

Image Analysis via Applied Harmonic Analysis: Perceptual Image Quality Assessment, Visual Servoing, and Feature Detection

DISSERTATION

zur Erlangung des Grades
Doktor der Naturwissenschaften

– Dr. rer. nat. –

vorgelegt im
Fachbereich 3 (Mathematik & Informatik)
der Universität Bremen

von

Rafael Reisenhofer

Datum des Kolloquiums: 17. September 2018

Erstgutachterin: Prof. Dr. Emily J. King

Zweitgutachter: Prof. Dr. Armin Iske

Abstract

Certain systems of analyzing functions developed in the field of applied harmonic analysis are specifically designed to yield efficient representations of structures which are characteristic of common classes of two-dimensional signals, like images. In particular, functions in these systems are typically sensitive to features that define the geometry of a signal, like edges and curves in the case of images. These properties make them ideal candidates for a wide variety of tasks in image processing and image analysis. This thesis discusses three recently developed approaches to utilizing systems of wavelets, shearlets, and α -molecules in specific image analysis tasks. First, a perceptual image similarity measure is introduced that is solely based on the coefficients obtained from six discrete Haar wavelet filters but yields state of the art correlations with human opinion scores on large benchmark databases. The second application concerns visual servoing, which is a technique for controlling the motion of a robot by using feedback from a visual sensor. In particular, it will be investigated how the coefficients yielded by discrete wavelet and shearlet transforms can be used as the visual features that control the motion of a robot with six degrees of freedom. Finally, a novel framework for the detection and characterization of features such as edges, ridges, and blobs in two-dimensional images is presented and evaluated in extensive numerical experiments. Here, versatile and robust feature detectors are obtained by exploiting the special symmetry properties of directionally sensitive analyzing functions in systems created within the recently introduced α -molecule framework.

Kurzfassung

Im Feld der angewandten harmonischen Analysis wurde eine Reihe spezieller Funktionensysteme entwickelt, die sich besonders gut zur Darstellung häufig auftretender Strukturen in zweidimensionalen Bildern eignen. Elemente in diesen Funktionensystemen zeichnen sich vor allem durch eine hohe Sensitivität gegenüber Merkmalen aus, die die Geometrie eines Bildes definieren, wie etwa Kanten oder Kurven. In der vorliegenden Arbeit betrachten wir drei neuartige Verfahren zur Anwendung derartiger Funktionensysteme im Bereich der digitalen Bildverarbeitung. Zunächst wird ein perzeptuelles Maß für Bildähnlichkeit entwickelt, welches auf der Faltung der zu vergleichenden Bilder mit sechs diskreten Haar-Wavelet-Filtern beruht. In der experimentellen Analyse zeigt sich, dass das neu definierte Maß, trotz seiner Einfachheit, auf umfangreichen Testdatenbanken eine erstaunlich hohe Korrelation mit menschlichen Urteilen aufweist. Die zweite Anwendung betrifft die Steuerung von Robotern, die mit visuellen Sensoren ausgestattet sind. Konkret wird untersucht, wie die Koeffizienten diskreter Wavelet- und Shearlet-basierter Transformationen der gelieferten Bilder genutzt werden können, um die Bewegung eines Roboters mit sechs Freiheitsgraden zu kontrollieren. Schließlich wird eine neuartige Methode zur Erkennung und Charakterisierung von Kanten, Linien und Kreisflächen in zweidimensionalen Bildern vorgestellt und in ausführlichen numerischen Experimenten evaluiert. Zur Erkennung von Strukturen werden hierbei vor allem die speziellen Symmetrieeigenschaften von orientierungssensitiven zweidimensionalen Funktionen verwendet, die unter Verwendung der erst kürzlich entwickelten α -Molekül-Theorie konstruiert wurden.

Danksagung

Mein Dank gilt allen Freund*innen, Wegbegleiter*innen und Mentor*innen, die mich auf der langen akademischen Reise zur Fertigstellung dieser Dissertation begleitet und unterstützt haben. Ich denke, Leistungen können stets nur bedingt individuell zugeschrieben werden. Jeder gute Gedanke, der auf den kommenden Seiten zu finden sein mag, wäre vor allem der Verdienst eines Umfelds, das mir so großzügig die inhaltlichen und materiellen Möglichkeiten zur Durchführung der vorliegenden Untersuchungen gegeben hat. Für jede Ungenauigkeit und jeden Irrtum übernehme ich hingegen gerne selbst die Verantwortung.

Besonders danken möchte ich meiner Betreuerin, Prof. Emily J. King, für ihr Vertrauen in mich und die über Jahre hinweg gewissenhafte Anleitung meiner Arbeit, sowie meinen Eltern, Ingrid und Martin, für ihre unbedingte Liebe und Unterstützung.

Acknowledgments

I want to thank all my friends, colleagues, and mentors, who accompanied and supported me on the long academic journey towards this dissertation. I believe that most achievements can only be in part attributed to single individuals. Any interesting thought that might be contained in this thesis is mostly due to an environment that generously provided me with the support and inspiration necessary to conduct the present investigations. For possible mistakes and errors, however, I gladly take responsibility myself.

In particular, I am grateful to my advisor, Prof. Emily J. King, for her trust and guidance, and to my parents, Ingrid and Martin, for their unconditional love and support.

Contents

1	Introduction	1
1.1	Notation	7
2	Perceptual Image Quality Assessment	11
	<i>Prelude</i> The Haar Wavelet	11
2.1	Introduction	16
2.1.1	Contributions	18
2.1.2	The Feature Similarity Index	19
2.2	The Haar Wavelet-Based Perceptual Similarity Index	21
2.2.1	Parameter Selection	25
2.3	Experimental Results	26
2.4	Conclusion	31
3	Visual Servoing	37
	<i>Prelude</i> Wavelet- and Shearlet-Based Transforms	37
3.1	Introduction	53
3.1.1	Contributions	54
3.1.2	Photometric Visual Servoing	55
3.2	Wavelet-Based Visual Servoing	57
3.2.1	Computing the Interaction Matrix via Derivative Wavelets	59
3.3	Shearlet-Based Visual Servoing	62
3.4	Numerical and Experimental Validation	64
3.4.1	Numerical Simulations	66
3.4.2	Experimental Validation	69
3.5	Conclusion	75
4	Edge, Ridge, and Blob Detection	79
	<i>Prelude</i> α -Molecules	79
4.1	Introduction	86
4.1.1	Contributions	88
4.2	Related Work	88
4.2.1	Derivative of Gaussian Wavelets	89
4.2.2	Feature Detection via Phase Congruency	90
4.2.3	Maximum Point Estimation From Discrete Samples	92

4.3	Symmetric Molecule-Based Feature Detection	93
4.3.1	Symmetric α -Molecules for Edge, Ridge, and Blob Detection	102
4.3.2	Edge Detection	105
4.3.3	Ridge Detection	107
4.3.4	Blob Detection	110
4.3.5	Estimation of Local Tangent Directions for Edges and Ridges	112
4.4	Implementation for Digital Images	113
4.5	Performance Evaluation on Synthetic Images	116
4.5.1	Edges	119
4.5.2	Ridges	122
4.5.3	Blobs	125
4.6	Applications	126
4.6.1	Determining Widths and Orientations of Retinal Blood Vessels . . .	127
4.6.2	Automated Cell Colony Counting	131
4.6.3	Detection and Characterization of Flame Fronts	134
4.6.3.1	Synthetic Inputs	137
4.6.3.2	Edge and Ridge Detection on PLIF Images	141
4.7	Discussion of Numerical Results	141
4.8	Conclusion	143
4.9	Acknowledgments	146
	Bibliography	147
	Appendix	169
	A Pearson Correlations for Different Image Quality Assessment Metrics	169
	B Examples of Visual Feature Vectors Used for Visual Servoing	171
	C Additional Visual Servoing Experiments	173
	D Synthetic Images Used for the Evaluation of SymFD	176
	E Visual Comparison of Feature Detectors on Synthetic Flame Data	179
	List of Acronyms	185

CHAPTER 1

Introduction

Over the course of this thesis, we will develop and analyze novel approaches to three diverse tasks from the fields of computer vision, robotics, and digital image processing. The common denominator of these applications is that at their core, all of them boil down to the processing and analysis of two-dimensional digital images. The unifying theme of the present thesis is that in each of the three newly developed methods, the respective image analysis will be carried out by applying certain tools and methods that originate from the field of applied harmonic analysis.

Applied harmonic analysis is a thriving and innovative area in applied mathematics, whose contributions have been highly influential in theoretical and applied areas of research alike. In the past five years alone, methods from applied harmonic analysis have played a significant role in groundbreaking works such as the first observation of gravitational waves [1], Martin Hairer’s seminal theory of regularity structures [77], and in making first steps towards a rigorous mathematical understanding of the intriguing success of modern deep learning architectures [19].

Broadly speaking, the fundamental research goal of applied harmonic analysis is the development and rigorous analysis of systems of functions that, not unlike sets of Lego bricks, can be used to efficiently represent arbitrary elements from certain classes of functions or signals. Its origins can be traced back to the famous French mathematician Jean Baptiste Joseph Fourier, who claimed in his book *Théorie analytique de la chaleur* (The Analytical Theory of Heat) [62], first published in 1822, that any 2π -periodic function that is integrable over $[-\pi, \pi]$ could be written as

$$f(x) = \sum_{n \in \mathbb{Z}} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int} dt \right) e^{inx}, \quad \text{for all } x \in [-\pi, \pi). \quad (1.1)$$

In particular, Fourier had a profound impact on 19th century physics by showing how expansions of this form, which yield representations of arbitrary functions in terms of linear combinations of the eigenfunctions e^{inx} of the Laplace operator, can be applied to solve linear partial differential equations such as the heat equation.

While Fourier's original claim that the right hand side of (1.1) would yield pointwise convergence for *any* 2π -periodic function was eventually refuted¹, it turns out that the set of normalized complex exponentials $(2\pi)^{-1/2} \{e^{inx}\}_{n \in \mathbb{Z}}$ yields an orthonormal basis (ONB) for $L^2(-\pi, \pi)$.

A next important step was the construction of expansions similar to (1.1) that are defined on the whole real line and could also be used to define ONBs for $L^2(\mathbb{R})$. One can easily observe that, due to their lack of decay, the complex exponentials considered by Fourier are not elements of $L^2(\mathbb{R})$ and thus cannot simply be used in this setting. In particular, the lack of localization also means that the coefficients of the Fourier series (1.1), that are defined as the L^2 -inner products of an analyzed function f with differently modulated complex exponentials, only yield information about *which* frequencies are dominant in f , but not *where* they occur. A simple but elegant solution to this issue was introduced by Dennis Gábor in the late 1940's in his seminal work on the theory of communication [64]. There, he proposed to multiplicatively combine complex exponentials pointwise with a Gaussian window function, which provided additional localization in the time domain, thereby introducing a methodology which is nowadays known as *time-frequency analysis*. Analogous to modulating the complex exponential, the thusly obtained analyzing functions could also be shifted, such that the obtained set of functions would cover both the complete frequency spectrum and the whole real line. Furthermore, by considering different types of window functions, such as the indicator function of a closed interval, this approach can indeed be used to obtain ONBs for $L^2(\mathbb{R})$ (see, e.g., [80]).

A significant drawback of considering window functions as a means of introducing spatial localization is that this approach is not adaptive, that is, the degree of localization remains fixed throughout the whole domain of analysis. In practice, however, signals are often alternating between smooth regions, on which large windows would be preferable, and regions that are governed by high frequencies and transient features, such as singularities, which would require highly localized analyzing functions. A brilliant way of obtaining systems of functions whose localization in the time domain is in a sense adaptive can be found in the theory of *wavelets*, which has roots in the works of early pioneers like Haar [76], Gábor [64], or Calderón [23], but rose to prominence in the late 1980's through the contributions of people like Ingrid Daubechies [43, 45], Stéphane Mallat [120], and Yves Meyer [128]. Systems of wavelets are constructed by dilating and shifting so-called mother wavelets, which can be imagined as functions with a localized oscillatory behavior. The main advantage of considering dilation instead of simple frequency modulation is that by squeezing or stretching a generator function, the dilation operator is simultaneously

1 A continuous counterexample was given by Paul du Bois-Reymond in 1873 (see, e.g., [98]).

adjusting the frequency *and* the degree of spatial localization of a function. This approach not only yields highly efficient representation systems for certain types of functions, such as piecewise smooth functions, but can also be used to obtain a wide class of ONBs for $L^2(\mathbb{R})$ (see, e.g., the preludes to Chapters 2 and 3).

