

University of New Hampshire

## University of New Hampshire Scholars' Repository

---

Master's Theses and Capstones

Student Scholarship

---

Spring 2010

### Bio-inspired log-polar based color image pattern analysis in multiple frequency channels

Xiaolu Li

*University of New Hampshire, Durham*

Follow this and additional works at: <https://scholars.unh.edu/thesis>

---

#### Recommended Citation

Li, Xiaolu, "Bio-inspired log-polar based color image pattern analysis in multiple frequency channels" (2010). *Master's Theses and Capstones*. 553.

<https://scholars.unh.edu/thesis/553>

This Thesis is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Master's Theses and Capstones by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact [Scholarly.Communication@unh.edu](mailto:Scholarly.Communication@unh.edu).

## **NOTE TO USERS**

**This reproduction is the best copy available.**

**UMI<sup>®</sup>**



**BIO-INSPIRED LOG-POLAR BASED COLOR IMAGE  
PATTERN ANALYSIS IN MULTIPLE FREQUENCY CHANNELS**

**BY**

**Xiaolu Li**

**BS in Communication Engineering,  
Beijing University of Posts and Telecommunications, China, 2008**

**THESIS**

**Submitted to the University of New Hampshire  
in Partial Fulfillment of  
the Requirements for the Degree of**

**Master of Science  
in  
Electrical Engineering**

**May, 2010**

UMI Number: 1485435

All rights reserved

**INFORMATION TO ALL USERS**

The quality of this reproduction is dependent upon the quality of the copy submitted.

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



UMI 1485435

Copyright 2010 by ProQuest LLC.

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



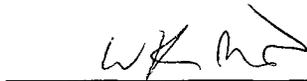
ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

This thesis has been examined and approved.



---

Thesis Director, Richard A. Messner  
Associate Professor, Electrical and Computer Engineering



---

W. Thomas Miller, III  
Professor, Electrical and Computer Engineering



---

L. Gordon Kraft, III  
Professor, Electrical and Computer Engineering

05/06/10

---

Date

## ACKNOWLEDGEMENTS

I would like to express my gratitude to my research adviser Dr. Richard Messner whose stimulating suggestions helped me in the research for this thesis. It was he who always gave me great encouragements that made me more confident in overcoming difficulties and making improvements. The research experience in the Synthetic Vision and Pattern Analysis Lab is a precious wealth for all my life.

Many thanks to the Department of Electrical Engineering for giving me two years of Teaching Assistantship position which made my master study here possible. I would like to thank my thesis committee members, Dr. Thomas Miller and Dr. Gordon Kraft for their careful review of my thesis document. I also would like to acknowledge Dr. Kent Chamberlin for giving me his full support during my graduate journey at UNH.

I want to thank Beijing University of Posts and Telecommunications (BUPT) for giving me excellent undergraduate education. Thank you to my good friends from both BUPT and UNH CSSA, for all their continuous help, support and valuable suggestions both on academics and my everyday life.

Special thanks to my parents and my boyfriend Jianwen, whose patient love enabled me to complete this work.

# TABLE OF CONTENTS

|  |     |
|--|-----|
| ACKNOWLEDGEMENTS .....   | iii |
| TABLE OF CONTENTS .....  | iv  |
| LIST OF TABLES .....   | vi  |
| LIST OF FIGURES .....  | vii |
| ABSTRACT .....   | ix  |
| INTRODUCTION.....  | 1   |
| CHAPTER I.....   | 8   |
| BACKGROUND .....   | 8   |
| 1.1 Introduction of Log-Polar Mapping.....   | 8   |
| 1.2 The Lateral Subtraction Inhibition (LSI) Network Model .....   | 13  |
| CHAPTER II.....  | 18  |
| THE LSI BASED GRAYSCALE IMAGE FEATURE EXTRACTION .....   | 18  |
| 2.1 Pre-Whitening Filter Adopting LSI Recurrent Network Model.....   | 18  |
| 2.1.1 Solutions of LSI model applied in log-polar domain .....   | 19  |
| 2.1.2 Decorrelation in the adaptive recurrent networks.....  | 22  |
| 2.2 The Introduction to ICA for Feature Extraction and Encoding .....  | 24  |
| CHAPTER III.....   | 28  |
| THE BIO-INSPIRED METHODOLOGY OF COLOR IMAGE PATTERN<br>RECOGNITION.....  | 28  |
| 3.1 A Methodology for Color Image Pattern Recognition.....   | 28  |
| 3.2 Fundamentals of Color Image Processing.....  | 32  |
| 3.3 Object Centroid Calculation and the Log-Polar Mapping.....   | 35  |
| 3.4 Experimental Results of Color Image Object Recognition .....   | 40  |
| CHAPTER IV .....   | 44  |
| IMAGE PREPROCESSING IN MULTIPLE SPACIAL FREQUENCY CHANNELS<br>.....  | 44  |
| 4.1 Multiple Spatial Frequency Filters Applied on Grayscale Images .....   | 44  |
| 4.2 Experimental Results of Multi-Frequency Image Filtering .....  | 54  |
| 4.3 The Feature Extraction Exploration of Different Frequency Channels.....  | 63  |
| CHAPTER V.....   | 66  |
| CONCLUSIONS AND FUTURE DIRECTIONS.....   | 66  |
| LIST OF REFERENCES .....   | 71  |
| APPENDIX A.....  | 77  |
| REAL TIME IMAGE PATTERN RECOGNITION IMPLEMENTATIONS IN AN<br>OPTICAL SYSTEM.....   | 77  |
| A.1 The Significance of the Real Time Implementations of Distortion Invariant<br>Image Pattern Recognition in Optical System ..... | 77  |
| A.2 Various Models of Position, Scale and Rotation Invariant Optical<br>Correlation.....   | 78  |
| A.2.1 A brief review of different implementation models .....  | 78  |
| A.2.2 Initial models of rotation and scale invariant optical correlations .....  | 79  |

|   |    |
|---|----|
| A.2.3 Distortion invariant optical system models using coordinate transformations .....                     | 82 |
| A.3 Some Advanced Technologies and Algorithms.....  | 86 |
| A.3.1 Advanced computational algorithms for distortion invariant pattern recognition .....                  | 86 |
| A.3.2 The synthetic discriminant functions (SDF) .....  | 87 |
| A.3.3 Advances of spatial filters .....   | 89 |
| APPENDIX B.....   | 95 |
| SOME DERIVATIONS OF INTRODUCING MULTIPLE LINEAR CHANNELS IMPROVES THE PATTERN RECOGNITION PERFORMANCE ..... | 95 |
| APPENDIX C .....  | 97 |
| THE ORGANIZATION OF THE CODES IN CD .....   | 97 |
| APPENDIX D .....  | 99 |
| THE OPENCV 2.0 ENVIRONMENT SETUP ON MICROSOFT VISUAL STUDIO 2008 .....                                      | 99 |

## LIST OF TABLES

|   |    |
|---|----|
| Table 2. 1 The convolution kernel for coupling model $b_{ijkl} = e^{-1.55d}$ .....  | 20 |
| Table 3.1 The results of non-rotated color image object recognition in RGB channels.....  | 41 |
| Table 3.2 The results of rotated color image object recognition in RGB channels .....   | 41 |
| Table 3.3 The recognition rate comparison of the original blue channel (non-rotated) and the stretched one .....                              | 43 |
| Table 4.1 The results of non-rotated grayscale image object recognition in multi-frequency bands channels by four different classifiers ..... | 64 |
| Table 4.2 The results of rotated grayscale image object recognition in multi-frequency bands channels by hybrid classifiers .....             | 64 |

## LIST OF FIGURES

|  |    |
|--|----|
| Fig. 1 The proposed color image pattern recognition methodology .....  | 4  |
| Fig. 2 The proposed grayscale image pattern recognition methodology .....  | 5  |
| Fig. 3 The proposed cascaded color image pattern recognition methodology .....   | 6  |
|  |    |
| Fig. 1.1 A foveated image (Fovead on the pistil of the left flower).....   | 9  |
| Fig. 1.2 Circular receptive fields with different radii are distributed along<br>concentric rings. ....  | 10 |
| Fig. 1.3 Illustration of the Cartesian plane and the log-polar plane. ....   | 11 |
| Fig. 1.4 The log-polar mapping applied to regular patterns.....  | 12 |
| Fig. 1.5 Mach Bands - The perceived brightness is enhanced at the edges .....  | 13 |
| Fig. 1.6 How photoreceptors and receptive fields are arranged [14].....  | 14 |
| Fig. 1.7 Lateral Subtractive Inhibition using the feed-backward coupling model.  | 15 |
|  |    |
| Fig. 2.1 The result of the network, the iterative solutions for the 2-D grid with<br>lateral inhibition; original image (left), resulting image (middle); Resulting signal<br>profiles: network versus iterative solution (right); NRMSE = 0. 0038, Image size<br>used is 32 x 32 [6]..... | 21 |
| Fig. 2.2 The recurrent network with adaptive connections .....   | 23 |
| Fig. 2.3 An image in log-polar space (left) with outlined pseudo-salient locations<br>(right) [6].....   | 26 |
|  |    |
| Fig. 3.1 The methodology of color image feature extraction and pattern<br>recognition .....  | 31 |
| Fig. 3.2 The 24-bit RGB cube.....  | 33 |
| Fig. 3.3 An RGB image example and original R, G, B channels .....  | 34 |
| Fig. 3.4 An RGB image example and the intensities of R, G, B channels .....  | 34 |
| Fig. 3.5 (a) The monochrome object sitting in black background .....   | 36 |
| Fig. 3.6 The example of chromatic images and their centered objects on RGB<br>channels.....  | 38 |
| Fig. 3.7 Left to right are the R, G, B channels of the log-polar mapped image of<br>Fig. 3.6 (e).....  | 39 |
| Fig. 3.8 Blue component (left) and the same blue component stretched to [0 255]<br>(right).....  | 43 |
|  |    |
| Fig. 4.1 The single channel model of grating detection .....   | 46 |
| Fig. 4.2 Block diagram of a three channels filtering model.....  | 46 |
| Fig. 4.3 Part of the methodology of multi-frequency channels image pattern<br>recognition .....  | 47 |
| Fig. 4.4 The cascaded preprocessing scheme of multi-frequency channels .....   | 48 |
| Fig. 4.5 (a) Gaussian LPF1, $\sigma_1 = 0.25$ , height unity; (b) Gaussian LPF2, $\sigma_2 = 0.1$ ,<br>height 0.5; (c) Gaussian BPF, DOG filter (a)-(b); (d) Gaussian HPF 1-(b), (e)<br>Profiles the above filters; (f) Another filter designs .....                                       | 57 |

|   |     |
|---|-----|
| Fig. 4.6 Filtering process on a synthetic image of cosine functions in both x and y directions.....   | 58  |
| Fig. 4.7 The frequency spectrum of the filtered artificial images and their corresponding filters.....  | 59  |
| Fig. 4.8 (a) The original object image (centered), (b) Filtered image of Gaussian LPF, (c) Filtered image of Gaussian BPF, (d) Filtered image of Gaussian HPF .     | 60  |
| Fig. 4.9 The log-polar mapped example of Fig. 4.8 (b), (c), (d).....  | 60  |
| Fig. 4.10 (a) The spectrum of the original object, (b) Gaussian LPF filtered spectrum, (c) Gaussian BPF filtered spectrum, (d) Gaussian HPF filtered spectrum ..... | 61  |
| Fig. A.1 Schematic diagram of a general optical.....  | 81  |
| Fig. A.2 Scale and rotation invariant optical correlation system .....  | 83  |
| Fig. A.3 Real-time optical correlator system schematic.....   | 84  |
| Fig. A.4 Experimental setup of this CT implementation.....  | 86  |
| Fig. D.1 .....  | 99  |
| Fig. D.2 .....  | 99  |
| Fig. D.3 .....  | 100 |
| Fig. D.4 .....  | 100 |
| Fig. D.5 .....  | 101 |
| Fig. D.6 .....  | 101 |
| Fig. D.7 .....  | 102 |
| Fig. D.8 .....  | 102 |

## **ABSTRACT**

### **BIO-INSPIRED LOG-POLAR BASED COLOR IMAGE PATTERN ANALYSIS IN MULTIPLE FREQUENCY CHANNELS**

by

Xiaolu Li

University of New Hampshire, May, 2010

The main topic addressed in this thesis is to implement color image pattern recognition based on the lateral inhibition subtraction phenomenon combined with a complex log-polar mapping in multiple spatial frequency channels. It is shown that the individual red, green and blue channels have different recognition performances when put in the context of former work done by Dragan Vidacic. It is observed that the green channel performs better than the other two channels, with the blue channel having the poorest performance. Following the application of a contrast stretching function the object recognition performance is improved in all channels. Multiple spatial frequency filters were designed to simulate the filtering channels that occur in the human visual system. Following these preprocessing steps Dragan Vidacic's methodology is followed in order to determine the benefits that are obtained from the preprocessing steps being investigated. It is shown that performance gains are realized by using such preprocessing steps.

## INTRODUCTION

Humans can recognize a face, an apple, or identify car keys with ease. This is accomplished by taking in the image information through our visual system and then making a cognitive evaluation of extracted information from that image to form a judgment on the particular image. The question we ask is how do we construct a system to sense and interpret images? This leads us to the problem addressed by pattern recognition in general. In a paper written by Ben Mauk [1], "Humans read and comprehend with an ease that belies the immense difficulty of computer pattern recognition. It is one of the central problems in artificial intelligence. One example of computer pattern recognition application is the handwritten address-reading prototype deployed by USPS since 1997. Today, current reading success rates are above 90 percent."

Although a 90 percent reading success is commendable, it is desired to obtain 100 percent reading success. We can thus turn to nature to possibly understand how we can achieve robust performance for many tasks that cannot easily be performed by a machine. Biologically inspired image processing and pattern recognition has been an attractive scientific research topic for the past several decades. Image pattern recognition is widely used in applications which include finger print identification, face recognition, character recognition, signature recognition, and the classification of objects in scientific/research areas such as:

astronomy, engineering, statistics, medical and machine learning [2].

The computer-image analysis process is often a long sequence of elementary steps that is similar to a recipe for an exquisite restaurant dish. First, images must be acquired and often times sequentially stored. This is followed by any preprocessing that may be required prior to subsequent analysis. Finally after a feature extraction phase, image content is analyzed and interpreted or classified. In and of itself, this open loop paradigm does not, support real-time processing, even for the simplest of tasks that humans easily perform [3]. The reason for exploiting the human visual system (HVS) is that there is no better or more efficient real time pattern recognition mechanism known than the HVS. Robotic vision systems and artificial intelligence knowledge are ongoing research areas that benefit from the study of the HVS [4].

This thesis is based on, and extended from, some prior research that was performed in The Synthetic Vision and Pattern Analysis Laboratory at the University of New Hampshire. In 1984, Richard Messner designed a smart sensor system for an automatic pattern recognition system, adopting the simultaneous image processing and feature extraction capability of the HVS to enable operation in real time [4]. In 1995 Li Mu investigated algorithms which effectively simulated the lateral inhibition mechanism and explored their applications in image processing [5]. In 2009, Dragan Vidacic utilized lateral inhibition (more generally linear networks with recurrent connectivity) along with complex-log conformal mapping in machine based implementations of image information encoding, feature extraction and pattern recognition [6]. Vidacic's

dissertation integrated the biological concepts of spatially non-uniform image processing with signal processing in various stages of biological vision. These techniques were embedded in a pattern recognition framework. The methodology Vidacic proposed performs well for image feature extraction and pattern recognition in the presence of both rotation and scale in the input image plane. This thesis is consistent with all the above mentioned research. The methodology introduced by Vidacic is modified in several key ways in this research in an attempt to improve the back-end recognition capability while keeping as true as possible to human visual system characteristics.

Two new preprocessing procedures based on Vidacic's image recognition methodology are introduced in order to improve the object recognition rate.

One new method is to incorporate the use of color into the system. By extending Vidacic's work into color space we can more closely look into a model of the HVS where color plays a very important role. Since human eyes can perceive about 10,000,000 different colors, we predict that a lot of useful feature information is contained in color. We anticipate that using the color information can make the pattern recognition system more robust.

To accomplish this we need to acquire a color image and then decompose it into the individual red, green and blue channels. After decomposing the color image into RGB channels, the centroid of the object must be determined and passed through all the later processing stages that were developed by Vidacic. Following this we may compare our color based results with those grayscale results demonstrated by Vidacic. If by using a three channel color representation

the recognition performance is better than what was demonstrated by Vidacic, then this decomposing processing can be adopted as another necessary preprocessing stage in the rotation and scale invariant color image pattern recognition process. The proposed methodology of the color image pattern recognition is shown in Fig. 1.

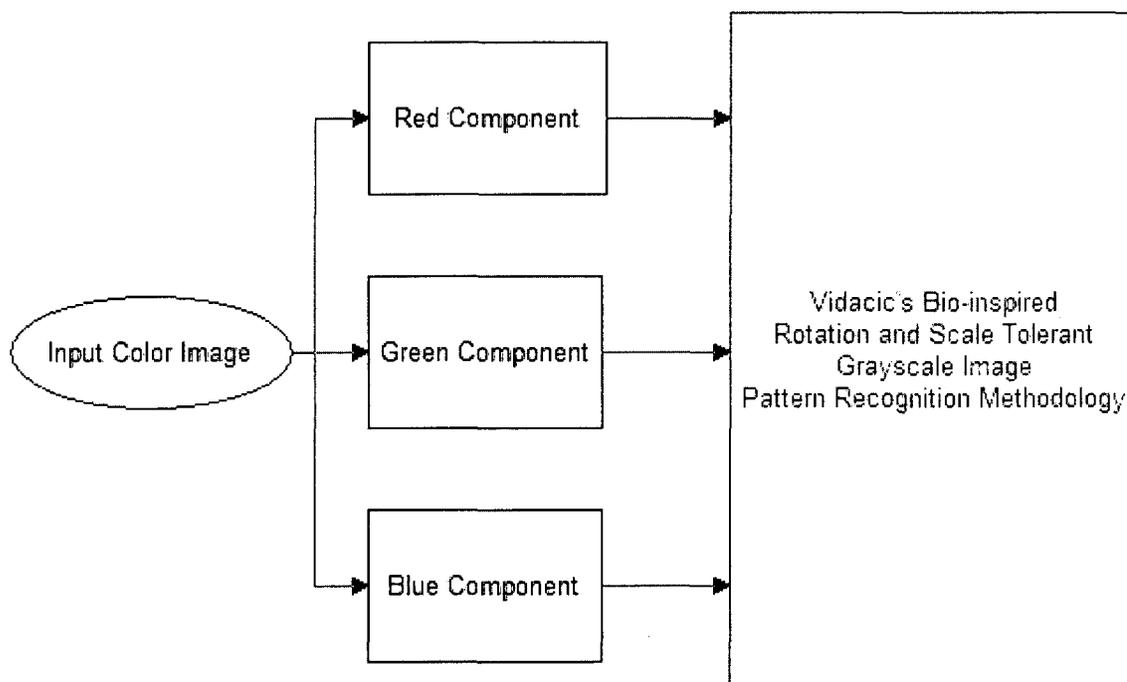


Fig. 1 The proposed color image pattern recognition methodology

It is commonly accepted that the HVS possesses several tuned spatial frequency filters in order to extract relevant feature information. To explore such a possibility in the context of Vidacic's work we consider processing the image by filtering with three spatial filters (i.e., LPF, BPF and HPF) prior to any subsequent processing stages. The reason we investigate further the image processing after filtering in multiple spatial frequency bands is that from a biological point of view the human visual system contains several different classes of spatial channels.

The design of the three filters is based on the frequency range that can be distinguished by the HVS. This is another preprocessing step that is proposed for the purpose of improving the object recognition rate in the image pattern recognition system developed by Vidacic. It included the low pass, band pass, high pass filters design, which are consistent with the multiple frequency bands filtering processes that exist in mammal's visual system. The proposed methodology of the multiple spatial frequencies filtering approach is shown in Fig. 2. Through the performance of this research it will be determined if any recognition improvement can be gained by adding these two methods to those developed by Vidacic.

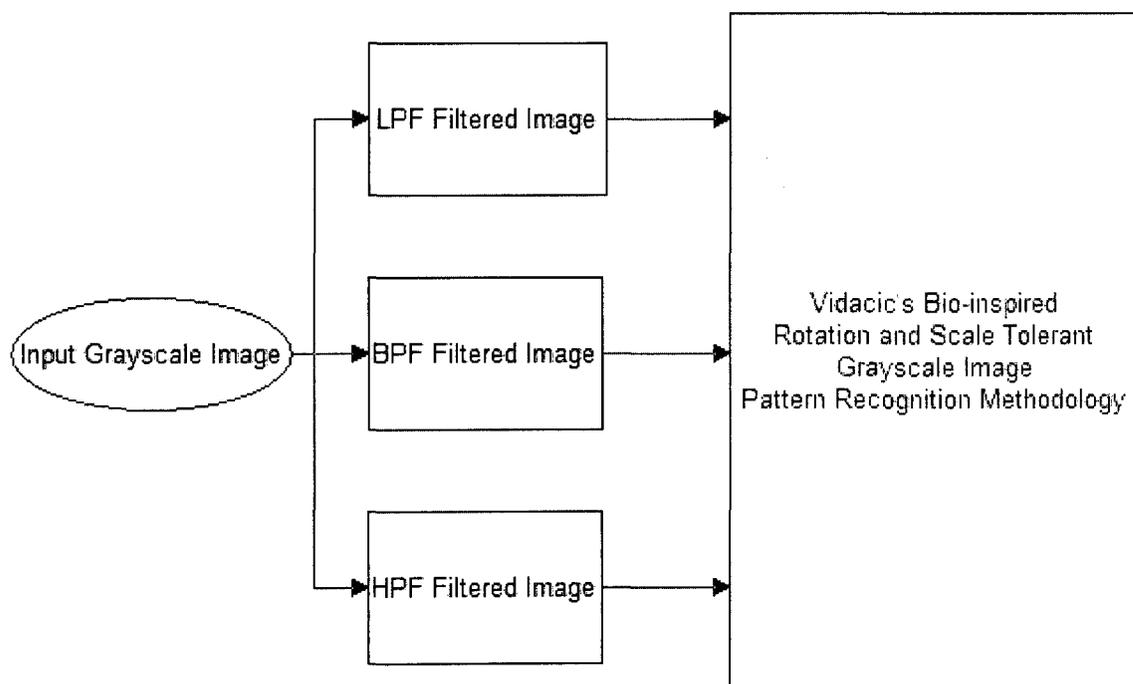


Fig. 2 The proposed grayscale image pattern recognition methodology in multiple spatial frequency bands

Based on the ideas outlined above, it is natural for us to think deeper about these preprocessing filters. Is it possible to cascade the above two preprocessing

stages together before performing the later pattern recognition tasks? This proposed methodology of combining the color image with spatial frequency filtering prior to pattern recognition is shown in Fig. 3. The question is: will the recognition results be better than that of only performing one alone?

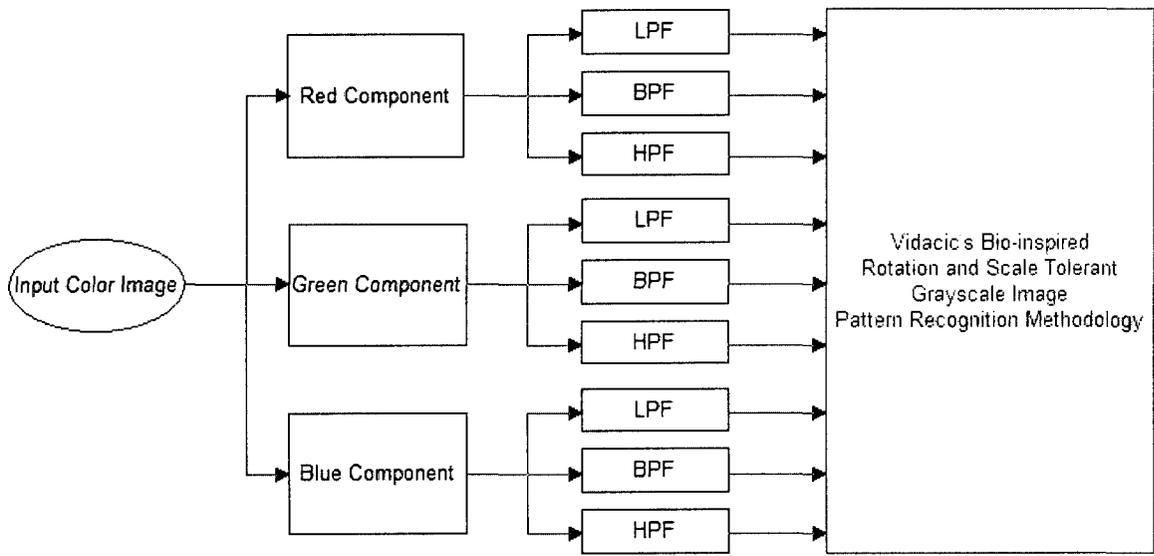


Fig. 3 The proposed cascaded color image pattern recognition methodology in multiple spatial frequency bands

The following contents of this thesis will verify the improvement of the color image pattern recognition performance compared with the Vidacic's grayscale image pattern recognition by simulation.

