1-29-2009

AM-FM methods for image and video processing

Victor Manuel Murray Herrera

Victor Manuel Murray Herrera
Candidate

Electrical and Computer Engineering
Department

This dissertation is approved, and it is acceptable in quality and form for publication on microfilm:

Approved by the Dissertation Committee:

Miguel Matias, Chairperson

Accepted:

Dean, Graduate School

Date

AM-FM Methods for Image and Video Processing

by

Víctor Manuel Murray Herrera

Electrical Engineer, Pontificia Universidad Católica del Perú, 2004
M.S., Electrical Engineering, University of New Mexico, 2005

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Engineering

The University of New Mexico

Albuquerque, New Mexico

December, 2008

Dedication

To Jesus Christ, God, Holy Spirit; Virgin Mary, Saint Champagnat; my Guardian Angel; Saint Joseph, John Paul II, Saint Jude Thaddeus. For listening to me in every prayer, every night and every morning.

To my parents: Aurea del Rosario Herrera de Murray and Victor E. Murray Benavides, for all their love, support, encouragement. Without them I would not have done anything.

To the memory of Ati, more than my nanny my grandmother.

To all my family, specially the closest ones to me: aunt Doris, uncles Ernesto, Daniel, Walter, cousins Ronny, Giuly, Ivan, Tito, Milagros, Violeta, Zaid, nephews Christian and Farid, nieces Tani, Jime, Yamile, Alessia, Vania. For making me feel that my home is in Peru with you, for your support and love.

Thank you all.

Acknowledgments

I would like to thank my advisor, Marios Pattichis, for his support, advices and friendship since the day of the acceptance in the grad program. Without his support as both advisor in the research and boss in the assistantship, this work would not have been possible. He originally proposed this research topic, and he has been advising me to the right way everyday. Thanks for been a great advisor but a better person.

I would like to thank to my dissertation committee for their time and support in the finishing of this work. Thanks to professors Ramiro Jordan, Christos Christodoulou and Tom Hagstrom.

I would like to acknowledge to the Department of Electrical and Computer Engineering (ECE) of UNM for the support and the grad assistantship since the acceptance in the grad program, to FMAC and DARPA for the research assistantship, and to VisionQuest Biomedical for trusting on me for the internship during the Summer 2008.

I want to acknowledge DARPA for suggesting and supporting the research, presented in this dissertation, related to Motion Estimation and Activity Recognition. I would also like to acknowledge Peter Soliz, Bert Davis, Sergio Murillo and Carla Agurto for the research described in the Retinal Image Analysis chapter.

I would also like to acknowledge Ramiro Jordan (and ISTEC) for all the support to bring *latinos* to UNM. For his friendship, help and advice to us all the time.

To Paul Rodriguez, Simon Barriga, and Alonzo Vera, for their friendship and for trusting on me when I applied to UNM.

Thanks to my friends at UNM, specially to those in the ISTEC lab and the Peruvians, for helping me to feel like in South America many times. Special thanks to Ivan Lopez, Sergio Murillo and Alexandre Franco, without their (maybe forced at the beginning) friendship since the first day at UNM this would have been much more difficult. Also, thanks to Orlando Arbildo for his his friendship since the undergrad studies. Thanks to my friends in *Universitarios Futbol Club*, because all the exciting soccer games (and fights, sometimes) that helped me to do not think only in research.

To the Aquinas Newman Center, for helping me spiritually and for the friendship.

During these four years, there have been many friends close to me in the classes, in the exams, during the Qualifying exam, at parties, at buffets, in the indoor soccer tournament, in the 4-shot coffee cups, in my vacations in Peru every year, in my home neighborhood, in my home university, etc. All of you are really important to me, and I will be thankful to you all my life.

AM-FM Methods for Image and Video Processing

by

Víctor Manuel Murray Herrera

ABSTRACT OF DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Engineering

The University of New Mexico

Albuquerque, New Mexico

December, 2008

AM-FM Methods for Image and Video Processing

by

Víctor Manuel Murray Herrera

Electrical Engineer, Pontificia Universidad Católica del Perú, 2004
M.S., Electrical Engineering, University of New Mexico, 2005

Ph.D., Engineering, University of New Mexico, 2008

Abstract

This dissertation is focused on the development of robust and efficient Amplitude-Modulation Frequency-Modulation (AM-FM) demodulation methods for image and video processing[1]. The motivation for this research lies in the wide number of image and video processing applications that can significantly benefit from this research. A number of potential applications are developed in the dissertation.

First, a new, robust and efficient formulation for the instantaneous frequency (IF) estimation: a variable spacing, local quadratic phase method (VS-LQP) is presented. VS-LQP produces much more accurate results than current AM-FM methods. At significant noise levels ($SNR < 30dB$), for single component images, the VS-LQP

[1]There is currently a patent pending that covers the AM-FM methods and applications described in this dissertation.

method produces better IF estimation results than methods using a multi-scale filterbank. At low noise levels ($SNR > 50dB$), VS-LQP performs better when used in combination with a multi-scale filterbank. In all cases, VS-LQP outperforms the Quasi-Eigen Approximation algorithm by significant amounts (up to $20dB$).

New least squares reconstructions using AM-FM components from the input signal (image or video) are also presented. Three different reconstruction approaches are developed: (i) using AM-FM harmonics, (ii) using AM-FM components extracted from different scales and (iii) using AM-FM harmonics with the output of a low-pass filter. The image reconstruction methods provide perceptually lossless results with image quality index values bigger than 0.7 on average. The video reconstructions produced image quality index values, frame by frame, up to more than 0.7 using AM-FM components extracted from different scales.

An application of the AM-FM method to retinal image analysis is also shown. This approach uses the instantaneous frequency magnitude and the instantaneous amplitude (IA) information to provide image features. The new AM-FM approach produced ROC area of 0.984 in classifying Risk 0 versus Risk 1, 0.95 in classifying Risk 0 versus Risk 2, 0.973 in classifying Risk 0 versus Risk 3 and 0.95 in classifying Risk 0 versus all images with any sign of Diabetic Retinopathy.

An extension of the 2D AM-FM demodulation methods to three dimensions is also presented. New AM-FM methods for motion estimation are developed. The new motion estimation method provides three motion estimation equations per channel filter (AM, IF motion equations and a continuity equation). Applications of the method in motion tracking, trajectory estimation and for continuous-scale video searching are demonstrated. For each application, we discuss the advantages of the AM-FM methods over current approaches.

Contents

List of Figures **xv**

List of Tables **xxiii**

I Introduction **1**

1 Introduction **2**

1.1 AM-FM demodulation theory 2

1.2 An overview of AM-FM theory and Applications 4

1.3 Motivation of current work 7

1.4 Thesis statement 10

1.5 Contributions 10

1.6 Dissertation Overview 11

2 Background on AM-FM Demodulation Methods **13**

2.1 AM-FM demodulation methods 13

2.1.1 Continuous Space AM-FM Demodulation based on extensions of the Analytic Signal 13

2.1.2 Uniqueness of the Hilbert-based approaches 15

2.1.3 Quasi-Eigen Approximation (QEA) algorithm 16

2.1.4 Quasi-Local Method (QLM) 17

II AM-FM methods for image processing 19

3 Robust Discrete Image AM-FM Demodulation Methods 20

3.1 2-D Multi-scale filterbank design for applications in discrete images . 21

3.2 Robust multi-scale implementation of the QEA algorithms 23

3.3 Robust multi-scale implementation of the Quasi Local Method (QLM) 24

3.4 A robust Instantaneous Frequency estimation based on a Local Quadratic Phase (LQP) model 24

3.4.1 Pre-filtering and post-filtering 28

3.4.2 Modulation for High Frequencies 29

3.5 Results for Robust AM-FM estimation 29

3.5.1 Gaussian amplitude modulated (Gaussian AM) 30

3.5.2 Gaussian amplitude-modulated Quadratic frequency-modulated (Gaussian AM Quadratic FM) 30

3.5.3 Error in IF estimation 31

3.5.4 Results for synthetic images 31

3.6 Discussion for Robust AM-FM estimation 33

4 Image Reconstructions using AM-FM **40**

4.1 Least-Squares Reconstructions using AM-FM harmonics (LESHA) . . 41

4.2 Least-Squares Reconstructions using AM-FM harmonics and the LPF (LESCA) 42

4.3 Multi-scale least-squares reconstructions (MULTILES) 43

4.4 Results for Image Reconstructions 44

4.5 Discussion on Image Reconstructions 45

5 Retinal Image Analysis **56**

5.1 Introduction 56

5.2 Background on Related Work 59

5.3 Methods 60

5.3.1 Filterbank design 60

5.3.2 Multi-scale AM-FM analysis 62

5.3.3 Prediction of DR state based on the AM-FM histogram estimates 64

5.3.4 Validation using a Leave-one-out Method 67

5.4 Results 68

5.4.1 Analysis of Detection 68

5.4.2 Results for Diabetic Retinopathy 70

5.5 Discussion for Diabetic Retinopathy 71

III AM-FM methods for video processing 74

6 Robust Discrete Video AM-FM Demodulation Methods 75

6.1 3D Multi-scale filterbank design for applications in discrete videos . . 75

6.2 3D VS-LQP method 77

6.3 Results on 3D AM-FM estimation 80

6.3.1 3D Chirp 80

6.3.2 3D Gaussian amplitude-modulated frequency-modulated . . . 81

6.3.3 Results for synthetic 3D signals 81

6.4 Discussion for 3D AM-FM estimation 82

7 Video Reconstructions using AM-FM Decompositions 83

7.1 3-D Multi-scale reconstructions 83

7.2 Results for Video Reconstructions 86

7.3 Discussion for Video Reconstructions 87

8 Motion Estimation (ME) methods based on AM-FM 89

8.1 Introduction 89

8.2 Background on Motion Estimation 92

8.2.1 Optical Flow 92

8.2.2 Continuous Formulation 92

8.2.3 Phase-Based Methods for Motion Estimation 93

8.3 AM-FM based Motion Estimation Method 94

8.3.1 An AM-FM model for motion estimation 94

8.3.2 Discrete Optimization 95

8.4 Results for Motion Estimation 99

8.4.1 Results for a synthetic example 100

8.4.2 Results for atherosclerotic plaque videos 102

8.4.3 Results for tracking of Taxi Video with AM-FM based Motion Estimation 106

8.5 Discussion 108

8.6 Continuous-scale video search 110

8.6.1 Continuous-scale video search Dataset 110

8.6.2 Methods 112

8.6.3 Results for Activity Recognition 113

8.6.4 Discussion for Activity Recognition 114

IV Conclusions and Future Work 116

9 Concluding Remarks and Future Work 117

9.1 Summary of AM-FM methods in Image processing 117

9.2 Summary of AM-FM methods in Video processing 119

9.3 Future Work 120

References **121**

List of Figures

1.1 2D Multi-scale AM-FM demodulation. Dominant AM-FM components are selected over different image scales. The bandpass filter selector (upper left) is used to define the bandpass filters that correspond to each scale. The dominant AM-FM component is selected using the maximum amplitude at every pixel (see lower right). . . . 5

2.1 Block diagram of extended QLM method. 18

3.1 Multi-scale filterbanks designed using optimal separable filter design versus traditional filterbanks with support in four quadrants. For AM-FM analysis we use: (a) Single-scale, (b) two-scale and (c) three-scale filterbanks. For traditional Wavelet analysis we show: (d) Single-scale, (e) two-scale and (f) three-scale filterbanks. 21

3.2 Min-max specifications for 1D filter design. Multidimensional filters can be efficiently implemented using 1D filters. (a) Lowpass filter. (b) Bandpass filter. (c) Highpass filter. 22

3.3 Block diagrams of the basic QEA methods. (a) Based on FFT. The Hilbert filter is applied along the column direction. (b) Using direct, discrete space convolutions. Here, h_m denotes the 2D impulse response for filter m. Note that for the FFT approach in (a), we will first need to zero-pad. Also, since our filters are separable, note that we can compute the convolution result using separable, 1D convolutions along the rows and columns. Similarly, we can use 1D FFTs for more efficient implementations. 23

3.4 Absolute value of the condition number for the arccos (solid line) and arcsin (dash line) is shown. From this graph, it is clear that we have the most stable evaluation for the arccos function $|x| \approx 0$. 26

3.6 Pre-filtering and post-filtering using a median filter to improve accuracy in the IF estimation. The median filter is applied to the input signal, to reject impulsive noise without sacrificing bandwidth, as well as to the IF estimates, to improve continuity. 29

3.5 VS-LQP block diagram using a general bandpass filter in (a). Since separable 1D filters are used, the spectrum of the bandpass filter is showed in 1D and it is generic for x and y. IFx_block is showed in (b) (IFy_block is similar). 35

3.7 PSNR for IA estimation for all the methods used. Improvements in the estimation due to the filterbanks used. (a), (b), (c) Gaussian amplitude modulated signals. (d), (e), (f) Gaussian amplitude-modulated frequency-modulated. (a), (d) QEA when no filterbank is used (dotted line) versus using a two-scale filterbank (solid line). (b), (e) QEA when no filterbank is used (dotted line) versus using a three-scale filterbank (solid line). (c), (f) QLM method using a two-scale filterbank (solid line) versus using a three-scale filterbank (dotted line). 36

3.8 PSNR for IF estimation. Results computed using: (i) no filterbank (first column), (ii) the two-scale filterbank (second column) and (iii) the three-scale filterbank (last column). We only show the results in the x direction since results in the y direction were very similar. (a)-(e) Gaussian amplitude modulated signals. (f)-(j) Gaussian amplitude-modulated frequency-modulated. (a)-(c), (f)-(h) VS-LQP versus QEA. (d), (e), (i), (j) VS-LQP versus QLM. 37

3.9 PSNR for IF estimation: worst case. Results computed using: (i) no filterbank (first column), (ii) the two-scale filterbank (second column) and (iii) the three-scale filterbank (last column). We only show the results in the x direction. Gaussian amplitude modulated signals in the first row and Gaussian amplitude-modulated frequency-modulated in the last one. 38

3.10 PSNR for IF estimation. VS-LQP versus QEA using pre- and post-filtering using median filters without using filterbanks in (a) and (b). Medium-low frequency signals used. (a) Gaussian amplitude modulated results. (b) Gaussian amplitude-modulated frequency-modulated results. Modulation technique applied to high frequency ($\varphi_x = \varphi_y = \{5\pi/6, 7\pi/10, 4\pi/6, 5\pi/8\}$) input signals in (c) and (d) for Gaussian amplitude modulated signals. (c) VS-LQP versus QEA using a three-scale filterbank. (d) VS-LQP versus QLM using a three-scale filterbank. 39

4.1 Modified Gram-Schmidt (MGS) Algorithm for computing an orthonormal basis over the space of the AM-FM harmonics. 42

4.2 Multi-scale AM-FM reconstructions for the *Lena* image. (a) Original image 512x512 pixels. Single-scale filterbank results for: (b) LESHA, (c) MULTILES and (d) LESCA. Two-scale filterbank results for: (e) LESHA, (f) MULTILES and (g) LESCA. Three-scale filterbank results for: (h) LESHA, (i) MULTILES and (j) LESCA. 49

4.3 Multi-scale AM-FM reconstructions for the *Mandrill* image. (a) Original image 512x512 pixels. Single-scale filterbank results for: (b) LESHA, (c) MULTILES and (d) LESCA. Two-scale filterbank results for: (e) LESHA, (f) MULTILES and (g) LESCA. Three-scale filterbank results for: (h) LESHA, (i) MULTILES and (j) LESCA. . 50

4.4 Best and worst *Lena* and *Mandrill* reconstructions between a two-scale filterbank and a three-scale filterbank based on the image quality index from Table 4.8. (a) Original *Lena* image 512x512 pixels. (b) MULTILES two-scale filterbank: $Q = 0.8142$ ($MSE = 52.7913$). (c) LESHA three-scale filterbank: $Q = 0.5299$ ($MSE = 121.4774$). (d) Original *Mandrill* image 512x512 pixels. (e) MULTILES three-scale filterbank: $Q = 0.8104$ ($MSE = 198.5808$). (f) LESHA three-scale filterbank: $Q = 0.3224$ ($MSE = 563.9774$). 51

4.5 High amplitude FM Analysis using QEA method for *Lena*. For (b), (c), (g) and (h) we only show the FM component pixels for which the instantaneous amplitude is above the IA mean. (a) Original image. FM analysis in (b) using a two-scale filterbank and in (c) using a three-scale filterbank. IF using a two-scale filterbank in (d) and using a three-scale filterbank in (e). Zoom on *Lena*'s hair in (f). FM analysis in (g) using a two-scale filterbank and in (h) using a three-scale filterbank. IF using a two-scale filterbank in (i) and using a three-scale filterbank in (j). 52

4.6 High amplitude FM Analysis using QEA method for a fingerprint image. Here, we show results for which the instantaneous amplitude is greater than the IA statistical mode. (a) Original image. (b) FM using a two-scale filterbank. (c) FM using a three-scale filterbank. (d) IF using a two-scale filterbank. (e) IF using a three-scale filterbank. 53

4.7 Method to produce an image from a fixed column in a real video. . . 54

4.8 Frames of videos with a column clearly highlighted. (a) Frame of *Akiyo* video with column 79 clearly highlighted. (b) Frame of *Foreman* video with column 85 clearly highlighted. 54

4.9 FM Analysis for *Akiyo* and *Foreman*. Single-scale filterbank, two-scale filterbank and three-scale filterbank in rows two, three and four, respectively. Original Images: (a) *Akiyo* and (e) *Foreman*. *Akiyo*'s results in column one. *Foreman*'s results in column two. 55

5.1 Decision Matrix. *Figure by Bert Davis from VisionQuest Biomedical.* 58

5.2 Four-scale filterbank used for retinal image analysis. (a) Complete frequency spectrum of the filterbank. (b) Zoom on the low frequency bandpass filters. 61

5.3 Example of the separable design. An 1D filterbank response is shown. The filter corresponds to a bandpass filter with frequency support in $[\pi/8, \pi/4]$. 61

5.4 Four different scales (see Table 5.6) for analyzing a retinal image that was graded as Risk 0. (a) Original image. (b) Region of Interest (ROI) for green channel only. (c) $a_6 \cos \varphi_6$. (d) $a_5 \cos \varphi_5$. (e) $a_4 \cos \varphi_4$. (f) $a_1 \cos \varphi_1$. 63

5.5 Four different cases of scales (see Table 5.6) for analyzing a retinal image that was graded as Risk 1. (a) Original image. (b) Region of Interest (ROI) for green channel only. (c) $a_6 \cos \varphi_6$. (d) $a_5 \cos \varphi_5$. (e) $a_4 \cos \varphi_4$. (f) $a_1 \cos \varphi_1$. 64

5.6 Jacknife algorithm. 67

5.7 ROC for detection of Risk 0 (normal) in a set of normals versus different forms of DR: Risk {1, 2, 3}. (a) Risk 0 versus Risk 1. (b) Risk 0 versus Risk 2. (c) Risk 0 versus Risk 3. (d) Risk 0 versus all patients with any form of DR. 73

6.1 Numbering of the bandpass filters for the time variable for, from left to right, 3D single-, two- and three-scale filterbanks, respectively. . . 76

6.2 Frequency spectrum decomposition for 3D two-scale filterbank. (a) 2D two-scale filterbank for a single w_x slice. (b) Complete 3D frequency spectrum decomposition for the 3D two-scale filterbank. . . . 77

7.1 3D-MULTILES method for video reconstruction using a 3D two-scale filterbank. 85

8.1 Atherosclerotic plaques. (a) Internal carotid artery. (b) Surgery in the carotid artery. Images from Adam ©: *http://www.adam.com* . . 91

8.2 Chirp used for our first test in motion estimation. For the simulation, from (8.19), we use square images of size $N = 128$, $\alpha_1 = \alpha_2 = 9/40(N-1)$, and $\beta_1 = \beta_2 = 18/40(N-1)$. This gives an instantaneous frequency magnitude range interval of $[-9\pi/20, 9\pi/20]$ for both directions. For the motion, we set $A_h = A_v = 2$ and $f_h = f_v = 1.101$

8.3 First frame of atherosclerotic plaque videos that were used. We display an 'x' over pixels where the Kalman filter was able to track throughout the video. The individual video characteristics are as follows: (a) video of size 125 rows $\times$ 250 columns $\times$ 67 frames, (b) video of size 125 rows $\times$ 150 columns $\times$ 111 frames, (c) video of size 100 rows $\times$ 240 columns $\times$ 87 frames, and (d) video of size 125 rows $\times$ 250 columns $\times$ 67 frames. We would like to thank Maura Griffin (PhD) and Ms. Niki Georgiou (S.R.N.) for selecting the case studies and carefully collecting the ultrasound videos that have collected the atherosclerotic videos for this study. 103

8.4 Pixel tracking results from frames 1, 21, 41 and 61 for each video from Figs. 8.3 (a)-(b). For the video shown in Fig. 8.3(a), we have extracted a Region of Interest (ROI) of $30-140\times30-210$. We then show frame 1 in (a), frame 21 in (b), frame 41 in (c), frame 61 in (d). Similarly, we show the same frames for the video shown in Fig. 8.3 (b) for ROI: $90-160\times120-270$ in (e)-(h). 104

8.5 Pixel tracking results from frames 1, 21, 41 and 61 for each video from Figs. 8.3 (c)-(d). For the video showed in Fig. 8.3(c), we have extracted a Region of Interest (ROI) of $100-260\times80-315$. We then show frame 1 in (a), frame 21 in (b), frame 41 in (c), frame 61 in (d). Similarly, we show the same frames for the video of Fig. 8.3 (d) for ROI: $40-160\times100-300$ in (e)-(h). 105

8.6 Video reconstructions for the *taxi* video (we only show frame 1) using dominant components from each scale. (a) 3D Single-scale filterbank. (b) 3D Two-scale filterbank. (c) 3D Three-scale filterbank. 106

8.7 Frame 1 of the *Hamburg taxi video* for tracking with the motion estimated during the video. (a) Zoom on the person. (b) Zoom on the taxi. 107

8.8 Examples of the three different versions for a *fight* video: (a) original video, (b) rotated video and (c) rotated and downsampled video. Frame 300 is shown. 111

8.9 Content-based Video Retrieval System. 112

8.10 Steps for continuous scale video search using AM-FM features. . . . 113

List of Tables

3.1 Modulation frequency factors for high frequency signals. 30

4.1 Bandpass filters corresponding to different image scales. 44

4.2 MSE using LESHA (section 4.1). MSE values less than 150 are shown in **bold typeface**. 45

4.3 MSE using MULTILES (section 4.3, see Table 4.1 also). LPF denotes the low-pass filter reconstruction. L denotes the low-frequency scale AM-FM component. Similarly, M and H denote the medium- and high-frequency scale AM-FM component. MSE values less than 150 are shown in **bold typeface**. 45

4.4 MSE using LESCA (section 4.2). MSE values less than 150 are shown in **bold typeface**. 46

4.5 Image Quality Index using LESHA (section 4.1). Q-values above 0.75 are shown in **bold typeface**. 46

4.6 Image Quality Index using MULTILES (section 4.3), see Table 4.1 also). LPF denotes the low-pass filter reconstruction. L denotes the low-frequency scale AM-FM component. Similarly, M and H denote the medium- and high-frequency scale AM-FM component. Q-values above 0.75 are shown in **bold typeface**. 46

4.7 Image Quality Index using LESCA (section 4.2). Q-values above 0.75 are shown in **bold typeface**. 47

4.8 MSE and Image Quality Index (Q) for the reconstructions shown in Figs. 4.2 and 4.3. 47

5.1 Retinopathy Grade Distribution. 58

5.2 Image Quality Distribution. Highest grade is 5. Lowest grade is 1. . 59

5.3 Risk of Macular Edema (ME) Distribution. 59

5.4 Results from other research studies. Sensitivity and Specificity are given in percentage (%). 60

5.5 Bandpass filters used for estimating AM-FM in a four-scale filterbank. 63

5.6 Scales used for the nine cases in retinal image analysis. 63

5.7 Number of Factors in Normal versus Risk {1, 2, 3} Retinopathy. . . 70

5.8 Distribution of image quality by Risk level. Higher values indicate better image quality. 70

6.1 MSE and PSNR for IA estimation using 3D synthetic signals. 81

6.2 PSNR (dB) for ||IF|| estimation using 3D synthetic signals. 82

6.3 MSE for ||IF|| estimation using 3D synthetic signals. 82

7.1 MSE and Average Image Quality Index in the reconstructions of the videos using 3D-MULTILES method (multi-scale). LPF denotes low-pass filter reconstruction, M denotes the medium frequencies and H denotes high frequencies. 86

7.2 MSE and Average Image Quality Index in the reconstructions of the videos using 3D-LESHA method. 87

7.3 MSE and Average Image Quality Index in the reconstructions of the videos using 3D-LESCA method (same as 3D-LESHA + LPF). . . . 87

8.1 MSE and density in the velocity estimation of the synthetic chirp signal. 102

8.2 MSE for the taxi video using 3D-MULTILES method. 107

8.3 MSE for the taxi video using 3D-LESHA method. 107

8.4 Coefficients used for the taxi video reconstructions using 3D-MULTILES method. 108

8.5 An example demonstrating the computation of the table of distances. 112

8.6 Numeration for the videos used in Table 8.7. 114

8.7 CBVR table for the first approach using 15 videos. Correct matches are shown underlined. The average precision is 82%. 115

Part I

Introduction

Chapter 1

Introduction

In this first chapter we present an introduction to this dissertation work. We start with an introduction to the basis of AM-FM demodulation theory. Then, we present an overview of the AM-FM demodulation theory together with AM-FM applications already developed.

Next, we present the motivation for this dissertation followed by the thesis statement. We finish this chapter stating the contributions of this dissertation and presenting an overview of this document.

1.1 AM-FM demodulation theory

We consider multi-scale AM-FM representations, under least-square approximations, for images given by

$$I(x,y) \simeq \sum_{n=1}^{M} a_n(x,y)\cos\varphi_n(x,y), \tag{1.1}$$

and for videos given by

$$I(x,y,t) \simeq \sum_{n=1}^{M} a_n(x,y,t)\cos\varphi_n(x,y,t), \tag{1.2}$$

where $n = 1, 2, \ldots, M$ denote different scales. In (1.1), a continuous image $I(x,y)$ is a function of a vector of spatial coordinates (x,y). Similarly, in (1.2), a continuous video $I(x,y,t)$ is a function of a vector of spatiotemporal coordinates (x,y,t). A collection of M different scales are used to model essential signal modulation structure. The amplitude functions $a_n(\cdot)$ are always assumed to be positive.

Our interest in AM-FM model is for modeling non-stationary image and video content in terms of amplitude and phase functions (see for example [1]). The basic idea is to let the frequency-modulated (FM) components $\cos\varphi_n(\cdot)$ capture fast-changing spatial variability in the image intensity. For each phase function $\varphi_n(\cdot)$ we define the instantaneous frequency (IF), $\nabla\varphi_n(\cdot)$, in terms of the gradient. Thus, for images

$$\nabla\varphi_n(x,y) = \left(\frac{\partial\varphi_n}{\partial x}(x,y),\, \frac{\partial\varphi_n}{\partial y}(x,y)\right) \tag{1.3}$$

and for videos

$$\nabla\varphi_n(x,y,t) = \left(\frac{\partial\varphi_n}{\partial x}(x,y,t),\, \frac{\partial\varphi_n}{\partial y}(x,y,t),\, \frac{\partial\varphi_n}{\partial t}(x,y,t)\right). \tag{1.4}$$

The instantaneous frequency vector $\nabla\varphi_n(\cdot)$ can vary continuously over the spatial domain of the input signal.

Fig. 1.1 depicts the basic AM-FM demodulation method approach. Given the input image $I(k_1,k_2)$, we first apply the Hilbert transform to form a *2D extension* of the 1D analytic signal: $I_{AS}(k_1,k_2)$. $I_{AS}(k_1,k_2)$ is processed through a collection of bandpass filters (to be described in the section 3.1) with the desired scale. Each processing block will produce the instantaneous amplitude, the instantaneous phase, and the instantaneous frequencies in both x and y directions by means of either the

QEA method (sub-section 2.1.3) or the QLM method (sub-section 2.1.4). The processing for a discrete video $I(k_1, k_2, k_3)$ is similar, applying first a Hilbert transform to form a *3D extension* of the 1D analytic signal: $I_{AS}(k_1, k_2, k_3)$. The instantaneous frequencies are estimated in all x, y and t directions. For each pixel, we select the AM-FM demodulation estimates from the processing block that gives the largest instantaneous amplitude estimate. Hence, the algorithm adaptively selects the estimates from the bandpass filter with the maximum response. This approach does not assume spatial continuity, and allows the model to quickly adapt to singularities in the input signal.

For the reconstruction methods presented here, the AM-FM estimations are computed scale by scale and then added as in (1.1) and (1.2), for images and videos, respectively (chapters 4 and 7, respectively). Similar scale by scale estimation is used for the application in retinal images (chapter 5). For AM-FM based motion estimation methods, the velocity vectors are computed based on (1.2) in chapter 8. Also, an extension of this approach is used for activity recognition in videos.

1.2 An overview of AM-FM theory and Applications

Multidimensional Amplitude-Modulation Frequency-Modulation (AM-FM) models provide methods that allow for continuous-scale analysis. There is strong interest in the further development of AM-FM models, with potential applications in various areas in signal, image and video processing. Also, in [2], the authors show a class of multidimensional orthogonal FM transforms, related with AM-FM. Multidimensional AM-FM demodulation methods have been studied in [3, 4, 5, 6, 7, 8, 9, 10]. Also, in [11, 12], the author presents an axiomatic approach that suggests that the

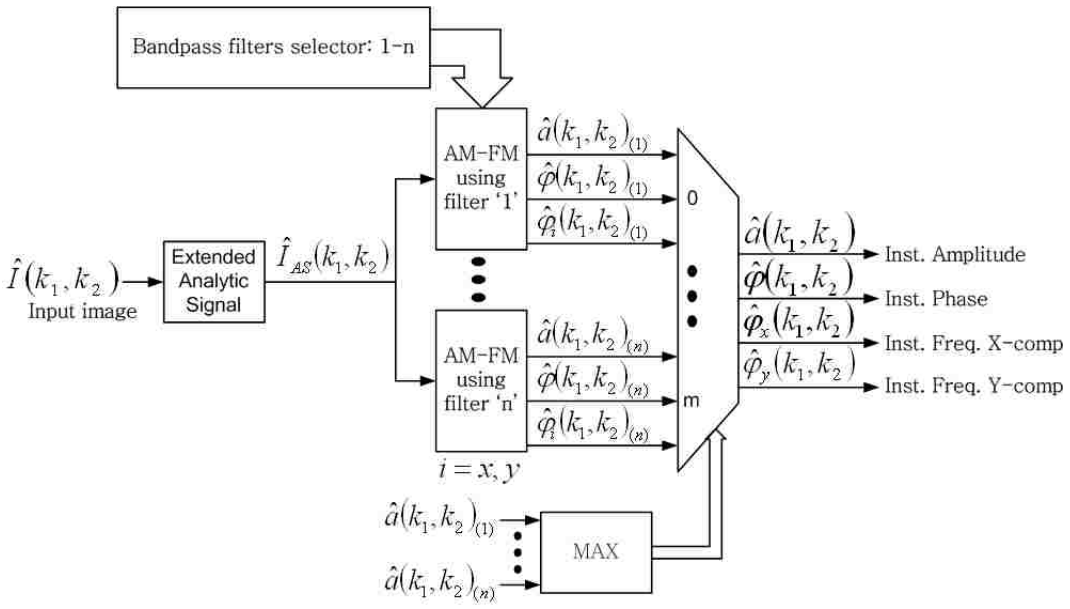


Figure 1.1: 2D Multi-scale AM-FM demodulation. Dominant AM-FM components are selected over different image scales. The bandpass filter selector (upper left) is used to define the bandpass filters that correspond to each scale. The dominant AM-FM component is selected using the maximum amplitude at every pixel (see lower right).

analytic-signal demodulation method is the only algorithm that satisfies certain intuitive conditions.

Applications in one-dimensional (1D) signals have been done in [13, 14, 15, 16, 17, 18], and for speech signal analysis in [19, 20, 21, 22, 23, 24]. Recent interest in one-dimensional AM-FM demodulation work is usually attributed to research associated with the Teager-Kaiser energy operator and its applications in speech signal analysis [19, 20, 21, 22]. Multicomponent AM-FM analysis for 1D signals and speech processing have been done in [3, 14, 25, 26], and recently in [27]. In [28], the authors provided an AM-FM model for images.

Two-dimensional (2D) applications include image pattern analysis including shape

from shading in [29], image interpolation in [30], fingerprint classification in [31], the treatment of analytic signals in [32], image retrieval in digital libraries in [33], applications in image segmentation in [34], [35], [36], for repairing of damaged image textures in [37], component tracking in [38] and in reconstruction in [39, 40, 41, 42, 43, 44]. In [1], the authors use it for analyzing the structure of textured images. In [38, 28, 45, 42] multi-component AM-FM image representations are studied.

For video image analysis, [46] describes an application in cardiac image segmentation. Earlier, Fleet and Jepson developed phase-based methods for motion estimation in [47]. Recently, AM-FM based methods for reconstruction of videos have been given in [48] and for motion estimation in medical videos in [49] and general videos in [48].

The multidimensional extension of the one-dimensional analytic-signal demodulation algorithm was first proposed by Havlicek in his dissertation [50]. A nice tutorial introduction to multidimensional AM-FM demodulation techniques can be found in [51]. More recently, an overview of advances made in the area of multidimensional AM-FM models and methods can be found in [52]. Similarly, multidimensional AM-FM demodulation methods can be found in [53]. A two-dimensional extension to the Teager-Kaiser energy operator can be found in [54]. Early applications of the multidimensional Teager-Kaiser operator and a sign-correction to the instantaneous frequency vector estimation is given in [55, 56].