The present thesis mainly deals with methods for the analysis of two-dimensional images. In the continuous realm, such images are often modeled as elements of the space $L^2(\mathbb{R}^2)$. Accordingly, we will predominantly consider transforms of two-dimensional functions that are defined on $L^2(\mathbb{R}^2)$. There are indeed straightforward approaches to generalize the theory of wavelets to the two-dimensional setting, which yield practical and efficient multiscale representations of two-dimensional images (see, e.g., the prelude to Chapter 3). However, it can also be shown that wavelet-based transforms are in a sense not optimal for dealing with curvilinear features, such as edges or lines, which typically govern the geometry of two-dimensional images. The main reason for this is that such features are *anisotropic*, that is, they are locally associated with a specific direction (e.g., the tangent direction at an edge point), while the dilation operator used in the definition of two-dimensional wavelet transforms is *isotropic* in the sense that it dilates equally along both dimensions. This motivated the development of various types of multiscale representation systems for two-dimensional signals that are based on anisotropic dilation in the sense that the employed dilation operator affects one dimension more than the other. Examples of such systems include curvelets [25], contourlets [49], ridgelets [26, 27], and shearlets [104, 110]. Anisotropic dilation yields high-frequency analyzing functions that have a strong directional bias. Analogous to the cases of time-frequency analysis, and one-dimensional wavelet analysis, where the use of localized generators required the application of a shift operator to cover the whole real line, directionally sensitive analyzing functions require the application of an additional operator which is capable of changing the preferred orientation of a function. For instance, this can be achieved by applying rotation matrices, in the case of curvelets, or shear matrices, in the case of shearlets, to the argument of a generating function. It turns out that for specific model classes of two-dimensional images, systems that are based on anisotropic dilation are in fact capable of providing in a sense optimal approximation rates (see, e.g., the preludes to Chapters 3 and 4). Furthermore, wavelet- and shearlet-based transforms in particular can be associated with faithful transitions between the continuous and the discrete realm, in the sense that the continuous-time coefficients that are based on L^2 -inner products can directly be linked to discrete convolutions of a given digital image with certain two-dimensional filters (see, e.g., [107, 120]).

The provable efficiency in representing piecewise smooth one- and two-dimensional signals as well as the availability of fast discrete-time transforms make the aforementioned multiscale representation systems great candidates for a wide range of tasks in image

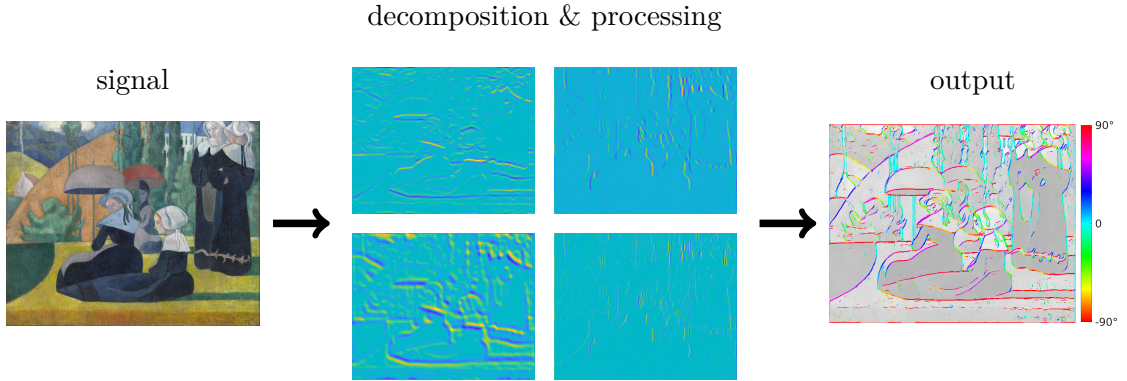


Figure 1.1: The methods presented in this thesis all follow the basic pipeline depicted above. In a first step, a transform is applied to a given input image, which was specifically designed to yield analysis coefficients that carry the significant information with respect to the considered task. Then, a processing step is performed on the coefficients, which yields the desired output. In the example presented above, the output highlights the edges of the input image along with their local tangent directions. The output was obtained by utilizing systems of anisotropically scaled and rotated generator functions with special symmetry properties. The respective edge and tangent direction measures will be defined in Chapter 4. (Input image: *Les bretonnes aux ombrelles* (Émile Bernard, 1892). Source: Wikimedia Commons.)

processing and image analysis. In many cases, the respective applications are strongly related to the notion of *sparsity*. Broadly speaking, a sequence is called sparse if it only contains a small number of non-zero coefficients. The concept of sparsity has recently gained a lot of attention in the wake of the theory of *compressed sensing* [24, 61], which provides guarantees for the exact recovery of a signal from only a small number of measurements under the assumption that the signal satisfies certain sparsity constraints. This concept of *sparse regularization* is also widely utilized in the context of digital image processing, where an additional sparsity constraint can be applied to define tractable optimization problems whose solutions yield, for example, denoised, or inpainted versions of a given signal [95], or which separate distinct morphological components in an image (see, e.g., [155]).

In the methods presented in this thesis, however, the capability of the considered multiscale systems to yield sparse representations will only be used implicitly. In fact, in most cases, we will not even require the considered function systems to provide complete descriptions of the analyzed functions in the sense that it can be fully recovered from the obtained analysis coefficients. In contrast, we will directly utilize the ability of specifically tailored transforms to be particularly sensitive towards the structures and properties of an image that are most significant in the considered application. The analysis coefficients yielded by the respective transforms can then directly be processed to yield the desired results. The basic pipeline of this approach to applied harmonic analysis-based image

analysis, which will be followed by all of the methods discussed in the three main chapters of this thesis, is schematically depicted in Figure 1.1.

The thesis is structured such that each of the following three chapters is devoted to a specific application. In all three cases, we will discuss and analyze a novel approach to solving the respective tasks via measures, or algorithms, which are based on specific types of multiscale image transforms. To allow for a progressive build-up of the associated theory, the chapters are ordered with respect to the complexity of the considered transforms. Furthermore, each chapter is preceded by a prelude, which aims to provide a comprehensive introduction to the respective tools and ideas from applied harmonic analysis that will subsequently be used. Note that the three preludes are written such that they could also be read independently of the other contents of this thesis.

Chapter 2 introduces the Haar wavelet-based perceptual similarity index (HaarPSI), a novel full reference image similarity measure that is solely based on the analysis coefficients obtained from convolving the considered input images with six discrete Haar wavelet filters. Given a pair of images, where one image is a distorted version of the other, the goal of such a measure is to quantitatively assess the similarity between both images such that the result matches the human perception of similarity. The HaarPSI incorporates various important principles from the early visual system, such as orientation, and frequency selectivity, or visual saliency, and can be considered as a structurally, and computationally simplified version of the well-known feature similarity index (FSIM) [174]. Despite its simple computational structure, the HaarPSI achieves state of the art correlations with human opinion scores on several large benchmarking databases.

In Chapter 3, we discuss a novel framework for applying wavelet- or shearlet-based multiscale image representations in the context of vision-based robot control, which is also known as visual servoing. Classical visual servoing methods typically rely on the precise detection and tracking of specific features within the observed domain. To increase the stability in unfavorable conditions, so-called *direct* visual servoing schemes have recently been introduced, which only consider the observed image intensities. The newly proposed wavelet- and shearlet-based visual servoing control laws can be considered as a best of both worlds approach in the sense that the detail coefficients of a wavelet or shearlet transform are sensitive to structures and thereby implicit feature detectors, while the coarse approximation yielded by a multiscale image representation is closely related to the feature vectors used in direct visual servoing schemes. Besides an analytical derivation of the so-called *interaction matrix*, which linearly relates the motion of the robot with changes in the wavelet- and shearlet-based feature vectors, we discuss a number of simulated and real-world experiments that aim to evaluate the applicability and respective strengths and weaknesses of the newly proposed control laws.

In Chapter 4, we present a novel approach to the detection and characterization of edges, ridges, and blobs in two-dimensional images which exploits the symmetry properties of directionally sensitive analyzing functions in multiscale systems that are constructed in the framework of α -molecules [71, 73]. The proposed feature detectors are inspired by the notion of phase congruency, stable in the presence of noise, and by definition invariant to changes in contrast. We also show how the behavior of coefficients corresponding to differently scaled and oriented analyzing functions can be used to obtain a comprehensive characterization of the geometry of features in terms of local tangent directions, widths, and heights. The accuracy and robustness of the proposed measures are validated and compared to various state of the art algorithms in extensive numerical experiments in which we consider sets of clean and distorted synthetic images that are associated with reliable ground truths. To further demonstrate the applicability, we show how the proposed measures can be used to characterize the geometry of flame fronts, to detect and characterize blood vessels in digital retinal images, and how the proposed blob measure can be applied to automatically count the number of cell colonies in a Petri dish.

All methods presented in this thesis aim to fully exploit the vast theoretical knowledge that has been developed within the field of applied harmonic analysis to this date. However, the works themselves are undoubtedly more of an applied than of a theoretical nature. The following list summarizes the main theoretical contributions:

- Sections 3.2 and 3.3 contain an analytical derivation of interaction matrices that provide a linear relationship between the velocity of a visual sensor and the associated wavelet- and shearlet-based visual feature vectors.
- Theorem 3.2.1 gives a simple rule for obtaining wavelet and scaling filters associated with a *derivative wavelet* multiresolution analysis (MRA).
- Theorem 4.3.4 shows that the systems of two-dimensional symmetric molecules considered in the definition of the respective feature detectors in Chapter 4 indeed fall into the framework of α -molecules.
- Lemma 4.3.1 provides a closed-form expression for the L^1 -norms of arbitrary derivatives of the Gaussian based on the zeros of the associated Hermite polynomials. This result is helpful for defining L^1 -normalized symmetric molecules which play an integral part in the definition of the newly proposed edge, ridge, and blob measures. Furthermore, it is used in Theorem 4.3.5, which associates the second and fourth derivatives of the Gaussian with a notion of a radius, which is required for the definition of the newly proposed ridge and feature width measures.

Table 1.1: List of commonly used symbols

Symbol	Description
x	scalar
\mathbf{x}	vector; elements in $\ell^2(\mathbb{Z})$
x_n or $x[n]$	n -th entry in \mathbf{x}
\mathbf{x}_n or $\mathbf{x}[n]$	n -th element in a sequence of vectors
\mathbf{A}	matrix; elements in $\ell^2(\mathbb{Z}^2)$
$A_{n,m}$ or $A[m,n]$	entries of the matrix \mathbf{A}
\mathbf{A}_n or $\mathbf{A}[n]$	n -th element in a sequence of matrices
$f(x)$	continuous-time function
$\mathbf{f}(x)$	vector-valued function
$\mathbf{H}(x)$	matrix-valued function
\bar{x}	complex conjugate
$ x $	absolute value
$\lceil x \rceil$	ceiling function; $\lceil x \rceil = \min\{n \in \mathbb{Z}: n \geq x\}$
$\lfloor x \rfloor$	floor function; $\lfloor x \rfloor = \max\{n \in \mathbb{Z}: n \leq x\}$
\mathbb{N}	$\{1, 2, 3, \dots\}$
\mathbb{N}_0	$\{0, 1, 2, 3, \dots\}$
\mathbb{R}_+	positive real numbers
$\mathbb{R}_{\geq 0}$	non-negative real numbers

1.1 Notation

To ensure a consistent discrimination of continuous-time and discrete-time objects, we will use bold symbols to denote vectors, matrices, and vector-, or matrix-valued functions, and non-bold symbols to denote scalars and scalar-valued functions (see Table 1.1).

We use $L^p(\mathbb{R}^d)$ with $p, d \in \{1, 2\}$ to denote the Banach space of equivalence classes of functions $f: \mathbb{R}^d \rightarrow \mathbb{C}$ that are equal almost everywhere and satisfy

$$\|f\|_p = \left(\int_{\mathbb{R}^d} |f(\mathbf{x})|^p \, d\mathbf{x} \right)^{1/p} < \infty. \quad (1.2)$$

Furthermore, for $p = 2$, the space $L^2(\mathbb{R}^d)$ is a separable Hilbert space equipped with the inner product

$$\langle f, g \rangle_{L^2} = \int_{\mathbb{R}^d} f(\mathbf{x}) \overline{g(\mathbf{x})} \, d\mathbf{x}, \quad f, g \in L^2(\mathbb{R}^d). \quad (1.3)$$

We denote the Fourier transform of a function $f \in L^1(\mathbb{R}^d)$ as

$$\widehat{f}(\boldsymbol{\xi}) = \int_{\mathbb{R}^d} f(\mathbf{x}) e^{-2\pi i \langle \mathbf{x}, \boldsymbol{\xi} \rangle_{\ell^2}} d\mathbf{x}, \quad \boldsymbol{\xi} \in \mathbb{R}^d, \quad (1.4)$$

with the ℓ^2 -inner product $\langle \cdot, \cdot \rangle_{\ell^2}$ as defined below, and consider the usual extension to $L^2(\mathbb{R}^d)$.

