulated in pixel domain, where authors have designed elaborate spatio-temporal formulations that aim at estimating the correct statistics and associated predictors (see for example [56, 57, 58, 59, 60] and references therein). Despite their general formulations, many techniques are eventually designed for simplified transitions. This is because, in the absence of significant modeling assumptions, dealing with sophisticated transitions requires accurate inter-picture correlation information over sizeable neighborhoods and the solution of large and sometimes badly-conditioned systems of equations (see Appendix I in this chapter). Accuracy can also be limited as obtaining accurate correlations over image regions with non-stationary statistics is itself a difficult estimation problem. Due to such issues many established techniques restrict themselves to simplified transitions (typically, to additive noise only) and derive practical pel-recursive estimators (e.g., [56, 61]), utilize spatio-temporal Kalman filters (e.g., [57, 62]), or concentrate on Wiener filters (e.g., [58]).

Temporal prediction research in video compression has mainly concentrated on simple pixel domain parametrization that are geared towards accounting for interpolation errors, aliasing noise, and global brightness changes that may be present in the reference frames (see for e.g., [63, 64, 65]). While these techniques do form better predictors, the temporal evolution models they are targeted at are very limited compared to the predictor proposed here[1]. As the author of [66] suggested, the proposed predictor may also be related to multi-hypothesis prediction approaches. Note that filter-based techniques typically result in a small set of filters having low-pass characteristics which are not applicable on frames rich in spatial frequencies and textures (see Section 5.3 and Figure 5.4). As shown later, by only using causal information over limited spatial neighborhoods, the proposed technique can result in effective prediction over widely varying spatial statistics. Earlier research on the other hand needs to signal per-block choice of filters with overhead bits and typically requires much larger neighborhoods for the design of filters. One of the strengths of the proposed predictor is its effective transform domain parameterization which allows adaptation and accurate prediction with little training data. This is a desirable characteristic when operating over non-stationary frame statistics with localized singularities.

In the very rich image/video restoration research, authors have targeted specific models of distortions and have obtained very good results for the targeted scenarios using different forms of regularization (see for example, [67] for a recent survey of restoring common degradations in film, [68, 69, 70, 71] for prediction under noise, [72] for camera focus correction, [73] for rain-like noise removal, and [74, 75, 76] for deconvolution). While such

[1] Tools like weighted prediction [54] can target fades but require blending scenes to be available among previously decoded frames in isolated form because of motion.

techniques excel at their niche scenarios, they are computationally expensive, they typically need many correlated pictures, and they require the estimation of model parameters. They are also not robust to deviations from their niche models and to the presence of structured interference and clutter. The algorithm constructed in this chapter on the other hand is not narrowly committed to a specific corruption model in that the particular nature of the corruption (i.e., whether it is structured interference, noise, clutter, linear distortions, etc.) does not necessitate any changes in the steps of the algorithm. The proposed algorithm also does not perform any estimation of interference parameters and accomplishes high performance results that remain valid under a competitive compression setting.

Being close to the temporal prediction techniques cited above in terms of application area, the proposed predictor is conceptually closer to deblending [77, 78], denoising [79, 80, 81, 82, 83, 84], and recovery techniques [85, 36, 14] that similarly use sparse representations and exploit the non-convex structure of the sets that natural images lie in. The inverse blending techniques typically operate by using two images that are different blends of two target natural images. Their goal is to recover the two natural images assuming simple blending functions. The proposed framework is very different as one image is predicatively encoded based on the other. The proposed predictor is also more powerful as it can handle blends involving more than two images, spatially varying blending parameters, cross-fades depicting a transition from one blend to a different blend so that the image to be predicted is a blend itself, brightness and focus changes, clutter, and many other inter-picture transitions as well as their combinations. In comparison to regularized denoising setups, the "noise" that the proposed predictor removes from the reference frame is highly structured and cannot be dealt with using simple denoising techniques. Furthermore, the proposed predictor straightforwardly handles intensity modulations, linear distortions, blends, clutter, etc., which are difficult to address with simple thresholding iterations used in sparse recovery techniques. Despite the many substantial differences however, the fundamental similarity between the proposed predictor and these denoising methods is the reliance on the non-convex structure of natural image sets. As shown later this chapter, these sets are so structured that given two signals x and z from such sets, a reference signal of the form $y = x + z$ (or in fact, much more complicated reference) can be used to form accurate predictors of x or z (or much more complicated targets as well).

This chapter is organized as follows. Section 5.2 and 5.3 formulate the problem and illustrate the basic ideas of the proposed technique respectively. The main algorithm implementation for hybrid video coding is introduced in Section 5.4 and simulation results are provided in Section 5.5. In section 5.6, the proposed prediction method is slightly modified and applied to the image registration problem. This chapter concludes in Section 5.7 with some remarks.

5.2 Problem Formulation

Let $x_n \in \mathbb{R}^N$ denote the n-th frame to be predicted (and coded) and let y be the reference frame to be used in its prediction. For notational convenience assume that only one reference frame will be used in prediction and that motion compensation has taken place, i.e., if $\tilde{x}_{n-1}$ denotes the decoded reference frame, $y = \mathcal{MC}(\tilde{x}_{n-1})$, where $\mathcal{MC}$ denotes the motion compensation operation. In the following, the subscript n of x_n is dropped for convenience when there is no confusion.

The motion compensated reference frame y is assumed to be consisted of two parts: one prediction-relevant part ($l \odot (s * x)$) and one prediction-irrelevant part ($z + w$):

$$y = l \odot (s * x) + z + w \tag{5.1}$$

where l denotes a band-limited intensity modulation signal, $\odot$ denotes per component multiplication, s is a linear spatial filter, $*$ denotes linear convolution of appropriate dimensions, z is structured interference caused by a signal with similar characteristics as x, and w is white noise.

In typical scenes, l can be used to model spatial lighting variations such as shadows and diffuse light that manifest themselves as intensity modulation of the signal; s models the variations of the point spread functions of image capture devices, post-processing operations, and optical lens focus (when nearby objects are focused, far away objects are blurred, and vice versa); z can be, for example, due to specular lighting, due to other scenes fading in or out, or due to special visual effects introduced into the video in post processing. In the simulations of this chapter, w will be white. However, it can easily be generalized to include quantization noise in y.

The problem formulated in equation 5.1 is the composition of a signal denoising or separation problem and a blind linear inverse problem. The available information includes y and the spatial causal part (already coded) of x. Note that straightforward application of denoising techniques [79, 86, 38, 77] can deal only with w, and restoration techniques [87, 75, 74] with w and with known s. As shown later, the proposed method provides a solution that is substantially more general. The algorithm proposed in the following section can handle spatial variations in both w and in s, provides completely blind and automated operation with no parameter estimation, and perhaps most importantly, allows operation when one or more interfering signals like z are present.

5.3 The Basic Ideas

To construct a good prediction of x given y and the spatially causal part of x, we need to separate the prediction-relevant part ($l \odot (s * x)$) and the prediction-irrelevant parts (z and w), identify the relevant and reject the irrelevant, learn the smooth lighting map l and spatial filter s, and invert both of them.

The key steps of the proposed predictor are as follows:

1. Separate the prediction-relevant and prediction-irrelevant parts with an appropriate over-complete sparse transform.

2. Learn the lighting map and spatial filtering in the sparse transform domain with the causal neighborhood information.

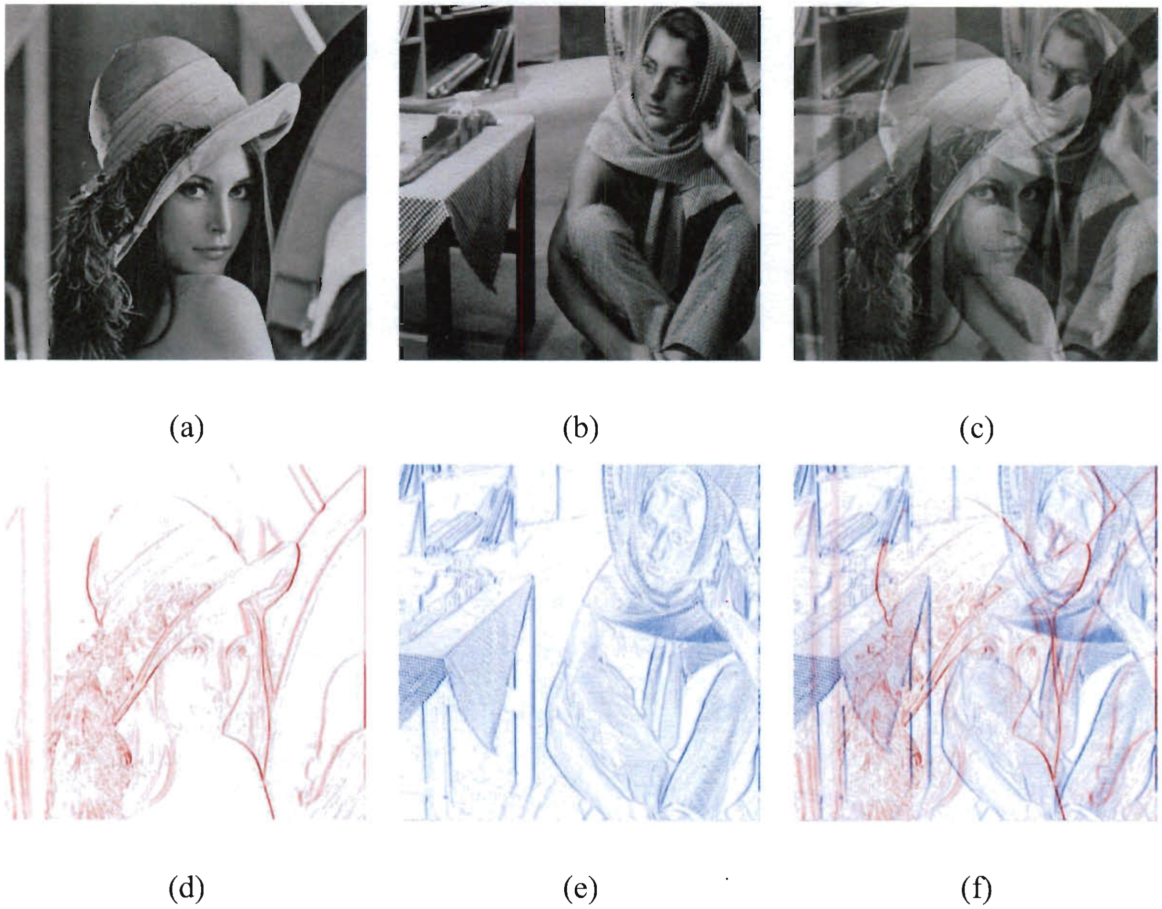
3. Construct a temporal linear predictor in the sparse transform domain to invert the lighting map and spatial filtering.

4. Form the image domain prediction with proper iteration and progression in the sparse transform domain.

5.3.1 The Separation of the Relevant and Irrelevant in a Sparse Transform Domain

As formulated in the previous section, we assume that (1) the relevant part $l \odot (s * x)$ is a natural image (a modulated and filtered version of another natural image x), (2) the irrelevant part z is also a natural image, and (3) the irrelevant part w is white Gaussian noise. So, the relevant part and the irrelevant part z are typically sparse under certain over-complete sparse transforms, such as the (complex) wavelet and the over-complete DCT. Because of their sparsity, those parts will rarely overlap with each other in the sparse transform domains. Also, the white Gaussian noise w is typically spread out evenly in the sparse transform domains. It overlaps with the sparse relevant part with only a very small fraction of its total energy. Therefore, the three parts $l \odot (s * x)$, z and w can be separated automatically by the sparse transforms. This automatic separation of the relevant and irrelevant parts by the sparse transforms is illustrated with a pictorial example in Figure 5.4.

