

1-1-2003

# An optimized SVM kernel for texture classification and its application in microcalcification detection

Mahdi Sabri  
*Ryerson University*

Follow this and additional works at: <http://digitalcommons.ryerson.ca/dissertations>



Part of the [Electrical and Computer Engineering Commons](#)

---

## Recommended Citation

Sabri, Mahdi, "An optimized SVM kernel for texture classification and its application in microcalcification detection" (2003). *Theses and dissertations*. Paper 196.

This Thesis is brought to you for free and open access by Digital Commons @ Ryerson. It has been accepted for inclusion in Theses and dissertations by an authorized administrator of Digital Commons @ Ryerson. For more information, please contact [bcameron@ryerson.ca](mailto:bcameron@ryerson.ca).

# AN OPTIMIZED SVM KERNEL FOR TEXTURE CLASSIFICATION AND ITS APPLICATION IN MICROCALCIFICATION DETECTION

By  
Mahdi Sabri  
Bachelor of Electrical Engineering  
Sharif University of Thecnology  
Tehran, Iran

A thesis  
submitted to Ryerson University  
in Partial fulfilment of the  
requirement for the degree of  
Master of Applied Science  
in the program of  
Electrical and Computer Engineering.

Toronto, Ontario, Canada, 2003

©Mahdi Sabri 2003

MAHDI SABRI  
RYERSON UNIVERSITY LIBRARY

UMI Number: EC52894

## INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

**UMI**<sup>®</sup>

---

UMI Microform EC52894

Copyright 2008 by ProQuest LLC.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest LLC  
789 E. Eisenhower Parkway  
PO Box 1346  
Ann Arbor, MI 48106-1346

## Acknowledgment

I would like to express my gratitude to my supervisor Prof. Javad Alirezaie. Not only this work, but also my higher education was not possible without his support. I am also thankful to Prof. Krishnan for introducing me to the field of statistical signal processing and helpful discussion during the course of my studies at Ryerson University. I would like to thank the thesis exam committee members for their valuable comments and discussions.

I would like to acknowledge Prof. Manuel Davy and Dr. Kim for useful discussion and provided codes.

I would like to thank my parents who raised me in a way that I always believe that knowledge is a value that cannot be replaced even by great wealth.

Also to my wife who highly supported me and patiently accompanied me through all the difficulties we had during my MASc education.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	A CAD system based on texture features . . . . .	2
1.3	Summary of contributions . . . . .	3
1.4	Thesis outlines . . . . .	4
<b>2</b>	<b>SVM Learning Classifier</b>	<b>5</b>
2.1	Linear SVM Classifier . . . . .	5
2.2	Non-Linear SVM Classifier . . . . .	8
<b>3</b>	<b>Time series and Time-frequency features</b>	<b>12</b>
3.1	Local discriminant basis . . . . .	12
3.1.1	Wavelet and Wavelet Packet Transforms . . . . .	13
3.1.2	LDB Algorithm . . . . .	17
3.2	Linear predictive coding . . . . .	20
3.2.1	The LPC Model . . . . .	20
3.2.2	LPC analysis equation . . . . .	22
3.2.3	The Autocorrelation method . . . . .	24
3.3	Cohen's Space-Frequency signal representation . . . . .	28

<b>4</b>	<b>A Novel SVM Kernel for Texture Classification</b>	<b>30</b>
4.1	Motivation for texture classification . . . . .	30
4.2	Applications . . . . .	30
4.3	Texture Classification Challenges and Previous works . . . . .	31
4.3.1	Feature Extraction . . . . .	31
4.3.2	Classification . . . . .	32
4.4	External features to build a new kernel for SVM . . . . .	33
4.5	Classification and segmentation Results . . . . .	35
4.5.1	Comparison with Original SVM method . . . . .	36
4.5.2	Comparison with Other Classification methods . . . . .	37
4.5.3	Segmentation Results and Comparison . . . . .	38
4.6	Gabor Filter Banks . . . . .	43
<b>5</b>	<b>Detection of Abnormalities In Mammograms</b>	<b>45</b>
5.1	Motivation for a CAD system . . . . .	46
5.2	Previous work . . . . .	48
5.2.1	Shape Analysis . . . . .	48
5.2.2	Iterative feature extraction . . . . .	48
5.2.3	Thresholding and morphology . . . . .	49
5.2.4	Multiscale matched filters . . . . .	49
5.2.5	Statical feature of surrounding region . . . . .	50
5.2.6	Laplacian scale-space . . . . .	51
5.2.7	Spatial statistical features and wavelet features . . . . .	51
5.2.8	Support Vector Machine . . . . .	51
5.2.9	Multi-resolution Based Segmentation . . . . .	52

5.3	Proposed Algorithm . . . . .	55
5.3.1	Image segmentation . . . . .	56
5.3.2	Enhancement of the mammograms . . . . .	57
5.3.3	Preprocessing . . . . .	59
5.3.4	Reduction of False positives through supervised learning . . . . .	65
5.4	Experimental Results . . . . .	66
5.4.1	Database . . . . .	66
5.4.2	Selected Data Set . . . . .	69
5.4.3	Results . . . . .	71
<b>6</b>	<b>Conclusion and future work</b>	<b>75</b>
6.1	SVM in texture classification and external texture features . . . . .	75
6.2	Incorporating the Gabor filter Bank to Increase the performance of LPC-SVM texture classification algorithm . . . . .	76
6.3	Microcalcification detection . . . . .	76
6.4	Research contribution . . . . .	78
6.5	Future work . . . . .	78
	<b>Vita</b>	<b>89</b>

# List of Figures

3.1	Wavelet and Sine wave . . . . .	14
3.2	Computing wavelet coefficients via subband coding method . . . . .	15
3.3	Signal Decomposition to wavelet coefficients . . . . .	15
3.4	Generalization of subband coding for 2D Signal . . . . .	16
3.5	Wavelet Packet transform: An extension to wavelet transform . . . . .	16
3.6	Wavelet Packet decomposition (quad-tree for image signal) . . . . .	17
3.7	LDB best basis selection scheme for a two class(A,B) problem . . . . .	19
3.8	Linear Prediction Coding Model . . . . .	20
3.9	Signal spectra is compared with its estimated spectra using LPC of order a) 6 b) 12 . . . . .	26
3.10	Spectral estimation of sample speech signal, with a LPC of order a) 24 b) 48 . . . . .	27
4.1	Average VC-Dimension for three kernel functions based on the three feature extraction reviewed in chapter 3 . . . . .	36
4.2	Two-texture images used in experiments (a) D4, D84 (b) D5, D92( Brodatz album) . . . . .	38
4.3	multitexture images used in experiments . . . . .	39



4.4	multitexture images segmentation:(a) Five texture Brodatz image originally published by Jain and Farrokhnia(b) segmentation with LPC-SVM before post processing (c) segmentation with LPC-SVM after post processing (d) segmentation using Optimized Gabor Filter (e)segmentation using Optimized Gabor Filter (sigmoidal activation) . . . . .	42
5.1	Sample of microcalcification in a mammogram. The microcalcification is shown in the right after zooming and enhancement. . . . .	47
5.2	Overall flow of the developed algorithm for detection of microcalcification . . . . .	55
5.3	Preprocessing steps . . . . .	60
5.4	Orthogonal subbands at different resolutions . . . . .	61
5.5	Mother wavelet for DAUB 4, a narrow and tall wavelet for analyzing high-frequency characteristics . . . . .	63
5.6	Mother wavelet for DAUB 20,(a) a wide and short wavelet for analyzing low-frequency characteristics(b) a narrow and tall wavelet for analyzing high-frequency characteristics . . . . .	64
5.7	Sample ics file . . . . .	67
5.8	Chain codes values and directions . . . . .	68
5.9	Sample Overlay File . . . . .	68
5.10	ROC curves show that the proposed algorithm outperforms SVM . . . . .	72
5.11	ROC curve for the most recent published research on DDSM . . . . .	73

# List of Tables

4.1	Brief comparison between SVM and NN . . . . .	33
4.2	Source of test images . . . . .	38
4.3	Error Rates (percent) for two-texture image (Fig.4.2a). . . . .	39
4.4	Error Rates (percent) for two-texture image(Fig.4.2b). . . . .	39
4.5	Error Rates (percent) for multitexture image(Fig4.3.a). . . . .	40
4.6	Error Rates (percent) for multitexture image(Fig4.3.b). . . . .	40
4.7	Comparison of LPC-SVM with SGLDM,FPS,TWT,LAWS, GABOR method	40
4.8	Comparison of Correct Classification Rate in Logic Operator Method and LPC-SVM . . . . .	41
4.9	Comparison of Correct Classification Rate in Wavelet Transform Method and LPC-SVM . . . . .	41
4.10	Error Rates (percent) for two-texture image (Fig.4.2a). . . . .	44
5.1	Rate of Breast Cancer mortality in some developed countries. . . . .	45
5.2	Summary of related research in calcification detection . . . . .	54
5.3	Coefficients of the DAUB filters. . . . .	62
5.4	Summary of ground truth information for the test images . . . . .	70
5.5	The number of false positives and correct true positives in each test image .	74

# Chapter 1

## Introduction

### 1.1 Motivation

THE most frequently diagnosed cancer for Canadian women is by far breast cancer. This is applicable to the most of the developed countries. An average Canadian woman has a 1 in 9 chance to develop breast cancer in her life time. Each day more than 10 women die in Canada from breast cancer. Because the Canada population is aging, the number of women who will die from breast cancer will increase [1].

There is no cure for breast cancer. The only effective way to save the patient life is early detection and removing cancerous cells. X-ray mammography is a very important tool and the most commonly used method for early detection. For women aged over 40 years old, screening mammography is recommended once a year . This result in huge number of mammograms that they need to be examined and interpreted by trained radiologists.

Unfortunately, mammograms are one of the most difficult medical images to interpret. Specifically, visual assessment of microcalcification turns out to be a challenging task. Approximately 25% of the cancers that are visible on retrospective review are failed to detect by radiologists [2].

Thus, computer aided diagnosis (CAD) systems are motivated to help the radiologists with this tedious task. In fact, the goal of a CAD systems is not to create a high-tech radiologist or automate the detection procedure. CAD systems are intended to act as the first cut in the process and a second opinion for the radiologists. The term "second Opinion" means that the radiologists can use the results of a computer analysis of the mammogram in making a diagnosis. A CAD system tries to locate suspicious regions in the mammogram for more detailed examination by the radiologists.

## 1.2 A CAD system based on texture features

Over last two decades several approaches have been used by image processing scientists and researchers to develop a CAD system for microcalcification detection. A series of different methods and research results are briefly addressed in chapter five. Texture features are important and commonly used. It is not only used in breast cancer detection, but also is extensively used in other applications which deal with abnormality detection in medical images. Any texture classification method involves two major steps:

- **Feature Extraction**
- **Classification**

In this thesis both concepts are studied and a new method is proposed. Researchers have found the feature extraction more challenging and numerous different methods have been proposed and applied. Recently, Kim *et. al.* [3] showed the effectiveness of the Support Vector Machine (SVM) in texture classification and its advancement over neural networks. Since SVM performs well in high dimensional space such as the space spanned by texture images, they successfully employed SVM without any external features. In fact, the kernel function in SVM algorithm, implicitly performs feature extraction. Since SVM is basically

suited for two-class classification problems, it is potentially a good choice for several different medical imaging which deal with abnormality detection.

The main contribution of this thesis in the sense of texture classification is proposing a new texture classification algorithm by effectively employing external features within SVM kernel and introducing a new feature extraction method for texture classification.

### 1.3 Summary of contributions

- Developing a new texture classification algorithm by proposing a new SVM kernel which incorporates external features.
- Proposing an extension to aforementioned algorithm by using Gabor filter banks to enrich the external features, for the case of highly non-stationary signal and image classification.
- Achieving promising microcalcification rate by applying the new texture classification algorithm.
- Overcoming the asymmetry of training data in mammograms by applying a wavelet based pre-process stage to detect suspicious sites.
- Discussing the clinically assessed difficulty level of mammograms in interpreting of classification rates.

## 1.4 Thesis outlines

SVM is reviewed in Chapter two. Three feature extraction method are considered to be used within SVM kernel. These feature algorithm are presented in Chapter three. In Chapter four, texture classification challenges are discussed and previous works are cited. The three feature extraction algorithm presented in Chapter three, are compared by their estimated VC-dimension. LPC has been found superior and is used within a new kernel for SVM. The new texture classification algorithm is compared with several traditional texture classifications and some recently reported novel methods. In Chapter five, previous works in microcalcification detection are addressed and the new algorithm is applied to mammograms following a pre-process stage. Chapter six is dedicated to conclusion and intended future works.

## Chapter 2

# SVM Learning Classifier

**S**UPPORT vector machine is basically proposed for the case of binary classification. The main idea of support vector machine is to construct a hyperplane as the decision surface in such a way that the margin of separation between positive and negative examples is maximized. This goal is motivated by principles of statistical learning theory and the method of structural risk minimization [4], [5]. Indeed, the support vector machine is an approximate implementation of the structural risk minimization. According to the statistical learning theory, the error rate of a learning machine on the test data (i.e. the generalization error rate) is bounded by the sum of the training-error rate and a term that depends on Vapnik-Chervonenkis(VC) dimension [4]. In the case of separable patterns, SVM produces a value of zero for the first term and minimize the second term. Accordingly, The optimal hyperplane which is sought by SVM is equivalent to minimum bound on the VC-dimension. In the following, we first review the formulation of SVM in the simple case of separable patterns and then we discuss the case of non-separable case.

### 2.1 Linear SVM Classifier

Let vector  $\mathbf{x} \in \chi$  denote a pattern to be classified, and let scalar  $y$  denotes its class label  $y \in \{\pm 1\}$ . In addition, let  $\{(\mathbf{x}_i, y_i), i = 1, 2, \dots, l\}$  denote a given set of  $l$  training examples.

The problem is how to construct a classifier [i.e., a decision function  $f(x)$ ] that can correctly classify an input pattern that is not necessarily from the training set. On the assumption of linearly separable case, there exist a linear function of the form

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (2.1)$$

Such that for each training example  $\mathbf{x}_i$ , the function yields  $f(\mathbf{x}_i) \geq 0$  for  $y_i = +1$ , and  $f(\mathbf{x}_i) < 0$  for  $y_i = -1$ . In other words, training examples from the two different classes are separated by the hyperplane  $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = 0$ .

For a given training set, while there may exist many hyperplane that separate the two classes, the SVM classifier is based on the hyperplane that maximizes the separating margin between the two classes. In other words, SVM finds the hyperplane that causes the largest separation between the decision function values for the borderline members of the two classes referred as support vectors. This hyperplane can mathematically be found by minimizing the following cost function:

$$\Psi(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} = \frac{1}{2} \|\mathbf{w}\|^2 \quad (2.2)$$

Subject to separability constraints:

$$\mathbf{w}^T \mathbf{x}_i + b \geq +1, \quad \text{for } y_i = +1$$

and

$$\begin{aligned} \mathbf{w}^T \mathbf{x}_i + b &\leq -1, \quad \text{for } y_i = -1 \\ i &= 1, 2, \dots, l. \end{aligned} \quad (2.3)$$



Equivalently, these constraints can be written more compactly as

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad \text{for } i = 1, 2, \dots, l. \quad (2.4)$$

This specific problem formulation may not be useful in practice because the training data may not be completely separable by a hyperplane. In this case, slack variables, denoted by  $\xi_i$ , can be introduced to relax the separability constraints in (2.4) as follows:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0; i = 1, 2, \dots, l. \quad (2.5)$$

Accordingly, the cost function in (2.2) can be modified as follows:

$$\Psi(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \quad (2.6)$$

where  $C$  is a user-specified, positive, regularization parameter. In (2.6), the variable  $\xi$  is a vector containing all the slack variables  $\xi_i, i = 1, 2, \dots, l$ .

The modified cost function in (2.6) constitutes the so-called **structural risk**, which balances **empirical risk** (i.e., the training error reflected by second term) with **model complexity** (the first term) [6]. The regularization parameter  $C$  controls this trade-off. The purpose of using model complexity to constrain the optimization of empirical risk is to avoid overfitting, a situation in which the decision boundary too precisely corresponds to the training data, and thereby fails to perform well on data outside the training set.

## 2.2 Non-Linear SVM Classifier

In practice, input patterns are unlikely to be linearly separable. Non-linear SVM is motivated by Cover's theorem [7] on the separability of patterns. The theorem states that such a linearly non-separable pattern space can be transformed into a new feature space where patterns are linearly separable with a high probability. According to the theorem, the transform must be non-linear and the dimension of feature space must be high enough.

Let  $\mathbf{x}$  denote a vector drawn from the input space  $\chi$ , assumed to be of dimension  $m_0$ . Let  $\{\Phi_j(\mathbf{x})\}_{j=1}^{m_1}$  denote a set of non-linear transformations from the input space to the feature space of dimension  $m_1$ . It is assumed that  $\{\Phi_j(\mathbf{x})\}$  is defined *a priori* for all  $j$ . Then, a hyperplane acting as decision surface is defined as:

$$\sum_{j=1}^{m_1} w_j \Phi_j(\mathbf{x}) + b = 0 \quad (2.7)$$

where  $\{w_j\}_{j=1}^{m_1}$  denotes a set of linear weights connecting the feature space to output space, and  $b$  is the bias. The equation (2.7) can be simplified by writing:

$$\sum_{j=0}^{m_1} w_j \Phi_j(\mathbf{x}) = 0 \quad (2.8)$$

where it is assumed that  $\Phi_0(\mathbf{x}) = 1$  for all  $\mathbf{x}$ , so that  $w_0$  denotes the bias  $b$ . Equation (2.8) defines the decision surface computed in the feature space in terms of the linear weights of the machine. The quantity  $\Phi_j(\mathbf{x})$  represent the input supplied to the weight  $w_j$  via the feature space. In fact, the vector  $[\Phi_0(\mathbf{x}), \Phi_1(\mathbf{x}), \dots, \Phi_{m_1}(\mathbf{x})]^T$  can be considered the "image" induced in the feature space due to the input vector  $\mathbf{x}$ .

The problem of finding weight coefficients  $\mathbf{w}$  can be formulated as an optimization problem with constraint. It can be shown that finding optimal hyperplane is equal to minimizing the cost function [6]:

$$\Psi(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i \quad (2.9)$$

The parameter  $C$  is regarded as the regularization parameter and is selected by user. A large  $C$  corresponds to assigning a higher penalty to the training errors.

Given the training samples as  $\{(\mathbf{x}_i, d_i)\}_{i=1}^N$ , the constraints which must be satisfied are:

$$d_i(\mathbf{w}^T \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \text{ for } i = 1, 2, \dots, N \quad (2.10)$$

This constrained optimization problem can be solved using the lagrange multiplier. The Lagrangian function is constructed as:

$$\mathbf{J}(\mathbf{w}, \alpha, \xi, b) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i [d_i(\mathbf{w}^T \Phi(\mathbf{x}_i) + b) - 1 + \xi_i] \quad (2.11)$$

where the nonnegative variables  $\alpha_i$  are called Lagrange multipliers. The solution to constrained optimization problem is determined by the saddle point of the Lagrange function  $\mathbf{J}(\mathbf{w}, b, \alpha)$ , which has to be minimized with respect to  $\mathbf{w}$  and  $b$ . It also has to be maximized with respect to  $\alpha$ . Applying the optimality conditions:

$$\frac{\partial \mathbf{J}(\mathbf{w}, \alpha, \xi, b)}{\partial \mathbf{w}} = \mathbf{0} \quad (2.12)$$

$$\frac{\partial \mathbf{J}(\mathbf{w}, \alpha, \xi, b)}{\partial b} = \mathbf{0} \quad (2.13)$$

$$\frac{\partial \mathbf{J}(\mathbf{w}, \alpha, \xi, b)}{\partial \xi} = \mathbf{0} \quad (2.14)$$

to Lagrange function (2.11 ) yields:

$$\mathbf{w} = \sum_{i=1}^N \alpha_i d_i \Phi(\mathbf{x}_i) \quad (2.15)$$

Then, the non-linear SVM classifier is obtained as:

$$f(x) = \sum_{i=1}^N \alpha_i d_i \Phi^T(\mathbf{x}_i) \Phi(\mathbf{x}) + b = \sum_{i=1}^N \alpha_i d_i \mathbf{K}(\mathbf{x}_i, \mathbf{x}) + b \quad (2.16)$$

where the function  $\mathbf{K}(\cdot, \cdot)$  is defined as:

$$\mathbf{K}(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^T \mathbf{x}_2 \quad (2.17)$$

$\mathbf{K}$  is referred as kernel function.