In the discrete-time setting, we use $\ell^p(\mathbb{Z})$ for $p > 0$ to denote the space of sequences that satisfy

$$\|\mathbf{x}\|_p = \left(\sum_{n \in \mathbb{Z}} |x_n|^p \right)^{1/p} < \infty. \quad (1.5)$$

Note that for $p \in (0, 1)$, $\|\cdot\|_p$ is only a quasi-norm and that $\ell^2(\mathbb{Z})$ is again a Hilbert space equipped with the inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\ell^2} = \sum_{n \in \mathbb{Z}} x_n y_n, \quad \mathbf{x}, \mathbf{y} \in \ell^2(\mathbb{Z}). \quad (1.6)$$

Furthermore, for sequences \mathbf{x} and \mathbf{y} , we define the discrete convolution operator as

$$(\mathbf{x} * \mathbf{y})[m] = \sum_{n \in \mathbb{Z}} x_n \cdot y_{m-n}, \quad m \in \mathbb{Z}, \quad (1.7)$$

and the dyadic down-, and upsampling operators via

$$(\mathbf{x}_{\downarrow 2})[n] = x_{2n}, \quad (\mathbf{x}_{\uparrow 2})[n] = \begin{cases} x[n/2] & n \text{ is even,} \\ 0 & \text{else,} \end{cases}, \quad n \in \mathbb{Z}, \quad (1.8)$$

respectively.

Furthermore, with a measurable space X , we define the indicator function $\mathbb{1}_B: X \rightarrow \{0, 1\}$ of a measurable set $B \subset X$ via

$$\mathbb{1}_B(x) = \begin{cases} 1 & x \in B, \\ 0 & \text{else.} \end{cases} \quad (1.9)$$

The edge, ridge, and blob measures introduced in Chapter 4 rely heavily on the use of functions with special symmetry properties. We call a function $f: \mathbb{R} \rightarrow \mathbb{R}$ *even-symmetric*, if it satisfies

$$f(x) = f(-x), \quad \forall x \in \mathbb{R}, \quad (1.10)$$

odd-symmetric, if it holds that

$$f(x) = -f(-x), \quad \forall x \in \mathbb{R}, \quad (1.11)$$

and *symmetric* if either (1.10), or (1.11) is true.

For functions f and g , we write $f(x) \lesssim g(x)$ to indicate that there exists a constant $C > 0$ such that

$$f(x) \leq C \cdot g(x) \quad \forall x. \quad (1.12)$$

If both $f(x) \lesssim g(x)$ and $f(x) \gtrsim g(x)$, we write $f(x) \asymp g(x)$.

In Chapter 3, we will sometimes consider displacements in three-dimensional space as elements of the special Euclidean group $SE(3)$. An element of $SE(3)$ is a pair (\mathbf{R}, \mathbf{t}) , where $\mathbf{R} \in SO(3)$ describes a rotation in three-dimensional space and $\mathbf{t} \in \mathbb{R}^3$ a translation. For simplicity, we will sometimes write elements in $SE(3)$ as six-dimensional vectors $[x, y, z, \omega_x, \omega_y, \omega_z]^\top$, where the first three components describe the translational and the latter three components the angular displacement with regards to the respective axis.

Throughout this thesis, we will often use dot-notation to denote functions that were obtained via an operator that acts on the argument of another function. For instance, for a translation parameter $t \in \mathbb{R}$, and a function f , we would write $f(\cdot - t)$ to denote the function obtained from shifting f by t .

CHAPTER 2

Perceptual Image Quality Assessment

The contents of this chapter are based on:

- REISENHOFER, R., BOSSE, S., KUTYNIOK, G., and WIEGAND, T.: “A Haar wavelet-based perceptual similarity index for image quality assessment”. *Signal Process. Image Comm.* (2018), vol. 61: pp. 33–43
-

Prelude The Haar Wavelet

A historically interesting example of an ONB for $L^2(0, 1)$ that can easily be extended to an ONB for $L^2(\mathbb{R})$ was already proposed in 1910 by Alfreed Haar [76]. In his dissertation, which was supervised by none other than David Hilbert, Haar showed that such a basis can be defined by considering a set of shifts and dilates of the simple step function

$$\psi^{\text{haar}}(x) = \begin{cases} 1 & 0 \leq x < \frac{1}{2}, \\ -1 & \frac{1}{2} \leq x < 1, \\ 0 & \text{else.} \end{cases} \quad (2.1)$$

To be precise, Haar considered the set of normalized dyadic dilates and their integer shifts

$$\Psi^{\text{haar}} = \left\{ \psi_{j,m}^{\text{haar}} = 2^{-j/2} \psi^{\text{haar}}(2^{-j} \cdot -m) : j, m \in \mathbb{Z} \right\}. \quad (2.2)$$

The Haar functions $\psi_{0,0}^{\text{haar}}$, $\psi_{-1,0}^{\text{haar}}$, and $\psi_{-2,0}^{\text{haar}}$ are plotted in Figure 2.1(a). It is in fact easy to show that Ψ^{haar} is an orthonormal system by exploiting the binary tree structure induced by the support sets of the elements $\psi_{j,m}^{\text{haar}}$, as illustrated by the following lemma.

Lemma 2.0.1 ([76]). *The set Ψ^{haar} is an orthonormal system.*

Proof. Note that for any $j, m \in \mathbb{Z}$, the function $\psi_{j,m}^{\text{haar}}$ is supported on the interval $I_{j,m} = [2^j m, 2^j(m+1))$ with $\psi_{j,m}^{\text{haar}}(x) = 2^{-j/2}$ for $x \in [2^j m, 2^j(m+1) - 2^{j-1})$, and $\psi_{j,m}^{\text{haar}}(x) =$

$-2^{-j/2}$ for $x \in [2^j m + 2^{j-1}, 2^j(m+1))$. Let $j_1, j_2, m_1, m_2 \in \mathbb{Z}$. A simple computation yields

$$\left\langle \psi_{j_1, m_1}^{\text{haar}}, \psi_{j_1, m_1}^{\text{haar}} \right\rangle_{L^2} = \int_{2^{j_1} m_1}^{2^{j_1}(m_1+1)} 2^{-j_1} dx = 1. \quad (2.3)$$

In the case of $j_1 = j_2$, the support sets of differently shifted Haar wavelets never overlap, that is, $I_{j_1, m_1} \cap I_{j_1, m_2} = \emptyset$ and therefore

$$\left\langle \psi_{j_1, m_1}^{\text{haar}}, \psi_{j_1, m_2}^{\text{haar}} \right\rangle_{L^2} = 0, \quad (2.4)$$

for $m_1 \neq m_2$. Let us finally assume without loss of generality that $j_2 < j_1$. In this case, the function $\psi_{j_1, m_1}^{\text{haar}}$ is constant on the interval I_{j_2, m_2} and

$$\left\langle \psi_{j_1, m_1}^{\text{haar}}, \psi_{j_2, m_2}^{\text{haar}} \right\rangle_{L^2} = \psi_{j_1, m_1}^{\text{haar}}(2^{j_2} m_2) \int_{I_{j_2, m_2}} \psi_{j_2, m_2}^{\text{haar}}(x) dx = 0, \quad (2.5)$$

independently of the choices of m_1 and m_2 . \square

A nice way of proving that Ψ^{haar} is also a basis is by showing that its orthogonal complement in $L^2(\mathbb{R})$ only contains the zero function. This is also the approach Haar used in his original work to show that a subset of Ψ^{haar} , in which only non-negative scaling parameters j are considered, together with the indicator function on the unit interval yields an ONB for $L^2(0, 1)$. Here, we will consider a different technique, which was also used for illustrative purposes by Ingrid Daubechies in [43], and explicitly construct linear combinations of Haar functions $\psi_{j, m}^{\text{haar}}$ that yield arbitrarily close approximations of a given function in $L^2(\mathbb{R})$. In particular, we will use the Haar functions to encode the difference between increasingly coarse approximations of the considered L^2 -function, that are based on functions that are piecewise constant on the intervals $I_{j, m}$ (see proof of Lemma 2.0.1). In doing so, we will implicitly construct a structure that defines a so-called multiresolution analysis (MRA) [120]. The MRA framework, which will be discussed in more detail in the prelude to Chapter 3, is a powerful tool that can be used to define an extensive class of orthonormal bases for $L^2(\mathbb{R})$. It furthermore establishes a faithful transition between the transform of an L^2 -function with respect to an ONB in the continuum and transforms of digital signals that are based on convolutions with finite and discrete filters.

Theorem 2.0.2 ([43, 76]). *The set Ψ^{haar} (see (2.2)) is an ONB for $L^2(\mathbb{R})$. In particular, for each function $f \in L^2(\mathbb{R})$ and every $\varepsilon > 0$, there exist positive integers $J, K \in \mathbb{N}$ and a*

finite sequence $\mathbf{c}_{-J} = \{c_{-J,m}\}_{m=-2^{2J}}^{2^{2J}-1}$ such that the linear combination of Haar functions

$$\tilde{f} = \sum_{j=-J+1}^J 2^{j/2} \sum_{m=-2^{J-j}}^{2^{J-j}-1} d_{j,m} \psi_{j,m}^{\text{haar}} + \sum_{j=J+1}^{J+1+K} 2^{j/2} \left(\frac{c_{J,-1}}{2^{j-J}} \psi_{j,-1}^{\text{haar}} - \frac{c_{J,0}}{2^{j-J}} \psi_{j,0}^{\text{haar}} \right), \quad (2.6)$$

where for $j \in \{-J+1, \dots, J\}$ the finite sequences \mathbf{d}_j and \mathbf{c}_j are defined as

$$\mathbf{d}_j = \left\{ d_{j,m} = -\frac{1}{2} c_{j-1,2m} + \frac{1}{2} c_{j-1,2m+1} \right\}_{m=-2^{J-j}}^{2^{J-j}-1}, \quad (2.7)$$

$$\mathbf{c}_j = \left\{ c_{j,m} = \frac{1}{2} c_{j-1,2m} + \frac{1}{2} c_{j-1,2m+1} \right\}_{m=-2^{J-j}}^{2^{J-j}-1}, \quad (2.8)$$

satisfies

$$\|f - \tilde{f}\|_2 < \varepsilon. \quad (2.9)$$

Proof. It was already shown in Lemma 2.0.1 that Ψ^{haar} is an orthonormal system. Let $f \in L^2(\mathbb{R})$ and $\varepsilon > 0$. The set of piecewise constant functions with compact support is dense in $L^2(\mathbb{R})$. Retaining the notation for intervals $I_{j,m} = [2^j m, 2^j(m+1))$ from the proof of Lemma 2.0.1, this means that there exists an integer $J \in \mathbb{N}$ and a sequence $\mathbf{c}_{-J} = \{c_{-J,m}\}_{m=-2^{2J}}^{2^{2J}-1}$ such that the piecewise constant function

$$f_{-J}(x) = \begin{cases} c_{-J,m} & x \in I_{-J,m}, m \in \{-2^{2J}, \dots, 2^{2J}-1\} \\ 0 & \text{else,} \end{cases} \quad (2.10)$$

which is compactly supported on the interval $[-2^J, 2^J]$, satisfies

$$\|f - f_{-J}\|_2 < \varepsilon/2. \quad (2.11)$$

For $j \in \{-J+1, \dots, J\}$, we now define sequences \mathbf{d}_j and \mathbf{c}_j as in (2.7) and (2.8), respectively, and piecewise constant functions f_j analogous to (2.10). Note that the piecewise constant functions f_j represent increasingly coarse approximations of f . For $j \in \{-J, \dots, J-1\}$, each function f_j can be written as the sum of its successor f_{j+1} and a linear combination of Haar wavelets weighted by the entries of the sequence \mathbf{d}_{j+1} :

$$f_j = \sum_{m=-2^{J-j}}^{2^{J-j}-1} c_{j,m} \mathbb{1}_{I_{j,m}} \quad (2.12)$$

$$= \sum_{m=-2^{J-j-1}}^{2^{J-j-1}-1} \frac{c_{j,2m} + c_{j,2m+1}}{2} \mathbb{1}_{I_{j+1,m}} + \frac{c_{j,2m+1} - c_{j,2m}}{2} (\mathbb{1}_{I_{j,2m}} - \mathbb{1}_{I_{j,2m+1}}) \quad (2.13)$$

$$= \sum_{m=-2^{J-j-1}}^{2^{J-j-1}-1} c_{j+1,m} \mathbb{1}_{I_{j+1,m}} + \sum_{m=-2^{J-j-1}}^{2^{J-j-1}-1} d_{j+1,m} (\mathbb{1}_{I_{j,2m}} - \mathbb{1}_{I_{j,2m+1}}) \quad (2.14)$$

$$= f_{j+1} + 2^{(j+1)/2} \sum_{m=-2^{J-j-1}}^{2^{J-j-1}-1} d_{j+1,m} \psi_{j+1,m}^{\text{haar}}. \quad (2.15)$$