Analytic image methods for AM-FM demodulation are based on extending the notion of the 1D analytic signal to 2D or simply to provide a Hilbert-based extension of the 1D Hilbert-based demodulation approach. Early work can be found in [50]. In [57] and [58], the authors introduce the phase quadrature transform for extending the Hilbert transform into two dimensions. In [9], Felsberg and Sommer introduced the monogenic signal, a nice extension of the analytic signal to images. The monogenic signal is a three-dimensional representation that combines an image with its Riesz

transform to yield a sophisticated 2D analytic signal. Some related work on image demodulation based on the analytic image includes [4, 5, 6, 9].

1.3 Motivation of current work

The analysis of non-stationary images and videos presents many unique challenges to current digital image and video processing methods. In one-dimensional analysis of non-stationary signals, we often use the short-time Fourier Transform (STFT). When using the STFT, we window in time the signal at different points and then evaluate its fast Fourier Transform (FFT) over that short window of pixels (time in the 1D case) to determine the frequency and phase content. Thus, the STFT is a convenient two-dimensional representation that provides only frequency information at different intervals in time. Unfortunately, this approach cannot be generalized to images and videos. To generalize this STFT approach to images, we would have to evaluate the 2D FFT at a large collection of two-dimensional windows, producing a four-dimensional representation. Furthermore, a six-dimensional representation will be produced for digital videos.

Since the late 1990s, the discrete Wavelet transform (DWT) has been used for transform-based image processing. Unlike the Fourier transform, Wavelet transforms are based on small waves of varying frequency and limited duration. Thus, the DWT is a space-frequency representation of the input signal and it is related to harmonic analysis as in Fourier. While FFT uses equally spaced frequency division, DWT uses logarithmic divisions of the frequency. Thus, Wavelets do not measure the frequency content directly but through scales.

The motivation of using AM-FM methods came from the desire of analyzing and representing non-stationary, yet locally coherent signal structures. These non-stationarities are often related with visual perception and interpretation. AM-FM

methods provide pixel-based information in terms of the instantaneous amplitude, instantaneous frequency and instantaneous phase. Again, while STFT give us only frequency information within a window in time and Wavelets give us the space and frequency information through matching to fixed scales, AM-FM gives us these instantaneous amplitude, phase and frequency information directly from the input signal and at a pixel-based resolution. In contrast with Fourier transforms, AM-FM functions are complex sinusoids admitting non-stationary amplitude and frequency modulations which are allowed to change or adapt arbitrarily throughout the signal domain at a pixel-based resolution.

In AM-FM models, each AM-FM component is inherently and naturally capable of capturing, extracting and representing significant non-stationary structure within the signal, these non-stationary amplitude and frequency modulations are allowed to change or adapt arbitrarily throughout the signal domain at a pixel-based resolution.

Recently, there have been a number of extensions to the one-dimensional AM-FM energy operator algorithms, including an algebraic method that avoids the use of bandpass filters for demodulating multiple components [14], and different approaches for improving the accuracy of the estimation [17], [24], [18]. Similarly, there has been significant research in extending the one-dimensional algorithms based on the analytic signal. A recent extension to the one-dimensional demodulation methods based on the analytic signal can be found in [23]. Recent, multidimensional extensions of the analytic signal demodulation techniques can be found in [3], [50]. Alternative AM-FM methods have also been proposed in [15], [16], and [59]. A modulation bound for discrete modulated signals is given in [60] and one-dimensional Cramer-Rao bounds for a non-linear AR model are given in [61]. For general problems, the one-dimensional, analytic signal methods have been widely accepted as referenced by two recent books [11] and [32]. Strong arguments for the use of one-dimensional analytic signal method are given by Cohen in [32] and Vakman in [11], [12]. Similar

to the one-dimensional uniqueness arguments of Vakman, Havlicek proposed three criteria to support the use of multidimensional, analytic image methods in his PhD dissertation (pages 68-74 in [50]). However, this work has not been extended to three dimensions. A recent example of the standard use of Gabor filters is discussed in [62], where the use of multiple scales for image classification is motivated by models of the human visual system.

The focus of this dissertation is on 2D and 3D extensions of the 1D analytic methods for AM-FM demodulation. In this dissertation, we develop new and efficient AM-FM methods for image and video processing. A new, robust and efficient formulation for the instantaneous frequency (IF) estimation is developed. The new IF estimation method produces more accurate results than current methods and also works in the presence of high levels of noise. We develop algorithms based on AM-FM features for applications in image and video processing. We consider the analysis of digital images and videos using their AM-FM components through the least-square reconstructions of the input signal produced by different frequency scales. An application for the analysis of retinal images for classification of diabetic retinopathy diseases in people is developed based on AM-FM features produced in different frequency scales. A new method for pixel-based motion estimation based on AM-FM methods is developed. We also consider an application of this model in analyzing atherosclerotic plaque motion, general video motion and video activity recognition. This study is motivated from a desire to extend traditional motion estimation methods into the development of reliable methods for video trajectory estimation. This method produces three equations per pixel per scale (AM, FM, and continuity equations) allowing us to produce more accurate results and full dense motion estimates. We extend the application to new AM-FM methods for content based video retrieval for activity recognition.

1.4 Thesis statement

The primary thesis of this dissertation is that the development of new AM-FM methods can lead to a wide range of new and improved image and video processing applications. First, we introduce new, more accurate methods for multidimensional, multi-scale AM-FM demodulation. Second, we demonstrate applications in accurate image and video reconstructions. Third, we discuss new applications in motion estimation, image classification, content-based image retrieval and video activity detection.

1.5 Contributions

The most important contributions include:

- Development of a new, robust and efficient IF demodulation method, taking advantage of new separable, optimal (in the min-max sense), multi-scale filterbanks. Preliminary results using 2D filterbanks were presented in [39]. For video applications, preliminary results using 3D filterbanks were presented in [48] and [49].
- Development of new and efficient AM-FM demodulation methods for image analysis and classification for retinal images. The results have been accepted to be presented in Asilomar 2008 [63] and [64]. This approach led to outstanding classification results in retinal image analysis. This effort has been supported by *VisionQuest Biomedical* through a grant from NIH.
- New 2D/3D reconstruction methods using AM-FM models. Least-squares reconstructions using AM-FM models provide effective 2D/3D signals representations. Results have been presented in [39], [48] and will also be presented in [65].

- New AM-FM demodulation-based motion estimation method. An extension of the original phase model approach in [47], the new AM-FM demodulation method produces better estimates over different scales. Results have been presented in [48] and [49].
- New AM-FM methods for content based video retrieval for activity recognition have been developed. This effort has been supported by DARPA.

1.6 Dissertation Overview

This section contains a brief overview of all that is contained within this dissertation. The dissertation is organized into four parts.

Part I consists of two chapters. Chapter 1 presents a brief introduction to AM-FM demodulation theory together with a literature review of the theory, approaches and applications. The thesis motivation and statement, and the expected contributions for doing this dissertation are also presented here. In chapter 2, we present the foundations of AM-FM demodulation theory.

Then, part II presents the theory and applications of this dissertation related with still image processing. First, in chapter 3, we present robust discrete image AM-FM demodulation methods. Second, chapter 4 presents three image reconstruction methods using least-square approximations based on the AM-FM demodulation methods. Results and discussions are presented for both chapters 3 and 4. Finally, chapter 5 presents the theory, applications and results for retinal image analysis.

Video processing methods are presented in part III. Starting with robust discrete video AM-FM demodulation methods in chapter 6, we then present video reconstructions in chapter 7. A new AM-FM based motion estimation theory is presented in chapter 8. An activity recognition application is presented in this chapter too. The

latter chapters conclude with results and discussions.

Concluding remarks, for image and video processing and future work are presented in the final chapter.

Chapter 2

Background on AM-FM Demodulation Methods

2.1 AM-FM demodulation methods

2.1.1 Continuous Space AM-FM Demodulation based on extensions of the Analytic Signal

Consider a single component AM-FM image defined by

$$I(x,y) = a(x,y)\cos\varphi(x,y). \tag{2.1}$$

In this case, the AM-FM demodulation problem is to estimate: (i) the amplitude function $a(x,y)$, (ii) the phase function $\varphi(x,y)$ and (iii) the instantaneous frequency function $\nabla\varphi(x,y)$ from $I(\cdot)$.

Analytic image methods for AM-FM demodulation are based on extending the notion of the 1D analytic signal to 2D or simply to provide a Hilbert-based extension of the 1D Hilbert-based demodulation approach. Early work can be found in [50].

Bedrosian in [66] presented the Analytic Signal representation of modulated waveforms. Nuttall and Bedrosian in [67] presented a quantitative measure of the discrepancy between the Hilbert transform and the quadrature version of a carrier with amplitude and/or phase modulation. This measure was based on the Hilbert transform theorem. Bedrosian in [68] stated this theorem and Stark in [69, 70] extended it to n-dimensions. It states that, in the 1D single case, for well-behaved functions f and g, the Hilbert transform of the product fg is given by $\mathcal{H}[f(x)g(x)] = f(x)\mathcal{H}[g(x)]$ if one of the following conditions holds: (i) the functions f and g are low-pass and high-pass functions respectively, with non-overlapping spectra, or (ii) the functions f and g are analytic, i.e., their real and imaginary parts are Hilbert parts. Picinbono in [71, 72] defined that for a 1D signal, the function $z(t)$ is the analytic signal of $x(t)$ if and only if it can be written as $z(t) = x(t) + jy(t)$, where $y(t)$ is the Hilbert transform of $x(t)$. Then, the canonical pair associated with $x(t)$ is $[a(t), \varphi(t)]$ with $z(t) = a(t) \exp \varphi(t)$.
In [57, 58], the authors introduce the phase quadrature transform for extending the Hilbert transform into two dimensions. In [9], Felsberg and Sommer introduced the monogenic signal, a nice extension of the analytic signal to images. The monogenic signal is a three-dimensional representation that combines an image with its Riesz transform to yield a sophisticated 2D analytic signal. Some related work on image demodulation based on the analytic image includes [4, 5, 6, 9].

The AM-FM demodulation methods in this work are based on extensions of the one-dimensional analytic signal. The first step is to compute the extended analytic signal associated with $I(x, y)$. The extended analytic signal is computed using [50]:

$$I_{AS}(x, y) = I(x, y) + j\mathcal{H}_{2d}[I(x, y)], \tag{2.2}$$

where $\mathcal{H}_{2d}$ denotes a two-dimensional extension of the one-dimensional Hilbert transform operator. The two dimensional operator is defined in terms of the one dimen-

sional operator, operating in either the x or the y direction:

$$\mathcal{H}_{2d}[I(x,y)] = \frac{1}{\pi x} * I(x,y). \tag{2.3}$$

For the algorithm to work, we must have that $I_{AS}(\cdot)$ is equal, or at-least approximately equal, to the complex AM-FM harmonic associated with (2.1) (see [32] for the 1D case):

$$I_{AS}(x,y) \approx a(x,y) \exp\left(j\varphi(x,y)\right). \tag{2.4}$$

When the approximation holds, it is possible to estimate the amplitude, the phase and the instantaneous frequency using

$$a(x,y) = |I_{AS}(x,y)|, \tag{2.5}$$

$$\varphi(x,y) = \arctan\left(\frac{\text{imag}(I_{AS}(x,y))}{\text{real}(I_{AS}(x,y))}\right) \tag{2.6}$$

and

$$\omega(x,y) = \text{real}\left[-j\frac{\nabla I_{AS}(x,y)}{I_{AS}(x,y)}\right]. \tag{2.7}$$

The algorithm can thus be summarized into two steps. First, compute the extended analytic signal using (2.2). Second, compute of all the estimates using (2.5), (2.6) and (2.7).

The procedure for working with 3D signals is similar and will be explained in chapter 6.

2.1.2 Uniqueness of the Hilbert-based approaches

This work focuses on two- and three-dimensional extensions of the one-dimensional concept of an analytic signal (or equivalent concepts), usually computed using the Hilbert transform. In [11, 12] an axiomatic approach that shows that the analytic-signal demodulation method is the only approach that satisfies certain intuitive conditions was presented:

- Amplitude continuity and differentiability. Small variation in the input signals should lead to small variations in the instantaneous amplitude (IA).
- Phase independence of scaling and homogeneity. If the input signal is scaled by a constant factor, its instantaneous frequency (IF) and instantaneous phase (IP) should be the same as the ones in the original signal.
- Harmonic correspondence. An input signal with constant amplitude and frequency (e.g., a sinusoid) should retain its value, i.e., estimation should be theoretically exact.

Similar to the one dimensional, uniqueness arguments of Vakman, Havlicek proposed three arguments for the multidimensional, analytic image methods, based on an adjusted Hilbert transform, in his PhD dissertation (pages 68-74 and Appendix A in [50]).

2.1.3 Quasi-Eigen Approximation (QEA) algorithm

For discrete-space signals, we consider the discrete-space input signal given by

$$I(k_1, k_2) = a(k_1, k_2) \cos \varphi(k_1, k_2), \tag{2.8}$$

where k_1 and k_2 represent the discrete versions of x and y, respectively. We define the discrete-space extended estimate of the 1D analytic signal $\hat{I}_{AS}(k_1, k_2)$. This is accomplished via the use of a discrete-time Hilbert Transform or using the Fast Fourier Transform (FFT), as given in (2.2).

The instantaneous frequency (IF) can be estimated using (see [13]):

$$\hat{\varphi_x}(k_1,k_2) = \arcsin\left[\frac{\hat{I}_{AS}(k_1+1,k_2)-\hat{I}_{AS}(k_1-1,k_2)}{2j\hat{I}_{AS}(k_1,k_2)}\right], \tag{2.9}$$

$$\hat{\varphi_y}(k_1,k_2) = \arcsin\left[\frac{\hat{I}_{AS}(k_1,k_2+1)-\hat{I}_{AS}(k_1,k_2-1)}{2j\hat{I}_{AS}(k_1,k_2)}\right], \tag{2.10}$$

$$\hat{\varphi_x}(k_1,k_2) = \arccos\left[\frac{\hat{I}_{AS}(k_1+1,k_2)+\hat{I}_{AS}(k_1-1,k_2)}{2\hat{I}_{AS}(k_1,k_2)}\right] \tag{2.11}$$

and

$$\hat{\varphi_y}(k_1,k_2) = \arccos\left[\frac{\hat{I}_{AS}(k_1,k_2+1)+\hat{I}_{AS}(k_1,k_2-1)}{2\hat{I}_{AS}(k_1,k_2)}\right]. \tag{2.12}$$

The IA and IP are estimated using $\hat{I}_{AS}(k_1,k_2)$ as described in sub-section 2.1.1.

2.1.4 Quasi-Local Method (QLM)

The Quasi Local Method (QLM) was introduced, for 1D signals and half the discrete frequency spectrum, by Girolami and Vakman in [23]. In 2005, Rodriguez in [13] extended the method for 2D signals, the entire, discrete frequency spectrum and any finite number of dimensions without the need of resampling.

The discrete-time IA can be estimated using:

$$\hat{a}(k_1,k_2) = \sqrt{2\breve{g}_{(0,0)}(k_1,k_2)}, \tag{2.13}$$

where $\breve{g}_{(\epsilon_1,\epsilon_2)}(k_1,k_2) = h_{LP}(k_1,k_2) * g_{(\epsilon_1,\epsilon_2)}(k_1,k_2)$, $g_{(\epsilon_1,\epsilon_2)_x}(k_1,k_2) = f(k_1+\epsilon_1,k_2)f(k_1-\epsilon_2,k_2)$, $g_{(\epsilon_1,\epsilon_2)_y}(k_1,k_2) = f(k_1,k_2+\epsilon_1)f(k_1,k_2-\epsilon_2)$ ($\epsilon_1,\ \epsilon_2 \geq 0$), and $h_{LP}(k_1,k_2)$ is a discrete lowpass filter.

The IF through the x direction can be estimated using:

$$\hat{\varphi_x}(k_1,k_2) = \arccos\left(\frac{R_x(k_1,k_2)+\sqrt{R_x^2(k_1,k_2)+8}}{4}\right), \tag{2.14}$$

where

$$R_x(k_1,k_2) = \frac{2\breve{g}_{(1,1)_x}(k_1,k_2)}{\breve{g}_{(1,0)_x}(k_1,k_2) + \breve{g}_{(0,1)_x}(k_1,k_2)}. \tag{2.15}$$

Similar analysis is done for the y direction. In order to avoid aliasing, the IF of the input signal must be restricted to $0 < \varphi_i(k_1,k_2) < \pi f_s/2$, for φ_x or φ_y, where f_s is the sampling frequency. When $\pi f_s/2 \leq \varphi_i(k_1,k_2) < \pi f_s$, for φ_x or φ_y, [13], the IF through the x direction can be estimated using:

$$\hat{\varphi_x}(k_1,k_2) = \pi - \arccos\left(\frac{-R_x(k_1,k_2) + \sqrt{R_x^2(k_1,k_2)+8}}{4}\right), \tag{2.16}$$

Note that this method does not use the Hilbert transform for the estimations and the estimated instantaneous frequency is done only in magnitude. Thus, the Hilbert transform will be used for the sign estimation as it is done in QEA methods. Moreover, the QLM does not estimate the instantaneous phase (IP), so the QEA method is used for the IP. The basic approach presented in Rodriguez's dissertation is summarized in Fig. 2.1.

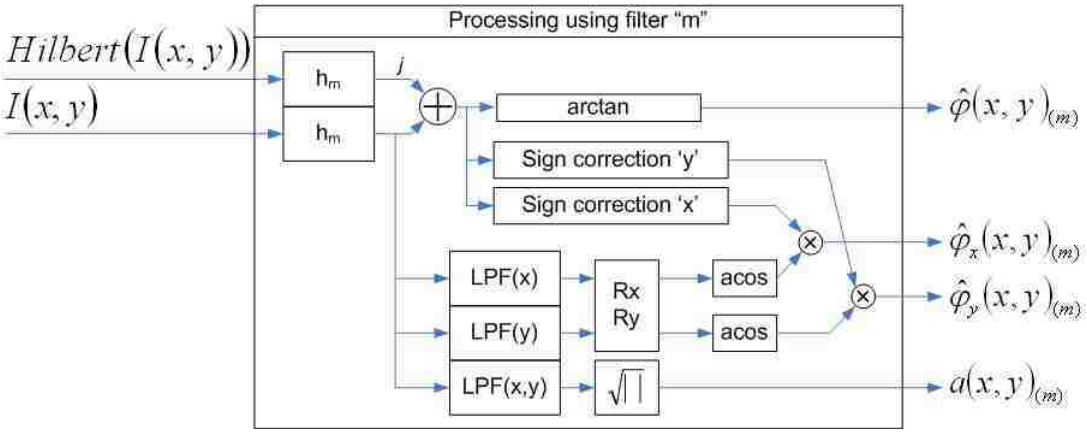


Figure 2.1: Block diagram of extended QLM method.

Part II

AM-FM methods for image processing

Chapter 3

Robust Discrete Image AM-FM Demodulation Methods

In this chapter we develop robust AM-FM demodulation methods. We begin with the design of an efficient filterbank (section 3.1) to be used in all the 2D methods and applications described here. Next we present a robust approach to use the QEA and QLM methods in sections 3.2 and 3.3, respectively. Then we present a new and robust method for the IF estimation: VS-LQP (in section 3.4). We begin with a discussion on the stable evaluation of functions and extend it to the stable evaluation of inverse trigonometric functions. Also, an optimum value of the displacement for low frequency values of IF is introduced. Finally, results and discussion are presented in the last sections.

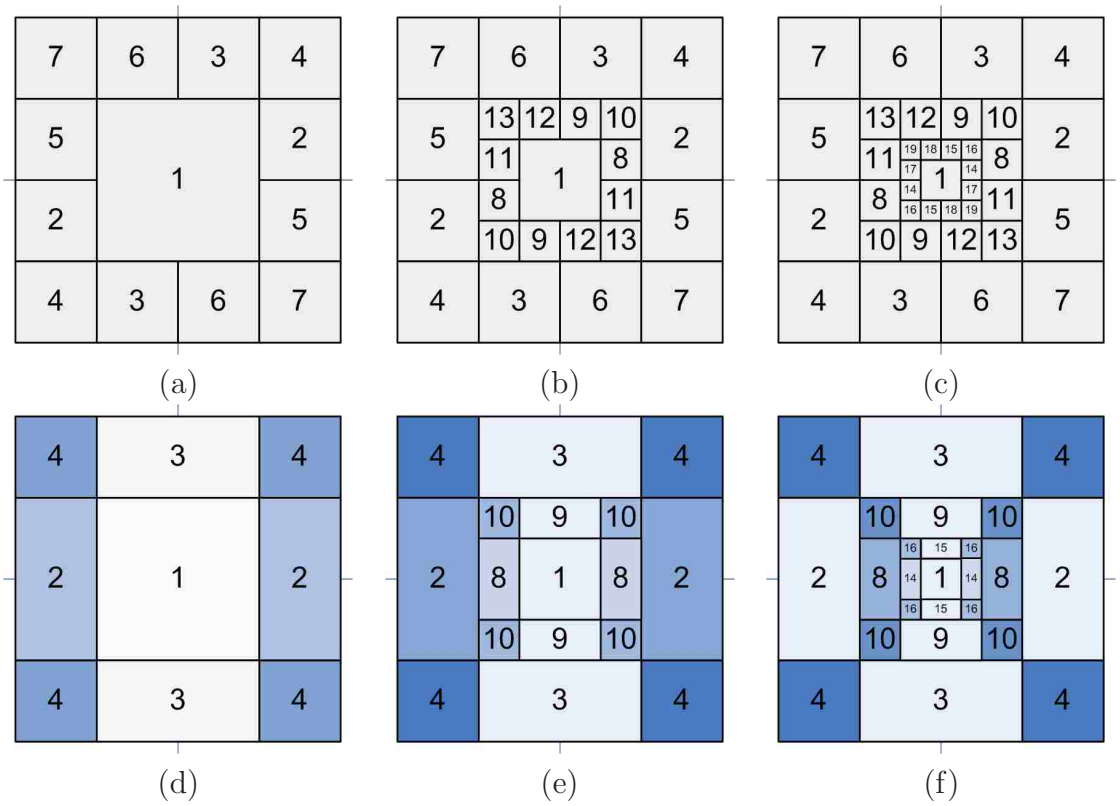


Figure 3.1: Multi-scale filterbanks designed using optimal separable filter design versus traditional filterbanks with support in four quadrants. For AM-FM analysis we use: (a) Single-scale, (b) two-scale and (c) three-scale filterbanks. For traditional Wavelet analysis we show: (d) Single-scale, (e) two-scale and (f) three-scale filterbanks.

3.1 2-D Multi-scale filterbank design for applications in discrete images

We use separable multi-scale filterbanks covering the whole frequency spectrum (see Fig. 3.1 (a)-(c)). Also, in this new design each bandpass filter has frequency support in only two quadrants of the frequency spectrum. This approach produces a correct sign estimation compared with the traditional Wavelet-like bandpass filters with

support in four quadrants (see Fig. 3.1 (d)-(f)).

The filters were designed using a min-max, equiripple approach. Passband ripple was set at $0.017dB$ and the stopband attenuation was set to $66.02dB$. For all filters, the transition bandwidths were fixed to $\pi/10$. Based on this approach, the unit gain over the passband eliminated the need for amplitude correction, as required by a Gabor filterbank approach (see [3]).

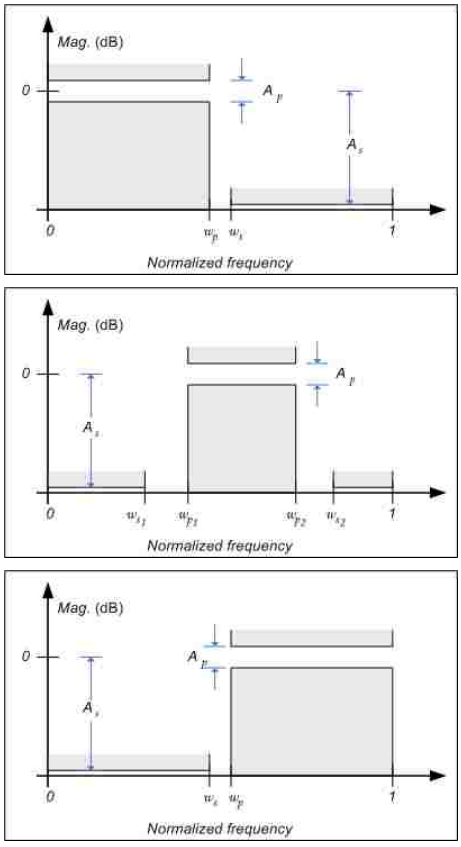


Figure 3.2: Min-max specifications for 1D filter design. Multidimensional filters can be efficiently implemented using 1D filters. (a) Lowpass filter. (b) Bandpass filter. (c) Highpass filter.

3.2 Robust multi-scale implementation of the QEA algorithms

We can implement the filterbanks in either the spatial or the discrete frequency domain. The two methods are summarized in Fig. 3.3.

For both algorithms, a sign correction process for the estimated instantaneous frequency has been added. This sign correction process is based on the arcsin of (2.9) and (2.10).

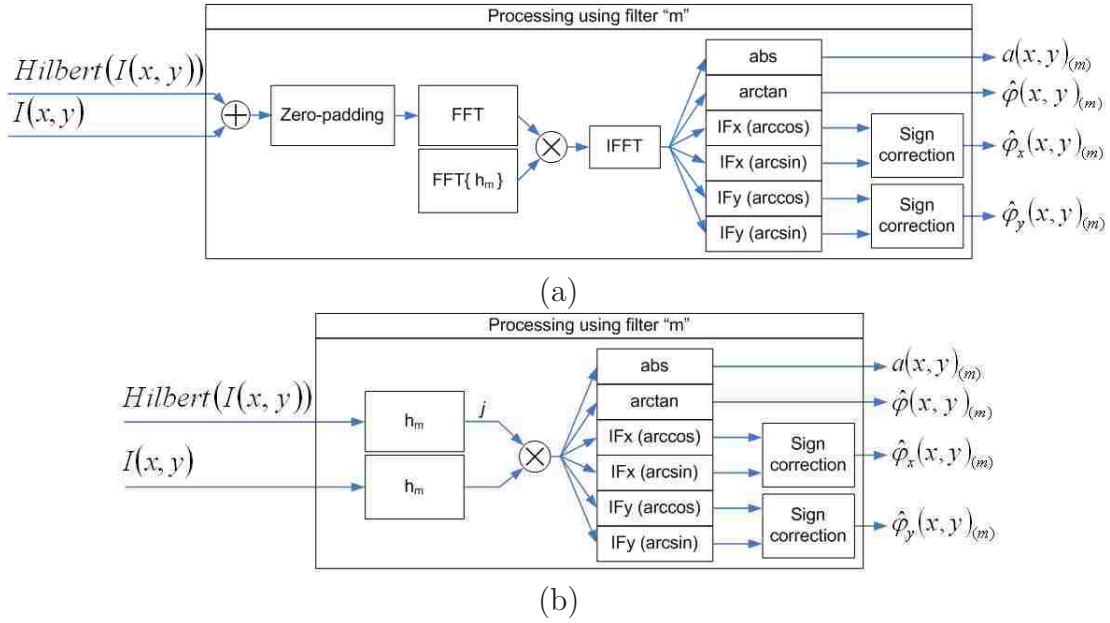


Figure 3.3: Block diagrams of the basic QEA methods. (a) Based on FFT. The Hilbert filter is applied along the column direction. (b) Using direct, discrete space convolutions. Here, h_m denotes the 2D impulse response for filter m. Note that for the FFT approach in (a), we will first need to zero-pad. Also, since our filters are separable, note that we can compute the convolution result using separable, 1D convolutions along the rows and columns. Similarly, we can use 1D FFTs for more efficient implementations.

3.3 Robust multi-scale implementation of the Quasi Local Method (QLM)

The original QLM was described in the sub-section 2.1.4. We note that a bad Lowpass Filter (LPF) design will significantly reduce the accuracy of the estimates. Then, the first step to have a better estimation is with the design of an efficient LPF. The design of the LPF was done using min-max designs as in section 3.1 with a cutoff frequency set to $\pi/10$, transition band set to $\pi/100$, with the same ripple and attenuation to the other cases. Next, depending on the central frequency support of the filter used in the filterbank, the range of R_x and R_y must be checked to avoid estimation problems. The ranges are checked in each processing block. The valid ranges are:

$$R_x \in \begin{cases} [-\infty, 1], & \text{if } |w_{c_x}^i| \leq \pi f_s/2 \\ [-1, \infty], & \text{otherwise} \end{cases}$$

$$R_y \in \begin{cases} [-\infty, 1], & \text{if } |w_{c_y}^j| \leq \pi f_s/2 \\ [-1, \infty], & \text{otherwise} \end{cases}$$

where $w_{c_x}^i$ and $w_{c_y}^j$ are the central frequencies of the filter used in the processing block i through the horizontal direction and j through the vertical direction, respectively. In addition to the factor corrections, the use of the filterbanks designed in section 3.1 significantly improves the estimation accuracy.

3.4 A robust Instantaneous Frequency estimation based on a Local Quadratic Phase (LQP) model

We present a new method for robust IF estimation based on a local quadratic phase model [1]. We use the term variable spacing, local quadratic phase methods

(VS-LQP) to describe the approach the we develop in this section. Our objective is to develop and accurate method.

In our development of AM-FM demodulation methods, we have encountered three inverse trigonometric functions: arcsin, arccos and arctan. Their condition numbers are given by

$$(\text{cond arccos})(x) = \frac{|x|}{|\arccos(x)||\sqrt{1-x^2}|}, \tag{3.1}$$

$$(\text{cond arcsin})(x) = \frac{|x|}{|\arcsin(x)||\sqrt{1-x^2}|} \quad \text{and} \tag{3.2}$$

$$(\text{cond arctan})(x) = \frac{|x|}{|\arctan(x)||1+x^2|}. \tag{3.3}$$

Here, we note that large condition numbers reflect numerical instability (see Fig. 3.4). The condition numbers become infinite for both the arcsin and the arccos for $x = \pm 1$. Otherwise, we have finite condition numbers for all values of x, including the cases when $\arcsin(x) = 0$, $\arccos(x) = 0$, or $\arctan(x) = 0$.
From Fig. 3.4, it is clear that for $|x| < 0.6$ evaluating the arccos is more stable than evaluating the arcsin. Furthermore, the most stable evaluations occur for small values of $|x|$, when using the arccos function.

Even though the arccos function provides the most stable evaluations, we also need to use the arcsin function to estimate the proper IF quadrant. On the other hand, we also find that there are also significant problems in accurate estimation at very high frequencies. As it is well-known, the difference operator of (2.9)-(2.10) greatly amplifies high-frequency noise (see [73]). Thus, in what follows, we chose to use the arccos formulas of (2.11)-(2.12), despite the fact that they will not yield accurate estimates at high-frequencies.

To derive a more stable algorithm for the low instantaneous frequency values, we return to the QEA equations (see [13]), and consider the case for arbitrary spacings

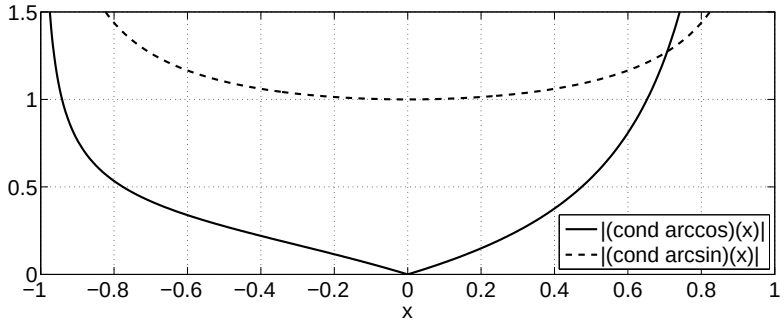


Figure 3.4: Absolute value of the condition number for the arccos (solid line) and arcsin (dash line) is shown. From this graph, it is clear that we have the most stable evaluation for the arccos function $|x| \approx 0$.

n_1 and n_2. For n_1, we have:

$$\begin{aligned}
&\gamma_{\arccos}(n_1) \\
&= \frac{\hat{I}_{AS}(k_1+n_1,k_2)+\hat{I}_{AS}(k_1-n_1,k_2)}{2\hat{I}_{AS}(k_1,k_2)} \\
&= \frac{a(k_1+n_1,k_2)\exp(j\varphi(k_1+n_1,k_2))+a(k_1-n_1,k_2)\exp(j\varphi(k_1-n_1,k_2))}{2a(k_1,k_2)\exp(j\varphi(k_1,k_2))},
\end{aligned} \tag{3.4}$$

where $\gamma_{\arccos}(n_1)$ is the input for the arccos function for φ_x estimation in (2.11). Now, if we can guarantee that $\gamma_{\arccos}(n_1)$ is small, then the application of the arccos function will be stable (see (3.1) and Fig. 3.4).

In the original QEA, we assume that $a(k_1+n_1,k_2) \approx a(k_1-n_1,k_2) \approx a(k_1,k_2)$ to get

$$\begin{aligned}
\frac{\hat{I}_{AS}(k_1+n_1,k_2)+\hat{I}_{AS}(k_1-n_1,k_2)}{2\hat{I}_{AS}(k_1,k_2)} &\approx \frac{\exp(j\varphi(k_1+n_1,k_2))+\exp(j\varphi(k_1-n_1,k_2))}{2\exp(j\varphi(k_1,k_2))} \\
&= \cos(n_1\varphi_x(k_1,k_2)).
\end{aligned}$$

Clearly, this assumption will be violated for any non-constant amplitude. To improve on the accuracy of the estimation, we will first estimate the instantaneous amplitude

without requiring the prior estimate of the instantaneous frequency. Here, it is important to note that this approach is now possible because we use unit gain filterbanks (see section 3.1).