Practically Lagrange Multipliers are solved from dual form of (2.9), which is expressed as:

$$\sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_j \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) \quad (2.18)$$

Subject to:

$$0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, N \quad (2.19)$$

$$\sum_{i=1}^N \alpha_i d_i = 0 \quad (2.20)$$

The dual problem is solved numerically through quadratic programming. The Karush-Kuhn-

Tucker optimality conditions for (2.18) lead to the following three cases for each  $\alpha_i$ : a)  $\alpha_i = 0$ . This corresponds to  $d_i f(\mathbf{x}_i) > 1$ . In this case, the data element  $\mathbf{x}_i$  is outside the decision margin of the function  $f(\mathbf{x})$  and is correctly classified. b)  $0 < \alpha_i < C$ . In this case,  $d_i f(\mathbf{x}_i) = 1$ . The data element  $\mathbf{x}_i$  is strictly located on the decision margin of  $f(\mathbf{x})$ . Hence,  $\mathbf{x}_i$  is referred as a margin support vector for  $f(\mathbf{x})$ . c)  $\alpha_i = C$ . In this case,  $d_i f(\mathbf{x}_i) < 1$ .

The data element  $\mathbf{x}_i$  is inside decision margin (though it may still be correctly classified). Accordingly,  $\mathbf{x}_i$  is called an error support vector of  $f(\mathbf{x})$ . The most of the training examples in a typical problem are correctly classified by the trained classifier (case a), i.e., only a few training examples will be support vectors. For simplicity, let  $\mathbf{s}_j, \alpha_j^*, j = 1, 2, \dots, N_s$ , denotes these support vectors and their corresponding nonzero Lagrange multipliers respectively. The decision function in (2.16) can be simplified as:

$$f(x) = \sum_{j=1}^{N_s} \alpha_j^* d_j \mathbf{K}(\mathbf{s}_j, \mathbf{x}) + b \quad (2.21)$$

As it can be seen in (2.18) and (2.21), The nonlinear mapping  $\Phi(\cdot)$  never appears explicitly in either dual form of SVM training problem or the resulting decision function. The mapping  $\Phi(\cdot)$  only enters the problem implicitly through the kernel function  $\mathbf{K}(\cdot, \cdot)$ , thus it is only necessary to define  $\mathbf{K}(\cdot, \cdot)$  which implicitly defines  $\Phi(\cdot)$ . However, when choosing a kernel function, it is necessary to check that it is associated with the inner product of some non-linear mapping. Mercer's theorem [6] states that such a mapping indeed underlies a kernel  $\mathbf{K}(\cdot, \cdot)$  provided that  $\mathbf{K}(\cdot, \cdot)$  is a positive integral operator, that is, for every square-integrable function  $g(\cdot)$  defined on  $R^n$  the kernel  $\mathbf{K}(\cdot, \cdot)$  satisfies the following condition:

$$\int \int K(\mathbf{x}, \mathbf{y}) g(\mathbf{x}) g(\mathbf{y}) d\mathbf{x} d\mathbf{y} \geq 0 \quad (2.22)$$

A new kernel is proposed in Chapter 4 based on feature extraction methods reviewed in Chapter 3.

## Chapter 3

# Time series and Time-frequency features

**F**EATURE selection is a crucial step in signal and image classification and pattern recognition. The goal of feature selection is having large "between class distance" and small "within-class variance" in the feature vector space.

SVM classification method basically does not incorporate any external feature. In fact the kernel function implicitly extract features within the learning scheme.

The main focus of current and following chapter is investigating the effectiveness of employing external features within SVM kernel. In the following sections three feature extraction methods are reviewed and they will be compared in the next chapter.

### 3.1 Local discriminant basis

Local discriminant basis (LDB) [8] is a supervised scheme for feature extraction and is the classification counterpart of the Best Basis algorithm developed by Coifman and Wickerhauser [9] for signal and image compression. The appeal of LDB algorithm lies in the supervised selection of basis functions from redundant and structured basis that are well localized both in time and frequency. The tree structured dictionary of the basis, which is provided by wavelet packet transform, is pruned to find the most discriminative basis for the

purpose of the classification.

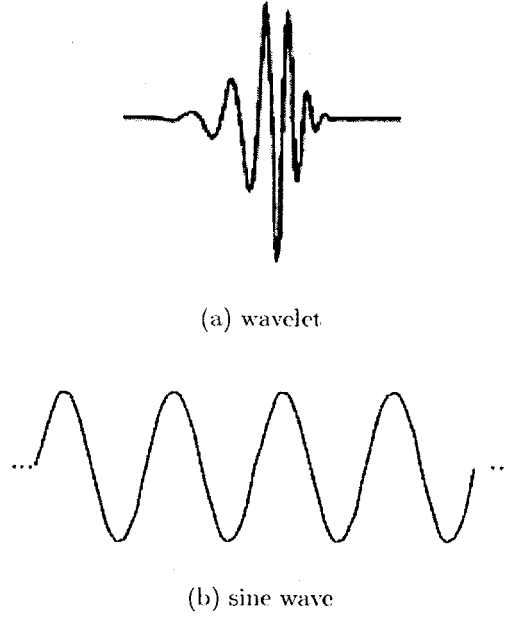
In the following subsections, wavelet, wavelet packet, and LDB algorithm are briefly reviewed.

### 3.1.1 Wavelet and Wavelet Packet Transforms

Fourier transform has been the principle tool for signal analysis for a long time. It constructs a sinusoidal basis to describe the energy distribution of a signal in frequency domain. But Fourier transform has a drawback that the time (or spatial) information is lost due to the integration over entire real axis. Thus by looking at the Fourier Transform domain it is impossible to say when (or where) a particular phenomena took place. However most of the practical signal and images contain quasi-periodic and quasi-stationary characteristics that capturing them is crucial for classification purpose. Windowing the signal and applying Fourier Transform to each window (Short Time Fourier Transform(STFT)) was the first attempt to overcome this problem. STFT in turn, reduces the frequency resolution which is in compliance with Heisenberg uncertainty principal [10]. Wavelet transform, which is based on multiresolutional analysis, were proposed and developed to overcome fixed resolution problem in STFT.

Similar to the Fourier transform, that describes a signal in terms of the sinusoidal waves of various frequencies. wavelet analysis is an attempt to describe a signal according to shifted and scaled versions of a basis function. This basis function is referred as mother wavelet. By considering the wavelets and sine waves, depicted in Fig 3.1, we can see intuitively that signals with sharp changes might be better analyzed with an irregular wavelet than with a smooth sinusoid. It also makes sense that local features can be described better with wavelets that have local extent.

The discrete wavelet transform is defined as [10]:



**Figure 3.1:** Wavelet and Sine wave

$$W_{\Phi}(j, k) = \frac{1}{\sqrt{N}} \sum_n s(n) \Phi_k(n) \quad (3.1)$$

$$W_{\Psi}(j, k) = \frac{1}{\sqrt{N}} \sum_n s(n) \Psi_{j,k}(n) \quad (3.2)$$

$$n = 0, \dots, N - 1$$

$$j = 0, \dots, J - 1$$

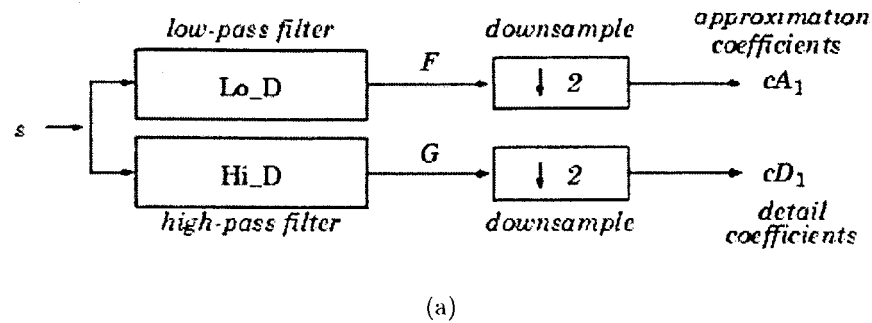
$$k = 0, \dots, 2^j - 1$$

$$J = \log_2 N$$

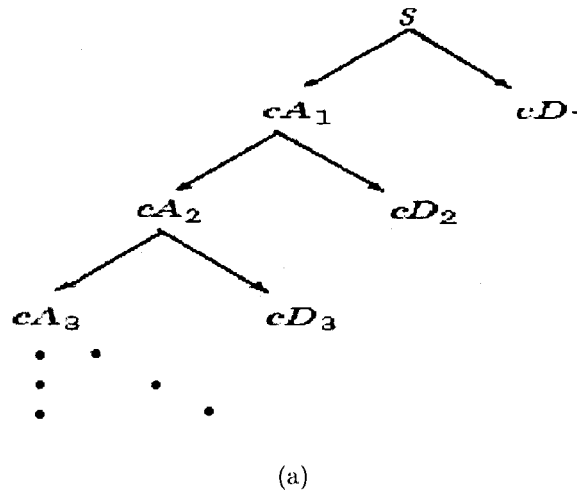
where  $N$  is signal length and  $W_{\Phi}$  and  $W_{\Psi}$  are approximation and detailed coefficients respectively.  $j$  and  $k$  denote scale and translation.  $\Psi$  and  $\Phi$  are wavelet and scale function. Mallat [11] implemented an algorithm known as two-channel subband coder to obtain the discrete wavelet coefficient which is described in Fig. 3.2. The expansion can be continued



$J$  times (Fig. 5.4) for complete signal decomposition resulting in the complete set of wavelet coefficients. For the sake of image processing applications the subband coding scheme of Fig. 3.2 is generalized as depicted in Fig. 3.4.



**Figure 3.2:** Computing wavelet coefficients via subband coding method



**Figure 3.3:** Signal Decomposition to wavelet coefficients

The wavelet transform is extended to wavelet packet decomposition which offers a richer range of possibilities for signal analysis. In wavelet packet transform after splitting the signal to approximation and detail in first stage we continue the splitting over the detail part as

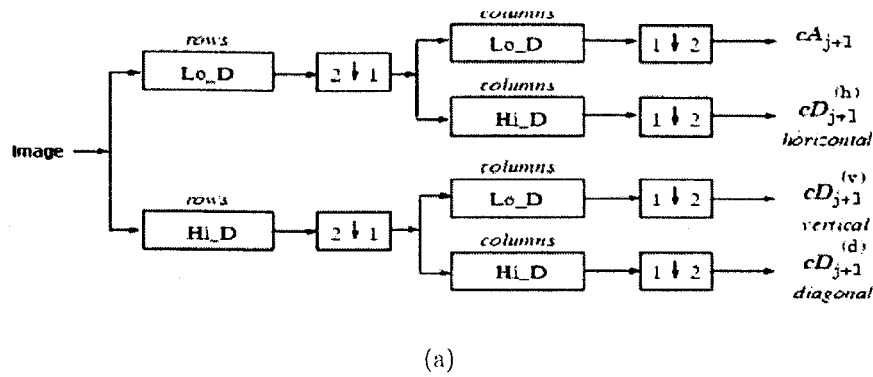


Figure 3.4: Generalization of subband coding for 2D Signal

well as approximation part. Thus, for  $n$  level decomposition there will be  $n + 1$  possibilities. These different possibilities construct the redundant library, which is used by LDB.

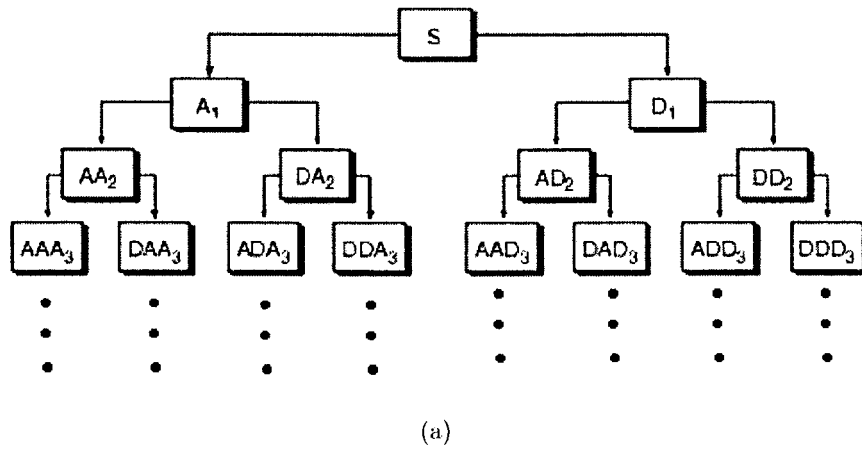
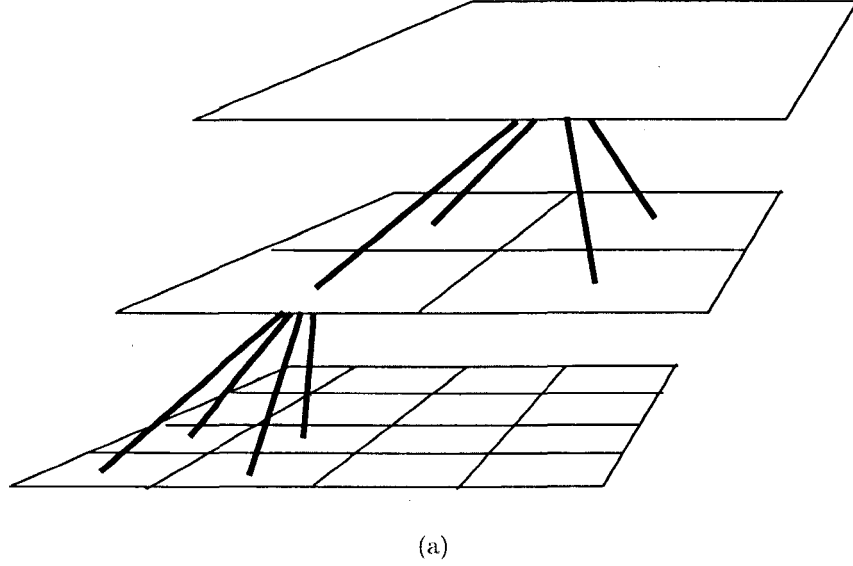


Figure 3.5: Wavelet Packet transform: An extension to wavelet transform



**Figure 3.6:** Wavelet Packet decomposition (quad-tree for image signal)

### 3.1.2 LDB Algorithm

The graphical LDB scheme is shown in Fig. 3.7. In the following a formal statement of the algorithm is presented.

Assume that set of  $N$  training signals from  $L$  different classes are given and  $N_l$  is the number of signals in class  $l$ . Let  $\mathbf{s}_i^{(l)}$  denote the collection of training signals in class  $l$ , in which the superscript  $(l)$  indicates the class that the signal belongs to. Suppose that each  $\mathbf{s}$  belongs to a unique class so that  $N = N_1 + N_2 + \dots + N_L$ . LDB uses this training set of signals to search for a best basis in available libraries of bases respect to its cost function.

**Definition I** Given a sequence of vectors  $\{\mathbf{s}^{(c)}\}_{c=1}^C$  ( $C$  is number of classes), their J-divergence is defined as in (3.3) where the summation is taken over all pairs of  $i$  and that are not equal.

$$J(\{\mathbf{s}^{(c)}\}_{c=1}^C) = \sum \sum J(\mathbf{s}^{(i)}, \mathbf{s}^{(j)}) \quad (3.3)$$

**Definition II** Let  $\{s_i^{(c)}\}_{i=1}^{N_c}$  be a set of training signals belonging to class  $c$ . Then the time-

frequency energy map of class  $c$ , denoted by  $\Gamma_c$ , is a table of real numbers specified by triplet  $(j, k, l)$ :

$$\Gamma_c(j, k, l) = \frac{\sum_{i=1}^{N_c} (\omega_{j,k,l}^T s_i^{(c)})^2}{\sum_{i=1}^{N_c} \|s_i^{(c)}\|} \quad (3.4)$$

where  $\omega$  is a basis function and  $j = 0, 1, \dots, J$ ,  $k = 0, 1, \dots, 2^j - 1$ , and  $l = 0, 1, \dots, 2^{n_0-j} - 1$ .

In other word,  $\Gamma_c$  is computed by accumulating the squares of expansion coefficients of the signals at each position in the binary tree followed by the normalization of the total energy of the signals belonging to class  $c$ . This normalization is important especially if there is significant difference in the number of samples among classes. The following notation is used in LDB algorithm stated below,

$$\Delta_{j,k} = J(\{\Gamma_c(j, k, \cdot)\}_{c=1}^C) = \sum_l J(\Gamma_1(j, k, l), \dots, \Gamma_C(j, k, l)) \quad (3.5)$$

Here is an algorithm to select an orthonormal basis ( from the library), which maximizes the discriminant measure on the time-frequency energy distributions of classes. Let  $B_{j,k}$  denote a set of basis vectors at the subspace  $\Omega_{j,k}$  arranged as a matrix:

$$B_{j,k} = \{\Omega_{j,k,0}, \dots, \Omega_{j,k,2^{n_0-j}-1}\} \quad (3.6)$$

Let the  $A_{j,k}$  represent the LDB (which the algorithm is searching for) restricted to  $B_{j,k}$ .

**Algorithm (The Local Discriminant Basis Selection Algorithm)** . Given a training database of classes of signals:

**Step 0.** Choose a library of orthonormal bases(i.e. specify QMFs for wavelet packet).

**Step 1** Construct time-frequency energy map  $\Gamma_c$  for  $c = 1, \dots, C$

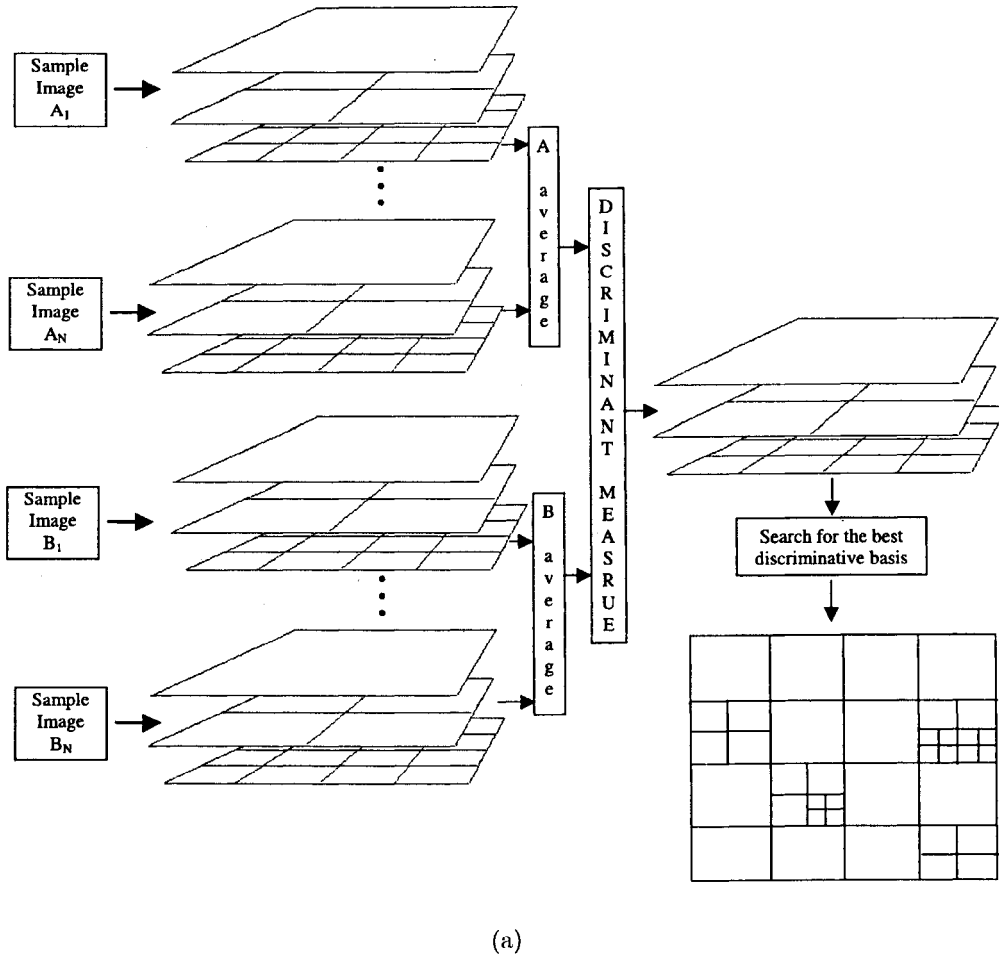
**Step 2** Set  $A_{j,k} = B_{j,k}$  for  $k = 0, 1, \dots, 2^j - 1$ .