In the coarsest case of f_{J-1} , Equation (2.15) reduces to a weighted sum of two pairs of indicator functions and Haar wavelets which are compactly supported on the intervals $[-2^J, 0]$ and $[0, 2^j]$, respectively:

$$f_{J-1} = c_{J,-1} \mathbb{1}_{J,-1} + c_{J,0} \mathbb{1}_{J,0} + 2^{J/2} (d_{J,-1} \psi_{J,-1}^{\text{haar}} + d_{J,0} \psi_{J,0}^{\text{haar}}). \quad (2.16)$$

To obtain an arbitrarily precise approximation of f , this process can be continued by smearing out the two remaining averages $c_{J,-1}$ and $c_{J,0}$ beyond the original support of f_{-J} . For $j \in \{J, \dots, J+K\}$, we define

$$f_j = \frac{c_{J,-1}}{2^{j+1-J}} \mathbb{1}_{I_{j+1,-1}} + \frac{c_{J,0}}{2^{j+1-J}} \mathbb{1}_{I_{j+1,0}} + 2^{(j+1)/2} \left(\frac{c_{J,-1}}{2^{j+1-J}} \psi_{j+1,-1}^{\text{haar}} - \frac{c_{J,0}}{2^{j+1-J}} \psi_{j+1,0}^{\text{haar}} \right). \quad (2.17)$$

By finally choosing the integer K large enough such that

$$\frac{1}{2^{K+1}} \left(c_{J,-1} \|\mathbb{1}_{I_{J+K+1,-1}}\|_2 + c_{J,0} \|\mathbb{1}_{I_{J+K+1,0}}\|_2 \right) = 2^{\frac{J-K+1}{2}} (|c_{J,-1}| + |c_{J,0}|) < \varepsilon/2, \quad (2.18)$$

the Haar wavelet-based approximation

$$\tilde{f} = \sum_{j=-J+1}^J 2^{j/2} \sum_{m=-2^{J-j}}^{2^{J-j}-1} d_{j,m} \psi_{j,m}^{\text{haar}} + \sum_{j=J+1}^{J+1+K} 2^{j/2} \left(\frac{c_{J,-1}}{2^{j-J}} \psi_{j,-1}^{\text{haar}} - \frac{c_{J,0}}{2^{j-J}} \psi_{j,0}^{\text{haar}} \right) \quad (2.19)$$

satisfies

$$\|f - \tilde{f}\|_2 = \left\| f - f_{-J} + \frac{c_{J,-1}}{2^{K+1}} \mathbb{1}_{I_{J+K+1,-1}} + \frac{c_{J,0}}{2^{K+1}} \mathbb{1}_{I_{J+K+1,0}} \right\|_2 \quad (2.20)$$

$$\leq \|f - f_{-J}\|_2 + \left\| \frac{c_{J,-1}}{2^{K+1}} \mathbb{1}_{I_{J+K+1,-1}} \right\|_2 + \left\| \frac{c_{J,0}}{2^{K+1}} \mathbb{1}_{I_{J+K+1,0}} \right\|_2 < \varepsilon. \quad (2.21)$$

□

The linear combination of Haar functions (2.6) is defined such that for any fixed scale $j \in \{-J+1, \dots, J\}$, the shifts of the function $\psi_{j,0}^{\text{haar}}$ are used to encode the difference

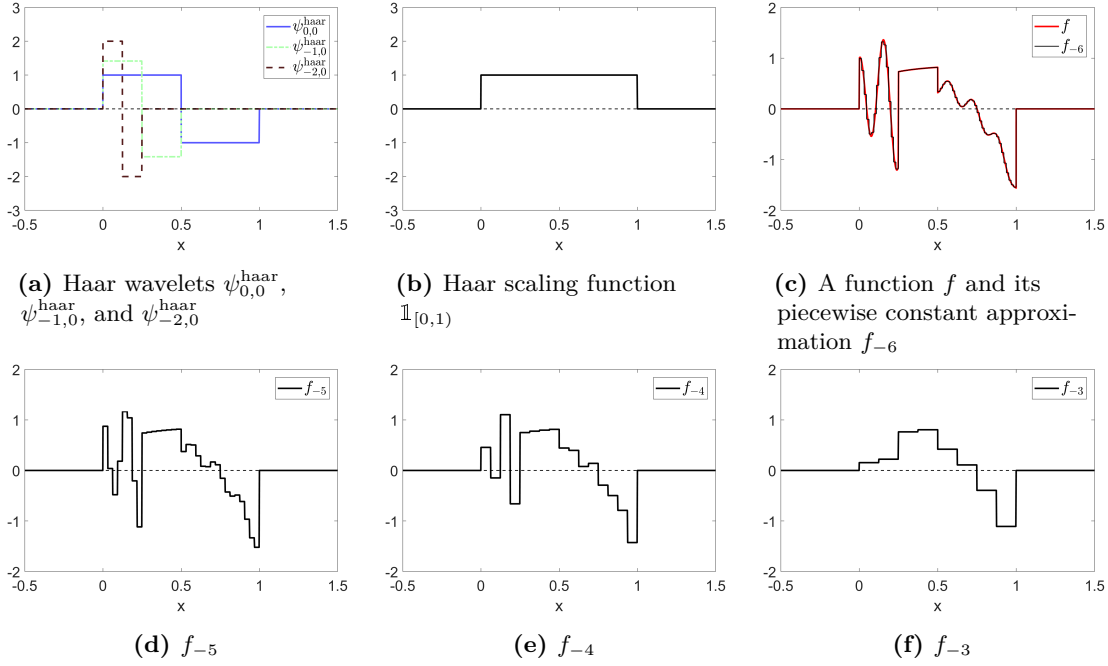


Figure 2.1: (a) and (b) show examples of three differently scaled Haar wavelets and the associated scaling function. (c)-(f) show plots of a function f and the respective piecewise constant approximations f_j for $j \in \{-6, \dots, -3\}$ (cf. proof of Theorem 2.0.2).

between the piecewise constant approximations f_{j-1} and f_j , that are themselves linear combinations of shifts of the dilated indicator functions $\mathbb{1}_{[0,1)}(2^{-j+1}\cdot)$ and $\mathbb{1}_{[0,1)}(2^{-j}\cdot)$, respectively. In this setup, the linear spans of shifts of dilated indicator functions implicitly define a nested structure of approximation spaces where the orthogonal complement of the space at scale j within the space at scale $j-1$ is given by the linear span of shifts of the dilated Haar function $\psi_{j,0}^{\text{haar}}$. This specific structure can in fact be generalized to yield the powerful MRA framework, in which the function taking the role of $\mathbb{1}_{[0,1)}$ is typically denoted as the *scaling function*, and ψ^{haar} as the *mother wavelet* (see prelude to Chapter 3). The term *wavelet* here refers to a certain class of locally oscillating functions that will also be discussed in more detail in the prelude to Chapter 3.

Another intriguing aspect of Theorem 2.0.2 is that the sequence \mathbf{c}_{-J} yields a finite and discrete description of an arbitrarily close approximation of a given function f when its entries are interpreted as the coefficients of a linear combination of translates of $\mathbb{1}_{[0,1)}(2^{-J}\cdot)$. Examples of approximations f_j for $j \in \{-6, \dots, -3\}$ of an L^2 -function f are shown in Figure 2.1. In particular, a closer look at Equations (2.7), (2.8), and (2.15) reveals that the approximation f_j at a fixed scale j can be written in terms of a linear combination of Haar functions $\psi_{j+1,m}^{\text{haar}}$ and a linear combination of indicator functions $\mathbb{1}_{[0,1)}(2^{-j-1}\cdot - m)$, where the respective coefficients are defined as the dyadically subsampled convolutions of

the sequence \mathbf{c}_j with the finite and discrete filters $\{-\frac{1}{2}, \frac{1}{2}\}$, and $\{\frac{1}{2}, \frac{1}{2}\}$, respectively. In other words, the proof of Theorem 2.0.2 suggests that, when interpreting a discrete and finite signal as the coefficients of a linear combination of translates of the indicator function $\mathbb{1}_{[0,1)}(2^{-J}\cdot)$, the associated sequences of Haar coefficients \mathbf{d}_j can directly be computed in the discrete setting by convolving the signal with a simple discrete and finite filter. This provides a faithful transition between the continuum, where the coefficients representing f in terms of the Haar ONB are given via the respective L^2 -inner products, and the discrete setting, which we can also exploit to justify the application of the so-called Haar wavelet filters (2.32) in the definition of the Haar wavelet-based perceptual similarity index (HaarPSI) later in this chapter. In particular, we will see in the prelude to Chapter 3 that each MRA uniquely defines a so-called *scaling filter* that assumes the role of $[\frac{1}{2}, \frac{1}{2}]$ in Equation (2.8). Due to the fact that the sequences \mathbf{c}_j define approximations of a given function f , they are often denoted as *approximation coefficients* while the sequences \mathbf{d}_j are termed *detail coefficients*.

2.1 Introduction

Digital images and videos are omnipresent in daily life and the importance of visual data is still growing. By 2020, nearly a million minutes of video content is estimated to cross the internet every second [34].

Typically, video and image signals are compressed before storage or transmission and ultimately intended to be viewed by humans. However, both compression and transmission errors can introduce distortions to video or image signals that are visible to human viewers. For evaluating or optimizing a transmission system or parts of it, for example by controlling the rate-distortion trade-off of a video encoder, it is crucial to measure the severity of distortions in a perceptually meaningful way. Quality in a perceptually meaningful way can only be measured reliably in psychometric tests. In such tests, participants are asked to rate the subjectively perceived quality of images or videos that have previously been subject to some kind of distortion introducing processing. The quality ratings of individual participants can eventually be averaged to obtain a single mean opinion score (MOS) for each stimulus. However, although being the gold standard for assessing perceived quality, such studies are expensive and time-consuming and not feasible for real-time tasks like optimizing or monitoring transmission systems.

Image quality assessment (IQA) methods typically belong to one of three categories with different challenges and scopes of applications: Full reference image quality assessment (FR IQA) approaches require and utilize the availability of a reference image. Reduced reference image quality assessment (RR IQA) methods exploit a small set of features

extracted from the reference image. No reference image quality assessment (NR IQA) approaches estimate the perceived quality of a possibly distorted image solely from the image itself [113]. Unconstrained NR IQA has the notion of being the holy grail of IQA and, when successful, essentially replicates human abilities. It is, however, not a feasible approach for some applications such as, for example, encoder control for video compression. An NR IQA metric used for rate-distortion optimization in a video encoder would steer the optimization towards coding decisions that remove any type of noise or artifacts. However, there are videos in which noise and artifacts were intentionally added to create a certain visual effect. As an example, the reader is invited to imagine a video encoder that removes film grain from the Quentin Tarantino movie *The Hateful Eight* due to the application of an NR IQA metric that penalizes in a sense noisy coding decisions. Such an encoder would change a deliberate artistic decision made by the filmmakers and thus deteriorate the viewing experience.

The simplest FR IQA metric is the mean squared error (MSE), which is defined as the average of the squared differences of the reference and the distorted image. Although being widely used, it does not correlate well with perceived visual quality [67]. More sophisticated approaches towards perceptually accurate IQA typically follow one of three strategies. *Bottom-up* approaches explicitly model various processing mechanisms of the human visual system (HVS), such as masking effects [167], contrast sensitivity [41], or just-noticeable-distortion [90, 118] in order to assess the perceived quality of images. For instance, the adaptivity of the HVS to the magnitude of distortions is modeled explicitly by the concept of most apparent distortion (MAD) [112] in order to apply two different assessment strategies for supra- and near-threshold distortions.

However, the method proposed in this chapter as well as most image quality metrics developed recently follow a *top-down* approach. There, general functional properties of the HVS are assumed in order to identify and to exploit image features corresponding to the perceived quality. Prominent examples are structural similarity index (SSIM) [165], visual information fidelity (VIF) [149], the gradient gradient similarity measure (GSM) [116], spectral residual-based similarity (SR-SIM) [172], and the visual saliency-induced index (VSI) [173]. The SSIM [165] aims at taking into account the sensitivity of the HVS towards structural information. This is done by pooling three complementary components, namely luminance similarity (comparing local mean luminance values), contrast similarity (comparing local variances) and structural similarity, which is defined as the local covariance between the reference image and its perturbed counterpart. Although being criticized [53], it is highly cited and among the most popular image quality assessment metrics. The SSIM was generalized for a multi-scale setting by the multi-scale structural similarity index (MS-SSIM) [166]. One of the first information theoretic approaches to FR IQA was the VIF [149].