In the original QEA, the IF estimate was used to correct the instantaneous amplitude estimate due to pre-filtering by a Gabor filterbank (with non-unit gain over the passband, see [3, 50]). In our proposed approach, we use a plug-in rule to produce a new signal $\bar{I}_{AS}$ with unit instantaneous amplitude using:

$$\begin{aligned}\bar{I}_{AS}(k_1,k_2) &= \frac{\hat{I}_{AS}(k_1,k_2)}{|\hat{I}_{AS}(k_1,k_2)|} \\ &= \frac{a(k_1,k_2)\exp(j\varphi(k_1,k_2))}{a(k_1,k_2)} \\ &= \exp(j\varphi(k_1,k_2)). \end{aligned} \tag{3.5}$$

Let's consider a quadratic approximation for $\varphi(\mathbf{k})$ (see [1]) of order $\|\mathbf{k}-\mathbf{k}_0\|^3$, such that

$$\varphi(\mathbf{k}) \approx \varphi(\mathbf{k}_0) + (\mathbf{k}-\mathbf{k}_0)^T\nabla\varphi(\mathbf{k}_0) + \frac{1}{2}(\mathbf{k}-\mathbf{k}_0)^T\mathbf{F}(\mathbf{k}_0)(\mathbf{k}-\mathbf{k}_0),$$

where $\mathbf{F}(\mathbf{k})$ denotes the Hessian of $\varphi(\mathbf{k})$. Then

$$\begin{aligned}\frac{\bar{I}_{AS}(k_1+n_1,k_2)+\bar{I}_{AS}(k_1-n_1,k_2)}{2\bar{I}_{AS}(k_1,k_2)} &= \frac{\exp(j\varphi(k_1+n_1,k_2))+\exp(j\varphi(k_1-n_1,k_2))}{2\exp(j\varphi(k_1,k_2))} \\ &= \cos(n_1\varphi_x(k_1,k_2)). \end{aligned} \tag{3.6}$$

This gives

$$\varphi_x(k_1,k_2) = \frac{1}{n_1}\arccos\left(\frac{\bar{I}_{AS}(k_1+n_1,k_2)+\bar{I}_{AS}(k_1-n_1,k_2)}{2\bar{I}_{AS}(k_1,k_2)}\right). \tag{3.7}$$

The analysis for φ_y is similar. The goal is to have $\gamma_{\arccos}(n_1)$ as close to 0 as possible. We consider four possible values: $n_1 = 1, 2, 3$ and 4. We do not increase more the value of n_1 since we could go to an unstable zone for (3.1) (see Fig. 3.4).

Note that for the argument of the arccos(·) to be zero, we require that

$$\bar{I}_{AS}(k_1 - n_1, n_2) = -\bar{I}_{AS}(k_1 + n_1, n_2).$$

Thus, for $n_1 = 1$, the maximum frequency that we can attain, without requiring interpolation, will be $w_1 = \pi/2$. In what follows, we will only consider integer values for n_1. In particular, for $n_1 = 4$, we can consider low IF down to $w_1 = \pi/8$.

Next, using a bandpass filter with cutoff frequencies given by $w_{p_{1x}}$ and $w_{p_{2x}}$ (see section 3.1), given the input signal $\bar{I}_{AS}(x, y)$, we estimate the sign of φ_x using (2.9) as it is done using QEA method. To estimate the IF magnitude, see Fig. 3.5, we look first at the estimate $\gamma_{\arccos}(n_1)$ for $n_1 = 1, 2, 3, 4$. Each value for n_1 produces an estimate for φ_x ($\hat{\varphi}_{1x}$, $\hat{\varphi}_{2x}$, $\hat{\varphi}_{3x}$ and $\hat{\varphi}_{4x}$ in Fig. 3.5(b)). Finally, the magnitude of φ_x is given by the estimate with n_1 such that both $|\gamma_{\arccos}(n_1)|$ is minimum and φ_x, for the same n_1, belongs to the frequency support of the bandpass filter:

$$\min_{n_1=1,2,3,4} |\gamma_{\arccos}(n_1)| \quad \text{subject to} \quad \varphi_x \in [w_{p_{1x}}, w_{p_{2x}}]. \tag{3.8}$$

Similar for φ_y.

3.4.1 Pre-filtering and post-filtering

To reject impulsive noise without sacrificing bandwidth, as in the case of using filterbanks, we apply a median filter to the input signal. Also, the same median filter is applied to the IF estimation outputs (for both x and y directions) to provide continuity in the estimates. Fig. 3.6 depicts this approach.

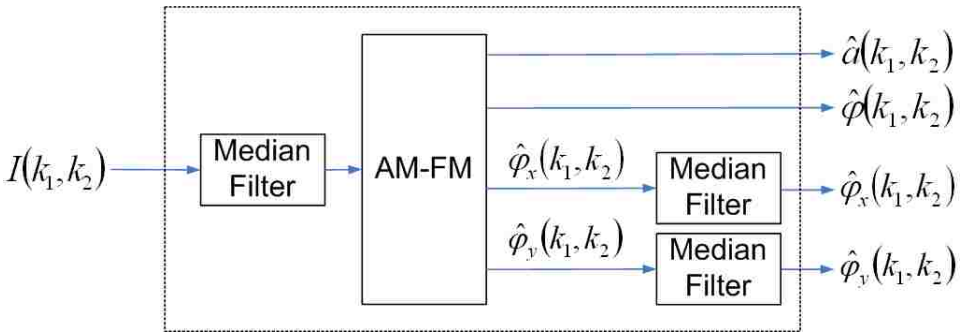


Figure 3.6: Pre-filtering and post-filtering using a median filter to improve accuracy in the IF estimation. The median filter is applied to the input signal, to reject impulsive noise without sacrificing bandwidth, as well as to the IF estimates, to improve continuity.

3.4.2 Modulation for High Frequencies

Recall from the VS-LQP method that, for spacing $n_1 = 1$, the maximum frequency that we can estimate accurately, without requiring interpolation, is $w_1 = \pi/2$. Thus, for high frequencies, we apply a modulation to the input signal given by $\exp(jw_{M_1}k_1 + jw_{M_2}k_2)$ to modulate the input signal to the medium-low frequencies. The frequencies w_{M_1} and w_{M_2} are determined by the frequency support of the band-pass filter where the dominant amplitude is located. Table 3.1 shows the values of w_{M_1} and w_{M_2} depending on the frequency support of the signal.

3.5 Results for Robust AM-FM estimation

In this section, we present results from all methods using synthetic images. In the case of using the QEA method (section 3.2), we filter the signal through each channel using direct, discrete space convolutions.

We first define the synthetic images used, then define the metric to measure the

Table 3.1: Modulation frequency factors for high frequency signals.

Bandpass filter	w_{M_1}	w_{M_2}
2	$+\pi/2$	0
3	0	$-\pi/2$
4	$+\pi/2$	$-\pi/2$
5	$-\pi/2$	0
6	0	$-\pi/2$
7	$-\pi/2$	$-\pi/2$

error in terms of the IF. Finally, we present the results.

3.5.1 Gaussian amplitude modulated (Gaussian AM)

We use

$$I(k_1,k_2) = 100\ \exp\left[-\frac{1}{2}\alpha^2\left(\left(\frac{k_1+\frac{1}{2}}{\frac{N}{2}}\right)^2+\left(\frac{k_2+\frac{1}{2}}{\frac{N}{2}}\right)^2\right)\right]\cos\left(w(k_1+k_2)\right),$$

for $k_1,k_2 = -\frac{N}{2},\ldots,\frac{N}{2}-1$, with $N=512$, $\alpha=2.5$. Here, $\varphi_x(\cdot)=\varphi_y(\cdot)=w$, for $w=\{\pi/3,\pi/5,\pi/6,\pi/8,\pi/12,\pi/14,\pi/16\}$.

3.5.2 Gaussian amplitude-modulated Quadratic frequency-modulated (Gaussian AM Quadratic FM)

For this signal, we use

$$I(k_1,k_2) = 100\ \exp\left[-\frac{1}{2}\alpha^2\left(\left(\frac{k_1+\frac{1}{2}}{\frac{N}{2}}\right)^2+\left(\frac{k_2+\frac{1}{2}}{\frac{N}{2}}\right)^2\right)\right]\cos\varphi(k_1,k_2)$$

for $k_1,k_2 = -\frac{N}{2},\ldots,\frac{N}{2}-1$ and $N=512$. The phase $\varphi(k_1,k_2)$ has a quadratic form

$$\varphi(k_1,k_2) = \alpha_x k_1 + \frac{\beta_x}{2}k_1^2 + \alpha_y k_2 + \frac{\beta_y}{2}k_2^2,$$

where $\beta_i = \left(\frac{f_{B_i} - f_{A_i}}{N-1}\right)$, $\alpha_i = \frac{\beta_i N}{2} + f_{A_i}$, for $i = 1, 2$. f_{B_i} represents the desired maximum instantaneous frequency in the i direction (x or y), and f_{B_i} represents the desired minimum instantaneous frequency. Setting f_{B_i} and f_{A_i} will produce $\varphi_x(k_1, k_2)$ and $\varphi_y(k_1, k_2)$ to be in desired ranges. We consider four cases in terms of the instantaneous frequency $\nabla\varphi(k_1, k_2)$: (i) $\varphi_x(k_1, k_2) \in [0.001\pi, 0.01\pi]$, (ii) $\varphi_x(k_1, k_2) \in [0.01\pi, 0.1\pi]$, (iii) $\varphi_x(k_1, k_2) \in [0.166\pi, 0.2\pi]$ and (iv) $\varphi_x(k_1, k_2) \in [0.2\pi, 0.3\pi]$. For all cases, $\varphi_y(k_1, k_2) = -\varphi_x(k_1, k_2)$.

3.5.3 Error in IF estimation

Due to the ambiguity of the instantaneous frequency vectors (coming from $\cos\varphi(k_1, k_2) = \cos(-\varphi(k_1, k_2))$, we compute the mean-squared error using the closest estimation between $\nabla\hat{\varphi}(k_1, k_2)$ and $-\nabla\hat{\varphi}(k_1, k_2)$, to $\nabla\varphi(k_1, k_2)$:

$$\begin{aligned} e_i = & \frac{1}{N \cdot N} \cdot \sum_{k_1,k_2=1}^{N,N} \mathcal{S}(k_1, k_2) \left(\varphi_i(k_1, k_2) - \hat{\varphi}_i(k_1, k_2)\right)^2 \\ & + \frac{1}{N \cdot N} \cdot \sum_{k_1,k_2=1}^{N,N} \left(1 - \mathcal{S}(k_1, k_2)\right) \left(\varphi_i(k_1, k_2) + \hat{\varphi}_i(k_1, k_2)\right)^2, \end{aligned} \tag{3.9}$$

where:

$$\mathcal{S}(k_1, k_2) = \begin{cases} 1, & \text{if } \left|\left|\hat{f}_{sub}(k_1, k_2)\right|\right|_2 \leq \left|\left|\hat{f}_{add}(k_1, k_2)\right|\right|_2 \\ 0, & \text{otherwise} \end{cases},$$

with $\hat{f}_{sub} = \nabla\varphi - \hat{\nabla\varphi}$ and $\hat{f}_{add} = \nabla\varphi + \hat{\nabla\varphi}$.

3.5.4 Results for synthetic images

In Figure 3.7 we show the PSNR (given by $20\log_{10}\left(100/\sqrt{MSE}\right)$) for IA estimation. Since QEA and VS-LQP use the same approach for the IA estimation, we only

present the results for the QEA and QLM methods. We compare the results in terms of the PSNR using: (i) no filterbank, (ii) the two-scale filterbank and (iii) the three-scale filterbank. We compare the methods in the presence of white Gaussian noise ($SNR = \{20.00, 30.46, 40.00, 50.46\}dB$). We use all the signals described at the beginning of this section.

Figure 3.8 compares the error in the IF estimation among all the methods. We compare the results in terms of the PSNR (given by $20\log_{10}\left(\pi/\sqrt{MSE}\right)$) using: (i) no filterbank (first column), (ii) the two-scale filterbank (second column) and (iii) the three-scale filterbank (last column). We compare the methods in the presence of white Gaussian noise ($SNR = \{20.00, 30.46, 40.00, 50.46, 60.00\}dB$). Since the results in x and y directions are similar, we only show the results in the x direction. In Figure 3.9 we show the results that produced the worst PSNR for the proposed methods.

Results with median filter and modulation

We compare VS-LQP versus QEA applying the pre- and post-filtering using median filters without the use of filterbanks in Figs. 3.10 (a) and (b) for Gaussian amplitude modulated signals and Gaussian amplitude-modulated frequency-modulated signals, respectively. We compare the methods in the presence of white Gaussian noise from $SNR = 1.9dB$ to $SNR = 40dB$ applied to the input signal.

Figs. 3.10 (c) and (d) show the results for high frequency ($\varphi_x = \varphi_y = \{5\pi/6, 7\pi/10, 4\pi/6, 5\pi/8\}$) Gaussian amplitude modulated signals using the modulation technique in the presence of Gaussian noise from $SNR = 1.9dB$ to $SNR = 40dB$ applied to the input signal. A three-scale filterbank was used. We compare VS-LQP versus QEA in (c) and versus QLM in (d).

3.6 Discussion for Robust AM-FM estimation

As we stated in section 3.4, since the VS-LQP method is designed for low instantaneous frequencies signals, we do not consider input signals with IF bigger than, or closer to, $\pi/2$ (for either x or y direction). Also, we consider synthetic signals with quadratic phase to compare the new methods.

For IA estimation, and considering the QEA method (see Figs. 3.7 (a), (b), (d) and (e)), we can see that using the multi-scale filterbank improved the accuracy of the estimation in $\sim 10dB$. Note that the three-scale filterbank produced better results than the two-scale filterbank. These results are clear when Gaussian AM signals were used. However, when a Gaussian AM-FM signal was used, and the frequencies were very close to DC, the IA estimation was not good when a filterbank was used. The reason is that the frequency spectrum of the IA is close or overlapped with the frequency support of the FM part, producing instability for the AM-FM demodulation approach (see sub-section 2.1.1 and [67, 69, 68, 70]).
When we use the QLM, the three-scale filterbank results were very similar to the ones using the two-scale filterbank. The same problem as in QEA of the very low frequencies using a Gaussian AM-FM is seen.

Now, we continue our discussion for IF estimation. From the results, we show that the VS-LQP approach yields significant improvements. Without using a filterbank, we can see from Figs. 3.8 (a) and (f), how the VS-LQP method is robust in the presence of noise, always giving the best results. Typical results without using a filterbank range from $10dB$ to $20dB$ better than the QEA method for noise levels from $20dB$ to $60dB$. When the QLM was used (see Figs. 3.8 (d), (e), (i) and (j)), the improvement of using a three-scale filterbank versus a two-scale filterbank was around $25dB$ for Gaussian AM signals and around $22dB$ for Gaussian AM-FM signals in the presence of $60dB$ of noise.

At significant noise levels ($SNR < 30dB$), for single component images, the VS-LQP method produces the best results, which are also better than when using a multi-scale filterbank. At low noise levels ($SNR > 50dB$), VS-LQP performs better when used in combination with a multi-scale filterbank. In all cases, VS-LQP outperforms the QEA algorithm by significant amounts (up to $20dB$).

Fig. 3.9 shows the worst cases for IF estimation errors when the input signals have frequencies close to $\pi/2$. In Figs. 3.9 (a) and (d) we can see how VS-LQP is not better than QEA when the presence of noise was reduced. However, note that as the noise is increased, the VS-LQP produces better results. In Figs. 3.9 (b), (c), (e) and (f), we can see how the QEA and VS-LQP produced very similar good results whereas QLM always produced the lower PSNR.

Discussion for median filter and modulation

The use of pre- and post-filtering using median filters in the presence of high levels of noise for medium and low frequency input signals does improve the accuracy of the IF estimation (see Figs. 3.10 (a) and (b)). VS-LQP always produces the best results. In the presence of noise from $SNR = 1.9dB$ to $SNR = 40dB$, VS-LQP was better than QEA from $5dB$ to $20dB$ in the PSNR sense.

When the input signal is a high frequency signal, VS-LQP produced better IF estimations than QEA and QLM in the presence of high levels of noise (see Figs. 3.10 (c) and (d)). Using a three-scale filterbank, VS-QLP produced high PSNR values in the presence of noise from $SNR = 1.9dB$ to $40dB$. VS-QLP was up to $14dB$ better than QEA and up to $20dB$ better than QLM.

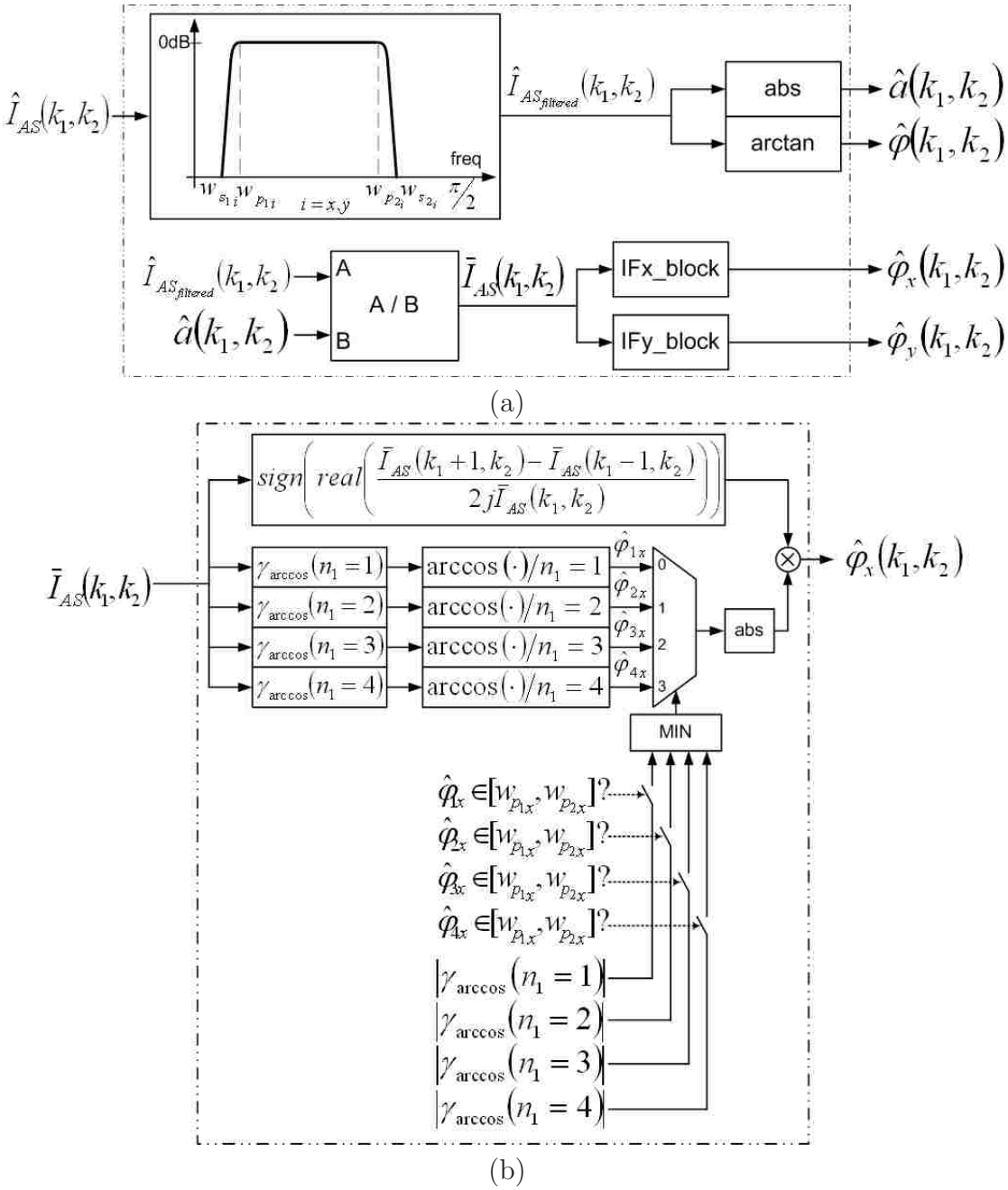


Figure 3.5: VS-LQP block diagram using a general bandpass filter in (a). Since separable 1D filters are used, the spectrum of the bandpass filter is showed in 1D and it is generic for x and y. IFx_block is showed in (b) (IFy_block is similar). $\gamma_{\arccos}(n_1) = \left(\bar{I}_{AS}(k_1+n_1,k_2) + \bar{I}_{AS}(k_1-n_1,k_2)\right) / \left(2\bar{I}_{AS}(k_1,k_2)\right)$.

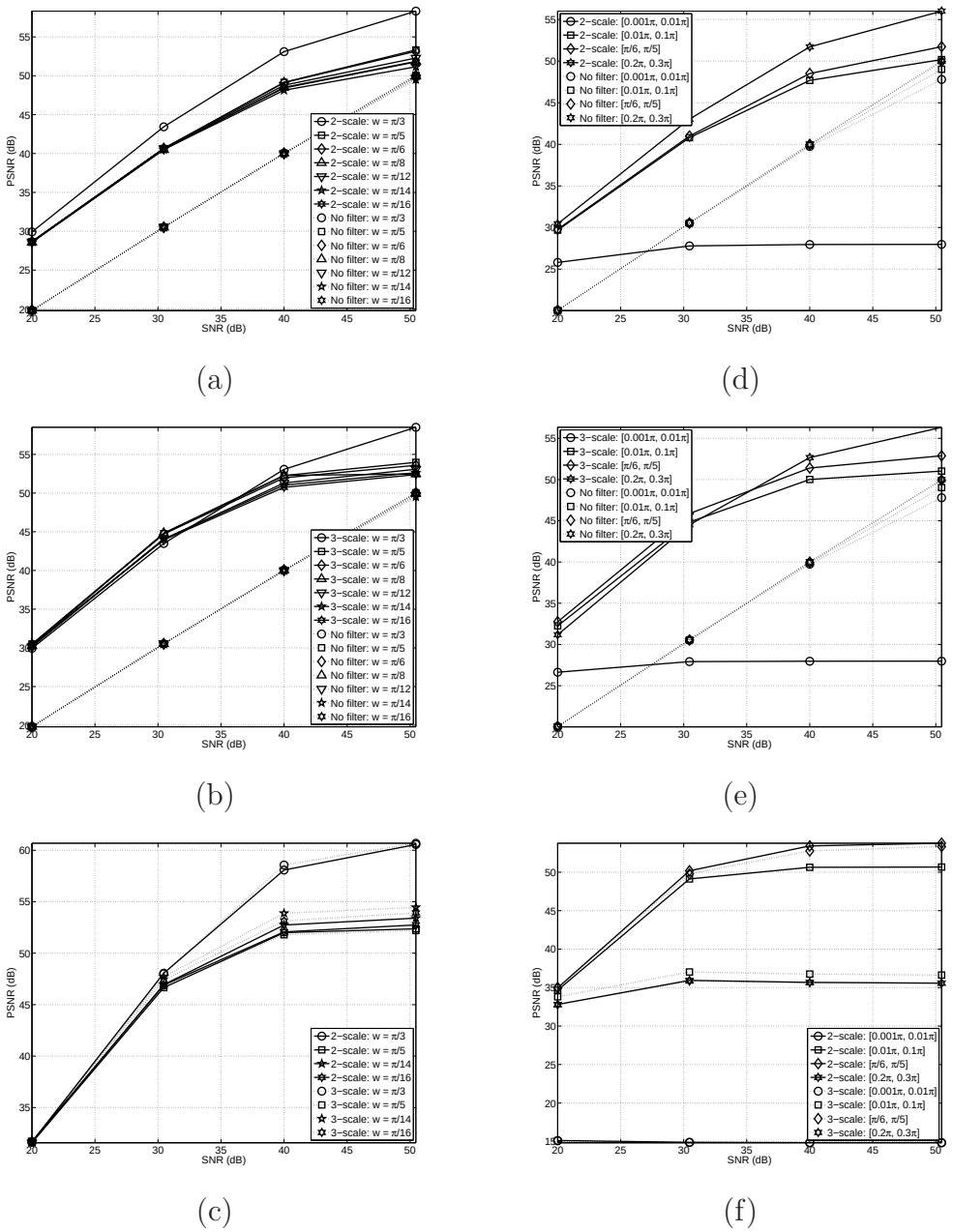


Figure 3.7: PSNR for IA estimation for all the methods used. Improvements in the estimation due to the filterbanks used. (a), (b), (c) Gaussian amplitude modulated signals. (d), (e), (f) Gaussian amplitude-modulated frequency-modulated. (a), (d) QEA when no filterbank is used (dotted line) versus using a two-scale filterbank (solid line). (b), (e) QEA when no filterbank is used (dotted line) versus using a three-scale filterbank (solid line). (c), (f) QLM method using a two-scale filterbank (solid line) versus using a three-scale filterbank (dotted line).

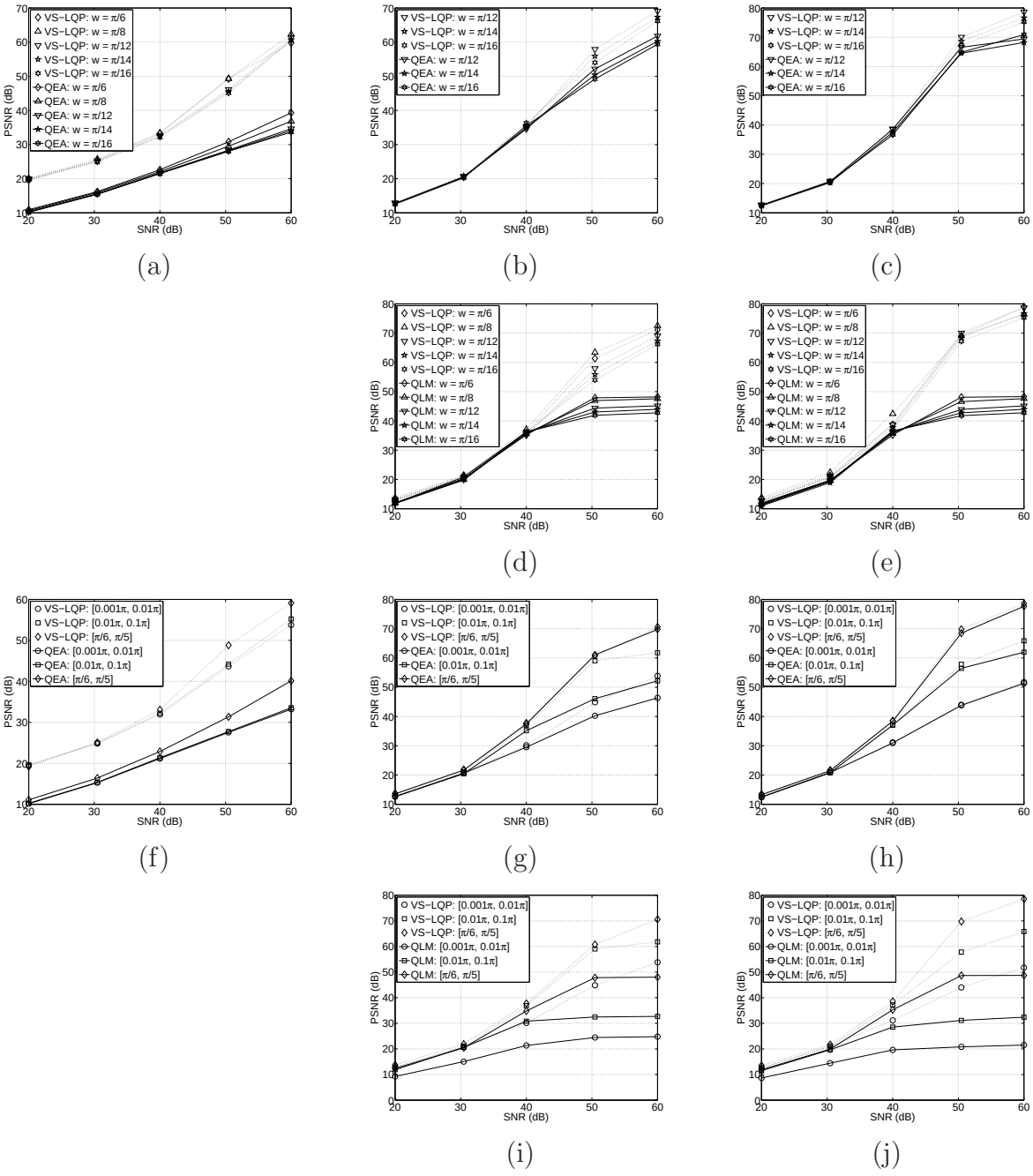


Figure 3.8: PSNR for IF estimation. Results computed using: (i) no filterbank (first column), (ii) the two-scale filterbank (second column) and (iii) the three-scale filterbank (last column). We only show the results in the x direction since results in the y direction were very similar. (a)-(e) Gaussian amplitude modulated signals. (f)-(j) Gaussian amplitude-modulated frequency-modulated. (a)-(c), (f)-(h) VS-LQP versus QEA. (d), (e), (i), (j) VS-LQP versus QLM.

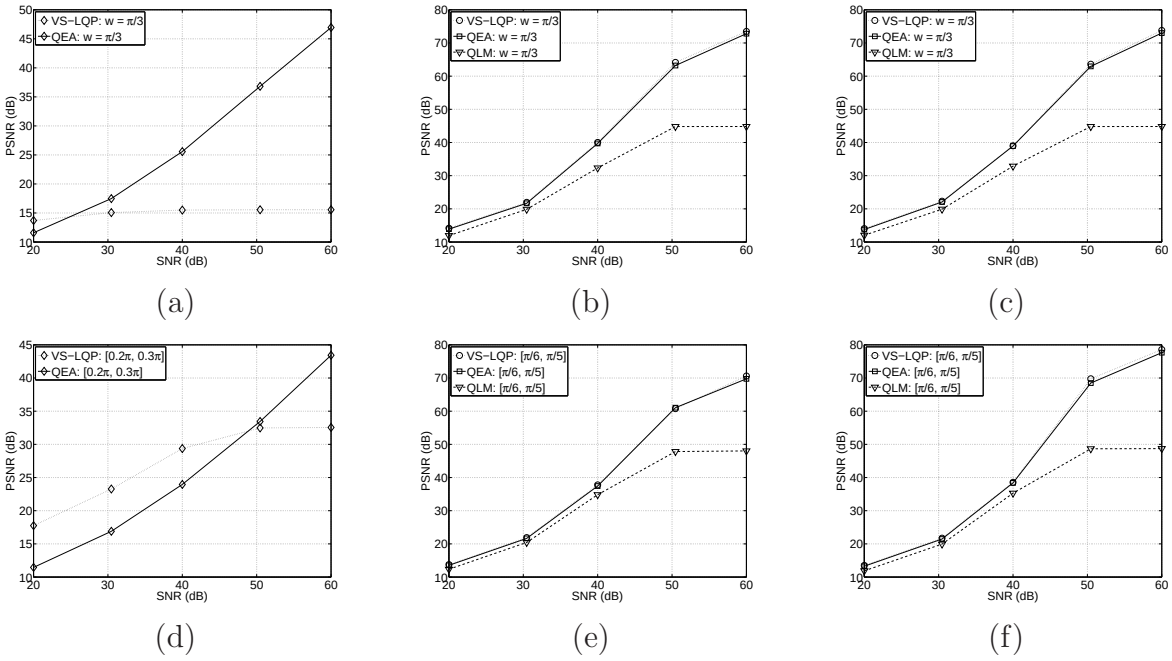


Figure 3.9: PSNR for IF estimation: worst case. Results computed using: (i) no filterbank (first column), (ii) the two-scale filterbank (second column) and (iii) the three-scale filterbank (last column). We only show the results in the x direction. Gaussian amplitude modulated signals in the first row and Gaussian amplitude-modulated frequency-modulated in the last one.

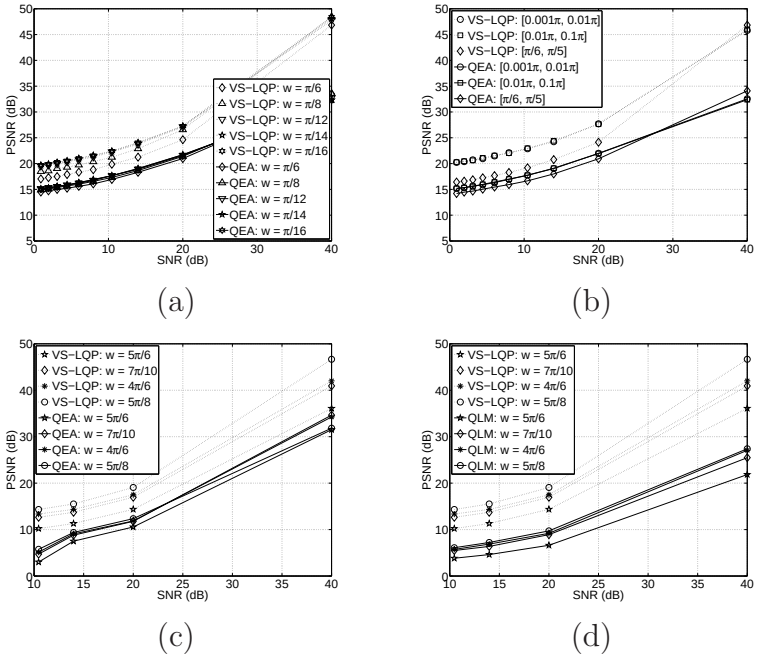


Figure 3.10: PSNR for IF estimation. VS-LQP versus QEA using pre- and post-filtering using median filters without using filterbanks in (a) and (b). Medium-low frequency signals used. (a) Gaussian amplitude modulated results. (b) Gaussian amplitude-modulated frequency-modulated results. Modulation technique applied to high frequency ($\varphi_x = \varphi_y = \{5\pi/6, 7\pi/10, 4\pi/6, 5\pi/8\}$) input signals in (c) and (d) for Gaussian amplitude modulated signals. (c) VS-LQP versus QEA using a three-scale filterbank. (d) VS-LQP versus QLM using a three-scale filterbank.

Chapter 4

Image Reconstructions using AM-FM

In this chapter, we discuss how to reconstruct an image using its AM-FM components. We consider three different approaches: using AM-FM harmonics (section 4.1), using AM-FM components extracted from different scales (section 4.3) and a combined approach (section 4.2).

For our reconstructions we use least squares methods derived from the proposed multi-scale decomposition that is computed using QEA (only method that estimates both the phase and amplitude).

4.1 Least-Squares Reconstructions using AM-FM harmonics (LESHA)

We consider reconstructing an image using its AM-FM harmonics (see [31]):

$$\hat{I}(k_1, k_2) \approx d + \sum_{n=1}^{h} c_n a(k_1, k_2) \cos\left(n\varphi(k_1, k_2)\right), \tag{4.1}$$

In (4.1), we assume that the instantaneous amplitude $a(k_1, k_2)$ and the instantaneous phase $\varphi(k_1, k_2)$ have been computed using a single dominant component estimated from all scales (see Fig. 1.1).

