university of groningen

University Medical Center Groningen

University of Groningen

COSFIRE (Combination of Shifted Filter Responses)

Azzopardi, George

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2013

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Azzopardi, G. (2013). *COSFIRE (Combination of Shifted Filter Responses): A trainable filter approach to visual pattern recognition*. s.n.

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

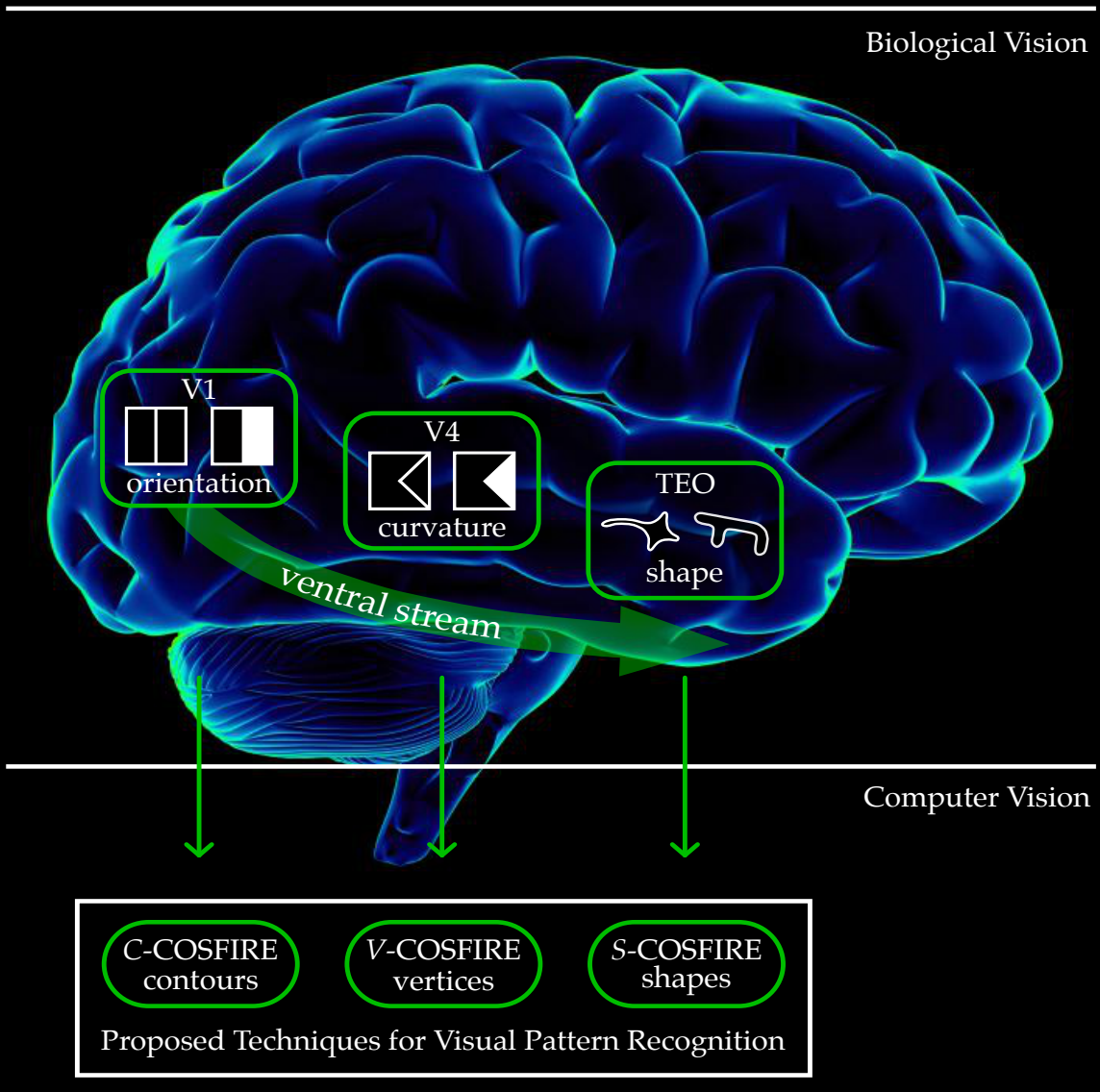
Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 03-06-2022

COSFIRE

Combination Of Shifted Filter Responses

A Trainable Filter Approach to Visual Pattern Recognition



George Azzopardi

COSFIRE

Combination Of Shifted Filter Responses

A Trainable Filter Approach to Visual Pattern Recognition

George Azzopardi

RIJKSUNIVERSITEIT GRONINGEN

COSFIRE
Combination Of Shifted Filter Responses
A Trainable Filter Approach to Visual Pattern Recognition

Proefschrift

ter verkrijging van het doctoraat in de
Wiskunde en Natuurwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
vrijdag 26 april 2013
om 14.30 uur

door

George Azzopardi

geboren op 4 juli 1981
te Qormi, Malta

Promotor:	Prof. dr. N. Petkov
Beoordelingscommissie:	Prof. dr. B. ter Haar Romeny Prof. dr. M. Vento Prof. dr. P. Campisi

ISBN: 978-90-367-6128-4

To my loving wife Charmaine Borg,
and my dear parents

Contents

List of Figures **iv**

List of Tables **v**

Acknowledgements **vii**

1 Introduction **1**
1.1 Scope . . . 2
1.2 Thesis Organization . . . 5

2 A CORF Model for Contour Detection **7**
2.1 Introduction . . . 8
2.2 Computational Model . . . 9
2.2.1 Overview . . . 9
2.2.2 Model LGN Cells with Center-Surround RFs . . . 10
2.2.3 Sub-Unit and its Parameters . . . 11
2.2.4 Sub-Unit Responses . . . 13
2.2.5 Combining Sub-Unit Responses . . . 13
2.2.6 Treating Different Orientations . . . 14
2.3 Results: CORF vs GF Model Comparison . . . 16
2.3.1 Is CORF a Simple Cell Model? . . . 16
2.3.2 Performance Evaluation for Contour Detection . . . 19
2.4 Discussion . . . 23
2.5 Summary . . . 26
2.6 Conclusions . . . 27

3 COSFIRE Filters for Keypoint Detection **31**
3.1 Introduction . . . 32
3.2 Method . . . 35
3.2.1 Overview . . . 35

3.2.2 Detection of Orientations by 2D Gabor Filters 36
3.2.3 Configuration of a COSFIRE Filter 36
3.2.4 Blurring and Shifting Gabor Filter Responses 39
3.2.5 Response of a COSFIRE Filter 40
3.2.6 Achieving Invariance 41
3.2.7 Detection of More Complex Patterns 44
3.3 Applications 45
3.3.1 Recognition of Handwritten Digits 45
3.3.2 Detection and Recognition of Traffic Signs 48
3.4 Discussion 50
3.5 Conclusions 55

4 Detection of Vascular Features in Retinal Images using COSFIRE filters 57
4.1 Introduction 58
4.2 Method 60
4.2.1 Overview 60
4.2.2 Detection of Dominant Orientations by 2D Gabor Filters . . . 61
4.2.3 Configuration of a COSFIRE Filter 63
4.2.4 Blurring and Shifting Gabor Filter Responses 65
4.2.5 COSFIRE Filter Response 67
4.2.6 Tolerance to Rotation, Scale and Reflection 67
4.3 Experimental Results 69
4.3.1 Data Sets and Ground Truth 69
4.3.2 Configuration of COSFIRE Filters 69
4.3.3 Performance Evaluation 74
4.3.4 Effects of the Number of COSFIRE Filters 74
4.3.5 Effects of the Training Data 75
4.3.6 Evaluating the Generalization Ability of COSFIRE Filters . . . 76
4.4 Discussion 78
4.5 Conclusions 80

5 Trainable *S*-COSFIRE Filters for Shape Recognition and Localization 81
5.1 Introduction 82
5.2 Method 87
5.2.1 Overview 87
5.2.2 Detection of Vertex Features by *V*-COSFIRE Filters 88
5.2.3 Configuration of a *S*-COSFIRE Filter 89
5.2.4 Blurring and Shifting *V*-COSFIRE Responses 91
5.2.5 Response of a *S*-COSFIRE Filter 92

5.2.6 Tolerance to Geometric Transformations 94
5.3 Applications 97
5.3.1 Spotting Letters and Keywords in Handwritten Manuscripts . 97
5.3.2 Vision for Home Tidying Pickup Robot 102
5.4 Discussion 106
5.5 Conclusions 108

6 Summary and Outlook **109**
6.1 Summary 109
6.2 Outlook 111

Samenvatting **115**

Bibliography **119**

Research Activities **135**

Curriculum Vitae **139**

List of Figures

1.1 Relationship of the proposed techniques to biological vision 4

2.1 Sketch of the proposed CORF computational model of a simple cell . 9
2.2 Model LGN responses to synthetic edge 11
2.3 Configuration of a CORF model cell 12
2.4 Responses of a CORF model to a synthetic image 13
2.5 Orientation selectivity of the CORF model 14
2.6 Achieving rotation invariance 15
2.7 Response to a shifting sinusoidal grating 16
2.8 Receptive field of an antisymmetric CORF model 17
2.9 CORF model exhibits contrast invariant orientation tuning 18
2.10 AND-type model 18
2.11 CORF model exhibits cross orientation suppression 19
2.12 Contour detection of natural objects from the RuG dataset 20
2.13 Contour detection of natural objects from the Berkeley dataset 21
2.14 CORF outperforms Gabor in contour detection task 24
2.15 Robustness to noise 25
2.16 Achieving response saturation 26
2.17 Linear models do not exhibit contrast invariant orientation tuning . . 29
2.18 Linear models do not exhibit cross orientation suppression 30

3.1 Examples of corners and junction patterns 32
3.2 Shortcomings of models that are based on similarity of descriptors . 33
3.3 Overview of the analysis of a specified prototype pattern 35
3.4 Configuration of a COSFIRE filter 37
3.5 Application demonstration of a COSFIRE filter 38
3.6 Selectivity of a COSFIRE filter 40
3.7 Orientation selectivity of a COSFIRE filter 43
3.8 Configuration by a more complex prototype pattern 45
3.9 Examples of handwritten digits from the MNIST data set 46
3.10 Examples of handwritten-digit-selective COSFIRE filters 47
3.11 Experimental results achieved on the MNIST data set 48
3.12 COSFIRE filters selective for reference traffic signs 49

3.13 Detecting and recognizing two different traffic signs in a complex scene 51

4.1 Selecting a prototypical bifurcation 60
4.2 Example of a 2D Gabor function 62
4.3 Configuration of a bifurcation-selective COSFIRE filter 64
4.4 Detecting vascular bifurcations with a COSFIRE filter 66
4.5 Example of a retinal fundus image from the DRIVE data set 70
4.6 A set of ten prototype bifurcations 71
4.7 Results of a rotation-, scale-, reflection-tolerant COSFIRE filter 72
4.8 Enlargement of the detected features 73
4.9 Performance results 75
4.10 A COSFIRE filter is robust to detect incomplete bifurcations 77

5.1 Human perception of shape 82
5.2 Face-selective neurons in inferotemporal cortex 83
5.3 Some neurons in inferotemporal cortex are shape-selective 84
5.4 Object recognition by parts 84
5.5 Schematic view of the proposed composite filter 86
5.6 Example of a prototype shape of interest 88
5.7 A data set of 60 synthetic vertices 89
5.8 Responses of vertex-selective V-COSFIRE filters 90
5.9 Configuration of a S-COSFIRE filter 91
5.10 Example of the application of a S-COSFIRE filter 93
5.11 Experiment to evaluate the selectivity of a S-COSFIRE filter 95
5.12 Spotting instances of the letter 'e' in handwritten manuscripts 98
5.13 More spotting instances of the letter 'e' in handwritten manuscripts . 99
5.14 Spotting the combined letters 'to' in handwritten manuscripts 100
5.15 Spotting the keyword 'Germany' in handwritten manuscripts 101
5.16 Configuration of a shoe-selective S-COSFIRE filter 103
5.17 Detection of shoes in indoor scenes 105

6.1 Inhibitory contours may increase discrimination ability 112
6.2 Directions for future research 113

List of Tables

4.1 Performance results for separate sets of COSFIRE filters 76

Acknowledgements

It goes without saying that this work would not have been possible without the support of many people. The journey leading to this thesis was a roller coaster of emotions. While there have been moments of difficulty, I am overwhelmed by the pleasant and enriching experiences I gained during my doctoral studies.

I start by thanking my promotor, or as the Germans say "Doktorvater" Nicolai Petkov, for his invaluable guidance throughout my studies. He has always believed in me and has been enthusiastic about this project from the beginning. To me he has been more than a research advisor, his advice on topics ranging from philosophy, travelling, sports and appreciating wine have benefited and enriched me in many ways. Most of the ideas that we discussed about my studies originated either during pleasant bicycle trips around the charming villages of the province of Groningen or during dinners that we shared when we were abroad for conferences. His broad technical knowledge and his meticulous reviewing were key to improve the quality of my work. I am grateful for teaching me the tools to simplify complex concepts and to write clearly. I feel privileged for having been under his mentorship and for extending our collaborations further.

I would like to express my gratitude to the members of the reading committee, the professors Bart ter Haar Romeny, Mario Vento and Patrizio Campisi, for their thorough review of this thesis.

I also feel fortunate and honoured for having spent four years in an academic environment surrounded by other intelligent people. The professors Marco Aiello, Michael Biehl, Michael Wilkinson, Paris Avgeriou and Alexander Lazovik are examples of true scientists that are passionate about their work and are always ready for a discussion. In particular, I thank Michael Biehl for sharing an interesting experience in my home country, Malta. Together with Nicolai Petkov we gave a seminar in 2010 at St. Martin's Institute of IT and also held a meeting with the honorouble health minister of Malta about potential scientific collaborations. A year later, a similar seminar was organized at the same place with another renowned speaker,

professor Thomas Lippert who is the head of Jülich Supercomputing Centre in Germany. Those two seminars were motivated by Charles Theuma, who is the principal of the institute and who has been influential in my academic career path since my undergraduate studies.

My Greek colleague Ioannis Giotis with whom I shared the office has also helped me many times during my studies. Although our studies never really crossed paths, it was always of benefit discussing different technical problems. Besides scientific work, it used to be a pleasure discussing things about life as well as arguing about football especially on Monday mornings. From him, I also finally learned that the Greek letter π is not pronounced as "pie".

I am also thankful for having shared an enjoyable atmosphere at the office with my other colleagues from different computer science groups: Ugo Moschini, Aree Witeolar, Ernest Mwebaze, Heerko Groefsema, Kerstin Bunte, Petra Schneider, Eirini Kaldeli, Andrea Pagani, Elie Khoury, Mahir Can Doganay, Fred Kiwanuka, Pavel Bulanov, Giuseppe Papari, Ehsan Ullah Warriach, Viktoriya Degeler, Ilche Georgievski, Ando Emerencia, Ahmad Waqas, Andreas Neocleous, Tuan Anh Nguyen, Faris Nizamic and others. With such an international mix of people it was always fun discussing cultural diversity and impressing each other with traditional food of different countries during "International Nights". A special thanks goes to Heerko and Ando for translating the summary of this thesis into the Dutch samenvatting.

Also, a sincere thanks is due to other student collaborators Zhe Sun, Nicola' Strisciuglio, Vincenzo Capozzoli, Chenyu Shi, Jiapan Guo and Laura Fernandez Robles for being fun people to work with on interesting projects.

Moreover, I would like to thank the administrative staff who made my life at the university easier. I thank the cheerful team of Esmee Elshof, Desiree Hansen, Ineke Schelhaas and Helga Steenhuis for assisting me with paperwork and other practical matters. I also thank Janieta de Jong and Yvonne van der Weerd for making the beginning and the final procedures of my studies as smooth as possible.

Long academic days were partially balanced by enjoying playing football with a group of diverse people, including Janniko, Richard, Anoup, Jimmy, Martijn, Robin, Alicio, Danny, Kai, Anil, Marc, Henk, Joel, Winot and others. This activity was, however, interrupted due to an unfortunate injury which left the anterior cruciate ligament of my right knee torn. During that period the presence of many friends was essential. I thank Lars-Uwe Lahoda for making the necessary arrangements for a surgery in a private hospital in Gelsenkirchen, Germany. I am very grateful to Daan Franken, Joachim Wöhrle and Annika Tovote for travelling long distances to visit me during my hospital stay. I can never forget Steffie van der Steen, Elisa Kupers and Annika Tovote for bringing me delicious food when I had my leg casted.

They made the fairy tale of red riding hood seem real. Also, a big thank you is due to Marieke Boelhouwer for taking care of our dog. The visits by the lovely couples Nico Jan and Madelon van Hemel-Ruiter, Ishay and Meirav Halmut-Dagan, and David and Karolina Sikorska as well as by Claudi Bockting also made my knee feel better. Thank you all so much! Thankfully, my knee is back in shape and I started playing football again.

Among other friends, Diederick, Mads, Ruud, Jean Pia, Patrick, Ben, Kor, Jolinda, Leo and Janneke have formed part of our extended family in Groningen. Leo and Janneke allowed us to meet our adorable dog, Okker, who always gives us warm welcomes when we return home from work.

Gratitude is also due to my research manager Gert Kruithof who offered me the position of a research scientist at TNO (Netherlands organization for applied scientific research) and for agreeing to a formal collaboration with University of Groningen where I work one day per week with Nicolai Petkov at the Intelligent Systems group.

I want to give special thanks to my paranymphs, my sister Maria and my dear friend Joachim, for being the brave front liners standing by my side on the *D*-Day.

Naturally, I am indebted to my parents, John and Josephine, for raising me up to become who I am today. While it has been difficult for them to understand what I am doing in Groningen, they knew that I am chasing a dream and their constant moral support has been very important. The same goes to my three sisters and their husbands, Rita and Stephan Casha, Josianne and Alfred Chetcuti, and Maria and Michael Zaguirre Gutiza as well as their young children Jerome, Jordan, Althea and Leanne who only know me studying abroad. My parents-in-law, Joseph and Mary Borg, who always treated me like their son, have always encouraged me to persist no matter what. I also thank my brother-in-law Josef Borg, his wife Ruth and their daughters Leanne and Cassidy.

Most importantly, I would like to thank my wife Charmaine who was also conducting PhD studies at the same time at University of Groningen, but in the faculty of Psychology. Despite being busy with her studies, Charmaine has always been of immense support. Her love, smile and positive attitude are the secret ingredients which kept me going, especially in difficult times.

George Azzopardi
Groningen
February 27, 2013

Chapter 1

Introduction

At the age of three my niece already knew what dogs are and could recognize them from their visual appearance. She could also recognize breeds that she had never seen before, under various illumination and environmental conditions, and she could do so in a fraction of a second. This visual task, which seems to be effortless for a young child, is still very difficult for computer vision.

The fact that nearly one-third of the human brain is devoted to visual processing – more than to any of the other four senses – speaks volumes about the complexity of object recognition. The remarkable human visual system has fascinated many researchers and has been extensively studied from different disciplines such as visual perception and neuroscience. One major goal is to understand how a finite number of neurons can process visual information for the detection and recognition of a large number of possible objects types. The ongoing findings of such research studies have inspired the computer simulation of the enviable visual processing of the brain. This has been the focus for many research groups in the computer vision community, the findings of which also contribute to achieve a better understanding of how the brain processes visual information.

The pioneering studies of Hubel and Wiesel (1962, 1974) on the primary visual cortex of cats revealed that most neurons in this area of the brain respond to edges or bars of specific orientation. Their Nobel prize-winning discovery gave new insights about visual processing and motivated a large body of neurophysiological studies to unveil the selectivity of neurons in other areas of visual cortex. To date there seems to be an agreement that processing of visual information in the brain is performed through what is called the visual pathway. It has two parts, called the dorsal and the ventral stream. The dorsal stream is concerned with processing the spatial location of a given object from the viewpoint of an observer. In this thesis, I focus on the processing of the ventral stream, which takes an input signal from the retina through the lateral geniculate nucleus (LGN) and transforms it to achieve object detection and recognition. This stream comprises cortical areas V1, V2, V4, TEO and TE (Ungerleider and Mishkin, 1982; Goodale and Milner, 1992).

1.1 Scope

This thesis is inspired by the visual information processing performed by some neurons in the ventral stream of the brain. It starts by revisiting a speculation of Hubel and Wiesel (1962). They hypthesized that the orientation selectivity of a simple cell is achieved by combining the responses of a group of center-surround LGN cells, the receptive fields of which are aligned in a colinear manner. This hypothesis had instigated the first two research questions of this thesis:

1. To what extent can a filter, which receives input from model LGN cells, exhibit properties that are typical of simple cells?

2. How would it perform in contour detection, which is assumed to be the primary biological role of simple cells?

These questions led to the construction of a novel computational model of a simple cell. It is implemented in a contour detection operator that takes as input the responses of appropriately aligned center-surround filters.

A computational model that has the ability to simulate the receptive fields of simple cells in primary visual cortex is the Gabor function model (Daugman, 1985). It has gained particular popularity due to its mathematical elegance and has been extensively used in many image processing and computer vision applications. The following is, therefore, the next question that is posed in this context:

3. How would the proposed model compare to the Gabor function model?

The above questions are answered in Chapter 2 of this thesis. Area V4 in visual cortex receives input from areas V1/V2 and is known to comprise neurons that are selective for various aspects of visual information, such as color (Zeki, 1973), texture (Hanazawa and Komatsu, 2001) and shape (Pasupathy and Connor, 1999). Here, focus is given to the latter type of neuron, which respond to curved contour features rather than to simple edges or bars (Pasupathy and Connor, 1999). Such a neuron integrates information about the curvature properties of a contour. The properties of these neurons instigated the following research questions:

4. To what extent can the curvature properties of contours and their position be modelled by a filter that would achieve selectivity similar to some V4 neurons?

5. In what ways would such a filter be useful in practical computer vision applications?

These questions are addressed in Chapter 3 by proposing a new filter that takes as input the responses of edge- and bar-detectors at certain positions. Such a filter is configured to give responses at particular points, called keypoints, surrounded by specific geometrical arrangements of contour parts. In this regard, the proposed filter is referred to as a keypoint detector.

The complexity of shape tuning increases in ventral stream areas subsequent to V4. For instance, some neurons in area TEO integrate information about the properties of multiple concave and convex curvatures and their mutual spatial position. It is thought that such TEO neurons have as afferents collections of V4 neurons that respond to independent curvatures (Brincat and Connor, 2004). The combined activity of the afferent V4 neurons triggers the concerned TEO neuron to fire. The involved connectivity and the responsiveness to complex shapes achieved in this area of the ventral stream motivated the last research questions of this thesis:

6. How can preference to shape of increasing complexity be modelled by a filter?

7. How effective would it be for the detection and recognition of deformable objects embedded in complex scenes?

This topic is dealt with in Chapter 5. These questions are addressed by configuring a more elaborative composite filter. By using the responses of the proposed keypoint detectors mentioned above instead of simple contour detectors, a more sophisticated filter can be constructed to give responses for more complex patterns and for deformable objects.

The three types of filters that are introduced in this thesis share the same approach for the computation of their responses. They combine input from simpler filters by a multivariate function. Taking the responses of simpler filters at different positions can be implemented by appropriately shifting the responses in such a way that they meet in some point, which is considered the center of support of the concerned filter. It is called Combination of Shifted Filter Responses or COSFIRE for brevity.

In order to distinguish the proposed filters, the contour-, vertex-and shape-selective filters are referred to as *C*-COSFIRE, *V*-COSFIRE and *S*-COSFIRE, respectively. An alternative name to COSFIRE is CORF. It stands for Combination of Receptive Fields and it is a term that is more appropriate to neuroscience literature. Fig. 1.1 illustrates the relationship between the visual processing in the ventral stream of the visual system of the brain and the types of computer vision COSFIRE filters that are proposed in this thesis.

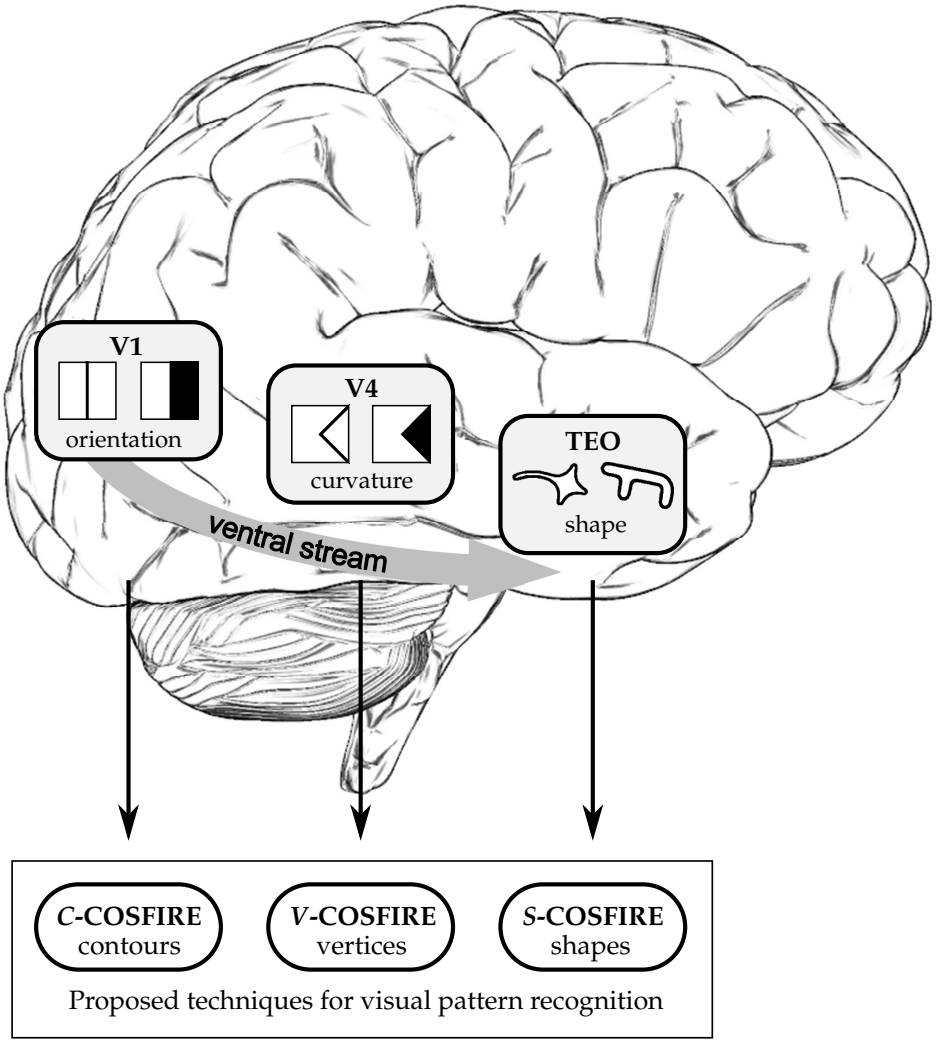


Figure 1.1: Object detection and recognition is thought to be the biological role of the ventral stream of the visual system of the brain. Area V1 contains neurons that are orientation-selective, neurons in area V4 are selective for curvatures, and area TEO contains neurons that are selective for shapes. The arrows indicate the relationship to the computer vision COSFIRE filters that are proposed in this thesis: *C*, *V*, and *S* stand for contour-, vertex- and shape-selective, respectively. They are inspired by and share some properties with neurons of the corresponding areas of the brain.

1.2 Thesis Organization

The remaining chapters of this thesis are organized as follows.

Chapter 2 describes the proposed computational model of a simple cell in area V1 of visual cortex. Extensive computational experiments are reported to demonstrate that the proposed model that is implemented in a contour-selective COSFIRE or *C*-COSFIRE filter possesses properties that are typical of simple cells. Moreover, quantitative analysis is provided that compares the proposed model and the Gabor function model in contour detection in images of natural scenes that are taken from two public data sets: RuG (Grigorescu et al., 2003a) and Berkeley (Martin et al., 2001).

Chapter 3 presents a trainable vertex-selective COSFIRE or *V*-COSFIRE filter and demonstrates how it can be applied for keypoint detection. A *V*-COSFIRE filter gives a response to a preferred set of contour parts and their preferred mutual spatial positions that are learned from a prototype feature. It enables a "veto-power" as a response is only achieved when all parts of the concerned prototype feature are present. This chapter also highlights the aspects where the proposed *V*-COSFIRE filter differs, for instance, from the SIFT detector (Lowe, 2004) and other state-of-the-art techniques. Qualitative experiments are reported to demonstrate the similarity of a vertex-selective keypoint filter to a neuron in area V4 of visual cortex. As to computer vision, the effectiveness of *V*-COSFIRE filters is investigated in two applications: the recognition of the segmented MNIST handwritten digits (LeCun et al., 1998) and the detection and recognition of traffic signs in complex scenes.

Chapter 4 addresses an important application in the area of medical imaging analysis. Here, *V*-COSFIRE filters of the type introduced in Chapter 3 are applied in rotation-, scale- and reflection-invariant mode for the detection of vascular bifurcations in segmented retinal fundus images. The proposed method is evaluated on two publicly available data sets: STARE (Hoover et al., 2000) and DRIVE (Staal et al., 2004). Results are also compared to other related methods.

Chapter 5 provides a description of the proposed shape-selective COSFIRE or *S*-COSFIRE filters for the recognition and localization of deformable objects embedded in complex scenes. This is achieved by combining the responses of certain keypoint filters that are tuned for different parts of an object of interest. A descriptive comparison to other object detection techniques that are inspired by the brain is also provided. In order to test their effectiveness, the proposed shape filters are applied in two practical applications: letter and keyword spotting in handwritten manuscripts and object spotting in complex scenes for the computer vision system of a domestic robot.

The above mentioned four chapters have been submitted to peer reviewed sci-

entific journals: the first three are published and the fourth one is currently under review.

Finally, Chapter 6 provides a summary of the thesis and gives an outlook of how the proposed techniques can be developed further and applied in different computer vision applications.

Published as:

G. Azzopardi, N. Petkov, *"A CORF computational model that relies on LGN input outperforms the Gabor function model,"* Biological Cybernetics, vol. 2(3), pp. 177–89, 2012.

Chapter 2

A CORF Model for Contour Detection

Abstract

Simple cells in primary visual cortex are believed to extract local contour information from a visual scene. The 2D Gabor function (GF) model has gained particular popularity as a computational model of a simple cell. However, it short-cuts the LGN, it cannot reproduce a number of properties of real simple cells, and its effectiveness in contour detection tasks has never been compared with the effectiveness of alternative models. We propose a computational model that uses as afferent inputs the responses of model LGN cells with center-surround receptive fields and we refer to it as a CORF (Combination of Receptive Fields) model. We use shifted gratings as test stimuli and simulated reverse correlation to explore the nature of the proposed model. We study its behavior regarding the effect of contrast on its response and orientation bandwidth as well as the effect of an orthogonal mask on the response to an optimally oriented stimulus. We also evaluate and compare the performances of the CORF and GF models regarding contour detection, using two public data sets of images of natural scenes with associated contour ground truths. The receptive field map of the proposed CORF model, determined with simulated reverse correlation, can be divided in elongated excitatory and inhibitory regions typical of simple cells. The modulated response to shifted gratings that this model shows is also characteristic of a simple cell. Furthermore, the CORF model exhibits cross orientation suppression, contrast invariant orientation tuning and response saturation. These properties are observed in real simple cells but are not possessed by the GF model. The proposed CORF model outperforms the GF model in contour detection with high statistical confidence (RuG data set: $p < 10^{-4}$, and Berkeley data set: $p < 10^{-4}$). The proposed CORF model is more realistic than the GF model and is more effective in contour detection, which is assumed to be the primary biological role of simple cells.

2.1 Introduction

The majority of neurons in primary visual cortex (area V1) exhibit orientation selectivity (Hubel and Wiesel, 1962). Typically, such a neuron would respond to an edge or a line of a given orientation in a given area of the visual field, called its receptive field (RF). Hubel and Wiesel identified two main classes of neuron that they called simple and complex cells. The RFs of simple cells can be divided in excitatory and inhibitory regions while no such division is possible in complex cells. These findings gave rise to an active area of research (Hubel and Wiesel, 1974; Macleod and Rosenfeld, 1974; Tyler, 1978; De Valois et al., 1978, 1979; Albrecht et al., 1980; Von Der Heydt, 1987) and the pioneers Hubel and Wiesel were later awarded a Nobel prize (Hubel, 1982).

The above electrophysiological findings had instigated the development of various computational models. Simple cells are typically modeled by linear spatial summation followed by half-wave rectification (Movshon et al., 1978; Maffei et al., 1979; Andrews and Pollen, 1979; Glezer et al., 1980; Kulikowski and Bishop, 1981; Daugman, 1985; Jones and Palmer, 1987). Marcelja (1980) has suggested that the elementary one-dimensional signals studied by Gabor (1946) can be used to model the structure of the RFs of simple cells. Later Daugman (1985) extended this idea to a family of two-dimensional (2D) Gabor functions that have been claimed to fit well the 2D RF profiles of cat simple cells (Jones and Palmer, 1987). The validity of Gabor functions for modeling of simple cells has also been questioned (Stork and Wilson, 1990) and alternative functions have been proposed, such as differences of offset Gaussians.

The Gabor function (GF) model does not take into account the anatomical structure of the visual system as it uses as inputs the intensity values of an image as these are projected on the retina and bypasses the lateral geniculate nucleus (LGN) within the thalamus (Chung and Ferster, 1998; Ferster et al., 1996; Reid and Alonso, 1995). Furthermore, it fails to reproduce some properties of simple cells, such as cross orientation suppression, independence of orientation tuning on contrast, and response saturation.

Hubel and Wiesel (1962) speculated that the RF profile of a simple cell and its orientation selectivity are the result of the specific alignment of the RFs of LGN cells that provide input. In this study we consider such a computational model of a simple cell and call it the CORF (Combination of Receptive Fields) model.

The biological role of orientation-selective cells is believed to be the extraction of local contour information, which is a fundamental step for further, more complex visual tasks, such as object recognition (Morrone and Owens, 1987; Morrone and Burr, 1988; Rosenthaler et al., 1992; Mehrotra et al., 1992; Heitger, 1995; Kovesi,

1999). The performance of various computational models of a simple cell in contour detection tasks has, however, not been quantified and they have not been compared in that respect. In the following, we evaluate and compare the performances of the proposed CORF model and the GF model, using two public data sets of images of natural scenes with associated contour ground truths. We also compare their ability to reproduce other properties of simple cells, such as cross orientation suppression, independence of orientation tuning on contrast, and response saturation.

The chapter is organized as follows. In Section 2.2 we present the CORF model. In Section 2.3, we demonstrate that it is a model of a simple cell and compare it with the GF model in different respects. Section 2.4 contains a discussion of some aspects of the model. We give a brief summary in Section 2.5 and draw conclusions in Section 2.6. We explain simulated reverse correlation in Appendix A. In Appendix B, we elaborate on the choice of an output function for the proposed model and in Appendix C we give details about the evaluation method of contour localization that we use in our experiments.

2.2 Computational Model

2.2.1 Overview

Fig. 2.1 illustrates the general set-up of the proposed CORF computational model. Each of the light and dark disks in Fig. 2.1 represents the RF of a sub-unit that receives input from a pool of center-on ('+') or center-off ('-') model LGN cells. Such model LGN cells detect contrast changes. A sub-unit computes the sum of the weighted responses of the model LGN cells it receives input from. These model LGN cells have the same polarity (on or off) and RF size and neighboring RFs. The RF of the sub-unit is the union of the RFs of the involved model LGN cells and it has

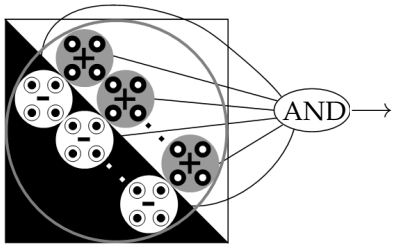


Figure 2.1: Sketch of the proposed CORF computational model of a simple cell. Orientation selectivity is achieved by combining the responses of two parallel sets of co-linear sub-units: one of center-on ('+') and the other of center-off ('-') type. Each sub-unit computes a sum of the weighted responses of a local group of model LGN cells.

the same polarity (on or off) as these cells. In this way, a sub-unit detects contrast changes, similar to a model LGN cell, but it does so in a wider area. A sub-unit can be thought of as a dendrite branch of a simple cell which receives input from a pool of adjacent LGN cells.

The orientation selectivity of a CORF model cell is achieved by combining the responses of given sub-units with appropriate polarities and alignment of their RFs, as illustrated by Fig. 2.1. The model parameters are determined in an automatic configuration process in which an example edge of a given orientation and polarity is presented. This input stimulus gives rise to a certain local configuration of model LGN cell activities in the RF of the concerned CORF model cell. This local configuration is used to determine the polarity of the involved sub-units and their mutual spatial arrangement. The response of the considered CORF model cell is computed as the weighted geometric mean of the sub-unit responses. In this way, a CORF model cell generates a response only when all its afferent inputs are stimulated. In the following sub-sections we explain the model configuration process in more detail.

2.2.2 Model LGN Cells with Center-Surround RFs

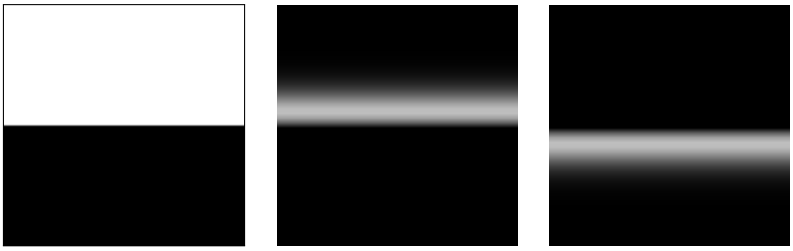
We use a difference of 2D Gaussian functions to model an LGN cell (Rodieck, 1965; Croner and Kaplan, 1995):

$$DoG_{\sigma}^{+}(x,y) \stackrel{\text{def}}{=} \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) - \frac{1}{2\pi(0.5\sigma)^2}\exp\left(-\frac{x^2+y^2}{2(0.5\sigma)^2}\right) \tag{2.1}$$

where σ is the standard deviation of the outer Gaussian function. We fix the standard deviation of the inner Gaussian function to 0.5σ, which is in accordance with electrophysiological findings of LGN cells in mammals (Irvin et al., 1993; Xu et al., 2002). A $DoG_{\sigma}^{+}(x,y)$ function corresponds to a center-on RF, such that the central region is excitatory and the surrounding is inhibitory. A center-off RF is denoted by $DoG_{\sigma}^{-}(x,y)$ and is defined as the negative of $DoG_{\sigma}^{+}(x,y)$:

$$DoG_{\sigma}^{-}(x,y) \stackrel{\text{def}}{=} -DoG_{\sigma}^{+}(x,y) \tag{2.2}$$

The response of a model LGN cell with a RF centered at image coordinates (x,y) is computed by linear spatial summation of the intensity distribution $I(x',y')$ in the input image, weighted with the function $DoG(x-x',y-y')$, followed by half-wave



(a) Input image (b) $c_\sigma^+(x,y)$ (c) $c_\sigma^-(x,y)$

Figure 2.2: (a) Synthetic input image of a sharp edge. Responses of (b) center-on and (c) center-off model LGN cells. The response of a model LGN cell is rendered as the brightness of a pixel on which the RF of that cell is centered.

rectification[1]:

$$c_\sigma^\delta(x,y) \stackrel{\text{def}}{=} |I \star DoG_\sigma^\delta|^+ \tag{2.3}$$

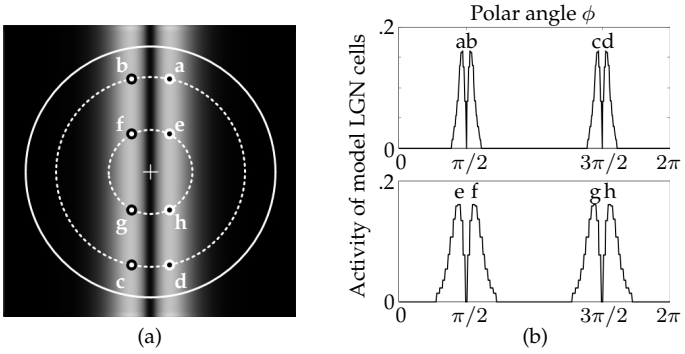
where the symbol δ represents the polarity (+ for center-on and - for center-off) of the DoG function used.

Fig. 2.2a shows a synthetic input image of a sharp step edge and in Fig. 2.2(b-c) we show the corresponding responses of center-on and center-off model LGN cells.

2.2.3 Sub-Unit and its Parameters

Fig. 2.3 illustrates the process of configuring a CORF model cell. The outer circle in Fig. 2.3a demarks the RF of that cell. Its center is positioned on an edge (in the input image) which gives rise to model LGN cell responses rendered in Fig. 2.3a. The eight small spots represent the RF centers of eight sub-units, four of center-on and four of center-off type. The inclusion of such sub-units in the computational model is decided on as follows. We take two (in general k) concentric circles centered on the RF center and consider the responses of model LGN cells along these circles, Fig. 2.3b. The positions along these circles at which these responses reach significant local maxima are the positions at which we include sub-units. For the considered example, there are four such positions for each of the two circles, which results in the inclusion of eight sub-units in the model at hand. The number of sub-units depends on the number of circles we consider and the specific input pattern presented at the time of configuration. In our model, each sub-unit included is represented in parametric form by a tuple $(\delta, \sigma, \rho, \phi)$ where the parameters represent the polarity δ of the sub-unit, the scale parameter σ of the involved model LGN cells in its pool,

[1]We use $\star$ and $|.|^+$ to denote convolution and half-wave rectification, respectively.



(a) (b)

Figure 2.3: Configuration of a CORF model cell. (a) The largest circle depicts the RF boundary of the model, the center of which is illustrated by the '+' marker and lies on a vertical edge. (b) The top and bottom plots show the responses of model LGN cells along the outer and inner interrupted circles in (a), respectively. The labeled small spots represent the RF centers of eight sub-units, four center-on (white spots with black boundary) and four center-off (black spots with white boundary) with coordinates that are determined by the corresponding labeled local maxima in the considered responses of model LGN cells.

the radius ρ and the polar angle ϕ of the RF center of the sub-unit relative to the RF center of the CORF model cell, respectively.

We denote by $S = \{(\delta_i, \sigma_i, \rho_i, \phi_i) \mid i = 1 \dots n\}$ the set of 4-tuples that represent the configured sub-units above. For the concerned model in Fig. 2.3, for $\sigma = 5$, and two configuration circles ($\rho \in \{18, 34\}$) for an image of size 100×100 pixels, the method described above results in eight sub-units, which are specified by the tuples in the following set:

$$S = \left\{ \begin{array}{l} (\delta_1 = -,\ \sigma_1 = 5,\ \rho_1 = 34,\ \phi_1 = 1.48), \\ (\delta_2 = +,\ \sigma_2 = 5,\ \rho_2 = 34,\ \phi_2 = 1.66), \\ (\delta_3 = +,\ \sigma_3 = 5,\ \rho_3 = 34,\ \phi_3 = 4.62), \\ (\delta_4 = -,\ \sigma_4 = 5,\ \rho_4 = 34,\ \phi_4 = 4.80), \\ (\delta_5 = -,\ \sigma_5 = 5,\ \rho_5 = 18,\ \phi_5 = 1.41), \\ (\delta_6 = +,\ \sigma_6 = 5,\ \rho_6 = 18,\ \phi_6 = 1.74), \\ (\delta_7 = +,\ \sigma_7 = 5,\ \rho_7 = 18,\ \phi_7 = 4.55), \\ (\delta_8 = -,\ \sigma_8 = 5,\ \rho_8 = 18,\ \phi_8 = 4.88) \end{array} \right\}$$

The first tuple in the set S, for instance, describes a sub-unit that collects its afferent inputs from the responses of center-off ($\delta_1 = -$) model LGN cells with a RF size characterized by a standard deviation of ($\sigma_1 =$) 5 pixels, around a position at a radius of ($\rho_1 =$) 34 pixels and an angle of ($\phi_1 =$) 1.48 radians with respect to the RF center of the CORF model cell; the RF center of this sub-unit is marked by 'a' in Fig. 2.3a.