The VIF models wavelet coefficients as Gaussian scale mixtures and quantifies the mutual information shared between reference and test images. The information theoretic measure of mutual information was furthermore shown to be correlated to perceived image quality. In [116], changes in contrast and structure are captured by considering local gradients, while the squared difference in pixel values between the reference image and the distorted image is used to measure luminance variations. This approach thus follows the basic framework of combining complementary feature maps originally introduced in [165]. Additionally, masking effects are estimated, based on the local gradient magnitude of the reference image and incorporated when the two feature maps are combined. The SR-SIM [172] takes into account changes in the local horizontal and vertical gradient magnitudes. Additionally, it incorporates changes in a spectral residual-based visual saliency estimate. The VSI [173] is similar to the SR-SIM in the sense that it assesses similarities in the gradient magnitude and utilizes visual saliency maps. However, it further exploits the visual saliency map by using its entries as weights in the spatial similarity pooling. Furthermore, [173] also explores the influence of different saliency models on the performance of the proposed image quality measure. A combination of two feature maps is also applied successfully by the feature similarity index (FSIM) [174]. Due to its conceptual similarity to the proposed method, it will be discussed in more detail in Section 2.1.2.

Adopting the advances in machine learning and data science, IQA methods following a third, purely *data driven* strategy have been proposed recently. So far, data driven approaches were mainly developed for the domain of NR IQA [17, 93, 170, 175], but they have also been adapted in the context of FR IQA [16].

2.1.1 Contributions

This work introduces the HaarPSI, a novel and computationally inexpensive measure yielding FR IQAs. The HaarPSI utilizes the magnitudes of high-frequency Haar wavelet coefficients to define local similarities and low-frequency Haar wavelet coefficients to weigh the importance of (dis)similarities at specific locations in the image domain.

The six discrete two-dimensional Haar wavelet filters used in the definition of the HaarPSI respond to horizontal and vertical edges on different frequency scales. The HaarPSI is thus based on elementary implementations of functional properties known to be exhibited by neurons in the primary visual cortex, namely orientation selectivity and spatial frequency selectivity. We aim to demonstrate that such a simple model already suffices to define a similarity measure that yields state of the art correlations with human opinion scores.

The HaarPSI can also be seen as a drastic simplification of the FSIM [174], which is

based on a similar combination of similarity and weight maps. In the definition of the FSIM, local similarities as well as the weights rely on the phase congruency measure [100]. Computing this measure requires images to be convolved with 16 complex-valued filters and contains several non-trivial steps such as adaptive thresholding. For the HaarPSI on the other hand, the two maps are computed from the responses of only six discrete Haar wavelet filters and are cleanly separated in the sense that local similarities and weights are based on different frequency scales. Surprisingly, these simplifications not only decrease the required computational effort but also lead to consistently higher correlations with human mean opinions scores.

In Section 2.3, we evaluate the consistency of the HaarPSI with the human quality of experience and compare its performance to state of the art similarity measures like the SSIM [165], the FSIM [174], and the VSI [173]. As depicted in Tables 2.1 and 2.2, the HaarPSI achieves higher correlations with human opinion scores than all other considered FR IQA metrics in all test cases except one, where it only comes second to the VSI. In addition, the HaarPSI can be computed significantly faster than the metrics yielding the second and third highest correlations with human opinion scores, namely the VSI and the FSIM. In order to facilitate reproducible research, a MATLAB and a Python implementation of the HaarPSI are publicly available at <http://www.haarpsi.org/> and github.com/rgcda/haarpsi.

It is both convenient and surprising that the promising experimental results of the HaarPSI are based on the responses of Haar filters, which are arguably the simplest and computationally most efficient wavelet filters existing. The results of a numerical analysis of the applicability of other wavelet filters in the newly proposed similarity measure can be found in Table 2.4.

2.1.2 The Feature Similarity Index

The FSIM [174], proposed in 2011, is currently one of the most successful and influential FR IQA metrics. The FSIM combines two feature maps derived from the phase congruency measure [100] and the local gradients of the reference and the distorted image to assess local similarities between two images. For a grayscale image $\mathbf{I} \in \ell^2(\mathbb{Z}^2)$, the gradient map is defined by

$$G_{\mathbf{I}}[\mathbf{m}] = \sqrt{((\mathbf{g}^{\text{hor}} * \mathbf{I})[\mathbf{m}])^2 + ((\mathbf{g}^{\text{ver}} * \mathbf{I})[\mathbf{m}])^2}, \quad \mathbf{m} \in \mathbb{Z}^2, \quad (2.22)$$

where \mathbf{g}^{hor} and \mathbf{g}^{ver} denote horizontal and vertical gradient filters (e.g. Sobel or Scharr filters), and $*$ denotes the discrete two-dimensional convolution operator. The method used

in the implementation of the FSIM to compute the phase congruency map was developed by Peter Kovessi [101] and contains several non-trivial operations, such as adaptive soft thresholding. In its essence, the phase congruency map of a grayscale image \mathbf{I} is given by

$$\text{PC}_{\mathbf{I}}[\mathbf{m}] \approx \frac{|\sum_n (\mathbf{g}_n^c * \mathbf{I})[\mathbf{m}]|}{\sum_n |(\mathbf{g}_n^c * \mathbf{I})[\mathbf{m}]|}, \quad \mathbf{m} \in \mathbb{Z}^2, \quad (2.23)$$

where \mathbf{g}_n^c denotes differently scaled and oriented complex-valued wavelet filters. The basic idea behind (2.23) is that if the obtained complex-valued wavelet coefficients have the same phase at a location \mathbf{m} , taking the absolute value of the sum is the same as taking the sum of the absolute values. If this is the case, $\text{PC}_{\mathbf{I}}[\mathbf{m}]$ will be close to or precisely 1. The phase congruency measure is also strongly related to the measures for the detection of edges, ridges, and blobs introduced in Chapter 4. It is therefore discussed in more detail in Section 4.2.2.

To assess local similarities between two images with respect to the maps defined in (2.22) and (2.23), the FSIM – like many other image quality metrics – uses a simple similarity measure for non-negative scalar values that already appeared in [165], namely

$$S(a, b, C) = \frac{2ab + C}{a^2 + b^2 + C}, \quad a, b \geq 0, \quad (2.24)$$

with a constant $C > 0$. The graph of $S(a, b, C)$ for values ranging from 0 to 100 and $C = 30$ is shown in Figure 2.2(b). The local feature similarity map for two grayscale images $\mathbf{I}_1, \mathbf{I}_2 \in \ell^2(\mathbb{Z}^2)$ is defined as

$$\text{FSIM}_{\mathbf{I}_1, \mathbf{I}_2}[\mathbf{m}] = S(G_{\mathbf{I}_1}[\mathbf{m}], G_{\mathbf{I}_2}[\mathbf{m}], C_1)^\beta \cdot S(\text{PC}_{\mathbf{I}_1}[\mathbf{m}], \text{PC}_{\mathbf{I}_2}[\mathbf{m}], C_2)^\gamma, \quad \mathbf{m} \in \mathbb{Z}^2, \quad (2.25)$$

with constants $C_1, C_2 > 0$ and exponents $\beta, \gamma > 0$. Based on the assumption that the HVS is especially sensitive towards structures at which the phases of the Fourier components are in congruency (see, e.g., [130]), the phase congruency map is not only used in (2.25) but also applied to determine the relative importance of different image areas with respect to human perception. The feature similarity index is computed by taking the weighted mean of all local feature similarities, where the phase congruency map is used as a weight function, that is

$$\text{FSIM}_{\mathbf{I}_1, \mathbf{I}_2} = \frac{\sum_{\mathbf{m}} \text{FSIM}_{\mathbf{I}_1, \mathbf{I}_2}[\mathbf{m}] \cdot \text{PC}_{\mathbf{I}_1, \mathbf{I}_2}[\mathbf{m}]}{\sum_{\mathbf{m}} \text{PC}_{\mathbf{I}_1, \mathbf{I}_2}[\mathbf{m}]}, \quad (2.26)$$

where

$$\text{PC}_{\mathbf{I}_1, \mathbf{I}_2}[\mathbf{m}] = \max \{\text{PC}_{\mathbf{I}_1}[\mathbf{m}], \text{PC}_{\mathbf{I}_2}[\mathbf{m}]\}, \quad \mathbf{m} \in \mathbb{Z}^2. \quad (2.27)$$

The FSIM was also generalized to a color-sensitive feature similarity index (FSIMC) that is

defined in the YIQ color space [174]. In the YIQ space, the Y channel encodes luminance information, while the I and Q channels encode chromatic information. Color images defined in the RGB color space can easily be transformed to the YIQ space with a linear mapping, namely

$$\begin{bmatrix} \mathbf{I}^Y \\ \mathbf{I}^I \\ \mathbf{I}^Q \end{bmatrix} \approx \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{I}^R \\ \mathbf{I}^G \\ \mathbf{I}^B \end{bmatrix}. \quad (2.28)$$

FSIMC simply incorporates the chroma channels I and Q into the local feature similarity measure (2.25) while the gradient maps as well as the phase congruency maps are purely derived from the luminance channel Y. Formally, the FSIMC is defined for color images \mathbf{I}_1 and \mathbf{I}_2 , where the single channels are denoted as $\mathbf{I}_1 = [\mathbf{I}_1^Y, \mathbf{I}_1^I, \mathbf{I}_1^Q]^\top$, and $\mathbf{I}_2 = [\mathbf{I}_2^Y, \mathbf{I}_2^I, \mathbf{I}_2^Q]^\top$, respectively, by

$$\text{FSIMC}_{\mathbf{I}_1, \mathbf{I}_2} = \frac{\sum_{\mathbf{m}} \text{FSC}_{\mathbf{I}_1, \mathbf{I}_2}[\mathbf{m}] \cdot \text{PC}_{\mathbf{I}_1^Y, \mathbf{I}_2^Y}[\mathbf{m}]}{\sum_{\mathbf{m}} \text{PC}_{\mathbf{I}_1^Y, \mathbf{I}_2^Y}[\mathbf{m}]}, \quad (2.29)$$

where the local feature similarity measure is now defined as

$$\text{FSC}_{\mathbf{I}_1, \mathbf{I}_2}[\mathbf{m}] = S(\mathbf{G}_{\mathbf{I}_1^Y}[\mathbf{m}], \mathbf{G}_{\mathbf{I}_2^Y}[\mathbf{m}], C_1)^\alpha \cdot S(\text{PC}_{\mathbf{I}_1^Y}[\mathbf{m}], \text{PC}_{\mathbf{I}_2^Y}[\mathbf{m}], C_2)^\beta \quad (2.30)$$

$$\cdot S(\mathbf{I}_1^I[\mathbf{m}], \mathbf{I}_2^I[\mathbf{m}], C_3)^\gamma \cdot S(\mathbf{I}_1^Q[\mathbf{m}], \mathbf{I}_2^Q[\mathbf{m}], C_4)^\gamma, \quad (2.31)$$

with $\mathbf{m} \in \mathbb{Z}^2$, constants $C_1, C_2, C_3, C_4 > 0$, and exponents $\alpha, \beta, \gamma > 0$.

2.2 The Haar Wavelet-Based Perceptual Similarity Index

The basic idea of the HaarPSI is to construct feature maps in the spirit of (2.22) as well as a weight function similar to (2.23) by considering a single wavelet filterbank. The response of any high-frequency wavelet filter will look similar to the response yielded by a gradient filter like the Sobel operator. Furthermore, the phase congruency measure used as a weight function in the FSIM is computed directly from the output of a multi-scale complex-valued wavelet filterbank, as illustrated by Equation (2.23). This gives a strong intuition that it should be possible to define a similarity measure derived from the response of a single set of discrete wavelet filters that at least matches the performance of the FSIM on benchmark databases but requires significantly less computational effort.