Figure 5.4 shows the mixture of the relevant part (the Lena image) and the irrelevant part (the Barbara image). In the image domain, the two parts (Lena and Barbara) are mixed together everywhere (Figure 5.4 (a,b,c)); Therefore, it is very hard to separate them from each other without knowing exactly how the two images are mixed together. In contrast, in the over-complete sparse transform domain, the two signals rarely overlap with each other spatially. In Figure 5.4 (d,e,f), the signal component of Lena is color-coded as the red channel and the signal component of Barbara is color-coded as the blue channel. When Lena and Barbara are mixed together in Figure 5.4 (f), the overlapping of the two is color-coded as the mixture of the red channel and the blue channel, so that it will be a color different from either red or blue (the color is close to some shade of purple depending on the relative strength of the two signal components). As we can see in the figure, at most places the color is still either red or blue, because the signals of Lena and Barbara rarely overlap. The color becomes purple, only at very limited locations where the two signals



(a) (b) (c)

(d) (e) (f)

Figure 5.4 : Example images (a) Lena, (b) Barbara, and (c)Lena-Barbara average. The pictures in (d), (e) and (f) illustrate synthetic images of transform coefficient magnitudes for (a), (b) and (c) respectively. The pictures in (d) and (e) are obtained by applying a translation invariant decomposition of 4 by 4 block DCTs to the top row images, taking the coefficients with block index (0, 1), obtaining their magnitude, and spatially arranging the resulting values to reflect the relative translations. Darker colors show larger coefficients. (f) shows (d) and (e) as two color channels to demonstrate overlaps.

overlap, because of the sparsity of the two signals in the transform domain.

In a short summary, sparse transforms automatically separate the prediction-relevant and prediction-irrelevant parts. However, we still need to identify whether a transform coefficient at a particular location is from the relevant part or from the irrelevant part.

To distinguish the relevant part $l \odot (s * x)$ from the irrelevant part z, the spatial smoothness of x and z under the used transform has to be exploited. That is, in the over-complete sparse transform domain, the signal (x or z) at the any location is assumed to have similar energy density as in the causal neighborhood. This assumption has been used in the previous chapter for inpainting. It can be intuitively justified by the smoothness of the transform coefficients of Lena and Barbara shown in Figure 5.4 (d) and (e). Therefore, the transform coefficient at the current location is the relevant part, if in the causal neighborhood the relevant part has significant energy; otherwise, it is the irrelevant part [2]. In the following, this idea is generalized to construct an adaptive linear temporal predictor in the sparse transform domain which can not only distinguish the relevant part, but also lean and invert the spatial modulation and filtering effects.

5.3.2 The Proposed Linear Temporal Predictor in the Sparse Transform Domain

The problem formulation of Equation 5.1 can be greatly simplified in the over-complete sparse transform domain with some mild assumptions on the spatial modulation (l) and filtering (s) effects.

First, the spatial modulation (l) is assumed to vary slowly in space, so that it can be considered a constant in any spatial neighborhood which is sufficiently larger than the support of the sparse transform basis vectors. This is a very mild assumption especially for the over-complete DCT employed for video coding implementation later in this chapter. For example, when the 4x4 DCT is chosen, the spatial modulation only needs to be approximately constant in every 4x4 pixel neighborhood. Therefore, under this assumption, the

[2] If the relevant part has exactly the same shape as the neighborhood boundary, we may erroneously treat the current coefficient as irrelevant. However, for real life images , it rarely happens.

modulation effect of l becomes a constant scaling of the coefficients in the sparse transform domain.

Second, we assume that the filter s has a relatively smooth frequency response and it also varies slowly in space. In the transform domain, while forming $(s * x)$, s modifies the transform coefficients of x that correspond to a particular "frequency band" by modulating the coefficients with a locally smooth factor. Therefore, the transform coefficients of x in the same frequency band and over spatially close regions are modulated with similar factors to form the coefficients of $(s * x)$. Hence, overall, we assume that the filtering effect of s can be approximately diagonalized by the over-complete sparse transform (e.g., the complex wavelet or the over-complete DCT). In summary, if a linear filter s has smooth frequency response and varies slowly in space, then s can be approximately diagonalized by the CWT or the over-complete DCT. The implication and limitation of this assumption is analyzed later in Equation 5.6.

With the above two assumptions, we can use the following equation to approximate Equation 5.1 in the over-complete sparse transform domain:

$$c_y(k) = \alpha(k)c_x(k) + c_z(k) + c_w(k) \tag{5.2}$$

where c_x, c_y, c_z and c_w are the transform coefficients of x, y, z and w respectively, α approximates the overall influences of l and s, k denotes the spatial location within each "frequency band", and the frequency band index is dropped for clarity and simplicity.

The most important property of Equation 5.2 is that α can be assumed to vary slowly in space in each frequency band. According to the discussion above, both l and s vary slowly in space. Therefore, we are able to assume that $\alpha(k)$ also varies slowly with location k

in every frequency band. This smoothness property of $\alpha(k)$ greatly simplifies the learning and inverting of the spatial modulation of l and the filtering effect of s.

I propose the following adaptive linear temporal predictor $P_L\left(c_y(k)\,;\,\gamma(k),\Delta(k)\right)$ to estimate $c_x(k)$ with given $c_y(k)$ based on Equation 5.2.

$$P_L\left(c_y(k)\,;\,\gamma(k),\Delta(k)\right) = \gamma(k)c_y(k) + \Delta(k) \tag{5.3}$$

The predictor parameters $\gamma(k)$ and $\Delta(k)$ can be trained in the causal neighborhood, since $\alpha(k)$ can be assumed to be smooth in the neighborhood.

In the following, I will introduce a simple solution to the above equation along this line.

5.3.3 The Learning and Inverting of the Spatial Modulation and the Filtering Effect

With the above assumptions, I propose to solve the problem in Equation 5.2 by estimating the transform coefficients $c_x(k)$ with the following linear predictor:

$$\begin{aligned} \widehat{c}_x(k) &= P_L\left(c_y(k)\,;\,\gamma(k),\Delta(k)\right) \\ &= \gamma(k)c_y(k) + \Delta(k) \\ \{\gamma(k),\Delta(k)\} &= \arg\min_{\gamma,\Delta} \sum_{q\in Q_k} \|c_x(q) - \gamma c_y(q) - \Delta\|^2 \end{aligned} \tag{5.4}$$

where Q_k is the causal neighborhood at location k within which $c_x(q)$ and $c_y(q)$ are available. That is, the optimal parameters $\{\gamma(k),\Delta(k)\}$ can be determined via well known least-squares estimation technique.

With the estimated transform coefficients $\widehat{c}_x(k)$ for every k in every frequency band, the prediction $\widehat{x}$ can be constructed with the iterative inversion of the over-complete sparse transform in a similar way as the inpainting algorithm discussed in the previous chapter.

Suppose $\{\phi_k\}_{k=1}^{K}$ and $\{\phi'_k\}_{k=1}^{K}$ are respectively the sets of analysis and reconstruction basis functions of the sparse transform used (e.g., the DCT or the CWT), then,

$$\widehat{x} = \frac{1}{K} \sum_{k=1}^{K} \phi'_q * \widehat{c}_x(k) \tag{5.5}$$

I also observed that starting from $\widehat{c}_x(k)$ and iterating on the magnitude $|\widehat{c}_x(k)|$ may sometimes improve the reconstruction quality.

To see the implications on the spatial filter s, consider the simple deconvolution problem of $y = s * x$. Suppose that coefficients in sub-band k of y are multiplied with a prediction weight $\gamma(k)$ in order to obtain the predictor.

$$\begin{aligned}
\widehat{x} &= \frac{1}{K} \sum_{k} \phi'_k * (\gamma(k)\, \phi_k * y) \\
&= \frac{1}{K} \sum_{k} \gamma(k)\, \phi'_k * \phi_k * (s * x) \\
&= \frac{1}{K} \sum_{k} (\gamma(k)\, \phi'_k * \phi_k) * s * x
\end{aligned}$$

Hence, as long as the optimal deconvolution filter, s^{-1}, can be closely approximated by the following equation

$$s^{-1} = \frac{1}{K} \sum_{k} \gamma(k)\, \phi'_k * \phi_k \tag{5.6}$$

for some $\{\gamma(k)\}_{k=1}^{K}$, the optimal filter will be in the span of predictors that can be constructed by the proposed algorithm above. It also has to be noted that there are structural limitations on such s^{-1}. For example, with localized orthonormal transforms, Equation 5.6 corresponds to symmetric filters of limited support.

5.3.4 A Motivating Experiment Demonstrating the Basic Ideas

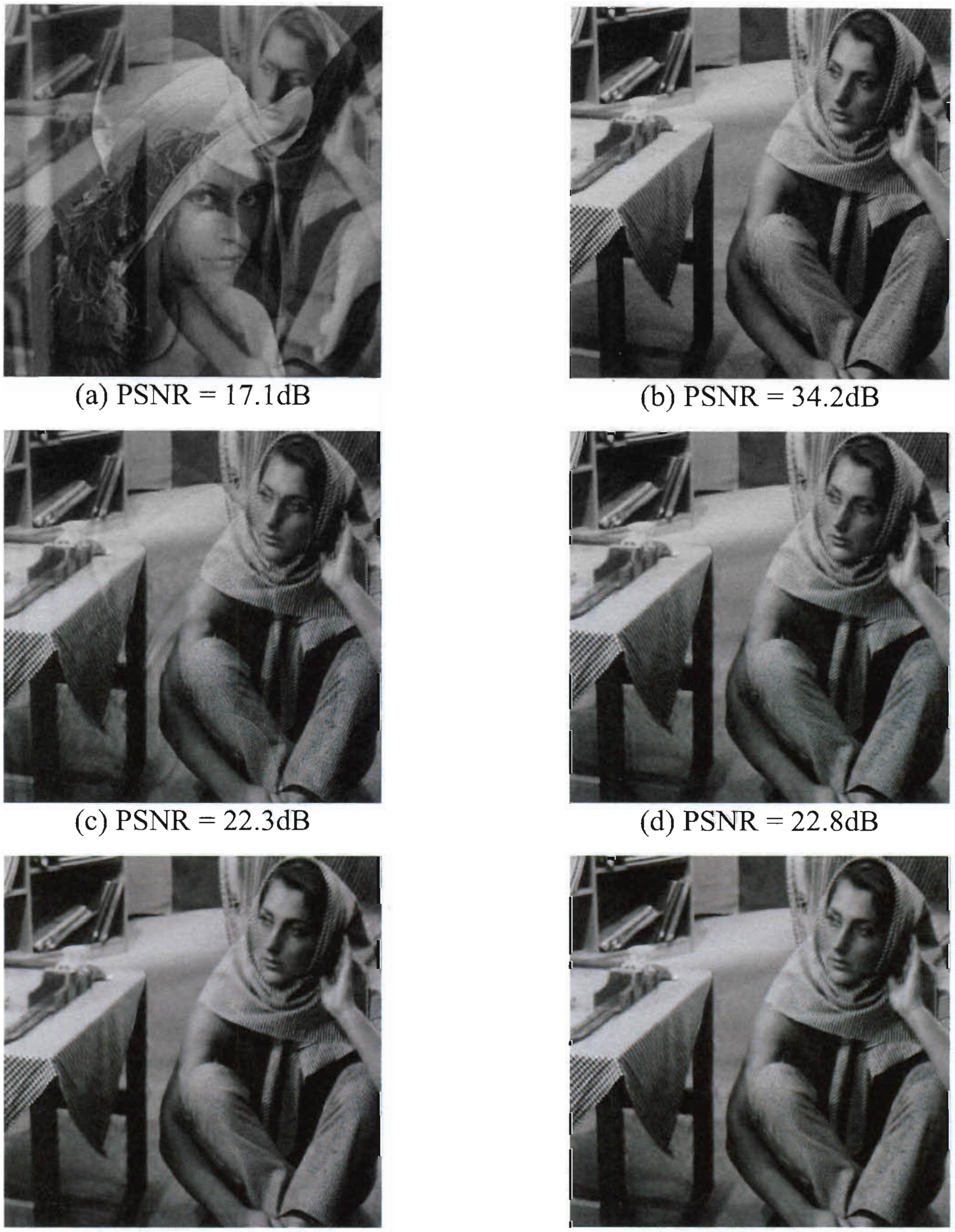
The above ideas about the proposed predictor is demonstrated by the experiment shown in Figure 5.5. In this experiment, we ignore the complexity in implementation with video coding and focus on showing the effectiveness of the proposed ideas. The full algorithm implementation for video coding will be introduced in the next section.

