

University of New Hampshire University of New Hampshire Scholars' Repository

Doctoral Dissertations

Student Scholarship

Fall 2009

Biologically inspired feature extraction for rotation and scale tolerant pattern analysis

Dragan Vidacic

University of New Hampshire, Durham

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

Recommended Citation

Vidacic, Dragan, "Biologically inspired feature extraction for rotation and scale tolerant pattern analysis" (2009). *Doctoral Dissertations*. 507.

<https://scholars.unh.edu/dissertation/507>

This Dissertation 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 Doctoral Dissertations by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.

**BIOLOGICALLY INSPIRED FEATURE EXTRACTION FOR
ROTATION AND SCALE TOLERANT PATTERN ANALYSIS**

BY

DRAGAN VIDACIC

**B.S. University of Novi Sad, 2000
M.S. University of New Hampshire, 2003**

DISSERTATION

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

**Doctor of Philosophy
in
Electrical Engineering**

September, 2009

UMI Number: 3383330

INFORMATION TO USERS

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

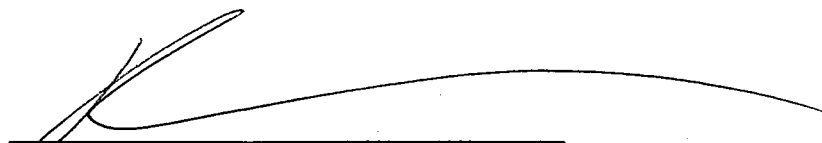
In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.




UMI Microform 3383330
Copyright 2009 by ProQuest LLC
All rights reserved. This microform edition 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 dissertation has been examined and approved.



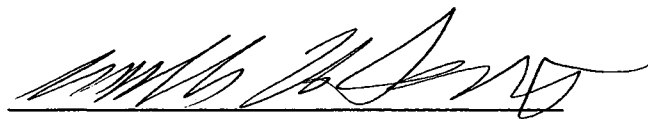
Dissertation Director, Richard A. Messner,
Associate Professor, Department of Electrical and Computer Engineering



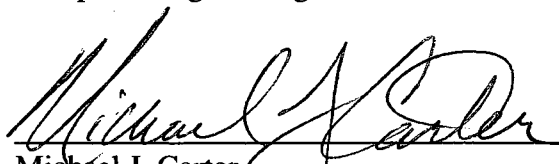
W. Thomas Miller, III,
Professor, Department of Electrical and Computer Engineering



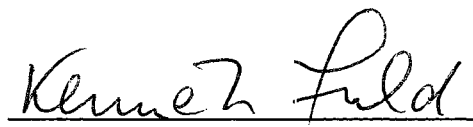
Andrew L. Kun,
Associate Professor, Department of Electrical and Computer Engineering



William H. Lenharth,
Affiliate Associate Research Professor, Department of Electrical and
Computer Engineering



Michael J. Carter,
Associate Professor, Department of Electrical and Computer Engineering



Kenneth Fuld,
Professor, Department of Psychology

8/2/2009
Date

ACKNOWLEDGEMENTS

I would like to thank my advisor Professor Richard A. Messner for his support, guidance and advice during the research and writing of this dissertation. The experience and knowledge I gained from him during our work, is, and will be, a great value for me throughout my career.

Also, my thanks goes to Dr. Thomas Miller, Dr. William Lenharth, Dr. Andrew Kun, Dr. Michael Carter and Dr. Kenneth Fuld for being members of my committee, for their advice and insights.

Finally, I would like to thank my family and friends for their continuous and unselfish support during this very special journey.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS.....	iv
LIST OF FIGURES	vi
LIST OF TABLES	ix
ABSTRACT	x
I. INTRODUCTION.....	1
1.1 Statement of the Problem	1
1.2 Importance of the Problem	7
1.3 Specific Project Objectives	7
1.4 Background.....	9
II. SOLUTIONS OF RECURRENT LINEAR NETWORKS APPLIED IN LOG-POLAR DOMAIN	20
2.1 Solutions of Network with LSI via Iterative and Convolution Models	21
2.2 Convolution in Log-Polar Domain	26
2.3 Non-Uniform Lateral Inhibition - the Exact Network Solution	30
2.4 Characterization of the Space Invariant Network with Lateral Connectivity in Log-Polar Space	33
III. CONSTRUCTION OF BIOLOGICALLY INSPIRED FILTERS UTILIZING SPECTRAL PROPERTIES OF TOEPLITZ-BLOCK-TOEPLITZ MATRICES	37
3.1 Remarks on Recurrent Linear Networks and Toeplitz-Block-Toeplitz Systems	37
3.2 Construction and Efficient Solution of Recurrent Neural Network Characterized by TBT Matrix	41
3.3 Construction of Connection Matrix (Five Practical Models).....	45
3.3.1 Gaussian model.....	46
3.3.2 DOG model.....	47
3.3.3 Difference of Offset Gaussians (DOOG) model.....	48
3.3.4 Second Order Derivative of Gaussian (2ODG) model.....	49
3.4 Steerability of the Recurrent Network with Lateral Inhibition-Excitation	51
3.5 Simulation Results	53
IV. SEMI-BLIND PRE-WHITENING BY SPATIALLY INVARIANT NETWORKS WITH LATERAL INHIBITION/EXCITATION	60
4.1 Decorrelation and Whitening by Recurrent Neural Networks	60
4.2 Decorrelation Role of Non-Adaptive Networks with Lateral Connectivity	66
4.3 Category-Specific Pre-Whitening Filters	67
4.4 Experimental Results	73
V. BIOLOGICALLY INSPIRED INFORMATION PRE-PROCESSING AND OBJECT RECOGNITION.....	82
5.1 Independent Component Analysis as Biologically Inspired Signal Encoding	82
5.2 Recognition of Objects Through Log-Polar Transform and Local ICA Analysis ..	89
5.3 Experimental Results	107
VI. CONCLUSION AND DIRECTIONS	112
REFERENCES	116

APPENDIX A. ADDITIONAL RESULTS DESCRIBING DECORELATION AND PRE-WHITENING BY FILTERS BASED ON NETWORKS WITH RECURRENT CONNECTIVITY	124
APPENDIX B. SPARSENESS OF THE ICA BASED FEATURE ENCODERS	129

LIST OF FIGURES

Figure 1.1 Biologically inspired computer vision system for pattern analysis; the lateral inhibition/excitation adapted for log-polar environment is implemented through Retino-Cortical Filter.	6
Figure 1.2 Log-polar mapping	10
Figure 1.3 Log-polar mapping (discrete implementation).....	11
Figure 2.1 The result of the network, the iterative and the convolution-based solutions for the 2-D grid with lateral inhibition; original image (<i>upper left</i>), resulting image (<i>upper right</i>); Resulting signal profiles: network versus iterative solution (<i>bottom left</i>), network versus convolution model (<i>bottom right</i>); Normalized Root Mean Squared Error between resulting signals: $NRMSE_{netw,iter} = 0.0038$, and $NRMSE_{netw,conv} = 0.0589$; Image size used is 32 x 32	23
Figure 2.2 The convolution kernels for model A (<i>left</i>) and model B (<i>right</i>).....	24
Figure 2.3 Linear combination of four Gaussian kernels (<i>left</i>); corresponding profile (<i>right</i>)	25
Figure 2.4 Linearly space-scaled kernels and their mapped versions - kernel size is increasing with distance from mapping origin	28
Figure 2.5 Distance parameter used for calculation of coupling coefficients in space-variant inhibition model;	31
Figure 2.6 Non-uniform lateral inhibition in log-polar space; original image (<i>left</i>) and processed image (<i>right</i>); Model B is used.....	32
Figure 2.7 Resulting receptive field formed as a linear combination of nine Gaussian functions (with spatial offset) with each function multiplied by corresponding coefficient from Model B kernel shown in Figure 2. 2 (<i>left</i>); Resulting field profile (<i>right</i>);	35
Figure 3.1 Uniform receptor-neuron grid	41
Figure 3.2 Excitatory/inhibitory feed-backward interaction between two arbitrary sensory units.....	42
Figure 3.3 Generating functions (<i>left</i>) and corresponding connection functions (<i>right</i>); Gaussian model (<i>top</i>), DOG model (<i>bottom</i>);.....	50
Figure 3.4 Generating functions (<i>left</i>) and corresponding connection functions (<i>right</i>); DOOG model (<i>top</i>), 2ODG model (<i>bottom</i>);.....	50
Figure 3.5 Steerable impulse response of the network with 2ODG coupling model in continuum; $h_i^\theta(x, y)$ is an arbitrarily rotated filter while $k_i(\theta)$ and $h_i^{\theta_i}(x, y)$ are the interpolating and basis functions respectively.....	53
Figure 3.6 The number of sufficient filters versus α ; Particular parameters chosen are: for Gaussian model $\sigma = \pi/3$, for the DOOG model $\sigma = \pi/3$, $u_0 = 2$, for the 2DOG model $\sigma = \pi/5$ while for the DOG model $\alpha_1 = -0.11$, $\sigma = \pi/4$, $\sigma_1 = \pi/5$	55
Figure 3.7 One pass convolution with pre-computed kernel (connection model is DOG with α_1, σ and σ_1 identical to those from Figure 3. 6 and $\alpha = -0.13$); filtered square pattern (<i>top</i>), mid-line resulting profiles for iterative procedure and single-pass convolution (<i>bottom</i>); pre-computed kernel constructed by using $N=7$	56

Figure 3.8 Steerable network impulse response for 2ODG connection model ($\alpha = 0.12$, $\sigma = \pi/5$, $N = 7$); $h(m, n)$ portion (top), $h_i(m, n)$ (bottom);.....	58
Figure 3.9 Steering the network with 2ODG connection model ($\alpha = 0.12$, $\sigma = \pi/5$, $N = 7$); linear combination of 15 filters used to steer the response to three different orientations.	58
Figure 4.1 The recurrent network with adaptive connections (self inhibition/excitation can potentially be allowed) used in [63]	61
Figure 4.2 Transformation of the image patch into single column vector.....	65
Figure 4.3 The partition of TBT matrix into groups of identical entries	71
Figure 4.4 Learning the pre-whitening filter.....	72
Figure 4.5 Images of nature (category 1).....	74
Figure 4.6 Images of log-polar warped objects (category 2)	74
Figure 4.7 Images of predominantly horizontal gratings/texture (category 3)	74
Figure 4.8 Performance of learned pre-whitening filter when processing three groups of data from Category 1	78
Figure 4.9 Performance of learned pre-whitening filter when processing three groups of data from Category 2	79
Figure 4.10 Performance of learned pre-whitening filter when processing three groups of data from Category 3	80
Figure 5.1 Distribution of activity of code coefficients for two cases: sparse code (full line) and non-sparse code (dotted line)	84
Figure 5.2 Feature extraction and object recognition based on pre-whitening and ICA filters	90
Figure 5.3 Four objects from image database	91
Figure 5.4 The result of centering of an object prior to log-polar mapping	93
Figure 5.5 Centered object image extraction	94
Figure 5.6 Log-polar mapping of centered objects	95
Figure 5.7 The problem of fixed grid with sequential patch sampling in log-polar space; example where rotation of object causes completely different content of patches (right); only for specific rotation angles this content is preserved (left).....	96
Figure 5.8 Selection of patches based on points of interest – sampling locations are not a <i>priori</i> defined.....	97
Figure 5.9 The spatial response of the even log-Gabor function; 2-D grey scale intensity model (left); center-line profile (right).....	100
Figure 5.10 The spatial response of the odd log-Gabor Function; 2-D grey scale intensity model (left); center-line profile (right).....	100
Figure 5.11 An image in log-polar space (left) with outlined pseudo-salient locations (right)	102
Figure 5.12 Collection of patches centered at pseudo-salient points from Figure 5. 11.	103
Figure A.1 Averaged performance of learned pre-whitening filter when processing 15 individual images from Category 1	126
Figure A.2 Averaged performance of learned pre-whitening filter when processing 24 individual images from Category 2	127
Figure A.3 Averaged performance of learned pre-whitening filter when processing 15 individual images from Category 3	128

Figure B.1 Warped image (top); histogram approximations of the pre-processed signal (bottom left) and of the post-processed signal by ICA feature encoding (bottom right).....	130
--	-----

LIST OF TABLES

Table 3.1 Steering coefficients for the network with 2ODG connection model ($\alpha = 0.12$, $\sigma = \pi/5$, $N = 7$).	57
Table 4.1 The results of decorrelation with filters based on non-adaptive uniform networks with lateral inhibition.	76
Table 5.1 The results of non-rotated object recognition by four different classifiers	107
Table 5.2 The results of object recognition by hybrid classifiers (rotated objects as input)	108
Table 5.3 The results of object recognition by hybrid classifiers and for x and y feature components normalized to range (0, 5).	109
Table 5.4 The results of non-rotated object recognition by hybrid classifiers and for eliminated x and y feature components.	110
Table A.1 The results of decorrelation with filters based on non-adaptive uniform networks with lateral inhibition – images from category 2 used for processing	124
Table A.2 The results of decorrelation with filters based on non-adaptive uniform networks with lateral inhibition – images from category 3 used for processing	125

ABSTRACT

BIOLOGICALLY INSPIRED FEATURE EXTRACTION FOR ROTATION AND SCALE TOLERANT PATTERN ANALYSIS

By

Dragan Vidacic

University of New Hampshire, September, 2009

Biologically motivated information processing has been an important area of scientific research for decades. The central topic addressed in this dissertation is utilization of lateral inhibition and more generally, linear networks with recurrent connectivity along with complex-log conformal mapping in machine based implementations of information encoding, feature extraction and pattern recognition. The reasoning behind and method for spatially uniform implementation of inhibitory/excitatory network model in the framework of non-uniform log-polar transform is presented. For the space invariant connectivity model characterized by Topelitz-Block-Toeplitz matrix, the overall network response is obtained without matrix inverse operations providing the connection matrix generating function is bound by unity. It was shown that for the network with the inter-neuron connection function expandable in a Fourier series in polar angle, the overall network response is steerable. The decorrelating/whitening characteristics of networks with lateral inhibition are used in order to develop space invariant pre-whitening kernels specialized for specific category of input signals. These filters have extremely small memory footprint and are successfully utilized in order to improve performance of adaptive neural whitening

algorithms. Finally, the method for feature extraction based on localized Independent Component Analysis (ICA) transform in log-polar domain and aided by previously developed pre-whitening filters is implemented. Since output codes produced by ICA are very sparse, a small number of non-zero coefficients was sufficient to encode input data and obtain reliable pattern recognition performance.

CHAPTER I

INTRODUCTION

Throughout history, humans have always been inspired by nature when trying to solve practical problems. Biologically motivated models in engineering and the evolution of technology have lead to more and more difficult challenges to be considered resulting in development of fields such as artificial intelligence, artificial muscles, artificial vision and many others [1]. One of the most intriguing senses that many biological entities possess is vision. The way in which humans acquire, process and, finally, interpret visual information has been a topic of scientific study for years. As our knowledge about the physiology of biological vision evolved, efforts to create or mimic adequate models of the visual information processing pathway followed. The effectiveness, speed and accuracy of certain aspects of information processing system in humans are still superior in many ways over machine implementations. Questions arise whether, and how, we can further reduce the gap between the biology and man made machines so that our “smart computers” can sense, process and perform as well as a human.

1.1 Statement of the Problem

The central issue addressed by the research presented herein is the efficient exploitation of specific biological functionality which is observed in many mammalian sensory subsystems. The exploitation of such functionality is directed toward solving

complex engineering problems related to signal processing and pattern classification. An important aspect of the work addressed in this dissertation is development of adequate information processing models that can find direct implementation in the domain of machine vision. In the quest for such knowledge, the inevitable question arises: how can one create a computationally efficient, biologically inspired, image processing framework suitable for pattern recognition that can be utilized on an inexpensive PC platform. Two requirements are common for machine vision applications. The first one is maintaining a large field of view while preserving the detailed information from the scene at the point of interest. The second one is providing the capability of the system to process the data at a very fast rate. These requirements directly correlate with the way humans (and many other organisms) process the visual information. The non-uniform distribution of visual sensors in the retina corresponds to the spatially variant sampling strategy of the Human Visual System (HVS). The response of a given retinal ganglion cell depends on the light intensity falling on the photoreceptors within the small, more or less, circular area of the retina called the ganglion cell receptive field [2]. The retinal receptors are densely packed at the fovea and their density decreases as the radial distance from the fovea increases. This directly allows for data reduction and wide field of view while preserving the detailed information at the gaze point. The diameter of the receptive field size of the retinal ganglion cells as reported by [2], [3] and [4] increases linearly with eccentricity. The distribution of ganglion cells has been reported to be similar to the cone distribution [2], [5], [6] and [7]. At the fovea there are at least three ganglion cells per cone [7], while at the periphery there is one ganglion cell per two cones. Obvious data reduction performed at the retinal sensory level through non-uniform sampling is only one aspect of

information processing occurring in frontal stages of the HVS. Early discoveries have shown that the inverse of cortical magnification factor¹ is linearly dependent on eccentricity [9] and [10]. This lead to the introduction of a conformal logarithmic mapping (or sometimes call merely a complex-log, or log-polar mapping) as the model that agrees well with the rearrangement of retinal signals in the visual cortex (this signal rearrangement or mapping is usually referred as retino-cortical projection) [11], [12], [13] and [14]. Additionally, the lateral inhibition mechanism is known to occur in the early stages of visual information processing in biological systems. This is especially seen in the retina, lateral geniculate nucleus (LGN) and certain areas of the visual cortex [15] - [19]. According to [20] in the mid-1960's the linear retinal receptive field model was introduced and antagonistic center-surround profile of retinal receptive fields was modeled as a Difference of Gaussian (DOG) function. The DOG model also describes the characteristics of LGN receptive fields since they show a great level of similarity to retinal ganglion cell receptive fields [21], [22]. This dissertation describes the effort aimed towards integrating the biological concepts of spatially non-uniform image processing and data reduction models with signal processing principles inherent to various stages of biological vision. These techniques are eventually embedded in a computationally efficient signal processing and pattern recognition framework.

In the frontal processing stage of the proposed model, the effort has concentrated on developing computationally effective strategies that incorporate networks with lateral feedback connectivity in conjunction with a log-polar mapping. The importance of lateral inhibitory connectivity in the context of the visual processes is very often

¹ The cortical magnification factor is defined as the linear extent of the striate cortex to which each degree of the retinal field projects.

perceived as edge enhancement (or even interpreted as Mach-band effect). The complete role of inhibitory mechanisms can be revealed only after relating the statistical properties of the visual information at the input and output of the processing framework. More importantly, shaping the chosen signal processing tools (i.e. filters effectively performing the lateral-inhibitory/excitatory process on the particular signal) for foveated and/or a warped environment leads towards attractive models that correlate very closely with biological principles of information extraction and their artificial implementation in the field of machine vision. From the biological perspective, it can be said that the signal processed by mechanisms inherent to the retino-cortical pathway might be a suitable (and very probably close to optimal) input for higher cognitive modules of the pattern recognition system. However, the implementation of additional information processing stages, similar to those occurring in the visual cortex, is likely to provide a more beneficial data structure for feature extraction before actual pattern classification is performed. For example, Gabor type functions were proposed as a suitable model for cortical simple cell responses as shown in [23] - [26]. Techniques presented in this dissertation try to incorporate cortical information processing including, but not limited to, approximate modeling of the response of such Gabor-like cortical cells within the context of logarithmic conformal mapping. The ultimate goal of the system development is its practical implementation. The significant body of the work presented in this dissertation addresses theoretical concepts and provides experimental results obtained by utilizing biologically motivated techniques of signal processing as de-facto pre-processing tools in pattern classification tasks. It also attempts to provide detailed insight into the significance and usefulness of these computational tools as well as

revealing their role from a much broader perspective. In other words it treats them as an integral part of the inherently biologically motivated data processing chain capable of deduction processes.

It is reasonable to assume that the evolution of mammalian brain would have resulted in highly efficient usage of natural resources. This would inevitably imply that the human visual system has developed principles of information encoding representing the most effective and most useful mechanisms for representation of natural images. In the well known work of Barlow [27] it is argued that the role of sensory coding is the development of input stream representations that generate factorial code and reduce redundancy. The importance of redundancy reduction is very clearly stated by Olshausen and Field [28], [29] who point out that the information structure present in the world is actually embedded within complex statistical dependencies among photoreceptor activities. According to these researchers, the goal of a biological visual sensory system is to extract intrinsic information from the input signal, or by reducing input redundancy, reveal collections of independent events that actually compose the image. Bell and Sejnowski [30] point out that the problem equivalent to Barlow's problem of redundancy reduction is Independent Component Analysis (ICA). ICA is the signal transformation technique aiming to find the data components that are as statistically independent from each other as possible and is becoming very popular for deployment in pattern recognition systems. Furthermore, same researchers showed that natural images contain statistical regularities suitable for sparse encoding which is tightly related to ICA [31].

One of the questions that remains unanswered is whether principles of redundancy reduction (for example through ICA or sparse coding) can be embedded into

a wider image processing and pattern recognition framework with the front end module performing a retino-cortical type of signal transformation through log-polar mapping and which also incorporates lateral inhibition/excitation directly in the network. At the same time it is necessary to obtain efficient and meaningful (potentially adapted for specific sensor topology) implementation of lateral inhibitory/excitatory mechanisms in log-polar environment. These problems are addressed in detail in the text that follows. The block diagram shown in Figure 1.1 depicts the strategy of information processing studied herein. The block named Retino-Cortical Filter plays the central role in information preprocessing by this system. It represents the stage of signal processing based on a network with recurrent connectivity adapted for the log-polar environment.



Figure 1.1 Biologically inspired computer vision system for pattern analysis; the lateral inhibition/excitation adapted for log-polar environment is implemented through Retino-Cortical Filter.

It must be noted that this research is not aimed towards development of some sort of new pattern classification engine. As it will be presented later in this text, the recognition itself is performed by utilizing well known classification tools. As is the case with any engineering field, the significant effort in this work is made to discover potential usability of the proposed models through solving practical problems such as object recognition.

1.2 Importance of the Problem

Efficient image processing algorithms have always been of great importance for computer vision and pattern recognition applications due to the requirement of real time performance. Reduction of the computational costs while sacrificing very little in signal detail is an acceptable strategy of data processing if reliable decisions can be made for future system actions. Utilization of characteristics of biological systems for solving practical problems inevitably leads towards reducing the gap between manmade machines and live systems. When it comes to image processing and pattern recognition, the aspects of foveation and warping still deserve great attention from academic and industrial viewpoints. This is especially true for the implementation of efficient data processing strategies in such framework(s), as well as recognizing the practical limitations of those techniques which have far reaching consequences in the domain of machine vision and artificial intelligence.

1.3 Specific Project Objectives

The specific project objectives can be now listed:

- 1) The first objective is to obtain the response of the receptor network with pre-determined (not necessarily purely inhibitory, but potentially a combined inhibitory/excitatory inter-neuron coupling model) and correlate it with conventional center-surround (DOG-like) response of the retinal receptive field. It is also desired to describe the response of the neural recurrent grid with a closed form or simple numerical solution.

2) The obtained response of the network with recurrent inhibitory/excitatory connectivity should be incorporated into the model of a retino-cortical filter that essentially performs non-uniform input data processing. The aim is to avoid all spatially non-uniform operations in the linear input space with the exception of the log-polar mapping itself. Instead, adequate spatially uniform signal processing in the cortical domain will be investigated. The desired processing technique should be consistent with, and adapted for the topology of log-polar warping. The main problem is to find out how this filter adaptation can be achieved?

3) One of the objectives is to find and evaluate alternative, network-based models of non-uniform neural coupling, consistent with log-polar mapping. The spatially varying strength of inhibition between receptor fields and the implementation of a microscopic neuron network solution on such a model will be investigated.

4) The front end processing realized through the implementation of biological filters in the warped domain will be succeeded by the cognitive system module responsible for pattern classification. The specific role of the frontal filters based on networks with recurrent lateral connections within the broader context of a signal processing and pattern recognition framework will be addressed. The feature selection/extraction will be performed based on the specific pattern recognition model. The actual pattern classification technique will be based on already available and well known classifiers, i.e. the detailed study of various classification engines potentially deployable in this framework is out of scope of this dissertation.

1.4 Background

Log-polar mapping (complex log mapping) is the well known conformal coordinate transformation where the coordinates from the transformation domain determined by the complex number $z = x + iy$ are mapped into the complex number $l = u + iv$ by applying the following mapping rule:

$$u = \ln \frac{r}{a}, \quad (1.1)$$

$$v = \theta, \quad (1.2)$$

where

$$r = \sqrt{x^2 + y^2} \quad (1.3)$$

$$\theta = \arg(x + iy). \quad (1.4)$$

The pair (x, y) represents the coordinates in the input domain (input image) while the pair (u, v) represents coordinates from the mapped region. This region is often referred to as the log-polar or cortical domain. Constant a represents the minimal radius that is mapped (the logarithmic function is singular at $r=0$). In some instances, the points with radii smaller than a are mapped through a linear transformation function. The log-polar coordinate mapping is shown in Figure 1.2.

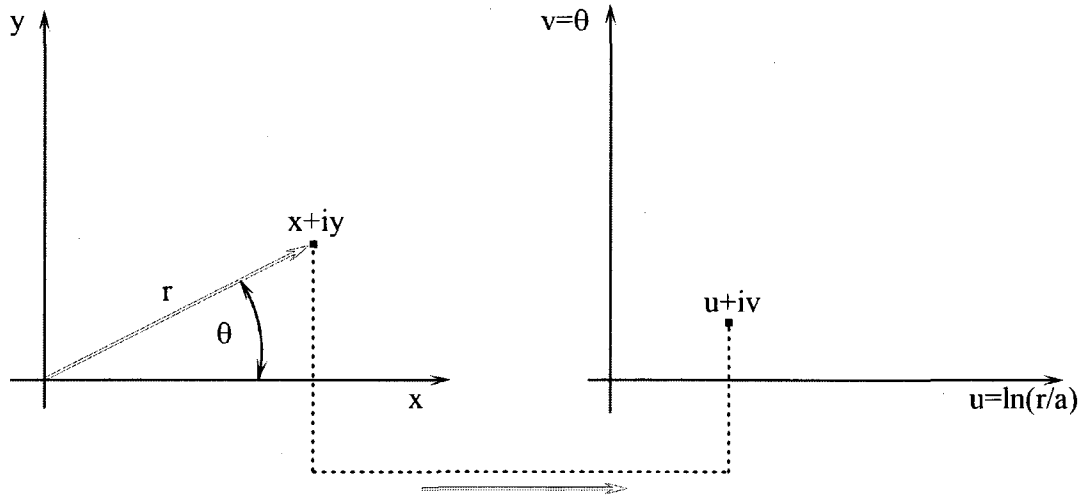


Figure 1.2 Log-polar mapping

In the case of machine vision, the mapping is discrete. The Polar Exponential Grid (PEG) is placed over the input image for non-uniform sampling. Log-polar samples are obtained by applying a specific sampling method (area sampling for example) and the resulting image is formed by rearranging pixels in the new uniform (u, v) space. The PEG radii exponentially increase with the radial index. The discrete version of complex log mapping is shown in Figure 1.3.

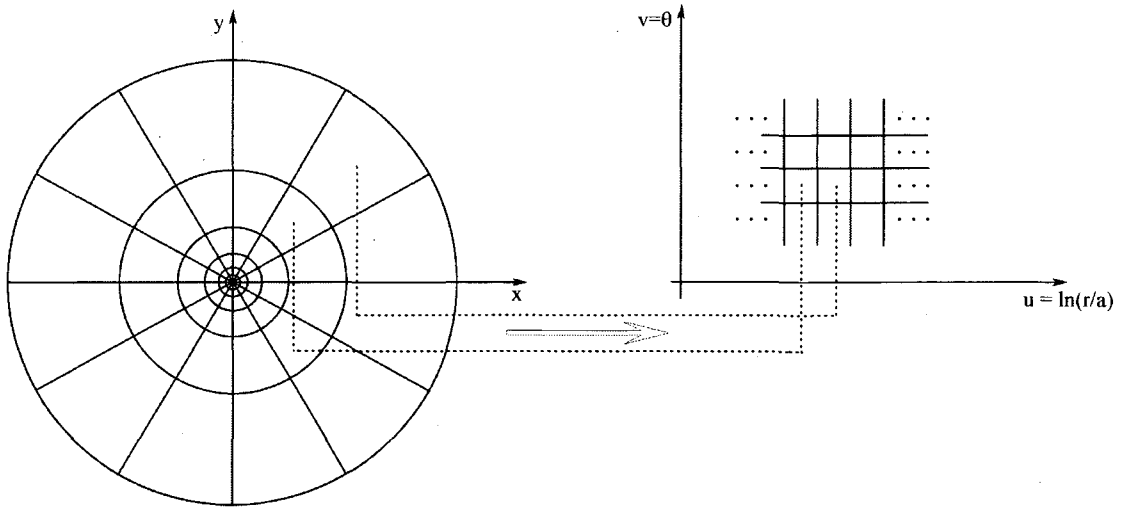


Figure 1.3 Log-polar mapping (discrete implementation)

The usefulness of log-polar mapping comes from the rotation and size/scale invariance achieved in the resulting domain [32] making the transform suitable for pattern recognition. Scaling or rotating the input image causes the output to shift along u and v axis respectively. The resulting data sets are significantly reduced enabling faster and computationally efficient processing.