We then want to compute the AM-FM harmonic coefficients c_n, $n = 1, 2, \ldots, h$, so that $\hat{I}(k_1, k_2)$ is a least-squares estimate of $I(k_1, k_2)$ over the space of the AM-FM harmonics. We compute c_n using:

$$\begin{bmatrix} d \\ c_1 \\ \vdots \\ c_M \end{bmatrix} = \left(A^T A\right)^{-1} \left(A^T b\right), \tag{4.2}$$

where the columns of A contain the basis functions. Thus, the first column of A is filled with 1's, while the i^{th} column is filled with the values of the $(i-1)^{th}$ AM-FM harmonic, and b is a column vector of the input image. We also compute an orthonormal basis over the space of the AM-FM harmonics using the Modified Gram-Schmidt (MGS) Algorithm [74] (see Fig. 4.1).

Modified Gram-Schmidt (MGS) Algorithm:
Given the matrix 'A', with the input vectors in its columns, such that the first column of A is filled with 1's, while the i^{th} column is filled with the values of the $(i-1)^{th}$ AM-FM harmonic, compute the number of columns and rows of A:
rows, columns $\leftarrow$ A
for i = 0 to columns - 1
    $q_i = A_{(:,i)}$
    for j = 0 to i - 2
        $r_{j,i} = q_j^T \cdot q_i$
        (In classical Gram-Schmidt algorithm: $r_{j,i} = q_j^T \cdot A_{(:,i)}$)
        $q_i = q_i$ - $r_{j,i} \cdot q_j$
    end
    $r_{i,i} = \| q_i \|_2$
    if $r_{i,i} \leq$ tolerance
        quit
    end if
    $q_i = q_i / r_{i,i}$
end for

Figure 4.1: Modified Gram-Schmidt (MGS) Algorithm for computing an orthonormal basis over the space of the AM-FM harmonics.

4.2 Least-Squares Reconstructions using AM-FM harmonics and the LPF (LESCA)

We extend LESHA by simply adding the low-pass filter (LPF) output to the DC and the AM-FM harmonics. We have:

$$\hat{I}(k_1, k_2) \approx d + c_0 G(k_1, k_2) + \sum_{n=1}^{h} c_h a(k_1, k_2) \cos\left(n\varphi(k_1, k_2)\right), \tag{4.3}$$

where h is the number of AM-FM harmonics and:

d:	denotes a constant DC image,
$G(k_1, k_2)$:	denotes the LPF output,

$a\cos\varphi$:	denotes the dominant AM-FM component estimated across scales and
$a\cos h\varphi$:	denotes the h^{th} AM-FM harmonic.

4.3 Multi-scale least-squares reconstructions (MULTILES)

The third method uses AM-FM estimates extracted from different scales. We describe the correspondence between scales and bandpass filters in Table 4.1 (see Fig. 3.1 (a)-(c) also). For QLM methods, filters 5, 6, 7, 11, 12, 13, 17, 18 and 19 are not applicable for the IP estimation (see sub-section 2.1.4). Define:

d:	Global DC image estimate,
$G(k_1, k_2)$:	low-pass filter output,
$a_1\cos\varphi_1$:	high-frequency scale AM-FM component,
$a_2\cos\varphi_2$:	medium-frequency scale AM-FM component and
$a_3\cos\varphi_3$:	low-frequency scale AM-FM component.

In this case, we consider least squares reconstructions given by:

$$\hat{I}(k_1, k_2) \approx d + c_0 G(k_1, k_2) + \sum_{n=1}^{s} c_n a_n(k_1, k_2) \cos\left(\varphi_n(k_1, k_2)\right), \tag{4.4}$$

where s is the number of scales used.

We then compute the AM-FM multi-scale coefficients c_n, $n = 0, 1, \ldots, s$, so that $\hat{I}(k_1, k_2)$ is a least-squares estimate of $I(k_1, k_2)$. We also compute an orthonormal basis over the space of the AM-FM estimations scale by scale (see previous section).

Table 4.1: Bandpass filters corresponding to different image scales.

Scale	**Single-scale**	**Two-scale**	**Three-scale**
LPF	1	1	1
High frequencies	2, 3, 4, 5, 6, 7	2, 3, 4, 5, 6, 7	2, 3, 4, 5, 6, 7
Medium frequencies	NA*	8, 9, 10, 11, 12, 13	8, 9, 10, 11, 12, 13
Low frequencies	NA*	NA*	14, 15, 16, 17, 18, 19

NA* = Not Applicable.

4.4 Results for Image Reconstructions

In this section, we present results for real images from [75]: 38 aerial images, 64 texture images and 44 miscellaneous images (including the standards *Lena* and *Mandrill*). The images were either 256x256, 512x512 or 1024x1024 pixels. We present results based on both the MSE and the universal image quality index [76].

Tables 4.2, 4.3 and 4.4 summarize results for MSE for LESHA, MULTILES and LESCA (sections 4.1, 4.3 and 4.2, respectively).

The image quality index values (Q) are given in Tables 4.5, 4.6 and 4.7. Figs. 4.2 and 4.3 shows examples of the reconstructions for *Lena* and *Mandrill* images using QEA. In Fig. 4.4, we show the best and the worst reconstructions in terms of Q for both Lena and Mandrill, we use a two-scale and a three-scale filterbank.

In Figs. 4.5 and 4.6 we present an analysis of the FM results for *Lena* and a fingerprint, respectively. The IF vectors are shown. The analysis is performed without the lowpass information using a two-scale and a three-scale filterbank. Note that we do not consider pixels with very low IA values because those could be computational errors. We use the values of the IA greater or equal than the mean for *Lena* and greater or equal than the statistical mode for the fingerprint.

We also apply the FM analysis to images created from digital videos. For this,

Table 4.2: MSE using LESHA (section 4.1). MSE values less than 150 are shown in **bold typeface**.

	Single-scale		***Two-scale***		***Three-scale***	
Harmonics	1	5	1	5	1	5
Aerial	**60.4431**	**60.3892**	**133.6635**	**133.4294**	188.8712	187.9637
Miscellaneous	**114.0075**	**103.2501**	306.3992	291.3852	421.8333	394.3852
Textures	220.9811	215.8769	702.4578	670.5156	1107.4	1028.9

Table 4.3: MSE using MULTILES (section 4.3, see Table 4.1 also). LPF denotes the low-pass filter reconstruction. L denotes the low-frequency scale AM-FM component. Similarly, M and H denote the medium- and high-frequency scale AM-FM component. MSE values less than 150 are shown in **bold typeface**.

	Single-scale		***Two-scale***		***Three-scale***	
Scales	LPF	LPF+H	LPF	LPF+M+H	LPF	LPF+L+M+H
Aerial	**60.3962**	**36.8965**	**133.4740**	**79.7860**	187.5334	**76.0427**
Miscellaneous	153.4645	**63.0507**	298.2355	**129.1126**	406.8230	**111.8693**
Textures	210.4859	**129.7632**	651.1088	393.9337	1020.0157	356.3910

we generate 2D images by sampling a line of pixels through time (see Fig. 4.7). The standard videos of *Akiyo* and *Foreman* were used. Fig. 4.8 shows frames from these videos. For *Akiyo*, we sampled column 79 through time. For *Foreman*, we sampled column 85 through time, respectively. Fig. 4.9 shows the results for *Akiyo* and *Foreman*, in (a)-(d) and (e)-(h), respectively.

4.5 Discussion on Image Reconstructions

As expected, from Tables 4.2, 4.3, 4.4, 4.5, 4.6 and 4.7, as we increase the number of scales for the filterbank, the MSE is increased and Q is decreased. This is due to the fact that each additional scale removes lowpass information captured in the previous

Table 4.4: MSE using LESCA (section 4.2). MSE values less than 150 are shown in **bold typeface**.

	Single-scale		***Two-scale***		***Three-scale***	
	LPF	LPF+5h*	LPF	LPF+5h*	LPF	LPF+5h*
Aerial	**60.3962**	**36.8759**	**133.4740**	**99.8574**	187.5334	**129.3556**
Miscellaneous	153.4645	**52.7000**	298.2355	141.8296	406.8230	189.8731
Textures	210.4859	129.4057	651.1088	500.3847	1020.0157	621.6269

LPF+5h* = LPF + 5 AM-FM harmonics.

Table 4.5: Image Quality Index using LESHA (section 4.1). Q-values above 0.75 are shown in **bold typeface**.

	Single-scale		***Two-scale***		***Three-scale***	
Harmonics	1	5	1	5	1	5
Aerial	**0.7523**	**0.7523**	0.5224	0.5225	0.3255	0.3258
Miscellaneous	**0.7664**	**0.7637**	0.5775	0.5764	0.4126	0.4129
Textures	**0.8469**	**0.8469**	0.6421	0.6474	0.4063	0.4239

scale. In other words, a single-scale filterbank uses a larger part of the frequency spectrum than is used for the two-scale filterbank (and so forth).

Note also that for comparing methods, as more AM-FM harmonics and/or components are used, both MSE and Q are improved. As an example, for LESHA

Table 4.6: Image Quality Index using MULTILES (section 4.3), see Table 4.1 also). LPF denotes the low-pass filter reconstruction. L denotes the low-frequency scale AM-FM component. Similarly, M and H denote the medium- and high-frequency scale AM-FM component. Q-values above 0.75 are shown in **bold typeface**.

	Single-scale		***Two-scale***		***Three-scale***	
Scales	LPF	LPF+H	LPF	LPF+M+H	LPF	LPF+L+M+H
Aerial	**0.752**	**0.866**	0.522	**0.775**	0.325	**0.779**
Miscellaneous	**0.763**	**0.837**	0.579	0.747	0.416	0.744
Textures	**0.848**	**0.902**	0.650	**0.804**	0.425	**0.814**

Table 4.7: Image Quality Index using LESCA (section 4.2). Q-values above 0.75 are shown in **bold typeface**.

	Single-scale		***Two-scale***		***Three-scale***	
	LPF	LPF+5h*	LPF	LPF+5h*	LPF	LPF+5h*
Aerial	**0.7523**	**0.8667**	0.5225	0.7028	0.3259	0.6110
Miscellaneous	**0.7630**	**0.8355**	0.5791	0.6889	0.4161	0.5820
Textures	**0.8489**	**0.9013**	0.6504	0.7459	0.4255	0.6683

LPF+5h* = LPF + 5 AM-FM harmonics.

Table 4.8: MSE and Image Quality Index (Q) for the reconstructions shown in Figs. 4.2 and 4.3.

		Lena		***Mandrill***	
Filterbank	***Method***	**MSE**	**Q**	**MSE**	**Q**
Single-scale	LESHA	15.5002	0.8336	221.4102	0.7880
	MULTILES	10.2028	0.9000	119.0642	0.8917
	LESCA	10.1955	0.9001	119.0505	0.8917
Two-scale	LESHA	71.4186	0.6721	439.1018	0.5273
	MULTILES	52.7913	0.8142	220.5814	0.8004
	LESCA	63.3575	0.7443	279.9986	0.7338
Three-scale	LESHA	121.4774	0.5299	563.9774	0.3224
	MULTILES	48.8083	0.8058	198.5808	0.8104
	LESCA	90.7965	0.6263	328.6932	0.6612

(section 4.1), we can see that the more AM-FM harmonics we use, the higher Q we get and the lower MSE. MULTILES produces the best results, whereas LESHA produces the worst results.

The use of multiple scales produced the best results as shown in Figs. 4.2 and 4.3, where we present examples of the reconstructions for *Lena* and *Mandrill* images. These two examples visually demonstrate that even if the reconstruction looks very good, the difference between the MSE of *Lena* and *Mandrill* can be very big. However, reflecting good perceptual agreement, Q is really close between them in almost

all of the cases, except for the LESHA method. The reconstructed images also show that the bigger the number of scales that the filterbank has, the bigger the MSE and the lower the value of Q is obtained. This can be seen in the resolution of the images.

We present the best and the worst reconstructions in Fig. 4.4. In the case of *Lena*, we can see artifacts in her shoulder for the best reconstruction. The best reconstruction of *Mandrill* has problems in the lower zone of the beard. For both images, the worst reconstructions look like a smoothed version of the original.

When the FM and the IF information is analyzed, we can see from Figs. 4.5 and 4.6 how AM-FM captures the fast changes from one pixel to another. Figs. 4.5 (f)-(g) show clearly how the frequency component adapt to *Lena*'s hair. In Fig. 4.6 we can see this adaptation more clear. Similar results we can see when images from digital videos are used (see Fig. 4.9).

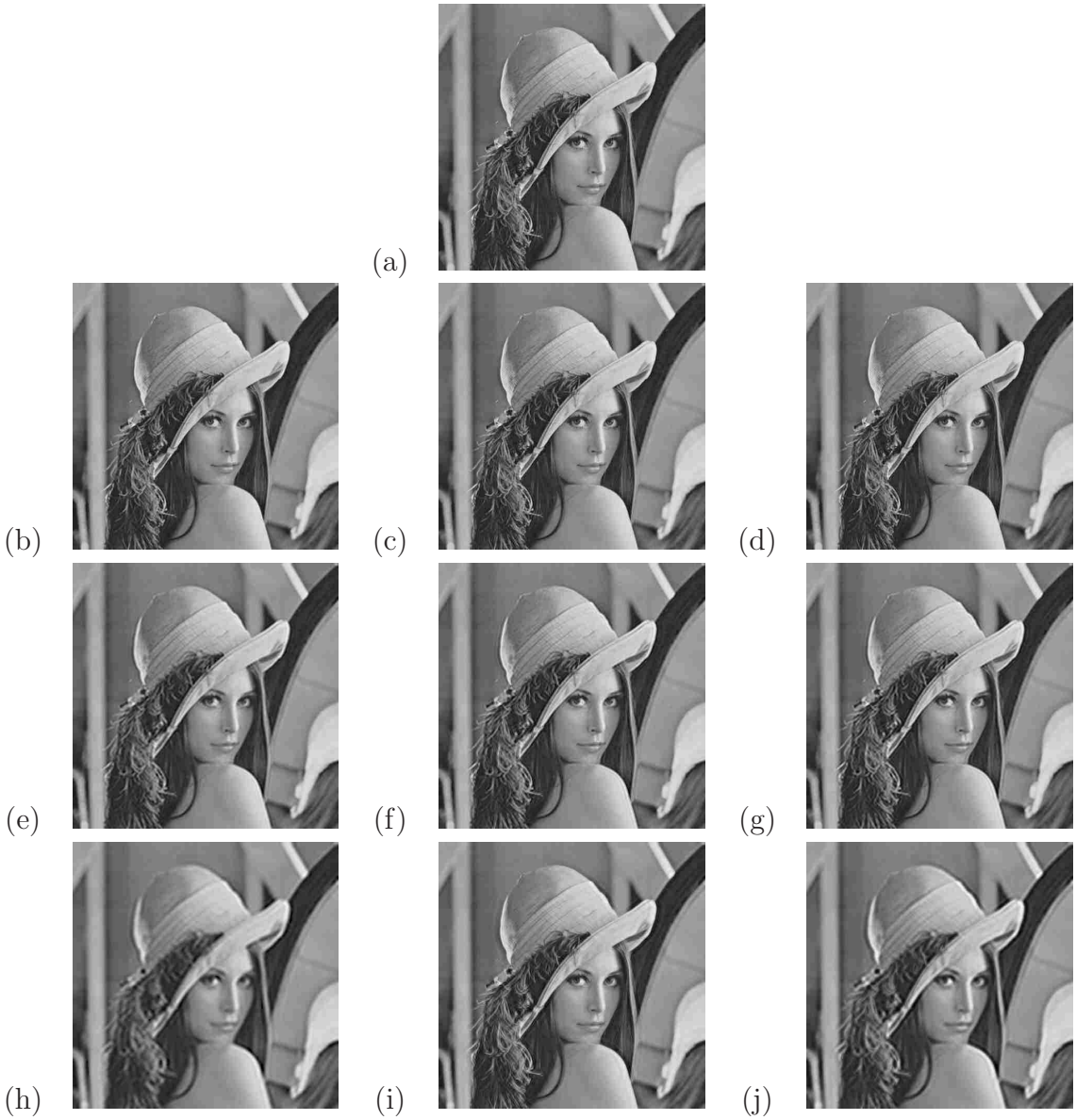


Figure 4.2: Multi-scale AM-FM reconstructions for the *Lena* image. (a) Original image 512x512 pixels. Single-scale filterbank results for: (b) LESHA, (c) MULTILES and (d) LESCA. Two-scale filterbank results for: (e) LESHA, (f) MULTILES and (g) LESCA. Three-scale filterbank results for: (h) LESHA, (i) MULTILES and (j) LESCA.

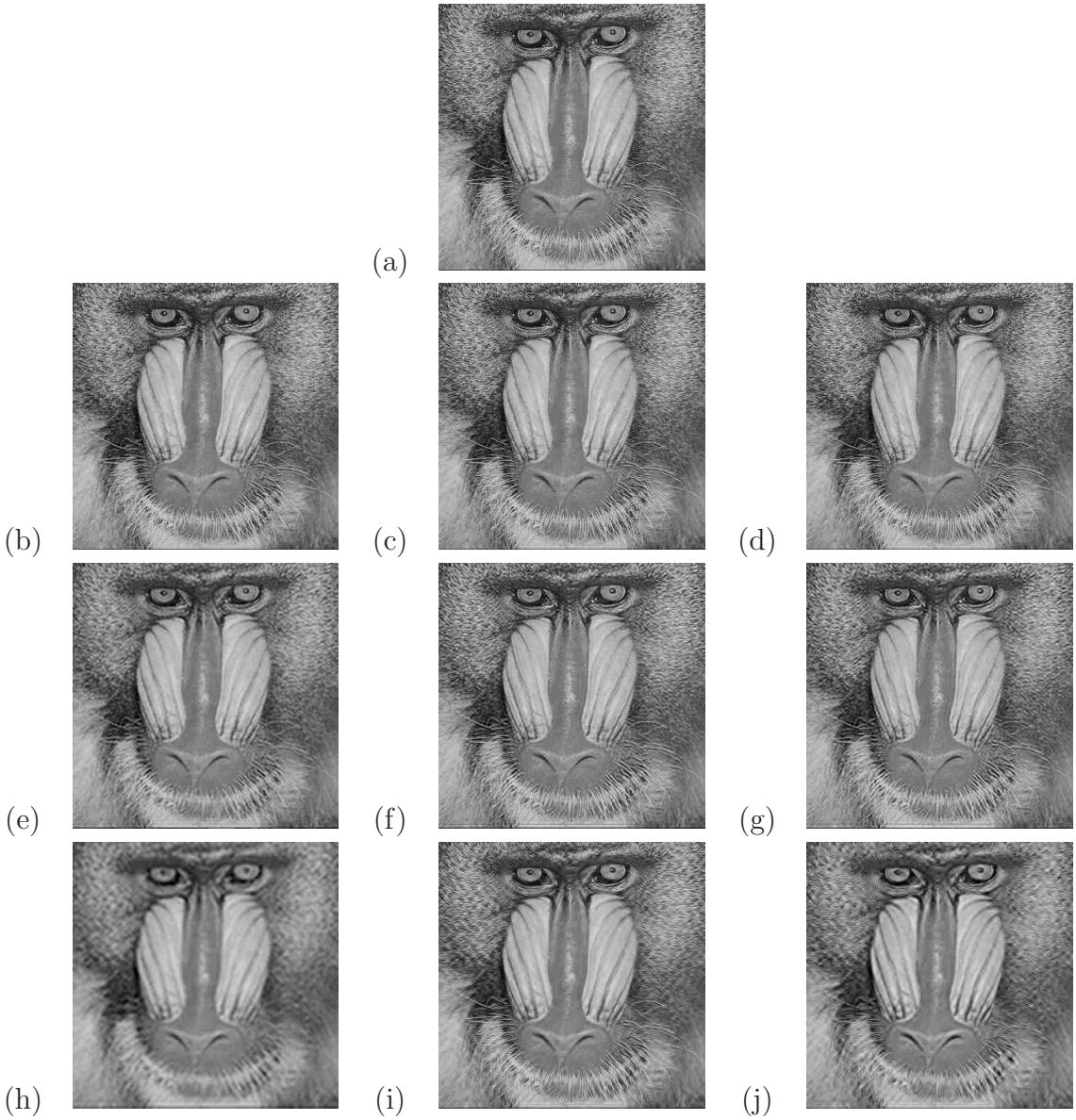


Figure 4.3: Multi-scale AM-FM reconstructions for the *Mandrill* image. (a) Original image 512x512 pixels. Single-scale filterbank results for: (b) LESHA, (c) MULTILES and (d) LESCA. Two-scale filterbank results for: (e) LESHA, (f) MULTILES and (g) LESCA. Three-scale filterbank results for: (h) LESHA, (i) MULTILES and (j) LESCA.

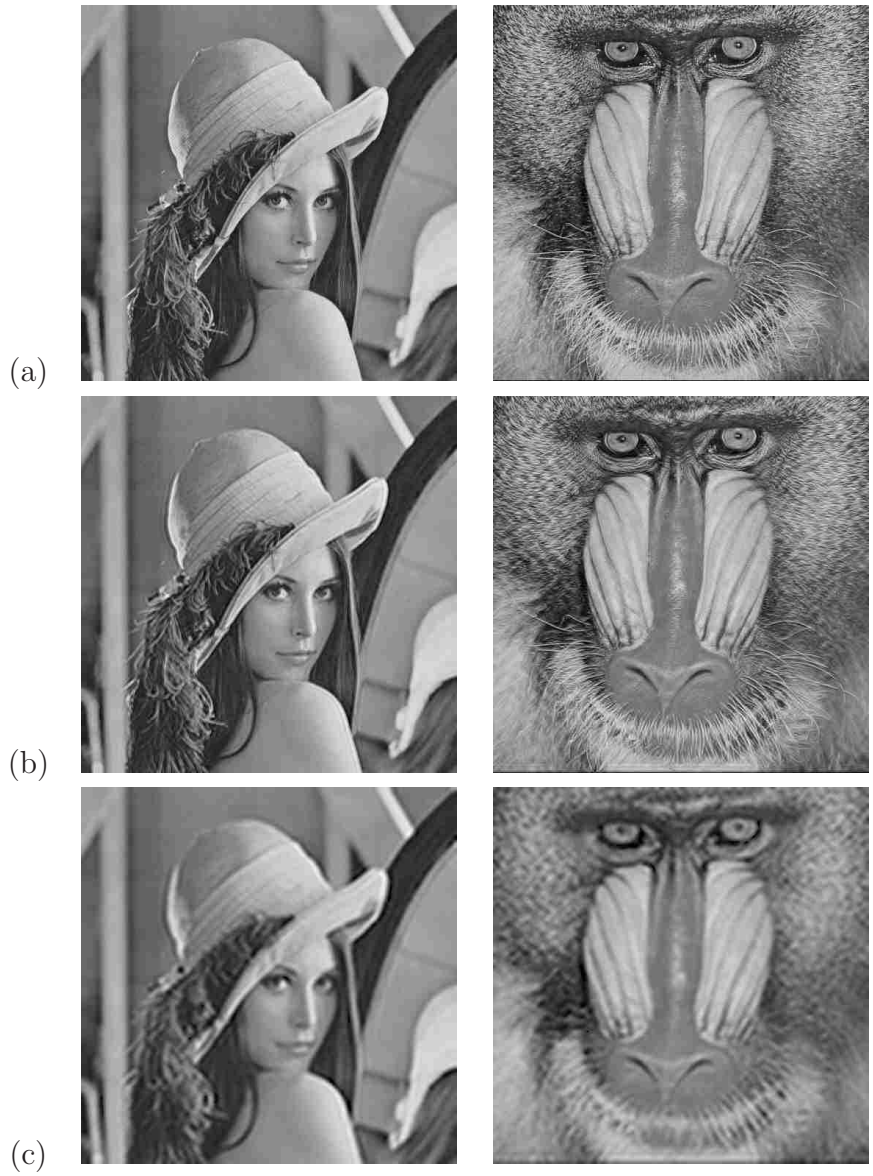


Figure 4.4: Best and worst *Lena* and *Mandrill* reconstructions between a two-scale filterbank and a three-scale filterbank based on the image quality index from Table 4.8. (a) Original *Lena* image 512x512 pixels. (b) MULTILES two-scale filterbank: $Q = 0.8142$ ($MSE = 52.7913$). (c) LESHA three-scale filterbank: $Q = 0.5299$ ($MSE = 121.4774$). (d) Original *Mandrill* image 512x512 pixels. (e) MULTILES three-scale filterbank: $Q = 0.8104$ ($MSE = 198.5808$). (f) LESHA three-scale filterbank: $Q = 0.3224$ ($MSE = 563.9774$).

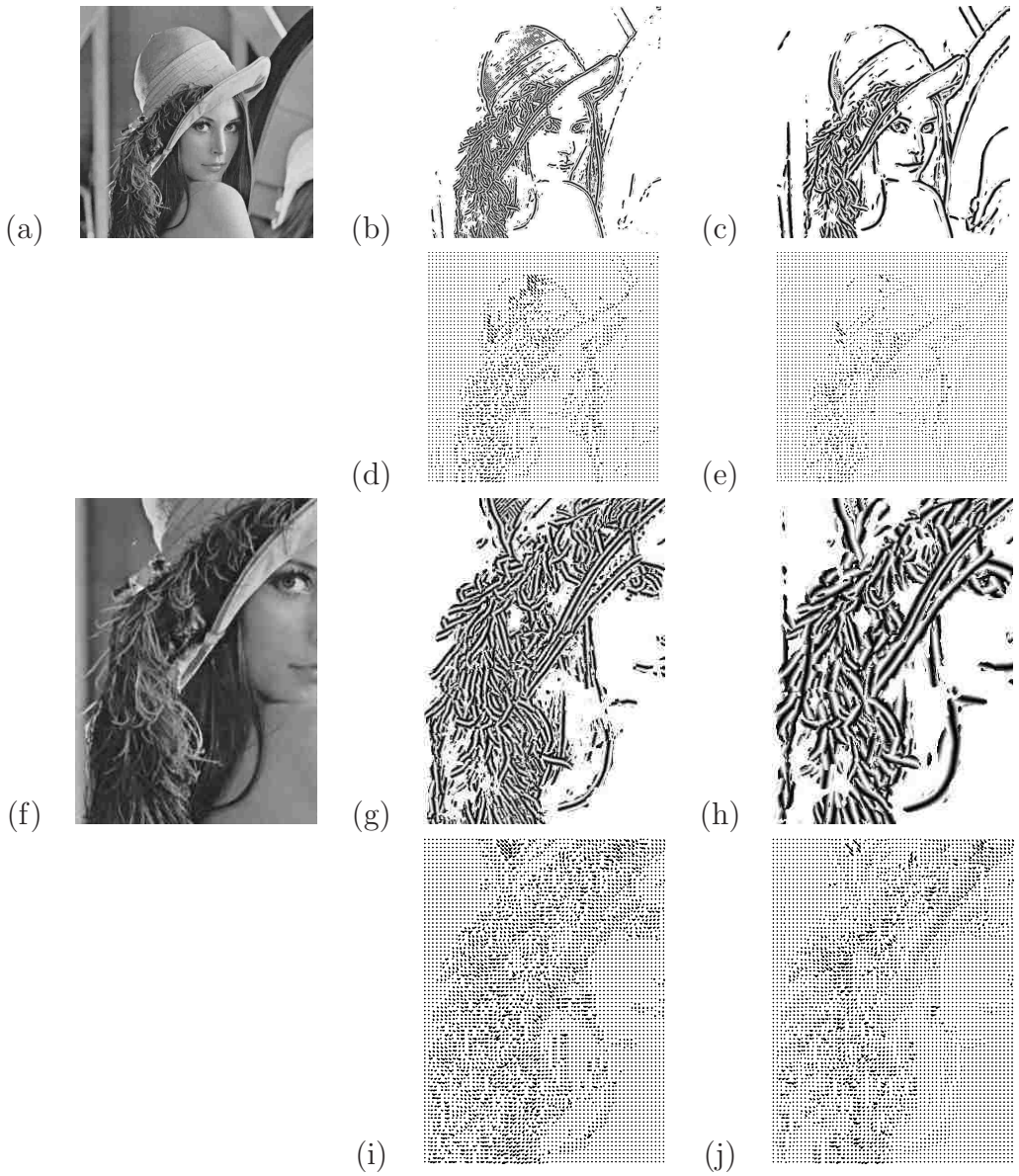


Figure 4.5: High amplitude FM Analysis using QEA method for *Lena*. For (b), (c), (g) and (h) we only show the FM component pixels for which the instantaneous amplitude is above the IA mean. (a) Original image. FM analysis in (b) using a two-scale filterbank and in (c) using a three-scale filterbank. IF using a two-scale filterbank in (d) and using a three-scale filterbank in (e). Zoom on *Lena*'s hair in (f). FM analysis in (g) using a two-scale filterbank and in (h) using a three-scale filterbank. IF using a two-scale filterbank in (i) and using a three-scale filterbank in (j).

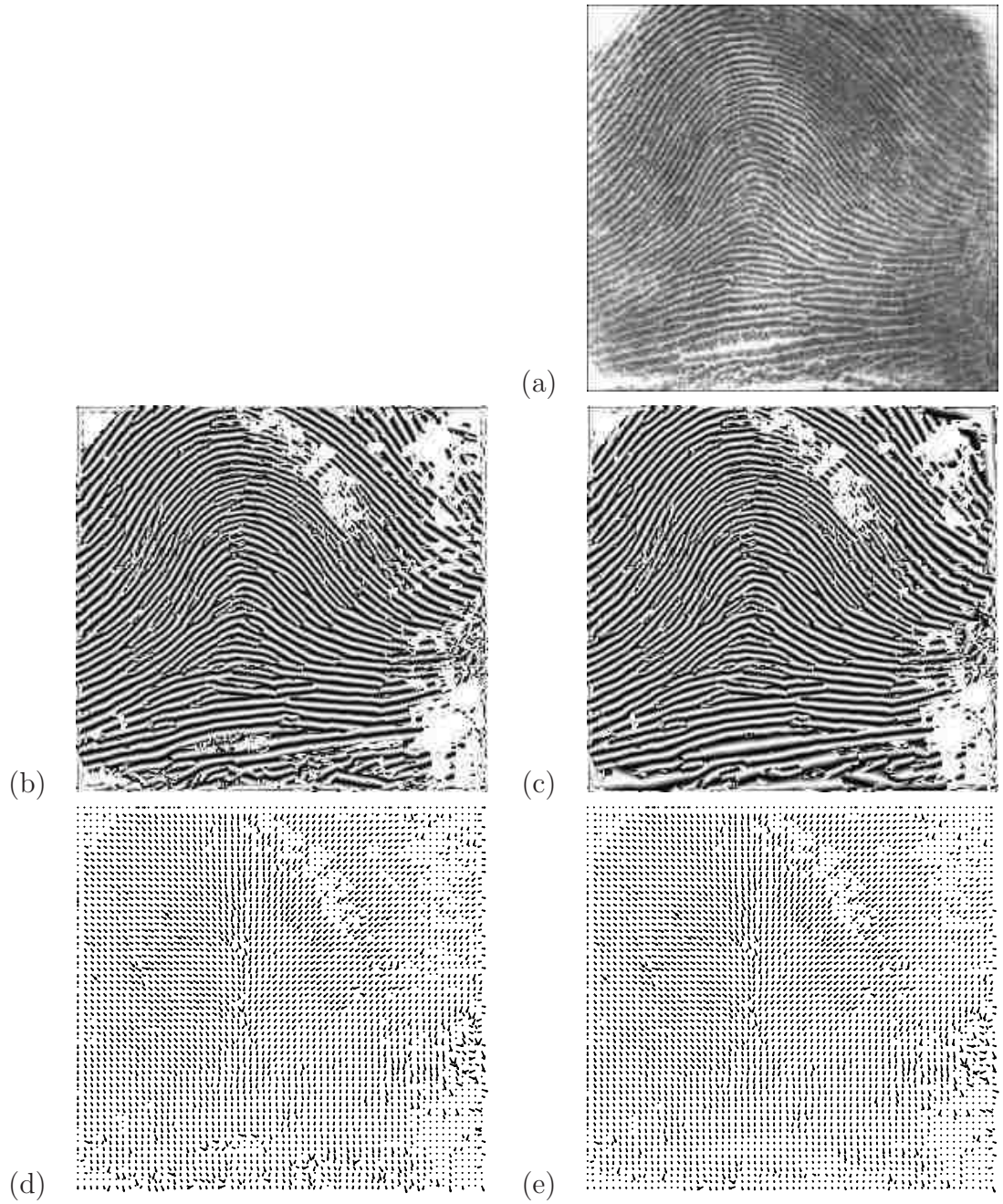


Figure 4.6: High amplitude FM Analysis using QEA method for a fingerprint image. Here, we show results for which the instantaneous amplitude is greater than the IA statistical mode. (a) Original image. (b) FM using a two-scale filterbank. (c) FM using a three-scale filterbank. (d) IF using a two-scale filterbank. (e) IF using a three-scale filterbank.

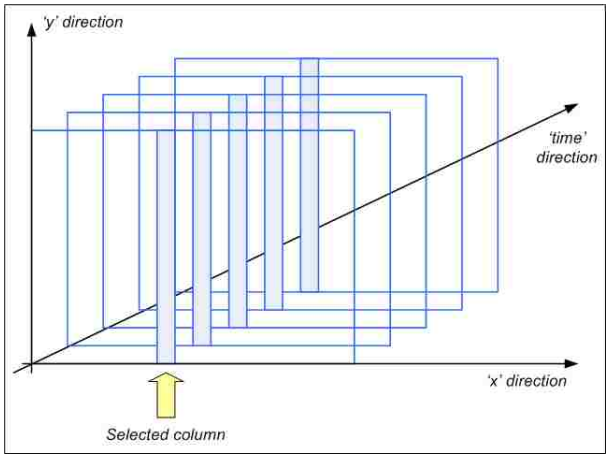


Figure 4.7: Method to produce an image from a fixed column in a real video.