This thesis is partitioned into six chapters. The thesis structure is as follows:

Chapter I presents the background knowledge, prior researches, basic ideas and the introduction of the major areas addressed in this thesis.

Chapter II introduces Vidacic's methodology of biological inspired (LSI based) feature extraction for rotation and scale tolerant pattern analysis. This

methodology is used following the preprocessing steps outlined in the later chapters.

Chapter III proposes the color image preprocessing of the biologically inspired deformation invariant image pattern recognition system. Object classification simulations and results analysis are included.

Chapter IV exploits the multiple spatial frequency bands filtration process applied on the grayscale images. The experimental results of pattern recognition are also shown and explanations to these results are given.

Chapter V includes the conclusion and future work explorations.

Appendix A provides a view which exploits the optical implementations of rotation and scale distortion invariant pattern recognition which can be accomplished by using optical correlators.

Appendix B demonstrates the theoretical derivations of the statement that multiple channels processing in parallel can lead to smaller recognition failure rate in the filtering point of view.

Appendix C provides the organization of the simulation codes in the CD.

Appendix D introduces the installation and environment setup of OpenCV2.0 on Microsoft Visual Studio 2008.

# CHAPTER I

## BACKGROUND

### 1.1 Introduction of Log-Polar Mapping

At the highest levels of evolution, living beings have developed vision systems that are both active and space-variant. In the vision system of primates and specifically in the human eye, a space-variant image sensor structure exists and is characterized by a non-uniform pixel geometry. In biological systems such as these, at least two different areas can be recognized: the fovea, characterized by a relatively uniform sensor distribution of fine resolution, and the periphery, where the sensor distribution is sparser and is non-uniformly distributed. An example of a space variant resolution image is presented in Fig. 1.1. Details can be detected clearly in the central area, but the periphery allows a more coarsely sampled area which covers a large visual area. Pixel or sensor dimensions which follow such a methodology follow a monotonic decreasing law, which is a function of the distance from the center radial focus point.

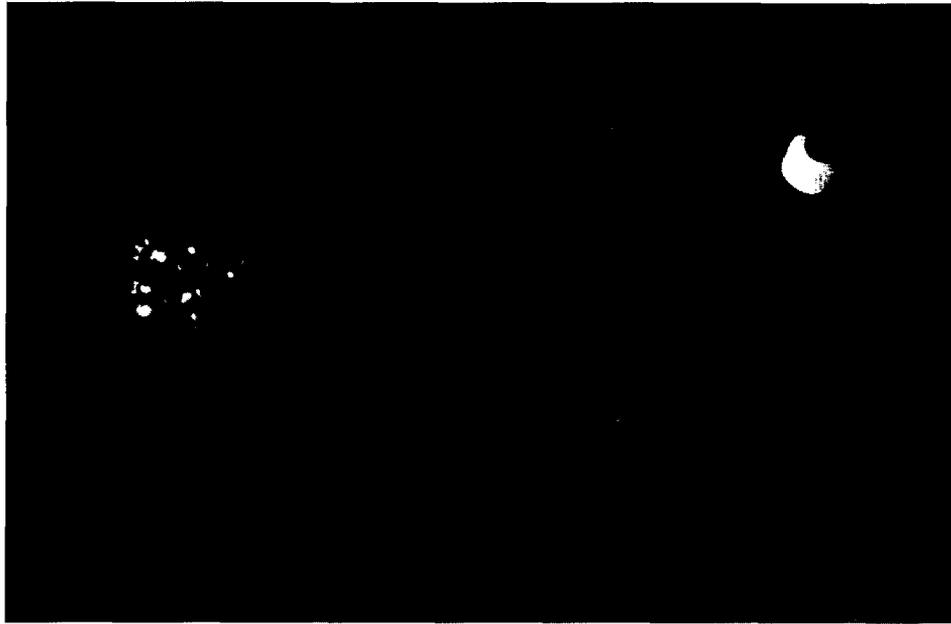


Fig. 1.1 A foveated image (Fovead on the pistil of the left flower).

The log-polar mapping was first motivated by its resemblance with the structure of the retina of mammalian biological vision systems. It has been found that the excitation of the cortex can be approximated by a log-polar mapping of the eye's retinal image. In other words, the real world projected onto the retinas of our eyes, is reconfigured onto the striate cortex of the brain by a process similar to a log-polar mapping before it is processed by our brain [7]. In the human visual system, the cortical mapping is performed through a space-variant sampling strategy, with the sampling period increasing almost linearly with the distance from the fovea. Within the fovea the sampling period becomes almost constant. This retino-cortical mapping can be described through a transformation from the retinal plane onto the cortical plane [8]. The foveated eye model which is inspired by biological solutions is adopted here for low and medium level image analysis. In particular, log-polar maps result in a reduced amount of data

required to be processed, while preserving details in the fovea necessary for robust pattern recognition [9].

The structure of the human retina-like model is based on a discrete distribution of elements whose radii increase linearly with eccentricity in the visual field. In the early 1980s [10], the technology of solid-state video cameras and scanning procedures were used to implement a proposed retina that was thought to be similar to the actual human visual system. Adopting the log mapping foveated system solution, the receptive fields' structure is composed of concentric rings at increasing distance from the center and receptive fields have different radii [11]. Moreover, a linear relationship between the reciprocal of acuity and retinal eccentricity has been found at least up to  $30^\circ$  of visual field. This linear relationship is in good agreement with the receptive field size of retinal ganglion cells [12, 13], as shown in Fig. 1.2.

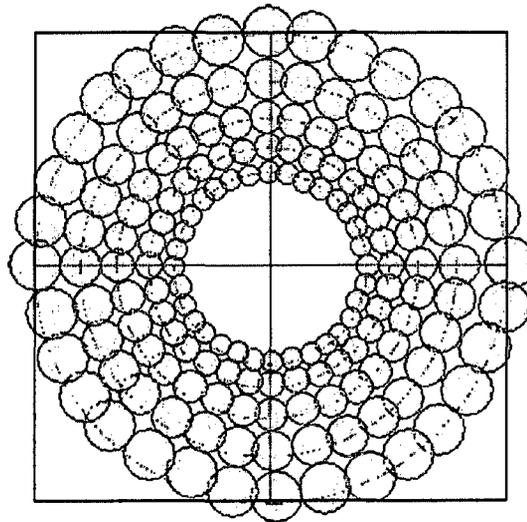


Fig. 1.2 Circular receptive fields with different radii are distributed along concentric rings.

The conformal logarithmic mapping (CLPM) was first discussed in detail by Weiman and Chaikin [14] who provided an in-depth mathematical treatment of it

and also suggested how it might be useful in image processing techniques. Several other authors have used the CLPM for different applications, but it was Messner and Szu [15] who were the first to discuss the leading edge scale and rotation invariance properties of the CLPM. The CLPM is a non-linear mapping of the Cartesian plane  $(x, y)$  to the log-polar plane  $(r, \theta)$  as shown in Fig. 1.3.

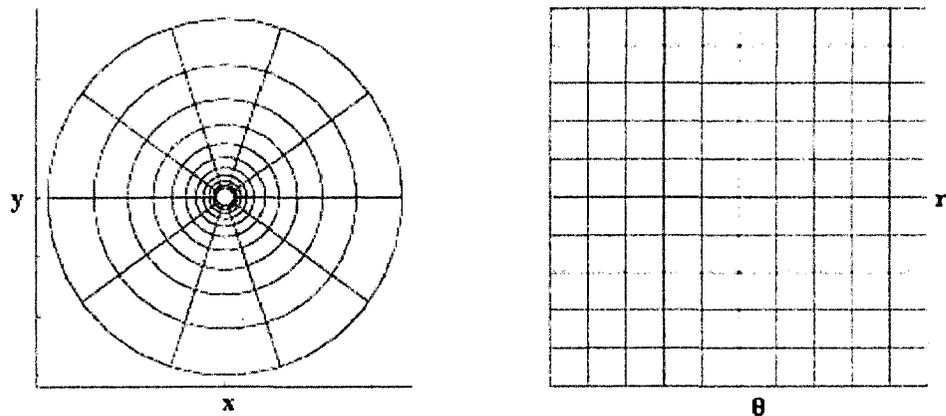


Fig. 1.3 Illustration of the Cartesian plane and the log-polar plane.

In the case of machine vision, the mapping is actually discrete. The Polar Exponential Grid (PEG) is placed over the input image for non-uniform sampling.

The equations for the log-polar mapping are as follows:

$$r = \ln \sqrt{x^2 + y^2} \quad (1.1)$$

$$\theta = \arctan\left(\frac{y}{x}\right) \quad (1.2)$$

This transformation presents some interesting properties for the scale and rotation invariance about the origin in the Cartesian plane. It can be seen from Fig. 1.4, that a single circle maps to a single horizontal line in the transform space since the radius of the circle at all angles is given by a constant value  $r$  coordinate for all  $\theta$  coordinates. Similarly an image of radial lines which have

constant angle but variable radius, results in a map of vertical lines in the transform space [16]. These mapping characteristics are well known and are fundamental to many rotation and scaling invariant pattern recognition algorithms. Rotation and scaling change result in rotational shifts along the  $\theta$  axis and translational shifts along the  $r$  axis, respectively. This results in a transformation of rotations and scale in the input image field into translational shifts in the log polar space. If an object is rotated around the origin, its mapping stays on the same horizontal line moving from the left to the right. If a point moves out from the origin along a radial line, this will result in only a vertical displacement of the mapped image. This has been clearly documented and accepted by researchers [10].

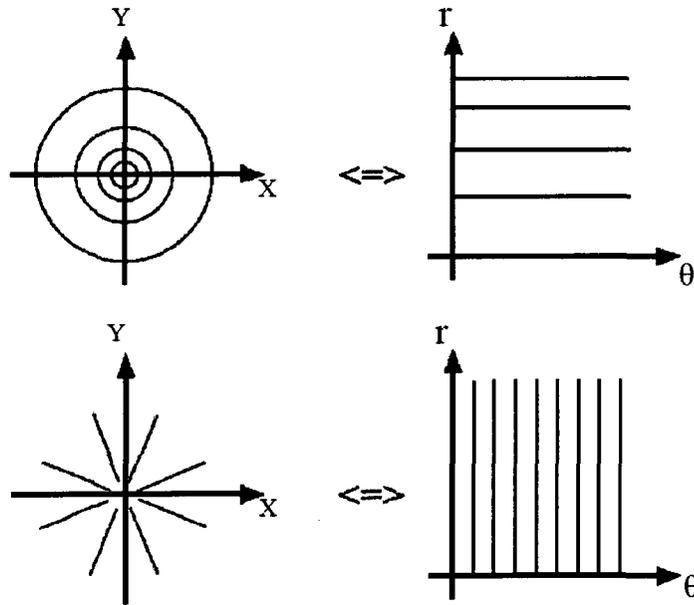


Fig. 1.4 The log-polar mapping applied to regular patterns (concentric circles and radial lines in the image plane).

## 1.2 The Lateral Subtraction Inhibition (LSI) Network Model

A very interesting phenomenon that occurs in the human eye (i.e., visual plane) might be able to be exploited for our purposes. It is observed that an object is perceived to be darker if the surrounding area is bright, and vice versa. Such a phenomenon can be modeled in various ways and there are many physiological and psychological tests that can support the existence of what is now commonly referred to as the Lateral LSI phenomenon in the mammalian retina [17, 18]. Lateral inhibition is a competitive interaction between photoreceptive neurons. In the human visual system (HVS), light falls onto both light (cone) and dark photoreceptors (rod) causing the two regions to compete with one another. This increases the contrast and sharpness in the visual response and can cause two effects: simultaneous contrast and the Mach band effect [19]. Mach band effect is shown in Fig. 1.5. The dash line is the intensity values perceived by the HVS.

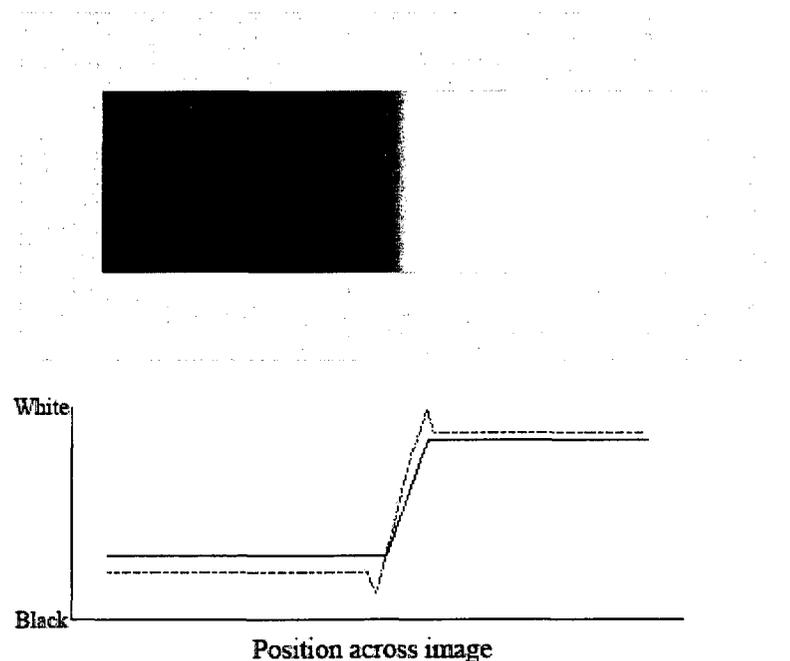


Fig. 1.5 Mach Bands - The perceived brightness is enhanced at the edges

For instance, when the eyes are focusing, several sensory neurons in the fovea next to one another are stimulated. Neurons that are firing suppress the stimulation of neighboring neurons. In the face of inhibition, only the neurons that are most stimulated and least inhibited will fire, so the firing pattern tends to concentrate at stimulus peaks. This contrast between the light and dark creates a sharper image.

Fig. 1.6 is a depiction of an extreme close-up to show how photoreceptors are arranged along the retina. The lighter photoreceptors are the ones that activate in response to light while the others are the photoreceptors that activate in the absence of light. As you can see, some of the dark photoreceptors encircle the lighter ones and vice versa. In reality, however, these two types of photoreceptors look the same. Each photoreceptor circle creates what is called a receptive field and connects to a cell that relays information from its connected photoreceptors to the brain.

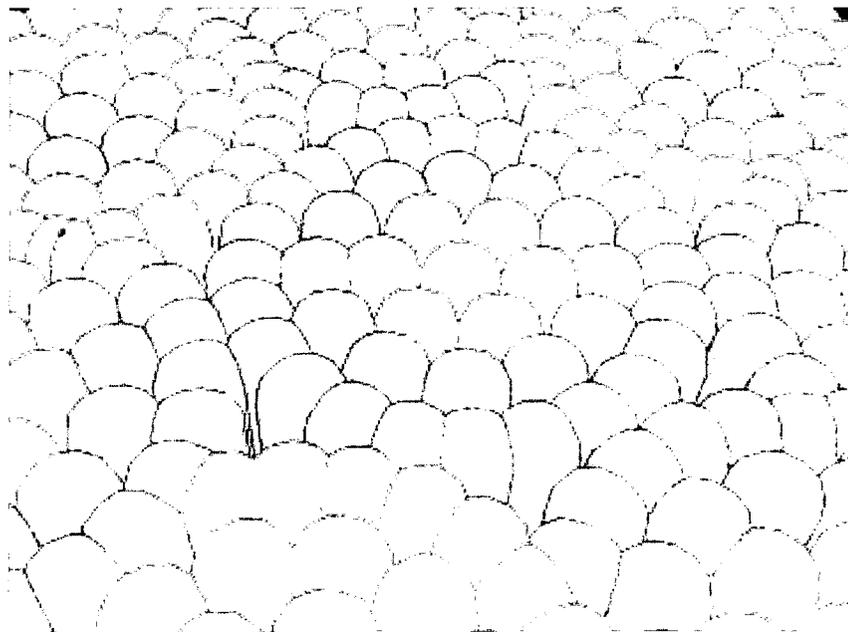


Fig. 1.6 How photoreceptors and receptive fields are arranged [20]

According to Furman [21], there are two types of lateral inhibition interaction: subtractive and shunting. Each of them has a lateral information flow coupling model defined as either feed-forward or feed-backward. The subtractive inhibition is more intuitive in that it behaves as a common feedback process between the input photo luminance signal and the neural output signal. In this case the linear relationship can be expressed in a matrix form. In addition, since the feed-backward model more closely resembles the biological system, only the subtractive type of lateral inhibition and the feed-backward information flow model is adopted in this thesis.

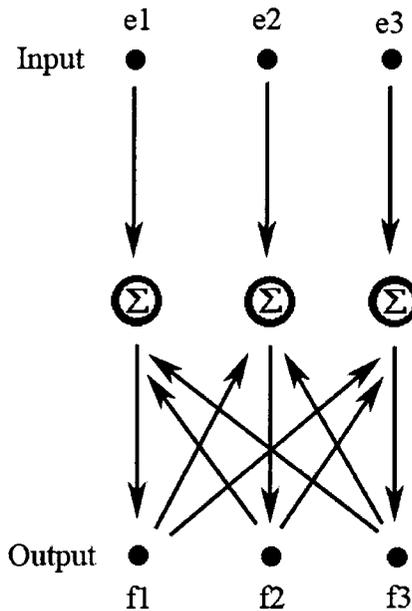


Fig. 1.7 Lateral Subtractive Inhibition using the feed-backward coupling model.

For the 1-D case, the following equation expresses the LSI using the feed-backward coupling model:

$$f_i = e_i - \sum_{\substack{j=1 \\ i \neq j}}^N b_{ij} f_j \quad (1.3)$$

Here  $e_i$  and  $f_i$  respectively represent for the input luminance over the receptor field and output neural firing rate in response to that illumination. The coupling coefficient,  $b_{ij}$  represents the interaction between the photoreceptors with index  $i$  and  $j$ . When  $i = j$ ,  $b_{ij} = 0$  which implies that there is no self-inhibition for photoreceptor  $i$ .

For the 2-D case (N-by-N), the corresponding equation can be extended from 1-D case to the following 2-D case as follows:

$$f_{ij} = e_{ij} - \sum_{\substack{l=1 \\ i \neq j}}^N \sum_{\substack{k=1 \\ k \neq l}}^N b_{ijkl} f_{kl} \quad (1.4)$$

(i,j) and (k,l) are the two coordinates of two photoreceptors. Similarly,  $b_{ijkl}$  is the coupling coefficient between these two receptors. When  $k = i$  and  $l = j$ ,  $b_{ijkl} = 0$ .

Reformulating into a Matrix form, we have:

$$[F] = [E] - [B] * [F] \quad (1.5)$$

To understand this equation it is advantageous to take a 3-by-3 receptor field to explore, if we consider the following:

$$[F] = [[F_1][F_2][F_3]]^T \quad (1.6)$$

$$[E] = [[E_1][E_2][E_3]]^T \quad (1.7)$$

$$\text{Where: } [F_1] = [f_{11} f_{12} f_{13}]^T$$

$$[F_2] = [f_{21} f_{22} f_{23}]^T$$

$$[F_3] = [f_{31} f_{32} f_{33}]^T$$

$[B_{ik}]$ : is a 9-by-9 matrix consisting of the feed-backward coupling coefficients.

For  $b_{ijk}$ , representing the interaction between receptors  $b_{ij}$  and  $b_{kl}$ , intra-row and

inter-row receptor coupling coefficient.  $i, j$  and  $k$  vary from 1 to 3 respectively in this 3-by-3 receptor field case. The final solution to the Eq. 1.7 is:

$$[F] = [S]^{-1} * [E] \quad (1.8)$$

$$\text{where: } [S] = [I] + [B] \quad (1.9)$$

Here the Matrix  $[I]$  is the identity matrix. One important characteristic of the matrix  $[S]$  is its symmetry. There are many different classes of coupling models including Furman's inverse-cube/inverse/inverse-square/neighbor coupling [22]. Here we adopt Messner's exponential coupling model  $b_{ijkl} = e^{-1.55d}$  [4], where  $d$  is defined as the Euclidian distance between receptors.

Adopting the above mentioned background knowledge, the input image can be mapped to the log-polar domain for scale and rotation invariant image processing. The biological inspired LSI model is utilized as the pre-whitening filter for better performance of whitening and subsequent feature extraction and encoding. In Chapter III and Chapter IV the corresponding pattern recognition methodology for color image and multiple spatial frequency filters are investigated, respectively.

## CHAPTER II

### THE LSI BASED GRAYSCALE IMAGE FEATURE EXTRACTION

In this chapter, we are going to introduce Vidacic's grayscale image pattern recognition methodology. In this methodology, the complex log-Gabor filtering is applied on the log-polar mapped image to extract pseudo saliency patches. Pre-whitening filters are applied based on network with lateral recurrent connectivity. Then the whitening by Plumbley algorithm is followed and multi-scale ICA technique is adopted to encode the features for training or pattern classifications. The purpose of introducing this chapter is to get the readers familiar with the theory behind the steps in the Vidacic's bio-inspired log-polar based image pattern analysis methodology. Most of the algorithms and descriptions in the log-polar domain feature extraction process are followed without alteration from Vidacic's methodology [6]. The later preprocessing designs are extended from the basic methodology in this chapter.

#### 2.1 Pre-Whitening Filter Adopting LSI Recurrent Network Model

The pre-whitening process is essential for the final pattern recognition process for extracting of meaningful feature sets. Neural networks with recurrent

connectivity have widely been used for modeling visual information processing in biological systems. After the color image/multiple frequency preprocessing, the pre-whitening (i.e., decorrelation) process is usually applied following the collection of patches extracted by the log-Gabor filtering process.

### ***2.1.1 Solutions of LSI model applied in log-polar domain***

As we stated in the previous chapter, the LSI model is adopted to play the role of a retino-cortical filter in the recurrent linear network because of its resemblance to the structure of retina cells. The microscopic neuron network solution is usually used to describe the model, by implementing Eq. 1.8 on the visual plane. The linear model of the recurrent linear network can be described as follows:

$$y = (I + B)^{-1}x \quad (2.1)$$

The microscopic neuron network solution was found by approximating the inverse of the following matrix by a convergent power series:

$$(I + B)^{-1} = I - B + B^2 - B^3 + \dots \quad (2.2)$$

However, this model requires an extremely large amount of computation on matrix inversions when dealing with relatively large images. An important result is that for some special low pass filters, the spatially non-uniform filtering in foveation can be achieved by uniform low-pass filtering in log-polar space [36]. Some alternative models that are introduced in [6] can be adopted to solve the networks with LSI and relate them to log-polar coordinate mapping.

The iterative-based and convolution-based models are introduced to avoid the large matrix computation and thus save computation time. The output of neighboring inhibitory receptors in the network can be described as:

$$f_{ij}^n = e_{ij}^n - \sum_{\substack{k=1 \\ k \neq i}}^N \sum_{\substack{l=1 \\ l \neq i}}^N b_{ijkl} f_{kl}^p \quad (2.3)$$

Where  $n$  represents the current time instance and  $f^p$  represents the most recently obtained value of output. The coupling model  $b_{ijkl} = e^{-1.55d}$  mentioned in the previous chapter is adopted for the later simulation.

|           |           |           |           |           |
|-----------|-----------|-----------|-----------|-----------|
| 0.0011361 | 0.0036029 | 0.0068373 | 0.0036029 | 0.0011361 |
| 0.0036029 | -0.041248 | -0.1828   | -0.041248 | 0.0036029 |
| 0.0068373 | -0.1828   | 1.1714    | -0.1828   | 0.0068373 |
| 0.0036029 | -0.041248 | -0.1828   | -0.041248 | 0.0036029 |
| 0.0011361 | 0.0036029 | 0.0068373 | 0.0036029 | 0.0011361 |

Table 2. 1 The convolution kernel for coupling model  $b_{ijkl} = e^{-1.55d}$

The network response can also be obtained by performing a convolution operation with a determined small kernel (For example: 5 x 5).

The convolution based model is more practical than the iterative model for the large image case, while the direct network solution can be easily utilized for small networks. The convolution by kernel  $g(x, y)$  applied to the pre-mapped space is defined as:

$$t(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\alpha, \beta) g(x - \alpha, y - \beta) d\alpha d\beta \quad (2.4)$$

while the convolution in mapped space is defined as:

$$t^*(u, v) = \int_{-\infty}^{\infty} \int_0^{2\pi} h^*(u_\alpha, v_\beta) g^*(u - u_\alpha, v - v_\beta) du_\alpha dv_\beta \quad (2.5)$$

where  $(x, y) \leftrightarrow (u, v)$  and  $(\alpha, \beta) \leftrightarrow (u_\alpha, v_\beta)$  represent coordinate mapping pairs. It can be shown that the convolution in the pre-mapped domain is identical to convolution in the log-polar space, i.e.  $t(x, y) = t^*(u, v)$ .