The wavelet chosen for this endeavor is the so-called Haar wavelet, which we already got to know in the prelude to this chapter. The one-dimensional ℓ^2 -normalized Haar filters are given by

$$\mathbf{h}_1^{1D} = \frac{1}{\sqrt{2}} \cdot \{1, 1\} \quad \text{and} \quad \mathbf{g}_1^{1D} = \frac{1}{\sqrt{2}} \cdot \{-1, 1\}, \quad (2.32)$$

where \mathbf{h}_1^{1D} denotes the low-pass scaling filter and \mathbf{g}_1^{1D} the corresponding high-pass wavelet filter (cf. Equations (2.8) and (2.7)). For any scale $j \in \mathbb{N}$, we can construct two-dimensional Haar filters by setting

$$\mathbf{g}_j^{(1)} = \mathbf{g}_j^{1D} \otimes \mathbf{h}_j^{1D}, \quad (2.33)$$

$$\mathbf{g}_j^{(2)} = \mathbf{h}_j^{1D} \otimes \mathbf{g}_j^{1D}, \quad (2.34)$$

where \otimes denotes the outer product and the one-dimensional filters \mathbf{h}_j^{1D} and \mathbf{g}_j^{1D} are given for $j > 1$ by

$$\mathbf{g}_j^{1D} = \mathbf{h}_1^{1D} * (\mathbf{g}_{j-1}^{1D})_{\uparrow 2}, \quad (2.35)$$

$$\mathbf{h}_j^{1D} = \mathbf{h}_1^{1D} * (\mathbf{h}_{j-1}^{1D})_{\uparrow 2}, \quad (2.36)$$

where $\uparrow 2$ is the dyadic upsampling operator, and $*$ denotes the one-dimensional convolution operator. Note that $\mathbf{g}_j^{(1)}$ responds to horizontal structures, while $\mathbf{g}_j^{(2)}$ picks up vertical structures. The six Haar filters used to define the HaarPSI are shown in Figure 2.2(a). The local similarity map $\mathbf{FS}_{\mathbf{I}_1, \mathbf{I}_2}$ multiplicatively combines gradient-based and phase congruency-based similarities whose contributions are weighted by the exponents $\alpha, \beta > 0$. The HaarPSI does not consider different types of similarities. However, to correctly predict the perceptual similarity experienced by human viewers, it can be useful to apply an additional non-linear mapping to the local similarities obtained from high-frequency Haar wavelet filter responses. This non-linearity is chosen to be the logistic function, which is widely used as an activation function in neural networks for modeling thresholding in biological neurons and is given for a parameter $\alpha > 0$ as

$$l_\alpha(x) = \frac{1}{1 + e^{-\alpha x}}. \quad (2.37)$$

For two grayscale images $\mathbf{I}_1, \mathbf{I}_2 \in \ell^2(\mathbb{Z}^2)$, the local similarity measure used to compute the HaarPSI is based on the first two stages of a two-dimensional discrete Haar wavelet transform and given by

$$\mathbf{HS}_{\mathbf{I}_1, \mathbf{I}_2}^{(k)}[\mathbf{m}] = l_\alpha \left(\frac{1}{2} \sum_{j=1}^2 S \left(\left| (\mathbf{g}_j^{(k)} * \mathbf{I}_1)[\mathbf{m}] \right|, \left| (\mathbf{g}_j^{(k)} * \mathbf{I}_2)[\mathbf{m}] \right|, C \right) \right), \quad \mathbf{m} \in \mathbb{Z}^2, \quad (2.38)$$

where $C > 0$, $k \in \{1, 2\}$ selects either horizontal or vertical Haar wavelet filters, S denotes the similarity measure (2.24), and $*$ is the two-dimensional convolution operator. The local similarity measure $\mathbf{HS}_{\mathbf{I}_1, \mathbf{I}_2}^{(k)}$ can be seen as an analog to $\mathbf{FS}_{\mathbf{I}_1, \mathbf{I}_2}$. However, $\mathbf{HS}_{\mathbf{I}_1, \mathbf{I}_2}^{(k)}$ does

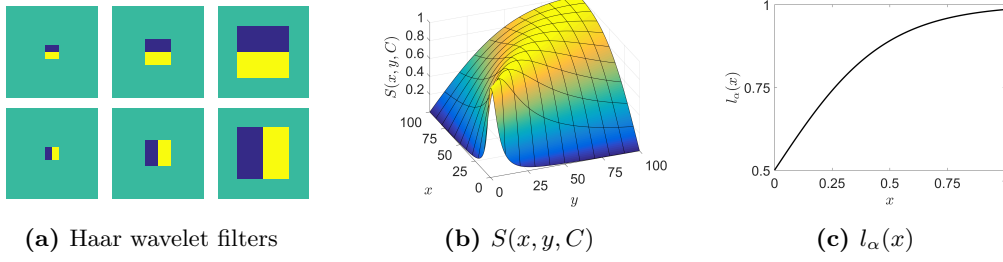


Figure 2.2: Main components of the HaarPSI. (a) The six Haar wavelet filters whose responses build the core of the HaarPSI. (b) The function $S(x, y, C)$ for $C = 30$. (c) The logistic function $l_\alpha(x)$ for $\alpha = 4.2$.

not mix different different concepts like gradients and phase congruency and is computed straightforwardly on the responses of two high-frequency discrete Haar wavelet filters. A visualization of the local similarity map $\mathbf{HS}_{\mathbf{I}_1, \mathbf{I}_2}^{(k)}$ is shown in Figure 2.3.

Analogous to the phase congruency map $\mathbf{PC}_{\mathbf{I}}$ in the definition of the FSIM, the HaarPSI considers a weight map which is derived from the response of a single low-frequency Haar wavelet filter:

$$\mathbf{W}_{\mathbf{I}}^{(k)}[\mathbf{m}] = \left| (g_3^{(k)} * \mathbf{I})[\mathbf{m}] \right|, \quad \mathbf{m} \in \mathbb{Z}^2, \quad (2.39)$$

where $k \in \{1, 2\}$ again differentiates between horizontal and vertical filters. Figure 2.3 shows an example of the weight map $\mathbf{W}_{\mathbf{I}}^{(k)}$ computed from a natural image.

The HaarPSI for two grayscale images $\mathbf{I}_1, \mathbf{I}_2$ is finally defined as the weighted average of the local similarity map $\mathbf{HS}_{\mathbf{I}_1, \mathbf{I}_2}^{(k)}$, that is,

$$\text{HaarPSI}_{\mathbf{I}_1, \mathbf{I}_2} = l_\alpha^{-1} \left(\frac{\sum_{\mathbf{m}} \sum_{k=1}^2 \mathbf{HS}_{\mathbf{I}_1, \mathbf{I}_2}^{(k)}[\mathbf{m}] \cdot \mathbf{W}_{\mathbf{I}_1, \mathbf{I}_2}^{(k)}[\mathbf{m}]}{\sum_{\mathbf{m}} \sum_{k=1}^2 \mathbf{W}_{\mathbf{I}_1, \mathbf{I}_2}^{(k)}[\mathbf{m}]} \right)^2, \quad (2.40)$$

with

$$\mathbf{W}_{\mathbf{I}_1, \mathbf{I}_2}^{(k)}[\mathbf{m}] = \max \left\{ \mathbf{W}_{\mathbf{I}_1}^{(k)}[\mathbf{m}], \mathbf{W}_{\mathbf{I}_2}^{(k)}[\mathbf{m}] \right\}, \quad \mathbf{m} \in \mathbb{Z}^2, \quad (2.41)$$

for $k \in \{1, 2\}$. The function $l_\alpha^{-1}(\cdot)$ maps the weighted average from the interval $[\frac{1}{2}, l_\alpha(1)]$ back to $[0, 1]$. Applying $(\cdot)^2$ further spreads the HaarPSI in the unit interval and helps to linearize the relationship between the HaarPSI and human opinion scores. In particular, this procedure aims to increase the readability of the HaarPSI in the sense that a single value should be meaningful on its own and not only relative to other HaarPSI values. Note that, due to the monotonicity of the logistic function, applying $l_\alpha^{-1}(\cdot)^2$ cannot improve or worsen the rank order-based correlations with human opinion scores reported in Section 2.3.

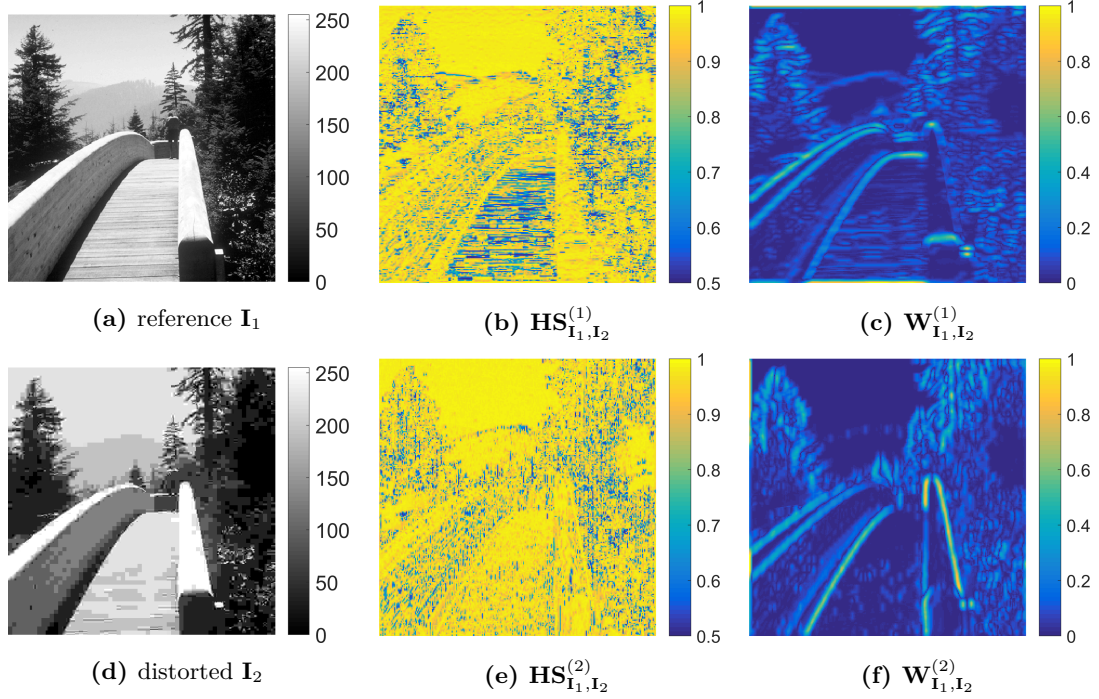


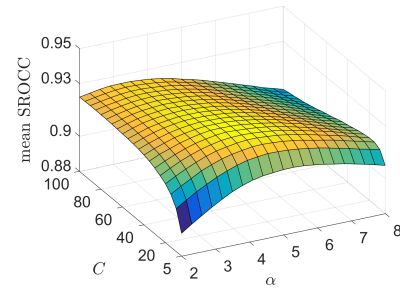
Figure 2.3: Examples for HaarPSI similarity and weight maps. (a) An undistorted reference image. (b) The reference image distorted by the JPEG compression algorithm. (c) The horizontal local similarity map $\mathbf{HS}_{\mathbf{I}_1, \mathbf{I}_2}^{(1)}$. (d) The vertical local similarity map $\mathbf{HS}_{\mathbf{I}_1, \mathbf{I}_2}^{(2)}$. (e) The (normalized) horizontal weight function $\mathbf{W}_{\mathbf{I}_1, \mathbf{I}_2}^{(1)}$. (f) The (normalized) vertical weight function $\mathbf{W}_{\mathbf{I}_1, \mathbf{I}_2}^{(2)}$. The images (a) and (b) are part of the CSIQ database [112].

Analogous to the FSIM, the HaarPSI can be extended to color images in the YIQ color space by considering a third local similarity map based on the chroma channels I and Q. The map $\mathbf{HS}_{\mathbf{I}_1, \mathbf{I}_2}^{(3)}$ is computed analogous to (2.38) by averaging local similarities obtained from comparing \mathbf{I}_1^I with \mathbf{I}_2^I and \mathbf{I}_1^Q with \mathbf{I}_2^Q . In contrast to $\mathbf{HS}_{\mathbf{I}_1, \mathbf{I}_2}^{(1)}$ and $\mathbf{HS}_{\mathbf{I}_1, \mathbf{I}_2}^{(2)}$, the chromatic information used for $\mathbf{HS}_{\mathbf{I}_1, \mathbf{I}_2}^{(3)}$ is not based on orientation sensitive filters. The corresponding weight map $\mathbf{W}_{\mathbf{I}_1^Y, \mathbf{I}_2^Y}^{(3)}$ is thus also computed by averaging $\mathbf{W}_{\mathbf{I}_1^Y, \mathbf{I}_2^Y}^{(1)}$ and $\mathbf{W}_{\mathbf{I}_1^Y, \mathbf{I}_2^Y}^{(2)}$. Formally, the generalization of the HaarPSI to color images is given by

$$\text{HaarPSIC}_{\mathbf{I}_1, \mathbf{I}_2} = l_\alpha^{-1} \left(\frac{\sum_{\mathbf{m}} \sum_{k=1}^3 \mathbf{HS}_{\mathbf{I}_1, \mathbf{I}_2}^{(k)}[\mathbf{m}] \cdot \mathbf{W}_{\mathbf{I}_1^Y, \mathbf{I}_2^Y}^{(k)}[\mathbf{m}]}{\sum_{\mathbf{m}} \sum_{k=1}^3 \mathbf{W}_{\mathbf{I}_1^Y, \mathbf{I}_2^Y}^{(k)}[\mathbf{m}]} \right)^2, \quad (2.42)$$

	All databases	TID 2008 & 2013	LIVE & CSIQ
<i>Parameters</i>			
C	30	30	20
α	4.2	4.2	5.8
<i>SROCC</i>			
LIVE	0.9683	0.9683	0.9677
TID 2008	0.9097	0.9097	0.9031
TID 2013	0.8732	0.8732	0.8651
CSIQ	0.9604	0.9604	0.9625

The highest correlation in each row is written in **bold-face**.