(a)  (b) 

Figure 4.8: Frames of videos with a column clearly highlighted. (a) Frame of *Akiyo* video with column 79 clearly highlighted. (b) Frame of *Foreman* video with column 85 clearly highlighted.

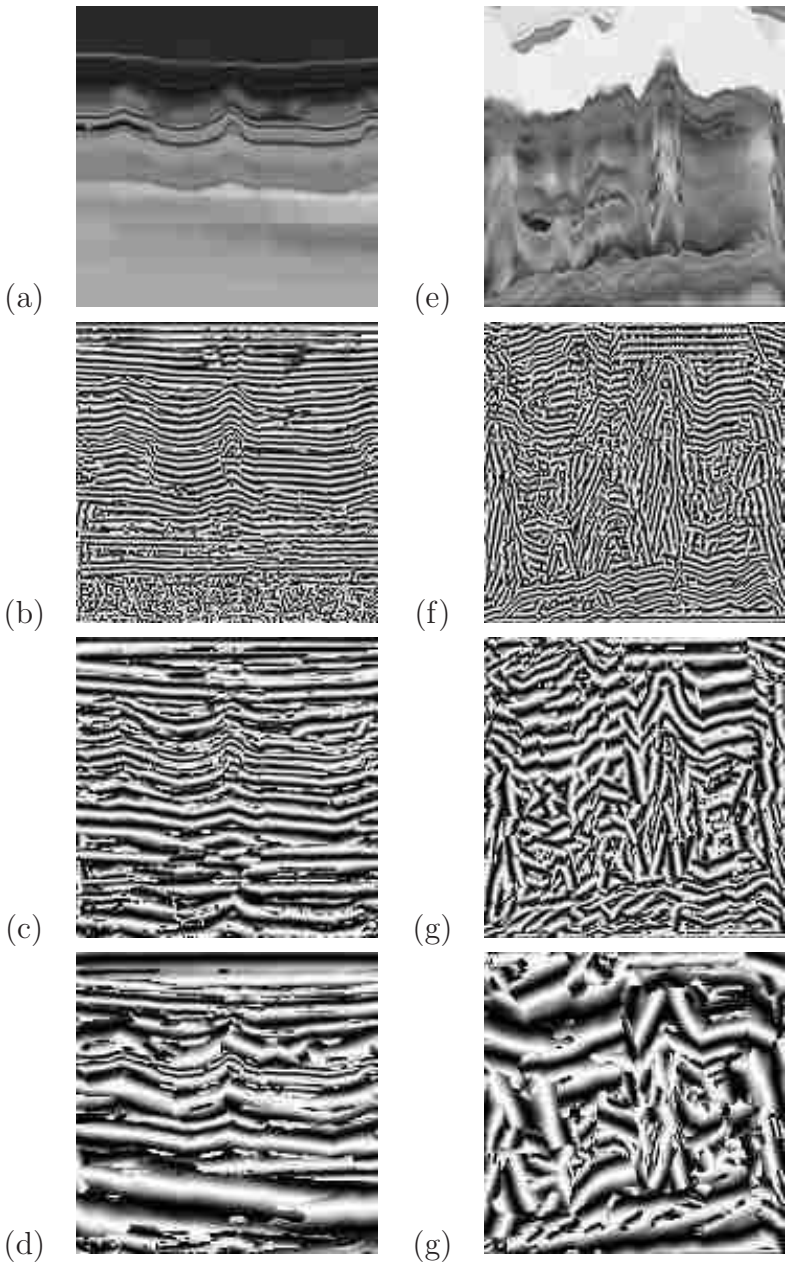


Figure 4.9: FM Analysis for *Akiyo* and *Foreman*. Single-scale filterbank, two-scale filterbank and three-scale filterbank in rows two, three and four, respectively. Original Images: (a) *Akiyo* and (e) *Foreman*. *Akiyo*'s results in column one. *Foreman*'s results in column two.

Chapter 5

Retinal Image Analysis

In this chapter, we present results for retinal image analysis using methods AM-FM methods. This work was motivated, proposed and supported by VisionQuest Biomedical. For this work, we established a collaboration that includes ophthalmologists and technologists from the University of Iowa and the Texas Retina Institute of South Texas in San Antonio, TX.

We present an introduction in section 5.1. Next, we present a background in retinal image analysis in section 5.2. We present our method in section 5.3. Finally, the results and discussion are described in the last two sections. The results and procedures described in the next sections are based on the reports presented to VisionQuest Biomedical.

5.1 Introduction

We start this sub-section defining terms that are used in this section. To assess image classification performance we refer to Fig. 5.1. In Fig. 5.1, we use the term true positive to refer to a case when both the computer and the clinician(s)

classified an image as exhibiting disease. On the other hand, a false positive refers to a normal case (as viewed by the clinician(s)) that was classified as abnormal by the computer. Similarly, true negative refers to the case when they both agree on a normal classification, while a false negative refers to a normal classification by computer of an abnormal case. We define:

- Sensitivity: Fraction of diseased population correctly diagnosed.

$$Sensitivity = \frac{TP}{TP + FN}$$

- Specificity: Fraction of normal population correctly diagnosed.

$$Sensitivity = \frac{TN}{TN + FP}$$

Note that $1 - Specificity = FAF$ is often used, where FAF is the fraction of false alarms.

For the purposes of this application, we sought a distribution of "normal" (controls) images that was much less than is typically found in a random population of diabetics. AM-FM methods were evaluated using the MESSIDOR[1] database [77]. Tables 5.1, 5.2 and 5.3 give a summary of the cases found in the subset (N = 400 images) of the MESSIDOR database. This database contains the closest approximation to the distribution of disease severity proposed for use in this application. For all but one category, this database exceeds the proposed number of 50 cases.

The MESSIDOR fundus image database contains 1200 color digital images of the posterior pole (back of the eye, the retina between the optic disc and the macula) [77]. These data were acquired at three ophthalmologic departments using a digital color CCD camera with a Topcon TRC NW6 non-mydriatic ophthalmoscope. Images

[1] *Méthodes d'Evaluation de Systèmes de Segmentation et d'Indexation Dédies à l'Ophtalmologie Rétinienne*, in English: *Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology.* Kindly provided by the MESSIDOR program partners.

were taken with a 45 degree field of view. The image format is 8 bits per color plane at 1440x960, 2240x1488 and 2304x1536 pixels. 800 images were acquired with pupil dilation (one drop of Tropicamide at 10%) and 400 without dilation [77]. The 1200 images are packaged in three sets, one per ophthalmologic department.

We selected 265 images for our study. Each image has been diagnosed as indicated in the Tables 5.1, 5.2 and 5.3 and an image quality value assigned where 0 is the poorest quality and 5 is the best quality. All of the images contained in the database were used for making actual clinical diagnoses.

In the next sub-section, we present a background about related work for detection and phenotyping of retinal disease. Next we present the methods used in sub-section 5.3. In sub-sections 5.4 and 5.5 we present the results and the discussion, respectively.

	Diseased	Normal
Positive	TP True Positive	FP False Positive
Negative	FN False Negative	TN True Negative

Figure 5.1: Decision Matrix. *Figure by Bert Davis from VisionQuest Biomedical.*

Table 5.1: Retinopathy Grade Distribution.

Grade	**Number of images (percentage)**
3	52 (20%)
2	50 (19%)
1	71 (27%)
0	92 (34%)
Total	265 (100%)

Table 5.2: Image Quality Distribution. Highest grade is 5. Lowest grade is 1.

Grade	**Number of images (percentage)**
5	90 (34%)
4	139 (52%)
3	24 (9%)
2	10 (4%)
1	2 (1%)

Table 5.3: Risk of Macular Edema (ME) Distribution.

Grade	**Number of images (percentage)**
2	37 (14%)
1	15 (6%)
0	213 (80%)

5.2 Background on Related Work

Professor Abramoff, a collaborator at the University of Iowa, has approached the problem using Gabor filters. He evaluated the performance of the system on 10,000 exams from 5692 unique patients with diabetes (2 images from each eye). The images were collected from diabetics as part of an actual (diabetic retinopathy) DR screening program at 10 different clinics with 4 different types of retinal nonmydriatic camera. These diabetic patients had not been previously diagnosed with DR. The performance of the DR screening algorithm had an area under the curve of 0.85 on the first visit, and an optimal sensitivity of 0.84 and specificity of 0.64. These results show the potential of this approach. That the system is relatively stable is illustrated by the fact that the area under the curve for the second visit is 0.84.

At this point, 7689/10000 (77%) of the exams had acceptable image quality,

4648/7689 (60%) were true negatives, 59/7689 (0.8%) false negatives, 319/7689 (4%) true positives, and 2581/7689 (33%) false positives. 27% of false negatives contained large hemorrhages and/or neovascularizations. By improving the image feature classification step, they have been able to perform additional validation on an additional set of 10,000 exams, with an area under the ROC[2] curve of 0.90 (see Table 5.4).

Table 5.4: Results from other research studies. Sensitivity and Specificity are given in percentage (%).

Researcher	**Reference**	**Se**[‡]	**Sp**[§]	**AUC**	**# Samples**
Larson (2002)	[78, 79]	93.1	71.4	–	231
Abramoff & Niemeijer	[80, 81, 82, 83]	90.0	–	–	10000
Lee	[84]	100.0	87.0	0.90	369
Gardner	[85]	88.4	83.5	–	–
AM-FM Study	(unpublished)	99.0	92.0	0.93 to 0.98	265
Abramoff	(unpublished)	–	–	0.90	400

Se[‡] = Sensitivity. **Sp**[§] = Specificity.

5.3 Methods

5.3.1 Filterbank design

Extending the filterbank design of section 3.1, we designed a new four-scale filterbank (see Fig. 5.2). This filterbank uses separable bandpass filters also. Fig. 5.3 shows an example of the frequency response of one 1D bandpass filter.

In Fig. 5.2, filter 1 corresponds to a low pass filter (LPF) with frequency support in $[-\pi/16, \pi/16]$ for both x and y directions. For all the other filters, the bigger the label number of the filter, the lower the frequency support that it has. The filters in the highest frequencies (filters from 2 to 7 in Fig. 5.2), have a bandwidth of $\pi/2$

[2]ROC = Receiver Operating Characteristic.

for both x and y directions. The bandwidth is decreased by a factor of 0.5 for each added scale.

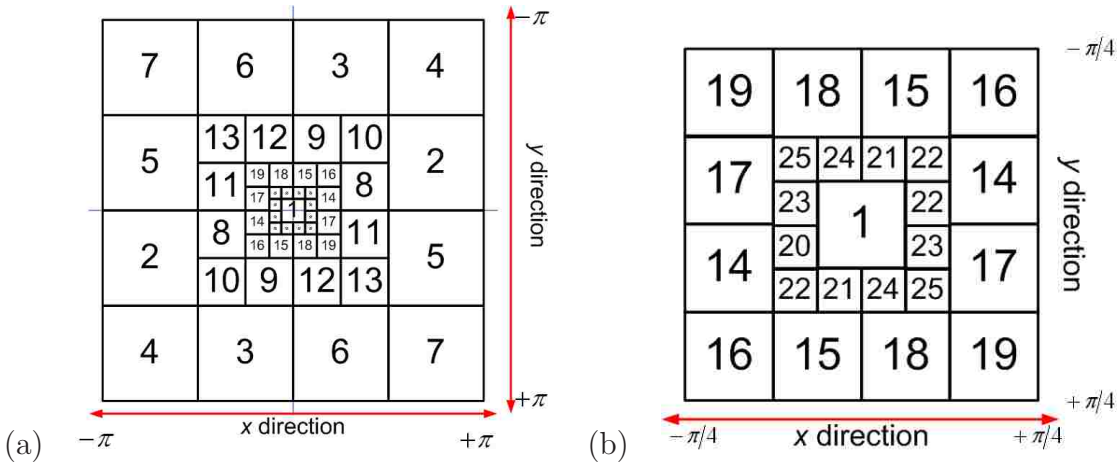


Figure 5.2: Four-scale filterbank used for retinal image analysis. (a) Complete frequency spectrum of the filterbank. (b) Zoom on the low frequency bandpass filters.

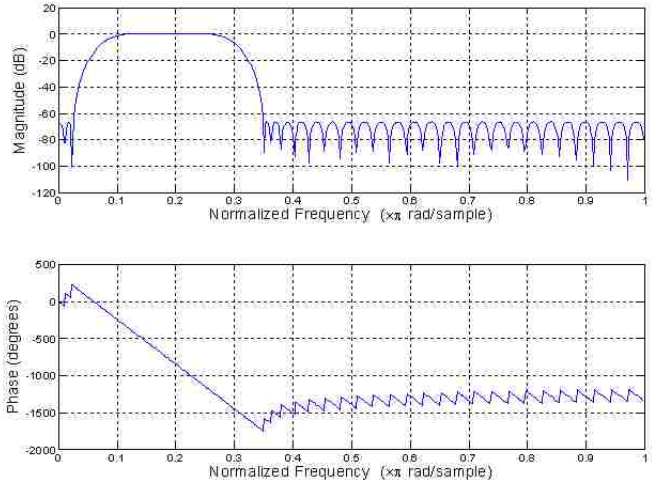


Figure 5.3: Example of the separable design. An 1D filterbank response is shown. The filter corresponds to a bandpass filter with frequency support in $[\pi/8, \pi/4]$.

5.3.2 Multi-scale AM-FM analysis

We use AM-FM components extracted from different scales. We describe the correspondence between the scales and bandpass filters in Table 5.5 (see Fig. 5.2 also). For retinal applications, we consider nine different cases of extracting dominant AM-FM component from different scales (see Table 5.6). From each case i, $i = 1, \ldots, 9$, we use the histograms of both the instantaneous amplitude a_i and the magnitude of the instantaneous frequency ($\nabla\varphi_i$) given by $||\nabla\varphi_i||$ as image features. Note that high-frequency bandpass filters (filters from 2 to 7 in Fig.5.2) are not used for the analysis because the information in there is the high frequency noise of the image. Reconstructions from theses cases are given in Figs. 5.4 and 5.5 for a patient with Risk 0 and for a patient with Risk 1, respectively.

We use histograms of IA and the magnitude of the IF, ||IF||, to create a feature vector for detection of DR. Using histograms at different scales (see Table 5.6) the information extracted with AM-FM can be analyzed to find differences among retinal images with DR and healthy images. A region containing micro-aneurysms, hemorrhages, and exudates will have different estimates for IA than a region lacking these features. Using these histograms, we can find if a certain frequency component that encodes a feature is present at the image.

Both histograms, of a_i and $||\nabla\varphi_i||$, for $i = 1, 2, \ldots$ or 9, are computed using forty bins, leading to one histogram of eighty bins. Histograms are computed for each image at all nine filterbank cases and analyzed separately. Thus, each image has nine histograms, one per filterbank case.

Table 5.5: Bandpass filters used for estimating AM-FM in a four-scale filterbank.

	Scales	**Bandpass filters**
LPF	Low pass filter	1
VL	Very low frequencies	20, 21, 22, 23, 24, 25
L	Low frequencies	14, 15, 16, 17, 18, 19
M	Medium frequencies	8, 9, 10, 11, 12, 13

Table 5.6: Scales used for the nine cases in retinal image analysis.

Case #	**Scales used for AM-FM estimation**	**AM-FM component**
1	VL, L, M	$a_1 \cos \varphi_1$
2	LPF	$a_2 \cos \varphi_2$
3	VL	$a_3 \cos \varphi_3$
4	L	$a_4 \cos \varphi_4$
5	M	$a_5 \cos \varphi_5$
6	LPF, VL, L, M	$a_6 \cos \varphi_6$
7	LPF, VL	$a_7 \cos \varphi_7$
8	VL, L	$a_8 \cos \varphi_8$
9	L, M	$a_9 \cos \varphi_9$

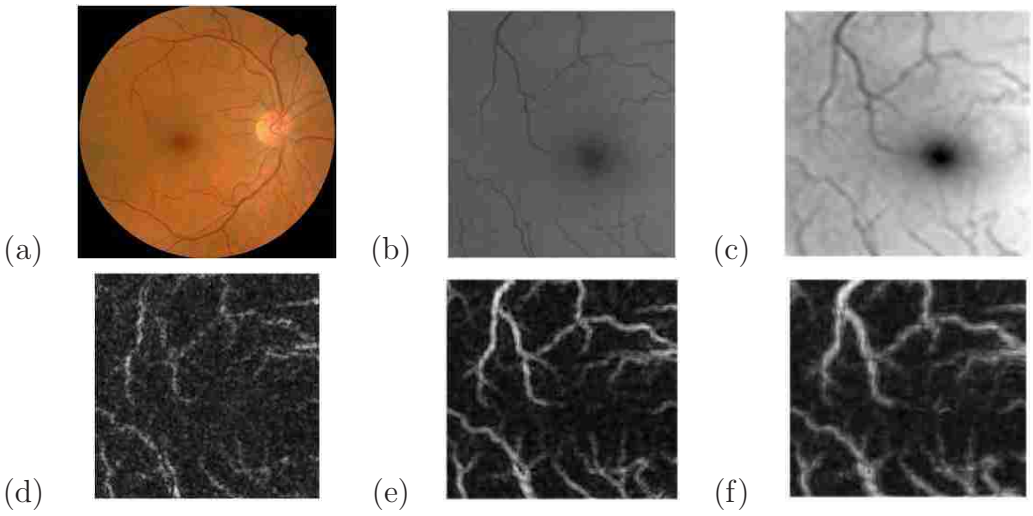


Figure 5.4: Four different scales (see Table 5.6) for analyzing a retinal image that was graded as Risk 0. (a) Original image. (b) Region of Interest (ROI) for green channel only. (c) $a_6 \cos \varphi_6$. (d) $a_5 \cos \varphi_5$. (e) $a_4 \cos \varphi_4$. (f) $a_1 \cos \varphi_1$.

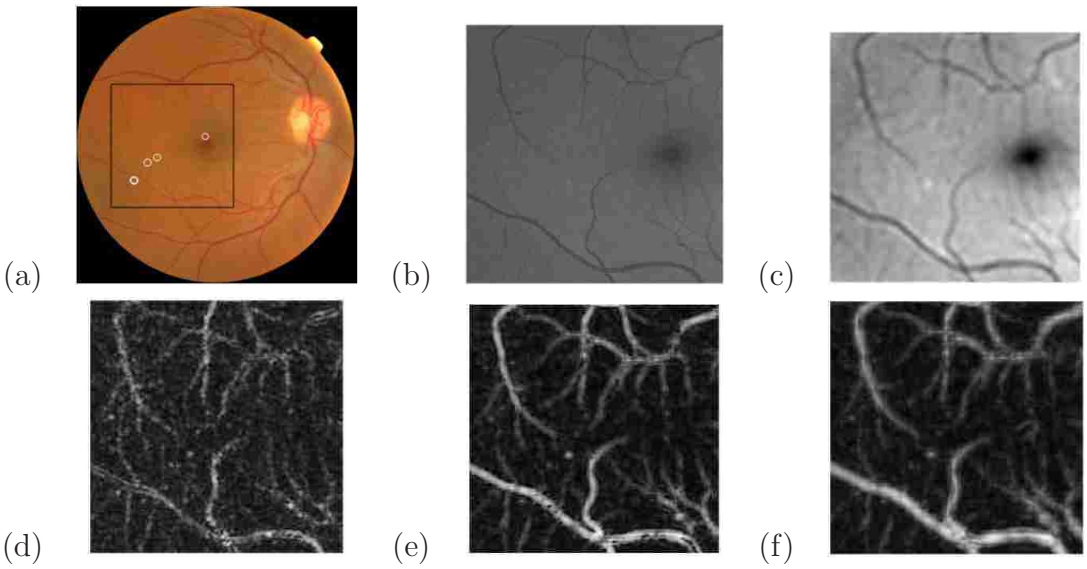


Figure 5.5: Four different cases of scales (see Table 5.6) for analyzing a retinal image that was graded as Risk 1. (a) Original image. (b) Region of Interest (ROI) for green channel only. (c) $a_6 \cos \varphi_6$. (d) $a_5 \cos \varphi_5$. (e) $a_4 \cos \varphi_4$. (f) $a_1 \cos \varphi_1$.

5.3.3 Prediction of DR state based on the AM-FM histogram estimates

The purpose of this sub-section is to develop a predictor of disease state based on the histogram bins counts generated. The dependent variable, disease state, was coded with 0's for normal and 1's for the disease state. The normal images were separately compared with disease Risks 1, 2 and 3. Thus, we considered three cases: (i) 0 versus 1, (ii) 0 versus 2 and (iii) 0 versus 3.

Each of the nine cases for estimating the AM-FM features, see previous sub-section and Table 5.6, has an 80-bin histogram. Then, we create a joined histogram vector containing the histogram of the nine case. Thus, 9 cases $\times$ 80 bins each = 720 bins used for each image. This number is reduced by first solving the detection problem for each filter case, then combining the nine detectors into one. The regression

model is given by

$$Y = X\beta + \varepsilon, \tag{5.1}$$

where Y is the vector of diagnoses such as 0's and 1's for normal and diseased, X is the matrix of independent variables (bin counts of IA and ||IF|| histograms), β is the vector of weights and ε is a vector of random errors. The Gauss-Markov theorem assures that the least-squares estimate of β given by the so called normal equations

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{5.2}$$

is the Best Linear Unbiased Estimator (BLUE) of β and hence the BLUE of Y as given by $\hat{Y} = X\hat{\beta}$.

Several issues arise that make the least-squares estimator undesirable or difficult to compute accurately. One instance is where the number of cases, the number of rows of the X matrix, is less than the number of variables, the number of columns of the X matrix. In this case, an inverse of $X'X$ does not exist. Another common problem is when the columns of the X matrix are highly correlated which is called multi-colinearity. In this case the inverse of $X'X$ may exist, but is numerically unstable and the normal equations cannot be solved accurately. Unfortunately, both circumstances exist in the current data sets. The number of histogram bins is frequently larger than the number of eyes in the study, and adjacent histogram bins are obviously very correlated if the underlying frequency distribution is continuous. A number of methods referred to as shrinkage methods exist that forgo the unbiased criterion and in these cases decrease the variance more than the loss of accuracy due to some added bias.

One shrinkage method that is in common use is principal components regression (PCR). PCR is based on principal components analysis (PCA) which factors the nxp matrix X into two parts

$$X = TL,$$

where the matrix T is an orthogonal matrix of size nxp containing the p principal components and L is a pxp matrix of loading factors. The advantage of PCA is that the principal components are ordered in their ability to reconstruct X. That is X can be approximated parsimoniously by the first columns of T and rows of L. This leads to a reduction of the regression problem called PCR:

$$X \approx T_1 L_1$$
$$Y = X\beta + \varepsilon \approx T_1(L_1\beta) + \varepsilon = T_1\gamma + \varepsilon,$$

where now the number of variables has been reduced from p to the number of columns of T_1 and the multi-colinear columns of X have been replaced by the uncorrelated columns of T_1.

The difficulty of PCR is that in the linear regression application it is the wrong optimization. PCA is based only on the covariance of the columns of X. In the regression problem, it is the covariance of X with Y that is of concern rather than the columns of X with themselves. While the first columns of the T matrix frequently will serendipitously have all of the information on Y contained in X, numerous applications have been incurred where the first columns of T contained only information about X with the information about Y being in the eigenvectors corresponding to the smaller eigenvalues, i.e., the later columns of T [86]. In that case, PCR misses the information X contains about Y completely.

A second shrinkage method very similar to PCR solves the problem discussed in the previous paragraph. The method, called **Partial Least Squares** (**PLS**), factors the X matrix exactly like PCA, only while PCA optimally explains the covariance of the columns of X with themselves, the PLS optimization explains the covariance of the columns of X with Y as parsimoniously as possible. PLS results in a T matrix of so called latent factors which are orthogonal, and the first columns T_1 use as few columns as possible to explain the variability of Y. PLS will be used in the analysis of the data and predictions of diagnoses.

5.3.4 Validation using a Leave-one-out Method

When a model is fit to a data set using any of the methods discussed in the previous section, the predictor is then applied to that data set to assess the accuracy of the predictor. It is inevitable that when the model is used to make predictions for a new data set not used in the construction of the model, the accuracy is not as good. The difference between the two predictions is referred to as the optimism of the model. Several methods are used to minimize the optimism, one such method being the jackknife, or leave-one-out method.

The concept of the jackknife is to serially remove each case of the data, fit the model without that case, and then predict the y value for that case using the model. What this is accomplishing is to make predictions using models that were not constructed using the case being predicted. The algorithm is summarized in Fig. 5.6.

Using this algorithm a less optimistic prediction error can be constructed. This algorithm and other variants of it are used to assess all results in this study.

For i = 1 to n
    $X_i = X$ with the i^{th} row removed.
    $Y_i = Y$ with the i^{th} element removed.
    Fit the model Y_i and X_i.
    Predict the i^{th} value of Y using the i^{th} row of X.
    Accumulate the sum of the squared prediction errors.
end

Figure 5.6: Jacknife algorithm.

5.4 Results

5.4.1 Analysis of Detection

The purpose of this section is to develop a predictor of disease state based on the histogram bins counts generated. The dependent variable, disease state, was coded with 0's for normal and 1's for the disease state. The normal images were separately compared with disease Risks 1, 2 and 3. For each of the nine cases (see Table 5.6), histograms of the IA and ||IF|| are generated for every image. This is a total of 720 histogram bins for each image. This number is reduced by first solving the detection problem for each case, then combining the nine detectors into one. The regression model is given by (5.1).

The first step was to develop a robust estimator for each case. To accomplish this, each case were analyzed separately fitting PLS models with 2 to 20 latent factors. In each combination, a jackknife estimate of the predictors was constructed and the **area under the ROC curve** (**AUC**) estimated using these predictions. Jackknifed predictions were used for robustness. The number of latent factors giving the highest AUC was retained as the optimal number of latent factors for that filter case. To increase the robustness of each model, only cases using 10 or fewer factors were included in the next step since a larger number of factors generally indicates that the model is fitting noise. Table 5.7 shows the resulting number of latent factors used in the experiments that produced the results which will be presented in the sub-section 5.4.

For example, in the detection of Risk 1 in the data containing Risk 1 as the only level of retinopathy and Risk 0 (normals), cases 4 and 9 (from Table 5.6) provided the best model for the detection of the Risk 1 patients. These cases appear to indicate that the spatial information, as encoded by the associated scales, serve to

differentiate the two classes of images, i.e., normal versus Risk 0. Similarly, one can select appropriate combinations of scale (cases from Table 5.6) for the detection of the Risk 2 and Risk 3. Likewise, detecting certain lesion types, such as NVE[3]/NVD[4], will be performed using specific scales.

Based on this factor-based analysis, the next step was to use the information to produce a matrix of independent variables that is parsimonious, well conditioned and robust. To accomplish this, the data for each case with 10 or fewer factors was used in a PLS model using the optimal number of factors as shown in the tables above. For the normal's vs. Risk 1 retinopathy, for example, case 1 was fitted with a PLS model using 9 factors producing a T matrix of t-scores, T_1. The case 2, which requires 16 factors, was not fitted since the number of factors is greater than 10. Finally, case 9 was fitted using 9 factors producing a t-score matrix T_9. From this, a matrix of independent variables is constructed as

$$X = [T_1 \mid T_3 \mid T_5 \mid T_6 \mid T_8 \mid T_9] .$$

From the regression model of (5.1), X is constructed from the T matrices. This model was fitted using PLS with 2 factors and jackknifed predictions of the diagnoses obtained. The resulting ROC curves and AUC's estimated from these predictions are presented in the next sub-section.

[3]NVE = Neovascularization.
[4]NVD = New Vessels of the Disc.

Table 5.7: Number of Factors in Normal versus Risk {1, 2, 3} Retinopathy.

Risk 1		**Risk 2**		**Risk 3**	
Case	# Factors	Case	# Factors	Case	# Factors
1	9	1	8	1	8
2	16	2	4	3	8
3	7	3	11	3	3
4	18	4	4	4	4
5	5	5	2	5	9
6	2	6	4	6	8
7	16	7	4	7	9
8	2	8	7	8	10
9	9	9	3	9	4

Table 5.8: Distribution of image quality by Risk level. Higher values indicate better image quality.

		Image Quality				
Risk	**Total**	1	2	3	4	5
0	92	2	3	5	43	39
1	71	0	1	4	37	29
2	50	0	3	9	23	15
3	52	0	3	6	36	7

5.4.2 Results for Diabetic Retinopathy

We first consider the application of AM-FM feature extraction and PLS classification to the four categories of DR severity (0 = none; 1 = few MAs[5]; and 2 = MAs and hemorrhages present, and 3 = extensive MAs, hemorrhages, possible ME[6] and neovascularization). The sensitivity, specificity, and area under the ROC curve are

[5]MAs = Microaneurysms.
[6]ME = Macular Edema.

given for both the testing processes described above. First, we determine the ability to correctly detect those images with signs of DR in a set of images composed of Risk 0 (normal N = 92) and Risk 1 (N = 71). A total of 163 images were selected from the total available (N = 265). As with all the experiments, AM-FM features were calculated for the nine cases (see Table 5.6). The PLS-based classifier was tested using all combinations to determine the best model (as measured by AUC). Fig. 5.7 (a) shows the ROC curve for detecting and classifying Risk 1 images in a set of Risk 1 and Risk 0. Next, we tested the same sample of Risk 0 (normals) versus Risk 2, versus Risk 3 and versus all images with any sign of DR in Figs. 5.7 (b), (c) and (d), respectively.

In our second application, we consider the application of AM-FM/PLS processing to images sets with vascular abnormalities and risk for macular edema (DR level 3 and macular edema level 2). Also, the effects of image quality are addressed here. Table 5.8 shows the distribution of image quality for the 265 test images. Here, we combine 36 images from Risks 2 and 3 with risk of macular edema and attempt to separate them from 11 images with vascular abnormalities. All (100%) of these images were correctly classified.

5.5 Discussion for Diabetic Retinopathy

In classifying Risk 0 versus Risk 1 (Fig. 5.7 (a)), we got a sensitivity of 95% with a specificity is 85%. Area under the ROC is 0.984. This data set of 265 includes 4.5% images of less than optimal quality, i.e., worse than image quality grade 3. A sensitivity of 98% and specificity of 93% would be achieved if 4-5% of the worse quality images are removed. The rationale for eliminating poor quality images is found in papers by Zimmer-Galler [87] and Abramoff [88], where 11% and 12% were found to be ungradeable in their studies.

Fig. 5.7 (b) presents the ROC curve for comparing Risk 0 versus Risk 2. For a sensitivity of 90%, the specificity is 80%. Area under the ROC was 0.95. Risk 0 versus Risk 3 from Fig. 5.7 (c) presents a sensitivity of 100%, specificity is 82%. Area under the ROC was 0.973. Finally, Risk 0 (normals) versus all images with any sign of DR (Figure 5.7 (d)) produced a sensitivity of 100%, specificity is 82% and an area under the ROC equal to 0.95.

Our preliminary conclusion is that the AM-FM approach is somewhat insensitive to image quality. This finding was not entirely unexpected. Our approach required no preprocessing of the images. It is common to "flatten" images to remove uneven lighting artifacts which hamper most segmentation approaches for explicitly detecting lesions, such as MA and hemorrhages. With the AM-FM approach, the frequency-based features are not seriously affected.
For the 265 image data set, the sensitivity and specificity statistics quoted above are improved to 99% and 92%, respectively when removing the 4.5% worse quality images as is done in many other studies.

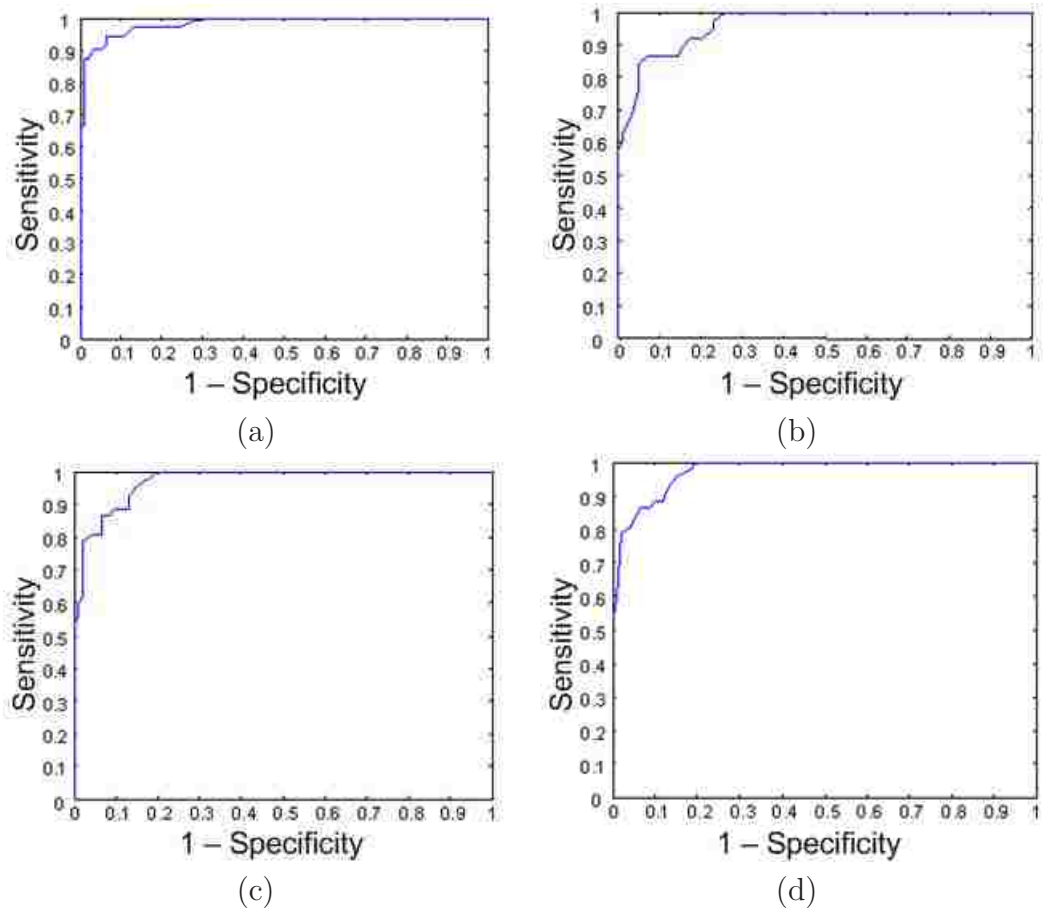


Figure 5.7: ROC for detection of Risk 0 (normal) in a set of normals versus different forms of DR: Risk {1, 2, 3}. (a) Risk 0 versus Risk 1. (b) Risk 0 versus Risk 2. (c) Risk 0 versus Risk 3. (d) Risk 0 versus all patients with any form of DR.

