Lappeenrannan teknillinen yliopisto
Lappeenranta University of Technology

Joni-Kristian Kämäräinen

FEATURE EXTRACTION USING GABOR FILTERS

Thesis for the degree of Doctor of Science (Technology) to be presented with due permission for public examination and criticism in the Auditorium of the Student Union House at Lappeenranta University of Technology, Lappeenranta, Finland on the 28th of November, 2003, at noon.

Acta Universitatis
Lappeenrantaensis
165

Supervisor Professor Heikki Kälviäinen
Laboratory of Information Processing
Department of Information Technology
Lappeenranta University of Technology
Finland

Reviewers Professor Jussi Parkkinen
Department of Computer Science
University of Joensuu
Finland

Professor Jouko Lampinen
Laboratory of Computational Engineering
Department of Electrical and Communications Engineering
Helsinki University of Technology
Finland

Opponents Professor Jouko Lampinen
Laboratory of Computational Engineering
Department of Electrical and Communications Engineering
Helsinki University of Technology
Finland

Professor Olli Silven
Information Processing Laboratory
Department of Electrical and Information Engineering
University of Oulu
Finland

ISBN 951-764-816-2
ISSN 1456-4491

Lappeenrannan teknillinen yliopisto
Digipaino 2003

Preface

The work presented in this thesis was done in 2000-2003 at the Laboratory of Information Processing, Department of Information Technology, Lappeenranta University of Technology, Finland.

I wish to express my gratitude to my supervisor Professor Heikki Kälviäinen, who guided my research, for his support and encouragement during my work on the thesis. I gratefully acknowledge the contributions of the co-authors in the publications, Professors Josef Kittler and Jarmo Partanen, researchers Miroslav Hamouz, Tuomo Lindh and Jero Ahola, and especially my close colleague Ville Kyrki. Without their help the thesis would not have been completed. I also thank all my fellow researchers in the Laboratory of Information Processing.

I want to thank Professors Jussi Parkkinen and Jouko Lampinen for reviewing the thesis and providing me with most valuable suggestions concerning the text.

The financial support from the East Finland Graduate School in Computer Science and Engineering, Academy of Finland, Foundation of Lappeenranta University of Technology, Emil Aaltonen Foundation, Nokia Foundation, and Tekniikan Edistämissäätiö is gratefully acknowledged.

Finally, I owe my dearest thanks to my nearest relatives and especially to you, my love, Päivi.

Lappeenranta, October 2003

Joni-Kristian Kämäräinen

Abstract

Joni-Kristian Kämäräinen
FEATURE EXTRACTION USING GABOR FILTERS
Lappeenranta, 2003
150 p.

Acta Universitatis Lappeenrantaensis 165
Diss. Lappeenranta University of Technology

ISBN 951-764-816-2
ISSN 1456-4491

Signals are of great interest since they are the form used to describe natural phenoma and provide information from engineering systems. However, a signal must often be processed to mine important properties for further tasks, e.g., recognizing words from a speech or detecting objects from an image. The process that refines a signal to more meaningful pieces of information is often referred to as feature extraction.

For decades time-frequency analysis has played a central role in signal processing as it combines two fundamental domains and allows simultaneous representation of the signals in both. Time-frequency methods should provide something novel not present in either of the domains solely. Intuitively one could consider frequency to represent "What" and time "Where" and the combined representation gives a description of the kind of events a signal contains and where they occur. If events, whatever they are, are salient subparts, they can be considered analogous to features and respectively any time-frequency representation to feature extraction.

Gabor filters have attracted researchers since they extract information quanta in forms of time and frequency, two physically measurable quantities, combined in the most elegant way by the Heisenberg's uncertainty relation. In addition, the physiology of our own perception systems seems to present similar characteristics. In this thesis, the main results of Gabor research are collated to track the evolution from the very first paper by Dennis Gabor in 1946 to the present state of art. As new results object detection and recognition methods based on the Gabor filtering theory are presented and applied to several applications: induction motor bearing damage detection, recognition of line-drawing symbols and electric components, and detection of faces.

Keywords: feature extraction, Gabor, time-frequency analysis, signal detection, object recognition, image processing, computer vision, face detection

UDC 004.93'1 : 004.932

$\mathfrak{F}\{\cdot\}$	Fourier transform
$\mathfrak{F}^{-1}\{\cdot\}$	Inverse Fourier transform
α	Sharpness (Gabor function major axis)
β	Sharpness (Gabor function minor axis)
γ	Sharpness of Gabor filter (major axis)
η	Sharpness of Gabor filter (minor axis)
θ	Orientation angle of Gabor filter
$\xi(t)$	1-d signal
$\xi(x,y)$	2-d signal
ϕ	Gabor filter phase shift
$\psi(t)$	1-d Gabor filter in time domain
$\psi(t;f)$	1-d Gabor filter in time domain
$\psi(n)$	Discrete 1-d Gabor filter in time domain
$\psi(x,y)$	2-d Gabor filter in spatial domain
$\psi(x,y;f,\theta)$	2-d Gabor filter in spatial domain
$\psi^*(t)$	Complex conjugate of $\psi(t)$
$\psi(t)*\xi(t)$	Convolution of $\psi(t)$ and $\xi(t)$
$\psi_{kl}(t)$	Gabor function in Gabor expansion
Δ	Uncertainty
$\Psi(f)$	1-d Gabor filter in frequency domain
$\Psi(u,v)$	2-d Gabor filter in frequency domain

a_{kl}	Expansion coefficient in Gabor expansion
f	Frequency (Gabor filter frequency)
f_0	Frequency of Gabor function
j	Imaginary unit
$r_\xi(t; f)$	Response of 1-d Gabor filter
$r_\xi(x, y; f, \theta)$	Response of 2-d Gabor filter
t	Time
t_0	Location of Gabor function
x	Spatial coordinate
y	Spatial coordinate

1-d	one dimensional
2-d	two dimensional
GEF	Gabor elementary function
ICA	Independent component analysis
PCA	Principal component analysis
STFT	Short-time Fourier transform

List of Publications

I. Kamarainen, J.-K., Kyrki, V., Kälviäinen, H., "Noise Tolerant Object Recognition Using Gabor Filtering", In *Proceedings of the 14th International Conference on Digital Signal Processing (Santorini, Greece, 2002)*, vol. 2, pp. 1349–1352.

II. Kamarainen, J.-K., Kyrki, V., Kälviäinen, H., "Fundamental Frequency Gabor Filters for Object Recognition", In *Proceedings of the 16th International Conference on Pattern Recognition (Quebec City, Canada, 2002)*, vol. 1, pp. 628–631.

III. Kamarainen, J.-K., Kyrki, V., Kälviäinen, H., "Robustness of Gabor Feature Parameter Selection", In *Proceedings of the IAPR Workshop on Machine Vision Applications (Nara, Japan, 2002)*, pp. 132–135.

IV. Kamarainen, J.-K., Kyrki, V., Hamouz, M., Kittler, J., Kälviäinen, H., "Invariant Gabor Features for Face Evidence Extraction", In *Proceedings of the IAPR Workshop on Machine Vision Applications (Nara, Japan, 2002)*, pp. 228–231.

V. Kamarainen, J.-K., Kyrki, V., Lindh, T., Ahola, J., Partanen, J., "Statistical Signal Discrimination for Condition Diagnosis", In *Proceedings of the Finnish Signal Processing Symposium (Tampere, Finland, 2003)*, pp. 195–198.

VI. Hamouz, M., Kittler, J., Kamarainen, J.-K., Kälviäinen, H., "Hypotheses-Driven Affine Invariant Localization of Faces in Verification Systems", In *Proceedings of the 4th International Conference on Audio- and Video-Based Person Authentication (Guildford, UK, 2003)*, pp. 276–284.

VII. Lindh, T., Ahola, J., Kamarainen, J.-K., Kyrki, V., Partanen, J., "Bearing Damage Detection Based on Statistical Discrimination of Stator Current", In *Proceedings of the 4th International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives (Atlanta, Georgia, USA, 2003)*, pp. 177-181.

VIII. Kyrki, V., Kamarainen, J.-K., Kälviäinen H. "Simple Gabor Feature Space for Invariant Object Recognition", accepted for publication in *Pattern Recognition Letters*, Elsevier.

The publications are in the chronological order in which they have been published and in this thesis these publications are referred to as *Publication I*, *Publication II*, *Publication III*, *Publication IV*, *Publication V*, *Publication VI*, *Publication VII*, and *Publication VIII*.

Contents

Chapter I

Introduction

It would not be appropriate to begin this thesis with anything else but a reference to the original work published by Nobel laureate Dennis Gabor in 1946 where he proposed representing signals as a combination of elementary functions [49]. Those particular elementary functions are now known as Gabor functions. Gabor's work was a continuation and partly parallel to the famous works of Nyquist [101] and Shannon [127] founding the theory of communication. Since then Gabor functions have been deployed in many areas of signal processing and this thesis studies Gabor functions particularly in feature extraction. To list some of the most influential works, milestones which contributed to this research, credit must be given to the detailed analysis of Gabor's expansion by Bastiaans [6, 8, 9], the introduction of a general picture processing operator, the first two dimensional (2-d) counterpart of the Gabor elementary function, by Granlund [52], the generalization to two dimensions and the 2-d function as a model of simple cell in the mammalian visual cortex by Daugman [38], and wavelet treatment of Gabor filters by Tai Sing Lee [83]. Utilizing the results provided by the above-mentioned and many more researchers, this thesis provides new information on feature extraction based on Gabor filters.

This thesis can also be viewed as a study of features based on time-frequency representations. Time and frequency are two fundamental domains and physically measurable quantities, but still idealizations if one is considered from the other's perspective. Frequency is a simple waveform in the time domain, but to be sharply defined in the frequency domain it must be infinite in the time domain; a waveform always existed and remains forever. A duality between the domains holds and a sharp time instant consequently needs components from all frequencies having an infinite frequency representation. Neither phenomenon is encountered in everyday life but signals rather tend to have properties from the both; they certainly have some frequency characteristics, but they also start to raise after some time instant and correspondingly start to fade before some further point. The description needed is a function of both time and frequency. For example, in music frequency could describe a tone and time an instant the tone is played. A combined representation, time-frequency representation, would thus be similar to the notation in the music composition. Furthermore, as the notation can be used to recognize a song, so the time-frequency representation can be used to recognize a signal. A process that refines a signal to more meaningful pieces of information is often referred to as feature extraction. There are many different time-frequency representations from which features can be extracted [33, 119], but in this thesis one particular based on Gabor filters is considered. It is intriguing to study why a function that can describe phenomena in quantum mechanics also benefits engineering tasks, such as image processing.

This thesis has been made bearing three main objectives in mind: (i) to cover the history of the fundamental theories and innovations involved in the discovery and development of Gabor filters in signal processing; (ii) to revisit existing and examine new filter properties useful in problems encountered in feature extraction; and (iii) to place the theory utilized in publications *Publication I* – *Publication VIII* into a correct research frame and evaluate its novelty. It may seem that the history considerations in this thesis receive almost too much attention, but the Gabor research community lacks a good survey and the author has thus found it necessary to gather all information into this work and enlighten the basic principles behind the theories. Different Gabor researchers have combined theories from many different contexts and as a result the Gabor function remains an alien concept, confusing researchers who are not aware of the context they should work with; Gabor filter, Gabor expansion, Gabor transform, Gabor jet, or Gabor wavelet? These questions are covered in Chapter 2 but in a broad sense the author's contribution to the first objective is scattered all around the thesis where attempts are made to encapsulate the most prominent theories. Despite the fact that this thesis also discusses Gabor filters in one dimension it is evident that the results are mainly relevant to image processing. Gabor filters in general seem to be more familiar to the image processing community than other research areas, which is probably due to the physiological fact that the 2-d Gabor filter is an accurate model of a simple cell in the striate cortex of the mammalian visual system [38]. This qualitative fact has promoted the filters for almost twenty years and is still mentioned in almost every single article on the subject. Nevertheless, the physiological issues are only briefly considered in this study where attempts are made to provide the motivations with a more quantitative rationale. The main contribution to the second objective is in Chapter 3 and *Publication VIII* where the most important properties of features based on Gabor filter responses are presented. The final objective is the most difficult one since there is no general research frame covering Gabor filters, some things seem to have been borrowed from frame theory [42] and some things from wavelet research [36]. In this sense even this thesis cannot be exclusive but leaves a lot of room for future work. The research moves toward Gabor feature spaces in the last sections of Chapter 3, where a simple feature space is presented, robustness and distortion tolerances are analyzed, and finally a connection to the shiftability concept [129] is pointed out. The research has been of bottom-up type, starting from the early experiments utilizing only several filter properties [76, 77] and concluding in this thesis where the fundamental theories are covered and more comprehensively combined to understand and realize experiments in *Publication I* — *Publication VIII*. A contribution to applications of feature extraction is present since the theory has been applied in real applications, i.e. recognition of electric components (*Publication II*), face detection (*Publication IV*, *Publication VI*), and induction motor bearing damage detection (*Publication V*,*Publication VII*).

Summary of publications

In *Publication I* are experimentally analyzed noise and distortion tolerances of the proposed Gabor filter based features. Specifically, Gaussian and salt-and-pepper noise, pixel displacement distortion, and gradient type illumination variation are included. The specific features, global Gabor features and fundamental frequency Gabor features, by the author et al. are analyzed in the tasks first introduced in [76] and *Publication II*. A conclusion is made that Gabor filters seem to provide a remarkable noise and distortion tolerance. The author participated in the development and implementation of the methods, writing the publication, and performed the experiments.

Publication II introduces a novel use of Gabor filters utilizing low frequencies which are shown to be capable of representing the object size information. In addition, algorithms are provided to learn new objects and detect them and their pose in images. The author participated in the development and implementation of the methods, writing the publication, and performed the experiments.

In *Publication III* the analysis initiated in *Publication I* is continued while the focus is changed to the filter parameters. This consideration is stressed to provide information about the performance of the features when only a sub-optimal set of the parameters can be realized, which is often the case in real applications. The author participated in the development and implementation of the methods, writing the publication, and performed the experiments.

Publication IV presents a system used in the first phase of a complete face authentication system. The author of this thesis developed a translation, scale, and rotation invariant face evidence extraction method based on Gabor features. The author developed and implemented the method and participated in performing the experiments and writing the publication.

In *Publication V* the method introduced in *Publication VII* is analyzed in more detail and improved by replacing the divergence measure with a more stable one. While *Publication VII* describes the specific application area and the preliminary results *Publication V* studies the advantages and pitfalls of the method itself. The author developed and implemented the method and participated in performing the experiments and writing the publication.

Publication VI describes the first results of a face detection system into which the face evidence extraction module introduced in *Publication IV* is integrated. The publication provides results in terms of the face detection accuracy, where the system performs very well being superior to the previous implementation in [56]. The author developed and implemented the face evidence extraction module and participated in performing the experiments and writing the publication.

Publication VII introduces a method for analysis of diagnostic signals. The method is applied to stator current signals of inductive motors to detect characteristic frequencies distinguishing normal condition motors and motors with a bearing fault. The detected frequencies are consistent with the theory and the method also reveals information about conditions where the theory cannot be applied. The author developed and implemented the method and participated in performing the experiments and writing the publication.

Publication VIII summarizes and analyses in more detail the feature space used in *Publication IV* and *Publication VI*. The author contributes as the other main author of the presented theory and in writing the publication.

Chapter II

Gabor Filters

In this chapter the theory and properties of Gabor filters, being the basic building blocks of the feature extraction in the following chapters, are recalled. The first consideration is what are the reasons for restricting the analysis in the two predefined domains, time and frequency in one dimension and space and frequency (spatial-frequency) in two dimensions. Finally, if one can agree that these two domains are of special interest, the ultimate goal of feature extraction and object recognition, "Knowing what is where?", can be connected to Heisenberg's uncertainty principle and especially to Gabor filters as elements extracting the minimal amount of information.

2.1 Signals in two domains

For decades there have been two alternative approaches to describe one dimensional (1-d) signals, the first one represents signals as a function of time and the second as a function of frequency. A representation can be transformed from one to another via the Fourier or inverse Fourier transforms and they thus carry the same information but in different forms. Both of the representations are somewhat idealizations since the first one operates on sharply defined time instants and the second with infinite waveform trains on rigorously defined frequencies. Unsurprisingly the idealizations conflict even in descriptions of simple phenomena such as a change of frequency which is encountered in everyday life, e.g., the tone of a note and a time instant the note should be played are both essential information for music composition. However, the contradiction of the terms is evident as the statement involves both time and frequency. In the analysis of such signals their basic properties should be characterized; what kind of events a signal contains and when do they occur. At least in the simple example of music composition frequency may represent the type of event, tone, and time represents occurrence of the event. It is thus motivated to construct an approach that represents signals as a function of both time and frequency. This was also the motivation of Dennis Gabor in 1946 when he proposed the use of special elementary functions, later named after him as Gabor elementary functions (GEF), to represent signals simultaneously in time and frequency

[49]. Gabor's work was based on the recently discovered wave mechanical theories and in particular on Heisenberg's uncertainty principle, which led to the derivation of the Gabor elementary function. The function represents the minimal quantum of information, that is, it occupies the minimal area, a rectangle, in the time-frequency plane; the minimal amount of simultaneous information in time and frequency. Since then Gabor's studies have had an important role in the analysis of windowed Fourier transforms [119], in the development of wavelet theory [36], and in feature extraction in image analysis [38, 52].

There is a symmetry between synthesizing signals as small information pieces, a combination of elementary functions, and analyzing signals where important information pieces are extracted from a signal. Instead of signal analysis, Gabor's original work was intended for signal synthesis, i.e., how a signal can be constructed from a linear combination of Gabor elementary functions. Gabor gave an iterative solution for expanding signals into elementary functions [49] and later the solution was derived in a closed form by Bastiaans [6, 8, 9]. In the expansion, signals are first analyzed by a biorthogonal function set, biorthogonal to the set of Gabor functions, to compute expansion coefficients that are used to synthesize original signals from the Gabor functions. To be consistent with Gabor's work, biorthogonal functions should be used in analysis [6, 8], but often Gabor elementary functions themselves are used. This is mainly due to the convenience of a system which is based on the Gabor functions and not the biorthogonals. The system can be easily extended without reconstruction of a new biorthogonal set, the analyzing functions are themselves optimally localized in the time-frequency plane, and since all basis functions are generated from a single function form, important properties for the feature extraction can be realized. These issues will be revisited in the next chapter.

In signal analysis and especially in image processing, which is the main application area in this thesis, feature extraction plays a central role in "Knowing what is where?". If one can be convinced that frequency content provides the information about "what", and time, spatial coordinates in 2-d, provides the information about "where", a connection to the uncertainty principle and Gabor's work can be established and their results utilized.

2.1.1 Frequency

First it should be considered what makes these two domains, time and frequency, special and is it reasonable to restrict the study to them alone, or are there other alternative domains where the same properties could be realized. This issue was of interest to Gabor as well, but his application area, data communication, set fundamental requirements to prefer time and frequency [49, Appendix (9.1)]. With one-dimensional (1-d) signals it is natural to analyze their variation in time since this is the case with the most of the signals encountered and time itself is fundamental. However, it should be noted that for some other signals, e.g., two dimensional (2-d) images, the characteristics of the dimensions are not necessarily comparable to time. Still, the importance of time or spatial coordinates can always be justified, but the frequency representation via the Fourier transform is much more complicated as it considers phenomena in an infinite time interval which is far from our everyday point of view.

In general, the selection between proper domains can be restricted to ones which can carry same signals as the time domain, but in different forms, to guarantee that no important information is lost. To accommodate useful mathematical properties a mapping between domains should preferably be performed by linear transforms. Characteristically one uses the transformation as a mathematical or physical tool to alter the problem into one that can be solved, which is the case for example for the Fourier and Laplace transforms as techniques for solving linear equations. In this thesis the frequency domain has an essential role and the following Fourier transform pairs are used

$$
\begin{aligned}
G(f) = \mathfrak{F}\{g(t)\} &= \int_{-\infty}^{\infty} g(t)e^{-j2\pi tf}dt \\
g(t) = \mathfrak{F}^{-1}\{G(f)\} &= \int_{-\infty}^{\infty} G(f)e^{j2\pi tf}df
\end{aligned} \tag{2.1}
$$

in 1-d space and

$$
\begin{aligned}
G(u,v) = \mathfrak{F}\{g(x,y)\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)e^{-j2\pi(ux+vy)}dxdy \\
g(x,y) = \mathfrak{F}^{-1}\{G(u,v)\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(u,v)e^{j2\pi(ux+vy)}dudv
\end{aligned} \tag{2.2}
$$

in 2-d space [19]. Since the Fourier transform has a bounded continuous inverse transform pair the transform can be represented by orthogonal eigen functions, complex frequencies $e^{-j2\pi tf}$, which construct a continuous spectrum of the transform operator $\mathfrak{F}$ [72]. Thus, any bounded function which has a Fourier transform counterpart can be represented by spectral values of the corresponding eigen functions, amplitudes of frequency spectrum. While the Fourier transform can be regarded as a purely mathematical functional it should be noted that many other transforms also have a spectrum, but spectrum characteristics are distinctly different to Fourier. In this study the main focus is in feature extraction and its subsequent object recognition, and thus, the self-similarity of frequency wave forms, which would not be the case for example with a space spanned by polynomials of different orders, will be essential for further considerations in the following chapters. Furthermore, frequency content of a Fourier spectrum is semantically the same regardless of signal shifts in the time domain due to the quadrature nature of complex frequencies. This would not be the case for example with splines or wavelets because they also have time dependent characteristics and non-continuous spectrum.

Certainly the impact of the original background of the frequency representation, physical phenomena, is strong. Waveforms, optical, electrical, or acoustical and their spectra are appreciated equally as physically picturable and measurable entities. Furthermore, the linear transforms, such as the Fourier transform, provide tools for solving linear equations used to describe dynamics encountered in the physical world. The Fourier transform itself was a result from the original study of Fourier considering solutions for heat conduction and diffusion differential equations [19, Chapt. 17]. In addition, a strong and comprehensive mathematical foundation has been established within the long term research of the Fourier analysis. Gabor was also concerned about reasons other than the elementary mathematical properties and strong theory of the harmonic wave functions. In his study he cleverly pointed out that some other selection of orthogonal basis, like Bessel functions, would not provide spectral components with a number proportional to a specific time interval [49]; Bessel frequencies are proportional to the function roots in the given time interval. This is the case especially in the discrete domain. Still, it seems that the mathematical advantages of the waveforms is the reason that the fundamental information theoretical bounds and theories are formulated in the time and frequency domains [101, 127].