In this experiment, the spatial modulation l is assumed to be constant; the spatial filtering effect s is a Gaussian low-pass filter; the frame to code x is Barbara (Figure 5.5 (b)); the irrelevant part z is Lena; the reference frame (Figure 5.5 (a)) is the average of the filtered Barbara and Lena. Both the reference frame and the frame to code contain additive wight Gaussian noise with $\sigma = 5$. We use the same CWT as in the previous chapter and perform coefficient-wise prediction as in Equation 5.4.

To illustrate the influence of the magnitude and phase, the second row of Figure 5.5 shows the performance of a similar solution to Equation 5.4 which estimates the magnitude of the current frame transform coefficient $|\widehat{c}_x(k)|$ from the magnitude of reference frame coefficient $|c_y(k)|$.

$$\begin{aligned} |\widehat{c}_x(k)| &= \gamma(k)|c_y(k)| + \Delta(k) \\ \{\gamma(k), \Delta(k)\} &= \arg\min_{\gamma,\Delta} \sum_{q \in Q_k} \| |c_x(q)| - \gamma |c_y(q)| - \Delta \|^2 \end{aligned} \qquad (5.7)$$

This method is similar to inpainting with the CWT magnitude in the previous chapter. Its performance shall be the same as Equation 5.4 where c_x is significant and c_y is negligible. Where c_x is negligible and c_y is significant, it keeps the phase of c_y and leaves an apparent trace of the irrelevant part (Lena) in the prediction (see Figure 5.5 (c)). After iterated reconstruction from the estimated CWT magnitude, the trace of the Lena goes away as explained



(a) PSNR = 17.1dB

(b) PSNR = 34.2dB

(c) PSNR = 22.3dB

(d) PSNR = 22.8dB

(e) PSNR = 26.5dB

(f) PSNR = 26.8dB

Figure 5.5 : An experiment to demonstrate the prediction ideas: (a) the reference frame y (the average of Lena and blurred Barbara plus white Gaussian noise, PSNR = 17.1dB; blurred Barbara relative to clean Barbara has PSNR = 24.3dB) (b) the frame to be coded x (Barbara plus white Gaussian noise, PSNR = 34.2dB); (c) magnitude prediction result (PSNR = 22.3dB); (d) magnitude prediction and iterated reconstruction result (PSNR = 22.8dB); (e) proposed prediction result: denoised and sharpened (PSNR = 26.5dB); (f) proposed prediction and iterated reconstruction result (PSNR = 26.8dB)

in the previous chapter (Figure 5.5 (d)). The overall PSNR performance is improved from 22.3dB to 22.8dB.

In contrast, the solution of Equation 5.4 takes advantage of the decorrelated phases of the two parts and estimates the phase of c_x as much as possible (this can remove the adverse influence of the phase of c_y even when the phase of c_x is not predictable with the simple training in the causal neighborhood). Therefore, in Figure 5.5 (e), it is much harder to find any trace of Lena. The PSNR of the prediction in Figure 5.5 (e) is 26.5dB and the PSNR of the blurred Barbara (relative to the clean Barbara to be predicted) in the reference frame is 24.3dB. It shows that the predictor successfully isolate the relevant part and invert the blurring filter s by at least 2.2dB (from 24.3dB to 26.5dB). If iterative reconstruction from magnitude is performed, the PSNR can be further improved slightly to 26.8dB (Figure 5.5 (f)).

The above experiment does not consider the macroblock structure of video codecs. For ease of implementation, I choose the over-complete DCT in the next section and describe the details about constructing a predictor for state-of-the-art hybrid video coding.

5.3.5 A Short Summary of the Proposed Temporal Predictor

1. The temporal (or inter-picture) image prediction problem considered in this chapter is formulated in the image domain by Equation 5.1

$$y = l \odot (s * x) + z + w$$

2. With some mild conditions, the problem is translated to a sparse transform domain

as in Equation 5.2.

$$c_y(k) = \alpha(k)c_x(k) + c_z(k) + c_w(k)$$

3. An adaptive linear prediction solution is proposed in Equation 5.4.

$$P_L\left(c_y(k)\,;\,\gamma(k),\Delta(k)\right) = \gamma(k)c_y(k) + \Delta(k)$$

$$\{\gamma(k),\Delta(k)\} = \arg\min_{\gamma,\Delta}\sum_{q\in Q_k}||c_x(q) - \gamma c_y(q) - \Delta||^2$$

4. With $\widehat{c}_x(k) = P_L\left(c_y(k)\,;\,\gamma(k),\Delta(k)\right)$, the image domain prediction can be easily constructed with simple inverse transform or iterated reconstruction.

5.4 The Main Algorithm with the DCT for Video Coding

In this section, the main algorithm implementation with the DCT for video coding is introduced in details. The implementation follows the basic ideas in the previous section and also takes into account the macroblock structure required by video encoders and other implementation details. The iterated reconstruction is implemented as progressive estimation and reconstruction because of the limitation imposed by video encoding.

Let $\mathbf{H}_1,\ldots,\mathbf{H}_M$ $(N \times N)$ denote an over-complete set of linear transforms. For issues of implementation simplicity and computational speed, in this section $\mathbf{H}_i,\ i = 1,\ldots,M,$ are given by a $p \times p$ block DCT transform and its $p^2 - 1$ shifts with $M = p^2$. Please note however that it is straightforward to utilize the proposed predictor with different transforms and different redundancy factors.

We would like to implement M temporal predictors of x_n which are denoted as $\hat{x}_n^i,\ i =$

$1, \ldots, M$. Each predictor is simple and is composed via

$$c_i = \mathbf{H}_i^T y \tag{5.8}$$

$$d_i(j) = \gamma_i(j) c_i(j), \; j = 1, \ldots, N \tag{5.9}$$

$$\hat{x}_n^i = \mathbf{H}_i d_i, \tag{5.10}$$

where $\gamma_i(j)$ are linear prediction weights that are determined to minimize mean squared prediction error. The final prediction of x_n is obtained as an average of the M predictors via

$$\hat{x}_n = \frac{1}{M} \sum_{i=1}^{M} \hat{x}_n^i, \tag{5.11}$$

though weighted averaging techniques similar to [6] can also be used. Of course, if desired, it is straightforward to augment the basic prediction setup in (5.8) through (5.11) to account for further correlations among coefficients.

Modern high-performance video compression involves many steps and optimizations that aim at satisfying several important requirements. Macroblock structure utilized in codecs and the order in which macroblocks are coded present practical issues for any temporal prediction algorithm. In this section, we discuss our main algorithm that is geared towards operation inside a high performance video codec (h264/AVC reference software - JM10.2 [54]) in macroblock units. The main concern is the design of an algorithm that obtains the prediction weights in (5.9) using available information during the decoding of each macroblock.

Since the utilized transforms are block based, the transform domain prediction operation can be thought of as predicting $p \times p$ blocks in x_n using the corresponding blocks in y

and then averaging the predictions suitably to satisfy (5.11). In order to help notation, let us adopt a block viewpoint and deviate slightly from the frame-wide notation of Section 5.3. Suppose b_x ($p^2 \times 1$) represents a $p \times p$ block in x_n that is to be predicted and let b_y be the corresponding block in y. We know the transform coefficients of b_y which will be used via (5.11) to determine predictions of the transform coefficients of b_x. We need to determine the per coefficient prediction weights[3].

Figure 5.6 (a) illustrates b_x within the frame to be coded (current frame), inside the macroblock to be coded (current macroblock). In determining the prediction weights we would like to utilize the information provided by previously decoded macroblocks inside the current frame[4]. Let Λ_{b_x} denote a $L \times L$ spatial neighborhood around b_x. Consider all blocks formed by shifting a $p \times p$ mask inside Λ_{b_x}. Denote those blocks that completely overlap known data (training data) by $t_1, \ldots, t_Q$. Let $u_1, \ldots, u_Q$ be the corresponding blocks in y.

Block prediction: Suppose we are performing the prediction for the k^{th} DCT coefficient of b_x. Let h_k ($p^2 \times 1$) denote the k^{th} DCT basis function. We derive the prediction weight γ_k as the weight that minimizes the mean squared prediction error on the training data, i.e.,

$$\gamma_k = \operatorname*{argmin}_{\gamma} \sum_{q=1}^{Q} ||h_k^T t_q - \gamma h_k^T u_q||^2. \tag{5.12}$$

[3] For the moment we leave issues related to motion estimation to Section 5.5 and keep assuming that y is such that translations are properly accounted for.

[4] If previously decoded macroblocks are not available one can set the prediction weights to an appropriate value (such as 1) or determine optimal weights and send them using overhead bits.

The prediction for the k^{th} DCT coefficient in b_x is then set as

$$\widehat{h_k^T b_x} = \gamma_k h_k^T b_y. \tag{5.13}$$

Once the prediction is carried out for all DCT coefficients of b_x, an inverse block DCT yields the pixel domain prediction $\hat{b}_x$. Observe that the solution of (5.12) is straightforward and the computations for its calculation can be reused when predicting other blocks so that complexity remains contained.