(a) Input image (b) $r_S(x, y)$

Figure 2.4: (a) A synthetic image (512×512 pixels) and (b) the corresponding responses of a CORF model that is selective for vertical edges.

2.2.4 Sub-Unit Responses

We denote by $s_{\delta_i,\sigma_i,\rho_i,\phi_i}(x, y)$ the response of a sub-unit, which we compute by linear spatial summation of the half-wave rectified responses $c^{\delta_i}_{\sigma_i}(x, y)$ of model LGN cells with preferred polarity δ_i and scale σ_i around position (ρ_i, ϕ_i) with respect to the RF center of the CORF model cell, weighted with a 2D Gaussian function $G_{\sigma'}$:

$$s_{\delta_i,\sigma_i,\rho_i,\phi_i}(x, y) \stackrel{\text{def}}{=} \sum_{x'} \sum_{y'} \{c^{\delta_i}_{\sigma_i}(x + \Delta x_i - x', y + \Delta y_i - y') G_{\sigma'}(x', y')\} \qquad (2.4)$$
$$\Delta x_i = \rho_i \cos\phi_i, \Delta y_i = \rho_i \sin\phi_i, -3\sigma' \leq x', y' \leq 3\sigma'$$

The standard deviation σ' of the Gaussian function $G_{\sigma'}$ is a linear function[2] of the parameter ρ which is consistent with neurophysiological evidence for the relationship between the eccentricity and the average RF diameter of LGN cells (Xu et al., 2001). Eq. 2.4 presents a convolution of the weighting function $G_{\sigma'}$ with the function $c^{\delta_i}_{\sigma_i}(x, y)$ that is shifted by $(\Delta x_i, \Delta y_i)$ where this shift vector is determined by the sub-unit parameters (ρ_i, ϕ_i).

2.2.5 Combining Sub-Unit Responses

We define the response r_S of a CORF model cell as the weighted geometric mean of all sub-unit responses that belong to the specific selection determined by the set S:

[2]The standard deviation σ' of the 2D Gaussian function $G_{\sigma'}$ is computed as $\sigma' = (d_0 + \alpha\rho)/6$. We set $d_0 = 2$ and $\alpha = 0.9$.

$$r_S(x,y) \stackrel{\text{def}}{=} \left(\prod_{i=1}^{|S|} \left(s_{\delta_i,\sigma_i,\rho_i,\phi_i}(x,y) \right)^{\omega_i} \right)^{1/\sum_{i=1}^{n}\omega_i} \tag{2.5}$$
$$\omega_i = \exp^{-\frac{\rho_i^2}{2\sigma'^2}},\ \sigma' = \frac{1}{3} \max_{i\in\{1\ldots|S|\}} \{\rho_i\}$$

Computing the product of sub-unit responses means that the concerned CORF model cell is activated only when all its afferent sub-units are active. The input contribution of sub-units decreases with an increasing distance of their RF centers from the RF center of the CORF model cell.

As illustrated in Fig. 2.4 the configured CORF model cell responds to a vertical edge in the given synthetic image.

Fig. 2.5 illustrates the orientation selectivity of the concerned model. A maximum response is obtained for the orientation for which the model was configured. The response declines with the deviation of the orientation of the input stimulus from the optimal one and practically disappears when this deviation is greater than $\pi/4$ radians. Qualitatively, this is in line with the orientation tuning recorded by De Valois et al. (1982) where the majority of simple cells in macaque visual cortex have an orientation bandwidth of 0.7 radians at half amplitude.

2.2.6 Treating Different Orientations

The orientation preference of a CORF model cell as defined above depends on the orientation of the edge used for the model configuration. One can create models with different orientation preference by presenting different edges. Alternatively, one can manipulate the parameters in the set S, which corresponds to orientation

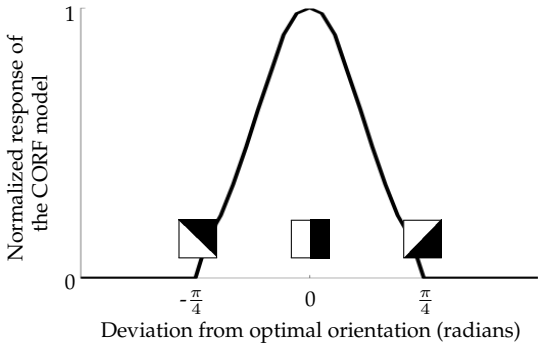


Figure 2.5: Orientation selectivity of the CORF model.

preference for 0° to obtain a new set $\Re_\psi(S)$ with orientation preference ψ. We define $\Re_\psi(S)$ as follows:

$$\Re_\psi(S) \stackrel{\text{def}}{=} \{(\delta_i, \sigma_i, \rho_i, \phi_i + \psi) \mid \forall\ (\delta_i, \sigma_i, \rho_i, \phi_i) \in S\} \tag{2.6}$$

For the detection of contours of any orientation, we merge the responses of CORF models with different orientation preference by taking the maximum value at a given location (x, y):

$$\hat{r}_S(x, y) \stackrel{\text{def}}{=} \max_{\psi \in \Psi} \{r_{\Re_\psi(S)}(x, y)\} \tag{2.7}$$

where Ψ is a set of n_θ equidistant orientations given as $\Psi = \{\frac{2\pi}{n_\theta} i \mid 0 \leq i < n_\theta\}$.

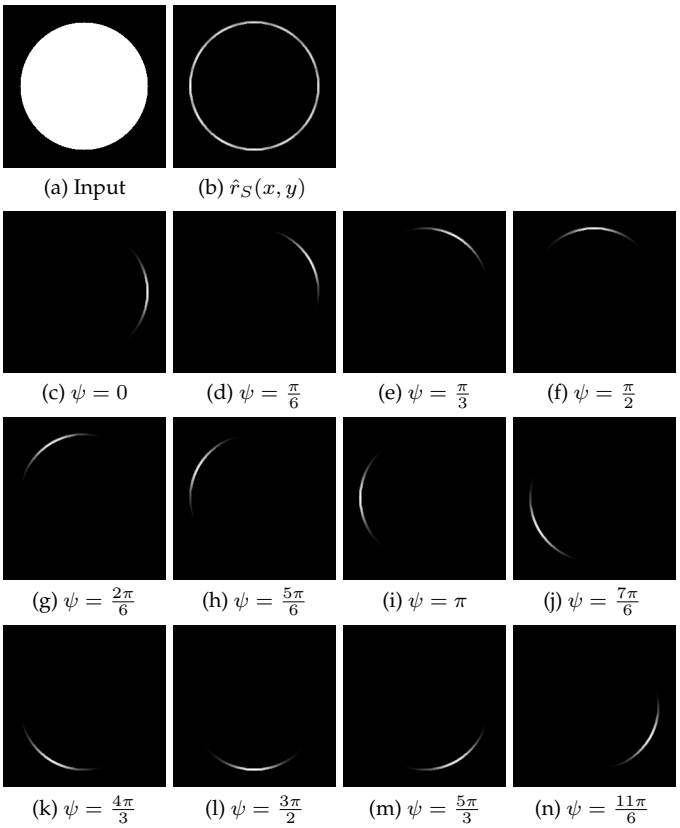


Figure 2.6: (a) Synthetic input image (512×512 pixels). (b) The maximum superposition of the (c-n) responses of CORF models that are selective for 12 equidistant orientations.

Fig. 2.6a illustrates a synthetic input image of a bright disk. Fig. 2.6b shows the maximum superposition $\hat{r}_S(x, y)$ of the responses, shown in Fig. 2.6(c-n), of a set of CORF models that are selective for $(n_\theta =)$ 12 equidistant orientations. As demonstrated in Fig. 2.6b, the choice of $n_\theta = 12$ ensures sufficient response for all orientations.

2.3 Results: CORF vs GF Model Comparison

2.3.1 Is CORF a Simple Cell Model?

A classical means of testing an orientation selective cell for its type, simple vs. complex, is to record its response to a grating of its preferred orientation and spatial frequency and observe if the response changes when the grating is shifted. A complex cell shows unmodulated response while the response of a simple cell alternates between high and low values with the grating shift. We computed the responses of the concerned CORF model cell to a grating with a changing spatial phase offset and observed that this response is qualitatively typical of the behaviour of a simple cell, Fig. 2.7.

An important feature of a simple cell that distinguishes it from a complex cell is that its RF can be divided into excitatory and inhibitory regions. Since the CORF model cell that we propose above computes the response as a product of weighted LGN responses, it is not clear that its RF can be divided in such regions. To explore this aspect of the model, we apply to it *simulated* reverse correlation as explained in detail in Appendix A. Fig. 2.8 shows a map determined in this way for a CORF model cell. The map exhibits two elongated regions: an excitatory one (rendered bright) in which the values are positive and an inhibitory one (shown dark) in which the values are negative, characteristic of the RF profiles of real simple cells.

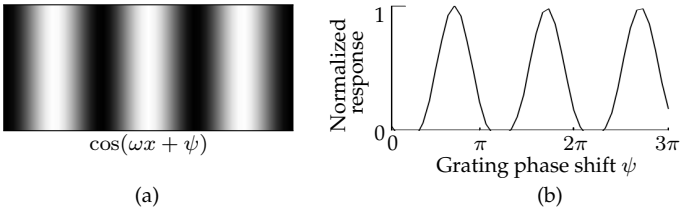


Figure 2.7: (a) Vertical sinusoidal grating and (b) normalized response (in a given point) of a CORF model cell to this grating as a function of the phase offset ψ. The response of the model alternates between high and low values with a changing phase offset ψ (that corresponds to shifting the grating) as the response of a real simple cell would do.

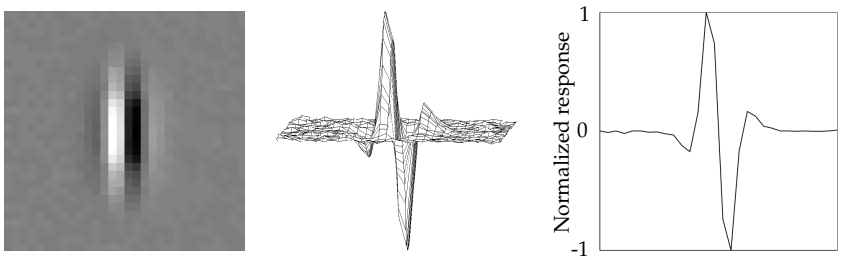


Figure 2.8: RF map of a CORF model cell determined with simulated reverse correlation, rendered as (a) intensity map and (b) 2D function. (c) One-dimensional profile of the function along a RF cross section that is perpendicular to the preferred orientation.

We should point out that the map shown in Fig. 2.8 is not an impulse response that can be used for weighted summation of the input stimulus followed by half-wave rectification or thresholding, as this is the case with the GF model. The GF model and any other semi-linear model based on weighted spatial summation will produce a non-zero response that grows with stimulus contrast for any edge orientation but the one that is strictly orthogonal to the preferred orientation. This is a drawback of such models because it is in contrast to neurophysiological evidence that the orientation bandwidth of simple cells is not affected by the contrast of the stimuli, Fig. 2.9a (*see* e.g. Finn et al. (2007) and references therein). The orientation bandwidth of the CORF model cell can be controlled by appropriate selection of the model parameters, such as the given ρ and σ values. Similar to real simple cells, the resulting orientation bandwidth of the CORF model cell is invariant to contrast, Fig. 2.9b. On the other hand, the orientation bandwidth of the GF model can only be constrained by using a threshold, which will lead to a broadening orientation tuning with increasing contrasts (the so-called iceberg effect), Fig. 2.9c.

The key to understanding the orientation tuning of the CORF model cell is its multiplicative character: the response is computed as a *product* of the afferent inputs and if any of these is zero, the response will be zero, independent of how strong the other inputs are. For instance, the CORF model cell, the structure of which is depicted in Fig. 2.10, gives a zero response to an edge stimulus of high contrast with an orientation that differs substantially from the preferred orientation due to the zero input that it receives from some of the afferent groups (sub-units) of model LGN cells.

Another property of real simple cells, that cannot be reproduced by semi-linear spatial summation models such as the GF model, is cross orientation suppression (Priebe and Ferster, 2006). This means that an oriented mask stimulus that is or-

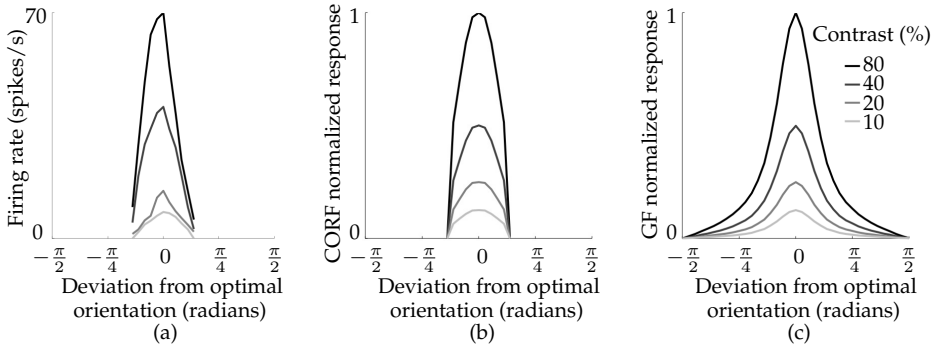


Figure 2.9: Orientation tuning curves at four different contrasts obtained for (a) a real simple cell (redrawn from Sclar and Freeman (1982)), (b) a CORF model cell with similar bandwidth and (c) a GF model. The GF model gives response at any stimulus orientation but the one that is strictly orthogonal to the preferred orientation. If a threshold is used, the orientation bandwidth of the GF model becomes dependent on stimulus contrast, which is not the case with real simple cells.

thogonal to a simultaneously presented test stimulus of the orientation preferred by the cell inhibits the response. The proposed CORF model cell also has this property as illustrated by Fig. 2.11. Linear summation models, such as the GF, do not possess this property, because an orthogonal mask signal contributes equally to the excitatory and inhibitory areas and its net contribution is zero, so that the response to the main stimulus is not affected.

These two important properties that are exhibited by the proposed CORF model are attributable to the multiplicative nature (weighted geometric mean) of the output function (Eq. 2.5). In Appendix B we demonstrate that if we use an addition output function instead of multiplication, the resulting model would lose these two properties.

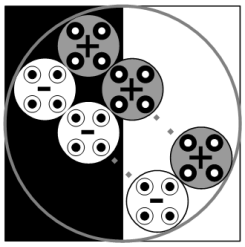


Figure 2.10: Model LGN cells with RFs that fall in homogeneous areas of the stimulus provide zero input and therefore the response of the CORF model cell is zero. This property causes the CORF model cell to have a band-limited orientation tuning curve.

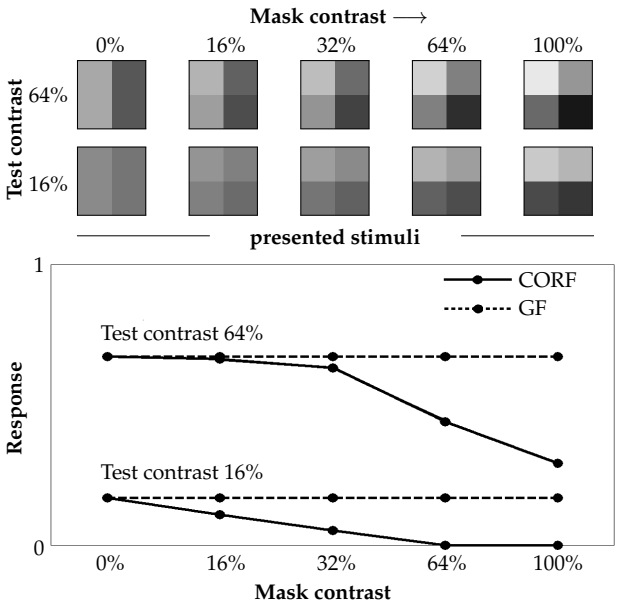


Figure 2.11: The CORF model cell exhibits cross orientation suppression while the GF model does not. The plots illustrate the responses of the CORF and the GF models to the images shown at the top. In these experiments, the RFs of the models are placed in the center of a given image. Every image is generated by superimposing a horizontal-edge stimulus (mask) on a vertical-edge stimulus (test) which is the preferred orientation stimulus of the concerned models. Similar to real simple cells, the response of the CORF model cell decreases with an increasing contrast of the mask stimulus. The linear GF model is not affected by any of the orthogonal mask stimuli.

2.3.2 Performance Evaluation for Contour Detection

In this section, we present a procedure for the evaluation of the performance of the CORF computational model in a contour detection task. We use the same procedure to evaluate the performance of the GF model and subsequently we compare the performances of the two models.

Data Sets and Ground Truth

Fig. 2.12 (first row) shows four images of natural objects and associated ground truth contour maps (second row) defined by a human observer. These images belong to a data set that has been developed at the University of Groningen (RuG)[3]. It comprises 40 such images along with their corresponding ground truths, and was first used

[3]The complete data set is available at `http://www.cs.rug.nl/~imaging`

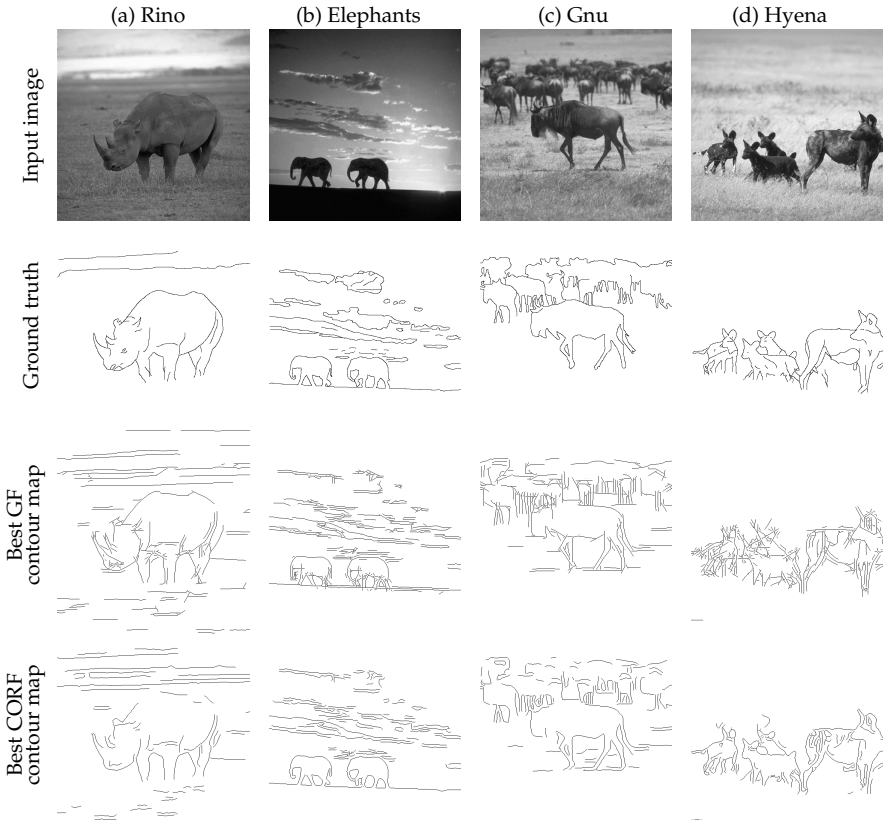


Figure 2.12: (First row) Images of natural objects from the RuG dataset. (Second row) The corresponding contour maps designed by a human observer. Optimal contour maps obtained by (third row) the GF and by (fourth row) the proposed CORF models, respectively.

by Grigorescu et al. (2003a). The ground truth images depict only the contours of objects, while omitting edges that are caused by texture.

Fig. 2.13 illustrates further examples of four input images taken from a data set of Berkeley (Martin et al., 2001). This data set comprises 300 images of natural objects, and for each image the ground truth is provided as a collection of multiple (5 to 10) contour maps that are drawn by different human observers, Fig. 2.13 (second row).

In the following we explain how we obtain contour maps from outputs of the concerned GF and CORF models. We then define performance measures that we use to evaluate a contour map that is obtained by a model with a corresponding ground truth contour map made by a human observer (Zhang, 1996; Shin et al., 1998; Grigorescu et al., 2003a).

Figure 2.13: (First row) Images of natural objects from the Berkeley data set. (Second row) Ground truth illustrated as a superimposition of the collection of multiple contour maps drawn by different observers. Darker contours correspond to the agreement of multiple hand drawn contours. Optimal contour maps obtained by (third row) the GF and by (fourth row) the proposed CORF models, respectively.

Binary Contour Map

To obtain a binary contour map for a given input image we apply to the output of the concerned model (CORF or GF) a procedure widely used in computer vision.

It consists of edge thinning by non-maxima suppression followed by binarization by hysteresis thresholding (Canny, 1986; Sonka et al., 1999). The former step essentially determines the ridges in the operator response. The latter step requires two parameter values, referred to as the low and high thresholds. We work with a value of the low threshold that is a fixed fraction (0.5) of the value of the high threshold. The resulting contour map depends on the value of the (high) threshold used: the lower that value, the larger the number of contour pixels in the map because weaker responses can pass the threshold.

The third and the fourth rows of Fig. 2.12 and Fig. 2.13 show the contour maps obtained in this way from the outputs of the GF and CORF models, respectively. These maps are obtained for certain values of the threshold parameter that are explained below.

Quantitative Performance Measure

We use performance indicators called recall R and precision P to measure the similarity between the contour map obtained by a given model and the ground truth. Recall R is defined as the fraction of true contour pixels (according to the ground truth) that are successfully detected by a given model. Precision P is defined as the fraction of true contour pixels from all the detected ones. They are formally defined as:

$$R \stackrel{\text{def}}{=} \frac{n_{TP}}{n_{TP} + n_{FN}}, \qquad P \stackrel{\text{def}}{=} \frac{n_{TP}}{n_{TP} + n_{FP}} \tag{2.8}$$

where n_{TP}, n_{FP}, and n_{FN} stand for the numbers of true positives, false positives, and false negatives, respectively. The values of the recall and precision depend on the threshold used for binarization: precision increases and recall decreases with an increasing value of that threshold.

In order to come to a single performance measure that would allow for results comparison, as suggested by Martin et al. (2004), we compute the harmonic mean of recall and precision, commonly referred to as the F-measure:

$$F \stackrel{\text{def}}{=} \frac{2PR}{P + R} \tag{2.9}$$

and consider as an optimal result the filter output for which this harmonic mean reaches its maximum for a given image.[4] The contour maps shown in Fig. 2.12 correspond to maximum values of the F-measure for the corresponding images.

[4]We refer to Appendix C for specification of the way in which we deal with inexact contour localization.

Experimental Results

In our experiments we apply both CORF and GF computational models to every input image. Below, we specify the parameter values that were used by the concerned models and then we present a statistical comparison between their performances.

The CORF and the GF models that we compare share two parameters: the number of orientations, n_θ, which we set to 12 (intervals of $\pi/6$) and a scale parameter σ. The GF model requires two other parameters, which we set as suggested by Petkov (1995): the wavelength $\lambda = \sigma/0.4$ and the spatial aspect ratio $\gamma = 0.5$. These parameter values ensure that the RF maps of the GF and CORF models are similar.

For every input image, we only consider the two maximum F-measure values, one for each of the two models, across all nine scales ($\sigma \in \{1, 1.5, \ldots, 5\}$) used. For every given value of σ, a CORF model is configured with different values[5] of the parameter ρ.

Fig. 2.14 shows a scatter plot that illustrates a comparison between the performances of the two models for the RuG data set. The names of the images are shown on the x-axis in a descending order of the performance achieved with the CORF model. For 30 out of 40 input images, the maximum F-measure value that is achieved by the proposed CORF model is greater than the maximum F-measure value that is achieved by the GF model. For the images where the GF model performs better, the difference in the corresponding F-measure values is minimal.

We apply a right-tailed paired-samples t-test to the set of pairs of F-measure values that are achieved by the proposed CORF model and the GF model and obtain as a result that the proposed CORF model outperforms the GF model for both the RuG ($t(39) = 4.39$, $p < 10^{-4}$) and the Berkeley ($t(299) = 3.88$, $p < 10^{-4}$) data sets.

Notable is the fact that, compared to the GF model, the best contour maps obtained by the proposed CORF model, such as the ones shown in the fourth rows of Fig. 2.12 and Fig. 2.13, contain less texture and the high curvature points are better preserved.

2.4 Discussion

The ground truth maps of the Berkeley data set were designed to evaluate performance for region segmentation rather than object contour detection. For instance, the ground truth maps shown in Fig. 2.13 omit the contours of the stripes in the shirt (Fig. 2.13a) and the patterns in the hat (Fig. 2.13c) of the the two persons.

[5]For $\sigma \in \{1, 1.5, 2\}$ we use three radii ($\rho \in \{3, 7, 14\}$), for $\sigma \in \{2.5, 3, 3.5\}$ we use four radii ($\rho \in \{3, 6, 13, 25\}$) and for $\sigma \in \{4, 4.5, 5\}$ we use 5 radii ($\rho \in \{3, 5, 9, 18, 34\}$).

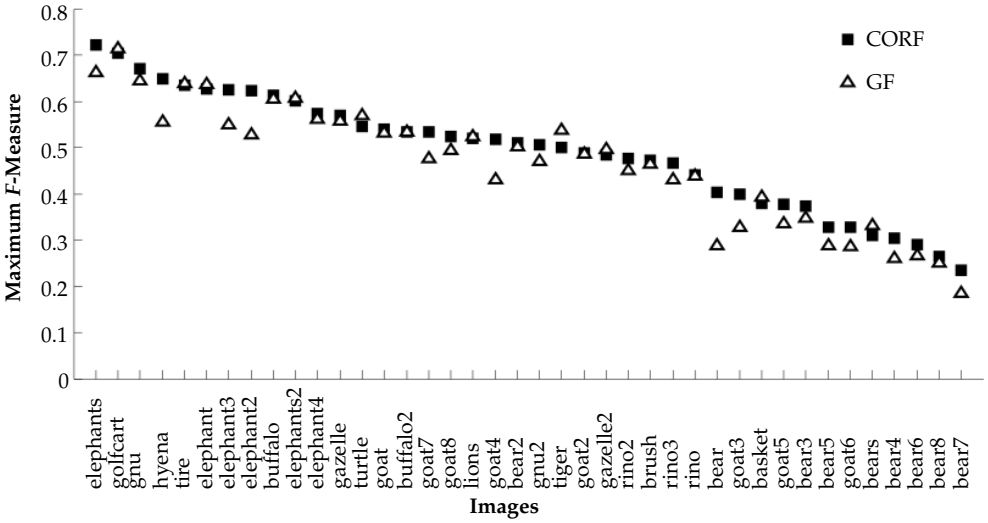


Figure 2.14: Maximum F-measure values obtained with the CORF and GF models for the images of the RuG data set.

Fig. 2.15 shows the optimal contour maps achieved by the proposed CORF model and the GF model for a synthetic stimulus without and with noise. The CORF model is more robust to noise and achieves better edge localization than the GF model. As pointed out by DuBuf (1993), curved edges that are detected by the GF model are orientation-quantized and appear as line segments in the resulting contour maps. In contrast, the proposed CORF model is able to preserve the smoothness and orientation of contours. While a high curvature contour is correctly detected by the CORF model, the GF model detects such a contour as multiple extended edges that cross each other, see, for instance, the difference in how the contours of the stripes in the shirt (Fig. 2.13a) and the pattern in the hat (Fig. 2.13c) are detected by the two models.

The class of simple cells is rather broad (DeAngelis et al., 1995). For instance, many simple cells have non-classical RFs which exhibit inhibition (Grigorescu et al., 2003a). These effects are outside the scope of this study. Furthermore, the scope of the proposed CORF computational model is limited to simple cells that respond to static stimuli.

The responses of real simple cells are often not proportional to stimulus contrast. This means that the response does not double if the contrast of the stimulus grows, for instance, from 40% to 80%. This property is known as response saturation (Sclar et al., 1990). We can also incorporate this nonlinear property in our model by applying a sublinear function with saturation, such as the sigmoid function, to the model

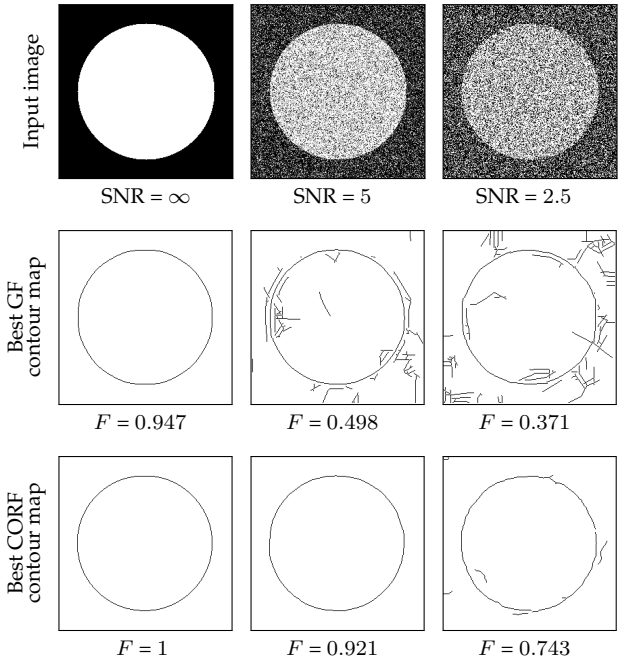


Figure 2.15: (First row) One noiseless and two noisy stimuli with different signal-to-noise ratios (SNR). Optimal contour maps achieved by the (second row) GF model and the (third row) proposed CORF model.

LGN responses. The resulting CORF model cell shows dependence of response on contrast that is qualitatively similar to the one shown by real simple cells, Fig. 2.16.

While we use geometric mean (essentially multiplication) as an output function of the proposed CORF computational model, we did not name it *Product* but more generally *Combination* of Receptive Fields. We did that in order to keep the possibility open to use other output functions in the future.

The implementation of the proposed CORF model is straightforward: it includes convolutions of computed model LGN responses to compute sub-unit responses[6], shifting appropriately these sub-unit responses in order to take into account that they need to be taken in different points, and a pixel-wise evaluation of a multivariate output function. Elsewhere, we refer to this type of filter, built on top of simpler filters, as Combination Of Shifted Filter Responses, abbreviated as COSFIRE (Azzopardi and Petkov, 2013).

The popularity of the GF model is largely due to the fact that a Gabor function

[6]This is equivalent to blurring the computed LGN responses.

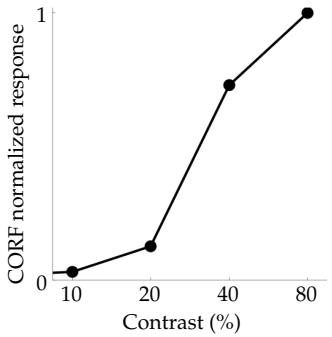


Figure 2.16: Responses of a CORF model cell as a function of stimulus contrast for the orientation preferred by the model. Here, the responses of the constituent model LGN cells are modified by a sigmoid function.

is a product of two well-known basic functions, a cosine function and a Gaussian function, that we all know since secondary school. Consequently, it can be easily remembered and its properties are easily understood. At a higher level, its elegance comes from the joint maximization of its compactness in the space domain and the spatial frequency domain (Daugman, 1985). It is amazing that such a conceptually simple and elegant function is a good first approximation to the properties of real neurons. The fact that it cannot reproduce all properties of real simple cells is just another evidence that the real world is less Platonic than we might wish it to be.

2.5 Summary

In this chapter we proposed a CORF computational model of an orientation-selective cell in area V1 of visual cortex. Orientation selectivity is achieved by combining in an AND-type operation the responses of a collection of model LGN cells with center-surround RFs that are properly aligned.

By means of simulated reverse correlation, we demonstrated that the RF of the proposed model cell can be divided in parallel elongated excitatory and inhibitory regions, an organization that is characteristic of simple cells. We also showed that, similar to real simple cells, the response of a CORF model cell to a drifting grating alternates between high and low values with a changing phase offset of the grating. Moreover, the proposed CORF model cell exhibits important nonlinear properties, namely cross orientation suppression, contrast invariant orientation tuning and response saturation, which are found in real simple cells. Such properties are not possessed by models that are based on linear spatial summation, such as the GF model. Therefore, the proposed CORF computational model should be seen as a

more realistic model of a simple cell.

We demonstrated the effectiveness of the proposed CORF model in a contour detection application and compared its performance with that of the popular GF model. For performance evaluation, we used two public data sets of natural images (RuG and Berkeley) with associated ground truth contour images. In both cases, the proposed CORF model outperforms the GF model (RuG: $t(39) = 4.39$, $p < 10^{-4}$ and Berkeley: $t(299) = 3.88$, $p < 10^{-4}$).

2.6 Conclusions

The proposed CORF model of a simple cell that relies on input from model LGN cells is anatomically more realistic than the GF model. Furthermore, it shares more properties with real simple cells than the GF model, such as cross orientation suppression, contrast invariant orientation tuning and response saturation. Finally, the proposed CORF model is more effective than the GF model in contour detection, which is assumed to be the primary biological task of simple cells.

Appendix A. Simulated Reverse Correlation

Reverse correlation is an electrophysiological technique in which a randomly generated binary or, more generally, white noise image is presented and the response of a cell to this stimulus is measured (de Boer and Kuyper, 1968; Jones and Palmer, 1987; Ringach and Shapley, 2004). Many different such images are presented and the images which elicit spikes from a given neuron are added together. Finally the average intensity value is subtracted from this accumulated image. The intensity of a pixel in the resulting image is indicative for the contribution of that position in the visual field to increase (if the pixel value is positive) or decrease (if it is negative) the probability of firing of the concerned cell. Reverse correlation can be applied to determine the spatio-temporal profiles of visual neurons by taking into account the delay of a spike after a stimulus image is presented and computing a separate summation image for each delay bin. We are not concerned with temporal aspects here.

We use *simulated* reverse correlation by which we mean that we present to our CORF model cell random binary images and measure the computed response of the model cell in a given location. This response is however not binary (spike or no spike) as with real neurons but a graded value which is intended to represent an approximation of the firing rate of a neuron. To take this aspect into account, we weight the presented binary stimulus image with the response which the model cell produces in the considered point and sum up these response-weighted binary input images. We subtract from the response-weighted sum of binary input images its mean across all pixels. This is the way in which we obtained the 2D function shown in Fig. 2.8(a-b).

Appendix B. Nonlinear vs. Linear CORF Response

Fig. 2.17 shows the orientation tuning curves of a linear CORF model in which the response is computed by summing up the responses of all afferent sub-units. This model produces response for all orientations of the presented stimulus, in contrast to the properties of real simple cells. If a threshold is used to constrain the response, the orientation bandwidth of the model becomes dependent on the contrast of the stimulus, while the orientation bandwidth of real simple cells does not depend on the contrast.

In these experiments, the RFs of the models are placed in the center of a given image. Every image is generated by superimposing a horizontal-edge stimulus (mask) on a vertical-edge stimulus (test) which is the preferred orientation stimulus of the

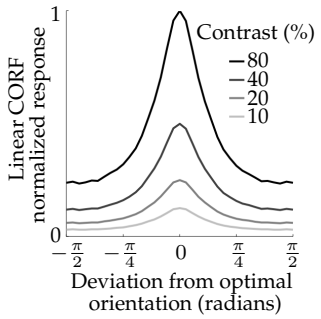


Figure 2.17: Orientation tuning curves at four different contrasts obtained for a linear CORF model, the output of which is computed as the sum of the responses of all afferent sub-units. This linear CORF model gives response even at orientations that are orthogonal to the preferred orientation.

concerned models.

Moreover, a linear CORF model does not possess the property of cross orientation suppression. On the contrary, a mask stimulus that is orthogonal to a simultaneously presented test stimulus of the orientation preferred by the model increases rather than suppresses the response, Fig. 2.18.

Appendix C. Evaluation of Inexact Contour Localization

Localization of detected contours may deviate slightly from the desired position specified in the corresponding ground truth. This might either be due to the inaccurate ground truth itself that is designed by a human observer, or otherwise by the inaccurate detection localization of the operator. For this reason, we use the tolerance method proposed by Grigorescu et al. (2004), where a detected pixel is considered correct if a corresponding ground truth contour pixel is present in a 5×5 neighborhood. Every ground truth contour pixel is considered only once, such that no two or more detected contour pixels can be matched to the same ground truth contour pixel. The false negatives and false positives can then be determined from the remaining unmatched contour pixels.

In the case of the Berkeley data set, where the ground truth is provided as a collection of multiple contour maps, we apply the following procedure. First, we match the operator contour map separately with every contour map in the corresponding ground truth collection. Subsequently, we compute the number of true positives as the total sum of the correctly matched boundary pixels with the set of ground truth

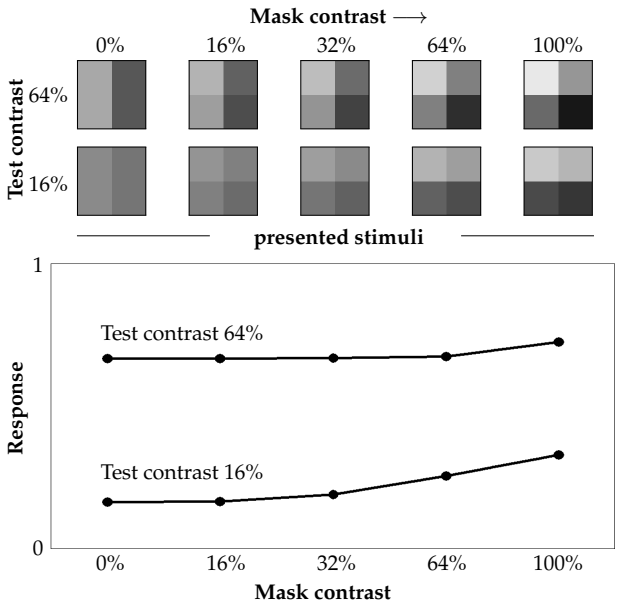


Figure 2.18: A linear CORF model does not exhibit cross orientation suppression, but rather it shows the opposite effect. The plots illustrate the responses of a linear CORF model to the images shown at the top, compare with Fig. 2.11.

contour maps. Similarly, the number of false negatives is the total sum of incorrectly unmatched boundary pixels. However, the false positives are only those boundary pixels that do not match any of the human observer maps. For further details we refer to Martin et al. (2004).

Published as:

G. Azzopardi, N. Petkov, *"Trainable COSFIRE filters for keypoint detection and pattern recognition,"* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35(2), pp. 490-503, 2013[1]

Chapter 3

COSFIRE Filters for Keypoint Detection

Abstract

Keypoint detection is important for many computer vision applications. Existing methods suffer from insufficient selectivity regarding the shape properties of features and are vulnerable to contrast variations and to the presence of noise or texture. We propose a trainable filter which we call COSFIRE (Combination Of Shifted FIlter REsponses) and use for keypoint detection and pattern recognition. It is automatically configured to be selective for a local contour pattern specified by an example. The configuration comprises selecting given channels of a bank of Gabor filters and determining certain blur and shift parameters. A COSFIRE filter response is computed as the weighted geometric mean of the blurred and shifted responses of the selected Gabor filters. It shares similar properties with some shape-selective neurons in visual cortex, which provided inspiration for this work. We demonstrate the effectiveness of the proposed filters in two applications: the recognition of handwritten digits (MNIST data set: 99.48% correct classification), and the detection and recognition of traffic signs in complex scenes (100% recall and precision). The proposed COSFIRE filters are conceptually simple and easy to implement. They are versatile keypoint detectors and are highly effective in practical computer vision applications.

[1]Section 3.1 is omitted because the application to the automatic detection of vascular bifurcations in retinal images is treated in more detail in Chapter 5.

3.1 Introduction

The detection of perceptually salient features, often referred to as keypoints or landmarks, is an important task in many computer vision applications, such as image registration, stereo camera calibration, object tracking and object recognition.

A substantial body of work has been done in this area and several methods have been proposed for the detection, description and matching of keypoints. These methods characterize a keypoint by a specific data structure derived from the image content in the surroundings of the concerned point. In this sense, the terms keypoint and landmark refer to a local pattern rather than a single point. The typical patterns of interest range from simple edges to corners and junctions, Fig 3.1. The Harris detector (Harris and Stephens, 1988), for instance, detects corner-like structures and achieves rotation invariance by using the eigenvalues of the Hessian matrix. This detector, which aroused much interest, was extended by including local gray-level invariants based on combinations of Gaussian derivatives (Schmid and Mohr, 1997). Later, scale-invariant approaches were proposed by selecting keypoints as the maxima points in a Laplacian (Lindeberg, 1998) or Difference-of-Gaussian (DoG) (Lowe, 1999) scale space. The Laplacian-based scale selection and the Harris detector were also combined into the so called Harris-Laplace operator (Mikolajczyk and Schmid, 2001).

A salient feature or keypoint is often characterized by a local image descriptor which may vary from a simple scalar value to a rich description, such as a feature vector, a bag of values, or some other data structure. An extensive survey of local descriptors can be found in (Mikolajczyk and Schmid, 2005). It compares a number of descriptors including derivatives of Gaussians (Florack et al., 1994), moment invariants (Mindru et al., 2004), complex features (Baumberg, 2000), responses of steerable filters (Freeman and Adelson, 1991), phase-based local features (Carneiro and Jep-

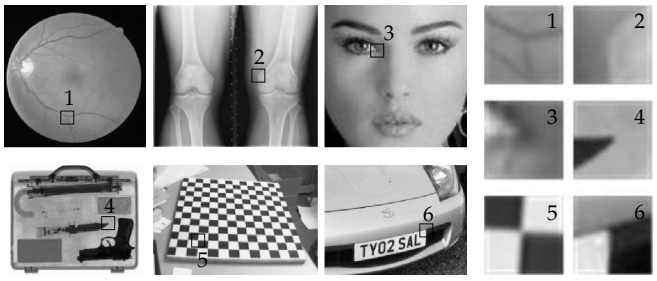


Figure 3.1: (a) Examples of corners and junction patterns marked in (a) photographic images and (b) their enlargements.

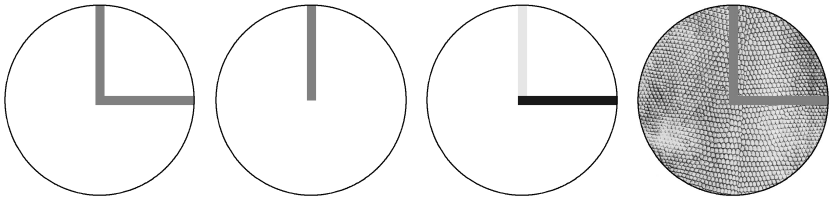


Figure 3.2: (a) Prototype pattern. (b) Test pattern which has 50% similarity (computed by template matching) to the prototype. (c-d) Test patterns that have only 30% similarity to the prototype due to (c) contrast differences and (d) presence of texture. From a shape detection point of view, the patterns in (c) and (d) are more similar to the prototype in (a) than the pattern in (b). This example shows the shortcomings of other models that are based on distance or dissimilarity of descriptors. The local image pattern is used as a descriptor in this example. Methods that compute local descriptors only shift the problem to a feature space.

son, 2003), and shows that the best performance is achieved with the SIFT descriptor (Lowe, 2004). Various extensions of the SIFT descriptor have been proposed including the PCA-SIFT (Yan Ke and Sukthankar, 2004) and the GLOH (Mikolajczyk and Schmid, 2005) which use principal component analysis for dimensionality reduction. Nevertheless, the original SIFT approach outperforms both mentioned variants and seems to be the most popular keypoint descriptor currently. Recently, another operator called SURF (Bay et al., 2008) has been introduced which is somewhat similar to SIFT and speeds up the efficiency of keypoint selection.