**Step 3** Determine the best subspace  $A_{j,k}$  for  $j = J - 1, \dots, 0$  and  $k = 0, 1, \dots, 2^j - 1$  by the following rule:

if  $\Delta_{j,k} \leq \Delta_{j+1,2k} + \Delta_{j+1,2k+1}$   
 then  $A_{j,k} = B_{j,k}$   
 else  $A_{j,k} = A_{j+1,2k} \oplus A_{j+1,2k+1}$

**Step 4.** Order the basis functions by their power of discrimination (see Step 5).

**Step 5.** Use  $k(< n)$  most discriminant basis functions for constructing classifiers.



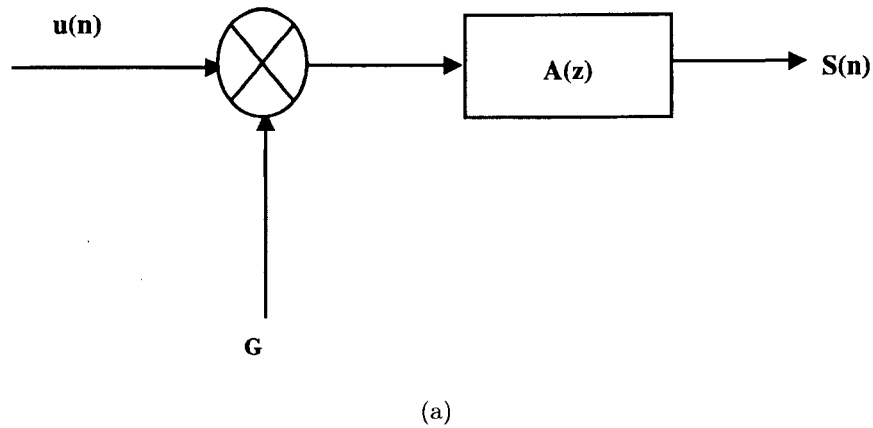
**Figure 3.7:** LDB best basis selection scheme for a two class(A,B) problem

## 3.2 Linear predictive coding

Linear predictive coding (LPC) is a popular and effective technique for signal compression [12], [13]. In particular, it has been employed successfully for speech coding [14], [15], [16], [17] and image compression [18].

Usually, in a linear predictive coder, a sample of the image signal to be coded is first predicted. If the sample is predicted as a weighted sum of other samples of this signal, the LPC is called the autoregressive (AR). If the sample is predicted from a known excitation, the LPC is referred to as the moving average (MA) model. If the sample is predicted using both the signal to be coded and an excitation signal, the LPC is known as the autoregressive moving average (ARMA) model. In this thesis, the (AR) model has been used and it is reviewed in the following subsections.

### 3.2.1 The LPC Model



**Figure 3.8:** Linear Prediction Coding Model

The basic idea behind the LPC model is that a given signal  $s(n)$ , can be approximated as a linear combination of the previous  $p$  signal samples, such that

$$s(n) \approx a_1 s(n-1) + a_2 s(n-2) + \dots + a_p s(n-p) \quad (3.7)$$

The equation 3.7 can be converted to an equality by including an hypothetical input term,  $Gu(n)$ , which is referred as excitation in speech analysis literature. The equation can be written as:

$$s(n) = \sum_{i=1}^p a_i s(n-i) + Gu(n) \quad (3.8)$$

where  $u(n)$  is a normalized hypothetical input and  $G$  is its gain. By expressing Eq. 3.8 in the z-domain we get the relation:

$$S(z) = \sum_{i=1}^p a_i z^{-i} S(z) + GU(z) \quad (3.9)$$

leading to the transfer function:

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} = \frac{1}{A(z)} \quad (3.10)$$

Fig. 3.8 is presenting the transfer function in 3.10. The linear prediction model indicates that a signal  $s(n)$  can be estimated by a system of order  $p$  with a scaled input  $u(n)$ . While the signal is quasi-stationary the input can be a random noise or arbitrary quasi-periodic signal. Accordingly, the  $A(z)$  coefficients can provide an efficient feature for presenting the signal  $s(n)$ .

### 3.2.2 LPC analysis equation

Based on the model of Fig. 3.8 the estimate  $\tilde{s}(n)$  is defined as:

$$\tilde{s} = \sum_{k=1}^p a_k s(n-k) \quad (3.11)$$

and the prediction error is:

$$e(n) = s(n) - \tilde{s} = s(n) - \sum_{k=1}^p a_k s(n-k) \quad (3.12)$$

Now a set of coefficients must be sought to minimize mean square error in a short segment of the signal. To set up the equations that must be solved to determine the predictor coefficients, the short-term signal and error segments are defined at signal sample  $n$  as:

$$s_n(m) = s(n+m); e_n(m) = e(n+m) \quad (3.13)$$

where  $m$  is the segment length. Then the mean square error which must be minimized can be written as:

$$E_n = \sum_m [s_n(m) - \sum_{k=1}^p a_k s_n(m-k)]^2 \quad (3.14)$$

To solve Eq. 3.14, for the predictor coefficients, we differentiate  $E_n$  with respect to each  $a_k$  and set the results to zero,

$$\frac{\partial E_n}{\partial a_k} = 0, \quad k = 1, 2, \dots, p \quad (3.15)$$

giving

$$\sum_m s_n(m-i) s_n(m) = \sum_{k=1}^p \hat{a}_k \sum_m s_n(m-i) s_n(m-k) \quad (3.16)$$



Knowing that terms of the form  $\sum s_n(m-i)s_n(m-k)$  are terms of the short-term covariance of  $s_n(m)$ , i.e.,

$$\phi_n(i, k) = \sum_m s_n(m-i)s_n(m-k) \quad (3.17)$$

The Eq. 3.16 can be expressed in the compact notation:

$$\phi_n(i, 0) = \sum_{k=1}^p \hat{a}_k \phi_n(i, k) \quad (3.18)$$

which describe a set of  $p$  equations in  $p$  unknowns. It is readily shown that the minimum mean-squared error,  $\hat{E}_n$ , can be expressed as:

$$\hat{E}_n = \sum_m s_n^2(m) - \sum_{k=1}^p \hat{a}_k \sum_m s_n(m)s_n(m-k) \quad (3.19)$$

$$\phi_n(0, 0) - \sum_{k=1}^p \hat{a}_k \phi_n(0, k) \quad (3.20)$$

Thus the minimum mean-squared error consists of a fixed term ( $\phi_n(0, 0)$ ) and terms that depend on the predictor coefficients.

To solve Eq. 3.18 for the optimum predictor coefficients ( $\hat{a}_k$ ) the  $\phi_n(i, k)$  must be computed for  $1 \leq i \leq p$  and  $0 \leq k \leq p$ , and then the resulting set of  $p$  simultaneous equations. Practically, the method of solving aforementioned equations is a strong function of the range of  $m$  used in defining both the section of signal for analysis and the region over which the mean-squared error is computed. In this thesis the autocorrelation method was used and is reviewed in the next subsection.

### 3.2.3 The Autocorrelation method

In this approach, it is assumed that the signal segment,  $s_n(m)$ , is identically zero outside the interval  $0 \leq m \leq N - 1$ . This is equivalent to assuming that the signal,  $s(m + n)$ , is multiplied by a finite length window,  $w(m)$ , which is identically zero outside the range  $0 \leq m \leq N - 1$ . Thus the signal sample for minimization can be expressed as:

$$s_n(m) = s(m + n).w(n), \quad 0 \leq m \leq N - 1 \quad (3.21)$$

$$s_n(m) = 0, \quad \text{otherwise}$$

According to this equation, for  $m < 0$ , the error signal  $e_n(m)$  is zero since  $s_n(m) = 0$  for all  $m < 0$  and therefore there is no prediction error. Furthermore, for  $m > N - 1 + p$  there is again no prediction error because  $s_n(m) = 0$  for all  $m > N - 1$ . However, in the region of  $m = 0$  (i.e., from  $m = 0$  to  $m = p - 1$ ) the windowed signal  $s_n(m)$  is being predicted from previous samples, some of which are arbitrarily zero. Hence the potential for relatively large prediction errors exists in this region. Furthermore, in the region of  $m = N - 1$  (i.e., from  $m = N - 1$  to  $m = N - 1 + p$ ) the potential of large prediction errors again exists because the zero-valued (weighted) signal is being predicted from at least some nonzero previous samples. The purpose of the window of Eq. 3.21 is to taper the signal near  $m = 0$  and near  $m = N - 1$  so as to minimize the errors at section boundaries.

Based on using the weighted signal of Eq. 3.21 the mean-squared error becomes:

$$E_n = \sum_{m=0}^{N-1+p} e_n^2(m) \quad (3.22)$$

and  $\phi_n(i, k)$  can be expressed as:

$$\phi_n(i, k) = \sum_{m=0}^{N-1+p} s_n(m - i)s_n(m - k), \quad 1 \leq i \leq p, \quad 0 \leq k \leq p \quad (3.23)$$

or

$$\phi_n(i, k) = \sum_{m=0}^{N-1-(i-k)} s_n(m) s_n(m + i - k), \quad 1 \leq i \leq p, \quad 0 \leq k \leq p \quad (3.24)$$

Since Eq. 3.24 is only a function of  $i - k$  (rather than the two independent variables  $i$  and  $k$ ), the covariance function,  $\phi_n(i, k)$ , reduces to the simple autocorrelation function, i.e.,

$$\phi_n(i, k) = r_n(i - k) = \sum_{m=0}^{N-1-(i-k)} s_n(m) s_n(m + i - k) \quad (3.25)$$

Since the autocorrelation function is symmetric, i.e.  $r_n(-k) = r_n(k)$ , the LPC equations can be expressed as:

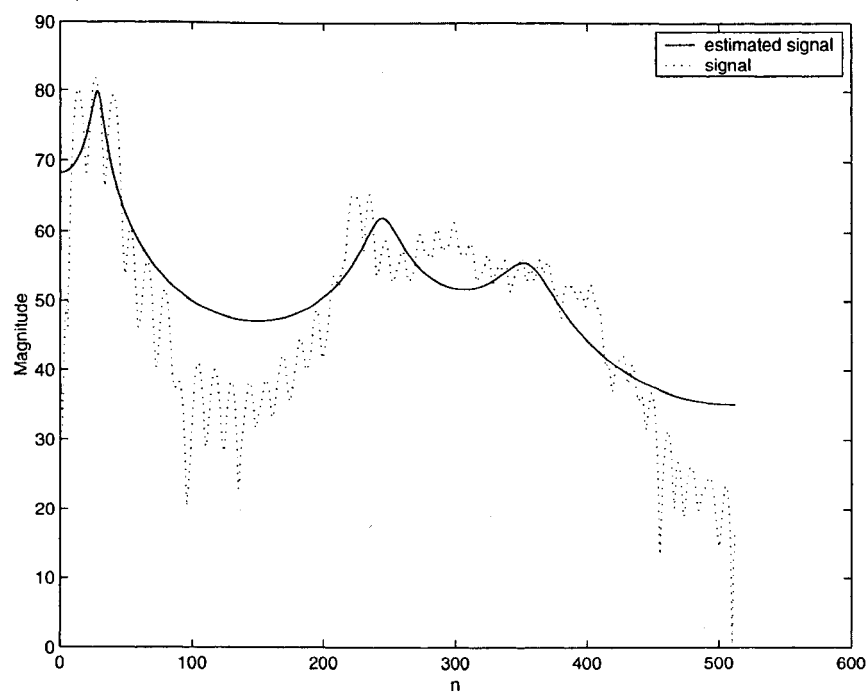
$$\sum_{k=1}^p r_n(|i - k|) \hat{a}_k = r_n(i), \quad 1 \leq i \leq p \quad (3.26)$$

this equation can be expressed in matrix form as:

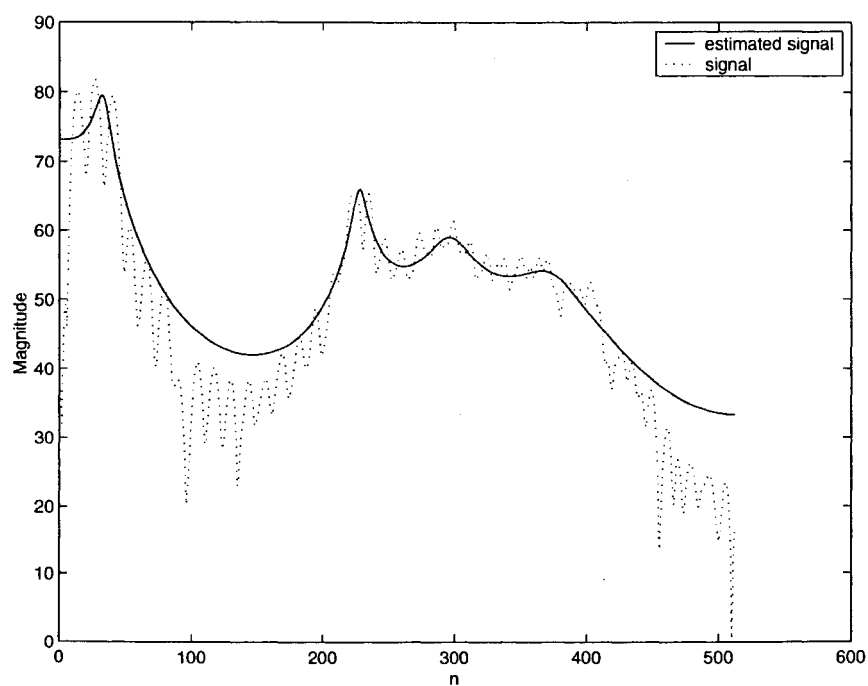
$$\begin{pmatrix} r_n(0) & r_n(1) & r_n(2) & \cdots & r_n(p-1) \\ r_n(1) & r_n(0) & r_n(1) & \cdots & r_n(p-2) \\ r_n(2) & r_n(1) & r_n(0) & \cdots & r_n(p-3) \\ \vdots & \vdots & \vdots & & \vdots \\ r_n(p-1) & r_n(p-2) & r_n(p-3) & \cdots & r_n(0) \end{pmatrix} \begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \vdots \\ \hat{a}_p \end{pmatrix} = \begin{pmatrix} r_n(1) \\ r_n(2) \\ r_n(3) \\ \vdots \\ r_n(p) \end{pmatrix} \quad (3.27)$$

The  $p \times p$  matrix of autocorrelation values is a Toeplitz matrix (symmetric with all diagonal elements equal) and hence can be solved efficiently through several well-known procedures. Levinson-Durbin algorithm implemented in matlab toolbox have been used in this thesis.

The spectral estimation for a given speech sample is shown in Fig. 3.9 and Fig. 3.10 for LPC model of different orders.

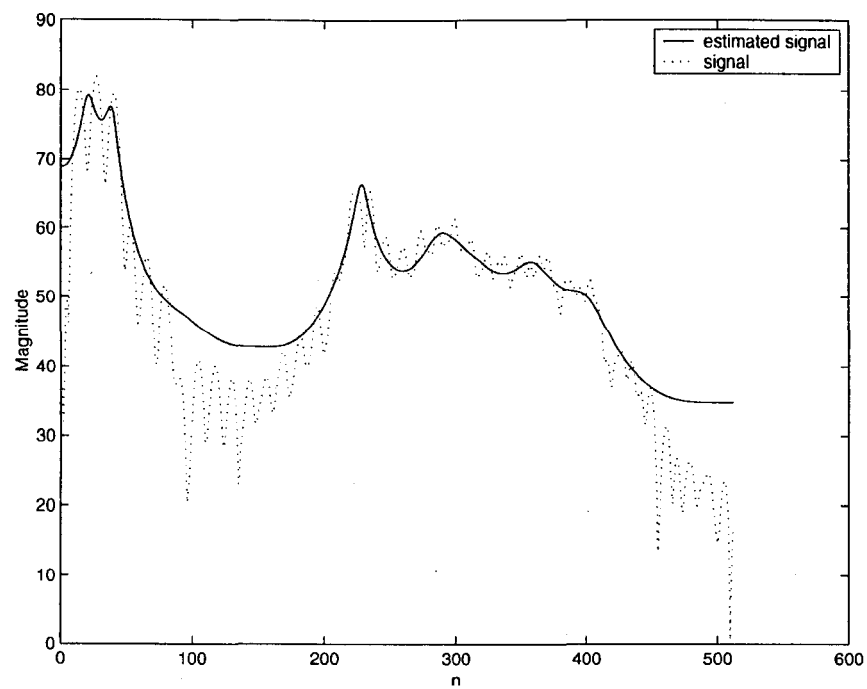


(a)

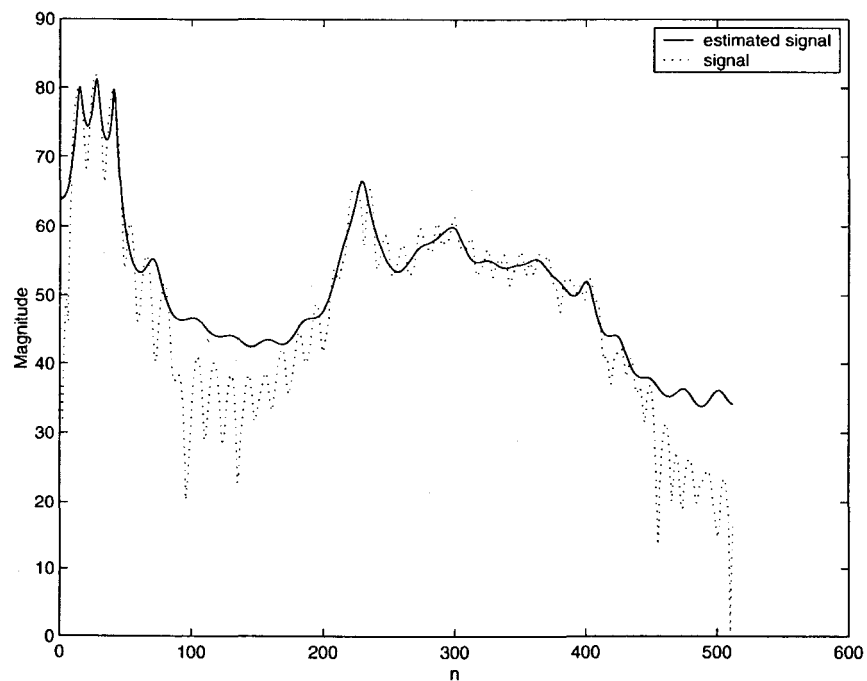


(b)

**Figure 3.9:** Signal spectra is compared with its estimated spectra using LPC of order a) 6 b) 12



(a)



(b)

**Figure 3.10:** Spectral estimation of sample speech signal, with a LPC of order a) 24 b) 48

### 3.3 Cohen's Space-Frequency signal representation

The Cohen's group space-frequency [19] representation of signal  $s(n)$  is shown by  $C_s^\phi(n, \Omega)$  and is parameterized by its SFR kernel  $\phi$ . Where  $\Omega$  is discrete frequency.