(a)

(b)

Figure 2.4: HaarPSI parameter optimization. (a) Values for the parameters C and α which maximize the mean SROCC with respect to randomly selected subsets of the considered databases. The values in the first column were obtained by including all four databases in the optimization procedure. For the results depicted in columns 2 and 3, the optimization was restricted to the TID 2008 & TID 2013 respectively the LIVE & CSIQ databases. The SROCC values shown in the last four rows are with respect to the full databases. (b) The mean SROCC with respect to the subsets of all four databases plotted as a function of the parameters C and α .

with $\mathbf{HS}_{\mathbf{I}_1, \mathbf{I}_2}^{(1)}$ and $\mathbf{HS}_{\mathbf{I}_1, \mathbf{I}_2}^{(2)}$ defined as in (2.38),

$$\mathbf{HS}_{\mathbf{I}_1, \mathbf{I}_2}^{(3)}[\mathbf{m}] = l_\alpha \left(\frac{1}{2} \left(S(|(\mathbf{f} * \mathbf{I}_1)[\mathbf{m}]|, |(\mathbf{f} * \mathbf{I}_2)[\mathbf{m}]|, C) + S(|(\mathbf{f} * \mathbf{I}_1^Q)[\mathbf{m}]|, |(\mathbf{f} * \mathbf{I}_2^Q)[\mathbf{m}]|, C) \right) \right), \quad (2.43)$$

with a 2×2 mean filter \mathbf{f} and

$$\mathbf{W}_{\mathbf{I}_1^Y, \mathbf{I}_2^Y}^{(3)}[\mathbf{m}] = \frac{1}{2} \left(\mathbf{W}_{\mathbf{I}_1^Y, \mathbf{I}_2^Y}^{(1)}[\mathbf{m}] + \mathbf{W}_{\mathbf{I}_1^Y, \mathbf{I}_2^Y}^{(2)}[\mathbf{m}] \right), \quad \mathbf{m} \in \mathbb{Z}^2. \quad (2.44)$$

2.2.1 Parameter Selection

The HaarPSI as well as the color-sensitive Haar wavelet-based perceptual similarity index (HaarPSIC) require only two parameters to be selected, namely C and α . Both parameters were optimized on randomly chosen subsets of four large publicly available databases, where each subset was a quarter the size of the original database. Each of the databases, which will be described in more detail in Section 2.3, contains large numbers of differently distorted images and their corresponding MOS values. The parameters C and α were selected to maximize the mean of the Spearman's rank-order correlation coefficient (SROCC) obtained from comparing HaarPSIC and MOS values from subsets of the Tampere image database (TID) 2008 [138], the TID 2013 [139], the Laboratory for Image & Video Engineering image quality assessment database (LIVE) [150] and the Laboratory of Computational and Subjective Image Quality image quality database (CSIQ) [112]. The optimization was carried out in two steps. First, a grid search was performed in which the parameter C took

values in the interval $[5, 100]$ and α in the range between 2 and 8. The best (C, α) -pair was then used as the initial value of the Nelder-Mead algorithm. The thus refined parameters were eventually rounded to the nearest integer in the case of C and to the nearest tenth in the case of α . This procedure resulted in the choices of $C = 30$ and $\alpha = 4.2$. To verify the generality of the HaarPSI, the same optimization procedure was repeated once only considering the TID 2008 and TID 2013 databases and once restricted to the LIVE and the CSIQ image databases. The results of all three optimizations are compiled in Figure 2.4.

2.3 Experimental Results

The consistency of the HaarPSI with the human perception of image quality was evaluated and compared with most of the image quality metrics discussed in Section 2.1 on four large publicly available benchmark databases of quality-annotated images. Those databases differ in the number of reference images, the number of distortion magnitudes and types, the number of observers, the level of control of the viewing conditions, and the stimulus presentation procedure.

The LIVE database [150] contains 29 reference color images and 779 distorted images that were perturbed by JPEG compression, JPEG 2000 compression, additive Gaussian white noise, Gaussian blurring as well as JPEG 2000 compressed images that have been transmitted over a simulated Rayleigh fading channel. Each distortion is introduced at five to six different levels of magnitude. On average, about 23 subjects evaluated the quality of each image with respect to the reference image. The viewing conditions were fairly controlled for in terms of viewing distance. Ratings were collected in a double stimulus manner.

The TID 2008 database [138] comprises 25 colored reference images and 1700 degraded images, that had been subject to a wide range of distortions, including various types of noise, blur, JPEG and JPEG 2000 compression, transmission errors, local image distortions, as well as luminance and contrast changes. Subjective ratings were gathered by comparisons. The results from several viewing conditions of experiments in three different labs and on the internet were averaged. TID 2008 was later extended to TID 2013 [139], which added new types of distortions, which are mostly of a chromatic nature. In total, TID 2013 contains 3000 differently distorted images.

The CSIQ database [112] is based on 30 reference color images and contains 866 distorted images. Six different types of distortions (JPEG compression, JPEG 2000 compression, global contrast decrements, additive pink Gaussian noise, and Gaussian blurring) at four to five different degradation magnitudes were applied to the reference images. The viewing distance was controlled. Images were presented on a monitor array and subjects were asked

to place all distorted versions of one reference image according to its perceived quality.

The main goal of most computational image similarity measures is to yield a monotonic relationship with human mean opinion scores across different databases and distortion types. To ensure a fair evaluation, different computational measures are typically compared with respect to rank order-based correlations or after performing nonlinear regression. Throughout the numerical evaluation of the HaarPSI, we apply the rank order-based SROCC to measure correlations between human mean opinion scores and different computational similarity and distortion indexes. We also considered applying Kendall's τ and the Pearson product-moment correlation after performing a four parameter logistic regression as alternatives for the SROCC. We found that these correlation coefficients essentially duplicate the results reported in this section. The corresponding versions of Tables 2.1 and 2.3 were thus not included here but can be found at www.haarpsi.org.

Following the ITU guidelines for evaluating quality prediction models [85], we also tested the statistical significance of the results reported in this section. Correlation coefficients for which the H_0 hypothesis that they are not significantly different than the respective HaarPSI correlation can be refuted with $p < 0.05$ are highlighted in color in Tables 2.1, 2.3 and 2.4. In accordance with [59], the variance of the z-transforms were approximated by $1.06/(N - 3)$, where N denotes the degrees of freedom (i.e., the number of samples in the considered database or distortion-specific subset).

The four databases used in the numerical evaluation only contain color images. However, out of the metrics considered in our experiments, only the FSIM and the HaarPSI are defined for both grayscale and color images, while the VSI was specifically designed for color images. All other similarity measures considered in our experiments only accept grayscale images as input or perform an RGB to grayscale conversion as a first processing step. To reflect these differing designs, all methods were tested on all databases once with the original color images and once with grayscale conversions obtained from the MATLAB *rgb2gray* function. To obtain the VSI for pairs of grayscale images, corresponding RGB images were created by setting the values for all three color channels to the values of the given grayscale channel. The correlation coefficients of all ten considered similarity measures with the human mean opinion scores for the LIVE image database, TID 2008, TID 2013 and the CSIQ database are compiled in Table 2.1.

Table 2.2 provides a quick impression of the overall performance of each metric. It depicts the average SROCC of each metric with respect to all four databases as well as the mean execution time in milliseconds. The average execution time was measured on a Intel Core i7-4790 CPU clocked at 3.60 GHz. To measure the execution time, each quality measure was computed ten times for ten different pairs of randomly generated 512×512 pixel images. All computations and measurements were carried out in MATLAB using

Table 2.1: Spearman rank order correlations of IQA metrics on benchmark databases.

	PSNR	VIF	SSIM	MS-SSIM	GSM	MAD	SR-SIM	FSIM	VSI	HaarPSI
<i>Grayscale Images</i>										
LIVE	0.8756	0.9636	0.9479	0.9513	0.9561	0.9672	0.9619	0.9634	0.9534	0.9690
TID 2008	0.5531	0.7491	0.7749	0.8542	0.8504	0.8340	0.8913	0.8804	0.8830	0.9043
TID 2013	0.6394	0.6769	0.7417	0.7859	0.7946	0.7807	0.8075	0.8022	0.8048	0.8094
CSIQ	0.8058	0.9195	0.8756	0.9133	0.9108	0.9466	0.9319	0.9242	0.9372	0.9546
<i>Color Images</i>										
LIVE	0.8756	0.9636	0.9479	0.9513	0.9561	0.9672	0.9619	0.9645	0.9524	0.9683
TID 2008	0.5531	0.7491	0.7749	0.8542	0.8504	0.8340	0.8913	0.8840	0.8979	0.9097
TID 2013	0.6394	0.6769	0.7417	0.7859	0.7946	0.7807	0.8075	0.8510	0.8965	0.8732
CSIQ	0.8058	0.9195	0.8756	0.9133	0.9108	0.9466	0.9319	0.9310	0.9423	0.9604

Lower and higher correlations than HaarPSI (statistically significant with $p < 0.05$).
The highest correlation in each row is written in **boldface**.

Table 2.2: Mean SROCC and average execution time.

	Color Images		Grayscale Images	
	Mean SROCC	Avg. time	Mean SROCC	Avg. time
HaarPSI	0.9279	24 ms	0.9093	10 ms
VSI	0.9223	79 ms	0.8946	80 ms
FSIM	0.9076	142 ms	0.8925	121 ms
SRSIM	0.8982	10 ms	0.8982	10 ms
MAD	0.8821	892 ms	0.8821	891 ms
GSM	0.8780	8 ms	0.8780	7 ms
MS-SSIM	0.8762	30 ms	0.8762	24 ms
SSIM	0.8350	6 ms	0.8350	5 ms
VIF	0.8273	459 ms	0.8273	453 ms
PSNR	0.7185	2 ms	0.7185	1 ms

implementations made freely available by the respective authors. Note that due to an additional conversion step, metrics that are only defined for grayscale images can have slightly higher execution times when evaluated on color images.

A high correlation with the mean opinion scores annotated to the distorted images of a large database containing many different types and degrees of distortions is arguably the best indicator of an image quality measure's consistency with human perception. However, for certain applications like compression or denoising, it could be more important to know if an image quality metric has a high correlation with the human experience *within* a single distortion class. Table 2.3 depicts the SROC coefficients for all image quality metrics when only subsets of databases containing specific distortions like Gaussian blur or JPEG transmission errors are considered.

Table 2.3: Distortion-specific SROCC for different IQA metrics.