The detection of keypoints that are similar to some keypoint which is selected as a prototype is typically done by computing a similarity (or dissimilarity) measure that is usually based on the Euclidean (or some other) distance between the respective keypoint descriptors. These methods are not robust to contrast variations and as a result they suffer from insufficient selectivity to the shape properties of features. This issue is illustrated by Fig. 3.2. The pattern in Fig. 3.2a that is formed by two lines that make a right-angle vertex is, as a shape, very different from a pattern that is formed by just one of the constituent lines, Fig. 3.2b. Approaches that are based on the dissimilarity between keypoint descriptors such as the ones mentioned above may find these two patterns similar to a considerable extent. On the other hand, such methods might produce lower similarity scores for patterns that are regarded similar from the aspect of shape by a human observer, but show differences in contrast and/or contain texture, Fig. 3.2(c-d).

In this article we are interested in the detection of contour-based patterns. We introduce trainable keypoint detection operators that are configured to be selective for given local patterns defined by the geometrical arrangement of contour segments. The proposed operators are inspired by the properties of a specific type of shape-

selective neuron in area V4 of visual cortex, which exhibit selectivity for parts of (curved) contours or for combinations of line segments (Pasupathy and Connor, 1999, 2002).

We call the proposed keypoint detector COSFIRE (Combination Of Shifted FIlter REsponses) filter as the response of such a filter in a given point is computed as a function of the shifted responses of simpler (in this case orientation-selective) filters. Using shifted responses of simpler filters – Gabor filters in this study – corresponds to combining their respective supports at different locations to obtain a more sophisticated filter with a bigger support. The specific function that we use here to combine filter responses is weighted geometric mean, essentially multiplication, which has specific advantages regarding shape recognition and robustness to contrast variations. Such a model design decision is mainly motivated by the better results obtained using multiplication vs. addition. It gets further support by psychophysical evidence (Gheorghiu and Kingdom, 2009) that curved contour parts are likely detected by a neural mechanism that multiplies the responses of afferent subunits (sensitive for different parts of the curve pattern). Due to the multiplicative character of the output function, a COSFIRE filter produces a response only when all constituent parts of a pattern of interest are present.

A COSFIRE filter is conceptually simple and straightforward to implement: it requires the application of selected Gabor filters, Gaussian blurring of their responses, shifting of the blurred responses by specific, different vectors, and multiplying the shifted responses. The questions of which Gabor filters to use, how much to blur their responses and how to shift the blurred responses are answered in a COSFIRE filter configuration process in which a local pattern of interest that defines a keypoint is automatically analysed. The configured COSFIRE filter can then successfully detect the same and similar patterns. We also show how the proposed COSFIRE filters can achieve invariance to rotation, scale, reflection and contrast inversion.

The rest of the chapter is organized as follows: in Section 3.2 we present the COSFIRE filter and demonstrate how it can be trained and used to detect local contour patterns. In Section 3.3, we demonstrate the effectiveness of the proposed trainable COSFIRE filters in two practical applications: the recognition of handwritten digits and the detection and the recognition of traffic signs in complex scenes. Section 3.4 contains a discussion of some aspects of the proposed approach and highlights the differences that distinguish it from other approaches. Finally we draw conclusions in Section 3.5.

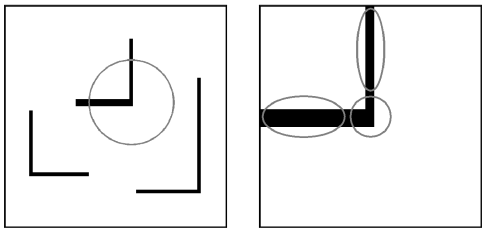
3.2 Method

3.2.1 Overview

The following example illustrates the main idea of our method. Fig. 3.3a shows an input image containing three vertices. We consider the encircled vertex, which is shown enlarged in Fig. 3.3b as a (prototype) pattern of interest and use it to automatically configure a COSFIRE filter that will respond to the same and similar patterns.

The two ellipses shown in Fig. 3.3b represent the dominant orientations in the neighbourhood of the specified point of interest. We detect such lines by symmetric Gabor filters. The central circle represents the overlapping supports of a group of such filters. The response of the proposed COSFIRE detector is computed by combining the responses of these Gabor filters in the centers of the corresponding ellipses by multiplication. The preferred orientations of these filters and the locations at which we take their responses are determined by analysing the local prototype pattern used for the configuration of the concerned COSFIRE filter. Consequently, the filter is selective for the presented local spatial arrangement of lines of specific orientations and widths. Taking the responses of Gabor filters at different locations around a point can be implemented by shifting the responses of these Gabor filters by different vectors before using them for the pixel-wise evaluation of a multivariate function which gives the COSFIRE filter output.

In the next sub-sections, we explain the automatic configuration process of a COSFIRE filter that will respond to a given prototype feature of interest and similar patterns. The configuration process determines which responses of which Gabor filters in which locations need to be multiplied in order to obtain the output of the filter.



(a) (b)

Figure 3.3: (a) Synthetic input image (of size 256×256 pixels). The circle indicates a prototype feature of interest that is manually selected by a user. (b) Enlargement of the selected feature. The ellipses represent the support of line detectors that are identified as relevant for the concerned feature.

3.2.2 Detection of Orientations by 2D Gabor Filters

We build the proposed COSFIRE filter using as input the responses of Gabor filters which are known for their orientation selectivity.

We denote by $g_{\lambda,\theta}(x,y)$ the response of a Gabor filter of preferred wavelength λ and orientation θ to a given input image. Such a filter has other parameters, such as spatial aspect ratio, bandwidth and phase offset, that we skip here for brevity. The responses of a symmetrical and an antisymmetrical filter can be combined in a Gabor energy filter. Surround suppression can also be applied to Gabor (energy) filter responses to reduce responses to texture and improve the detectability of object contours. For brevity of presentation, we do not consider all these aspects of Gabor filters here and we refer to (Petkov, 1995; Petkov and Kruizinga, 1997; Kruizinga and Petkov, 1999; Grigorescu et al., 2002; Petkov and Westenberg, 2003; Grigorescu et al., 2003c,b) for technical details and to our online implementation[2]. We normalize[3] all Gabor functions that we use in such a way that all positive values of such a function sum up to 1 and all negative values sum up to -1.

We threshold the responses of Gabor filters at a given fraction t_1 ($0 \leq t_1 \leq 1$) of the maximum response of $g_{\lambda,\theta}(x,y)$ across all combinations of values (λ,θ) used and all positions (x,y) in the image, and denote these thresholded responses by $|g_{\lambda,\theta}(x,y)|_{t_1}$. We comment on the choice of the value of t_1 in Sections 3.3 and 3.4.

3.2.3 Configuration of a COSFIRE Filter

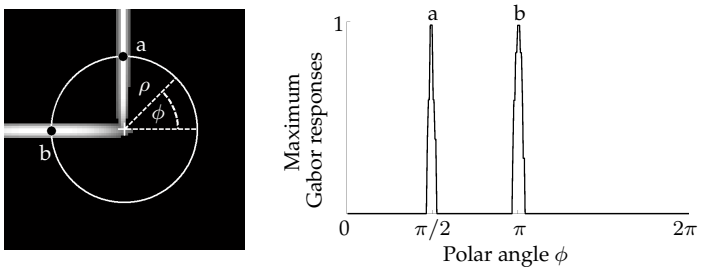
A COSFIRE filter uses as input the responses of some Gabor filters, each characterized by parameter values (λ_i,θ_i), around certain positions (ρ_i,ϕ_i) with respect to the center of the COSFIRE filter. A set of four parameter values $(\lambda_i,\theta_i,\rho_i,\phi_i)$ characterizes the properties of a contour part that is present in the specified area of interest: $\lambda_i/2$ represents the width, θ_i represents the orientation and (ρ_i,ϕ_i) represents the location. In the following we explain how we obtain the parameter values of such contour parts around a given point of interest.

We consider the responses of a bank of Gabor filters along a circle[4] of a given radius ρ around a selected point of interest, Fig. 3.4. In each position along that circle, we take the maximum of all responses across the possible values of (λ,θ) used in the filter bank. The positions that have values greater than the corresponding values of the neighbouring positions along an arc of angle $\pi/8$ are chosen as the

[2] `http://matlabserver.cs.rug.nl`

[3] This normalization ensures that the response to an image of constant intensity is 0. Without normalization, this is true only for antisymmetrical filters. It also ensures that the response to a line of width w will be largest for a symmetrical filter of preferred wavelength $\lambda = 2w$. We mention this explicitly because line detection is essential in one application that we present in Section 3.3.

[4] For $\rho = 0$ we only consider the point of interest.



(a) (b)

Figure 3.4: Configuration of a COSFIRE filter. (a) The gray-level of a pixel represents the maximum value superposition of the thresholded (at $t_1 = 0.2$) responses of a bank of Gabor filters (5 wavelengths $\lambda \in \{4, 4\sqrt{2}, 8, 8\sqrt{2}, 16\}$ and 8 orientations $\theta \in \{\frac{\pi i}{8}, i = 0 \dots 7\}$) at that position. The white cross indicates the location of the point of interest selected by a user and the bright circle of a given radius (here $\rho = 30$) indicates the locations considered around the point of interest. (b) Values of the maximum value superposition of thresholded Gabor filter responses along the concerned circle. The labeled black dots in (a) mark the positions (relative to the center of the filter) at which the respective strongest Gabor filter responses are taken. These two positions correspond to the two local maxima in the plot in (b).

points that characterize the dominant orientations around the point of interest. We determine the polar coordinates (ρ_i, ϕ_i) for each such point with respect to the center of the filter. For such a location (ρ_i, ϕ_i) we then consider all combinations of (λ, θ) for which the corresponding responses $g_{\lambda,\theta}(x, y)$ are greater than a fraction $t_2 = 0.75$ of the maximum of $g_{\lambda,\theta}(x, y)$ across the different combinations of values (λ, θ) used. For each value θ that satisfies this condition, we consider a single value of λ, the one for which $g_{\lambda,\theta}(x, y)$ is the maximum of all responses across all values of λ. For each distinct pair of (λ, θ) and for location (ρ_i, ϕ_i) we obtain a tuple $(\lambda_i, \theta_i, \rho_i, \phi_i)$. Thus, multiple tuples can be formed for the same location (ρ_i, ϕ_i). In Section 3.4, we provide further comment on the choice of the value of t_2.

We denote by $S_f = \{(\lambda_i, \theta_i, \rho_i, \phi_i) \mid i = 1 \dots n_f\}$ the set of parameter value combinations, which fulfill the above conditions. The subscript f stands for the local prototype pattern around the selected point of interest. Every tuple in the set S_f specifies the parameters of some contour part in f.

For the point of interest shown in Fig. 3.4a, with two values of the parameter ρ ($\rho \in \{0, 30\}$), the selection method described above results in four contour parts with parameter values specified by the tuples in the following set:

$$S_f = \left\{ \begin{array}{l} (\lambda_1 = 8, \ \ \theta_1 = 0, \ \ \rho_1 = 0, \ \ \phi_1 = 0), \\ (\lambda_2 = 8, \ \ \theta_2 = 0, \ \ \rho_2 = 30, \ \phi_2 = \frac{\pi}{2}), \\ (\lambda_3 = 16, \ \theta_3 = \frac{\pi}{2}, \ \rho_3 = 0, \ \ \phi_3 = 0), \\ (\lambda_4 = 16, \ \theta_4 = \frac{\pi}{2}, \ \rho_4 = 30, \ \phi_4 = \pi) \end{array} \right\}$$

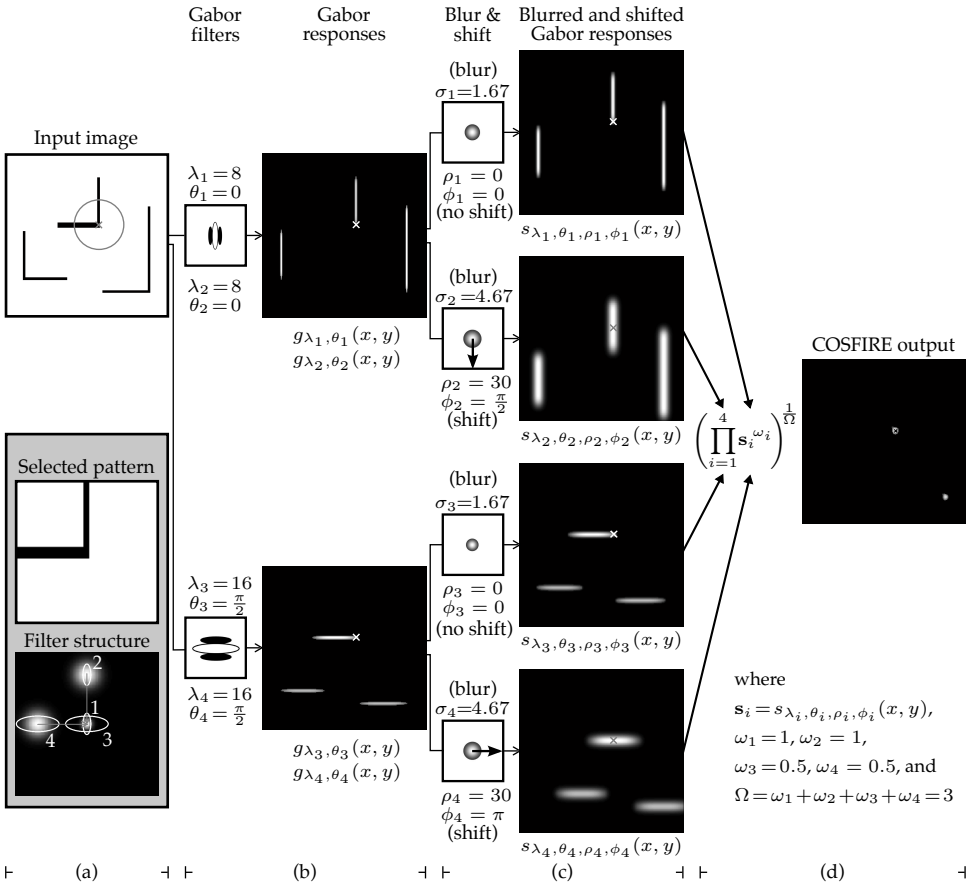


Figure 3.5: (a) Input image (of size 256×256 pixels). The enframed inlay images show (top) the enlarged prototype feature of interest, which is the vertex encircled in the input image and (bottom) the structure of the COSFIRE filter that is configured for this feature. This filter is trained to detect the spatial local arrangement of four contour parts. The ellipses illustrate the wavelengths and orientations of the Gabor filters, and the bright blobs are intensity maps for Gaussian functions that are used to blur the responses of the corresponding Gabor filters. The blurred responses are then shifted by the corresponding vectors. (b) Each contour part of the input pattern is detected by a Gabor filter with a given preferred wavelength λ_i and orientation θ_i. Two of these parts ($i = \{1, 2\}$) are detected by the same Gabor filter and the other two parts ($i = \{3, 4\}$) are detected by another Gabor filter; therefore, only two distinct Gabor filters are selected from the filter bank. (c) We then blur the thresholded (here at $t_1 = 0.2$) response $|g_{\lambda_i,\theta_i}(x, y)|_{t_1}$ of each concerned Gabor filter and subsequently shift the resulting blurred response images by corresponding polar coordinate vectors $(\rho_i, \phi_i + \pi)$. (d) Finally, we obtain the output of the COSFIRE filter by computing the weighted geometric mean (here $\sigma' = 25.48$) of all the blurred and shifted thresholded Gabor filter responses. The '$\times$' marker indicates the location of the specified point of interest. The two local maxima in the output of the COSFIRE filter correspond to the two similar vertices in the input image.

The last tuple in S_f, $(\lambda_4 = 16, \theta_4 = \pi/2, \rho_4 = 30, \phi_4 = \pi)$, for instance, describes a contour part with a width of $(\lambda_4/2 =)$ 8 pixels and an orientation $\theta_4 = \pi/2$ that can be detected by a Gabor filter with preferred wavelength $\lambda_4 = 16$ and orientation $\theta_4 = \pi/2$, at a position of $\rho_4 = 30$ pixels to the left ($\phi_4 = \pi$) of the point of interest; this location is marked by the label 'b' in Fig. 3.4. This selection is the result of the presence of a horizontal line to the left of the center of the feature that is used for the configuration of the filter.

3.2.4 Blurring and Shifting Gabor Filter Responses

The above analysis of the considered local pattern of interest f indicates that this pattern produces four strong responses $g_{\lambda_i,\theta_i}(x, y)$ of Gabor filters with parameters $(\lambda_1 = 8, \theta_1 = 0)$, $(\lambda_2 = 8, \theta_2 = 0)$, $(\lambda_3 = 16, \theta_3 = \pi/2)$, and $(\lambda_4 = 16, \theta_4 = \pi/2)$ in the corresponding positions with polar coordinates (ρ_i, ϕ_i) with respect to the filter center. Next, we use these responses to compute the output of the COSFIRE filter. Since the concerned responses are in different positions (ρ_i, ϕ_i) with respect to the filter center, we first shift them appropriately so that they come together in the filter center. The COSFIRE filter output can then be evaluated as a pixel-wise multivariate function of the shifted Gabor filter responses.

Before these shift operations, we blur the Gabor filter responses in order to allow for some tolerance in the position of the respective contour parts. We define the blurring operation as the computation of maximum value of the weighted thresholded responses of a Gabor filter. For weighting we use a Gaussian function $G_\sigma(x, y)$, the standard deviation σ of which is a linear function of the distance ρ from the center of the COSFIRE filter:

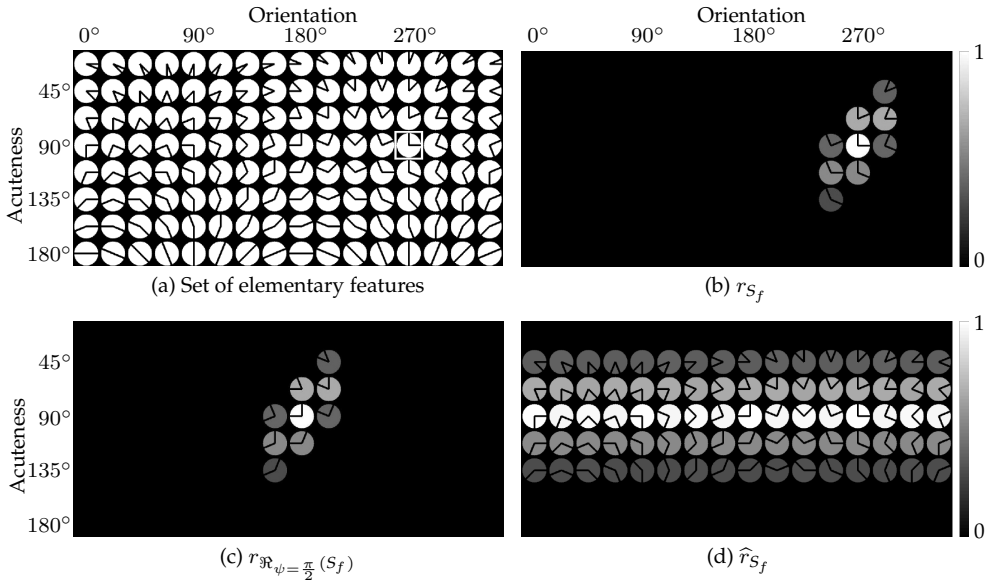
$$\sigma = \sigma_0 + \alpha\rho \tag{3.1}$$

where σ_0 and α are constants. The choice of the linear function in Eq. 3.1 is explained in Section 3.4. The value of the parameter α determines the orientation tuning of the COSFIRE filter: the orientation bandwidth becomes broader with an increasing value of α.

Next, we shift the blurred responses of each selected Gabor filter (λ_i, θ_i) by a distance ρ_i in the direction opposite to ϕ_i. In polar coordinates, the shift vector is specified by $(\rho_i, \phi_i + \pi)$. In cartesian coordinates, it is $(\Delta x_i, \Delta y_i)$ where $\Delta x_i = -\rho_i \cos\phi_i$, and $\Delta y_i = -\rho_i \sin\phi_i$. We denote by $s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x, y)$ the blurred and shifted response of the Gabor filter that is specified by the i-th tuple $(\lambda_i, \theta_i, \rho_i, \phi_i)$ in the set S_f:

$$s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x, y) \stackrel{\text{def}}{=} \max_{x',y'} \{ |g_{\lambda_i,\theta_i}(x - x' - \Delta x_i, y - y' - \Delta y_i)|_{t_1} \, G_\sigma(x', y') \} \tag{3.2}$$

where $-3\sigma \leq x', y' \leq 3\sigma$.



(a) Set of elementary features (b) r_{S_f}

(c) $r_{\Re_{\psi=\frac{\pi}{2}}(S_f)}$ (d) $\widehat{r}_{S_f}$

Figure 3.6: (a) A set of elementary features. The enframed feature is used as a prototype for configuring a COSFIRE filter. (b) Responses of the configured filter rendered by shading of the features. (c) Responses of a rotated version ($\psi = \frac{\pi}{2}$) of the filter obtained by manipulation of the filter parameters and (d) rotation-invariant responses for sixteen discrete orientations.

Fig. 3.5 illustrates the blurring and shifting operations for this COSFIRE filter, applied to the image in Fig. 3.3a. For each of the four contour parts detected in the prototype feature pattern we first compute the corresponding Gabor filter responses and then we blur and shift these responses accordingly.

In practice, the computation of one blurred response (for the same values of the parameters λ, θ and ρ), for instance with $s_{\lambda,\theta,\rho,\phi=0}(x,y)$, is sufficient: the result of $s_{\lambda,\theta,\rho,\phi}(x,y)$ for any value of ϕ can be obtained from the result of the output of $s_{\lambda,\theta,\rho,\phi=0}(x,y)$ by appropriate shifting.

3.2.5 Response of a COSFIRE Filter

We define the response $r_{S_f}(x,y)$ of a COSFIRE filter as the weighted geometric mean of all the blurred and shifted thresholded Gabor filter responses $s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y)$ that correspond to the properties of the contour parts described by S_f:

$$r_{S_f}(x,y) \stackrel{\text{def}}{=} \left| \left(\prod_{i=1}^{|S_f|} \left(s_{\lambda_i,\sigma_i,\rho_i,\phi_i}(x,y) \right)^{\omega_i} \right)^{1/\sum_{i=1}^{|S_f|}\omega_i} \right|_{t_3} \tag{3.3}$$

where $\omega_i = \exp^{-\frac{\rho_i^2}{2\sigma'^2}}$, $0 \leq t_3 \leq 1$, and $|.|_{t_3}$ stands for thresholding the response at a fraction t_3 of its maximum across all image coordinates (x, y). For $1/\sigma' = 0$, the computation of the COSFIRE filter becomes equivalent to the standard geometric mean, where the s-quantities have the same contribution. Otherwise, for $1/\sigma' > 0$, the input contribution of s-quantities decreases with an increasing value of the corresponding parameter ρ. In our experiments we use a value of the standard deviation σ' that is computed as a function of the maximum value of the given set of ρ values: $\sigma' = (-{\rho_{\max}}^2/2\ln 0.5)^{1/2}$, where $\rho_{\max} = \max_{i\in\{1\ldots|S_f|\}}\{\rho_i\}$. We make this choice in order to achieve a maximum value $\omega = 1$ of the weights in the center (for $\rho = 0$), and a minimum value $\omega = 0.5$ in the periphery (for $\rho = \rho_{\max}$).

Fig. 3.5 shows the output of a COSFIRE filter which is defined as the weighted geometric mean of four blurred and shifted images from the responses of two Gabor filters. Note that this filter responds at points where a pattern is present which is identical or similar to the prototype pattern f at and around the selected point of interest, which was used in the configuration of the filter. In this example, the COSFIRE filter reacts strongly in a given point to a local pattern that contains a horizontal line to the left of that point, a vertical line above it, together with a horizontal and a vertical line at the point.

Fig. 3.6a shows a set of elementary features that are angles of different acuteness and orientations. For the illustration in Fig. 3.6 we configure a COSFIRE filter using the enframed local pattern in Fig. 3.6a where the point of interest is positioned on the corner of the vertex. The structure of the filter is determined by using three values of ρ ($\rho \in \{0, 12, 30\}$). Fig. 3.6b shows the responses of this COSFIRE filter where the strength of the maximum filter response to a given feature is rendered as a gray level shading of that feature. The maximum response is reached at or near the corner. In this case, the COSFIRE filter achieves the strongest response to the local prototype pattern that was used to configure it, but it also reacts, with less than the maximum response, to angles that differ slightly in acuteness and/or orientation. This example illustrates the selectivity and the generalization ability of the proposed filter.

3.2.6 Achieving Invariance

In the following we explain how we achieve invariance to rotation, scale, reflection and contrast inversion.

Rotation Invariance

Using the set S_f that defines the concerned filter, we form a new set $\Re_\psi(S_f)$ that defines a new filter, which is selective for a version of the prototype feature f that is rotated by an angle ψ:

$$\Re_\psi(S_f) \stackrel{\text{def}}{=} \{(\lambda_i, \theta_i + \psi, \rho_i, \phi_i + \psi) \mid \forall\, (\lambda_i, \theta_i, \rho_i, \phi_i) \in S_f\} \tag{3.4}$$

For each tuple $(\lambda_i, \theta_i, \rho_i, \phi_i)$ in the original filter S_f that describes a certain local contour part, we provide a counterpart tuple $(\lambda_i, \theta_i + \psi, \rho_i, \phi_i + \psi)$ in the new set $\Re_\psi(S_f)$. The orientation of the concerned contour part and its polar angle position with respect to the center of the filter are off-set by an angle ψ relative to the values of the corresponding parameters of the original part.

Fig. 3.6c shows the responses $r_{\Re_\psi(S_f)}$ of the COSFIRE filter that correspond to $\Re_\psi(S_f)$ to the set of elementary features shown in Fig. 3.6a. This filter responds selectively to a version of the original prototype feature f rotated counterclockwise at an angle of (ψ =) $\pi/2$. It is, however, configured by manipulating the set of parameter value combinations, rather than by computing them from the responses to a rotated version of the original prototype pattern f.

A rotation-invariant response is achieved by taking the maximum value of the responses of filters that are obtained with different values of the parameter ψ:

$$\hat{r}_{S_f}(x, y) \stackrel{\text{def}}{=} \max_{\psi \in \Psi} \{r_{\Re_\psi(S_f)}(x, y)\} \tag{3.5}$$

where Ψ is a set of n_ψ equidistant orientations defined as $\Psi = \{\frac{2\pi}{n_\psi} i \mid 0 \leq i < n_\psi\}$. Fig. 3.6d shows the maximum superposition $\hat{r}_{S_f}(x, y)$ for $n_\psi = 16$. The filter according to Eq. 3.5 produces the same response to local patterns that are versions of each other, obtained by rotation at discrete angles $\psi \in \Psi$.

As to the response of the filter to patterns that are rotated at angles of intermediate values between those in Ψ, it depends on the orientation selectivity of the filter S_f that is influenced by the orientation bandwidth of the involved Gabor filters and by the value of the parameter α in Eq. 3.1. Fig. 3.7 illustrates the orientation selectivity of the COSFIRE filter, which is configured with the enframed local prototype pattern in Fig. 3.6a using $\alpha = 0.1$. A maximum response is obtained for the local prototype pattern that was used to configure this filter. The response declines with the deviation of the orientation of the local input pattern from the optimal one and practically disappears when this deviation is greater than $\pi/8$. When the deviation of the orientation is $\pi/16$, the response of the filter is approximately half of the maximum response. This means that the half-response bandwidth of this COSFIRE filter is $\pi/8$. Thus, $n_\psi = 16$ distinct preferred orientations (in intervals of $\pi/8$) ensure

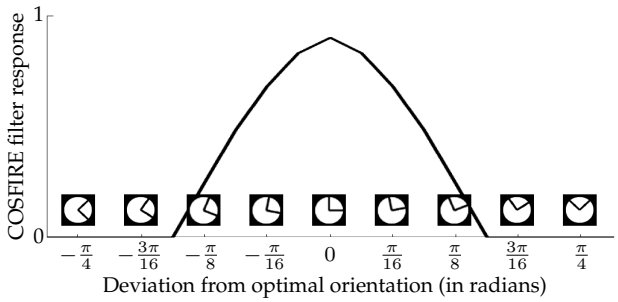


Figure 3.7: Orientation selectivity of a COSFIRE filter that is configured with a right-angle vertex.

sufficient response for any orientation of the feature used to configure the filter.

As demonstrated by Fig. 3.6d, when the concerned filter is applied in rotation-invariant mode ($n_\psi = 16$), it responds selectively to the prototype pattern, a right angle, independently of the orientation of the angle.

Scale Invariance

Scale invariance is achieved in a similar way. Using the set S_f that defines the concerned filter, we form a new set $T_\upsilon(S_f)$ that defines a new filter which is selective for a version of the prototype feature f that is scaled in size by a factor υ:

$$T_\upsilon(S_f) \stackrel{\text{def}}{=} \{(\upsilon\lambda_i, \theta_i, \upsilon\rho_i, \phi_i) \mid \forall\, (\lambda_i, \theta_i, \rho_i, \phi_i) \in S_f\} \tag{3.6}$$

For each tuple $(\lambda_i, \theta_i, \rho_i, \phi_i)$ in the original filter S_f that describes a certain local contour part, we provide a counterpart tuple $(\upsilon\lambda_i, \theta_i, \upsilon\rho_i, \phi_i)$ in the new set $T_\upsilon(S_f)$. The width of the concerned contour part and its distance to the center of the filter are scaled by the factor υ relative to the values of the corresponding parameters of the original part.

A scale-invariant response is achieved by taking the maximum value of the responses of filters that are obtained with different values of the parameter υ:

$$\tilde{r}_{S_f}(x, y) \stackrel{\text{def}}{=} \max_{\upsilon \in \Upsilon} \{r_{T_\upsilon(S_f)}(x, y)\} \tag{3.7}$$

where Υ is a set of υ values equidistant on a logarithmic scale defined as $\Upsilon = \{2^{\frac{i}{2}} \mid i \in \mathbb{Z}\}$.

Reflection Invariance

As to reflection invariance we first form a new set $\acute{S}_f$ from the set S_f as follows:

$$\acute{S}_f \stackrel{\text{def}}{=} \{(\lambda_i, \pi - \theta_i, \rho_i, \pi - \phi_i) \mid \forall\ (\lambda_i, \theta_i, \rho_i, \phi_i) \in S_f\} \tag{3.8}$$

The new filter, which is defined by the set $\acute{S}_f$, is selective for a reflected version of the prototype feature f about the $y-$axis. A reflection-invariant response is achieved by taking the maximum value of the responses of the filters S_f and $\acute{S}_f$:

$$\acute{r}_{S_f}(x,y) \stackrel{\text{def}}{=} \max\{r_{S_f}(x,y), r_{\acute{S}_f}(x,y)\} \tag{3.9}$$

Combined Invariance to Rotation, Scale and Reflection

A combined rotation-, scale-, and reflection-invariant response is achieved by taking the maximum value of the rotation- and scale-invariant responses of the filters S_f and $\acute{S}_f$ that are obtained with different values of the parameters ψ and υ:

$$\bar{r}_{S_f}(x,y) \stackrel{\text{def}}{=} \max_{\psi\in\Psi, \upsilon\in\Upsilon} \{\hat{r}_{\Re_\psi(T_\upsilon(S_f))}(x,y), \hat{r}_{\Re_\psi(T_\upsilon(\acute{S}_f))}(x,y)\} \tag{3.10}$$

Invariance to Contrast Inversion

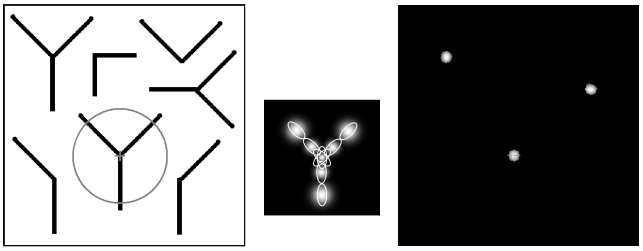
Next to the above geometric invariances, we can achieve invariance to contrast inversion by using Gabor filters with inverse polarity.

We do not elaborate further on this possibility because we do not use it in the applications presented above.

3.2.7 Detection of More Complex Patterns

The filter considered above is selective for a local pattern that consists of two lines forming an angle. However, in the configuration of the COSFIRE filter we made no assumptions about the specific type of local pattern it should detect. The configuration result is determined by the local prototype pattern presented.

Next, we illustrate the configuration a filter that can detect a bifurcation pattern formed by three lines of different orientations, Fig. 3.8(a-b). In Fig. 3.8c we show the rotation-invariant response of the concerned COSFIRE filter for the input image in Fig. 3.8a. Besides the original local prototype pattern that was used to configure this filter, it correctly detects the presence of another two similar features: one in a vertical orientation and the other pattern in a horizontal orientation.



(a) (b) (c)

Figure 3.8: (a) Synthetic input image (of size 256×256 pixels). (b) The structure of a COSFIRE filter that is configured using the encircled pattern in (a) with three values of ρ ($\rho \in \{0, 12, 30\}$) and $\sigma_0 = 2.5$. (c) Rotation-invariant response $\widehat{r}_{S_f}$ of the COSFIRE filter (here $\sigma' = 25.48$).

3.3 Applications

In the following we demonstrate the effectiveness of the proposed COSFIRE filters by applying them in two practical applications: the recognition of handwritten digits and the detection and recognition of traffic signs in complex scenes.

3.3.1 Recognition of Handwritten Digits

Handwritten digit recognition is a challenging task in the community of pattern recognition, which has various commercial applications, such as bank cheque processing and postal mail sorting. It has been used as a benchmark for comparing shape recognition methods. Feature extraction plays a significant role in the effectiveness of such systems. A detailed review of the state-of-the-art methods is given in (Liu et al., 2003).

In the following, we show how the proposed trainable COSFIRE filters can be configured to detect specific parts of handwritten digits. Consequently, the collective responses of multiple such filters can be used as a shape descriptor of a given handwritten digit. We use the well known modified NIST (MNIST) data set (LeCun et al., 1998) to evaluate the performance of this approach. This data set comprises 60000 training and 10000 test digits[5] where each digit is given as a grayscale image of size 28×28 pixels, Fig. 3.9.

In the configuration step, we choose a random subset of digit images from each digit class. For each such digit image we choose a random location in the image and use the local stroke pattern around that location to configure a COSFIRE filter. We use a given randomly selected location for the configuration of a COSFIRE filter

[5]The MNIST data set is available online: `http://yann.lecun.com/exdb/mnist`

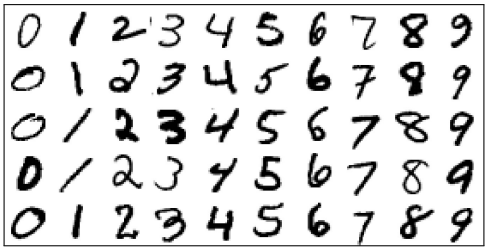


Figure 3.9: Examples of handwritten digits from the MNIST data set.

only if that filter consists of at least four tuples, otherwise we choose a different location. We impose this restriction in order to avoid the selection of small digit fragments as prototype patterns, which may consequently result in filters with low discriminative power. We provide further comments on the discriminative abilities of these COSFIRE filters in Section 3.4. For this application, we use three values of ρ ($\rho \in \{0, 3, 8\}$), $t_2 = 0.75$, $\sigma_0 = 0.83$, $\alpha = 0.1$, and a bank of antisymmetric Gabor filters with 16 equidistant orientations ($\theta \in \{\frac{\pi i}{8} \mid i = 0 \dots 15\}$) and one wavelength of ($\lambda =$) $2\sqrt{2}$. Fig. 3.10 illustrates the configuration of four such COSFIRE filters using local prototype patterns (parts of digits) that are randomly selected from four handwritten digits.

We perform a number of experiments with different values of the threshold parameter t_1 ($t_1 \in \{0, 0.05, 0.1, 0.15\}$). The values of the other parameters mentioned above are kept fixed for all experiments. For each value of t_1 we run an experiment by configuring up to 500 COSFIRE filters per digit class. We repeat such an experiment five times and report the average recognition rate. Repetition of experiments is necessary in order to compensate for the random selection of training digit images and the random selection of locations within these images that are used to configure the concerned filters.

After the configuration of a certain number of COSFIRE filters, every digit to which the set of these filters is applied can be described by a vector where each element corresponds to the maximum response of a COSFIRE filter across all locations in the input image. For instance, with 500 filters per digit class and 10 digit classes, a digit image to which this set of 5000 COSFIRE filters is applied is described by a vector of 5000 elements. For this application, the responses of the concerned Gabor filters provide equal contribution ($1/\sigma' = 0$) to the output of the corresponding COSFIRE filter.

The feature vectors obtained for the digit images of the training set are then used to train an all-pairs multi-class (with majority vote) support vector machine (SVM) classifier with a linear kernel. In Fig. 3.11a we plot the recognition rates that we

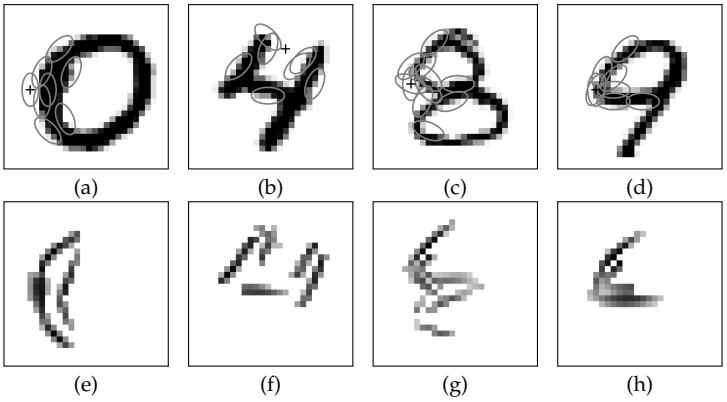


Figure 3.10: Example of the configuration of four COSFIRE filters, one for each of the handwritten digits 0, 4, 8, and 9. (a-d) The '+' markers show the randomly selected locations. The ellipses around the marked locations represent the support of the Gabor filters that are determined in the configuration of the concerned COSFIRE filters. (e-h) The corresponding reconstructions of the local patterns that are illustrated as a superposition of the thresholded ($t_1 = 0.1$) Gabor filter (inverted) responses, which contribute to the responses of the respective COSFIRE filters.

achieve for different values of the threshold t_1 and for different numbers of COSFIRE filters used. We achieve a maximum recognition rate of 99.40% with 4500 COSFIRE filters, where the filters are used in a non-invariant mode; i.e. without compensation for possible pattern reflection, rotation and scaling ($\psi = 0$, $\upsilon = 1$). In Fig. 3.9 one can observe, however, that some of the handwritten digits given in the MNIST data set differ slightly in orientation. We consider this fact and repeat the five experiments for the threshold $t_1 = 0$ (that contributed to the best performance so far), but this time applying the same COSFIRE filters in a partially rotation-invariant mode with five values of the rotation tolerance angle ψ ($\psi \in \{-\frac{\pi}{4}, -\frac{\pi}{8}, 0, \frac{\pi}{8}, \frac{\pi}{4}\}$). The plots in Fig. 3.11b show that the performance that is achieved with the partially rotation-invariant filters is improved to a maximum recognition rate of 99.48% with 4000 filters. This means that the error rate is decreased by 13.33% and that 500 less filters are required.

The recognition rate of 99.48% that we achieve is comparable to the best results obtained with other approaches[6] applied on the MNIST data set. In particular, our method outperforms the shape context approach (99.37% (Belongie et al., 2002)), and three other approaches (94.2% (Oberhoff and Kolesnik, 2008), 97.62% (Borji et al., 2008) and 98.73% (Hamidi and Borji, 2010)) that use biologically inspired features

[6] A list of results obtained by state-of-the-art approaches is maintained `http://yann.lecun.com/exdb/mnist/`

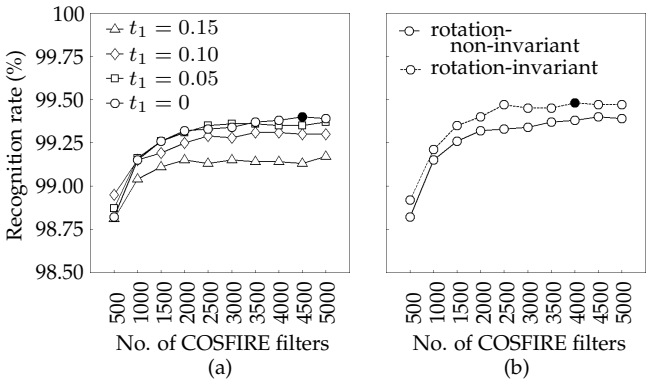


Figure 3.11: Experimental results achieved on the MNIST data set. (a) The four plots show the recognition rates achieved for different values of the threshold t_1 as a function of the number of COSFIRE filters used. The filled-in circle represents the maximum recognition rate of 99.40% achieved for $t_1 = 0$ with 4500 filters. In these experiments the COSFIRE filters are used in rotation-non-invariant mode. (b) Performance comparison between the same set of COSFIRE filters and $t_1 = 0$ that are first applied in rotation-non-invariant mode and then in a partially rotation-invariant mode. Here partial rotation-invariance is based on five values of the rotation tolerance angle ψ ($\psi \in \{-\frac{\pi}{4}, -\frac{\pi}{8}, 0, \frac{\pi}{8}, \frac{\pi}{4}\}$). The performance improves with partial rotation-invariant filters that achieve a maximum recognition rate of 99.48% (shown as a filled-in circle) with 4000 filters.

combined with a multi-layer perceptron (MLP) (Oberhoff and Kolesnik, 2008) and a linear SVM classifier (Borji et al., 2008; Hamidi and Borji, 2010). The highest recognition rate achieved to date is 99.61% (Ranzato et al., 2006). The approach used to achieve that result extends the original training data set by elastically distorting the training images.

Notable is the fact that we achieve the above result without any optimization regarding the COSFIRE filters and the parameters used. Moreover, we do not perform any pre-processing and/or post-processing operations, neither do we use an extended training data set with elastic distortion. The fact that we achieve a result that is very close to the best result ever achieved is remarkable because our method is versatile and has not been developed with the specific handwritten digit recognition application in mind, while the best methods are results of long-lasting research effort in which elaborate application-specific techniques have been developed.

3.3.2 Detection and Recognition of Traffic Signs

The detection and recognition of specific objects in complex scenes is one of the most challenging tasks in computer vision. Here, we show how the proposed COSFIRE

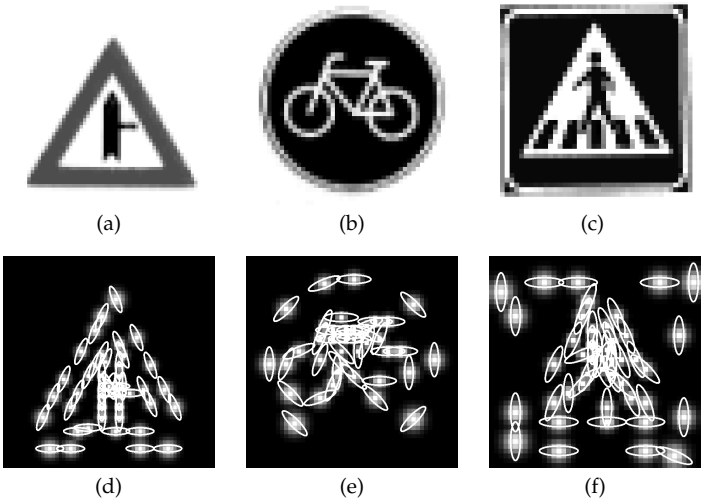


Figure 3.12: Three reference traffic signs: (a) an intersection, (b) compulsory giveway for bikes and (c) a pedestrian crossing. (d-f) The structures of the corresponding COSFIRE filters determined by the following parameter values: $\rho \in \{0, 2, 4, 7, 10, 13, 16, 20, 25\}$, $\sigma_0 = 0.67$, $\alpha = 0.04$, $\lambda = 4$, and $\theta \in \{\frac{\pi i}{8} \mid i = 1 \dots 15\}$.

filters can be used for the detection of traffic signs in complex scenes.