In Gabor's time the wavelet theory had not yet been introduced and even now wavelets lack as comprehensive studies as the Fourier transform. However, wavelets research has now increased in popularity in signal and image analysis and even the uncertainty theorem is considered in that context [35, 66]. The reasons given for using the frequency domain are qualitative rather than quantitative, and thus, the frequency domain and other possible domains, such as the scale domain of the wavelets or splines are alternatives rather than competitors. Based on the reasons mentioned it is motivated to assume that there are advantageous properties present in time and frequency or space and frequency, and thus, the concern about the bias of selecting these two specific domains can be allayed.

2.1.2 Uncertainty

Having established that time and frequency are the proper analysis domains, one would like to have an operator that analyzes signals simultaneously in both domains and provides information of localized time and frequency events. To work in both domains it can be assumed that the operator is based on a kernel that has a form of function. Gabor's original work was about synthesizing signals, but signal analysis is of interest in this thesis. However, due to the dual relationship, symmetry, between these two tasks, the same results can be applied. Gabor's main goal was to find if there were elementary information units on the time-frequency plane carrying the minimal amount of information. The minimal amount of information is bounded by the uncertainty principle, but what is the form of a function that occupies a minimal area on the time-frequency plane?

Let $\psi(t)$ denote a function in terms of time and $\Psi(u)$ its Fourier transform in terms of frequency. These two functions are connected via the Fourier transform pairs

$$\begin{aligned} \Psi(f) &= \mathfrak{F}\{\psi(t)\} = \int_{-\infty}^{\infty} \psi(t) e^{-j2\pi t f} dt \\ \psi(t) &= \mathfrak{F}^{-1}\{\Psi(f)\} = \int_{-\infty}^{\infty} \Psi(f) e^{j2\pi t f} df \end{aligned} \tag{2.3}$$

Clearly function ψ operates on both domains, but since it is well known that any compactly supported function cannot have a finite Fourier transform and vice versa [19], there must always exist uncertainty in the time and frequency location of ψ as well. While the goal is to inspect frequency events evolving in time, instantaneous frequencies, by ψ, measures of its time duration and frequency bandwidth are needed. Instead of sharp time instants and exact frequencies, less exact measures should be used to measure the location of ψ in the time-frequency plane. By employing new definitions, time duration and bandwidth, the presence of uncertainty can be allowed since they define only the effective widths of signals in time and frequency where outside the effective area the function values are negligible. The definitions allow the functions ψ and Ψ to spread to infinities while still having computational compact support in the sense of the bandwidth and time duration measures; infinite functions which actually concentrate to a finite area. It is clear already from the definition that there is no unique representation for the duration or bandwidth but the definitions vary depending on the application requirements. It should be noted that the selection of measure will also induce a bias to the uncertainty principle and consequently the optimal ψ may vary. Bandwidth has been widely studied in communication systems and it seems that even in this field no single definition suffices [3, 57, 130]. The bandwidth definition that Gabor found useful was the root mean square (r.m.s.) bandwidth, defined as the square root of the second centralized moment of a properly normalized form of the squared amplitude spectrum about a suitably chosen point. Since the r.m.s. bandwidth represents deviation from a mean value it can be accepted as a measure of uncertainty. Furthermore, a similar r.m.s. measure can be applied to uncertainty in time, the r.m.s. time duration. In literature both or a combination of these two measures have been referred to as Gabor bandwidths [3, 100]. Now uncertainties in time, the time duration Δt, and in frequency, the bandwidth Δf, can be defined as (e.g. [38])

$$\begin{aligned}
\Delta t &= \sqrt{\frac{\int_{-\infty}^{\infty}(t-\mu_t)^2\psi(t)\psi^*(t)dt}{\int_{-\infty}^{\infty}\psi(t)\psi^*(t)dt}} \\
\Delta f &= \sqrt{\frac{\int_{-\infty}^{\infty}(f-\mu_f)^2\Psi(f)\Psi^*(f)df}{\int_{-\infty}^{\infty}\Psi(f)\Psi^*(f)df}} \\
\mu_t &= \frac{\int_{-\infty}^{\infty}t\psi(t)\psi^*(t)dt}{\int_{-\infty}^{\infty}\psi(t)\psi^*(t)dt} \quad \mu_f = \frac{\int_{-\infty}^{\infty}f\Psi(f)\Psi^*(f)df}{\int_{-\infty}^{\infty}\Psi(f)\Psi^*(f)df}
\end{aligned} \tag{2.4}$$

where μ_t and μ_f can be interpreted as the mass centroids or means of ψ in time and frequency, and ψ^* and Ψ^* denote the complex conjugates of ψ and Ψ. For functions of time and frequency, these two definitions of uncertainty are connected via Heisenberg's uncertainty principle defined in the rigorous form as

$$\Delta t \Delta f \geq \frac{1}{4\pi} \tag{2.5}$$

where $1/(4\pi)$ corresponds the result obtained by using definitions in Eqs. (2.1) and (2.2) for the Fourier transform. There are several other forms of the Fourier transform, and thus, the uncertainty value is sometimes replaced by 1/2, but having an identical interpretation [19, 38, 49].

For any function ψ the uncertainties in time and frequency can be calculated with the r.m.s. bandwidth and duration measures in Eq. (2.4) and connected together via the uncertainty principle in Eq. (2.5). Now it is possible to derive a function ψ, which has a shape for which the product $\Delta t \Delta f$ actually assumes the smallest possible value, i.e., for which the inequality in Eq. (2.5) turns into an equality. The form of the solution can be derived from the uncertainty inequality and it is equivalent to a solution of a special case of the harmonic oscillator differential equation. *"The signal which occupies the minimum area $\Delta t \Delta f = \frac{1}{2}(= \frac{1}{4\pi})$ is the modulation product of a harmonic oscillation$^{(*)}$ of any frequency with pulse of the form of a probability function$^{(**)}$"* [49]

$$\psi(t) = \underbrace{e^{-\alpha^2(t-t_0)^2}}_{(**)}\underbrace{e^{j2\pi f_0 t+\phi}}_{(*)} \tag{2.6}$$

where α is the sharpness (time duration and bandwidth) of the Gaussian, t_0 denotes the centroid of the Gaussian, f_0 is the frequency of the harmonic oscillations, and ϕ denotes the phase shift of the oscillation. The Gabor elementary function in Eq. (2.6) has a Fourier spectrum of analytical form

$$\Psi(f) = \sqrt{\frac{\pi}{\alpha^2}} e^{-(\frac{\pi}{\alpha})^2 (f-f_0)^2} e^{-j2\pi t_0 (f-f_0)+\phi} \tag{2.7}$$

For the functions in Eqs. (2.6) and (2.7) it is straightforward to show that $\mu_t = t_0$, $\mu_f = f_0$, $\Delta t = \frac{1}{2\alpha}$, $\Delta f = \frac{\alpha}{2\pi}$, and $\Delta t \Delta f = \frac{1}{4\pi}$. The form of the function, the Gabor elementary function ψ, that minimizes the uncertainty and turns the inequality Eq. (2.5) into equality is now defined and the use and properties of the function will be considered next.

2.2 Gabor expansion

In literature it may sometimes be difficult to establish how the Gabor elementary functions are used in signal processing and how the use is related to Gabor's work. Sometimes the functions are used in signal analysis, as linear filters, and sometimes in signal synthesis, as reconstruction basis functions. These two approaches are dual and either can be used in both analysis and synthesis. In this thesis the focus is on signal analysis, feature extraction, using Gabor filters, but for a better understanding of the dual nature of the two approaches a brief introduction to signal synthesis, Gabor expansion or Gabor transform, is first given.

In expansions it is desired to represent a signal as a combination of elements or kernels of an expansion space. The expansion space is typically dense in the original space of signals, and thus, capable of representing any signal with an arbitrary accuracy. This is the case for example with the Fourier transform where a signal is represented as a sum of orthogonal wave forms on different frequencies, expansion functions. Similarly Gabor functions may form an expansion space, where the distinct advantage is a representation by optimally localized time-frequency kernels. The expansion functions do not necessarily have to constitute an orthogonal basis, such as in the wavelet or Fourier transforms, but a frame as is the case with Gabor functions [36]. In the linear case, any sufficiently well behaving signal can be represented as a linear combination, sum, of expansion functions, e.g., wavelets [36, 92]. This was also the main motivation of Gabor, when he proposed that the Gabor elementary functions in Eq. (2.6) can be used as the expansion functions and this kind of transform would represent time-frequency content of signals. The expansion would be optimal at least in the sense of the uncertainty principle; the expansion functions can provide the minimal information quantum. In practice, a signal $s(t)$ is represented as a sum of finite number of the Gabor elementary functions $\psi_{kl}(t)$ multiplied with specific expansion coefficients a_{kl}, Gabor coefficients, as

$$s(t) = \sum a_{kl} \psi_{kl}(t) \tag{2.8}$$

where

$$\psi_{kl}(t) = e^{-\alpha^2 (t-kt_0)^2} e^{j2\pi l f_0 t+\phi} \tag{2.9}$$

where k denotes time shifts and l frequency shifts of Gabor function. For an efficient computation of the expansion coefficients a_{kl} a biorthogonal function set is usually used in the analysis [6, 36], which is trivial for the orthogonal wavelets but more complicated for the non-orthogonal Gabor functions. It has been shown that the iterative method Gabor proposed for the calculation is impractical [50] and more stable solutions have been introduced by Bastiaans [6, 8, 9]. A unique solution can be found for only a special case of the time and frequency step parameters (k,l), that is, when the signal is critically sampled ($t_0 f_0 = 2\pi$), but even then the convergence to the correct solution may be unstable [6, 8, 58, 117, 141]. For the values $t_0 f_0 < 2\pi$, when the frame constituted by the Gabor functions, the Gabor lattice, is overcomplete, the solution is no longer unique, but still various methods converging to proper biorthogonal functions have been proposed [9, 85, 120, 152, 155]. In the case of the overcomplete representation there are always infinitely many solutions and some of them may even be very undesired as analysis functions since the localization is lost [85].

Despite the difficulty of the calculation of the biorthogonal counterparts, the Gabor expansion has some advantages compared to other time-frequency methods [119] and it has been successfully applied to time-frequency signal analysis [48, 153] and used in applications [29, 89]. Lately, inspired by wavelet theory the Gabor expansion has been analyzed in the context of the filter banks [16, 17] and the frame theory related to the Gabor expansion (e.g. [42]) has been extended toward the wavelets by Daubechies et al. in 2003 by introducing framelets [37].

2.3 Gabor filter

In signal processing the Gabor elementary functions can be considered as the optimal time-frequency synthesizing elements. The optimality is also a physical fact since the Gaussian form in the frequency domain can be physically implemented with simple hardware as was proposed already by Gabor himself [49, Appendix (9.1),Appendix (9.2)]. On the other hand, if the synthesis is not of interest but one needs optimally time-frequency localized analysis functions, the solution by definition has the form of the Gabor elementary function and again the sensing elements can be implemented with simple hardware (e.g. [70]). The same theories utilized in the Gabor expansion can be applied to the signal analysis with Gabor functions, but usually a more common approach is from the context of the short-time Fourier transform (STFT). The Gabor expansion and the reconstruction from the short-time Fourier transform lead to the equivalent reconstruction formulas [119].

Next the Gabor elementary function is utilized as a signal analyzing filter. It is natural to operate with a linear filter when the convolution operator describes the action for observing the signal. To be consistent with the linear filter theory a few justifications must be made to Eq. (2.6) in order to define a proper form of Gabor filter, or more precisely, a Gabor filter function since the continuous domain is still assumed. First, the Gabor function is concentrated near the time instant t_0 and for the convolution an origin centered filter form is preferred ($t_0 = 0$)

$$\psi(t) = e^{-\alpha^2 t^2} e^{j2\pi f_0 t + \phi} \tag{2.10}$$

By omitting t_0 the localization is not lost but can still be achieved via the convolution which can be calculated at any location t_0. In addition, there is no evidence that any specific phase (ϕ) would be more beneficial than any other, and furthermore, a response on an arbitrary filter phase can be constructed due to the quadrature relationship between the real and imaginary parts. Moreover, for the functions to be similar at all locations, the phase shift ϕ should depend on the location (t_0), and thus, the phase shift can be removed from the origin centered filter ($\phi = 0$) to have the Gabor filter function in a more compact form

$$\psi(t) = e^{-\alpha^2 t^2} e^{j2\pi f_0 t} \tag{2.11}$$

Now, if Eq. (2.11) is used as a linear filter, the filter response at some location t_1 can be calculated with the convolution

$$\begin{aligned} resp(t_1) = \psi(t_1) * \xi(t_1) = \int_{-\infty}^{\infty} \psi(t_1 - t)\xi(t)dt = \int_{-\infty}^{\infty} e^{-\alpha^2(t_1-t)^2} e^{j2\pi f_0(t_1-t)} \xi(t)dt \\ = \int_{-\infty}^{\infty} \underbrace{e^{j2\pi f_0 t_1} e^{-\alpha^2(t_1-t)^2}}_{\text{window}} \xi(t) e^{-j2\pi f_0 t} dt \end{aligned} \tag{2.12}$$

which is actually similar to the equation of the continuous short-time Fourier transform on frequency f_0 with a Gaussian window function at location t_1 [2, 64]. The only difference is the time instant specific phase factor $e^{j2\pi f_0 t_1}$ in the window function, which is a consequence of the filter definition where a 0° phase angle is assumed always in the filter centroid ($t_0 = 0$). In other words, in the STFT only the window is moved, but in the convolution with the Gabor filter also the frequency waveforms are shifted. As a result, the response magnitude of the Gabor filter function is the same as the magnitude of the Gaussian window short-time Fourier transform, but the phase is different. In practice, STFT window functions are compactly supported, but computationally properties, such as the uncertainty principle, often hold also for the short-time Fourier transform with the Gaussian window and its magnitude, spectrogram [64] (Gabor spectrogram [45]). Symmetrically with the Gabor expansion a reconstruction of a signal can be achieved from the STFT coefficients [1, 2] and, an interesting result, also some of the lost phase information in the spectrogram can be compensated in an approximate reconstruction from the magnitude information only [53, 150]. For the STFT with the Gaussian window a wavelet type structure, S transform, have been introduced [132], which overcomes some of the problems in the STFT, e.g., a constant window size which has been criticized by Daubechies in a comparison of the Gabor expansion and wavelets [36].

2.3.1 Filter normalization

More justifications can be made to define the Gabor filter function in Eq. (2.11) properly. First, the Gabor filter function in Eq. (2.11) has the same effective width, the duration and bandwidth, regardless of the central frequency f_0. Equal duration on all frequencies is not necessary but a similar approach as used in multi-resolution analysis and wavelets can be introduced to make the filters on different frequencies behave as scaled versions of each other [36]. Intuitively this makes sense since it is reasonable to inspect the same events but in different scales; the higher frequency the finer the details. Later this issue will be revisited when scale invariant properties of the Gabor filters are inspected. As a result, the time duration of the filter should depend on the central frequency f_0 to guarantee that the filters on different frequencies are scaled versions of each other [55]. This can be accomplished by substituting

$$\alpha = \frac{|f_0|}{\gamma} \tag{2.13}$$

where f_0 is the central frequency of the filter and γ controls the sharpness of the filter (time duration and bandwidth). Now the time duration of the filter is connected to the filter frequency. The same relationship between the time duration and the frequency has been reported for the S transform where $\gamma = \sqrt{2}$ is fixed [132], but in Eq. (2.13) γ can be adjusted depending on the application.

Now, if the scale factor Eq. (2.13) is substituted to the Gabor filter function in Eq. (2.11)

$$\psi(t) = e^{-(\frac{|f_0|}{\gamma})^2 t^2} e^{j2\pi f_0 t} \tag{2.14}$$

it is evident that the maximum response depends on the selected frequency f_0. From the Fourier transformed Gabor function in Eq. (2.7) it can be seen that the maximum response for a complex signal is $\sqrt{\pi/\alpha^2}$, and thus,

$$\sqrt{\frac{\alpha^2}{\pi}} \tag{2.15}$$

can be used as a normalization term for the response. Finally, a normalized 1-d Gabor filter function can be defined

$$\psi(t) = \sqrt{\frac{\alpha^2}{\pi}} e^{-\alpha^2 t^2} e^{j2\pi f_0 t} = \frac{|f_0|}{\gamma\sqrt{\pi}} e^{-(\frac{|f_0|}{\gamma})^2 t^2} e^{j2\pi f_0 t} \tag{2.16}$$

and by substituting $f_0 = f$ in a more simpler form as

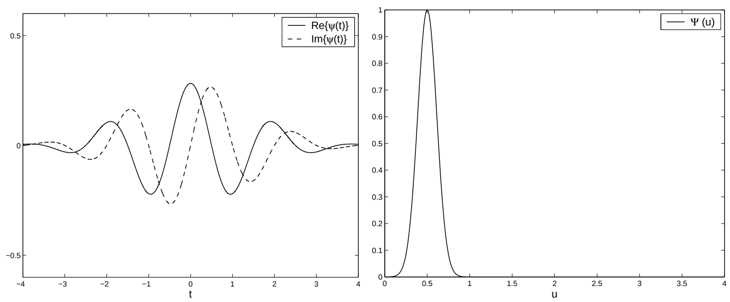
$$\psi(t) = \frac{|f|}{\gamma\sqrt{\pi}} e^{-(\frac{f}{\gamma})^2 t^2} e^{j2\pi f t} \tag{2.17}$$

which is the function referred to as the 1-d Gabor filter throughout this thesis. The 1-d Gabor filter function has an analytical form

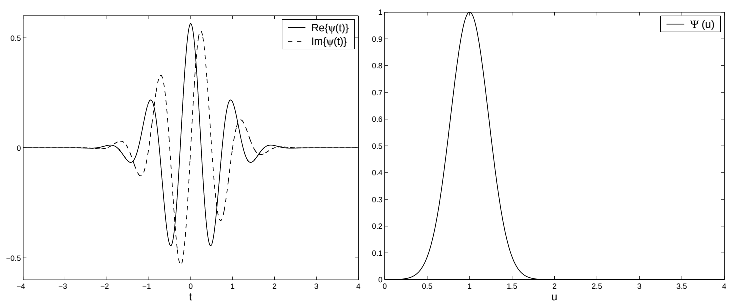
$$\Psi(u) = e^{-(\frac{\gamma\pi}{f})^2 (u-f)^2} \tag{2.18}$$

in the Fourier domain, where u denotes frequency. Note that u has replaced f in the Fourier transform in Eq. (2.1) since f denotes here the filter frequency.

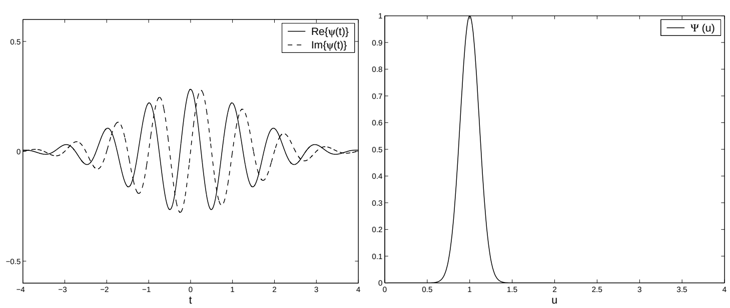
1-d Gabor filter functions in the time and frequency domains for various values of the parameters f and γ are shown in Fig. 2.1. Fig. 2.1(a) and Fig. 2.1(b) illustrate how the filters are scaled versions of each other for a constant γ. On the other hand, γ can be used to adjust the effective width of the filter; by increasing γ the filter spreads in the time domain and shrinks in the frequency domain as shown in Fig. 2.1(b) and Fig, 2.1(c).



(a)



(b)



(c)

Figure 2.1: Gabor filter functions in time and frequency domains for different values of parameters f and γ: (a) $f = 0.5,\ \gamma = 1$; (b) $f = 1.0,\ \gamma = 1$; (c) $f = 1.0,\ \gamma = 2$.

The 1-d Gabor filter function is a non-causal linear filter with an infinite support, that is, all the past and forthcoming inputs affect the response. This break of causality prevents the use of the filter in some specific applications but the filter can be used for stored signals, such as digital images. The filter has an infinite impulse response, but the response fades toward to zero rapidly on both sides of the origin $t = 0$. Furthermore, the filter function has a real Fourier transform in Eq. (2.18), the frequency response, and thus, it does not affect the phase of the input signal but has a constant phase response.

2.4 2-d Gabor filter

It is straightforward to generalize the previously defined theories to two dimensions, where the time variable t is replaced by the spatial coordinates (x, y) in space and the frequency variable u by the frequency variable pair (u, v), but this was not the case until the late 70's when the 2-d Gabor filters started to attain a considerable amount of attention. There were again two branches in the development of the Gabor methods, Gabor expansion and Gabor filters, the early works mainly utilized the expansion and later the Gabor filters have received more attention. If similar studies in the optics are not considered (e.g. [7, 26, 34, 90, 99]), the major impact on the development and use of 2-d Gabor filters has been image processing and especially 2-d feature extraction. The development of the 2-d Gabor elementary functions began from Granlund in 1978, when he defined some fundamental properties and proposed the form of a general picture processing operator. The general picture processing operator had a form of the Gabor elementary function in two dimensions and it was derived directly from the needs of the image processing without a connection to Gabor's work [52]. It is noteworthy that Granlund addressed many properties, such as the octave spacing of the frequencies, that were reinvented later for the Gabor filters. After Granlund's groundwork, the results were revisited in the context of the uncertainty principle in continuum of his work in the same research group but accompanied with Wilson and Knuttson [71, 145, 146, 147].