This recurrent network filtering will enhance the edge information in the input image. The impulse response of this network is similar to a circular Gabor filter, which can be constructed as a linear combination of Gaussian filters (DOG kernels) [37]. These filters have a characteristic center-surround shape. If the circular kernels are scaled linearly with their distance from the log-polar mapping origin, the filters in the mapped domain will remain spatially uniform. For faster execution, the mapping can be implemented through a log-polar look up table.

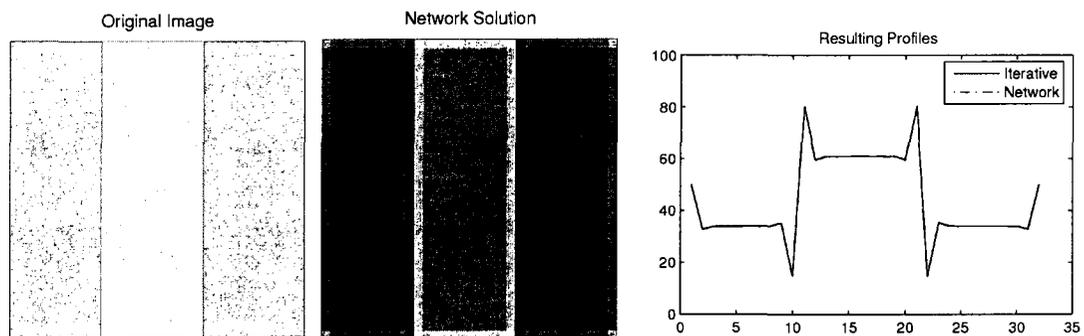


Fig. 2.1 The result of the network, the iterative solutions for the 2-D grid with lateral inhibition; original image (left), resulting image (middle); Resulting signal profiles: network versus iterative solution (right); NRMSE = 0.0038, Image size used is 32 x 32 [6]

The exact network solution is non-uniform lateral inhibition. However the actual processing on natural image was performed in the uniform log-polar space. In the localized region, the edge information was enhanced near the fovea, while the peripheral area is not inhibited. Since in a machine vision system the foveation point can move from one spot to the other, the non-uniform LSI model is not employed for later use.

Inter-neuron coupling models that inspired by functions used in stages of biological vision need to be defined prior to constructing the recurrent network. By utilizing spectral properties of its connection matrix, large recurrent network can be solved. According to [6], several practical models can be chosen as the connection matrix. They are the Gaussian model, Difference of Gaussian (DOG) model, Difference of Offset Gaussian (DOOG) model and Second Order Derivative of Gaussian (2ODG) model [38]. The steerabilities of these models were also investigated in [39, 40]. The guideline to choosing a proper inter-neuron coupling model to avoid large matrix inverse was revealed in [4]. The simulation showed the circular symmetric inter-neuron coupling models (Gaussian and DOG) are much less orientation sensitive than the non-circular models (e.g. DOOG and 2ODG).

### ***2.1.2 Decorrelation in the adaptive recurrent networks***

The previous mentioned LSI filters are learned during the network training. Barlow and Földiák [41] have shown that the linear model of network with recurrent inhibition stores information about the covariance matrix corresponding to the input data. As a result of such learning, the outputs of the system become uncorrelated. The adaptation algorithm for this network given in a matrix form can be represented as:

$$\Delta B = \alpha (yy^T - \beta I) \quad (2.6)$$

where,  $\alpha$  is the small learning constant and  $\beta$  is the parameter that determines the value of the final output signal variance – the output covariance matrix becomes diagonal matrix with value  $\beta$  on the main diagonal.

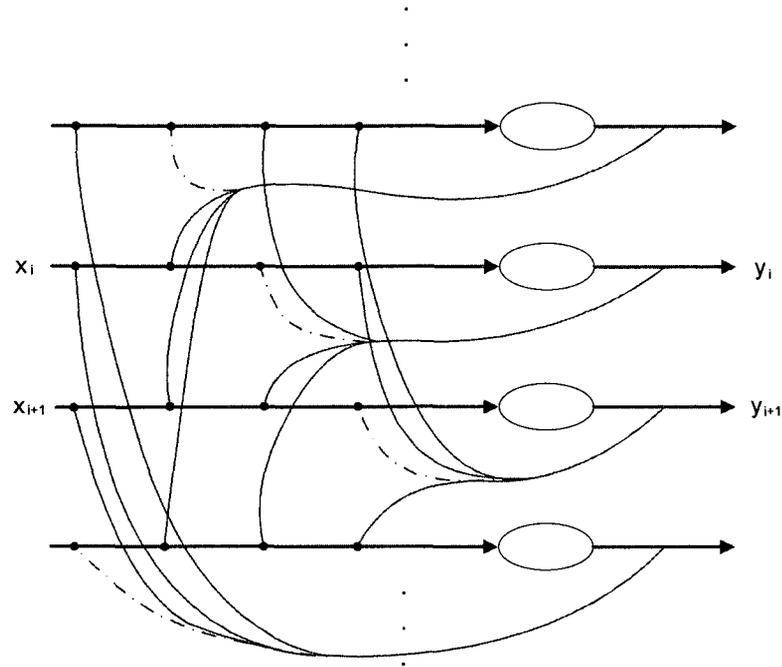


Fig. 2.2 The recurrent network with adaptive connections (self inhibition/excitation can potentially be allowed) used in [41]

The expectation is that this filter has a faster learning speed using the neural adaptive whitening process than when the weights are initially zero or random values.

The simulation results show that the data preprocessing with filters formed by spatially invariant connectivity models is almost as effective as preprocessing the particular category of data with the network whose connections are previously adopted to completely whiten the input. Certain prior knowledge about the specific category of data to be processed is necessary. This method of signal processing can be classified into the category of semi-blind data pre-whitening

since only partial information about input is available. In the following text, the use of category specific, spatially uniform pre-whitening networks in ICA based pattern recognition tasks is addressed. Specifically, the usefulness of such filters in obtaining feature sets of data in the log-polar domain is investigated.

Plumbley has shown that by adding a simple self-inhibitory connection, the fully connected recurrent network with lateral inhibition can be utilized to perform data whitening [25]. In that work, the author describes the learning algorithm for maximization of transmitted information through the network when the outputs are corrupted by additive noise. When the outputs of the network are decorrelated, the optimization of information transfer is achieved. This characterization of decorrelation is equivalent to whitening of the output signals. The input data is whitened by category-specific adaptive model. This method can be classified into category of semi-blind data pre-whitening since only the image category is available. If the wrong filter is chosen, the whitening becomes much slower and relatively inefficient. In this thesis, the category specific pre-whitening filters in ICA based pattern recognition tasks are used.

## **2. 2 The Introduction to ICA for Feature Extraction and Encoding**

Over the past decade, the ICA (Independent Component Analysis) technique has gained a lot of attention. ICA tries to find the linear combination of statistically independent sources. After the whitening of the input, the ICA processing becomes easier. Whitening of the input data is required by some ICA algorithms

while with others, it improves their convergence properties [22] [42, 43]. Here it is necessary to reveal some of the basic characteristics and significance of ICA. Olshausen and Field demonstrated that the characterization of mammalian visual cortex receptive fields is consistent with the sparse coding of natural images [44, 45]. Such sparse distributions reduce statistical dependence among output units, providing the link between sparse coding and the ICA [46, 47]. ICA can be used for many applications like feature extraction, signal separation or reducing noise in natural images.

First, the rotational and scaling variations in input patterns are converted to horizontal and perpendicular shifts in log-polar space. This is the reason why the log-polar mapping is usually referred as a very efficient image processing step towards scale and rotation tolerant feature extraction. The pre-whitening filters described in Chapter 2.1 act upon signals formed from localized image patches.

After the ICA based feature extraction techniques are applied, the rotation and scale tolerant feature space is formed and encoded via ICA filters by using a limited number of locations in the image. This stage is called pseudo-saliency detection. These locations are chosen based on the responses of biologically inspired complex log-Gabor kernels [48]. Locations selected based on the highest responses of such filters are collected as pseudo-saliency points. The motivation for usage of log-Gabor instead of the most commonly used Gabor filters here comes from Field's work [49]. Field use log-Gabor filters for feature encoding because that the symmetry of their frequency response is consistent with symmetry of spatial-frequency response of visual neurons. The log-Gabor

function has the ability to capture broad spectral information with a very compact spatial filter [48].

If the system input's dimension is  $n^2$ , where  $n$  is relatively small patch size (8-16), the matrix inversion problem is not significant for computation today. This network gives us advantage when we process the local image patches instead of processing the entire image. In the simulation example we extract 16 pseudo salient local patches of the image (An example is shown in Fig. 2.3). The data collection from all images was performed by using patches of size 16 x 16. The feature vector can be represented as a matrix containing 16 columns, each of which encoding one salient location. The first two components in the column are the x and y coordinates of the location center. Then the resulting data set was stored as a matrix with each row representing one component of the input.

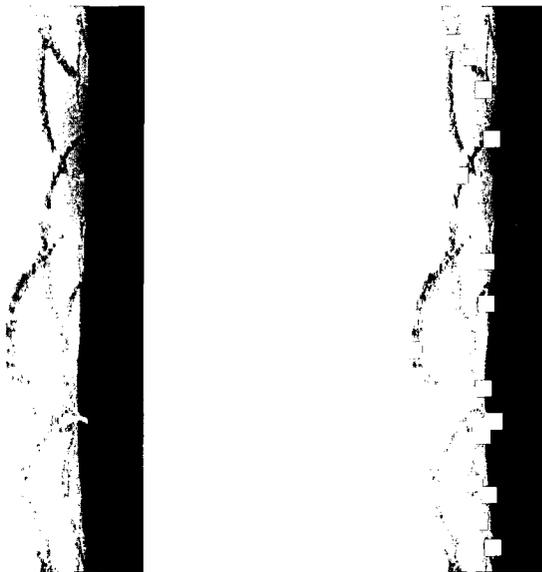


Fig. 2.3 An image in log-polar space (left) with outlined pseudo-salient locations (right) [6]

At the final stage, object recognition based on multiple classifiers is performed. Different types of classifiers were tested: nearest neighbor, 3-nearest neighbor

[50] and a hybrid nearest neighbor model of the previous two types of classifiers. For each test image there would be a voting procedure based on number of features/patches identified to belong to the same class. In the case of hybrid 3-nearest neighbor classifier, each patch was characterized as belonging to the particular classes based on three nearest patches in the training set. The final winner is the class which has the highest votes of patches.

If the recognition is performed on a rotated object, a variable offset is added to the coordinate  $x$  of each column vector. When implementing the hybrid nearest neighbor classifier, the winner class of each pre-determined rotation offset can be found. The pattern resulting in minimum cumulative distance parameter is chosen as the input for final classification. When implementing the hybrid 3-nearest neighbor classifier, besides deciding the winner class, a confidence value is determined representing the number of times that any of the feature sub-vectors was classified to belong to winning object. The overall winner is the object with highest confidence count.

In brief, the complex log-Gabor filtering is utilized to extract pseudo saliency patches on the log-polar mapped image. The LSI based pre-whitening filters are applied subsequently. Then the whitening and the multi-scale ICA analysis were followed to encode the features for training or pattern classifications. Finally we implement pattern classification of the non-rotated and rotated objects by using different pattern recognition classifiers.

## CHAPTER III

### THE BIO-INSPIRED METHODOLOGY OF COLOR IMAGE PATTERN RECOGNITION

When we process real world objects, our pattern recognition system should have the ability to deal with color information. In this chapter we are going to examine color image pattern recognition - extending the capability of grayscale image pattern recognition methodology explored by Vidacic [6]. First the fundamentals of color image models and processing are introduced. Secondly we present two methods to perform the centroid calculation for color image. The experimental results of the proposed color image object recognition methodology are followed together with the results analysis and discussion.

#### **3.1 A Methodology for Color Image Pattern Recognition**

In a typical image pattern recognition system the first steps will include image acquisition and enhancement, prior to any subsequent feature extraction. There are many different methodologies that have been proposed by different authors on the topic of color image pattern recognition emphasized in different

processing stages [23, 24]. In this thesis, we propose a methodology based on Dragan Vidacic's PhD research [6]. The methodology of the color image pattern recognition can be achieved by a number of processes which can be outlined as follows.

1. First is the color image sensing and acquisition process. Since this is accomplished in the spectral frequency domain, the color image can be decomposed to three RGB channels: red, green and blue channels (multi-channels). The signal of each individual color channels can be processed individually in the later processes.
2. Secondly, object detection and image segmentation. Calculation of the center of the individual channel of the color (RGB) object on a black background scene.
3. Then log-polar mapping (a non-uniform sampling) is applied to convert the centered image from visual plane to cortical plane (spatial domain to log-polar domain).
4. A complex log-Gabor filter is followed to process the mapped image. After filtering the mapped image, a certain number of the pseudo salient patches are collected and restored as a series of eigenvectors for the particular image.
5. A pre-whitening filtering is followed using the spatially invariant network model with lateral inhibition subtraction. The real whitening process implemented by Plumley algorithm [25] is combined with the pre-whitening filtering.
6. Feature encoding is applied to extract the feature vectors by using the Independent Component Analysis (ICA).

7. Judgment is made based on whether the image is for training or for classification. If the image is used for training, then restoration of the feature vectors is done. If the image is for classification, then it is ready to compare the extracted feature vectors with the restored ones in the classification database to make the judgment.

Chromatic digital image acquisition requires two elements: a physical device capable of chromatic sensing and a digitizer. The sensing process can be implemented either remotely or locally by using a digital camera (CCD sensor), or other optical sensor arrays.

In this thesis, emphasis is not put on the image acquisition. However, the digitizing process of the data is important since we need to restore the datasets of each pixel of the image and process these data by performing feature encoding.

Although here the RGB decomposition process is put right after the color image acquisition process, it is possible to be put in the system as a later step if desired. The advantage of placing it at the beginning of all the processes is that it requires less computation than other schemes. For example, if we put this step after log-polar mapping, we have to perform the log-polar mapping on a color image rather than on grayscale R, G, B intensity channels individually. For some categories of images one channel of the image is sufficient to achieve substantial pattern recognition performance (e.g., HVS is more sensitive to green channel information). Other schemes can be further studied and compared for particular emphasis.

The block diagram of the methodology mentioned above is shown in Fig. 3.1. All later steps following the color decomposition process are the same as those proposed by Vidacic [6].

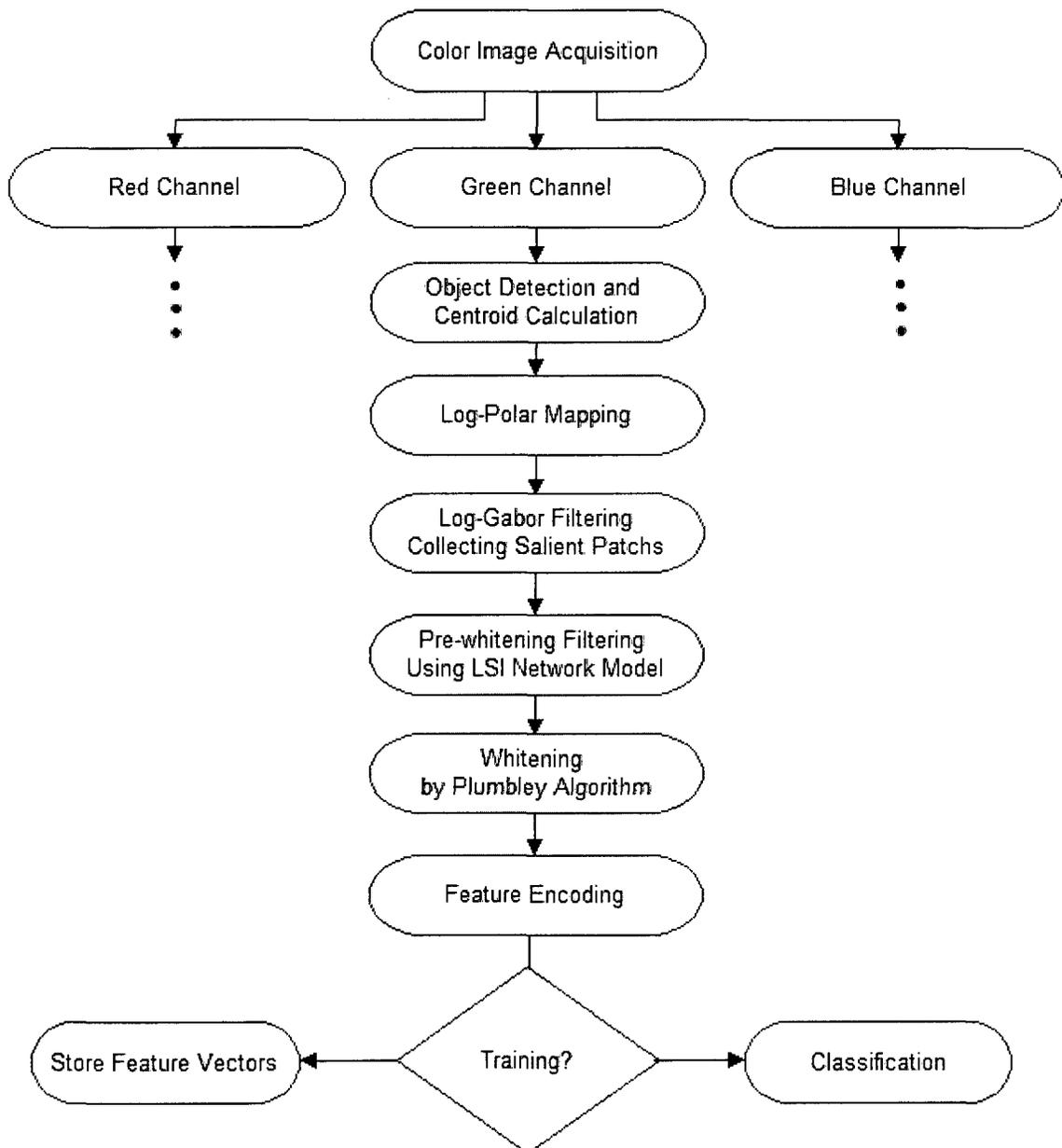


Fig. 3.1 The methodology of color image feature extraction and pattern recognition

### 3.2 Fundamentals of Color Image Processing

Humans can discern hundreds of thousands of colors and saturations. Color images can be generally quantized by using three (or more) color channels. It is possible to acquire the color image by using three independent sensors that are tuned for specific spectral bands of light (commonly used bands in today's cameras are called the red, green and blue channels). Similarly for displaying color images, we can use three emitters, each of which can generate a particular spectral band of light (again red, green and blue). After we mix the three channels together to produce an output we can display it on a color monitor where all colors covered by those three color channels can be shown. For the following discussion, the primary focus is on the processing of intensity images represented by a matrix of vectors which correspond to multiple channels. The luminance is the main concern here other than the chromaticity (hue and saturation). Generally, most processing can be extended to a color image space simply by applying it to every color component of the image [19].

Due to the absorption characteristics of the human eye, colors can be represented by the combination of the "primary colors": red (R), green (G) and blue (B). An RGB color image can be viewed as three monochrome intensity images (representing red, green and blue). There are numerous other color models in use today, for example: the CMY (cyan, magenta, yellow) and CMYK (cyan, magenta, yellow, black) models are often used in color printing; the YUV model for color TV; the HSI (hue, saturation, intensity) model. In this thesis, we

adopt the basic RGB model for further analysis. The conversion between different models can be achieved by using simple matrix operations [19].

Taking a full-color image for example, each of the red, green and blue images is an 8-bit image, the total number of the colors that can be represented in a 24-bit deep RGB image is  $(2^8)^3 = 16,777,216$ . Each red, green and blue channel is represented as an eight bit integer that can take on the values between zero and 255 (0 represents black and 255 represents white). The diagonal joining the black and white vertices is balanced by the three individual channels and represents the grey level intensity.

The color pixels of a full-color image can be interpreted as vectors. Here  $c$  is an arbitrary vector in RGB space, the size of the image is  $M * N$ , the spatial variables  $x$  and  $y$  indicate the coordinate of pixel location, for  $x=[0 M-1]$ ;  $y=[0 N-1]$ .

$$c(x, y) = \begin{bmatrix} R(x, y) \\ G(x, y) \\ B(x, y) \end{bmatrix} \quad (3.1)$$

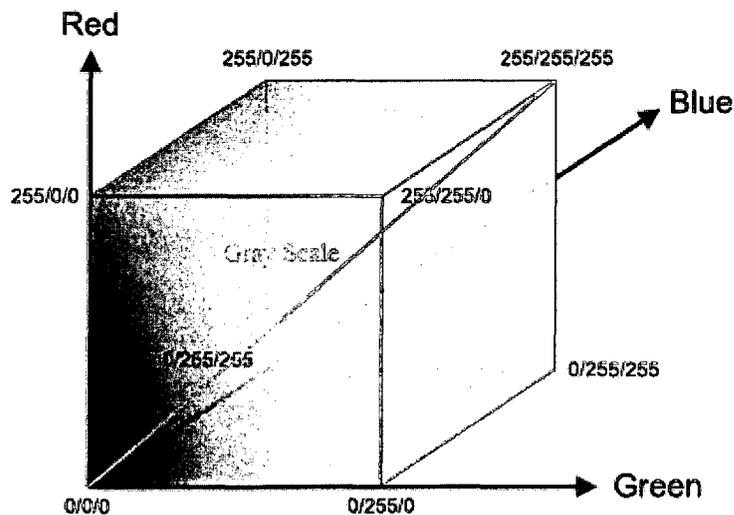
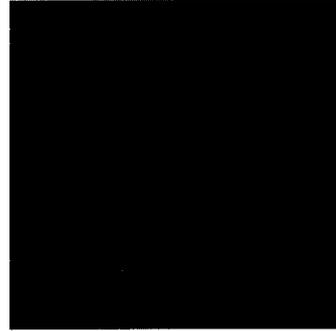


Fig. 3.2 The 24-bit RGB cube



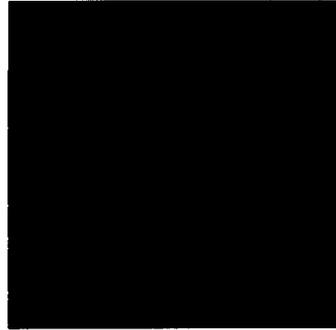
(a) RGB image



(b) Red component image



(c) Green component image



(d) Blue component image

Fig. 3.3 An RGB image example and original R, G, B channels



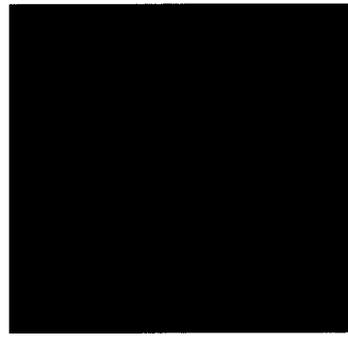
(a) RGB image



(b) Red component image(Intensity)



(c) Green component image(Intensity)



(d) Blue component image(Intensity)

Fig. 3.4 An RGB image example and the intensities of R, G, B channels

### 3.3 Object Centroid Calculation and the Log-Polar Mapping

In this thesis, the images we process are color objects sitting on a black background. Thus it is assumed that some sort of segmentation has been done prior to this study. Object centering is an essential step since the later procedure (especially the log-polar transformation) requires an image with an accurate calculated centroid. A little displacement of the centroid will introduce a large offset in the log-polar domain based on the scale and rotation change which correspond to the center point of the transform. Here we introduce two different approaches to perform the color image object centering.

The first approach is based on the segmentation applied in RGB vector space (Fig. 3.5). Suppose the image we are processing does not contain severe noise components or has already been processed to reduce any noise. This method can be adopted on the condition that the color object is monochrome. We need to classify each RGB pixel, to compare the Euclidean distance between an arbitrary point and the average column vector. Here let  $z$  denote an arbitrary point in RGB space and the RGB vector  $m$  denote the average color. If the distance is within the threshold  $T$ , the pixel will be stored as a part of the object.

The Euclidean distance between  $z$  and  $m$  is given by

$$\begin{aligned} D(z, m) &= \|z - m\| \\ &= \left[ (z_R - m_R)^2 + (z_G - m_G)^2 + (z_B - m_B)^2 \right]^{1/2} \end{aligned} \quad (3.2)$$

Generalizing Eq. 3.2 produces Eq. 3.3

$$D(z, m) = \left[ (z - m)^T C^{-1} (z - m) \right]^{1/2} \quad (3.3)$$

Here  $C$  is the covariance matrix of the samples representative of the color we want. For computation efficiency we use  $D$  squared. Furthermore, we can generate a bounded box centered on  $m$  which requires less computation than using a sphere centered on  $m$  directly as related to from Eq. 3.5. Following segmentation, the chromatic object is converted to grayscale for the center calculation.

In MATLAB, the function “rgb2gray” converts RGB values to grayscale values by forming a weighted sum of the R, G, and B components (Eq. 3.4).

$$\text{Grayscale } c[X \ Y] = 0.2989 * R + 0.5870 * G + 0.1140 * B \quad (3.4)$$

Here  $c$  is an arbitrary vector in RGB space, the size of the image is  $X * Y$ .