Concepts of non-uniform retina-like image sampling and log-polar mapping have been widely applied in the area of machine vision and image processing. An extensive review of biologically motivated data reduction models related to log-polar mapping can be found in [2]. Further more, hardware implementations of foveated vision chips including log-polar sensors are reported in [33] and [34]. There are numerous examples of complex-log mapping and foveation-based techniques in image processing and pattern recognition research. Some of this research concentrates on detection of specific shapes such as straight lines and circles [35]. Others have been used to solve problems such as image registration [36] and efficient implementation of edge detection or Hough

transform in a foveated environment [37] and [38]. A comparative analysis of two different feature extraction operators in log-polar context (one model is based on neural network learning capability and the other relies on the mathematical model of object features) presents superiority of neural network-based feature extraction [39]. The non-uniform retinotopic image sampling and complex-log mapping applied for face recognition are demonstrated by [40] and [41]. The applications of tracking, visual attention, license plate reading and video-telephone communications utilizing sensors with logarithmically structured space-variant pixel geometry are shown in [42]. Examples of spacecraft tracking and docking aided by the log-polar warping are presented in [43] and [44]. Log-polar based sensing is also used as the foveation method in applications of distributed surveillance [45].

Non-uniform processing of visual information in the frontal pathway of the HVS is accompanied by lateral inhibition between neuron receptors at different stages of information flow [15] - [19]. The feed-backward form of the lateral subtractive inhibition (LSI), happening at the neural level, can be mathematically modeled by the following 2-D equation for the grid with $N \times N$ receptors:

$$f_{ij} = e_{ij} - \sum_{k=1}^N \sum_{l=1}^N b_{ijkl} f_{kl} . \quad (1.5)$$

where, e_{ij} represents the input over the receptor field at location i, j , f_{ij} , and f_{kl} correspond to output of receptors at location i, j and k, l respectively. The term b_{ijkl} is the coupling coefficient describing the inhibition between receptors at locations i, j and k, l (if self-inhibition or self-excitation is prohibited: $b_{ijkl} = 0$ for $i = j$ and $k = l$). For

systems without adaptive capability, the weighting factor is usually dictated by the distance between receptor fields.

The above equations can be written in the matrix form:

$$F = E - BF \quad (1.6)$$

where F and E are the sensory output and the input respectively presented in matrix form, while B is the coupling coefficient matrix (usually named the network connection matrix). The F and E matrices are actually vector columns and they are given as (example for 3x3 receptor grid):

$$F = [[F_1] [F_2] [F_3]]^T, \quad (1.7)$$

$$E = [[E_1] [E_2] [E_3]]^T, \quad (1.8)$$

where:

$$[F_i] = [f_{i1} f_{i2} f_{i3}], \text{ and } [E_i] = [e_{i1} e_{i2} e_{i3}], i=1,2,3. \quad (1.9)$$

The coupling coefficient matrix B is given as:

$$B = [[B_{ik}]], i=1,2,3; k=1,2,3. \quad (1.10)$$

The matrix B is 9 by 9 matrix (in this example of 3 by 3 receptor grid), where each sub matrix B_{ik} represents the interaction between receptors in the i -th and k -th row. For example, the interaction between receptors in the 1st and 2nd row would be described by:

$$B_{12} = \begin{bmatrix} b_{1121} & b_{1122} & b_{1123} \\ b_{1221} & b_{1222} & b_{1223} \\ b_{1321} & b_{1322} & b_{1323} \end{bmatrix} \quad (1.11)$$

with b_{ijkl} representing the interaction between receptors b_{ij} and b_{kl} .

The final solution to the equation 1.6 is:

$$F = S^{-1}E \quad (1.12)$$

where:

$$S = I + B. \quad (1.13)$$

Matrix I is the identity matrix and it is presumed that S^{-1} exists. If the coupling coefficients depend only on distance between receptors, and the distribution of receptors is uniform in the form of a square grid, the sub-matrices $[B_{ij}]$ will be Toeplitz matrices and the matrix B will be symmetric. Detailed simulations describing the effect of networks with lateral inhibition and addressing fifteen different receptor coupling models can be found in [46].

One of the early biologically inspired models of networks with recurrent connectivity that fitted the experimental data was indeed the negative feedback model [47] and [48]. The microscopic neuron network solution form of a 2-D grid array with subtractive inhibition described by equations 1.6 - 1.13 was presented earlier in [49]. The same paper reveals the connection between network response characterized with LSI and processing of visual information through several spatially tuned band-pass filters. In this approach, the LSI was represented as the process of passing the input image through the set of different DOG filter kernels. Similar results are shown in [50]. Almost identical model of the network solution for the grid with recurrent DOG type connectivity (the model accounting for the inhibition and the excitation) can be found in [51] where the authors characterize the impulse response of the grid as a circular Gabor filter. The uniform implementation of lateral inhibition in a log-polar and pre-mapped linear domains with different computational models of LSI has also been investigated in [52].

Foveation and warping introduce various difficulties to image processing. When considering implementations of a pure lateral inhibitory networks in the context of log-polar mapping like the ones seen in [48] and [52], it is still necessary to answer if the application of uniform, space invariant model of lateral connectivity is consistent with inherently spatially non-uniform coordinate mapping. A significant step towards answering such a question was made by [21] where an example of non-uniform image processing by center-surround DOG kernels in a linear input space that is potentially followed by the log-polar mapping was presented as a model of early stages of biological vision. As presented later in this dissertation, the research described herein utilizes extremely important results of Tabernero et. al. given in [53], who reveal the foveation method based on characterization of non-uniform foveation filters from input linear space as uniform low-pass kernels in the log-polar domain. More precisely, these results demonstrated how the foveation process can be achieved by uniform low-pass filtering in the log-polar domain if the log-polar transform preserves the energy of the original image. This concept is used as a basis for justification of spatially uniform image operations in the cortical space as implemented herein.

Most of the work presented in this dissertation addresses (and is limited to) methods of space invariant signal processing in log-polar domain based on operations performed either by direct spatial domain computations or by finding the system response in frequency domain. However, it is worth noting that besides uniform image processing methods, very common techniques associated with space-variant vision (log-polar architecture considered) are operations performed on retinal image (the result of inverse log-polar mapping is sometimes called retinal image), those utilizing Fourier-Mellin

transform (successive applications of Fourier transform, log-polar mapping and Fourier transform) and image processing based on connectivity graphs [54]. The exponential chirp transform described in [54] introduces shift invariance like that of standard Fourier transform achieved in log-polar space by applying the exponential chirp kernel. However, computing the transform requires additional processing time. The operations based on connectivity graph representing neighbor relations in an arbitrary architecture [55] demand the generation of separate structure (graph) for image manipulation. Similar concept where the pre-computed look-up tables are used on the top of the connectivity graph for space-variant processing is addressed in [38]. Contrary to such computation models, the signal processing approach adapted in this research tries to extract significant signal information from the log-polar image based on solely spatially uniform operators and without generating any additional, potentially demanding data structures.

Previously mentioned findings and techniques related to image processing based on retino-cortical transformation (log-polar mapping and lateral inhibition in particular) can not be fully appreciated unless they are placed in the broader information processing framework. One of the most fascinating findings of different theoretical as well as experimental studies is that neural activity in the primate cortex is based on sparse signal coding [56], [57] (sparse codes are characterized by output units firing very rarely – a very small number of neurons is active at the same time). By imposing the objectives of information preservation and sparseness while learning the neural code, Olshausen and Field showed that such algorithm develops a set of localized, oriented, bandpass receptive fields similar to the receptive fields of simple cells from the primary visual cortex [28],

[31]. The resulting sparse code yields a higher level of statistical independence between its outputs [28].

The idea that the sensory processing in the brain is characterized by the principle of redundancy reduction which implies that input patterns are composed of features that are statistically independent events [27], [58] directly correlates with the relatively novel signal processing technique called Independent Component Analysis (ICA). On the other hand, the direct correlation between redundancy reduction (i.e. statistical independence of transformed signal components) and sparse signal coding was demonstrated by different researchers [30], [31] and [58]. Interestingly enough, the ICA filters of Bell and Sejnowski emerging from the learning algorithm that maximizes information transfer between input and output of the non-linear neural network [30] are very similar to those produced by Olshausen and Field and their sparseness maximization network. The redundancy reduction process itself is possible because most of the real world analog signals have sparse structures and can be represented with few active descriptors [31], [59]. Although being relatively novel technique, ICA is quite well described and many different algorithms exist for achieving the transformation with statistically independent outputs. The reader is pointed to excellent reviews on, and characterization of major ICA techniques by Hyvärinen [58] and Hyvärinen and Oja [60]. The most interesting utilization of ICA from the perspective of this work is feature encoding where the signal components are actually features and the given ICA transform yields coefficients of each feature. Very appealing method for performing the actual ICA is the so-called FastICA or Fixed Point ICA as given in [61]. This algorithm is based on maximization of negentropy (measure of non-Gaussianity) of transformed variable which at the same time

yields the minimization of mutual information between the network outputs. The algorithm is very fast (the convergence is cubic), there are no critical user-defined parameters to choose, it is computationally very simple and not too demanding on memory resources. Additionally, the MATLAB software package for this ICA technique is freely available from [62].

When implementing the ICA algorithms, it is very important that the input data is whitened [58] in order for algorithm to achieve better convergence. The whitening itself means that the covariance matrix (in this research the data are assumed to be zero-mean) is equal to the unity matrix. Another extremely important property of data whitening as the ICA pre-processing step is that the final transformation matrix that reveals the independent signal components (the transformation acting upon whitened data) actually becomes orthogonal. It is quite remarkable that the lateral inhibition as a process inherent to biologically inspired information processing can be utilized for data whitening. The most important and ground breaking work that reveals decorrelating property as a first step towards data whitening in networks with lateral inhibition was presented by Barlow and Földiák in [63]. In that text, the data decorrelation is achieved by symmetric anti-Hebbian learning in the network with recurrent connections without self-inhibition. The algorithm presented in [63] does not necessarily produce whitened outputs. However, Plumbley [64] 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. The explicit details on learning rules utilized in [63] and [64] will be presented in the text that follows.

The lateral inhibition, phenomenon inherent to mammal nervous system plays a very important rule in data whitening – the pre-processing step for ICA - a technique that again has so many flavors of biologically inspired signal encoding. To the best of the knowledge of the author of this text, no research has treated visual information processing from the unified perspective encompassing the retino-cortical transform characterized by the log-polar mapping in conjunction with the lateral inhibition/excitation (i.e. the family of filters arising from networks with recurrent connectivity) and ICA in a broader signal processing framework. It is reasonable to assume that addressing the information processing in warped cortical space through the development of environment specific tools based on neural models of signal whitening might yield an effective platform for practical implementation of ICA for feature encoding and pattern recognition.

After elaborating the techniques of efficient and natural implementation of networks with lateral connectivity in log-polar domain, the theoretical concepts revealing construction of filters arising from such networks is addressed based on the analysis of the network connection matrix spectral properties. Also, novel result revealing the steering property of special subclass of such networks is presented. Finally, the method of learning and development of very efficient biologically inspired kernels based on networks with recurrent connectivity (with lateral inhibition/excitation) and demonstration of their practical usability for signal whitening when acting as pre-processors in pattern recognition systems are addressed.

CHAPTER II

SOLUTIONS OF RECURRENT LINEAR NETWORKS APPLIED IN LOG-POLAR DOMAIN

Before an attempt is made to implement and use biologically inspired principles of signal processing inherent to retino-cortical filter, it is necessary to investigate how LSI (and potentially lateral excitation) and complex-log mapping should be used in order to efficiently complement one another. Obviously, the non-uniform nature of the log-polar mapping raises the question of whether the implementation of inhibitory process (at least from the standpoint of machine, i.e a computer) should be performed prior to, or after the mapping itself. Also, it is not clear how the outcome of the process of lateral inhibition is to be obtained. In other words, what is the adequate, efficient computational model? One of the possible solutions is to implement equations given in previous chapter. This is sometimes referred to as microscopic neuron network solution [48]. However, it is evident that such a model requires direct matrix inversion and for large matrices this can be computationally very demanding and time consuming. This fact becomes more than obvious for large images. For example, an image of size $N \times N$ would require the inversion of a matrix of size $N^2 \times N^2$. This chapter presents how alternative models can be used when solving networks with lateral feed-backward interactions and especially how those solutions relate to log-polar coordinate mapping. It is also argued that spatially non-uniform type of inhibitory network with inhibition decreasing as the distance from the fovea increases does not produce desired signal

processing outputs. The reasoning behind uniform implementation of networks with recurrent connectivity in log-polar domain is also addressed.

2.1 Solutions of Network with LSI via Iterative and Convolution Models

In order to avoid the computation of large matrix inversion when seeking the solution of networks with lateral inhibition, it is possible to use two alternative methods to obtain the solution: the iterative approach based on the algorithm similar to what was previously seen in [52] and the convolution-based method.

The representation of the exact solution for the networks with LSI is represented by equations 1.12 and 1.13. However, it is possible to regard the entire network as a dynamic system where the response of the network is not instantaneous. If it is assumed that forming the final output of the system requires a certain transitional period, then, the current firing rate of the particular neural receptor can be expressed in terms of its input and the output of neighboring inhibitory receptors at current and previous time instances:

$$f_{ij}^n = e_{ij}^n - \sum_{\substack{k=1 \\ k \neq i}}^N \sum_{\substack{l=1 \\ l \neq j}}^N b_{ijkl} f_{kl}^p \quad (2.1)$$

with index n representing the current time instance and p being equal to n or $n-1$, i.e. the most recently obtained value of particular output f_{kl} . It is important to point out that equation 2.1 uses the most recently obtained values of the system output (current time instance) that are available in combination with output at previous time instance. This is different from the model used in [52] which computes the output based solely on the previous time instance values. If the method converges for particular network

connectivity model, it is not necessary to compute the large matrix S . Thus, the matrix inversion is avoided. In order to validate this iterative model described by equation 2.1, several inter-neuron coupling models for simulations were used. Here, only two of them are presented. Assuming that d is the distance between neurons (i,j) and (k,l) , the model A is represented by coupling:

$$b_{ijkl} = 0.25 \frac{1}{d^3} \quad (2.2)$$

while the model B is represented by:

$$b_{ijkl} = e^{-1.55d}. \quad (2.3)$$

Equations 2.2 and 2.3 represent such inhibitory coupling models where inter-neuron interactions are almost non existent when receptors are far apart. This introduces an additional advantage when implementing the iterative algorithm. The summation in equation 2.1 can be performed over the limited, inhibitory significant region. Figure 2.1 depicts the agreement between the exact network and the iterative solution for model B coupling. The visual appearance of the resulting images is identical for all discussed methods for solving the recurrent network - one more method will be described in the text that follows. This is why only one resulting image is displayed. Similar results were obtained for model A inter-neuron coupling. Normalized root mean squared error (NRMSE) parameter for an $N \times N$ image is calculated as:

$$NRMSE = \sqrt{\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \frac{(x(i,j) - y(i,j))^2}{y(i,j)^2}} \quad (2.4)$$

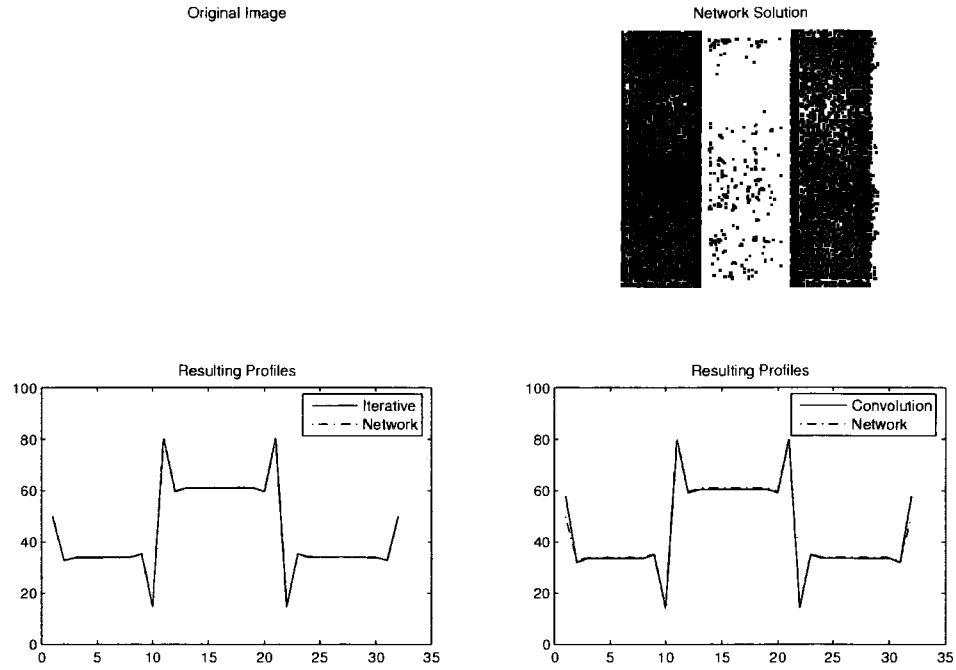


Figure 2.1 The result of the network, the iterative and the convolution-based solutions for the 2-D grid with lateral inhibition; original image (*upper left*), resulting image (*upper right*); Resulting signal profiles: network versus iterative solution (*bottom left*), network versus convolution model (*bottom right*); Normalized Root Mean Squared Error between resulting signals: $NRMSE_{netw,iter} = 0.0038$, and $NRMSE_{netw,conv} = 0.0589$; Image size used is 32 x 32

The iterative approach does not require the inversion of large matrices, but several passes through the computational space are necessary in order for the solution to properly converge. The process is interrupted once the NRMSE between successive results is less than a predefined threshold. Since the computational demands of this iterative model are still significant, the alternative, convolution-based solution for the grid with lateral inhibition was explored.

The network solution presented by equation 1.12 reflects the linear system characteristics. In order to compute the impulse response of the grid with lateral inhibition, the system can be excited with the unit pulse and either the direct network or the iterative solution models presented earlier can be used to obtain the output. It turns

out that the impulse response, when a suitable inter-neuron coupling model is used, can be efficiently represented by kernels of fairly small size. Instead of using the computationally demanding methods, it is possible to obtain the response of the network to any input by performing a single convolution operation with already determined kernel. Figure 2.1 depicts the result of the convolution-based lateral inhibition where the model B coupling was used and kernel size was chosen as 5x5. The increase in system size does not influence the resulting kernel. This was confirmed when the response was calculated for 32x32, 40x40 and 48x48 2-D systems for both inter-neuron coupling models. The most significant coefficients of Model A and model B kernels are given in Figure 2.2.

Model A					Model B				
0.0011085	-0.011144	0.048301	-0.011144	0.0011085	0.0011361	0.0036029	0.0068373	0.0036029	0.0011361
-0.011144	0.040454	-0.2909	0.040454	-0.011144	0.0036029	-0.041248	-0.1828	-0.041248	0.0036029
0.048301	-0.2909	1.2729	-0.2909	0.048301	0.0068373	-0.1828	1.1714	-0.1828	0.0068373
-0.011144	0.040454	-0.2909	0.040454	-0.011144	0.0036029	-0.041248	-0.1828	-0.041248	0.0036029
0.0011085	-0.011144	0.048301	-0.011144	0.0011085	0.0011361	0.0036029	0.0068373	0.0036029	0.0011361

Figure 2.2 The convolution kernels for model A (*left*) and model B (*right*)

These filters represent high-frequency detectors. The output edge information is enhanced while preserving the relative luminance levels in the image. If the organization of the 2-D grid with lateral inhibition is such that all neighboring receptors are inhibitory, the whole network will act as the filter tuned for high-frequency detection and enhancement. If the inter-neuron coupling model is for example an inverted Difference of Gaussian function, immediate neighbors of a particular receptor have excitatory role while more distant neurons act as inhibitors. It was previously claimed that the impulse response of this network was similar to a circular Gabor filter [51]. These filters have a

characteristic center-surround shape with decaying "lobes" of response at the periphery. It is interesting that the outcome of lateral inhibition was earlier represented as performing filtering by multiple band-pass DOG filters [48] and [50]. Actually, the circular Gabor-like kernels can be constructed as linear combination of more Gaussian filters (or DOG kernels). The example is shown in Figure 2.3 where four Gaussian (or two DOG) kernels are used. Extremely reducing the variance of Gaussian functions constituting the filter yields the edge detector similar to that presented in Figure 2.2. This directly correlates with the result described by Messner [48] and Messner and Szu [50] since their work interprets lateral inhibitory process with series of DOG bandpass filters only after approximating one of the Gaussian kernels as a unit pulse which is the extreme case of reducing the Gaussian function variance.

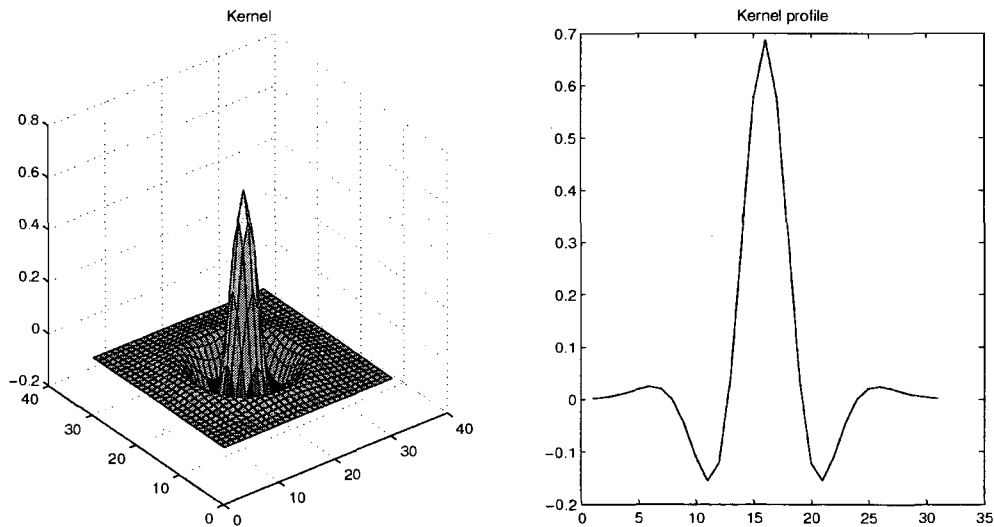


Figure 2.3 Linear combination of four Gaussian kernels (*left*); corresponding profile (*right*)

From this discussion we can see that the process of lateral inhibition can be performed through the convolution of the input with kernels similar in form to the filters from Figure 2.2. Additionally, two representations of the network impulse response as given previously by different authors in the form of circular Gabor-like filters or series of DOG filters are shown to correlate well with each other.

2.2 Convolution in Log-Polar Domain

As previously stated, log-polar mapping representing an approximation for the rearrangement of retinal signals in the visual cortex is characterized by spatial non-uniformity. When applying specific processing tasks in such an environment it is important to obtain a natural and efficient way of data manipulation. As discussed in the previous section, the actions of networks with lateral interactions can indeed be represented in terms of direct convolution. If one is to be consistent with, or at least guided by biological principles of visual information processing, the non-uniform nature of log-polar transformation has to be accompanied by the processing kernels that are spatially variant and increase in size with distance from the fovea. This yields direct analogy with the structure and actions of retinal ganglion cell receptive fields.

Conventional convolution with spatially variant filters is computationally very demanding. Interestingly, for the continuous domain and a special class of low-pass filters, it was shown that the spatially non-uniform filtering resulting in foveation can be achieved by uniform low-pass filtering in log-polar space if the mapping preserves the energy of the original signal [53]. This result is extremely important because it points out one more property of the warped computational space when a complex-log function is

used for the mapping method. The log-polar computations reflecting non-uniform operations in the input domain are performed not only in a smaller space, which preserves the details from the center of transformation, but additionally, they can potentially be done by using uniform filters. If circular, isotropic kernels are space-scaled linearly with their distance from the log-polar mapping origin (consistent with the structure of the PEG, the diameter of the receptive field is increased linearly with distance from the fovea), the mapped (warped) filters will remain spatially uniform. This fact can be used to uniformly process the signal in the mapped (i.e. cortical) domain and achieve a result similar to the one obtained by non-uniform filtering in the pre-mapped space followed by the log-polar warping. The 1-D analogy of space-scaled kernels and their mapped counterparts is shown in Figure 2.4. In this case, the log mapping is applied to space-variant DOG function and a uniform function is obtained in the resulting domain.

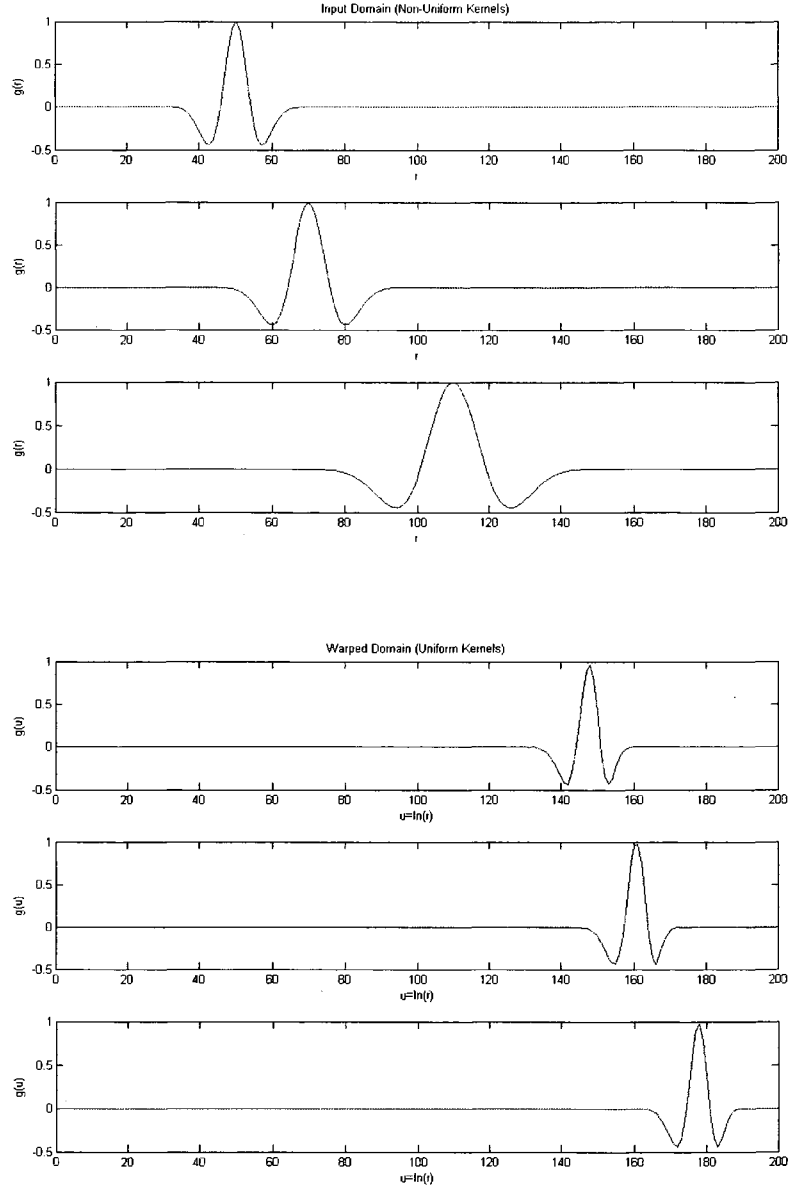


Figure 2.4 Linearly space-scaled kernels and their mapped versions - kernel size is increasing with distance from mapping origin

Assuming the computation is performed in continuum, the convolution process in log-polar domain can be correlated to the convolution in pre-mapped space as follows.

The inverse of the complex-log transform (assuming $a = 1$) is given as:

$$x = e^u \cos(v) \quad (2.5)$$

$$y = e^u \sin(v) \quad (2.6)$$

The convolution by kernel $g(x, y)$ in pre-mapped space is defined as:

$$t(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\alpha, \beta) g(x - \alpha, y - \beta) d\alpha d\beta \quad (2.7)$$

while the convolution in mapped space is defined as:

$$t^*(u, v) = \int_{-\infty}^{\infty} \int_0^{2\pi} h^*(u_\alpha, v_\beta) g^*(u - u_\alpha, v - v_\beta) du_\alpha dv_\beta \quad (2.8)$$

where $(x, y) \leftrightarrow (u, v)$ and $(\alpha, \beta) \leftrightarrow (u_\alpha, v_\beta)$ represent coordinate mapping pairs. It can be shown that for any coordinate pair, the convolution in the pre-mapped domain is identical to convolution in the log-polar space, i.e. $t(x, y) = t^*(u, v)$ if the kernel satisfies (see also [53] for low-pass version of filtering):

$$g^*(u - u_\alpha, v - v_\beta) = g(e^u \cos v - e^{u_\alpha} \cos v_\beta, e^u \sin v - e^{u_\alpha} \sin v_\beta) \quad (2.9)$$

and the 2-D signal is mapped as:

$$h^*(u_\alpha, v_\beta) = e^{2u_\alpha} h(e^{u_\alpha} \cos v_\beta, e^{u_\alpha} \sin v_\beta). \quad (2.10)$$

Equation 2.10 shows that the warped image must be modified by the factor e^{2u_α} (i.e., transformation Jacobian) in order to force the preservation of the convolved signal energy. Alternatively, it is possible to consider applying the energy preservation factor to the kernel itself instead to the warped signal.