Despite the original contribution of Granlund et al. the most referred works seem to be those conducted by Daugman [38, 39]. Daugman was the first who exclusively derived the uncertainty principle in two dimensions and showed the surprising equivalence between a structure based on the 2-d Gabor functions and the organization and the characteristics of the mammalian visual system. Daugman defined similar uncertainty measures as in Eq. (2.4) for the 2-d space and spatial-frequency domain in terms of Δx, Δy, Δu, and Δv for which it holds that

$$\Delta x \Delta u \geq \frac{1}{4\pi}, \quad \Delta y \Delta v \geq \frac{1}{4\pi}, \quad \text{and } \Delta x \Delta u \Delta y \Delta v \geq \frac{1}{16\pi^2} \tag{2.19}$$

The most general form of the functions that achieve the lower bound of the uncertainty inequalities in Eq. (2.19) is

$$e^{-(Ax^2+Bxy+Cy^2+Dx+Ey+F)} \tag{2.20}$$

where $B^2 < 4AC$ and D,E, and F are complex [38], but for practical use certain simplifications can be made. The following simplifications are particularly motivated if the construction of the 2-d Gabor filters is done according to biological evidences (e.g. [38]), but here they will be justified only based on the regulations in the linear filter theory. First, a filter form of the Gabor elementary function is preferred, which is the function centered to the origin ($Re\{D\} = Re\{E\} = Re\{F\} = 0$) and as in the 1-d case there is no phase shift ($Im\{F\} = 0$). Furthermore, since the 2-d Gabor filter is a product of an elliptical Gaussian in any rotation times a complex exponential representing harmonic modulation on any spatial frequency and any orientation, it can be assumed for simplicity that the orientation of the Gaussian and the harmonic modulation are the same ($B = 0$). Every orientation and all filter shapes can be spanned despite the assumed simplifications. By applying the given simplifications, a proper form of the 2-d Gabor elementary function can be defined as

$$\begin{aligned} \psi(x, y) &= e^{-(\alpha^2 x'^2 + \beta^2 y'^2)} e^{j 2\pi f_0 x'} \\ x' &= x \cos\theta + y \sin\theta \\ y' &= -x \sin\theta + y \cos\theta \end{aligned} \tag{2.21}$$

To provide a similar shape of the filter function regardless of the frequency f_0 substitutions $\alpha = |f_0|/\gamma$ and $\beta = |f_0|/\eta$ can be made, where the sharpness of the filter along the major and minor axes are now controlled by γ and η. The filter response can be normalized to have a compact closed form of the normalized 2-d Gabor filter function

$$\begin{aligned} \psi(x, y) &= \frac{f^2}{\pi\gamma\eta} e^{-(\frac{f^2}{\gamma^2} x'^2 + \frac{f^2}{\eta^2} y'^2)} e^{j 2\pi f x'} \\ x' &= x \cos\theta + y \sin\theta \\ y' &= -x \sin\theta + y \cos\theta \end{aligned} \tag{2.22}$$

where f is the central frequency of the filter, θ the rotation angle of the Gaussian major axis and the plane wave, γ the sharpness along the major axis, and η the sharpness along the minor axis (perpendicular to the wave). In the given form, the aspect ratio of the Gaussian is $\lambda = \eta/\gamma$. The normalized 2-d Gabor filter function has an analytical form in the frequency domain

$$\begin{aligned} \Psi(u, v) &= e^{-\frac{\pi^2}{f^2}(\gamma^2 (u' - f)^2 + \eta^2 v'^2)} \\ u' &= u \cos\theta + v \sin\theta \\ v' &= -u \sin\theta + v \cos\theta \end{aligned} \tag{2.23}$$

The effects of the parameters, interpretable via the Fourier similarity theorem, are demonstrated in Figs. 2.2(a)-2.3(b) and gathered to Table 2.1.

It is interesting that of the first contributors only Granlund used the filtering approach in the beginning [52] when in many famous papers, such as the one by Daugman [39] or by Porat et al. [113], the Gabor expansion was used and maybe confusingly compared to the human visual system, where the Gabor filters should be used.

2.5 Discrete filters

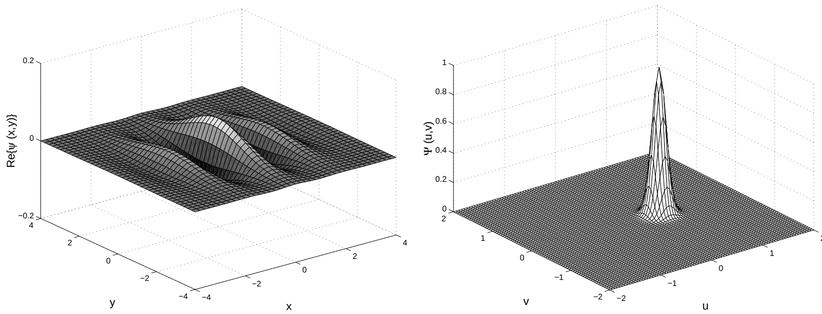
In the transform of continuous domain entities to the discrete domain one always needs to be sure that entities are represented with sufficient accuracy in order to reapply the continuous domain theories. This is the case also with the Gabor filters and their discrete domain representations. Discrete versions of the Gabor expansion have been addressed in several studies (e.g. [8, 9, 85, 117, 120, 141]), but the main concern has been biorthogonal

Table 2.1: Corresponding filter properties in space and spectral domains. [38]

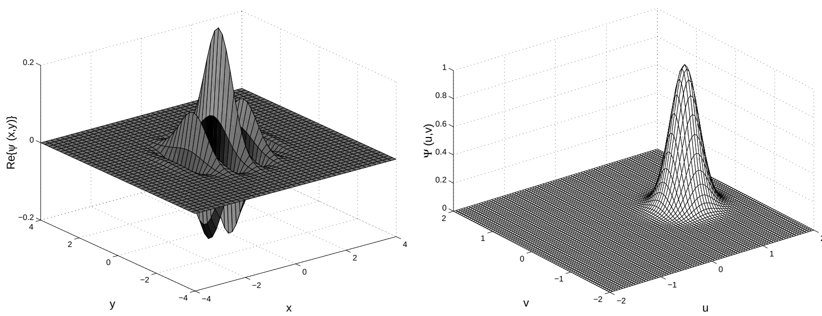
2-d space domain	**2-d spatial-frequency domain**
1. Set the frequency of the plane-wave to ω_0 and orientation θ_0 (Fig. 2.2(a) and Fig. 2.2(b)).	1. Position spectral centroid at Fourier plane coordinates (u_0, v_0), where $u_0 = \omega_0 cos(\theta_0)$ and $v_0 = \omega_0 sin(\theta_0)$ (Fig. 2.2(a) and Fig. 2.2(b)).
2. Stretch (compress) filter in x direction ($\theta = 0$) by factor γ (Fig. 2.3(a)).	2. Compress (stretch) spectrum in u direction by factor γ (Fig. 2.3(a)).
3. Stretch (compress) filter in y direction ($\theta = 0$) by factor η (Fig. 2.3(a)).	3. Compress (stretch) spectrum in v direction by factor η (Fig. 2.3(a)).
4. Rotate filter through angle θ around origin of coordinates (Fig. 2.3(b)).	4. Rotate spectrum through angle θ around origin of coordinates (Fig. 2.3(b)).

function construction implementations, not the sampling and quantization of the Gabor and biorthogonal functions and the expansion coefficients. It should be noted that some misleading concepts used in the studies, such as critical sampling, undersampling, and oversampling, often refer to the spacing of the Gabor functions and not to the sampling of the functions themselves. These sampling schemes are related to the frame properties of a space spanned by the elementary functions and affecting the computation of the biorthogonal functions [9]. The expansion methods have been transfered to even more exotic sampling schemes, such as quincunx lattices [140]. The more traditional meanings of sampling and quantization in the discrete representation of the continuous functions may seem trivial for the Gabor filters, but in this thesis where the most of the results are based on real experiments (*Publication I*–*Publication VIII*) and it is thus sound to have an insight into the discrete Gabor filters as well.

A proper number of sampling points and adequate quantization accuracy must be established to represent the Gabor functions, biorthogonal functions, filter responses and the expansion coefficients with a sufficient accuracy. The quantization of the coefficients and responses has received some attention and particularly it has been shown that an accurate quantization of the phase information is more important than the magnitude information [113], which is a known result in the signal reconstruction [106, 139] and appears to contribute in the recognition and detection results as well (e.g. *Publication IV*). However, recently it has been shown how reconstructions succeed also from local magnitudes [150]. In practice, the signals are usually strictly amplitude limited and the quantization levels are not thus an issue, but always a sufficient amount of sample points must be used in the discrete representation of the functions and signals in order to achieve reliable results. A proper construction is an implication from sampling theorem, but it has been only briefly visited for the Gabor filters (e.g. [18, 151], *Publication III*). In applications it is sometimes necessary to explicitly apply discrete domain restrictions as for example is the case in optimization of the Gabor filter parameters using gradient information in *Publication V*.



(a)



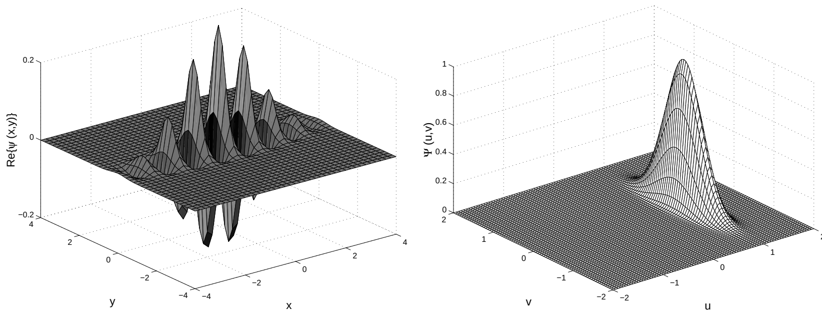
(b)

Figure 2.2: 2-d Gabor filter functions with different values of the parameters f, θ, γ, and η in the space and spatial-frequency domains: (a) $f = 0.5,\ \theta = 0°,\ \gamma = 1.0,\ \eta = 1.0$; (b) $f = 1.0,\ \theta = 0°,\ \gamma = 1.0,\ \eta = 1.0$.

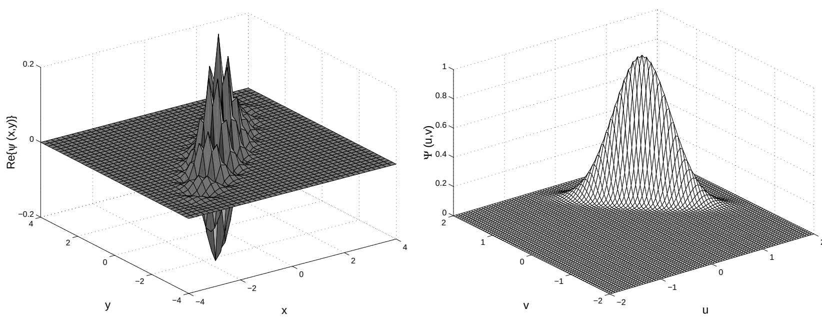
The sampling theorem states that any band-limited signal can be reconstructed from a discrete representation if the sampling frequency is at least twice the highest frequency component in the signal [127]. Respectively an accurate representation of a discrete Gabor filter can be guaranteed if the aliasing is negligible, that is, in the discrete frequency domain the filter does not spread over the valid frequency range from -0.5 to 0.5. However, the avoidance of the aliasing in the frequency domain only guarantees a proper sampling of the filter, but it does not say anything about a proper filter in the time domain. In the time domain it would be beneficial if the filter envelope falls completely in discrete bins of a filter representation, that is, the filter values are negligible outside the filter vector.

As any continuous function, the Gabor filter in Eq. (2.17) can be expressed as a discrete sequence

$$\psi(n) = \psi(Tn) \tag{2.24}$$



(a)



(b)

Figure 2.3: 2-d Gabor filter functions with different values of the parameters f, θ, γ, and η in the space and spatial-frequency domains: (a) $f = 1.0,\ \theta = 0°,\ \gamma = 2.0,\ \eta = 0.5$; (b) $f = 1.0,\ \theta = 45°,\ \gamma = 2.0,\ \eta = 0.5$.

If only discrete signals are considered and the relation between the discrete and actual frequencies can be avoided, unit sampling $T = 1$ can be assumed and the discrete filter can be written as

$$\psi(n),\ n \in \{-\frac{L-1}{2}, \ldots, 0, \ldots, \frac{L-1}{2}\} \tag{2.25}$$

for an odd number L of filter bins and

$$\psi(n),\ n \in \{-\frac{L}{2}, \ldots, 0, \ldots, \frac{L}{2} - 1\} \tag{2.26}$$

for even L. The minimum and maximum indices of the filter can be obtained from $n_{min} = (-\frac{L-1}{2} \ \vee\ -\frac{L}{2})$ and $n_{max} = (\frac{L-1}{2} \ \vee\ \frac{L}{2} - 1)$ respectively depending on whether L is odd or even. Now, a discrete 1-d normalized Gabor filter can be defined as

$$\psi(n) = \frac{|f|}{\gamma\sqrt{\pi}} e^{-(\frac{f}{\gamma})^2 n^2} e^{j2\pi f n} \tag{2.27}$$

where f denotes discrete frequencies between $\frac{n_{min}}{L}$ and $\frac{n_{max}}{L}$. Since the normalized form of the filter is zero at the zero frequency and on the other hand, the maximum and minimum frequencies are not attractive due to the aliasing, it is reasonable to use filters either on the positive frequencies $\{\frac{1}{L}, \ldots, \frac{n_{max}-1}{L}\}$ or the negative frequencies $\{\frac{n_{min}+1}{L}, \ldots, -\frac{1}{L}\}$. To distinguish between the positive and negative frequencies and to obey the sampling theorem the filter must have negligible values outside the range from 0 to Nyquist frequency ($\frac{n_{min}}{L}$ or $\frac{n_{max}}{L}$). To measure the aliasing of the filter in Eq. (2.27) a corresponding Fourier domain filter can be utilized

$$\Psi(n) = e^{-(\frac{\gamma\pi}{f})^2(\frac{n}{L}-f)^2} \tag{2.28}$$

In Fig. 2.4 are shown the positive ($A_2 :\ 0 < f < 0.5$) and negative ($A_1 :\ -0.5 < f < 0$) ranges of the allowable discrete frequencies and the aliasing effect ($A_3 \rightarrow A_2$ and $A_4 \rightarrow A_1$) is illustrated. To ensure that a filter is properly inside the frequency range, that is,

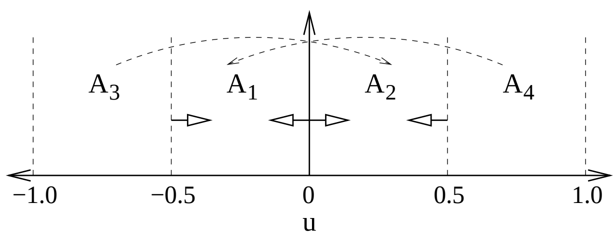


Figure 2.4: Aliasing in the discrete frequency domain.

at most the proportion $1 - p_f$ is aliased, it must hold that

$$\begin{aligned} \sum_{n=1}^{n_{max}-1} \Psi(n) \geq p_f \sum_{n=-\infty}^{\infty} \Psi(n),\ 0 < f < 0.5 \\ \sum_{n=n_{min}+1}^{-1} \Psi(n) \geq p_f \sum_{n=-\infty}^{\infty} \Psi(n),\ -0.5 < f < 0 \end{aligned} \tag{2.29}$$

Discrete constraints can be defined in cases when the positive and negative frequencies must be distinguished. For positive frequencies $f = \frac{n_f}{L}$

$$\begin{aligned} &\sum_{n=1}^{n_{max}-1} \Psi(n) \geq p_f \sum_{n=-\infty}^{\infty} \Psi(n) \\ \Rightarrow &\sum_{n=1}^{n_f} \Psi(n) + \sum_{n=n_f+1}^{n_{max}-1} \Psi(n) \geq p_f \sum_{n=-\infty}^{\infty} \Psi(n) \end{aligned} \tag{2.30}$$

which can be lower bound estimated with the continuous integrals

$$\begin{aligned} &\int_0^{\frac{n_f}{L}} \Psi(u)du + \int_{\frac{n_f}{L}}^{\frac{n_{max}}{L}} \Psi(u)du \geq p_f \int_{-\infty}^{\infty} \Psi(u)du \\ \Rightarrow &\int_0^{\frac{n_{max}}{L}} \Psi(u)du \geq p_f \int_{-\infty}^{\infty} \Psi(u)du \end{aligned} \tag{2.31}$$

Furthermore, a lower bound estimation which holds for both, even and odd L, can be obtained by setting $\frac{n_{max}}{L} = \frac{L/2-1}{L} = \frac{1}{2} - \frac{1}{L}$ when the inequality can be simplified to

$$\frac{1}{2}erf(\gamma\pi) + \frac{1}{2}erf\left(\frac{\gamma\pi}{|f|}(\frac{1}{2} - \frac{1}{L}) - \gamma\pi\right) \geq p_f \tag{2.32}$$

where erf is the error function

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \tag{2.33}$$

Eq. (2.32) holds also for the negative frequencies ($-0.5 < f < 0$).

In addition to proper construction in the limits of the frequency domain, which is achieved by satisfying the inequality Eq. (2.32), also proper construction in the time domain should be realized. The equation Eq. (2.32) indeed provides an accurate representation of the filter in the time domain, but the filter envelope may spread beyond the discrete size of the filter. To prevent the overflow, the same approach can be applied. The effective filter envelope is specified by the modulating Gaussian, and thus,

$$\sum_{n_{min}}^{n_{max}} e^{-(\frac{f}{\gamma})^2 n^2} \geq p_t \sum_{-\infty}^{\infty} e^{-(\frac{f}{\gamma})^2 n^2} \tag{2.34}$$

By lower bound estimations a general constraint that holds for odd and even L can be derived and the time domain constraint can be written as

$$erf\left(\frac{|f|}{\gamma}(\frac{L}{2} - 1)\right) \geq p \tag{2.35}$$

Equation Eq. (2.35) can also be used to estimate the minimum size L of the filter in the time domain.

In *Publication III* the constraints Eq. (2.32) and Eq. (2.35) are generalized also to two dimensions. To define a strict lower boundary L can be set equal to the size of the smallest dimension of the 2-d filter, $L = \min\{L_x, L_y\}$, and the filter can be considered in the standard pose where the major axis is along the x-axis ($\theta = 0$). Now, the frequency constraint for the 2-d filter is

$$\frac{1}{4}\left[erf(\gamma\pi) + erf\left(\frac{\gamma\pi}{|f|}(\frac{1}{2} - \frac{1}{L}) - \gamma\pi\right)\right]\left[erf(\eta\pi) + erf\left(\frac{\eta\pi}{|f|}(\frac{1}{2} - \frac{1}{L}) - \eta\pi\right)\right] \geq p_f \tag{2.36}$$

and the spatial constraint

$$erf\left(\frac{|f|}{\gamma}(\frac{L}{2} - 1)\right) erf\left(\frac{|f|}{\eta}(\frac{L}{2} - 1)\right) \geq p_t \tag{2.37}$$

Finally, using the discrete forms of the Gabor filter in Eq. (2.27) or Eq. (2.28) or corresponding 2-d forms, and by satisfying the given constraints Eq. (2.32) and Eq. (2.35) or Eq. (2.36) and Eq. (2.37) a proper construction of the Gabor filters can be achieved and reliable results in applications can be expected.

Chapter III

Gabor Features

In the previous chapter signal analysis was considered in the joint time-frequency domain and Gabor filters were introduced as basic operator kernels extracting the minimal amount of information. This chapter concentrates on the use of the filters and selection of the filter parameters for recognizing and localizing events in signals. Gabor filters are the first step in processing and refining raw data to more informative features which can be utilized by upper processing layers to extract higher level or even abstract information. It has been already shown how the two domain representation is an advantage, but in addition, the smooth differentiable form of the Gabor filter itself and distinguishing properties of the features provide an intriguing framework for feature extraction in the varying conditions encountered in many detection, recognition, and classification problems.

3.1 Time-frequency features of 1-d signals

As has been already discussed, many signals have characteristics which cannot be processed in time or frequency separately, but a joint time-frequency representation must be used. In general, joint time-frequency representations include a wide variety of methods, such as the short-time Fourier transform, spectrogram, Gabor transform, Wigner-Ville distribution, time-frequency kernel methods, and even the wavelet transform [31, 33, 59]. All these methods have their advantages and disadvantages, but all of them compound time and frequency simultaneously and are applicable to 1-d signal processing. The short-time Fourier transform, spectrogram, Gabor transform, and wavelet transform are linear time-frequency representations while the methods derived from the general time-frequency distribution, such as Wigner-Ville and the kernel distributions, are quadratic. The quadratic distributions aim to overcome some of the problems of linear techniques and provide a high-resolution time-frequency representation [33]. It is no surprise that there is a trade-off and the better resolution can be achieved only at the price of cross term artifacts that often appear at time-frequency locations where no signal exists and can also overlap actual signal components with a magnitude that can be larger than the actual signals that created the cross terms [33, 67]. Different filtering kernel approaches can be used to mitigate these cross-term artifacts, but always at the expense of degrading the time-frequency resolution and eventually the main advantage is lost [67]. The Gabor expansion has been proposed as a filtering kernel for the Wigner-Ville distribution [118, 119]. However, all of the methods mentioned can generate a time-frequency representation of a signal, but the representation itself is useless without further processing, e.g., the feature extraction. The advantages and disadvantages of the different time-frequency representations may become more evident in the feature extraction context, and thus, the main advantages which have made Gabor filters attractive in applications are addressed next.

The normalized 1-d Gabor filter can be used to extract time-frequency features, Gabor features. The experimental part of this thesis utilizing the most central properties of the Gabor features, reported in the publications, is mainly done in two dimensions (*Publication I*, *Publication II*, *Publication III*, *Publication IV*, *Publication VI*, and *Publication VIII*), but here the properties are first demonstrated for 1-d case for simplicity. Gabor features of a signal $\xi(t)$ can be generated with the normalized 1-d Gabor filter in Eq. (2.17) via the convolution

$$r_\xi(t;f) = \psi(t;f) * \xi(t) = \int_{-\infty}^{\infty} \psi(t - t_\tau; f)\xi(t_\tau)dt_\tau \tag{3.1}$$

where the frequency f is included as a feature variable to the representation of the Gabor filter. Via the convolution the Gabor filter works as any linear filter and the filter response $r_\xi(t;f)$ in Eq. (3.1) can be calculated at any time instant t and for various frequencies f to extract features at different times and on different frequencies. The Gabor features can carry all signal information, but in a different form since an arbitrary amount of the original signal can be recovered from the filter responses as reported by Bastiaans for the Gabor expansion [6, 8, 9]. In addition to this reconstruction property, operations to perform a deformation, such as translation and scaling, independent detection of signal events can be established.

3.1.1 Translation property

The first and most obvious use of the time-frequency representation is to detect time varying changes in the frequency content of a signal. The changes can be temporal short duration events or static events which last over a particularly long duration. The division to the temporal and long-lasting events is not of importance but the change itself in the frequency content is important. Such salient events can, for example, be the heart beat pulses in ECG signals [84], phonemes in speech [138], mechanical fault impulses in machinery diagnostics ([156], *Publication V*, *Publication VII*), and in general, transient sub-parts of signals [48]. The time varying changes are undoubtedly important features in many problems since they provide information of the instantaneous content [25, 33, 48]. There are many successful applications utilizing different time-frequency features and it is not probable that the Gabor filters would be superior in every application, but the utilization of the uncertainty principle can always be used as a starting point providing the highest reachable resolution jointly in time and frequency. If the bandwidth or bandwidths of an event to be inspected are known, the smallest time window for the detection of the event can be defined by the uncertainty principle. If the time axis is tightly covered with the windows and overlap of them is sufficient, an event in an arbitrary position will always fall on the effective area of some of the windows. On this basis a translation invariant search can be established for the Gabor features. The translation property is one of the central properties used with Gabor features and it generally holds for all time invariant systems.