Many parametric SFR kernel shapes have been proposed in literature, and an efficient choice is the family of radially Gaussian kernels, defined in the ambiguity plane as:

$$\phi(\xi, \tau) = \exp - \frac{\rho^2}{2c(\varphi)^2} \quad (3.28)$$

Where  $\rho$  and  $\phi$  are the polar coordinates given by:

$$\rho = \sqrt{\xi^2 + \tau^2} \quad (3.29)$$

$$\varphi = \arctan^{-1}\left(\frac{\xi}{\tau}\right) \quad (3.30)$$

and the contour function is:

$$c(\varphi) = a_0 + \sum_{p=1}^{P_{max}} [a_p \cos(2p\varphi) + b_p \sin(2p\varphi)] \quad (3.31)$$

in which  $a_0$  is chosen such that  $c(\varphi) \geq a'_0 > 0$  for all  $\varphi$ . The SFR kernel parameters are then  $a'_0, a_p$  and  $b_p$ , with  $p = 1, \dots, p_{max}$ . In the following, this set of parameters is denoted:

$$\theta = [a'_0, a_1, b_1, \dots, a_{p_{max}}, b_{p_{max}}]$$

Given the training set of signals  $O_N$  for a two-class classification problem, an optimization procedure is applied to obtain the best discriminative set of parameters  $\theta$ . In the optimization procedure,  $O_L$  and  $O_T$  are considered to be subsets of approximately equal size obtained by randomly partitioning training set  $O_N$  in two parts, each containing elements from both classes. There are  $T_{+1}$  samples in  $O_T$  labelled +1, and  $T_{-1}$  samples labelled -1 (It is assumed that the two classes of data are labelled by  $-1, 1$ ). Also, There is  $L_{+1}$  samples in  $O_L$  labelled +1, and  $L_{-1}$  samples labelled -1. In order to optimally obtain  $\theta$ , an optimization criterion  $P(\theta | O_N)$  is introduced which is minimized with respect to  $\theta$  via a standard optimization

procedure. For a given training set  $O_N$  and given kernel parameters  $\theta$ ,  $P(\theta | O_N)$  is calculated as follows:

Step 1) Use the set  $O_L$  to train the classifier.

Step 2) Classify each element in  $O_T$ . This results in two sets of values:

$$f_{+1}^{(i)} = f_{O_L}(\mathbf{x}_i | y_i = +1), \quad i = 1, \dots, T_{+1}$$

and

$$f_{-1}^{(i)} = f_{O_L}(\mathbf{x}_i | y_i = -1), \quad i = 1, \dots, T_{-1}$$

Step3) Compute the empirical mean  $m_{+1}, m_{-1}$  and standard deviation  $s_{+1}, s_{-1}$  of  $f_{+1}^{(i)}$

and  $f_{-1}^{(i)}$  where

$$m_{+1} = \frac{1}{T_{+1}} \sum_{i=1}^{T_{+1}} f_{+1}^{(i)}, \quad m_{-1} = \frac{1}{T_{-1}} \sum_{i=1}^{T_{-1}} f_{-1}^{(i)} \quad (3.32)$$

$$s_{+1}^2 = \frac{1}{T_{+1}} \sum_{i=1}^{T_{+1}} (f_{+1}^{(i)} - m_{+1})^2, \quad s_{-1}^2 = \frac{1}{T_{-1}} \sum_{i=1}^{T_{-1}} (f_{-1}^{(i)} - m_{-1})^2 \quad (3.33)$$

Step 4) Compute the criterion introduced in [20]:

$$P(\theta | O_T, O_L) = \frac{1}{2} Q\left(\frac{m_{+1}}{s_{+1}}\right) + \frac{1}{2} Q\left(\frac{m_{-1}}{s_{-1}}\right) \quad (3.34)$$

where

$$Q(v) = \int_v^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \quad (3.35)$$

steps one to four are repeated  $R$  times with different subsets  $O_L$  and  $O_T$ , and the final criterion  $P(\theta | O_N)$  is obtained as the average of  $P(\theta | O_T, O_L)$  over all the subsets tested. Under Gaussian assumption over the distributions of  $f_{+1}^{(i)}$  and  $f_{-1}^{(i)}$ , it can be shown [20]  $P(\theta | O_N)$  is an estimate of the classification probability of error for the classifier implemented with parameter  $\theta$ . In practice,  $\theta$  is selected to minimize this estimated probability of error  $P(\theta | O_N)$  using a standard numerical optimization procedure. When the optimal  $\theta' = \arg \min_{\theta} P(\theta | O_N)$  is obtained, the classifier is trained over full learning set  $O_N$  using the optimal SFR kernel.

## Chapter 4

# A Novel SVM Kernel for Texture Classification

### 4.1 Motivation for texture classification

Researchers have been working for several years to implement image understanding algorithms that can duplicate the excellent ability of human brain to recognize objects. In addition to shape of an object, it is believed that texture characteristics play an important role in human brain and visual system to recognize and interpret objects in perceived images. Consequently, texture analysis has been an active research field and large number of diverse algorithms have been proposed and tried. Texture classification currently is used as an essential part in a variety of image processing applications.

### 4.2 Applications

A wide range of different industrial applications have been successfully implemented based on texture analysis algorithms. Marti *et. al.* [21] used texture features in conjunction with other modalities for object recognition in industrial environments. Kumar and Pang [22]



applied texture analysis for defect detection in textured materials. Li *et. al.* [23] classified tough and tender beef by texture analysis.

Texture characteristics have been also used extensively in medical imaging to detect abnormalities. Texture features and maximum likelihood classifier were used by Horng *et. al.* [24] to classify ultrasonic liver images into three category of normal liver, liver hepatitis and cirrhosis. A statistical model based on autoregressive periodic random field model in conjunction with conventional texture analysis parameters were used by Bleck *et. al.* [25] to detect microfocal lesions in ultrasound liver images. Wang *et. al.* [26] used texture features to detect infected tissues by ulcer and coli in endoscopic images. Texture features and analysis is also used for breast cancer detection. Gurcan *et. al.* [27] used higher order statistical texture features for detection of microcalcifications in mammograms. They used skewness and kurtosis to discriminate between normal and abnormal tissues. In another work, textural features was used by Kim and Park [28] to detect microcalcification in digitized Mammography. A computer aided diagnosis (CAD) system for automatic abnormalities detection in chest radiographs was introduced by Ginneken *et. al.* [29] using local texture analysis.

## 4.3 Texture Classification Challenges and Previous works

Texture classification algorithms generally include two crucial steps: 1) feature extraction and 2) classification. In the following subsections the aforementioned and the previous work on each of them are addressed.

### 4.3.1 Feature Extraction

In feature extraction stage, a set of features are sought that can be efficiently computed and embody as much discriminative information as possible about the textural characteristics.

There are several methods for extracting textural feature which can be loosely classified as statistical, model-based and signal processing methods [3]. In statistical methods, textures are modelled using statistical measures. These methods are motivated by the fact that psychophysical experiments have demonstrated human visual system ability to extract some statistics of an order higher than two [30]. The major drawback to this type of method is the enormous amount of data involved in higher order statistics. In Model-based methods, textures are characterized based on probability distribution in random fields, such as Markov chains and Markov random fields (MRFs) [31], [32]. MRF based methods are widely used but they are computationally expensive. Signal processing methods (also referred as multi-channel filtering) are popular due to their simplicity of implementation. In these methods a textured input image is decomposed into feature images using filter banks such as Gabor, wavelet or neural network based filters [33], [34], [35]. The main issue in these methods is selection of optimum set of filters for the problem in hand.

### 4.3.2 Classification

In the second stage of a texture classification algorithm, a classification paradigm is constructed to distinguish between texture features correspondent to different texture classes. Several different classifiers have been reported. Minimum distance classifier based on Euclidean or Mahalanobis distance was used in [36] and [37]. Manian *et. al.* used K-nearest neighbor in [38]. Fisher linear discriminant (FLD) was used in [39] by Clausi and Jernigan. Raghu and Yegnanarayana used artificial neural network (ANN) in [40]. Bayesian classifier was used in [41], [42]. Bayesian classifier is known as optimal classifier, but calculation of underlying probability distribution of the problem under study is not practically possible, specially in the absence of adequate number of training samples. In this thesis, the SVM has been chosen since it was shown that SVM outperforms other classification methods [6].

Superiority of SVM originates from its ability to generalize in high dimensional spaces, such as the space which is spanned by texture patterns. The generalization ability of SVM is based on its profound relation to the underlying statistical learning theory. In fact, SVM is an approximate implementation of Structural Risk Minimization method(SRM) since it imposes a theoretical bound on generalization error and sparseness of the solution. In SVM, instead of minimizing an objective function based on the training samples ( such as mean square error), it is attempted to minimize a bound on generalization error ( i.e., the error made by the learning machine on test data not used during training). Therefore, an SVM tend to perform well when applied to data outside the training set. SVM achieves this advantage by focusing on the training examples that are most difficult to classify. These "borderline" training examples are called support vectors.

Since SVM is a learning based classifier, it is mostly comparable with Neural Network. A brief comparison is given in Table 4.1.

**Table 4.1:** Brief comparison between SVM and NN

SVM	NN
Minimize Structural Risk	Minimize Empirical Risk
Less over fit	More over fit
Convex QP can always find a global optimum	gradient descent may stick at local optima
Challenge: choice of kernel functions	Challenge: Structures of network
Faster training	Slower training

## 4.4 External features to build a new kernel for SVM

Unlike other texture classification methods, SVM based method dose not necessarily incorporate any external feature extraction method. Kim *et. al.* [3] showed the effectiveness of

SVM in texture classification problem. In fact, in an SVM, feature extraction is implicitly performed by a kernel, which is defined as the dot product of two mapped patterns. There are some basic and popular kernels which are widely used in different research including Kim *et. al.*'s work [3]. It is believed that the proper selection of SVM kernel significantly affect the overall performance of the algorithm.

In this thesis, we are seeking a transformation within SVM kernel to transform input patterns to a new space with emphasizing on differences between classes and deemphasizing on similarities. In practice, this is important when we face small learning set and there is a need of model-free representation space. A variety of approaches has been used such as wavelet packet based algorithms [43] (selection of the best discriminative basis) and methods based on time-frequency representations (TFR) [44], [45] (selection of the best TFR within Cohen's class [19]). Also, LPC has been widely used in speech recognition and coding applications [46]. In this section, the effectiveness of the aforementioned features are studied and compared. The most effective feature is employed in developing a new kernel for SVM.

The main contribution of this thesis, in the sense of texture classification, is incorporating an optimized external feature into a new kernel function for the SVM algorithm. The proposed kernel has the form of:

$$\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) = \exp - \frac{1}{2\sigma^2} \left[ \sum_{n=1}^{w^2} \sum_{\Omega=1}^{w^2} |NT_{x_i}(n, \Omega) - NT_{x_j}(n, \Omega)|^2 \right] \quad (4.1)$$

where  $x_i$  and  $x_j$  are the vector forms of the subimages sampled by a windowing operator of the size  $w \times w$ .  $n$  and  $\Omega$  are presenting spatial and frequency domain variables. The notation  $NT_{x_i}(n, \Omega)$  emphasize the normalization of the transform  $T$ :

$$NT_x(n, \Omega) = \frac{|T_x(n, \Omega)|}{\sum_1^{w^2} \sum_1^{w^2} |T_x(n, \Omega)|} \quad (4.2)$$

The transformation  $T$  in 4.1 can be any of the three feature extraction reviewed in

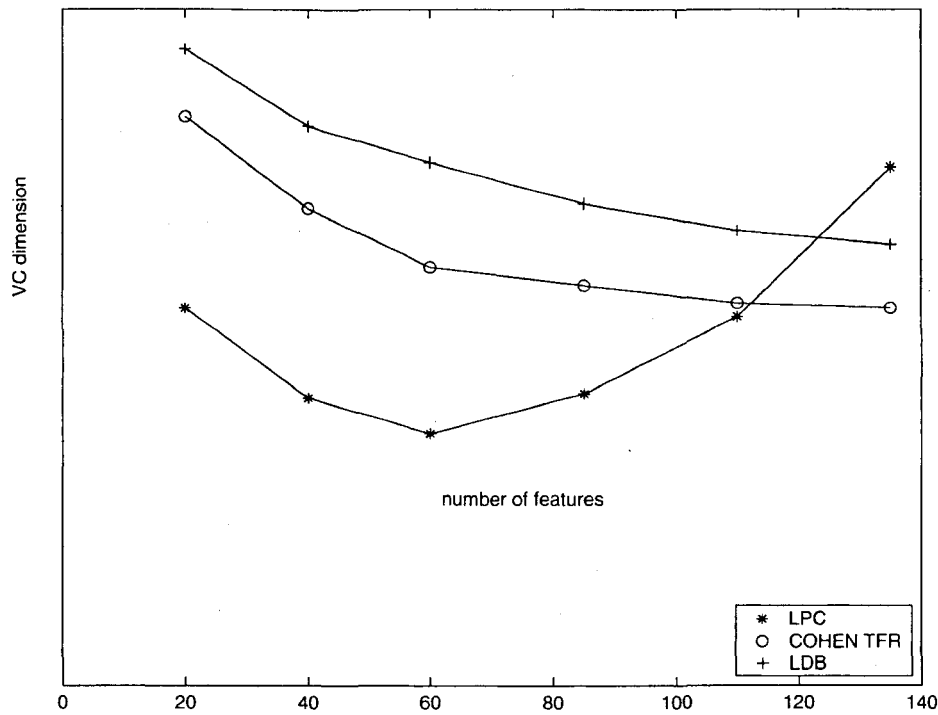
Chapter 3. To select the optimized feature extraction for texture classification the VC-dimension is estimated [6] for each of the three feature extraction. The VC-dimension is principally indicates the largest number of points that can be separated in all possible ways using the functions of the given class [4]. The VC-dimension of the kernel and its parameters can be estimated by the radius  $r$  of the smallest ball containing all the data points in the feature space [47]. The proposed strategy is to select the kernel and parameters that minimize  $r$  and eventually VC-dimension.

The VC-dimension is estimated for each possible two-texture classification among the four texture shown in Fig. 4.1. According to this estimation LPC feature extraction with optimal feature number is adopted to develop the SVM with optimized kernel texture classification. This algorithm is referred as optimized SVM (OSVM) or LPC-SVM algorithm in the rest of this thesis.

## 4.5 Classification and segmentation Results

To verify the effectiveness of the proposed method, experiments were performed on classification and segmentation of several test images. The test images were drawn from two different commonly used texture sources: the Brodatz album [48] and MIT vision texture (VisTex) database [49]. Table 4.2 summarize the source of the test images. All textures are gray-scale images with 256 levels. The classifiers were trained on randomly selected portions of subimages of texture images that are not included in the test subimages. Both the training and test images were globally histogram equalized before being used. Gray scales were linearly normalized into  $[-1, 1]$ .

The results of classification are compared with some other texture classification methods including original SVM [3] as well as logic operators method [38], wavelet transform method



**Figure 4.1:** Average VC-Dimension for three kernel functions based on the three feature extraction reviewed in chapter 3

[37] and optimal Gabor filter method [39].

#### 4.5.1 Comparison with Original SVM method

Images in Fig.4.2 are  $256 \times 256$ . Classifiers were trained by 1000 patterns from each texture. This corresponds to about 1.7 percent of the total available input patterns. The results are compared at different window sizes of  $9 \times 9$ ,  $13 \times 13$ ,  $17 \times 17$ , and  $21 \times 21$ . The original SVM shows the optimal classification rate at window size  $17 \times 17$ . In the proposed LPC-SVM the classification error rate decreases by increasing window size. Classification error rates are presented in Table 4.10 and 4.4. The proposed method outperforms the original SVM specifically in larger window size. This is due to the ability of LPC kernel in inducing more

space-frequency characteristic into feature space at larger window size.

#### 4.5.2 Comparison with Other Classification methods

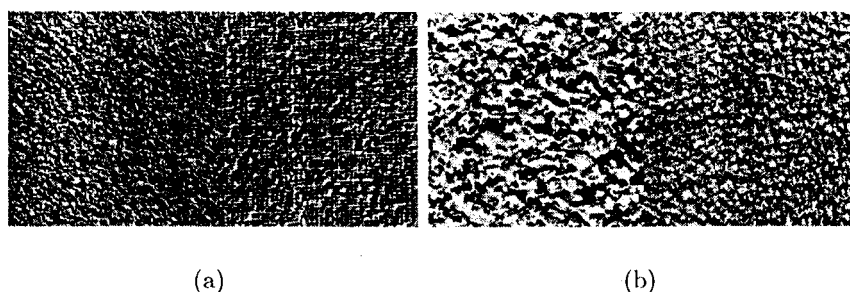
In order to establish the superiority of the LPC-SVM, its performance is compared with the most common techniques in texture classification. The spatial gray level dependence method (SGLDM) or co-occurrence matrix method, Fourier power spectrum method, tree-structured wavelet transform method, Laws texture features and Gabor method are used. The SGLDM estimates the second-order joint conditional probability density functions, written in matrix form and are called co-occurrence matrices. Haralick [50] proposed fourteen statistical features that can be computed from these matrices. Although SGLDM has been proven to perform well for texture classification, the selection of the appropriate distance between pixels and angle for the co-occurrence matrix computation poses a problem and it is also computationally intensive. The Fourier spectrum method has not performed well even in earlier comparisons [51]. The statistical Fourier features of average magnitude, maximum magnitude, energy and features using the zonal masks used in [51] are computed. In the tree-structured wavelet transform (TWT) method, the texture samples are decomposed into multiresolution hierarchy only at nodes where the energy of the decomposed subimages is not significantly smaller than the other subimages at that level. The energy map of the channels is used as a feature vector for classification [52]. Energy features up to four level of decompositions are considered. This method has the drawback of becoming noisy at higher levels of decomposition. Four of the most powerful Laws mask are used to compute the texture energy measures [53]. Gaussian window function is used to compute Gabor transform and the Gabor coefficients are approximated using an optimization criteria [54]. Average energy and residual features are computed with this method. The feature selection process is applied with all methods in order to obtain the best feature set for each algorithm

The Euclidean distance classifier used for all the aforementioned methods. Six different experiment with six texture classes in each case are conducted. The results are given in Table 4.7.

The proposed method is compared with two new method, Logical Operators [38] and wavelet co-occurrence features method [37]. Results are listed in Tables 4.8 and 4.9.

### 4.5.3 Segmentation Results and Comparison

One of the most important application of texture analysis in image processing is segmentation. The results of segmentation using proposed method are shown and compared with optimized Gabor filter method in Fig.4.4.