Part III

AM-FM methods for video processing

Chapter 6

Robust Discrete Video AM-FM Demodulation Methods

In this chapter we present robust 3D AM-FM demodulation methods. We begin with the design of an efficient filterbank (section 6.1) to be used in all the 3D methods and applications described here. Next, in section 6.2 we present a robust approach to use the VS-LQP methods in 3D. Finally, results and discussion are presented in the last sections.

6.1 3D Multi-scale filterbank design for applications in discrete videos

We extended the 2D filterbank design from section 3.1 to generate a 3D multi-scale filterbank. The same equiripple design, with the same specifications are used. The third dimension, *time*, will increase the total number of bandpass filters.

Fig. 6.1 shows, from left to right, the numbering system for 1D filters through

the time direction, to be used for a 3D single-, two- and three-scale filterbank, respectively. Note that the bandpass filters always have frequency support in only half of the spectrum. For the filters through time, we adopt the following notation:

- Filter 1 is the low pass filter (LPF).
- Filters 2 and 3 are high frequency filters.
- Filters 4 and 5 are medium frequency filters.
- Filters 6 and 7 are low frequency filters.

Thus, considering that for 2D applications the designed filterbanks have 7, 13 and 19 bandpass filters for a single-, two- and three-scale filterbank, respectively, for the 3D filterbanks we have 21, 65 and 133 3D bandpass filters.

Recall the 2D two-scale filterbank from section 3.1 (see Fig. 3.1(b)). Fig. 6.2(a) shows the same filterbank but now with the added frequencies associated with the time variable. Fig. 6.2(b) shows the 3D frequency-domain decomposition for a 3D two-scale filterbank.

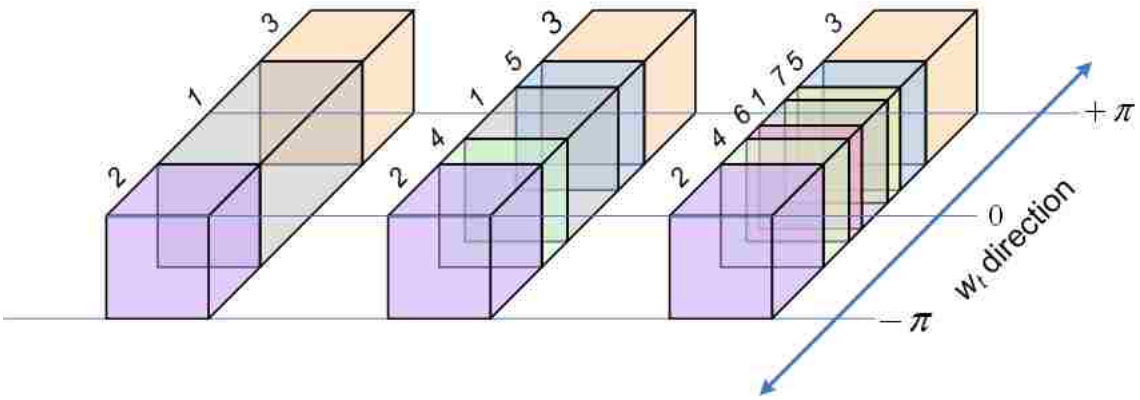
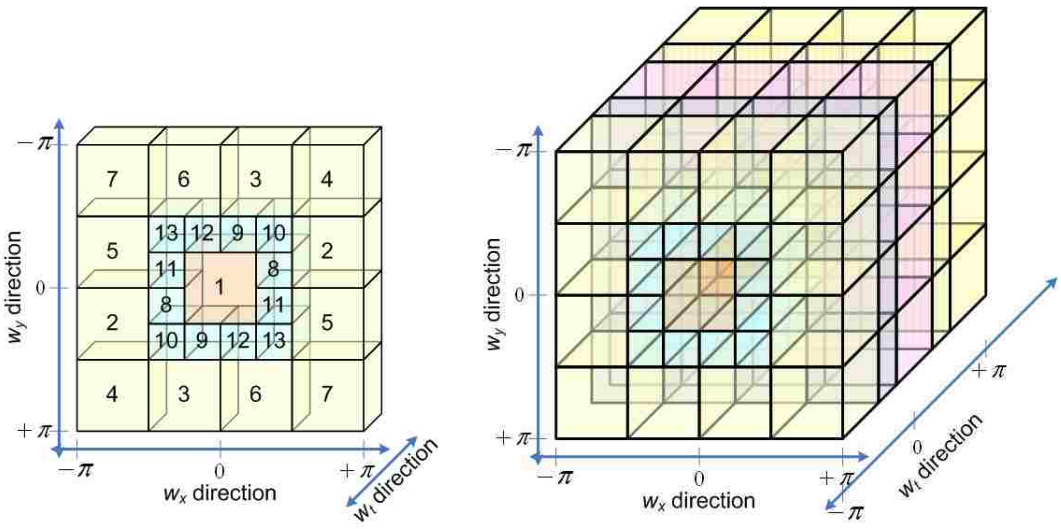


Figure 6.1: Numbering of the bandpass filters for the time variable for, from left to right, 3D single-, two- and three-scale filterbanks, respectively.



(a) (b)

Figure 6.2: Frequency spectrum decomposition for 3D two-scale filterbank. (a) 2D two-scale filterbank for a single w_x slice. (b) Complete 3D frequency spectrum decomposition for the 3D two-scale filterbank.

6.2 3D VS-LQP method

We extend to 3D the 2D formulas from chapter 3 (see chapter 2 also). We consider 3D multi-scale AM-FM representations for videos given by (1.2), with the IF defined by (1.4).

Consider a single component AM-FM video defined by

$$I(x, y, t) = a(x, y, t) \cos \varphi(x, y, t). \tag{6.1}$$

In this case, the AM-FM demodulation problem is to estimate: (i) the amplitude function $a(x, y, t)$, (ii) the phase function $\varphi(x, y, t)$, and (iii) the instantaneous frequency function $\nabla\varphi(x, y, t)$ from $I(.)$.

The 3D AM-FM demodulation methods in this work are based on extensions

of the one-dimensional analytic signal. The first step is to compute the extended 3D version of the one-dimensional analytic signal associated with $I(x, y, t)$. The extended analytic signal is computed using:

$$I_{AS}(x, y, t) = I(x, y, t) + j\mathcal{H}_{3d}[I(x, y, t)], \tag{6.2}$$

where $\mathcal{H}_{3d}$ denotes a three-dimensional extension of the one-dimensional Hilbert transform operator. The three dimensional operator is defined in terms of the one dimensional operator, operating in either the x, the y or the t direction:

$$\mathcal{H}_{3d}[I(x, y, t)] = \frac{1}{\pi x} * I(x, y, t), \tag{6.3}$$

for the x direction. For the algorithm to work, we must have that $I_{AS}(\cdot)$ is equal, or at-least approximately equal, to the complex AM-FM harmonic associated with (6.1) (see [32] for the 1D case):

$$I_{AS}(x, y, t) \approx a(x, y, t) \exp\left(j\varphi(x, y, t)\right). \tag{6.4}$$

When the approximation holds, it is possible to estimate the amplitude, the phase and the instantaneous frequency using

$$\hat{a}(x, y, t) = |I_{AS}(x, y, t)|, \tag{6.5}$$

$$\hat{\varphi}(x, y, t) = \arctan\left(\frac{\text{imag}(I_{AS}(x, y, t))}{\text{real}(I_{AS}(x, y, t))}\right), \tag{6.6}$$

and

$$\hat{\nabla\varphi}(x, y, t) = \text{real}\left[-j\frac{\nabla I_{AS}(x, y, t)}{I_{AS}(x, y, t)}\right]. \tag{6.7}$$

The algorithm can thus be summarized into two steps. First, compute the extended analytic signal using (6.2). Second, compute all the estimates using (6.5), (6.6) and (6.7).

For discrete videos, we define the discrete-space-time input signal as

$$I(k_1, k_2, k_3) \approx a(k_1, k_2, k_3) \cos\varphi(k_1, k_2, k_3), \tag{6.8}$$

where k_1 and k_2 represent the discrete versions of x and y, respectively, and k_3 represents the discrete version of t. In what follows, we will not consider the conversion of the estimated discrete frequencies back to continuous space. It is also important to note that the term "analytic" here is inaccurate. Our approach is meant to simply enable us to produce an estimate of $a(k_1, k_2, k_3) \exp(j\varphi(k_1, k_2, k_3))$ from $a(k_1, k_2, k_3) \cos \varphi(k_1, k_2, k_3)$ *very fast*. Our approach is not intended to produce an analytic signal in the mathematical sense. We define the discrete-space-time extended estimate of the 1-D analytic signal $\hat{I}_{AS}(k_1, k_2, k_3)$. This is accomplished via the use of a discrete-time Hilbert Transform or using the Fast Fourier Transform (FFT), as given in (6.2).

Recall sections 3.2 and 3.4 to extend the 2D robust QEA method to 3D. Define $\bar{I}_{AS}$ such that

$$\begin{aligned}
\bar{I}_{AS}(k_1, k_2, k_3) &= \frac{\hat{I}_{AS}(k_1, k_2, k_3)}{|\hat{I}_{AS}(k_1, k_2, k_3)|} \\
&= \frac{a(k_1, k_2, k_3) \exp(j\varphi(k_1, k_2, k_3))}{a(k_1, k_2, k_3)} \\
&= \exp(j\varphi(k_1, k_2, k_3)).
\end{aligned}$$

Then, for the IF (assuming $\delta x = \delta y = \delta t = 1$) we get:

$$\hat{\varphi}_x(k_1, k_2, k_3) = \frac{1}{n_1} \arcsin \left[\frac{\bar{I}_{AS}(k_1 + n_1, k_2, k_3) - \bar{I}_{AS}(k_1 - n_1, k_2, k_3)}{2j\bar{I}_{AS}(k_1, k_2, k_3)} \right], \quad (6.9)$$

$$\hat{\varphi}_y(k_1, k_2, k_3) = \frac{1}{n_2} \arcsin \left[\frac{\bar{I}_{AS}(k_1, k_2 + n_2, k_3) - \bar{I}_{AS}(k_1, k_2 - n_2, k_3)}{2j\bar{I}_{AS}(k_1, k_2, k_3)} \right], \quad (6.10)$$

$$\hat{\varphi}_t(k_1, k_2, k_3) = \frac{1}{n_3} \arcsin \left[\frac{\bar{I}_{AS}(k_1, k_2, k_3 + n_3) - \bar{I}_{AS}(k_1, k_2, k_3 - n_3)}{2j\bar{I}_{AS}(k_1, k_2, k_3)} \right], \quad (6.11)$$

$$\hat{\varphi}_x(k_1, k_2, k_3) = \frac{1}{n_1} \arccos \left[\frac{\bar{I}_{AS}(k_1 + n_1, k_2, k_3) + \bar{I}_{AS}(k_1 - n_1, k_2, k_3)}{2\bar{I}_{AS}(k_1, k_2, k_3)} \right], \quad (6.12)$$

$$\hat{\varphi}_y(k_1, k_2, k_3) = \frac{1}{n_2} \arccos \left[\frac{\bar{I}_{AS}(k_1, k_2 + n_2, k_3) + \bar{I}_{AS}(k_1, k_2 - n_2, k_3)}{2\bar{I}_{AS}(k_1, k_2, k_3)} \right] \quad (6.13)$$

and

$$\hat{\varphi}_t(k_1,k_2,k_3)=\frac{1}{n_3}\arccos\left[\frac{\bar{I}_{AS}(k_1,k_2,k_3+n_3)+\bar{I}_{AS}(k_1,k_2,k_3-n_3)}{2\bar{I}_{AS}(k_1,k_2,k_3)}\right], \quad (6.14)$$

with similar constraints as in the 2D case in (3.8).

6.3 Results on 3D AM-FM estimation

In this section, we present results for the 3D AM-FM methods using synthetic 3D signals. We filter the signal through each channel using the 3D FFT.
We first define the synthetic signals used and then we present the results.

6.3.1 3D Chirp

For this signal, we use $I(k_1,k_2,k_3)=\cos\varphi(k_1,k_2,k_3)$, for $k_1,k_2,k_3=-\frac{N}{2},\ldots,\frac{N}{2}-1$ and $N=256$. The phase $\varphi(k_1,k_2,k_3)$ has a quadratic form

$$\varphi(k_1,k_2,k_3)=\alpha_x k_1+\frac{\beta_x}{2}k_1^2+\alpha_y k_2+\frac{\beta_y}{2}k_2^2+\alpha_t k_3+\frac{\beta_t}{2}k_3^2,$$

where $\beta_i=\left(\frac{f_{B_i}-f_{A_i}}{N-1}\right)$, $\alpha_i=\frac{\beta_i N}{2}+f_{A_i}$, for $i=1,2,3$. f_{B_i} represents the desired maximum instantaneous frequency in the i direction (x, y or t), and f_{B_i} represents the desired minimum instantaneous frequency. Setting f_{B_i} and f_{A_i} will produce $\varphi_x(k_1,k_2,k_3)$, $\varphi_y(k_1,k_2,k_3)$ and $\varphi_t(k_1,k_2,k_3)$ to be in desired ranges. We consider the instantaneous frequency $\varphi_x(k_1,k_2,k_3)\in[-0.4\pi,0.4\pi]$, with $\varphi_x(k_1,k_2,k_3)=\varphi_y(k_1,k_2,k_3)=\varphi_t(k_1,k_2,k_3)$.

Table 6.1: MSE and PSNR for IA estimation using 3D synthetic signals.

	3D Chirp		**3D Gauss AM-FM**	
Filterbank	**MSE**	**PSNR** (dB)	**MSE**	**PSNR** (dB)
Single-scale	2.127x10^{-2}	16.721	29.21	25.344
Two-scale	2.732x10^{-2}	15.636	33.05	24.809
Three-scale	3.324x10^{-2}	14.784	44.10	23.555

6.3.2 3D Gaussian amplitude-modulated frequency-modulated

We use

$$I(k_1,k_2,k_3) = 100\ \exp\left[-\frac{1}{2}\alpha^2\sum_{i=1}^{i=3}\left(\frac{k_i+\frac{1}{2}}{\frac{N}{2}}\right)^2\right]\cos\varphi(k_1,k_2,k_3)$$

for $k_1,k_2,k_3 = -\frac{N}{2},\ldots,\frac{N}{2}-1$ and $N=256$. The phase $\varphi(k_1,k_2,k_3)$ has a quadratic form equal to the 3D Chirp signal from the previous sub-section.

6.3.3 Results for synthetic 3D signals

In Table 6.1 we show both the MSE and the PSNR (given by $10\log_{10}\left(\frac{IA_{max}^2}{MSE}\right)$, with $IA_{max}=1$ when a 3D chirp is used, and $IA_{max}=100$ in the other case) for the IA estimation of both synthetic images.

In Table 6.2, we show the errors, in the IF estimation, in terms of the PSNR for the 3D Chirp and 3D Gaussian amplitude-modulated frequency-modulated. Results in terms of the MSE are showed in Table 6.3. The errors are computed in terms of the magnitude of the IF in each direction. Thus, the PSNR is computed using $10\log_{10}\left(\frac{\pi^2}{MSE}\right)$.

Table 6.2: PSNR (dB) for ||IF|| estimation using 3D synthetic signals.

	3D Chirp			**3D Gauss AM-FM**		
Filterbank	x	y	t	x	y	t
Single-scale	43.806	40.403	42.780	39.137	35.862	39.907
Two-scale	43.819	42.459	43.386	41.517	41.168	42.767
Three-scale	42.030	41.972	39.467	41.057	41.210	39.939

Table 6.3: MSE for ||IF|| estimation using 3D synthetic signals.

	3D Chirp			**3D Gauss AM-FM**		
Filterbank	x	y	t	x	y	t
Single-scale	$4.11\text{x}10^{-4}$	$9.00\text{x}10^{-4}$	$5.20\text{x}10^{-4}$	$1.20\text{x}10^{-3}$	$2.56\text{x}10^{-3}$	$1.01\text{x}10^{-3}$
Two-scale	$4.10\text{x}10^{-4}$	$5.60\text{x}10^{-4}$	$4.53\text{x}10^{-4}$	$6.96\text{x}10^{-4}$	$7.54\text{x}10^{-4}$	$5.22\text{x}10^{-4}$
Three-scale	$6.18\text{x}10^{-4}$	$6.27\text{x}10^{-4}$	$1.12\text{x}10^{-3}$	$7.74\text{x}10^{-4}$	$7.47\text{x}10^{-4}$	$1.00\text{x}10^{-3}$

6.4 Discussion for 3D AM-FM estimation

As expected, from Table 6.1, we can see how in the IA estimation as we increase the number of scales in the filterbank (in this noise-free example), the error increases. In terms of ||IF|| estimation, both signals produced very similar results (see Tables 6.2 and 6.3). Note how when we change the use of a single-scale filterbank to a two-scale filterbank, the MSE is significantly reduced.

Chapter 7

Video Reconstructions using AM-FM Decompositions

In this chapter, we discuss how to reconstruct a video using its AM-FM components. We extend the theory presented for 2D signals in chapter 4 to get 3D versions of those three methods.

7.1 3-D Multi-scale reconstructions

We use the 3D multi-scale filterbank from section 6.1 to extend the 2D reconstruction methods from section 4 to reconstruct videos using: (i) 3D Least-Squares Reconstructions using AM-FM harmonics (3D-LESHA), (ii) 3D Multi-scale least-squares reconstructions (3D-MULTILES) and (iii) 3D Least-Squares Reconstructions using AM-FM harmonics and the DCA (3D-LESCA). Since the theory is similar to the one explained in section 4, we only present 3D-MULTILES.

3D-MULTILES is based on the scales of the filterbanks designed in section 6.1.

Define:

d:	Global DC image estimate.
$G(k_1, k_2, k_3)$:	Low pass filter output.
$a_1 \cos \varphi_1$:	High-frequency scale AM-FM component.
$a_2 \cos \varphi_2$:	Medium-frequency scale AM-FM component.
$a_3 \cos \varphi_3$:	Low-frequency scale AM-FM component.

In this case, we consider least squares reconstructions given by:

$$\hat{I}(k_1, k_2, k_3) \approx d + c_0 G(k_1, k_2, k_3) + \sum_{n=1}^{s} c_n a_n(k_1, k_2, k_3) \cos\left(\varphi_n(k_1, k_2, k_3)\right), \quad (7.1)$$

where s is the number of scales used.

We then compute the AM-FM multi-scale coefficients c_n, $n = 0, 1, \ldots, s$, so that $\hat{I}(k_1, k_2, k_3)$ is a least-squares estimate of $I(k_1, k_2, k_3)$. We also compute an orthonormal basis over the space of the AM-FM estimations scale by scale (see section 4.3 for a 2D reference). Fig. 7.1 shows an example of the 3D bandpass filters used for this method when a two-scale filterbank is used.

It is important to recall that adding decomposition levels also reduces the total amount of video signal energy that is captured by the decomposition. First, let us note that for a single scale decomposition, video signal energy is captured by the low-pass filter component and the dominant high-frequency components, selected from the high frequency 3D bandpass filters. Then, in two-scale decompositions, the 3D spectrum captured by the low-pass filter is further decomposed into two new scales. We again find the dominant components in this second scale while the lowest frequency components are captured by the new low-pass filters. Similarly, for three-scales, we decompose the frequency spectrum of the 3D low-pass filter.

The extracted dominant components from each scale allow us to provide decompositions using an independent AM-FM component per scale. Furthermore, the

corresponding dominant channel filters allow us to extract local spatiotemporal content over each pixel. This approach allows us, e.g., to re-formulate the classical motion estimation problem with several independent equations over each scale. It is also important to note that the AM-FM decomposition also allows us to track both continuous and discontinuous motions since at every pixel we can associate three different dominant channels from three different scales.

For the 3D-LESHA reconstruction method, we consider reconstructing the input video using AM-FM harmonics (see [31]):

$$\hat{I}(k_1, k_2, k_3) \approx d + \sum_{n=1}^{h} c_n a(k_1, k_2, k_3) \cos\left(n\varphi(k_1, k_2, k_3)\right), \tag{7.2}$$

where d is a scalar and h is the maximum number of AM-FM harmonics to use.

For 3D-LESCA, we also use $G(k_1, k_2, k_3)$ (the LPF output). Thus, we consider least squares video reconstructions using:

$$\hat{I}(k_1, k_2, k_3) = d + c_0 G(k_1, k_2, k_3) + \sum_{n=1}^{h} c_h a(k_1, k_2, k_3) \cos\left(n\varphi(k_1, k_2, k_3)\right). \tag{7.3}$$

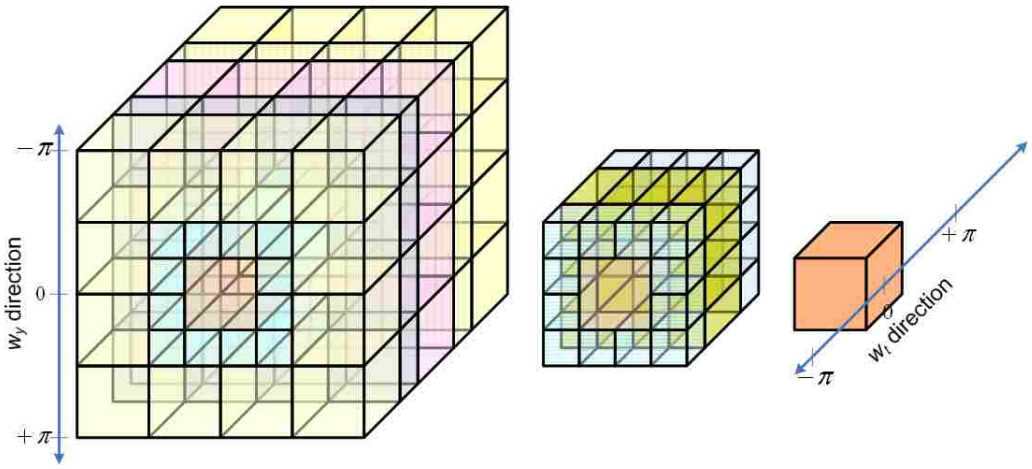


Figure 7.1: 3D-MULTILES method for video reconstruction using a 3D two-scale filterbank.

Table 7.1: MSE and Average Image Quality Index in the reconstructions of the videos using 3D-MULTILES method (multi-scale). LPF denotes low-pass filter reconstruction, M denotes the medium frequencies and H denotes high frequencies.

	Scales		
Mean-Square Error	LPF	LPF+M	LPF+M+H
Average MSE	379.09	330.70	249.64
Minimum MSE	334.01	284.58	216.00
Maximum MSE	467.06	413.54	316.28
Average Image Quality Index	LPF	LPF+M	LPF+M+H
Average Qv	0.543	0.618	**0.712**
Minimum Qv	0.535	0.607	**0.706**
Maximum Qv	0.552	0.626	**0.717**

7.2 Results for Video Reconstructions

We use the videos from [89], provided for detecting human activity. The video resolution is at 384x288 pixels per frame, 25 frames per second and compressed using MPEG2 (half the resolution of the PAL standard). The file sizes are mostly between 6 and 12 MB, a few up to 21 MB. We used nine videos in total, corresponding to three different scenarios: a person browsing at information displays (3 videos), groups of people encountering (3 videos) and people fighting (3 videos).

We reconstructed the videos using the three methods described in section 7.1 and using a two-scale filterbank (see section 6.1).

In Table 7.1 we show the average MSE obtained in the reconstruction of the nine videos using the 3D-MULTILES method. We present both the minimum MSE and the maximum MSE for each case. We also present the average image quality index value Qv (see section 4.4. Tables 7.2 and 7.3 show the MSE and Qv for the 3D-LESHA and 3D-LESCA methods, respectively.

Table 7.2: MSE and Average Image Quality Index in the reconstructions of the videos using 3D-LESHA method.

	Number of AM-FM harmonics				
Mean-Square Error	1	2	3	4	5
Average MSE	433.20	429.64	429.20	428.65	428.64
Minimum MSE	401.15	397.82	397.32	396.83	396.82
Maximum MSE	475.92	472.70	472.27	471.70	471.69
Average Image Quality Index	1	2	3	4	5
Average Qv	0.558	0.560	0.558	0.557	0.557
Minimum Qv	0.551	0.552	0.550	0.550	0.550
Maximum Qv	0.566	0.567	0.565	0.565	0.565

7.3 Discussion for Video Reconstructions

Similarly to the 2D results (see section 4.4), 3D-MULTILES produce the best results in the MSE sense, and 3D-LESHA the worst ones. 3D-LESHA does not improve considerably its results when the number of harmonics is increased. Note that 3D-LESCA improved its results from MSE ~ 379, when only the LPF was used, to MSE

Table 7.3: MSE and Average Image Quality Index in the reconstructions of the videos using 3D-LESCA method (same as 3D-LESHA + LPF).

Mean-Square Error	LPF	LPF+1h$^\flat$	LPF+5h$^\sharp$
Average MSE	379.09	287.14	284.8553
Minimum MSE	334.01	250.63	247.1343
Maximum MSE	467.06	357.53	355.3691
Average Image Quality Index	LPF	LPF+1h$^\flat$	LPF+5h$^\sharp$
Average Qv	0.543	0.667	0.668
Minimum Qv	0.535	0.661	0.662
Maximum Qv	0.552	0.674	0.674

LPF+1h$^\flat$ = LPF + LPF + 1 AM-FM harmonic.

LPF+5h$^\sharp$ = LPF + 5 AM-FM harmonics.

~ 284, when 5 harmonics were added. Furthermore, 3D-MULTILES improved its results up to MSE ~ 249 when all the scales where used.

For 3D-LESCA, note that video reconstruction was improved when a single AM-FM harmonic was added to the LPF (see Table 7.3). When 3D-MULTILES is used, we get almost good quality videos with Qv around 0.712 on the average. The 3D AM-FM reconstructions methods are useful for analyzing the importance of each scale, in terms of contribution of information, for applications such as motion estimation. Indeed, 3D-MULTILES will be used in the next chapter as the first processing step for analyzing the input videos to compute the motion estimation in them.

Chapter 8

Motion Estimation (ME) methods based on AM-FM

8.1 Introduction

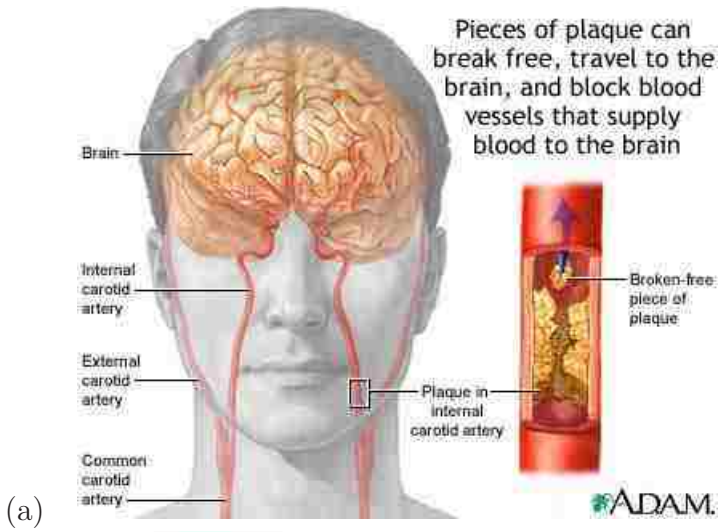
In this chapter, we present a new method for pixel-level based motion estimation using an Amplitude-Modulation Frequency-Modulation (AM-FM) model for digital video. We also consider an application of this model in analyzing atherosclerotic plaque motion, general video motion and video activity recognition.

Our study is motivated from a desire to extend traditional motion estimation methods into the development of reliable methods for video trajectory estimation. To accomplish this for atherosclerosis, we are interested in developing realistic plaque motion models that are motivated from clinical experience. We expect that accurate motion estimation will help us develop more accurate models that can predict plaque rupture (see Fig. 8.1). If we can predict the breakup, we should be able to reduce the number of stentings (or surgeries) in patients (Fig. 8.1 (b)).

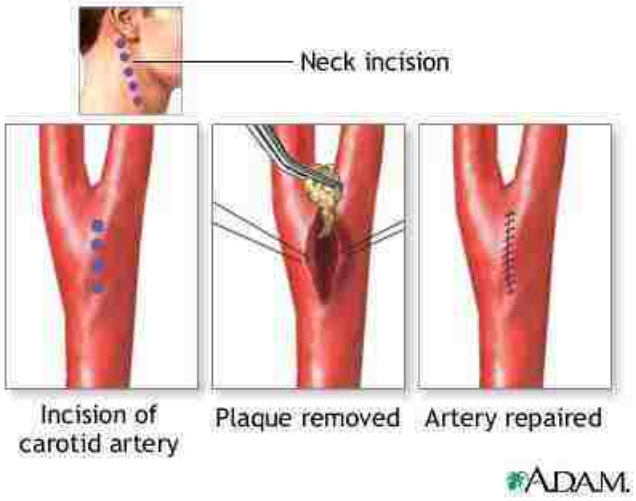
In related work [90], the authors computed optical flow estimates from 45 patients and reported a significant increase in the maximal discrepant surface velocity for the symptomatic cases, as compared to the asymptomatic cases. In [91], the authors used 3D intravascular ultrasound to provide a computational analysis of stress distribution. Clearly, the development of accurate motion estimation methods can also benefit this related research.

Using realistic motion models, we propose a new, amplitude and phase based motion estimation method based on robust, multidimensional Amplitude-Modulation Frequency-Modulation (AM-FM) methods. Extending out prior work reported in [92], we investigate the limits of the new method as compared against traditional motion estimation methods.

We provide background information about related work in section 8.2. We describe our new method in section 8.3. Results are shown in section 8.4 and finally, the discussion is given in section 8.5. In section 8.6, we consider an application in video activity recognition.



(a)



(b)

Figure 8.1: Atherosclerotic plaques. (a) Internal carotid artery. (b) Surgery in the carotid artery. Images from Adam ©: *http://www.adam.com*

8.2 Background on Motion Estimation

8.2.1 Optical Flow

Fundamental to the concept of actual object motion, but nevertheless different, is optical flow. Optical flow is the instantaneous motion of image intensities. This is not the same as the motion of the objects being imaged. This is easily understood by these examples:

- A moving camera taking pictures of a stationary object will produce image (optical flow) motion (despite the lack of object motion).
- A rotating uniform intensity sphere will not produce any image (optical flow) motion.

Still, optical flow is all the motion information that the image supplies. Therefore, most methods of motion estimation, motion compensation, etc. depend on it.

8.2.2 Continuous Formulation

The image intensity at a point in space and time is $I(x, y, t)$. After a sufficiently small time interval $\triangle t$, the intensity at (x, y) will move to a point $(x + \triangle x, y + \triangle y)$. In other words: $I(x+\triangle x, y+\triangle y, t+\triangle t) = I(x, y, t)$. This assumes that the intensity does not change, just its position.

Expanding the LHS in a Taylor's series:

$$I(x + \triangle x, y + \triangle y, t + \triangle t) = I(x, y, t) + \triangle x \cdot \frac{\partial I}{\partial x} + \triangle y \cdot \frac{\partial I}{\partial y} + \triangle t \cdot \frac{\partial I}{\partial t} + h.o.t.$$

If we ignore the higher order term ($h.o.t.$), divide by $\triangle t$, and then take the limit as $\triangle t \to 0$, we get:

$$\frac{\partial I}{\partial x} \cdot \frac{\partial x}{\partial t} + \frac{\partial I}{\partial y} \cdot \frac{\partial y}{\partial t} + \frac{\partial I}{\partial t} = 0.$$

Now, define the motion velocities u and v using: $u(x,y,t) = \partial x/\partial t$ and $v(x,y,t) = \partial y/\partial t$.

Putting this all together gives the optical flow constraint equation (OFCE):

$$I_x u + I_y v + I_t = 0. \tag{8.1}$$

In (8.1), note that at every pixel we have a single equation in two unknowns.

8.2.3 Phase-Based Methods for Motion Estimation

Fleet and Jepson [47] proposed the use of an AM-FM model for modeling digital video based on

$$\mathbf{R}(\mathbf{x},t) = \rho(\mathbf{x},t)\exp(j\varphi(\mathbf{x},t)) \tag{8.2}$$

where $\mathbf{x}$ denotes the spatial variables $\mathbf{x} = (x,\, y)$, t denotes time, $\rho(\mathbf{x},t)$ denotes the amplitude and $\varphi(\mathbf{x},t)$ denotes the phase component. For estimating the AM-FM components, Fleet and Jepson used a set of band-pass complex valued Gabor filters.

For pixel velocity estimation, the basic model of (8.2) can only be used to estimate the projected component velocities that are in the direction of the instantaneous frequency. We express this using $\mathbf{v}_n = \alpha\ \mathbf{n}(\mathbf{x},t)$ where $\varphi_{\mathbf{x}}(\mathbf{x},t) = (\varphi_x, \varphi_y)^T$, $\mathbf{n}(\mathbf{x},t) = \varphi_{\mathbf{x}}(\mathbf{x},t)/\|\varphi_{\mathbf{x}}(\mathbf{x},t)\|$ and $\alpha = -\varphi_t(\mathbf{x},t)/\|\varphi_{\mathbf{x}}(\mathbf{x},t)\|$.

For estimating the instantaneous frequency, they proposed to use:

$$\nabla\varphi(\mathbf{x},t) = \frac{Im\left[\mathbf{R}^*(\mathbf{x},t)\,\nabla\mathbf{R}(\mathbf{x},t)\right]}{\rho^2(\mathbf{x},t)}, \tag{8.3}$$

where $\mathbf{R}^*$ denotes the complex conjugate of $\mathbf{R}$, $Im[\mathrm{z}] \equiv (Im[z_1], Im[z_2], Im[z_3])$.