Although the implementation of this basic 2-D calculus while computing integrals within coordinate system change reveals techniques to be used for signal processing in the cortical domain, an important limitation remains to be considered. For a large class of signals and filters one needs extremely high sampling rates even at the periphery of the computational space in order to be able to appropriately represent warped filters without

significant signal distortion. Using an actual network with lateral connectivity as shown in Figure 2.2 proves that the warping of such discrete filters is not even practically plausible. Even for filters given as linear combination of multiple Gaussian filters, the resulting log-polar space necessary to represent warped versions of these functions would often be too large to achieve effective computational results. That is why throughout this dissertation it is assumed that the loss of data due to coarse sampling of the PEG is an acceptable strategy while direct implementation of convolution as shown in equations 2.7 – 2.10 is not practical. However, it is possible to show that implementation of space invariant recurrent network (possibly through convolution with center-surround type filters emerging as solution to such networks) in the log-polar discrete space is equivalent to sampling the input with space-variant fields/kernels of center-surround structure. Before elaborating on this concept, the next section investigates one more possibility of implementation of non-uniform lateral inhibition in cortical space.

2.3 Non-Uniform Lateral Inhibition - the Exact Network Solution

In order to achieve spatially non-uniform lateral inhibition the model where the inter-neuron coupling coefficients depend on distances between samples in the PEG is investigated. The actual processing is performed in the uniform log-polar (cortical) space. Receptors positioned near the fovea are close to each other and they are strongly influenced by their neighboring inhibitors. The inhibition between neighboring receptors becomes weaker as their distance from the center of the mapping is increased. Figure 2.5 depicts how the distance for calculation of inter-neuron coupling coefficients as given by

equations 2.2 and 2.3 is calculated. Parameters r_0 and θ represent radial and angular sampling units of PEG.

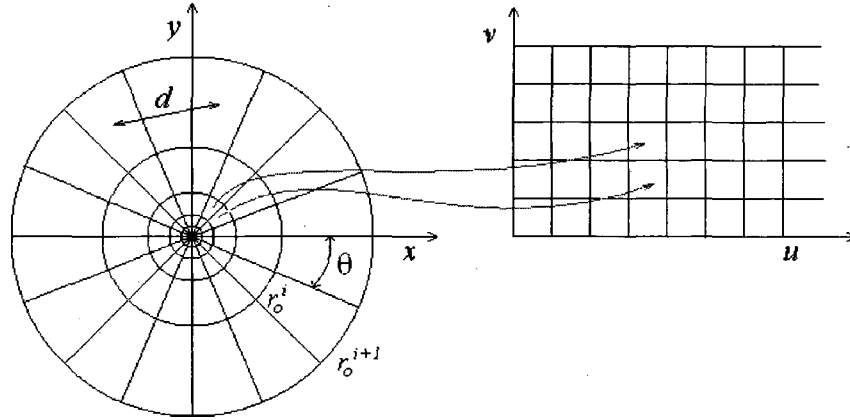


Figure 2.5 Distance parameter used for calculation of coupling coefficients in space-variant inhibition model;

The experiments show that the presented model of non-uniform inhibition results in edge enhancement localized in the region near the fovea. The peripheral information content is not affected at all by the inhibition. It is interesting that only the coupling model B gave the edge enhancement result. When using exponential coupling models (i.e. the model where the coefficients are given as $b_{ijmn} = e^{-kd}$), it is possible to modulate the inhibition extent by changing the parameter k . The resulting signal processed by the network with the non-uniform inhibition is shown in Figure 2.6 (the exact network solution is implemented).



Figure 2.6 Non-uniform lateral inhibition in log-polar space; original image (*left*) and processed image (*right*); Model B is used

For this model of lateral inhibition, the overall image intensity adjustment was made in order to normalize the DC gain to unity. It was obtained by processing the DC signal and calculating the per-element ratio between the original and resulting signal intensities. A similar process of intensity adjustment is described in [46]. The adjustment factor is not spatially uniform and it depends on the radial distance from the fovea. Strong artifacts are present in the fovea (notice left edge of the processed image in Figure 2.6) and they might be due to the very significant coupling combined with the absence of neighboring neurons. This effect is not present at the periphery since the inhibition in that region is practically non-existent.

The inability of the addressed network to process and affect the peripheral image information does not correspond well with the notion of spatially variant receptive fields of the HVS present throughout the entire processing domain. Also, from the standpoint of machine vision peripheral information can be very significant – the system for example

has to focus on new events registered in the visual field and very likely they happen at the periphery. That is why the presented model of spatially non-uniform lateral inhibition was rejected from further consideration in the biologically inspired signal processing framework described herein.

2.4 Characterization of the Space Invariant Network with Lateral Connectivity in Log-Polar Space

The convolution operation in the log-polar domain representing the counterpart to the convolution in the input pre-mapped space is described by equations 2.7 – 2.10. However, the exact solution of networks with lateral connectivity as described by the microscopic neuron network solution or alternative models in the form of iterative or convolution-based algorithms yields filters that are practically either impossible or at the very least very difficult to properly warp. The mapping of these filters is questionable since they exhibit center-surround properties on the level of adjacent filtering coefficients. A natural possibility for how actions of such networks might be implemented in effective manner is in the application of direct convolution with a pre-computed network impulse response in the mapped domain. Practical implementations of the log-polar transformation require a sampling grid (like PEG) effectively containing non-uniform receptive fields that are centered at each grid node and appropriately amplitude normalized while performing the low-pass filtering. This means that the first stage of warping itself contains spatially variant filtering that can be executed in a fast manner through the implementation of look-up tables. The low-pass filtering models do not have to be limited to simple area averaging – they merely reflect

the type of the receptive field chosen. In other words, one can for example use a Gaussian function as a mask to determine the resulting value of a particular pixel in log-polar space. These receptive fields can overlap. The construction of sampling structures in the log-polar environment along with implementation of different types of receptive field functions is well summarized in [2].

In the case of log-polar mapping, if one implements the convolution in the mapped cortical space with a discrete kernel of center-surround type (like those from Figure 2.2) the resulting operation as seen in input pre-mapped space is effectively the implementation of non-uniform processing with receptive fields affected by the function chosen for sampling of the input signal (e.g. Gaussian kernel). The resulting, non-uniform kernels have also a center-surround property since each low-pass type PEG receptive field is multiplied by the adequate coefficient from the kernel representing the impulse response of the network with lateral connectivity. The overall responses of such filters can effectively be described as difference of offset Gaussian (DOOG) functions where each Gaussian can have different spatial properties. In order to visualize the resulting receptive field, the local portion of the PEG containing several receptive fields can be closely approximated with a uniform square grid if the radial sampling unit r_0 is small enough (close to unity) and the angular sampling unit θ is chosen such that the sampling areas from Figure 2.5 have the shape that is as close as possible to a square. In such a case the cumulative response of the group of nine overlapping Gaussian receptive fields effectively forming DOOG function with multiplying coefficients given by the nine most significant values in Model B kernel from Figure 2.2 is shown in Figure 2.7.

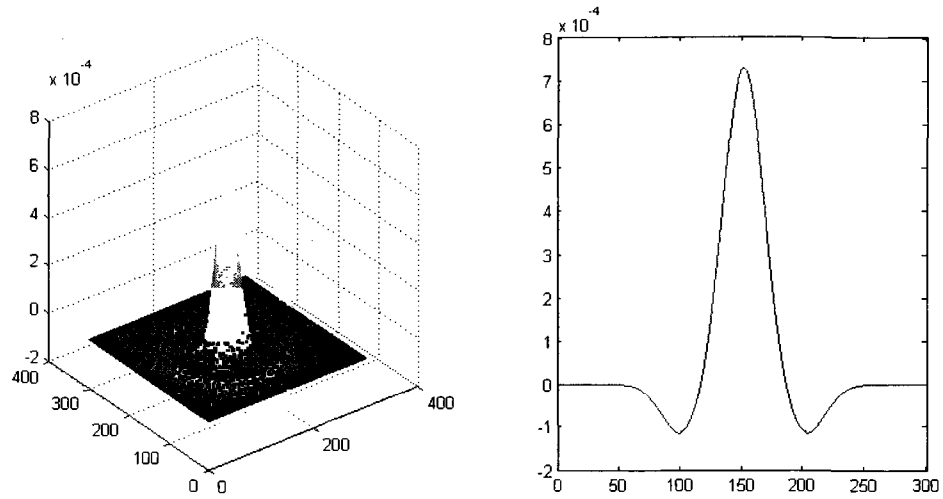


Figure 2.7 Resulting receptive field formed as a linear combination of nine Gaussian functions (with spatial offset) with each function multiplied by corresponding coefficient from Model B kernel shown in Figure 2. 2 (left); Resulting field profile (right);

The profile of the filter demonstrates the center-surround characteristic. The spatial extent as well as the amplitude of the filter is location dependent, i.e. non-uniform. Note that the given model directly incorporates the nature of the convolution kernel representing the action of the network with lateral connectivity as well as the averaging receptive field inherent to log-polar mapping. This way, the antagonistic center-surround property of the resulting function in the form of DOOG filter is associated directly with the network with lateral inhibition. In contrast to the networks with strictly Gaussian inter-neuron coupling yielding DOG filters as described in [48] and [50], the presented model reveals that far broader classes of networks with lateral inhibition in combination with low-pass filters inherent to complex-log mapping yield overall responses similar to those observed in frontal stages of biological vision (e.g. DOOG type receptive fields [65]). An important property of the inhibitory network impulse response that yields such

DOOG-like behavior of overall receptive field is the center-surround type characteristic as shown in Figure 2.2.

Based on the discussion presented in this chapter, one can conclude that the implementation of spatially uniform operations reflecting the response of lateral inhibitory networks in the log polar environment is an adequate strategy for implementation of lateral inhibition in cortical space. When analyzed from the perspective of being in line with biologically motivated principles of signal processing, the model itself produces non-uniform center-surround receptive fields similar to those found in the frontal stages of mammalian vision systems. This concept is also in agreement with the notion that the “exact” uniform convolution in the log-polar domain is equivalent to the spatially non-uniform filtering in input space as described in [53]. From a practical standpoint, the action of the network with recurrent connectivity can be applied either through convolution, the iterative model or by direct network solution. Which method is to be implemented is naturally dependent on the problem to be solved. In the case of processing large images, convolution is more practical. However, for small size networks, direct network solution can be easily utilized.

CHAPTER III

CONSTRUCTION OF BIOLOGICALLY INSPIRED FILTERS UTILIZING SPECTRAL PROPERTIES OF TOEPLITZ-BLOCK-TOEPLITZ MATRICES

This chapter addresses the construction of filters arising from neural networks with feed-backward connections accounting for lateral inhibition and excitation. It provides important theoretical results revealing how to choose the inter-neuron coupling model so that the network solution can be obtained through Taylor matrix series expansion (this expansion can also be interpreted as a feed-forward filter) without computing a potentially large matrix inverse. A novel, steering property of specific class of network filters is also demonstrated.

3.1 Remarks on Recurrent Linear Networks and Toeplitz-Block-Toeplitz Systems

As mentioned in previous text, the essential motivation for studying and deploying networks with lateral connectivity comes from more or less simplified models of biological vision. The well known example is the Hartline-Ratliff equation [16] which describes the mutual inhibitory influences of visual receptor units of the limulus. The discussion presented herein is restricted to the simplified, linear model of the limulus equation and its steady state solution as described by:

$$\mathbf{y} = (\mathbf{I} + \mathbf{B})^{-1} \mathbf{x} . \quad (3.1)$$

This relation essentially summarizes equations 1.13 and 1.14 with modified notation for input and output. Vectors \mathbf{x} and \mathbf{y} from equation 3.1 are time invariant input and system output respectively while \mathbf{B} represents the network connection matrix. Gutkin and Smith [67] studied the dynamics of linearized recurrent networks in the presence of additive noise and pointed out that the asymptotically attractive steady state in the mean described by equation 3.1 is achieved if all eigenvalues of the connection matrix \mathbf{B} have real parts greater than -1. The linearity between input and output as well as the contrast enhancement effects produced by the 1-D network with lateral inhibition were demonstrated by Furman [47]. Following Furman's work, the 2-D model of such a network was addressed by Messner [48] and later used by Szu and Messner [50] when deriving multiple-channel novelty filters of associative memory. In [48] and [50] the same authors suggested a strong correlation between actions of networks with recurrent Gaussian connectivity and spatial frequency theory of early visual processing [68]. The authors of [52] used the iterative method to solve the network characterized by 3.1 without identifying the proper network weights that would guarantee convergence of such solution. The linear model of recurrent inhibition with DOG inter-neuron coupling was used to account for a Mach band effect, Herman grid illusion and White's effect [51]. Similar model effectively explained the modified Poggendorff illusion [69]. The network used in [51] yields a simple feed forward filter resulting from the matrix inverse operation. In this chapter it is demonstrated how to obtain such filters without explicitly inverting the usually large matrix $\mathbf{I} + \mathbf{B}$.

Now, consider networks with a spatial distribution of sensory units producing the system characterized by the TBT (Toeplitz-Block-Toeplitz) matrix, i.e. the matrix with

block Toeplitz structure where each block is of Toeplitz type. The network response can be found by using specialized algorithms for such systems. Some of those techniques utilize the persymmetry property of the TBT matrix [70], [71] and in the case of large matrices can still be computationally demanding. The solution of a TBT system reformulated as a 2-D deconvolution seen in [72] requires the smoothness of the TBT matrix entries. If the inter-neuron coupling is spatially limited or is extremely small for mutually distant units, the connection matrix becomes block banded with banded blocks and can be solved by techniques presented in [73] and [74]. Alternatively, the convolution with a kernel computed iteratively from the coupling model can be used to obtain a close approximation of the network output. Solving the system by this method can reveal important characteristics of the network impulse response filter and as such is implemented herein. In order to achieve the convergent result, the estimation of spectral properties of TBT matrix becomes critical.

The continuous function $f : [-\pi, \pi]^2 \rightarrow R$, generates the TBT matrix $\mathbf{B}_{n,m}$ of structure:

$$\mathbf{B}_{n,m} = \begin{bmatrix} \mathbf{B}_0 & \mathbf{B}_1 & \cdots & \mathbf{B}_{n-1} \\ \mathbf{B}_{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{B}_1 \\ \mathbf{B}_{-n+1} & \cdots & \mathbf{B}_{-1} & \mathbf{B}_0 \end{bmatrix} \quad (3.2)$$

with Toeplitz blocks :

$$\mathbf{B}_l = \begin{bmatrix} b_{l,0} & b_{l,1} & \cdots & b_{l,m-1} \\ b_{l,-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & b_{l,1} \\ b_{l,-m+1} & \cdots & b_{l,-1} & b_{l,0} \end{bmatrix} \quad (3.3)$$

if:

$$b_{p,q} = \frac{1}{4\pi^2} \int_{[-\pi,\pi]^2} f(x,y) e^{-j(px+qy)} dx dy, \text{ where } p,q = 0, \pm 1, \pm 2, \dots \quad (3.4)$$

The relationship between the spectra of Toeplitz matrix and its generating function was introduced by Grenader and Szego [75]. Following their work, Serra [76] and Tilli [77] showed that the interval containing the eigenvalues of the TBT matrix is closely related to the properties of the matrix generating function. More precisely, Serra [76] states that for the function f continuous on the interval $[-\pi, \pi]^2$, and not identically constant, for any n and m , all the eigenvalues of $\mathbf{B}_{n,m}$ lie in the interval $(\min_{[-\pi,\pi]^2} f, \max_{[-\pi,\pi]^2} f)$.

This allows one to characterize the spectral properties of the TBT matrix $\mathbf{B} = \mathbf{B}_{n,m}$ based on the corresponding generating function. Such characterization can be used when choosing the neural network inter-neuron coupling functions.

In [48] and [50] the microscopic neuron network solution for the 2-D receptor grid with recurrent lateral inhibition was found by approximating the inverse of the following matrix by a convergent power series:

$$(\mathbf{I} + \mathbf{B})^{-1} = \mathbf{I} - \mathbf{B} + \mathbf{B}^2 - \mathbf{B}^3 + \dots \quad (3.5)$$

Identifying a proper connection matrix \mathbf{B} guaranteeing convergence of this series is essential for valid calculation.

The main theoretical concepts presented in the following sections of this chapter are as follows. Based on conditions for convergence of the matrix power series in equation 3.5 and the spectral properties of TBT matrices, the method for construction of inter-neuron connection models is identified and applied to obtain parameters of five different coupling functions. Once the convergence of 3.5 is guaranteed, the equivalent filter representing the network impulse response is pre-computed explicitly based solely

on the connection model. Finally, it is shown that for the inter-neuron coupling in the form of a function expandable in a Fourier series in polar angle, the network behaves as a steerable [78] filter.

3.2 Construction and Efficient Solution of Recurrent Neural Network Characterized by TBT Matrix

The simplified form of Hartline-Ratliff equation for the $n \times m$ 2-D system case with solution given by equation 3.1 can be written in the form:

$$\mathbf{y} = \mathbf{x} - \mathbf{B}\mathbf{y} \quad (3.6)$$

where \mathbf{x} and \mathbf{y} are $nm \times 1$ vectors representing input and output of the system and \mathbf{B} is the $nm \times nm$ connection matrix. Note that in the case of input in the form of an $n \times m$ image, the vector \mathbf{x} from equation 3.6 is formed by sequentially aligning rows of the input image into single vector column. Discussion presented in this dissertation considers the network with a uniform distribution of receptor-neurons in a 2-D square grid with the structure as presented in Figure 3.1.

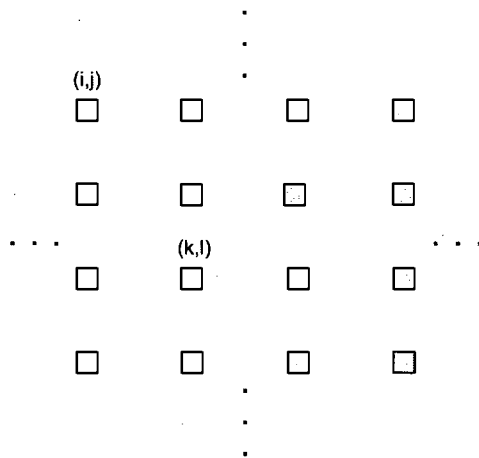


Figure 3.1 Uniform receptor-neuron grid

It is assumed that the lateral interaction between units is in the form of additive/subtractive feed-backward connections yielding the recurrent network. The basic model of such interaction between two arbitrary sensory units is shown in Figure 3.2.

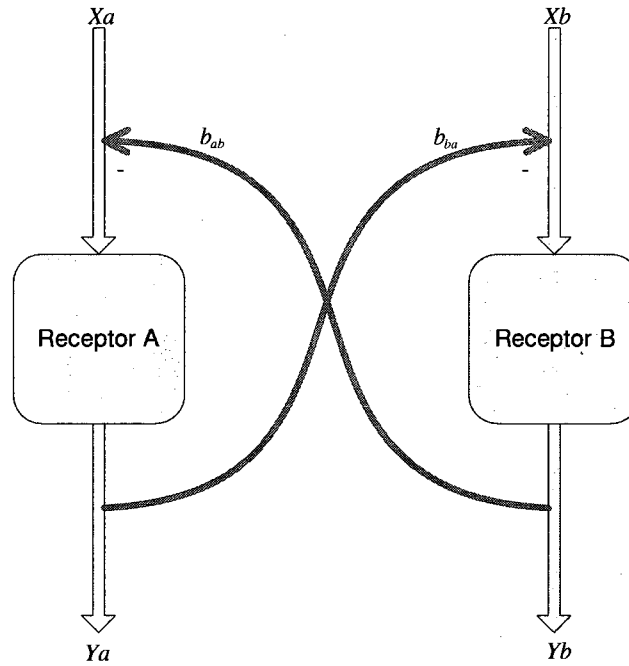


Figure 3.2 Excitatory/inhibitory feed-backward interaction between two arbitrary sensory units

The equations governing the model presented in Figure 3.2 are:

$$\begin{aligned} Y_a &= X_a - b_{ab} Y_b \\ Y_b &= X_b - b_{ba} Y_a \end{aligned} \tag{3.7}$$

where coefficients b_{ab} and b_{ba} governing the amount of inhibition/excitation can take both positive and negative values. When the equations from 3.7 are expanded to a 2-D model, the resulting representation in matrix form is given by equation 3.6 with coupling coefficients being elements of the matrix **B**. For the spatially invariant inter-neuron

coupling and sensory distribution as given in Figure 3.1, the connection matrix is of TBT type. Furthermore, if the coupling depends only on distances between receptors, the TBT connection matrix is symmetric with symmetric blocks. Following the notation from (3.2) and (3.3), each block \mathbf{B}_l of this matrix represents interactions between units in rows i and k so that $l = k - i$.

The TBT form of matrix \mathbf{B} can efficiently be utilized when solving the system defined by equation 3.6. If all the eigenvalues of such a matrix satisfy the condition $|\lambda_i| < 1$, the series given in 3.5 converges [79]. When choosing the inter-neuron coupling model for the receptor grid with recurrent lateral inhibition-excitation, the direct implementation of Serra's result can ensure adequate spectral properties of the TBT connection matrix and allow for the system solution to be found by approximating the matrix inverse in the form of 3.5. After inspecting the equation 3.3 more closely, it becomes evident that entries of every sub-matrix \mathbf{B}_l are taken along the second frequency coordinate, while keeping the first one constant. Throughout the remainder of this dissertation it is assumed that when forming the matrix \mathbf{B} , the sampling in Fourier domain is actually undertaken after the function f is rotated by 90 degrees yielding the new definition for the generating function of the TBT matrix: the continuous function $f : [-\pi, \pi]^2 \rightarrow R$ generates the TBT matrix $\mathbf{B}_{n,m}$ of structure given in equation 3.2 and 3.3 if

$$b_{p,q} = \frac{1}{4\pi^2} \int_{[-\pi,\pi]^2} f^{90^\circ}(x,y) e^{-j(px+qy)} dx dy, \text{ where } p,q = 0, \pm 1, \pm 2, \dots \quad (3.8)$$

In the previous equation, $f^{90^\circ}(x, y)$ represents the rotation of the function $f(x, y)$ by 90 degrees about the origin. Note that the 2-D convolution can be approximated by multiplying the 1-D vector of size $nm \times 1$ by the TBT matrix of size $nm \times nm$ where the 1-D vector is obtained by serializing the input image or 2-D signal into a single column vector [48]. Keeping this in mind, one can see that the reason for the “redefinition” of the generating function is the fact that the multiplication by the TBT matrix generated according to equation 3.8 closely approximates the 2-D convolution with the discrete kernel $b(m, n)$ (assuming this kernel is center-symmetric) formed by sampling the Fourier transform of the (original, not rotated) function f in the form:

$$b(m, n) = \frac{1}{4\pi^2} \int_{[-\pi, \pi]^2} f(x, y) e^{-j(mx+ny)} dx dy, \text{ where } m, n = 0, \pm 1, \pm 2, \dots \quad (3.9)$$

That is why, for the purposes of the work presented here, it is also said that the function f generates the discrete function $b(m, n)$ (or as later referred in the text connection function) if equation 3.9 is satisfied.

Proposition 3.1 (Direct interpretation of Serra’s result for TBT matrices):

If the connection matrix \mathbf{B} from equation 3.6 is generated by the function f that is continuous and not identically constant on $[-\pi, \pi]^2$, and if:

$$\max_{[-\pi, \pi]^2} |f(x, y)| \leq 1 \quad (3.10)$$

the series 3.5 will converge and the system response can be written in the form:

$$\mathbf{y} = (\mathbf{I} - \mathbf{B} + \mathbf{B}^2 - \mathbf{B}^3 + \dots) \mathbf{x}. \quad (3.11)$$

This proposition practically defines the method for construction of the network that can be solved without explicit matrix inversion. Provided convergence is guaranteed, the system response can be realized by the sum of finite number of filters constituting the series 3.5 [48]. The resulting filter can be pre-computed before actual processing takes place. Based on the rotation property of the Fourier transform [80] connection weights from 3.8 can be related to the convolution kernel $b(m, n)$ as:

$$b_{m,n} = b(n, -m). \quad (3.12)$$

The overall impulse response of the network is given as:

$$h_i(m, n) = \delta(m, n) + \sum_{i=1}^N (-1)^n b^{(i)}(m, n) \quad (3.13)$$

where $b^{(i)}(m, n)$ is the discrete connection function convolved with itself $i - 1$ times and $\delta(m, n)$ representing the 2-D unit impulse. The size of kernel $b(m, n)$ can be quite small if the extent of the inter-neuron coupling is spatially limited. This is in agreement with all practical models addressed herein. As demonstrated in the subsequent text, the choice of generating functions with absolute values of their extremes significantly less than unity improves the speed of convergence and requires fewer components in the series 3.13 to be used.

3.3 Construction of Connection Matrix (Five Practical Models)

This section focuses on construction of the recurrent network with specific, *a priori* defined inter-neuron coupling models. Such models are inspired by functions commonly used to describe receptive fields at various stages of biological vision [65].

The Gaussian kernel as defined herein can be represented in the spatial and frequency domains as follows:

$$g_{\alpha,\sigma}(x, y) = \alpha e^{-\frac{x^2+y^2}{2\sigma^2}} \xrightarrow{F} G_{\alpha,\sigma}(u, v) = \frac{2\pi\alpha}{\sigma^2} e^{-\frac{u^2+v^2}{2\sigma^2}}. \quad (3.14)$$

Consistent with equations 3.1 and 3.8, the generating function is denoted as $f(x, y)$ while the corresponding connection matrix is labeled as \mathbf{B} . This matrix is formed by adequately sampling the Fourier transform of the generating function also named connection function. From this point on, it is assumed that the generating function is defined on the entire R^2 domain, i.e. $f: R^2 \rightarrow R$, but is also narrow-width, with values outside the interval $[-\pi, \pi]^2$ practically close to zero. When this is the case, based on 3.12 the weights of the network can be approximated as:

$$b_{-n,m} = b(m, n) = b(u, v) \Big|_{\substack{u=mU \\ v=nV}} = \frac{1}{4\pi^2} \int_{(-\infty, \infty)^2} f(x, y) e^{-j(ux+vy)} dx dy \Big|_{\substack{u=mU \\ v=nV}} \quad (3.15)$$

where $U = V = 1$ and $m, n = 0, \pm 1, \pm 2, \dots$. Note that while integrating on $[-\pi, \pi]^2$ the integral from 3.15 integrates the same function as that given in 3.9. In all cases except for the DOG model it is assumed that parameter $\alpha > 0$. Since the matrix \mathbf{B} is real, it is necessary that generating functions have real Fourier transforms.

3.3.1 Gaussian model

When dealing with Gaussian coupling, cases with and without self-inhibition are studied. The generating function $f(x, y)$ and the corresponding inter-neuron coupling are given as:

$$f(x, y) = G_{\alpha,\sigma}(x, y) \xrightarrow{\text{Generate}} b(u, v) = g_{\alpha,\sigma}(u, v) \quad (3.16)$$

where, operation ‘*Generate*’ is defined by Eq. 3.15. For the network with self-inhibition, the condition for convergence of the series given in 3.5 becomes:

$$\frac{2\pi\alpha}{\sigma^2} \leq 1 \quad (3.17)$$

By eliminating self-inhibition, the elements on the main diagonal of \mathbf{B} are set to zero. In this case the connection matrix spectrum is characterized by:

$$\lambda_i \in \left(-\alpha, \frac{2\pi\alpha}{\sigma^2} - \alpha \right). \quad (3.18)$$

Finally, for narrow-width generating functions with $\sigma < \sqrt{\pi}$ the condition for convergence of 3.5 becomes:

$$\frac{2\pi\alpha}{\sigma^2} - \alpha \leq 1. \quad (3.19)$$

3.3.2 DOG model

For the DOG inter-neuron coupling model, the generating and corresponding connection functions are:

$$f(x, y) = G_{\alpha, \sigma}(x, y) - G_{\alpha_1, \sigma_1}(x, y) \xrightarrow{\text{Generate}} b(u, v) = g_{\alpha, \sigma}(u, v) - g_{\alpha_1, \sigma_1}(u, v). \quad (3.20)$$

Conditions for the convergence of series given in 3.5 are set by limiting the extreme values of the generating function or:

$$\max \left(\left| 2\pi \left(\frac{\alpha}{\sigma^2} - \frac{\alpha_1}{\sigma_1^2} \right) \right|, \left| \frac{2\pi\alpha}{\sigma^2} e^{-\frac{\sigma_1^2}{\sigma_1^2 - \sigma^2} \ln \frac{\alpha \sigma_1^4}{\alpha_1 \sigma^4}} - \frac{2\pi\alpha_1}{\sigma_1^2} e^{-\frac{\sigma^2}{\sigma_1^2 - \sigma^2} \ln \frac{\alpha \sigma_1^4}{\alpha_1 \sigma^4}} \right| \right) \leq 1 \quad (3.21)$$

The pure Gaussian connection model with unity DC component as used in [48] and [50] yields very slow convergence of the series given in 3.5. If Gaussian or DOG coupling as described by relations 3.16 or 3.20 are used, the network response can still be

characterized as the linear combination of Gaussian filters. For example, in the case of Gaussian connection profile without self-inhibition, and assuming the finite number of terms in the series 3.5 is used, the resulting filter can be rewritten as:

$$h_i = \sum_{i=1}^N \mu_i \left(g^{(i-1)} - \frac{1}{\xi} g^{(i)} \right) + \mu_{N+1} g^{(N)} \quad (3.22)$$

where g is a Gaussian kernel with DC component equal to ξ . Coefficients μ_i can easily be obtained from equations 3.11 and 3.13 by replacing the unity matrix with the narrow width Gaussian $g^{(0)}$. Equation 3.22 represents the result similar to that shown by [50], thus confirming that the recurrent network with Gaussian (or DOG) lateral inhibition/excitation behaves as a collection of multiple channel spatial frequency filters.