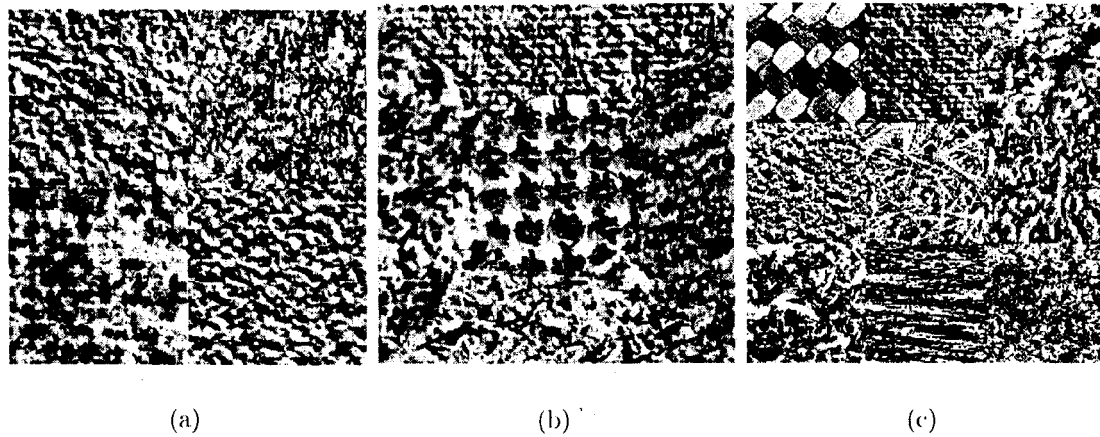


**Figure 4.2:** Two-texture images used in experiments (a) D4, D84 (b) D5, D92( Brodatz album)

**Table 4.2:** Source of test images

Image	Source
Fig4.2a	D4 and D84 from [48]
Fig4.2b	D5 and D92 from [48]
Fig4.3a	D4,D9,D19,and D57 from [48]
Fig4.3b	Fabric.0007, Fabric.0009, Leaves.0003, Misc0002, and Sand.0000 from [49]
Fig4.3c	Fabric.0000, Fabric.0007, Flowers.0005, Food.0005, Grass.0001, Metal.0002, Sand.0000, and Stone.0004 from [49]





**Figure 4.3:** multitexture images used in experiments

**Table 4.3:** Error Rates (percent) for two-texture image (Fig.4.2a).

Parameter	Error Rate %	
	Original SVM	LPC-SVM
9 × 9	12.7	9.6
13 × 13	9.4	7.6
17 × 17	8.6	4.1
21 × 21	13.0	1.2

**Table 4.4:** Error Rates (percent) for two-texture image(Fig.4.2b).

Parameter	Error Rate %	
	Original SVM	LPC-SVM
9 × 9	14.6	14.2
13 × 13	12.1	11.2
17 × 17	11.9	7.3
21 × 21	15.6	5.0

**Table 4.5:** Error Rates (percent) for multitexture image(Fig4.3.a).

Parameter	Error Rate %	
window size	Original SVM	LPC-SVM
$9 \times 9$	22.3	15.2
$13 \times 13$	17.3	11.9
$17 \times 17$	16.1	8.7
$21 \times 21$	21.8	7.1

**Table 4.6:** Error Rates (percent) for multitexture image(Fig4.3.b).

Parameter	Error Rate %	
window size	Original SVM	LPC-SVM
$9 \times 9$	21.8	14.5
$13 \times 13$	20.0	10.3
$17 \times 17$	18.5	7.2
$21 \times 21$	19.7	4.3

**Table 4.7:** Comparison of LPC-SVM with SGLDM,FPS,TWT,LAWS, GABOR method

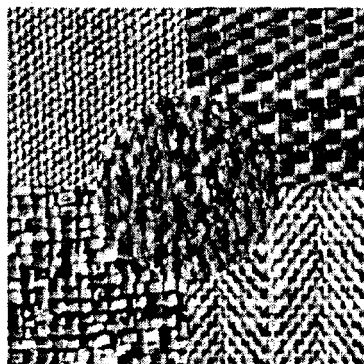
Texture Image r	% Correct Classification with different methods					
Mosaic of 6 textures	LPC-SVM	SGLDM	FPS	TWT	Laws	Gabor
D94,D101,D36,D84,D103,D56	97	67	54	63	52	62
D28,D20,D9,D38,D50,D57	98	84	75	62	84	70
D90,D74,D93,D34,D65,D53	96	64	56	62	47	67
D105,D79,D82,D52,D19,D78	99	59	54	53	46	55
D28,D9,D57,D24,D4,D38	99	67	56	65	77	63
D103,D105,D12,D78,D79,D82	97	77	56	58	58	59

**Table 4.8:** Comparison of Correct Classification Rate in Logic Operator Method and LPC-SVM

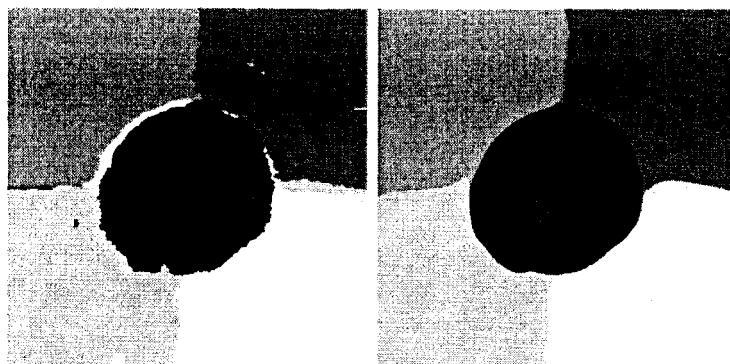
Texture	Logic Operator Method	LPC-SVM
D15	89	89
D19	97	100
D52	81	100
D65	84	100
D74	73	81
D79	92	98
D82	86	98
D84	72	95

**Table 4.9:** Comparison of Correct Classification Rate in Wavelet Transform Method and LPC-SVM

Texture	Wavelet transform Method	LPC-SVM
Bark.0006	92.86	89.0
Clouds.0001	94.0	100
Fabric.0017	97.6	100
Grass.0001	78.6	96.1
Leaves.0012	91.7	94.2
Misc.0002	97.7	100
Sand.0002	96.4	100



(a)



(b)

(c)



(d)

(e)

**Figure 4.4:** multitexture images segmentation:(a) Five texture Brodatz image originally published by Jain and Farrokhnia(b) segmentation with LPC-SVM before post processing (c) segmentation with LPC-SVM after post processing (d) segmentation using Optimized Gabor Filter (e)segmentation using Optimized Gabor Filter (sigmoidal activation)

## 4.6 Gabor Filter Banks

Multichannel filtering is an effective in texture classification and is widely used. Filter banks have the ability of decompose an image into relevant texture features suitable for the purpose of classification. Multichannel filtering is motivated by its ability to mimic Human Visual System (HVS). Hubel and Wiesel [55] showed that simple cells in retina are sensitive to specific orientations with approximate bandwidth of  $30^\circ$ . In addition to sensitivity, the HVS has spatial frequency sensitivity [56]. This consequently led to a HVS model consist of independent detector mechanism each preceded by a relatively narrow band Filter tuned to a different frequency. Experiments indicate that the frequency bandwidth of simple cells in the visual cortex is about one octave [57].

In this way, Gabor filters are motivated to be used in the filter bank due to their ability to be tuned into various oriental and spatial, spatial frequency characteristic. Spatially, a Gabor function is a Gaussian modulated by a sine or cosine:

$$h(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{-\frac{1}{2}\left[\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right]\right\} \cos(2\pi Fx) \quad (4.3)$$

And in the frequency domain:

$$H(u, v) = \exp\{-2\pi^2[(u - F)^2\sigma_x^2 + v^2\sigma_y^2]\} + \exp\{-2\pi^2[(u + F)^2\sigma_x^2 + v^2\sigma_y^2]\} \quad (4.4)$$

We introduce Gabor Filter bank in this section to note that the ability of LPC-SVM algorithm in dealing with highly non-stationary and non-periodic signals, can be improved by adding gabor filter bank features. It is believed that gabor filter banks have an excellent ability in extracting non-stationary features. The main challenge in deploying Gabor filter

bank is proper and efficient feature selection. In the proposed method, LPC is used and the end of each filter for feature extraction. Such a feature extraction is not efficient for most of the classifiers and learning algorithms. Considering the ability of SVM in high dimensional feature spaces, the resulting algorithm is an excellent pattern recognition algorithm.

**Table 4.10:** Error Rates (percent) for two-texture image (Fig.4.2a).

Parameter	Error Rate %		
	Original SVM	LPC-SVM	LPC-SVM with Gabor
9 × 9	12.7	9.6	5
13 × 13	9.4	7.6	3.9

## Chapter 5

# Detection of Abnormalities In Mammograms

BREAST cancer is a major problem of public health in the western world, where it is by far the most common cancer among women. Each day, breast cancer calls for approximately one hundred death in North America. As shown in Table 5.1 the rate of mortality is also considerably high in other developed countries.

**Table 5.1:** Rate of Breast Cancer mortality in some developed countries.

Country	Mortality ( per 100 000)
Canada	65.1
France	35.5
Germany	44.5
Japan	10.7
Korea	3.9
Singapore	13.8

On out of eight women over 40 years of age, develops one type of breast cancer. There is no way to prevent or cure breast cancer and the etiologies of breast cancer are unclear and no single dominant cause has emerged. The only way to save the life of patient is early detection and removing cancerous tissues before spreading to the other parts of body. The risk of

mortality increases if the cancerous tumors are not detected early. Thus, early diagnosis and treatment has a critical role in increasing the chance of survival. X-ray mammography is the most effective, low cost and highly sensitive method capable of detection of breast cancer in asymptomatic woman at its early stage [58].

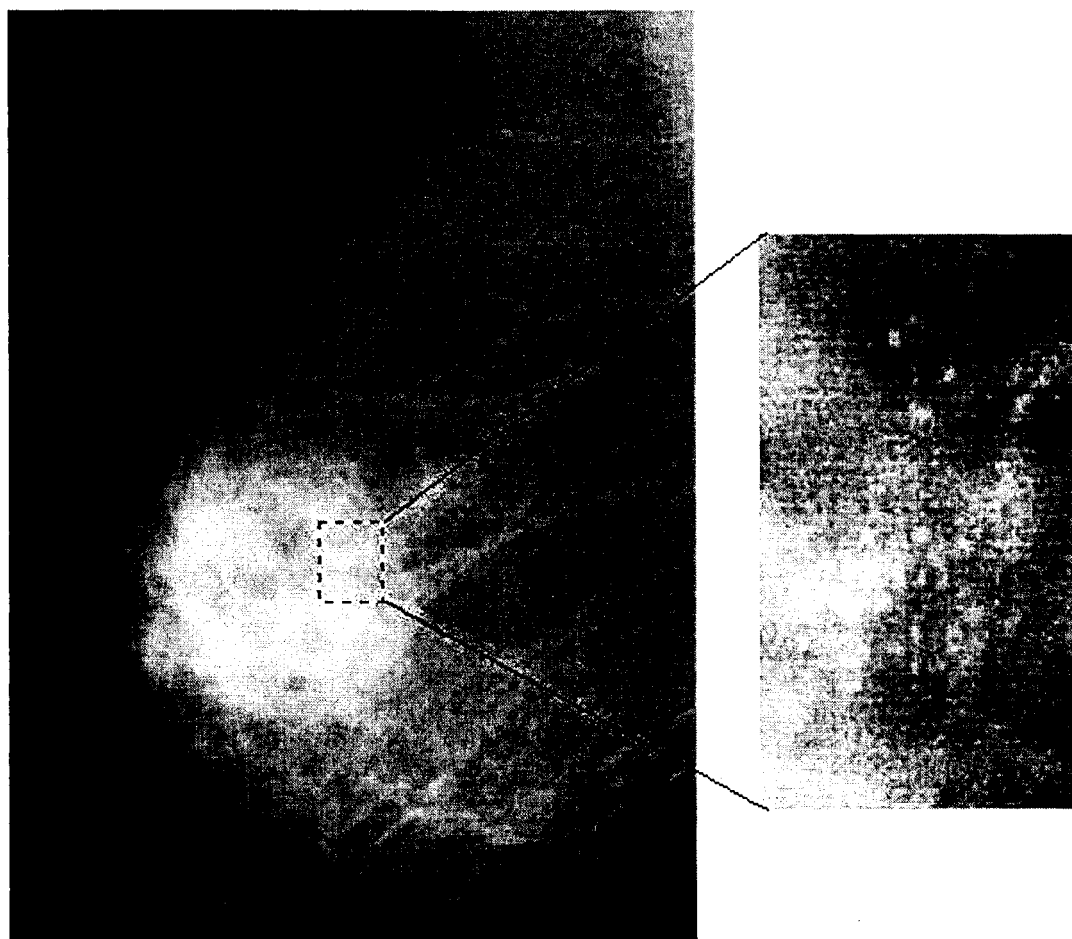
There are two types of mammography: screening and diagnostic. The objective of screening mammography is to detect cancer when it is still small in asymptomatic woman. The diagnostic mammography is an X-ray examination of the breast of a woman who has either breast cancer symptoms or an abnormality in her screening mammography. An increasing number of developed countries have started screening mammography. Women over 40 years of age are recommended to obtain mammograms regularly for screening purpose. This results to a huge number of mammogram that needs to be examined and interpreted by expert radiologists.

## 5.1 Motivation for a CAD system

Among the various types of breast abnormalities that are visible in the mammograms, clustered microcalcification and mass lesions are the most important ones. Masses and clustered microcalcifications often characterize early breast cancer that can be detected in mammograms before a woman or physician can palp them. Among these two early signs, the cluster microcalcification is more difficult to be detected.

A microcalcification is a small granule-like calcium deposit that has accumulated in the breast tissue, and it appears as a small bright spot embedded within an inhomogeneous background of the mammogram. Clustered microcalcification is defined by radiologist as the presence of three or more visible microcalcifications within a square of  $1\text{ cm}^2$ . It is very difficult to interpret X-ray mammograms because of the aforementioned tiny size of microcalcification and the very small differences in the image densities of various breast tissues, specially in dense breast. It is estimated that radiologists fail to detect approximately





**Figure 5.1:** Sample of microcalcification in a mammogram. The microcalcification is shown in the right after zooming and enhancement.

25% of the cancers that are visible on retrospective review [2]. This significant failure rate is due to visual fatigue, inexperienced radiologist and noise and lack of contrast in mammograms. Screening mammography also results in a huge number of mammograms which must be examined by experienced radiologist. Manual reading, the current examination procedure, is labor intensive, time consuming and demands great concentration. To help radiologist to overcome this huge burden more accurately and in a reasonable time, several CAD (Computer Aided Diagnosis) systems have been proposed and tried.

Indeed, the goal of CAD systems is not to create a high-tech radiologist or automate the

detection procedure. CAD systems are intended to act as the first cut in the process and a second opinion for the radiologists. The term "second Opinion" means that the radiologists can use the results of a computer analysis of the mammogram in making a diagnosis. We try to locate suspicious regions in the mammogram for more detailed examination by the radiologists.

## 5.2 Previous work

Several CAD systems have been designed for extracting abnormalities in breast X-ray images. A few number of the most popular and recently proposed methods are briefly reviewed in the following. The reviewed methods and algorithm are summarized and listed in Table 5.2.

### 5.2.1 Shape Analysis

Shen *et. al.* [59] developed a set of shape factors to measure the roughness of contours of calcifications in mammograms and for use in their classification as malignant or benign. The analysis of mammograms is performed in three stages. First, a region growing technique is used to obtain the contours of calcifications. Then, three measures of shape features, including compactness, moments, and Fourier descriptors are computed for each region. Finally, their applicability for classification is studied by using the three shape measures to form feature vectors. Classification of 143 calcifications from 18 biopsy-proven cases as benign or malignant using the three measures with the nearest-neighbor method was reported 100% accurate.

### 5.2.2 Iterative feature extraction

Karssemeijer [60] implemented an adaptive noise equalization and statistics-based model for the detect of microcalcification. As a pre-processing step, a robust rescaling was applied

as a filter adaptive noise equalization, which estimated noise characteristic from the image at hand. This system also took into account the impact of the variability in the tissue across the image. Microcalcifications were detected in these rescaled images using an initial segmentation and an iterative process used to update the pixel labels. The labelling process used filtered versions of the mammogram, each of which depicted a local image feature thought to be important in distinguishing a microcalcification. This work used 25 training images and another 40 for testing, from the Nijmegen database. These images all contained one or more known clusters of microcalcifications labelled by an expert radiologist. The images were digitized at 12 bits of gray-scale and to a size of  $2048 \times 2048$  pixels.

### 5.2.3 Thresholding and morphology

Nishikawa *et. al.* [61] proposed an algorithm including three basic steps. First, the digital mammogram is enhanced using a spatial filter. This step tries to enhance the signals from microcalcification and suppress the signals from background structures of the breast. Second, global and local gray-level thresholding and a morphological erosion is used to extract potential microcalcifications from the image. Third, the feature analysis are used to reduce the number of false positive (FP) detections. The algorithm has been tested on a database (10bits/pixel, 0.1 mm/pixel) of 78 mammograms with size of  $800 \times 1000$  pixels. Half of the images without microcalcification and half of them with at least one microcalcification for a total of 41 microcalcifications. Visual interpretation was used for determining the true positive rate and the average number of false positives per image. A true positive rate of %85 were obtained at an average false positive detection of one per image.

### 5.2.4 Multiscale matched filters

Strickland and Hahn [62] developed a wavelet transform algorithm which acted as a bank of multiscale matched filters for the clustered microcalcification. By studying the cross

section of clustered microcalcifications, Strickland and Hahn concluded that it is reasonable to model a microcalcification using a circularly-symmetric Gaussian function. Thus, the matched filters were designed for detecting Gaussian objects in correlated Markov noise. The Markov noise were adopted as the model of background and normal tissues. The method was tested on a database of 40 mammograms digitized at 12 bits/pixel and  $100\text{ }\mu\text{m/pixel}$ . The data based consisted of 40 mammograms with size  $2048 \times 2048$  pixels. Each mammograms contained at one or more clusters microcalcifications. A cluster was considered detected if two or more microcalcifications were found within the truth circle. An False Positive (FP) was counted if two or more erroneous detections were made within an empty, closed region of 0.5 in width. The results showed that the method could achieve a True Positive(TP) rate of about 55% with 0.7 FP detection per image. Alternatively, a TP rate of about 85% with a corresponding 3.2 FP per image.

### 5.2.5 Statical feature of surrounding region

Kim and Park [28] proposed a method based on statistical textural features. In their method, four features named horizontal-weighted sum, vertical-weighted sum, diagonal-weighted sum and grid-weighted sum are computed based on the surrounding region matrix. Surrounding region matrix presents softness or coarseness for each Region Of Interest (ROI). The extracted features are feed into a three-layer back-propagation neural network for classification. 120 X-ray mammograms were selected from the patient files based on visual criteria and biopsy results. The mammograms were digitized with pixel size of  $100\text{ }\mu\text{m}$  and 12 bits per pixel. 172 ROI's were selected to evaluate the method. 72 ROI's contained microcalcification and 100 ROI's without microcalcification. 86 ROIs were used for training including 36 ROI's with microcalcifications. A fraction of 0.8 TP were reported at 0.4 FP.

### 5.2.6 Laplacian scale-space

Netsch and Peitgen [63] developed a method based on the Laplacian scale-space representation of the mammogram. Their motivation for applying Laplacian filtering originates on the fact that microcalcifications have an small bright circular appearance or slightly elongated spots in the mammogram. Bright spots corresponds to local maxima in Laplacian filtered images if the size of the filter kernel is chosen appropriately. Netsch and Peitgen applied their method on database of 40 mammograms digitized at 12 bits/pixel and 100  $\mu\text{m}$ /pixel. The database consisted of 40 mammograms with size  $2048 \times 2048$  pixels. Each mammograms contained one or more clusters microcalcifications. They achieved 85% TP detection at 1 FP per image.

### 5.2.7 Spatial statistical features and wavelet features

Yu and Guan [64] used a mixture of wavelet features and gray level statistical features. A set of 15 features were selected among 31 features by measuring the discriminatory power of the features using general regression neural networks via sequential forward and sequential backward selection method. The method were applied to a database of 40 digitized mammograms of 21 patients. The database included 105 microcalcification and were digitized with 12 bit per pixels. The true positive of 90% were reported at 0.5 false positive per image. It must be noted that test data were used in feature extraction step. This must be considered, when the results of this work is compared with other above mentioned methods.

### 5.2.8 Support Vector Machine

El-Naqa *et al.* [65] applied SVM for detection of clustered microcalcification in digital mammograms. In this approach SVM was used without deploying any external feature

extraction. In preprocess phase, a linear-phase finite impulse response high-pass filter with 3-dB cutoff frequency  $\omega_c = 0.125$  was applied to digital mammograms. The gray level pixels were directly applied to SVM after preprocessing on the digital mammograms. Using the gray level pixels without prior pre-process stage causes problem in training of SVM due to the very large number of training sample presenting normal tissue in comparison with the few number of training samples presenting microcalcifications. El-Naqa *et al.* [65] proposed an algorithm referred as successive enhancement-learning(SEL) to reduce the training samples without microcalcification.

The algorithm were evaluated using a database provided by the Department of Radiology at The University of Chicago. The data set contained 76 clinical mammograms, all containing multiple microcalcifications. The mammograms were of dimension  $1000 \times 700$  pixels, with a spatial resolution of 0.1mm/pixel and 10 bits gray scale. The database included 1120 microcalcifications detected by expert radiologists at the Department of Radiology of University of Chicago. A correct rate of 94% were reported at one false positive per image.

The authors compared their algorithm with several other methods and claimed its superiority.