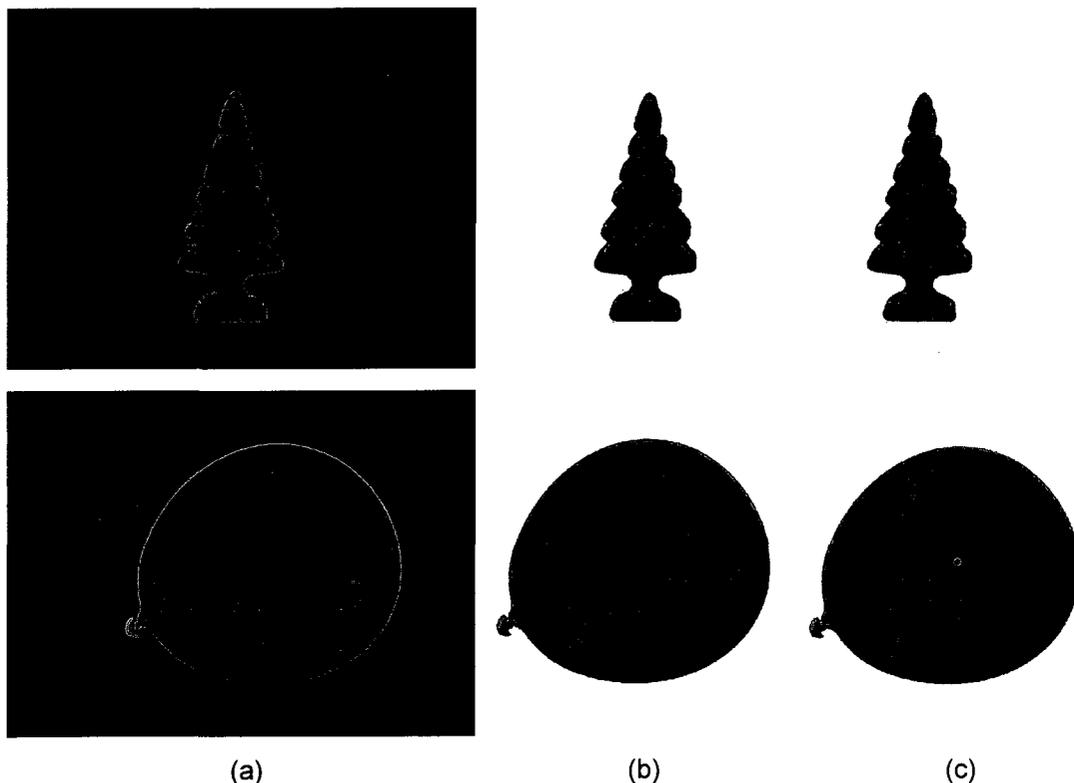


Fig. 3.5 (a) The monochrome object sitting in black background  
 (b) The object obtained after the segmentation  
 (c) The centered grayscale object

The second method decomposes the chromatic image into RGB planes and then operates on the grayscale image corresponding to the individual color channel. If we process in the spatial domain (i.e. image plane), we can perform image segmentation by setting up the threshold according to the ratio of the current pixel's gray level value to the maximum gray level value. The most practical way to center the object is to first make a binary image then use the centroid calculation. The centroid function is accomplished by using the "regionprops" function in the MATLAB image toolbox. Coordinate conversions are calculated and applied to the image object to center it within the image. Based on Otsu's threshold selection from gray level histograms algorithm which starts from the calculation of the first order moment, we calculate the object's centroid [26]. After finding the centroid of the object, the background is removed to a predefined width which surrounds the object. Finally the image is expanded by replicating the background to form images that are the same dimension.

This algorithm has limitations when being applied to certain types of objects. For objects where individual channels have very low dynamic range, the centering process may cut some parts of the object off. The same loss will happen when dealing with complicated objects with dark shapes inside of the object. Observing Eq. 3.4, the blue component has the lowest contrast while the green component has the highest contrast. Thus it can be seen that if any object cut-off loss happens, in most cases it will occur in the blue channel rather than the other two. This limitation implies that objects would necessarily need to be pre selected which is not a constraint that can be tolerated in an unstructured system.

An example of an acceptable object and an example of unacceptable objects are show in Fig. 3.6.

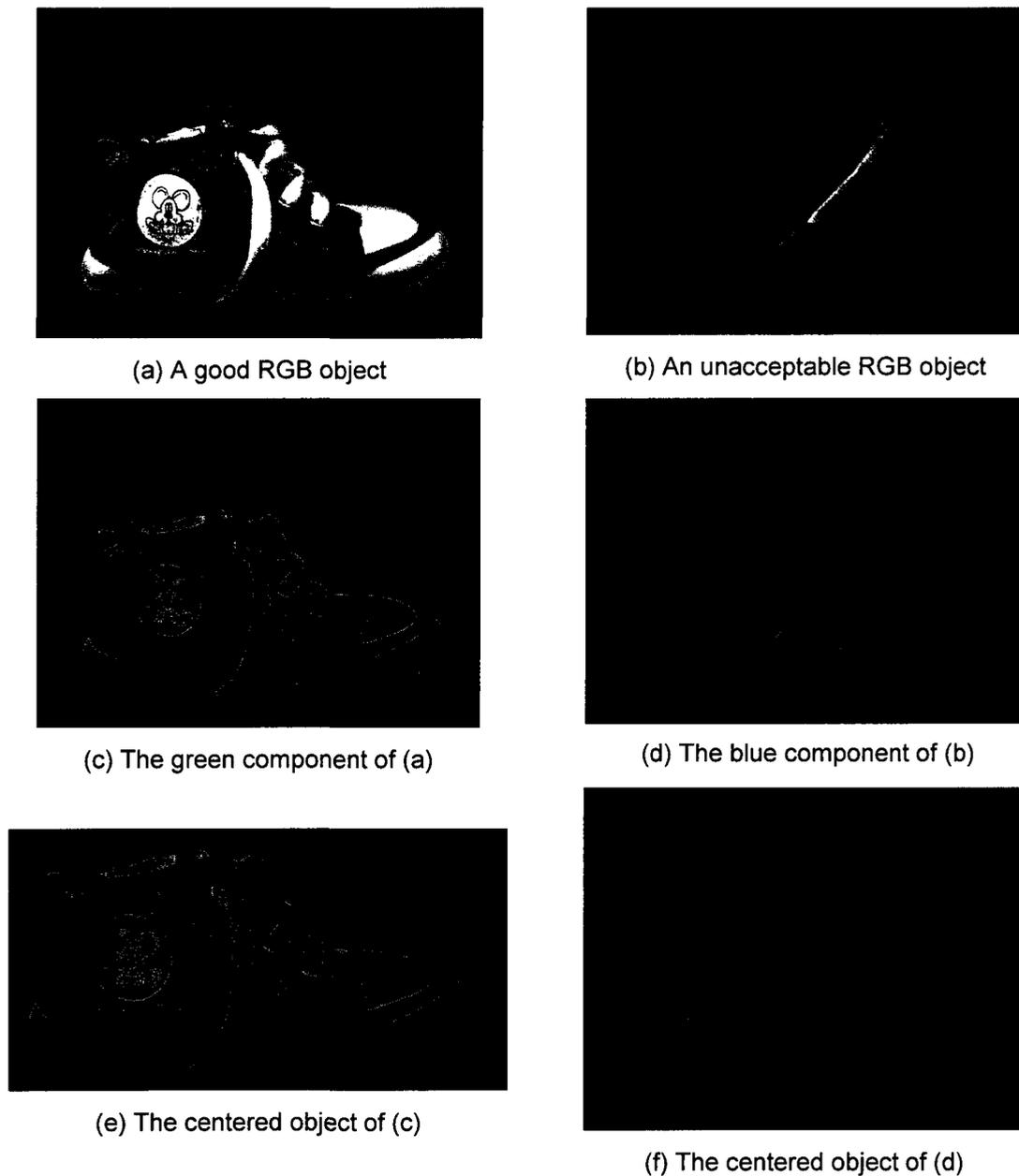


Fig. 3.6 The example of chromatic images and their centered objects on RGB channels

After computing the centroid of the three images corresponding to the R, G, B channels, we come to the conclusion that the centroid of a chromatic image can be found by calculating the center of either intensity image of the R, G, B

channels. The first method will cause a color information loss since it is transformed into grayscale. The RGB model contained all the color information, so the second method which retains more information after the decomposing and recomposing process, will be adopted for the later processes.

The log-polar mapping is applied after the centroid calculation. The mapped image is generated in the cortical plane based on the centered images in the visual plane. This log-polar mapping is accomplished using the “cvLogPolar” function in OpenCV. The advantage of log-polar was explained in detail in Chapter 1.2. For later pattern recognition use, it is necessary to process on all the three color channels. This implies processing on a batch of images at the same time. MATLAB can achieve the same desired log-polar function. However C-based OpenCV is more efficient in computation compared with MATLAB. Fig. 3.7 shows an example of log-polar mapped image in RGB channels.



Fig. 3.7 Left to right are the R, G, B channels of the log-polar mapped image of Fig. 3.6 (e)

### 3.4 Experimental Results of Color Image Object Recognition

The algorithms of pre-whitening, whitening, and pattern classification are discussed in Chapter II. The datasets we used for testing were retrieved from [27]. Forty color objects were chosen from the database. Eight images of each same object in different illuminations were chosen for the training process before the real classification test. Four images of each same object were chosen to perform the real classification test. That makes twelve images for the same object in total. The number of test images of all the 40 objects in the database is 4 times 40, that is 160 in total. Additionally, four out of the twelve images were rotated artificially through software about the log-polar transform origin by some random angle in the range  $(-10^\circ, 10^\circ)$ . The scale-invariant feature was not tested since the scale change in the spatial domain only manifested a horizontal shift in the log-polar domain. The log-polar mapping was implemented in OpenCV while the rest of the algorithm was implemented in MATLAB. The actual computations were executed in off-line and batch mode, and the speed of execution was not considered as an important parameter. In all of the experiments the number of non-zero ICA components was 256. Former research shows that reducing the number of non-zero ICA components from 256 down to 2 will not affect the recognition rate significantly [6]. Different classification engines were tested: nearest neighbor, 3-nearest neighbor [28] and two corresponding hybrid nearest neighbor classifiers, all of which are introduced in detail in Chapter II. The experimental results obtained by application of four different classification

techniques are summarized in Table 3.1. The rotated color image recognition rate is shown in Table 3.2 (only the hybrid classifiers are tested).

| Three frequency bands filters(256 non-zero ICA coefficients) | Recognition result pass/fail for specific classifier |                    |                         |                           |
|--|--|--------------------|-------------------------|---------------------------|
|  | Nearest Neighbor                                     | 3-Nearest Neighbor | Hybrid Nearest Neighbor | Hybrid 3-Nearest Neighbor |
| Red Component  | 140/20   | 128/32             | 154/6                   | 158/2                     |
| Green Component  | 142/18   | 139/21             | 158/2                   | 160/0                     |
| Blue Component   | 130/30   | 105/55             | 130/20                  | 145/15                    |
| Results without preprocessing (grayscale)                    | 147/13   | 144/16             | 158/2                   | 160/0                     |

Table 3.1 The results of non-rotated color image object recognition in RGB channels by four different classifiers

| Three frequency bands filters(256 non-zero ICA coefficients) | Recognition result pass/fail for specific classifier |                           |
|--|--|---------------------------|
|  | Hybrid Nearest Neighbor                              | Hybrid 3-Nearest Neighbor |
| Red Component  | 149/11   | 158/2                     |
| Green Component  | 156/4  | 159/1                     |
| Blue Component   | 114/46   | 142/18                    |
| Results without preprocessing (grayscale)                    | 155/5  | 160/0                     |

Table 3.2 The results of rotated color image object recognition in RGB channels by hybrid classifiers

From the results in the tables, some interesting conclusions are outlined below:

1. The hybrid method is superior to the traditional nearest neighbor classification. The reason for this might be that potential variations in the selection of salient locations lead to large distances between various feature vectors. However the hybrid method execution time is about two times more than the non-hybrid one.

The hybrid method can thus be adopted in the case when better recognition performance has priority over time to process.

2. Comparing Table 3.2 to Table 3.1, the rotated image recognition rate is slightly lower than the original image, though within an acceptable range (7%). This confirms the methodology's feasibility in rotation-invariant pattern recognition.

3. It is observed that green component works best among all the three channels, while blue channel has poorer performance. Compared with the same objects recognition performance by applying Vidacic's grayscale object recognition methodology in the last line of each table, we can see the green channel has almost as much as the grayscale images recognition rates. This implies a relative potential efficient color image pattern recognition scheme – to extract the green channel component of the color image and perform the whole methodology on it to obtain sufficient recognition rate.

4. One possible reason that leads to the undesirable recognition rate of the blue channel is that the blue channel exhibits a lower range of contrast intensity, about [0 20]. This is smaller than the green channel whose contrast range is about [0 255]. This drawback is caused by Eq. 3.4 in MATLAB which puts lower weights to blue channel. This drawback might be overcome by stretching out the histogram of the blue component to [0 255] (Fig. 3.8). In MATLAB simulation, the image is rescaled to 0-225 by the difference of the maximum and minimum values of matrix. For example, the image matrix generated by ten columns of intensity value 0, ten columns of intensity value 128 and ten columns of intensity

value 255 shows exactly the same with intensity value (2, 3, 4) matrix. The blue channel can be improved by stretching out the intensity range by Eq. 3.5.

$$c[X Y]_{\text{Blue Stretched}} = 255 * (c[X Y]_{\text{Blue}} - [\text{Min}]) / ([\text{Max}] - [\text{Min}]) \quad (3.5)$$

Here c is an arbitrary vector in RGB space, the size of the image is X \* Y.

In the blue [0 20] case, [Max] is an X \* Y matrix with all elements set to value 20 and [Min] is an X \* Y matrix with all elements equal to 0. Referring to Table 3.3, the pattern recognition results corresponding to the intensity stretched blue channel performs better than the original un-stretched one. This gives us a thought that after performing a similar stretch algorithm to the other channels we may be able to improve the recognition performance. This theory also works very well in the preprocessing filtering case in Chapter IV.

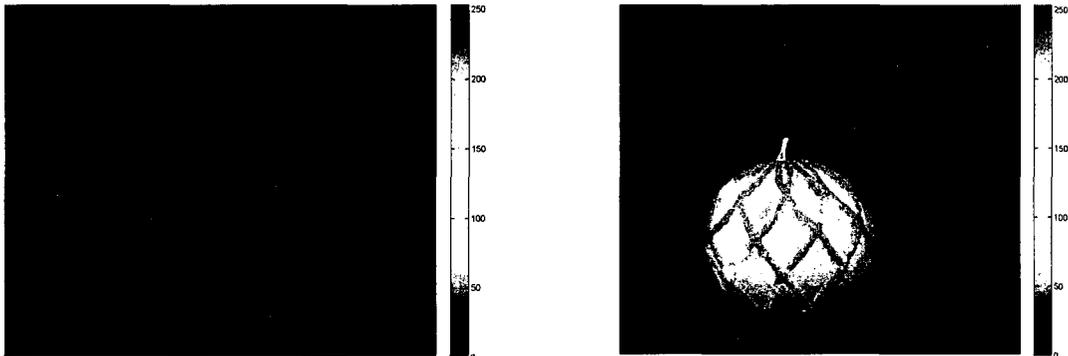


Fig. 3.8 Blue component (left) and the same blue component stretched to [0 255] (right)  
The pseudo color map 'jet' is used for visualization

| Three frequency bands filters(256 non-zero ICA coefficients) | Recognition result pass/fail for specific classifier |                           |
|--|--|---------------------------|
|  | Hybrid Nearest Neighbor                              | Hybrid 3-Nearest Neighbor |
| Blue Component   | 114/46   | 142/18                    |
| Stretched Blue Component                                     | 158/2  | 160/0                     |
| Results without preprocessing (grayscale)                    | 155/5  | 160/0                     |

Table 3.3 The recognition rate comparison of the original blue channel (non-rotated) and the stretched one

## **CHAPTR IV**

### **IMAGE PREPROCESSING IN MULTIPLE SPACIAL FREQUENCY CHANNELS**

In this chapter our focus is on another preprocessing step in the image pattern recognition system – the step of multiple spatial frequency filtering. In this methodology, prior to the log-polar mapping and feature extraction step, the input grayscale images are filtered by the pre-designed low pass filter, mid pass filter and high pass filter. In order for the reader to understand the results we will first introduce the theory of multiple spatial frequency filters design. Secondly examples of LPF, BPF and HPF filtered images are presented. The object recognition experimental results for the multiple spatial frequency filtered images are presented together with the results analysis and discussion.

#### **4. 1 Multiple Spatial Frequency Filters Applied on Grayscale Images**

The human visual system contains several different classes of spatial mechanisms or channels (five to seven spatially-tuned frequency bands) that process the retinal image in parallel [29]. Neuro-physiological results support

psychophysical evidence for the existence of multiple frequency channels in the visual system [30]. Enroth-Cugell and Robson found that the retinal ganglion cells of the cat are optimally tuned to particular ranges of spatial frequency. They also found that the more narrowly tuned cells there are, the broader range in the nonlinear area will be tuned [30]. This suggests that the linear transmission of spatial information is essential for visual processing. In other experiments, the spatial frequency was found between 1.9 and 22.4 Cycles/Deg [31]. All results are compatible with the assumption that the human visual system contains sensory channels, each selectively sensitive to different narrow bands of spatial frequencies, whose outputs can be detected independently [32]. The bandwidth determined from different psychophysical methods is about 1 octave to 2 octaves. Experiments showed [31] that 1 octave is an important general bandwidth for filtering image. Orientation bandwidths average  $\pm 15$  degrees. The filter choices and shapes of the channels in the HVS are worth further research.

Quite a bit of research had been done on the visual spatial characterization of striate cortex (V1) neurons, both in animals and humans [32, 33]. The visual channel's response profiles in both species are very similar: narrowly tuned, with a peak at about 1 Cycles/Deg and cut-off acuity of about 7 or 8 Cycles/Deg [33]. In the single-channel model (Fig. 4.1), the image is passed through a single spatial filter. Noise is added to the filter output to represent the uncertainty of the detection process. The output is then passed through a threshold device.

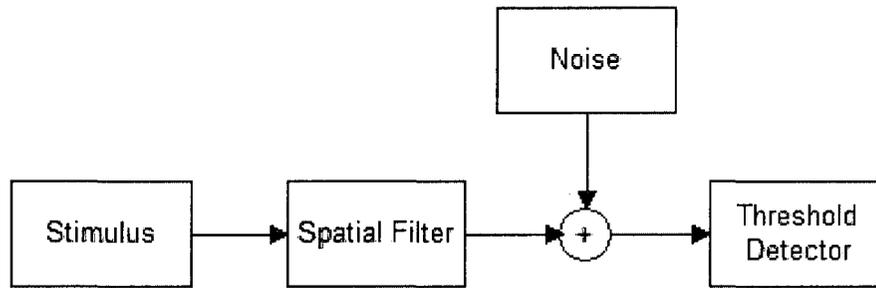


Fig. 4.1 The single channel model of grating detection

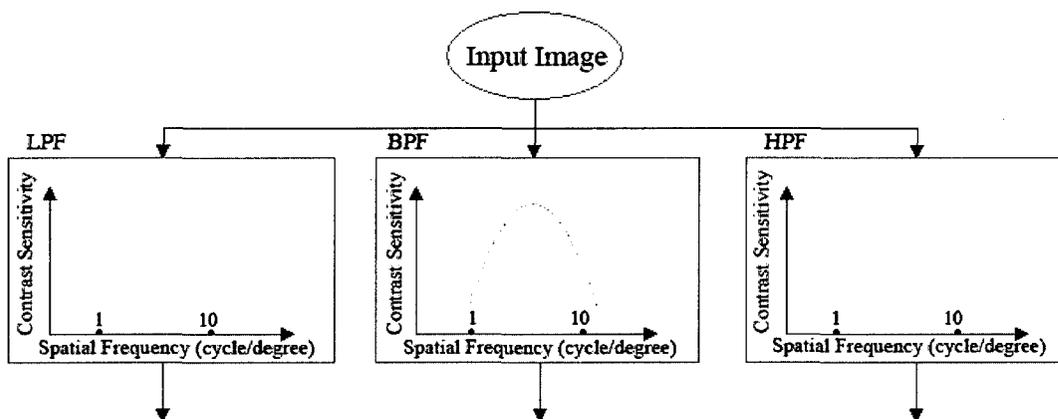


Fig. 4.2 Block diagram of a three channels filtering model

In order to determine the spatial frequency and orientation selectivity of channels in the human visual system, several techniques such as selective adapting, masking (used for measuring orientation selectivity) and threshold summation are used. In the multiple-channel model (Fig. 4.2), the input image is passed through a parallel bank of narrowband filters. The outputs of the various channels are stochastically independent. These filters can pass through most of the image frequency bands while introducing limited distortion.

In this chapter, the three spatial frequency filters were created as a preprocessing step of the whole image pattern recognition system. Three spatial frequency channels are considered in this dissertation. The three channels are

low pass, band pass and high pass in the spatial frequency domain. We can certainly design five to seven filters, whose combination can cover most of the spatial frequency range and may be more precise than the three channel model. However in order to test the idea that a multiple spatial frequency approach can yield better recognition performance it was decided to use just three channels. For now, we focus on the intensity of the image rather than the hue or saturation component, so that the examples adopted here are all grayscale images. The block diagram which outlines the multi-frequency bands preprocessing steps performed in this chapter is shown in Fig. 4.3. The later steps are the same as the color image pattern recognition diagram shown in Fig. 3.1.

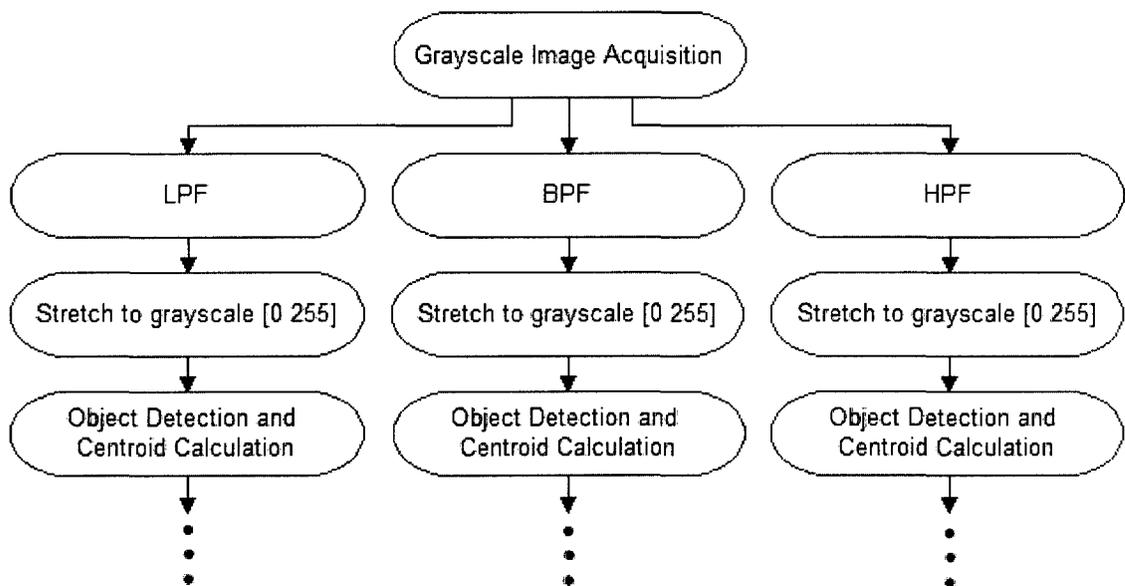


Fig. 4.3 Part of the methodology of multi-frequency channels image pattern recognition

For further research, we can combine the theory developed in Chapter III and Chapter IV. One possible scheme for implementing the multi-frequency channels color image pattern recognition scheme is shown in Fig. 4.4. The decomposition

of the color image into RGB channels is accomplished prior to the multi-frequency band processing. Different pre-whitening filters are dedicated to different categories of images. In this scheme, the pre-whitening filter for log-polar mapped images is used for the color image feature extraction. Though the pattern recognition rates of this scheme are not tested, we predict that the recognition rate may be a little worse than the single preprocessing scheme, especially for the blue channel plus the HPF branch. It is also reasonable to predict the green channel plus the LPF can give the best performs in all these branches due to prior observations regarding the green channel and its performance.