3.3.3 Difference of Offset Gaussians (DOOG) model

The generating and the connection functions for the DOOG model used herein are:

$$\begin{aligned} f(x, y) &= G_{\alpha, \sigma}(x, y) [2\mu \cos(xu_0 + yv_0) - 1] \xrightarrow{\text{Generate}} \\ b(u, v) &= \mu [g_{\alpha, \sigma}(u - u_0, v - v_0) + g_{\alpha, \sigma}(u + u_0, v + v_0)] - g_{\alpha, \sigma}(u, v) \end{aligned} \quad (3.23)$$

By setting $\mu = 0.5$, the model similar to the one from [65] consistent with the construction of the DOOG profile based on central differencing of Gaussian functions is obtained. In this case:

$$f(x, y) = -2G_{\alpha, \sigma}(x, y) \sin^2 \left(\frac{xu_0 + yv_0}{2} \right). \quad (3.24)$$

Assuming the offset is in the x direction ($v_0 = 0$) the series given in 3.5 converges if:

$$f(\pm \sqrt{t}, 0) \geq -1 \quad (3.25)$$

where

$$t = \frac{\left(\frac{u_0}{2\sigma^2} + \frac{u_0^3}{8}\right) - \sqrt{\left(\frac{u_0}{2\sigma^2} + \frac{u_0^3}{8}\right)^2 - \frac{u_0^4}{12\sigma^2}}}{\frac{u_0^3}{24\sigma^2}} \quad (3.26)$$

approximates the location of the generating function minimum. Expression. 3.26 is obtained through the application of a Taylor series expansion.

3.3.4 Second Order Derivative of Gaussian (2ODG) model

The last type of the inter-neuron coupling addressed is the Second Order Derivative of Gaussian. The analysis is based on the derivative in the x direction. Rotation of the generating function by an arbitrary angle (construction of derivative in different direction) causes the rotation of its Fourier transform while the extreme values remain intact yielding the same conditions for convergence of the series given in 3.5. The 2ODG generating function and its coupling model are defined as:

$$f(x, y) = -x^2 G_{\alpha, \sigma}(x, y) \xrightarrow{\text{Generate}} b(u, v) = \sigma^2 (\sigma^2 u^2 - 1) g_{\alpha, \sigma}(u, v) \quad (3.27)$$

The condition for convergence of series 3.5 for this connectivity model is:

$$\frac{4\pi\alpha}{e} \leq 1. \quad (3.28)$$

Figure 3.3 and Figure 3.4 depict examples of generating functions and the corresponding coupling models that were used.

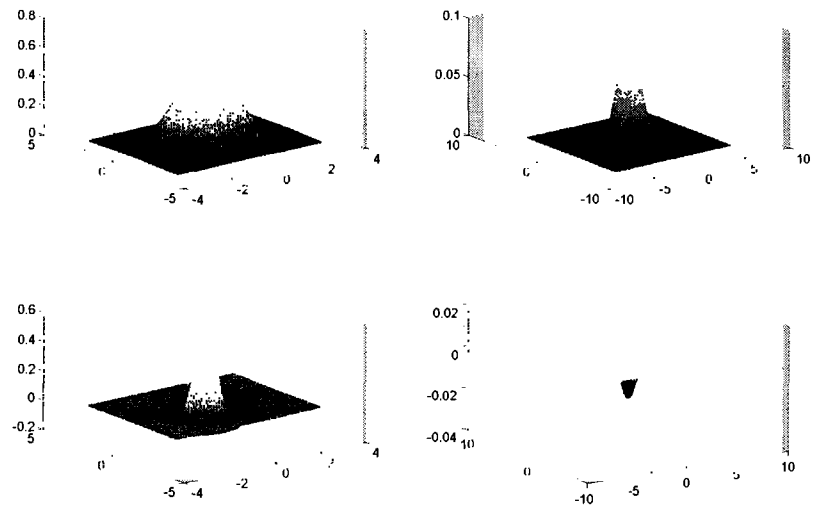


Figure 3.3 Generating functions (left) and corresponding connection functions (right); Gaussian model (top), DOG model (bottom);

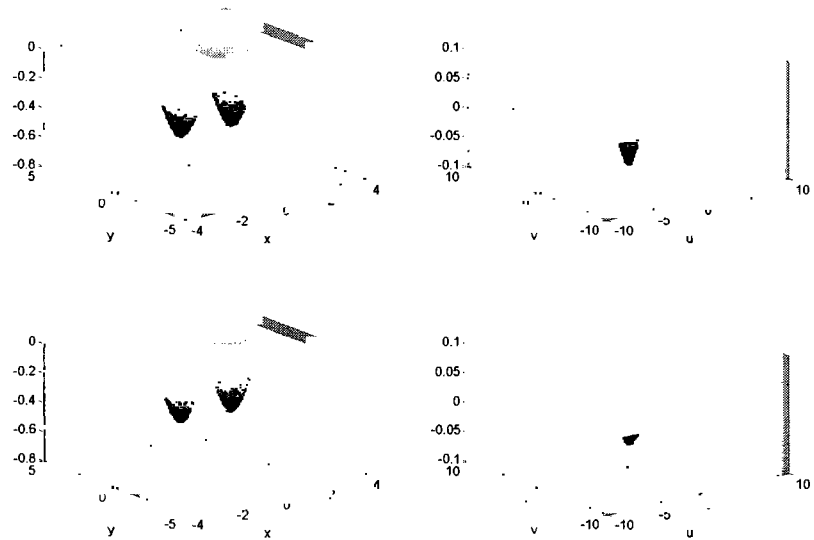


Figure 3.4 Generating functions (left) and corresponding connection functions (right); DOOG model (top), 2ODG model (bottom);

3.4 Steerability of the Recurrent Network with Lateral Inhibition-Excitation

The network filter given in equation 3.13 represents the sum of $N + 1$ components with each function $b^{(i)}(n, m)$ constructed as i cascades of kernel $b(n, m)$. Consider the connection function $b(x, y)$ (note the replacement of variables u, v with x, y) that can be represented in polar coordinates $r = \sqrt{x^2 + y^2}$ and $\varphi = \arg(x, y)$ as a Fourier series in polar angle:

$$b(r, \rho) = \sum_{n=-N}^N a_n(r) e^{jn\varphi} \quad (3.29)$$

The Fourier transform of such kernel is a function that can also be expanded in a similar way:

$$B(R, \theta) = \sum_{n=-N}^N A_n(R) e^{jn\theta} \quad (3.30)$$

where $R = \sqrt{u^2 + v^2}$, $\theta = \arg(u, v)$ and $A_n(R) = 2\pi(-j)^n \int_0^\infty a_n(r) J_n(2\pi Rr) r dr$, J_n being

the n th order Bessel function [81]. Cascading filters represented in the frequency domain by 3.30 produces a steerable filter, i.e. an arbitrarily rotated function can be synthesized by using the linear combination of basis kernels (for details about steerable functions see [78] and [82]). The steering property holds also for the entire resulting filter given by 3.13 since the overall sum remains expandable in Fourier decomposition in polar angle.

It is well known that a 2ODG kernel can be steered by using the linear combination of three basis kernels [78]. As a special case of a kernel given in 3.29, consider the order n

directional derivative of a function $b_c(x, y)$ that is circularly symmetric in spatial and frequency domains. The $m-1$ convolutions of such a kernel with itself yields:

$$b_c^{n,m}(x, y) = \underbrace{\frac{\partial^n b_c(x, y)}{\partial x^n} * \frac{\partial^n b_c(x, y)}{\partial x^n} * \dots * \frac{\partial^n b_c(x, y)}{\partial x^n}}_{m-1 \text{ convolutions}} \xrightarrow{F} (ju)^{nm} B_c^m(u, v) \quad (3.31)$$

The basis for steering of this cascade consists of $nm+1$ functions since the pure real (or imaginary) frequency response is the product of polynomial in u of order nm containing only odd or even terms and circularly symmetric function. The sum of filters:

$$h(x, y) = \sum_{i=1}^N (-1)^i \underbrace{\frac{\partial^n b_c(x, y)}{\partial x^n} * \frac{\partial^n b_c(x, y)}{\partial x^n} * \dots * \frac{\partial^n b_c(x, y)}{\partial x^n}}_{i-1 \text{ convolutions}} \quad (3.32)$$

can similarly be represented as a product of polynomial in x of order nN and another circularly symmetric function where the polynomial coefficients may depend on the radius r . In this, more general case $2nN+1$ basis functions are sufficient to steer the function $h(x, y)$. This implies that when $b_c^{n,1}(x, y)$ is a connection function, the overall network impulse response given as:

$$h_i(x, y) = \delta(x, y) + h(x, y) \quad (3.33)$$

is steerable as well and represents the continuous equivalent of the filter given in 3.13. For n even, the number of basis functions that steer 3.33 is reduced to $nN+1$. For example, in the case of 2ODG coupling, $2+1$ filters steer the network. Steerability can efficiently be used to determine the response of the network for arbitrary orientation of the connection function without performing its actual rotation. One just needs to calculate the linear combination of basis function responses. Limiting the basis set to a relatively small number of functions as well as enabling fast convergence of the series given in 3.5 is essential for practical implementations. Figure 3.5 depicts components of

the steerable impulse response of a network with the 2ODG coupling model assuming $N = 3$ and continuous space.

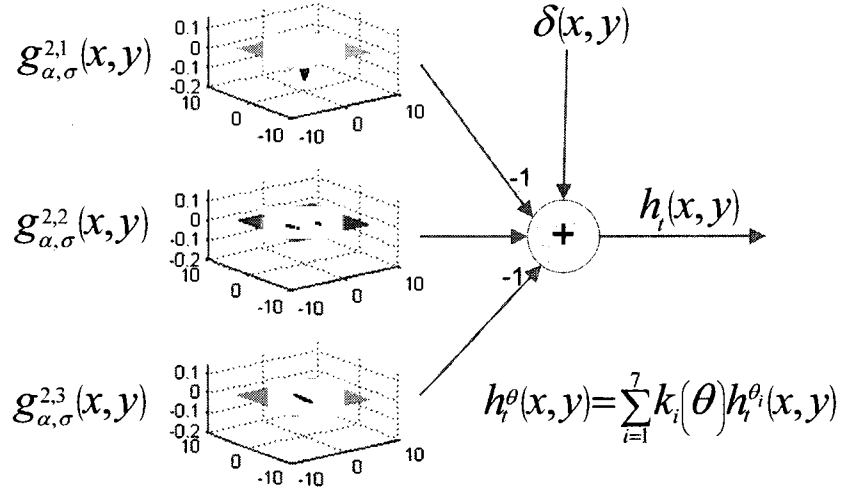


Figure 3.5 Steerable impulse response of the network with 2ODG coupling model in continuum; $h_i^{\theta}(x, y)$ is an arbitrarily rotated filter while $k_i(\theta)$ and $h_i^{\theta_i}(x, y)$ are the interpolating and basis functions respectively.

3.5 Simulation Results

This section presents simulations obtained when processing synthetic images with networks whose connection weights are constructed using models of inter-neuron coupling addressed in sections 3.3 and 3.4. By bounding the generating function, the spectral radius of the connection matrix is maintained in the interval $(0,1)$. This guarantees convergence of the power series given by equation 3.5 but it also enables finding of the network solution via the iterative scheme:

$$\mathbf{y}_{n+1} = \mathbf{x} - \mathbf{B}\mathbf{y}_n \quad (3.34)$$

In order to validate the network construction and solution introduced via Proposition 1 and equations 3.11 and 3.13, the solution obtained by direct convolution with a pre-computed kernel (application of equation 3.13) of relatively small size is compared with the solution obtained by application of the iterative scheme given by equation 3.34. The estimated number of filters N from equation 3.13 necessary to produce the convergent result is reached once the energy of the term $b^{(N)}(n,m)$ is extremely small (0.0002%) when compared to the energy of the filter $h_i(n,m)$. The iterative procedure given by equation 3.34 is terminated when the average absolute difference in pixel values between successive solutions is less than a predefined small threshold (for the grayscale images with levels of gray between 0 and 255 this threshold was set to 0.5).

When changing parameters of the connection model(s), extremes of generating functions are varied. This in turn influences the rate of convergence of the series given by equation 3.5. For arbitrarily chosen parameters, the resulting sufficient number of filters N for all coupling models discussed in this paper is presented in Figure 3.6.

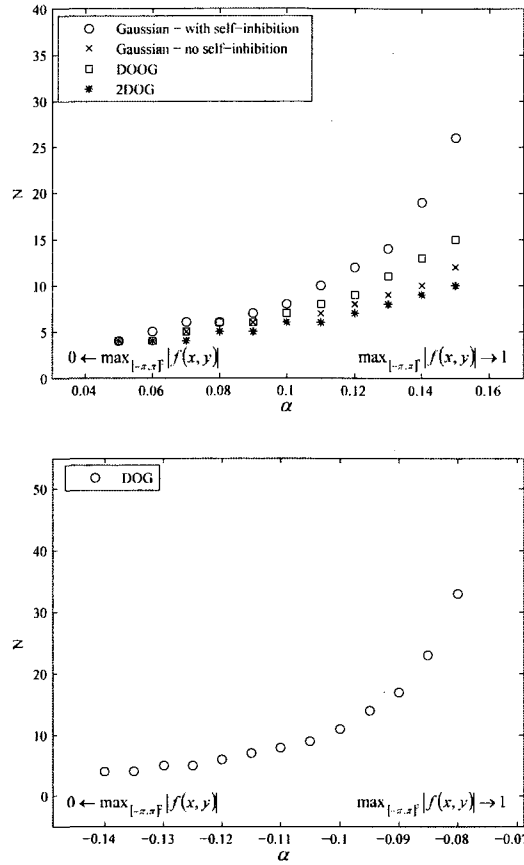


Figure 3.6 The number of sufficient filters versus α ; Particular parameters chosen are: for Gaussian model $\sigma = \pi/3$, for the DOOG model $\sigma = \pi/3$, $u_0 = 2$, for the 2DOG model $\sigma = \pi/5$ while for the DOG model $\alpha_1 = -0.11$, $\sigma = \pi/4$, $\sigma_1 = \pi/5$.

Based on Figure 3.6, appropriately chosen parameters of the connection function allow the use of a reasonably small N . Networks with circular symmetric inter-neuron coupling models (e.g. Gaussian and DOG) produce visually similar results since the response is not orientation sensitive. Similarly, when non-circular models (e.g. DOOG and 2DOG) are used, the response is orientation sensitive. Figure 3.7 depicts the result obtained by processing the square test pattern of uniform intensity placed on a dark background when using the DOG connection model.

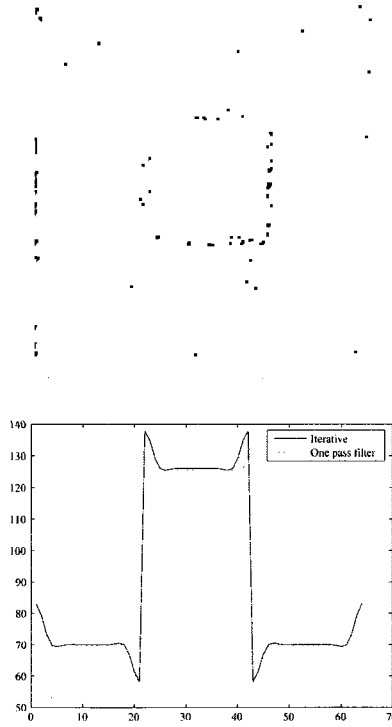


Figure 3.7 One pass convolution with pre-computed kernel (connection model is DOG with α_1, σ and σ_1 identical to those from Figure 3.6 and $\alpha = -0.13$); filtered square pattern (top), mid-line resulting profiles for iterative procedure and single-pass convolution (bottom); pre-computed kernel constructed by using $N=7$

As shown by Figure 3.7 the result of processing the input pattern with a simple feed-forward filter is almost identical to the one obtained by finding the iterative solution to the system. As an example for the orientation sensitive model, processing by using a steerable network filter constructed from the 2ODG connection function is also demonstrated. The number of sufficient filters is chosen based on results described in Figure 3.6. Interpolating coefficients necessary for steering are obtained from [78] while using basis functions evenly spaced between 0 and π . These coefficients are provided in the Table 3.1.

i	$k_i(\pi/6)$	$k_i(\pi/2)$	$k_i(5\pi/6)$
1	0.1333	-0.0667	0.1333
2	-0.2157	0.0682	-0.0996
3	0.6378	-0.073	0.0824
4	0.6378	0.0824	-0.073
5	-0.2157	-0.0996	0.0682
6	0.1333	0.1333	-0.0667
7	-0.0996	-0.2157	0.0682
8	0.0824	0.6378	-0.073
9	-0.073	0.6378	0.0824
10	0.0682	-0.2157	-0.0996
11	-0.0667	0.1333	0.1333
12	0.0682	-0.0996	-0.2157
13	-0.073	0.0824	0.6378
14	0.0824	-0.073	0.6378
15	-0.0996	0.0682	-0.2157

Table 3.1 Steering coefficients for the network with 2ODG connection model ($\alpha = 0.12$, $\sigma = \pi/5$, $N = 7$).

The mesh plots of the network response $h_i(m, n)$ and its portion $h(m, n)$ are shown in Figure 3.8. Results of processing the hexagonal test pattern by steering the filter $h(m, n)$ are depicted in Figure 3.9.

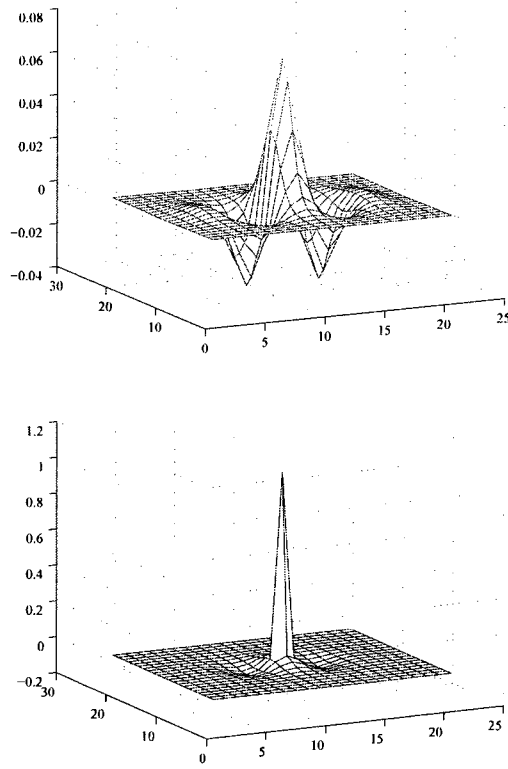


Figure 3.8 Steerable network impulse response for 2ODG connection model ($\alpha = 0.12$, $\sigma = \pi/5$, $N = 7$); $h(m, n)$ portion (top), $h_l(m, n)$ (bottom);



Figure 3.9 Steering the network with 2ODG connection model ($\alpha = 0.12$, $\sigma = \pi/5$, $N = 7$); linear combination of 15 filters used to steer the response to three different orientations.

The theoretical and experimental results presented in this chapter clearly demonstrate when and how the large recurrent network with *a priori* defined lateral connectivity model can be solved by utilizing spectral properties of its connection matrix. The solution does not require large matrix inversion, nor calculation of the matrix eigenvalues in order to explicitly determine the matrix spectral radius. This advantage is very important for systems of considerable size. Biologically inspired receptive fields are very often characterized as steerable and discussion presented in this chapter revealed that the recurrent linear networks can also behave as steerable kernels. This is the case when the connection function of the neural coupling model is expandable in a Fourier series in polar angle. Practical construction and utilization of filters based on the responses of recurrent linear networks with lateral connections is the central topic of the following two chapters.

CHAPTER IV

SEMI-BLIND PRE-WHITENING BY SPATIALLY INVARIANT NETWORKS WITH LATERAL INHIBITION/EXCITATION

One of the most difficult tasks in pattern recognition is effective extraction of meaningful feature sets. The choice of signal processing tools that prepare and shape the information along the processing chain before the final pattern classification is made becomes critical. This chapter addresses the role of networks with lateral recurrent connectivity as pre-whitening filters. In particular, the computationally efficient kernels with relatively small memory footprint representing spatially uniform model of neural coupling are developed as pre-processors for category-specific signals. They are later utilized in practical feature extraction process.

4.1 Decorrelation and Whitening by Recurrent Neural Networks

Neural networks with recurrent connectivity have widely been used for modeling various information processing phenomena present in biological systems. A very attractive method of signal processing emerges from a specific type of adaptation of recurrent neural networks to input patterns. In the very influential paper published by Barlow and Földiák [63] it was shown that the linear model of network with recurrent inhibition can learn about the input signal statistics. The network actually stores the information about the covariance matrix of input data if the interactions between units are adapted according to anti-Hebbian learning rule. As a result of such learning, the outputs

of the system become uncorrelated and this happens even in the case of strong correlations between input signal components. The linear network model with multiple outputs connected to multiple inputs via adaptive connections is shown in Figure 4.1. It represents the extension of the model shown in Figure 3.2 allowing for modification of lateral weights, and having more than two inputs/outputs.

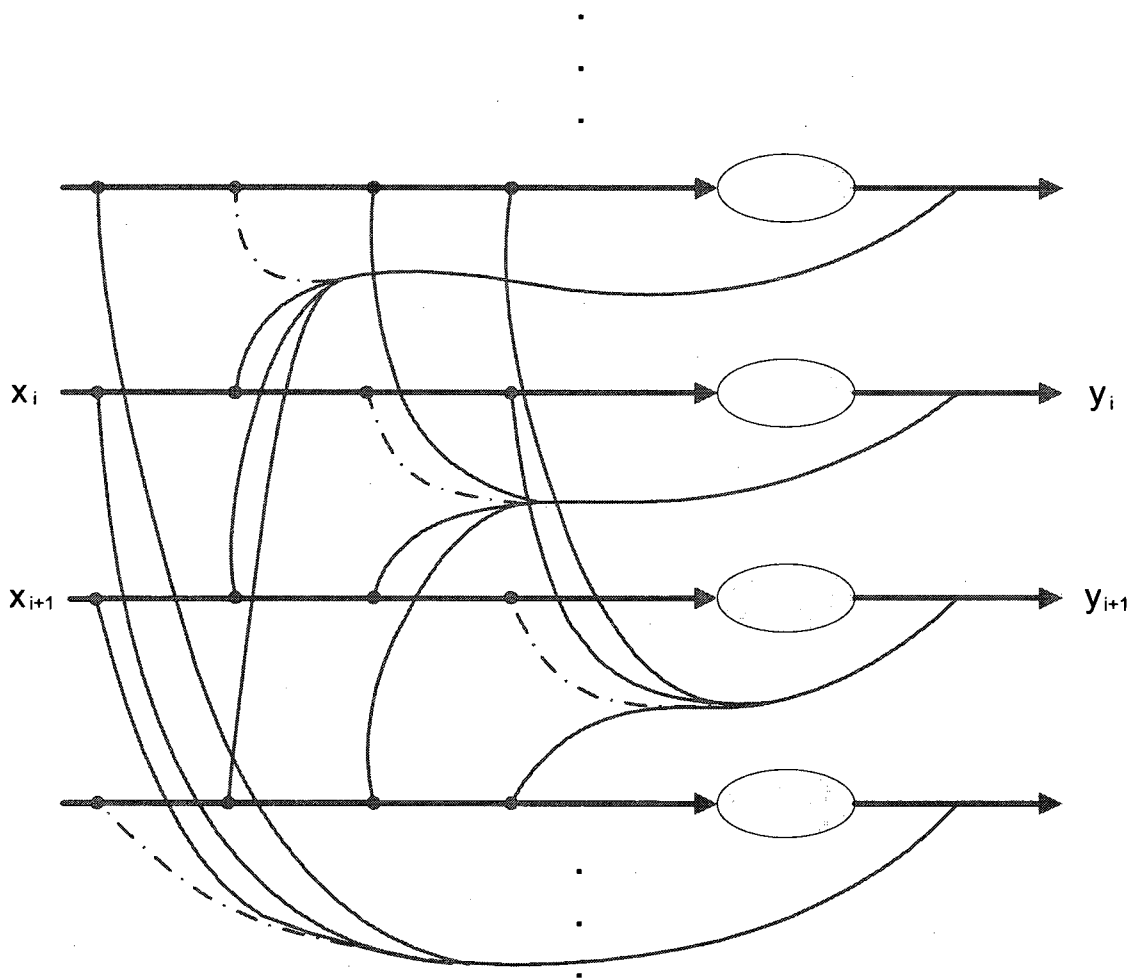


Figure 4.1 The recurrent network with adaptive connections (self inhibition/excitation can potentially be allowed) used in [63]

The equation governing the output of the network from Figure 4.1 can be written as:

$$y_i = x_i - \sum_j b_{i,j} y_j \quad (4.1)$$

where, $b_{i,j}$ represents the synaptic weight between units i and j ². The dash – dot connection pattern symbolizes the fact that the self-inhibition/excitation might be present in the model, although it was not allowed in the particular network of Barlow and Földiák. The anti-Hebbian learning rule for this network is:

$$\begin{aligned} \Delta b_{i,j} &= \alpha y_i y_j & \text{if } i \neq j \\ \Delta b_{i,j} &= 0 & \text{otherwise} \end{aligned} \quad (4.2)$$

where, α is the small positive constant which influences the rate of adaptation. The higher the correlation between two outputs, the stronger the inhibition between them. This in turn produces the stronger decorrelating action of the particular synapse in the network. The result of the learning process described by equation 4.2 is the network output data set with purely diagonal covariance matrix. The data decorrelation can essentially be achieved by various transforms, i.e. the decorrelation problem has generally more than one solution [30]. For example, one of the very common and well studied algorithms that decorrelates the input data is the Principal Component Analysis (PCA). When the neural approach is considered, several different architectures and accompanying learning rules are being used for decorrelation and/or PCA [63], [64], [83] - [90]. Many of these algorithms employ generalizations of classic Oja's Principle Component Analysis neuron and Hebbian learning algorithm [91], as well as lateral inhibition between neurons for decorrelation of signal components. The algorithm for decorrelation addressed in this chapter has its impetus in the model proposed by

² Note that the inhibition in equation 4.1 is represented by positive values of synapses.

Plumbley [64]. 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 and under the limitation on available power for transmission. The optimization of information transfer for such a case is achieved when outputs of the network are decorrelated and at the same time have equal variance. This characterization of decorrelation and equal variance is equivalent to whitening of the output signals. Assuming the zero-mean process, the output covariance matrix of the whitened signal is given as:

$$C_y = E\{yy^T\} = \beta I \quad (4.3)$$

where, output vector y is obtained in accordance with equation 3.1 In equation (4.3), the operator E denotes the expected value. The linear recurrent network with lateral connections between output and input units that achieves this type of signal processing is almost identical to the one used by Barlow and Földiák in [63]. It has an additional feedback from the output of each neural unit to its input – this is so-called self-inhibitory connection. The adaptation algorithm for this network given in a matrix form can be represented as:

$$\Delta B = \alpha(yy^T - \beta I) \quad (4.4)$$

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

The decorrelation and whitening has an important role in signal processing and pattern recognition disciplines. As outlined by Palmeiri et. al. [86] the decorrelation corresponds to “enforcing at least a first-order independence among the system

variables.” As such, it can become the first step towards the detection of completely statistically independent events (components) within the processed signal. Over the past decade or so, the technique of ICA has gained a lot of attention since it represents the model which tries to find exactly such a representation of the input - the linear combination of statistically independent sources that comprise the analyzed signal. After the whitening of the input, the ICA processing becomes easier, or well-posed. This is because once whitening is achieved, the un-mixing of the signal components (the extraction and separation of independent components) for real-valued signals is achieved by using the orthogonal matrix [92] for data transform. In fact, whitening of the input data is required by some ICA algorithms while with others, it improves their convergence properties [58]. For example, the pre-whitening of the input signal is implemented in [30] and [93] before the actual ICA algorithm is executed.

The input data to the network with lateral connectivity used throughout the remainder of this dissertation consists of patches extracted from various images. This method is based on the fact that correlations among pixels in pictures decrease rapidly with distance [94] and for practical information encoding purposes it is sufficient to observe correlations between pixels 4-5 distance units apart [95]. The patch of size $n \times n$ is transformed into an input vector x of size $n^2 \times 1$. This operation is achieved by successively transposing patch rows and concatenating them into resulting vector. The process of 2-D signal transformation into 1-D vector is shown in Figure 4.2.