### 5.2.9 Multi-resolution Based Segmentation

S. Sentelle *et al.* [66] proposed a multi-resolution approach combined wavelet analysis to provide a segmentation of potential calcifications. An Initial multi-resolution approach to fuzzy c-means(FCM) segmentation was employed. Some tissue areas were chosen in each image and were broken into multiple windows. Within each window, wavelet analysis was used to generate a contrast image, and a local FCM segmentation generated an estimate of local intensity. A simple two-rule fuzzy system thereafter combined intensity and contrast information to derive fuzzy memberships of pixels in the high-contrast, bright pixel class. A double threshold is applied at the end to this fuzzy membership to detect and segment

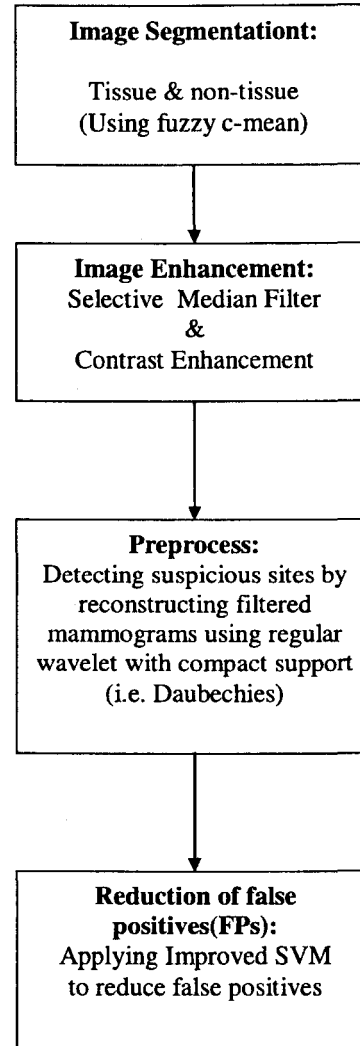
calcification. The algorithm was applied to 25 images obtained from the Digital Database for Screening Mammography(DDMS) provided by the University of Southern Florida.

**Table 5.2:** Summary of related research in calcification detection

Authors	Relevant techniques	Image sources
L. Shen <i>et al.</i> [59]	Shape Analysis	MIAS Database
Karssemeijer [60]	Iteratively segmenting filtered mammograms after denoising the original image	65 image from Nijmegen database
Nishikawa [61]	Detection of potential calcification by global and local gray-level thresholding and a morphological erosion and reducing false positive rate by feature analysis.	78 mammograms from department of radiology, University of Michigan
Strickland and Hahn [62]	Designing matched filters for detecting Gaussian objects in correlated Markov noise	40 images from Nijmegen database
Kim and Park [28]	Extracting four texture features from surrounding region matrix and using three-layer back-propagation neural network for classification.	Department of radiology, Asan Medical centre, Korea
Netsch and Peitgen [63]	Obtaining local maxima (bright spots) in the Laplacian scale-space representation of the mammogram.	40 image from Nijmegen database
Yu and Guan [64]	Selecting the best of 15 features among 31 features by measuring the discriminatory power of the features using general regression neural networks and using sequential forward and sequential backward classification method.	40 images of 21 patient from Nijmegen database.
El-Naqa <i>et al.</i> [65]	Applying support vector machine to high-pass filtered mammograms and reducing the samples presenting normal tissues by successive enhancement-learning method.	Database of mammograms from the Department of Radiology, University of Michigan
S. Sentelle <i>et al.</i> [66]	Using fuzzy c-mean segmentation on the multiresolution images followed by reconstructing the filtered image by bio-orthogonal wavelet.	25 images from Digital Database of Screening Mammography(DDSM) at the University of Southern Florida.



### 5.3 Proposed Algorithm



**Figure 5.2:** Overall flow of the developed algorithm for detection of microcalcification

The raw data in digital mammogram are not appropriate for learning algorithms such as SVM. The very high resolution and noisy images result in huge computation and inaccurate classification. Noise suppression and reduction methods are targeting image enhancement while trying to preserve microcalcification. On the other hand, while the image enhancement techniques try to ease detection of tiny calcification embodied in dense tissues, they are

potential of enhancing noise as well. To reduce the number of samples prior to classification, a highly sensitive algorithms can be used to detect suspicious sites in the mammogram. Our proposed methods includes three steps:( as shown in Fig. 5.2)

- a) Segmentation of image to tissue and non-tissue regions.
- b) Image enhancement
- c) Detection of suspicious sites
- d) Applying the improved SVM to sample of windows centered at the suspicious sites.

In the following subsections the aforementioned steps are discussed in more details.

### 5.3.1 Image segmentation

The proposed algorithm begins by separating the mammographic tissue from non-tissue regions in the image. The fuzzy c-mean (FCM) [67] is used in this thesis, which is basically an unsupervised least-squares clustering algorithm.

The FCM is similar to hard c-mean (HCM) approach but employs fuzzy membership when labelling data points rather than assigning crisp labels as in HCM. In FCM, a set of initial cluster centres are specified and an iterative process begins to adjust initial centres and calculate the fuzzy membership of all data points to the corresponding centres. At the end, maximal membership is employed for determination of crisp labelling from the fuzzy membership.

FCM is initiated by assigning an initial class label to each data point determined from the initialization routine. This assignment is stored in membership matrix ( equation 5.1). In this equation,  $k$  is an index into the image for data point  $x_k$ , and  $(i,j)$  are the class indices for initial centres  $v$ . The number of classes is indicated by  $c$ . The  $m$  is the fuzzification value.

$$u_{i,k} = \left[ \sum_{j=1}^c \left( \frac{\|x_k - v_i\|^2}{\|x_k - v_j\|^2} \right)^{\frac{2}{m-1}} \right]^{-1} \quad \forall i, k \quad (5.1)$$

This  $\mathbf{U}$  matrix is then used to calculate cluster centres of the data points for each class label, which are stored in matrix  $\mathbf{V}$  (Eq. 5.2). In this equation,  $i$  is the cluster centre number,  $n$  is the number of pixels,  $x_k$  is the pixel value, and  $m$  is the fuzzification index.

$$v_i = \left( \frac{\sum_{k=1}^n u_{ik}^m x_k}{\sum_{k=1}^n u_{ik}^m} \right) \quad \forall i \quad (5.2)$$

From the new cluster centres, the  $\mathbf{U}$  membership matrix is recalculated. Once the  $\mathbf{U}$  matrix has been recalculated, the cluster centres are then recalculated. This process continues until the measure of the distance between old cluster centres and each set of new cluster centres meets a threshold criterion. At termination, the cluster centres for newly discovered data clusters as well as labels for the data points can be retrieved.

however, this algorithm is computationally expensive due to its memory requirements, iterative nature, and slow convergence for large data sets such as high resolution mammograms. In this thesis, to reduce the data points processed by the algorithm, a multi-resolution approach is employed. The image is first down sampled by a factor of 16 and the algorithm is applied to the down sampled image. This decrease the convergence time of the procedure significantly, while the accuracy of the segmentation is still acceptable for the purpose of segmenting image to tissue and non-tissue. Local and global thresholding is finally applied to the image.

### 5.3.2 Enhancement of the mammograms

The fundamental enhancement needed in mammography is an increase in contrast, especially for dense breast. Contrast between malignant tissues and normal dense tissue may be present on a mammogram but below the threshold of human perception. Similarly, micro-

calcifications in a sufficiently dense mass may not be readily visible because of low contrast. As a result, defining the characteristics of microcalcifications is difficult.

Conventional image processing techniques do not perform well on mammographic images [68]. In this thesis, selective median filtering is used to remove background noise while preserving the edge information of suspicious areas. This approach was proposed by Lai *et al.* [69]. A selective median filter is defined as follows:

Given a window  $W(i, j)$ , centered at image coordinates  $(i, j)$ , the output of the selective median filter is:

$$\hat{x}_{i,j} = \text{median}\{x_{r,s} \in N(i, j), \text{ and } |x_{r,s} - x_{i,j}| < T\} \quad (5.3)$$

where  $x_{i,j}$  is the image intensity at  $(i,j)$ ,  $N(i,j)$  is the area in the image covered by window  $W(i, j)$ , and  $T$  is the threshold. In computing the median, the set of pixels are restricted to those with a difference in gray level no greater than the threshold  $T$ . The amount of edge smearing can be controlled by adjusting the parameter  $T$ . If  $T$  is small, the edge preserving power of the filter is strong, but its smoothing effect is small. If  $T$  is large, the filter behaves the other way around.

In the next step of enhancement the image's contrast is improved by contrast-to-noise ratio method [70]:

$$\hat{x}(i, j) = \frac{x(i, j) - \text{mean}(y(r, s), r, s \in \text{Window})}{\text{std}(y(l, m), l, m \in \text{Window})} \quad (5.4)$$

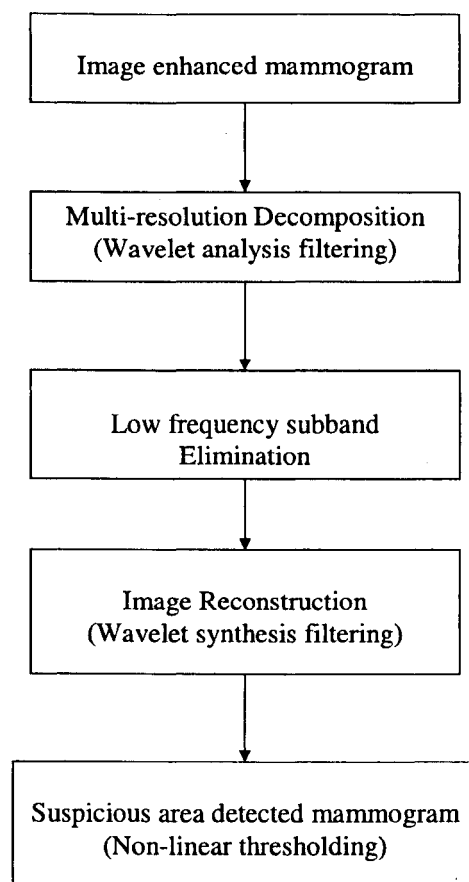
where  $x(i, j)$  is the pixel value at the position  $(i, j)$ , and "Window" is an  $9 \times 9$  square area centered at position  $(i, j)$ ,  $\text{std}$  is the standard deviation of the pixel values in the "Window",  $\hat{x}(i, j)$  is the normalized gray level value at position  $(i, j)$ .

### 5.3.3 Preprocessing

The preprocessing detection of suspicious area is based on the hypothesis that the microcalcification present in mammograms can be preserved under a transform which can localize the signal characteristics in the original and transform domain. In a time signal the harmonic frequency components are present but they are hidden, whereas in its frequency spectrum the time information is hidden. Therefore, transforms with basis functions other than the complex sinusoids must be used. In addition, these basis functions must be able to localize the signal in both spatial and frequency domains. A suitable transform that satisfies the above requirements is the wavelet transform. The wavelet transform uses basis functions that can dilate in scale and translate in position according to the signal characteristics. Given that the microcalcifications correspond to high frequency components of the image spectrum and wavelets can localize signal characteristics in both frequency and scale, our hypothesis is that the resolution and scale of the microcalcifications in the spatial domain can be preserved if we use wavelet filters to decompose the mammogram into different frequency subbands. Accordingly, microcalcifications can be extracted from mammograms by suppressing the subband of the wavelet-decomposed image that carries the lowest frequencies and contains smooth (background) information, before the reconstruction of the image.

The proposed system is described in the block diagram shown in Fig. 5.3. The original mammogram is decomposed into a set of orthogonal subbands of different resolution and frequency content. The decomposition is based on wavelet analysis filtering and downsampling along the rows and columns of the image. Fig. 5.4 shows the seven subbands of resulting after two levels of wavelet decomposition of the image.

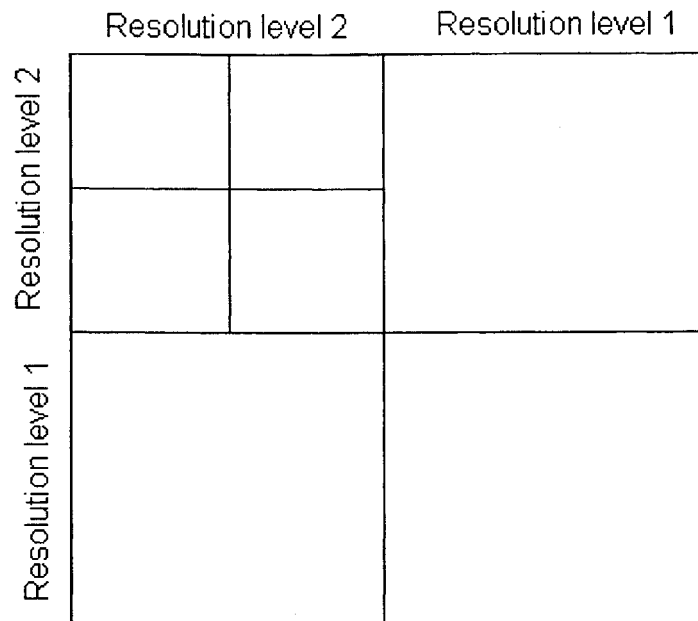
The four subbands at resolution 1 are produced by the decomposition scheme described in Section 3.1.1. The application of the same decomposition scheme to the upper-left subband that carries the lowest frequencies at resolution 1 results in the two level subband decomposition shown in Fig. 5.4. In the wavelet decomposed image shown in Fig. 5.4 the upper-left



**Figure 5.3:** Preprocessing steps

subband at resolution level 2 contains the background intensity of the original image and, thus, carries the lowest frequencies of the image spectrum. The microcalcifications, which corresponds to the highest frequencies, are carried by other subbands. The detection of microcalcifications is accomplished by setting the wavelet coefficients of the upper-left subband to zero in order to suppress the image background information before the reconstruction of the image. The reconstructed mammogram is expected to contain only high-frequency components, including the microcalcifications. The reconstruction consists of wavelet synthesis filtering and up-sampling along the rows and columns of the image.

The wavelet filters used in analysis and synthesis stages, are maximally flat wavelet filter constructed by Daubechies [71]. These wavelets are compactly supported and regular.



**Figure 5.4:** Orthogonal subbands at different resolutions

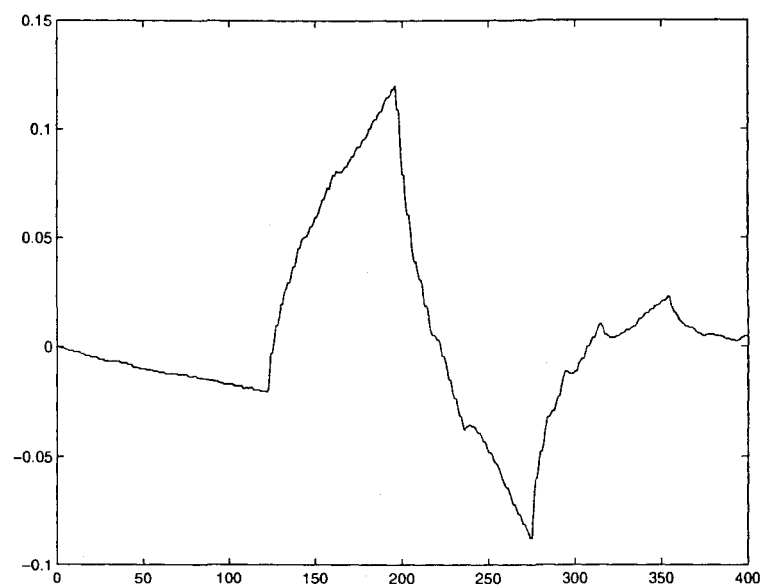
Wavelets are compactly supported if they have finite support with maximum number of vanishing moments for their support width. The compact support improves the time resolution of wavelets. Table 5.3 shows the filter coefficients of the two wavelets from Daubechies 4-coefficient (DAUB 4) filter and Daubechies 20-coefficient (DAUB 20) filter. In this table,  $g(n)$  is the low-pass filter and  $h(n)$  is the high-pass filter as shown in Fig.3.4. The high-pass filter can be obtained as  $h(n) = (-1)^n g(n)(-n + 2N - 1)$  where the  $N$  is the length of the filter. Fig 5.5 shows the amplitude plot of the mother wavelet  $\psi$  for the family of DAUB 4 filters. Fig.5.5(a) represent the long window used to analyze long term behavior of a signal, whereas Fig. 5.5 (b) is the scaled and translated version of the same wavelet used to analyze the instantaneous behavior of a signal. Fig. 5.6 shows the amplitude plot of the mother wavelet  $\psi$  for the family of DAUB 20 filters. In each case, note that the stretched wavelets have higher amplitudes while the dilated wavelets have lower amplitudes. The results of the preprocess stage is shown in Fig. 5.6. All microcalcifications present in the original mammogram are visible in the images produced by the proposed preprocess. The performance of

the preprocess depends on the length of the wavelet filters used in the decomposition of the mammograms. This is a direct sequence of the different shapes of the corresponding mother wavelets. According to Fig. 5.5 and 5.6, the mother wavelet of the DAUB4 filter is more spike-like compared with that of smoother DAUB20 filter. It is clear from these images that the DAUB4 filter detects more pixels of high spatial frequency compared with the DAUB20 filter. These pixels may belong to microcalcifications, breast boundary, or background noise. Thus, shorter wavelet filters are more sensitive to existing microcalcifications but they tend to produce more false positives. Therefore, the DAUB4 is used in this thesis to have all potential microcalcifications in the next step( supervised learning).

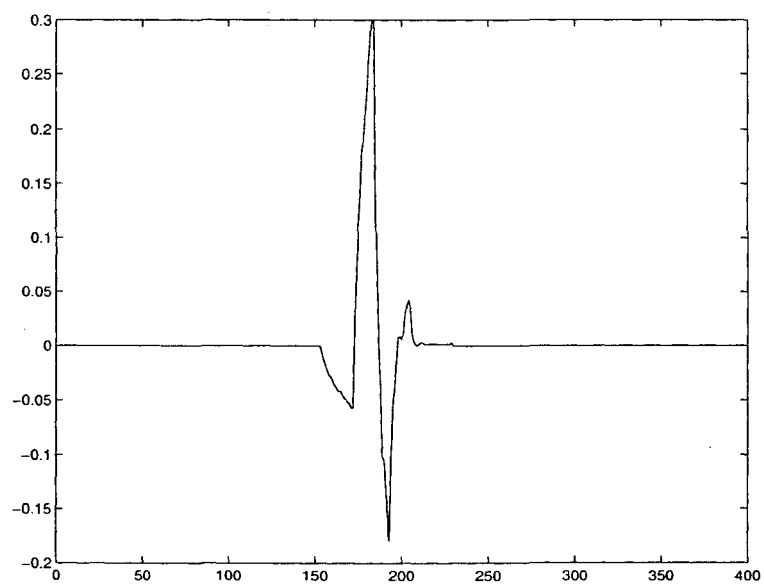
**Table 5.3:** Coefficients of the DAUB filters.