We use a public data set[7] of 48 color images (of size 360×270 pixels) that was originally published in (Grigorescu et al., 2003a). Each of these images contains (at least) one of three possible traffic signs illustrated in Fig. 3.12(a-c).

For this application we configure filters for patterns that are more complex than the ones involved in the previous two applications. We configure one COSFIRE filter for each of the three traffic signs, Fig. 3.12(d-f). We use a bank of antisymmetric Gabor filters with one wavelength ($\lambda = 4$) and 16 equidistant orientations ($\theta \in \{\frac{\pi i}{8} \mid i = 0 \dots 15\}$), and threshold their responses with $t_1 = 0.1$. The reference traffic signs that are used to configure the filters and the signs embedded in the complex scenes have approximately the same viewpoint and their sizes differ only by at most 10%. For such rigid objects, it is more appropriate to configure COSFIRE filters that achieve high selectivity. With this requirement in mind, we choose to configure the filters with a large number of ρ values ($\rho \in \{0, 2, 4, 7, 10, 13, 16, 20, 25\}$) and a small value of the parameter α ($\alpha = 0.04$) that allows little tolerance in the position of the involved edges.

[7] Traffic sign data set is online:
http://www.cs.rug.nl/~imaging/databases/traffic_sign_database/traffic_sign_database.html

We use the three COSFIRE filters to detect and recognize the corresponding traffic signs in the entire data set of 48 images. For each color image, we first convert it to grayscale and subsequently apply the filters. The antisymmetric Gabor filters that we use to provide inputs to the COSFIRE filters are applied with isotropic surround suppression (Grigorescu et al., 2004) (using an inhibition factor of 2) in order to reduce responses to the presence of texture in these complex scenes. Rather than using the parameter t_3 to threshold the filter responses at a given fraction of the maximum filter response, we choose to threshold the responses at a given absolute value. Moreover, we also threshold responses that are smaller than a fraction of the maximum value of all the responses produced by the three filters. We call this threshold validity ratio. For an absolute threshold of 0.04 and a validity ratio of 0.5 we obtain perfect detection and recognition performance for all the 48 traffic scenes. This means that we detect all the traffic signs in the given images with no false positives and correctly recognize every detected sign. Fig. 3.13 illustrates the detection and recognition of two different traffic signs, shown encircled[4], in one of the input images. For this application, we apply the COSFIRE filters in a non-invariant mode ($\psi = 0$, $\upsilon = 1$) and compute their output by a weighted geometric mean of the concerned Gabor filter responses ($\sigma' = 21.23$).

3.4 Discussion

When presenting the method in Section 3.2, we indicated that a prototype feature used for the configuration of a COSFIRE filter is selected by a user. The detection recognition of traffic signs presented in Section 3.3.2 is an example of such an application. The method is, however, not restricted by this aspect: there exists the possibility that a system 'discovers' patterns to be used for configuration and Section 3.3.1 provides an example of such an application.

We use Gabor filters for the detection of lines and edges. Gabor filters, however, are not intrinsic to the proposed method, and other orientation-selective filters can also be used.

The configuration of a COSFIRE filter is based on the spatial arrangement of contour parts that lie along concentric circles of given radii around a specified point of interest. In the first two applications that we present we choose to configure COSFIRE filters with three values of the radius parameter ρ as they provide sufficient coverage of the corresponding features. However, for the third application we use nine values of the parameter ρ in order to configure COSFIRE filters that are selective for more complex patterns. The choice of the number of ρ values is related to the size and complexity of the local prototype pattern that is used to configure a

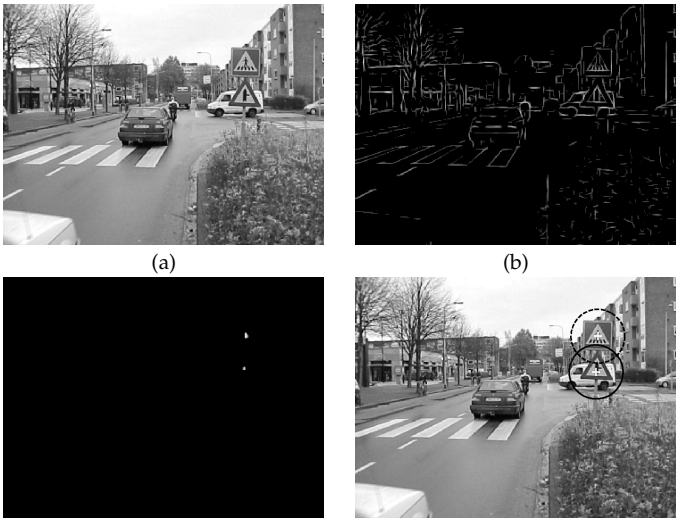


Figure 3.13: (a) Input image (crossing_004.png). (b) Superposition of thresholded responses ($t_1 = 0.1$) of a bank of antisymmetric Gabor filters ($\lambda = 4$ and $\theta \in \{\frac{\pi i}{8} \mid i = 0 \dots 15\}$) with isotropic surround suppression (inhibition factor is 2). (c) Superposition of the thresholded responses of the three COSFIRE filters. (d) Correct detection and recognition of two traffic signs. The cross markers indicate the locations of the two local maxima responses each surrounded with a circle that represents the support of the corresponding COSFIRE filter (continuous circle represents the intersection sign and the dashed circle represents the pedestrian crossing sign).

filter. The number of ρ values used also controls the tradeoff between the selectivity and generalization ability of a filter: a COSFIRE filter becomes more selective and more discriminative with an increasing number of ρ values.

A COSFIRE filter uses three threshold parameters: t_1, t_2 and t_3. The value of parameter t_1 depends on the contrast of the image material involved in a given application and the presence of noise. It controls the level at which the response of a Gabor filter is supposed to indicate the presence of a line or an edge at a given position. For the first application, which concerns binary input images, we achieved good results for $t_1 = 0.2$. Yet, for the second and third applications, which use grayscale input images, we obtained the best results for $t_1 = 0$ and $t_1 = 0.1$, respectively. The threshold parameter t_2, which is used only in the configuration phase, is application-independent. It implements a condition that the selected responses are significant and comparable with the strongest possible response. We fix the value of this threshold to ($t_2 =$) 0.75. The parameter t_3 is optional. It may be used to suppress

the responses of the COSFIRE filter that are below a given fraction of the maximum response value across all locations of the input image. For instance, in the first application we evaluate the performance of the COSFIRE filters with different values of the parameter t_3, while for the second application we do not threshold ($t_3 = 0$) the responses. In the third application we threshold the responses at a given absolute value rather than use this threshold parameter.

The proposed COSFIRE filters can be applied in various modes. In the next chapter we demonstrate that rotation-, scale-, and reflection-invariant COSFIRE filters are effective for the detection of vascular bifurcations in retinal images. For the recognition of handwritten digits we only made use of partial rotation-invariance and for the detection and recognition of traffic signs in complex scenes we used non-invariant COSFIRE filters.

In the following we highlight three main aspects in which the proposed COSFIRE filters can be distinguished from other keypoint detectors. First, a COSFIRE filter gives a response only when all parts of the filter-defining prototype feature are present. In contrast, dissimilarity based approaches give response also to parts of the prototype pattern. Second, while a COSFIRE filter combines the responses of Gabor filters at different scales, typical scale-invariant approaches, such as SIFT, use the same scale, the one at which the concerned keypoint is an extremum in a given scale space. Third, the area of support of a COSFIRE filter is adaptive. It is composed of the support of a number of orientation-selective filters whose geometrical arrangement around a point of interest is learned from a given local contour prototype pattern. On the contrary, the area of support of other operators is typically related to the appropriate scale rather than to the shape properties of the concerned pattern. To the best of our knowledge the proposed filters are the first ones which combine the responses of orientation-selective filters with their main area of support outside the point of interest. The presence of added noise around a pattern of interest has little or no affect on a COSFIRE filter response. For other operators, any added noise in the surroundings of a pattern of interest results in a descriptor that may differ substantially from the descriptor of the same but noiseless pattern.

The computational cost of the configuration of a COSFIRE filter is proportional to the maximum value of the given set of ρ values and to the size of the bank of Gabor filters used. In practice, for the parameter values that we used in the three applications, a COSFIRE filter is configured in less than half of a second for a Matlab implementation that runs on a 3GHz processor. The computational cost of the application of a COSFIRE filter depends on the computations of the responses of a bank of Gabor filters and their blurring and shifting. In practice, in the first application, a handwritten digit of size 28×28 pixels is described by 5000 rotation-non-invariant

COSFIRE filters in less than 10 seconds on a computer cluster[8]. In the second application, a complex scene of size 360×270 pixels is processed in less than 10 seconds on a standard 3GHz processor by three non-invariant COSFIRE filters. For this application we achieve the same performance as reported in (Grigorescu et al., 2003a) but with a much lower computational cost. We used Matlab implementation[9] for all the experiments.

The application of the proposed method to the recognition of handwritten digits contains an interesting aspect from a machine learning point of view. In traditional machine learning, the features to be used are fixed in advance and the machine learning aspect concerns the classification of observed feature vectors. If traditional machine learning is concerned with features at all, this is typically limited to the selection of predefined features or using them to derive 'new' features as (linear) combinations of the original ones. Examples are principle component analysis and generalized matrix learning vector quantization (Bunte et al., 2011). Traditional machine learning is typically not concerned with the question of how the original features are defined. This aspect of the problem is, however, crucial for the success: almost any machine learning method will perform well with good features. The interesting aspect we would like to point out is that in the proposed approach the appropriate prototype features are learned in the filter configuration process when a feature of interest is presented.

In our experiments, we do not analyse the discriminative ability of the individual COSFIRE filters because in this work we are not concerned with the optimization of the filters, but rather to show their versatility. As a consequence some of the configured filters that we used for the handwritten digit application might be redundant due to being selective for correlated patterns or for patterns with low distinctiveness. One way of dealing with such redundancy is to compute a dissimilarity measure between the prototype patterns used for the configuration of different COSFIRE filters. Moreover, a prototype feature selection method may also be incorporated in a machine learning algorithm, such as relevance learning vector quantization (Hammer and Villmann, 2002) or a support feature machine (Klement and Martinetz, 2010), to identify the most relevant COSFIRE filters.

The COSFIRE filters that we propose are inspired by the properties of one class

[8]We executed the experiments for the MNIST data set on a computer cluster of 255 multi-core nodes (`http://www.rug.nl/cit/hpcv/faciliteiten/HPCCluster/`). We split the MNIST data set of 70000 images (60000 training and 10000 test digits) into 250 batches of 280 images each, and processed the 250 batches in parallel. In this way, the digit descriptors of one experiment using 5000 rotation-non-invariant COSFIRE filters takes approximately (9.5 seconds $\times$280 images =) 45 minutes. An experiment with 5000 partial rotation-invariant COSFIRE filters (five values of the parameter ψ) takes five times as much.

[9]Matlab scripts for the configuration and application of COSFIRE filters can be downloaded from `http://matlabserver.cs.rug.nl/`.

of shape-selective neuron in area V4 of visual cortex (Pasupathy and Connor, 1999, 2001, 2002; Hegde and Van Essen, 2007). The selectivity that is exhibited by a COSFIRE filter, which we configured in Section 3.2, to a data set of elementary features (Fig. 3.6) is qualitatively similar to the selectivity of some V4 neurons studied in (Pasupathy and Connor, 1999). The way we determine the standard deviation of the blurring function in Eq. 3.1 is also motivated by neurophysiological evidence that the average diameter of receptive fields[10] of V4 neurons increases with the eccentricity (Gattass et al., 1988). Since there is a considerable spread in the behavior across neurons of the concerned type, different computational models may be needed to adequately cover the diversity of functional properties in that empirical space. In this respect, the proposed COSFIRE filter can be considered as a computational model of shape-selective V4 neurons that is complementary to other models (Riesenhuber and Poggio, 1999; Serre et al., 2005, 2007; Cadieu et al., 2007; Fidler and Leonardis, 2007).

The specific type of function that we use to combine the responses of afferent (Gabor) filters for the considered applications is a weighted geometric mean. This output function proved to give better results than various forms of addition. Furthermore, there is psychophysical evidence that human visual processing of shape is likely performed by multiplication (Gheorghiu and Kingdom, 2009). In future work, we plan to experiment with functions other than (weighted) geometric mean.

The proposed COSFIRE filters are particularly useful due to their versatility and selectiveness, in that a COSFIRE filter can be configured for any given local feature and is built on top of other - here orientation-selective - simpler filters. Elsewhere, we have used other type of simple filters (mexican hat operators) to build a contour operator, which we call CORF (Combination of Receptive Fields) (Azzopardi and Petkov, 2012). We use the terms COSFIRE and CORF for the same design principle in an engineering and neuroscience context, respectively.

There are various directions for future research. One direction is to apply the proposed trainable COSFIRE filters in other computer vision tasks, such as geometric stereo calibration, image retrieval, the recognition of handwritten characters, architectural symbols, and pedestrians. Another direction is to enrich the properties of a COSFIRE filter by including information about the color and texture distribution in a given local prototype pattern. A third direction is to extend the proposed approach to 3D COSFIRE filters that can be applied, for instance, to tubular organ registration and bifurcation detection in X-ray computed tomography medical images or to video sequences.

[10] In neurophysiology a receptive field refers to an area in the visual field which provides input to a given neuron. Its mathematical counterpart is the support of an operator.

3.5 Conclusions

We demonstrated that the proposed COSFIRE filters provide effective machine vision solutions in two practical applications: the recognition of handwritten digits (99.48% correct classification) and the detection and recognition of traffic signs in complex scenes (100% recall and precision). For the first application, the proposed COSFIRE filters achieve a performance that is close to the performance of the best application-specific method. For the second, it gives the same performance as another method which has much higher computational complexity.

The novel COSFIRE filters are conceptually simple and easy to implement: the filter output is computed as the product of blurred and shifted Gabor filter responses. They are versatile detectors of contour related features as they can be trained with any given local contour pattern and are subsequently able to detect identical and similar patterns. The COSFIRE approach is not limited to the combination of Gabor filter responses: more generally, it can be applied to the responses of filters that provide information about texture, color, contours and motion.

Published as:

G. Azzopardi, N. Petkov, *"Automatic detection of vascular bifurcations in segmented retinal images using trainable COSFIRE filters,"* Pattern Recognition Letters, in press, doi: http://dx.doi.org/10.1016/j.bbr.2011.03.031, 2012

Chapter 4

Detection of Vascular Features in Retinal Images using COSFIRE filters

Abstract

The vascular tree observed in a retinal fundus image can provide clues for cardiovascular diseases. Its analysis requires the identification of vessel bifurcations and crossovers. We use a set of trainable keypoint detectors that we call Combination Of Shifted FIlter REsponses or COSFIRE filters to automatically detect vascular bifurcations in segmented retinal images. We configure a set of COSFIRE filters that are selective for a number of prototype bifurcations and demonstrate that such filters can be effectively used to detect bifurcations that are similar to the prototypical ones. The automatic configuration of such a filter selects given channels of a bank of Gabor filters and determines certain blur and shift parameters. The response of a COSFIRE filter is computed as the weighted geometric mean of the blurred and shifted responses of the selected Gabor filters. The COSFIRE approach is inspired by the function of a specific type of shape-selective neuron in area V4 of visual cortex. We ran experiments on three data sets and achieved the following results: a) recall of 97.88% *at precision of* 96.94% *on 40 manually segmented images provided in the DRIVE data set, b) a recall of* 97.32% *at precision of* 96.04% *on 20 manually segmented images provided in the STARE data set, and c) a recall of* 97.02% *at precision of* 96.53% *on a set of 10 automatically segmented images obtained from images in the DRIVE data set. The COSFIRE filters that we use are conceptually simple and easy to implement: the filter output is computed as the weighted geometric mean of blurred and shifted Gabor filter responses. They are versatile keypoint detectors as they can be configured with any given local contour pattern and are subsequently able to detect the same and similar patterns.*

4.1 Introduction

Retinal fundus images provide a unique possibility to take a non-invasive look at the eye and the systemic status of the human body. Besides ocular diseases, such as age-related macular degeneration and glaucoma that are two of the leading causes of blindness, other systemic diseases are also manifested in the retina. Complications of such diseases include diabetic retinopathy from diabetes mellitus (Frank, 2004), hypertension (Tso and Jampol, 1982) and atherosclerosis from cardiovascular disease (Chapman et al., 2002), as well as brain diseases and neuropathies, such as multiple sclerosis (Parisi et al., 1999) and Huntington's disease (Paulus et al., 1993). The retina can thus be considered as a mirror of the health status of a person.

Here we focus on retinal image analysis for the diagnosis of cardiovascular diseases. For a thorough review on retinal image analysis we refer to Abramoff et al. (2010). The vascular geometrical structure in the retina is known to conform to structural principles that are related to certain physical properties (Zamir et al., 1979; Sherman, 1981). For instance, the studies of Murray (1926a,b) revealed that the most efficient blood circulation is achieved when the blood flow is proportional to the cubed power of the vessel's radius; this is known as Murray's law. The branching angle between the two child vessels is also important to optimize the efficiency of the entire vascular network (Zamir et al., 1992).

The analysis of the geometrical structure of the vessel tree is thus important as deviations from the optimal principles may indicate (increased risk of) signs for vascular pathology; a thorough review is given by Patton et al. (2006). The detection of junctions in the vessel tree of a retinal fundus image, commonly referred to as vascular bifurcations and crossovers, is one of the basic steps in this analysis, and it is typically carried out in a time-consuming manual procedure (Chapman et al., 2002). The automation of such a tedious process is thus important to improve the efficiency and to avoid inaccuracies due to human fatigue.

The existing attempts to automate the detection of retinal vascular bifurcations can be categorized into two classes; geometrical-feature based and model based approaches. The former methods are highly dependent on the segmentation and skeletonization techniques. They also involve extensive local pixel processing and branch point analysis (Martinez-Perez et al., 2002; Chanwimaluang and Guoliang, 2003; Jung and Hong, 2006; Bhuiyan et al., 2007; Ardizzone et al., 2008; Aibinu et al., 2010; Calvo et al., 2011). Incomplete bifurcations, which are commonly produced by automatic segmentation techniques, are generally not detected by such skeleton based approaches. On the other hand, model-based approaches are usually more adaptive and have smaller computational complexity that makes them more appropriate for real-time applications (Ali et al., 1999; Shen et al., 2001; Tsai et al., 2004).

However, model based approaches suffer from insufficient generalization ability as they are usually unable to model all the features of interest. Consequently, these methods may fail to detect atypical bifurcations.

Besides the diagnosis of pathologies, retinal fundus images have also been used for person verification as the geometrical arrangement of the vascular bifurcations is an effective biometric (Bevilacqua et al., 2005; Ortega et al., 2009). Moreover, vascular bifurcations may also be used as key features to find correspondences between the retinal images of the same eye taken from different views, i.e. for registration of retinal images (Becker et al., 1998; Shen et al., 2003; Tsai et al., 2004; Li Chen et al., 2011).

We use the COSFIRE (Combination Of Shifted FIlter REsponses) filters, which we have introduced elsewhere (Azzopardi and Petkov, 2013), for the detection of bifurcations in segmented retinal images. COSFIRE filters are trainable keypoint detection operators, which are selective for given local patterns that consist of combinations of contour segments. These operators are inspired by the properties of some neurons in area V4 of visual cortex, which are selective for parts of (curved) contours or for combinations of line segments (Pasupathy and Connor, 1999, 2002). In this work we focus on one application and provide more elaborate experiments to demonstrate the robustness and generalization capability of the COSFIRE filters.

The response of a COSFIRE filter in a given point is computed as a function of the shifted responses of simpler filters. Using shifted responses of simpler filters – Gabor filters in the concerned application – corresponds to combining their respective supports at different locations to obtain a more sophisticated filter with a bigger support. The specific function that we use here to combine the responses of simpler filters is weighted geometric mean, essentially multiplication, which has specific advantages regarding shape recognition.

Two-dimensional (2D) Gabor filters (Daugman, 1985) that we use as input to our COSFIRE filters have been extensively used to detect oriented structures (lines and/or edges) in many computer vision applications, including retinal image analysis. For instance, these filters have been found effective in detecting signs of glaucoma (Bodis-Wollner and Brannan, 1997; Sun et al., 2006; Muramatsu et al., 2009) detecting the optic nerve head of the retina (Rangayyan et al., 2010), and mostly for the segmentation of the vessel tree in retinal fundus images (Soares et al., 2006; Qin Li et al., 2006; Usman Akram et al., 2009; Moin et al., 2010; Xiaojun Du and Bui, 2010; Yavuz and Koumlse, 2010; Fraz et al., 2011; Selvathi and Lalitha Vaishnavi, 2011; Yavuz and Kose, 2011). Apart from Gabor filters, other state-of-the-art methods have also been found effective for the segmentation of the vessel tree (Chauduri et al., 1989; Jiang and Mojon, 2003; Staal et al., 2004; Niemeijer et al., 2004; Mendonca and Campilho, 2006; Ricci and Perfetti, 2007).

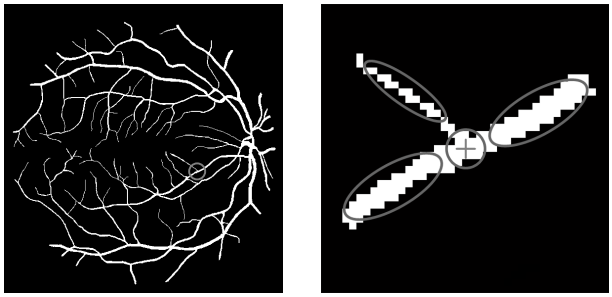
The rest of the chapter is organized as follows: In Section 4.2 we present our method and demonstrate how it can be used to detect retinal vessel features. In Section 4.3, we evaluate the effectiveness of the COSFIRE filters on manually and automatically segmented retinal images from the DRIVE and STARE data sets. In Section 4.4 we provide a discussion of some aspects of our approach and finally we draw conclusions in Section 4.5.

4.2 Method

A COSFIRE filter for the detection of local combinations of lines is conceptually simple and straightforward to implement: it requires the application of selected Gabor filters, blurring of their responses, shifting the blurred responses by specific, different vectors, and multiplying the shifted responses. The questions of which Gabor filters to use, how much to blur them and how far to shift them are answered in a filter configuration process in which a local pattern of interest that defines a keypoint is automatically analysed. The configured COSFIRE filter can then detect the same and similar patterns.

4.2.1 Overview

In Fig. 4.1a we illustrate a typical vascular bifurcation encircled in a segmented retinal fundus image[1]. We use this feature as a prototype bifurcation, which is shown enlarged in Fig. 4.1b, to automatically configure a COSFIRE filter that will respond to the same and similar bifurcations.



(a) (b)

Figure 4.1: (a) The circle indicates a bifurcation that is selected by a user. (b) Enlargement of the selected feature. The ellipses represent the dominant orientations around the specified point of interest that is indicated by a cross marker.

[1]The image used in this example is named 40_manual1.gif in the DRIVE data set (Staal et al., 2004).

The three ellipses shown in Fig. 4.1b represent the dominant orientations in the neighbourhood of the specified point of interest. We detect such orientations by Gabor filters. The central circle represents the overlapping support of a group of such filters. The response of a COSFIRE filter is computed by combining the responses of the concerned Gabor filters by a weighted geometric mean. The preferred orientations of these filters and the locations at which we take their responses are determined by automatically analysing the local prototype pattern used for the configuration of the concerned COSFIRE filter. Consequently, the COSFIRE filter is selective for the presented local spatial arrangement of lines of specific orientations and widths. Taking the responses of Gabor filters at different locations around a point can be implemented by shifting the responses of these Gabor filters by different vectors before using them for the pixel-wise evaluation of a function which gives the COSFIRE filter output.

Such a design is inspired by electrophysiological evidence that some neurons in area V4 of visual cortex are selective for moderately complex stimuli, such as curvatures, that receive inputs from a group of orientation-selective cells in areas V1 and V2 (Pasupathy and Connor, 1999, 2001, 2002). Moreover, in a psychophysical experiment, Gheorghiu and Kingdom (2009) show that curved contour parts are likely detected by a nonlinear operation that combines the responses of afferent orientation-selective filters by multiplication. Since a COSFIRE filter makes use of such multiplication, it produces a response only when all its afferent inputs from Gabor filters are stimulated; i.e. all constituent parts (in this case lines) of a vascular bifurcation are present.

In the following sections we explain the automatic configuration process of a COSFIRE filter that will be selective for the prototype bifurcation shown in Fig. 4.1b. The configuration process determines which responses of which Gabor filters in which locations need to be combined in order to obtain the output of the filter.

4.2.2 Detection of Dominant Orientations by 2D Gabor Filters

We build the COSFIRE filter using as input the responses of 2D Gabor filters, which are known to serve as line and edge detectors.

We denote by $h_{\lambda,\theta}(x,y)$ a Gabor function of preferred wavelength λ and orientation θ:

$$h_{\lambda,\theta}(x,y) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda}+\zeta\right) \tag{4.1}$$
$$x' = x\cos\theta + y\sin\theta,\;\; y' = -x\sin\theta + y\cos\theta$$

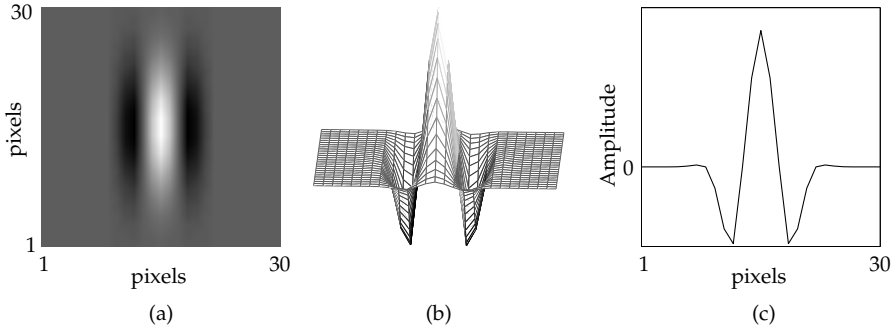


Figure 4.2: A 2D Gabor function rendered as (a) a 2D intensity map and (b) a 3D surface, with the following parameter values: $\theta = 0$, $\lambda = 8$, $\gamma = 0.5$, $\sigma = 0.31\lambda$ and $\zeta = 0$. (c) 1D profile of the Gabor function along a horizontal cross section. The corresponding filter is selective for lines with preferred vertical orientation and a preferred width of $\lambda/2 = 4$ pixels.

where γ represents the aspect ratio that specifies the ellipticity of the support of the Gabor function, σ is the standard deviation of the Gaussian envelope and ζ is the phase offset of the sinusoidal wave function. Here, we set $\gamma = 0.5$, as suggested by Petkov (1995), and we set the other parameters to $\sigma = 0.31\lambda$ and $\zeta = 0$.

We denote by $g_{\lambda,\theta}(x, y)$ the response[2] of a symmetric Gabor filter to a given grayscale image I:

$$g_{\lambda,\theta}(x, y) = I * h_{\lambda,\theta}(x, y) \tag{4.2}$$

The responses of a symmetrical ($\zeta \in \{0, \pi\}$) and an antisymmetrical ($\zeta \in \{\frac{\pi}{2}, \frac{3\pi}{2}\}$) Gabor filter can be combined in a quadrature filter, commonly referred to as Gabor energy filter. Moreover, surround suppression can be applied as a post-processing step to Gabor (energy) filter responses to reduce responses to texture and improve the detection of object contours. For brevity of presentation we do not consider all these aspects of Gabor filters here and we refer to Petkov and Kruizinga (1997); Kruizinga and Petkov (1999); Grigorescu et al. (2002); Petkov and Westenberg (2003); Grigorescu et al. (2003c,b) for technical details and to our online implementation[3].

We re-normalize all Gabor functions that we use in such a way that all positive values of such a function sum up to 1 while all negative values sum up to -1. Among other things, this ensures that the response to a line of width w will be largest for a symmetrical filter of preferred wavelength $\lambda = 2w$. It also ensures that the response to an image of constant intensity is 0. Without re-normalization, this is true only for antisymmetrical filters. For the specific application at hand, we use symmetrical

[2]We use * to denote convolution.

[3]`http://matlabserver.cs.rug.nl`

Gabor functions ($\zeta = 0$) as they respond best to line structures which makes them appropriate to detect the presence of vessels in retinal fundus images.

Fig. 4.2 illustrates the area of support of a symmetrical Gabor filter, that is selective for elongated line structures with a preferred vertical orientation and a preferred width of four pixels.

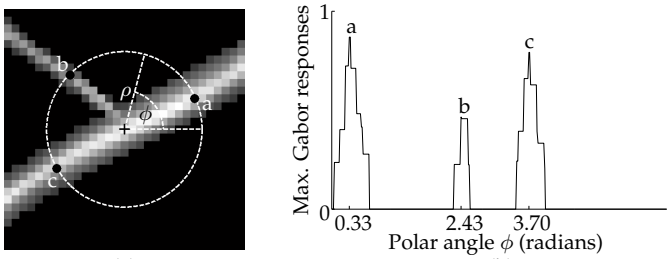
In our experiments, we apply a bank of Gabor filters with six wavelengths equidistantly spaced on a logarithmic scale ($\lambda \in \{2(2^{\frac{i}{2}}) \mid i = 0 \dots 5\}$) and eight equidistant orientations ($\theta = \{\frac{\pi}{8} i \mid i = 0 \dots 7\}$) on segmented retinal fundus images of size 565×584 pixels and of size 605×700 pixels. In such images, the blood vessels have widths that vary from 1 to 7 pixels.

We threshold the responses of Gabor filters at a given fraction t_1 ($0 \leq t_1 \leq 1$) of the maximum response of $g_{\lambda,\theta}(x, y)$ across all combinations of values (λ, θ) used and all positions (x, y) in the given image, and denote these thresholded responses by $|g_{\lambda,\theta}(x, y)|_{t_1}$. The choice of the threshold value depends on the contrast of the input images. For the binary segmented images that we use, a threshold value of $t_1 = 0.2$ proves sufficient to eliminate responses achieved on the background while preserving responses to features of interest. Fig. 4.3a illustrates the maximum value superposition of the thresholded responses of the concerned bank of Gabor filters obtained for the vascular bifurcation shown in Fig. 4.1b.

4.2.3 Configuration of a COSFIRE Filter

A COSFIRE filter uses as input the responses of a number of Gabor filters, each characterized by a pair of parameter values (λ_i, θ_i), around certain positions (ρ_i, ϕ_i) with respect to the center of the COSFIRE filter. A set of four parameter values $(\lambda_i, \theta_i, \rho_i, \phi_i)$ characterizes the properties of a contour part that is present in the specified area of interest ($\lambda_i/2$ represents the width, θ_i represents the orientation and (ρ_i, ϕ_i) represents the location in polar coordinates). In the following we explain how we obtain the parameter values of such contour parts around a given point of interest.

We consider the responses of the bank of Gabor filters along a circle of a given radius ρ around a point of interest (x, y) that is specified by a user, Fig. 4.3. In each position along that circle, we take the maximum of all responses across the possible values of (λ, θ). The positions that have values greater than the corresponding values of the neighbouring positions along an arc of angle $\pi/8$ are chosen as the points that characterize the dominant orientations around the point of interest. We determine the polar coordinates (ρ_j, ϕ_j) for such point with respect to the center of the filter. For each such location (ρ_j, ϕ_j), we then consider all combinations of (λ, θ) for which the corresponding thresholded Gabor responses $|g_{\lambda,\theta}(x + \rho_j \cos\phi_j, y + \rho_j \sin\phi_j)|_{t_1}$



(a) (b)

Figure 4.3: (a) The gray-level of a pixel is the maximum value superposition of the thresholded responses ($t_1 = 0.2$) of a bank of Gabor filters at that position, $\max_{\lambda,\theta} |g_{\lambda,\theta}(x,y)|_{t_1}$. The cross marker indicates the location of the point of interest selected by a user, while the bright circle of a given radius ρ (here $\rho = 10$) indicates the locations considered in the analysis phase. (b) Values of the maximum value superposition of Gabor filter responses along the concerned circle. The labeled black dots in (a) mark the positions (relative to the center of the filter) at which the respective strongest Gabor filter responses are taken. These three positions correspond to the three local maxima in the plot in (b).

are greater than a fraction $t_2 = 0.75$ of the maximum of $|g_{\lambda,\theta}(x',y')|_{t_1}$ across the different combinations of values (λ,θ) used and across all locations in the given image. The threshold parameter t_2 implements a condition that the selected Gabor responses are significant and comparable to the strongest possible response. For each value θ that satisfies this condition, we consider a single value of λ, the one for which $|g_{\lambda,\theta}(x+\rho_j\cos\phi_j, y+\rho_j\sin\phi_j)|_{t_1}$ is the maximum of all responses across all values of λ. For each distinct pair of (λ,θ) and location (ρ_j,ϕ_j), we obtain a tuple $(\lambda_i,\theta_i,\rho_i,\phi_i)$. Multiple tuples can thus be formed for the same location (ρ_j,ϕ_j).

We denote by $S_f = \{(\lambda_i,\theta_i,\rho_i,\phi_i) \mid i = 1\ldots n_f\}$ the set of parameter value combinations, where n_f represents the number of contour parts which fulfill the above conditions. The subscript f stands for the prototype bifurcation. Every tuple in the set S_f specifies the parameters of some contour part in f.

For the point of interest shown in Fig. 4.3a and two given values of the radius ρ ($\rho \in \{0, 10\}$), the selection method described above results in the following set:

$$S_f = \left\{ \begin{array}{l} (\lambda_1 = 8\sqrt{2},\ \theta_1 = 5\pi/8,\ \rho_1 = 0,\ \ \phi_1 = 0), \\ (\lambda_2 = 4\sqrt{2},\ \theta_2 = 5\pi/8,\ \rho_2 = 10,\ \phi_2 = 0.3316), \\ (\lambda_3 = 4,\ \ \ \theta_3 = \pi/4,\ \ \rho_3 = 10,\ \phi_3 = 2.4260), \\ (\lambda_4 = 4,\ \ \ \theta_4 = 3\pi/8,\ \rho_4 = 10,\ \phi_4 = 2.4260), \\ (\lambda_5 = 4\sqrt{2},\ \theta_5 = 5\pi/8,\ \rho_5 = 10,\ \phi_5 = 3.7001), \\ (\lambda_6 = 4\sqrt{2},\ \theta_6 = 3\pi/4,\ \rho_6 = 10,\ \phi_6 = 3.7001) \end{array} \right\}$$

The second tuple in the above set S_f, ($\lambda_2 = 4\sqrt{2},\ \theta_2 = 5\pi/8,\ \rho_2 = 10,\ \phi_2 = 0.3316$), for instance, describes a contour part with a width of $\lambda_2/2 = 2.83$ pixels and an orientation of $\theta_2 = 5\pi/8$ that can be detected by a Gabor filter with preferred wavelength of $\lambda_2 = 4\sqrt{2}$ pixels and preferred orientation of $\theta_2 = 5\pi/8$, at a position of $\rho_2 = 10$ pixels to the north east, $\phi_2 = 0.3316$, of the point of interest; this location is marked by the label 'a' in Fig. 4.3. This selection is the result of the presence of a diagonally oriented blood vessel to the north east from the center of the prototype bifurcation that is used for the configuration of the COSFIRE filter.

4.2.4 Blurring and Shifting Gabor Filter Responses

The above analysis of the given prototype bifurcation f indicates that this pattern produces six strong responses $g_{\lambda_i,\theta_i}(x,y)$ of Gabor filters with parameters (λ_i, θ_i) in the corresponding positions with polar coordinates (ρ_i, ϕ_i) with respect to the center of the COSFIRE filter. Next, we use these responses to compute the output of the COSFIRE filter. Since the concerned responses are in different positions (ρ_i, ϕ_i) with respect to the filter center, we first shift them appropriately so that they come together in the filter center. The COSFIRE filter output can then be evaluated as a pixel-wise multivariate function of the shifted Gabor filter responses.

Before these shift operations, we blur the Gabor filter responses in order to allow for some tolerance in the position of the respective contour parts. We define the blurring operation as the computation of maximum value of the weighted responses of a Gabor filter. For weighting we use a Gaussian function $G_{\hat{\sigma}}(x,y)$, the standard deviation $\hat{\sigma}$ of which is a linear function of the distance ρ from the center of the COSFIRE filter: $\hat{\sigma} = \hat{\sigma}_0 + \alpha\rho$. Here we use $\hat{\sigma}_0 = 0.67$ and $\alpha = 0.1$. The choice of the linear function that we use to determine the standard deviation $\hat{\sigma}$ of the blurring function is explained in Section 4.4.

Next, we shift the blurred responses of each selected Gabor filter (λ_i, θ_i) by a distance ρ_i in the direction opposite to ϕ_i. In polar coordinates, the shift vector is specified by $(\rho_i, \phi_i + \pi)$, while in cartesian coordinates it is $(\Delta x_i, \Delta y_i)$ where $x_i = -\rho_i \cos\phi_i$ and $y_i = -\rho_i \sin\phi_i$. We denote by $s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y)$ the blurred and shifted thresholded response of the Gabor filter that is specified by the i-th tuple $(\lambda_i, \theta_i, \rho_i, \phi_i)$ in the set S_f:

$$s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y) \stackrel{\text{def}}{=} \max_{x',y'}\{|g_{\lambda_i,\theta_i}(x - x' - \Delta x_i, y - y' - \Delta y_i)|_{t_1} G_{\hat{\sigma}}(x',y')\} \qquad (4.3)$$

where $-3\hat{\sigma} \leq x', y' \leq 3\hat{\sigma}$.

Fig. 4.4 illustrates the blurring and shifting operations for this COSFIRE filter, for

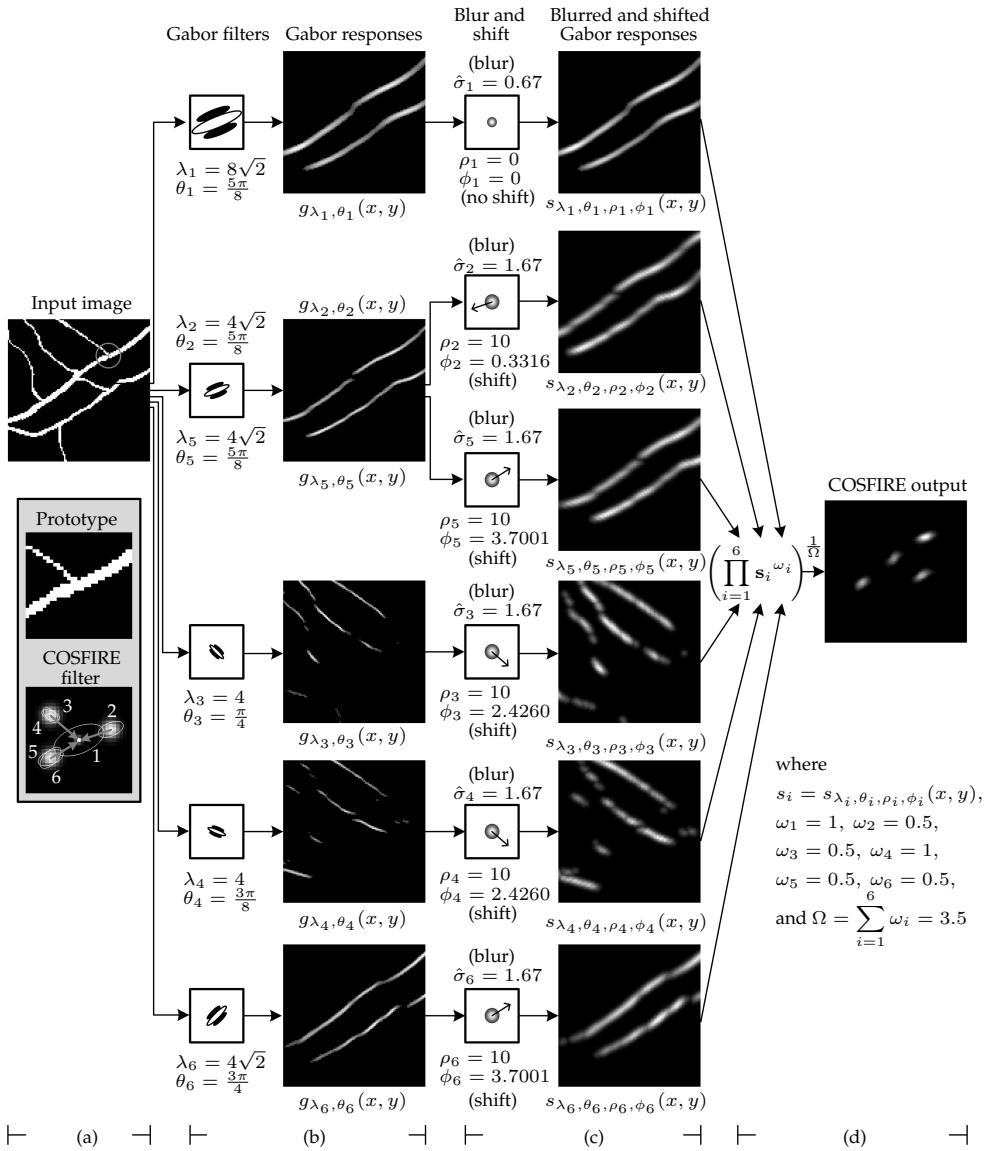


Figure 4.4: (a) Part of the segmented retinal image shown in Fig. 4.1a. The enframed inlay images show (top) the enlarged bifurcation encircled in the given image and (bottom) the structure of the COSFIRE filter that is configured by this prototype bifurcation. This COSFIRE filter is trained to detect the local spatial arrangement of six contour parts. The ellipses illustrate the orientations and wavelengths of the selected Gabor filters, and the bright blobs are intensity maps of 2D Gaussian functions that are used to blur the responses of the corresponding Gabor filters. The blurred responses are then shifted by the corresponding vectors. (b) Each contour part of the prototype pattern is detected by a Gabor filter with a given preferred wavelength λ_i and orientation θ_i. Two of these parts ($i = \{2, 5\}$) are detected by the same Gabor filter. (c) We then blur the thresholded response (here at $t_1 = 0.2$) of each concerned Gabor filter $|g_{\lambda_i,\theta_i}(x, y)|_{t_1}$ and subsequently shift the resulting blurred image by a polar coordinate vector $(\rho_i, \phi_i + \pi)$. (d) Finally, we obtain the output of the COSFIRE filter (here $t_3 = 0$) by computing the weighted geometric mean of all the blurred and shifted Gabor filter responses $s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x, y)$. This COSFIRE filter gives local maximum responses in four points; the point with the strongest response corresponds to the prototype bifurcation and the other three points correspond to three bifurcations that are similar to the prototype bifurcation.

an input pattern that is cropped from the central region of the image in Fig. 4.1a. For each tuple in the configured set S_f we first compute the thresholded responses of the corresponding Gabor filters and then we blur and shift these responses accordingly.