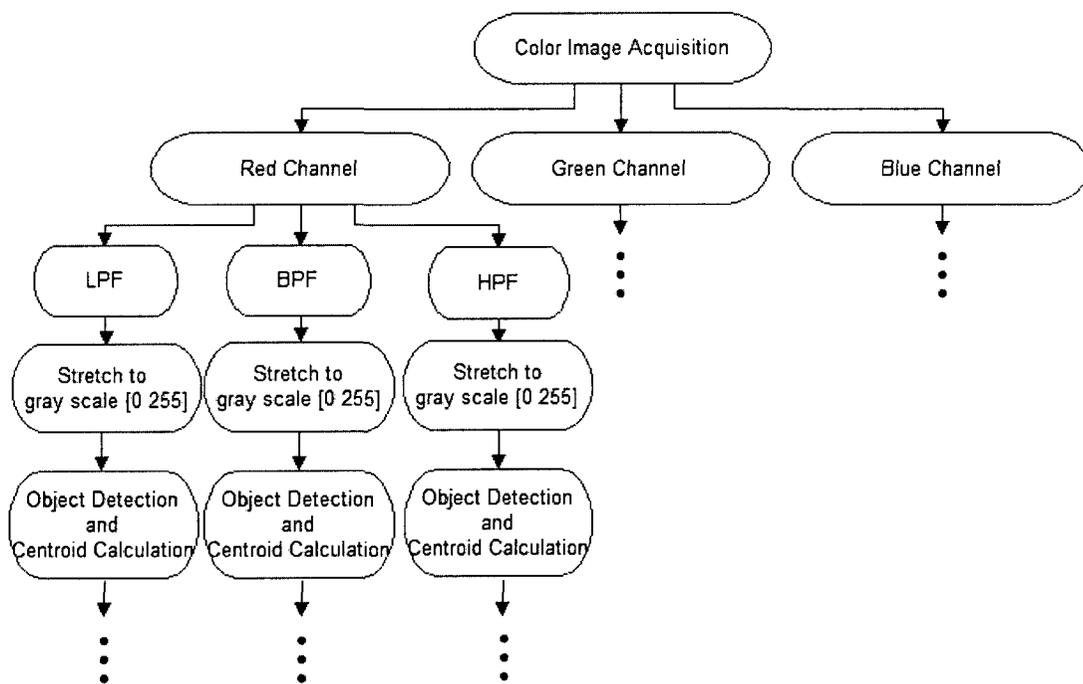


Fig. 4.4 The cascaded preprocessing scheme of multi-frequency channels color image pattern recognition

In order to obtain the three individual channels' information, we can consider two approaches [34]. The first one is to process the image through spatial filtering.

Spatial convolution is usually simplified by using small masks that attempt to capture the salient features of their frequency domain counterparts. The linear spatial filtering process consists of multiplying each pixel in the neighborhood by a corresponding coefficient (known as a mask, template or kernel) and summing the results to obtain the response at each point (x, y). Two-dimensional masks (Kernels) are made for performing either convolution (Eq. 4.1) or correlation (Eq. 4.2). This 2D mask is then applied to the image. Consider  $h(x,y)$  being a filter of size  $m \times n$  applied to an image  $f(x,y)$  (where  $m$  and  $n$  are odd integers). Thus two equations are possible as written in Eq. 4.1 and 4.2 below. Notation  $*$  stands for correlation while notation  $\bar{*}$  stands for convolution.

$$h(x, y) * f(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b h(s, t) f(x + s, y + t) \quad (4.1)$$

$$h(x, y) \bar{*} f(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b h(s, t) f(x - s, y - t) \quad (4.2)$$

Box Smoothing, also known as "sliding average", "moving average", or "boxcar smoothing" is commonly used as a first order approach to accomplish low pass filtering in the image plane. This simplest form of smoothing is the commonly referred to as a "moving average" which simply replaces each data value with the average of neighboring values. In equation form, the moving average is calculated by Eq. 4.3.

$$\bar{x}[i] = \frac{1}{2M + 1} \sum_{j=-M}^M x[i + j] \quad (4.3)$$

It can be implemented by convolving the input data with a box-shaped pulse of  $2M+1$  values all equal to  $1/(2M+1)$ . These coefficients are called smoothing kernels. In this method, the essential design is to determine the size of the mask.

For the low pass 1 filter, the original image is convolved with a rect function of size 51 by 51 pixels and height unity to produce. The band pass filter can be generated by taking the difference of two low pass filters. One as mentioned for the low pass 1 filter, the other low pass 2 filter designed is a rect function of size 51 by 51 pixels and height 0.5. The high pass filtered image can be achieved by taking the difference of the original image and the low pass 1 filtered image.

The second method of image filtering is to perform the filtering process in the frequency domain. Filtering in the frequency domain consists of modifying the Fourier transform of an image and then computing the inverse transform to obtain the processed results.

$$F(u, v) = DFT[f(x, y)] \quad (4.4)$$

$$g(x, y) = IDFT[H(u, v)F(u, v)] \quad (4.5)$$

Here  $f(x,y)$ ,  $g(x,y)$ , both  $M \times N$  arrays, are the input and output image, respectively.  $H(u,v)$  is the filter transfer function. The essential step of the frequency filter design is the choice of filter type and the choice of parameters.

A better approach for filtering is to use a smooth function as opposed to a rectangular function which has a step discontinuity. Gaussian filters are smooth filters which adopt values based on the Gaussian function. Both the forward and inverse Fourier transforms of a Gaussian function are Gaussian functions.

Consider the 1-D frequency domain Gaussian filter, where  $\sigma$  is the standard deviation of the Gaussian curve, and can be considered the cut-off frequency:

$$H(u) = Ae^{-u^2/2\sigma^2} \quad (4.6)$$

Extending the 1-D case to 2-D we obtain the low pass filter transfer function:

$$H_{LP}(\mu, \nu) = Ae^{-(\mu^2 + \nu^2)/2\sigma^2} \quad (4.7)$$

More complex filters can be constructed using the basic Gaussian function of Eq. 4.5. A high pass filter can be obtained from a given low pass filter using the following equation:

$$H_{HP}(\mu, \nu) = 1 - H_{LP}(\mu, \nu) \quad (4.8)$$

A band pass (mid pass) filter can be constructed as the difference of Gaussians as shown in Eq. 4.9. The LPF, BPF and HPF share the same center.

$$H_{BP}(\mu, \nu) = H_{LP1}(\mu, \nu) - H_{LP2}(\mu, \nu) = Ae^{-(\mu^2 + \nu^2)/2\sigma_1^2} - Be^{-(\mu^2 + \nu^2)/2\sigma_2^2} \quad (4.9)$$

Filters based on Gaussian functions are of particular interest because of their following properties:

1. The 2-D Gaussian function has rotational symmetry. This makes the prior judgment bias unlikely to happen in the subsequent edge detection process.
2. A Gaussian function is a monochrome function. This property makes the smoothing process remain undistorted, since the weight is monotonically decreasing according to the increasing of the distance from the center point.
3. The Fourier transform of a Gaussian function is again Gaussian. This LP nature ensures the filtered image not contaminated by the high frequency noise components, while retaining most of the signal components.
4. The  $\sigma$  characterizes the bandwidth (the extent of smoothing) of the Gaussian filter. The frequency band increases with an increase in  $\sigma$ . A smaller  $\sigma$  makes the sum of the LPF, BPF and HPF filters more similar to the modulation transfer function in the HVS. If we increase  $\sigma$ , the reconstructed image is more similar to the input image but more overlap between filters will

be apparent leading to less relevant information being preserved in the individual channels. Thus careful attention must be given to the choice of  $\sigma$ .

5. Because of the separability of the Gaussian function, a relatively larger size of Gaussian filters can be practically implemented. The computation of 2D Gaussian filters increase linearly with the width of the smoothing mask instead of increasing quadratically.

By reason of the above properties, Gaussian filters were widely used in image processing since 1970s. Similar to Butterworth filter, Gaussian filters avoid the “ringing” effect that occurs by using an ideal low pass filter (i.e., Box Smooth filter, ILPF). Gaussian low pass (smooth) filter is the same functional form in both the image domain and spatial frequency domain, making it very desirable for implementation.

Using the filters in the spatial domain requires less computation (spatial convolution) and can be practically implemented using real time hardware based on Finite Impulse Response (FIR) filters.

Filtering in the frequency domain requires the computation of the DFT and IDFT. Working in the frequency domain we can take advantage of the correspondence between frequency content and image appearance directly. In this domain some tasks could be difficult or impossible to formulate directly in the spatial domain.

Many achievable literatures on this topic employ the spatial filtering for multi-channel processing. In this thesis, we perform filtering in the frequency domain for better flexibility and transportability. It also avoids the mask generating process in spatial filtering. In MATLAB simulation, an inverse filter can be

applied to the frequency domain when we want to restore the input image. The frequency filters can also be implemented in a real-time optical system, as described in the Appendix A.

The low pass filter has a blurring effect which can provide for better analysis of an image's low-frequency details by removing high frequency edges and patterns. The high pass filter can be applied to assist in the detection of object edges or to sharpen an image. The high frequency details are accentuated in a filtered image and render the edge detail in the original image more visible.

Based on the above discussion, the three spatial frequency filters of low pass, band pass and high pass possess independent usefulness as an intra-class, inter-class, and membership identification classification scheme. Low pass spatial frequency information determines the "faint curvatures" for intra-class classification. Band pass (mid pass) spatial frequency information determines the membership identification through the use of "bright features" [4]. Band pass filtering can be used to enhance edges (suppressing low frequencies) while reducing the noise at the same time (attenuating high frequencies). High pass spatial frequency information determines the inter-class classification through the use of "faint edge" features.

## 4. 2 Experimental Results of Multi-Frequency Image Filtering

The implementation of all filters was done in the spatial frequency domain. Fig. 4.5 depicts the filters realized: (a) the Gaussian LPF 1's parameters are  $\sigma = 0.25$ , height unity. In (b) the Gaussian LPF 2's parameters are  $\sigma = 0.1$ , height 0.5. (c) shows a Difference of Gaussian filter. By taking the difference of (a)-(b), we can obtain a band pass filter. The product of a high pass filter and a low pass filter in the frequency domain can also achieve the band pass filter, in which case the intensity value of the BPF starts from zero. A Difference of Gaussian (DOG) filter for the traditional BPF is preferred to start from the intensity value of 0.5 (Or we can call it pseudo band pass filter). The Gaussian high pass filter in (d) is generated by 1-(b). (e) shows the profiles of all filters in (a)-(d). From a biological point of view, it can be shown as Cycle/Degree VS. Contrast Sensitivity instead of Cycle/Unit Distance VS. Intensity. (f) shows another filter design with different parameters ( $\sigma_1 = 0.5$ ,  $\sigma_2 = 0.2$ ). In this case the curve of the sum of all three filters is more flat, which means in this design better reconstructed images can be generated. However this time each individual channel contains less information in its particular LP, BP or HP frequency bands. This is a trade-off in the filters design.

The normalized root mean squared error (NRMSE) parameter for an  $N \times N$  image is calculated as:

$$\text{NRMSE} = \sqrt{\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \frac{(x(i, j) - y(i, j))^2}{y(i, j)^2}} \quad (4.10)$$

In a real object example, design (e)'s NRMSE = 0.0487 is while design (f)'s NRMSE = 0.0077. The parameter in (f) is the one that performs best in the range of  $\sigma_1 = [0.1 \ 0.5]$ ,  $\sigma_2 = [0.01 \ \sigma_1]$ .

Fig. 4.6 shows the filtering process on an artificial image. In this example we generated the pixel matrix in the spatial domain by using the following formula:

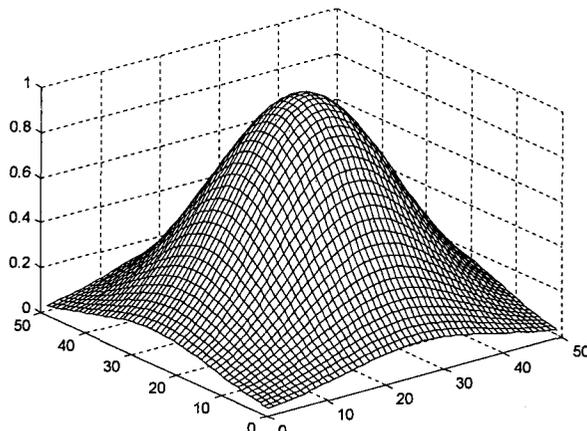
$$Z = \cos(5 * X + 10 * Y) + \cos(10 * X + 10 * Y) + \cos(5 * X + 5 * Y) + 3 \quad (4.11)$$

X and Y make a 41 x 41 matrix. Fig. 4.6 (a) shows the surface of Z. The range of Z is [0, 6]. Fig. 4.6 (b) shows the cross section intensity of Z. Fig. 4.6 (c) is the grayscale intensity that is displayed by the "imshow" function in the MATLAB image processing toolbox. Fig. 4.6 (d)-(f) are the filtered image (in the spatial domain) by applying the LPF, BPF, HPF of Fig. 4.5 to Fig. 4.6 (c). Fig. 4.6 (d) shows that most of the information contained in the original image is contained in the designed LPF frequency band. The gray border is generated by the padding of zeros before the convolution operation, which relived the wraparound interference error. Fig. 4.6 (g) is the sum of Fig. 4.6 (d)-(f), i.e., the reconstructed original image. Fig. 4.6 (h) shows the NRMSE between the original image and the reconstructed image, NRMSE = 0.1318. The mean of the difference = 0.1194, both are relatively small and within the tolerance range (0.06%).

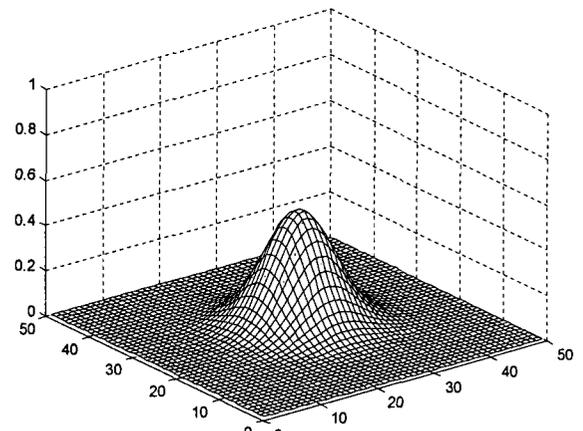
Fig. 4.7 (a) is the spectrum of Fig. 4.6 (a), the original image in the Fourier domain. Fig. 4.7 (b) is the surface of (a), which shows the three delta pulses on both sides of axis and the DC bias. Fig. 4.7 (c)-(h) show the frequency spectrums of the original image filtered by LPF, BPF, HPF and their corresponding filters, respectively. In the low pass case, the high frequency

components are mostly removed, leaving part of the  $5 * Y+5 * Y$  components visible. In the band pass case, the components out of the filter ring are more suppressed compared with the ones inside the pass band. DC bias is blocked out in both the band pass and high pass case. Other visible grids are caused by the non-ideal filters (composed by the superposition of the Sinc functions' side lobes). This image is assumed to be periodically extended in the image space. Thus there are strong edge effects between the neighbors of such a periodic array [35]. The frequency spectrum of this synthetic image can illustrate the filter's properties better. The synthetic image we generated has higher frequency components than is found in the real objects used in the study.

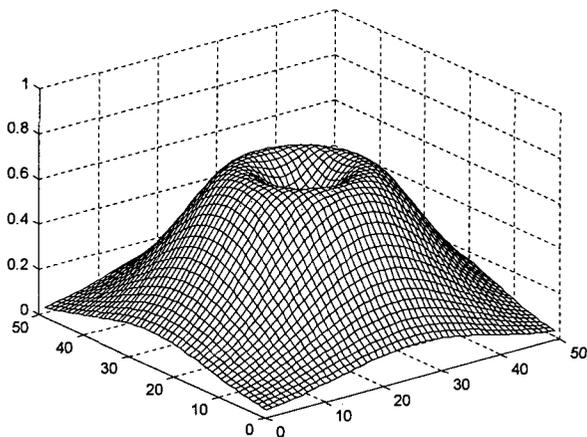
Fig. 4.8 shows a practical object image filtering achieved by MATLAB simulation. Fig. 4.8 (a) is the original image; (b) to (d) is the LPF, BPF and HPF filtered image, respectively. In the LPF filtered image (b), the edge is blurred and the object is smoother, which makes the filtered image more adaptive for intra-class classification. In the BPF filtered image (c), the brightness of the image is enhanced so that it can be used for membership identification. In the HPF filtered image (d), the high frequency components demonstrate an emphasis on the edge components, enhancing their contrast and details. This property is fit for the inter-class classification. All the practical results verified our theoretical anticipation of the filter designs. Since the grayscale range of the three filtered images are within the range [0 60], they are not well suited for visualization and subsequent processing. Similar to Eq. 3.5, the contrast stretching function is applied on all channels to stretch the image grayscale range to [0 255].



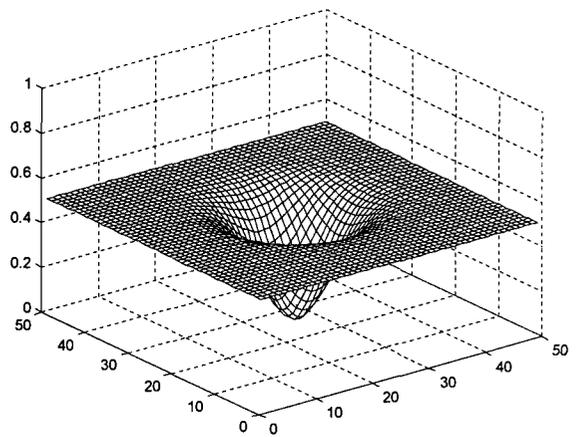
(a) LPF 1



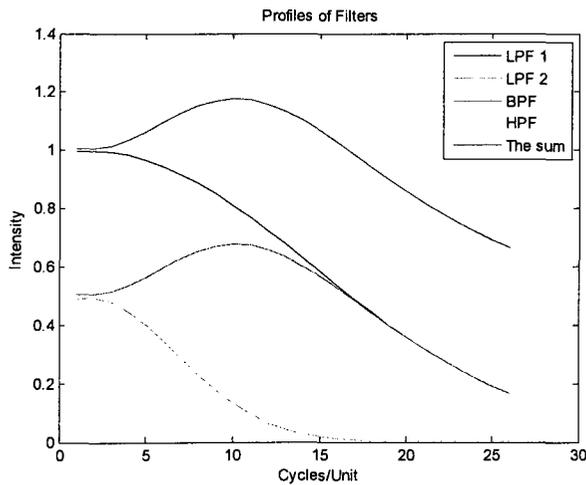
(b) LPF2



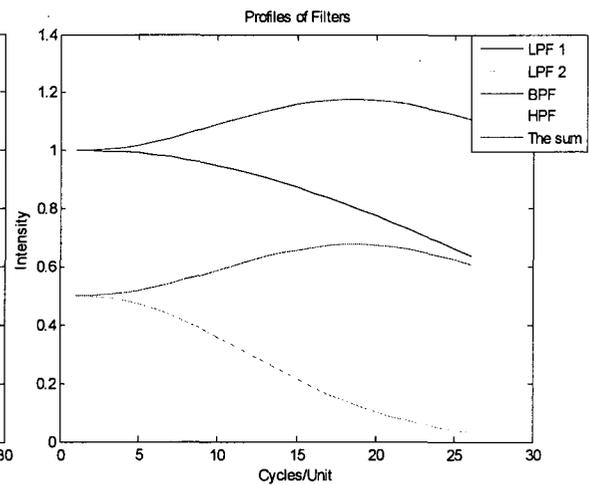
(c) BPF



(d) HPF



(e) Profiles of the above filters



(f) Another filter designs ( $\sigma_1 = 0.5$ ,  $\sigma_2 = 0.2$ )

Fig. 4.5 (a) Gaussian LPF1,  $\sigma_1 = 0.25$ , height unity; (b) Gaussian LPF2,  $\sigma_2 = 0.1$ , height 0.5; (c) Gaussian BPF, DOG filter (a)-(b); (d) Gaussian HPF 1-(b), (e) Profiles the above filters; (f) Another filter designs

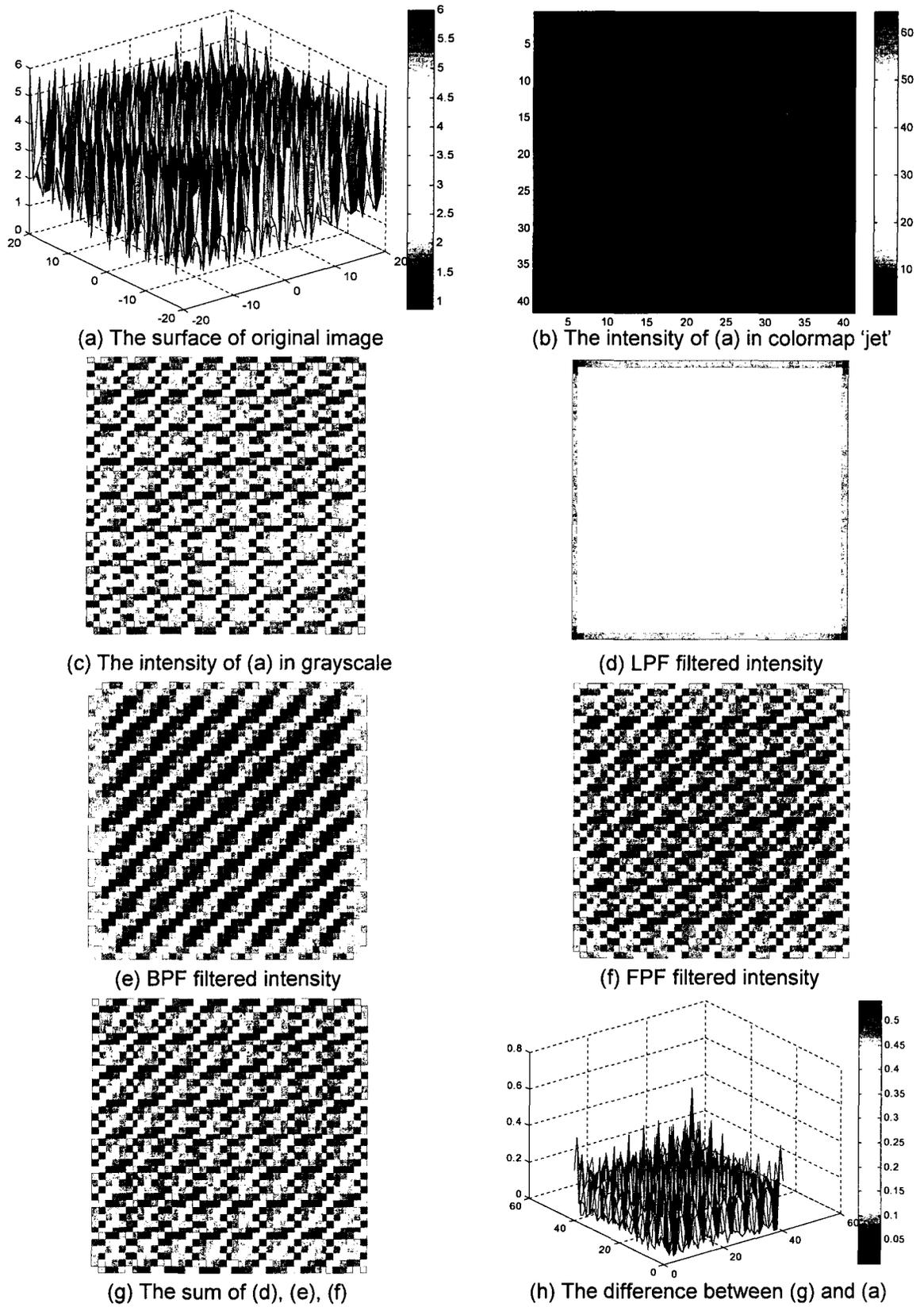


Fig. 4.6 Filtering process on a synthetic image of cosine functions in both x and y directions