For the 1-d Gabor feature in Eq. (3.1) and a translated version $\xi_1(t)$ of some signal $\xi(t)$,

$$\xi_1(t) = \xi(t - t_1) \tag{3.2}$$

it can be shown that

$$\begin{aligned} r_{\xi_1}(t; f) &= \int_{-\infty}^{\infty} \psi(t - t_\tau; f)\xi_1(t_\tau)dt_\tau \\ &= \int_{-\infty}^{\infty} \psi(t_\tau; f)\xi_1(t - t_\tau)dt_\tau \\ &= \int_{-\infty}^{\infty} \psi(t_\tau; f)\xi(t - t_1 - t_\tau)dt_\tau \\ &= \int_{-\infty}^{\infty} \psi((t - t_1) - t_\tau; f)\xi(t_\tau)dt_\tau \\ &= r_\xi(t - t_1; f) \end{aligned} \tag{3.3}$$

which is the translation property of the Gabor filter allowing a translation invariant detection of events. The translation property is common for most linear time-frequency representations and in general it holds for functions with a continuous translation parameter.

3.1.2 Scale property

While the translation property was recognized as an advantage already in the earliest studies of the short-time Fourier transform, the scale property did not gain any significant attention until the introduction of wavelets. The scale property is an equivalence of the transformation property but in the frequency space. To be consistent with the previous chapter, the term frequency should be used instead of scale since they are controlled by the same parameter, f, which denotes the central frequency of the filter, but actually the scale and frequency concepts are shown to become equivalent. The scale property was not recognized within the Gabor functions since a linear rectangular lattice and a constant sharpness, effective duration, of the filter were dominating in the time-frequency methods. The rectangular lattice illustrated in Fig. 3.1 was the same as proposed by Gabor [49] and used by Bastiaans [6, 8, 9] and it has had a strong impact on the development of Gabor analysis [42]. Actually it is exactly this selection of parameters that makes Gabor filtering merely a special case of the short-time Fourier transform rather than a time-scale method [36]. In a division of the mathematical properties of the spaces and frames both, the Gabor filtering and the wavelet transforms, belong to the Lie group where a more strict subdivision puts the short-time Fourier transforms in the Weyl-Heisenberg group and the wavelets in the "ax+b" (affine) group [36, 42]. The name coherent states, denoted by the small circles in Fig. 3.1, is adopted from physics where the time-frequency space is often referred to as a phase space [36].

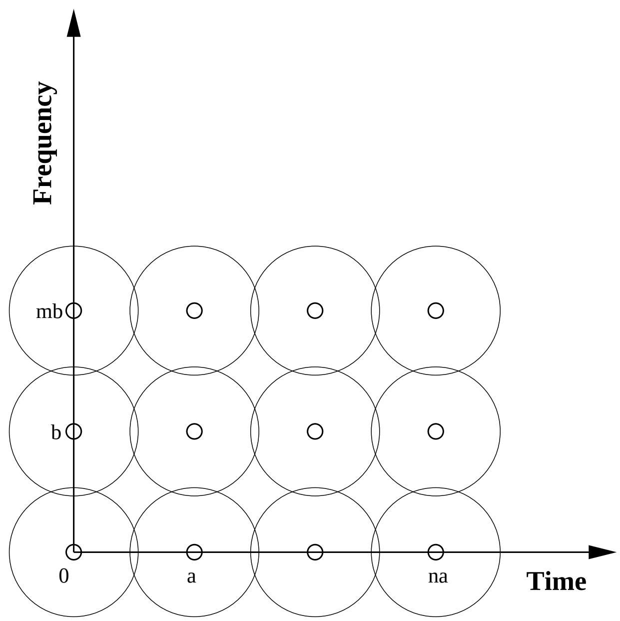


Figure 3.1: Rectangular lattice indicating spacing of Gabor functions $\psi(t - na; mb)$ in time-frequency plane.

If the sharpness α in the Gabor function in Eq. (2.11) is selected to be constant and not dependent on the frequency f, the envelope size of the Gabor filter is the same regardless of the frequency f, that is, there are more waves inside the filter envelope on higher frequencies; filters do not represent scale information and f cannot be interpreted as the scale. That is the case with most short-time Fourier transform approaches and also the original form of the Gabor expansion. The representation was strongly argued by Daubechies who claimed distinct advantages for the affine group as compared to the Weyl-Heisenberg group [36]. However, it does not have to be necessarily the case with the Gabor functions and a condition moving the Gabor functions toward the affine group was made by connecting the sharpness α, denoting also the wavelength, and the frequency f as shown in Eq. (2.13) [55], a condition which is applied in the given form of the Gabor filter. In the forms in Eqs. (2.17) and (2.18) the filters are scaled versions of each other as in the affine group; the filters extract exactly the same information but in different scales, f denoting the scale parameter. Before a scale property of the Gabor filters can be utilized in scale invariant detection, the rectangular lattice structure is typically discarded and spacing of the frequencies f must be selected in a more proper manner. For a scaled signal

$$\xi_2(t) = \xi(at) \tag{3.4}$$

it holds for the filter response that ([55])

$$\begin{aligned} r_{\xi_2}(t;f) &= \int_{-\infty}^{\infty} \psi(t-t_\tau;f)\xi_2(t_\tau)dt_\tau \\ &= \int_{-\infty}^{\infty} \psi(t_\tau;f)\xi_2(t-t_\tau)dt_\tau \\ &= \int_{-\infty}^{\infty} \psi(t_\tau;f)\xi(at-at_\tau)dt_\tau \\ \hat{t}_\tau &= at_\tau \;\; dt_\tau = \frac{d\hat{t}_\tau}{a} \\ &\Rightarrow \int_{-\infty}^{\infty} \psi(\hat{t}_\tau;\frac{f}{a})\xi(at-\hat{t}_\tau)dt_\tau \\ &= r_\xi(at;\frac{f}{a}) \;\; . \end{aligned} \tag{3.5}$$

The interpretation of the above result, well known for the wavelets, is straightforward: a response for a signal is the same as the response of a similarly scaled filter for a scaled version of the signal. The result itself holds in the rectangular lattice as well, but to maintain homogeny spacing between the scales a logarithmic relation between the frequencies f must be established ([21, 36, 55] and *Publication VIII*)

$$f_k = c^{-k} f_{max}, \text{ for } k = 0, \ldots, m-1 \tag{3.6}$$

where f_{max} is the maximum frequency used and c is the frequency scale factor. Some useful values for c include $c = 2$ for octave spacing and $c = \sqrt{2}$ for half-octave spacing. A logarithmic spacing of the frequencies is illustrated in Fig. 3.2 where there is a logarithmic frequency lattice marked by the small circles and examples of logarithmically spaced Gabor filters. It should be noted that now filters on lower frequencies spread over a substantially larger time duration than filters on high frequencies, and thus, the logarithmic relation can be reflected also to the time domain spacing of the functions as with the wavelets (marked circles and ellipses in Fig. 3.2) [36, 92, 133].

It is interesting that the logarithmic spacing did not receive attention in the Gabor expansion research until the wavelets were introduced. With the given condition and the logarithmic spacing the Gabor expansion can be considered a multi-resolution analysis method [36] or even a Gabor wavelet [55] despite of the non-orthogonality of the Gabor functions. Hitherto it seems awkward whether and when the wavelet theory [133] or the frame theory of time-frequency methods [42] should be applied to Gabor filtering since the wavelet properties of the Gabor functions have only briefly been considered [83]. Results from the first studies on theory encapsulating the non-orthogonal Gabor functions and the multi-resolution analysis, introducing framelets, has been only recently reported [37]. Inspired by the wavelets the time-frequency results were expanded to the scale space as in the S transform [132] or even to the joint of the three quantities time, frequency, and scale [32]. Despite the fact that the wavelet structure within the 1-d Gabor filter has mostly been neglected, similar properties have been mentioned already in the first study of 2-d Gabors by Granlund [52] and later, inspired by the Laplacian pyramid [23] and biological evidences, in many studies (e.g. [78, 81, 113]).

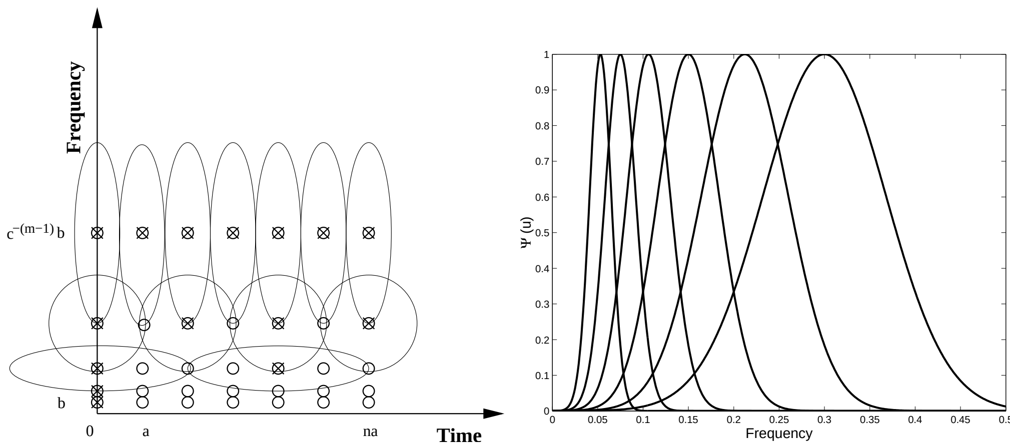


Figure 3.2: Logarithmic lattice indicating spacing of Gabor functions $\psi(t - na; c^{-k}b)$ in the time-frequency plane and an example of Gabor filter functions in the frequency domain.

3.1.3 Applications

It is not a surprise that the possible application areas of Gabor filtering are the same as for all time-frequency and time-scale methods. However, the optimal method always depends on the application and may vary even though some automation can be applied to the selection [122, 123]. Sometimes it is only the analytical, differentiable form of the Gabor filter that makes it attractive to the applications. *Publication V* and *Publication VII* describe an application for detection of bearing faults in electric motors based on the frequency content of the stator current. Generally many faults are impulse-type short duration events whose occurrence time can be detected from the time-frequency representation, but in many systems, such as in motors containing rolling elements, signals are periodic and faults may also be seen in the global frequency content. The proposed approach in *Publication V* and *Publication VII* uses the global frequency content and the Gabor filter mainly contributes to the mathematical simplicity when features are extracted and optimized to maximize the discrimination between two types of signals. However, the given method can be extended to utilize time information, which is beneficial in the segmentation of the stationary parts from the quasi-stationary stator current signal [156]. The wavelets have been used in similar problems and also in short duration fault detection, but often only the translation property is utilized and similar events are not searched for over the scales [86, 87, 116]. Studies of the Gabor expansion for extracting the short duration events, signal transients, etc., can be applied to the feature extraction with the Gabor filters as well [25, 48]. The time-frequency representation may also be used to remove noise from the signal [25, 153] and in some cases a system identification can be improved using a chirp signal and a prefiltering step with time-frequency filters [153].

The use of the scale property to search events in time and scale is rarely used in 1-d signal processing, but it will be revisited in the next section for 2-d images. One application where events are inspected in different scales is the detection of ECG signal characteristics [84]. If image processing is the main application of 2-d Gabor filters, speech processing cannot be bypassed in the context of 1-d Gabor filters. The human hearing system seems to consist of nerves that have an organization of the logarithmic frequency scale and a varying bandwidth very similar to those shown in Fig. 3.2 [121]. Since the sound itself is conveniently processed, synthesized and analyzed, using the sinusoidal waves [95] and the biological sensors describe similar characteristics as the Gabor filters, it is not a surprise that the time-frequency features obeying the logarithmic scale have been so successful in speech processing [51, 91, 138].

Of the given application examples Gabor filters are used only in *Publication V*, *Publication VII*, and [156], but there is no evidence that Gabor filters cannot be used as one general feature extraction framework in all of the mentioned applications by utilizing the translation and scale properties.

3.2 Space–frequency features of 2-d signals

Many of the 1-d methods and theories of time-frequency signal processing can be generalized to 2-d signal processing, but especially in image processing Gabor filters have received a considerable amount of attention. The groundwork in image processing was made in the late 70's and early 80's first by Granlund, who initiated the idea of a general image processing operator [52], and in a continuum connected to the uncertainty by Wilson and Knutsson [71, 145, 147]. Their contemporary John Daugman participated in an active vision system research group and became known to the image processing community for his studies claiming the 2-d Gabor filter as an accurate model of a simple cell in the striate cortex belonging to the mammalian visual system [38, 39], undoubtedly the reason for the great impact of Gabor research in image processing. Motivated by the physiological findings Porat and Zeevi were the first to introduce a scheme for object recognition consisting of a set of Gabor filters with properly selected parameters [113]. It is an interesting detail that Daugman and contemporaries who were well aware of the receptive field profile of the simple cell employed the Gabor expansion and biorthogonals while it is the Gabor filter itself which resembles the simple cell receptive field profile [39, 113]. Since the foundations of the Gabor filters in feature extraction have been established, they have been successfully used in many applications, especially in texture analysis and face recognition. Lately more theoretical works have contributed to Gabor research by studying the filters in the context of wavelets [83].

The normalized 2-d Gabor filter in Eq. (2.22) can be used to extract space-spatial-frequency features as demonstrated in the experiments in *Publication I*, *Publication II*, *Publication III*, *Publication IV*, *Publication VI*, and *Publication VIII*. The feature extraction can be done via the convolution as

$$\begin{aligned} r_\xi(x, y; f, \theta) &= \psi(x, y; f, \theta) * \xi(x, y) \\ &= \int\int_{-\infty}^{\infty} \psi(x - x_\tau, y - y_\tau; f, \theta)\xi(x_\tau, y_\tau)dx_\tau dy_\tau \end{aligned} \tag{3.7}$$

3.2.1 Physiology of vision

It seems that in 1-d signal processing the Gabor functions are mostly applied via the Gabor expansion, but in 2-d processing Gabor filtering has received much more attention. For 1-d signals this can be partly explained by an active use of other time-frequency methods and for 2-d signals one dominating reason to prefer Gabor filters over other possible features has been correspondences found in the physiology of mammalian vision. The physiology claim has promoted Gabor filters and they have been very popular in feature extraction from 2-d images, especially in texture analysis and face authentication, which will be reviewed later. The relevance of these considerations is not emphasized in this thesis but the reasoning is rather based on present theory and achieved properties of the Gabor filters. Still, a short introduction to the topic is here covered for consistency. The following is mainly based on Palmer's book giving a comprehensive introduction and references to the latest results [108].

The vision system is based on a complex, loosely layered, feedback structure where specialized cells in the retina and brain process and transmit the visual information. The Gabor filters are related to one specific type of specialized cells, simple cells, in the striate cortex of the brain as the filters resemble the receptive field profile of simple cells. Even if the receptive field profile is taken as meaning the exhibitive and inhibitive spatial areas in the visual field of an eye, the simple cells are not directly connected to the photo receptive cells in the retina. Actually there are several layers of different kinds of cells already in the retina and correspondingly the responses of the simple cells are passed to upper level cells, such as complex cells, in the striate cortex. This kind of structure can be seen as a chain of processing elements which combine information from the preceding layer and provide more refined and informative input to the next processing level. Still, by measuring directly the receptive field profile and not in correspondence to the previous layer an insight can be gained for a better understanding of the mnemonic signal events that stimulate our brains. It should be noted that responses of some cells cannot be simulated using static structures, and furthermore, some connections may bypass layers, and some connections in the brains even provide feedback from the upper layers. There has been intensive research on the physiology of vision, but at least in the image processing community the most referenced paper seems to be the one by Daugman where he presented a picture of the measured receptive field profiles of the cat's simple cells and showed the significant accuracy of the Gabor filter to embody the cat's simple cell [38]. Daugman also considered the parameters of the Gabor filter from the point of view of the physiology: quadrature phase relation of the real and imaginary components, frequency and orientation bandwidths, octave frequency spacing, and equal orientation of the Gaussian and sinusoidal. The Gabor filter is evidently a too simple model to explain all the details of the simple cells, but it does at least provide an approximate model to be used in simulations of parts of the visual system. Computational models that could simulate the physiology of the mammalian visual system are of great interest for research in many different fields [14, 102, 107].

In addition to the physiological findings, one interesting aspect is also the rules or cost functions that could explain the solutions provided by the evolution. This holds for the organization and structure of the visual system as well; what kind of computational model could mimic the strategies employed in the evolution or learning in the visual system; what are feature extraction tasks where the simple cells would be optimal. For example certain statistical methods, roughly similar to principal component analysis (PCA) and independent component analysis (ICA), have been proposed and they seem to find some kind of sparse codes for the natural images. The principal components seem to resemble global oriented frequencies and independent components local oriented frequencies, similar to Gabor filters [103, 104, 105]. In these specific cases the necessity of the sparse codes have been argued since the number of cells is actually much larger than the size of the input [5], but further studies may provide answers to more intriguing questions, such as what is the function of our visual system.

3.2.2 Translation, scale, and rotation properties

Feature extraction from 2-d signals to capture events, for example, to detect or recognize objects in images, is a challenging task since events may vary in many ways in their presence. Even if considering only rigid objects, objects may be translated to any spatial location, rotated to any orientation, and scaled to any size. In general, a system which tolerates all these variations is referred to as geometrically invariant. Invariance is not necessarily a property of a feature but a feature space. 2-d Gabor filter based structures capable of translation, rotation or scale invariant searches of objects have been introduced, but hitherto it seems that *Publication IV* is the first one directly utilizing all the properties in invariant search of objects. In general, any object recognition method having a sufficient degree of robustness can be used in the geometrically invariant detection by searching objects in a standard pose and by rotating and scaling input images. This would work for Gabor features as well without any further consideration of invariance properties, but it may however require much more computational resources than by applying the existing invariant properties. The most well-known method utilizing invariant properties of Gabor features is probably the one introduced by Buhmann et al. [22], where the scale and rotation properties are used to estimate the size and pose of a known object and then a labeled graph based spatial search is applied to a correspondingly normalized image [151]. Besides the translation property, also the scale property can be directly used in the search [21], but due to the computational complexity of the labeled graph matching the initial estimation provides more attractive results [151]. The translation property is the one which is at least implicitly used in the most studies, e.g., segmentation of textures [18, 44] and as already mentioned in the face localization by labeled graph matching [22, 78, 149]. Filter structures for scale invariant object recognition have also been introduced already in the earliest papers (e.g. [113]), but have rarely been explicitly utilized in object detection ([21], *Publication IV*). It should be noted that the Gabor filter responses are robust against a certain degree of distortions, and thus, methods may perform accurately in only coarsely same object scales and orientations [22, 115, 148, 149]. Filters tuned on highest frequencies are scale invariant representing image edges whereas lower frequencies are tightly tied to the size of an object, e.g., fundamental frequencies in *Publication II.*

The translation and scale properties of the Gabor features derived in one dimension can be generalized also to two dimensions. In addition, in order to search objects in the presence of geometric manipulations, one have to cope with a new variant, the orientation, which has no analog in one dimension. A rotation property which can be used as the previously defined translation and scale properties has not been utilized in most studies since for example the rotation invariance is not usually preferred in the texture segmentation. The effect of the rotation of an object on the filter responses has been studied in line detection [28], but first introduced in a more general form by Würtz from Buhmann's group [151] and derived from that as a simple response matrix shift operation by the author in [77] and *Publication VIII* and by others in [109].

A rotated version $\xi_3(x,y)$ of a 2-d signal $\xi(x,y)$, an image, rotated anti-clockwise around a spatial location (x_0,y_0) by an angle ϕ can be written as

$$\begin{aligned}\xi_3(x,y) &= \xi(\hat{x},\hat{y})\\ \hat{x} &= (x-x_0)\cos\phi + (y-y_0)\sin\phi + x_0\\ \hat{y} &= -(x-x_0)\sin\phi + (y-y_0)\cos\phi + y_0\end{aligned} \tag{3.8}$$

The filter response using Eq. (3.7) for the rotated image is

$$r_{\xi_3}(x_0,y_0;f,\theta) = \int\int_{-\infty}^{\infty} \psi(x_0-x_\tau, y_0-y_\tau; f,\theta)\xi_3(x_\tau,y_\tau)dx_\tau y_\tau \tag{3.9}$$

which is in terms of the rotated parameters

$$\begin{aligned}&r_{\xi_3}(x_0,y_0;f,\theta) =\\ &\int\int_{-\infty}^{\infty} \psi([(x_0-x_\tau)\cos\theta + (y_0-y_\tau)\sin\theta], [-(x_0-x_\tau)\sin\theta + (y_0-y_\tau)cos\theta])\\ &\quad \xi([(x_\tau-x_0)\cos\phi + (y_\tau-y_0)\sin\phi + x_0], [-(x_\tau-x_0)\sin\phi + (y_\tau-y_0)\cos\phi + y_0])\\ &\quad dx_\tau dy_\tau\end{aligned} \tag{3.10}$$

and can be rewritten by changing the integration axes to (x'_τ, y'_τ) which are correspondingly rotated clockwise around the point (x_0,y_0) by the angle ϕ

$$\begin{aligned}&\int\int_{-\infty}^{\infty} \psi(\hat{x}_\tau,\hat{y}_\tau;f,\theta)\xi(x'_\tau,y'_\tau)dx'_\tau dy'_\tau\\ &\hat{x}_\tau = (x_0-x'_\tau)\cos(\theta-\phi) + (y_0-y'_\tau)\sin(\theta-\phi)\\ &\hat{y}_\tau = -(x_0-x'_\tau)\sin(\theta-\phi) + (y_0-y'_\tau)\cos(\theta-\phi)\end{aligned} \tag{3.11}$$

It is easy to notice that the previous form equals to

$$\begin{aligned}&\int\int_{-\infty}^{\infty} \psi(x_0-x'_\tau, y_0-y'_\tau; f,\theta-\phi)\xi(x'_\tau,y'_\tau)dx'_\tau dy'_\tau\\ &= r_\xi(x_0,y_0;f,\theta-\phi)\end{aligned} \tag{3.12}$$

Finally, utilizing the translation, scale, and rotation properties of the 2-d normalized Gabor filter responses it can be concluded that for a 2-d signal $\xi'(x,y)$ which is translated from a location (x_0,y_0) to a location (x_1,y_1), scaled by a factor a and rotated anti-clockwise by an angle ϕ around the location (x_1,y_1) it holds that

$$r_{\xi'}(x_1,y_1;f,\theta) = r_\xi(x_0,y_0;\frac{f}{a},\theta-\phi) \tag{3.13}$$

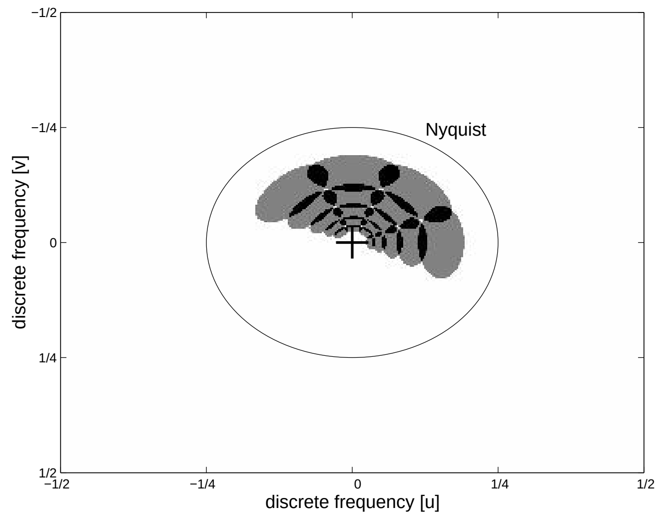
which is a central result for a rotation, scale, and translation invariant search of objects in images ([151], *Publication VIII*).