4.2.5 COSFIRE Filter Response

We define the response $r_{S_f}(x,y)$ of a COSFIRE filter as the weighted geometric mean of all the blurred and shifted thresholded responses of Gabor filters $s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y)$ that correspond to the properties of the contour parts described by set S_f:

$$r_{S_f}(x,y) \stackrel{\text{def}}{=} \left| \left(\prod_{i=1}^{|S_f|} \left(s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y)\right)^{\omega_i} \right)^{1/\sum_{i=1}^{|S_f|}\omega_i} \right|_{t_3} \tag{4.4}$$

$$\omega_i = \exp\left(\frac{-\rho_i^2}{2\acute{\sigma}^2}\right),\ \acute{\sigma} = \sqrt{\frac{-\rho_{\max}^2}{2\ln 0.5}},\ \rho_{\max} = \max_{i \in \{1\ldots|S_f|\}}\{\rho_i\}$$

where $|.|_{t_3}$ stands for thresholding the response at a fraction t_3 ($0 \leq t_3 \leq 1$) of its maximum across all image coordinates (x,y). The input contribution of $s-$quantities decreases with an increasing value of the corresponding parameter ρ. Here, we use a value of the standard deviation $\acute{\sigma}$ that is computed as a function of the maximum value $\rho_{\max}$ of the given set of ρ values. We make this choice in order to achieve a maximum value $\omega = 1$ of the weights in the center (for $\rho = 0$), and a minimum value $\omega = 0.5$ in the periphery (for $\rho = \rho_{\max}$).

Fig. 4.4 shows the output of a COSFIRE filter, which is defined as the weighted geometric mean of six blurred and shifted images from the thresholded responses of five Gabor filters. Here, the maximum response is reached in the center of the prototype bifurcation that was used to configure this COSFIRE filter and the other three local maxima points correspond to bifurcations that are similar to the prototype bifurcation.

4.2.6 Tolerance to Rotation, Scale and Reflection

In the following we provide details on how we achieve tolerance to rotation, scale and reflection.

We achieve rotation and scale tolerance by manipulating the set of parameter values in S_f, rather than by computing them from the responses to rotated and resized versions of the prototype pattern. Using the set S_f that defines the concerned filter, we form a new set $\Re_{\psi,\upsilon}(S_f)$ as follows:

$$\Re_{\psi,\upsilon}(S_f) \stackrel{\text{def}}{=} \{(\upsilon\lambda_i, \theta_i + \psi, \upsilon\rho_i, \phi_i + \psi) \mid (\lambda_i, \theta_i, \rho_i, \phi_i) \in S_f)\} \tag{4.5}$$

For each tuple $(\lambda_i, \theta_i, \rho_i, \phi_i)$ in the original filter S_f that describes a certain local contour part, we provide a counterpart tuple $(\upsilon\lambda_i, \theta_i + \psi, \upsilon\rho_i, \phi_i + \psi)$ in the new set $\Re_{\psi,\upsilon}(S_f)$. The preferred orientation of the concerned contour part and its polar angle position with respect to the center of the filter are offset by an angle ψ relative to the values of the corresponding parameters of the original tuple. Moreover, the width of the concerned contour part and its distance to the center of the filter are scaled by a factor υ relative to the values of the corresponding parameters of the original tuple.

A rotation- and scale-tolerant response is then achieved by taking the maximum value of the responses of filters that are obtained with different values of the parameters ψ and υ:

$$\hat{r}_{S_f}(x, y) \stackrel{\text{def}}{=} \max_{\psi\in\Psi, \upsilon\in\Upsilon} \{r_{\Re_{\psi,\upsilon}(S_f)}(x, y)\} \tag{4.6}$$

where Ψ is a set of n_ψ equidistant orientations defined as $\Psi = \{\frac{2\pi}{n_\psi} i \mid 0 \leq i < n_\psi\}$ and Υ is a set of υ scale factors equidistant on a logarithmic scale defined as $\Upsilon = \{2^{\frac{i}{2}} \mid i \in \mathbb{Z}\}$. According to Eq. 4.6 a COSFIRE filter will produce the same response for local patterns that are versions of each other, obtained by rotation at discrete angles $\psi \in \Psi$ and resizing by given scale factors $\upsilon \in \Upsilon$. As to the response of the filter to patterns that are rotated at angles of intermediate values between those in Ψ, it depends on the orientation selectivity of the filter that is influenced by the orientation bandwidth of the involved Gabor filters and by the value of the parameter α, mentioned above. For $\alpha = 0.1$ that we use, $n_\psi = 16$ equidistant preferred orientations ensure sufficient response for any orientation of the local prototype pattern that is used to configure the COSFIRE filter.

As to reflection tolerance, we first form a new set $\acute{S}_f$ from the set S_f as follows:

$$\acute{S}_f \stackrel{\text{def}}{=} \{(\lambda_i, \pi - \theta_i, \rho_i, \pi - \phi_i) \mid \forall\, (\lambda_i, \theta_i, \rho_i, \phi_i) \in S_f\} \tag{4.7}$$

The new filter, which is defined by the set $\acute{S}_f$, is selective for a reflected version of the prototype pattern f about the $y-$axis.

Finally, a combined rotation-, scale-, and reflection-tolerant response is computed as the maximum value of the rotation- and scale-tolerant responses of the filters S_f and $\acute{S}_f$ that are obtained with different values of the parameters ψ and υ:

$$\bar{r}_{S_f}(x, y) \stackrel{\text{def}}{=} \max\{\hat{r}_{S_f}(x, y), \hat{r}_{\acute{S}_f}(x, y)\} \tag{4.8}$$

4.3 Experimental Results

We use the DRIVE (Staal et al., 2004) and the STARE (Hoover et al., 2000) data sets of retinal fundus images to evaluate the performance of the COSFIRE filters by quantifying their effectiveness in detecting vessel features.

4.3.1 Data Sets and Ground Truth

The DRIVE data set, which was obtained from a screening programme of diabetic retinopathy in the Netherlands, comprises 40 colour images each of size 565×584 pixels. The STARE data set contains 20 retinal colour images each of size 605×700 pixels, ten of which are of patients with no pathology and the other ten contain pathology.

These data sets are mainly used for the evaluation of algorithms for segmentation of the vessel tree. For this reason, next to the original photographic images the data sets also include the corresponding manually segmented binary images. The ground truth[4] data, which comprises the coordinates of bifurcations and crossovers of the manually segmented retinal images[5], was defined by the authors of this study.

Fig. 4.5(a-b) show the grayscale version of the colour retinal image with filename 40_training.tif and its segmentation in blood vessels and background, respectively, which are both taken from the DRIVE data set (Staal et al., 2004). Fig. 4.5b contains 101 blood vessel features, shown encircled, which present Y- or T-form bifurcations or crossovers. Fig. 4.5d illustrates the superposition of the thresholded superimposed responses ($t_1 = 0.2$) of the bank of Gabor filters applied to the manually segmented image shown in Fig. 4.5b.

4.3.2 Configuration of COSFIRE Filters

We configure COSFIRE filters in an iterative mode. In the first iteration, we use the prototype bifurcation f_1 illustrated in Fig. 4.1 to configure a COSFIRE filter denoted by S_{f_1} with three values of the radius parameter ρ ($\rho \in \{0, 4, 10\}$). Fig. 4.6f illustrates the structure of the filter S_{f_1}.

We then apply filter S_{f_1} to the same training image and show the obtained results in Fig. 4.7. We set the value of its threshold parameter $t_3(S_{f_1})$ to a fraction that produces the largest number of correctly detected bifurcations and no false positives. For COSFIRE filter S_{f_1} this criterium is satisfied for $t_3(S_{f_1}) = 0.39$. The encircled re-

[4]The ground truth data can be downloaded from the following website:
`http://www.cs.rug.nl/~imaging/databases/retina_database`

[5]Named in DRIVE *_manual1.gif, and named in STARE *.ah.ppm

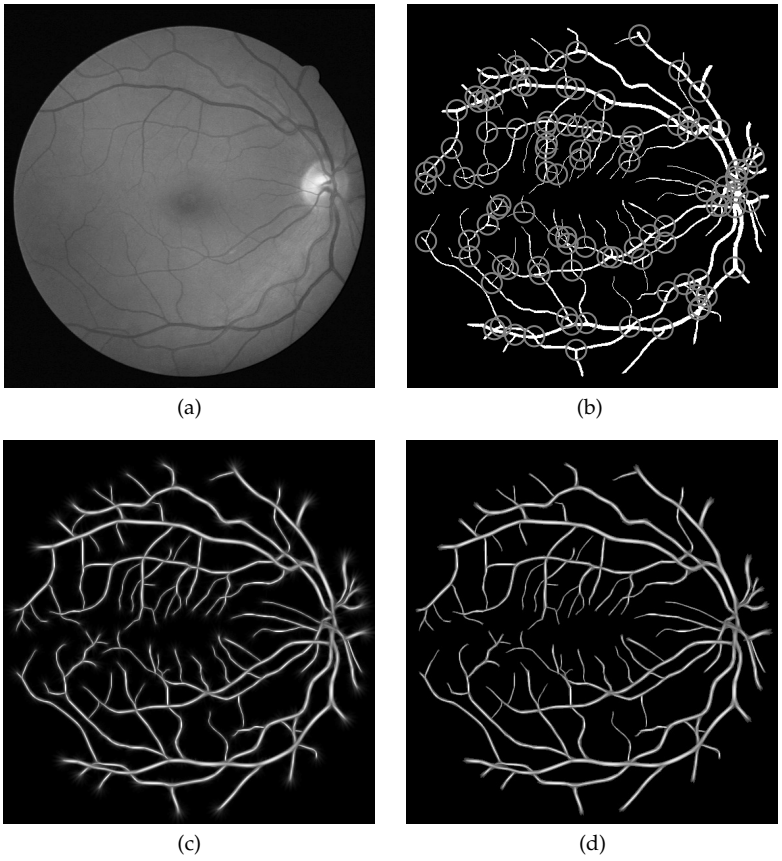
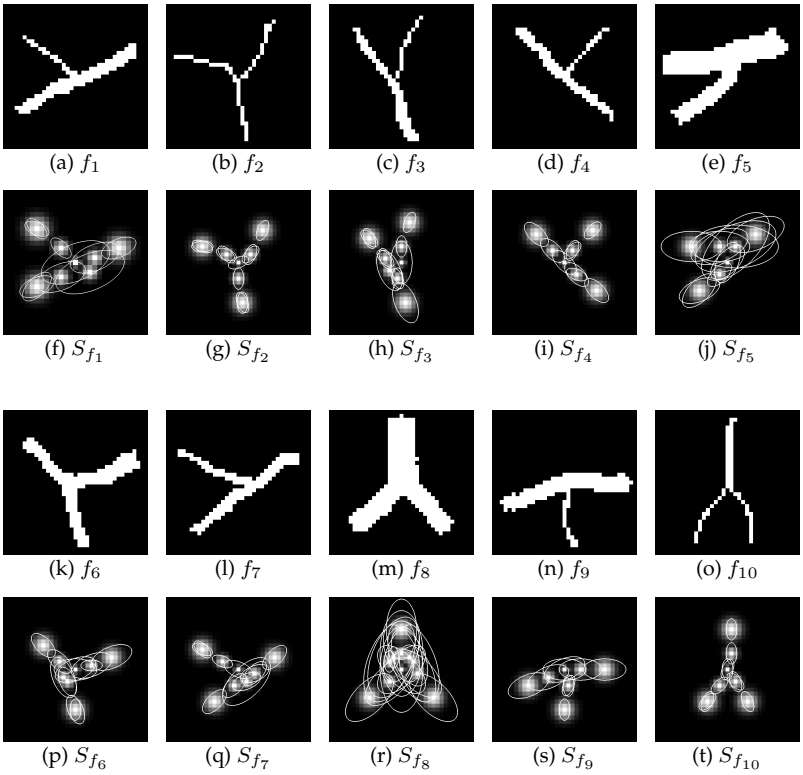


Figure 4.5: Example of a retinal fundus image from the DRIVE data set. (a) Original image (of size 565×584 pixels) with filename 40_training.tif. (b) The manual binary segmentation of vessels and background, also available in DRIVE. The typical widths of blood vessels vary between 1 and 7 pixels. This range of width values determines our choice of the set Λ of wavelength values used in the bank of Gabor filters. The circles surround Y- and T-formed vessel bifurcations and crossings. (c-d) Superposition of the responses of a bank of symmetric Gabor filters with a threshold (c) $t_1 = 0$ and (d) $t_1 = 0.2$.

gions[6] are centered on the local maxima of the filter response and if two such regions overlap by more than 75%, only the one with the stronger response is shown.

Without provisions for rotation and scale tolerance ($\Psi = \{0\},\ \Upsilon = \{0\}$) COSFIRE filter S_{f_1} detects five vascular bifurcations, one of which is the prototype bifurcation that was used to configure this filter, Fig. 4.7a. When rotation tolerance is introduced

[6]The radius of the circle is the sum of the maximum value of the radial parameter ρ and blur radius used at this value of ρ.



(a) f_1 (b) f_2 (c) f_3 (d) f_4 (e) f_5

(f) S_{f_1} (g) S_{f_2} (h) S_{f_3} (i) S_{f_4} (j) S_{f_5}

(k) f_6 (l) f_7 (m) f_8 (n) f_9 (o) f_{10}

(p) S_{f_6} (q) S_{f_7} (r) S_{f_8} (s) S_{f_9} (t) $S_{f_{10}}$

Figure 4.6: (a-e, k-o) A set of ten prototype bifurcations and (f-j, p-t) the structures of the corresponding ten COSFIRE filters. These bifurcations are all taken from the same retinal image with filename 40_training.gif. The following are the learned threshold values: $t_3(S_{f_1}) = 0.39,\ t_3(S_{f_2}) = 0.21,\ t_3(S_{f_3}) = 0.36,\ t_3(S_{f_4}) = 0.37,\ t_3(S_{f_5}) = 0.33,\ t_3(S_{f_6}) = 0.19,\ t_3(S_{f_7}) = 0.33,\ t_3(S_{f_8}) = 0.27,\ t_3(S_{f_9}) = 0.34$, and $t_3(S_{f_{10}}) = 0.32$.

($\Psi = \{\frac{\pi}{8}i \mid i = 0 \dots 15\},\ \Upsilon = \{0\}$), the number of correctly detected bifurcations is increased to 38, Fig. 4.7b. With the addition of scale tolerance ($\Psi = \{\frac{\pi}{8}i \mid i = 0 \dots 15\},\ \Upsilon = \{-1, 0, 1\}$), filter S_{f_1} correctly detects 62 bifurcations, Fig. 4.7c. Finally, a total of 85 (out of 101) bifurcations are correctly detected when the filter is applied in rotation-, scale- and reflection-tolerant mode, Fig. 4.7d.

This high number of detected features illustrates the strong generalization capability of our approach because 84.16% of the features of interest are detected by a single COSFIRE filter. Notable is the fact that for the selected threshold $t_3(S_{f_1}) = 0.39$ the filter does not produce any false positive responses, which means that this result is achieved at precision of 100%. Fig. 4.8(a-d) show the features that are detected by

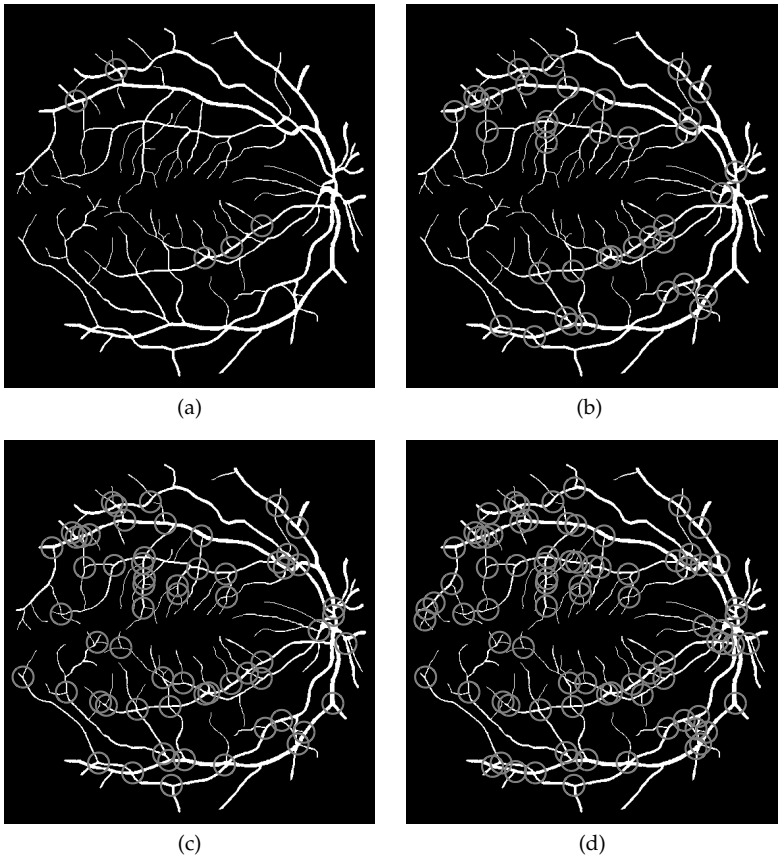
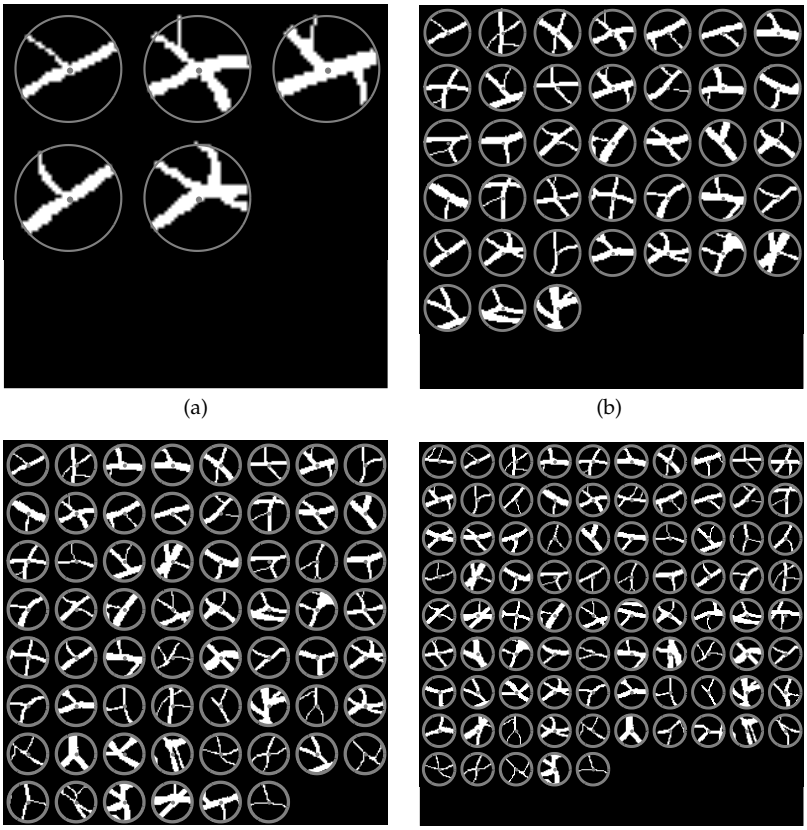


Figure 4.7: (a) Results of applying the filter S_{f_1} to the training retinal image (40_manual1.gif) in different modes. The number of detected true positive (TP) features increases as the filter achieves tolerance to geometric transformations: (a,e) non-tolerant (TP $= 5$), (b,f) rotation-tolerant (TP $= 38$), (c,g) rotation- and scale-tolerant (TP $= 62$), and (d,h) rotation-, scale- and reflection-tolerant (TP $= 85$). (e-f) Enlargement of the detected features in the corresponding retinal images. The gray dots indicate the positions at which the COSFIRE filter achieves local maximum responses. The surrounding circles correspond to the area of support of the concerned COSFIRE filter.

the concerned COSFIRE filter rendered as enlarged and isolated patterns. Each gray dot indicates the location at which filter S_{f_1} achieves a local maximum response. Qualitatively, the indicated locations closely match the real branching points of interest.

In the second iteration of the training process, we randomly choose one of the bifurcations, which we denote by f_2, that has not been detected by filter S_{f_1} and



(a) (b) (c) (d)

Figure 4.8: Enlargement of the detected features in the corresponding retinal images shown in Fig. 4.7. The gray dots indicate the positions at which the COSFIRE filter achieves local maximum responses. The surrounding circles correspond to the area of support of the concerned COSFIRE filter.

use it to configure a second COSFIRE filter S_{f_2}. With this second filter we detect 82 features of interest of which 70 coincide with the features detected by the first filter S_{f_1} and 12 are newly detected features ($t_3(S_{f_2}) = 0.21$). Merging the responses of both filters results in the detection of 97 distinct bifurcations.

We repeat this procedure of configuring filters using features that have not been detected by the previously trained filters until all the vessel features in the training image are detected. In this case, a set of five filters that are configured by the prototype bifurcations shown in Fig. 4.6(a-e) proves sufficient to detect all 101 retinal features in the concerned training image. An important aspect of this training

procedure is that the values of the threshold parameters $t_3(S_{f_i})$ are chosen in such a way to produce a recall of 100% and a precision of 100%[7] for the training retinal image.

4.3.3 Performance Evaluation

We evaluate the effectiveness of the above five COSFIRE filters, which were configured from bifurcations selected from a single retinal image, on the entire data set of 40 manually segmented retinal images taken from the DRIVE data set. For this data set and for the determined values of the threshold parameters $t_3(S_{f_i})$ of the corresponding filters we achieve a total recall R of 97.38% and a total precision P of 95.75%. Changing the values of the corresponding parameters $t_3(S_{f_i})$ yields different performance results: P increases and R decreases with increasing values of $t_3(S_{f_i})$. For each COSFIRE filter we add to (or subtract from) the corresponding learned threshold value $t_3(S_{f_i})$ the same offset value in steps of 0.01. With the referred five COSFIRE filters, the harmonic mean $(2PR/(P+R))$ of the precision and recall reaches a maximum at a recall R of 98.26% and a precision P of 94.98% when each value of the threshold parameter $t_3(S_{f_i})$ is offset by the same value of -0.01 from the corresponding learned threshold value, Fig. 4.9b.

4.3.4 Effects of the Number of COSFIRE Filters

In the above experiments we made no assumptions on the required number of COSFIRE filters. The performance results may, however, change with a different number of such filters. We consider this aspect and configure more COSFIRE filters as follows. First, we use the same training retinal image and apply to it the above set of five COSFIRE filters, but this time with the value of their respective threshold parameters $t_3(S_{f_i})$ offset by $+0.1$. For this offset value the COSFIRE filters become more selective and achieve a recall of 81.19% at a precision of 100% for the training retinal image. Then, we use the iterative training procedure described above to configure more COSFIRE filters until all the features of interest are detected in the training retinal image. It turns out that by configuring another five COSFIRE filters, Fig. 4.6(k-t), and applying them together with the first five filters to the concerned image we achieve a recall of 100% at a precision of 100%.

We evaluate the effectiveness of the set of 10 COSFIRE filters by executing 10 experiments on the same data set of 40 retinal images. In the first experiment we only apply the first COSFIRE filter S_{f_1} and in the subsequent experiments we increment

[7]Recall is the percentage of true bifurcations or crossovers that are successfully detected. Precision is the percentage of correct bifurcations or crossovers from all local maximum filter responses that exceed the corresponding t_3 threshold values.

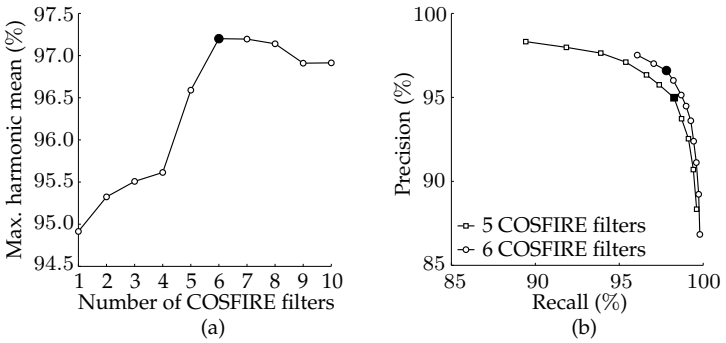


Figure 4.9: (a) The plot illustrates the maximum harmonic mean achieved by using different number of COSFIRE filters. The best performance is achieved with only six COSFIRE filters for a maximum harmonic mean of 97.20%. This point is indicated by the filled-in circle. (b) Precision-recall plots obtained with five and six COSFIRE filters. For each plot the threshold values of parameters $t_3(S_{f_i})$ are varied by adding the same offset value, ranging between -0.05 to $+0.05$ in intervals of 0.01, to the corresponding learned threshold values. The precision P increases and the recall R decreases with an increasing offset value. The harmonic mean of precision and recall reaches a maximum at $R = 97.81\%$ and $P = 96.60\%$ for six filters and at $R = 98.26\%$ and $P = 94.98\%$ for five filters. These points are indicated by a filled-in circle and a filled-in square marker, respectively.

the number of COSFIRE filters one by one, in the same order that they are configured, Fig. 4.6. For each experiment we compute the maximum harmonic mean and show it in the plot in Fig. 4.9a. The maximum harmonic mean is achieved with only six COSFIRE filters for a recall of 97.81% at a precision of 96.60%, Fig. 4.9b. This result is achieved when the learned value of each threshold parameter $t_3(S_{f_i})$ is offset by the same value of $+0.03$. In Fig. 4.9b we show the corresponding precision-recall plot for these six filters together with the one for five filters.

4.3.5 Effects of the Training Data

We achieve the above performance results by applying six COSFIRE filters that were configured by using bifurcations that we selected from a single retinal image. Here we test whether the training retinal image that we use has an effect on the performance of the COSFIRE filters. For this purpose we repeat the entire experimental procedure for another two training retinal images.

For the first of the two new experiments we use the training retinal image with filename 21_manual1.gif that we randomly choose from the DRIVE data set. In this case, we achieve a maximum harmonic mean for a recall of 98.46% and a precision of 96.66% for the entire data set of 40 images with only five COSFIRE filters. For

Table 4.1: Performance results for separate sets of COSFIRE filters that are configured by using bifurcations selected from different training retinal images. Although the number of resulting COSFIRE filters may vary, the random selection of the training retinal image does not effect the global performance results. The harmonic mean H of recall R and precision P is comparable across the three experiments.

Exp. No.	Filename of training image	No. of COSFIRE filters	R (%)	P (%)	H (%)
1	40_manual1.gif	6	97.81	96.60	97.20
2	21_manual1.gif	5	98.46	96.66	97.58
3	38_manual1.gif	5	97.36	97.57	97.46
		Average performance	97.88	96.94	97.41

the second experiment, in which we use the training retinal image with filename 38_manual1.gif, it turns out that the best performance is also achieved with five COSFIRE filters: recall of 97.36% at a precision of 97.57%.

Our conclusion is that the selection of the training retinal image does not effect the overall performance of the COSFIRE filters. Table. 4.1 shows the comparable performance results that we achieve for the three experiments.

4.3.6 Evaluating the Generalization Ability of COSFIRE Filters

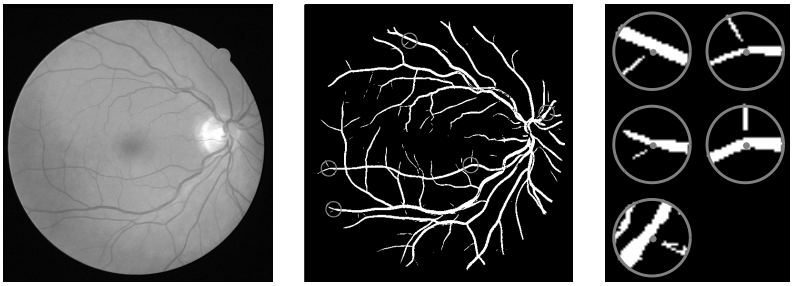
We use the STARE data set to evaluate the generalization ability of the six COSFIRE filters that we configured above. We apply the concerned six filters and evaluate their performance by varying the originally learned corresponding threshold values $t_3(S_{f_i})$ by given offsets.

For the 20 manually segmented retinal images of the STARE data set we reach a maximum harmonic mean at a recall of 97.83% and a precision of 95.64% when the threshold values $t_3(S_{f_i})$ are offset by $+0.02$. For the same threshold offset value of $+0.03$ which contributed to the maximum harmonic mean that we achieved for the DRIVE data set, we achieve a recall of 97.32% at a precision of 96.04% for the STARE data set.

We carry out further experiments on binary retinal images that are automatically segmented by a program. We use the state-of-the-art algorithm[8] proposed by (Soares et al., 2006) to automatically segment the fundus images of the DRIVE data set. We randomly select 10 binary images[9] that are automatically segmented with the concerned algorithm and use them to evaluate the generalization ability of the same six COSFIRE filters. Fig. 4.10b illustrates an example of one vessel tree

[8]The Matlab code of (Soares et al., 2006) can be downloaded from
`http://sourceforge.net/apps/mediawiki/retinal/index.php?title=Software`

[9]The segmented images that we use are produced from the color retinal images named: 02_test.tif, 04_test.tif, 07_test.tif, 09_test.tif, 10_test.tif, 12_test.tif, 13_test.tif, 16_test.tif, 18_test.tif, 19_test.tif,



(a) (b) (c)

Figure 4.10: (a) Original retinal fundus image (of size 565×584 pixels) with filename 02_test.tif that is taken from the DRIVE data set. (b) The corresponding segmented image that is automatically obtained by the algorithm proposed by (Soares et al., 2006). The encircled bifurcations that are shown enlarged in (c) are few examples of incomplete bifurcations that are correctly detected by the COSFIRE filters. The gray dots in (c) indicate the locations at which the COSFIRE filters achieve local maximum responses.

that is automatically extracted from the original retinal image shown in Fig. 4.10a. For these 10 images and for the same six COSFIRE filters we achieve a maximum harmonic mean at a recall of 97.02% and a precision of 96.53% when the threshold values $t_3(S_{f_i})$ are also offset by a value of $+0.03$.

Similar to other algorithms, the automatic segmentation technique that we use suffers from insufficient robustness to preserve the connectedness of certain bifurcations. This problem is typically addressed by applying some morphological operations, such as closing. However, such enhancements do not consider the orientation of the involved contours before filling the gaps, and thus they usually introduce artefacts. Bifurcation detection approaches that are based on the skeleton of the vessel tree, such as the ones proposed in (Martinez-Perez et al., 2002; Chanwimaluang and Guoliang, 2003; Jung and Hong, 2006; Bhuiyan et al., 2007; Ardizzone et al., 2008; Aibinu et al., 2010; Calvo et al., 2011) fail to detect incomplete bifurcations. Here, we do not apply any post-processing techniques to the resulting binary image that is obtained by the concerned segmentation algorithm because the COSFIRE filters are robust to such incomplete bifurcations. This robustness is attributable to the tolerance that we allow in the position of the involved contour parts. Fig. 4.10c shows few examples of incomplete bifurcations that are nevertheless correctly detected by the COSFIRE filters.

The results that we achieve for the STARE images and for the automatically segmented DRIVE images are similar to the best results that we achieve on the DRIVE data set of manually segmented images. These results demonstrate the generalization ability of the six COSFIRE filters that we configured on a single manually segmented retinal image that we selected from the DRIVE data set.

4.4 Discussion

In our work we are concerned with the shape-recognition performance of the proposed COSFIRE approach and we choose to isolate this from low contrast effects that come with colour/grayscale retinal images. We evaluate the performance of the COSFIRE filters that we use on manually segmented retinal images provided in the DRIVE (Staal et al., 2004) and in the STARE (Hoover et al., 2000) data sets but also on segmented retinal images that are automatically produced by the state-of-the-art algorithm that was introduced by (Soares et al., 2006).

We showed that by using a single training retinal image we could configure a small set of COSFIRE filters, in this case six, and subsequently use them to effectively detect the vessel features in segmented retinal images. We demonstrated that the training retinal image that we randomly chose did not effect the performance of the resulting COSFIRE filters. We also showed that the COSFIRE filters have strong generalization ability as they achieved comparable performance across three data sets. As to the localization accuracy, we showed some qualitative results but did not compute it quantitatively because the ground truth was not provided by medical experts. Nevertheless, the precision and the localization accuracy can be improved by performing additional analysis of the features that are detected by the COSFIRE filters.

The performance results (harmonic mean of 97.20%: recall of 97.81% and precision of 96.60%) that we achieve on the manually segmented images of the DRIVE data set outperform the ones (harmonic mean of 96.83%: recall of 98.52% and precision of 95.19%) that we obtained with a preliminary method (Azzopardi and Petkov, 2011). Moreover, the results that we report are comparable to the ones obtained by other studies but they do not report results for the entire data sets of DRIVE and STARE. For instance, Bhuiyan et al. (2007) report a recall of 95.82% on a small data set of five retinal images only. Aibinu et al. (2010) report a total recall of 98.35% and a total precision of 95.22% on a small set of five retinal images taken from the STARE data set. Other studies (Ali et al., 1999; Shen et al., 2001; Martinez-Perez et al., 2002; Chanwimaluang and Guoliang, 2003; Tsai et al., 2004; Jung and Hong, 2006; Ardizzone et al., 2008; Calvo et al., 2011) that also investigated the detection of vessel features do not provide results on the DRIVE and on the STARE data sets.

In the iterative training process we start by randomly selecting a typical bifurcation to configure the first filter and in every subsequent iteration we configure a filter by a bifurcation that is randomly selected from the bifurcations that are not detected by the previously configured filters. The selection of a different order of prototype bifurcations may yield to more (or less) than six COSFIRE filters to achieve the same or similar performance. By systematically analysing the orientation and scale statis-

tics of given training bifurcations one may configure an optimal set of COSFIRE filters. Such an approach was applied by Wang et al. (2005) to select an optimal bank of Gabor filters for the recognition of characters.

The configuration of a COSFIRE filter models the spatial arrangement of dominant contour parts that lie along concentric circles of given radii around a specified point of interest. A COSFIRE filter becomes more selective with an increasing number of such circles. Here we use vascular bifurcations to configure COSFIRE filters that receive input from selected channels of a bank of symmetric Gabor filters. The prototype patterns and Gabor filters are, however, not intrinsic to the method. Elsewhere we show that COSFIRE filters can be configured by other types of prototype patterns (Azzopardi and Petkov, 2013). Moreover, we demonstrated that by using Mexican hat operators, rather than Gabor filters, we can effectively configure a contour operator that we call CORF (Azzopardi and Petkov, 2012).

We implemented the proposed approach in Matlab[10] and ran all experiments on a personal computer with a 2.3GHz processor. We also made this application available on the internet[11]. The number of computations that is required to process a retinal image does not depend on the number of COSFIRE filters used. This is because multiple COSFIRE filters may share many computations. In practice, the computation of one blurred and shifted response (for the same values of the parameters λ, θ and ρ), for instance with $s_{\lambda,\theta,\rho,\phi=0}(x, y)$, is sufficient: the result of $s_{\lambda,\theta,\rho,\phi}(x, y)$ for any value of ϕ can be obtained from the result of the output of $s_{\lambda,\theta,\rho,\phi=0}(x, y)$ by appropriate shifting. Therefore, the number of computations required depend on the total number of combinations of values (λ, θ, ρ) involved in all the COSFIRE filters used. For six values of λ, eight values of θ and three values of ρ that we use to configure the COSFIRE filters it results in a maximum of 144 blurring operations. The remaining computations for shifting and computing the responses of the COSFIRE filters are negligible. The six COSFIRE filters that we use involve a total number of 41 unique combinations of (λ, θ, ρ) and they take less than 50 seconds to process one retinal image of size 565×584 pixels. We note that the implementation that we use here is based on sequential processing. The COSFIRE approach can, however, be easily implemented in a parallel mode as most of the computations are independent of each other.

The multivariate function that we use to combine the responses of Gabor filters is weighted geometric mean. In future work, we plan to experiment with other functions. The way we determine the standard deviation of the blurring function is inspired by neurophysiological evidence that the average diameter of receptive

[10]Matlab scripts for the configuration and application of COSFIRE filters can be downloaded from `http://matlabserver.cs.rug.nl/`

[11]`http://matlabserver.cs.rug.nl/RetinalVascularBifurcations`

fields[12] of V4 neurons increases with the eccentricity (Gattass et al., 1988).

The COSFIRE filters that we use differ from other bifurcation detection approaches in two main aspects. First, in comparison to skeleton-based approaches our method is more robust for the detection of incomplete junctions that are typically incorrectly produced by automatic segmentation algorithms. Skeleton-based approaches also involve some pre-processing techniques, such as morphological operations and skeletonization, which are not required by the COSFIRE approach. Second, the model-based approaches suggested in the literature (Ali et al., 1999; Shen et al., 2001; Tsai et al., 2004) design a fixed model with *a priori* knowledge and predefine it in the implementation. This is also the case for skeleton-based approaches that use a pre-defined set of template kernels. On the contrary, a COSFIRE filter is trainable as it is configured with any local pattern that is specified by a user. Thus, the COSFIRE approach is more versatile and can be used to detect patterns other than bifurcations and crossovers.

In future work we will evaluate the COSFIRE filters on grayscale retinal images and we will extend this work to create a descriptor of certain properties, such as angle measurement, of the detected bifurcations. Another direction for future research is to extend this approach to 3D COSFIRE filters that can be applied, for instance, to tubular organ registration and bifurcation detection in X-ray computed tomography (CT) medical images.

4.5 Conclusions

We demonstrated that the trainable COSFIRE filters that we use are an effective means to automatically detect vascular bifurcations in manually and automatically segmented retinal fundus images. We achieve an average recall of 97.88% at an average precision of 96.94% on the DRIVE data set of 40 manually segmented retinal images, a recall of 97.32% at a precision of 96.04% on the STARE data set of 20 manually segmented retinal images, and a recall of 97.02% at a precision of 96.53% on 10 automatically segmented retinal images from the DRIVE data set.

The COSFIRE filters are versatile detectors of contour related features because they can be trained with any local contour prototype pattern and are subsequently able to detect patterns that are identical or similar to the prototype pattern. Furthermore, they are conceptually simple and easy to implement: the filter output is computed as the weighted geometric mean of blurred and shifted Gabor filter responses.

[12] In neurophysiology a receptive field refers to an area in the visual field which provides input to a given neuron. Its mathematical counterpart is the support of an operator.

Submitted as:

G. Azzopardi, N. Petkov, *"Trainable shape-selective S-COSFIRE filters for object recognition and localization,"*, 2012.

Chapter 5

Trainable *S*-COSFIRE Filters for Shape Recognition and Localization

Abstract

Recognition and localization of objects embedded in complex scenes is important for many computer vision applications. Most existing methods require prior segmentation of the objects from the background which on its turn requires recognition. We propose trainable shape-selective filters which we call S-COSFIRE and use to recognize and localize objects of interest that are embedded in complex scenes. Such a filter is automatically configured to be selective for an arrangement of contour-based features that belong to a prototype shape specified by an example. The configuration comprises selecting relevant vertex detectors and determining certain blur and shift parameters. The response is computed as the weighted geometric mean of the blurred and shifted responses of the selected vertex detectors. S-COSFIRE filters share similar properties with some neurons in inferotemporal cortex, which provided inspiration for this work. We demonstrate the effectiveness of S-COSFIRE filters in two applications: letter and keyword spotting in handwritten manuscripts and object spotting in complex scenes for the computer vision system of a domestic robot. S-COSFIRE filters are effective to recognize and localize (deformable) objects in images of complex scenes without requiring prior segmentation. They are versatile trainable shape detectors, conceptually simple and easy to implement.

5.1 Introduction

Shape is perceptually the most important visual characteristic of an object. Although there is no mathematical definition – as with most perceptual related concepts – there seems to be an agreement that the two-dimensional shape of an object is characterized by the relative spatial positions of a collection of contour-based features.

Let us consider, for instance, the square in Fig. 5.1a, which we refer to as a reference or prototype object. From the point of view of visual perception the incomplete object in Fig. 5.1b is very similar to the prototype even though it is composed of only 25% of the contour pixels of the reference object. On the contrary, the closed polygon in Fig. 5.1c, which has the bottom half equivalent to that of the prototype is perceptually less similar to it. Furthermore, there is little perceptual similarity between the prototype and its scrambled contour parts shown in Fig. 5.1d.

As a matter of fact, there is neurophysiological evidence that objects, such as faces, are recognized by detecting certain features that are spatially arranged in a certain way (Kobatake and Tanaka, 1994). By means of single-cell recordings in adult monkeys it was, for instance, found that a neuron in inferotemporal cortex gives similar responses for the two images shown in Fig. 5.2(a-b). The icon presented in Fig. 5.2b is a simplified version of the monkey's face shown in Fig. 5.2a. It only consists of a circle that surrounds a horizontally-aligned pair of spots on top of a horizontal bar. Removing one of these features, Fig. 5.2(c-d), causes the concerned cell to give very small response.

Another neurophysiological study (Brincat and Connor, 2004) reveals that some neurons in inferotemporal cortex integrate information about the curvatures, orientations and positions of multiple (typically two to four) simple contour elements, such as angles or curved contour segments. Fig. 5.3 shows a few such stimuli. In

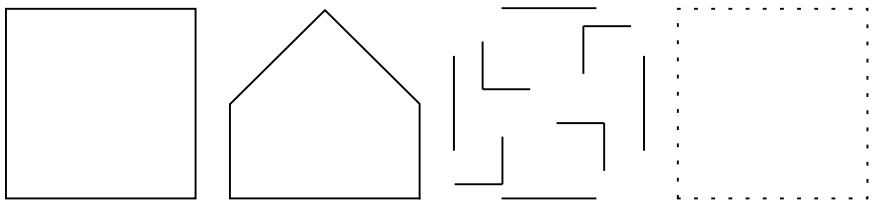


Figure 5.1: (a) A prototype shape. (b) A test pattern that has only 25% similarity (computed by template matching) to the prototype is perceptually more similar to the prototype than the polygon in (c) and the set of contour parts in (d), both of which have 50% similarity (computed by template matching) to the prototype.

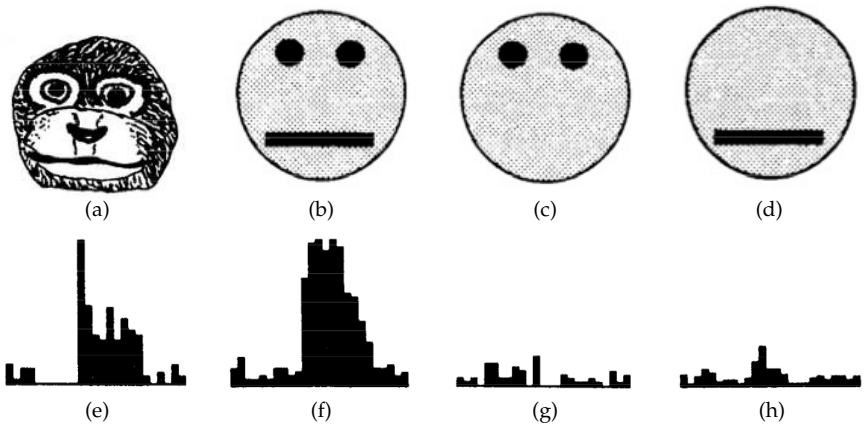


Figure 5.2: (a-d) A set of stimuli used in an electrophysiological study (Kobatake and Tanaka, 1994) to test the selectivity of a neuron in inferotemporal cortex. (e-h) The activity of the concerned neuron for the corresponding stimuli. The neuron gives high response only when the stimulus contains a detailed or simplified representation of the face boundary that surrounds a pair of eyes on top of a mouth. If any of these features is missing, the neuron gives negligible response.

that study the authors argue that their findings are in line with other studies that support parts-based shape representation theories (Marr and Nishihara, 1978; Biederman, 1987; Salinas and Abbott, 1997; Wilson et al., 1997; Mel and Fiser, 2000; Edelman and Intrator, 2003; Riesenhuber and Poggio, 1999), and suggest that nonlinear integration in the inferotemporal cortex might help to extend sparseness of shape representation along the ventral stream of the visual system of the brain.

The fact that the visual system of the brain deploys information about the relative positions of features in shape recognition is in contrast with some approaches in computer vision which use the presence of certain features but do not take into account their spatial relations. For instance, in the bag-of-visual words approach (Fei-Fei et al., 2007), the similarity between two images is computed by comparing the frequencies of occurrence of certain simple features. This method does not consider the spatial relationships between the features and as a result it may find the collection of elementary features in Fig. 5.1d similar to the prototype in Fig. 5.1a to a considerable extent. Furthermore, this approach requires the test object to be segmented from the background and encounters severe problems with recognizing objects that are embedded in an environment which contains features of the concerned types.