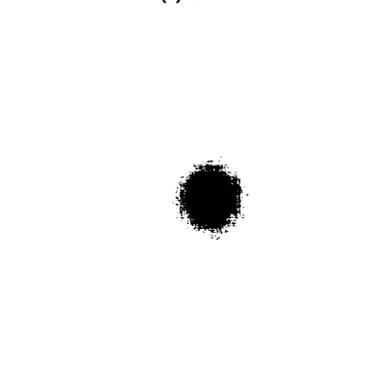
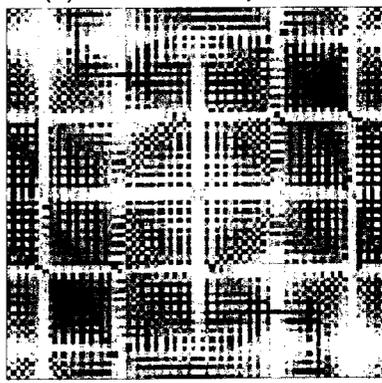
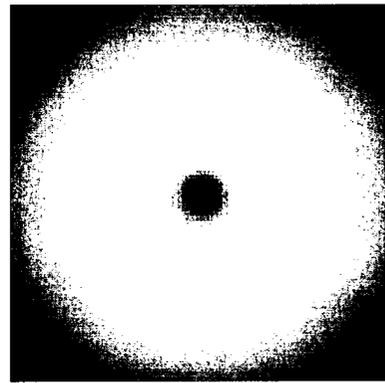
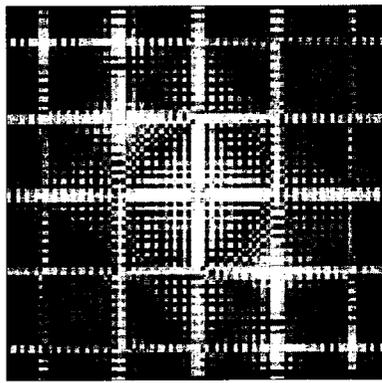
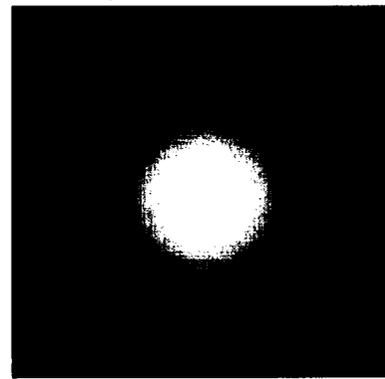
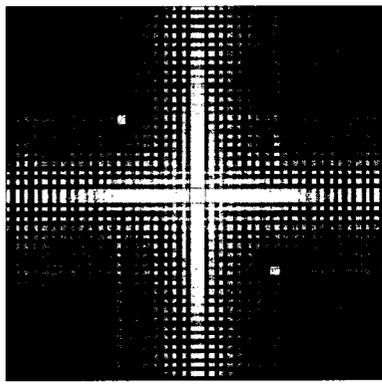
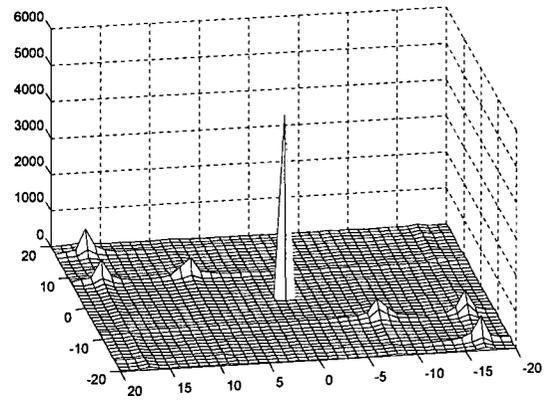
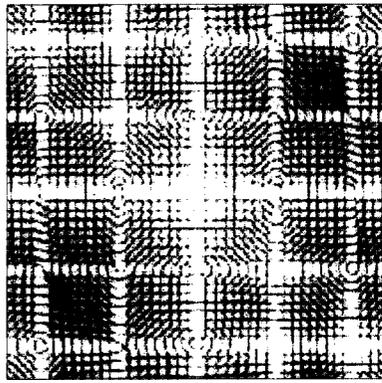


Fig. 4.7 The frequency spectrum of the filtered artificial images and their corresponding filters

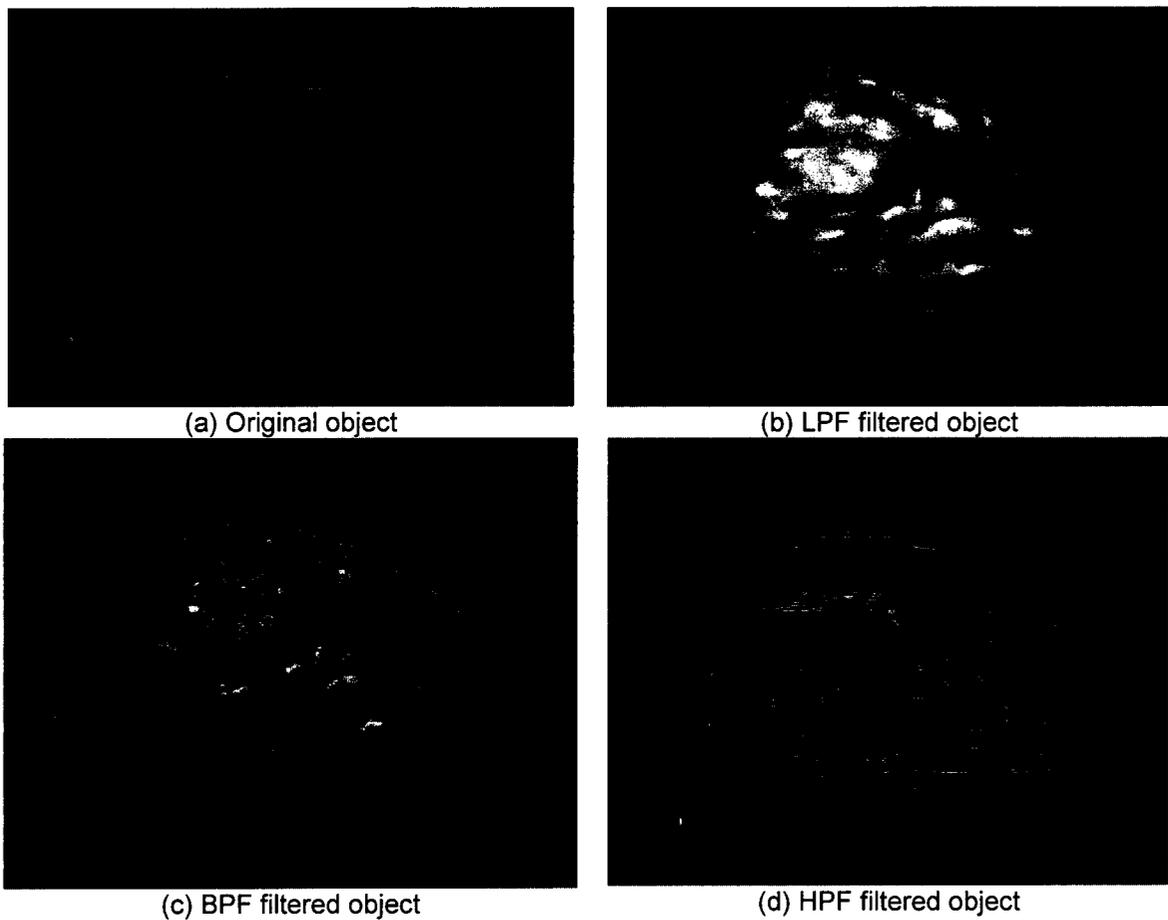


Fig. 4.8 (a) The original object image (centered), (b) Filtered image of Gaussian LPF, (c) Filtered image of Gaussian BPF, (d) Filtered image of Gaussian HPF

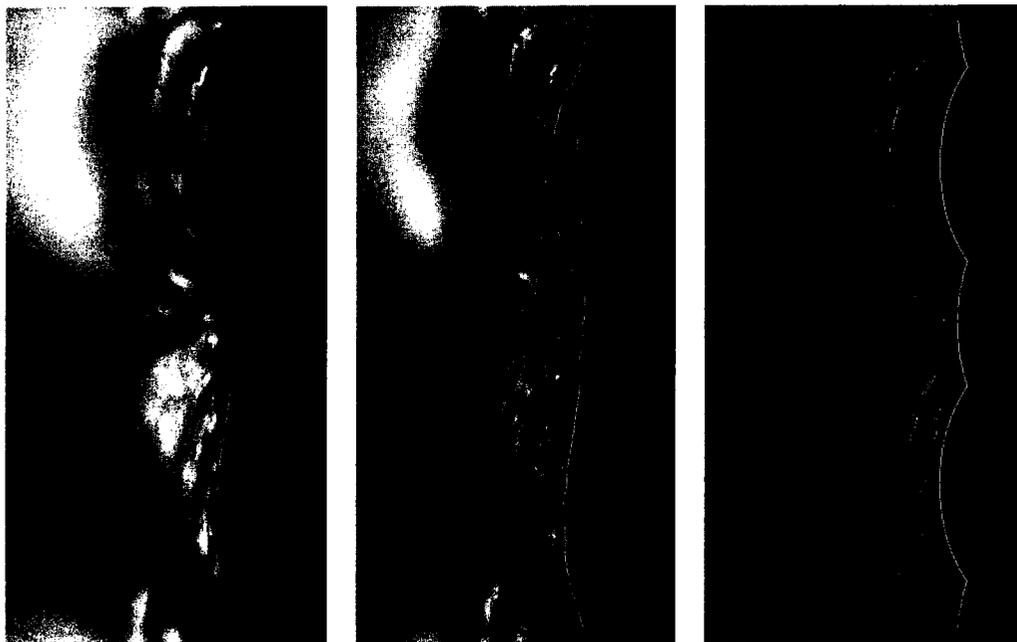


Fig. 4.9 The log-polar mapped example of Fig. 4.8 (b), (c), (d)

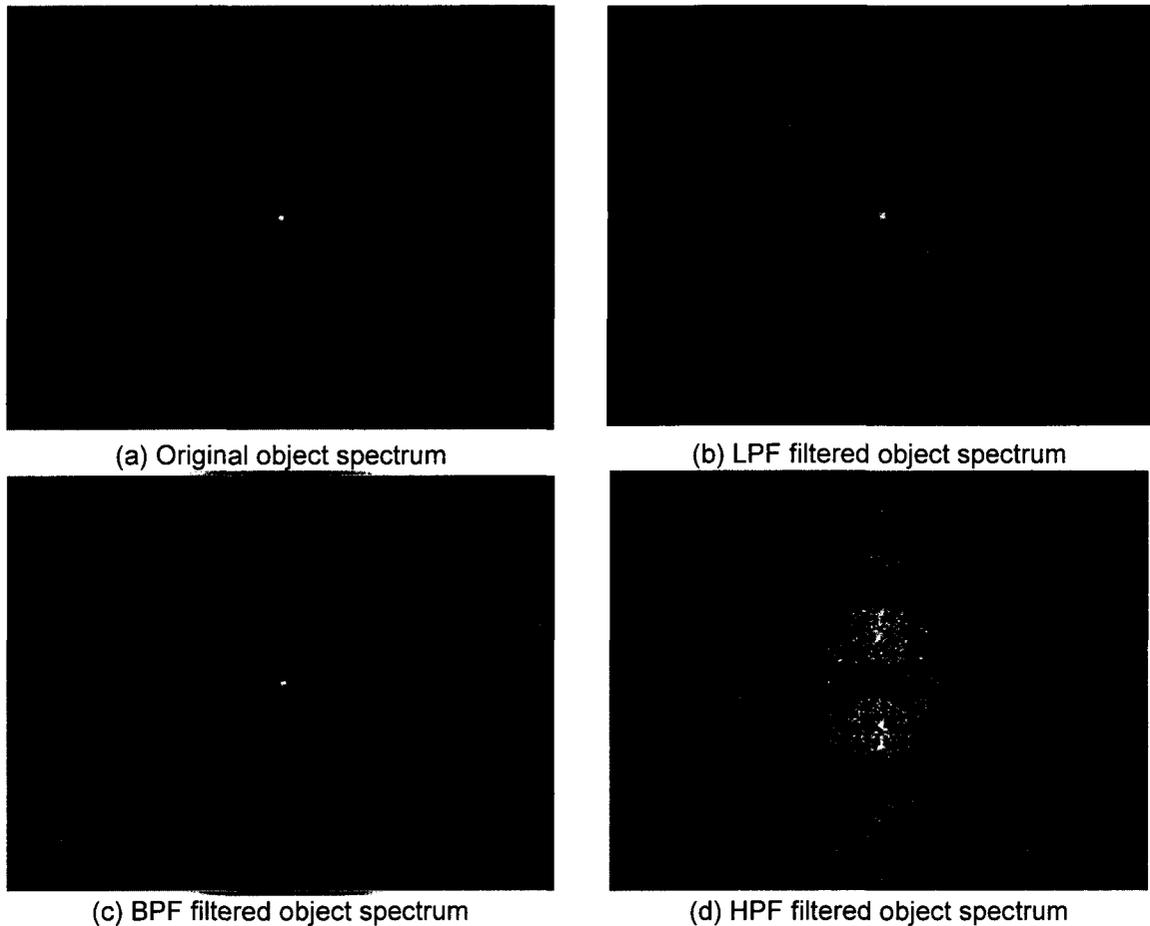


Fig. 4.10 (a) The spectrum of the original object, (b) Gaussian LPF filtered spectrum, (c) Gaussian BPF filtered spectrum, (d) Gaussian HPF filtered spectrum

Fig. 4.9 contains the three filtered images after mapped to the log-polar plane, corresponding to Fig. 4.8. These filtered images are the input images for the later object recognition processes. From this log-polar mapped image sixteen local pseudo salient patches are to be extracted.

Fig. 4.10 shows the frequency spectrum of the three filtered images corresponding to Fig. 4.8. The real object image contains numerous low frequency components as opposed to high frequency components. The properties of these frequency spectrums are consistent with Fig. 4.7. Here we

have some interesting discussions about several special concerns of the above experimental results.

1. The experimental results show that this multi-frequency channel scheme works for object image, texture image, as well as natural scene image. Typically, these three categories of images are the most common cases in image recognition systems.

2. Otsu's centering algorithm [35] showed its limitations on the HPF filtered image, even if it is stretched to a grayscale [0 255]. False edge replication will interfere with the centering and rotation introducing processes. In some cases, the centroid of three channels for the same image may not be consistent. However, within the same channel, the center is consistent so that the later pattern recognition process is possible. Due to these limitations, better centroid calculating algorithms can be investigated in the future work.

3. A PNG format image does not remain undistorted following filtering by low pass or high pass filters. The BMP format image gives results of the best fidelity but it requires the most hard disk storage to restore. The compressed image of a JPG format is a compromise of less storage requirement and better image fidelity. Furthermore, the numbers in the filtered image datasets matrix are floating format, which need to be converted to integer format when writing or illustrating the images.

### 4. 3 The Feature Extraction Exploration of Different Frequency Channels

The algorithms of pre-whitening, whitening, and object recognition are the same as described in Chapter II. The datasets we used for testing and the rotated datasets are also the same with Chapter III. The difference of the two systems is that after the LPF, BPF or HPF, all the resulted images' grayscale are stretched to [0 255]. The pattern recognition results are presented in Table 4.1. In all the four classifiers, the pattern recognition results of LPF filtered images and BPF filtered images are better than the HPF filtered images. The better two filtered images have comparable recognition performance compared with the pure grayscale image recognition performance. These results show us the potential usability of the multiple frequency bands preprocessing approach applied in the grayscale image pattern classification system. If the 2/3 voting scheme is applied in the multiple spatial channels (or only two applicable channels are considered in this case), we predict that the pattern recognition performance can be improved in some extent. Both the LPF and BPF filtered images retain most of the frequency information in the image. It is observed that most of the information in the image is in the designed LPF and BPF filtered images. The HPF filtered results are not as desirable to the less information being present. It is observed that Vidacic's methodology doesn't work well on the HPF filtered image. One reason is that most frequency components in a real object image are within the designed range of the LPF and BPF filters. Even though the edge information is enhanced, the best recognition rate we can get is still about 75

percent in this case. Another possible reason is that Vidacic's methodology might not work well on the edge information that extracted by the HPF.

| Three frequency bands filters(256 non-zero ICA coefficients) | Recognition result pass/fail for specific classifier |                    |                         |                           |
|--|--|--------------------|-------------------------|---------------------------|
|  | Nearest Neighbor                                     | 3-Nearest Neighbor | Hybrid Nearest Neighbor | Hybrid 3-Nearest Neighbor |
| LPF filtered image   | 149/11   | 144/16             | 159/1                   | 160/0                     |
| BPF filtered image   | 150/10   | 145/15             | 157/3                   | 160/0                     |
| HPF filtered image   | 117/43   | 104/56             | 67/93                   | 105/55                    |
| Results without preprocessing                                | 147/13   | 144/16             | 158/2                   | 160/0                     |

Table 4.1 The results of non-rotated grayscale image object recognition in multi-frequency bands channels by four different classifiers

| Three frequency bands filters(256 non-zero ICA coefficients) | Recognition result pass/fail for specific classifier |                           |
|--|--|---------------------------|
|  | Hybrid Nearest Neighbor                              | Hybrid 3-Nearest Neighbor |
| LPF filtered image   | 157/3  | 160/0                     |
| BPF filtered image   | 155/5  | 160/0                     |
| HPF filtered image   | --   | --                        |
| Results without preprocessing                                | 155/5  | 160/0                     |

Table 4.2 The results of rotated grayscale image object recognition in multi-frequency bands channels by hybrid classifiers

The pattern recognition rates for the LPF and BPF filtered rotated images are shown in Table 4.2. HPF filters for rotated images are not tested since in the non-rotated case they did not give desirable results. Similar with Chapter III, for hybrid nearest neighbor model, the recognition results of rotated objects are a little worse than the recognition results of non-rotated objects. The hybrid 3-

nearest neighbor model performs as well as the non-rotated image. These results confirm that the multiple spatial filtering preprocessing can be very useful in scale and rotation invariant image pattern recognition.

One possible future research direction is to extend our two dimensional color image / grayscale image to three dimensions. Only the illuminations of the eight training images of the same object are different from each other. For the 3D case, all images viewed in different angles ( $0^\circ - 360^\circ$ ) of an object can be adopted as training datasets. Theoretically, the more training images we have the better recognition rate we can achieve. 3D image pattern recognition may be utilized in the machine / robot vision system to achieve more feature extraction information at a faster rate.

The main content of this chapter is introducing a preprocessing stage by applying the low pass, mid pass and high pass filters to the input image in parallel. This is done for the purpose of improving the object recognition performance. Simulations show that the LPF and BPF filtered images can achieve substantial recognition results, while the recognition performances of the HPF filtered images are not so desirable. For further study, 2/3 voting scheme can be applied on these decomposed images in multiple channels in order to improve the recognition rate. Better feature extraction algorithm can be studied to apply on the HPF filtered images.

## CHAPTER V

### CONCLUSION AND FUTURE DIRECTIONS

In this thesis we proposed two preprocessing stages with the purpose of improving Vidacic's LSI based feature extraction methodology for rotation and scale tolerant pattern recognition analysis. The input color image is first decomposed into individual red, green and blue channels before the subsequent stages of recognition. The simulation results show that the recognition performance in the green channel can be as good as that of merely using a grayscale representation of the image. The object recognition performances of the three channels are coarsely related to the proportion to the weighing coefficients in the RGB to gray transfer function in MATLAB simulation. In MATLAB the green component has the most contrast while the blue component has the lowest contrast. However, after we map the blue channel's intensity values to [0 255] by applying a stretch function, the recognition performance is improved. This means the intensity values in each individual channel has enough information for good feature extraction as long as the intensity range is from 0 to 255.

Another preprocessing we presented is to filter the input grayscale image through multiple spatial frequency filters: LPF, BPF and HPF. The reason we investigate feature extraction after filtering in multiple spatial frequency bands is that such frequency channels or mechanisms are existing in biological vision systems. Furthermore, it is clear that different features occur in different spatial frequency bands. Thus, by filtering in such a fashion it is possible to extract features from each spatial frequency band in an optimal way. The design of the three filters is based on the frequency range that can be distinguished in the HVS.

This is another preprocessing filtering process that should show promise for the purpose of improving the object recognition performance in the proposed image pattern recognition system. It included the pass bands and the cut-off frequencies designs of low pass, band pass and high pass filters. It is shown that the pattern recognition rates of LPF and BPF filtered images of different classifiers are all within the tolerant range of ten percent. It is our prediction that by applying these combined two preprocessing steps on the same datasets the recognition performance will be improved compared with the grayscale images' pattern recognition performance presented by Vidacic. It is noted that the HPF filtered images have the lowest recognition pass rates. We believe it is caused by most of the frequency information in a testing object image is contained in low pass and mid pass range and Vidacic's methodology's limitation on the edge information.

After the proposed preprocessing, the centroids of the input images are calculated and log-polar mapping is adopted to map the image from the visual

plane to the cortical plane. This allow for the recognition system to work for rotation and scale changes of the input image. This in turn allow for invariant feature extraction. The log-polar mapped image is then pre-whitened by the biologically inspired LSI model (uniform model) leading to the edge enhancement. This step produces an output image that is more orthogonal to others thus providing for better recognition rates when using a neural based approach. The whitening process is then implemented by the Plumbley algorithm for the later ICA analysis. Several pattern classifiers that follow Vidacic's work are tested in the later pattern recognition stages. The Hybrid 3-Nearest Neighbor classifier is shown to have the best performance though it is the most time consuming model in terms of computation.

As a guide for further work in this area we offer the following suggestions for investigation:

1. For objects where intensity of any individual channel has very low dynamic range, the centering process may cut some parts of the object off. The same loss will happen when there are dark shapes inside of a complex object. The limitation of color image segmentation need to be further investigated in order to avoid the pre-selection of the image objects.
2. The color decomposition and different spatial filters can be applied at different stages in the pattern recognition system. Different methodology schemes with different emphasis can be further studied.
3. The spatial frequency filter is now in the scale of Cycle/Unit Distance VS. Intensity, however the more biological flavored "Cycle/Degree VS. Contrast

Sensitivity” scale can be further investigated to design a system that has image distance parameters involved (i.e., the Z axis distance from the sensor site to the object). Furthermore, different types and shapes of spatial frequency filters other than Gaussian are worth further study.

4. Instead of extracting the pseudo salience patches by applying the log-Gabor filter, different filters can be studied to realize different feature extraction aiming at LPF/BPF/HPF filtered image for inter-class, intra-class and membership identification. Multiple feature extraction filters might work better than only one filter for all frequency bands.
5. One of the interesting areas for future work is to determine which pre-whitening filters to deploy at a particular time. It would be related to identifying the optimal set of characteristics of the input image that would classify the input into the proper signal category. One possible solution would most likely investigate characteristics of the input data covariance matrix.
6. The existence of a highly sparse information code provides a very promising basis for future improvements of the presented feature extraction model.
7. The movement of the foveation point is worth further research. The main feature of a fovea-based system is a spatial, hierarchically ordered structure referred to central zone. In this system attention concentrated on small areas of visual scene, can be solved by focusing and zooming procedures. The retina has to scan the image to acquire the information we want. However, in order to move the retina, the visual system has to recognize the feature of interest and their spatial coordinates [8].

8. After decomposing a color image into individual channels (for example RGB mode), the 2/3 voting scheme applying on the three channels processing concurrently will make the system perform better. Because the time is limited for this thesis study, the pattern recognition performance of 2/3 voting scheme can be simulated in the future work. The theoretical proof of the statement that in the filtering point of view, more channels with random noises lead to smaller recognition failure rate is introduced in Appendix B.

Visual recognition of patterns is a basic capability of most advanced species in nature. A relatively large percentage of the human brain is devoted to visual processing with a substantial portion of this area used for pattern recognition. Vision provides a variety of cues about the environment (motion, color, shape, etc.) with a single sensor. It enables tasks such as objects and targets recognition, navigation, grasping and manipulation [68]. Therefore, we consider visual pattern recognition to be an important, cost-efficient primitive for computer vision and digital image processing. This thesis has given rise to a deformation invariant pattern recognition methodology of color image based on extracting a set of characteristic features from a log-polar mapped image. It is our hope that this methodology of rotation and scale tolerant image pattern recognition can be used in a practical application in the field of robotics, automation and many other branches of artificial intelligent in the future.