3.2.3 Simple feature space

Hitherto only a single filter and its invariance properties have been considered, but in practice, applications utilizing only a single filter are unusual (e.g. [137]) and it is more common to combine the responses from several filters. However, as the number of the filters increases, the computational and space complexities increase by at least a linear factor and for many cases by a factor of a higher degree polynomial (e.g. in *Publication IV*). It seems that the trade-off of using more features is the increasing complexity. The Gabor features in Eq. (3.7) are calculated at finite sampling points of the filter parameters, spatial coordinates $(x, y)_i$, frequencies f_j, and orientations θ_k, and to perform an efficient computation and store features within a reasonable space the indices i, j, and k should run over as small number of elements as possible. The features based on the Gabor filters represent a 2-d signal as a grid in a four dimensional feature space. The advantages of the higher dimensional representation are the simple operations for an invariant search of objects. A proper sampling grid and invariant search operations have not been introduced to cover all the filter parameters but only discussed in several studies (e.g. in [113]) due to the complex interconnections of the parameters, e.g., a scale change affects the spatial coordinates not only the frequency. Nevertheless, a simple feature space is introduced in *Publication VIII* and successfully applied to the face detection application in *Publication IV* and *Publication VI*. The feature space is reviewed next as it provides a basis for development of more general feature spaces.

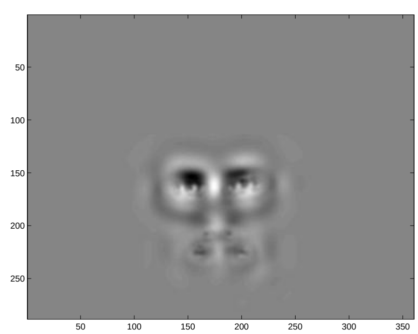
First, one crucial condition must be relaxed to simplify analysis in this stage; the given translation, scale and rotation properties guarantee that the responses are equal but correspondingly shifted in the feature space, which is obviously a relevant result only if the geometric manipulation is the same as the sampling step between the parameters of two filters, but does not prove anything if the manipulation falls somewhere between. This condition will be reassessed later when the shiftability concept is considered.

A significant simplification made in the proposed feature space in *Publication VIII* is the use of only one spatial location (x, y) to represent an object. The assumption is justified if the objects are simple or at least their appearances in feature space are distinguishable from each other. This is not the case with, for example, the human faces, but seems to hold between salient sub-parts, such as nostrils, eyes, mouth corners, etc. It should be noted that the effective width of the Gabor filters spread over a considerably larger area than one point, the exact size depending on the sharpness (bandwidth) factors γ and η. In addition, many filters in one location tuned to various frequencies and orientations span a sub-space which has an accuracy which decreases from the filter origin. This is demonstrated in Fig. 3.3 where an original facial image is reconstructed using filter responses from 10 locations and four orientations and five frequencies. The reconstruction from a few filters can be further improved by optimization [73] but evidently filters capture information from a larger area than just the one point for which they are computed. If given objects can be distinguished using the responses at a single spatial location, a translation invariant search can be performed by simply inspecting the responses at every possible location. This is the approach proposed in many studies and for a computational improvement the complexity can be drastically decreased by utilizing a non-uniform sampling scheme which is more dense near the region of interest [112, 113, 158]. Non-uniform sampling raises a new problem since the region of interest must first be found and for which task heuristic algorithms, such as the saccadic search [131] or hierarchical search [151], can be useful.



(a)

(b)



(c)

Figure 3.3: Example of local reconstruction using filter responses at 10 different locations: (a) used filters; (b) original image; (c) reconstruction.

If only one location is searched and the frequencies are drawn from the logarithmic scale in Eq. (3.6), the search can be done independently in the scale and location as objects appear the same in the feature space [113]. To search objects over various rotations the filter orientation angles can be drawn from

$$\theta_k = \frac{k2\pi}{n}, \quad k = \{0, \ldots, n-1\} \tag{3.14}$$

but since for real signals the responses at $[\pi, 2\pi[$ are 90° phase shifted from responses on $[0, \pi[$ it is usual to compute responses only for the half plane

$$\theta_k = \frac{k\pi}{n}, \quad k = \{0, \ldots, n-1\} \tag{3.15}$$

as illustrated in Fig. 3.3(a). Due to the symmetry of the responses Eq. (3.15) is particularly valid if the magnitude information is used [44, 52].

Now, by using the filter responses in Eq. (3.7) at location (x_0, y_0) with the parameters drawn from Eq. (3.6) and Eq. (3.15) a feature matrix $\mathbf{G}$ can be constructed as

$$\mathbf{G} = \begin{pmatrix} r(x_0, y_0; f_0, \theta_0) & r(x_0, y_0; f_0, \theta_1) & \cdots & r(x_0, y_0; f_0, \theta_{n-1}) \\ r(x_0, y_0; f_1, \theta_0) & r(x_0, y_0; f_1, \theta_1) & \cdots & r(x_0, y_0; f_1, \theta_{n-1}) \\ \vdots & \vdots & \ddots & \vdots \\ r(x_0, y_0; f_{m-1}, \theta_0) & r(x_0, y_0; f_{m-1}, \theta_1) & \cdots & r(x_0, y_0; f_{m-1}, \theta_{n-1}) \end{pmatrix} \tag{3.16}$$

Operations for rotation and scale invariant searches of objects can be defined as the column-wise circular shift of the response matrix corresponding to the rotation of the object around the location (x_0, y_0) and a row-wise shift corresponds to the scaling of an object by a factor c (see Eq. 3.6, *Publication VIII*). Furthermore, an illumination invariance can be achieved by normalizing the feature matrix.

3.2.4 Applications

Gabor features in some form have been used in many applications, but there are two important application areas which have had a major impact on the development and research of Gabor features on a practical basis: texture segmentation and face recognition. On the other hand applications requesting invariance over all, translation, scale, and rotation, are very few, which is evident since, for example, in discrimination of textures no invariance is preferred and in face recognition images usually contain only small alignment variation. However, in many applications the robustness, distortion and noise tolerances examined in *Publication I* and *Publication III*, are beneficial.

Textures as statistical and structural patterns are a promising application area for Gabor filters. Textures can often be simplified to a set of oriented localized narrow band frequencies [18] and discrimination always involves spatial, frequency, and orientation information. A texture is typically classified to a different class if frequency or rotation parameters differ, and thus no invariance is assumed. The first Gabor reference related to textures seems to be in 1985 when Caelli and Moraglia experimented the human observer's ability to discriminate simple textures generated by Gabor functions [24]. Later, inspired by the work of Caelli and Moraglia, the first study of automatic texture analysis was made utilizing the response magnitudes in discrimination and segmentation of different artificial textures [44]. In the same study an optimistic comparison between the method and human observers was conducted, but with poor correlations. It became evident that phase information is also important in the segmentation and after several minor studies the first more comprehensive treatment of Gabor filters in texture analysis was made by Bovik, Clark, and Geisler [18]. They inspected the response of the Gabor filter in detail for a simple texture model and pointed out the necessity of both amplitude and phase information for discrimination. For a more complex segmentation problem where an unknown number of unknown textures must be segmented Jain and Farrokhnia provided a complete algorithm based on features formed from the real part of the Gabor filters [61]. Jain and Farrokhnia were the first to address one crucial problem, how to select a set of filters, and they choose a strategy where filters providing the maximal amount of reconstruction information were selected iteratively with a well lower-bounded solution by a statistical measure. In addition, the filter responses were subjected to a nonlinear transformation. Following their research selection of the transform type has been a research subject in its own right [54]. Jain's method was particularly successful and he reused it in several other applications, e.g., in object detection [62] and locating address blocks on envelopes [63]. The problem of finding an optimal set of filters, maximizing discrimination and segmentation accuracy, has been among the central topics in texture analysis. The problem can be more easily formulated and solved if there are two known textures [41, 144], but in a more general case where there is no a priori information, textures are not known or the number of textures is unknown, more complicated approaches are involved [15, 20, 134, 137]. A comparison of some of the methods can be found from [30]. As originally introduced by Jain, post-processing of the filter responses is also an important consideration. In most of the methods the responses are at least smoothed in the spatial domain, but also non-linear post-processing methods, such as transfer function [61, 69], complex moments [13], and grating cells [74], have been proposed to improve the quality of the features for discrimination. Comparison of the proposed non-linear post-processing methods can be found from [54]. By applying rotation invariant post-processing [79] or constructing an orientation insensitive filter from a combination of Gabor filters in different orientations [159] rotation invariant discrimination of textures can be also achieved. Textural Gabor features can be used also for an efficient search in textural databases [93]. Texture segmentation is a difficult yet very important task in many image analysis or computer vision applications and the good results achieved utilizing Gabor filters encourages exploration of the proposed techniques whenever the texture segmentation is needed.

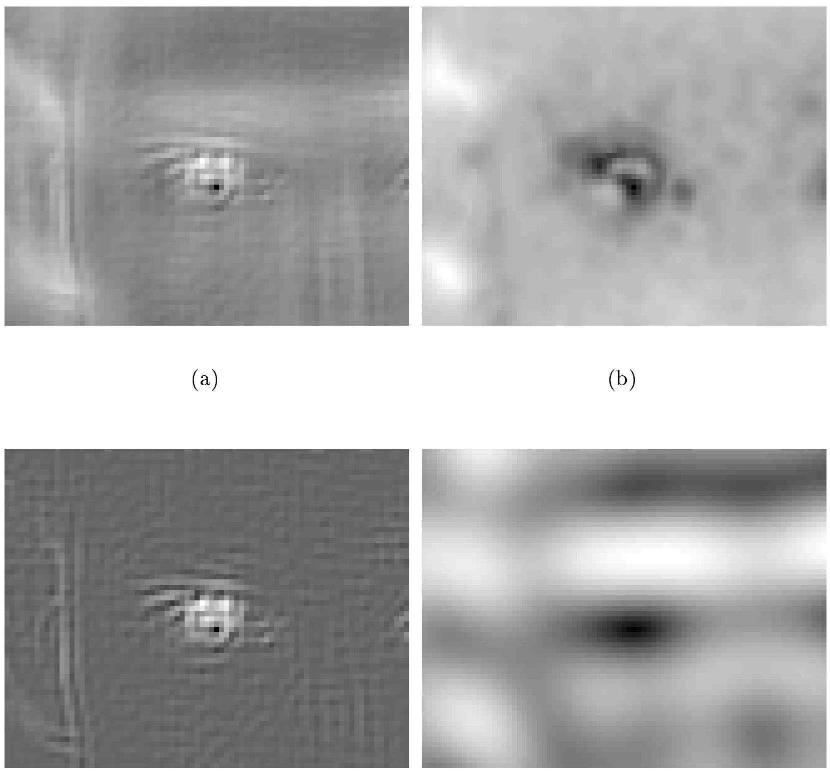
In face detection and recognition based on Gabor filters typically a set of filters is selected to represent salient sub-parts of faces and various approaches are applied to describe spatial relationships between these sub-parts. An approach with a strong impact on later research was the one initiated by Buhmann, Lange and von der Malsburg by introducing a dynamic link architecture [22]. In the dynamic link architecture objects, faces, are represented by sparse graphs, whose vertices are labeled by a multi-resolution description, a set of Gabor features called a Gabor jet, and whose edges are labeled by geometrical distance vectors. The object recognition can be formulated as elastic graph matching by optimizing a matching cost function [22, 78]. Variations of this architecture have been given in several studies [73, 82, 115, 126, 136, 149]. The dynamic link architecture has spurred researchers to evaluate reasons why some parts of faces are more important [68]. A method comparable to those for texture segmentation was proposed by Lampinen and Oja, where unsupervised clustering was used to specialize features and generate a feature map, and a feature histogram was used for recognition [81]. In the most recent studies novel methods have been proposed for face detection and recognition in the presence of various disturbances and distortions [4, 88, 109, 128, 131]. One of the most promising methods is the one introduced by Park and Yang [109], where the same rotation invariant search as in *Publication VIII* and originally in [77] is used. The method in *Publication IV* developed by the author represents an approach, which directly utilizes all the translation, scale, and rotation properties of Gabor filters. The method acts as a pre-processing step to the face detection system, which provides a geometrically invariant detection of faces (*Publication VI*). Using the system in *Publication VI* faces can be accurately aligned and any of the existing face recognition methods can be applied.

Gabor filters have moreover been used in detection of edges [96, 98] (detailed analysis in [96]), junctions from curve shapes in the feature space [28], object dimensions (*Publication II*), and curvatures [27]. Their frequency tuning property has been utilized in fingerprint image enhancement [60, 154] and, an interesting detail, they have been used in detection of simple objects in infrared images [143] decades after Casasent introduced an optical solution of the same recognition scheme (e.g. in [26]). One notorious application is the extremely accurate authentication system utilizing Daugman's iris code generated from Gabor responses to an eye iris image [40]. Gabor filters seem to receive credits in many different type of applications, and thus, perhaps Granlund's idea of the general image processing operator was not too optimistic at all [52].

3.3 Similarity measures for Gabor features

One important consideration with any type of features is how to to compare them, i.e., how to establish a proper similarity measure. Often features, such as responses extracted using the Gabor feature matrix in Eq. (3.16), are extracted from predefined object locations in training images and trained to a selected classifier, which learns internal Gabor feature representations of object classes. If the training is successful the classifier can be used to detect and recognize objects in new unseen images. The classifier may not directly depend on any chosen similarity measure. However, since almost all classifiers have implicit assumptions of their input and target domains, the structure and behavior of a feature space must be analyzed in order to achieve reliable and accurate results. Often it is desired that features should smoothly vary among examples of a specific class to form a clear feature cluster or clusters in the feature space, e.g., in the face detection approaches by Lampinen et al. [80, 81] and in *Publication VI*. However, it is not uncommon that a method is directly based on a similarity measure, such as the face detection and recognition methods by Lampinen et al. [82], Park and Yang [109], and Buhmann's group [22, 78]. In these methods, success of the recognition clearly depends on the selection of a similarity measure, which again depends on behavior of features in the feature space.

Since Gabor filter is sensitive to both magnitude and phase, its response typically oscillates on a frequency related to the filter frequency f and local image frequencies. The oscillation makes the object search by similarity maximizing a difficult task. This is the case for example in the labeled graph matching algorithm which is based on the comparison of Gabor jets [22]. A Gabor jet is a vector of Gabor filter responses on different frequencies and orientations that corresponds the feature matrix in Eq. (3.16) reshaped into a vector form. The difficulty of comparing jets is demonstrated in Fig. 3.4, where is shown the similarity value map between filter responses in the left eye center of the facial image in Fig. 3.3(b) and its neighborhood (dark values denote high similarity). The rapid oscillation (Fig. 3.4(a)) is due to the phase differences in responses of adjacent pixels, and thus, response magnitudes behave more smoothly as illustrated in Fig. 3.4(b) providing a more robust similarity measure than the Euclidean distance of complex numbers [22, 78]. It should be noted that the oscillation is much more intensive on high frequencies (Figs. 3.4(c) and 3.4(c)).



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

Figure 3.4: Distance map between Gabor filter response vectors (Gabor jets) in the left eye center and its neighborhood ($f = \frac{1}{3}, \frac{1}{6}, \frac{1}{12}, \frac{1}{24}$, $\theta = 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$) (a) Euclidean distance of complex responses; (b) Euclidean distance of response magnitudes; (c) Euclidean distance of complex responses on the highest frequency ($\frac{1}{3}$); (d) Euclidean distance of complex responses on the lowest frequency ($\frac{1}{24}$).

There are several problems related to the Euclidean distance itself. The value of the Euclidean distance depends on the magnitudes and is thus very sensitive to illumination changes. The effect of illumination can be suppressed by normalizing the responses. The normalization can be applied separately to filters on different frequencies [22], which is motivated by the fact that the power spectrum of "natural images" has more energy on low than on high frequencies [78]. After normalization the Euclidean distance provides a proper similarity measure ([77, 109], *Publication II*, *Publication III*), but since relationships of responses are considered more informative than single magnitudes the dot product representing a projection between two vectors is typically used as a similarity measure having a convenient similarity value between 0 and 1 [78]. Unfortunately the normalization is not harmless either; it emphasizes small responses generated, for example, by noise on a constant background, that typically increases the number of false alarms. The effect of small responses, which are considered non-informative, can be reduced by applying non-linear terms into the similarity measure based on the dot product [21, 22] or eliminated by neglecting locations where the responses are too small ([148, 151], *Publication IV*). In addition, for small values of sharpness parameters γ and η in Eq. (2.22), the filter envelope may have a significant overlap over the zero frequency (violation of the restriction in Eq. (2.36)), in which cases the removal of zero response is typically used [21, 82, 151]. In practice, stability of low frequencies (e.g. 3.4(d)) can be utilized in a hierarchical object localization scheme where a coarse localization is performed on low frequencies and refined using high frequency information [21, 151].

The phase sensitivity produces the oscillation in Gabor filter responses, but it also provides information of a more accurate location of objects. Thus, to improve the above similarity measures, also the phase information must be included. Malsburg's research group has proposed several similarity measures combining the magnitude and phase [136, 148, 149, 151]. The proposed similarity measures are originally based on disparity measures for binocular camera systems [43, 135] and are most useful when objects in images are roughly in the same scale and orientation. The proposed similarity measures are based on local phase difference estimates and stability conditions of phase information [43, 135, 148]. Lately, Lampinen et al. have proposed a similarity measure based on an ad hoc conditional likelihood distribution of filter responses [82].

In Fig. 3.5 there are shown examples of complex Gabor filter responses for real objects used in *Publication IV* and *Publication VI*. In Fig. 3.5 are plotted filter responses from left and right eye centers, which should be roughly symmetric, on two frequencies at zero orientation. Complex responses in Figs. 3.5(a) and 3.5(b) form an approximate Gaussian clusters, which assumption has been used in the classifier selection in *Publication IV* and *Publication VI*. The phase warping affects a discontinuation to the classes as shown in the magnitude-phase images in Figs. 3.5(c) and 3.5(d). From Figs. 3.5(c) and 3.5(d) it can be seen how the phase variation decreases on lower frequencies. In addition, the robustness and stability of phase information increases for small values of sharpness parameters (wider bandwidth).

Furthermore, the shift operations in *Publication VIII* can be included to similarity measures, e.g., to establish a rotation invariant similarity measure [77, 109].

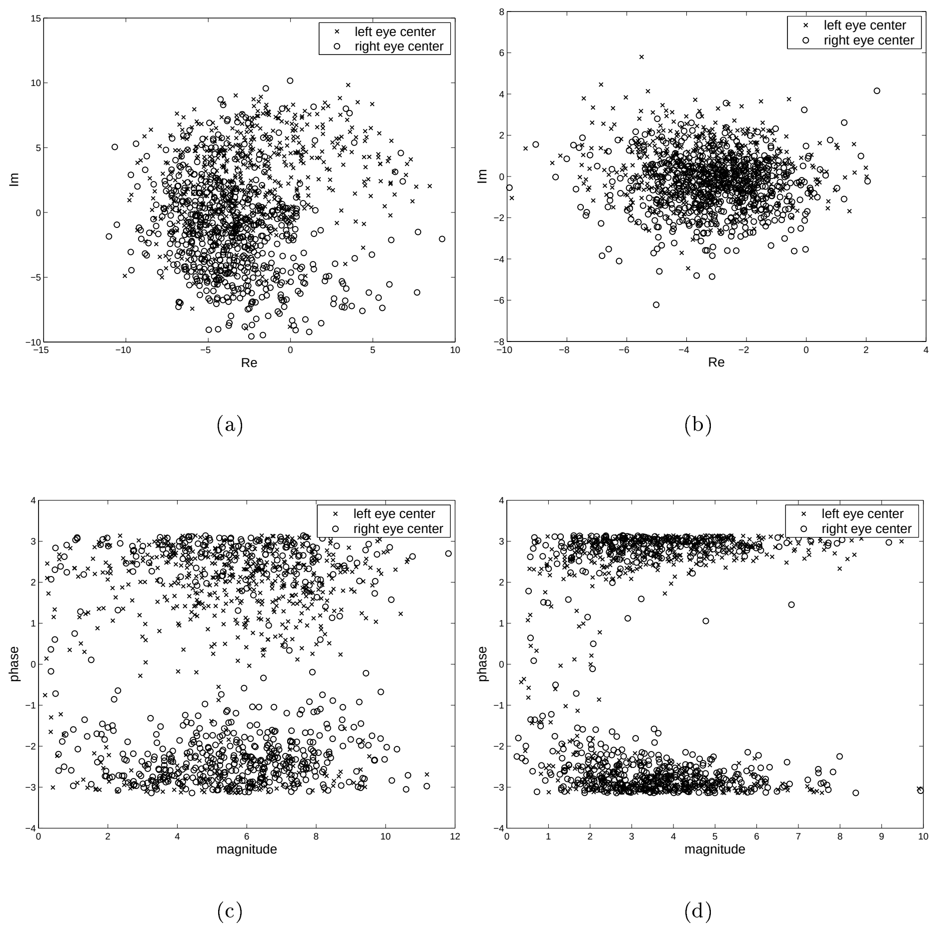


Figure 3.5: Gabor filter responses in the left and right eye centers from 600 training images in XM2VTS database (*Publication IV*): (a) real-imaginary plot on high frequency ($f = \frac{1}{3}, \theta = 0$); (b) real-imaginary plot on low frequency ($f = \frac{1}{12}, \theta = 0$); (c) magnitude-phase plot on high frequency ($f = \frac{1}{3}, \theta = 0$); (d) magnitude-phase plot on low frequency ($f = \frac{1}{12}, \theta = 0$).