Existing approaches that consider the spatial relationship of features include the

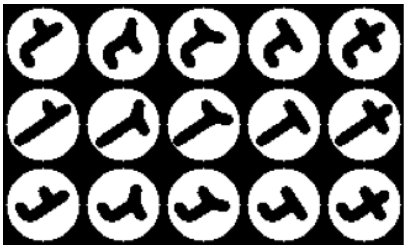


Figure 5.3: A subset of systematically designed stimuli that were used in an electrophysiological experiment in (Brincat and Connor, 2004) to study the selectivity of some neurons in inferotemporal cortex of macaque monkeys. The concerned neurons give high responses only for stimuli that are composed of a number of (typically two to four) curve segments, the relative spatial positions of which are in a specific geometrical arrangement.

so-called standard model (Serre et al., 2007), some probabilistic techniques, such as the generative constellation model (Fergus et al., 2003; Fei-Fei et al., 2007) and a hierarchical model of object categories (Fidler and Leonardis, 2007; Fidler et al., 2008). These approaches rely on summation of the responses of elementary feature detectors and may find the images in Fig. 5.1(c-d) quite similar to the prototype in Fig. 5.1a. For instance, such a technique may consider a circle with a horizontal line within it as a face even though the representations of the eyes are missing, Fig.5.2(c-d). Other examples are shown in Fig. 5.4 where a centaur can be mistakenly classified as a horse or a siren confused with a fish.

In this study we introduce an object detection technique which is motivated by the shape selectivity of some neurons in inferotemporal cortex. The principal idea of this approach is to construct a shape-selective filter that combines the responses of some simpler filters that detect some partial features of the concerned shape in

(a) (b)

Figure 5.4: Object recognition models that sum up the responses of certain filters which are selective to different parts of an object may confuse (a) a centaur with a horse or a (b) siren with a fish.

specific positions that are characteristic of that shape. In Fig. 5.5 we illustrate briefly this idea by the following example. Let us assume that we have three filters, one of which detects a circle, another one detects a small disk and the third one detects a horizontal bar. Applying these three filters to the face icon shown in Fig. 5.2b results in feature-specific responses in different positions. We then shift these responses in different directions at different distances in such a way that they meet in some point that we consider as the center of a composite filter. We compute the output of this composite filter as the (weighted) geometric mean of the inputs received from the costituent filters.

We call this approach to the construction of filters Combination Of Shifted FIlter REsponses (COSFIRE). Previously, we successfully applied this approach to the construction of line and edge detectors (Azzopardi and Petkov, 2012) and simple contour-related features, such as vascular bifurcations (Azzopardi and Petkov, 2011). In (Azzopardi and Petkov, 2011) we demonstrated how the collective responses of multiple COSFIRE filters to segmented patterns, such as handwritten digits, can be used to form a shape descriptor with high discrimination ability. That descriptor, however, does not take into account the relative spatial arrangement of the concerned features. Similar to other shape descriptors (Belongie et al., 2002; Grigorescu et al., 2003a; Sebastian et al., 2004; Latecki et al., 2005; Ghosh and Petkov, 2005; Demirci et al., 2006; Lauer et al., 2007; Ling and Jacobs, 2007; Goh, 2008; Xu et al., 2009; Almazan et al., 2012) that approach works well with segmented objects, but it is not effective for the detection of objects embedded in complex scenes. In order to distinguish the two types of filter, we refer to the composite shape-selective filter that we propose in this study as *S*-COSFIRE and to the filter proposed in (Azzopardi and Petkov, 2013) as *V*-COSFIRE (*S* and *V* stand for shape and vertex, respectively).

Due to its multiplicative character, a *S*-COSFIRE filter produces a response only when all the costituent filters give responses, in spatial positions that are characteristic for the prototype shape. In order to allow for some tolerance in the relative positions of the involved object parts we use a smoothing function to blur the responses of the costituent filters.

There are three aspects in which the *S*-COSFIRE filters that we propose differ from the other models that also consider the spatial geometric arrangement of parts. First, our model is implemented in a filter that gives a response, a scalar value between 0 and 1, for each position in the image. The higher the value the more similar the shape around the concerned location is to the prototype shape that was used to configure that filter. In this regard, a *S*-COSFIRE filter can be thought of a model of a shape-selective neuron in inferotemporal cortex of the type studied in (Kobatake and Tanaka, 1994; Brincat and Connor, 2004), which fires only when a specific ar-

rangement of contour-based features is present in its receptive field. A *S*-COSFIRE filter addresses object recognition and localization as a joint problem, which is in line with how David Marr (Marr, 1982) defined the sense of seeing: "... to know what is where by looking". In contrast, the other methods referred to above use

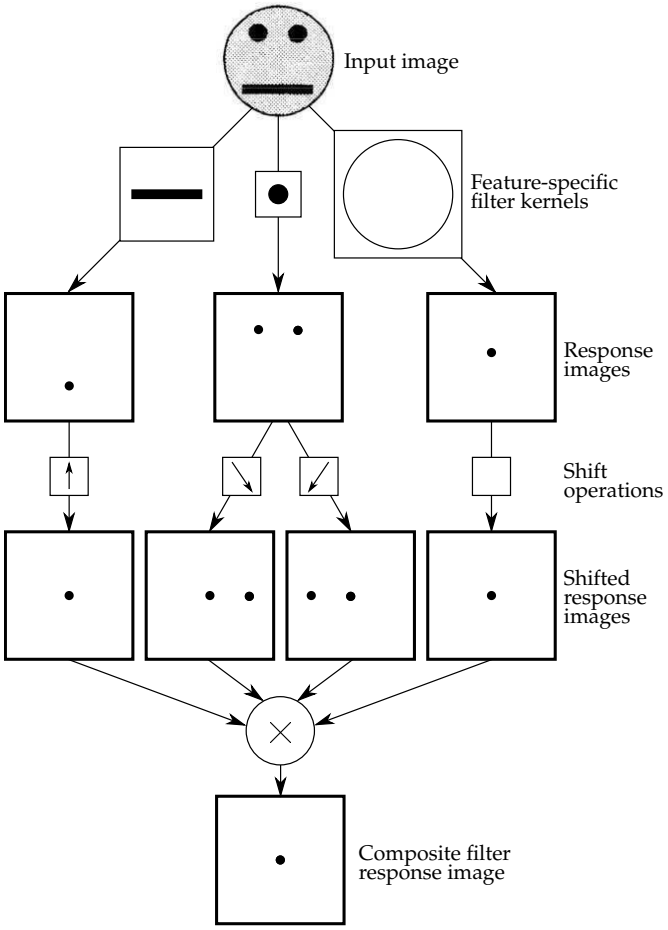


Figure 5.5: The feature-specific (disk-, bar- and circle-selective) filters achieve high responses in the centers of the eyes, the mouth and the face boundary of the input image. The response images of the concerned feature-specific filters are then translated by certain shift vectors, the magnitude and direction of which are automatically determined in a configuration process. Finally, the output of the composite shape-selective filter is computed as the (weighted) geometric mean, which is essentially multiplication, of the shifted responses. The final output indicates the presence of a face which is composed of a circle that surrounds a pair of horizontally-aligned disks on top of a horizontal bar.

multiple prototypes and consider several responses from different feature detectors to form a mixture of probability distributions or a vector of responses. For these methods, the geometrical spatial arrangement of the concerned prototype defining parts is achieved by training a supervised classifier and subsequently the similarity between a test pattern and a prototype is computed by a distance metric. Moreover, they suffer from insufficient robustness to localization because they treat this matter at a region level rather than at a pixel level.

Second, since the omission of an object part can radically change shape perception, we regard every feature (and its relative spatial position) that forms part of a prototype shape as essential. This aspect is implemented in the multiplicative character of a *S*-COSFIRE filter. On the contrary the other models, which rely on summing up the responses of some elementary feature detectors, may achieve a response even when any of the prototype-defining features is missing. These models may thus match objects that are perceptually different.

Third, while the *S*-COSFIRE approach that we present achieves invariance to rotation, scaling and reflection by simply manipulating some model parameters, the other techniques can only achieve invariance to such geometric transformations by extending the training set with example objects that are rotated, scaled and/or reflected versions of a prototype.

The remaining part of this chapter is organized as follows: In Section 5.2 we present the proposed *S*-COSFIRE filter and demonstrate how it can be configured to be selective for a given shape of interest. In Section 5.3, we demonstrate the effectiveness of the proposed approach in two applications: letter and keyword spotting in handwritten manuscripts and vision for home tidying pickup robot. Section 5.4 contains a discussion on the properties of the *S*-COSFIRE filters and finally we draw conclusions in Section 5.5.

5.2 Method

5.2.1 Overview

The following example illustrates the main idea of the method that we propose. We consider the triangle, shown in Fig. 5.6, as a shape of interest and we call it *prototype*. We use this prototype shape to automatically configure a *S*-COSFIRE filter that will respond to shapes that are identical with or similar to this prototype.

The shape-selective *S*-COSFIRE filter that we propose uses input from simpler filters, in this case filters that are selective for vertices. We use vertex-selective COSFIRE filters of the type proposed in (Azzopardi and Petkov, 2013) to detect the vertices of the prototype shape. Such a filter, which we refer to it as *V*-COSFIRE, com-

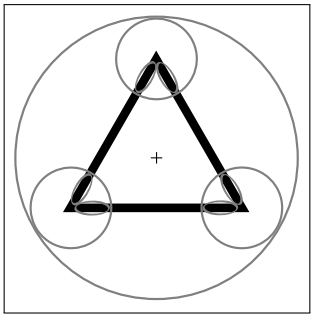


Figure 5.6: The triangle is the prototype shape of interest. The '+' marker indicates the center of the large circle that is specified by a user to select the triangle. The small circles indicate the supports of three vertex detectors that are identified as relevant for the concerned prototype shape. The small ellipses represent the supports of line detectors that are selective for the contour parts of the corresponding vertices.

bines the responses of line detectors, the areas of support of which are indicated by the small ellipses in Fig. 5.6.

The response of the proposed *S*-COSFIRE filter is computed by combining the responses of the concerned *V*-COSFIRE filters in the centers of the corresponding circles by a weighted geometric mean. The preferred orientations and the preferred apertures of these filters together with the locations at which we take their responses are determined by analysing the responses of a set of *V*-COSFIRE filters to the prototype shape. Consequently, the *S*-COSFIRE filter will be selective for the given spatial arrangement of vertices of specific orientations and apertures. Taking the responses of *V*-COSFIRE filters at different locations around a point can be implemented by shifting the responses appropriately before using them for the pixel-wise evaluation of a multivariate function which gives the *S*-COSFIRE filter output.

In the next sub-sections, we explain the automatic configuration process of a *S*-COSFIRE filter that will respond to a given prototype shape of interest and to similar shapes. The configuration process determines which responses of which *V*-COSFIRE filters in which locations need to be multiplied in order to obtain the output of the *S*-COSFIRE filter.

5.2.2 Detection of Vertex Features by *V*-COSFIRE Filters

We denote by $r_{V_{f_i}}(x, y)$ the response of a *V*-COSFIRE filter V_{f_i} that is selective for a vertex f_i. We threshold these responses at a given fraction t_1 ($0 \leq t_1 \leq 1$) of the maximum response across all image coordinates (x, y) and denote these thresholded

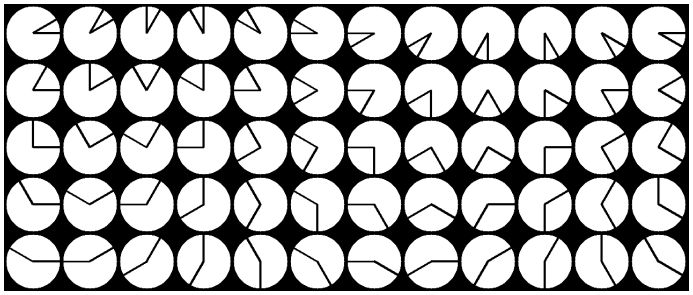


Figure 5.7: A data set of 60 synthetic vertices, $f_1, \ldots, f_{60}$ (left-to-right, top-to-bottom). A V-COSFIRE filter V_{f_k} is selective for a vertex f_k.

responses by $|r_{V_{f_i}}(x,y)|_{t_1}$. We use the publicly available Matlab implementation[1] of V-COSFIRE filters. Such a filter uses as input the responses of given channels of a bank[2] of Gabor filters. For further technical details about the properties of V-COSFIRE filters we refer to (Azzopardi and Petkov, 2013).

We use a bank of V-COSFIRE filters that are selective for vertices of different orientations (in intervals of $\pi/6$ radians) and different apertures (in intervals of $\pi/6$ radians), Fig. 5.7. Fig. 5.8a illustrates the responses of this bank of V-COSFIRE filters to the prototype shape shown in Fig. 5.6a. The strongest responses are obtained by three V-COSFIRE filters that are selective for vertices of the types f_{13}, f_{17}, and f_{21}, shown in Fig. 5.7. The corresponding locations, (x_1, y_1), (x_2, y_2), (x_3, y_3), at which they obtain the maximum responses are indicated in Fig. 5.9.

5.2.3 Configuration of a *S*-COSFIRE Filter

A S-COSFIRE filter uses as input the responses of selected V-COSFIRE filters $V_{f_{j_i}}$, $i = 1 \ldots n$, each selective for some vertex f_{j_i}, around a certain position (ρ_i, ϕ_i) with respect to the center of the S-COSFIRE filter. A 3-tuple $(V_{f_{j_i}}, \rho_i, \phi_i)$ that consists of a V-COSFIRE filter specification $V_{f_{j_i}}$ and two scalar values (ρ_i, ϕ_i) characterizes the properties of a vertex that is present in the given prototype shape: $V_{f_{j_i}}$ represents a V-COSFIRE filter that is selective for a vertex f_{j_i} and (ρ_i, ϕ_i) are the polar coordinates of the location at which its response is taken with respect to the center of the S-COSFIRE filter. In the following we explain how we obtain the parameter values of such vertices around a given point of interest.

[1] The Matlab implementation of a V-COSFIRE filter can be downloaded from `http://matlabserver.cs.rug.nl/`

[2] Here we use a bank of Gabor filters with five wavelengths $\lambda = \{4, 4\sqrt{2}, 8, 8\sqrt{2}, 16\}$ and six equidistant orientations $\theta \in \{0, \frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}, \frac{5\pi}{6}\}$

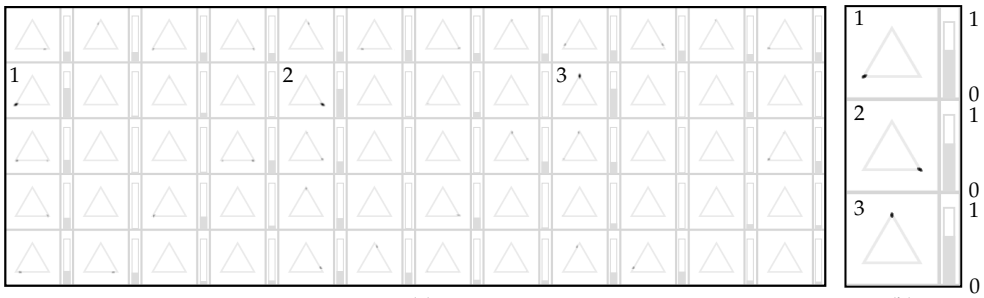


Figure 5.8: (a) Inverted response images (thresholded at $t_1 = 0$) of the *V*-COSFIRE filters that are selective for the corresponding vertices shown in Fig. 5.7. The rendering of each response image overlays the input image. Each bar plot indicates the strength of the maximum response of the corresponding *V*-COSFIRE filter to the input image. (b) The images marked by the labels '1', '2' and '3' illustrate the responses of three *V*-COSFIRE filters that achieve the highest significant responses ($t_2 = 0.75$). The concerned *V*-COSFIRE filters are selective for the vertices f_{13}, f_{17} and f_{21} in Fig. 5.7, respectively.

For each location in the input image of the prototype shape we take the maximum value of all responses achieved by the bank of *V*-COSFIRE filters mentioned above. The positions that have values greater than those of their corresponding 8-neighbours[3] are chosen as the points that have local maximum responses. For each such point (x_i, y_i) we determine the polar coordinates (ρ_i, ϕ_i) with respect to the center of the *S*-COSFIRE filter, Fig. 5.9. Then we determine the *V*-COSFIRE filters, the responses of which are greater than a fraction $t_2 = 0.75$ of the maximum response $r_{V_{f_i}}(x, y)$ for all $i \in \{1, \dots n_f\}$ where n_f is the number of *V*-COSFIRE filters used across all locations in the input image. Thus, multiple *V*-COSFIRE filters can be significantly activated for the same location (ρ_i, ϕ_i). The selected points characterize the dominant vertices in the given prototype shape of interest.

We denote by $S_{\mathsf{S}} = \{(V_{f_{j_i}}, \rho_i, \phi_i) \mid i = 1 \dots n_f\}$ the set of parameter value combinations, which describes the properties and locations of a number of vertices. The subscript S stands for the prototype shape of interest. Every tuple in set S_{S} specifies the parameters of some vertex in prototype S.

For the prototype shape of interest in Fig. 5.6a, the selection method described above results in three vertices with parameter values specified by the tuples in the following set:

[3]We use the function `imregionalmax` in Matlab to identify the locations that have local maxima responses

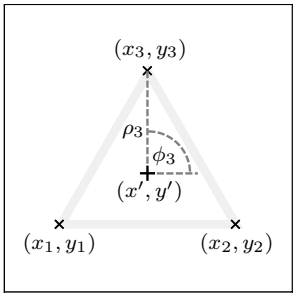


Figure 5.9: Configuration of a *S*-COSFIRE filter. The '$\times$' markers indicate the locations, (x_1, y_1), (x_2, y_2), (x_3, y_3), where the corresponding three *V*-COSFIRE filters, $V_{f_{13}}$, $V_{f_{17}}$, $V_{f_{21}}$, achieve the maximum responses. These locations correspond to the three vertices of the prototype shape, which is rendered here with low contrast. The Cartesian coordinates of each point (x_i, y_i) are converted into the polar coordinates (ρ_i, ϕ_i) with respect to the given point of interest (x', y'), indicated by the '+' marker.

$$S_{\mathsf{S}} = \left\{ \begin{array}{l} (V_{f_{j_1=21}},\ \rho_1 = 50,\ \phi_1 = \pi/2), \\ (V_{f_{j_2=13}},\ \rho_2 = 50,\ \phi_2 = 7\pi/6), \\ (V_{f_{j_3=17}},\ \rho_3 = 50,\ \phi_3 = 5\pi/3) \end{array} \right\}$$

The first tuple in S_{S}, $(V_{f_{j_1=21}}, \rho_1 = 50,\ \phi_1 = \pi/2)$, for instance, indicates the presence of vertex f_{21}, at a position of $(\rho_1 =)\ 50$ pixels to the north ($\phi_1 = \pi/2$) of the support center of the *S*-COSFIRE filter at hand.

5.2.4 Blurring and Shifting *V*-COSFIRE Responses

The above configuration results in a *S*-COSFIRE filter that is selective for a preferred spatial arrangement of three vertices forming an equilateral triangle. Next, we use the responses of the *V*-COSFIRE filters that are selective for the corresponding vertices to compute the output of the *S*-COSFIRE filter as follows.

First, we *blur* the responses of the *V*-COSFIRE filters in order to allow for some tolerance in the position of the respective vertices. This increases the generalization ability of the *S*-COSFIRE filter under construction. We define the blurring operation as the computation of maximum value of the weighted thresholded responses of a *V*-COSFIRE filter. For weighting we use a Gaussian function $G_\sigma(x, y)$, the standard deviation σ of which is a linear function of the distance ρ from the center of the *S*-COSFIRE filter:

$$\sigma = \sigma_0 + \alpha\rho \tag{5.1}$$

where σ_0 and α are constants. The choice of the linear function in Eq. 5.1 is inspired

by the visual system of the brain for which we provide more detail in Section 5.4. For $\alpha > 0$, which we use, the tolerance to the position of the respective vertices increases with an increasing distance ρ from the support center of the concerned *S*-COSFIRE filter.

Second, we *shift* the blurred responses of each *V*-COSFIRE filter by a distance ρ_i in the direction opposite to ϕ_i. With this shifting the concerned *V*-COSFIRE filter responses, which are located at different positions (ρ_i, ϕ_i) meet at the support center of the *S*-COSFIRE filter. The output of the *S*-COSFIRE filter can then be evaluated as a pixel-wise multivariate function of the shifted and blurred responses of *V*-COSFIRE filter responses. In polar coordinates, the shift vector is specified by $(\rho_i, \phi_i + \pi)$, and in Cartesian coordinates, it is $(\Delta x_i, \Delta y_i)$ where $\Delta x_i = -\rho_i \cos\phi_i$, and $\Delta y_i = -\rho_i \sin\phi_i$. We denote by $s_{V_{f_{j_i}},\rho_i,\phi_i}(x,y)$, the blurred and shifted thresholded response of a *V*-COSFIRE filter that is specified by the i-th tuple $(V_{f_{j_i}}, \rho_i, \phi_i)$ in the set S_{S}:

$$s_{V_{f_{j_i}},\rho_i,\phi_i}(x,y) \stackrel{\text{def}}{=} \max_{x',y'}\{\left|r_{V_{f_{j_i}}}(x-x'-\Delta x_i, y-y'-\Delta y_i)\right|_{t_1} G_\sigma(x',y')\} \tag{5.2}$$

where $-3\sigma \leq x', y' \leq 3\sigma$.

Fig. 5.10 illustrates the blurring and shifting operations for this *S*-COSFIRE filter, applied to the image shown in Fig. 5.10a.

5.2.5 Response of a *S*-COSFIRE Filter

We define the response $r_{S_{\mathsf{S}}}(x,y)$ of a *S*-COSFIRE filter as the weighted geometric mean of the blurred and shifted thresholded responses of the selected *V*-COSFIRE filters $s_{V_{f_{j_i}},\rho_i,\phi_i}(x,y)$:

$$r_{S_{\mathsf{S}}}(x,y) \stackrel{\text{def}}{=} \left| \left(\prod_{i=1}^{|S_{\mathsf{S}}|} \left(s_{V_{f_{j_i}},\rho_i,\phi_i}(x,y) \right)^{\omega_i} \right)^{1/\sum_{i=1}^{|S_{\mathsf{S}}|}\omega_i} \right|_{t_3} \tag{5.3}$$

$$\omega_i = \exp^{-\frac{\rho_i^2}{2\sigma'^2}},\ 0 \leq t_3 \leq 1$$

where $|.|_{t_3}$ stands for thresholding the response at a fraction t_3 of its maximum across all image coordinates (x,y). For $1/\sigma' = 0$, the computation of the *S*-COSFIRE filter is equivalent to the standard geometric mean, where the s-quantities have the same contribution. Otherwise, for $1/\sigma' > 0$, the input contribution of s-quantities decreases with an increasing value of the corresponding parameter ρ. In our exper-

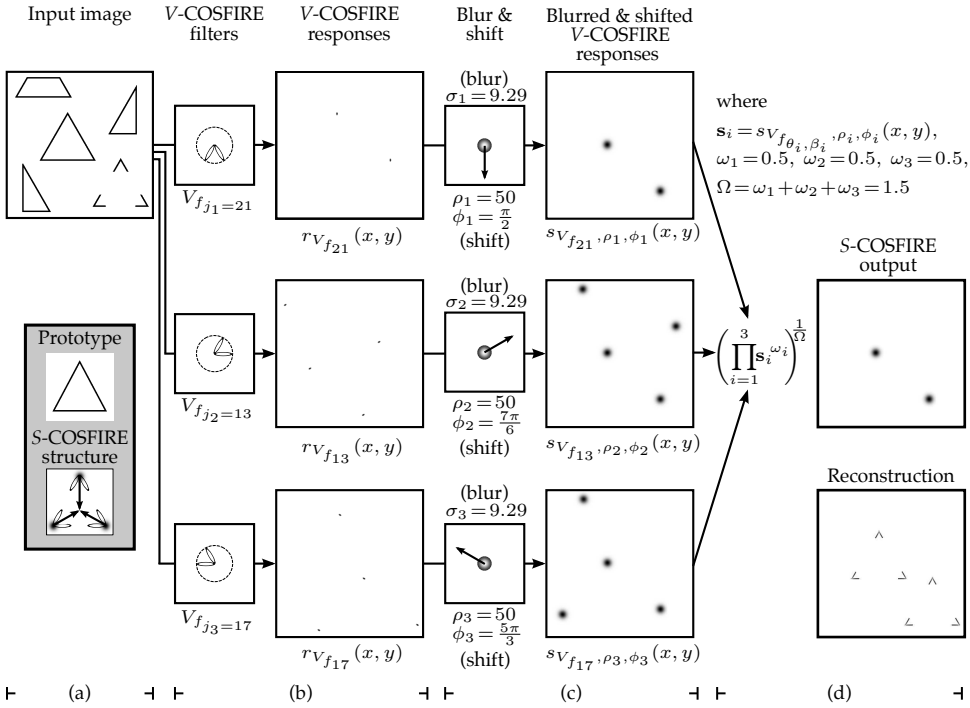


Figure 5.10: (a) Input image (of size 512×512 pixels). The enframed inlay images show (top) the enlarged prototype shape of interest, which is identical to the equilateral triangle in the input image and (bottom) the structure of the *S*-COSFIRE filter that is configured by this prototype. This *S*-COSFIRE filter is trained to detect the spatial arrangement of three vertices, each of which is detected by a separate *V*-COSFIRE filter. The ellipses illustrate the wavelengths and orientations of the Gabor filters that are used by the *V*-COSFIRE filters, and the dark blobs are intensity maps for Gaussian functions that are used to blur the responses of the corresponding *V*-COSFIRE filters. The blurred responses are then shifted by the corresponding vectors. (b) The *V*-COSFIRE filters that are automatically identified from the prototype shape and the corresponding response images to the input image. (c) We then blur (here we use $\sigma_0 = 0.1$ and $\alpha = 0.0853$ to compute σ_i in Eq. 5.1) the thresholded (here at $t_1 = 0$) response $|r_{V_{f_i}}(x,y)|_{t_1}$ of each concerned *V*-COSFIRE filter and subsequently shift the resulting blurred response images by corresponding polar-coordinate vectors $(\rho_i, \phi_i + \pi)$. (d) We use weighted geometric mean (here $\sigma' = 91.44$) of all the blurred and shifted *V*-COSFIRE filters to compute (top) the output of the *S*-COSFIRE filter and show (bottom) the reconstruction of the detected features. The reconstruction is achieved by superimposing the Gabor filter responses which contribute to the corresponding *V*-COSFIRE filters that give input to the *S*-COSFIRE filter. The two local maxima in the output of the *S*-COSFIRE filter correspond to the triangle and to the perceived one in the input image. For better clarity we use inverted gray-level rendering to show the images in the right of the columns (b), (c) and (d).

iments we use a value of the standard deviation σ' that is computed as a function of the maximum value of the given set of ρ values: $\sigma' = (-{\rho_{\max}}^2/2\ln 0.5)^{1/2}$, where $\rho_{\max} = \max_{i\in\{1\ldots|S_{\mathsf{S}}|\}}\{\rho_i\}$. We make this choice in order to achieve a maximum value $\omega = 1$ of the weights in the center (for $\rho = 0$), and a minimum value $\omega = 0.5$ in the periphery (for $\rho = \rho_{\max}$).

Fig. 5.10d shows the output of a *S*-COSFIRE filter which is defined as the weighted geometric mean of three blurred and shifted response images obtained by the three concerned *V*-COSFIRE filters. Note that this filter responds in the middle of a spatial arrangement of three vertices that is identical with or similar to that of the prototype shape S, which was used for the configuration of the *S*-COSFIRE filter. In this example, the *S*-COSFIRE filter reacts strongly in a given point that is surrounded by three vertices each having an aperture of $\pi/3$ radians: one northward-pointing, another one south-west-pointing and a south-east-pointing vertex to the north, south-west and south-east of that point, respectively. Besides the complete triangle that was used for configuration, the concerned filter also detects the Kanizsa-type (Kanizsa and Gerbino, 1976) illusory triangle. This is in line with neurophysiological and psychophysical evidence, in that the visual system is capable of detecting a shape with illusory contours, based on its visible salient parts. A thorough review of this phenomenon is provided in (Roelfsema, 2006).

Fig. 5.11a shows a systematically designed data set of isosceles triangles with different acute angles and different orientations. For the illustration in Fig. 5.11 we configure a *S*-COSFIRE filter using the enframed shape in Fig. 5.11a. Fig. 5.11b shows the responses of this *S*-COSFIRE filter where the strength of the maximum filter response to a given shape is rendered as a gray level shading of that image. The maximum response is reached at or near the center. In this case, the *S*-COSFIRE filter achieves the strongest response to the prototype shape that was used to configure it. However, it also reacts, with less than the maximum response, to triangles, the vertices of which differ slightly in acuteness. This example illustrates the selectivity (for triangle shapes) and the generalization ability of the proposed *S*-COSFIRE filter.

5.2.6 Tolerance to Geometric Transformations

Here we explain how a *S*-COSFIRE filter can be modified to achieve tolerance to the following geometric transformations: rotation, scaling and reflection. Similar to a *V*-COSFIRE filter, such a tolerance is achieved by manipulating the values of some parameters rather than by configuring separate filters by rotated, scaled and reflected versions of the prototype shape of interest.

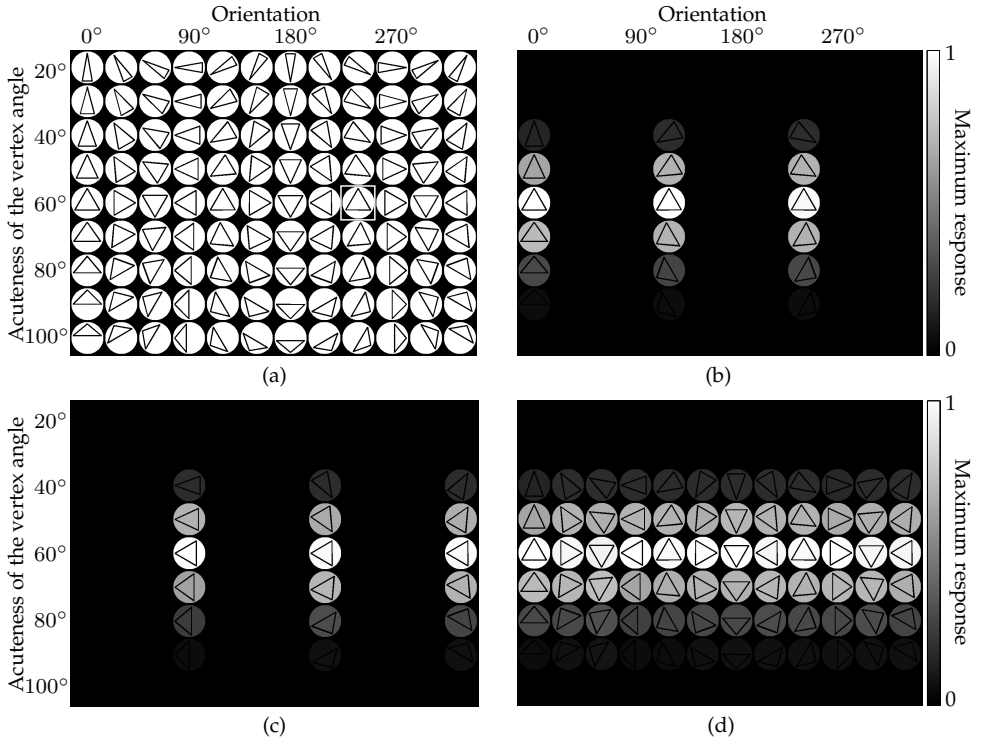


Figure 5.11: (a) A systematically designed data set of isosceles triangles that vary in orientation (right-to-left: rotated anticlockwise in intervals of $\frac{\pi}{6}$) as well as in the aperture of the vertex angle (top-to-bottom: increases in intervals of $\frac{\pi}{18}$). The enframed triangle is used as a prototype shape for configuring a *S*-COSFIRE filter. (b) Responses of the configured *S*-COSFIRE filter rendered by shading of the shapes. (c) Responses of a rotated version ($\psi = \frac{\pi}{2}$) of the filter obtained by manipulating the filter parameters and (d) rotation-tolerant responses for 12 discrete orientations.

Tolerance to Rotation

Using the set S_{S} that defines the concerned *S*-COSFIRE filter, we form a new set $\Re_{\psi}(S_{\mathsf{S}})$ that defines a new filter, which is selective for a version of the prototype shape S that is rotated by an angle ψ:

$$\Re_{\psi}(S_{\mathsf{S}}) \stackrel{\text{def}}{=} \{(\Re_{\psi}(V_{f_{j_i}}), \rho_i, \phi_i + \psi) \mid \forall\, (V_{f_{j_i}}, \rho_i, \phi_i) \in S_{\mathsf{S}}\} \tag{5.4}$$

For each tuple $(V_{f_{j_i}}, \rho_i, \phi_i)$ in the original filter S_{S} that describes a certain vertex of the prototype shape, we provide a counterpart tuple $(\Re_{\psi}(V_{f_{j_i}}), \rho_i, \phi_i + \psi)$ in the

new set $\Re_\psi(S_\mathsf{S})$. The set $\Re_\psi(V_{f_{j_i}})$ defines[4] a *V*-COSFIRE filter that is selective for vertex f_{j_i} that is also rotated by an angle ψ. The orientation of the concerned vertex and its polar angle position ϕ_i with respect to the support center of the *S*-COSFIRE filter are off-set by an angle ψ relative to the values of the corresponding parameters of the original vertex.

Fig. 5.11c shows the responses $r_{\Re_\psi(S_\mathsf{S})}$ of the *S*-COSFIRE filter that correspond to the set $\Re_\psi(S_\mathsf{S})$ to the data set of shapes shown in Fig. 5.11a. This filter responds selectively to a version of the original prototype shape S rotated counterclockwise at an angle of ($\psi =$) $\pi/2$.

A rotation-invariant response is achieved by taking the maximum value of the responses of filters that are obtained with different values of the parameter ψ:

$$\hat{r}_{S_\mathsf{S}}(x,y) \stackrel{\text{def}}{=} \max_{\psi\in\Psi}\{r_{\Re_\psi(S_\mathsf{S})}(x,y)\} \tag{5.5}$$

where Ψ is a set of n_ψ equidistant orientations defined as $\Psi = \{\frac{2\pi}{n_\psi}i \mid 0 \leq i < n_\psi\}$. Fig. 5.11d shows the maximum superposition $\hat{r}_{S_\mathsf{S}}(x,y)$ for $n_\psi = 12$. The filter according to Eq. 5.5 produces the same response to equilateral triangles that are versions of each other, obtained by rotation at discrete angles $\psi \in \Psi$.

Tolerance to Scaling

Tolerance to scaling is achieved in a similar way. Using the set S_S that defines the concerned *S*-COSFIRE filter, we form a new set $T_\upsilon(S_\mathsf{S})$ that defines a new filter, which is selective for a version of the prototype shape S that is scaled in size by a factor υ:

$$T_\upsilon(S_\mathsf{S}) \stackrel{\text{def}}{=} \{(T_\upsilon(V_{f_{j_i}}), \upsilon\rho_i, \phi_i) \mid \forall\,(V_{f_{j_i}}, \rho_i, \phi_i) \in S_\mathsf{S}\} \tag{5.6}$$

For each tuple $(V_{f_{j_i}}, \rho_i, \phi_i)$ in the original *S*-COSFIRE filter S_S that describes a certain vertex of the prototype shape, we provide a counterpart tuple $(T_\upsilon(V_{f_{j_i}}), \upsilon\rho_i, \phi_i)$ in the new set $T_\upsilon(S_\mathsf{S})$. The set $T_\upsilon(V_{f_{j_i}})$ defines[4] a V-COSFIRE filters that responds to a version of the vertex f_{j_i} scaled by the factor υ. The size of the concerned vertex and its distance to the center of the filter are scaled by the factor υ relative to the original values of the corresponding parameters.

A scale-invariant response is achieved by taking the maximum value of the responses of filters that are obtained with different values of the parameter υ:

$$\tilde{r}_{S_\mathsf{S}}(x,y) \stackrel{\text{def}}{=} \max_{\upsilon\in\Upsilon}\{r_{T_\upsilon(S_\mathsf{S})}(x,y)\} \tag{5.7}$$

[4]We refer to (Azzopardi and Petkov, 2013) for the technical details about the invariance that is achieved by a *V*-COSFIRE filter.

where Υ is a set of υ values equidistant on a logarithmic scale defined as $\Upsilon = \{2^{\frac{i}{2}} \mid i \in \mathbb{Z}\}$.

Reflection Invariance

As to reflection invariance we first form a new set $\acute{S}_{\mathsf{S}}$ from the set S_{S} as follows:

$$\acute{S}_{\mathsf{S}} \stackrel{\text{def}}{=} \{(\acute{V}_{f_{j_i}}, \rho_i, \pi - \phi_i) \mid \forall\, (V_{f_{j_i}}, \rho_i, \phi_i) \in S_{\mathsf{S}}\} \tag{5.8}$$

The set $\acute{V}_{f_{j_i}}$ defines[4] a new *V*-COSFIRE filter that is selective for the corresponding vertex f_{j_i} reflected about the $y-$axis. Similarly, the new *S*-COSFIRE filter $\acute{S}_{\mathsf{S}}$ is selective for a reflected version of the prototype shape S also about the $y-$axis. A reflection-invariant response is achieved by taking the maximum value of the responses of the filters S_{S} and $\acute{S}_{\mathsf{S}}$:

$$\acute{r}_{S_{\mathsf{S}}}(x,y) \stackrel{\text{def}}{=} \max\{r_{S_{\mathsf{S}}}(x,y), r_{\acute{S}_{\mathsf{S}}}(x,y)\} \tag{5.9}$$

Combined Tolerance to Rotation, Scaling and Reflection

A *S*-COSFIRE filter achieves tolerance to all the above geometric transformations by taking the maximum value of the rotation- and scale-tolerant responses of the filters S_{S} and $\acute{S}_{\mathsf{S}}$ that are obtained with different values of the parameters ψ and υ:

$$\bar{r}_{S_{\mathsf{S}}}(x,y) \stackrel{\text{def}}{=} \max_{\psi \in \Psi, \upsilon \in \Upsilon} \{\hat{r}_{\Re_\psi(T_\upsilon(S_{\mathsf{S}}))}(x,y), \hat{r}_{\Re_\psi(T_\upsilon(\acute{S}_{\mathsf{S}}))}(x,y)\} \tag{5.10}$$

We use this type of transformation-invariant filter in an application that is presented in Section 5.3.2 below.

5.3 Applications

In the following we demonstrate the effectiveness of the proposed *S*-COSFIRE filters by applying them in two practical applications: the spotting of letters and keywords in handwritten manuscripts and the spotting of objects in complex scenes for the computer vision system of a domestic robot.

5.3.1 Spotting Letters and Keywords in Handwritten Manuscripts

The automatic recognition of letters and words in handwritten manuscripts is an application that has been extensively investigated for several decades (Plamondon

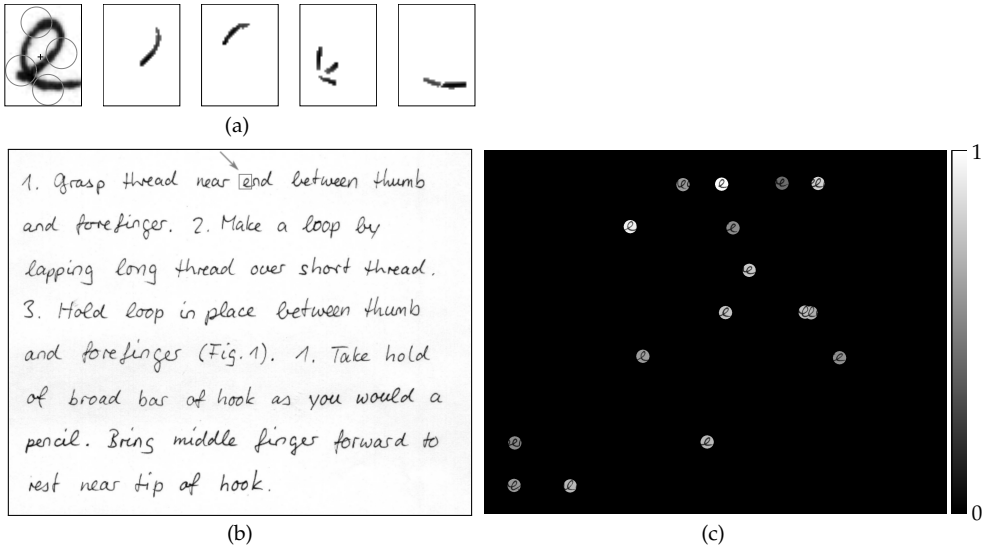


Figure 5.12: (a) The leftmost image is a prototype letter 'e' which is cropped from (b) a handwritten manuscript taken from the IAM offline database (Marti and Bunke, 2002). The circles represent the support of *V*- COSFIRE filters, the locations of which are randomly selected in a way that their support do not overlap each other. The '+' marker indicates the center of a *S*-COSFIRE filter with an area of support that is a combination of those of the *V*-COSFIRE filters. The rightmost four images in (a) depict the reconstructions of the local patterns that are illustrated as a superposition of the thresholded ($t_1 = 0.3$) Gabor filter (inverted) responses that contribute to the corresponding *V*-COSFIRE filters. (c) Normalized response of the concerned *S*-COSFIRE filter to the handwritten manuscript rendered by shading of the letters. The concerned *S*-COSFIRE filter responds to 16 out of 24 instances of the letter 'e' in the given manuscript image, of which the strongest response is achieved for the prototype letter that was used for its configuration.

and Srihari, 2000; Vinciarelli, 2002). Despite this effort the problem has not been solved completely yet.

Here, we show how the *S*-COSFIRE filters that we propose can be effectively used to detect single letters, combinations of letters, as well as complete keywords. The latter task is commonly referred to as keyword spotting in handwritten manuscripts and it is an active research area (Saykol et al., 2004; Leydier et al., 2005; Moghaddam and Cheriet, 2009; Gatos and Pratikakis, 2009; Frinken et al., 2012).

As a demonstration we show how to detect the letters 'e' in the handwritten manuscript[4] shown in Fig. 5.12a, which is taken from the IAM offline database[5]

[4]The manuscript image is extracted from the file named e01-113.png in the IAM offline database.

[5]`http://www.iam.unibe.ch/fki/databases/iam-handwriting-database`.

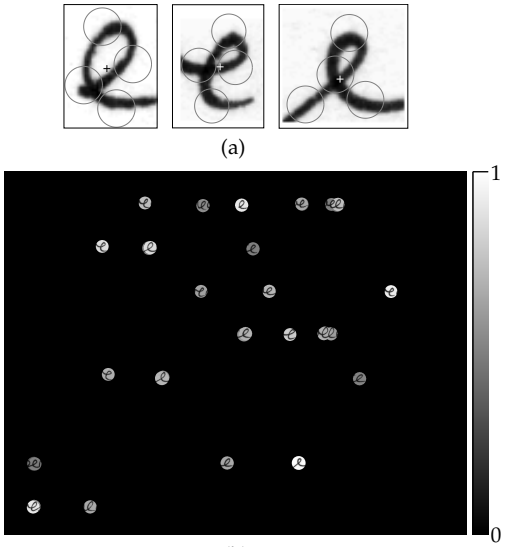


Figure 5.13: (a) Three prototype letters 'e' that are used to configure three *S*-COSFIRE filters. The circles represent the support of the *V*-COSFIRE filters used and the '+' markers indicate the support center of the concerned *S*-COSFIRE filters. (b) The maximum superposition of the normalized responses of the three *S*-COSFIRE filters to the handwritten manuscript image shown in Fig. 5.12b rendered by shading of the letters. In total, the concerned three filters detect all letters of interest and they do not produce any false positives.