## LIST OF REFERENCES

- [1] Ben Mauk, LiveScience.com, August 16, 2007
- [2] ARTIFICIAL INTELLIGENCE TOPICS -The AAI's virtual library of informational and pedagogical resources for students, teachers, journalists, and everyone who would like to learn about AI – Pattern Recognition, <http://www.aaai.org/aitopics/pmwiki/pmwiki.php/AITopics/PatternRecognition>
- [3] G. Sandini, and V. Tagliasco “*An anthropomorphic retina-like structure for scene analysis*”, Computer, Graphics and Image Processing, Vol. 14, 1980
- [4] R. A. Messner, “*Smart visual sensors for real-time image processing and pattern recognition based upon human visual system characteristics*”, Dissertation, Clarkson University, 1984
- [5] M. Li “*The lateral subtractive inhibition algorithm: simulation and application in image processing*”, Master thesis, Department of Electrical Engineering, University of New Hampshire, 1995
- [6] D. Vidacic, “*Biologically Inspired Feature Extraction for Rotation and Scale Tolerant Pattern Analysis*”, Ph.D. Proposal, Department of Electrical Engineering, University of New Hampshire, 2009
- [7] E. Schwartz “*Anatomical and Physiological Correlates of Visual Computation from Striate to Infero-Temporal Cortex*”, IEEE Trans.On Systems, Man, and Cybernetics, Vol. SMC-14, No. 2, March/April 1984
- [8] H. Araujo and J. M. Dias “*An Introduction to the Log-Polar Mapping*”, Cybernetic Vision, 1996
- [9] Gamba. P, Lombardi. L, Porta. M “*Log-map analysis*”, Parallel Computing, Vol. 12, 2008
- [10] G. Sandini, and V. Tagliasco “*An anthropomorphic retina-like structure for scene analysis*”, Computer, Graphics and Image Processing, Vol. 14, 1980
- [11] T. Baron, M. D. Levine, Y. Yeshurum, “*Exploring with a foveated robot eye system*”, Proceedings of the 12<sup>th</sup> International Conference on Pattern Recognition, vol. d, 1994
- [12] F. W. Weymouth, “*Visual sensory units and minimum angle of resolution*”, Amer. J. Ophthalmol Vol. 46, 1958

- [13] S. M. Anstis, "A chart demonstrating variations in acuity with retinal position", Vision Res. Vol. 14, 1974
- [14] C. F. R. Weiman and G. Chaikin, "Logarithmic spiral grids for image processing and display", Computer Graphics and Image Processing, November 1979.
- [15] R. Messner and H. Szu, "An image processing architecture for real time generation of scale and rotationinvariant patterns", Computer Vision, Graphics, and Image Processing, Vol. 31, July 1985.
- [16] B. K. Vadapally, Z. Rahman, "Image Registration using Conformal Log Polar Mapping", Proc. Of SPIE, Vol. 7341, 2009
- [17] C. Blackmore, E. A. Tobin, "Lateral inhibition between orientation detectors in the cat's visual cortex", Expl. Brain Res., Vol. 15 1972
- [18] J. A. Movshon, I. D. Thompson, D. J. Tolhurst, "Receptive field organization of complex cells in the cat's striate cortex", J. Physiology, Vol. 283, 1978
- [19] G. A. Baxes, Digital Image Processing Principles and Applications, pp. 16-20, 1994
- [20] The web page for Retinal Illusions, North Kentucky University, <http://www.Nku.Edu/~issues/illusions/Retina.Htm>
- [21] G. G. Furman, "Comparison of models for Subtractive and shunting Lateral-inhibition in receptor-neuron fields", Kybernetic, Vol. 2, 1965
- [22] M. G. Luniewicz and R. A. Messner, "Effects of lateral subtractive inhibition within the context of a polar-log spatial coordinate mapping," in Proceedings of SPIE Intelligent Robots and Computer Vision VII, vol. 1002, pp. 58-65, 1988
- [23] R. L. Hsu, M. Abdel-Mottaleb, AK. Jain, "Face detection in color images", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, issue. 5, Pages: 696-706, May 2002
- [24] C. Palm, "Color texture classification by integrative Co-occurrence matrices", Pattern Recognition, vol. 37, issue.5, Pages: 965-976, May 2004
- [25] M. D. Plumbley, "Efficient Information Transfer and anti-Hebbian neural Networks," Neural Networks, vol. 6, pp. 823-833, 1993
- [26] N. Otsu, "A threshold selection method from gray level histograms", IEEE Transactions on Systems Man and Cybernetics, vol. 9, pp. 62-66, 1979

- [27] The web page for Amsterdam Library of Object Images, Informatics Institute, Faculty of Science – University of Amsterdam, retrieved from <http://staff. Science. Uva.Nl/~aloi>
- [28] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, John Wiley and Sons, Inc. New York, Chichester, Weinheim, Brisbane, Singapore, Toronto, 2001.
- [29] Ginsburg, “*Visual information processing based on spatial filters constrained by biological data*”, PhD dissertation, Aerospace Medical Research Laboratory, Wright-Patterson AFB, 1979
- [30] H. R. Wilson, “*Psychophysical Evidence for Spatial Channels*”, *Physical and Biological Processing of Images*, Proceedings of an International Symposium, Springer-Verlag Berlin, 1982
- [31] R. L. Hsu, M. Abdel-Mottaleb, AK. Jain, “*Face detection in color images*”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, issue. 5, Pages: 696-706, May 2002
- [32] C. Palm, “*Color texture classification by integrative Co-occurrence matrices*”, *Pattern Recognition*, vol. 37, issue.5, Pages: 965-976, May 2004
- [33] M. P. Sceniak, M. J. Hawken, and R. Shapley, “*Visual Spatial Characterization of Macaque V1 Neurons*”, *Journal of Neurophysiology*, February 2001
- [34] R. C. Gonzalez R. E. Woods, “*Digital Image Processing*”, second edition, international edition, pp. 289, 2002
- [35] Introduction to Fourier Transforms for image processing –Rotation and edge effects, <http://www. cs. unm. edu/~brayer/vision/fourier. html>
- [36] A. Taberner, J. Portilla and R. Navarro, “*Duality of log-polar image representations in the space and spatial frequency domains*”, *IEEE Transactions on Signal Processing*, vol. 47, pp. 2469 – 2479, September 1999
- [37] Y. Yu, T. Yamauchi and Y. Choe, “*Explaining low-level brightness-contrast illusions using disinhibition*”, *Biologically Inspired Approaches to Advanced Information Technology: First International Workshop, BioADIT 2004*, Lausanne, Switzerland, 2004, Revised Selected Papers, Lecture Notes in Computer Science, vol. 3141, pp. 166-175, 2004.
- [38] R. A. Young, L. M. Lesperance and W. W. Meyer, “*The Gaussian derivative model for spatial-temporal vision: I. Cortical model*”, *Spatial Vision*, vol. 14, pp. 261-319, 2001

- [39] W. T. Freeman and H. H. Adelson, "*The design and use of steerable filters*", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, pp. 891-906, September 1991
- [40] J. B. Price and M. H. Hayes, "*Steerable filter cascades*", in Proceedings of IEEE International Conference on Image Processing ICIP 99, vol. 2, pp. 880-884, Kobe, Japan, October 1999
- [41] H. Barlow and P. Földiák, "*Adaptation and Decorrelation in Cortex*", in The Computing Neuron, Addison-Wesley Computation And Neural Systems Series, Addison-Wesley Longman Publishing Co. , Inc. , Boston, MA, USA, 1989
- [42] A. J. Bell and T. J. Sejnowski, "*The 'Independent Components' of Natural Scenes are Edge Filters*," Vision Research, vol. 37, pp. 3327-3338, 1997
- [43] A. Cichocki and S. Amari, Adaptive Blind Signal and Image Processing, John Wiley & Sons, LTD, pp. 129, 2002
- [44] B. A. Olshausen and D. J. Field, "*Emergence of Simple-cell Receptive Field Properties by Learning a Sparse Code for Natural Images*," Nature, vol. 381, pp. 607-609, 1996.
- [45] B. A. Olshausen and D. J. Field, "*Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?*" Vision Research, vol. 37, pp. 3311-3325, 1997
- [46] H. B. Barlow, "*Unsupervised learning*," Neural Computation, vol. 1, pp. 295-311, 1989
- [47] B. A. Olshausen and D. J. Field, "*Natural Image Statistics and Efficient Coding*," Network: Computation in Neural Systems, vol. 7, pp. 333-339, 1996
- [48] P. D. Kovesi, The web page of MATLAB and Octave Functions for Computer Vision and Image Processing, School of Computer Science & Software Engineering, The University of Western Australia. Available from: <http://www.csse.uwa.edu.au/~pk/research/matlabfns/>
- [49] D. J. Field, "*Relations between the statistics of natural images and the response properties of cortical cells*", Journal of Optical Society of America A, vol. 4, pp. 2379 – 2394, December 1987
- [50] R. O. Duda, P. E. Hart and D. G. Stork, Pattern Classification, John Wiley and Sons, Inc. New York, Chichester, Weinheim, Brisbane, Singapore, Toronto, 2001

- [51] D. Casasent and D. Psaltis "*Position, rotation and scale invariant optical correlation*", Applied Optics, July 1976
- [52] D. Casasent and D. Psaltis "*Multiple-invariant space-variant optical processors*", Applied Optics, August 1977
- [53] D. Casasent and D. Psaltis "*Deformation invariant optical processors using coordinate transformations*", Applied Optics, February 1978
- [54] Y. Saito, S. Komatsu and H. Ohzu, "*Scale and rotation invariant real-time optical correlator using computer generated hologram*", Optics Communications, August 1983
- [55] A. Moya, J. Esteve-Taboada, J. Garcí'a and C. Ferreira, "*Shift- and scale-invariant recognition of contour objects with logarithmic radial harmonic filters*", Applied Optics, October 2000
- [56] D. Casasent, SF. Xia, A. J. Lee and JZ. Song, "*Real-time deformation invariant optical pattern recognition using coordinate transformations*", Applied Optics, March 1987
- [57] A. Jensen, L. Lindvold and E. Rasmussen, "*Transformation of image positions, rotations, and sizes into shift parameters*", Applied Optics, May 1987
- [58] D. Asselin and H. Arsenault, "*Rotation and scale invariance with polar and log-polar coordinate transformations*", Optics Communications, January 1994
- [59] CF. Hester and D. Casacent, "*Inter-class discrimination using synthetic discriminant functions (SDFs)*", Proceedings of the society of photo-optical instrumentation engineers, 1981
- [60] J. Garcia, T. Szoplik, and C. Ferreira, "*Shift-and-scale-invariant pattern recognition using an elliptic coordinate-transformed phase-only filter*", Applied Optics, August 1992
- [61] J. Yao and G. Lebreton, "*Invariant pattern recognition smart algorithms for segmented two-dimensional patterns with an axis of symmetry*", Optical Engineering, December 1999
- [62] P. Ambs, Y. Fainman, S. H. Lee, and J. Gresser "*Computerized design and generation of space-variant holographic filters. 1: System design considerations and applications of space-variant filters to image processing*", Applied Optics, November 1988

- [63] D. Cojoc, M. Teresa Molina, J. Garcí'a, and C. Ferreira, "*Coordinate-transformed filter for shift-invariant and scale-invariant pattern recognition*", Applied Optics, July 1997
- [64] T. Szoplik and H. Arsenault "*Shift and scale-invariant anamorphic Fourier correlator using multiple circular harmonic filters*", Applied Optics, October 1985
- [65] T. Koukoulas, R. C. D. Young, C. R. Chatwin "*Quantised phase filters with binary low and high pass amplitude response the effect of quantization for scale and rotation input distortions*", Optics and Lasers in Engineering, 2005
- [66] P. Bone, R. Young and C. Chatwin "*Position-, rotation-, scale-, and orientation-invariant multiple object recognition from cluttered scenes*", Optical Engineering, July 2006
- [67] R. Enrique G. M and J. Álvarez-B, "*Nonlinear composite filter performance*", Optical Engineering, June 2009
- [68] M. E. Munich, P. Pirjanian, E. Di Bernardo, L. Goncalves, N. Karlsson, and D. Lowe, "*Break-Through Visual Pattern Recognition for Robotics and Automation*", Evolution Robotics, Inc. Pasadena, California, USA
- [69] A. Gelb, Applied Optimal Estimation, written by the technical staff, the Analytic Sciences Corporation, 1974

## **APPENDIX A**

### **REAL TIME IMAGE PATTERN RECOGNITION IMPLEMENTATIONS IN AN OPTICAL SYSTEM**

In this appendix, the theory of a multi-invariant optical system and different implementation models are introduced. Both the physical implementations and system designs for position, rotation and scale invariant optical correlations are discussed, including different system structures for particular focus of attention using different optical elements.

#### **A.1 The Significance of the Real Time Implementations of Distortion Invariant Image Pattern Recognition in Optical System**

Pattern recognition is a problem of identification and localization of targets of interest in a scene. Optical pattern recognition can be accomplished by using optical correlators and provide useful tools for target detection and recognition, target tracking, and machine vision. However, conventional correlators, suffer from high sensitivity to even small changes in the input pattern. Possible sources of distortion are position, scaling and rotation changes. Many efforts have been directed toward distortion-invariant pattern recognition since 1970s [51].

Different solutions for implementations have been proposed by various authors to solve the distortion problems.

The rotation and scale invariant pattern recognition adopted the log-polar model to compress data and decrease degrees of freedom. Many researches about realizing a log-polar mapping in a digital system had been performed [52]. However, pattern recognition using a coherent optical system has great advantages of high speed and parallel processing. Therefore optical implementations using holographic optical elements, laser beams and lenses have been used for implementing log-polar coordinate transformations. Such systems are attractive for their high processing speed. For most of the cases, the systems are composed of two parts: the coordinate transformation portion and the correlation portion. The coordinate transformation portion can be implemented either in the object plane or the correlation plane depending on the different system schemes. Some new techniques and system schemes are discussed too in the following contents.

## **A. 2 Various Models of Position, Scale and Rotation Invariant Optical Correlation**

### ***A.2.1 A brief review of different implementation models***

In the 1970s, David Casasent and Demetri Psaltis presented a new theory about optical transformation that combines geometrical coordinate transformations with the conventional optical Fourier transformation [51]. They had also generated

the general formulation of its properties, the optical synthesis and the existence of space variant optical processors using coordinate transformations in [52] and provided a theoretical analysis of multiple invariant optical correlators in [53]. In the 1980s, as a further approach to the real time deformation invariant optical correlator, Yoshiharu Saito implemented the log-polar mapping using a computer generated hologram and illustrated the typical devices used in such an optical system [54]. The traditionally used and preferred Fourier transform was substituted by other mathematical operations such as the Mellin transform, the fractional Fourier transform, the Hough transform, or the Hartley transform. In the first attempt Casasent and Psaltis obtained scale and rotation invariance with a method based on the Mellin transform but at the price of losing the shift invariance. Later, Duvernoy, described the optical Fourier spectra with statistical descriptors. Sheng and Duvernoy used circular-Fourier-radial-Mellin transform descriptors, obtained invariance to translation, rotation, and scale. The recognition of the target was implemented as a classification in a multidimensional feature space. More recently the recognition with these three main invariances was obtained by Fang and Häusler using an original transformation [55].

### ***A.2.2 Initial models of rotation and scale invariant optical correlations***

As an application in optical pattern recognition, the optical implementation of a new transformation which is invariant to both a scale and a rotational change in

the input image is presented by Casasent and Psaltis both theoretically and experimentally in 1976 [51].

For a positional, rotational and scale invariant (PRSI) transformation system, the transfer function is a Mellin transform in  $r$  due to its scale invariant property (FT[F(  $\exp\rho, \theta$  )=  $M(\omega_\rho, \omega_\theta)$  ]. So if the complete transformation of  $f(x,y)$  is represented by,

$$M(\omega_\rho, \omega_\theta) = M_1(\omega_\rho, \omega_\theta) + M_2(\omega_\rho, \omega_\theta) \quad (\text{A.1})$$

The scaled and rotated transformation of  $f'(x,y)$  is,

$$M'(\omega_\rho, \omega_\theta) = M_1(\omega_\rho, \omega_\theta) \exp[-j(\omega_\rho \ln a + \omega_\theta \theta_0)] + M_2(\omega_\rho, \omega_\theta) \exp\{-j[\omega_\rho \ln a - \omega_\theta (2\pi - \theta_0)]\} \quad (\text{A.2})$$

After taking the Fourier transform of the product  $M * M'$ , we can obtain the cross correlation of two functions that are scaled and rotated versions of one another. Its amplitude will be equal to the amplitude of the auto-correlation function itself.

The system can be realized using conventional optical correlators. This allows one to utilize the parallel processing, 2-D nature, high speed, and Fourier transform properties of a coherent optical processor while expanding its flexibility and practicality. This architectural design philosophy also allows the realization of a space-variant system using digital, CCD, solid-state and other correlation techniques. The schematic diagram of a general optical space variant correlator is shown in Fig. A.1. The optical matched spatial filter is formed at plane P1 and it is achieved by placing the coordinate transformed function at the input plane P0 and recording the interference of its Fourier transform with a plane wave reference beam. The term of interest in the subsequent transmittance of plane

P1 is  $F^*(u,v)$  or the complex conjugate of the Fourier transform which diffracts light at an angle  $\theta$  to the optical axis as shown by lens L2 and plane P2 [53].

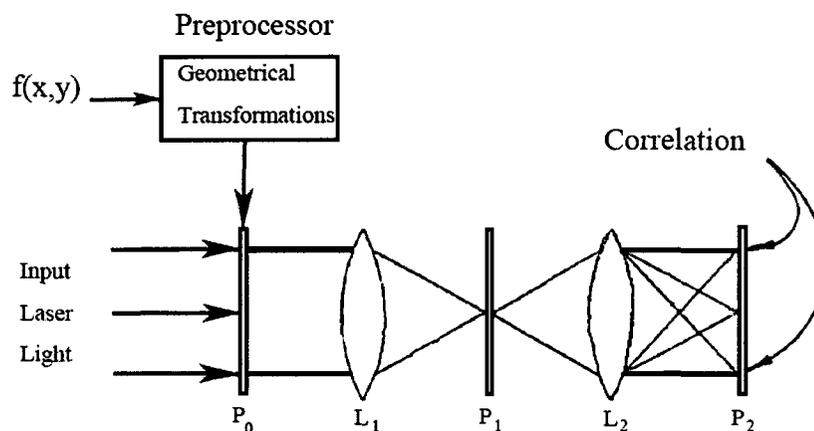


Fig. A.1 Schematic diagram of a general optical space-variant, deformation-invariant optical correlator.

The initial experimental confirmation used a simple square as the input function  $F(wx,wy)$ .  $F'(wx,wy)$  has a  $45^\circ$  rotation angle and a 200% scale change.

The central autocorrelation peak is shifted by the  $\theta_0 = 90^\circ$  rotation between the two inputs. The sum of the amplitudes of the cross correlation is found equal to the amplitude of the autocorrelation peak of the referenced one. The displacement of the correlation peak from the  $p'$  axis is found to be proportional to the 200% scale difference between the two inputs. Thus, input objects that differ in scale as well as rotation can be correlated optically. In addition, the scale change and amount of rotation can be determined from the coordinates of the correlation peak.

### ***A.2.3 Distortion invariant optical system models using coordinate transformations***

Optical coordinate transformations can be implemented by three different means: modification of the circuits of a TV receiver to display the coordinate transformation, the use of patented optical components and by means of holograms.

#### **(1) SAITO's Real Time Optical Correlator Model**

After Casasent and Psaltis used a polar camera and a scanned input spatial light modulator to perform the log-polar mapping, Saito reported a realization using a computer generated hologram (CGH) mask in 1983 [54]. Scale and rotation invariant optical pattern recognition for 2-D objects is accomplished by the coordinate transformation and succeeding optical matched filtering. By the use of BSO ( $Bi_{12}SiO_{20}$ ) device on the coordinate transformed plane, this processing is implemented in real time.

The schematic diagram of the system is shown in Fig. A.2. The coordinate transformed pattern written on the BSO device by an Ar laser is read out by an He-Ne laser and is treated as an input image for the matched filtering system. When the input pattern differs only in scale from the reference pattern which is used for producing the matched spatial filter (MSF), the resultant coordinate transformed pattern in the P1 plane moves laterally by the  $\ln r$  axis  $\propto a$ . Therefore the correlation peak produced in the output plane moves by the amount of  $-(f_2/f_1) \ln a$ , where  $f_1$  and  $f_2$  are the focal lengths of the lenses  $F_1$



often also the coordinate transformed pattern at P1. A device such as an LCTV is required to contain the P data from the system shown in Fig. A.3. The phase errors of the LCTV are corrected for by the phase conjugate hologram (PCH). A matched spatial filter of the coordinate transformed object to be recognized is formed at P2 to yield the optimal correlation SNR. The output correlation is produced at P3, where it is detected by a camera and displayed on an isometric display.

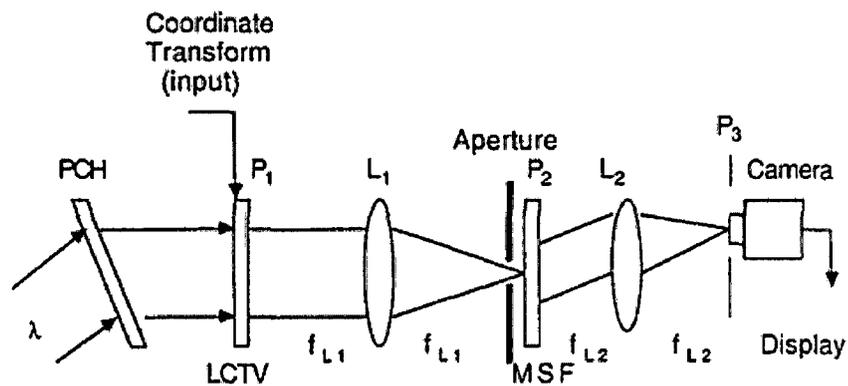


Fig. A.3 Real-time optical correlator system schematic.

### (3) Jensen's Multi-Channel Optical Robot System

Also in 1987, Jensen describes a novel method for recognizing the positions, orientations, and sizes of 2-D objects simultaneously by a two-channel coherent optical system [57]. The system opens the possibilities for an optical vision system that is able to detect positions, sizes, rotations, and shapes simultaneously for a number of objects. With several input images, the transformed output images are scaled by the mapping filter, and overlapping images can occur. How much the overlap will reduce the ability to recognize single objects is at the moment an open question which calls for further

experiments. The number of objects present simultaneously in the input plane does depend on the degree of overlap that can be tolerated. A reasonable estimate of this number would be in the range of 4-8 objects but it must be emphasized that this cannot as yet be substantiated with a classical optical correlator.

#### (4) Asselin's Juxtaposed CTs Model

In 1993, Daniel Asselin and Henri-H Arsenault proposed a new log-polar coordinate transformation (CT) optical real time implementation with computer generated holograms partitioned into non-uniformed facets [58]. The hologram recording is done with a commercial photo-typesetter at a very low cost. The image wrapping problem is alleviated by means of juxtaposed CTs. Because the FT is translation invariant, the intensity is invariant under changes of orientation or scale, and unknown objects may be classified by comparing their Fourier transforms. This differs from previous works where a correlator was used to classify the objects.

The holographic optical element they used was partitioned into many facets. Each one of them consists of a properly chosen diffraction grating which maps a small region of the hologram to the corresponding region of the CT. The liquid crystal light valve (LCLV) divides the setup into two parts: a writing part on the right and a reading part on the left. The LCLV can remove the complex phase information by the two polarizers and the beam splitter. The CT is formed on the writing side and a collimated argon laser beam trans-illuminates the object. The

LCLV, P1, P2 and the beam splitter were used to remove the complex phase information. Since the CCD camera is sensitive, a He-Ne laser beam is chosen to illuminate the reading side and reflects to P2. FT is performed on lens L2 (Fig. A.4).

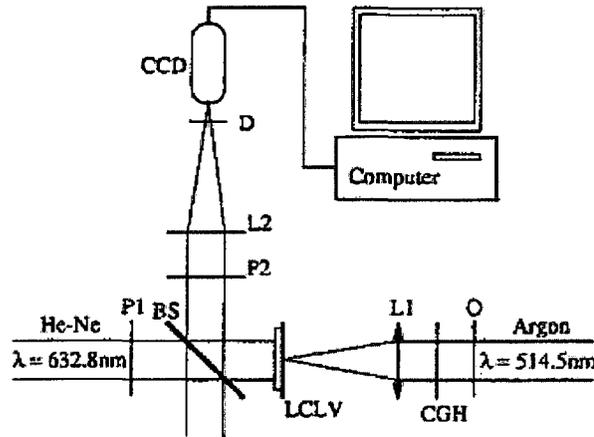


Fig. A.4 Experimental setup of this CT implementation

They also devised three different optical methods to implement the juxtaposition of the CTs, including successive redirection of facets, use of the twin diffracted order and multiplex holography on a facet.

### A.3 Some Advanced Technologies and Algorithms

#### ***A.3.1 Advanced computational algorithms for distortion invariant pattern recognition***

In the last decade, many new works had been done in realizing the log-polar mapping in an optical system. These topics include but not limited to the advanced invariant pattern recognition algorithms, the advanced spatial filters,

detecting noisy targets in environmental degradation/noisy environment, invariant object recognition under three dimensional rotations and changes of scale and automatic object identification irrespective of geometric change. In this section we will focus on the updated techniques on new computational algorithms and new filter structures.

Since the original work of Vander Lugt in optical matched filtering using 4f correlators and Wierer and Goodman in joint-transform correlators (JTCs), a relatively large number of techniques and configurations have been devised and investigated, all of which improve the tolerance of a system for a particular type of input distortion. These include the use of phase-only filters (POFs), circular harmonic expansion (CHE) functions, rotationally multiplexed holograms, synthetic discriminant functions (SDFs), minimum average correlation energy (MACE) filters, phase with constrained amplitude filters (PCMF), the Wiener filter, and the ternary phase-amplitude filter (TPAF).

### ***A.3.2 The synthetic discriminant functions (SDF)***

In 1980, Hester and Casasent proposed the idea of synthetic discriminant functions (SDF) for pattern recognition and pattern classification [59]. Making use of the idea of SDF, a correlator filter possessing shift, rotation and limited scale invariance of a certain range can easily be designed. According to the theory of SDF, the desired synthetic discriminant function is synthesized by adding up several mathematical functions algebraically, each of which is derived from a certain training pattern. The more training patterns used, the larger the

range of the invariance will be, while the more calculation is needed to obtain the synthetic discriminant function. Making use of the idea, a simple way to synthesize a shift, rotation and limited size correlation filter is proposed in [58]. The SDF is synthesized by superimposing four 2<sup>nd</sup> order circular harmonics of a training reference pattern in 4 different sizes.

The idea of the composite filter of Caulfield and Maloney has extended the matched-filter concept into a weighted sum of filters, which contain different views or sizes of an object, to obtain distortion-invariant object recognition [60]. The most important issue in composite filtering is the right choice of weights for a given composite filter that is a superposition of non-orthogonal components. This problem was also solved by the introduction of the synthetic discriminant functions (SDF), which included the expected distortions in the filter design to create invariance to distortions. The shift-and-distortion-invariant SDF is synthesized as a linear combination of a basis function set to achieve a constant correlation output for in-class images. The SDF theory was generalized to both spatial and frequency domains by using phase-only, binary, binary phase-only, and magnitude-only solutions. The practical implementation of the SDF exists in spatial-light modulators.