Given the desired predictor in (5.11), it is clear that we need to predict all $p \times p$ blocks in the current macroblock, i.e., all shifts of a $p \times p$ block that overlap the current macroblock. H.264/AVC utilizes a 4×4 transform (compression transform) to encode prediction errors. In forming predictors for our blocks, we try to utilize available data as much as possible so that compression transform coefficients of the prediction error are used to augment the known data regions of the current macroblock as soon as they become available. Note however that the exact pattern in which compression transform coefficients are sent depends on the motion modes determined at the motion estimation stage [54]. For example, the motion mode where motion blocks are 8×8 sends coefficients in a different order compared to the 16×16 mode. As such, our technique orders block predictions based on the motion mode in order to utilize prediction error updates as much as possible. In order to save space, further discussion of these detailed but conceptually straightforward implementation issues will be ignored in the presentation. The proposed implementation also utilizes predictions provided for previous blocks to augment the training data by scanning the current macroblock in layers (see Figure 5.6 (b) for an example scan). The procedure below outlines the proposed prediction of a macroblock as carried out by a hybrid video decoder:

5.6 Application in Image Registration

The proposed temporal image prediction method can also be successfully applied to the image registration/matching problems. In order to motivate this application, let us consider the case of predicting a target image (say, the image peppers) using an unrelated source image (say, the image barbara). Assuming sufficiently large neighborhoods are used in generating the training pairs, Equation (5.4) is expected to result in $\gamma(k) = 0$ (zero prediction coefficient), but in general the mean predictor Δ will be close to the mean of the target training region, especially for the DC sub-band. As a consequence, the algorithm will form nonzero predictors using information from the target image, even when the source image is completely unrelated. In a compression setting such predictors are beneficial and allow compressing target blocks around their causally calculated means. In registration/matching cases, however, it becomes a problem because one can declare a match over unrelated images. Therefore, for registration/matching applications, the fixed-mean predictor $P_F(c_y(k)\,;\,\gamma(k))$ below is proposed instead. Also, the training region is not limited to the causal region as in the coding scenario.

$$P_F(c_y(k)\,;\,\gamma(k)) = \gamma(k)c_y(k) + h_k^T\eta \tag{5.14}$$

$$\gamma(k) = \arg\min_{\gamma} \sum_{q\in\widetilde{Q}_k} \|c_x(q) - \gamma c_y(q)\|^2 \tag{5.15}$$

where $\widetilde{Q}_k$ is the non-causal training region, h_k is defined in Equation 5.12, and η is a constant vector with a given value that quantifies a generic average for image pixel values (e.g., 128 for 8-bit images).

In Figure 5.10, the proposed prediction method is applied in an affine registration setting where the goal is to find the affine warp parameter that quantify the geometric relation between the source and target images. The affine motion estimation technique of [88] is used as a baseline registration algorithm. The baseline algorithm calculates spatial and temporal derivatives in a coarse to fine, multi-resolution fashion to obtain the affine warp parameters. While only locally optimal, this technique provides high performance and robust results for many examples (see [88]). The aim in this section is to augmented the baseline algorithm with the proposed temporal prediction method in order to derive a new affine image registration technique. The new algorithm is identical to the baseline except that the spatial and temporal derivatives are calculated after prediction using the temporal prediction method proposed in this chapter.

Figure 5.10 (a) illustrates the case involving clean signals. The source image, target image, the source image registered with the ground-truth warp, the source registered using the warp calculated by the baseline technique, and finally the source image registered using the warp calculated by the proposed augmented algorithm are shown from left to right. Both algorithms obtain the correct warp and the source is correctly registered. The Frobenius norm of the affine warp error is shown for the calculated warps.

The second example involves a source that is heavily corrupted with Gaussian noise ($\sigma_w = 200$). The baseline technique diverges from the correct solution while the proposed method remains very close to it. The same can be observed in Figures 5.10 (c) and (d) which are corrupted with a sine wave and an unrelated image respectively. Again, proposed registration method obtains results that are very close to the ground truth while baseline by itself diverges.

a specific corruption model in that the particular nature of the corruption (i.e., whether it is structured interference, noise, clutter, linear distortions, etc.) does not necessitate any changes in the steps of the algorithm. The same steps that perform deconvolution reject highly structured interference, denoise, recover missing pixel values, separate and recombine scenes, and so on. The algorithm does not perform any estimation of corruption parameters and accomplishes high performance results that remain valid under a competitive compression setting. It also has to be noted that many other sophisticated image transitions cannot be dealt with using this version of the proposed algorithm (e.g., transitions involving complex motion effects such as transparent motion [89, 90]).

It is important to point out that some of the errors in the proposed block-recursive algorithm are naturally due to causality and can easily be avoided in a compression setting by sending prediction parameters as overhead. The results based on causal predictors are motivated by ease of integration within an established video codec and its established syntax. An optimized video coder specifically built around the proposed temporal predictor is expected to obtain significantly improved results.

The proposed predictor in this chapter can be improved by using more sophisticated decompositions and by allowing for more elaborate predictors. In terms of decompositions, beyond the transform optimization of Appendix I and various established designs (e.g., [91, 92]), one can also utilize adaptive transform optimizations that maximize the sparsity of the decomposition (e.g., [93, 94]) or recent work that provide various adaptive reconstructions from expansive decompositions (e.g., [83, 36]). The predictor can be generalized by using various kernel-based techniques (e.g., [95]) and also incorporate geometrical regularity of image singularities (e.g., [96, 94]) so that the interference is better rejected. Of course, when necessary, the reach of this work can be improved by employing high-level estimation of corruption parameters.

5.8 Appendix I: prediction-optimal transforms

In this section we derive transforms that optimize the prediction of target random vectors, denoted by v, using anchor random vectors, denoted by u. Assume that u and v are adjusted to have zero-mean. Let us first consider the general case where we would like to solve

$$\min_{\mathbf{G},\mathbf{\Lambda},\mathbf{H}} E[||v - \mathbf{G\Lambda H}^T u||_2^2, \tag{5.16}$$

where $\mathbf{H}$ is the analysis basis, $\mathbf{G}$ is the synthesis basis, and $\mathbf{\Lambda}$ is a diagonal matrix which encapsulates the scalar predictors applied in transform domain. Without any restrictions on $\mathbf{G}$, $\mathbf{\Lambda}$, $\mathbf{H}$ it is clear that the $\mathbf{G\Lambda H}^T$ product should match the optimal linear predictor [97],

$$\mathbf{G\Lambda H}^T = E[vu^T](E[uu^T])^{-1}, \tag{5.17}$$

which minimizes (5.16), i.e., one does not gain or lose anything by considering the problem in transform domain. For any such $\mathbf{G\Lambda H}^T$ product the optimal predictor becomes

$$\hat{v} = E[vu^T](E[uu^T])^{-1}u + \Delta, \tag{5.18}$$

where Δ is a vector that incorporates the target mean.

Suppose now that we are interested in orthonormal transforms, i.e., we would like solve

$$\min_{\mathbf{H},\mathbf{\Lambda}} E[||v - \mathbf{H\Lambda H}^T u||_2^2, \text{ subject to } \mathbf{H}^T\mathbf{H} = \mathbf{1}, \tag{5.19}$$

where $\mathbf{\Lambda}$ is again diagonal. Let $\mathbf{K} = \mathbf{H\Lambda H}^T$ and note that $\mathbf{K}$ is symmetric. Equation 5.19 is thus equivalent to

$$\min_{\mathbf{K}} E[||v - \mathbf{K}u||_2^2, \text{ subject to } \mathbf{K} = \mathbf{K}^T. \tag{5.20}$$

Noting that $\mathbf{K}(i,j) = \mathbf{K}(j,i)$, and setting the derivatives of (5.20) with respect to $\mathbf{K}(i,j)$ to zero we obtain

$$E[vu^T] + E[uv^T] = E[uu^T]\mathbf{K} + \mathbf{K}E[uu^T]. \tag{5.21}$$

Consider the eigen decomposition $E[uu^T] = \mathbf{F}\mathbf{U}\mathbf{F}^T$ where $\mathbf{F}$ is orthonormal and $\mathbf{U}$ is diagonal. Let $\mathbf{R} = E[vu^T]$ and let $\check{\mathbf{A}} = \mathbf{F}^T\mathbf{A}\mathbf{F}$ denote the similarity transformed version of a matrix $\mathbf{A}$. In the $\mathbf{F}$ coordinate system Equation 5.21 becomes

$$\check{\mathbf{R}} + \check{\mathbf{R}}^T = \mathbf{U}\check{\mathbf{K}} + \check{\mathbf{K}}\mathbf{U}, \tag{5.22}$$

which is straightforward to solve and yields

$$\mathbf{U}(i,i)\check{\mathbf{K}}(i,j) + \mathbf{U}(j,j)\check{\mathbf{K}}(i,j) = \check{\mathbf{R}}(i,j) + \check{\mathbf{R}}(j,i) \tag{5.23}$$

$$\check{\mathbf{K}}(i,j) = (\check{\mathbf{R}}(i,j) + \check{\mathbf{R}}(j,i))/(\mathbf{U}(i,i) + \mathbf{U}(j,j)) \tag{5.24}$$

after which we have $\mathbf{K}^* = \mathbf{F}\check{\mathbf{K}}\mathbf{F}^T$. The optimal predictor becomes

$$\hat{v} = \mathbf{K}^* u + \Delta, \tag{5.25}$$

and the optimal $\mathbf{H}^*,\ \mathbf{\Lambda}^*$ pair can be obtained via an eigen decomposition of $\mathbf{K}^*$. Note that while the orthonormal case is structured and slightly better conditioned compared to the general case, the resulting predictors are constrained.

Now let us assume that u and v are extracted from localized training regions within x and y respectively. Regardless of the structure of the prediction, it is clear that one needs accurate correlation/cross-correlation statistics and the conditioning of $E[uu^T]$ plays an important role in forming the predictors. When one has to learn these statistics over available data, the resulting learning problem is thus conflicting. On the one hand, one would like to have many training vector pairs which require large neighborhoods, on the other hand, one needs to quickly adapt to spatial variations which is only possible through using very small neighborhoods. The computational complexity of forming the predictors also involves conflicting goals. Observe that not only does one have to obtain substantial correlation statistics but one also has to perform inverses, accomplish other complex matrix operations, and matrix-vector multiplications in order to obtain the optimal predictors. Lowering the complexity may be possible by performing these operations infrequently or only once per-signal but this is again completely counter to the required adaptivity.

By making modeling assumptions one can alleviate these operational difficulties and, as long as the model is applicable, obtain accurate predictors. We conclude by noting the two assumptions our base algorithm will make in terms of the notation of this section. We will assume that $\mathbf{F}$ is approximately constant and can be approximated in terms of the DCT basis. We will also assume that the evolution from u to v is such that $\check{\mathbf{R}}$ is approximately diagonal. These assumptions by themselves are of course not particularly noteworthy. What is interesting is the fact that they can be made to work so well even on cases involving complex transitions.

5.9 Appendix II: differential compression with piecewise smooth processes

In this section we consider the one dimensional model of [98] in order to outline the benefits of inter-signal prediction and compression in terms of localized correlations within piecewise stationary processes. This model is stated in continuous time and in the unit interval for analytically tractable bounds on transform coefficient variances. Since it incorporates discontinuities, the $1 - D$ model can be considered appropriate for rows extracted from real-world images. Extensions of the discussion to two dimensions are straightforward and considered at the end. The main point of this section is that differential encoding accomplished by using simple predictors in transform domain obtains the optimal asymptotic performance.