3.4 Practical considerations

3.4.1 General

In order to have any but theoretical interest a method should perform stably in terms of parameters and input variation, and the computational and space complexities must be adequately low. A full invariant search of objects even in the proposed simple feature space requires an exhaustive search, which is always a time demanding operation and proportional to the size of the space. In such cases the feature extraction and the filtering itself do not tend to be the the bottlenecks but a classification. For example a full search in a two dimensional image needs at least

$$\text{Image width} \times \text{Image height} \times \text{Row (scale) shifts} \times \text{Column (orientation) shifts}$$

feature vectors to be classified. In practice heuristics and a priori information must be incorporated in order to reduce the computing time, e.g., in *Publication IV* and *Publication VI* face evidences can be searched only over a few orientations and scales, and not in every location.

Often, and no exception is made in this thesis, results are derived in the continuous domain and directly applied in experiments conducted in the discrete domain without stressing any problems that may arise in the domain transform. However, it is clear that there are several general issues that must be borne in mind in order to achieve reliable results with Gabor filters. As always in discrete representations of continuous signals the sampling and quantization are central considerations. This holds for Gabor filters as well and is of special importance if an accurate reconstruction is what is desired [18, 113]. The signal reconstruction is the main concern with Gabor expansion and in addition to the filter representation, attention must be paid also for sampling and quantization of the expansion coefficients [113]. On the other hand, in feature extraction standard computer arithmetics are usually available, filters are roughly of the same order of magnitude, and in well-posed problems the distinguishing properties should not be merely very small differences in the response of some filter. In the feature extraction it is thus reasonable to relax the concern above the effect of sampling and quantization errors; it is highly improbable that they will have any significant effect on performance. Much more essential considerations are the selection of the central frequencies f_k and the size of the filter envelopes controlled by γ and η.

Restrictions in the discrete domain, most importantly the Nyquist frequency, may prohibit an accurate discrete representation of a continuous domain filter. The filter must be tuned to a positive or negative central frequency f and the sharpness values γ and η must be selected to prevent aliasing, i.e., filter coefficients should be negligible outside the allowed discrete frequencies $]0, 1/2[$ or $]-1/2, 0[$. This problem has already been assessed and Eq. (2.32) (1-d) and Eq. (2.36) (2-d) were given to ensure the filter is properly sampled and Eq. (2.35) (1-d) and Eq. (2.37) (2-d) to ensure that the filter fits in the discrete filter bins. The given restrictions ensure an accurate and reliable results even with images of significantly different width and height.

Typically a cyclic convolution [19] is assumed in the computations and one has to be aware that this affects the filter responses making them unreliable near signal ends or image edges. To have a reliable response at a location (x, y) the effective area of the filter envelope should be inside the image at that point and the minimum filter size, $L_{min} = L$, can be resolved from Eq. (2.35) (1-d) or Eq. (2.37) (2-d) and the responses within a distance

$$\leq \frac{L_{min}}{2} \tag{3.17}$$

from edges should be neglected.

Proper construction of the filters can be achieved and unreliable filter responses can be avoided. Another important consideration is the selection of values for the frequency f, orientation θ, and sharpness γ and η. The selection of these values is the main problem in utilizing Gabor filters in feature extraction since they define the time-frequency division and the uncertainty of the extraction. While some adjustments can be made for the frequency and orientation, the sharpness parameters γ and η are completely application dependent. As a rule of thumb, the filter size should be adjusted to optimally encapsulate the event of interest, in time and frequency. For example, in texture segmentation a circular form of the filter ($\gamma = \eta$) is typically used, but to distinguish objects of different dimensions a not equal aspect ratio ($\eta < \gamma$) provides more accurate results (*Publication II*). On the other hand, selection of one parameter can be compensated in the selection of another [113]. This vague nature in the selection of the parameters is one of the reasons that an explicit optimization of the parameters has been applied (e.g. [41, 61, 144]), but for some applications this may be impossible or at least it causes inconvenience because the optimization must be repeated in new situations, e.g., introduction of a new object. To be more flexible it is possible to select the filter parameters to cover the whole parameter space within the limits of the computing time and resolving the importance of different filters is left to a learning system, a classifier, which is not in the scope of this thesis. Some adjustments can be made; the frequencies can be drawn from Eq. (3.6) where the computational resources limit the number of frequencies, m, and the maximal frequency, f_{max}, and the scale factor, c, can be selected to cover the frequencies of interest. Moreover, different orientations can be drawn from Eq. (3.15) where the only parameter is the number of orientations, n, which again depends on computational resources. In practice, as many as possible filters should be used within the limits of the computational resources, to extract as fine details as possible.

3.4.2 Shiftability

As mentioned for efficient computation and limited size representation of the features in Eq. (3.7), the features must be calculated only for a sufficiently small number of parameters. This is evident since the discretization of the spatial coordinates (x, y), frequency f, and orientation θ is a four dimensional feature space. There is a trade-off between the selection of the filter parameters and the continuity of the feature space. In other words, one would like to fill the feature space as efficiently as possible to decrease the computation, but at the same time the filters should overlap as much as possible to provide smooth behavior in situations where events fall somewhere between two filters. This trade-off is a known problem between orthogonal and non-orthogonal presentations since the continuous behavior, shiftability, can be achieved only by relaxing the orthogonality property [47, 129]. In shiftable systems a response for any parameter value can be constructed as a linear combination from the discrete values of the same parameter and if this holds for more than one parameter the system is said to be joint shiftable [129]. The shiftability concerns are addressed also for wavelets in studies of their translation invariance [12]. It should be noted that Gabor filters do not meet the requirements of exact shiftability, although they have the very beneficial property of being well localized, which is desirable in low-level image processing [110]. In order to decrease the orthogonality the overlap of the filters must be increased. The overlap cannot be accomplished by sharpness parameters γ and η but the number of filters must be increased to even approximate the shiftability conditions. The smooth behavior, approximate shiftability, is again a trade-off which is achieved at the expense of computational complexity. Approximate shiftability is currently an unexplored property of Gabor filters and spaces spanned by Gabor functions and only a few studies of the definition of approximate shiftability have been made [11, 142, 157].

3.4.3 Robustness and noise tolerance

Robustness and noise tolerance are essential considerations for practical applications. Robustness here means the stability of the responses when conditions, such as the filter parameters, are changed. There are many natural variations in the conditions of real applications which cannot be exclusively covered in the method development phase, and thus, the methods benefit if they are robust against these changes: a small change in the conditions induces only a small change in the system performance. Robustness against variation in the filter parameters is addressed in *Publication III* where the classification accuracy is experimentally shown to smoothly and continuously change proportional to the change of the parameter values. Robustness has been previously studied in [96] where the responses were analyzed in terms of the filter parameters for an edge function. It can be concluded that the accuracy decreases gradually as the deviation from the optimal parameter values increases; the phenomenon which can be assumed from the Gaussian shapes of Gabor filters in both domains. If an object is represented by different examples producing some kind of interpretable mean object and its variation, the Gabor features also have some mean and proportional variation in the feature space. Such behavior of the features is preferred by many classifiers producing smoothly changing decision boundaries where the classes can be represented, for example, by multi-modal Gaussian probability functions, e.g., the sub-cluster classifier in *Publication IV* and a Bayesian classifier assuming the Gaussian probability function (Mahalanobis distance) in *Publication V* and *Publication VII*.

Since some noises and distortions can also be described as variations of the filter parameters it is assumed that the behavior of the Gabor features is smooth in the presence of noise and distortions. Furthermore, the Gabor filters are optimally joint localized in time and frequency, and thus, distortions and noise present in distinct locations, time or frequency, do not significantly interfere with the filter responses. Noise and distortion tolerance is natural to Gabor filters and it was considered in *Publication I* where the Gabor filter based features perform outstandingly well in the presence of noise and distortions, namely Gaussian, salt-and-pepper and pixel displacement noise, and illumination gradient distortion. Of course noise removing methods can be applied and distortions can be eliminated, but in cases where these are not known a priori the Gabor filters still provide tolerance of a high degree.

Chapter IV

Application Examples

The application examples in this chapter, introduced in more detail in *Publication I* - *Publication VIII*, are based on the theory presented in the previous chapters. Here the results from the publications are briefly reviewed and new results are presented for clarification.

4.1 Induction motor bearing damage detection

A reliable method for an automatic bearing damage detection of induction motors would have a significant practical impact since harmful bearing problems often occur in industrial and commercial motor installations [125]. Methods which are based on the stator current signal would be especially attractive since they do not require additional instrumentation of measurement devices. The contemporary methods are based on the physical model of bearing damage characteristic frequencies (e.g. [124]). However, in industrial environments there are many factors which may make the detection on the characteristic frequencies impossible. For engineers in this field it would be beneficial to have a more general diagnosis tool which could be used to find frequencies and bandwidths where discriminative information between normal and failure conditions is present.

In *Publication V* and *Publication VII* is proposed a simple signal diagnosis method based on Gabor features. The method uses signal content on a frequency band as a feature and the level of discriminative information is measured by the standard first- and second-order statistics. In this particular case any band pass filter can replace the Gabor, but in the most popular approaches for the automatic bearing damage detection Gabor filters are preferred due to their useful properties in the preceding steps of classification, e.g., segmentation of stationary parts from stator current [156].

Two sets of signals, $x_k(t)$ and $y_k(t)$, represent examples from two classes, C_1 and C_2, respectively. The sub-index k denotes a measurement number, $k = 0, 1, \ldots, N_1 - 1$ for C_1 and $k = 0, 1, \ldots, N_2 - 1$ for C_2. Total power of a band-pass filtered signal can be used as feature

$$\int_{-\infty}^{\infty} |\psi(t) * x_k(t)|^2 dt \text{ and } \int_{-\infty}^{\infty} |\psi(t) * y_k(t)|^2 dt \tag{4.1}$$

and the discrimination capability on the pass band can be measured based on first-and second-order statistics. The first-order measure utilizes only the feature mean values

$$\begin{aligned} \mu_x &= E\left[\int_{-\infty}^{\infty} |\psi(t) * x_k(t)|^2 dt\right] \\ \mu_y &= E\left[\int_{-\infty}^{\infty} |\psi(t) * y_k(t)|^2 dt\right] \end{aligned} \tag{4.2}$$

and the second-order measure, assuming the normal distribution, also variances

$$\begin{aligned} \sigma_x^2 &= E\left[\left(\int_{-\infty}^{\infty} |\psi(t) * x_k(t)|^2 dt - \mu_x\right)^2\right] \\ \sigma_y^2 &= E\left[\left(\int_{-\infty}^{\infty} |\psi(t) * y_k(t)|^2 dt - \mu_y\right)^2\right] \end{aligned} \tag{4.3}$$

In the proposed method, the first-order statistics discriminative energy function is

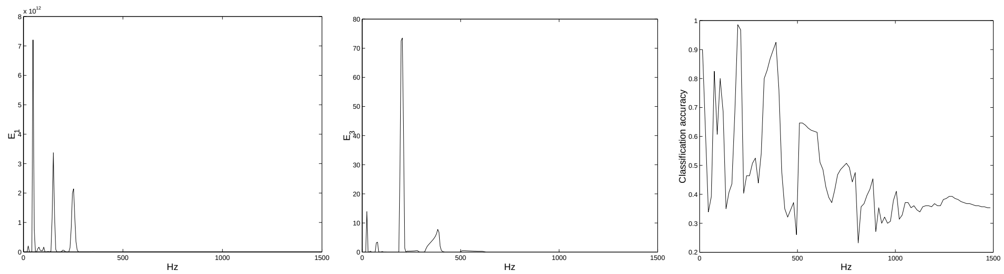
$$E_1 = \frac{1}{2} (\mu_x - \mu_y)^2 \tag{4.4}$$

and the second-order statistics discriminative energy function, based on the Fisher's discriminant ratio, is

$$E_3 = \frac{1}{2} \left(\frac{(\mu_x - \mu_y)^2}{\sigma_x^2 + \sigma_y^2} \right)^2 \tag{4.5}$$

The values of discriminative energy functions and classification results of Bayesian classifier with a normal distribution probability density function and equal priories are shown in Figs. 4.1 and 4.2 for a set of real measurements. From the experiments it is evident that the first-order statistics (E_1) do not provide sufficient information, but at least second-order statistics must be used. The results in Fig. 4.1 are from the experiment, where no load is connected to motors, bearings are large, and the failure is big, thus providing favorable conditions for presence of the characteristic frequencies. E_1 emphasizes only the motor supply frequency, but E_3 has its maximum on the first harmonic of the damage characteristic frequency ($202Hz$, Fig. 4.1(b)). It can be assumed that stator current signal on the actual characteristic frequency ($101Hz$) is saturated by other frequency factors of a rotating motor. The Bayesian classification yields its maximal accuracy also at the same band (Fig. 4.1(c)).

For more realistic data, where a load is connected to motors and affecting various disturbances, the results are quite different (Fig. 4.2). Other disturbance factors spread over a wider spectrum, and thus, the characteristic frequency and its several first harmonics provide almost the same discriminative information (Fig. 4.2(b)), but the classification results on any of them is significantly worse than in the previous experiment. This result indicates that the characteristic frequency and its first few harmonics will be distorted



(a) (b) (c)

Figure 4.1: Discriminative energy functions and Bayesian classification accuracy for stator current signals of induction motors with large bearings (minimum internal clearance $30 \mu m$), big bearing failure ($5mm$ hole), and without load (characteristic frequency approximately $101 Hz$): (a) E_1; (b) E_3; (c) Bayesian classifier.

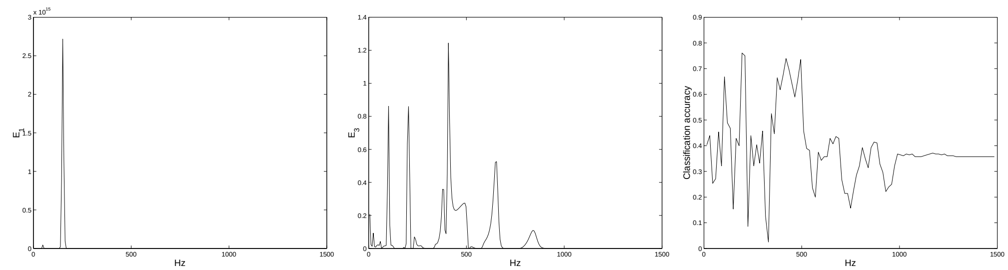
in actual industrial environments, pushing the method toward higher harmonics which are unfortunately less accurate in detection.

The proposed method in *Publication V* and *Publication VII*, and especially the E_3 discriminative energy function are shown to be useful for diagnosis. In this particular task the method was used to verify the existing theory of bearing fault characteristic frequencies and their applicability for damage detection in more realistic situations. The method can be useful also in diagnosis of any similar signals where the underlying physics of a phenomenon are not necessarily known since it provides a descriptive diagram of discriminative frequencies. Future development should include extensions to discriminate and diagnose several different type of failures and to measure discriminative information simultaneously using more than one frequency band.

4.2 Symbol recognition

Image databases have become increasingly popular in several application areas. For example, in medical imaging, engineering, and publishing large amounts of images need to be stored in image databases. *Publication I* and *Publication III* consider a Gabor feature based method for a recognition of electric symbols in binary images. The method could have its use in databases of engineering drawings, e.g., electrical circuits, maps, and architectural and urban plans. The method utilizing a global Gabor feature has originally been proposed by Kyrki et al. [75, 76]. Since the method itself is not novel the selection of the Gabor filter parameters has been the main consideration in *Publication I* and *Publication III*, i.e., how a non-optimal selection of parameters would affect to classification results.

In a global Gabor feature, the Gabor responses are summed over an image to form a global feature. The global Gabor feature can be considered as a histogram of the Gabor



(a) (b) (c)

Figure 4.2: Discriminative energy functions and Bayesian classification accuracy for stator current signals of induction motors with large bearings (minimum internal clearance $30\mu m$), big failure ($5mm$ hole), and with load (characteristic frequency approximately $101Hz$): (a) E_1; (b) E_3; (c) Bayesian classifier.

responses in Eq. (3.7) for different frequencies f and orientations θ as [75, 77]

$$G(f,\theta) = \sum_x \sum_y |r_\xi(x,y;f,\theta)| \tag{4.6}$$

Since edges appear on high frequencies and lines on their fundamental frequencies, a single properly selected frequency f is needed to extract sufficient information from line drawing images to represent a histogram of lines in different orientations. A feature vector consisting of global Gabor features on different orientations can be defined as

$$\mathbf{G} = (G(f,\theta_0) \;\; G(f,\theta_1) \;\; \ldots \;\; G(f,\theta_{n-1})) \tag{4.7}$$

The global Gabor feature in Eq. (4.6) is translation invariant as the responses are summed over the whole image. For higher frequencies representing edges, the object scale affects response magnitudes, but the ratio between the magnitudes remains over different scales. Thus, a scale invariant feature can be constructed by normalizing the feature vector. In addition, the normalization makes the feature also illumination invariant, that is, invariant to a constant multiplier. Rotational information of an object is stored in the columns of the vector and the smallest detectable rotation angle is limited by the number of different orientations N, unless interpolation of the responses is used. Now, using the column-wise shift operation from *Publication VIII* a rotation invariant similarity measure can be defined [77]

$$d(\mathbf{G}_1,\mathbf{G}_2) = \min_k \{ \sum_{\theta_i} \left[\mathbf{G}_1(f,\theta_i) - \mathbf{G}_2^{(\theta+k)}(f,\theta_i) \right]^2 \} \tag{4.8}$$

where $\mathbf{G}_1$ and $\mathbf{G}_2$ are feature vectors from two inspected images.

To test the performance of global Gabor features in image database queries an electric symbol recognition experiment was performed. The base classes were the symbols in

Fig. 4.3 and their randomly rotated variants were used in the classification. The classification was carried out by measuring distances to all base classes with the rotation invariant similarity measure in Eq. (4.8). The proposed features clearly outperformed

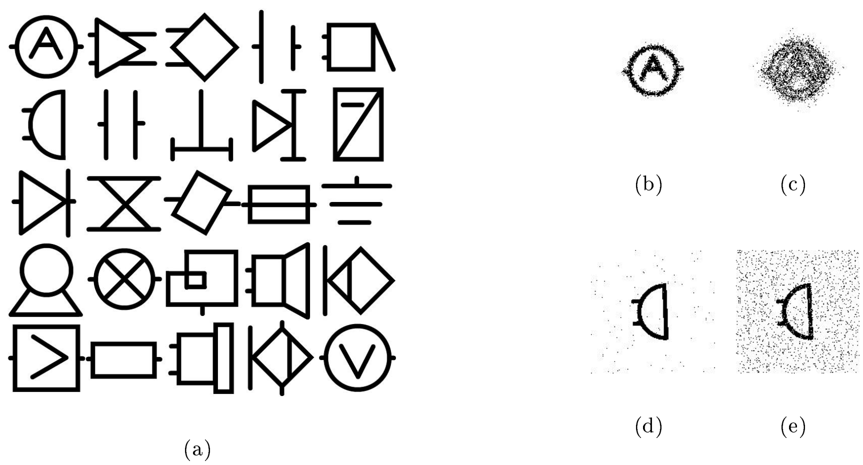


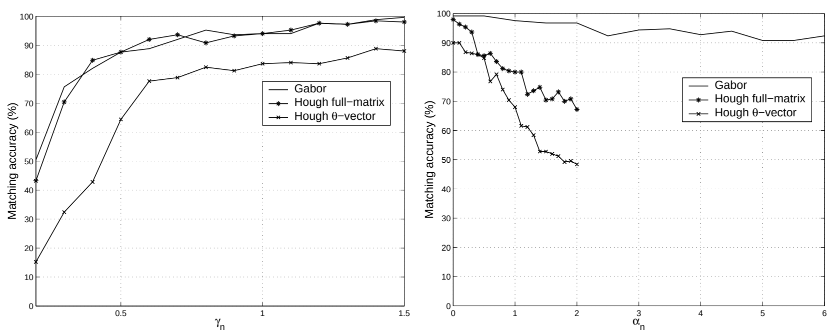
Figure 4.3: Electric symbols and noisy examples: (a) symbols; (b) displacement noise ($\gamma_n = 1$); (c) displacement noise ($\gamma_n = 0.2$); (d) salt-and-pepper noise ($\alpha_n = 0.4$); (e) salt-and-pepper noise ($\alpha_n = 2.0$).

the Hough transform based features proposed by Fränti et al. [46] by being comparable to that in the presence of displacement noise and significantly more tolerant to the salt-and-pepper noise (Fig. 4.4). Examples of noisy images are shown in Fig. 4.3. The results in Fig. 4.4 were achieved using the filter parameters $\gamma = 1$, $\eta = 1$, $f = 0.056$, and $N = 20$, but due to the smooth behavior of Gabor filter responses, the method is robust to a non-optimal selection of the filter parameters as can be seen by the results in Fig. 4.5 (*Publication III*).

The proposed method by Kyrki et al. [76] and analyzed in *Publication I* and *Publication III* is not directly applicable to large image databases since the global Gabor feature cannot provide a sufficient amount of discriminative information. However, the experiments provide evidences of representation power and robustness of Gabor filter based features which are useful information for future work where the local information should be incorporated into the features.

4.3 Electric component detection

Object detection from real scene images has been one of the most typical example of machine vision applications. An object detection method for electric components could be used for example to detect components and their pose from convey belts and to control



(a) (b)

Figure 4.4: Matching accuracies of symbols for the global Gabor feature and Hough transform based methods as functions of noise parameters: (a) displacement noise; (b) salt-and-pepper noise. [76]

a robot to pick the components at a packaging station. In this kind of object detection problems the object scale typically remains the same, but objects may appear in an arbitrary pose (orientation and translation) and there may be image distortions present, such as noise and illumination changes.

A method for electric component detection is introduced in detail in *Publication II* and also considered in *Publication I* in the sense of noise tolerance. In this particular application the components mainly differ in their size and Gabor filters on several frequencies and orientations are used to detect the components by a translation and rotation invariant manner. The method is based on a detection of the fundamental frequencies of objects, that is, frequencies which describe the overall shape of an object. In a 1-d case a rectangle function may represent an ideal shape where the width of the rectangle is the size information to be inspected. It can be shown that the response of Gabor filter has its maximum at the centroid of the rectangle on frequencies

$$f = \frac{1}{2w} + \frac{n}{w}, \quad n = 0, 1, 2, \ldots \tag{4.9}$$

where w is the rectangle width (*Publication II*). The same result can be generalized to two dimensions. Now, to represent an object, e.g., an electric component, frequencies with the maximum response are stored for all orientations describing the size variation of an object in different directions. To estimate also the pose of a detected object, the fundamental frequency features should be stored for objects in a standard pose. Algorithm 1 extracts fundamental frequency features at an approximate centroid (x', y') of object.