(Marti and Bunke, 2002). We select a prototype letter 'e' which we use to configure an *S*-COSFIRE filter, Fig. 5.12a. In practice, this selection is done by specifying a region of appropriate size centered at the concerned letter. The concerned 'e'-selective *S*-COSFIRE filter receives input from four *V*-COSFIRE filters. The number of *V*-COSFIRE filters used is a model parameter that is specified by the user. This value depends on the shape complexity (as represented by the number of vertex features) of the concerned prototype. The selectivity of an *S*-COSFIRE filter increases with an increasing number of involved *V*-COSFIRE filters. Here, the non-overlapping support areas of four *V*-COSFIRE filters are sufficient to cover the given prototype letter. We provide more details about this aspect of our method in Section 5.4. The involved *V*-COSFIRE filters are not predefined as this was done in Section 5.2. Here they are configured by using non-overlapping contour parts that are randomly selected by the system, Fig. 5.12a.

We apply the resulting *S*-COSFIRE filter to the handwritten manuscript image and show its responses in Fig. 5.12c. The shaded circles are centered on the local maxima of the filter response and their grayscale values correspond to the response

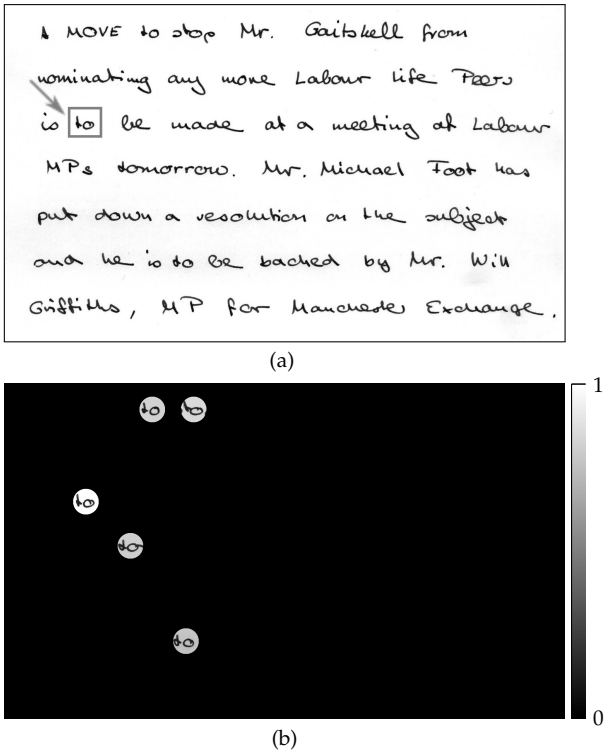


Figure 5.14: An example of spotting the combined letters 'to'. (a) A handwritten manuscript taken from the IAM offline database (Marti and Bunke, 2002). The indicated enframed two-letter word 'to' is used as a prototype to configure an *S*-COSFIRE filter. (b) Normalized response of the 'to'-selective *S*-COSFIRE filter to the handwritten manuscript rendered by shading of the letters. The concerned *S*-COSFIRE filter detects all five instances of the combined letters of interest. The strongest response is achieved for the combined letters that were used for the configuration of the *S*-COSFIRE filter.

strength of the filter. This filter detects 16 out of 24 letters 'e'. The strongest response is achieved for the letter 'e' that was used as a prototype to configure the concerned filter. This example demonstrates the generalization and discrimination abilities of the proposed *S*-COSFIRE filter because it detects 66.67% letters of interest and it does not produce any false positives. The detection of the double 'e' letters in the word 'between' in the fourth line also demonstrates the ability to detect closely adjacent shapes of interest.

As to the remaining letters 'e' that are not detected by this *S*-COSFIRE filter we proceed as follows: we take one of these letters and train a second *S*-COSFIRE filter using it. With this second filter we detect 10 out of 24 letters of interest, three of

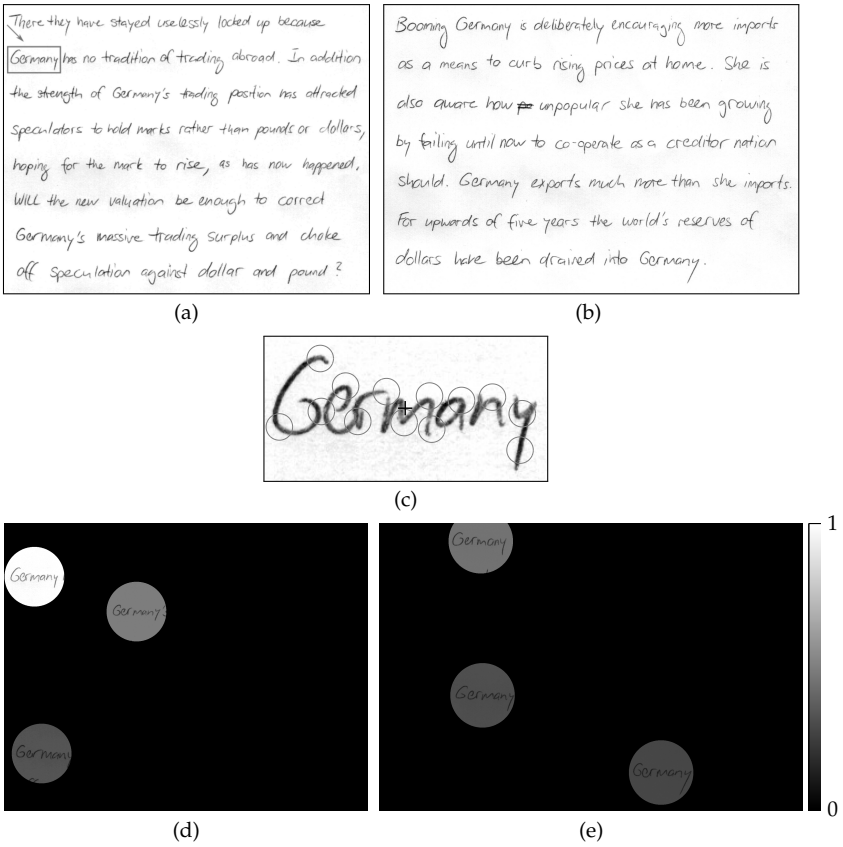


Figure 5.15: An example of spotting the keyword 'Germany' in (a-b) two handwritten manuscripts taken from the IAM offline database (Marti and Bunke, 2002). The indicated enframed keyword 'Germany' in (a) is used as a prototype to configure an *S*-COSFIRE filter. (c) The circles indicate the support areas of 13 *V*-COSFIRE filters that are used to provide input to the concerned *S*-COSFIRE filter, the support center of which is indicated by the '+' marker. (d-e) Normalized responses of this *S*-COSFIRE filter to the two handwritten manuscripts rendered by shading of the spotted words. The concerned *S*-COSFIRE filter detects all six instances of the keywords of interest in the given two images. The strongest response is achieved for the word that was used for the configuration of the *S*-COSFIRE filter.

which overlap with letters detected by the first *S*-COSFIRE filter and seven letters are newly detected ones. Applying the two filters together results in the detection of 23 distinct letters 'e'. By configuring a third *S*-COSFIRE filter with the remaining letter 'e', we detect all the letters of interest in the given handwritten manuscript, Fig. 5.13. Notable is the fact that this result is obtained without the need for prior

segmentation and that no false positives are produced by the filters. We configure and apply the mentioned three *S*-COSFIRE filters with the following parameter values: $t_1 = 0.3$, $t_2 = 0.75$, $t_3 = 0.3$, $\sigma_0 = 0.67$, and $\alpha = 0.1$.

Fig. 5.14a shows another handwritten manuscript[6] of a different author. We select the enframed two-letter prototype word 'to' and use it to configure an *S*-COSFIRE filter. Taking into account the fact that here the prototype 'to' has higher shape complexity than the previous prototype of a single letter 'e', we configure an *S*-COSFIRE filter that receives input from six *V*-COSFIRE filters, rather than four. For this example, one *S*-COSFIRE filter ($t_1 = 0.3$, $t_2 = 0.75$, $t_3 = 0.77$, $\sigma_0 = 0.67$, and $\alpha = 0.1$) was sufficient to detect all the five words of interest, three of which are independent words and the other two are parts of other words, Fig. 5.14b.

In Fig. 5.15 we demonstrate an example of keyword spotting. We use the keyword prototype 'Germany' that is shown enframed in Fig. 5.15a to configure an *S*-COSFIRE filter that receives input from 13 *V*-COSFIRE filters, Fig. 5.15c. Fig. 5.15(d-e) show the responses of the concerned *S*-COSFIRE filter ($t_1 = 0.1$, $t_2 = 0.75$, $t_3 = 0.1$, $\sigma_0 = 0.67$, and $\alpha = 0.1$.) to the two manuscript images[7] in Fig. 5.15(a-b). It spots all the six instances of the keyword 'Germany' and does not produce any false positives.

The *S*-COSFIRE filters that are selective for specific words may correspond to neurons or networks of neurons in a certain area in the posterior lateral-occipital cortex. This area receives input from V4 and is selective for combinations of vertices. It has been shown to play a role in the recognition of words and has been named Visual Word Form Area (Szwed et al., 2011).

5.3.2 Vision for Home Tidying Pickup Robot

Daily service robots that perform routine tasks are becoming popular as household appliances. Such tedious tasks include, but are not limited to, vacuum cleaning, setting up and cleaning up a dinner table, tidying up toys, and organizing closets. The design of domestic robots is a growing research area (Takahama et al., 2004; Chen et al., 2008; Lee and Park, 2009; Saeedi et al., 2011; Jiang et al., 2012; Kim et al., 2012; Bandera et al., 2012).

Here, we demonstrate how the *S*-COSFIRE filters that we propose can be used by a personal robot to visually recognize objects of interest in indoor environments. As an illustration we consider a task for a tidying pickup robot to detect shoes in different rooms of a home that match the prototype shoe shown in Fig. 5.16a.

[6]This image is extracted from the file named a01-000u.png in the IAM offline database.

[7]These images in Fig. 5.15(a-b) are extracted from the files named b01-049.png and b01-044.png, respectively, in the IAM offline database.

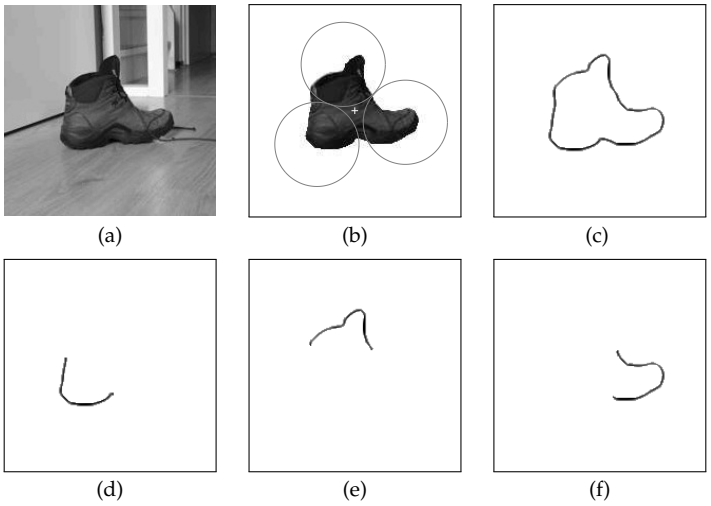


Figure 5.16: (a) A prototype shoe. Configuration of an *S*-COSFIRE filter using the manually segmented prototype shoe. The circles represent the non-overlapping supports of three *V*-COSFIRE filters, and the '+' marker indicates the center of support of the concerned *S*-COSFIRE filter. (c) The superimposed (inverted) thresholded responses ($t_1 = 0.3$) of a bank of Gabor energy filters with one wavelength ($\lambda = 4$) and 16 orientations in intervals of $\pi/8$. (d-f) Reconstructions of the local patterns for which the three resulting *V*-COSFIRE filters are selective. Each reconstruction is illustrated as a superposition of the (inverted) thresholded Gabor energy responses that contribute to the corresponding *V*-COSFIRE filter.

We use a segmented prototype image of the shoe to configure an *S*-COSFIRE filter, Fig. 5.16b. In a home application, this step can easily be automated by presenting the object of interest on a plain background. The concerned *S*-COSFIRE filter receives input from three *V*-COSFIRE filters that are selective for different parts of the shoe. These parts, which are represented by circles in Fig. 5.16b, are automatically chosen by the system from a circular local neighbourhood of a point of interest that is indicated by a '+' marker. In practice, the concerned point of interest and the radius of the corresponding local neighbourhood are manually specified by the user. The radii of the three circles are automatically computed in such a way that the circles touch each other. For the configuration of the concerned *V*-COSFIRE filters we use a bank of Gabor energy filters[8] with one wavelength ($\lambda = 4$) and 16 equidistant orientations ($\theta = \{\frac{\pi}{8}i \mid 0 \dots 15\}$), and we threshold the responses with $t_1 = 0.3$. Within each of the three circles, we consider a number of concentric circles, the radii of which increment in intervals of 4 pixels starting from 0. For the concerned three

[8]The response of a Gabor energy filter is computed as the L2-norm of the responses of a symmetric and anti-symmetric Gabor filters.

V-COSFIRE filters as well as the *S*-COSFIRE filter we use the same values of parameters α ($\alpha = 0.67$) and σ_0 ($\sigma_0 = 0.1$) in order to allow the same tolerance in the position of the involved edges and curvatures.

We created a data set that we call RUG-Shoes of 60 color images (of size 256×342 pixels) by taking pictures in different rooms of the same house. Of these images, 39 contain a pair of shoes of interest, another nine contain a single shoe and the remaining 12 do not contain any shoes. The distance above ground of the digital camera was varied between 50cm and 1m. All pictures of shoes were taken from the side view of the corresponding shoes. The shoes were, however, arranged in different orientations and their distances from the camera varied by at most 25% as compared to the distance which we used to take the image of the prototype shoe. We made the RUG-Shoes data set publicly available[9].

We use the configured *S*-COSFIRE filter to detect shoes in the entire data set of 60 images. For each color image, we first convert it to grayscale and subsequently apply the concerned *S*-COSFIRE filter in reflection-, scale- ($\upsilon \in \{\frac{3}{4}, 1, \frac{5}{4}\}$) and partially rotation-invariant ($\psi \in \{-\frac{\pi}{8}, 0, \frac{\pi}{8}\}$) mode. The Gabor energy filters that we use to provide inputs to the *V*-COSFIRE filters are applied with isotropic suppression (Grigorescu et al., 2004) in order to reduce responses to texture in these indoor scenes. We threshold the responses of the concerned *S*-COSFIRE filter with $t_3 = 0.1$ and for each image we consider only the highest two responses.

We obtain a perfect detection and recognition performance for all the 60 images in the RUG-Shoes data set. This means that we detect all the shoes in the given images with no false positives. Fig. 5.17 illustrates the detection of some shoes in four of the given images. Fig. 5.17b shows the detection of a single shoe that has a reflected arrangement – about the y-axis – with respect to the prototypical one. Fig. 5.17d and Fig. 5.17f each show the detection of a pair of shoes at different scales and at different orientations. Finally, Fig. 5.17g shows one of the 12 input images that do not contain any shoes. Rightly so, for these 12 images the *S*-COSFIRE filter does not respond.

[9]The RUG-Shoes data set can be downloaded from `http://matlabserver.cs.rug.nl/`

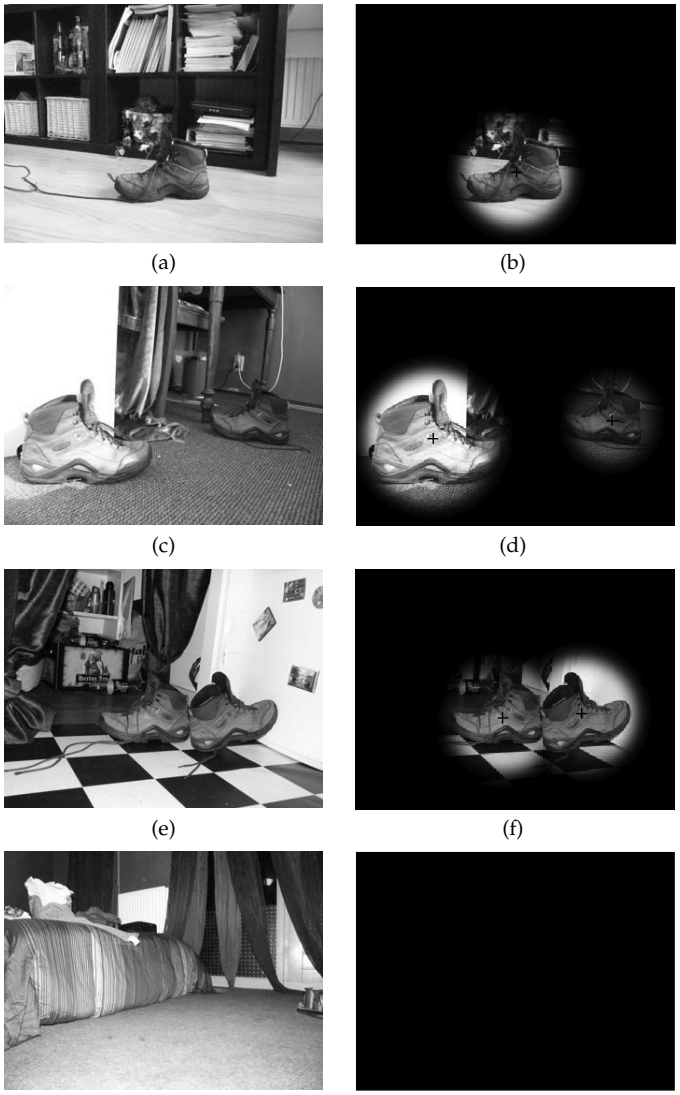


Figure 5.17: Detection of shoes in indoor scenes. (First column) Input images (of size 256×342 pixels) from the RUG-Shoes data set with filenames (a) Shoes03_1.jpg, (c) Shoes17_2.jpg, (e) Shoes58_2.jpg, and (g) Shoes38_1.jpg. (Second column) Results obtained for the corresponding images with a *S*-COSFIRE filter that is configured using the prototype shoe shown in Fig. 5.16a. Here we apply the *S*-COSFIRE filter in reflection-, scale- ($\upsilon \in \{\frac{3}{4}, 1, \frac{5}{4}\}$) and partially rotation-invariant ($\psi \in \{-\frac{\pi}{8}, 0, \frac{\pi}{8}\}$) mode. The '+' markers indicate the locations where the concerned *S*-COSFIRE filter achieves local maximum responses. The circular masks around the detected points correspond to the area of support of the "shoe"-selective *S*-COSFIRE filter. The radius of each mask is computed as follows. We first scale the maximum value of the parameter ρ by multiplying it with the υ value that contributed to the filter response at the concerned position. Then we sum the scaled ρ value and blur radius used at this scaled value of ρ.

5.4 Discussion

The trainable *S*-COSFIRE filters that we propose are part of a hierarchical object recognition approach that shares similarity with the ventral stream of visual cortex. In the first layer we detect lines and edges by Gabor filters, which are inspired by the function of orientation-selective cells in primary visual cortex (Daugman, 1985). Their responses are projected to a second layer and used by *V*-COSFIRE filters that detect vertices and curved contour segments. In our previous work (Azzopardi and Petkov, 2013), we showed that such filters give responses that are qualitatively similar to a class of cells in area V4 in visual cortex. Finally, in a third layer we have *S*-COSFIRE filters that combine the responses of certain *V*-COSFIRE filters. Such a filter is selective for a given spatial configuration of vertices and curved contour segments that defines a simple to moderately complex shape. *S*-COSFIRE filters share similar properties with shape-selective neurons in inferotemporal cortex, which provided inspiration for this work.

This hierarchical object recognition approach is, however, not restricted to three layers. The addition of further layers may be more appropriate for prototype objects of higher deformation complexity. For instance, let us consider a prototype shape of a simplistic human-body figure that is composed of a head, a pair of eyes, a nose, a mouth, two arms, two hands, a torso, two legs and two feet. We may configure a *S*-COSFIRE filter to be selective for the entire body with its center being at the center of mass of the body. Such a filter receives input from *V*-COSFIRE filters that are selective for distinct body parts. With this type of configuration the tolerance in the position of the body parts is computed with the same function that depends on the distance from the center of the *S*-COSFIRE filter. However, we know that certain body parts may require more tolerance or may be more correlated than others. For instance, the positions of the eyes, the nose and the mouth depend more on the position of the head than on the position of the legs. By taking this aspect in consideration it would be better to construct a hierarchical filter in the following way: configure a *S*-COSFIRE filter to be selective for the spatial arrangement of the head components (eyes, nose, and mouth), a *S*-COSFIRE filter for a hand and an arm, another one for a foot and a leg and a fourth one for the torso. Then, the responses of these four *S*-COSFIRE filters may be used as inputs to another, more complex *S*-COSFIRE filter.

A *S*-COSFIRE filter achieves a response when all parts of a shape of interest are present in a specific spatial arrangement around a given point in an image. The rigidity of this geometrical configuration may vary according to the application at hand. We use Gaussian functions to blur the responses of *V*-COSFIRE filters in order to allow for some tolerance in the position of the involved contour-based parts. The

standard deviation of such a Gaussian function depends on the distance from the center of the concerned *S*-COSFIRE filter: it grows linearly with a rate that is defined by the parameter α. Small values of α are more appropriate for the selectivity of rigid objects. Generalization ability increases with an increasing value of α. This mechanism is inspired by neurophysiological evidence that the average diameter of receptive fields of some neurons in visual cortex increases with the eccentricity (Gattass et al., 1988).

We applied the *S*-COSFIRE filters in reflection-, scale-, and partially rotation-invariant mode for the spotting of shoes embedded in complex scenes, while for the spotting of letters and keywords in handwritten manuscripts we did not allow any tolerance for such geometrical transformations.

The computational cost of the configuration of a *S*-COSFIRE filter is proportional to the number of costituent filters it uses. Here, we use *V*-COSFIRE filters of the type proposed in (Azzopardi and Petkov, 2013), to provide input to a *S*-COSFIRE filter. In general, the number of costituent filters needed grows with the shape complexity of the prototype object. In practice, for the parameter values that we used in the two applications, a *S*-COSFIRE filter is configured in less than half of a second for a Matlab implementation that runs on a 3GHz processor. Similarly, the computational cost of the application of a *S*-COSFIRE filter depends on the computations of the responses of the *V*-COSFIRE filters and their blurring and shifting operations. In practice, in the first application an image of a handwritten manuscript of size 3542×2479 pixels is processed in less than 30 seconds on a standard 3GHz processor with a non-invariant *S*-COSFIRE filter. For the second application, an image of an indoor scene of size 256×342 pixels is processed in less than five seconds by a reflection-, scale- and partially rotation-invariant *S*-COSFIRE filter.

The specific type of function that we use to combine the responses of costituent (*V*-COSFIRE) filters for the considered applications is a weighted geometric mean. This output function, which is also used to compute a *V*-COSFIRE filter response, proved to give better results than various forms of addition. Furthermore, there is psychophysical evidence that human visual processing of shape is likely performed by a nonlinear neural operation that multiplies afferent responses (Gheorghiu and Kingdom, 2009). In future work, we plan to experiment with functions other than (weighted) geometric mean.

The proposed *S*-COSFIRE filters are particularly useful due to their versatility and selectivity, in that a *S*-COSFIRE filter can be configured to be selective for any given deformable object and used to detect other objects embedded in complex scenes that are perceptually similar to it. This effectiveness is attributable to taking into account the mutual spatial positions of the responses of certain *V*-COSFIRE filters that are selective for simpler object parts. The *V*-COSFIRE filters which we in-

troduced in (Azzopardi and Petkov, 2013) are more appropriate for the detection of moderately complex – deformable and rigid – features or for the detection of highly-rigid objects. In that study we showed that the recognition of deformable objects by *V*-COSFIRE filters, such as handwritten digits, can be achieved by using a shape descriptor that is formed by the responses of multiple *V*-COSFIRE filters without taking in consideration the positions of these responses. While it is highly effective in segmented images, such a shape descriptor is not effective for the detection and recognition of deformable objects in complex scenes.

5.5 Conclusions

The *S*-COSFIRE filters that we propose are highly effective to detect and recognize deformable objects that are embedded in complex scenes without prior segmentation. This effectiveness is due to the deployment of both the presence of certain object-characteristic features and their mutual spatial arrangement. They are versatile shape detectors as they can be trained to be selective for any given visual pattern of interest.

A *S*-COSFIRE filter is conceptually simple and easy to implement: the filter output is computed as the weighted geometric mean of blurred and shifted responses of simpler *V*-COSFIRE filters.

Chapter 6

Summary and Outlook

6.1 Summary

This thesis proposes an innovative trainable filter approach to visual pattern recognition. It is inspired by neurophysiological evidence about the visual processing of contour, curvature and shape in the ventral stream of the brain.

The proposed approach is conceptually simple and easy to implement. The main idea is to configure a composite filter with an adaptive receptive field (or area of support) that combines the receptive fields of certain simpler filters at certain locations. For brevity, the modelling of receptive fields is referred to as CORF (Combination Of Receptive Fields). The afferent filters that provide input to a CORF model and the locations of their receptive fields are automatically determined in a single-step configuration phase that analyses the geometrical arrangement of the dominant responses of the involved simpler filters to a specified prototype pattern.

The output of a resulting composite filter in a given point is computed as the (weighted) geometric mean, essentially multiplication, of the responses of afferent filters at certain relative positions that are learned in the configuration step. Taking the responses of simpler filters at different locations around a point can be implemented by shifting the responses of the concerned filters by different vectors before using them for the pixelwise evaluation of the multivariate function. The implementation of a trainable composite filter is, therefore, called COSFIRE (Combination Of Shifted Filter Responses). Gaussian functions are used to blur the responses of costituent filters in order to allow for some tolerance in the relative positions of the involved input parts.

Due to the AND-type operation (weighted geometric mean) of a COSFIRE filter output a response is only achieved when all parts of the concerned prototype feature are present. This computation is mainly motivated by the better results obtained using multiplication rather than addition. It gets further support by psychophysical evidence (Gheorghiu and Kingdom, 2009) that some contour patterns are likely detected by a neural mechanism that multiplies the responses of afferent sub-units.

Chapter 2 shows that an orientation-selective filter can be configured by using

difference-of-Gaussians operators – that are models of LGN cells – as costituent filters with aligned center-surround receptive fields. The resulting filter is a CORF model of a simple cell in primary visual cortex. It is also referred to as contour-selective COSFIRE or *C*-COSFIRE filter. It is anatomically more realistic than the Gabor function model as it relies on model LGN input rather than using the intensity pixels of an image as they are projected on the retina. Moreover, it exhibits more properties than the Gabor function model that are typical of simple cells: contrast invariant orientation tuning and cross orientation suppression. By means of quantitative experiments it is shown that the proposed CORF model outperforms the Gabor function model in contour detection with high statistical significance.

This concept is extended further in Chapter 3, which shows that a trainable vertex detector can be configured by combining the responses of orientation-selective filters. Vertex-selective COSFIRE or *V*-COSFIRE filters are highly effective keypoint detectors. This ability is demonstrated in the application to the detection of vascular bifurcations in retinal fundus images in Chapter 4. Further experiments demonstrated that these filters are not limited to bifurcations only. The application to the detection and recognition of traffic signs in complex scenes in Chapter 3 is an example that *V*-COSFIRE filters are also effective detectors of complex but rather rigid patterns. Moreover, the application to the recognition of handwritten digits in Chapter 4 demonstrates that the responses of such filters can be used to form an effective descriptor for segmented or isolated patterns. *V*-COSFIRE filters are inspired by some shape-selective V4 neurons and have been demonstrated to reproduce responses that are similar to those of the concerned neurons.

Finally, in Chapter 5 it is shown that by using *V*-COSFIRE filters as afferent inputs, one can configure more complex filters that are selective for deformable shapes. These shape-selective COSFIRE or *S*-COSFIRE filters, are highly effective to recognize and localize (deformable) objects in images embedded in complex scenes without requiring prior segmentation. Their effectiveness is demonstrated in two applications: letter and keyword spotting in handwritten manuscripts and object spotting in complex scenes for the computer vision system of a domestic robot. *S*-COSFIRE filters are inspired by the properties of some TEO neurons.

The work proposed in this thesis contributes to the continuing trend of simulating biological vision to design more effective and robust computer vision solutions. The proposed techniques provide a foundation of an innovative trainable filter approach to visual pattern recognition. This approach is versatile as a filter can be configured with any given prototype pattern and is subsequently able to detect identical and similar ones.

6.2 Outlook

The following are ideas on directions how this work can be extended in future research.

The *C*-COSFIRE filter introduced in Chapter 2 is configured to be selective for a prototype edge and resulted in an effective contour detection operator. The prototype pattern is, however, not intrinsic to the method. If a bar structure is presented instead of an edge, a *C*-COSFIRE operator can be configured to be selective for elongated bar-like structures. A bar-selective COSFIRE or *B*-COSFIRE filter can then be applied, for instance, to detect blood vessels in medical images, such as retinal fundus and angiographic images.

There is neurophysiological evidence that simple cells receive strong inhibition with opposite polarity of their receptive fields (Hubel and Wiesel, 1962; Palmer, 1981; Borg-Graham et al., 1998). This is referred to as push-pull inhibition. One direction for future research is to incorporate this type of inhibition in the proposed model of a simple cell. It is expected that this enhancement will result in a *C*-COSFIRE filter that is more robust to contrast variations and more robust to noise.

The proposed trainable filter approach has been shown to be effective for the construction of filters by more complicated patterns other than simple edges. The recognition of segmented handwritten digits presented in Chapter 3 is one example of how *V*-COSFIRE filters can be effectively used to classify patterns that are isolated from the background. Other applications may include the recognition of handwritten characters of different alphabets, handwritten signatures, musical notes and sketched architectural symbols, which can, for instance, be applied to digital drawing tablets.

The *V*-COSFIRE filters that are proposed here take as input responses from orientation-selective filters that are only concerned with the contours of some patterns. The visual system of the brain, however, processes also color and texture information. One direction for future research is, therefore, to investigate the construction of *V*-COSFIRE filters to be tuned for patterns that are characterized by color or texture. Subsequently, the *S*-COSFIRE filter approach can be enriched into a multi-modal one by using various costituent *V*-COSFIRE filters that respond to contour, colour and texture. The resulting *S*-COSFIRE filter would be tuned to a pattern that is characterized by a combination of texture, color and contour-shape. As a matter of fact, this seems to be the function of some neurons in the anterior inferotemporal cortex (Tanaka et al., 1991). Such enhancements would widen the spectrum of applications where COSFIRE filters can be used.

Besides excitatory contour parts, some shape-selective neurons in area TEO of

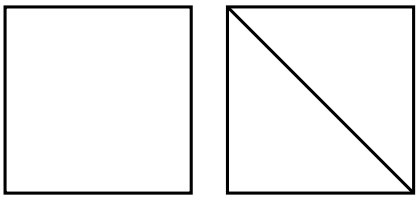


Figure 6.1: Example of two patterns that differ by a line segment. The diagonal line distinguishes the pattern in (b) from the one in (a).

visual cortex are also characterized by inhibitory contour parts (Brincat and Connor, 2004). This aspect instigates another direction for future investigation. In addition to contour parts which give excitatory input to a *S*-COSFIRE filter, other contour parts can be used to suppress the response of the filter. For instance, a *S*-COSFIRE filter that is configured to be selective for the square shown in Fig. 6.1a will also give a strong response to the pattern shown in Fig. 6.1b. By treating the diagonal line as an inhibitory contour part a *S*-COSFIRE filter can be constructed to be selective for squares without diagonal lines (of negative gradient) within them.

While feedforward connections are the main force of neuronal activity, there is neurophysiological evidence of feedback connections that mostly originate in areas V4, TEO and IT of visual cortex (Lamme and Roelfsema, 2000; Ungerleider et al., 2008). Such feedback processing are known to enhance the responses of the afferents input (Sillito et al., 2006). This is another concept which will be investigated for extending the properties of COSFIRE filters.

The effectiveness of the techniques that are proposed in this thesis was demonstrated in applications that involve 2D images. In further future work it will be studied how the proposed techniques can be extended by including the depth and time dimensions that will make them effective for the detection of patterns in 3D images and to track patterns in video sequences, respectively.

The application of the detection and recognition of traffic signs in Chapter 3 demonstrates the effectiveness of *V*-COSFIRE filters to detect rigid and complex patterns in natural scenes. In a similar way *V*- and *S*-COSFIRE filters may be used in other content-based image retrieval applications, such as biometric identification, medicine, crime prevention, historical research, among others.

The trainable COSFIRE approach that is presented here is not restricted to visual pattern recognition only. It may also be adopted to other signal processing applications. For instance, the similarity search in time series is an interesting topic with various applications such as discovering stocks with similar trends in financial markets or monitoring certain signal activity in sensor systems. In these signal process-

ing applications, methods such as piecewise linear regression and wavelet analysis, will be studied as costituent operators that provide input to COSFIRE filters.

As to computational efficiency, the proposed filters are implemented in a sequential programming scheme. The implementation of the filters can, however, be in a parallel or distributed mode as most of the computations are independent of each other. In future work, an implementation in parallel computing platforms, such as CUDA (NVIDIA, 2012), will be investigated. Parallel implementation will make the proposed filters effective for real-time applications.

Fig. 6.2 summarizes the above mentioned directions for future research to develop further the trainable COSFIRE filter approach that is proposed in this thesis.

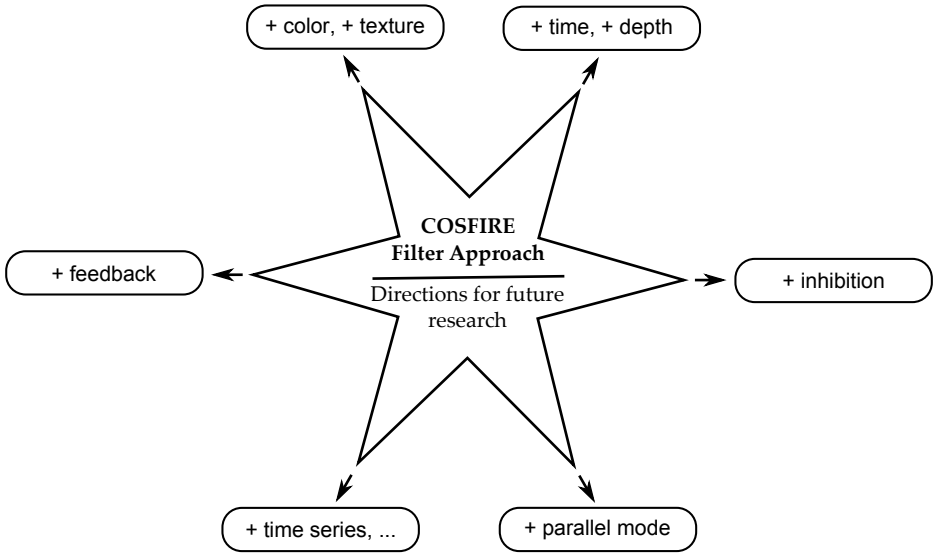


Figure 6.2: A schematic overview that illustrates six directions to develop further the proposed COSFIRE filter approach. Research directions include various aspects: addition of color and texture information; addition of time and depth properties, inhibitory pattern parts; feedback processing; adaptation of the proposed approach to other signal processing applications such as time series; and an implementation in parallel mode to make the technique suitable for real-time applications.

Samenvatting

Dit proefschrift draagt een innovatieve trainbare filterbenadering voor visuele patroonherkenning aan. De benadering is geïnspireerd door neurofysiologisch bewijs omtrent de visuele verwerking van contouren, krommingen en vormen in de ventrale route van het brein.

De voorgestelde benadering is conceptueel simpel en eenvoudig te implementeren. De hoofdgedachte is om een samengesteld filter te configureren met een adaptief receptief veld (of ondersteuningsgebied) dat de receptieve velden van bepaalde eenvoudige filters combineert op bepaalde locaties. Ten behoeve van beknoptheid wordt het modelleren van receptieve velden aangeduid met CORF (Combination Of Receptive Fields). De afferente filters die een CORF-model van invoer, en de locaties van hun receptieve velden voorzien, worden automatisch bepaald in een èènstapsconfiguratiefase die de geometrische inrichting van de dominante uitvoer van de betrokken simpelere filters analyseert tot een gespecificeerd prototype-patroon.

De uitvoer van een dergelijk samengesteld filter in een bepaald punt wordt berekend als het (gewogen) meetkundig gemiddelde, een feitelijke vermenigvuldiging, van de uitvoer van afferente filters op bepaalde relatieve posities die geleerd worden in de configuratiestap. Het bepalen van de uitvoer van simpelere filters op verschillende locaties rondom een punt kan geïmplementeerd worden door de uitvoer van de betrokken filters te verschuiven met verschillende vectoren alvorens ze te gebruiken voor het pixelwijs evalueren van de multivariabele functie. De implementatie van een trainbaar samengesteld filter wordt daarom een COSFIRE (Combination Of Shifted Filter Responses) genoemd. Gaussische functies worden gebruikt om de uitvoer van de inputfilters te vervagen om enige tolerantie in de relatieve posities van de betrokken invoerdelen toe te staan.

Door de AND-type bewerking (gewogen meetkundig gemiddelde) van een

COSFIRE-filteruitvoer wordt een niet-nul waarde alleen bereikt wanneer alle onderdelen van de desbetreffende prototypefunctie aanwezig zijn. Deze computatie is voornamelijk gemotiveerd door de betere resultaten die verkregen werden met het gebruik van vermenigvuldigingen ten opzichte van optellingen. Het wordt verder ondersteund door psychofysisch bewijs dat sommige contourpatronen waarschijnlijk worden gedetecteerd door een neuraal mechanisme dat de reacties van afferente sub-eenheden vermenigvuldigt.

Hoofdstuk 2 toont aan dat een oriëntatie-selectief filter geconfigureerd kan worden door difference-of-Gaussians operatoren, hetgeen modellen van LGN-cellen zijn, als invoerfilters met uitgelijnde center-surround receptieve velden. Het resulterende filter is een CORF-model van een simpele cel in de primaire visuele cortex, ook wel aangeduid als een contourselectief COSFIRE- of *C*-COSFIRE-filter. Het is anatomisch realistischer dan het Gabor-functiemodel gezien deze op model-LGN-invoer is gebaseerd in plaats van gebruik te maken van de intensiteitspixels van een afbeelding wanneer deze op de retina geprojecteerd worden. Bovendien vertoont het meer eigenschappen die kenmerkend zijn voor eenvoudige cellen dan het Gabor-functiemodel: contrast invariant orientation tuning en cross orientation suppression. Door middel van kwantitatieve experimenten wordt aangetoond dat het voorgestelde CORF-model het Gabor-functiemodel bij contourdetectie overtreft met hoge statistische significantie.

Dit concept wordt verder uitgebreid in hoofdstuk 3, waar aangetoond wordt dat een trainbare vertex-detector kan worden geconfigureerd door de uitvoer van oriëntatie-selectieve filters te combineren. Vertex-selectieve COSFIRE- of *V*-COSFIRE-filters zijn zeer effectieve keypoint-detectoren. Dit vermogen wordt gedemonstreerd door het filter in te zetten om vasculaire bifurcaties in retinale fundusafbeeldingen te detecteren in hoofdstuk 4. Aanvullende experimenten tonen aan dat deze filters niet gelimiteerd zijn tot bifurcaties alleen. De toepassing om verkeersborden te herkennen en te detecteren in complexe scènes in hoofdstuk 3 is een voorbeeld dat *V*-COSFIRE-filters ook effectieve detectoren zijn van complexe maar betrekkelijk vaste patronen. Bovendien geeft de toepassing om handgeschreven getallen te herkennen in hoofdstuk 4 aan dat de uitvoer van dergelijke filters gebruikt kan worden om een effectieve descriptor van gesegmenteerde of geïsoleerde patronen te vormen. *V*-COSFIRE-filters zijn geïnspireerd door enkele vorm-selectieve V4-neuronen en het is aangetoond dat ze uitvoer reproduceren die soortgelijk is aan de uitvoer van de neuronen in kwestie.

Tenslotte wordt in hoofdstuk 5 aangetoond dat, door *V*-COSFIRE-filters te gebruiken als afferente invoer, men complexere filters kan configureren die selectief zijn aangaande deformeerbare vormen. Deze shape-selectieve COSFIRE- of *S*-COSFIRE-filters zijn hoogst effectief om (deformeerbare) objecten ingebed in

complexe scènes te herkennen en te lokaliseren zonder voorafgaande segmentatie te vereisen. De doeltreffendheid wordt aangetoond door middel van twee toepassingen: letter- en trefwoordherkenning in handgeschreven manuscripten en objectherkenning in complexe scènes voor het computervisiesysteem van een huishoudelijke robot. *S*-COSFIRE filters zijn geïnspireerd door de eigenschappen van enkele TEO-neuronen.

Het werk dat in dit proefschrift voorgesteld wordt draagt bij aan de aanhoudende trend het biologisch gezichtsvermogen te simuleren om effectievere en robuustere computervisiesystemen te ontwerpen. De voorgestelde technieken verschaffen de basis voor een innovatieve trainbare filterbenadering voor visuele patroonherkenning. Deze benadering is veelzijdig gezien een filter met ieder gegeven prototype-patroon geconfigureerd kan worden en vervolgens in staat is identieke en vergelijkbare patronen te detecteren.

Bibliography

Abramoff, M., Garvin, M. and Sonka, M.: 2010, Retinal imaging and image analysis, *Biomedical Engineering, IEEE Reviews in* **3**, 169–208.

Aibinu, A. M., Iqbal, M. I., Shafie, A. A., Salami, M. J. E. and Nilsson, M.: 2010, Vascular intersection detection in retina fundus images using a new hybrid approach, *Computers in Biology and Medicine* **40**(1), 81–89.

Albrecht, D. G., De Valois, R. L. and Thorell, L. G.: 1980, Visual cortical-neurons - are bars or gratings the optimal stimuli, *Science* **207**(4426), 88–90.

Ali, C., Hong, S., Turner, J., Tanenbaum, H. and Roysam, B.: 1999, Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms, *IEEE Transactions on Information Technology in Biomedicine* **3**, 125–138.

Almazan, J., Fornes, A. and Valveny, E.: 2012, A non-rigid appearance model for shape description and recognition, *Pattern Recognition* **45**(9, SI), 3105–3113.

Andrews, B. W. and Pollen, D. A.: 1979, Relationship between spatial-frequency selectivity and receptive-field profile of simple cells, *Journal of Physiology-London* **287**(Feb), 163–176.

Ardizzone, E., Pirrone, R., Gambino, O. and Radosta, S.: 2008, Blood vessels and feature points detection on retinal images, *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pp. 2246 –2249.

Azzopardi, G. and Petkov, N.: 2011, Detection of Retinal Vascular Bifurcations by Trainable V4-Like Filters, *in* P. Real, D. Diaz-Pernil, H. Molina-Abril, A. Berciano and W. Kropatsch (eds), *Computer Analysis of Images and Patterns. Proceedings 14th International Conference, CAIP 2011*, pp. 451–9. Computer Analysis of Images and Patterns. 14th International Conference, CAIP 2011, 29-31 Aug. 2011, Seville, Spain.

Azzopardi, G. and Petkov, N.: 2012, A CORF computational model of a simple cell that relies on LGN input outperforms the Gabor function model, *Biological Cybernetics* **106**(3), 177–189.

Azzopardi, G. and Petkov, N.: 2013, Trainable COSFIRE Filters for Keypoint Detection and Pattern Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**(2), 490 –503.