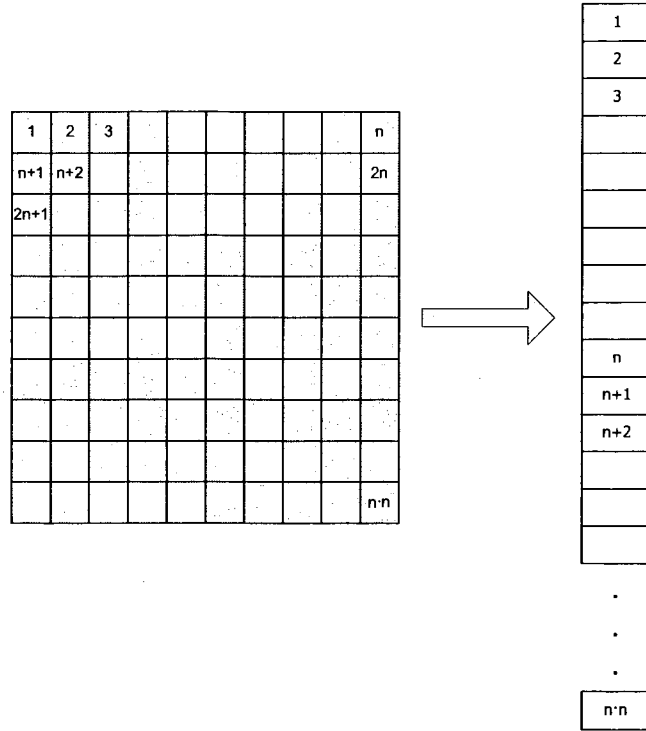


Figure 4.2 Transformation of the image patch into single column vector

In this chapter, specific methods of signal pre-processing are addressed in order to improve the performance of the whitening network with recurrent inhibition/excitation as described by equation 4.4. In particular, the attempt is made to construct efficient filters that are easy to store, i.e have small memory requirements, thus making them adequate for real-world (very likely embedded) implementations. These filters can be seen as special members of the family of kernels previously studied during the analysis of spatially invariant networks with lateral connectivity (Chapter II and Chapter III). However, their common characteristic is that the network connectivity matrix is not given *a priori* by the specific function. It is rather learned during the network training/adaptation but with certain constraints imposed on the connectivity pattern itself. Since the unsupervised model of learning described by equations 4.2 and 4.4 acquires the

knowledge about input signal statistics and spectrum during the adaptation itself, it can be expected that for similar input signals, similar decorrelating filters described by the matrix $(I + B)^{-1}$ arise. Construction of such filters for specific classes of images is addressed in this chapter. Prior to presenting this concept, it is necessary to address the usefulness of networks with spatially uniform lateral interactions with pre-defined and non-adaptive lateral connection models, i.e. models seen in Chapter II and Chapter III.

4.2 Decorrelation Role of Non-Adaptive Networks with Lateral Connectivity

Discussions related to spatially invariant lateral inhibitory/excitatory networks and their solution as given in Chapter II and Chapter III reveal and characterize techniques used to solve such systems. In particular, they outline some interesting concepts like steering properties or interpretation of such networks as collections of multiple bandpass filters. Obviously, the identical network architecture is presented in the model described in Figure 4.1 except that in this case, weights can be modified and are adaptive with no restrictions on spatial variability of the system (i.e. the connection matrix elements are not necessarily dependent on the distance between neural units or may not be *a priori* given in the form of some sort of explicit connection function). If the decorrelation of network outputs is desired, according to the adaptation models given in equations 4.2 and 4.4, it is necessary to increase overall inhibitory strength between two units that are more correlated. It is already shown that in the case of 2-D signals, the correlation in intensity between neighboring spatial locations is very strong [96]. Increasing the distance between pixels reduces that correlation [94], [96]. The behavior of such a correlation function is consistent with quite natural assumption that inhibitory

influence between neural receptors in biological systems decreases with distance too, i.e. it is always easier to inhibit your neighbor rather than some distant neuron. Note that models A and B or even the Gaussian model of inter-neuron coupling addressed previously assume such distance dependent decrease of inhibition between neural units. Considering these well known statistical properties of natural scenes and the general nature of learning rules given by 4.2 and 4.4, it is no surprise that certain networks with lateral inhibition (even with models that are non-adaptive) provide a significant level of signal decorrelation at the network output. That is why the first attempt aimed toward extraction of significant, meaningful features from the input signal is decorrelation via spatially invariant networks with pre-defined lateral connectivity function. The experimental results revealing the decorrelating characteristics of such systems are shown in section 4.4 of this chapter. Although significantly capable of reducing the correlation between input components, presented non-adaptive networks are limited by their potential to produce completely uncorrelated and what is more desired, whitened outputs. That is why alternative models of filters based on networks with recurrent connectivity are investigated.

4.3 Category-Specific Pre-Whitening Filters

Development of category-specific pre-whitening filters characterized by spatially invariant network with lateral connectivity is based on the adaptation algorithm given in equation 4.4. Essentially, the learning process has to remove the off-diagonal elements in the output covariance matrix when decorrelation of network outputs is desired. As a measure of decorrelation and whitening achieved, the difference between the target

covariance matrix βI and the matrix $E(yy^T)$ has to be monitored. The direct implementation of the adaptation rule (equation 4.4) would of course produce completely whitened signal at the output. The task presented herein is to determine whether the input image data characterized as a certain signal category (i.e. image of nature, image of an object on dark background or some different category) would yield category-specific filter that can be used to meaningfully pre-process the input before adaptation and learning described by equation 4.4 even begins. Additionally, it is desired for the pre-processing to be done as efficiently as possible with the filter containing a minimum number of different coefficients, thus yielding a small memory footprint for storage of such kernels. These pre-processing operations are expected to aid in efficiency of algorithm described by equation 4.4.

In order to develop the pre-processing filter in the form of linear network with lateral interactions between its inputs and outputs, it is the aim to experimentally find the adequate inhibition model that would be spatially uniform (characterized by single connectivity model) and that could still pre-process the input data in the manner of the whitening filter. This kernel does not necessarily have to strictly fit into any explicit form of a certain function (Gaussian, exponential, or similar). However, if such a connectivity model can be found, it should be stored in numerical form as a collection of coefficients representing a 2-D discrete function. If additional symmetry and space invariance assumptions for lateral connections between neural units are made, the connectivity matrix of the recurrent network becomes highly structured and easy to store. As mentioned earlier in the text, the spatially invariant convolution with the discrete kernel can be described as multiplication of Toeplitz-Block-Toeplitz matrix by the

vectorized input 2-D signal. Similarly, if the lateral inhibition between neural units is space-invariant, then the connection matrix is also of Toeplitz-Block-Toeplitz type. For systems where conditions presented by proposition 3.1 from Chapter III of this dissertation are satisfied, i.e. the matrix generating function is bounded by unity, the network filter response can be found without explicit matrix inversion. However, if the system input is effectively of dimension n^2 , where n is the relatively small patch size (usually 8,12,16 or similar) the matrix inversion problem is not a significant one for most of the computational platforms in use today. This is the major advantage of the network that processes local image patches compared to the system that would process the entire image of potentially large size. For the input signal patches of size $n \times n$, the resulting network connection matrix is of $n^2 \times n^2$ size. Full whitening in general requires n^4 coefficients in order to produce an ideal whitening filter. The proposed alternative to immediately learning and producing the filter requiring the connection matrix with essentially n^4 different entries is to try to find the typical whitening filter corresponding to a certain class of images by restricting the connection matrix to symmetric Toeplitz-Block-Toeplitz type with symmetric Toeplitz blocks. This condition corresponds to finding the adequate spatially uniform connectivity model for the neural network that can achieve only approximate whitening of the input signal. The connection function of this network is restricted to the function symmetric about both coordinate axes. If such connectivity matrix can be found, the storage requirements become far more relaxed since n^2 numbers are sufficient to describe the connection model. Of course, the anticipated network is expected to properly decorrelate/whiten similar types of signals and the only information that is needed before applying the particular type of filter is the

knowledge of the category the input belongs to. The proposed algorithm for learning the pre-whitening filter can be described as:

1. Acquire the input data from typical category of images the filter is designed for.
This phase assumes the collection of “large” number of patches taken from input images, their transformation into 1-D vectors as shown in Figure 4.2, and removing the mean from every component of the input signal.
2. Start the learning process according to equation 4.4.
3. After every update of the connection matrix, transform the matrix B into symmetric Toeplitz-Block-Toeplitz matrix with symmetric Toeplitz blocks by averaging all the corresponding diagonal/symmetric entries.
4. Based on the new connection matrix calculate the resulting output covariance matrix. The measure of algorithm convergence is the difference between desired and achieved covariance of output data. Assuming C_Y is the covariance of output data and βI is the desired covariance matrix, the difference between these matrices can be estimated as Frobenius norm:

$$E_F = \sqrt{\sum_i \sum_j |C_{Y_{i,j}} - \beta I_{i,j}|^2} \quad (4.5)$$

Once the norm E_F is less then the small pre-defined constant the learning procedure is terminated.

The algorithm presented herein requires the transformation of currently learned connection matrix into ideal symmetric Toeplitz-Block-Toeplitz form with symmetric Topelitz blocks. In order to achieve this structure, all elements of the matrix have to be

divided into groups with each group representing the unique entry in the resulting matrix. For illustration purposes, the simple form of such matrix partition is depicted in Figure 4.3. This example illustrates the situation where image data (i.e. patches) are of size 3×3 and there are nine different entries in the resulting Toeplitz-Block-Toeplitz matrix. Each group of identical elements is labeled by corresponding letter A-I. The sub-blocks of the model are depicted by different intensities of grey.

A	B	C	D	E	F	G	H	I
P	A	B	E	D	E	H	G	H
C	B	A	F	E	C	I	H	G
D	F	I	A	B	C	D	E	F
I	D	L	S	A	B	E	D	E
F	E	D	C	B	A	F	E	D
G	H	I	D	E	F	A	B	C
H	G	H	E	D	E	F	A	B
I	H	G	F	E	D	C	B	A

Figure 4.3 The partition of TBT matrix into groups of identical entries

The matrix depicted in Figure 4.3 can be considered as a pseudo-mask for generalizing the Toeplitz-Block-Toeplitz matrix given the arbitrary matrix Z of size 9×9 . Elements of matrix Z positioned at the location of particular letter are summed and scaled by the number of letters of that particular type in the matrix (averaging). The resulting matrix is obtained by placing such averaged coefficients into locations determined by particular letter after the summation takes place. The algorithm describing the learning of space invariant pre-whitening filters i.e. networks with uniform connection function is depicted in Figure 4.4.

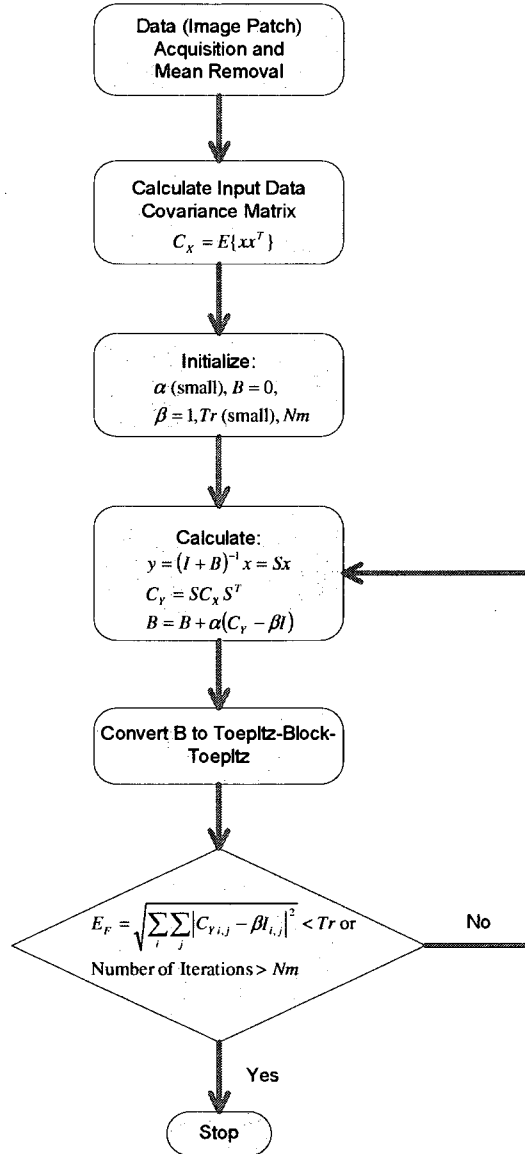


Figure 4.4 Learning the pre-whitening filter

It has to be noted that constraints imposed on entries of the connection matrix during the learning of the filter in some cases can not achieve ideally small E_F . If this threshold is chosen to be too small, the algorithm might never stop and it becomes necessary to impose additional termination criteria, e.g. the maximum number of iterations allowed (alternative might be the increase of E_F). The modification of connectivity coefficients

during the learning process is achieved by using the current output covariance matrix for the adaptation instead of individual network outputs as given by equation 4.4.

The ultimate test of the learned filter is the pre-processing of the image data from the same class the filter is learned on. Of course, the test images would not be the same as those used for learning the filter coefficients. The expectation is that the filter will improve the speed of learning in neural adaptive whitening process compared to the case when the weights are initially zero or random values. Also, the comparison in performance should be evaluated for such pre-processing filter (formed by using only n^2 different values) and the one that contains full $n^2 \times n^2$ connection model learned originally without imposing any particular constraints on the connection matrix B .

4.4 Experimental Results

This section presents two sets of experimental results. The first one is obtained when implementing filters based on networks with pre-defined, non-adaptive inter-neuron connectivity model on images of nature – also referred as category 1. The second set represents the outcome of the network adaptation and learning of pre-whitening filters as addressed in section 4.3 on three distinct categories of images. Those categories are: images of nature, log-polar warped images of objects placed on the dark background and images of predominantly horizontal texture. Two examples from each of the three image categories are represented in Figure 4.5, Figure 4.6 and Figure 4.7. Note that the actual size of images differs and they are resized for display purposes.



Figure 4.5 Images of nature (category 1)

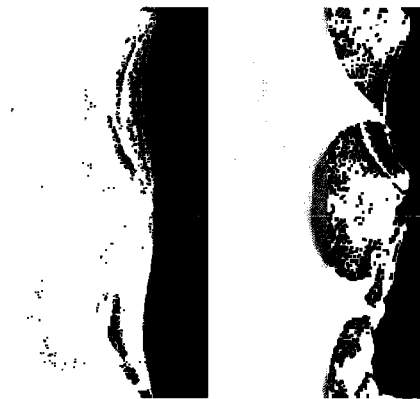


Figure 4.6 Images of log-polar warped objects (category 2)

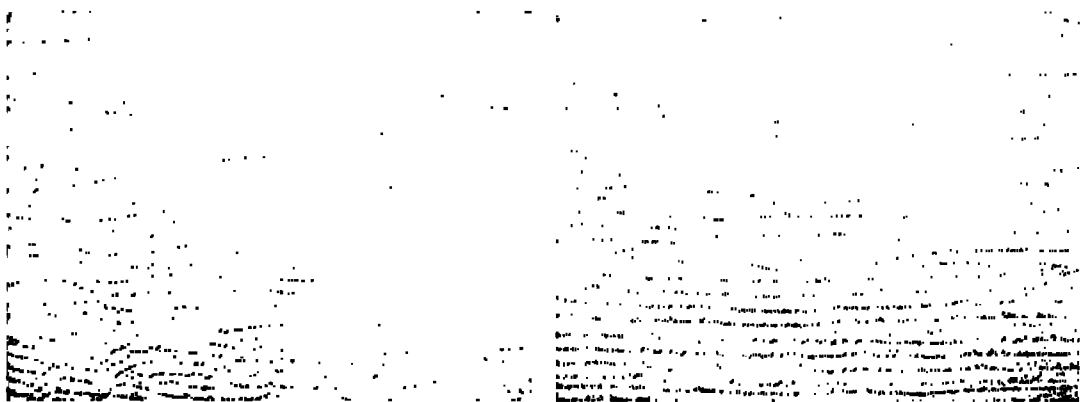


Figure 4.7 Images of predominantly horizontal gratings/texture (category 3)

The data collection from all images was performed by using patches of size 16×16 . In order to approximate the statistical characteristics of each image category, the patches were obtained by randomly sampling the collection of images from particular category and creating the data set of 25000 samples. After performing such data collection, the resulting data set was stored as a 256×25000 matrix with each matrix row representing one component of the input. Two additional steps are taken before actual learning of the filter takes place – the removal of the mean from each component as well as normalizing the data so that the global variance of each 256×25000 data set is unity. When forming the data sets, the size of input images varied. For category 1 the image size was 2048×1536 , for category 2 it was 256×512 while for category 3 it was 1024×768 . Each category of images was divided into four groups with each group containing certain number of images. For categories 1 and 3 five images per group were used and for category 2, each group contained eight images. The collection of data resulted in obtaining four data sets for each category of images.

In the case of data processing with non-adaptive recurrent network with lateral connectivity as addressed in chapters 3 and 4, results obtained when using data from image category 1 are presented herein. The filters selected for experiments were two kernels reflecting model A and model B coupling as given in Figure 2.2 as well as two filters reflecting Gaussian connection matrix with and without self-inhibition as given in section 3.3.1. In the latter case, the filter coefficients were calculated according to equation 3.13³. The data processing is performed in the form of convolution with particular kernel i.e. the direct network solution was not utilized. In order to calculate the

³ The coefficients from Gaussian connectivity model used in the case without self-inhibition were: $\alpha = 0.11, \sigma = \pi/3$, while in the case with self-inhibition were: $\alpha = 0.15, \sigma = \pi/3$.

measure of data decorrelation at the system output, it was necessary to obtain input and output covariance matrix (i.e. before and after particular filter is applied). The resulting image was sampled and processed in the identical manner as the input in order to obtain statistical properties of data. The results obtained by implementing filters characterizing non-adaptive recurrent networks with lateral inhibition are shown in Table 4.1.

Group Number (Category 1 used)	$\sum_{i \neq j} C_x(i, j) $	Measure of decorrelation for different connectivity models: $\sum_{i \neq j} C_y(i, j) $			
		Model A	Model B	Gaussian without self-inhibition	Gaussian with self-inhibition
1	4401.4613	665.9185	541.8871	1942.5685	1334.3708
2	5273.7644	796.4620	647.5157	2325.5166	1597.4860
3	3990.5288	603.1127	490.9116	1759.2033	1208.9130
4	4434.2161	668.4053	542.5185	1952.3204	1342.4003

Table 4.1 The results of decorrelation with filters based on non-adaptive uniform networks with lateral inhibition

As demonstrated in Table 4.1 all four models of lateral connectivity implemented in experimental work yielded significant reduction in correlation between image pixel intensities. This is actually expected result since both the correlation between pixels as well as inter-neuron coupling strength within addressed models fall off with inter-pixel/inter-neuron distance (i.e. the stronger the correlation, the stronger the inhibition). Additional results obtained when processing category 2 and category 3 of images with these non-adaptive filters are presented in Appendix A.

When learning category-specific pre-whitening filters, the algorithm described in Figure 4.4 is performed. Only one data set per category is used to actually learn the category-specific filter. Once the network filter is obtained by learning the spatially invariant lateral connection model (the Toeplitz-Block-Toeplitz connection matrix), it is

tested in three scenarios. The ultimate goal in all of those scenarios is to whiten the input data by adaptive model described by equation 4.4 but with implementation of data pre-processing by category-specific filters. The first test compares the behavior and effectiveness of the learned pre-whitening filter with the case when the adaptation starts from zero, i.e. instead of initializing the connection matrix with zeros, the pre-whitening filter learned from similar, but not identical data set is used. The second case is comparison of the whitening process when data is pre-processed by the learned filter characterized with symmetric TBT matrix with symmetric Toeplitz blocks on one hand and, on the other hand, the filter that is obtained by completely whitening the training data from the same category without enforcing the strict Toeplitz-Block-Toeplitz matrix connectivity form at every learning step. Finally, the third test compares how filters learned on the data from one category behave when applied in “wrong category” data pre-processing. Figure 4.8, Figure 4.9 and Figure 4.10 depict results obtained by preprocessing (pre-whitening) different data sets obtained from three groups from each image category. Data pre-processing by the specific filter is equivalent to initializing the network connectivity matrix B with the coefficients characteristic for the recurrent network constituting the given filter.

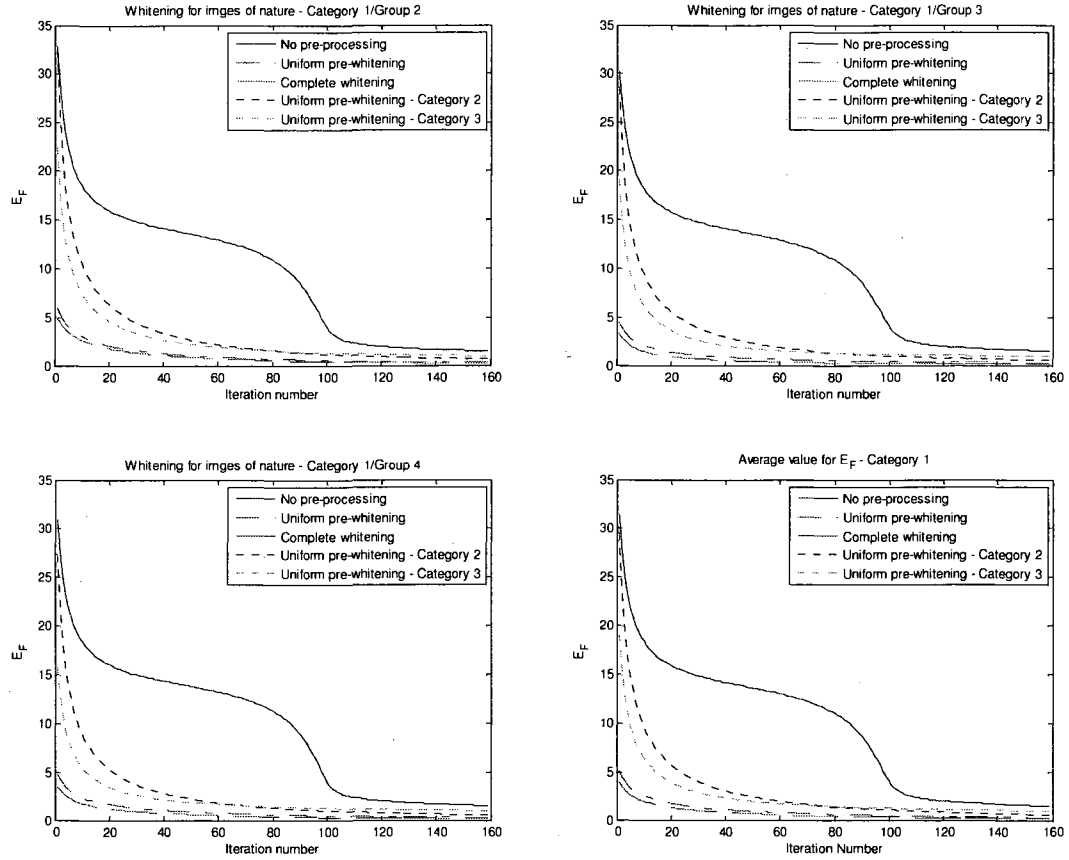


Figure 4.8 Performance of learned pre-whitening filter when processing three groups of data from Category 1

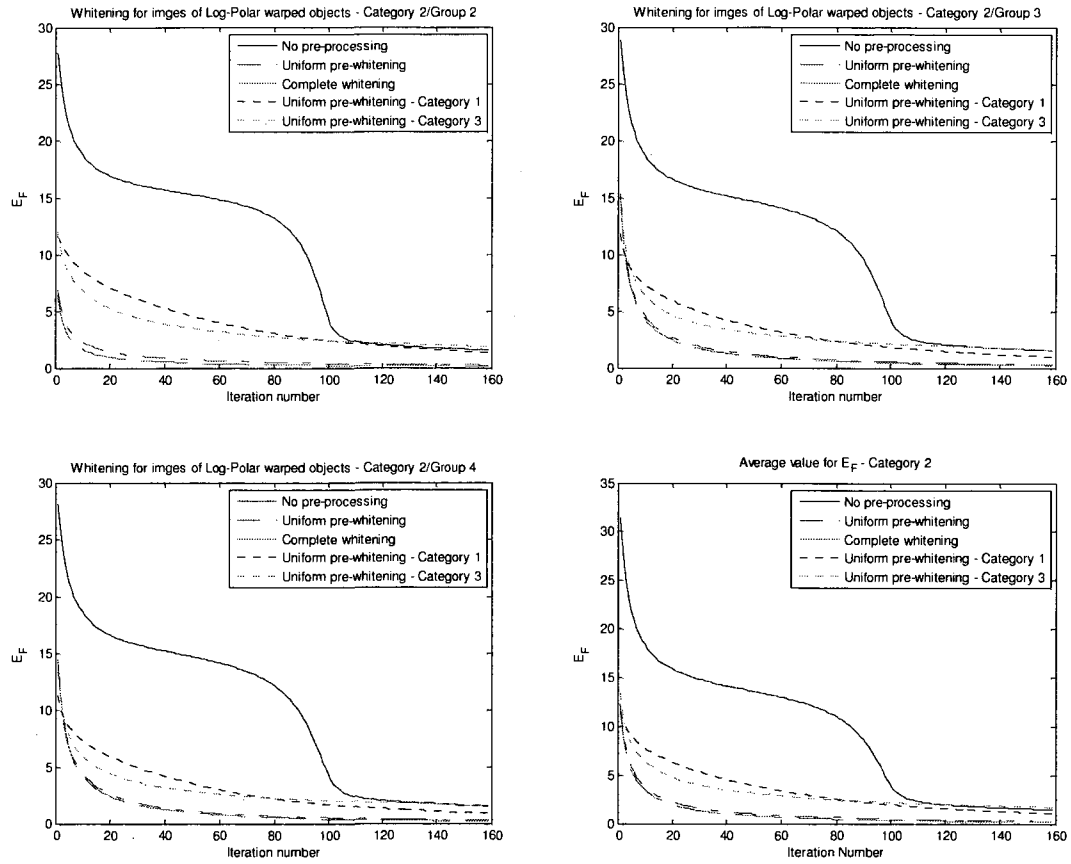


Figure 4.9 Performance of learned pre-whitening filter when processing three groups of data from Category 2

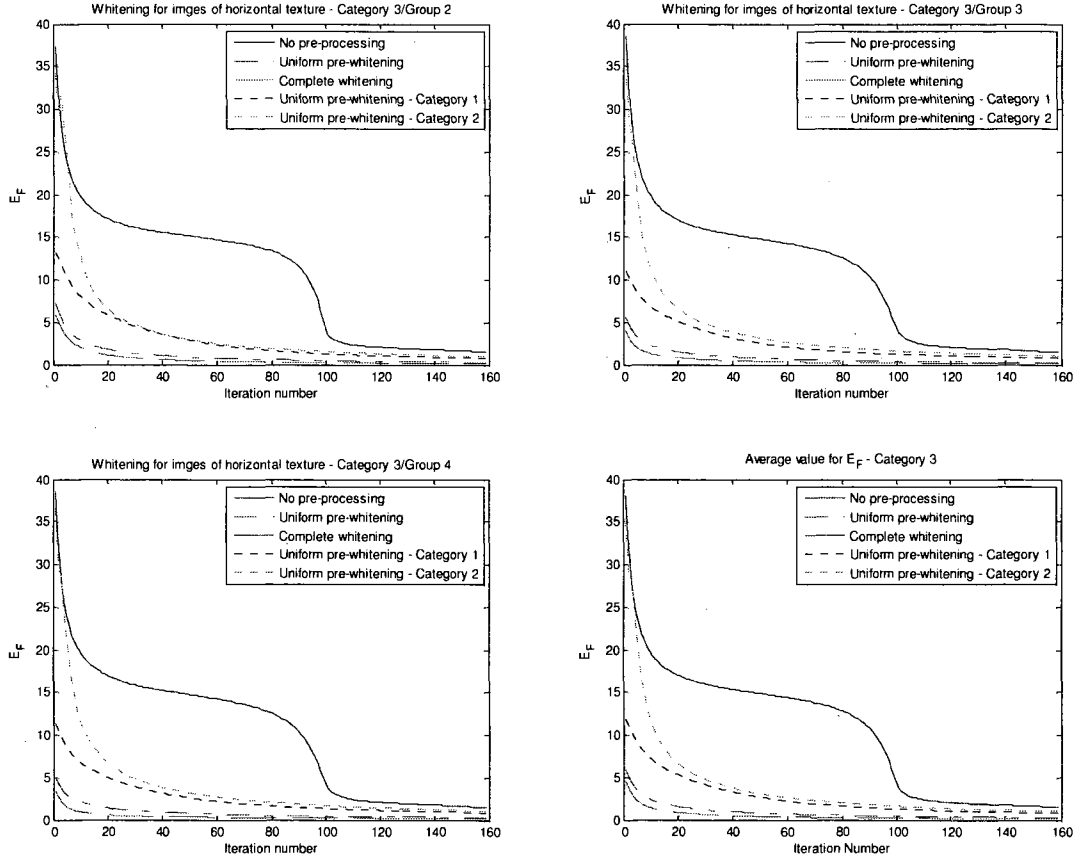


Figure 4.10 Performance of learned pre-whitening filter when processing three groups of data from Category 3

From Figure 4.8, Figure 4.9 and Figure 4.10 it is evident that learned pre-whitening filters can be efficiently used in order to prepare i.e. pre-process the data during the whitening operation as described in equation 4.4. In real-world scenarios, one would start from the pre-determined, reduced number of network coefficients forming the symmetric Toeplitz-Block-Toeplitz matrix with symmetric Toeplitz blocks and calculate the filter in the form given in equation 3.1. Once this operation is completed, the data whitening can continue based on Plumbley's model and without any restriction on the development (adaptation) of the connectivity matrix. As seen from these experimental results, the data pre-processing with filter formed by spatially invariant connectivity models (labeled

as uniform whitening) is almost as effective as pre-processing the particular category of data with the network whose connections are previously adapted to completely whiten the input. One should notice that 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 (for example one knows that the input to the system is the image of log-polar warped object, but one does not know which object in particular is in question). Presented results also demonstrate that if the wrong filter pre-processes the data, the whitening becomes much slower and relatively inefficient. In addition to the results presented in this section, the performance of the pre-whitening filters based on adaptation model depicted in Figure 4.4 when processing individual images is described in Appendix A. The role of category-specific pre-whitening kernels in solving pattern recognition problems is demonstrated in the text that follows.