Summary of PSM as defined in [98]: Let y be a zero-mean stochastic function defined on the unit interval. Assume that y is a realization from the piecewise smooth stochastic model (PSM) of [98]. This model realizes a finite set of discontinuity locations using a Poisson process and, conditioned on fixing these locations, obtains realizations so that function values at two points t_1 and t_2 are correlated via the autocorrelation function,

$$R(t_1, t_2) = \begin{cases} \mathcal{R}(|t_1 - t_2|), & \text{if there is no discontinuity point between } t_1 \text{ and } t_2, \\ 0, & \text{otherwise.} \end{cases}$$

In this section we will assume that realizations of the process have a random but bounded number of discontinuity points given by κ. As in [98] assume that $\mathcal{R}$ is of class C^{2r_1}, $r_1 \geq 1$, with r_1 quantifying the smoothness of the correlation function[5]. Let e_1 be the random vector that captures the location of the discontinuities. Given an orthonormal wavelet decomposition with compact basis functions of vanishing moments, let $\psi_{j,k}$ denote the

[5] For details please see the PSM definition in [98], page 1901. For two different ways in which unbounded κ can be accommodated please see [98], pages 1904 and 1908.

wavelet basis function at scale j and shift k. It can be shown that

$$E[| < y, \psi_{j,k} > |^2 | e_1] \leq \begin{cases} C_{s,1} 2^{-j(2r_1+1)}, & s : \text{if } support(\psi_{j,k}) \text{ does not overlap any discontinuities,} \\ C_{e,1} 2^{-j}, & e : \text{otherwise.} \end{cases} \tag{5.26}$$

Observe that wavelet coefficients over the smooth segments decay rapidly compared to those over the discontinuities. One of the main results shown in [98] is that transform coding with a wavelet decomposition obtains the operational distortion-rate function

$$D(R) \leq C_1 R^{-2r_1}, \tag{5.27}$$

for some constant $C_1 > 0$. Equation 5.27 is obtained despite the poor decay of coefficients over the discontinuities. In fact, one would obtain the same decay in the distortion-rate function had there been no discontinuities in the process. The gist of the result is that with a localized wavelet decomposition one can fully exploit the strong local correlations in the process (as determined by the smoothness of $\mathcal{R}$) regardless of the discontinuities. This is possible because it can be shown that if a realization has κ discontinuities, of the 2^j coefficients at scale j, only κ are of type e and have slow decay. An encoder can simply separate coefficients at each scale into two groups, an exponential ($\sim 2^j$) number of coefficients of type s and at most a constant number of coefficients of type e, encode each group optimally, and send the grouping information with comparatively negligible bits. The rapid decay of the coefficients over the smooth segments can be taken advantage of as long as the number of poorly decaying coefficients is small.

Simple Predictors in Transform Domain: With y generated as an instance of the PSM, let ι be generated independently via another instance of the PSM with autocorrelation function smoothness r_2 and discontinuities in e_2. Let $x = y + \iota$. Suppose that both the encoder and the decoder have x and we are interested in communicating y to the decoder. Given (5.27), it is clear that an encoder can communicate either y or ι depending on the associated smoothness to obtain $D(R) \leq C'R^{-2\max(r_1,r_2)}$, where $C' > 0$ is a constant. Let us see that one can obtain similar performance by predicting y using x and compressing the prediction error. We perform scalar prediction in transform domain so that $< y, \psi_{j,k} >$ is linearly predicted using $< x, \psi_{j,k} >=< y, \psi_{j,k} > + < \iota, \psi_{j,k} >$. Conditioned on knowing e_1 and e_2, a linear predictor $\alpha < x, \psi_{j,k} >$ of $< y, \psi_{j,k} >$, obtains the mean squared error (MSE),

$$(1-\alpha)^2 E[| < y, \psi_{j,k} > |^2 | e_1, e_2] + \alpha^2 E[| < \iota, \psi_{j,k} > |^2 | e_1, e_2]. \qquad (5.28)$$

Ideally we would like to use the prediction weight $\alpha^*(j, e_1, e_2)$ that minimizes (5.28), which can be obtained as

$$\alpha^*(j, e_1, e_2) = \frac{E[| < y, \psi_{j,k} > |^2 | e_1]}{E[| < y, \psi_{j,k} > |^2 | e_1] + E[| < \iota, \psi_{j,k} > |^2 | e_2]}. \qquad (5.29)$$

However, since we only have bounds for the expectations in (5.29), let us instead consider the upper-bound predictor, $\alpha^u(j, e_1, e_2)$, which heuristically replaces the variances with their upper-bounds obtained from (5.26) (and its equivalent for ι). $\alpha^u(j, e_1, e_2)$ assumes four different values in each scale depending on the coefficients involved in the prediction via

$$\alpha^u(j, e_1, e_2) = \begin{cases} \frac{C_{s,1}2^{-j(2r_1+1)}}{C_{s,1}2^{-j(2r_1+1)} + C_{s,2}2^{-j(2r_2+1)}}, & (s,s) \\ \frac{C_{s,1}2^{-j(2r_1+1)}}{C_{s,1}2^{-j(2r_1+1)} + C_{e,2}2^{-j}}, & (s,e) \\ \frac{C_{e,1}2^{-j}}{C_{e,1}2^{-j} + C_{s,2}2^{-j(2r_2+1)}}, & (e,s) \\ \frac{C_{e,1}2^{-j}}{C_{e,1}2^{-j} + C_{e,2}2^{-j}}, & (e,e) \end{cases}, \quad mse(\alpha^u(j, e_1, e_2)) \leq \begin{cases} C_{ss}2^{-j(2\max(r_1,r_2)+1)}. \\ C_{se}2^{-j(2r_1+1)}, \\ C_{es}2^{-j(2r_2+1)}. \\ C_{ee}2^{-j}, \end{cases}$$

(5.30)

where we have also substituted the upper-bounds into (5.28) to obtain the MSE for each case. These four cases reflect whether $< y, \psi_{j,k} >$ and $< \iota, \psi_{j,k} >$ overlap discontinuities or not, as determined via e_1 and e_2. Observe that in the worst-case scenario where the inequalities in (5.26) become equalities, $\alpha^u(j, e_1, e_2) = \alpha^*(j, e_1, e_2)$, i.e., the upper-bound predictor is optimal.

Differential Encoding: At first glance it seems as if we obtain the fastest decay $(2\max(r_1, r_2) + 1)$ in the prediction error only in a restricted case. However, even with the addition of ι, we have that of the 2^j coefficients of x at scale j, at most a constant number overlap the discontinuities. Hence, if $\alpha^u(j, e_1, e_2)$ can be realized at both the encoder and decoder, the bounds on the right side of (5.30) can be used to bound the variances of the prediction-error coefficients in order to encode them using the scalar coders discussed in [98]. This will accomplish

$$D(R) \leq C'' R^{-2\max(r_1, r_2)}, \tag{5.31}$$

with an appropriate $C'' > 0$. Realizing $\alpha^u(j, e_1, e_2)$ is also straightforward since one can again group coefficients into two with negligible bits, and further signal (s, e), (e, s), and (e, e) within the discontinuity group. An optimizing encoder will hence have performance equal or better than (5.31). Note also that for a worst-case Gaussian process with independent coefficients, (5.31) is asymptotically optimal.

Hard-thresholding Predictor: From a MSE point of view it is clear that the (s, s) branch in (5.30) is the one that primarily impacts asymptotic performance. For reference signals of the specific form $x = y + \iota$, it thus becomes possible to use even simpler predictors which can be realized using thresholding. Suppose $r_2 > r_1$ and consider the trivial predictor

$$\alpha^t(j, e_1, e_2) = \begin{cases} 1, & (s,s) \\ 1, & (s,e) \\ 1, & (e,s) \\ 1, & (e,e) \end{cases}, \quad mse(\alpha^t(j, e_1, e_2)) \leq \begin{cases} C'_{ss}2^{-j(2r_2+1)}, \\ C'_{se}2^{-j}, \\ C'_{es}2^{-j(2r_2+1)}, \\ C'_{ee}2^{-j}, \end{cases} \tag{5.32}$$

which obtains the desired asymptotic performance by encoding $-\iota$. For the case $r_2 \leq r_1$ one can turn off the prediction (so that y is encoded), and realize the "combined" predictor using hard-thresholding (using, say, an encoder that chooses a rate-distortion optimal threshold). Naturally, if the corruption also causes amplitude modulation of coefficients, if the smoothness of y and ι are varying in each segment, etc., the effectiveness of a thresholding or on/off predictor becomes limited and the desired performance cannot be realized.

Extension to two dimensions and non-smooth corruption: The above discussion can be extended to two-dimensions in a straightforward manner using decompositions that are appropriate for $2 - D$ signals with discontinuities along curves [99, 100, 101]. When x is further corrupted by white noise, inter-signal prediction will give improvements for scales in which the variances of the coefficients of y and ι are larger than the noise variance, i.e., the distortion-rate function will decay at $2\max(r_1, r_2)$ until one reaches a noise-floor where the decay will reduce to $2r_1$.

We conclude this section by noting that inter-signal compression is beneficial when the corruption caused by ι is small for at least some components in y, provided that the indices of these components can be predicted or signaled cheaply. For $1 - D$ piecewise smooth processes considering signals in wavelet domain and using simple predictors result in significant prediction benefits in segments where ι is smoother than y. It is clear that asymptotically optimal encoding becomes possible due to two factors, both of which are enabled by the wavelet transform.

- Efficient prediction that results in rapidly decaying prediction error coefficients. (In (5.30), except for few, if any, (e, e) overlaps, the variances of prediction error coefficients are rapidly decaying).

- Helper-information that realizes the required adaptive predictors using negligible bits.

Over general signals one must hence use a decomposition that highlights the sparsity in the data with the expectation that significant prediction gains are possible when the corruption is small on the information carrying parts of y, i.e., when larger (smaller) coefficients of ι corrupt smaller (larger) coefficients in y.

Chapter 6

Conclusions

In this dissertation, I looked at the spatial and temporal image prediction problems from the new perspective of the magnitude and phase representations of the CWT or the over-complete DCT. Under those representations, the formulation and modeling of complicated spatial and temporal image evolutions are greatly simplified: edges, textures, structured temporal interferences, and linear temporal distortions can be modeled easily within a small spatial and temporal neighborhood.

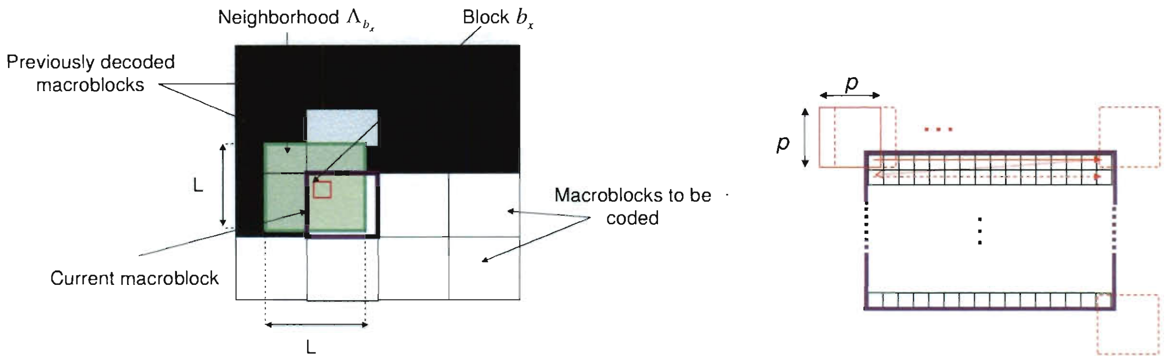
I investigated the theory of image reconstruction from the CWT magnitude and phase. I showed that an image is unique under certain conditions given its analytic magnitude or phase of the CWT, proposed iterative reconstruction algorithms, and presented results about the convergence of the proposed algorithms. The CWT phase is considered being close to the 2D spatial phase that may be used internally by the human visual system for encoding visual information. The investigation in this dissertation verified that the CWT phase is capable of representing images alone. It also sheds some new lights on the importance of the CWT magnitude by showing that the magnitude is also capable of representing images alone and encodes the visual information in a "dual" way to the phase.

Following the investigation about the magnitude and phase representations, I proposed simple geometrical models for interpolating the analytic CWT magnitude and phase, and constructed an iterative image inpainting algorithm to solve the image inpainting problem.

Under the magnitude and phase representation, important image features like edges and patterned textures become very simple to model and interpolate (linear models for edge location and phase). The proposed inpainting algorithm achieved high visual quality prediction results with low computational complexity.