Note that in order to recover the velocity components that are orthogonal to the instantaneous frequency vectors we will need to apply some type of smoothing over

the estimated velocities. Fleet and Jepson accomplish this by fitting a local linear model over 5×5 neighborhoods:

$$\mathbf{v}\left(\mathbf{x}, t\right) = \left(\alpha_0 + \alpha_1 x + \alpha_2 y, \beta_0 + \beta_1 x + \beta_2 y\right). \tag{8.4}$$

Then, from (8.4), the estimated 2D velocity is then taken to be (α_0, β_0).

To provide for a method of identifying accurate measures, Fleet and Jepson require that the estimated instantaneous frequency is within the range of the estimating filter. This is expressed as

$$\| \nabla\varphi\left(\mathbf{x}, t\right) - \left(k_i, w_i\right) \| \leq \tau\sigma_k \tag{8.5}$$

where (k_i, w_i) is the peak tuning frequency of the i-th filter, σ_k is the standard deviation of the filter's amplitude spectrum, and τ is a threshold used to reject unreliable estimates of instantaneous frequencies. Similarly, for the amplitude, they require that the local signal amplitude must be as large as the average local amplitude, and at least 5% of the largest response amplitude across all the filters at that frame. When either one of these two conditions is not met, the method does not provide velocity estimates.

8.3 AM-FM based Motion Estimation Method

8.3.1 An AM-FM model for motion estimation

In what follows, consider a single AM-FM component approximation to the input video:

$$I(x, y, t) = a(x, y, t) \exp(j\varphi(x, y, t)). \tag{8.6}$$

Recall the optical flow constraint equation:

$$I_x u + I_v v + I_t = 0. \tag{8.7}$$

We apply (8.7) to (8.6) and separate out the real from the imaginary parts to get the amplitude constraint equation

$$a_x u + a_y v + a_t = 0, \tag{8.8}$$

and the frequency modulation constraint equation

$$\varphi_x u + \varphi_y v + \varphi_t = 0. \tag{8.9}$$

The advantage of (8.8)-(8.9) is that they provide us with two equations per pixel. We also add a smoothness constrain to add a third equation. Collectively, the AM, FM and continuity constraints give:

$$E_s = \iint \left[u_x^2 + u_y^2 + v_x^2 + v_y^2\right] dxdy, \tag{8.10}$$

$$E_{AM} = \iint \left[a_x u + a_y v + a_t\right]^2 dxdy, \tag{8.11}$$

and

$$E_{FM} = \iint \left[\varphi_x u + \varphi_y v + \varphi_t\right]^2 dxdy. \tag{8.12}$$

We combine all constraints together to get

$$E = E_s + \lambda E_{FM} + \beta E_{AM}. \tag{8.13}$$

8.3.2 Discrete Optimization

For minimizing (8.13) we consider finite-difference approximations:

$$u_x \approx \left[u(i+1,j) - u(i,j)\right]/2,$$
$$u_y \approx \left[u(i,j+1) - u(i,j)\right]/2,$$
$$v_x \approx \left[v(i+1,j) - v(i,j)\right]/2,$$

and

$$v_y \approx [v(i, j+1) - v(i, j)]/2.$$

We then convert the sums to integrals and take derivatives with respect to $u(m, n)$ and $v(m, n)$ to get

$$\begin{aligned}\frac{\partial E}{\partial u(m,n)} &= 2\left[u(m,n) - u_{ave}(m,n)\right] \\ &\quad + 2\lambda\left[\varphi_x u + \varphi_y v + \varphi_t\right] \cdot \varphi_x \\ &\quad + 2\beta\left[a_x u + a_y v + a_t\right] \cdot a_x = 0\end{aligned}$$

and

$$\begin{aligned}\frac{\partial E}{\partial v(m,n)} &= 2\left[v(m,n) - v_{ave}(m,n)\right] \\ &\quad + 2\lambda\left[\varphi_x u + \varphi_y v + \varphi_t\right] \cdot \varphi_y \\ &\quad + 2\beta\left[a_x u + a_y v + a_t\right] \cdot a_y = 0.\end{aligned}$$

where

$$\begin{aligned}u_{ave}(m,n) &= \frac{1}{4}[u(m+1,n) + u(m-1,n) + u(m,n+1) \\ &\quad + u(m,n-1)]\end{aligned}$$

and

$$\begin{aligned}v_{ave}(m,n) &= \frac{1}{4}[v(m+1,n) \\ &\quad + v(m-1,n) + v(m,n+1) + v(m,n-1)].\end{aligned}$$

For optimal values, we require that the gradient should be zero

$$\frac{\partial E}{\partial u(m,n)} = \frac{\partial E}{\partial v(m,n)} = 0. \tag{8.14}$$

We re-write (8.14) as

$$\mathsf{A}\mathsf{x} = \mathsf{b},$$

where:

$$\mathsf{A} = \begin{bmatrix} (1+\lambda\varphi_x^2+\beta a_x^2) & (\lambda\varphi_x\varphi_t - \beta a_x a_y) \\ (\lambda\varphi_x\varphi_t - \beta a_x a_y) & \left(1+\lambda\varphi_y^2+\beta a_y^2\right) \end{bmatrix},$$

$$\mathsf{x} = \begin{bmatrix} u \\ v \end{bmatrix},$$

and

$$\mathsf{b} = \begin{bmatrix} (u_{ave} - \lambda\varphi_x\varphi_t - \beta a_x a_t) \\ (v_{ave} - \lambda\varphi_y\varphi_t - \beta a_y a_t) \end{bmatrix}.$$

Solving for $u(m,n)$ and $v(m,n)$ yields:

$$\begin{aligned} u = \quad u_{ave} \quad & -\lambda \frac{\left(\varphi_x + \beta a_y^2\varphi_x - \beta a_y a_x \varphi_y\right) u_{ave} + \varphi_y v_{ave} + \left(\beta a_y^2 + 1\right)\varphi_t - \beta a_y \varphi_y a_t}{1+\lambda\left(\varphi_y^2+\varphi_x^2\right) + \lambda\beta\left(\varphi_x a_y - \varphi_y a_x\right)^2 + \beta\left(a_x^2+a_y^2\right)} \varphi_x \\ & -\beta \frac{\left[\lambda\left(a_x\varphi_y - a_y\varphi_x\right)\varphi_y + a_x\right] u_{ave} + a_y v_{ave} - \lambda a_y\varphi_y\varphi_t + \left(\lambda\varphi_y^2+1\right) a_t}{1+\lambda\left(\varphi_y^2+\varphi_x^2\right) + \lambda\beta\left(\varphi_x a_y - \varphi_y a_x\right)^2 + \beta\left(a_x^2+a_y^2\right)} a_x \end{aligned} \tag{8.15}$$

and

$$\begin{aligned} v = \quad v_{ave} \quad & -\lambda \frac{\varphi_x u_{ave} + \left(\varphi_y + \beta a_x^2\varphi_y - \beta a_x a_y \varphi_x\right) v_{ave} + \left(\beta a_x^2 + 1\right)\varphi_t - \beta a_x \varphi_x a_t}{1+\lambda\left(\varphi_y^2+\varphi_x^2\right) + \lambda\beta\left(\varphi_x a_y - \varphi_y a_x\right)^2 + \beta\left(a_x^2+a_y^2\right)} \varphi_y \\ & -\beta \frac{a_x u_{ave} + \left[\lambda\left(a_y\varphi_x - a_x\varphi_y\right)\varphi_x + a_y\right] v_{ave} - \lambda a_x\varphi_x\varphi_t + \left(\lambda\varphi_x^2+1\right) a_t}{1+\lambda\left(\varphi_y^2+\varphi_x^2\right) + \lambda\beta\left(\varphi_x a_y - \varphi_y a_x\right)^2 + \beta\left(a_x^2+a_y^2\right)} a_y. \end{aligned} \tag{8.16}$$

Or, in a shorter form:

$$u = u_{ave} - \lambda\frac{u_{N1}}{D}\varphi_x - \beta\frac{u_{N2}}{D}a_x$$

and

$$v = v_{ave} - \lambda \frac{v_{N1}}{D} \varphi_y - \beta \frac{v_{N2}}{D} a_y$$

where:

$$\begin{aligned} u_{N1} &= \left(\varphi_x + \beta a_y^2 \varphi_x - \beta a_y a_x \varphi_y\right) u_{ave} + \varphi_y v_{ave} \\ &\quad + \left(\beta a_y^2 + 1\right) \varphi_t - \beta a_y \varphi_y a_t, \\ u_{N2} &= \left[\lambda \left(a_x \varphi_y - a_y \varphi_x\right) \varphi_y + a_x\right] u_{ave} + a_y v_{ave} \\ &\quad - \lambda a_y \varphi_y \varphi_t + \left(\lambda \varphi_y^2 + 1\right) a_t, \end{aligned}$$

and

$$\begin{aligned} D &= 1 + \lambda \left(\varphi_y^2 + \varphi_x^2\right) + \lambda \beta \left(\varphi_x a_y - \varphi_y a_x\right)^2 \\ &\quad + \beta \left(a_x^2 + a_y^2\right), \end{aligned}$$

and similarly for v_{N1}, v_{N2}, after exchanging x with y derivatives. The instantaneous frequency components are estimated using the AM-FM demodulation process described in section 6.

Instead of a 3D Gabor filterbank, we propose the use of a multiscale, separable 3D filterbank (see section 6.1). First, we take a 3D FFT of N frames of the input video. If we let $I(x, y, t)$ denote the input video, we then use $\tilde{I}(\omega_x, \omega_y, \omega_t)$ to denote its DFT. Similar to the 2D case, we zero-out all the frequency components that have negative ω_t. In other words, we set $\tilde{I}(\omega_x, \omega_y, \omega_t < 0) = 0$. Let $\hat{I}_{AS}$ denote the resulting video. Then, for each video, we compute the dominant AM-FM components over a three-scale 3D filterbank. To estimate the amplitude derivatives a_x, a_y, a_t, we take differences of Gaussian averages over 3x3 neighborhoods ($\sigma = 1$).

Using the derivative estimates, we iteratively compute velocity estimates at each iteration p. The relaxation algorithm is:

$$u^{(p+1)} = u_{ave}^{(p)} - \lambda \frac{u_{N1}}{D} \varphi_x - \beta \frac{u_{N2}}{D} a_x \tag{8.17}$$

and

$$v^{(p+1)} = v_{ave}^{(p)} - \lambda \frac{v_{N1}}{D} \varphi_y - \beta \frac{v_{N2}}{D} a_y. \tag{8.18}$$

We initialize the estimation by using zero-velocity estimates. The iterations are continued for both a prescribed number P of iterations and until further iterations do not change the solution. For example, for the u-component, we detect no change if $max_{(m,n)} \left| u^{(p+1)}(m,n) - u^{(p)}(m,n) \right| < TOL$, where TOL is a tolerance threshold.

Using the estimated motion vectors, we apply a Kalman filter to track the trajectories throughout the video. In summary, for the clinical videos, we establish the validity of the estimated trajectories based on: (i) the density of the estimated velocities, (ii) we require that trajectories remain valid throughout the video, (iii) estimation consistency and (iv) agreement with clinical expectations.

8.4 Results for Motion Estimation

We apply the developed AM-FM method for motion estimation using separable 3D filterbanks (see section 6.1). We use a dyadic decomposition with three-levels of decomposition. Furthermore, at each pixel, we compute motion estimates over the 3D channel that gives the maximum response (dominant component analysis). In this analysis, we did not use the VS-LQP. For IF estimation, the pixel spacing was fixed at a single pixel.

For the examples that we present we set the constraint parameters in (8.13) using:

- AM only: $\lambda = 0$, $\beta = 10$.
- FM only: $\lambda = 10$, $\beta = 0$.
- AM-FM: $\lambda = 10$, $\beta = 10$.

Here, we note that in (8.13), the continuity constraint is weighted by 1. Thus, in all of our experiments, we have placed a significantly more weight on the new AM-FM methods.

8.4.1 Results for a synthetic example

To test the implementations, we first present results using motion generated using a chirp image. A frame of this is shown in Fig. 8.2. The chirp image is given by

$$I(n_1, n_2) = \cos \varphi(n_1,\ n_2)$$

where

$$\varphi(n_1, n_2) = 2\pi \left[\alpha_1 n_1 + \beta_1 \frac{n_1^2}{2} + \alpha_2 n_2 + \beta_2 \frac{n_2^2}{2} \right], \tag{8.19}$$

$$\nabla\varphi(n_1, n_2) = 2\pi \left(\alpha_1 + \beta_1 n_1,\ \alpha_2 + \beta_2 n_2, \right) \tag{8.20}$$

with parameter values described in the caption of Fig. 8.2. Here, we note that the chirp image covers all the possible discrete values of the instantaneous frequency.

As documented in [93], we are interested in simulating periodic motion that closely resembles atherosclerotic plaque motion. This is accomplished by using:

$$x(t) = A_h \sin(\frac{2\pi}{N} f_h t) + \frac{A_h}{2} \sin(\frac{2\pi}{N}(2f_h)) + \frac{A_h}{3} \sin(\frac{2\pi}{N}(3f_h)) \tag{8.21}$$

and

$$y(t) = A_v \sin(\frac{2\pi}{N} f_v t) + \frac{A_v}{2} \sin(\frac{2\pi}{N}(2f_v)t) + \frac{A_v}{3} \sin(\frac{2\pi}{N}(3f_v)t). \tag{8.22}$$

As noted in [93], as measured in atherosclerotic videos, the harmonic amplitudes decay at a rate that is inversely proportional to the harmonic frequency. This is consistent with a discontinuity in the motion and has also been observed in the power spectra of the estimated trajectories.

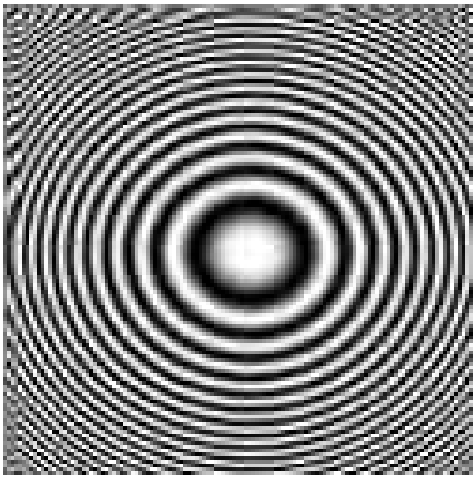


Figure 8.2: Chirp used for our first test in motion estimation. For the simulation, from (8.19), we use square images of size $N = 128$, $\alpha_1 = \alpha_2 = 9/40(N-1)$, and $\beta_1 = \beta_2 = 18/40(N-1)$. This gives an instantaneous frequency magnitude range interval of $[-9\pi/20, 9\pi/20]$ for both directions. For the motion, we set $A_h = A_v = 2$ and $f_h = f_v = 1$.

We compare the AM-FM results against optimal results obtained with Horn's method and also the phase-based method developed by Fleet and Jepson. We report the results in Table 8.1. For Horn, we compared against two different runs using a Lagrange multiplier of $\alpha = 10$, 20 for cases 1 and 2. In both cases, for the Gaussian smoothing parameter, we used $\sigma = 1.00$. For the Fleet and Jepson approach, to obtain a reasonable density, for the valid constraint: $\|\nabla\varphi - (k_i, w_i)\| \leq \tau\sigma_k$ and the amplitude constraint (see subsection 8.2.3), we used

- Case 1: $\sigma = 0.75$, $\text{Amp}_{threshold} = 1\%$, $\tau = 2$,
- Case 2: $\sigma = 1.00$, $\text{Amp}_{threshold} = 1\%$, $\tau = 2$,
- Case 3: $\sigma = 1.00$, $\text{Amp}_{threshold} = 5\%$, $\tau = 10$,
- Case 4: $\sigma = 1.25$, $\text{Amp}_{threshold} = 1\%$, $\tau = 2$.

In Table 8.1, we report the mean-square-error (MSE) for the magnitude of the esti-

mated velocity vectors and the density of the estimates.

Table 8.1: MSE and density in the velocity estimation of the synthetic chirp signal.

Method	**Motion Magnitude**	**Density**
Fleet (1)	Did not work	0%
Fleet (2)	0.0075	10.9851%
Fleet (3)	0.0140	33.4311%
Fleet (4)	0.0073	9.6752%
Horn (1)	0.0141	28.9235%
Horn (2)	0.0153	28.9235%
FM	0.0081	100%
AM	0.0090	100%
AM-FM	0.0085	100%

8.4.2 Results for atherosclerotic plaque videos

We present the first video frames in Fig. 8.3 and the tracked video frames in Figs. 8.4 and 8.5. For estimating the motion, we applied the AM, the FM, and the AM-FM method.

To verify tracking, we magnified the plaque regions that we wanted to analyze and carefully examined how each pixel is tracked through time. We also note that transducer motion can cause artificial plaque motion and this is something that we need to guard against. To avoid this problem we carefully examined the videos to confirm that: (i) we do not have significant intensity changes from frame to frame and (ii) there is a clear periodicity in the video frames, showing that there is no detectable drift between consecutive cardiac cycles. Now, having confirmed that there is no drift, we are led to believe that the motion estimates tended to be unbiased and thus any small, noisy variations were correctly filtered out by the Kalman filter trackers.

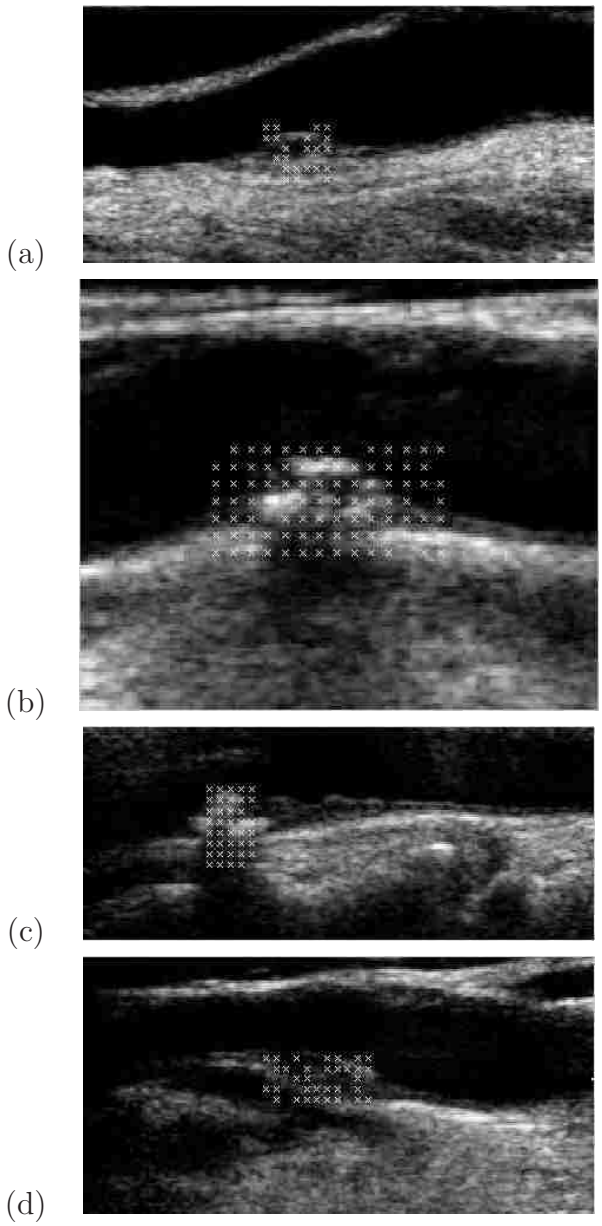


Figure 8.3: First frame of atherosclerotic plaque videos that were used. We display an 'x' over pixels where the Kalman filter was able to track throughout the video. The individual video characteristics are as follows: (a) video of size 125 rows $\times$ 250 columns $\times$ 67 frames, (b) video of size 125 rows $\times$ 150 columns $\times$ 111 frames, (c) video of size 100 rows $\times$ 240 columns $\times$ 87 frames, and (d) video of size 125 rows $\times$ 250 columns $\times$ 67 frames. We would like to thank Maura Griffin (PhD) and Ms. Niki Georgiou (S.R.N.) for selecting the case studies and carefully collecting the ultrasound videos that have collected the atherosclerotic videos for this study.

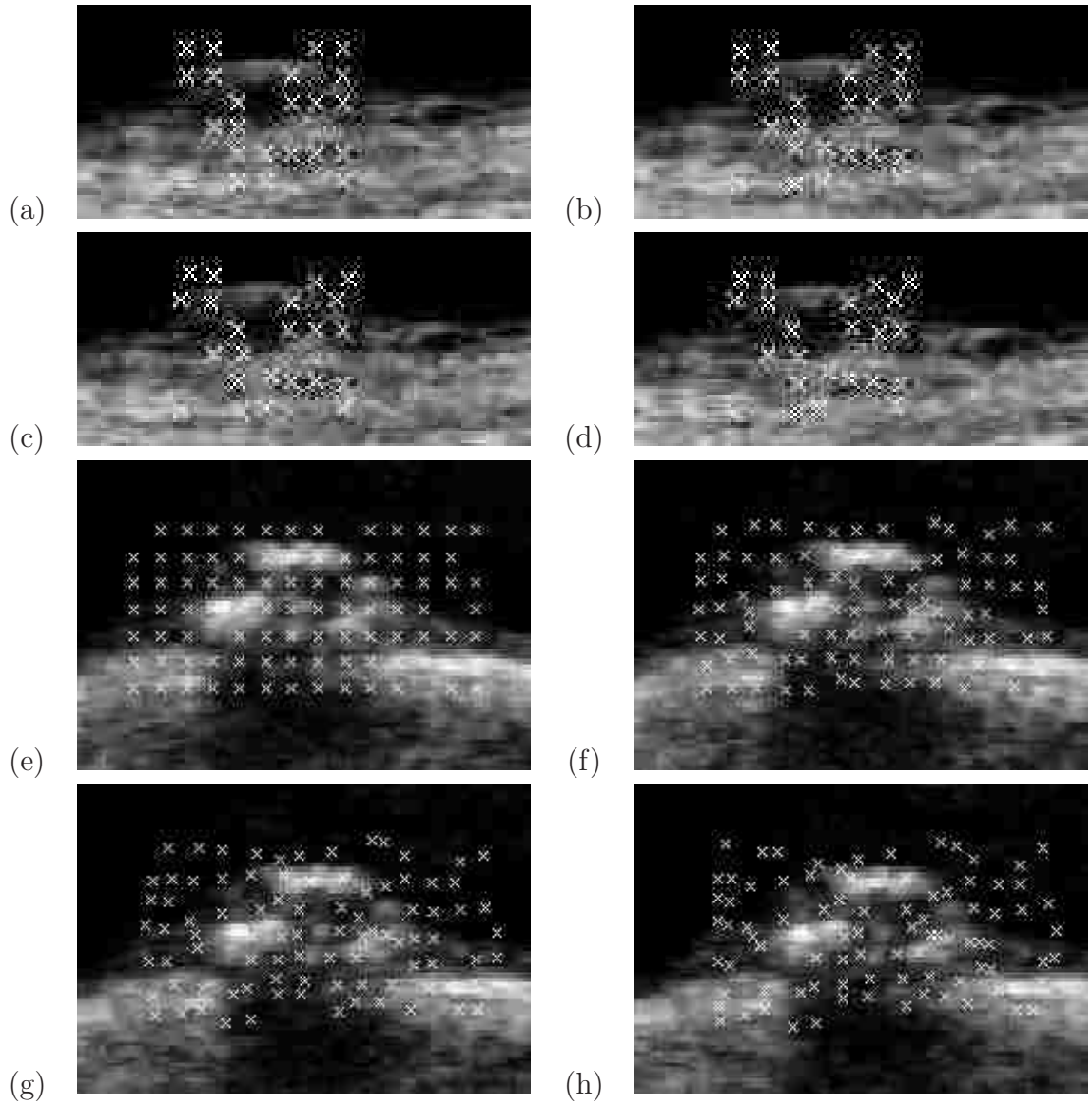


Figure 8.4: Pixel tracking results from frames 1, 21, 41 and 61 for each video from Figs. 8.3 (a)-(b). For the video shown in Fig. 8.3(a), we have extracted a Region of Interest (ROI) of $30-140 \times 30-210$. We then show frame 1 in (a), frame 21 in (b), frame 41 in (c), frame 61 in (d). Similarly, we show the same frames for the video shown in Fig. 8.3 (b) for ROI: $90-160 \times 120-270$ in (e)-(h).

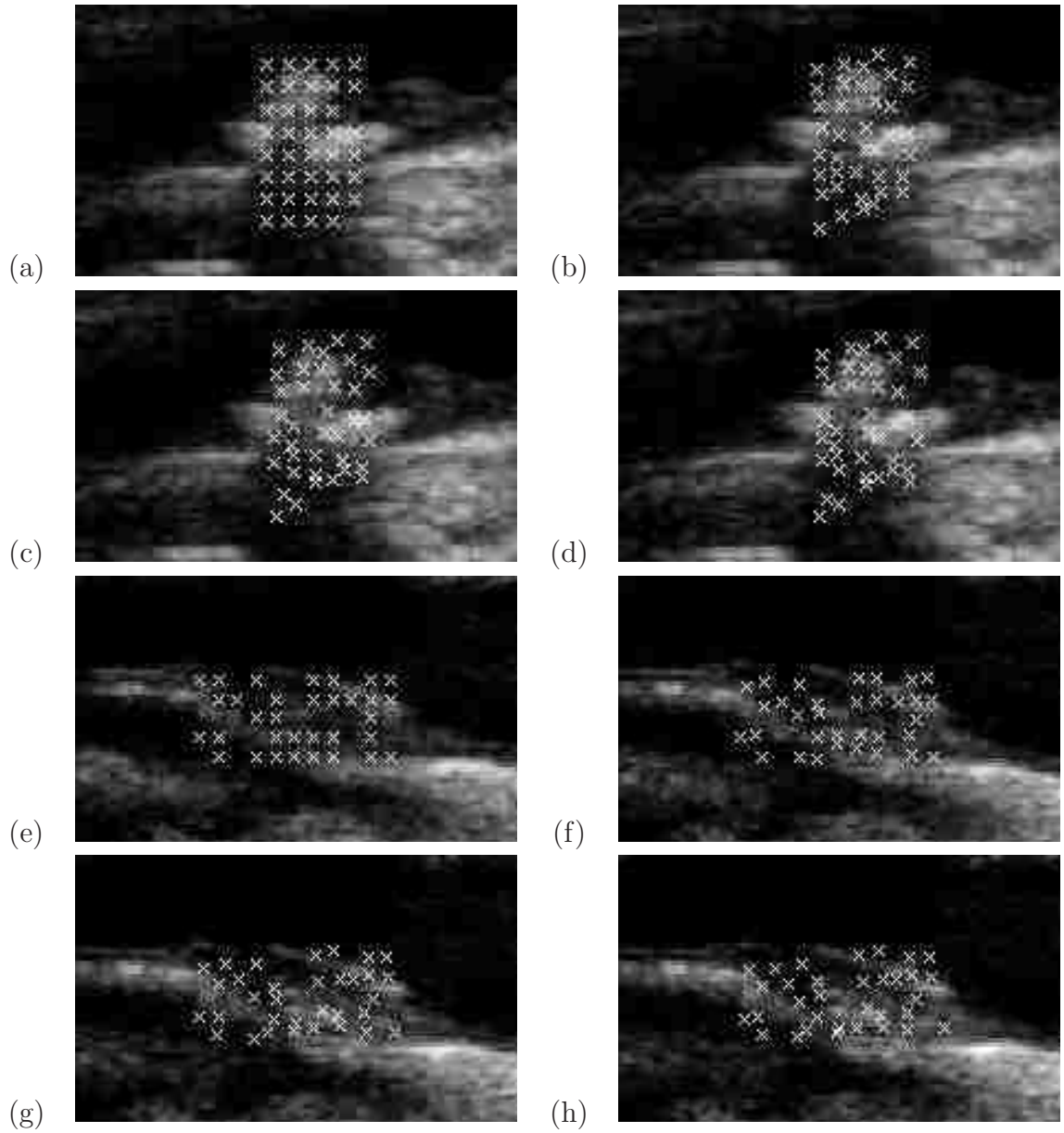


Figure 8.5: Pixel tracking results from frames 1, 21, 41 and 61 for each video from Figs. 8.3 (c)-(d). For the video showed in Fig. 8.3(c), we have extracted a Region of Interest (ROI) of $100 - 260 \times 80 - 315$. We then show frame 1 in (a), frame 21 in (b), frame 41 in (c), frame 61 in (d). Similarly, we show the same frames for the video of Fig. 8.3 (d) for ROI: $40 - 160 \times 100 - 300$ in (e)-(h).

8.4.3 Results for tracking of Taxi Video with AM-FM based Motion Estimation

In this sub-section, we present motion tracking results from the *taxi video* sequence (see Hamburg taxi video in [94]).

Reconstruction for taxi video: We start the analysis with a reconstruction of the video. We summarize video reconstruction results in Tables 8.2, 8.3, 8.4 and Fig. 8.6. We provide the MSE in Tables 8.2 and 8.3.

Tracking with AM-FM based ME: In Fig. 8.7 we show two different targets to demonstrate the independence among the AM, AM-FM and the FM estimates. We show a person tracking example in (a) and FM tracking results over the taxi region in (b).

(a) (b) (c)

Figure 8.6: Video reconstructions for the *taxi* video (we only show frame 1) using dominant components from each scale. (a) 3D Single-scale filterbank. (b) 3D Two-scale filterbank. (c) 3D Three-scale filterbank.

Table 8.2: MSE for the taxi video using 3D-MULTILES method.

Scales	**Single-scale**	**Two-scale**	**Three-scale**
0	34.7781	115.1141	186.2611
0, 1	34.5966	110.9505	185.2432
0, 1, 2	-	110.7431	180.1949
0, 1, 2, 3	-	-	179.9799

Table 8.3: MSE for the taxi video using 3D-LESHA method.

AM-FM Harmonics	**Single-scale**	**Two-scale**	**Three-scale**
1	34.7336	115.3749	202.3249
2	34.5546	115.3658	201.5240
3	34.2859	115.3632	197.0237
4	34.2842	115.2712	194.7118
5	34.2179	115.1972	194.6775

(a) (b)

Figure 8.7: Frame 1 of the *Hamburg taxi video* for tracking with the motion estimated during the video. (a) Zoom on the person. (b) Zoom on the taxi.

Table 8.4: Coefficients used for the taxi video reconstructions using 3D-MULTILES method.

Coefficients for AM-FM reconstructions using 3D-MULTILES.			
level	single-scale	two-scale	three-scale
0	0.9967	0.9999	1.0011
1	0.3459	1.0007	0.2205
2	-	0.3698	1.1030
3	-	-	0.3765

8.5 Discussion

For the synthetic video examples (sub-section 8.4.1), we note that the AM-FM approach provides reliable motion estimates at every pixel. In Table 8.1, this is demonstrated by the fact that AM, FM and AM-FM estimates have a density of 100%. On the other hand, the methods by Horn and Fleet and Jepson do not produce densities over 30%. This is attributed to the fact that the filter-banks used in the AM-FM approach cover the entire frequency plane.

In terms of accuracy, we note that the phase-based methods gave the best results. At 100% density, the proposed method achieved an accuracy that is comparable to what Fleet and Jepson achieved at 10%. It is also interesting to note that the AM method gave very good results and that the combined AM-FM method gave an accuracy that ranges between the AM and the FM methods. Here, we note that since the chirp has constant amplitude, the generated AM comes from processing the motion through the dyadic filter-bank.

For the atherosclerotic videos (sub-section 8.4.2) the best results were obtained by the AM method, closely followed by the full AM-FM and then the FM method. Apparently, ultrasound speckle dominated the FM while the AM captured the essential brightness variations without the noise. As a result, the Fleet and Jepson phased based method (similar to FM) was not able to track the pixel motion through the video. For Fleet and Jepson, the periodic motion resulted in stationary estimates

and the motion was not tracked throughout the periods. On the other hand, the optimized Horn's method appeared to track the pixels within the period, similar to the newly developed AM method. The correct pixel tracking is clearly visible in the examples in Fig. 8.4.

For tracking the person and the taxi in the *taxi* video, from Tables 8.2, 8.3 and 8.4, we note that the zeroth-scale refers to least-squares reconstruction using the low-pass filter output from each scale. This low-pass filter varies with scale as we detailed in the methods section. It is interesting to note the relative importance of the different scales as summarized in the optimal coefficients in Table 8.4. In Table 8.4, we note that using 3D-MULTILES, the second scale AM-FM component coefficient is equal to the low-pass filter component. As we shall see, the motion estimates from this level will prove very useful. On the other hand, the use of AM-FM harmonics (3D-LESHA method) did not reduce the MSE by any significant amount.

In Fig. 8.7 (a), both the AM and the dominant AM-FM component equations provide good trajectory tracking results. It is interesting to note that in this case, the person image is well-localized and its motion appears to be easier to follow.

From Fig. 8.7 (b), it is much more interesting to examine the FM tracking results over the taxi region. Tracking individual pixels over the taxi region proved to be much more challenging. To understand why, we simply note the uniform intensity regions over the surface of the taxi. Yet, the FM motion equation provided nice trajectory tracking results over the edges of the taxi images. We note that these nice tracking results could not have been reproduced with the AM motion equations.

In summary, we can see that AM motion estimation made a big difference and FM motion estimation will not work for small instantaneous frequencies.

8.6 Continuous-scale video search

In this section we present an AM-FM application in content-based video retrieval (CBVR) as an example of continuous-scale video search. We apply this CBVR retrieval to human activity recognition for the database given in [89].

Because of the fast growth of digital information, in this case of digital videos, searching video content is a challenging task. First approaches exploited low-level features such as colors, textures, shapes and lines orientation to retrieve the information. Later approaches look for matching of sketches in the objects of the video, analyzing the moving separately from the still components. Recently, research in CBVR has been focused on exploiting different approaches to recognize the video content. Some approaches are based on text-based video retrieval [95], others in matching trajectories [96, 97, 98] or learning methods [99]. The applications are in sports videos [100], broadcasting news [101], video surveillance [102] or in detecting human activity. We propose an AM-FM approach for retrieving videos of human activities.