Algorithm 1 *Extract fundamental frequency Gabor features* $\mathbf{F}$ *from the input image* $\xi(x, y)$ *at* (x', y')

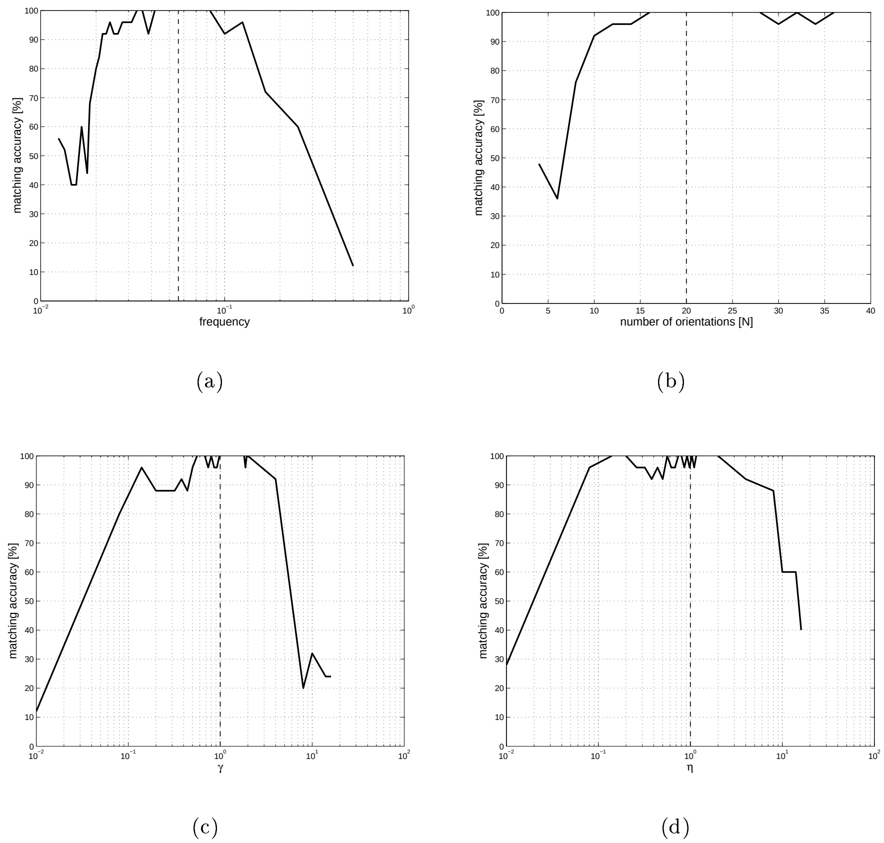


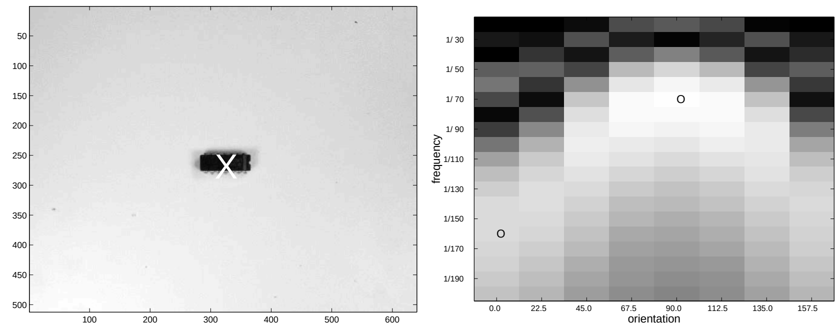
Figure 4.5: Matching accuracies of symbols for the global Gabor feature based method as functions of the Gabor filter parameters (values used in Fig. 4.4 marked): (a) frequency f; (b) number of orientations N; (c) sharpness γ; (d) sharpness η.

1: *Compute response matrix* $I_{M \times N}$ *for frequencies* $f \ldots f_{M-1}$ *and orientations* $\theta_0 \ldots \theta_{N-1}$ *at* (x', y'), $I(f_i, \theta_j) = r_\xi(x', y'; f_i, \theta_j)$.

2: *For all orientations find the maximal responses* $\mathbf{r}_{1 \times N}$ *and the frequencies* $\mathbf{f}_{1 \times N}$ *they appear in* $((f_l, r_l) = \max_{f_k} I(f_k, \theta_l))$.

3: *Create a feature matrix* $\mathbf{F}_{2 \times N} = \left(\mathbf{f}^T \; \mathbf{r}^T_{norm}\right)^T$ *using the normalized maximal responses* $\mathbf{r}_{norm} = \mathbf{r}/\|\mathbf{r}\|$ *and the corresponding frequencies* $\mathbf{f}$.

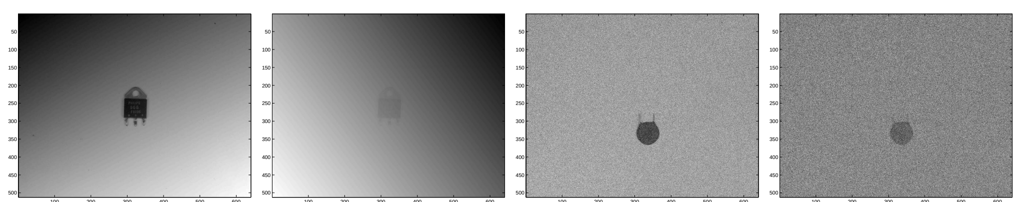
It should be noted that the spacing of frequencies is not necessarily logarithmic but linear to achieve a sufficient resolution for size discrimination. The feature constructed by the algorithm represents components by storing fundamental frequencies, which correspond the size of an object along different orientations as illustrated in Fig. 4.6. In the figure there is an object of width 80 and height 35 pixels when the corresponding fundamental frequencies are $\frac{1}{160}$ and $\frac{1}{70}$.



(a) (b)

Figure 4.6: Magnitudes of Gabor filter responses in the centroid of an electric component: (a) component and its centroid ($(x, y) = (324, 263)$); (b) diagram of the filter responses at the centroid.

It should be noted that in Algorithm 1 the normalization is not applied to responses until the maximum frequencies have been found, that is, higher energy content on lower frequencies is an advantage revealing the object size. However, before classification the responses are normalized to gain robustness against illumination changes. For this kind of problem for example the Fourier descriptors typically provide good results [111], but they need a successful object segmentation which may turn to a difficult task in presence of distortions and noise as in example images in Fig. 4.7. The detection method in



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

Figure 4.7: Examples of component images containing noise: (a) image gradient (slope 1.0); (b) image gradient (slope 2.5); (c) Gaussian noise (deviation 32); (d) Gaussian noise (deviation 100).

Publication II does not need any preprocessing before the classification. To recognize 8 different components at arbitrary poses in real images, a detection experiment was conducted using the fundamental frequency features and the rotation invariant similarity measure. The same experiment was also conducted using Hu's third order moments [10], but the results were significantly worse and for noise and distortions the moment invariant method completely failed. It is evident that the task is not trivial, but the Gabor feature based method succeeded particularly well in the detection of electric components from noisy and distorted images and achieved 100% accuracy for images without distortions (Fig. 4.8).

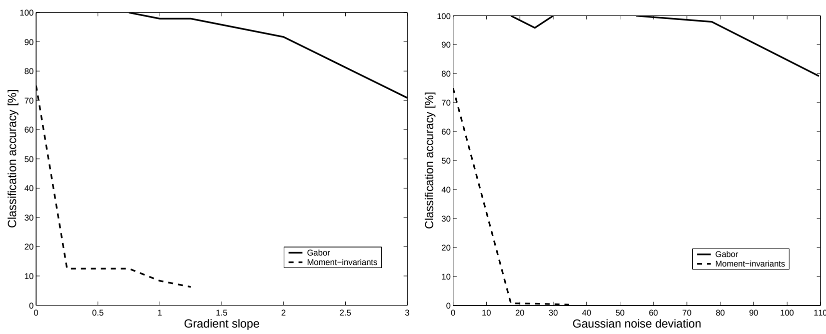


Figure 4.8: Electric component classification accuracies as functions of noise parameters.

The proposed method provides a framework for simple object recognition tasks. For training only one image from each class is needed and the only input information are the approximate centroids of objects. After training, a simple algorithm can be used to detect and recognize stored objects from unseen images in a rotation and translation invariant manner and it is also possible to estimate the object pose. The achieved noise and distortion tolerances are evidences of the reliability the fundamental frequency Gabor features may provide for applications.

4.4 Face evidence extraction

Face recognition is an important issue in developing of security applications for personal identification, and the recognition methods have been an active area in recent image processing research. In practical applications, face detection, localization, is often a prior step before the actual face recognition, but the success of the whole system may vitally depend on the detection accuracy. Recently, two authors of *Publication IV* and *Publication VI*, Hamouz and Kittler, have proposed a complete framework for more a general object detection based on discriminative regions [56, 94]. If a scale, rotation, and translation invariant detection of discriminative regions can be performed and spatial relationships between different discriminative regions of object are known, the detection of an object can be carried out. The success of the framework depends on successful selection and efficient extraction of discriminative regions, e.g., face evidences.

In the proposed face detection approach by Hamouz and Kittler the detected objects are frontal human faces and discriminative regions are salient sub-parts, such as nostrils and eyes. The proposed framework can be used as a front-end to a face authentication system as shown in Fig. 4.9.

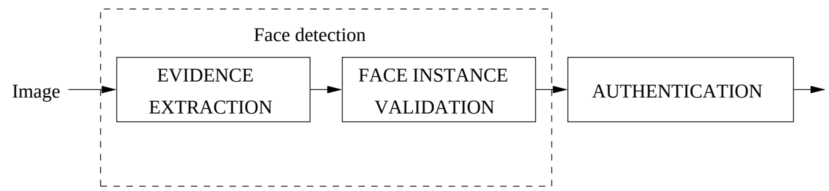


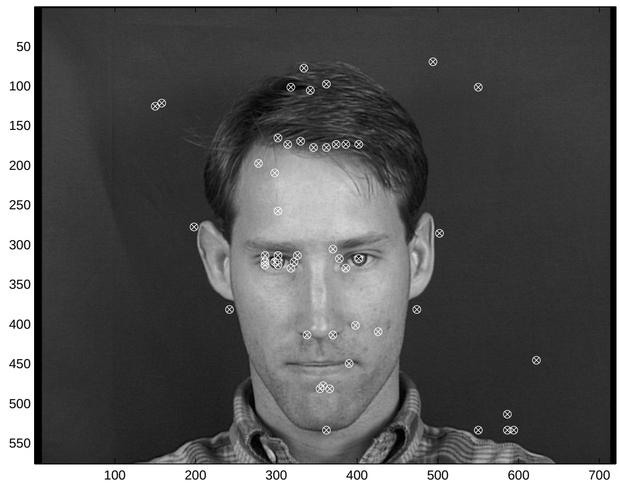
Figure 4.9: Authentication system.

In *Publication IV* and *Publication VI* the feature calculated for 10 different evidence classes (left/right outer eye corner, left/right eye center, left/right inner eye corner, left/right nostril, and left/right mouth corner) was the Gabor feature matrix in Eq. (3.16). The matrix responses were normalized to unity to achieve the illumination invariance. The classifier used in the publications was the simple clustering based method, subcluster classifier (SCC), but the proposed classifier can be replaced for example by the Bayesian classifier and the expectation maximization estimation of Gaussian mixture model probability distributions. The benchmark data set was XM2VTS facial image database of 600 training images and 560 test images [97]. In Fig. 4.10(a) is shown an example image and extracted evidence candidates which are passed to the face instance validation module described in *Publication VI* (see also Fig. 4.9). In Fig. 4.10(b) are shown the results of the face detection system as percents of images for which a specific error level was achieved. The error was defined as

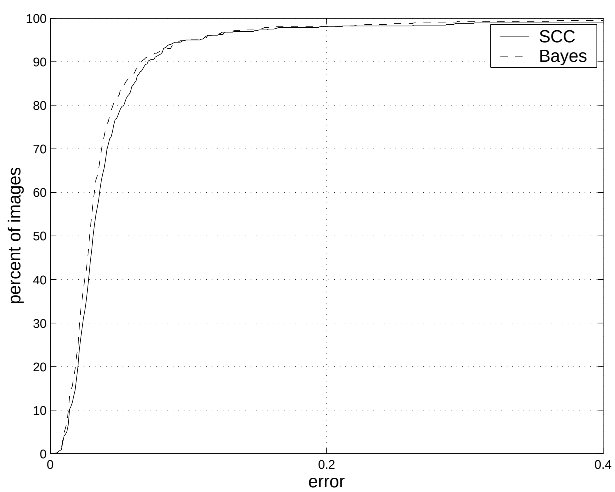
$$d_{eye} = \frac{\max(d_l, d_r)}{\parallel C_l - C_r \parallel} \tag{4.10}$$

where C_l and C_r are the correct eye center coordinates and d_l and d_r distances between the detected eye centers and the correct ones [65]. The translation, rotation, and scale invariant face evidence extraction was performed by the response matrix shift operations in the simple Gabor feature space.

The results for XM2VTS database in Fig. 4.10 were achieved using the method in *Publication VI* and it seems that the simple Gabor feature space outperforms the previous implementation of the same method with different features, Harris detectors, published in [94] and the detection is more accurate than the competitive method used in commercial systems [65]. Thus, a further development of the proposed method may provide a very robust and accurate face detection algorithm to be used in future authentication systems.



(a)



(b)

Figure 4.10: Face detection results: (a) examples of 5 best evidence candidates of the 10 classes. (b) face detection results for XM2VTS database using Gabor features with sub-cluster (SCC) and Bayesian classifiers.

CHAPTER V

Discussion

This thesis consists of publications presenting mainly practical results from experiments conducted with Gabor filter based features and of chapters tying the publications together under a common subject; feature extraction using Gabor filters. The common feature is Gabor filter, which have been utilized in all publications. During the sub-tasks no exclusive theoretical or even comparative analysis was applied, and thus, the purpose of this thesis is to study and understand the theory and properties of the features based on the Gabor filters; to present advantages that would prove Gabor filters to be optimal selections in the experiments conducted. Clearly, no distinct advantages have been shown, but on the other hand, the evolution of the Gabor features in the separate sub-tasks and the success in the experiments have been noteworthy for pointing out no insuperable disadvantages; this thesis is not a claim for the superiority of Gabor features but a humble study of the properties providing novelty value for feature extraction.

In Chapter 2 the basic works leading to the development of Gabor functions were reviewed and it was shown how the development branched out to Gabor expansion research, utilizing the theory of frames [42], and to Gabor filter research having analogies to the short-time Fourier transform [1, 114]. In Chapter 3 the parameters of the Gabor filter were considered in order to show how certain invariant properties can be achieved. Also, to explain the success in the reported experiments, the results in *Publication I - Publication VIII* were reviewed in Chapter 3 in the context of the most important properties of the Gabor features. As practical machine vision applications based on the results in this thesis the detection of electric components (*Publication II*) and face detection (*Publication IV* and *Publication VI*) were presented, and for 1-d signal processing a tool for industrial signal diagnosis (*Publication V* and *Publication VII*) was introduced.

There is one distinct goal, also an objective in this thesis, which has intrigued image processing researchers for decades: invariant object recognition. When considering invariance it is often the geometric invariance which is of greatest interest: a method should be capable of recognizing objects present in any location, orientation, and scale. Even for rigid 2-d objects, to have anything but merely theoretical interest, the method should have a certain degree of robustness to photometric disturbances, such as illumination change and image noise, and to natural image variations, such as backgrounds, and in addition, give even a partial recognition in the case of occlusion and to generalize over variation in the presence of the objects themselves. It is not possible to clearly define which of these requirements are in the scope of feature extraction research and which should be considered in upper layer processing and are, for example, problems of machine learning. Recognition in the biological system is not a feed-forward but a closed-loop mechanism providing an enormous amount of feedback to the preceding levels [108]. However, it is reasonable to prefer low-level methods that do not inhibit processing capabilities on upper layers. In these terms this thesis has also contributed to invariant object recognition research as noise and distortion tolerances and robustness issues were studied, and the feature space with the geometric invariant search operations was introduced. It seems that the invariance properties can be established in Gabor feature spaces and the Gabor features themselves tolerate many harmful distortions in images and objects where other methods may fail. Evidence supporting the theory were collected in the empirical parts in the publications. The theory and methods provided in this thesis can be applied to many application areas and in future it will be intriguing to develop the proposed methods towards genuine invariant object detection in the most complex problem domains where also other computational issues, such as classification of an enormous number of data points, will be relevant research issues.

This thesis is also a look at the time-frequency features. There are new methods, wavelets [36] being the most prominent, appearing in fields where time-frequency methods have traditionally been used and some may even claim that time has passed time-frequency representations by. However, in this study the invariance properties based on time-frequency information have been shown to be beneficial and new connections have been made to recently established concepts important in pattern recognition, such as the shiftability [129], which is evidence that there is still unrevealed potential in feature extraction with Gabor filters. This development can be seen also in wavelet research where despite the computational power of orthogonal wavelets the beneficial properties of non-orthogonal, redundant, wavelets have been noted and the wavelets are being further developed towards the theory of frames; for example framelets [37]. Future research will show if harmony between Gabor functions and wavelets can be established and whether a comprehensive framework for general feature extraction can be realized.

Bibliography

[1] Allen, J., and Rabiner, L. A unified approach to short-time Fourier analysis and synthesis. *Proceedings of the IEEE 65*, 11 (1977), 1558–1564.

[2] Allen, J. B. Short term spectral analysis, synthesis, and modification by discrete Fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing 25*, 3 (1977), 235–238.

[3] Amoroso, F. The bandwidth of digital data signals. *IEEE Communications Magazine 18*, 6 (1980).

[4] Ayinde, O., and Yang, Y.-H. Face recognition approach based on rank correlation of Gabor-filtered images. *Pattern Recognition 35*, 6 (2002), 1275–1289.

[5] Baddeley, R. An efficient code in V1? *Nature 381*, 13 (1996), 560–561.

[6] Bastiaans, M. J. Gabor's expansion of a signal into Gaussian elementary signals. *Proceedings of the IEEE 68*, 4 (1980), 538–539.

[7] Bastiaans, M. J. Optical generation of Gabor's expansion coefficients for rastered signals. *Optica Acta 29*, 10 (1982), 1349–1357.

[8] Bastiaans, M. J. Gabor's signal expansion and the Zak transform. *Applied Optics 33*, 23 (1994), 5241–5255.

[9] Bastiaans, M. J., and Geilen, M. C. On the discrete Gabor transform and the discrete Zak transform. *Signal Processing 49*, 3 (1996), 151–166.

[10] Belkasim, S., Shridhar, M., and Ahmadi, M. Pattern recognition with moment invariants: A comparative study and new results. *Pattern Recognition 24*, 12 (1991), 1117–1138.

[11] Benno, S., and Moura, J. Nearly shiftable scaling functions. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing* (1995), vol. 2, pp. 1097–1100.

[12] Benno, S., and Moura, J. Scaling functions robust to translations. *IEEE Transactions on Signal Processing 46*, 12 (1998), 3269–3281.

[13] Bigün, J., and du Buf, J. H. N-folded symmetries by complex moments in Gabor space and their application to unsupervised texture segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 16*, 1 (1994), 80–87.

[14] Blasdel, G., and Obermayer, K. Putative strategies of scene segmentation in monkey visual cortex. *Neural Networks 7*, 6/7 (1994), 865–881.

[15] Bodnarova, A., Bennamoun, M., and Latham, S. Optimal Gabor filters for textile flaw detection. *Pattern Recognition 35*, 12 (2002), 2973–2991.

[16] Bölcskei, H., and Hlawatcsh, F. Equivalence of DFT filter banks and Gabor expansion. In *SPIE Proceedings, Wavelet Applications in Signal and Image Processing III* (San Diego, CA, USA, 1995), vol. 2569, pp. 128–139.

[17] Bölcskei, H., Hlawatsch, F., and Feichtinger, H. Frame-theoretic analysis of oversampled filter banks. *IEEE Transactions on Signal Processing 46*, 12 (1998), 3256–3268.

[18] Bovik, A., Clark, M., and Geisler, W. Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence 12*, 1 (1990), 55–73.

[19] Bracewell, R. N. *The Fourier Transform and its Applications*, 2nd ed. McGraw-Hill, Inc., 1978.

[20] Bresch, M. Optimizing filter banks for supervised texture recognition. *Pattern Recognition 35*, 4 (2002), 783–790.

[21] Buhmann, J., Lades, M., and von der Malsburg, C. Size and distortion invariant object recognition by hierarchical graph matching. In *Proceedings of the International Joint Conference on Neural Networks* (1990), pp. 411–416.

[22] Buhmann, J., Lange, J., and von der Malsburg, C. Distortion invariant object recognition by matching hierarchically labeled graphs. In *Proceedings of the International Joint Conference on Neural Networks* (1989), vol. 1, pp. 155–159.

[23] Burt, P., and Adelson, E. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications 31*, 4 (1983), 532–540.

[24] Caelli, T., and Moraglia, G. On the detection of Gabor signals and discrimination of Gabor textures. *Vision Research 25*, 5 (1985), 671–684.

[25] Carmona, R., Hwang, W., and Torrésani, B. Multiridge detection and time-frequency reconstruction. *IEEE Transactions on Signal Processing 47*, 2 (1999), 480–492.

[26] Casasent, D., and Psaltis, D. New optical transforms for pattern recognition. *Proceedings of the IEEE 65*, 1 (1977), 77–84.

[27] Chan, W., Coghill, G., and Sivaswamy, J. A simple mechanism for curvature detection. *Pattern Recognition Letters 22*, 6-7 (2001), 731–739.

[28] Chen, J., Sato, Y., and Tamura, S. Orientation space filtering for multiple orientation line segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22*, 5 (2000), 417–429.

[29] Chen, V., and Ling, H. Joint time-frequency analysis for radar signal and image processing. *IEEE Signal Processing Magazine 16*, 2 (1999), 81–93.

[30] Clausi, D., and Jernigan, M. Designing Gabor filters for optimal texture separability. *Pattern Recognition 33*, 11 (2000), 1835–1849.

[31] Cohen, L. Time-frequency distributions – A review. *Proceedings of the IEEE 77*, 7 (1989), 941–981.

[32] Cohen, L. The scale representation. *IEEE Transactions on Signal Processing 41*, 12 (1993), 3275–3292.

[33] Cohen, L. *Time-Frequency Analysis*. Prentice-Hall, Inc., 1995.

[34] Cutrona, L., Leith, E., Palermo, C., and Porcello, L. Optical data processing and filtering systems. *IRE Transactions on Information Theory 6*, 3 (1960), 386–400.