CHAPTER V

BIOLOGICALLY INSPIRED INFORMATION PRE-PROCESSING AND OBJECT RECOGNITION

This chapter addresses the use of category-specific, spatially uniform pre-whitening networks in ICA-based pattern recognition tasks. Specifically, the usefulness of such filters in signal pre-processing when obtaining sparse feature sets of data in log-polar domain is investigated. It is demonstrated that reducing the number of non-zero components in feature vector representations yields promising pattern recognition performance.

5.1 Independent Component Analysis as Biologically Inspired Signal Encoding

After acquiring category-specific pre-whitening filters through learning technique addressed in Chapter IV of this dissertation, it is quite natural to investigate practical usefulness of such biologically motivated signal pre-processing. Previous efforts to answer this question could be categorized into two basic theories. As demonstrated in [47], the lateral inhibition among receptor/neurons improves the signal contrast at the network output. In practical terms, this can be interpreted as an edge enhancement operation [46]. On the other hand, when observing statistical properties of input signals, the adaptations of the recurrent network partially governed by lateral inhibition mechanisms lead to signal decorrelation and whitening [64], [83], [84], [89] which in turn can serve as common and important pre-processors for ICA [60]. The fundamental

objective of this chapter is to demonstrate how the category-specific pre-whitening filters can be combined with filters constructed for ICA signal encoding in the log-polar domain for feature extraction and pattern recognition purposes. Note that the entire signal processing chain has a strong biological flavor. Still, it has to be emphasized that the goal of this dissertation was not to pursue an ideal machine model of some specific biological vision system that would pre-process the input and perform the pattern recognition task in a purely biological manner. That is why as a pattern recognition technique, the form of a well known, not necessarily biologically motivated pattern classifier is used. Before describing the approach taken when solving the problem of object recognition by using this biologically motivated information processing framework, it is necessary to reveal some of the basic characteristics and significance of ICA.

It is well known that biological systems have evolved over long periods of time and that the development of mechanisms that process visual stimulus was driven towards effectively coping with signals received from the external environment. Bearing this in mind, it is reasonable to assume that the mammalian visual cortex (being part of image/signal processing chain) is no exception. The notion of discovering efficient signal encoding strategies that would be in agreement with known characteristics of information processing in visual cortex has been a major scientific effort in the last few decades. Significant progress in this direction was made by Olshausen and Field [28], [29] who demonstrated that characterization of mammalian visual cortex simple cell receptive fields as localized, oriented and bandpass [29] is consistent with the sparse

coding of natural images. The sparse coding of input images $I(x, y)$ can be described as signal representation in the form:

$$I(x, y) = \sum_i s_i a_i(x, y) \quad (5.1)$$

where, coefficients s_i have probability distributions (of their activity) with heavy tails highly peaked around zero. The functions a_i are called basis functions of the model. In practical terms, it is said that the particular neuron will be inactive most of the time and that it will fire rarely. This means that an image encoding is possible by only small number of active units. It is very important to point out that such sparse distributions reduce statistical dependence among output units [27], [31], thus providing the link between sparse coding and the ICA. The approximate illustration of the distribution of activity of code coefficients for sparse coding is shown in Figure 5.1.

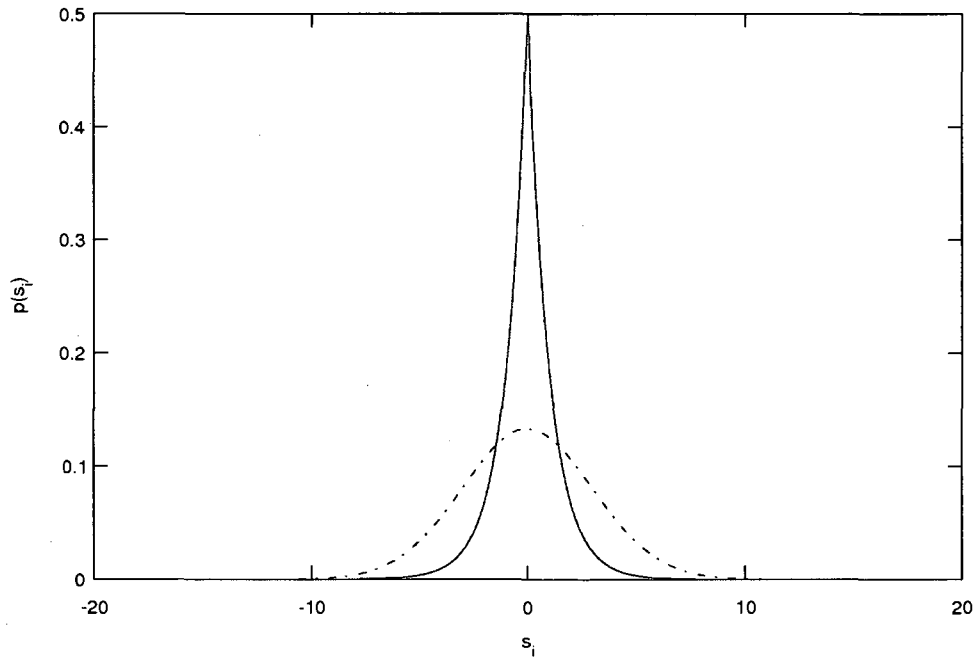


Figure 5.1 Distribution of activity of code coefficients for two cases: sparse code (full line) and non-sparse code (dotted line)

An example of the information processing incorporating models of complex cell responses as present in visual cortex [97] and the neural layer responsible for sparse coding of outputs of those complex cells is described by Hoyer and Hyvärinen [98]. Such a network learns contour coding from natural images where responses of complex cells are given as squared and summed outputs of quadrature Gabor filters.

Another important step in revealing efficient signal coding strategies potentially employed by the mammalian brain are techniques that are based directly on ICA. As stated by Bell and Sejnowski [30] the ICA is equivalent to Barlow's redundancy reduction problem [27] which postulates that the feature detectors of the HVS are developed and adapted with the goal of producing as statistically independent activations from each other as possible. Many different algorithms for ICA are published [58] and generally, their primary goal is to reveal statistically independent components present in the analyzed signal under the assumption that analyzed input is a mixture of some features (i.e. basis functions). One example of this direct link between algorithms for sparse code generation and ICA is given by Olshausen and Field [29]. They proved that the ICA algorithm of Bell and Sejnowski [99] based on maximization of mutual information between inputs and outputs of the encoding system can be interpreted as the model that finds the set of basis functions that maximize the average log-likelihood of images assuming coding coefficients are sparse and statistically independent. The decomposition described by relation 5.1 under the assumption of statistically independent and non-Gaussian coefficients s_i is called ICA and as stated by Hyvärinen and Hoyer it can be viewed from different perspective as sparse signal coding. The linear representation given by 5.1 has its counterpart in a matrix form:

$$x = As \quad (5.2)$$

which can be rewritten as:

$$x = a_1 s_1 + a_2 s_2 + \dots + a_n s_n \quad (5.3)$$

where, each basis function a_i represents the column of matrix A (also known as a mixing matrix) while s_i is the component of vector s (particular independent component). The input to the system is the column vector x which in the case of this dissertation can be seen as vectorized version of the 2-D signal obtained by serializing image rows into single column. For the particular set of different input signals, the coefficients s_i are different and in the case of practical ICA algorithms, they are as statistically independent as possible. For the given collection of images, the ICA algorithms attempt to find the best suited set of basis functions (or the mixing matrix A) that satisfies the model given by equation 5.2 while observing the independence of coefficients s_i . ICA captures the essential structure of the data and is suitable for many applications like feature extraction, signal separation (or more specifically for solving the well known cocktail party problem), separation of artifacts in MEG data, finding hidden factors in financial data or reducing noise in natural images [60]. ICA was recently successfully used for face recognition [100] where the authors showed that for such an application the ICA representations were superior to PCA ones. The utilization of ICA as a principle method for feature extraction from color and stereo images can be seen in [101]. This paper argues that treatment of chromatic and stereo information yields features and receptive fields that provide additional evidence for redundancy reduction process performed by the visual cortex.

In order to better understand the feature extraction model addressed herein, it is necessary to provide some detail about the specific ICA technique called FastICA [61], [102] since this method was used for ICA based image encoding for the purposes of this research. Originally, a fast fixed-point iteration algorithm for estimation of independent components from a given n -dimensional signal was proposed by Hyvärinen and Oja [102]. Many algorithms for ICA are performed by establishing an objective, or alternatively called contrast function with the ultimate goal of minimization or maximization of the same. In the case of the fast fixed-point ICA method from [102], the contrast function was kurtosis which for a zero-mean random variable s is defined as:

$$kurt(s) = E\{s^4\} - 3(E\{s^2\})^2 \quad (5.4)$$

Alternatively to using kurtosis, as shown in [61], it is possible to perform the maximization of an objective function of type:

$$J_G(s) = [E\{G(s)\} - E\{G(v)\}]^2 \quad (5.5)$$

where, v is a Gaussian variable of zero mean and unit variance and G is any non-quadratic function. This objective function represents the approximation of negentropy or the nongaussianity of underlying data. The one-unit FastICA algorithm yielding the maximization of contrast function given by equation 5.5 is based on the iteration scheme revealing the single independent component:

$$w(t+1) = E\{xg(w(t)^T x)\} - E\{g'(w(t)^T x)\}w(t) \quad (5.6)$$

where, the weight vector w is normalized after each iteration. The function g represents the first derivative of function G from equation 5.5. The so-called demixing matrix W that yields estimates of all independent components:

$$\bar{s} = Wx \quad (5.7)$$

is given as:

$$W = (w_1, w_2, \dots, w_n)^T. \quad (5.8)$$

In order to estimate all the different independent components, it is necessary to decorrelate outputs after each iteration. This can be done using several different schemes (see [61] for details).

In order to produce ICA decomposition of the input signal, the model given by equation 5.6 was used. The non-linearity $G(u)$ had the form:

$$G(u) = \log \cosh(u) \quad (5.9)$$

with derivative defined as:

$$g(u) = \tanh(u) \quad (5.10)$$

There are several reasons for the choice of the given ICA model. The algorithm is well known and proven to be extremely fast (the convergence is cubic). There are practically no user defined parameters. The function $G(u)$ given in 5.9 is a good general purpose choice for the contrast function. There is no need of prior knowledge or estimation of the probability distributions of signal components. The algorithm is computationally simple, has small memory requirements and the code implementation of FastICA is readily available and well documented [62].

As previously stated, many of ICA algorithms assume the data is whitened since their performance becomes significantly improved when dealing with such data structures. This is the case with most fastICA models. This fact emphasizes the importance of pre-whitening filters discussed in Chapter IV of this dissertation.

In the remainder of this chapter, the role of category-specific pre-whitening filters (based on networks with recurrent connectivity) in a complete chain of signal processing

for pattern recognition tasks will be revealed. The ICA based feature extraction techniques adapted for recognition of objects in the log-polar domain (i.e. cortical domain) are addressed. The rotation and scale tolerant feature space is formed and encoded via ICA filters by using only a limited number of locations in the image. These locations are chosen based on the responses of specific type of filters that are also biologically inspired – log-Gabor kernels. At the very end of the chapter, the results of object recognition based on multiple classifiers are reported. The experiments were conducted for images of objects taken under different lightning conditions.

5.2 Recognition of Objects Through Log-Polar Transform and Local ICA Analysis

The main problem addressed in this section can be described as recognition of objects in log-polar space based on localized ICA feature extraction. Essentially, the method of feature detection and encoding addressed herein relies on two groups of filters. The first group contains pre-whitening filters developed for specific class of input 2-D signals – log-polar warped objects sitting on a dark background. The second group contains ICA filters learned on the identical group/class of images. When working with log-polar images, the assumption that was constantly present throughout the entire research was that the center of the log-polar transform for a particular image is constant, i.e. the object is always properly centered and warped. This essentially allowed for rotation and scale tolerant information processing. The block diagram representing the algorithm used for feature extraction and pattern recognition is shown in Figure 5.2.

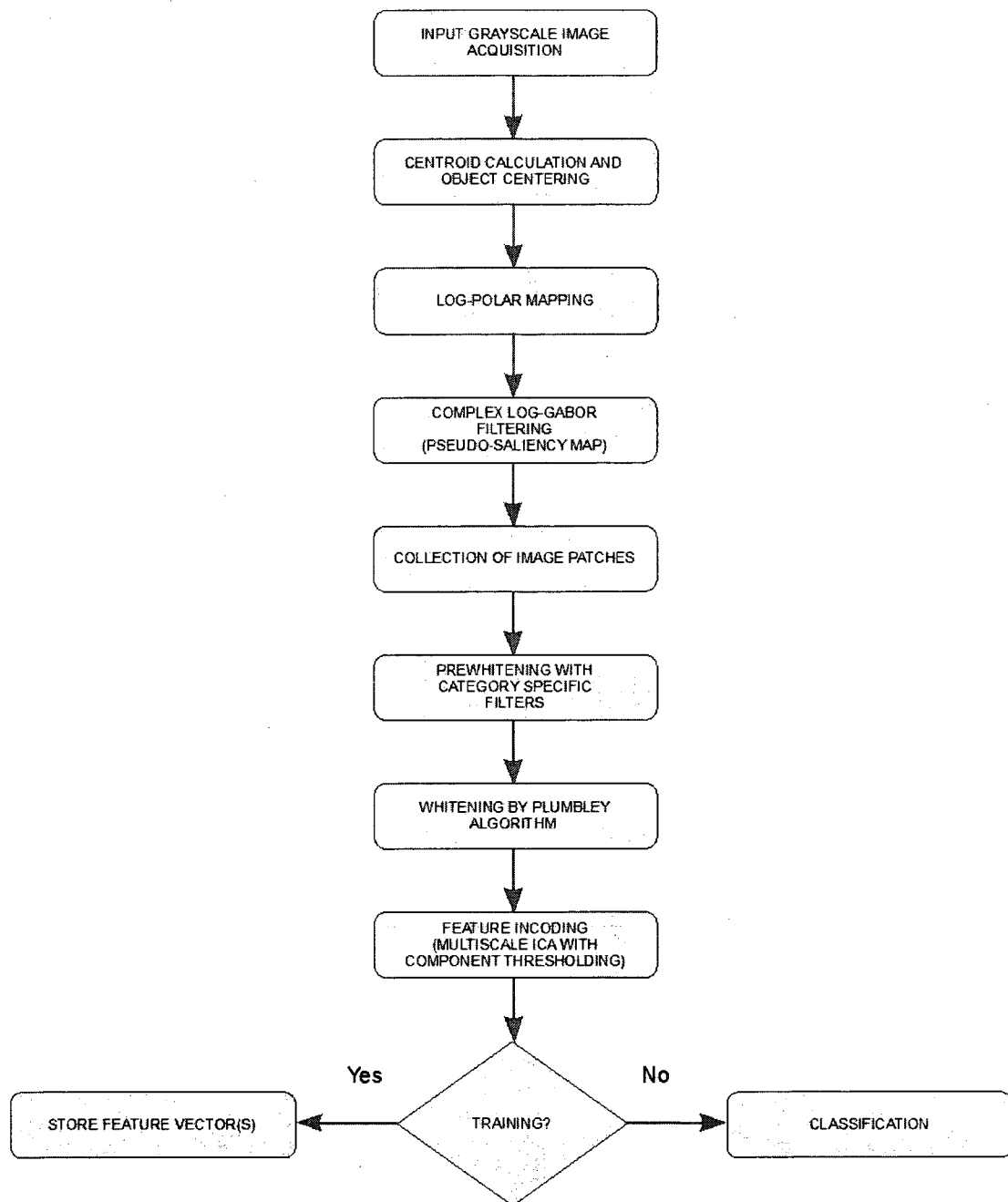


Figure 5.2 Feature extraction and object recognition based on pre-whitening and ICA filters

The input to the pattern recognition model depicted in Figure 5.2 was the set of grayscale images of objects sitting on a dark background. Each object (class) was represented by

twelve different grayscale images (or potentially their rotated versions), with each image captured under different lighting conditions. More information about the database used for experimental work (Amsterdam Library of Object Images) can be found in [103] and [104]. A few examples of object images from the database are shown in Figure 5.3.

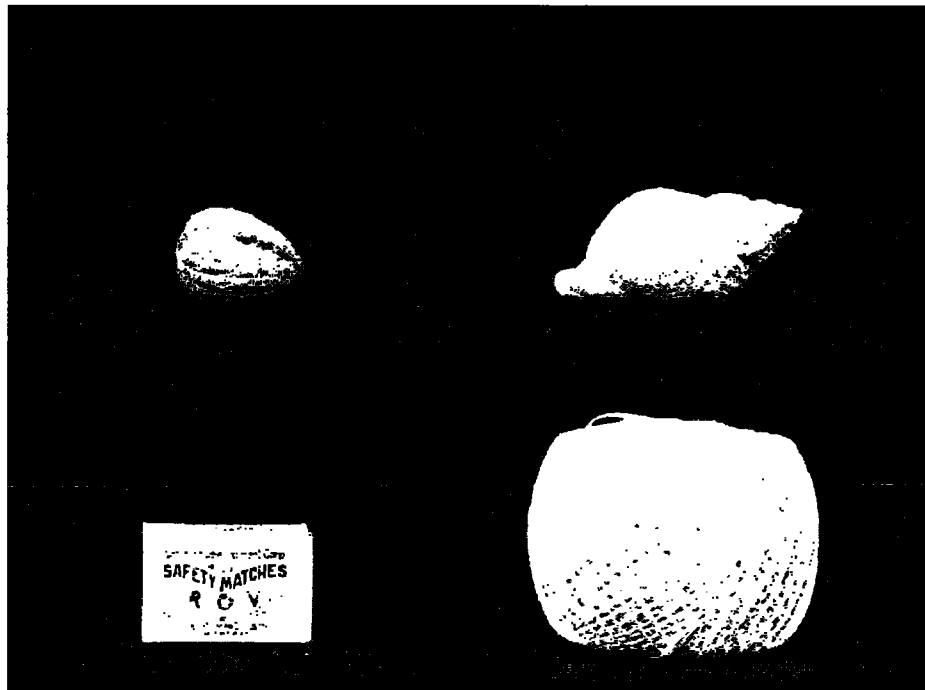


Figure 5.3 Four objects from image database

Figure 5.2 reveals that the algorithm contains the training mode and the actual recognition or test mode. Out of one thousand possible objects from the given database, forty of them were chosen for the experiment described in this dissertation. Out of twelve images of each object, eight were used in order to train the system and approximate the parameters necessary for accurate classification. Thus, the total number of training images was 320. Four images from each class were used for testing purposes, yielding a total of 160 test samples. After describing details of the feature extraction and

the pattern recognition algorithm the recognition rates are given for each implemented classifier.

The feature extraction phase starts with the input image acquisition which in this case is already performed by a third party. In order to perform accurate log-polar warping on every image, it is desired to compute the objective center of the complex-log transform. In the case of images from the given database the problem translates to separation of background from the object and computation of the specific point in the object that would serve as the center for log-polar mapping. The method used to separate the object from the background is thresholding based on Otsu's method [105]. Once the suitable threshold is obtained, it is assumed that each pixel with the value above the threshold belongs to the object itself, while others belong to the background. The next step in image processing is thresholding and forming the binary image based on the given threshold value. This binary image is only the temporary signal (a mask) that aids in detection of outer boundaries of a given object along horizontal and perpendicular axes as well as the center point of log-polar transform – the object centroid. The calculation of the object centroid itself is based on the calculation of moment of order one [106]. Once outer boundaries and object center-point are found based on this black and white image, they are adapted as basic boundaries of the new gray-scale image (the object is still surrounded by a small background). The result of these operations is the image that contains an object that is centered according to an objective and consistent algorithm. Examples of objects centered by the described method are presented in Figure 5.4.

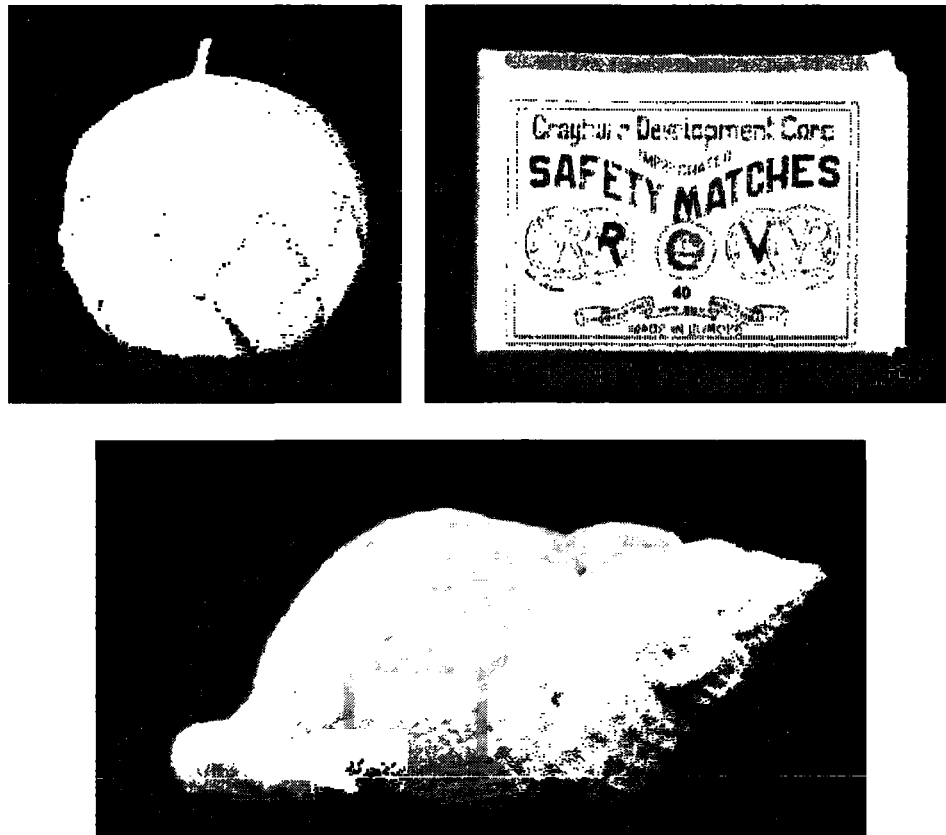


Figure 5.4 The result of centering of an object prior to log-polar mapping

The centering is performed in such way that the pixel corresponding to the centroid of an object is always in the center of the resulting image. The procedure of obtaining the grayscale image of the centered object is depicted in Figure 5. 5.

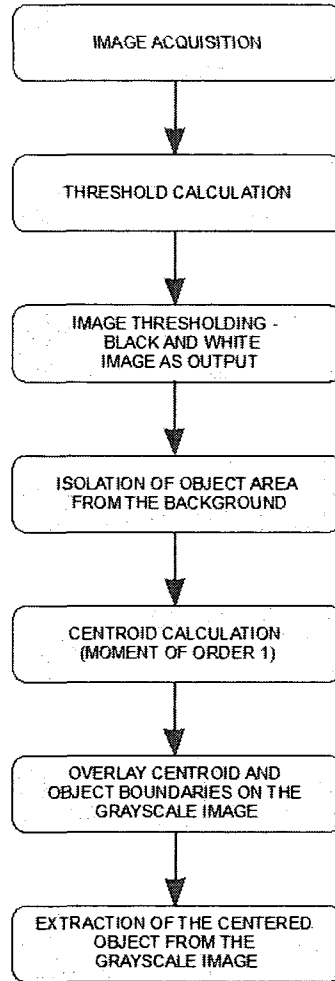


Figure 5. 5 Centered object image extraction

After object centering is performed, the log-polar mapping can take place. The mapping itself is performed by batch processing the entire set of input images using warping functions which are part of open source OpenCV library written in C/C++. Particular details about the entire library along with the help and sample codes can be obtained from [107]. For every image, the target size (the size of the image in log-polar domain) was set to 256×512 thus effectively representing 256 log-radial and 512 angular samples. The warped versions of objects from Figure 5.4 are shown in Figure 5.6.

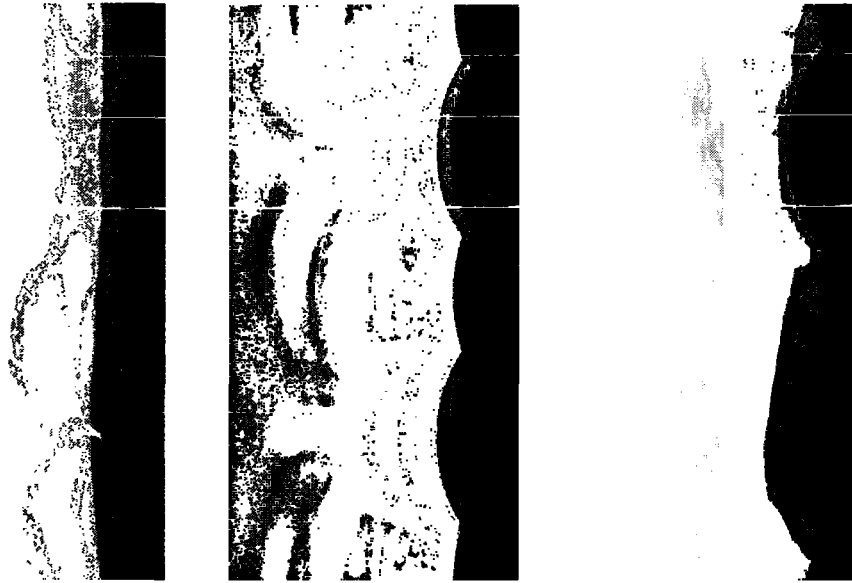


Figure 5.6 Log-polar mapping of centered objects

As shown in the block diagram depicted in Figure 5.2, after the images are converted to log-polar domain, the specific filtering with log-Gabor kernels is performed. This stage is also labeled as pseudo-saliency detection. The main reason why this step is chosen as a necessary part of the signal encoding process lies in the nature of log-polar mapping and the technique chosen for feature extraction.

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 very efficient image processing step towards scale and rotation tolerant feature representation. When dealing with ICA techniques in the log-polar domain one has to use special care. The pre-whitening filters described in Chapter IV are given in the form of networks with recurrent connectivity and they act upon signals formed from localized image patches. Similarly, one standard approach to describe the

effects of ICA algorithms (for example on natural images) is to sample great collection of image patches and treat them as specific signal outcomes. By following this model, this research attempts to identify efficient coding strategies for images already mapped to the log-polar domain while preserving rotation and scale tolerance of the mapping itself. Obviously, one possible method of feature encoding would be crude sequential scanning of the cortical image and taking patches of a specific size in a purely sequential order with fixed sampling grid. However, this has extremely negative effects in the case of object rotation or resizing. Only in the special case (that would most probably rarely happen in reality), when mapped object deformations in terms of scale and rotation angle are such that the content from one patch moves to represent the identical content contained in some other patch (i.e. rotations and scales are allowed only by multiple of certain discrete factor) the log-polar mapping would preserve it's size and scale tolerant feature encoding capacity. The demonstration of the effect of this fixed sampling grid in log-polar environment is depicted in Figure 5.7.

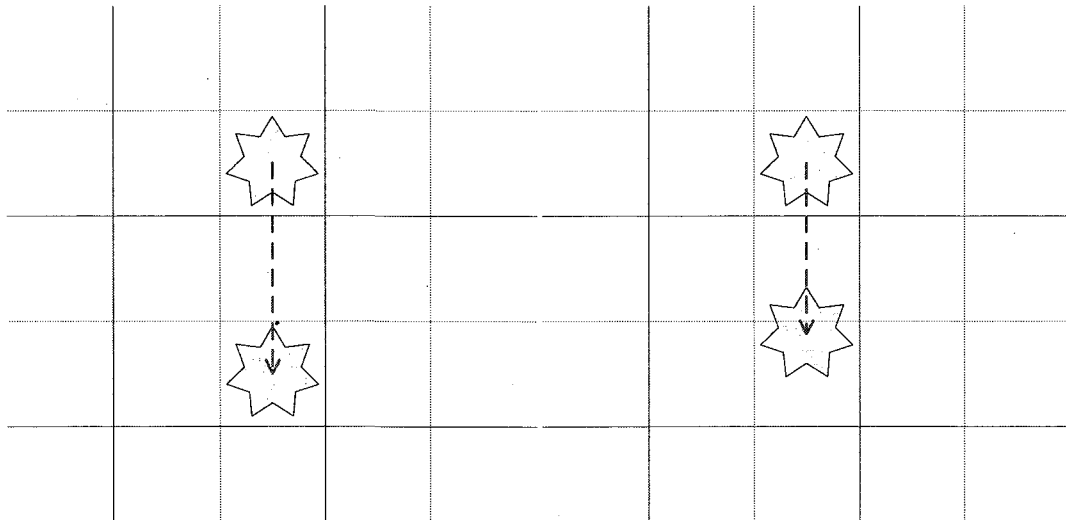


Figure 5.7 The problem of fixed grid with sequential patch sampling in log-polar space; example where rotation of object causes completely different content of patches (right); only for specific rotation angles this content is preserved (left)

This problem was solved by choosing only specific, localized regions of the log-polar image that would be interesting for further analysis. These locations and their immediate surrounding can be defined as interesting image patches whose content would be transformed by further stages of the algorithm and image features appropriately encoded. The advantage of this approach is that regardless of the pattern shift in the original or log-polar domain, the spatial location of sampling patches is not fixed as it is in the case presented in Figure 5.7. Rather, the patches and their locations are selected based on the points of interest in the image and they continuously move as the pattern itself moves due to rotation and/or scaling. This concept is shown in Figure 5.8.