8.6.1 Continuous-scale video search Dataset

For the database given in [89], we created two new versions from each video used: (i) a rotated version (30^o counterclockwise) and (ii) a downsampled (by 2) version of (i). Thus, each video is compared also versus its same version but rotated and also with a smaller size. We present examples of our video database in Fig. 8.8 for a *fight* video: (a) original video, (b) rotated video and (c) rotated and downsampled video. The video file sizes are mostly between 6 and 12 MB, a few up to 21 MB. A summary of the database that we used is also given in Table 8.7. We used five videos in total, corresponding to two different scenarios: groups of people meeting (3 videos) and people fighting (2 videos).

(a)

(b)

(c)

Figure 8.8: Examples of the three different versions for a *fight* video: (a) original video, (b) rotated video and (c) rotated and downsampled video. Frame 300 is shown.

8.6.2 Methods

The basic architecture of the system is shown in Fig. 8.9. We present the steps in greater detail in Fig. 8.10.

For this application we consider the use of motion velocity vectors computed using AM-FM methods.

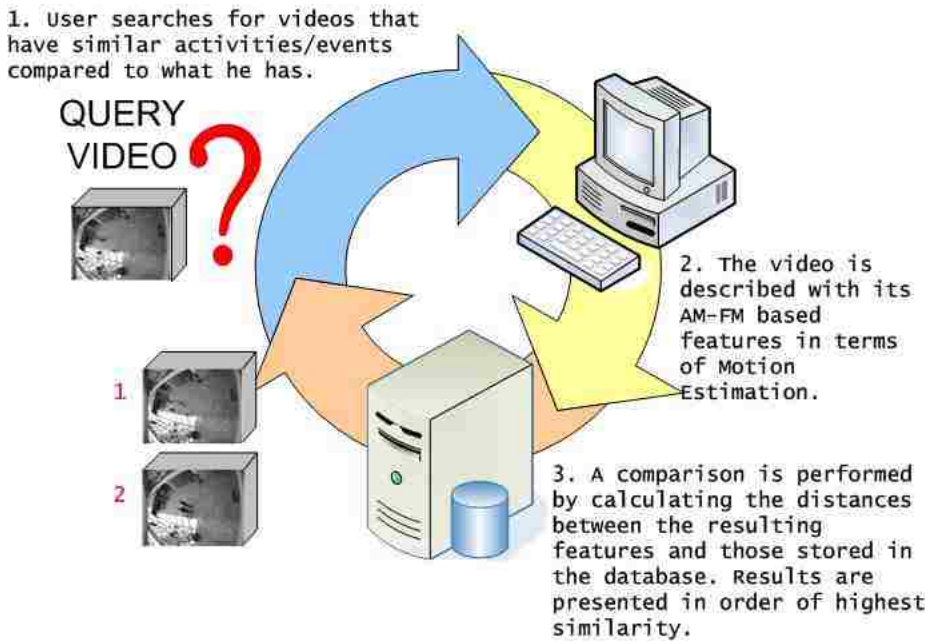


Figure 8.9: Content-based Video Retrieval System.

Table 8.5: An example demonstrating the computation of the table of distances.

	Video 1	**Video 2**	**Video 3**	**Video 4**
Video 1	0	distance{1,2}	distance{1,3}	distance{1,4}
Video 2	distance{2,1}	0	distance{2,3}	distance{2,4}
Video 3	distance{3,1}	distance{3,2}	0	distance{3,4}
Video 4	distance{4,1}	distance{4,2}	distance{4,3}	0

1. For an input video $I(k_1, k_2, k_3)$, we compute the AM-FM estimates IA and IF vectors using (6.5), (6.6), (6.9)-(6.14).
2. We compute the velocity vectors u and v using (8.17) and (8.18). Then, let V be the magnitude of the velocity vector given by $V = \sqrt{u^2 + v^2}$.
3. We compute the mean μ_V and the standard deviation σ_V of V and store them only for the frames were the action was presented. Note that since the number of frames where the action is presented differs from video to video, we only consider the central number frames with action such that the number of frames is the same for all the videos.
4. We create a feature vector for each video that contains the values μ_V and σ_V frame by frame.
5. We compute the distance among all the videos using **correlation** distance: given an $m-$by$-n$ matrix X, which is treated as m_{1xn} row vectors x_1, x_2, ..., x_m, the correlation distance between the vector x_r and x_s is defined as follows:

$$d_{rs} = 1 - \frac{(x_r - \bar{x}_r)(x_s - \bar{x}_s)^T}{\left[(x_r - \bar{x}_r)(x_r - \bar{x}_r)^T\right]^{\frac{1}{2}} \left[(x_s - \bar{x}_s)(x_s - \bar{x}_s)^T\right]^{\frac{1}{2}}}, \tag{8.23}$$

   where $\bar{x}_r = \frac{1}{n}\sum_j x_{rj}$ and $\bar{x}_s = \frac{1}{n}\sum_j x_{sj}$.
6. We create a table of distances as shown in Table 8.5.
7. We decide the number of videos relevant to the query using a k-nearest neighbor algorithm with $k = 3$.
8. The retrieved videos are presented in ascending order using the computed distances to the query video.

Figure 8.10: Steps for continuous scale video search using AM-FM features.

8.6.3 Results for Activity Recognition

In Table 8.7 we present the results for this approach (produced for a report to DARPA). In this case, we use five original videos (see sub-section 8.6.1), with their rotated and downsampled-rotated versions. The correct matches are shown underlined. Table 8.6 shows the numbering scheme used for the videos depending on the activity.

Table 8.6: Numeration for the videos used in Table 8.7.

	Video Type		
Activity	**Original**	**Rotated**	**Rotated and downsampled**
Fighting	1, 2	11, 12	21, 21
Meeting	3, 4, 5	13, 14, 15	23, 24, 25

Also, we compute the precision (p) and recall (r) rates as given in [103]:

$$p = \frac{\text{Number of correctly retrieved videos}}{\text{Number of videos returned by the method}},$$

and

$$r = \frac{\text{Number of correctly retrieved videos}}{\text{Number of videos relevant to the query}}.$$

Here, note that if the number of returned videos is the same as the number of relevant videos, we have that $p = r$. In our example, a k-nearest neighbor classifier ($k = 3$) was used to determine the number of videos that should be retrieved. Since the classifier always selected the correct cluster, we actually did get that $p = r$.

Also, in our example, for a *fight* video (see Table 8.6) $p = \{\text{number of correct retrieved videos}\}/5$ and for a *meeting* video $p = \{\text{number of correct retrieved videos}\}/8$.

8.6.4 Discussion for Activity Recognition

The approach considering σ_V and μ_V as the only features for the videos produced good results (see Table 8.7). Considering the eight retrieved videos shown, not only at least the first three retrieval were correct (perfect classification using k-nearest neighbor algorithm with $k = 3$) but also correct videos were retrieved in the other five positions. An average of 82% of precision obtained, with a maximum of 100% and a minimum of 60%. However, recall that the dataset used was small (15 videos with 2 activities).

Table 8.7: CBVR table for the first approach using 15 videos. Correct matches are shown underlined. The average precision is 82%.

Query video #	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	**Precision (%)**
1	11	12	2	3	13	23	21	5	60%
2	12	11	1	22	23	3	13	15	80%
3	13	23	14	4	25	5	15	11	88%
4	14	24	3	13	23	25	15	5	100%
5	15	25	3	13	23	12	14	11	75%
21	11	12	1	3	13	2	4	5	60%
22	2	12	11	1	21	3	13	23	100%
23	3	13	14	4	25	5	15	11	88%
24	14	4	3	13	23	25	15	5	100%
25	15	5	13	3	23	14	4	11	88%
11	1	12	2	3	13	23	21	5	60%
12	2	11	1	23	13	3	15	5	60%
13	3	23	14	4	25	15	5	11	88%
14	4	13	3	23	24	25	15	5	100%
15	5	25	13	3	23	14	12	4	88%

This approach does not need annotation and works without knowledge of the number of activities presented. Furthermore, it does not require a database with annotated features for the activities to compare. However, it is important to note that our video data-set was very small. Furthermore, the use of the k-nearest neighbor classifier required that we had prior knowledge on the number of videos that belong to each activity.

Part IV

Conclusions and Future Work

Chapter 9

Concluding Remarks and Future Work

We provide a brief summary of the new AM-FM methods and their applications in image and video processing. We also provide a short summary of proposed future work.

9.1 Summary of AM-FM methods in Image processing

The first step in developing new AM-FM methods was to design a new multi-scale filterbank. This design allowed us to have correct IF component sign estimation. The almost flat response in the bandpass frequency of the 1D filters eliminated errors due to the use of an amplitude correction as in the case of using Gabor filterbanks. The use of these filters in the AM-FM demodulation problem produced big improvements in the IA and IF estimations.

We developed a new method for robust instantaneous frequency estimation: VS-LQP (see section 3.4). For noisy signals, VS-LQP produced better results than previous methods such as QEA or QLM.

We developed new methods for reconstructing images based on their IA and IP information. Three least squares methods were presented: MULTILES, LESHA and LESCA (see chapter 4). These methods produced good image reconstructions and MULTILES showed that AM-FM components from different scales of the frequency spectrum contain important information that can be used to improve image quality in the reconstruction. Overall, we show that the system provides for accurate reconstructions of general images and show extracted AM-FM component parameters in several cases. Even though the AM-FM reconstructions were promising, our primary focus was on image analysis applications (as demonstrated in the case of *Lena*'s hair or the *fingerprint*).

For retinal image analysis (see chapter 5), we designed a new four-scale filterbank. For diabetic retinopathy (DR), we have developed an algorithm to detect the lesions' characteristics of DR in digital retinal images which have been graded (diagnosed) by retinal specialists. The algorithm can run without human supervision and will ultimately be used to identify those images suspected of containing diabetic retinopathy, macular degeneration, or other abnormalities. In this implementation, the computer-based screening system eliminates most (about 80 to 90%) of the normals, i.e., specificity, requiring the human experts to assess only the suspect images, reducing their workload significantly.

Clearly, all prior applications that were based on AM-FM demodulation can benefit from using the new filterbanks presented. Results using the new approach for AM-FM representations for the characterization of carotid plaque ultrasound images have been accepted in [104], where AM-FM representations provide new feature sets, which can be used for the classification of carotid plaques or other imaging

classification problems.

9.2 Summary of AM-FM methods in Video processing

As for 2D, the first step was to design new 3D multi-scale filterbanks with support in four octants of the frequency spectrum. Similar to the 2D case, the flat response in the bandpass frequency of the 1D filters eliminate errors due to the use of an amplitude correction as in the case of using Gabor filterbanks. We extended the 2D AM-FM formulation and theory to 3D.

We have shown video reconstructions using AM-FM from multiple-scales in three different forms based on the 2D methods. Even though these methods reconstructed the original videos within an acceptable quality, our focus was again concentrated on video analysis applications.

We presented new AM-FM methods for motion estimation, allowing us to estimate pixel motion (requiring two displacements vectors (u, v)) with up to three equations per pixel per scale (AM, FM, and continuity equations). The new filterbanks cover the entire frequency spectrum, allowing us to provide dense motion estimates. From the results on atherosclerotic plaque videos, we showed that the AM method can provide visibly correct tracking over the plaques.

The AM-FM based motion estimation method was used to characterize human activities in surveillance videos to retrieve videos. This content-based video retrieval (CBVR) method is unsupervised and does not require a learning processing or matching with known activities.

9.3 Future Work

The new AM-FM methods can benefit a wide range of applications. The retinal image analysis applications showed that AM-FM features can be used to characterize disease in eye images. The same approach can be used for applications such as detecting tumors in chest radiographs or for breast cancer (mammograms).

In the future, we will consider the implementation of AM-FM methods on an FPGA platform. This will enable us to explore real-time video processing applications.

The content based image/video retrieval applications presented in this dissertation are related with results that are based on subjective grading. The system needs to adapts its *hits* to the user. Thus, in future work, one can look for the best solution for the adaptation. A suggested solution is to follow the ideas of PageRank^{TM} [105, 106].

References

[1] M. S. Pattichis and A. C. Bovik, "Analyzing image structure by multidimensional frequency modulation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 5, pp. 753–766, 2007.

[2] M. S. Pattichis, A. C. Bovik, J. W. Havlicek, and N. D. Sidiropoulos, "Multidimensional orthogonal FM transforms," *IEEE Transactions on Image Processing*, vol. 10, no. 3, pp. 448–464, March 2001.

[3] J. P. Havlicek, D. S. Harding, and A. C. Bovik, "Multidimensional Quasi-Eigenfunction Approximations and multicomponent AM-FM models," *IEEE Transactions on Image Processing*, vol. 9, no. 2, pp. 227–242, February 2000.

[4] S. L. Hahn, "The analytic, quaternionic and monogenic 2-D complex delta distributions," Warsaw University Technology, Institute of Radioelectronics, Nowowiejska, report 3, April 2002.

[5] T. Bülow and G. Sommer, "Hypercomplex signals – a novel extension of the analytic signal to the multidimensional case," *IEEE Transactions on Signal Processing*, vol. 49, no. 11, pp. 2844–2852, November 2001.

[6] U. Köthe and M. Felsberg, "Riesz-Transforms vs. Derivatives: On the relationship between the boundary tensor and the energy tensor," in *Scale Space and PDE Methods in Computer Vision*, ser. LNCS, R. Kimmel, N. Sochen, and J. Weickert, Eds., vol. 3459. Springer, 2005, pp. 179–191.

[7] H. Knutsson, C. Westin, and G. Granlund, "Local multiscale frequency and bandwidth estimation," in *IEEE International Conference on Image Processing*, vol. 1, November 1994, pp. 36–40.

[8] M. Felsberg and U. Köthe, "GET: The connection between monogenic scale-space and gaussian derivatives," in *Scale-Space 2005, Springer LNCS*, vol. 3459, 2005, pp. 192–203.

[9] M. Felsberg and G. Sommer, "The Monogenic Signal," *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 3136–3144, December 2001.

[10] J. P. Havlicek, D. S. Harding, and A. C. Bovik, "Multicomponent multidimensional signals," *Multidimensional Systems and Signal Processing*, vol. 9, no. 4, pp. 391–398, 1998.

[11] D. Vakman, *Signals, Oscillations, and Waves: A Modern Approach*. Boston: Artech House, 1998.

[12] ——, "On the Analytic Signal, the Teager-Kaiser energy algorithm, and other methods for defining amplitude and frequency," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 791–797, April 1996.

[13] P. Rodriguez V., "Fast and accurate AM-FM demodulation of digital images with applications," Ph.D. dissertation, The University of New Mexico, 2005.

[14] B. Santhanam and P. Maragos, "Multicomponent AM-FM demodulation via periodicity-based algebraic separation and energy-based demodulation," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 473–490, March 2000.

[15] R. A. Baxter and T. F. Quatieri, "Shunting networks for multi-band AM-FM decomposition," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 1999, pp. 227–230.

[16] M. R. Morelande and A. M. Zoubir, "Model selection of random amplitude polynomial phase signals," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 578–589, March 2002.

[17] D. Dimitriadis and P. Maragos, "An improved energy demodulation algorithm using splines," in *IEEE Conf. on Acoustics, Speech, and Signal Processing*, May 2001.

[18] B. Santhanam, "Generalized energy demodulation for large frequency deviations and wideband signals," *IEEE Transactions on Signal Processing*, vol. 11, no. 3, pp. 341–344, March 2004.

[19] J. F. Kaiser, "On a simple algorithm to calculate the 'energy' of a signal," in *IEEE Conf. on Acoustics, Speech, and Signal Processing*, Albuquerque, NM, April 1990, pp. 381–384.

[20] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "On amplitude and frequency demodulation using energy operators," *IEEE Transactions on Signal Processing*, vol. 41, no. 4, pp. 1532–1550, April 1993.

[21] ——, "Energy separation in signal modulations with applications to speech analysis," *IEEE Transactions on Signal Processing*, vol. 41, no. 10, pp. 3024–3051, October 1993.

[22] A. Bovik, P. Maragos, and T. Quatieri, "AM-FM energy detection and separation in noise using multiband energy operators," *IEEE Transactions on Signal Processing*, vol. ASSP-41, no. 12, pp. 3245–3265, December 1993.

[23] G. Girolami and D. Vakman, "Instantaneous frequency estimation and measurement: a Quasi-Local Method," *Measurement Science Technology*, vol. 13, pp. 909–917, June 2002.

[24] D. Dimitriadis and P. Maragos, "Robust energy demodulation based on continuous models with application to speech recognition," in *European Conf. on Speech Communication. & Technology*, September 2003.

[25] D. Wei and A. C. Bovik, "On the instantaneous frequency of multicomponent AM-FM signals," *IEEE Signal Processing Letters*, vol. 5, no. 4, pp. 84–86, April 1998.

[26] S. Gazor and R. R. Far, "Adaptive maximum windowed likelihood multicomponent AM-FM signal decomposition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 479–491, 2006.

[27] F. Gianfelici, G. Biagetti, P. Crippa, and C. Turchetti, "Multicomponent AM-FM representations: An asymptotically exact approach," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, pp. 823–837, 2007.

[28] J. P. Havlicek, D. S. Harding, and A. C. Bovik, "The multi-component AM-FM image representation," *IEEE Transactions on Image Processing*, vol. 5, pp. 1094–1100, June 1996.

[29] B. J. Super and A. C. Bovik, "Shape from texture using local spectral moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 4, pp. 333–343, April 1995.

[30] S. T. Acton, D. P. Mukherjee, J. P. Havlicek, and A. C. Bovik, "Oriented texture completion by AM-FM Reaction-Diffusion," *IEEE Transactions on Image Processing*, vol. 10, no. 6, pp. 885–896, June 2001.

[31] M. Pattichis, G. Panayi, A. Bovik, and H. Shun-Pin, "Fingerprint Classification using an AM-FM model," *IEEE Transactions on Image Processing*, vol. 10, no. 6, pp. 951–954, June 2001.

[32] L. Cohen, *Time-Frequency Analysis*, ser. Prentice Hall Signal Processing, A. V. Oppenheim, Ed. Englewood Cliffs, New Jersey 07632: Prentice Hall PTR, 1995.

[33] J. Havlicek, J. Tang, S. Acton, R. Antonucci, and F. Quandji, "Modulation domain texture retrieval for CBIR in digital libraries," in *37th IEEE Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, November 2003.

[34] M. Pattichis, C. Pattichis, M. Avraam, A. Bovik, and K. Kyriakou, "AM-FM texture segmentation in electron microscopic muscle imaging," *IEEE Transactions on Medical Imaging*, vol. 19, no. 12, pp. 1253–1258, December 2000.

[35] T. Yap, J. Havlicek, and V. DeBrunner, "Bayesian segmentation of AM-FM texture images," in *35th IEEE Asilomar Conference on Signals, Systems and Computers*, vol. 2, Pacific Grove, CA, November 2001, pp. 1156–1160.

[36] N. Ray, J. Havlicek, S. T. Acton, and M. S. Pattichis, "Active contour segmentation guided by AM-FM Dominant Component Analysis," in *IEEE International Conference on Image Processing*, 2001, pp. 78–81.

[37] J. Havlicek, P. Tay, and A. Bovik, *Handbook of Image and Video Processing*. Elsevier Academic Press, 2005, ch. AM-FM Image Models: Fundamental Techniques and Emerging Trends, pp. 377–395.

[38] J. P. Havlicek and A. C. Bovik, "Multi-component AM-FM image models and wavelet-based demodulation with component tracking," in *IEEE International Conference on Image Processing*, 1994, pp. 41–45. [Online]. Available: citeseer.ist.psu.edu/531852.html

[39] V. Murray, P. Rodriguez V., and M. S. Pattichis, "Robust multiscale AM-FM demodulation of digital images," in *IEEE International Conference on Image Processing*, vol. 1, October 2007, pp. 465–468.

[40] R. A. Sivley and J. P. Havlicek, "Perfect reconstruction AM-FM image models," in *IEEE International Conference on Image Processing*, Oct 2006, pp. 2125–2128.

[41] K. Suri and J. P. Havlicek, "Phase algorithm for blocking artifact reduction in reconstructions from analysis-only AM-FM models," in *IEEE Southwest Symposium on Image Analysis and Interpretation*, 2006, pp. 6–10.

[42] J. P. Havlicek, D. S. Harding, and A. C. Bovik, "Reconstruction from the multi-component AM-FM image representation," in *IEEE International Conference on Image Processing*, vol. 2, Oct 1995, pp. 280–283.

[43] ——, "Multi-component signal demodulation and reconstruction using AM-FM modulation models," in *IEEE Workshop on Nonlinear Signal and Image Processing*, June 1995, pp. 41–45.

[44] R. A. Sivley and J. P. Havlicek, "A spline-based framework for perfect reconstruction AM-FM models," in *IEEE Southwest Symposium on Image Analysis and Interpretation*, 2006, pp. 198–202.

[45] J. P. Havlicek, D. S. Harding, and A. C. Bovik, "Computation of the multi-component AM-FM image representation," The University of Texas at Austin, Tech. Rep., 1996.

[46] P. Rodriguez V., M. Pattichis, and M. Goens, "M-mode echocardiography image and video segmentation based on AM-FM demodulation techniques," in *25th Intern. Conf. of the IEEE Engineering in Medicine and Biology Society*, vol. 2, September 2003, pp. 1176–1179.

[47] D. J. Fleet and A. D. Jepson, "Computation of component image velocity from local phase information," *Int. J. Comput. Vision*, vol. 5, no. 1, pp. 77–104, 1990.

[48] V. Murray and M. S. Pattichis, "AM-FM demodulation methods for reconstruction, analysis and motion estimation in video signals," in *IEEE Southwest Symposium on Image Analysis, and Interpretation*, March 2008, pp. 17–20.

[49] V. Murray, S. Murillo, M. Pattichis, C. Loizou, C. Pattichis, E. Kyriacou, and A. Nicolaides, "An AM-FM model for motion estimation in atherosclerotic plaque videos," in *41st IEEE Asilomar Conference on Signals, Systems and Computers*, 2007, pp. 746–750.

[50] J. P. Havlicek, "AM-FM image models," Ph.D. dissertation, The University of Texas at Austin, 1996. [Online]. Available: http://hotnsour.ou.edu/joebob/PdfPubs/JPHavlicekDiss.pdf

[51] J. Havlicek and A. Bovik, *Handbook of Image and Video Processing*, ser. Communications, Networking, and Multimedia. San Diego: Academic Press, 2000, ch. Image Modulation Models, pp. 305–316.

[52] M. Pattichis, J. Havlicek, S. Acton, and A. Bovik, *Advances in Image Processing and Understanding: A Festschrift for Thomas S. Huang*, ser. Machine Perception and Artificial Intelligence. Singapore: World Scientific Publishing, 2002, vol. 52, ch. Multidimensional AM-FM Models with Image Processing Applications, pp. 277–305.

[53] A. Bovik, N. Gopal, T. Emmoth, and A. Restrepo, "Localized measurement of emergent image frequencies by Gabor wavelets," *IEEE Transactions on Information Theory*, vol. IT-38, no. 3, pp. 691–712, March 1992.

[54] P. Maragos and A. Bovik, "Image amplitude and frequency demodulation using multidimensional energy separation," *Journal of the Optical Society of America*, vol. 12, no. 9, pp. 1867–1876, September 1995.

[55] M. S. Pattichis and A. C. Bovik, "A nonlinear fluid model for describing frequency modulation of image orientations," in *IEEE Workshop on Non-Linear Signal and Image Proc.*, Neos Marmaras, Halkidiki, Greece, June 20-22 1995, pp. 198–201.

[56] ——, "Multi-Dimensional frequency modulation in texture images," in *International Conference on Digital Signal Processing*, Limassol, Cyprus, June 26-28 1995, pp. 753–758.

[57] K. G. Larkin, D. J. Bone, and M. A. Oldfield, "Natural demodulation of two-dimensional fringe patterns. I. General background of the spiral phase quadrature transform," *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, vol. 18, no. 8, pp. 1862–1870, August 2001.

[58] K. G. Larkin, "Natural demodulation of two-dimensional fringe patterns. II. Stationary phase analysis of the spiral phase quadrature transform," *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, vol. 18, no. 8, pp. 1871–1881, August 2001.

[59] W. P. Torres and T. F. Quatieri, "Estimation of modulation based on fm-to-am transduction: Two-sinusoid case," *IEEE Transactions on Signal Processing*, vol. 47, no. 11, pp. 3084–3097, november 1999.

[60] A. C. Bovik, J. P. Havlicek, M. D. Desai, and D. S. Harding, "Limits on discrete modulated signals," *IEEE Transactions on Signal Processing*, vol. 45, no. 4, pp. 867–879, april 1997.

[61] W.-C. Pai and P. C. Doerschuk, "Statistical AM-FM models, extended Kalman filter demodulation, Cramer-Rao bounds, and speech analysis," *IEEE Transactions on Signal Processing*, vol. 48, no. 8, pp. 2300–2313, august 2000.

[62] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 411–426, 2007.

[63] C. Agurto, S. Murillo, V. Murray, M. Pattichis, S. Russell, M. Abramoff, and P. Soliz, "Detection and phenotyping of retinal disease using AM-FM processing for feature extraction," in *42nd IEEE Asilomar Conference on Signals, Systems and Computers*, 2008.

[64] V. Murray, M. Pattichis, and P. Soliz, "New AM-FM analysis methods for retinal image characterization," in *42nd IEEE Asilomar Conference on Signals, Systems and Computers*, 2008.

[65] V. Murray, P. Rodriguez V., and M. S. Pattichis, "Multi-scale image AM-FM demodulation methods," to be submitted to IEEE Transactions on Image Processing.

[66] E. Bedrosian, "The analytic signal representation of modulated waveforms," in *Proceedings of the Institute of Radio Engineers*, vol. 50, no. 10, Oct. 1962, pp. 2071–2076.

[67] A. Nuttall and E. Bedrosian, "On the quadrature approximation to the Hilbert transform of modulated signals," *Proceedings of the IEEE*, vol. 54, no. 10, pp. 1458–1459, Oct. 1966.

[68] E. Bedrosian, "A product theorem for Hilbert transforms," *Proceedings of the IEEE*, vol. 51, no. 5, pp. 868–869, May 1963.

[69] H. Stark, "An extension of the Hilbert transform product theorem," *Proceedings of the IEEE*, vol. 59, no. 9, pp. 1359–1360, September 1971.

[70] E. Bedrosian and H. Stark, "Comments on "an extension of the hilbert transform product theorem"," *Proceedings of the IEEE*, vol. 60, no. 2, pp. 228–229, February 1972.

[71] B. Picinbono, "On instantaneous amplitude and phase of signals," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 552–560, March 1997.

[72] ——, "Some remarks on instantaneous amplitude and frequency of signals," in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, Oct 1998, pp. 293–300.

[73] G. J. Hämmerlin and K. H. Hoffmann, *Numerical Mathematics*, 1st ed., ser. Undergraduate Texts in Mathematics Series. Springer Verlag, 1991, no. 1.

[74] J. W. Demmel, *Applied Numerical Linear Algebra*. SIAM, 1997, ch. Linear Least Square Problems, pp. 105–117.

[75] Signal and I. P. I. of the University of Southern California, "The USC-SIPI image database." [Online]. Available: http://sipi.usc.edu/database/

[76] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, March 2002.

[77] TECHNO-VISION Project, "MESSIDOR: methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology." [Online]. Available: http://messidor.crihan.fr/

[78] M. Larsen, J. Godt, N. Larsen, H. Lund-Andersen, A. K. Sjølie, E. Agardh, H. Kalm, M. Grunkin, and D. R. Owens, "Automated detection of fundus photographic red lesions in diabetic retinopathy," *Invest Ophthalmol Vis Sci*, vol. 44, pp. 761–766, February 2003.

[79] N. Larsen, J. Godt, M. Grunkin, H. Lund-Andersen, and M. Larsen, "Automated detection of diabetic retinopathy in a fundus photographic screening population," *Invest Ophthalmol Vis Sci*, vol. 44, no. 2, pp. 767–771, February 2003.

[80] M. D. Abramoff, M. Niemeijer, M. S. Suttorp-Schulten, M. A. Viergever, S. R. Russell, and B. van Ginneken, "Safe automated detection of diabetic retinopathy," *Diabetes Care 31*, 2008.

[81] M. Niemeijer, B. van Ginneken, S. R. Russell, M. S. A. Suttorp-Schulten, and M. D. Abramoff, "Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis." *Investigative Ophthalmology & Visual Science*, vol. 48, pp. 2260–2267, May 2007.

[82] M. Niemeijer, B. van Ginneken, J. Staal, M. Suttorp-Schulten, and M. D. Abramoff, "Automatic detection of red lesions in digital color fundus photographs," *IEEE Transactions on Medical Imaging*, vol. 24, no. 5, pp. 584–592, May 2005.

[83] M. D. Abramoff, B. van Ginneken, M. S. Suttorp-Schulten, S. R. Russell, and M. Niemeijer, "Improved computer aided detection of diabetic retinopathy evaluated on 10,000 screening exams," *Invest Ophthalm Vis Sci*, vol. 47, 2008, aRVO E-Abstract 2735.

[84] S. C. Lee, E. T. Lee, R. M. Kingsley, Y. Wang, D. Russell, R. Klein, and A. Warn, "Comparison of diagnosis of early retinal lesions of diabetic retinopathy between a computer system and human experts," *Arch Ophthalmol*, vol. 119, pp. 509–515, April 2001.

[85] G. G. Gardner, D. Keating, T. H. Williamson, and A. T. Elliott, "Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool," *Br J Ophthalmol*, vol. 80, pp. 940–944, November 1996.

[86] H. Martens and T. Naes, *Multivariate Calibration*, 1st ed. Wiley, July 1992.

[87] I. Zimmer-Galler and R. Zeimer, "Results of implementation of the digiscope for diabetic retinopathy assessment in the primary care environment," *Telemedicine and e-Health*, vol. 12, pp. 89–98, April 2006.

[88] M. D. Abramoff and M. S. Suttorp-Schulten, "Web-based screening for diabetic retinopathy in a primary care population: the EyeCheck project," *Telemedicine and e-Health*, vol. 11, pp. 668–674, December 2005.

[89] Context Aware Vision using Image-based Active Recognition, "Benchmark data for PETS-ECCV 2004." [Online]. Available: http://www-prima.imag.fr/PETS04/caviar_data.html

[90] S. Meairs and M. Hennerici, "Four-dimensional ultrasonographic characterization of plaque surface motion in patients with symptomatic and asymptomatic carotid artery stenosis," *Stroke*, vol. 30, no. 9, pp. 1807–1813, Sep. 1999.

[91] K. Imoto, T. Hiro, T. Fujii, A. Murashige, Y. Fukumoto, G. Hashimoto, T. Okamura, J. Yamada, K. Mori, and M. Matsuzaki, "Longitudinal structural determinants of atherosclerotic plaque vulnerability a computational analysis of stress distribution using vessel models and three-dimensional intravascular ultrasound imaging," *J. Am. Coll. Cardiol.*, vol. 46, pp. 1507–1515, Sept. 2005.

[92] S. Murillo, M. Pattichis, C. Loizou, C. Pattichis, E. Kyriacou, A. Constantinides, and A. Nicolaides, "Atherosclerotic plaque motion analysis from ultrasound videos," *Signals, Systems and Computers, 2006. ACSSC '06. Fortieth Asilomar Conference on*, pp. 836–840, 2006.

[93] S. Murillo, M. Pattichis, C. Loizou, C. Pattichis, and E. Kyriakou, "Atherosclerotic plaque motion trajectory analysis from ultrasound videos," in *IEEE Int. Special Topic Conf. on Inform. Technology in Biomedicine*, Ioannina, Greece, Oct. 2006, pp. 1–5.

[94] J. Barron, D. Fleet, and S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.

[95] T. Volkmer and A. Natsev, "Exploring automatic query refinement for text-based video retrieval," in *IEEE International Conference on Multimedia and Expo*, July 2006, pp. 765–768.

[96] J.-W. Hsieh, S.-L. Yu, and Y.-S. Chen, "Motion-based video retrieval by trajectory matching," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 3, pp. 396–409, March 2006.

[97] C.-W. Su, H.-Y. Liao, H.-R. Tyan, C.-W. Lin, D.-Y. Chen, and K.-C. Fan, "Motion flow-based video retrieval," *IEEE Transactions on Multimedia*, vol. 9, no. 6, pp. 1193–1201, October 2007.

[98] T.-J. Chin, L. Wang, K. Schindler, and D. Suter, "Extrapolating learned manifolds for human activity recognition," in *IEEE International Conference on Image Processing*, vol. 1, October 2007, pp. 381–384.

[99] C. Snoek, M. Worring, D. C. Koelma, and A. Smeulders, "A learned lexicon-driven paradigm for interactive video retrieval," *IEEE Transactions on Multimedia*, vol. 9, no. 2, pp. 280–292, February 2007.

[100] L. Huayong and Z. Hui, "A content-based broadcasted sports video retrieval system using multiple modalities: SportBR," in *The Fifth International Conference on Computer and Information Technology*, September 2005, pp. 652–656.

[101] A. Hauptmann, R. Yan, W.-H. Lin, M. Christel, and H. Wactlar, "Can high-level concepts fill the semantic gap in video retrieval? a case study with broadcast news," *IEEE Transactions on Multimedia*, vol. 9, no. 5, pp. 958–966, August 2007.

[102] W. Hu, D. Xie, Z. Fu, W. Zeng, and S. Maybank, "Semantic-based surveillance video retrieval," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 1168–1181, April 2007.

[103] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, December 2000.

[104] C. Christodoulou, C. Pattichis, V. Murray, M. Pattichis, and A. Nicolaides, "AM-FM representations for the characterization of carotid plaque ultrasound images," in *European Congress for Medical and Biological Engineering*, 2008.

[105] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Computer Networks and ISDN Systems*, vol. 30, no. 1-7, pp. 107–117, 1998.

[106] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the Web," Stanford Digital Library Technologies Project, Tech. Rep., 1998. [Online]. Available: http://citeseer.ist.psu.edu/page98pagerank.html