	DAUB4	DAUB20
n	g(n)	g(n)
0	0.482691	0.026670
1	0.836516	0.188177
2	0.224143	0.527207
3	-0.129409	0.688459
4		0.281172
5		-0.249816
6		-0.195946
7		0.127369
8		0.093057
9		-0.071394
10		-0.029457
11		0.033212
12		0.003606
13		-0.010733
14		0.001395
15		0.001992
16		-0.000686
17		-0.000116
18		0.000093
19		-0.00013



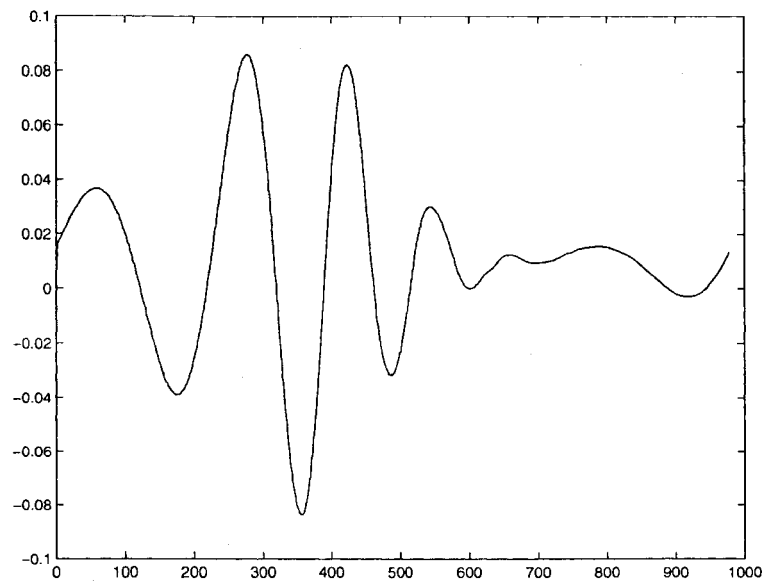


(a)

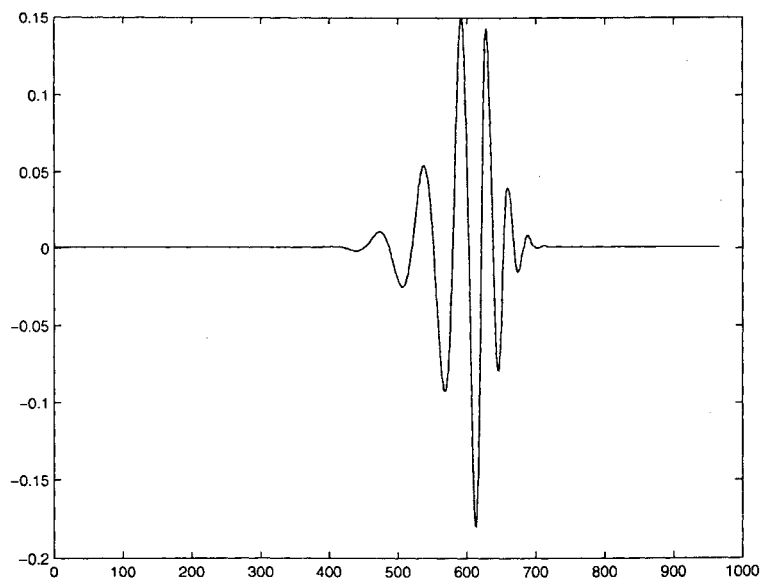


(b)

**Figure 5.5:** Mother wavelet for DAUB 4, a narrow and tall wavelet for analyzing high-frequency characteristics



(a)



(b)

**Figure 5.6:** Mother wavelet for DAUB 20, (a) a wide and short wavelet for analyzing low-frequency characteristics (b) a narrow and tall wavelet for analyzing high-frequency characteristics

### 5.3.4 Reduction of False positives through supervised learning

The pre-processed images contain false positives which are presenting noise and other high-frequency components rather than microcalcification. In this step, SVM in conjunction with LPC is used to implement a supervised learning machine to reduce the number of false positives. The detailed procedure is discussed in the following.

#### 5.3.4.1 Input features

For every pixel in the pre-processed image, we define the input pattern to LPC-SVM to be an  $M \times M$  window centered at the pixel detected in preprocess. These windows are selected from enhanced mammogram. The  $M \times M$  matrix is converted to a vector  $\mathbf{X}$  of length  $M^2$  and then its length is reduced to  $N < M^2$  by using LPC feature extraction method. During the training phase, each input feature  $\mathbf{X}$  is labelled with ( $y = +1$ ) for microcalcification present, or ( $y = -1$ ) for microcalcification absent.

#### 5.3.4.2 Model selection and SVM training

Once the training samples are gathered, the next step is to determine the SVM decision function. In this process, the following parameters must be determined:

- a) Window size,  $M$
- b) LPC features,  $N$
- c) The type of kernel function
- d) The regularization parameter in the structural risk function,  $C$

To optimize these parameters, m-fold cross validation [72] is applied to training data set. This procedure consists of the following steps.

First, divide randomly all the available training examples into  $m$  equal-sized subsets.

Second, for each model-parameter setting, train the SVM classifier  $m$  times; during each time of the  $m$  subsets is held out in turn while all the rest of the subsets are used to train the SVM. The trained SVM classifier is then tested using the held-out subset, and its classification error is recorded.

Third, the classification errors are averaged to obtain an estimate of the generalization error of the SVM classifier. In the end, the model with the smallest generalization error will be adopted.

In the next section, the performance of the overall algorithm is evaluated using Receiver Operating Characteristic (ROC) analysis.

## 5.4 Experimental Results

### 5.4.1 Database

The research results in this thesis is based on the data sets derived from *Digital Database for Screening Mammography (DDSM)*, located at the University of South Florida [73]. The DDSM contains approximately 2500 cases of fully annotated mammographic images. Each case contains two views of each breast along with patient information such as age at the time of study, American College of Radiology (ACR) breast density rating, subtlety rating for abnormalities, and ACR description of each abnormality. Information about how the mammograms were digitized is also included, such as relative spatial resolution and scanner used. The DDSM is organized into "cases" and "volumes". A "case" is a collection of images and information corresponding to one mammography exam of one patient.

A case consists of between 6 and 10 files. These are an "ics" file, an overview "16-bit PGM" file, four image files that are compressed with lossless JPEG encoding and zero to four overlay files. Normal cases will not have any overlay files.

### Structure of ".ics" files:

The ".ics" file provides information about a case as a whole. In ASCII format, it lists

```
ics_version 1.0
filename B-3024-1
DATE_OF_STUDY 2 7 1995
PATIENT_AGE 42
FILM
FILM_TYPE REGULAR
DENSITY 4
DATE_DIGITIZED 7 22 1997
DIGITIZER LUMISYS
SELECTED
LEFT_CC  LINES 4696 PIXELS_PER_LINE 3024 BITS_PER_PIXEL 12 RESOLUTION 50 OVERLAY
LEFT_MLO LINES 4688 PIXELS_PER_LINE 3048 BITS_PER_PIXEL 12 RESOLUTION 50 OVERLAY
RIGHT_CC LINES 4624 PIXELS_PER_LINE 3056 BITS_PER_PIXEL 12 RESOLUTION 50 OVERLAY
RIGHT_MLO LINES 4664 PIXELS_PER_LINE 3120 BITS_PER_PIXEL 12 RESOLUTION 50 OVERLAY
```

**Figure 5.7:** Sample ics file

important information such as the date of the study, the patients age, the ACR breast tissue density, the date of digitization of the films, the type of digitizer used and a list of the image files. The ".ics" file also gives an rating of 1 to 4 as assessed by an expert radiologist.

The size of each image file, number of bits per pixel, the scanning resolution (in microns) and information on the existence or lack of an overlay file for each image is provided. As it can be seen in Fig. 5.7, all four images have overlays. If the image description lines had "NON-OVERLAY" instead of "OVERLAY" then the images would not have overlay files.

### Structure of Overlay files

Abnormal cases have between one and four overlay files depending on the number of images in which the radiologist marked any abnormalities. Each overlay file may specify multiple abnormalities, so the first line of the file gives the total number of abnormalities. In the case of multiple abnormalities, each abnormality is then listed one after another. Each abnormality has information on the lesion type, the assessment, the subtlety, the pathology and at least one outline. The keywords that describe the lesion type are taken from the ACR Bi-RADS lexicon. The assessment code is a value from 1 to 5, and also comes from the ACR Bi-RADS standard. The outlines for the suspicious regions are derived from markings

made on the film by an experienced radiologist. Each boundary is specified as a chain code. This chain code is found on the line after the keyword "BOUNDARY" or "CORE". The first two values of each chain code are the starting column and row of the chain code that order. Following these two numbers, the chain code is given and a "#" character indicates the end of the chain code. The numbers correspond to the directions as shown in Fig. 5.8.

	<b>-X-&gt;</b>		
I	7	0	1
Y			
I	6	X	2
V	5	4	3

<b>Chain code value</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
<b>X Coordinate</b>	0	1	1	1	0	-1	-1	-1
<b>Y coordinate</b>	-1	-1	0	1	1	1	0	-1

Figure 5.8: Chain codes values and directions

```

TOTAL_ABNORMALITIES 1
ABNORMALITY 1
LESION_TYPE CALCIFICATION TYPE PLEOMORPHIC-FINE_LINEAR_BRANCHING DISTRIBUTION REGIONAL
ASSESSMENT 5
SUBTLETY 4
PATHOLOGY MALIGNANT
TOTAL_OUTLINES 4
BOUNDARY
8 1368 4 4 4 4 4 4 4 2 2 2 2 2 2 2 ... 0 0 0 0 0 0 0 0 1 #
    CORE
168 1824 2 2 2 2 2 2 2 2 2 2 2 2 2 ... 1 0 1 1 0 1 1 0 1 1 #
    CORE
384 1848 2 2 2 2 2 2 2 1 1 1 1 1 1 1 ... 0 0 0 0 0 0 0 0 0 #
    CORE
368 2192 6 6 6 6 6 6 6 6 0 0 0 0 0 0 0 ... 0 0 0 0 0 0 0 0 0 #

```

Figure 5.9: Sample Overlay File

### 5.4.2 Selected Data Set

In this thesis 60 images are used. The Images are submitted to database from mammography clinic, the Ann L. Baraco Centre for Women's Health, located at Sacred Heart Hospital in Pensacola, Florida. The images are divided to two sets of 40 and 20 images for training and testing. The 20 images used for test are listed in Table 5.4.

As described in Table 5.4, the 20 images with malignant microcalcifications were divided into four sets according to density and subtlety ratings: (1) high-density breast tissue with obvious abnormalities, (2) high density breast tissue with subtle abnormalities, (3) low-density breast tissue with obvious abnormalities, (4) low-density breast tissue with subtle abnormalities. In Table 5.4, column 1 indicates the case category. Column 2 indicates the case number and view. The breast density rating is shown in column 3. This rating is according to Breast Imaging Reporting and Data systems (BI-RADS) density ratings. A density rating in scale 1 to 4 is used as follows [74]: 1= The breast is almost entirely fat. 2 = There are scattered fibroglandular densities that could obscure a lesion on a mammogram. 3 = The breast is heterogeneously dense. This may lower the sensitivity of mammogram. 4= the breast is extremely dense which lowers the sensitivity of mammogram.

According to this metric a higher density rating implies that it is more difficult for radiologist ( and computer) to detect the abnormality. Column 4 indicates the BI-RADS assessment rating ranging from 1 to 5. 1 = Negative. 2 = Benign finding. 3 = Probably benign finding. 4 = Suspicious abnormality (biopsy should be considered). 5 = Highly suggestive of malignancy (appropriate action should be taken).

In column 5, subtlety rating is presented. The value of this measure range from 1-5, where 1 stands for "subtle" and 5 for "obvious". A higher subtlety rating indicates easier interpretation task for both radiologist and computer. Columns 5 and 6 are showing pathology and lesion type as well as number of abnormalities.

**Table 5.4:** Summary of ground truth information for the test images

	Case/view	Density	Assessment	Subtlety	Pathology	Lesion type/description
High density/ subtle	c0060LCC	3	4	2	Malignant	1 calcification/amorphous distribution: clustered
	c00060LMLO	3	4	2	Malignant	1 calcification/amorphous distribution: clustered
	c150LCC	3	4	1	Malignant	2 calcification/amorphous distribution: clustered
	c0150LMLO	3	4	1	Malignant	2 calcification/amorphous distribution: clustered
	c0198RCC	3	4	2	Malignant	1 calcification/amorphous distribution: segmental
Low density / subtle	c0198RMLO	3	4	2	Malignant	1 calcification/amorphous distribution: segmental
	c0002LCC	2	4	2	Malignant	1 calcification/pleomorphic distribution: segmental
	c0002LMLO	2	4	1	Malignant	1 calcification/pleomorphic distribution: segmental
	c0020LMLO	2	4	2	Malignant	1 calcification/pleomorphic distribution: regional
	c0169RCC	2	4	2	Malignant	1 calcification/amorphous distribution: clustered
High density / obvious	c0169RMLO	2	4	2	Malignant	1 calcification/amorphous distribution: clustered
	c0036RCC	4	4	4	Malignant	1 Calcification/punctate distribution: clustered
	c0036RMLO	4	4	4	Malignant	1 Calcification/punctate distribution: clustered
	c0120LCC	4	5	5	Malignant	1 Calcification/amorphus-pleomorphic distribution: segmental
	c0120LMLO	4	5	5	Malignant	1 Calcification/amorphus-pleomorphic distribution: segmental
Low density / obvious	c0045RMLO	2	4	5	Malignant	1 Calcification/pleomorphic distribution: clustered
	c0087LCC	2	5	5	Malignant	3 Calcifications/pleomorphic distribution: clustered
	c0087LMLO	3	5	5	Malignant	3 Calcifications/pleomorphic distribution: clustered
	c0214RCC	2	4	4	Malignant	1 Calcification/pleomorphic distribution: segmental
	c0214RMLO	2	4	4	Malignant	1 Calcification/pleomorphic distribution: segmental



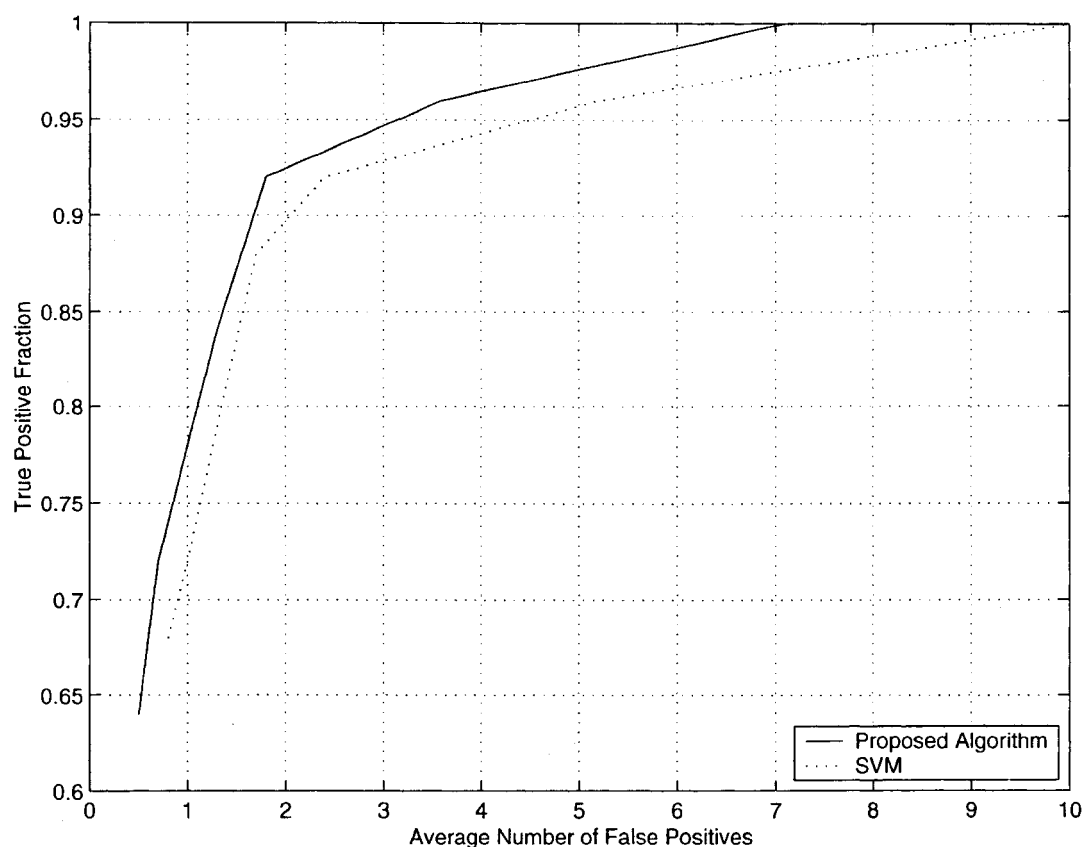
### 5.4.3 Results

The training data set were used to train the proposed algorithm. 40 images including 51 microcalcification were used in training phase. The preprocess algorithm on test data resulted in 5236 suspicious sites. The improved SVM then were applied to the test data. To detect the microcalcifications, the number of pixels which are classified as "Microcalcification Present", are counted in each  $1\text{cm}^2$  of the mammogram. A threshold is set to determine whether a site contains a microcalcification or not. If the number of pixels which are classified as "microcalcification present" is greater than the threshold, the site is considered as a microcalcification.

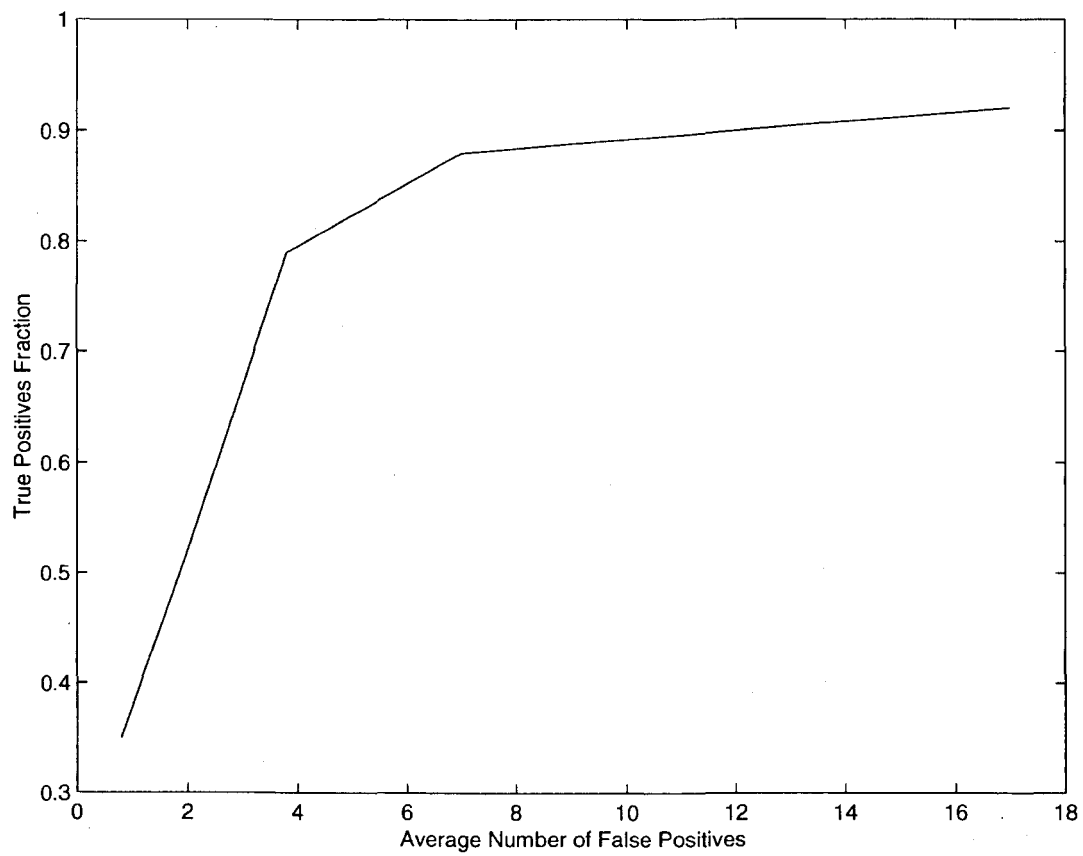
The ROC curves in Fig. 5.10 were generated by varying the threshold from 3 to 14. The proposed algorithm outperforms the previously applied SVM to the problem of microcalcification [65]. The El-Naqa *et al.* algorithm were applied to current data set to obtain the ROC curve. The area under the two curves are calculated as 7.22 for LPC-SVM and 5.68 for SVM. This indicates a 26% improvement. The numerical results which were obtained here for the SVM [65] is different from original report. The reason is that the data set is different here. It is obvious that the data set can significantly affect the calculated performance. i.e. if some difficult subtle images were excluded from the study, the performance was totally different. While the overall detection performance is very good, a more detailed analysis of the individual cases reveals additional insight into the performance. The ROC for the most recent published research on the same database is given in Fig.5.11. The are under curve is 4.86 which 46% less than proposes LPC-SVM method.