For inter-picture image prediction encountered in video coding and image registration, I proposed a novel temporal prediction algorithm which enables successful prediction under complicated scene transitions. Under the sparse and smooth representation of the over-complete DCT, the rejection of structured temporal interference and the learning and inverting of linear temporal distortion can be achieved by a set of simple predictors. The proposed temporal prediction algorithm has been successfully applied to the standard H.264/AVC video encoder and image registration.

```
Scan all $p \times p$ blocks overlapping the current macroblock.
  * Predict the $p \times p$ block.
  * Update current prediction.
  * Store prediction for final averaging.
  * Expand and update the training region.
Combine stored predictions for the macroblock.
```



(a) (b)

Figure 5.6 : Implementing the prediction algorithm for macroblock based operation. (a) Block b_x and neighborhood Λ_{b_x}. (b) Example prediction scan inside a macroblock in layers of horizontal, one pixel thick strips. Previous predictions are incorporated into the training data of later ones. At certain points in the scan, prediction of compression transform blocks are completed and corresponding prediction error updates are incorporated.

5.5 Simulation Results on Video Coding

The proposed predictor is implemented inside the JM h264/AVC reference software [54]. In generating the simulation results, the parameters in Figure 5.6 are set to $p = 4$ and $L = 12$ (corresponding to $+/-1$ block around each $p \times p$ block). For each macroblock, there is a 1 bit overhead information to determine if the proposed prediction is used or not. This bit was set within the rate-distortion optimization loop of the JM software. Only

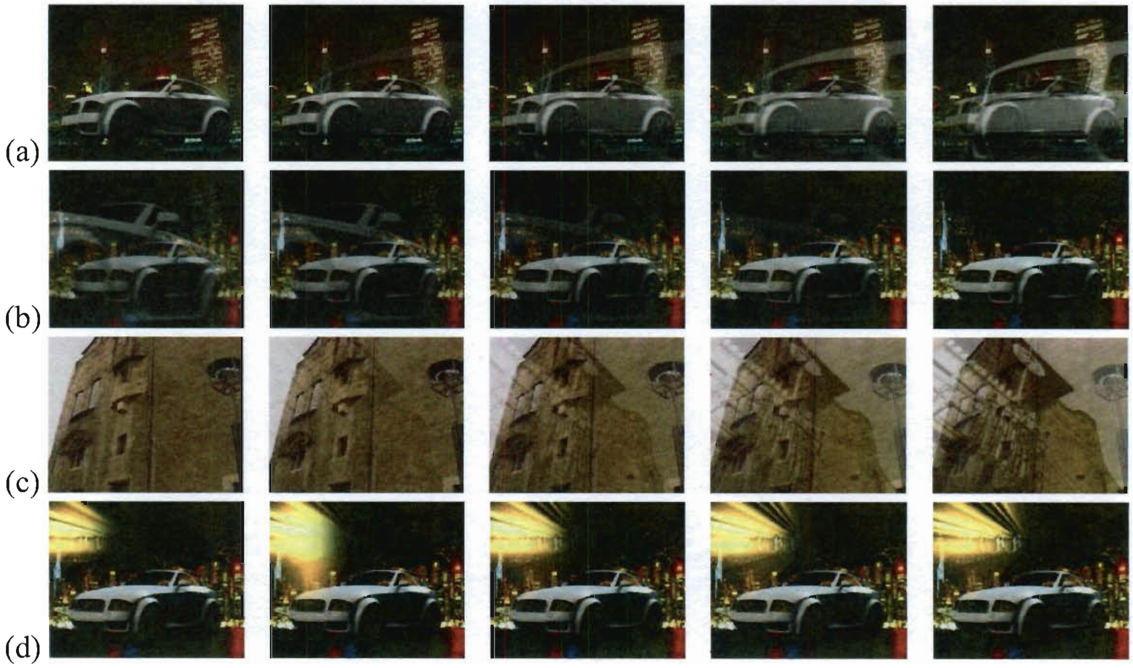


Figure 5.7 : Test set of five-frame transitions.

luminance macroblocks were predicted with the proposed technique. All test sequences are QCIF (144×176) resolution. Beyond the traditional motion search, a motion search on an integer grid is implemented to find the optimal integer motion vectors for the proposed predictor.

Figure 5.7 shows a set of five-frame transitions from the video sequences "car" (Figure 5.7 (a), (b), and (d)) and "glasgow" (Figure 5.7 (c)). In Figure 5.8, the corresponding rate-distortion results is provided using $QP = 22, 26, 30$. Each five-frame sequence was encoded in the $IPPPP$ pattern. The video frame rate is at 30 frames/sec. Both rate and distortion are averages for the four P frames since both h.264/AVC and the proposed predictor utilize the same INTRA frame. Figure 5.7 (a)-(c) correspond to fades, and (d) depicts a special effect with localized blurring, fades, and brightness changes. Figure 5.7

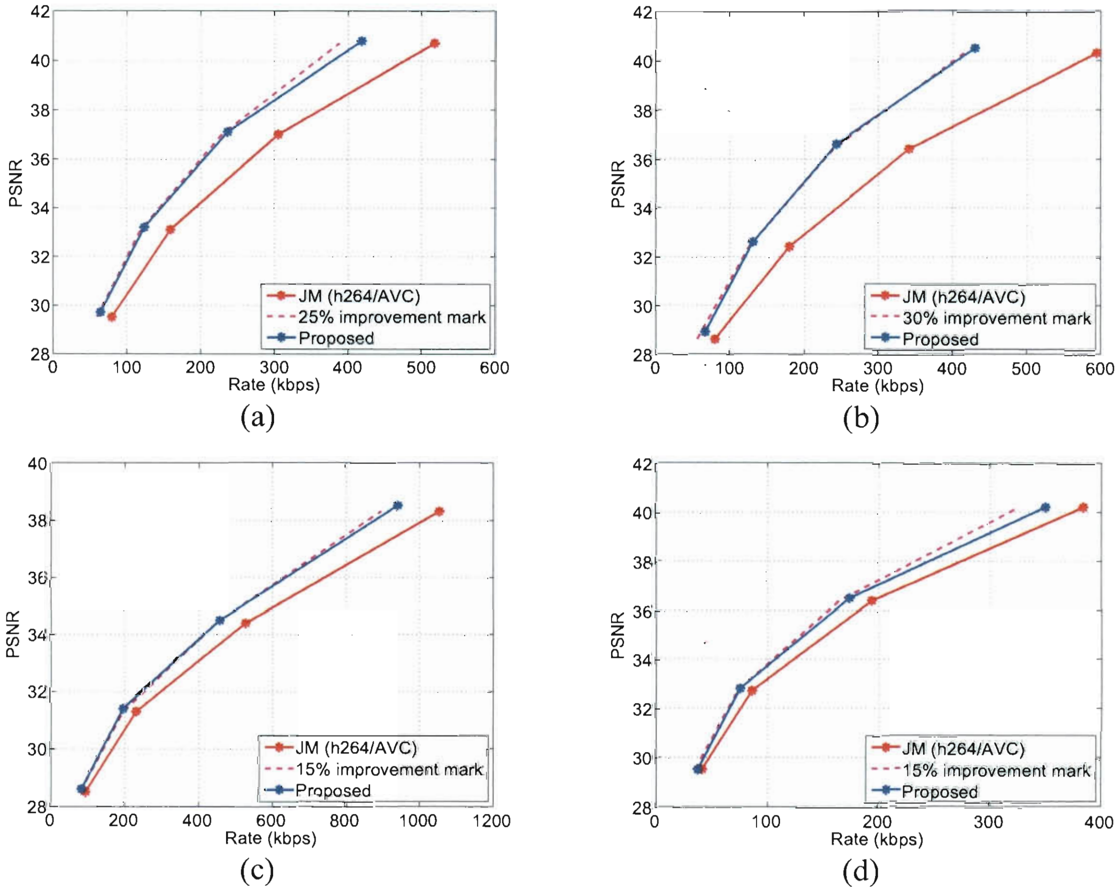
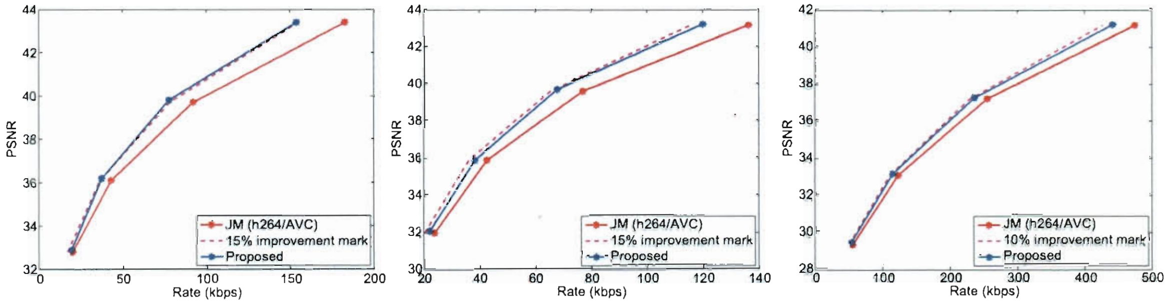


Figure 5.8 : Rate-distortion performance of a h264/AVC-based video coder (JM) without and with the proposed predictor for encoding the short sequences in Figure 5.7. The proposed predictor improves the performance by about 25%, 30%, 15% and 15% respectively.

(a) has mostly rotational motion, (b) and (c) have translations, and (d) is stationary. As seen in Figure 5.8, the proposed predictor obtains improvements for all cases even for scenes rich in spatial frequencies. Allowing the use of fractional motion is expected to improve these results.

Over entire sequences the gains provided by the proposed technique varies depending on the density of sophisticated temporal evolutions in the sequence. On simple sequences (such as "foreman", "car-phone", and "container"), gains are on the order of $2 - 5\%$ improvements in rate at constant distortion, whereas on more complicated sequences (such as movie trailers and commercials, see Figure 5.9), gains again become significant. A more compression-optimized implementation that determines and sends the prediction parameters in a rate-distortion optimal fashion is expected to improve these results.



(a) Movie trailer 1 (b) Movie trailer 2 (b) Commercial

Figure 5.9 : Rate-distortion performance of a h264/AVC-based video coder (JM) without and with the proposed predictor for encoding two movie trailers and a commercial. The sequences (150 frames) are encoded in the IPPPP profile. In each case, the distortion/rate involved in the first (INTRA) frame is not included since the encoding of this frame is the same for both coders

The bulk of the per-frame decoding complexity of the proposed work can be summarized as $\sim 3p^2$ multiplies $+p^2$ divides $+4p^2$ additions per-pixel (in order to solve (5.12) and apply (5.13)), and a translation invariant DCT decomposition. Note however that this complexity can be reduced significantly by reducing the size of the training set. Encoder complexity is more cumbersome due to the motion search. Complexity can be alleviated by restricting the technique to macroblocks where traditional prediction fails or is deemed inefficient in a rate-distortion sense. Encoder complexity may also be reduced by using fast motion search algorithms.