[35] Dahlke, S., and Maass, P. The affine uncertainty principle in one and two dimensions. *Computers & Mathematics with Applications 30*, 3–6 (1995), 293–305.

[36] Daubechies, I. The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory 36*, 5 (1990), 961–1005.

[37] Daubechies, I., Han, B., Ron, A., and Shen, Z. Framelets: MRA-based constructions of wavelet frames. *Applied and Computational Harmonic Analysis 14*, 1 (2003), 1–46.

[38] Daugman, J. G. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A 2*, 7 (1985), 1160–1169.

[39] Daugman, J. G. Complete discrete 2-d Gabor transform by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing 36*, 7 (1988), 1169–1179.

[40] Daugman, J. G. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence 15*, 11 (1993), 1148–1161.

[41] Dunn, D., and Higgins, W. Optimal Gabor filters for texture segmentation. *IEEE Transactions on Image Processing 4*, 7 (1995), 947–964.

[42] Feichtinger, H., and Strohmer, T., Eds. *Gabor Analysis and Algorithms*. Birkhäuser, 1998.

[43] Fleet, D., Jepson, A., and Jenkin, M. Phase-based disparity measurement. *CVGIP: Image Understanding 53*, 2 (1991), 198–210.

[44] Fogel, I., and Sagi, D. Gabor filters as texture discriminator. *Biological Cybernetics 61* (1989), 103–113.

[45] Franco, H., and Pauletti, R. Analysis of nonlinear oscillations by Gabor spectrograms. *Nonlinear Dynamics 12*, 3 (1997), 215–236.

[46] Fränti, P., Mednonogov, A., Kyrki, V., and Kälviäinen, H. Content-based mathing of line-drawing images using the Hough transform. *International Journal on Document Analysis and Recognition 3* (2000), 117–124.

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

[48] Friedlander, B., and Porat, B. Detection of transient signals by the Gabor representation. *IEEE Transactions on Acoustics, Speech, and Signal Processing 37*, 2 (1989), 169–180.

[49] Gabor, D. Theory of communications. *Journal of Institution of Electrical Engineers 93* (1946), 429–457.

[50] Genossar, T., and Porat, M. Can one evaluate the Gabor expansion using Gabor's iterative algorithm. *IEEE Transactions on Signal Processing 40*, 8 (1992), 1852–1861.

[51] Gopalan, K., Anderson, T., and Cupples, E. A comparison of speaker identification results using features based on cepstrum and Fourier-Bessel expansion. *IEEE Transactions on Speech and Audio Processing 7*, 3 (1999), 289–294.

[52] Granlund, G. H. In search of a general picture processing operator. *Computer Graphics and Image Processing 8* (1978), 155–173.

[53] Griffin, D., and Lim, J. Signal estimation from modified short-time Fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing 32*, 2 (1984), 236–243.

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

[55] Grossmann, A., and Morlet, J. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis 15*, 4 (1984), 723–736.

[56] Hamouz, M., Kittler, J., Matas, J., and Bílek, P. Face detection by learned affine correspondences. In *Joint international Workshop on Syntactical & Structural & Statistical Pattern Recognition* (Windsor, Ontario, Canada, 2002), pp. 566–575.

[57] Haykin, S. *Communication Systems*, 3rd ed. John Wiley & Sons, Inc., 1994.

[58] Helstrom, C. An expansion of a signal in Gaussian elementary signals. *IEEE Transactions on Information Theory 12*, 1 (1966), 81–82.

[59] Hlawatsch, F., and Boudreaux-Bartels, G. Linear and quadratic time-frequency signal representations. *IEEE Signal Processing Magazine 9*, 2 (1992), 21–67.

[60] Hong, L., Wan, Y., and Jain, A. Fingerprint image enhancement: Algorithm and performance evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 20*, 8 (1998), 777–789.

[61] Jain, A., and Farrokhnia, F. Unsupervised texture segmentation using Gabor filters. *Pattern Recognition 24*, 12 (1991), 1167–1186.

[62] Jain, A., Ratha, N., and Lakshmanan, S. Object detection using Gabor filters. *Pattern Recognition 30*, 2 (1997), 295–309.

[63] Jain, A. K., and Bhattacharjee, S. K. Address block location on envelopes using Gabor filters. *Pattern Recognition 25*, 12 (1992), 1459–1477.

[64] Janssen, A. Optimality property of the Gaussian window spectrogram. *IEEE Transactions on Signal Processing 39*, 1 (1991), 202–204.

[65] Jesorsky, O., Kirchberg, K., and Frischholz, R. Robust face detection using the Hausdorff distance. In *Proceedings of the Third International Conference on Audio- and Video-based Biometric Person Authentication* (2001), pp. 90–95.

[66] Jiang, Q. Orthogonal multiwavelets with optimum time-frequency resolution. *IEEE Transactions on Signal Processing 46*, 4 (1998), 830–844.

[67] Jr., S. L. M. Are quadratic time-frequency representations really necessary? In *Proceedings of the ICASSP-96 IEEE International Conference on Acoustics, Speech, and Signal Processing* (1996), vol. 5, pp. 2575–2578.

[68] Kalocsai, P., Neven, H., and Steffens, J. Statistical analysis of Gabor-filter representation. In *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition* (Nara, Japan, 1998), pp. 360–365.

[69] Kasparis, T., Charalampidis, D., Georgiopoulos, M., and Rolland, J. Segmentation of textured images based on fractals and image filtering. *Pattern Recognition 34*, 10 (2001), 1963–1973.

[70] Klapper, J., and Harris, C. On the response and approximation of Gaussian filters. *IRE Transactions on Audio 7*, 3 (1959), 80–87.

[71] Knutsson, H., Wilson, R., and Granlund, G. Anisotropic nonstationary image estimation and its applications: Part I – restoration of noisy images. *IEEE Transactions on Communications 31*, 3 (1983), 388–397.

[72] Kreyszig, E. *Introductory Functional Analysis with Applications*. John Wiley & Sons, 1989.

[73] Krüger, V., and Sommer, G. Gabor wavelet networks for efficient head pose estimation. *Image and Vision Computing 20*, 9-10 (2002), 665–672.

[74] Kruizinga, P., and Petkov, N. Nonlinear operator for oriented texture. *IEEE Transactions on Image Processing 8*, 10 (1999), 1395–1407.

[75] Kyrki, V. *Local and Global Feature Extraction for Invariant Object Recognition*. PhD thesis, Lappeenranta University of Technology, 2002.

[76] Kyrki, V., Kamarainen, J.-K., and Kälviäinen, H. Content-based image matching using Gabor filtering. In *Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems Theory and Applications* (Baden-Baden, Germany, 2001), pp. 45–49.

[77] Kyrki, V., Kamarainen, J.-K., and Kälviäinen, H. Invariant shape recognition using global Gabor features. In *12th Scandinavian Conference on Image Analysis* (Bergen, Norway, June 2001), pp. 671–678.

[78] Lades, M., Vorbrüggen, J., Buhmann, J., Lange, J., von der Malsburg, C., Würtz, R., and Konen, W. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers 42*, 3 (1993), 300–311.

[79] Lahajnar, F., and Kovačič, S. Rotation-invariant texture classification. *Pattern Recognition Letters 24*, 9-10 (2003), 1151–1161.

[80] Lampinen, J. *Neural Pattern Recognition: Distortion Tolerance by Self-Organizing Maps*. PhD thesis, Lappeenranta University of Technology, 1992.

[81] Lampinen, J., and Oja, E. Distortion tolerant pattern recognition based on self-organizing feature extraction. *IEEE Transactions on Neural Networks 6*, 3 (1995), 539–547.

[82] Lampinen, J., Tamminen, T., Kostiainen, T., and Kalliomäki, I. Bayesian object matching based on MCMC sampling and Gabor filters. In *Proceedings of the SPIE Intelligent Robots and Computer Vision XX: Algorithms, Techniques, and Active Vision* (2001), vol. 4572, pp. 41–50.

[83] Lee, T. S. Image representation using 2d Gabor wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence 18*, 10 (1996), 959–971.

[84] Li, C., Zheng, C., and Tai, C. Detection of ECG characteristic points using wavelet transforms. *IEEE Transactions on Biomedical Engineering 42*, 1 (1995), 21–28.

[85] Li, S., and M. Healy, J. A parametric class of discrete Gabor expansions. *IEEE Transactions on Signal Processing 44*, 2 (1996), 201–211.

[86] Lin, J. Feature extraction of machine sound using wavelet and its application in fault diagnosis. *NDT & E International 34*, 1, 25–30.

[87] Lin, J., and Qu, L. Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis. *Journal of Sound and Vibration 234*, 1 (2000), 135–148.

[88] Liu, C., and Wechsler, H. Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition. *IEEE Transactions on Image Processing 11*, 4 (2002), 467–476.

[89] Lu, Y., and Morris, J. Gabor expansion for adaptive echo cancellation. *IEEE Signal Processing Magazine 16*, 2 (1999), 68–80.

[90] Lugt, A. V. Signal detection by complex spatial filtering. *IEEE Transactions on Information Theory 10*, 2 (1964), 139–145.

[91] Magrin-Chagnolleau, I., Durou, G., and Bimbot, F. Application of time-frequency principal component analysis to text-independent speaker identification. *IEEE Transactions on Speech and Audio Processing 10*, 6 (2002), 371–378.

[92] Mallat, S. G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 11*, 7 (1989), 674–693.

[93] Manjunath, B., and Ma, W. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence 18*, 8 (1996).

[94] Matas, J., Bílek, P., Hamouz, M., and Kittler, J. Discriminative regions for human face detection. In *Proceedings of ACCV2002: The 5th Asian Conference on Computer Vision* (2002).

[95] McAulay, R., and Quatieri, T. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech, and Signal Processing 34*, 4 (1986), 744–754.

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

[97] Messer, K., Matas, J., Kittler, J., Luettin, J., and Maitre, G. XM2VTSDB: The extended M2VTS Database. In *Proceedings of Second International Conference on Audio and Video-based Biometric Person Authentication* (1999), R. Chellapa, Ed., pp. 72–77.

[98] Namuduri, K., Mehrotra, R., and Ranganathan, N. Efficient computation of Gabor filter based multiresolution responses. *Pattern Recognition 27*, 7 (1994), 925–938.

[99] Navarro, R., Vargas, A., and Campos, J. A parallel implementation of the optical Gabor-wavelet transform. *Journal of Optics A: Pure and Applied Optics 1* (1999).

[100] Nuttall, A., and Amoroso, F. Minimum Gabor bandwidth of M orthogonal signals. *IEEE Transactions on Information Theory 11*, 3 (1965), 440–444.

[101] Nyquist, H. Certain factors affecting telegraph speed. *Bell System Technical Journal 3* (1924), 324–346.

[102] Oğuztöreli, M. N. Modeling and simulation of vertebrate primary visual system: Basic network. *IEEE Transactions on Systems, Man, and Cybernetics 13*, 5 (1983), 766–781.

[103] Okajima, K. An infomax-based learning rule that generates cells similar to visual cortical neurons. *Neural Networks 14*, 9 (2001), 1173–1180.

[104] Olshausen, B., and Field, D. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature 381*, 13 (1996), 607–609.

[105] Olshausen, B., and Field, D. Sparse coding with an overcomplete basis set: A strategy employed by V1. *Vision Research 37*, 23 (1997), 3311–3325.

[106] Oppenheim, A., and Lim, J. The importance of phase in signals. *Proceedings of the IEEE 69*, 5 (1981), 529–541.

[107] Oram, M., and Perret, D. Modeling visual recognition from neurobiological constraints. *Neural Networks 7*, 6/7 (1994), 945–972.

[108] Palmer, S. E. *Vision Science – Photons to Phenomenology.* The MIT Press, 1999.

[109] Park, H., and Yang, H. Invariant object detection based on evidence accumulation and Gabor features. *Pattern Recognition Letters 22*, 8 (2001), 869–882.

[110] Perona, P. Deformable kernels for early vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence 17*, 5 (1995).

[111] Persoon, E., and Fu, K.-S. Shape discrimination using Fourier descriptors. *IEEE Transactions on Systems, Man, and Cybernetics 7*, 3 (1977), 170–179.

[112] Peterfreund, N., and Zeevi, Y. Nonuniform image representation in area-of-interest systems. *IEEE Transactions on Image Processing 4*, 9 (1995), 1202–1212.

[113] Porat, M., and Zeevi, Y. The generalized Gabor scheme of image representation in biological and machine vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence 10*, 4 (1988), 452–468.

[114] Portnoff, M. R. Time-frequency representation of digital signals and systems based on short-time Fourier analysis. *IEEE Transactions on Acoustics, Speech, and Signal Processing 28*, 1 (1980), 55–69.

[115] Pötzsch, M., Krüger, N., and von der Malsburg, C. Improving object recognition by transforming Gabor filter responses. *Network: Computation in Neural Systems 7*, 2 (1996), 341–347.

[116] Purkat, P., and Chakravorti, S. Pattern classification of impulse faults in transformers by wavelet analysis. *IEEE Transactions on Dielectrics and Electrical Insulation 9*, 4 (2002), 555–561.

[117] Qian, S., and Chen, D. Discrete Gabor transform. *IEEE Transactions on Signal Processing 41*, 7 (1993), 2429–2438.

[118] Qian, S., and Chen, D. Decomposition of the Wigner-Ville distribution and time-frequency distribution series. *IEEE Transactions on Signal Processing 42*, 10 (1994), 2836–2842.

[119] Qian, S., and Chen, D. Joint time-frequency analysis. *IEEE Signal Processing Magazine 16*, 2 (1999), 52–67.

[120] Qiu, S. Generalized dual Gabor atoms and best approximations by Gabor family. *Signal Processing 49*, 3 (1996), 167–186.

[121] Rossing, T. D. *The Science of Sound*, 2nd ed. Addison-Wesley Publishing Company, Inc., 1990.

[122] Saito, N. *Local Feature Extraction and Its Applications Using a Library of Bases*. PhD thesis, Yale University, 1994.

[123] Saito, N., Coifman, R., Geshwind, F., and Warner, F. Discriminant feature extraction using empirical probability density estimation and a local basis library. *Pattern Recognition 35*, 12 (2002), 2841–2852.

[124] Schiltz, R. Forcing frequency identification of rolling element bearings. *Sound and Vibration* (1990), 16–19.

[125] Schoen, R., Habetler, T., Kamran, F., and Bartfield, R. Motor bearing damage detection using stator current monitoring. *IEEE Transactions on Industry Applications 31*, 6 (1995), 1274–1279.

[126] Shams, L., Brady, M., and Schaal, S. Graph matching vs mutual information maximization for object detection. *Neural Networks 14*, 3 (2001), 345–354.

[127] Shannon, E. A mathematical theory of communication. *The Bell System Technical Journal 27* (1948), 397–423, 623–656.

[128] Sherrah, J., Gong, S., and Ong, E. Face distributions in similarity space under varying head pose. *Image and Vision Computing 19*, 12 (2001), 807–819.

[129] Simoncelli, E., Freeman, W., Adelson, E., and Heeger, D. Shiftable multiscale transforms. *IEEE Transactions on Information Theory 38*, 2 (1992), 587–607.

[130] Slepian, D. On bandwidth. *Proceedings of the IEEE 64*, 3 (1976), 292–300.

[131] Smeraldi, F., and Bigun, J. Retinal vision applied to facial features detection and face authentication. *Pattern Recognition Letters 23*, 4 (2002), 463–475.

[132] Stockwell, R., Mansinha, L., and Lowe, R. Localization of the complex spectrum: The S transform. *IEEE Transactions on Signal Processing 44*, 4 (1996), 998–1001.

[133] Strang, G., and Nguyen, T. *Wavelets and Filter Banks*, revised ed. Wellesley-Cambridge Press, 1997.

[134] Teuner, A., Pichler, O., and Hosticka, B. Unsupervised texture segmentation of images using tuned matched Gabor filters. *IEEE Transactions on Image Processing 4*, 6 (1995), 863–870.

[135] Theimer, W., and Mallot, H. Phase-based binocular vergence control and depth reconstruction using active vision. *CVGIP: Image Understanding 60*, 3 (1994), 343–358.

[136] Triesch, J., and von der Malsburg, C. Classification of hand postures against complex backgrounds using elastic graph matching. *Image and Vision Computing 20*, 13–14 (2002), 937–943.

[137] Tsai, D.-M., Wu, S.-K., and Chen, M.-C. Optimal Gabor filter design for texture segmentation using stochastic optimization. *Image and Vision Computing 19*, 5 (2001), 299–316.

[138] Umesh, S., Cohen, L., Marinovic, N., and Nelson, D. Scale transform in speech analysis. *IEEE Transactions on Speech and Audio Prcessing 7*, 1 (1999).

[139] Urieli, S., Porat, M., and Cohen, N. Optimal reconstruction of images from localized phase. *IEEE Transactions on Image Processing 7*, 6 (1998), 838–853.

[140] van Leest, A. J. The discrete Gabor transform and the discrete Zak transform on a quincunx lattice. In *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis* (Pittsburgh, Pennsylvania, USA, 1998), pp. 33–36.

[141] Wang, L., Chen, C.-T., and Lin, W.-C. An efficient algorithm to compute the complete set of discrete Gabor coefficients. *IEEE Transactions on Image Processing 3*, 1 (1994), 87–92.

[142] Wang, Q. The shiftability of some wavelet bases. *Computers and Mathematics with Applications 40*, 8-9 (2000), 957–964.

[143] Weber, D., and Casasent, D. Quadratic Gabor filters for object detection. *IEEE Transactions on Image Processing 10*, 2 (2001), 218–230.

[144] Weldon, T., Higgins, W., and Dunn, D. Efficient Gabor filter design for texture segmentation. *Pattern Recognition 29*, 12 (1996), 2005–2015.

[145] Wilson, R., and Granlund, G. The uncertainty principle in image processing. *IEEE Transactions on Pattern Analysis and Machine Intelligence 6*, 6 (1984), 758–767.

[146] Wilson, R., and Knutsson, H. Uncertainty and inference in the visual system. *IEEE Transactions on Systems, Man, and Cybernetics 18*, 2 (1988), 305–312.

[147] Wilson, R., Knutsson, H., and Granlund, G. Anisotropic nonstationary image estimation and its applications: Part II – predictive image coding. *IEEE Transactions on Communications 31*, 3 (1983), 398–406.

[148] Wiskott, L. *Labeled Graphs and Dynamic Link Matching for Face Recognition and Scene Analysis.* PhD thesis, Ruhr-Universität Bochum, 1995.

[149] Wiskott, L., Fellous, J.-M., Krüger, N., and von der Malsburg, C. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence 19*, 7 (1997), 775–779.

[150] Wundrich, I., von der Malsburg, C., and Würtz, R. Image representation by the magnitude of the discrete Gabor wavelet transform. Working paper.

[151] Würtz, R. P. *Multilayer Dynamic Link Networks for Establishing Image Point Correspondences and Visual Object Recognition.* PhD thesis, Ruhr-Universität Bochum, 1994.

[152] Xia, X.-G. On characterization of the optimal biorthogonal window functions for Gabor transforms. *IEEE Transactions on Signal Processing 44*, 1 (1996), 133–136.

[153] Xia, X.-G. System identification using chirp signals and time-variant filters in the joint time-frequency domain. *IEEE Transactions on Signal Processing 45*, 8 (1997), 2072–2084.

[154] Yang, J., Liu, L., Jiang, T., and Fan, Y. A modified Gabor filter design method for fingerprint image enhancement. *Pattern Recognition Letters 24*, 12 (2003), 1805–1817.

[155] Yao, J., Krolak, P., and Steele, C. The generalized Gabor transform. *IEEE Transactions on Image Processing 4*, 7 (1995), 978–988.

[156] Yazici, B., and Kliman, G. An adaptive statistical time-frequency method for detection of broken bars and bearing faults in motors using stator current. *IEEE Transactions on Industry Applications 35*, 2 (1999), 442–452.

[157] Yu, W., Daniilidis, K., and Sommer, G. Approximate orientation steerability based on angular Gaussians. *IEEE Transactions on Image Processing 10*, 2 (2001), 193–204.

[158] Zeevi, Y., and Shlomot, E. Nonuniform sampling and antialiasing in image representation. *IEEE Transactions on Signal Processing 41*, 3 (1993), 1223–1236.

[159] Zhang, J., Tan, T., and Ma, L. Invariant texture segmentation via circular Gabor filters. In *Proceedings of the International Conference on Pattern Recognition* (2002), vol. 2, pp. 901–904.

Publications

Publication I

Kamarainen, J.-K., Kyrki, V., and Kälviäinen, H.,
"Noise Tolerant Object Recognition Using Gabor Filtering",

Reprinted, with permission, from *Proceedings of the 14th International Conference on Digital Signal Processing (Santorini, Greece, 2002)*, vol. 2, pp. 1349–1352.

Publication II

KAMARAINEN, J.-K., KYRKI, V., AND KÄLVIÄINEN, H.,
"Fundamental Frequency Gabor Filters for Object Recognition",

Publication III

KAMARAINEN, J.-K., KYRKI, V., KÄLVIÄINEN, H.,
"Robustness of Gabor Feature Parameter Selection",

Reprinted, with permission, from *Proceedings of the IAPR Workshop on Machine Vision Applications (Nara, Japan, 2002)*, pp. 132–135.

Publication IV

Kamarainen, J.-K., Kyrki, V., Hamouz, M., Kittler, J., Kälviäinen, H.,
"Invariant Gabor Features for Face Evidence Extraction",

Reprinted, with permission, from *Proceedings of the IAPR Workshop on Machine Vision Applications (Nara, Japan, 2002)*, pp. 228–231.

Publication V

Kamarainen, J.-K., Kyrki, V., Lindh, T., Ahola, J., Partanen, J.
"Statistical Signal Discrimination for Condition Diagnosis",

Reprinted from *Proceedings of the Finnish Signal Processing Symposium (Tampere, Finland, 2003)*, pp. 195–198.

Publication VI

HAMOUZ, M., KITTLER, J., KAMARAINEN, J.-K., KÄLVIÄINEN, H.
"Hypotheses-Driven Affine Invariant Localization of Faces in Verification Systems",

Publication VII

LINDH, T., AHOLA, J., KAMARAINEN, J.-K., KYRKI, V., PARTANEN, J.
"Bearing Damage Detection Based on Statistical Discrimination of Stator Current",

 Proceedings of the 4th International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives (Atlanta, Georgia, USA, 2003), pp. 177–181.

Publication VIII

Kyrki, V., Kamarainen, J.-K., Kälviäinen, H.
"Simple Gabor Feature Space for Invariant Object Recognition",

Accepted for publication in *Pattern Recognition Letters*, Elsevier.