Bandera, J. P., Rodriguez, J. A., Molina-Tanco, L. and Bandera, A.: 2012, A survey of vision-based architectures for robot learning by imitation, *International Journal of Humanoid Robotics* **9**(1).

Baumberg, A.: 2000, Reliable feature matching across widely separated views, *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, number vol.1, pp. 774–781.

Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L.: 2008, Speeded-Up Robust Features (SURF), *Computer Vision and Image Understanding* **110**(3), 346–359.

Becker, D., Can, A., Turner, J., Tanenbaum, H. and Roysam, B.: 1998, Image processing algorithms for retinal montage synthesis, mapping, and real-time location determination, *IEEE Transactions on Biomedical Engineering* **45**(1), 105–118.

Belongie, S., Malik, J. and Puzicha, J.: 2002, Shape matching and object recognition using shape contexts, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(4), 509–522.

Bevilacqua, V., Cambo, S., Cariello, L. and Mastronardi, G.: 2005, A combined method to detect retinal fundus features", *Proceedings of European Conference on Emergent Aspects of Clinical Data Analysis*, pp. 1–6.

Bhuiyan, A., Nath, B., Chua, J. and Ramamohanarao, K.: 2007, Automatic detection of vascular bifurcations and crossovers from color retinal fundus images, *2007 Third International IEEE Conference on Signal-Image Technologies and Internet-Based System (SITIS)*, pp. 711–18. 2007 Third International IEEE Conference on Signal-Image Technologies and Internet-Based System (SITIS), ,, Shanghai, China.

Biederman, I.: 1987, Recognition-by-components - a theory of human image understanding, *Psychological Review* **94**(2), 115–147.

Bodis-Wollner, I. and Brannan, J.: 1997, Hidden visual loss in optic neuropathy is revealed using Gabor patch contrast perimetry, *Clinical Neuroscience* **4**(5), 284–291.

Borg-Graham, L., Monier, C. and Fregnac, Y.: 1998, Visual input evokes transient and strong shunting inhibition in visual cortical neurons, *Nature* **393**(6683), 369–373.

Borji, A., Hamidi, M. and Mahmoudi, F.: 2008, Robust handwritten character recognition with features inspired by visual ventral stream, *Neural Processing Letters* **28**(2), 97–111.

Brincat, S. L. and Connor, C. E.: 2004, Underlying principles of visual shape selectivity in posterior inferotemporal cortex, *Nature Neuroscience* **7**(8), 880–886.

Bunte, K., Biehl, M., Jonkman, M. F. and Petkov, N.: 2011, Learning effective color features for content based image retrieval in dermatology, *Pattern Recognition* **44**(9, SI), 1892–1902.

Cadieu, C., Kouh, M., Pasupathy, A., Connor, C. E., Riesenhuber, M. and Poggio, T.: 2007, A model of v4 shape selectivity and invariance, *Journal of Neurophysiology* **98**(3), 1733–1750.

Calvo, D., Ortega, M., Penedo, M. G. and Rouco, J.: 2011, Automatic detection and characterisation of retinal vessel tree bifurcations and crossovers in eye fundus images, *Computer methods and programs in biomedicine* **103**(1), 28–38.

Canny, J.: 1986, A computational approach to edge-detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**(6), 679–698.

Carneiro, G. and Jepson, A.: 2003, Multi-scale phase-based local features, *Proceedings 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, number vol.1, pp. I–736–743.

Chanwimaluang, T. and Guoliang, F.: 2003, An efficient blood vessel detection algorithm for retinal images using local entropy thresholding, *Proceedings of the 2003 IEEE International Symposium on Circuits and Systems (Cat. No.03CH37430)* **5**, 21–25.

Chapman, N., Dell'omo, G., Sartini, M. S., Witt, N., Hughes, A., Thom, S. and Pedrinelli, R.: 2002, Peripheral vascular disease is associated with abnormal arteriolar diameter relationships at bifurcations in the human retina, *Clinical Science* **103**(2), 111–116.

Chauduri, S., Chatterjee, S., Katz, N., Nelson, M. and Goldbaum, M.: 1989, Detection of blod-vessels in retinal images using tow-dimensional matched-filters, *IEEE Transactions on Medical Imaging* **8**(3), 263–269.

Chen, C., Li, H.-X. and Dong, D.: 2008, Hybrid control for robot navigation - A hierarchical Q-learning algorithm, *IEEE Robotics & Automation Magazine* **15**(2), 37–47.

Chung, S. and Ferster, D.: 1998, Strength and orientation tuning of the thalamic input to simple cells revealed by electrically evoked cortical suppression, *Neuron* **20**(6), 1177–1189.

Croner, L. J. and Kaplan, E.: 1995, Receptive-fields of P-ganglion and M-ganglion cells across the primate retina, *Vision Research* **35**(1), 7–24.

Daugman, J.: 1985, Uncertainty relation for resolution in space, spatial-frequency, and orientation optimized by two dimensionsl visual cortical filters, *journal of the Optical Society of America A-Optics Image Science and Vision* **2**(7), 1160–1169.

de Boer, R. and Kuyper, P.: 1968, Triggered correlation, *IEEE transactions on bio-medical engineering* **15**(3), 169–79.

De Valois, K. K., De Valois, R. L. and Yund, E. W.: 1979, Responses of striate cortex cells to grating and checkerboard patterns, *J.Physiol.(Lond.)* **291**(Jun), 483–505.

De Valois, R. L., Albrecht, D. G. and Thorell, L. G.: 1978, Cortical cells: bar and edge detectors, or spatial frequency filters?, *in* S. J. Cool and I. Smith, E. L. (eds), *Frontiers in Visual Science*, Springer-Verlag, Berlin, West Germany, pp. 544–56. Frontiers in Visual Science, March 1977, Houston, TX, USA.

De Valois, R. L., Yund, E. W. and Hepler, N.: 1982, The orientation and direction selectivity of cells in macaque visual-cortex, *Vision Research* **22**(5), 531–544.

DeAngelis, G. C., Ohzawa, I. and Freeman, R. D.: 1995, Receptive-field dynamics in the central visual pathways, *Trends in Neurosciences* **18**(10), 451–458.

Demirci, M. F., Shokoufandeh, A., Keselman, Y., Bretzner, L. and Dickinson, S.: 2006, Object recognition as many-to-many feature matching, *International Journal of Computer Vision* **69**(2), 203–222.

DuBuf, J. M. H.: 1993, Responses of simple cells - events, interferences, and ambiguities, *Biological Cybernetics* **68**(4), 321–333.

Edelman, S. and Intrator, N.: 2003, Towards structural systematicity in distributed, statically bound visual representations, *Cognitive Science* **27**(1), 73–109.

Fei-Fei, L., Fergus, R. and Perona, P.: 2007, Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories, *Computer Vision and Image Understanding* **106**(1), 59–70. 2nd International Workshop on Generative-Model Based Vision, Washington, DC, 2005.

Fergus, R., Perona, P. and Zisserman, A.: 2003, Object class recognition by unsupervised scale-invariant learning, *Proceedings 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Comput. Soc Tech. Committee on Pattern Anal. & Machine Intelligence (TCPAMI), pp. II–264–71 vol.2. CVPR 2003: Computer Vision and Pattern Recognition Conference, 18-20 June 2003, Madison, WI, USA.

Ferster, D., Chung, S. and Wheat, H.: 1996, Orientation selectivity of thalamic input to simple cells of cat visual cortex, *Nature* **380**(6571), 249–252.

Fidler, S., Boben, M. and Leonardis, A.: 2008, Similarity-based cross-layered hierarchical representation for object categorization, *2008 Ieee Conference on Computer Vision and Pattern Recognition, Vols 1-12* pp. 525–532.

Fidler, S. and Leonardis, A.: 2007, Towards scalable representations of object categories: Learning a hierarchy of parts, *2007 IEEE Conference on Computer Vision and Pattern Recognition, Vols 1-8* pp. 2295–2302.

Finn, I. M., Priebe, N. J. and Ferster, D.: 2007, The emergence of contrast-invariant orientation tuning in simple cells of cat visual cortex, *Neuron* **54**(1), 137–152.

Florack, L., ter Haar Romeny, B., Koenderink, J. and Viergever, M.: 1994, General intensity transformations and differential invariants, *Journal of Mathematical Imaging and Vision* pp. 171–87.

Frank, R.: 2004, Diabetic retinopathy, *New England Journal of Medicine* **350**(1), 48–58.

Fraz, M., Remagnino, P., Hoppe, A., Velastin, S., Uyyanonvara, B. and Barman, S.: 2011, A supervised method for retinal blood vessel segmentation using line strength, multiscale Gabor and morphological features, *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA 2011)*, pp. 410–15. 2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA 2011), 16-18 Nov. 2011, Kuala Lumpur, Malaysia.

Freeman, W. and Adelson, E.: 1991, The design and use of steerable filters, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**(9), 891–906.

Frinken, V., Fischer, A., Manmatha, R. and Bunke, H.: 2012, A Novel Word Spotting Method Based on Recurrent Neural Networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(2), 211–224.

Gabor, D.: 1946, Theory of communication, *Journal of the Institution of Electrical Engineers* **93**, 429–457.

Gatos, B. and Pratikakis, I.: 2009, Segmentation-free word spotting in historical printed documents, *2009 10th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 271–5. 2009 10th International Conference on Document Analysis and Recognition (ICDAR), 26-29 July 2009, Barcelona, Spain.

Gattass, R., Sousa, A. P. and Gross, C. G.: 1988, Visuotopic organization and extent of v3 and v4 of the macaque, *Journal of Neuroscience* **8**(6), 1831–1845.

Gheorghiu, E. and Kingdom, F. A. A.: 2009, Multiplication in curvature processing, *Journal of Vision* **9**(2).

Ghosh, A. and Petkov, N.: 2005, Robustness of shape descriptors to incomplete contour representations, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **27**(11), 1793–1804.

Glezer, V. D., Tsherbach, T. A., Gauselman, V. E. and Bondarko, V. M.: 1980, Linear and nonlinear properties of simple and complex receptive-fields in area-17 of the cat visual-cortex - a model of the field, *Biological Cybernetics* **37**(4), 195–208.

Goh, W.-B.: 2008, Strategies for shape matching using skeletons, *Computer Vision and Image Understanding* **110**(3), 326–345.

Goodale, M. A. and Milner, A. D.: 1992, Separate visual pathways for perception and action, *Trends in Neuroscience* **15**(1), 20–25.

Grigorescu, C., Petkov, N. and Westenberg, M. A.: 2003a, Contour detection based on nonclassical receptive field inhibition, *IEEE Transactions on Image Processing* **12**(7), 729–739.

Grigorescu, C., Petkov, N. and Westenberg, M. A.: 2003b, Contour detection based on nonclassical receptive field inhibition, *IEEE Transactions on Image Processing* **12**(7), 729–739.

Grigorescu, C., Petkov, N. and Westenberg, M. A.: 2003c, The role of non-crf inhibition in contour detection, *Journal of Computer Graphics, Visualization, and Computer Vision (WSCG)* **11**(2), 197–204.

Grigorescu, C., Petkov, N. and Westenberg, M. A.: 2004, Contour and boundary detection improved by surround suppression of texture edges, *Image and Vision Computing* **22**(8), 609–622.

Grigorescu, S. E., Petkov, N. and Kruizinga, P.: 2002, Comparison of texture features based on gabor filters, *IEEE Transactions on Image Processing* **11**(10), 1160–1167.

Hamidi, M. and Borji, A.: 2010, Invariance analysis of modified C2 features: case study-handwritten digit recognition, *Machine Vision and Applications* **21**(6), 969–979.

Hammer, B. and Villmann, T.: 2002, Generalized relevance learning vector quantization, *Neural Networks* **15**(8-9), 1059–1068.

Hanazawa, A. and Komatsu, H.: 2001, Influence of the direction of elemental luminance gradients on the responses of v4 cells to textured surfaces, *Journal of Neuroscience* **21**(12), 4490–4497.

Harris, C. and Stephens, M.: 1988, A combined corner and edge detector, *Proc. Fourth Alvey Vision Conference*, pp. 147–151.

Hegde, J. and Van Essen, D. C.: 2007, A comparative study of shape representation in macaque visual areas V2 and V4, *Cerebral Cortex* **17**(5), 1100–1116.

Heitger, F.: 1995, Feature detection using suppression and enhancement, Communication Technology Laboratory, Swiss Federal Inst. Technol., Lausanne, Tech. Rep. TR-163.

Hoover, A., Kouznetsova, V. and Goldbaum, M.: 2000, Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response, *IEEE Transactions on Medical Imaging* **19**(3), 203–210.

Hubel, D. H.: 1982, Exploration of the primary visual-cortex, 1955-78, *Nature* **299**(5883), 515–524.

Hubel, D. H. and Wiesel, T. N.: 1962, Receptive fields, binocular interaction and functional architecture in cats visual cortex, *J.Physiol.(Lond.)* **160**(1), 106–154.

Hubel, D. H. and Wiesel, T. N.: 1974, Sequence regularity and geometry of orientation columns in monkey striate cortex, *Journal of Comparative Neurology* **158**(3), 267–294.

Irvin, G. E., Casagrande, V. A. and Norton, T. T.: 1993, Center surround relationships of magnocellular, parvocellular, and koniocellular relay cells in primate lateral geniculate nucleus, *Visual Neuroscience* **10**(2), 363–373.

Jiang, X. and Mojon, D.: 2003, Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(1), 131–137.

Jiang, Y., Lim, M., Zheng, C. and Saxena, A.: 2012, Learning to place new objects in a scene, *InternationalJournal of Robotics Research* **31**(9), 1021–1043.

Jones, J. P. and Palmer, L. A.: 1987, An evaluation of the two-dimensional gabor filter model of simple receptive-fields in cat striate cortex, *Journal of Neurophysiology* **58**(6), 1233–1258.

Jung, E. and Hong, K.: 2006, Automatic retinal vasculature structure tracing and vascular landmark extraction from human eye image, *in* Lee, G and Slezak, D and Kim, TH and Sloot, P and Kim, HK and Szczuka, M and Ko, IS (ed.), *2006 International Conference on Hybrid Information Technology, Vol 2, Proceedings*, pp. 161–167.

Kanizsa, G. and Gerbino, W.: 1976, Convexity and symmetry in figure-ground organization, *In M. Henle (Ed.), Vision and artifact*, Springer, pp. 25–32.

Kim, S., Sim, J.-Y. and Yang, S.: 2012, Vision-based Cleaning Area Control for Cleaning Robots, *IEEE Transactions on Consumer Electronics* **58**(2), 685–690.

Klement, S. and Martinetz, T.: 2010, The support feature machine for classifying with the least number of features, *Proceedings of the 20th international conference on Artificial neural networks: Part II*, ICANN'10, pp. 88–93.

Kobatake, E. and Tanaka, K.: 1994, Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral-cortex, *Journal of Neurophysiology* **71**(3), 856–867.

Kovesi, P.: 1999, Image features from phase congruency, *Videre—Videre* **1**(3).

Kruizinga, P. and Petkov, N.: 1999, Non-linear operator for oriented texture, *IEEE Trans. on Image Processing* **8**(10), 1395–1407.

Kulikowski, J. J. and Bishop, P. O.: 1981, Fourier-analysis and spatial representation in the visual-cortex, *Experientia* **37**(2), 160–163.

Lamme, V. and Roelfsema, P.: 2000, The distinct modes of vision offered by feedforward and recurrent processing, *Trends in NeuroScience* **23**(11), 571–579.

Latecki, L., Lakaemper, R. and Wolter, D.: 2005, Optimal partial shape similarity, *Image and Vision Computing* **23**(2), 227–236. 11th International Conference on Discrete Geometry for Computer Imagery, Italian Inst Philosoph Studies, Naples, Italy, Nov 19-21, 2003.

Lauer, F., Suen, C. Y. and Bloch, G.: 2007, A trainable feature extractor for handwritten digit recognition, *Pattern Recognition* **40**(6), 1816–1824.

LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P.: 1998, Gradient-based learning applied to document recognition, *Proceedings of the IEEE* **86**(11), 2278–2324.

Lee, D. and Park, Y.: 2009, Vision-Based Remote Control System by Motion Detection and Open Finger Counting, *IEEE Transactions on Consumer Electronics* **55**(4), 2308–2313.

Leydier, Y., Le Bourgeois, F. and Emptoz, H.: 2005, Omnilingual segmentation-free word spotting for ancient manuscripts indexation, *Proceedings. Eighth International Conference on Document Analysis and Recognition*, pp. 533–7 Vol. 1. Proceedings. Eighth International Conference on Document Analysis and Recognition, 29 Aug.-1 Sept. 2005, Seoul, South Korea.

Li Chen, Yang Xiang, YaoJie Chen and XiaoLong Zhang: 2011, Retinal image registration using bifurcation structures, *2011 18th IEEE International Conference on Image Processing (ICIP 2011)*, IEEE Signal Process. Soc., pp. 2169–72. 2011 18th IEEE International Conference on Image Processing (ICIP 2011), 11-14 Sept. 2011, Brussels, Belgium.

Lindeberg, T.: 1998, Feature detection with automatic scale selection, *International Journal of Computer Vision* **30**(2), 79–116.

Ling, H. and Jacobs, D. W.: 2007, Shape classification using the inner-distance, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(2), 286–299.

Liu, C., Nakashima, K., Sako, H. and Fujisawa, H.: 2003, Handwritten digit recognition: benchmarking of state-of-the-art techniques, *Pattern Recognition* **36**(10), 2271–2285.

Lowe, D.: 2004, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision* **60**, 91–110.

Lowe, D. G.: 1999, Object recognition from local scale-invariant features, *Proceedings of the Seventh IEEE International Conference on Computer Vision*, number vol.2, pp. 1150–7 vol.2.

Macleod, I. D. G. and Rosenfeld, A.: 1974, Visibility of gratings - spatial frequency channels or bar-detecting units, *Vision Research* **14**(10), 909–915.

Maffei, L., Morrone, C., Pirchio, M. and Sandini, G.: 1979, Responses of visual cortical-cells to periodic and non-periodic stimuli, *Journal of Physiology-London* **296**(Nov), 27–47.

Marcelja, S.: 1980, Mathematical-description of the responses of simple cortical-cells, *Journal of the Optical Society of America* **70**(11), 1297–1300.

Marr, D.: 1982, *Vision: A computational investigation into the human respresentation and processing of visual information*, New York: Freeman.

Marr, D. and Nishihara, H. K.: 1978, Representation and recognition of spatial-organization of 3-dimensional shapes, *Proceedings of the Royal Society of London Series B-Biological Sciences* **200**(1140), 269–294.

Marti, U.-V. and Bunke, H.: 2002, The IAM-database: an English sentence database for offline handwriting recognition, *International Journal on Document Analysis and Recognition* **5**, 39–46.

Martin, D., Fowlkes, C., Tal, D. and Malik, J.: 2001, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, *Proc. 8th Int'l Conf. Computer Vision*, Vol. 2, pp. 416–423.

Martin, D. R., Fowlkes, C. C. and Malik, J.: 2004, Learning to detect natural image boundaries using local brightness, color, and texture cues, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(5), 530–549.

Martinez-Perez, M., Hughes, A., Stanton, A., Thom, S., Chapman, N., Bharath, A. and Parker, K.: 2002, Retinal vascular tree morphology: A semi-automatic quantification, *IEEE Transactions on Biomedical Engineering* **49**, 912–917.

Mehrotra, R., Namuduri, K. R. and Ranganathan, N.: 1992, Gabor filter-based edge-detection, *Pattern Recognition* **25**(12), 1479–1494.

Mel, B. W. and Fiser, J.: 2000, Minimizing binding errors using learned conjunctive features (vol 12, pg 247, 1999), *Neural Computation* **12**(4), 731–762.

Mendonca, A. M. and Campilho, A.: 2006, Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction, *IEEE Transactions on Medical Imaging* **25**(9), 1200–1213.

Mikolajczyk, K. and Schmid, C.: 2001, Indexing based on scale invariant interest points, *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, number vol.1, pp. 525–531.

Mikolajczyk, K. and Schmid, C.: 2005, A performance evaluation of local descriptors, *IEEE Transactions on Pattern Analysis & Machine Intelligence* **27**(10), 1615–1630.

Mindru, F., Tuytelaars, T., Van Gool, L. and Moons, T.: 2004, Moment invariants for recognition under changing viewpoint and illumination, *Computer Vision and Understanding* **94**(1-3), 3–27.

Moghaddam, R. and Cheriet, M.: 2009, Application of multi-level classifiers and clustering for automatic word spotting in historical document images, *2009 10th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 511–15. 2009 10th International Conference on Document Analysis and Recognition (ICDAR), 26-29 July 2009, Barcelona, Spain.

Moin, M., Tavakoli, H. and Broumandnia, A.: 2010, A new retinal vessel segmentation method using preprocessed Gabor and local binary patterns, *2010 6th Iranian Conference on Machine Vision and Image Processing (MVIP)*, IEEE Iran Sect., p. 6 pp. 2010 6th Iranian Conference on Machine Vision and Image Processing (MVIP), 27-28 Oct. 2010, Isfahan, Iran.

Morrone, M. C. and Burr, D. C.: 1988, Feature detection in human-vision - a phase-dependent energy-model, *Proceedings of the Royal Society of London Series B-Biological Sciences* **235**(1280), 221–.

Morrone, M. C. and Owens, R. A.: 1987, Feature detection from local energy, *Pattern Recognition Letters* **6**(5), 303–313.

Movshon, J. A., Thompson, I. D. and Tolhurst, D. J.: 1978, Receptive-field organization of complex cells in cats striate cortex, *Journal of Physiology-London* **283**(Oct), 79–99.

Muramatsu, C., Hayashi, Y., Sawada, A., Hatanaka, Y., Hara, T., Yamamoto, T. and Fujita, H.: 2009, Computerized Detection of Retinal Nerve Fiber Layer Defects in Retinal Fundus Images by Modified Polar Transformation and Gabor Filtering, *in* O. Dossel and W. Schlegel (eds), *11th International Congress of the IUPESM. Medical Physics and Biomedical Engineering. World Congress 2009. Biomedical Engineering for Audiology, Ophthalmology, Emergency & Dental Medicine*, pp. 124–6. 11th International Congress of the IUPESM. Medical Physics and Biomedical Engineering. World Congress 2009. Biomedical Engineering for Audiology, Ophthalmology, Emergency & Dental Medicine, 7-12 Sept. 2009, Munich, Germany.

Murray, C. D.: 1926a, The physiological principle of minimum work applied to the angle of branching of arteries, **9**, 835–841.

Murray, C. D.: 1926b, The physiological principle of minimum work: I. the vascular system and the cost of blood volume, *Proc Natl Acad Sci U S A*, Vol. 12, pp. 207–214.

Niemeijer, M., Staal, J., van Ginneken, B., Loog, M. and Abramoff, M.: 2004, Comparative study of retinal vessel segmentation methods on a new publicly available database, *Proceedings of the SPIE - The International Society for Optical Engineering*, pp. 648–56. Medical Imaging 2004. Image Processing, 16-19 Feb. 2004, San Diego, CA, USA.

NVIDIA: 2012, Cuda parallel computing platform.
URL: `http://www.nvidia.com/object/cuda_home_new.html`

Oberhoff, D. and Kolesnik, M.: 2008, Unsupervised shape learning in a neuromorphic hierarchy, *Pattern Recognition and Image Analysis* pp. 314–22.

Ortega, M., Penedo, M. G., Rouco, J., Barreira, N. and Carreira, M. J.: 2009, Personal verification based on extraction and characterisation of retinal feature points, *Journal of visual languages and computing* **20**(2), 80–90.

Palmer, LA, D. T.: 1981, Receptive-Field Structure in cat striate cortex, *Journal of Neurophysiology* **46**(2), 260–276.

Parisi, V., Manni, G., Spadaro, M., Colacino, G., Restuccia, R., Marchi, S., Bucci, M. and Pierelli, F.: 1999, Correlation between morphological and functional retinal impairment in multiple sclerosis patients, *Inverstigative Ophthalmology and Visual Science* **40**(11), 2520–2527.

Pasupathy, A. and Connor, C. E.: 1999, Responses to contour features in macaque area v4, *Journal of Neurophysiology* **82**(5), 2490–2502.

Pasupathy, A. and Connor, C. E.: 2001, Shape representation in area v4: Position-specific tuning for boundary conformation, *Journal of Neurophysiology* **86**(5), 2505–2519.

Pasupathy, A. and Connor, C. E.: 2002, Population coding of shape in area v4, *Nature Neuroscience* **5**(12), 1332–1338.

Patton, N., Aslam, T. M., MacGillivray, T., Deary, I. J., Dhillon, B., Eikelboom, R. H., Yogesan, K. and Constable, I. J.: 2006, Retinal image analysis: Concepts, applications and potential, *Progress in Retinal and Eye Research* **25**(1), 99–127.

Paulus, W., Schwarz, G., Werner, A., Lange, H., Bayer, A., Hofschuster, M., Muller, N. and Zrenner, E.: 1993, Impairment of retinal increment thresholds in Huntingtons-Disease, *Annals of Neurology* **34**(4), 574–578.

Petkov, N.: 1995, Biologically motivated computationally intensive approaches to image pattern-recognition, *Future Generation Computer Systems* **11**(4-5), 451–465.

Petkov, N. and Kruizinga, P.: 1997, Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: bar and grating cells, *Biological Cybernetics* **76**(2), 83–96.

Petkov, N. and Westenberg, M. A.: 2003, Suppression of contour perception by band-limited noise and its relation to non-classical receptive field inhibition, *Biological Cybernetics* **88**(10), 236–246.

Plamondon, R. and Srihari, S.: 2000, On-line and off-line handwriting recognition: A comprehensive survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(1), 63–84.

Priebe, N. J. and Ferster, D.: 2006, Mechanisms underlying cross-orientation suppression in cat visual cortex, *Nature Neuroscience* **9**(4), 552–561.

Qin Li, You, J., Lei Zhang and Bhattacharya, P.: 2006, A multiscale approach to retinal vessel segmentation using Gabor filters and scale multiplication, *2006 IEEE Conference on Systems, Man, and Cybernetics*, pp. 3521–7. 2006 IEEE Conference on Systems, Man, and Cybernetics, 8-11 Oct. 2006, Taipei, Taiwan.

Rangayyan, R. M., Zhu, X., Ayres, F. J. and Ells, A. L.: 2010, Detection of the Optic Nerve Head in Fundus Images of the Retina with Gabor Filters and Phase Portrait Analysis, *Journal of Digital Imaging* **23**(4), 438–453.

Ranzato, M., Poultney, C., Chopra, S. and LeCun, Y.: 2006, Efficient learning of sparse representations with an energy-based model, *in* J. P. et al. (ed.), *Advances in Neural Information Processing Systems (NIPS 2006)*, MIT Press.

Reid, R. C. and Alonso, J. M.: 1995, Specificity of monosynaptic connections from thalamus to visual-cortex, *Nature* **378**(6554), 281–284.

Ricci, E. and Perfetti, R.: 2007, Retinal blood vessel segmentation using line operators and support vector classification, *IEEE Transactions on Medical Imaging* **26**(10), 1357–1365.

Riesenhuber, M. and Poggio, T.: 1999, Hierarchical models of object recognition in cortex, *Nature Neuroscience* **2**(11), 1019–1025.

Ringach, D. and Shapley, R.: 2004, Reverse correlation in neurophysiology, *Cognitive Science* **28**(2), 147–166.

Rodieck, R. W.: 1965, Quantitative analysis of cat retinal ganglion cell response to visual stimuli, *Vision Research* **5**(11), 583–601.

Roelfsema, P. R.: 2006, Cortical algorithms for perceptual grouping, *Annual Review of Neuroscience* **29**, 203–227.

Rosenthaler, L., Heitger, F., Kubler, O. and von der Heydt, R.: 1992, Detection of general edges and keypoints, *Proc. Eur. Conf. Computer Vision (ECCV'92)*, G. Sandini, Ed., pp. 78–86.

Saeedi, S., Paull, L., Trentini, M. and Li, H.: 2011, Neural Network-Based Multiple Robot Simultaneous Localization and Mapping, *IEEE Transactions on Neural Networks* **22**(12, Part 2), 2376–2387.

Salinas, E. and Abbott, L. F.: 1997, Invariant visual responses from attentional gain fields, *Journal of Neurophysiology* **77**(6), 3267–3272.

Saykol, E., Sinop, A., Gudukbay, U., Ulusoy, O. and Cetin, A.: 2004, Content-based retrieval of historical Ottoman documents stored as textual images, *IEEE Transactions on Image Processing* **13**(3), 314–325.

Schmid, C. and Mohr, R.: 1997, Local grayvalue invariants for image retrieval, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(5), 530–535.

Sclar, G. and Freeman, R.: 1982, Orientation selectivity in the cats striate cortex is invariant with stimulus contrast, *Experimental brain research* **46**(3), 457–461.

Sclar, G., Maunsell, J. H. R. and Lennie, P.: 1990, Coding of image-contrast in central visual pathways of the macaque moneky, *Vision research* **30**(1), 1–10.

Sebastian, T., Klein, P. and Kimia, B.: 2004, Recognition of shapes by editing their shock graphs, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(5), 550–571.

Selvathi, D. and Lalitha Vaishnavi, P.: 2011, Gabor Wavelet based Blood Vessel Segmentation in Retinal Images Using Kernel Classifiers, *Proceedings 2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies (ICSCCN 2011)*, pp. 830–5. 2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies (ICSCCN 2011), 21-22 July 2011, Thuckafay, India.

Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G. and Poggio, T.: 2005, A theory of object recognition: Computations and circuits in the feedforward path of the ventral stream in primate visual cortex, AI Memo 2005-036/CBCL Memo 259, Massachusetts Inst. of Technology, Cambridge.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M. and Poggio, T.: 2007, Robust object recognition with cortex-like mechanisms, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(3), 411–426.

Shen, H., Roysam, B., Stewart, C., Turner, J. and Tanenbaum, H.: 2001, Optimal scheduling of tracing computations for real-time vascular landmark extraction from retinal fundus images, *IEEE Transactions on Information Technology in Biomedicine* **5**(1), 77–91.

Shen, H., Stewart, C., Roysam, B., Lin, G. and Tanenbaum, H.: 2003, Frame-rate spatial referencing based on invariant indexing and alignment with application to online retinal image registration, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(3), 379–384.

Sherman, T.: 1981, On connecting large vessels to small - the meaning of murray law, *Journal of General Physiology* **78**(4), 431–453.

Shin, M. C., Goldgof, D. and Bowyer, K. W.: 1998, An objective comparison methodology of edge detection algorithms using a structure from motion task, *Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No.98CB36231)*, pp. 190–5.

Sillito, A. M., Cudeiro, J. and Jones, H. E.: 2006, Always returning: feedback and sensory processing in visual cortex and thalamus, *Trends in Neuroscience* **29**(6, SI), 307–316. Meeting on Neural Substrates of Cognition, Madrid, Spain, 2005.

Soares, J. V. B., Leandro, J. J. G., Cesar, Jr., R. M., Jelinek, H. F. and Cree, M. J.: 2006, Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification, *IEEE Transactions on Medical Imaging* **25**(9), 1214–1222.

Sonka, M., Hlavac, V. and Boyle, R.: 1999, *Image processing, analysis, and machine vision*, Pacific Grove, CA: Brooks/Cole.

Staal, J., Abramoff, M., Niemeijer, M., Viergever, M. and van Ginneken, B.: 2004, Ridge-based vessel segmentation in color images of the retina, *IEEE Transactions on Medical Imaging* **23**(4), 501–509.

Stork, D. G. and Wilson, H. R.: 1990, Do gabor functions provide appropriate descriptions of visual cortical receptive-fields, *Journal of the Optical Society of America A-Optics Image Science and Vision* **7**(8), 1362–1373.

Sun, H., Dul, M. W. and Swanson, W. H.: 2006, Linearity can account for the similarity among conventional, frequency-doubling, and Gabor-based perimetric tests in the glaucomatous macula, *Optometry and Vision Science* **83**(7), 455–465.

Szwed, M., Dehaene, S., Kleinschmidt, A., Eger, E., Valabregue, R., Amadon, A. and Cohen, L.: 2011, Specialization for written words over objects in the visual cortex, *Neuroimage* **56**(1), 330–344.

Takahama, T., Nagatani, K. and Tanaka, Y.: 2004, Motion planning for dual-arm mobile manipulator - realization of "tidying a room motion", *2004 IEEE International Conference on Robotics and Automation (IEEE Cat. No.04CH37508)*, IEEE Robotics and Automation Soc, pp. 4338–43 Vol.5. 2004 IEEE International Conference on Robotics and Automation, 26 April-1 May 2004, New Orleans, LA, USA.

Tanaka, K., Saito, H., Fukada, Y. and Moriya, M.: 1991, Coding visual images of objects in the inferotemporal cortex of the macaque monkey, *Journal of Nuerophysiology* **66**(1), 170–189.

Tsai, C., Stewart, C., Tanenbaum, H. and Roysam, B.: 2004, Model-based method for improving the accuracy and repeatability of estimating vascular bifurcations and crossovers from retinal fundus images, *IEEE Transactions on Information Technology in Biomedicine* **8**(2), 122–130.

Tso, M. and Jampol, L.: 1982, Path-physiology of hypertensive retinopathy, *Opthalmology* **89**(10), 1132–1145.

Tyler, C. W.: 1978, Selectivity for spatial-frequency and bar width in cat visual-cortex, *Vision Research* **18**(1), 121–122.

Ungerleider, L. G., Galkin, T. W., Desimone, R. and Gattass, R.: 2008, Cortical connections of area V4 in the macaque, *Cerebral Cortex* **18**(3), 477–499.

Ungerleider, L. G. and Mishkin, M.: 1982, *Two cortical visual systems*, MIT Press.

Usman Akram, M., Tariq, A., Nasir, S. and Khan, S.: 2009, Gabor wavelet based vessel segmentation in retinal images, *2009 IEEE Symposium on Computational Intelligence for Image Processing (CIIP)*, p. 4 pp. 2009 IEEE Symposium on Computational Intelligence for Image Processing (CIIP), 30 March-2 April 2009, Nashville, TN, USA.

Vinciarelli, A.: 2002, A survey on off-line Cursive Word Recognition, *Pattern Recognition* **35**(7), 1433–1446.

Von Der Heydt, R.: 1987, Approaches to visual cortical function, *Reviews of Physiology Biochemistry and Pharmacology* **108**, 69–150.

Wang, X., Ding, X. and Liu, C.: 2005, Gabor filters-based feature extraction for character recognition, *Pattern Recognition* **38**(3), 369–379.

Wilson, H. R., Wilkinson, F. and Asaad, W.: 1997, Concentric orientation summation in human form vision, *Vision Research* **37**(17), 2325–2330.

Xiaojun Du and Bui, T.: 2010, Retinal Image Segmentation Based on Mumford-Shah Model and Gabor Wavelet Filter, *Proceedings of the 2010 20th International Conference on Pattern Recognition (ICPR 2010)*, IEEE Comput. Soc., pp. 3384–7. 2010 20th International Conference on Pattern Recognition (ICPR 2010), 23-26 Aug. 2010, Istanbul, Turkey.

Xu, C., Liu, J. and Tang, X.: 2009, 2D Shape Matching by Contour Flexibility, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31**(1), 180–186.

Xu, X. M., Bonds, A. B. and Casagrande, V. A.: 2002, Modeling receptive-field structure of koniocellular, magnocellular, and parvocellular LGN cells in the owl monkey (Aotus trivigatus), *Visual Neuroscience* **19**(6), 703–711.

Xu, X. M., Ichida, J. M., Allison, J. D., Boyd, J. D., Bonds, A. B. and Casagrande, V.: 2001, A comparison of koniocellular, magnocellular and parvocellular receptive field properties in the lateral geniculate nucleus of the owl monkey (Aotus trivirgatus), *Journal of Physiology-London* **531**(1), 203–218.

Yan Ke and Sukthankar, R.: 2004, PCA-SIFT: a more distinctive representation for local image descriptors, *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, number vol.2, pp. II–506–13.

Yavuz, Z. and Kose, C.: 2011, Retinal blood vessel segmentation using gabor filter and top-hat transform, *2011 IEEE 19th Signal Processing and Communications Applications Conference (SIU 2011)*, pp. 546–9. 2011 IEEE 19th Signal Processing and Communications Applications Conference (SIU 2011), 20-22 April 2011, Antalya, Turkey.

Yavuz, Z. and Koumlse, C.: 2010, Comparing 2D matched filter response and Gabor filter methods for vessel segmentation in retinal images, *2010 National Conference on Electrical, Electronics and Computer Engineering (ELECO 2010)*, pp. 648–52. 2010 National Conference on Electrical, Electronics and Computer Engineering (ELECO 2010), 2-5 Dec. 2010, Bursa, Turkey.

Zamir, M., Medeiros, J. and Cunningham, T.: 1979, Arterial bifurcations in the human retina, *Journal of General Physiology* **74**(4), 537–548.

Zamir, M., Sinclair, P. and Wonnacott, T.: 1992, Relation between diameter and flow in major branches of the arch of the aorta, *Journal of Biomechanics* **25**(11), 1303–1310.

Zeki, S. M.: 1973, Color coding in rhesus-monkey prestriate cortex, *Brain Research* **53**(2), 422–427.

Zhang, Y. J.: 1996, A survey on evaluation methods for image segmentation, *Pattern Recognition* **29**(8), 1335–1346.

Research Activities

Journal Papers

- G. Azzopardi, N. Petkov, *"Trainable COSFIRE filters for keypoint detection and pattern recognition"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35(2), pp. 490–503, 2013
- G. Azzopardi, N. Petkov, *"A CORF computational model that relies on LGN input outperforms the Gabor function model"*, Biological Cybernetics, vol. 2(3), pp. 177–89, 2012
- G. Azzopardi, N. Petkov, *"Automatic detection of vascular bifurcations in segmented retinal images using trainable COSFIRE filters"*, Pattern Recognition Letters, in press, doi: 10.1016/j.patrec.2012.11.002, 2012
- G. Azzopardi, N. Petkov, *"Trainable shape-selective S-COSFIRE filters for object recognition and localization"*, submitted, 2012

Conference Proceedings

- G. Azzopardi, N. Petkov, *"Contour Detection by CORF Operator,"* Artifical Neural Networks and Machine Learning - ICANN 2012. Proceedings of the 22nd International Conference on Artificial Neural Networks, Lausanne, Switzerland, Lectures Notes in Computer Science, vol. 7552, pp. 395-402, Springer, 2012
- George Azzopardi , Nicolai Petkov, *"Detection of retinal vascular bifurcations by rotation- and scale-invariant COSFIRE filters"*, Proceedings of the 9th international conference on Image Analysis and Recognition, Aveiro, Portugal, Lecture Notes in Computer Science, vol. 7325, pp. 363-371, Springer, 2012
- George Azzopardi , Nicolai Petkov, *"Detection of retinal vascular bifurcations by rotation-, scale- and reflection-invariant COSFIRE filters"*, Proceedings of the 25th international symposium on Computer-Based Medical Systems, Rome, Italy, IEEE Conference Publications, doi: 10.1109/CBMS.2012.6266338, pp. 1-4, 2012
- G. Azzopardi, N. Petkov: *"Detection of retinal vascular bifurcations by trainable V4-like filters"*, in Computer Analysis of Images and Patterns CAIP 2011, Seville, Spain, Lecture Notes in Computer Science, vol. 6854, pp. 451-459, Springer, 2011

Posters

- G. Azzopardi and N. Petkov, *"CORF: A computational model of a simple cell with application to contour detection"*, ECVP, Alghero, Sardinia, Italy, Sep 2012
- G. Azzopardi and N. Petkov, *"A CORF computational model of a simple cell with application to contour detection"*, AVA/BMVA Meeting on Biological and Computer Vision, Microsoft Research Center, Cambridge, UK, May 2012
- G. Azzopardi and N. Petkov, *"V4-like filters applied to the detection of retinal vascular bifurcations"*, AVA/BMVA Meeting on Biological and Computer Vision, School of Psychology, Cardiff University, Wales, May 2011

Research Fund

- Project Title: "EYE-CAD: Detection of early diabetic retinopathy by the automatic analysis of retinal fundus images using cutting edge technology based on artificial intelligence", The Malta Council for Science and Technology (MCST), Reference No. R&I-2012-016B (EUR200k), Malta, 2012

Summer Schools

- ICVSS, International Computer Vision Summer School, Siracusa, Sicily, 11-16th July 2011.
- AERFAI, Summer School on Pattern Recognition and Machine Learning in Multimedia Systems, Benicassim, Spain, 7-11th June 2010.

Talks

- "Brain-inspired computing for machine vision: Trainable COSFIRE filters for keypoint detection and pattern recognition", Juelich Forschungszentrum, Germany, July 2012
- "Brain-inspired computer vision: what comes beyond V2 in visual cortex?", A lecture as part of the ASCI course a8: Front-End Vision and Multi-Scale Image Analysis, Technische Universiteit Eindhoven (TU/e), Oct 2011
- "Automatic detection of retinal vascular bifurcations by trainable V4-like filters", Spring Meeting of the NVPHBV, Utrecht, The Netherlands, April 2011
- "Biologically Motivated Pattern Recognition", Meeting of the British Computer Science, Anglia Ruskin University in Chelmsford, UK, March 2011

Organized Seminars

- "Supercomputers and Brain-Inspired Machine Vision", St. Martin's Institute, Hamrun, Malta, 10th April 2012
  - "Computational models of the visual system", Prof. dr. Nicolai Petkov
  - "Supercomputers: the road to exascale computing", Prof. dr. dr. Thomas Lippert
  - "Brain-inspired algorithms for computer vision", George Azzopardi
- "Robot Intelligence: How do robots learn and see?", St. Martin's Institute, Hamrun, Malta, 14th April 2011
  - "Contour detection by surround suppression of texture", Prof. dr. Nicolai Petkov
  - "Matrix Relevance Learning: basic concepts and applications in medicine", Prof. dr. Michael Biehl
  - "Trainable V4-like filters for the detection of vascular bifurcations in retinal images", George Azzopardi

Curriculum Vitae

George Azzopardi was born on July 4th 1981, in Qormi, Malta. In May 2000, he joined Bank of Valletta (Malta) as a software developer where he was involved in the analysis, design and implementation of various projects for seven years. During that time, he took part-time lectures in the evenings and in 2004 he earned a certificate in Banking and Financial Services from the Chartered Institute of Bankers (Malta). Following that, in 2006, he received a BSc degree with honours (first class) in computer science from Goldsmiths University of London where he was awarded an academic achievement for his overall performance.

His interest in visual pattern recognition was discovered while working on the BSc dissertation which concerned the automatic recognition of handwritten signatures. In 2006, he was awarded a scholarship from the Government of Malta to pursue postgraduate studies at Queen Mary University of London. In September 2006, he moved to London (UK) for a year to read for a MSc degree focused on image processing and computer vision techniques. He graduated with distinction (ranked first) in 2008. In his MSc thesis he investigated the effectiveness of contrast invariant rank-based filters for the recognition of textures.

Between 2005 and 2008 he was a part-time SAP scientific developer at Iteanova Ltd. He did that work first in parallel to his full-time job at the Bank and then alongside his MSc studies in London.

In 2008, he was accepted for a four-year PhD position under the supervision of Prof. Nicolai Petkov at University of Groningen, the Netherlands. In October 2008, he and his wife Charmaine commenced their respective PhD journeys together in Groningen. His research interests are in the field of visual pattern recognition including computational modelling of the visual system of the brain. In his studies, he proposed novel brain-inspired algorithms for machine vision applications, including contour detection, feature detection, and shape recognition.

Since October 2012, George Azzopardi is a research scientist at TNO (Netherlands organisation for applied scientific research). Further to an agreement between TNO and University of Groningen he has a one-day-a-week position as a researcher at the Intelligent Systems group, which is headed by Prof. Nicolai Petkov.

He balances his life by fitness training, swimming, playing football and billiard with friends, visiting family in Malta, and travelling around the world.