The detail results of the proposed algorithm are given in Table 5.5. This table is presenting the case of 1.5 false positives at 92% of correct true positive detection. As described before, the test data has been selected for 4 classes mammograms: 1) low-density tissues with obvious calcifications, 2) low-density tissue with obvious calcifications, 3) high density tissues with obvious calcifications, 4) high-density tissue with subtle calcifications. The re-

sponse of the algorithm to these classes, are different, as can be seen in Table 5.5. The algorithm is quite successful in detection of microcalcification in low/high density-obvious and low density subtle and more research is required for high-density subtle. The application of the proposed algorithm on the high density images are resulted in relatively high number of false positives.



**Figure 5.10:** ROC curves show that the proposed algorithm outperforms SVM



**Figure 5.11:** ROC curve for the most recent published research on DDSM

**Table 5.5:** The number of false positives and correct true positives in each test image

	Case	True Positive Rate	False Positives	Subtlety
High density subtle	c0060LCC	1	2	2
	c0060LMLO	1	4	2
	c150LCC	1	3	1
	c0150LMLO	0.5	1	1
	c0198RCC	0.5	1	2
	c0198RMLO	1	2	2
Low density subtle	c0002LCC	1	1	2
	c0002LMLO	1	0	1
	c0020LMLO	1	1	2
	c0169RCC	1	1	2
	c0169RMLO	1	1	2
High density obvious	c0036RCC	1	6	4
	c0036RMLO	1	2	4
	c0120LCC	1	1	5
	c0120LMLO	1	1	5
Low density obvious	c0045RMLO	1	1	5
	c0087LCC	1	0	5
	c0087LMLO	1	1	5
	c0214RCC	1	1	4
	c0214RMLO	1	0	4

# Chapter 6

## Conclusion and future work

### 6.1 SVM in texture classification and external texture features

The SVM learning algorithm has been proven to outperform other texture classification algorithms which are based on supervised learning i.e. neural network.

In this thesis, the effectiveness of SVM in texture classification problem was investigated. It was shown that SVM is a powerful algorithm not only for dichotomy ( two-class) problems but also for multi-texture.

The effect of external features in performance of SVM were studied. The estimation of the VC-dimension of the external features were used to chose the best feature set among suggested features. The comparison of the VC-dimension resulted in LPC-SVM texture classification algorithm. LPC which provides a high dimensional feature space were chosen considering the excellent ability of SVM to set a learning hyperplane in a high dimension input features. In the study of VC-dimension, LPC provides the minimum VC-dimension and training error.

The discrimination ability of LPC features is reduced in high resolution classification

when the window size of the sampled textures and eventually the input feature dimension is reduced.

## **6.2 Incorporating the Gabor filter Bank to Increase the performance of LPC-SVM texture classification algorithm**

When we need to detect tiny feature in an image or to increase the resolution of a texture classification-based segmentation, we come across the problem of low dimension input features with poor texture information.

In the low dimension input texture feature space the VC-dimension and training errors of LPC is reduced. This is due to the poor discriminant features in a small texture window.

The Gabor filter bank is extensively and successfully used in several texture classification algorithms. To enrich the texture discrimination ability of input features in case of small window size, the Gabor filter bank were employed.

In the proposed algorithm, for each sample window in the original image, corresponding sample windows from filtered version of the image are extracted and concatenated to form an input feature prior to LPC feature selection. In this way the classification accuracy was improved slightly but the time consumption of the algorithm increased considerably.

## **6.3 Microcalcification detection**

SVM algorithm is an ideal learning machine for a supervised learning approach in a two-class classification (dichotomy) problem. The dichotomy problem is concerned in a wide range of

medical imaging applications. Detection of abnormalities in medical imaging is a dichotomy problem in the sense of classifying an image to normal and abnormal areas.

The high rate of breast cancer development in women over 40 years of age, and the mortality rate was the main motivation for applying the proposed algorithm to microcalcification detection in screening mammography. Microcalcification is the most difficult type of breast cancer to be detected by radiologist.

The major problem in application of LPC-SVM texture classification algorithm to microcalcification detection is the asymmetry of the training data. The huge number of pixels which are presenting normal tissue in compare with a very few number of pixels presenting the microcalcifications. A preprocess stage was added to overall algorithm to detect suspicious areas and reduce the number of the pixels presenting normal areas. The preprocess were applied prior to the proposed algorithm to increase the accuracy and time efficiency.

The test data for evaluating the proposed algorithm were evenly divided into four category of mammograms. The images were divided categorized according to tissue density and the difficulty of detection of microcalcification. The results were evaluated through ROC curve and were compared with SVM algorithm.

A correct rate of 92% were archived at 1.5 false positives per image. Considering the number of very difficult images in the data set used during the test, the overall performance is significant.

## 6.4 Research contribution

The main contribution of this work is in the field of medical imaging. The SVM algorithm were empowered by external features for texture classification. A number of external features were suggested and examined and through the study of VC-dimension the best were chosen. The performance of LPC features in high resolution classification, were improved by employing gabor filter bank features. The excellent ability of SVM in dichotomy problems were used in detection of abnormalities. The asymmetry problem of training data were solved with pre-process stage based on wavelet transform decomposition and reconstruction of the mammograms. The performance of the algorithm were discussed for the images clinically categorized into different levels of difficulty for assessment.

## 6.5 Future work

More research is needed to increase the accuracy of the algorithm in subtle cases of cancer and to reduce the number of false positives in high density images.

Incorporating information from other modalities such as ultra sound and MRI imaging must be considered to achieve a CAD system with acceptable performance for clinical use.



## Acronyms

<b>CAD:</b>	Computer Aided Diagnosis
<b>LDB:</b>	Local Discriminant Basis
<b>LPC:</b>	Linear Predictive Coding
<b>HVS:</b>	Human Visual System
<b>ROC:</b>	Receiver Operating Characteristic
<b>SVM:</b>	Support Vector Machine
<b>VC:</b>	Vapnic-Chervonenkis

# Bibliography

- [1] C. C. Society, "Cancer statistics," in <http://www.cancer.ca>, 2002.
- [2] N. Karssemeijer, "Computer assisted reading mammograms," *Journal of European Radiology*, vol. 7, pp. 743–748, 1997.
- [3] K. I. Kim, K. Jung, S. H. Park, and H. J. Kim, "Support vector machine for texture classification," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1542–1550, 2002.
- [4] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
- [5] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 1–47, 1998.
- [6] B. Scholkopf, K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, "Comparing support vector machines with gaussian kernels to radial basis function classifiers," *IEEE Transaction on Signal Processing*, vol. 45, pp. 2785–2765, 1997.
- [7] T. Cover, "Geometrical and statistical properties of systems of linear inequalities with application in pattern recognition," *IEEE Transaction on Electronic Computers*, vol. 14, pp. 326–334, 1965.

- [8] N. Saito and R. R. Coifman, "Local discriminant bases and their applications," *J. Mathematical Imaging and Vision*, vol. 4, pp. 337–358, 1995.
- [9] R.R.coifman and M. Wickerhauser, "Entropy-based algorithm for best basis selection," *IEEE Trans. Inform. Theory*, vol. 38, pp. 713–719, 1992.
- [10] S. G. Mallat, *A Wavelet Tour of Signal Processing*. San Diego: Academic Press, 1998.
- [11] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Pattern Anal. and Machine Intell*, vol. 11, pp. 674–693, 1989.
- [12] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*.
- [13] A. Gresho, "Advances in speech and audio compression," in *Proc. IEEE*, ser. pp. 900–918, 1994.
- [14] J. D. Gibson, "Adaptive prediction for speech differential encoding system," in *Proc. IEEE*, ser. pp. 1789–1797, vol. 68, 1974.
- [15] P. Kroon and E. F. Deprettere, "A class of analysis-by-synthesis predictive coders for high quality speech coding at rates between 4.8 and 16 kbits/s," *IEEE J. Select. Areas Commun*, vol. 6, pp. 353–363, 1988.
- [16] B. S. Atal, "Predictive coding of speech at low bit rates," *IEEE Trans. Commun.*, vol. COMM-30, pp. 600–614, 1982.
- [17] M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (celp): High quality speech at very low bit rates," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, ser. pp. 937–940, 1985.

- [18] J. H. Hu, Y. W., and P. T. Cahill, "Multispectral code excited linear prediction coding and its application in magnetic resonance images," *IEEE Trans. on Image Processing*, vol. 6, pp. 1555–1566, 1997.
- [19] F. Hlawatsch and G. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Processing Magazine*, vol. April, pp. 21–67, 2001.
- [20] M. Davy, C. Doncarli, and G. F. Boudreaux-Bartels, "Improved optimization of time-frequency based signal classifier," *IEEE Signal Processing Letters*, vol. 8, pp. 52–57, 2001.
- [21] J. Marti, J. Batlle, and A. Casals, "Model-based objects recognition in industrial environments for autonomous vehicles control," in *International Conference on Robotics and Automation*, ser. Proc. IEEE, 1997.
- [22] A. Kumar and G. Pang, "Defect detection in textured materials using gabor filters," in *Industry Applications Conference*, ser. Proc. IEEE, 2000.
- [23] J. Li, J. Tan, and P. Shatadal, "Classification of tough and tender beef by image texture analysis," *Meat Science*, vol. 45, pp. 341–346, 2001.
- [24] M. H. Horng, Y. N. Sun, and X. Z. Lin, "Texture feature coding method for classification of liver sonography," *Computerized Medical Imaging and Graphics*, vol. 26, pp. 33–42, 2002.
- [25] J. S. Bleck, U. Ranft, M. Gebel, H. Hecker, M. Westoff-Bleck, C. Thiesemann, S. Wagner, and M. Manns, "Random field models in the textural analysis of ultrasonic images of the liver," *IEEE Transaction on Medical Imaging*, vol. 15, pp. 796–801, 1996.

- [26] J. Wang, S. M. Krishnan, C. Kugean, and M. P. Tjoa, "Classification of endoscopic images based on texture and neural network," in *IEEE 23rd EBMS International Conference*, ser. Proc. IEEE, 2001.
- [27] M. N. Gurcan, Y. Yardmici, A. E. Cetin, and R. Ansari, "Detection of microcalcification in mammograms using higher order statistics," *IEEE Signal Processing Letters*, vol. 4, pp. 213–216, 1997.
- [28] J. K. Kim and H. W. Park, "Statistical texture features for detection of microcalcifications in digitized mammograms," *IEEE Transaction on Medical Imaging*, vol. 18, pp. 231–238, 1999.
- [29] B. V. Ginneken, S. Katsuragawa, B. M. T. H. Romeny, K. Doi, and M. A. Viergever, "Automatic detection of abnormalities in chest radiographs using local texture analysis," *IEEE Transaction on Medical Imaging*, vol. 21, pp. 139–149, 2002.
- [30] P. Diaconis and D. Freedman, "On the statistics of vision: The Julesz conjecture," *Journal of Math. and Psychology*, vol. 24, 1981.
- [31] S. Z. Li, *Markov Random Field Modeling in Computer Vision*. New York: Springer-Verlag, 1995.
- [32] S. C. Zhu, Y. Wu, and D. Mumford, "Filters, random field and maximum entropy(frame)- towards a unified theory for texture modeling," *International Journal of Computer Vision*, vol. 27, 1998.
- [33] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using gabor filters," *Pattern Recognition*, vol. 24, pp. 1167–1186, 1991.

- [34] C. Lu, P. Chung, and C. Chen, "Unsupervised texture segmentation via wavelet transform," *Pattern Recognition*, vol. 30, pp. 729–742, 1997.
- [35] A. K. Jain and K. Karu, "Learning texture discrimination masks," *IEEE Transaction Pattern Analysis and Machine Intelligence*, vol. 18, pp. 195–205, 1996.
- [36] J. W. Wang, C. Chen, W. M. Chien, and C. M. Tsai, "Texture classification using non-separable two-dimensional wavelets," *Pattern Recognition Letters*, vol. 19, pp. 1225–1234, 1998.
- [37] S. Arivazhagan and L. Ganesan, "Texture classification using wavelet transform," *Pattern Recognition Letters*, vol. 24, pp. 1513–1521, 2003.
- [38] V. Manian, R. Vasquez, and P. Katiyar, "Texture classification using logical operators," *IEEE Transaction on Image Processing*, vol. 9, pp. 1693–1703, 2000.
- [39] D. A. Clausi and M. E. Jerningan, "Designing gabor filters for optimal texture separability," *Pattern Recognition Letters*, vol. 33, pp. 1835–1849, 2000.
- [40] P. P. Raghu and B. Yegnanarayana, "Supervised texture classification using a probabilistic neural network and constraint satisfaction model," *IEEE Transaction on Neural Networks*, vol. 9, pp. 516–522, 1998.
- [41] L. M. Kaplan, "Extended fractal analysis for texture classification and segmentation," *IEEE Transaction on Image Processing*.
- [42] N. T. M. N. Shirazi, H. Noda, "Texture classification based on markov modeling in wavelet feature space," *Image and Vision Computing*, vol. 18, pp. 967–973, 2003.
- [43] N. Saito and R. Coifman, "Local discriminant bases and their applications," *Journal of Mathematical Imaging and Vision*, vol. 5, pp. 337–358, 1996.

- [44] C. Heitz, "Optimum time-frequency representations for the classification and detection of signals," *J. of Applied Signal Processing*, vol. 3, pp. 124–143, 1995.
- [45] B. Gillespie and L. Atlas, "Optimizing tf kernels for classification," *IEEE Transaction on Signal Processing*, vol. 48, pp. 485–496, 2001.
- [46] L. Rabiner and B. H. Juang, *Fundamentals of speech recognition*. Englewood Cliffs, N.J.: Prentice Hall, 1993.
- [47] C. B. B. Scholkopf and V. Vapnik, "Extracting support data for a given task," in *Proc. Int'l Conf. Knowledge Discovery Data Mining*, ser. pp. 252–257, 1995.
- [48] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*, New York, 1966.
- [49] M. Vision and M. Group, 1998.
- [50] R. M. Haralick, K. Shunmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transaction on Syst. Man. Cybern.*, vol. SMC-18, pp. 610–621, 1973.
- [51] R. W. Connors, "A theoretical comparison of texture algorithms," *IEEE Transaction on Pattern Anal. Machine Intel.*, vol. PAMI-2, pp. 204–222, 1980.
- [52] T. Chang and C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE Transaction on Image Processing*, vol. 2, pp. 429–441, 1993.
- [53] J. You and H. A. Cohen, "Classification and segmentation of rotated and scaled texture images using texture tuned masks," *Pattern Recognition*, vol. 36, pp. 245–258.
- [54] J. G. Daugman, "Complete discrete 2-d gabor transforms by neural networks for image analysis and compression," *IEEE Transaction On Speech, Signal Processing*, vol. 36, pp. 1169–1179.

- [55] D. Hubel and T. Wiesel, "Receptive fields and functional architecture in two nonstriate visual areas 18 and 19 of the cat," *J. Neurophysiol.*, vol. 28, pp. 229–289, 1965.
- [56] F. Campbell and J. Kulikowski, "Orientational selectivity of the human visual system," *J. Physiol.*, vol. 187, pp. 437–445, 1966.
- [57] D. Pollen and S. Ronner, "Visual cortical neurons as localized spatial frequency filters," *IEEE Trans. Syst. Man. Cybern.*, vol. 5, pp. 907–916, 1983.
- [58] S. Bruner and B. Langfeldt, *Recent Results in Cancer Research*. Berlin, Germany: Springer-Verlag, 1990.
- [59] L. Shen, R. M. Rangayyan, and J. E. L. Desautels, "Application of shape analysis to mammographic calcifications," *IEEE Trans. Medical Imaging*, vol. 13, pp. 263–274, 1994.
- [60] N. Karssemeijer, "Adaptive noise equalization and recognition of microcalcification clusters in mammograms," In : *Bowyer, K. W. and Astely, S. (eds), State of the Art in Digital Mammographic Image Analysis*, vol. Singapore: World Scientific, pp. 148–166, 1994.
- [61] R. M. Nishikawa, M. L. giger, K. Doi, C. J. Vyborny, and R. A. Schmidt, "Computer aaided detection of clustered microcalcification in digital mammograms," *Med. Biol. EnComput.*, vol. 33, pp. 174–178, 1995.
- [62] R. N. Strickland and H. I. Hahn, "Wavelet transform for detecting microcalcifications in mammograms," *IEEE Trans. In Med. Imag.*, vol. 15, pp. 218–229, 1996.
- [63] T. Netsch and H. Peitgen, "Scale-space signature for the detection of clustered microcal-



- cifications in digital mammograms," *IEEE Trans. In Med. Imag.*, vol. 18, pp. 774–786, 1999.
- [64] S. Yu and L. Guan, "A cad system for automatic detection of clustered microcalcifications in digitized mammogram films," *IEEE Trans On Med. Imag.*, vol. 19, pp. 115–126, 2000.
- [65] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. M. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Trans On Med. Imag.*, vol. 21, pp. 1552–1563, 2002.
- [66] S. Sentelle, C. Sentelle, and M. Sutton, "Multiresolution-based segmentation of calcification for the early detection of breast cancer," *Journal of Real-Time Imaging*, vol. 8, pp. 237–252, 2002.
- [67] J. C. Bezdek, J. Keller, R. Krisnapuram, and N. R. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Norwell, MA: Kluwer Academic Publisher, 1999.
- [68] T. C. Wang and N. B. Karayiannis, "Detection of microcalcifications in digital mammograms using wavelets," *IEEE Trans. on Medical Imaging*, vol. 17, pp. 498–509, 1998.
- [69] S. Lai, X. Li, and W. Bischof, "On techniques for detecting circumscribed masses in mammograms," *IEEE Trans. on Medical Imaging*, vol. 8, pp. 377–386, 1989.
- [70] P. F. Stetson, F. Sommer, and A. Macovski, "Lesion contrast enhancement in medical ultrasound imaging," *IEEE Trans. on Medical Imaging*, vol. 16, pp. 416–425, 1997.
- [71] I. Daubechies, *Ten Lecture on Waveletes*. Philadelphia, PA: SIAM.

- [72] K. R. Muller, S. Mika, G. Ratch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. on Neural Networks*, vol. 12, pp. 181–201, 2001.
- [73] H. Heath, K. Bowyer, D. Kopans, R. Moore, and P. Kegelmeyer, "The digital database for screening mammography," in *Yaffe, M.J.(ed.), IWDM 2000 5th International Workshop on Digital Mammography.*, ser. Madison, WI: Medical Physics Publishing, 2001.
- [74] A. C. of Radiology, *Breast Imaging Reporting and Data Systems(BIRADS)(2nd edn.)*.

## VITA

NAME: Mahdi Sabri

PLACE OF BIRTH: Tehran, Iran

YEAR OF BIRTH: 1972

POST-SECONDARY EDUCATION AND DEGREES: Sharif University of Technology  
Tehran, Iran  
1990-1995, BSc

HONORS AND AWARDS: Co-winner of President Award for Industrial Innovation  
1998

RELATED WORK EXPERIENCE: 1995-1997  
Software developer/designer, Maharan Corp., Tehran, Iran  
1997-2000  
Senior software designer, Maharan Corp., Tehran, Iran  
2000-2002  
Project Manager, Maharan Corp., Tehran, Iran

## PUBLICATIONS

1. M. Sabri, J. Alirezaie, "Support Vector Machine with Optimized Kernel for texture classification" submitted to Journal of Electronic Imaging (SPIE)
2. M. Sabri, J. Alirezaie, "Support Vector Machine with Optimized Time Frequency Kernel for texture classification" Accepted in Electronic Imaging 2003(SPIE)
3. M. Sabri, K. Hazaveh, J. Alirezaie, K. Raahemifar, Two Dimensional Local Discrimination Basis Algorithm for Texture Classification, in proceeding of IEEE Canadian conference in Electrical and computer Engineering 2003.
4. M. Sabri, J. Alirezaie Texture Classification Using Improved Local Discrimination Basis Algorithm in proceeding of IEEE SSP (Statistical Signal Processing) 2003.
5. M. Sabri, J. Alirezaie, S. Krishnan, Audio Noise Detection using Hidden Markov Model, to be published in proceeding of IEEE SSP2003.