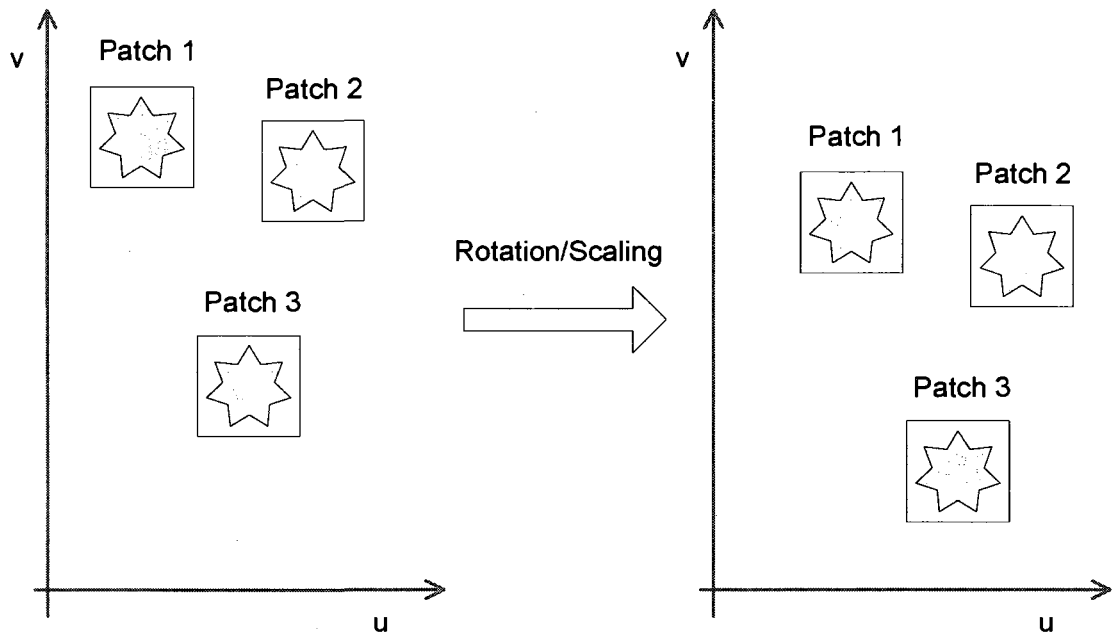


Figure 5.8 Selection of patches based on points of interest – sampling locations are not *a priori* defined

The model used for selection of pseudo-salient image regions is based on the responses of the complex log-Gabor filters [108]. Before describing the implemented pseudo-saliency selection process, it should be stated that alternative, relatively complex and biologically

inspired models for detection of salient image locations already exist and belong to the separate area of scientific research [109]. Such models are intended to mimic the processes of visual attention and are commonly implemented for revealing the conspicuous details in natural scenes (i.e. structures/objects that strongly differ or stand out from the background). The characterization of such saliency models in the warped, log-polar domain is a quite interesting and potentially very challenging task which is out of scope of this research.

The model for selection of points of interest presented herein is based on the application of the bank of multi-scale and multi-orientation log-Gabor filters. Particular locations, selected based on highest responses of such filters are voted as pseudo-saliency points. The motivation for usage of log-Gabor instead of the most commonly used Gabor filters for this stage of image processing comes from the work of Field [110]. Based on the analysis of natural image statistics and its spectra he concluded that encoding of such images is far superior with functions of log-Gabor type that have no DC components and are additionally characterized by extended tails in frequency domain. The commonly used regular Gabor functions would over-represent low frequency components since it would not be possible to maintain a low DC component and large filter bandwidth [108], [110]. Another argument Field used to justify usage of log-Gabor filters for image encoding is the symmetry of their frequency response on the log-frequency scale which as he claims 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 [108].

The frequency response of the log-Gabor function is described by:

$$G(f) = e^{-\frac{\log^2\left(\frac{f}{f_0}\right)}{2\log^2\left(\frac{\sigma}{f_0}\right)}} \quad (5.11)$$

where, f_0 is the filter center frequency while the term σ/f_0 dictates the filter bandwidth. The function $G(f)$ from expression 5.11 represents a Gaussian function of log-frequency. Since the log-Gabor filter has singularity at zero frequency, there is no analytic expression for the impulse response of the filter in the spatial domain. However, as described in [110] based on the analysis of their frequency response characteristic and by utilizing a numerical inverse Fourier transform, one can obtain the spatial form of these kernels. The computer code used for generating log-Gabor kernels for the purposes of this research is freely available [110]. The 2-D equivalent of the filter described by equation 5.11 has radial and angular components. The radial component is essentially identical to its 1-D counterpart while the angular component determines the orientation selectivity of the filter. Such a log-Gabor 2-D function with Gaussian model of angular component can be described as:

$$G(r, \theta) = e^{-\frac{\log^2\left(\frac{r}{r_0}\right)}{2\log^2\left(\frac{\sigma}{r_0}\right)}} e^{-\frac{(\theta - \theta_0)^2}{2\theta_\sigma^2}} \quad (5.12)$$

The first component of the filter is circular symmetric and is a function of the radial frequency r . The angular selectivity of the filter is controlled by parameters θ_0 and θ_σ . The inverse Fourier transform of the filter given by equation 5.12 is a complex 2-D signal with the real part representing the “even-symmetric component” of the resulting log-Gabor kernel, and the imaginary part representing the “odd-symmetric component” of the spatial complex response. When constructing the bank of log-Gabor filters one has

to consider various parameters depending on the desired application. The details of filter implementation as well as recommended filter bank parameters can be found in [108]. For the purposes of this research, two scale filter bank with four distinct orientations was used. The spatial response of the complex log-Gabor filter obtained numerically as the inverse Fourier transform of the frequency response given by equation 5.12 is shown in Figure 5.9 and Figure 5.10.

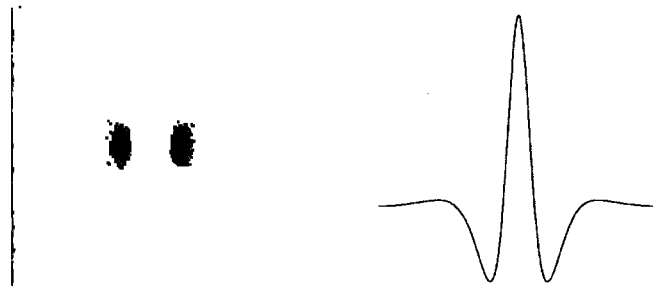


Figure 5.9 The spatial response of the even log-Gabor function; 2-D grey scale intensity model (left); center-line profile (right)

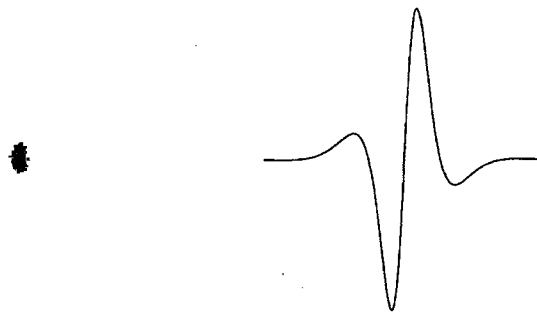


Figure 5.10 The spatial response of the odd log-Gabor Function; 2-D grey scale intensity model (left); center-line profile (right)

The spatial shape of the log-Gabor filter reminds us of the conventional Gabor function. The role of these filters in saliency detection as used in this research is primarily to detect image regions in the cortical space that would best stimulate cortical-like neural cells. Of course, one could also use Gabor-like kernels or even DOG type filters for this purpose. However, the choice to use log-Gabor functions was preferred because of their previously mentioned characteristics. The selection process of points of interest is based on the magnitude of the complex log-Gabor filter bank responses. For each location in the cortical space the overall magnitude of complex log-Gabor filter responses is calculated as the sum of magnitudes of all individual responses to filters belonging to the filter bank. For each scale and orientation the magnitude of filter response can be calculated as:

$$I_{s,\theta} = \sqrt{I_{even_{s,\theta}}^2 + I_{odd_{s,\theta}}^2}. \quad (5.13)$$

The overall response is given as:

$$I = \sum_{s=1}^2 \sum_{i=1}^4 I_{s,\theta_i}. \quad (5.14)$$

The selection is performed by searching for locations in the log-polar domain with maximum overall response, I . In order to achieve adequate separation of saliency locations as well as to allow the algorithm to properly advance from one salient location to the next one, it is necessary to effectively eliminate all already attended salient locations from further search. Once the maximum of I is found, the response function given by equation 5.14 at that location and the surrounding region is set to zero, i.e. it is locally inhibited. In this manner, the algorithm will search for the next significant response identifying the potential location of the new salient patch. The neighborhood size to be excluded from further search was 8 pixels. The number of pseudo-salient

locations chosen for processing was 16. Once the appropriate collection of locations in the cortical image with maximum overall response to the chosen log-Gabor filter bank is found, the corresponding coordinate pairs were sorted according to descending angular coordinate. In the case two locations with the same angular coordinate were found, the one with the lower log-radial coordinate preceded the one with the higher log-radial coordinate. The overall result of presented pseudo saliency detection process is the vector of 16 coordinate pairs with specific, ordered placement of salient locations in the vector itself. An example of the image in log-polar space with the resulting set of pseudo-salient locations is shown in Figure 5.11. Corresponding patches extracted from salient locations outlined in the right pane of Figure 5.11 are shown in Figure 5.12.

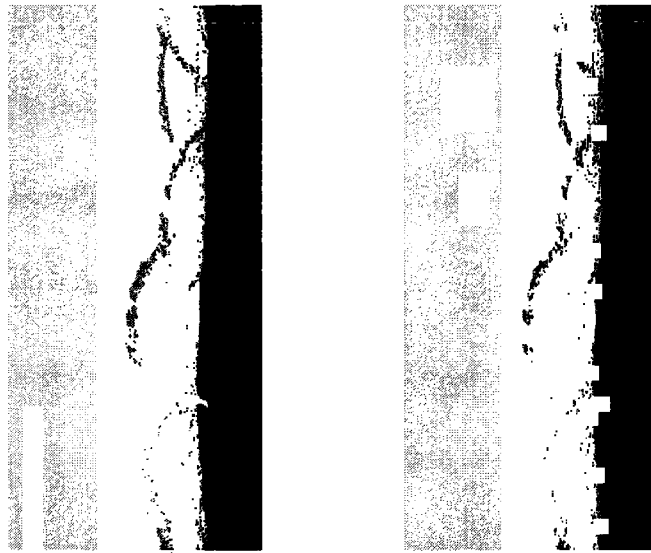


Figure 5.11 An image in log-polar space (left) with outlined pseudo-salient locations (right)

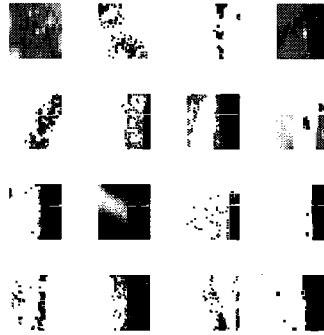


Figure 5.12 Collection of patches centered at pseudo-salient points from Figure 5.11

The actual response to log-Gabor filter bank is obtained by calculating the Fourier transform of given log-polar image, then multiplying this transform by the frequency response of particular filter member of the filter bank and finally the inverse Fourier transform of each of the products is taken.

The feature vector characterizing each object can be represented as a matrix containing 16 columns – each column encoding one salient location. The first two components in the column are the x and y coordinates of the location center. The next 256 components of each column represent an ICA decomposition of the higher scale patch while last 256 components are the ICA decomposition of the lower scale patch centered at the same salient coordinate. For each patch, the higher scale data was identified as a block of size 32×32 . This block is subsequently processed by a low-pass Gaussian filter, down-sampled to a 16×16 patch and then further processed as a standard data vector containing 256 components. In order to estimate the statistics of the given image, an additional 2000 patches of size 16×16 were extracted from random locations and added to the data representing salient locations. This data set was normalized to unit

variance and zero mean. The ICA decomposition as described by equation 5.7 is performed by first pre-processing the data with pre-whitening kernels learned on the set of log-polar warped images of objects placed on the dark background according to the algorithm described in Figure 4.4. Here, the efficiency of pre-whitening filters realized as recurrent networks with symmetric Toeplitz-Block-Toeplitz connection matrix (with symmetric Toeplitz blocks) can be fully utilized. By using a very small number of coefficients forming the filter (i.e. initial network weights for data whitening) it is possible to pre-process the data such that complete whitening by the adaptive algorithm defined by equation 4.4 is achieved in a small number of iterations as shown in Figure 4.9. The same principles of signal pre-processing were followed when obtaining (learning) the actual matrix W from 5.7. This matrix was obtained by implementing the fast ICA algorithm on the additional data from the same category and by implementing identical whitening procedures. The matrix W in this case can be described as:

$$W = B_1 B_{WT} \quad (5.15)$$

where, B_{WT} represents the entire two-phase whitening transformation of the input data. The estimated ICA representation of a given image patch is based on the modification of equation 5.7 affected by data whitening:

$$\bar{s} = B_1 \hat{x} \quad (5.16)$$

where, \hat{x} represents whitened data x . The overall extracted feature vector can be quite large – 16 salient locations are represented as a 514×16 matrix. However, given the nature of ICA and its tight relationship to sparse coding, it is natural to expect that many of the given feature vectors (actually matrices) can be sparse or can adequately encode the given object by taking into consideration only certain and most probably small

number of largest coefficients in the overall ICA representation. This led to an additional transform of the given feature vectors into structures with only pre-determined number of non-zero ICA coefficients. For each scale, only certain number of highest (by absolute value) ICA coefficients were kept intact while others were set to zero. Thus, a very sparse data structure was created. The resulting feature matrix FM can be described as:

$$FM = \begin{bmatrix} x_1 & x_2 & \cdots & x_{16} \\ y_1 & y_2 & \cdots & y_{16} \\ hs_{1,1} & hs_{2,1} & \cdots & hs_{16,1} \\ \vdots & \vdots & \vdots & \vdots \\ hs_{1,256} & hs_{2,256} & \cdots & hs_{16,256} \\ ls_{1,1} & ls_{2,1} & \cdots & ls_{16,1} \\ \vdots & \vdots & \cdots & \vdots \\ ls_{1,256} & ls_{2,256} & \cdots & ls_{16,256} \end{bmatrix} \quad (5.17)$$

where, x_i and y_i represent coordinates of the salient patch center in the log-polar image, $hs_{i,j}$ represent the ICA coefficient of the higher-scale patch while $ls_{i,j}$ are ICA coefficients of the lower-scale patch.

At the final stage of information processing, pattern classification is performed. Different types of distance based classification engines were tested: nearest neighbor, 3-nearest neighbor [111] and a specific modification of the previous two types of classifiers. It was noticed that the significant improvement over traditional nearest-neighbor type classification is achieved if the feature matrix is considered as a collection of un-ordered 16 sub-vectors each representing the salient image location. Each sub-vector was classified as belonging to a particular object/class based on minimum distance to the one of the sub-vectors representing salient patches in the training set. This way, for each test pattern there would be a voting procedure based on

number of features/patches identified to belong to the same class. The class with the highest vote would be the winner providing it has more than 8 votes (belonging features). This algorithm of classification is named hybrid nearest neighbor classifier. 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. Again, the final winner is the object/class characterized with the highest count of patches that belong to it. If the recognition is performed on potentially rotated object, a variable offset is added to the coordinate x of each column vector in FM thus achieving very fast rotation of the input test pattern. For each pre-determined rotation offset the winner class is found based on hybrid nearest neighbor classifier. Also, for each offset additional parameters are determined representing the number of features in the test pattern belonging to the particular class and the cumulative distance between all feature sub-vectors and corresponding closest feature sub-vectors in the training set. The pattern resulting in minimum cumulative distance parameter is the one chosen as input for final classification. If this pattern contains a majority of features belonging to the same class, then this class is proclaimed as winner. Otherwise, no decision is made. Alternatively, when implementing the hybrid 3-nearest neighbor classifier, for each rotation offset, besides deciding the winner class, the confidence value is determined representing the number of times any of the feature sub-vectors was classified to belong to the winning object based on 3-nearest neighbor rule. The overall winner is the object with highest confidence count.

5.3 Experimental Results

Finally, the results of signal processing and pattern classification algorithms described in the previous section are presented. The entire algorithm excluding the log- polar warping was implemented in MATLAB (the same stands for all experimental results presented in this paper). The actual computations were executed in off-line and batch mode, and no parameters were obtained that describe the system performance in terms of speed of execution. In all of the experiments the number of non zero ICA components was varied from a minimum 16 to a maximum of 256. Also, image rotations were introduced artificially through software by rotating the particular image about the log-polar transform origin by some random angle that was in the range $(-10^\circ, 10^\circ)$. The experimental results obtained by application of four different classification techniques are summarized in Table 5.1 and Table 5.2.

Number of non-zero ICA coefficients	Recognition result pass/fail for specific classifier			
	Nearest Neighbor	3-Nearest Neighbor	Hybrid Nearest Neighbor	Hybrid 3-Nearest Neighbor
256	143/17	137/23	155/5	159/1
80	143/17	136/24	156/4	159/1
64	143/17	134/26	156/4	159/1
48	143/17	135/25	156/4	159/1
32	143/17	135/25	156/4	159/1
16	143/17	135/25	155/5	159/1

Table 5.1 The results of non-rotated object recognition by four different classifiers

Number of non-zero ICA coefficients	Recognition result pass/fail for specific classifier	
	Hybrid Nearest Neighbor	Hybrid 3-Nearest Neighbor
256	153/7	159/1
80	153/7	159/1
64	152/8	159/1
48	152/8	159/1
32	152/8	159/1
16	154/6	159/1

Table 5.2 The results of object recognition by hybrid classifiers (rotated objects as input)

By comparing the results of object classification it can be concluded that the hybrid method is far superior to the traditional distance based classification. The reason to this might be in the fact that potential variations in the selection of salient locations (one or two salient locations differ or missed) might lead to very large distances between two feature vectors even when they contain large number of “correct” (i.e. identical) salient features. Also, certain experiments showed consistent pass/fail ratio for all possible variations of number of non-zero ICA components. Without going into detailed analysis of characteristics of each of the addressed classifiers, we point out few potential causes of this phenomenon. First, errors can happen in the pre-processing stages of feature extraction - if the crucial features are for example missed, classification based on any number of ICA components will fail. Second, since the signal encoding is sparse, only few “strong” non-zero coefficients in ICA representation are actually relevant and varying their number does not influence the result. The third factor would be related to potential dominance of x and y coordinates in the feature matrix over other feature components. If this dominance is significant one would require proper feature

normalization. While recognition rates for the case of recognition with hybrid classifiers⁴ when x and y components are normalized to range (0, 5) are presented in Table 5.3, we note that detailed feature normalization analysis was out of scope of this research.

Number of non-zero ICA coefficients	Recognition result pass/fail – non-rotated objects		Recognition result pass/fail – rotated objects	
	Hybrid Nearest Neighbor	Hybrid 3-Nearest Neighbor	Hybrid Nearest Neighbor	Hybrid 3-Nearest Neighbor
256	148/12	146/14	147/13	151/9
80	147/13	146/14	146/14	148/12
64	146/14	147/13	146/14	148/12
48	145/15	147/13	143/17	149/11
32	146/14	148/12	144/16	147/13
16	145/15	146/14	140/20	147/13

Table 5.3 The results of object recognition by hybrid classifiers and for x and y feature components normalized to range (0, 5)

In order to additionally test the potential of the local ICA based feature encoding, the experiments were performed for non-rotated object classification and for the case when x and y features components are eliminated, i.e. their value is set to zero. The results of these experiments are presented in Table 5.4. The location of salient patches is obviously not robustly encoded in the feature vector and higher, but not catastrophic misclassification rate is evident. Standard nearest neighbor and 3-nearest neighbor classifiers performed very poorly in this test.

⁴ Standard nearest neighbor and 3-nearest neighbor classifiers did not perform well in this case.

Number of non-zero ICA coefficients	Recognition result pass/fail for specific classifier	
	Hybrid Nearset Neighbour	Hybrid 3-Nearest Neighbour
256	148/12	146/14
80	147/13	146/14
64	146/14	147/13
48	145/15	147/13
32	146/14	148/12
16	145/15	146/14

Table 5.4 The results of non-rotated object recognition by hybrid classifiers and for eliminated x and y feature components

All results obtained in this section demonstrate that addressed methods of signal pre-processing and feature extraction are very promising in solving pattern recognition problems. One has to observe that the significant strength of the local ICA encoding technique is a relatively small number of non-zero elements in the feature representation of each pattern. When the number of non-zero components in the feature matrix is significantly reduced, the classification performance remains practically intact. This fact leads to the conclusion that each input is properly encoded by only the few most significant ICA coefficients in the resulting sparse representation – feature matrix itself is actually sparse. Effectiveness of the proposed feature encoding algorithm is additionally demonstrated in Appendix B by quantitatively describing how sparse resulting codes really are. The overall signal pre-processing before ICA encoding is achieved by a combined deployment of category-specific, pre-learned whitening filters as well as the adaptive recurrent neural network. Such data processing is effective only in cases when the system knows *a priori* what category of signal can be expected as the input from the environment at particular time instance so that proper pre-whitening filter is deployed. Rotation variations did not present any major challenge to the model, but it is noted that

these are obtained artificially by assuming the correct center of the log-polar transform for each object of interest. Scaling variations in the input pattern were not simulated since they represent analogous shifts via different axis to those of rotations once the signal is mapped to log-polar space.

CHAPTER VI

CONCLUSION AND DIRECTIONS

Biologically inspired signal processing techniques and their practical utilization have for a long time been a very exciting and attractive branch of scientific research. This dissertation addressed the problem of efficient utilization of recurrent linear networks with inhibitory/excitatory connectivity in the context of log-polar coordinate mapping for signal processing and pattern recognition purposes. Both recurrent neural connectivity and the complex-log transform are viewed as strongly biologically inspired signal processing operations. It is argued that the natural way of using the neural recurrent model as a supplement to the non-uniform log-polar image transformation is uniform processing of information in the post-warped, or cortical domain. Such information processing as seen from the input (non-warped) domain correlates to spatially non-uniform operations and is similar to actions observed in the early stages of the HVS. The resulting receptive fields in the case of lateral inhibition at the output are space-variant and of center-surround type. One of the very important results demonstrated is the characterization of the response of the spatially uniform inhibitory/excitatory neural models based on the synaptic connectivity of the network. The network filter response can be obtained through implementation of a Taylor series expansion providing the generating function of the connectivity matrix is bounded by the unity. Also, it was shown that if the neural connectivity function is expandable in a Fourier series in polar angle, the network filter response becomes steerable. This is a

very intriguing result since steerability is often associated with biologically inspired filters like “derivative of Gaussian” family of functions. Theoretical results revealed in Chapter III of this dissertation pertaining to the network steerability as well as the characterization of the network response based on the bounds of the connectivity matrix generating function can effectively be used to compute the true response of the large neural models where matrix inversion has to be avoided. Another powerful feature of networks with lateral inhibition/excitation lies in their capability to decorrelate and whiten the input data. This process is extremely important for the statistically based signal processing technique of ICA. It was shown that recurrent networks with symmetric TBT connectivity matrix (with symmetric Toeplitz blocks) can be used to develop compact pre-whitening filters with small memory requirements. These filters are learned for specific category of signal input and can significantly improve the performance of whitening algorithms based on adaptive networks with recurrent connectivity. Finally, their usage for localized ICA based image/signal encoding in a log-polar environment produced extremely desired feature of sparse code at the output. Such sparse codes when combined with a specific selection of interesting regions in the cortical domain for feature encoding produced very promising pattern recognition results.

The learning of the pre-whitening filter as described in this dissertation utilizes computations based on direct matrix inversion. Even with today’s computational platforms this operation becomes undesirable and it is necessary to find a less computationally demanding method for filter response calculations. This in turn directs one to investigate the possibilities of using an adaptation of the network connectivity matrix which would observe and incorporate results from Chapter III while learning the

pre-whitening filter. It is shown that this filter response can be computed without implementing matrix inversion.

It can be noticed that pre-whitening filters were developed, for a specific category of input images. It is also evident that there is no strict criterion that each image must satisfy in order to be classified as belonging to one of these categories. One of the interesting areas for future work that would provide a more distinct answer to the question of which pre-whitening filter to deploy at a particular time would be related to identifying the minimum or optimal set of characteristics of the input image/data that would classify that input into the proper signal category. One possible solution would most likely investigate characteristics of the input data covariance matrix. If the input data has predominant features emphasized along one or few directions, the usage of steerable filters (or recurrent steerable networks) can elegantly reveal additional information and provide an additional data set which may better characterize the input signal category (steerable filters would have one type of response to predominantly horizontal texture and completely different response to texture of predominantly different orientation).

To prove out some of the hypotheses presented in this dissertation a few (essentially all distance based) pattern recognition engines were used when validating the feature encoding model based on data pre-whitening and the local ICA transform. It is very likely that more robust techniques can be identified and deployed within the given framework. One obvious approach would be the deployment of neural pattern recognition models that are also biologically inspired. When searching for alternative, more efficient classifiers one has to take under consideration that although sparse the

current feature encoding model suffers from large dimensionality. The reduction of the feature vector/matrix has to be yet another focus of future efforts in order for presented information processing framework to be more practical. The existence of a highly sparse information code provides a very promising basis for future improvements of the presented feature extraction model. Finally, it is also highly desirable to assess the performance of the given model when deployed on more complex (and less restricted) pattern recognition problems. It is the author's hope that future work will help us to understand the complex processing that occurs in the most complex pattern recognition system known to man – man himself!

REFERENCES

- [1] Y. Bar-Cohen, "Biologically Inspired Intelligent Robots Using Artificial Muscles," in *Proc. of 2003 International Conference on MEMS, NANO and Smart Systems (ICMENS'03)*, Banff, Alberta, Canada, July 2003, pp. 2-8 2003.
- [2] M. Bolduk and M. D. Levine, "A review of biologically motivated space-variant data reduction models for robotic vision," *Computer Vision and Image Understanding*, vol. 69, no. 2, pp. 170-184, February 1998.
- [3] D. C. Van Essen and C. H. Anderson, "Information processing strategies and pathways in the primate retina and visual cortex," in *An Introduction to Neural and Electronic Networks*, S.F. Zornetzer, J.L. Davis, and C. Lau, Eds. Academic Press, San Diego, 1990, pp. 43-72 1990.
- [4] V. H. Perry, R. Oehler and A. Cowey, "Retinal ganglion cells that project to the dorsal lateral geniculate nucleus in macaque monkeys," *Neuroscience*, vol. 12, pp. 1101-1123, August 1984.
- [5] S. Shah and M. D. Levine, "Visual information processing in primate cone pathways: Part I, a model," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 26, pp. 259-274, 1995.
- [6] R. E. Kronauer and Y. Y. Zeevi, "Reorganization and diversification of signals in vision," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 15, pp. 91-101, 1985.
- [7] H. Wässle, U. Grünert, J. Röhrenbeck and B. B. Boycott, "Cortical magnification factor and the ganglion cell density of the primate retina," *Nature*, vol. 341, pp. 643-646, October 1989.
- [8] H. R. Wilson, D. Levi, L. Maffei, J. Rovamo, and R. DeValois, *The Perception of Form: Retina to Striate Cortex*. Academic Press, San Diego, 1990, pp. 103-127.
- [9] P. M. Daniel and D. Whitteridge, "The representation of the visual field on the cerebral cortex in monkeys," *The Journal of Physiology*, vol. 159, pp. 203-221, December 1961.
- [10] D. M. Hubel and T. N. Wiesel, "Uniformity of monkey striate cortex: a parallel relationship between field size, scatter and magnification factor," *Journal of Comparative Neurology*, vol. 158, pp. 295-302, 1974.
- [11] B. Fisher, "Overlap of receptive field centers and representation of the visual field in the cat's optic tract," *Vision Research*, vol. 13, no. 11, pp. 2113-2120, 1973.
- [12] E. Schwartz, "Spatial mapping in the primate sensory projection: analytic structure and relevance to perception," *Biological Cybernetics*, vol. 25, pp. 181-194, 1977.

- [13] E. Schwartz, "Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding," *Vision Research*, vol. 20, pp. 645-669, 1980.
- [14] E. L. Schwartz, *Computational Studies of the Spatial Architecture of Primate Visual Cortex*, vol. 10, chap. 9, pp. 359-411, Plenum, New York, 1994.
- [15] H. K. Hartline, H. G. Wagner and F. Ratliff, "Inhibition in the eye of limulus," *Journal of General Physiology*, vol. 39, pp. 651-673, 1956.
- [16] H. K. Hartline and F. Ratliff, "Inhibitory interaction of receptor units in the eye of Limulus," *Journal of General Physiology*, vol. 40, pp. 357-376, 1957.
- [17] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *Journal of Physiology*, vol. 160, pp. 106-154, 1962.
- [18] C. Blackmore and E. A. Tobin, "Lateral inhibition between orientation detectors in the cat's visual cortex," *Experimental Brain Research*, vol. 15, pp. 439-440, 1972.
- [19] K. Funke, Z. F. Kisvaraday, M. Volgushev and F. Worgotter, "Integrating Anatomy and Physiology of the Primary Visual Pathway: From LGN to Cortex," in *Models of Neural Networks IV – Early Vision and Attention*, J. L. van Hemmen, J. D. Cowan and E. Domany, Eds. New York: Springer-Verlag New York, Inc., pp. 97-171, 2001.
- [20] B. A. Wandell, *Foundations of Vision*. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts, pp.144-145, 1995.
- [21] C. Braccini, G. Gambardella, G. Sandini and V. Tagliasco, "A model of the early stages of the human visual system: Functional and topological transformations performed in the peripheral visual field," *Biological Cybernetics*, vol. 44, pp. 47-58, May 1982.
- [22] L. Maffei and A. Fiorentini, "The visual cortex as a spatial frequency analyzer," *Vision Research*, vol. 13, pp. 1255-1267, 1973.
- [23] S. Marcelja, "Mathematical description of the responses of simple cortical cells," *Journal of Optical Society of America*, vol. 70, pp. 1297-1300, November 1980.
- [24] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vision Research*, vol. 20, pp. 847-856, 1980.
- [25] J. G. Daugman, "Uncertainty relations for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *Journal of Optical Society of America*, vol. 2, pp. 1160-1169, July 1985.
- [26] J. P. Jones, and L. A. Palmer, "An evaluation of two dimensional Gabor filter model of simple receptive fields in cat striate cortex," *Journal of Neurophysiology*, vol. 58, pp. 1233-1258, December 1987.
- [27] H. B. Barlow, "Unsupervised learning," *Neural Computation*, vol. 1, pp. 295-311, 1989.
- [28] 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.