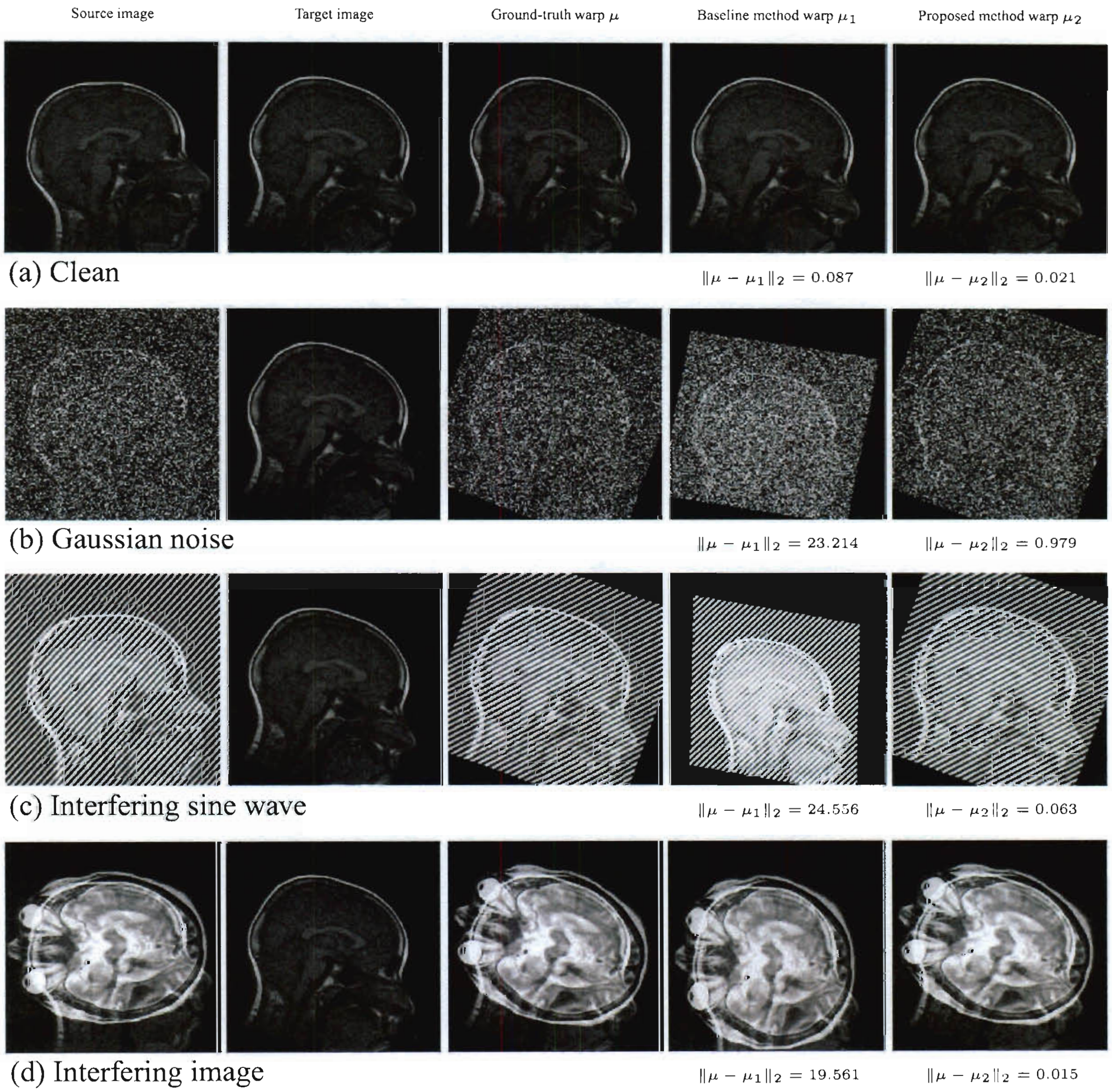

Source image
Target image
Ground-truth warp μ
Baseline method warp μ_1
Proposed method warp μ_2
(a) Clean
$\|\mu - \mu_1\|_2 = 0.087$
$\|\mu - \mu_2\|_2 = 0.021$
(b) Gaussian noise
$\|\mu - \mu_1\|_2 = 23.214$
$\|\mu - \mu_2\|_2 = 0.979$
(c) Interfering sine wave
$\|\mu - \mu_1\|_2 = 24.556$
$\|\mu - \mu_2\|_2 = 0.063$
(d) Interfering image
$\|\mu - \mu_1\|_2 = 19.561$
$\|\mu - \mu_2\|_2 = 0.015$

Figure 5.10 : Image registration examples using the proposed prediction method where the target image is clean and the source image may have interference (a) Clean; (b) Gaussian noise; (c) Interfering sine wave; (d) Interfering image.

5.7 Conclusions

This chapter constructed a temporal prediction algorithm that provides successful estimates over complex inter-picture transitions involving focus changes, cross-fades, intensity variations, noise, clutter, and so on. The proposed algorithm is not narrowly committed to

Bibliography

[1] D.C. Burr, “Sensitivity to spatial phase,” *Vision Res.*, vol. 20, pp. 391–396, 1980.

[2] T. Caelli and P. Bevan, “Visual sensitivity to two-dimensional spatial phase,” *J. Opt. Soc. Amer.*, vol. 72, pp. 1375–1381, 1982.

[3] J. Behar, M. Porat, and Y.Y. Zeevi, “Image reconstruction from localized phase,” *IEEE Trans. Signal Proc.*, vol. 40, no. 4, pp. 736–743, Apr. 1992.

[4] T. F. Chan and J. Shen, “Mathematical models of local non-texture inpaintings,” *SIAM J. Appl. Math.*, vol. 62, no. 3, pp. 1019–1043, 2001.

[5] M. Bertalmío, L. A. Vese, G. Sapiro, and S. Osher, “Simultaneous structure and texture image inpainting,” *IEEE Trans. Image Proc.*, vol. 12, no. 8, pp. 882–889, Aug. 2003.

[6] O. G. Guleryuz, “Weighted overcomplete denoising,” in *Proc. Asilomar Conference on Signals and Systems*, Nov. 2003, vol. 2, pp. 1992 – 1996.

[7] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, “Image inpainting,” in *Proc. SIGGRAPH*, 2000, pp. 417–424.

[8] C. Ballester, M. Bertalmío, V. Caselles, G. Sapiro, and J. Verdera, “Filling-in by joint interpolation of vector fields and gray levels,” *IEEE Trans. Image Proc.*, vol. 10, no. 8, pp. 1200–1211, Aug. 2001.

[9] A. Levin, A. Zomet, and Y. Weiss, “Learning how to inpaint from global image statistics,” in *Proc. Int. Conf. Computer Vision*, 2003, pp. 305–313.

[10] T. F. Chan, J. Shen, and Hao-Min Zhou, “Total variation wavelet inpainting,” *J. Math. Imaging Vis.*, vol. 25, no. 1, pp. 107–125, 2006.

[11] A. Efros and T. Leung, “Texture synthesis by non-parametric sampling,” in *Proc. Int. Conf. Computer Vision*, 1999, pp. 1033–1038.

[12] A. Criminisi, P. Perez, and K. Toyama, “Object removal by exemplar-based inpainting,” in *Proc. Conf. on Comp. Vision*, 2003, vol. 12, pp. 721–728.

[13] J. Wu and Q. Ruan, “Object removal by cross isophotes exemplar-based image inpainting,” in *Proc. Conf. on Comp. Vision*, 2003, vol. 12, pp. 721–728.

[14] O. G. Guleryuz, “Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising: Part i - theory,” *IEEE Trans. Image Proc.*, vol. 15, no. 3, pp. 539–554, March 2006.

[15] O. G. Guleryuz, "Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part ii: adaptive algorithms," *IEEE Trans. Image Proc.*, vol. 15, no. 3, pp. 555–571, March 2006.

[16] M. Elad, J.-L Starck, D. Donoho, and P. Querre, "Simultaneous cartoon and texture image inpainting using morphological component analysis (mca)," *Journal on Applied and Computational Harmonic Analysis ACHA*, vol. 19, pp. 340–358, 2005.

[17] M. J. Fadili, J. L. Starck, and F. Murtagh, "Inpainting and zooming using sparse representations," *The Comput. J.*, vol. 52, no. 1, pp. 64–79, 2009.

[18] A.V. Oppenheim and J.S. Lim, "The importance of phase in signals," *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529–541, May 1981.

[19] M.H. Hayes, J.S. Lim, and A.V. Oppenheim, "Signal reconstruction from phase or magnitude," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 28, no. 6, pp. 672–680, 1980.

[20] M.H. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 30, no. 2, pp. 140–154, 1982.

[21] P.L. Van Hove, M.H. Hayes, J.S. Lim, and A.V. OppenHeim, "Signal reconstruction from signed Fourier transform magnitude," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 31, no. 5, pp. 1286–1293, 1983.

[22] S.H. Nawab and J.S. Lim, "Signal reconstruction from short-time fourier transform magnitude," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 31, no. 4, pp. 986–998, 1983.

[23] D. Griffin and J.S. Lim, "Signal estimation from modified short-time fourier transform," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 32, no. 2, pp. 236–243, 1984.

[24] J. Weng, "Windowed fourier phase: completeness and signal reconstruction," *IEEE Trans. Signal Proc.*, vol. 41, no. 2, pp. 657–666, Feb. 1993.

[25] J.G. Proakis, *Digital Communications*, McGraw-Hill, 2000.

[26] T. Bulow and G. Sommer, "Hypercomplex signals-a novel extension of the analytic signal to the multidimensional case," *IEEE Trans. Signal Processing*, vol. 49, no. 11, pp. 2844 – 2852, Nov. 2001.

[27] H.F. Ates and M.T. Orchard, "A nonlinear image representation in wavelet domain using complex signals with single quadrant spectrum," in *Conference Record of 37th Asilomar Conference on Signals, Systems and Computers*, 2003, vol. 2, pp. 1966–1970.

[28] Gang Hua, "Noncoherent image denoising," M.S. thesis, Rice University, 2005.

[29] G. Hua and M. T. Orchard, “Image reconstruction from the phase or magnitude of its complex wavelet transform,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, April 2008, vol. 1, pp. 3261–3264.

[30] N.G. Kingsbury, “The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters,” in *Proc. European Signal Processing Conf.*, Sept. 1998, pp. 319–322.

[31] N.G. Kingsbury, “Image processing with complex wavelets,” *Phil. Trans. R. Soc. London*, vol. 357, pp. 2543–2560, Sept. 1999.

[32] I. W. Selesnick, R. G. Baraniuk, and N. Kingsbury, “The dual-tree complex wavelet transform - a coherent framework for multiscale signal and image processing,” *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, Nov. 2005.

[33] H. Ates, *Modeling location information for wavelet image coding*, Ph.D. thesis, Princeton University, 2003.

[34] W. Chan, H. Choi, and R. Baraniuk, “Directional hypercomplex wavelets for multidimensional signal analysis and processing,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, May 2004, vol. 3, pp. 393–396.

[35] W. Chan, H. Choi, and R. Baraniuk, “Quaternion wavelets for image analysis and processing,” in *Proc. Int. Conf. on Image Proc. (ICIP)*, 2004, vol. 5, pp. 3057–3060.

[36] G. Hua and M.T. Orchard, “Image inpainting based on geometrical modeling of complex wavelet coefficients,” in *Proc. Int. Conf. on Image Proc. (ICIP)*, Sept. 2007, vol. 1, pp. 553–556.

[37] H. Choi, J. Romberg, and R. Baraniuk, “Hidden markov tree modeling of complex wavelet transforms,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, June 2000, vol. 1, pp. 133–136.

[38] L. Sendur and I. W. Selesnick, “Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency,” *IEEE Trans. Signal Processing*, vol. 55, no. 11, pp. 2744–2756, Nov. 2002.

[39] L. Sendur and I. W. Selesnick, “Bivariate shrinkage with local variance estimation,” *IEEE Signal Proc. Letters*, vol. 9, no. 12, pp. 438–441, Dec. 2002.