The usual method for invariant pattern recognition is to design filters that present a constant correlation peak for all distorted versions of the target of interest, as the harmonic expansion approach and the SDF approach. In 1999, J. Yao and G. Lebreton proposed another method to estimate and normalize the distortion parameters. Invariant pattern recognition can be accomplished by a correlation

between the normalized input pattern and the undistorted reference [61]. The advantage of this approach is that one can simply use a classical matched filter to perform pattern recognition without the need to synthesize invariant filters. For 2-D patterns, they focus on estimating the four parameters: the in-plane rotation, the scale, the projection axis direction and the projection amount. Once all parameters are accurately estimated, they can be normalized and distortion invariant pattern recognition can then be achieved by an optical correlator.

In orientation estimation, the authors describe two algorithms. One is associative-memory-based orientation estimation algorithm in which N straight lines are stored as associated memories. It consists of two correlators in cascade, one for correlation and the other for convolving the nonlinearly operated correlation and recorded in the second hologram. This algorithm is efficient for orientation estimation, even with strong input noise, but it requires several iterations.

### ***A.3.3 Advances of spatial filters***

Several methods have been proposed for obtaining different kinds of distortion invariance. Most of them are based on the use of filters partially matched to the target. Apart from the methods concerning transformations of the input plane, much research has been done with different kinds of filters. Several modifications of the matched filter are introduced. A single circular-harmonic component (CHC) is used as impulse response of the matched filter, providing shift and rotation invariance in the correlation plane.

### (1) The space-variant holographic filters

In 1988, a computerized optical system has been constructed for the design and generation of space-variant holographic filters. The system was experimentally employed to generate holographic optical filters for the optical implementations of Hough transform, coordinate transformations [62]. The system hybrid and combines the advantages of both optical and electronic computers. Optical implementation of space-variant operations will overcome digital implementations limitation and allow real-time processing of large data arrays due to the inherent parallelism, connectivity, and speed of optics. The input pattern is displayed on the CRT/LCLV and imaged on the space-variant holographic filter. A camera is placed in the reconstruction plane of the holograms. The horizontal linearity of the CRT is adjusted to assure that the image projected on the tilted plane of the hologram is appropriately scaled. The space-variant processor can be operated in real time and will be very useful in applications to pattern recognition, image processing, e.g., Hough transforms and coordinates transform.

Optical rotation and shift-invariant pattern recognition using circular harmonic expansion has been used and other methods based on coordinate transform such as Mellin transform have been also investigated. Furthermore, the coordinate transform can be realized by a mechanical system or a holographic filter, generated by a computer or recorded optically. However, further processing does require high diffraction efficiency of the filter achievable if

dichromated gelatin (DCG) holograms are employed. Therefore, they generated in DCG plate a matrix of holograms to perform the log polar transform.

## (2) The Phase-Only Filters

Phase-only filters (POF) can be used to increase the light efficiency. Moreover these filters provide much greater discrimination ability than all matched filters.

In 1992, J. Garcia, T. Szoplik, and C. Ferreirap proposed a shift-and-scale-invariant elliptic coordinate-transformed phase-only filter. The filter is built in three steps: the complex conjugate of a basic-size target spectrum is calculated, its phase-only part is taken, and then the elliptic coordinate transformation is made. The pattern recognition system is solely composed of a 4f-type correlator with its shift invariance preserved. No separate coordinate transformer is necessary.

In 1997, D. Cojoc, M. Teresa Molina, J. Garcí'a, and C. Ferreira proposed a variable radial coordinate transformation of the phase-only filter that is dependent on the energy's angular distribution of the target spectrum to perform shift and scale invariant pattern recognition [63]. The filter function is then transformed by means of stretching along the radial coordinate so that the same energy contribution to the correlation peak is provided for any size target. A radially stretched filter (RSF) is prepared for a binary target and only the phase part is taken to serve as the RSF. Only one computer-generated hologram is needed for recognition instead of two, as is the case in systems based on coordinate

transformation in the object plane one hologram for coordinate transformation and the other for filtering are needed.

### (3) The Circular/Radial Harmonic Filters

In 1985, a version of the shift and scale-invariant matched filtering technique using a nonsymmetrical correlator was brought up by Tomasz Szoplik and Henri H. Arsenault [64]. Because of its rotation invariance property, the circular harmonic filter (CHF) was used instead of a classical matched filter with a wedge mask, though this method is proposed only as a digital scale invariant technique for pattern recognition.

In 2000, a modified phase-only logarithmic radial harmonic (LRH) filter that permits the realization of a shift and scale invariant optical recognition of contour objects was brought up. The modified LRH filter is a complex filter that compensates the energy variation resulting from the scaling of contour objects. Using an expansion analogous to the Mellin radial harmonics, although not orthogonal, into logarithmic radial harmonics (LRH) in the Fourier plane, they proposed a phase filter that provided good results in a broad range of scales.

The phase-only LRH filter has the form:

$$H^*(\rho, \phi) = \exp[i\Omega(\phi)](\rho/d)^{i(p/\omega)} \quad (\text{A.3})$$

The new modified LRH (MLRH) filter which will be valid for contour objects has the form:

$$H^*(\rho, \phi) = \exp[i\Omega(\phi)](\rho/d)^{i(p/\omega)-1} \quad (\text{A.4})$$

Using a classical convergent correlator, they optically obtained the correlations. To do that, the filters generated by computer were photoreduced on a lithographic film. The final correlation plane was captured with a Pulnix Model TM-765 CCD camera. The results of MLRH's the correlation peak is greater than the LRH's. This is due to relation which overloads the energy of object when in fact its energy is smaller. These optical results confirm those obtained digitally.

#### (4) The Quantized Phase Filters, MACH Filters and Nonlinear Composite Filters

After 2000, still more studies of filters had been done by different authors. The followings are a few examples of the new filter structures which are brought up in the recent years.

In 2005, low pass phase filtering and high pass phase filtering techniques were discussed to demonstrate the effect of quantization for scale and rotation input distortions [65]. Phase quantization proved to be very effective. In all cases, it seems to have very little effect on the invariance properties of the two filters. Even using binary amplitude information with just 4 levels of phase, a correlation peak is present. These filters require very little spectral information to be stored and may thus be suitable for systems where resolution should be kept to a minimum while still maintaining target identification capabilities.

In 2006, a maximum average correlation height (MACH) filter is used to create invariance to orientation and gives good tolerance to background clutter and noise [66]. They combined the two existing techniques. Out-of-plane rotation invariance and discriminating the target objects from cluttered or noisy

backgrounds are achieved using a MACH filter. A change in scale or rotation in the target object resulting in a horizontal or vertical shift makes the object detectable by correlation with the log-polar mapping.

The MACH filter, like the SDF, is another method of creating invariance to distortions in the target object by including the expected distortions in the construction of the filter. The MACH filter maximizes the relative height of the average correlation peak with respect to the expected distortions. Unlike the SDF, the MACH filter can be tuned to maximize the correlation peak height and noise suppression while also being tolerant to distortions in the target object.

In 2009, a K-Law nonlinear composite filter is used in object recognition with rotation, scale, noise, and illumination distortions [67]. The segmented filter improves the filter's performance when dealing with images rotated around the 360° range. The use of the nonlinear method SNM was proven as an effective method for better noise tolerance and illumination invariance. Nonlinear composite filters were used effectively with real-life images in the discrimination to identify and locate the CPU of a computer sound card even with extreme illumination variations and some rotation invariance.

## APPENDIX B

### SOME DERIVATIONS OF INTRODUCING MULTIPLE LINEAR CHANNELS IMPROVES THE PATTERN RECOGNITION PERFORMANCE

If one measurement of constant  $x$  with white noise

$$z = x + \omega$$

Then best guess of true value of  $x$  based on  $z$  is

$$\hat{x} = z$$

And error is

$$e = x - \hat{x} = x - (x + \omega) = -\omega$$

Variance of error is

$$E\{e^2\} = E\{(-\omega)^2\} = E\{\omega^2\} = \sigma_\omega^2$$

If two measurements of  $x$

$$z_1 = x + \omega_1 \quad \text{with} \quad E\{\omega_1^2\} = \sigma_1^2$$

$$z_2 = x + \omega_2 \quad E\{\omega_2^2\} = \sigma_2^2$$

Then best guess of  $x$  is combination of  $z_1$  and  $z_2$ , call it

$$\hat{x} = \alpha_1 z_1 + \alpha_2 z_2$$

Error is

$$x - \hat{x} = x - \alpha_1 z_1 - \alpha_2 z_2$$

$$e = x - \alpha_1(x + \omega_1) - \alpha_2(x + \omega_2) = x(1 - \alpha_1 - \alpha_2) - \alpha_1\omega_1 - \alpha_2\omega_2$$

If we want

$$E\{e\} = (1 - \alpha_1 - \alpha_2)x - \alpha_1E(\omega_1) - \alpha_2E(\omega_2) = 0$$

Then

$$\alpha_1 = 1 - \alpha_2$$

$$\text{Var}\{e\} = E\{e\} = E\{\alpha_1\omega_1 + \alpha_2\omega_2\}^2 = \alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2 = (1 - \alpha_2)^2\sigma_1^2 + \alpha_2^2\sigma_2^2$$

$$\text{When } \frac{\partial \text{Var}\{e\}}{\partial \alpha_2} = 0,$$

$$\alpha_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \quad \text{and} \quad \alpha_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

$$\hat{x} = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} z_1 + \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} z_2$$

The minimum  $\text{Var}\{e\}$  is

$$\alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2 = \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)^2\sigma_1^2 + \left(\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\right)^2\sigma_2^2 = \sigma_1^2 \frac{1}{1 + \frac{\sigma_1^2}{\sigma_2^2}} < \sigma_1^2$$

This result proves that if we have more than one channel processing simultaneously, the variance of recognition error will decrease compared with one channels case. For example, if there are 3 channels which have the same standard deviation, we may shrink the  $\text{Var}\{e\}$  to  $1/\sqrt{3}$  of the single channel case theoretically. So it is necessary to consider all channels' performances. These derivations can also be found in [69].

## APPENDIX C

### THE ORGANIZATION OF THE CODES IN CD

The main functions that used in the simulation process are listed below. The other side support codes are included in the CD but not listed.

#### Directory \MATLAB\

##### \Centroid\_Li\

```
%calculate the centroid of objects in the input color images
    centerForLP_color_red.m
    centerForLP_color_green.m
    centerForLP_color_blue.m
%calculate the centroid of objects in multiple spatial filtered grayscale images
    centerForLP_MultiBands.m
```

##### \ColorImageCode\

```
%decompose the color image into red, green, blue channels in batch mode.
    rgb_decompose_batch.m
%grayscale stretch function applied on blue channel in batch mode
    blue_stretch_batch.m
```

##### \LogGaborPR\_Li\

```
%encode features of the log-polar mapped image for ICA analysis in batch mode
    featureEncodeBatchLG_ICA.m
%encode features of the rotated log-polar mapped image for ICA analysis in
batch mode
    featureEncodeBatchLG_ICA_Rotated.m
```

##### \MultiBandsCode\Frequency Filtering

```
%process the image through low pass, mid pass and high pass filters
    freq_direct.m
    freq_direct_bandpass.m
    freq_direct_high.m
%multiple spatial frequency filtering in batch mode
    MultiBandsFiltered_Batch.m
%a filtering example including calculate the sum of filtered images
```

```
        construct_back.m
%an artificial image example of cosine function
        cos.m
%visualize all filters
        filter_show.m
%find the best parameters to simulate the multiple spatial channels exist in HVS
        find_best_filter.m
```

```
\MultiBandsCode\Spatial Filtering
```

```
% process the image through low pass, mid pass and high pass filters by spatial filtering
```

```
        high_gaussian_example.m
```

```
        low_gaussian_example.m
```

```
%transform the filtering from spatial filtering to frequency filtering
        spatial_freq.m
```

```
%color image datasets for test are in the following directory
```

```
\MATLAB\MyColorObjects\
```

```
%multiple frequency filtering grayscale image datasets for test are in the following directory
```

```
\MATLAB\MyMultiBandsObjects\
```

```
%HPF filtered image are in the following directory
```

```
\MATLAB\MyMultiBandsObjects1\
```

```
%Other called functions for decorrelation are in the following directory
```

```
\MATLAB\Decorrelate_Li\
```

```
\MATLAB\imageica_D_Li\
```

### **Directory \OpenCV\**

```
% log-polar wrapping for one test image
```

```
\OpenCV\LogPolarTest\
```

```
%log-polar wrapping for all test datasets in batch mode
```

```
\OpenCV\LogPolarBatch\
```

## APPENDIX D

# THE OPENCV 2.0 ENVIRONMENT SETUP ON MICROSOFT VISUAL STUDIO 2008

### PRE-INSTALLED SOFTWARES

VC++ 2008: <http://www.microsoft.com/express/product/default.aspx>

OpenCV: <http://www.opencv.org.cn/index.php/Download>

CMake: <http://www.cmake.org/cmake/resources/software.html>

### COMPILE OPENCV

#### 1. Use CMake to Export VC++ Project File

- Run cmake-gui, set the path as the installation path of OpenCV. Create new path \OpenCV2.0\vc2008 to restore the compiled running results.
- Click configure then choose Visual Studio 9 2008.

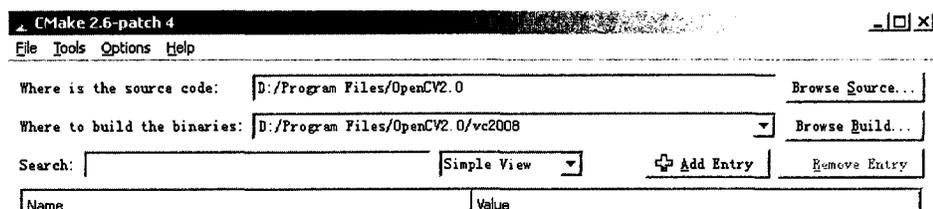


Fig. D.1

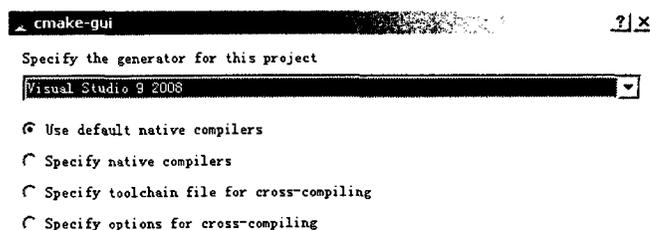


Fig. D.2

## 2. Compile OpenCV Debug and Release Version Dynamic Link Library

Generate the VC Solution File OpenCV. sln under \OpenCV2. 0\vc2008

- Under Debug, select Solution OpenCV under Solution Explorer, run "Rebuild Solution"; Build under the install tab.
- Under Release, select Solution OpenCV under Solution Explorer, run "Rebuild Solution"; Build under the install tab.

Then, file \*.d. dll of OpenCV (for debug) and file \*. dll (for release) will show under \OpenCV2. 0\vc2008\bin; File d. lib of OpenCV (for debug) and file \*. lib (for release) will show under \OpenCV2. 0\vc2008\lib; head file \*. h will show under \OpenCV2. 0\vc2008\include\opencv.

Configure the environment variables. Add path \OpenCV2. 0\vc2008\bin to Windows system environment Path. Then log off the current user or restart the computer.

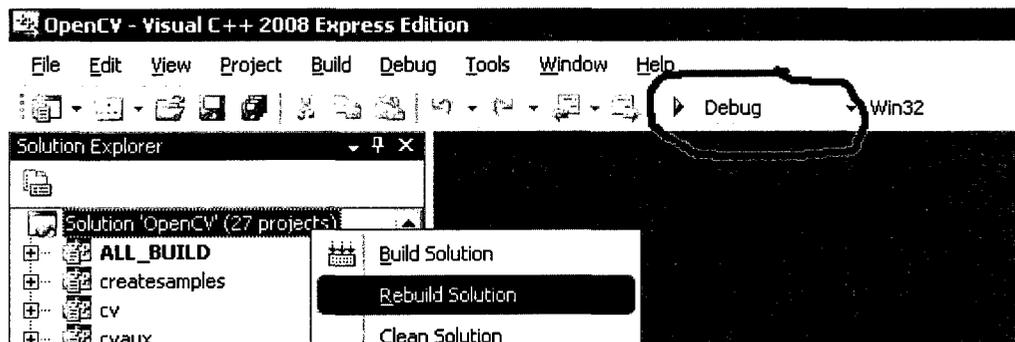


Fig. D.3

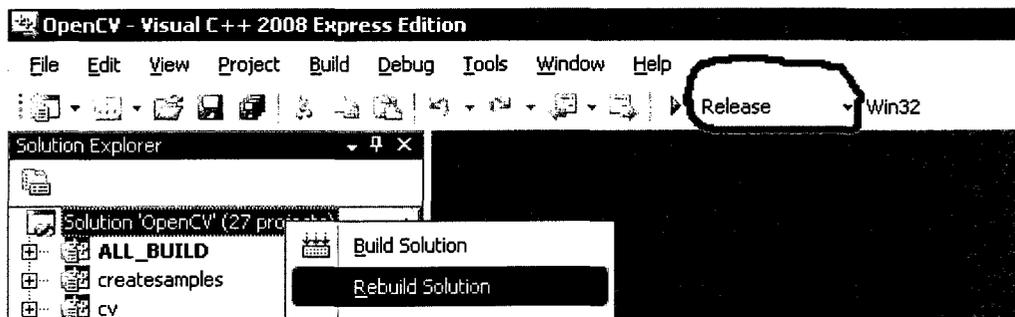


Fig. D.4

### 3. Setup the OpenCV Environment for VC++ 2008

Open VC++ 2008 Tools -> Options -> Projects and Solutions -> VC++ Directories

- Show directories for include files, add path \vc2008\include\opencv
- Show directories for library files, add path \vc2008\lib

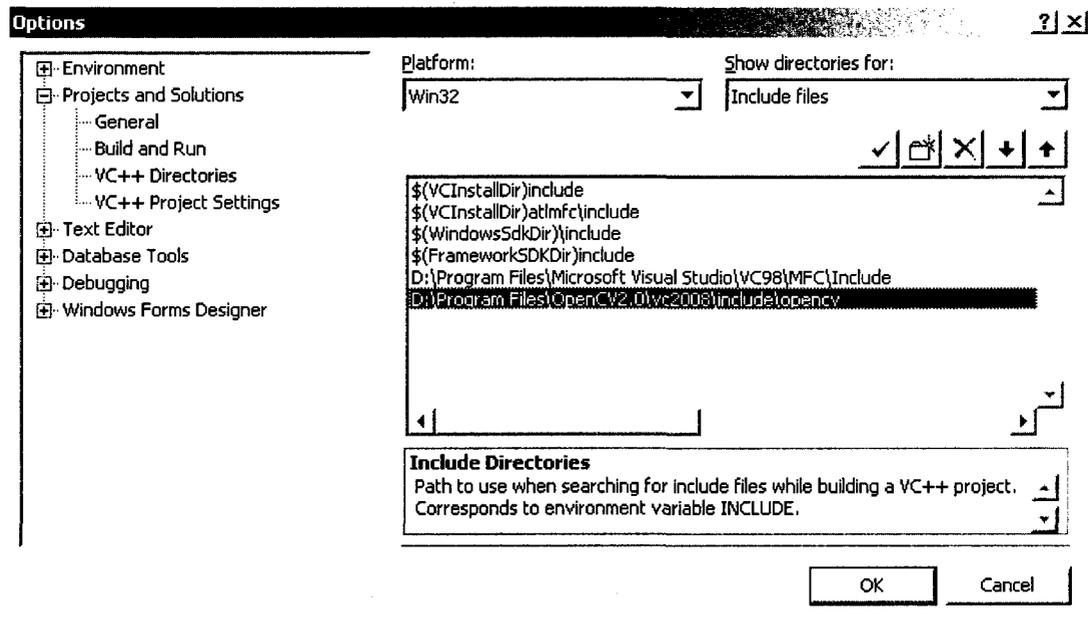


Fig. D.5

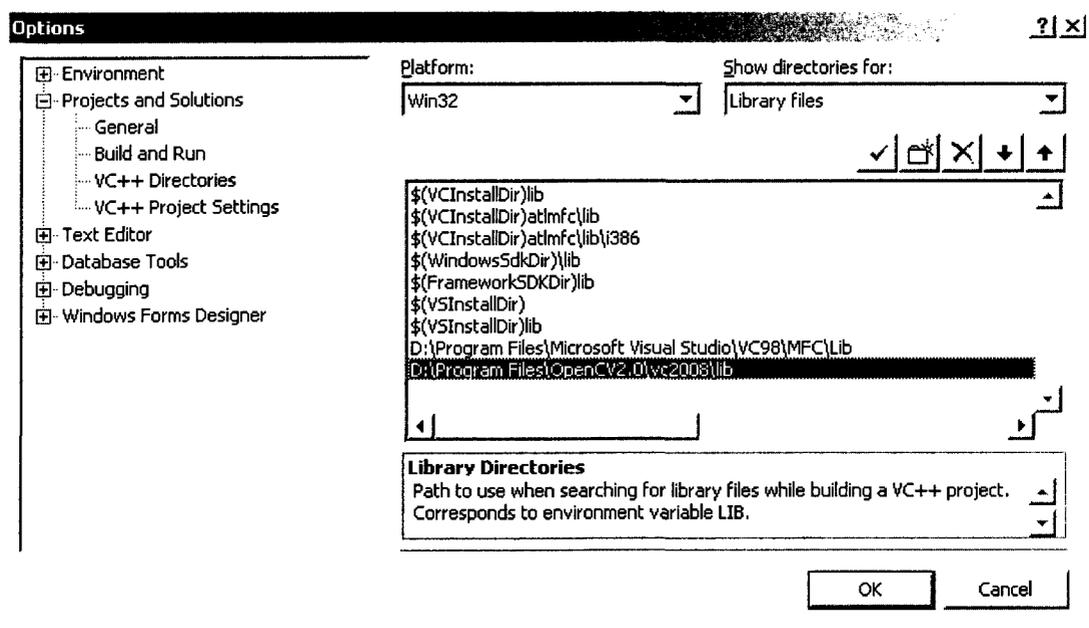


Fig. D.6

#### 4. Add library path to OpenCV 2. 0

- Open VC++ 2008, create a Win32 console project.
- Select the solution explorer properties.
- Add library to debug configure: `cxcore200d.lib cv200d.lib highgui200d.lib`
- Add library to release configure: `cxcore200.lib cv200.lib highgui200.lib`

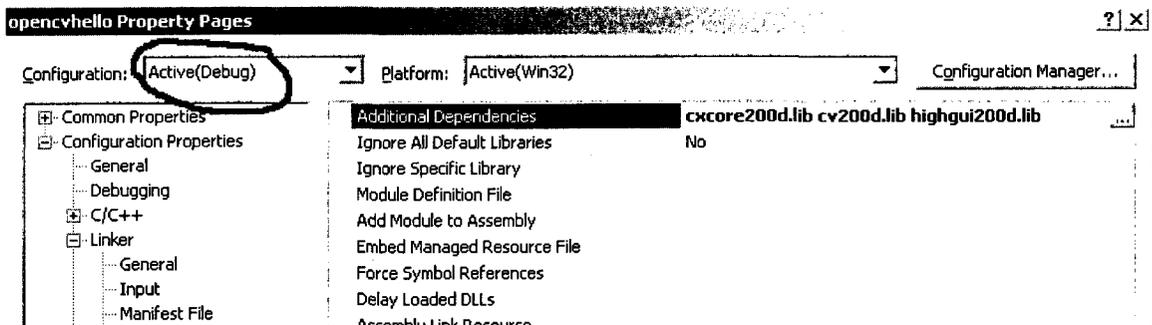


Fig. D.7

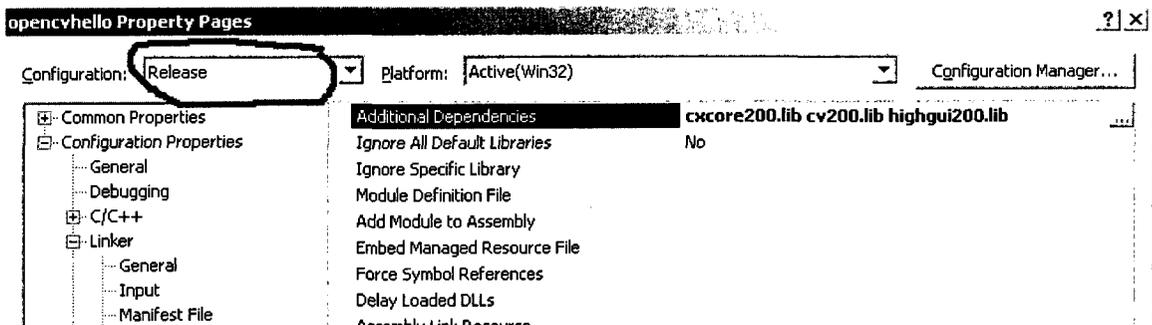


Fig. D.8