- [29] 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.
- [30] A. J. Bell and T. J. Sejnowski, "The 'Independent Components' of Natural Scenes are Edge Filters," *Vision Research*, vol. 37, pp. 3327-3338, 1997.
- [31] B. A. Olshausen and D. J. Field, "Natural Image Statistics and Efficient Coding," *Network: Computation in Neural Systems*, vol. 7, pp. 333-339, 1996.
- [32] R. A. Messner and H. H. Szu, "An image processing architecture for real time generation of scale and rotation invariant patterns," *Computer Vision, Graphics, and Image Processing*, vol. 31, pp. 50-66, July 1985.
- [33] G. Sandini, P. Questa, D. Scheffer, B. Diericx and A. Mannucci, "A retina-like CMOS sensor and its applications," in *Proc. of the 2000 IEEE Sensor Array and Multichannel Signal Processing Workshop. 2000*, Cambridge, USA, 2000, pp. 514-519.
- [34] A. Moini. (1997 March). Vision chips: seeing silicon [Online]. Available: http://www.see.et.tu-dresden.de/see/eb/analog/papers/mirror/visionchips/vision_chips/vision_chips.html
- [35] D. Young, "Straight lines and circles in the log-polar image," presented at The Eleventh British Machine Vision Conference, University of Bristol, September 11-14, 2000.
- [36] G. Wolberg and S. Zokai, "Robust image registration using log-polar transform," in *Proc. of IEEE Intl. Conference on Image Processing*, Vancouver, BC, Canada, 2000, pp. 493-496.
- [37] E. Nattel and Y. Yeshurun, "An Efficient Data Structure for Feature Extraction in a Foveated Environment," in *Proc. of Biologically Motivated Computer Vision: First IEEE International Workshop, BMCV 2000*, Seoul, Korea, 2000, *Lecture Notes in Computer Science*, vol. 1811, pp. 217-226, 2000.
- [38] E. Nattel and Y. Yeshurun, "Direct feature extraction in a foveated environment," *Pattern Recognition Letters*, vol.23 no.13, pp.1537-1548, November 2002.
- [39] H. M. Gomes and R. B. Fisher, "Learning-based versus model-based log-polar feature extraction operators: a comparative study," in *Proc XVI Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI'03)*, San Carlos, Brazil, October 2003, pp. 299 - 306.
- [40] F. Smeraldi and J. Bigun, "Retinal vision applied to facial features detection and face authentication," *Pattern Recognition Letters*, vol. 23, pp. 463-475, February 2002.
- [41] V. J. Traver, A. Bernardino, P. Moreno and J. Santos-Victor, "Appearance-based object detection in space-variant images: a multi-model approach," presented at International conference on Image Analysis and Recognition, Porto, Portugal, September 29-October 1, 2004. Available: <http://homepage.inf.ed.ac.uk/rbf/CAVIAR/PAPERS/04-iciar-faces.pdf>
- [42] B. B. Bederson, R. S. Wallace and E. Schwartz, "A miniaturized space-variant active vision system: Cortex I," *Machine Vision and Applications*, vol. 8, no. 2, pp. 101-109, February 1995.

- [43] J. G. Bailey and R. A. Messner, "Docking target design and spacecraft tracking system stability," *Proc. SPIE Intelligent Robots and Computer Vision VIII: Algorithms and Techniques*, vol. 1192, pp. 820-831, March, 1990.
- [44] C. F. Weiman and R. D. Juday, "Tracking algorithms using log-polar-mapped image coordinates," *Proc. SPIE Intelligent Robots and Computer Vision VIII: Algorithms and Techniques*, vol. 1192, pp. 843-853, , March, 1990.
- [45] D. Comaniciu, F. Berton and V. Ramesh, "Adaptive Resolution System for Distributed Surveillance," *Real-Time Imaging*, vol.8, pp. 427-437, October 2002.
- [46] L. Mu, "The lateral subtractive inhibition algorithm: Simulation and application in image processing," Master Thesis, University of New Hampshire, 1995.
- [47] G. G. Furman, "Comparison of models for subtractive and shunting lateral-inhibition in receptor-neuron fields," *Biological Cybernetics*, vol. 2, pp. 257-274, October 1965.
- [48] R.A. Messner, *Smart Visual Sensors for Real-Time Image Processing and Pattern Recognition Based Upon Human Visual System Characteristics*, PhD Dissertation, Clarkson University, 1984.
- [49] Op. Cit, R. A. Messner, 1984, pp. 59-69.
- [50] H.H. Szu and R.A. Messner, "Adaptive invariant novelty filters," *Proceedings of IEEE*, vol. 74, pp. 518-519, 1986.
- [51] Y. Yu, T. Yamauchi and Y. Choe, "Explaining low-level brightness-contrast illusions using disinhibition," in *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.
- [52] 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.
- [53] 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.
- [54] G. Bonmassar and E. L. Schwartz, "Fourier analysis and cortical architectures: the exponential chirp transform," *Real-time imaging*, vol. 3, no. 3, pp. 229-237, June 1997.
- [55] R. S. Wallace, P. W. Ong, B. B. Bederson and E. L. Schwartz, "Space variant image processing," *International Journal of Computer Vision*, vol. 13, no. 2, pp. 71-90, September 1994.
- [56] P. Földiák and M. P. Young, "Sparse Coding in the Primate Cortex," *The handbook of brain theory and neural networks*, pp. 895-898, MIT Press, Cambridge, MA, USA 1998.
- [57] B. A. Olshausen and D. J. Field, "Sparse Coding of Sensory Inputs," *Current Opinion in Neurology*, vol. 14, pp. 481-487, 2004.

- [58] A. Hyvärinen, "Survey on Independent Component Analysis," *Neural Computing Surveys*, vol. 2, pp. 94-128, 1999.
- [59] A. J. Bell and T. J. Sejnowski, "An Information-Maximization Approach to Blind Separation and Blind Deconvolution" Technical Report no. INC-9501, Institute for Neural Computation, UCSD, San Diego, CA, February 1995.
- [60] A. Hyvärinen, E. Oja, "Independent Component Analysis: Algorithms and Applications," *Neural Networks*, vol. 13, pp. 411-430, 2000.
- [61] A. Hyvärinen, "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis," *IEEE Transactions on Neural Networks*, vol. 10, pp. 626-634, 1999.
- [62] The web page for FastICA algorithm, The Department of Computer Science, University of Helsinki, retrieved from <http://www.cs.helsinki.fi/u/ahyvarin/papers/fastica.shtml>
- [63] 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, pp. 54-72.
- [64] M. D. Plumbley, "Efficient Information Transfer and anti-Hebbian neural Networks," *Neural Networks*, vol. 6, pp. 823-833, 1993.
- [65] 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.
- [66] H.K. Hartline and F. Ratliff, "Spatial summation of inhibitory influences in the eye of Limulus, and the mutual interaction of receptor units," *The Journal of General Physiology*, vol. 41, pp. 1049-1066, 1958.
- [67] B.S. Gutkin and C.E. Smith, "Conditions for noise reduction and stable encoding of spatial structure by cortical neural networks," *Biological Cybernetics*, vol. 82, pp. 469-475, May 2000.
- [68] S.E. Palmer, *Vision Science Photons to Phenomenology*, A Bradford Book, The MIT Press, Cambridge, Massachusetts, London, England, 1999.
- [69] Y. Yu and Y. Choe, "Angular disinhibition effect in a modified Poggendorff illusion," in *Proceedings of the 26th Annual Conference of the Cognitive Science Society, Mahwah, NJ, USA 2004*, pp. 1500-1505.
- [70] M. Wax and T. Kailath, "Efficient inversion of Toeplitz-block Toeplitz matrix," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 31, pp. 1218-1221, October 1983.
- [71] N. Kalouptsidis, G. Carayannis, and D. Manolakis, "On block matrices with elements of special structure," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '82*, vol. 7, pp. 1744-1747, Paris, France May 1982.

- [72] A.E. Yagle, "A fast algorithm for Toeplitz-block-Toeplitz linear systems," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '01*, vol. 3, pp. 1929-1932, Salt Lake City, USA, 2001.
- [73] S.J. Reeves, "Fast algorithm for solving block banded Toeplitz systems with banded Toeplitz blocks," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '02*, vol. 4, pp. 3325-3328, Orlando, USA, 2002.
- [74] D.A. Bini and B. Meini, "Solving block banded block Toeplitz systems with banded Toeplitz blocks," in *Proceedings of SPIE, Advanced Signal Processing Algorithms, Architectures, and Implementations IX*, vol. 3807, pp. 300-311, 1999.
- [75] U. Grenander and G. Szegő, *Toeplitz Forms and Their Applications*, 2nd ed., Chelsea Publishing, New York, 1984.
- [76] S. Serra, "Preconditioning strategies for asymptotically ill conditioned block Toeplitz systems," *BIT Numerical Mathematics*, vol. 34, pp. 579-594, December 1994.
- [77] P. Tilli, "On the asymptotic spectrum of Hermitian block Toeplitz matrices with Toeplitz blocks," *Mathematics of Computation*, vol. 66, pp. 1147-1159, 1997.
- [78] 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.
- [79] S. Barnett, *Matrices: Methods and Applications*, Oxford University Press, New York, 1990.
- [80] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley Publishing Company, 1992, pp. 97.
- [81] E.M. Stein and G. Weiss, *Introduction to Fourier analysis on Euclidean spaces*, Princeton University Press, Princeton, New Jersey, 1971.
- [82] 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.
- [83] P. Földiák, "Adaptive Networks for Optimal Linear Feature Extraction," in *Proceedings of IEEE International Conference on Neural Networks*, vol. 1, pp. 401-405, New York, 1989.
- [84] R. W. Brause, "A Symmetric Lateral Inhibition Network for PCA and Feature Decorrelation," in *Proceedings of International Conference on Artificial Neural Networks ICANN-93*, Springer Verlag, pp. 486-489, 1993.
- [85] F. Palmeieri and J. Zhu, "A Comparisson of Two Eigen-Networks," in *Proceedings of International Joint Conference on Neural Networks IJCNN-91*, vol. 2, pp. 193-199, Seattle, WA, USA, 1991.
- [86] F. Palmeiri, J. Zhu and C. Chang, "Anti-Hebbian Learning in Topologically Constrained Linear Networks: A Tutorial," *IEEE Transactions on Neural Networks*, vol. 4, pp. 748-761, September 1993.

- [87] D. Obradovic and G. Deco, "Linear Feature Extraction in Networks with Lateral Connections," in *Proceedings of 1994 International Conference on Neural Networks, IEEE World Congress on Computational Intelligence*, vol. 2, pp. 686-691, Orlando, FL, USA, 1994.
- [88] T. D. Sanger, "Optimal Unsupervised Learning in a Single-Layer feedforward neuronal Network," *Neural Networks*, vol. 2, pp. 459-473, 1989.
- [89] M. D. Plumbley, "A Hebbian/anti-Hebbian network which optimizes information capacity by orthonormalizing the principal subspace," in *Proceedings of Third International Conference on Artificial Neural Networks, 1993*, pp. 86-90, Brighton, UK, 1993.
- [90] E. Oja, "Principal Components, Minor Components and Linear Neural Networks," *Neural Networks*, vol. 5, pp. 927-935, 1992.
- [91] E. Oja, "A Simplified Neuron Model as a Principal Component Analyzer," *Journal of Mathematical Biology*, vol. 15, pp. 267-273, 1982.
- [92] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*, John Wiley & Sons, LTD, 2002, pp. 129.
- [93] P. O. Hoyer and A. Hyvärinen, "Independent Component Analysis Applied to Feature Extraction from Colour and Stereo Images," *Network: Computation in Neural Systems*, vol. 11, pp. 191-210, 2000.
- [94] R. W. Brause "Transform Coding by Lateral Inhibited Neural Nets," in *Proceedings of 1993 IEEE International Conference on Tools with AI*, pp. 14-21, Boston, Massachusetts, November 1993.
- [95] P. Wintz, "Transform Picture Coding," *Proceedings of IEEE*, vol. 60, pp. 809-820, 1972.
- [96] E. P. Simoncelli and B. A. Olshausen, "Natural Image Statistics and Neural Representation," *Annual Review of Neuroscience*, vol. 24, pp. 1193-1216, March 2001.
- [97] D. H. Hubel and T. N. Wiesel, "Receptive Fields and Functional Architecture of Monkey Striate Cortex," *The Journal of Physiology*, vol. 195, pp. 215-244, 1968.
- [98] P. O. Hoyer and A. Hyvärinen, "A Multi-layer Sparse Coding Network Learns Contour Coding from Natural Images," *Vision Research*, vol. 42, pp. 1593-1605, 2002.
- [99] A. J. Bell and T. J. Sejnowski, "An Information-maximization Approach to Blind Separation and Blind Deconvolution," *Neural Computation*, vol. 7, pp. 1129-1159, 1995.
- [100] M. S. Bartlett, J. R. Movellan and T. J. Sejnowski, "Face Recognition by Independent Component Analysis," *IEEE Transactions on Neural Networks*, vol. 13, pp. 1450-1464, 2002.
- [101] P. O. Hoyer and A. Hyvärinen, "Independent Component Analysis Applied to Feature Extraction from Color and Stereo Images," *Network: Computation in Neural Systems*, vol. 11, pp. 191-210, 2000.
- [102] A. Hyvärinen and E. Oja, "A Fast Fixed-Point Algorithm for Independent Component Analysis," *Neural Computation*, vol. 9, pp. 1483-1492, 1997.

- [103] M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders, "The Amsterdam Library of Object Images," *International Journal of Computer Vision*, vol. 61, pp. 103-112, January, 2005.
- [104] 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/>
- [105] N. Otsu, "A threshold selection method from gray level histograms," *IEEE Transactions on Systems Man and Cybernetics*, vol. 9, pp. 62-66, 1979.
- [106] Op. Cit, R. C. Gonzalez and R. E. Woods, pp. 514.
- [107] The web page of OpenCV library at SourceForge.net, obtained from: <http://sourceforge.net/projects/opencvlibrary/>
- [108] 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/>.
- [109] L. Itti, C. Koch and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, pp. 1254-1259, November 1998.
- [110] 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.
- [111] 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.

APPENDIX A

ADDITIONAL RESULTS DESCRIBING DECORRELATION AND PRE-WHITENING BY FILTERS BASED ON NETWORKS WITH RECURRENT CONNECTIVITY

In this section additional experimental results revealing decorrelation, as well as pre-whitening property of filters based on recurrent linear networks are provided. For non-adaptive network, the outcomes of processing the category 2 and category 3 images (image categories are defined in Chapter IV of this text) with filters identified by Model A and Model B as well as kernels based on Gaussian inhibitory coupling with and without the self-inhibition (kernel parameters are identical to those used in section 4.4) are given in Table A.1 and Table A.2.

Group Number (each group = 5 images)	$\sum_{i \neq j} C_X(i, j) $	$\sum_{i \neq j} C_Y(i, j) $ Model A	$\sum_{i \neq j} C_Y(i, j) $ Model B	$\sum_{i \neq j} C_Y(i, j) $ Gaussian without self- inhibition	$\sum_{i \neq j} C_Y(i, j) $ Gaussian with self- inhibition
1	1535.4207	227.2067	181.7874	668.3953	461.7766
2	3164.8795	471.7852	379.7187	1382.6587	953.6703
3	3185.9071	473.9958	381.0495	1391.9602	959.9482
4	2872.5191	427.1749	343.2601	1253.8520	865.2592

Table A.1 The results of decorrelation with filters based on non-adaptive uniform networks with lateral inhibition – images from category 2 used for processing

Group Number (each group = 5 images)	$\sum_{i \neq j} C_X(i, j) $	$\sum_{i \neq j} C_Y(i, j) $ Model A	$\sum_{i \neq j} C_Y(i, j) $ Model B	$\sum_{i \neq j} C_Y(i, j) $ Gaussian without self-inhibition	$\sum_{i \neq j} C_Y(i, j) $ Gaussian with self-inhibition
1	11190.6946	1709.1329	1401.3824	4973.1744	3409.2754
2	11311.0326	1713.9901	1396.9882	5001.6464	3433.9132
3	12286.4649	1866.9821	1525.0382	5442.4803	37344.6612
4	13129.3181	1988.2022	1619.7666	5803.6619	3983.5492

Table A.2 The results of decorrelation with filters based on non-adaptive uniform networks with lateral inhibition – images from category 3 used for processing

Coefficients representing the measure of decorrelation from Table A.1 and Table A.2 are highly in agreement with results shown in Table 4.1.

When performing data whitening while using category-specific pre-whitening filters learned according to algorithm addressed in Chapter IV, it is interesting to pre-process individual image data, i.e. when forming the data input to the network, randomly chosen patches from single image are acquired. The results shown in Figure A.1, Figure A.2 and Figure A.3 depict averaged algorithm performance over the set of multiple images from each of three categories. For each individual image 2000 16×16 patches were randomly sampled. The total number of images processed was 24 from category 2 and 15 from category 1 as well as from category 3.

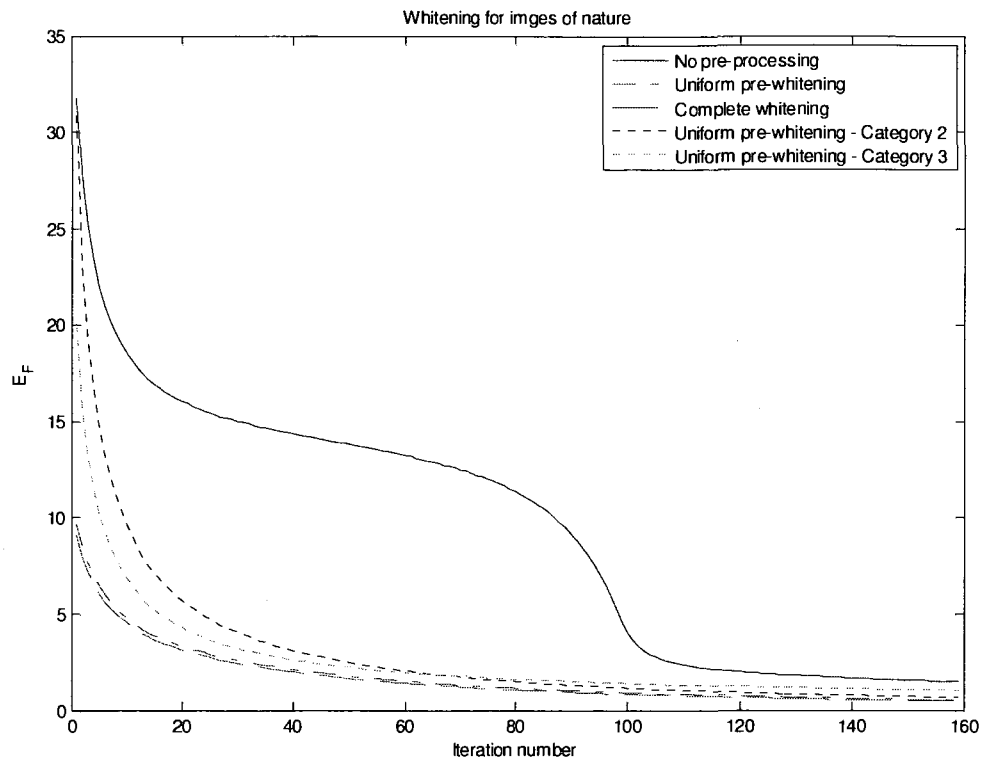


Figure A.1 Averaged performance of learned pre-whitening filter when processing 15 individual images from Category 1

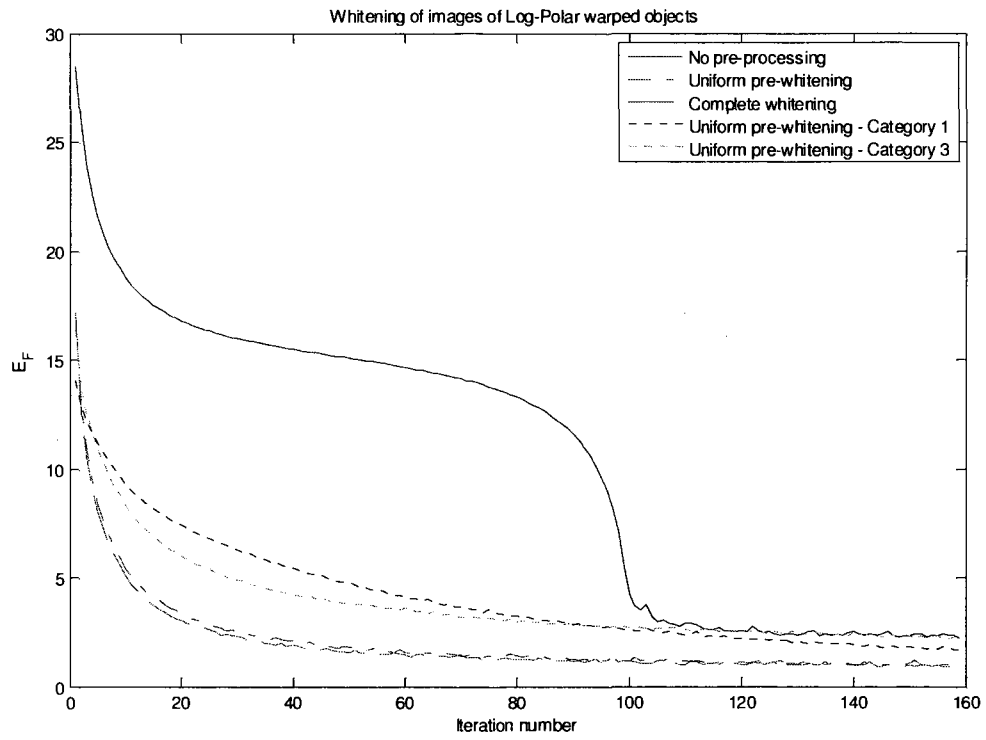


Figure A.2 Averaged performance of learned pre-whitening filter when processing 24 individual images from Category 2

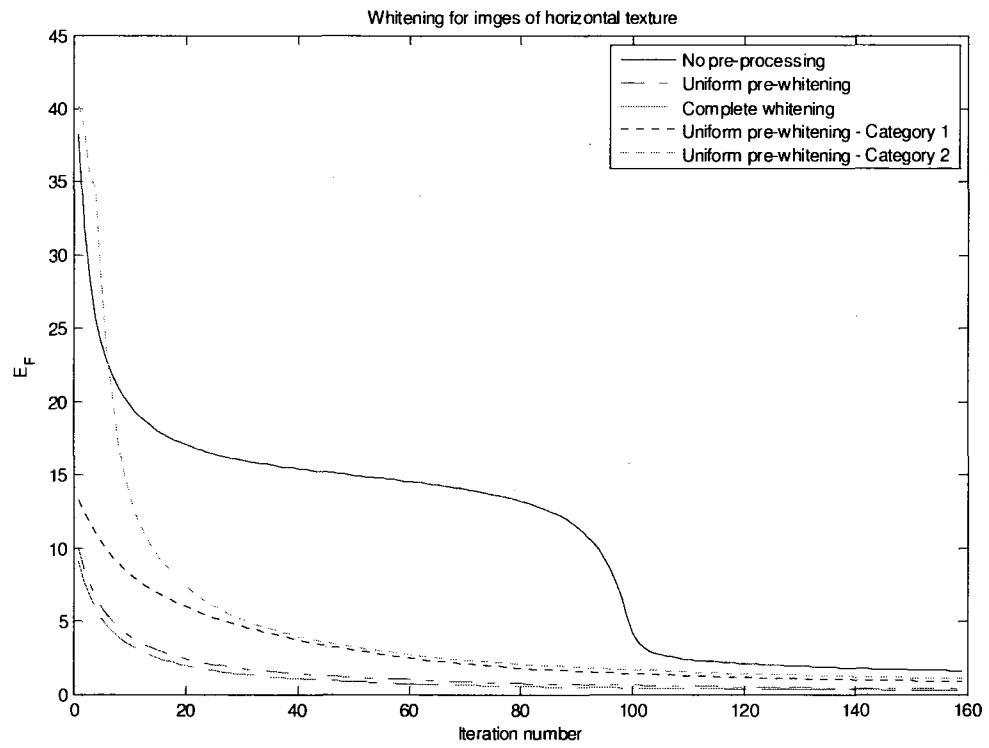


Figure A.3 Averaged performance of learned pre-whitening filter when processing 15 individual images from Category 3

APPENDIX B

SPARSENESS OF THE ICA BASED FEATURE ENCODERS

The image encoding process described in detail in Chapter V and depicted in Figure 5.2 yields highly sparse image code. Such code can be utilized for efficient information storage since it contains a very small number of non-zero coefficients. In order to demonstrate the effectiveness of deployed feature encoding techniques this section provides additional experimental evidence addressing the statistical properties of unprocessed and processed log-polar warped images of objects sitting on the dark background. In particular, 32 images were processed according to addressed feature encoding algorithm with the exception of locating pseudo-salient points in the image. Comparisons of the kurtosis of the resulting signal and the kurtosis of the signal prior to pre-whitening and ICA processing itself were made in order to estimate resulting code characteristics. One example of signal properties prior and post ICA feature encoding is given in Figure B.1. In this particular case the kurtosis of the original warped signal was 2.67 while the post-processed signal had kurtosis of 21.19. The average difference between kurtosis value (reflecting the sparseness of the signal code) between post and pre-processed image based on the entire 32 image/data set was 26.06. This fact confirms that the proposed pre-whitening algorithm along with the full ICA feature encoding technique produces highly sparse image code.

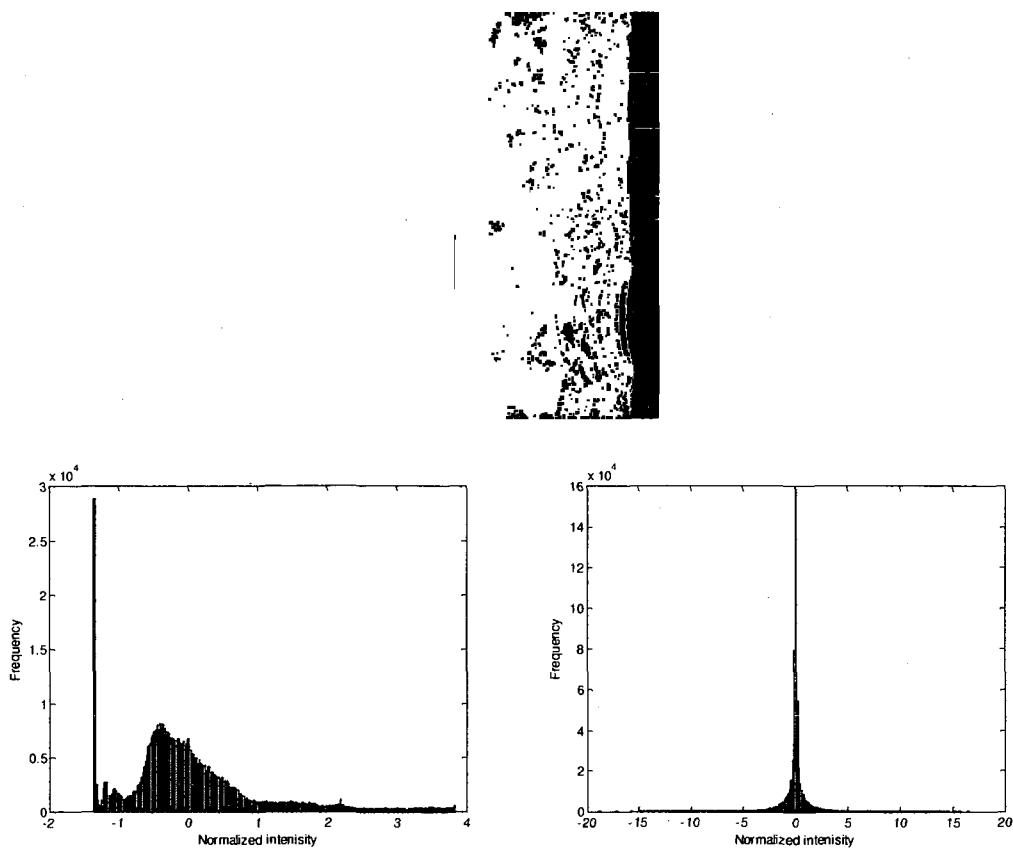


Figure B.1 Warped image (top); histogram approximations of the pre-processed signal (bottom left) and of the post-processed signal by ICA feature encoding (bottom right)