[40] W. Chan, H. Choi, and R. Baraniuk, “Coherent image processing using quaternion wavelets,” in *Wavelet Applications Signal Image Processing XI*, 2005, vol. 5914, pp. 59140z.1–59140z.10.

[41] W. Chan, H. Choi, and R. Baraniuk, “Coherent multiscale image processing using dual-tree quaternion wavelets,” *IEEE Trans. Image Proc.*, vol. 17, no. 7, pp. 1069 – 1082, July 2008.

[42] J. Bracamonte, M. Ansorge, F. Pellandini, and P. A. Farine, “Low complexity image matching in the compressed domain by using the dct-phase,” in *Proc. of the 6th COST 276 Workshop on Information and Knowledge Management for Integrated Media Communications*, May 2004, pp. 88–93.

[43] M. Porat and Y.Y. Zeevi, “The generalized gabor scheme of image reconstruction in biological and machine vision,” *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 10, pp. 452–468, 1988.

[44] S. Urieli, M. Porat, and N. Cohen, “Image reconstruction from localized phase,” *IEEE Trans. Image Proc.*, vol. 7, no. 6, pp. 838–853, June 1998.

[45] R. Balan, P. Casazza, and D. Edidin, “On signal reconstruction without phase,” *Appl. Comput. Harmon. Anal.*, vol. 20, pp. 345–356, 2006.

[46] G. Michael and M. Porat, “Image reconstruction from localized fourier magnitude,” in *Proc. Int. Conf. on Image Proc. (ICIP)*, 2001, vol. 1, pp. 213–216.

[47] D. C. Youla and H. Webb, “Image restoration by the method of convex projections: Part 1–theory,” *IEEEE Trans. Med. Imaging*, vol. MI-1, pp. 81–94, Oct.. 1982.

[48] David G. Luenberger, *Linear and Nonlinear Programming*, Springer, 2003.

[49] V.T. Tom, T.F. Quatieri, M.H. Hayes, and J.H. McClellan, “Convergence of iterative nonexpansive signal reconstruction algorithms,” *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 29, no. 5, pp. 1052–1058, 1981.

[50] Ut-Va Koc and K. J. R. Liu, “Dct-based subpixel motion estimation,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, 1996, vol. 4, pp. 1930–1933.

[51] I. Ito and H. Kiya, “Dct sign only correlation with application to image matching and the relationship with phase-only correlation,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, May 2007, vol. 1, pp. 1237–1240.

[52] W. Zeng and B. Liu, “Geometric-structure-based error concealment with novel applications in block-based low bit rate coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 648–665, June 1999.

[53] M. Aharon, M. Elad, and A. Bruckstein, “K-svd: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Trans. on Image Processing*, vol. 54, no. 11, pp. 4311 – 4322, 2006.

[54] Joint Video Team of ITU-T and ISO/IEC JTC 1, “Draft itu t recommendation and final draft international standard of joint video specification (itu-t rec. h.264 — iso/iec 14496-10 avc),” March 2003.

[55] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the h.264/avc video coding standard,” *IEEE Transactions on Circuits and Systems for Video technology*, vol. 13, no. 7, pp. 560–576, July 2003.

[56] J. Biemond, L. Looijenga, D.E. Boekee, and R.H. J.M. Plompen, “A pel-recursive wiener-based motion estimation algorithm,” *Signal Processing*, vol. 13, pp. 399–412, Dec. 1987.

[57] K. Jaemin and J.W. Woods, “Spatio-temporal adaptive 3-d kalman filter for video,” *IEEE Trans. on Image Processing*, vol. 6, no. 3, pp. 414–424, March 1997.

[58] M.K. Ozkan, A.T. Erdem, M.I. Sezan, and A. M. Tekalp, “Efficient multiframe wiener restoration of blurred and noisy image sequences,” *IEEE Trans. on Image Processing*, vol. 1, no. 4, pp. 453 – 476, Oct. 1992.

[59] A. Nosratinia and M. T. Orchard, “New relationships in operator-based backward motion compensation,” in *Proc. Data Compression Conference (DCC)*, March 1995.

[60] S.N. Efstratiadis and A.K. Katsaggelos, “A model-based pel-recursive motion estimation algorithm,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, April 1990, vol. 4, pp. 1973–1976.

[61] R. Rajagopalan, M. T. Orchard, and R. D. Brandt, “Motion field modeling for video sequences,” *IEEE Trans. on Image Processing*, vol. 6, no. 11, pp. 1503–1516, Nov. 1997.

[62] K. Jaemin and J.W. Woods, “3-d kalman filter for image motion estimation,” *IEEE Trans. on Image Processing*, vol. 7, no. 1, pp. 42–52, Jan. 1998.

[63] T. Wedi, “Adaptive interpolation filter for motion and aliasing compensated prediction,” in *Proc. Electronic Imaging 2002: SPIE Visual Communications and Image Processing (VCIP)*, Jan. 2002, vol. 4671, pp. 415–422.

[64] J. Vatis, B. Edler, I. Wassermann, D. T. Nguyen, and J. Ostermann, “Coding of coefficients of two-dimensional non-separable adaptive wiener interpolation filter,” in *Proc. Electronic Imaging 2005: SPIE Visual Communications and Image Processing (VCIP)*, 2005, vol. 5960, pp. 623–631.

[65] K. Kamikura, H. Watanabe, H Jozawa, H. Kotera, and S. Ichinose, “Global brightness-variation compensation for video coding,” *IEEE Transactions on Circuits and Systems for Video technology*, vol. 8, no. 8, pp. 988–1000, Dec. 1998.

[66] B. Girod, “Efficiency analysis of multihypothesis motion-compensated prediction for video coding,” *IEEE Trans. Image Processing*, vol. 9, no. 2, pp. 173–183, Feb. 2000.

[67] A. C. Kokaram, “On missing data treatment for degraded video and film archives: a survey and a new bayesian approach,” *IEEE Trans. on Image Processing*, vol. 13, no. 3, pp. 397–415, March 2004.

[68] K. A. Patwardhan, G. Sapiro, and M. Bertalmio, “Video inpainting under constrained camera motion,” *IEEE Trans. on Image Processing*, vol. 16, no. 2, pp. 545–553, Feb. 2007.

[69] A. C. Kokaram and S. J. Godsill, "Mcmc for joint noise reduction and missing data treatment in degraded video," *IEEE Trans. on Image Processing*, vol. 50, no. 2, pp. 189–205, Feb. 2002.

[70] X. Li and Y. Zeng, "Patch-based video processing: A variational bayesian approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, pp. 27–40, Jan. 2009.

[71] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture image inpainting," *IEEE Trans. on Image Processing*, vol. 12, no. 8, pp. 882 – 889, Aug. 2003.

[72] Y. Junlan, D. Schonfeld, and M. Mohamed, "Robust focused image estimation from multiple images in video sequences," in *Proc. Int. Conf. on Image Proc. (ICIP)*, Sept. 2007.

[73] K. Garg and S. K. Nayar, "Detection and removal of rain from videos," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004.

[74] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, pp. 1413–1457, 2004.

[75] M. Figueiredo and R. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Processing*, vol. 12, no. 8, pp. 906–916, July 2003.

[76] J. A. Guerrero-Colon, L. Mancera, and J. Portilla, "Image restoration using space-variant gaussian scale mixtures in overcomplete pyramids," *IEEE Trans. on Image Processing*, vol. 17, no. 1, pp. 27 – 41, Jan. 2008.

[77] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Processing*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[78] P. Kisilev, M. Zibulevshy, and Y. Y. Zeevi, "Blind separation of mixed images using multiscale transforms," *Proc. Int. Conf. Image Processing (ICIP)*, vol. 1, pp. 309–312, Sept. 2003.

[79] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Lecture Notes in Statistics: Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds. Springer-Verlag, Berlin, Germany, 1995.

[80] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. on Image Processing*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[81] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using gaussian scale mixtures in the wavelet domain," *IEEE Trans. Image Processing*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[82] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[83] O. G. Guleryuz, "Weighted averaging for denoising with overcomplete dictionaries," *IEEE Trans. Image Proc.*, vol. 16, no. 12, pp. 3020–3034, Dec. 2007.

[84] O. G. Guleryuz, "A nonlinear loop filter for quantization noise removal in hybrid video compression," in *Proc. Int. Conf. on Image Proc. (ICIP)*, Sept. 2005, vol. 2, pp. 333–336.

[85] J. L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Trans. on Image Processing*, vol. 14, no. 10, pp. 1570–1582, Oct. 2005.

[86] D. L. Donoho, "De-noising by soft thresholding," *IEEE Trans. Inform. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[87] D. L. Donoho, "Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition," *Appl. Comput. Harmon. Anal.*, vol. 2, pp. 101–126, 1995.

[88] S. Periaswamy, J.B. Weaver, D.M. Healy Jr., D. Rockmore, P.J. Kostelec, and H. Farid, "Differential affine motion estimation for medical image registration," in *Proceedings of the SPIE*, 2000, vol. 4119, pp. 1066 – 1075.

[89] V. Auvray, P. Bouthemy, and J. Lienard, "Multiresolution parametric estimation of transparent motions," in *Proc. Int. Conf. on Image Proc. (ICIP)*, Sept. 2005, pp. 141 – 144.

[90] L. Hongche, H. Tsai-Hong, M. Herman, and R. Chellappa, "Spatio-temporal filters for transparent motion segmentation," in *Proc. Int. Conf. on Image Proc. (ICIP)*, Oct. 1995, pp. 464 – 467.

[91] J. L. Starck, E. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. on Image Processing*, vol. 11, no. 6, pp. 670 – 684, Jun. 2002.

[92] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Trans. on Image Processing*, vol. 14, no. 12, pp. 2091 – 2106, Dec. 2005.

[93] M. Aharon, M. Elad, and A. M. Bruckstein, "K-svd and its non-negative variant for dictionary design," in *Proceedings of the SPIE conference wavelets*, 2005, vol. 5914, pp. 1066 – 1075.

[94] O. G. Sezer, O. Harmanci, and O. G. Guleryuz, "Sparse orthonormal transforms for image compression," in *Proc. Int. Conf. on Image Proc. (ICIP)*, Oct. 2008.

[95] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Trans. on Image Processing*, vol. 16, no. 2, pp. 349 – 366, Feb. 2007.

[96] E. Le Pennec and S. Mallat, "Sparse geometric image representation with bandelets," *IEEE Trans. on Image Processing*, vol. 14, no. 4, pp. 423 – 438, Apr. 2005.

[97] H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, Prentice Hall, 1986.

[98] A. Cohen, I. Daubechies, O. G. Guleryuz, and M. T. Orchard, "On the importance of combining wavelet-based nonlinear approximation with coding strategies," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1985–1921, July 2002.

[99] D. L. Donoho, "Wedgelets: Nearly-minimax estimation of edges," *Annals of Statistics*, vol. 27, pp. 859–897, 1999.

[100] R. Shukla, P. L. Dragotti, M. N. Do, and M. Vetterli, "Rate-distortion optimized tree-structured compression algorithms for piecewise polynomial images," *IEEE Trans. on Image Processing*, vol. 14, no. 3, pp. 343 – 359, 2005.

[101] M. B. Wakin, J. K. Romberg, C. Hyeokho, and R. G. Baraniuk, "Wavelet-domain approximation and compression of piecewise smooth images," *IEEE Trans. on Image Processing*, vol. 15, no. 5, pp. 1071 – 1087, 2006.