	PSNR	VIF	SSIM	MS-SSIM	GSM	MAD	SR-SIM	FSIM	VSI	HaarPSI
<i>LIVE</i>										
jpg2k	0.8954	0.9696	0.9614	0.9627	0.9700	0.9692	0.9700	0.9724	0.9604	0.9684
jpg	0.8809	0.9846	0.9764	0.9815	0.9778	0.9786	0.9823	0.9840	0.9761	0.9832
gwn	0.9854	0.9858	0.9694	0.9733	0.9774	0.9873	0.9812	0.9716	0.9835	0.9845
gblur	0.7823	0.9728	0.9517	0.9542	0.9518	0.9510	0.9660	0.9708	0.9527	0.9676
ff	0.8907	0.9650	0.9556	0.9471	0.9402	0.9589	0.9466	0.9519	0.9430	0.9527
<i>TID 2008</i>										
gwn	0.9070	0.8797	0.8107	0.8086	0.8606	0.8386	0.8989	0.8758	0.9229	0.9177
gwn	0.8995	0.8757	0.8029	0.8054	0.8091	0.8255	0.8957	0.8931	0.9118	0.8982
scn	0.9170	0.8698	0.8145	0.8209	0.8941	0.8678	0.9084	0.8711	0.9296	0.9271
mn	0.8515	0.8683	0.7795	0.8107	0.7452	0.7336	0.7881	0.8264	0.7734	0.7909
hfn	0.9270	0.9075	0.8729	0.8694	0.8945	0.8864	0.9195	0.9156	0.9253	0.9155
in	0.8724	0.8327	0.6732	0.6907	0.7235	0.0650	0.7678	0.7719	0.8298	0.8269
qn	0.8696	0.7970	0.8531	0.8589	0.8800	0.8160	0.8348	0.8726	0.8731	0.8842
gblr	0.8697	0.9540	0.9544	0.9563	0.9600	0.9196	0.9551	0.9472	0.9529	0.9001
den	0.9416	0.9161	0.9530	0.9582	0.9725	0.9433	0.9666	0.9618	0.9693	0.9711
jpg	0.8717	0.9168	0.9252	0.9322	0.9393	0.9275	0.9393	0.9294	0.9616	0.9417
jpg2k	0.8132	0.9709	0.9625	0.9700	0.9758	0.9707	0.9809	0.9780	0.9848	0.9860
jpgt	0.7516	0.8585	0.8678	0.8681	0.8790	0.8661	0.8881	0.8756	0.9160	0.8921
jpg2kt	0.8309	0.8501	0.8577	0.8606	0.8936	0.8394	0.8902	0.8555	0.8942	0.8963
pn	0.5815	0.7619	0.7107	0.7377	0.7386	0.8287	0.7659	0.7514	0.7699	0.8010
bdist	0.6193	0.8324	0.8462	0.7546	0.8862	0.7970	0.7798	0.8464	0.6295	0.8026
ms	0.6957	0.5096	0.7231	0.7338	0.7190	0.5163	0.5704	0.6554	0.6714	0.6051
ctrst	0.5859	0.8188	0.5246	0.6381	0.6691	0.2723	0.6475	0.6510	0.6557	0.6209
<i>TID 2013</i>										
gwn	0.9291	0.8994	0.8671	0.8646	0.9064	0.8843	0.9251	0.9101	0.9460	0.9368
gwn	0.8981	0.8299	0.7726	0.7730	0.8175	0.8019	0.8562	0.8537	0.8705	0.8593
scn	0.9200	0.8835	0.8515	0.8544	0.9158	0.8911	0.9223	0.8900	0.9367	0.9311
mn	0.8323	0.8450	0.7767	0.8073	0.7293	0.7380	0.7855	0.8094	0.7697	0.7858
hfn	0.9140	0.8972	0.8634	0.8604	0.8869	0.8876	0.9131	0.9040	0.9200	0.9069
in	0.8968	0.8537	0.7503	0.7629	0.7965	0.2769	0.8280	0.8251	0.8741	0.8656
qn	0.8808	0.7854	0.8657	0.8706	0.8841	0.8514	0.8497	0.8807	0.8748	0.8893
gblr	0.9149	0.9650	0.9668	0.9673	0.9689	0.9319	0.9622	0.9551	0.9612	0.9149
den	0.9480	0.8911	0.9254	0.9268	0.9432	0.9252	0.9398	0.9330	0.9484	0.9456
jpg	0.9189	0.9192	0.9200	0.9265	0.9284	0.9217	0.9396	0.9339	0.9541	0.9512
jpg2k	0.8840	0.9516	0.9468	0.9504	0.9602	0.9511	0.9672	0.9589	0.9706	0.9704
jpgt	0.7685	0.8409	0.8493	0.8475	0.8512	0.8283	0.8543	0.8610	0.9216	0.8938
jpg2kt	0.8883	0.8761	0.8828	0.8889	0.9182	0.8788	0.9165	0.8919	0.9228	0.9204
pn	0.6863	0.7720	0.7821	0.7968	0.8130	0.8315	0.7967	0.7937	0.8060	0.8154
bdist	0.1552	0.5306	0.5720	0.4801	0.6418	0.2812	0.4722	0.5532	0.1713	0.4471
ms	0.7671	0.6276	0.7752	0.7906	0.7875	0.6450	0.6562	0.7487	0.7700	0.7152
ctrst	0.4400	0.8386	0.3775	0.4634	0.4857	0.1972	0.4696	0.4679	0.4754	0.4382
ccs	0.0766	0.3099	0.4141	0.4099	0.3578	0.0575	0.3117	0.8359	0.8100	0.6735
mgn	0.8905	0.8468	0.7803	0.7786	0.8348	0.8409	0.8781	0.8569	0.9117	0.8902
cn	0.8411	0.8946	0.8566	0.8528	0.9124	0.9064	0.9259	0.9135	0.9243	0.9275
lcni	0.9145	0.9204	0.9057	0.9068	0.9563	0.9443	0.9608	0.9485	0.9564	0.9622
icqd	0.9269	0.8414	0.8542	0.8555	0.8973	0.8745	0.8810	0.8815	0.8839	0.8953
cha	0.8872	0.8848	0.8775	0.8784	0.8823	0.8310	0.8758	0.8925	0.8906	0.8599
ssr	0.9042	0.9353	0.9461	0.9483	0.9668	0.9567	0.9613	0.9576	0.9628	0.9651
<i>CSIQ</i>										
gwn	0.9363	0.9575	0.8974	0.9471	0.9440	0.9541	0.9628	0.9359	0.9636	0.9666
jpeg	0.8881	0.9705	0.9546	0.9634	0.9632	0.9615	0.9671	0.9664	0.9618	0.9695
jpg2k	0.9362	0.9672	0.9606	0.9683	0.9648	0.9752	0.9773	0.9704	0.9694	0.9815
gpn	0.9339	0.9511	0.8922	0.9331	0.9387	0.9570	0.9520	0.9370	0.9638	0.9594
gblr	0.9291	0.9745	0.9609	0.9711	0.9589	0.9682	0.9767	0.9729	0.9679	0.9783
ctrst	0.8621	0.9345	0.7922	0.9526	0.9354	0.9207	0.9528	0.9438	0.9504	0.9450

Lower and higher correlations than HaarPSI (statistically significant with $p < 0.05$).
The highest correlation in each row is written in **boldface**.

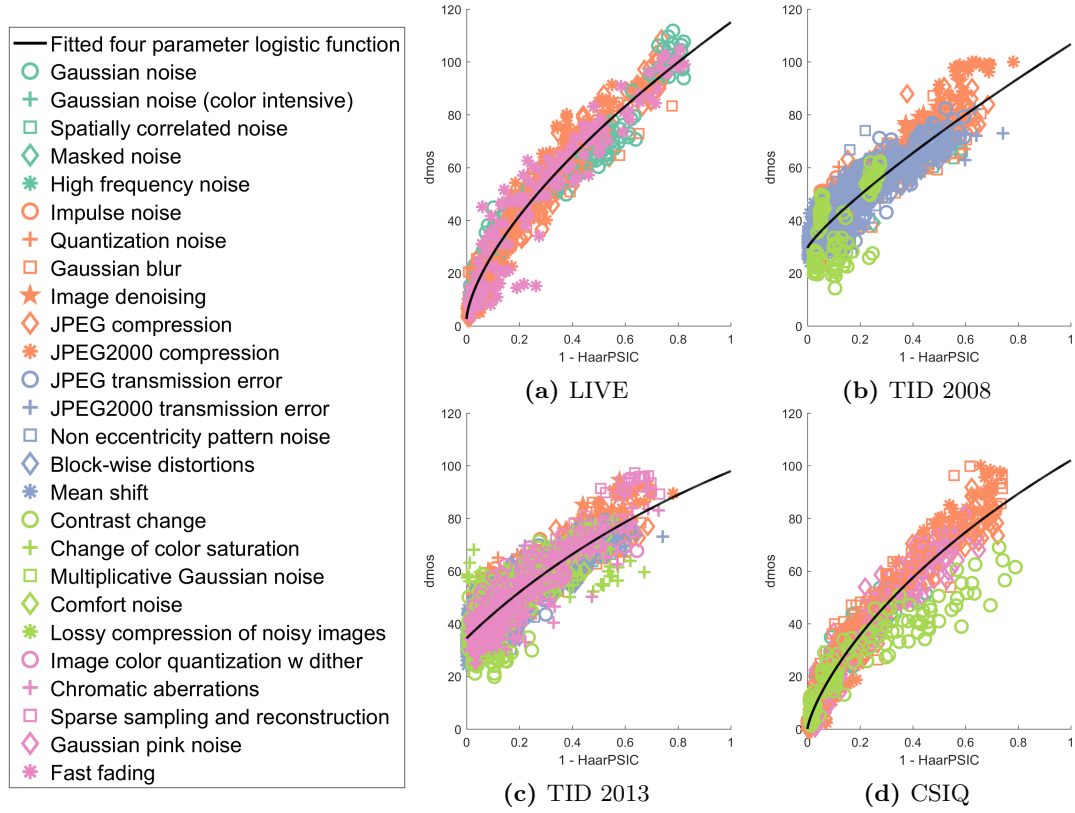


Figure 2.5: Scatter plots of HaarPSIC values against the difference mean opinion score (DMOS) from the LIVE, TID 2008, TID 2013 and CSIQ image databases.

Single correlation coefficients provide a useful means of objectively evaluating and comparing different computational models of image quality. However, they only measure a specific aspect of the relationship between an image similarity metric and human opinion scores, like linearity in the case of the Pearson correlation coefficient or monotonicity in the case of the SROCC. In an attempt to better visualize the relationship between the HaarPSI and human opinion scores, Figure 2.5 shows scatter plots of the HaarPSI against the DMOS for all four databases. To provide as much insight as possible, the plots are categorized by specific distortion types.

It should be noted that for all results reported in this section, the HaarPSI, as well as other image quality metrics such as the SSIM, the FSIM or the VSI, were preprocessing each image by convolving it with a 2×2 mean filter as well as a subsequent dyadic subsampling step. This preprocessing approximates the low-pass characteristics of the optical part of the human visual system [134] by a simple model.

2.4 Conclusion

The HaarPSI is a novel and computationally inexpensive image quality measure based solely on the coefficients of three stages of a discrete Haar wavelet transform. Its validity with respect to the human perception of image quality was tested on four large databases containing more than 5000 differently distorted images, with very promising results. In a comparison with 9 popular state of the art image similarity metrics, the HaarPSI yields significantly higher or statistically indistinguishable Spearman correlations when restricted to grayscale conversions. For color images, it only comes second to the VSI when tested on the TID 2013 (see Table 2.1). Along with its simple computational structure and its comparatively short execution time, this suggests a high applicability of the HaarPSI in real world optimization tasks. In particular, image quality metrics like PSNR, SSIM, or SR-SIM, that outperform the HaarPSI with respect to speed achieve considerably inferior correlations with human opinion scores (see Table 2.2). Regarding the applicability of the HaarPSI in specific optimization tasks, we would like to mention that the HaarPSI has consistently high correlations with human opinion scores throughout all databases with respect to distortions caused by the JPEG and JPEG 2000 compression algorithms (see Table 2.3).

The results reported in Tables 2.1 and 2.3 might seem contradictory at first glance. In many cases, the HaarPSI yields the highest SROCC for a complete database but is outperformed by other metrics like the VSI when restricting the same database to a single distortion type. However, taking into account statistical significance, it is apparent that only when tested on the TID databases restricted to Gaussian blur, the performance of the HaarPSI is consistently lower than the performance of other similarity metrics. This particular shortcoming can be explained by the fact that the HaarPSI is almost exclusively relying on high-frequency information and thus maybe too sensitive in the case of distortions purely based on low-pass filtering.

When only considering a specific type of distortion, the correlations yielded by the HaarPSI might be improved by tuning the constants C and α , which have originally been selected to optimize the overall performance. Increasing C decreases the sensitivity of the HaarPSI to changes in the high-frequency components measured by the similarity maps $HS_{I_1, I_2}^{(1,2)}$ relative to the weights $W_f^{(1,2)}$, which are based on a lower frequency band and serve as a rough model of attention-like processes. The effect of the parameter α on the HaarPSI is qualitatively similar when it is approaching zero. This could explain the roughly negative linear relationship between C and α in Figure 2.4. However, for larger choices of α , the function $l_\alpha(\cdot)$ is increasingly mimicking the behavior of a thresholding operator in the sense that only severe changes in the high-frequency components will have a significant effect

on the HaarPSI. To also provide a quantitative analysis of these relationships, Figure 2.6 depicts the influence of C and α on the correlation with human opinion scores in the case of TID 2013 with respect to six different distortions. Figure 2.6(c) indeed suggests that in the case of Gaussian blur, the performance of the HaarPSI can be improved by attenuating its sensitivity to changes in the high-frequency components via increasing C and choosing α close to 0. In contrast, Figure 2.6(a) indicates that the HaarPSI achieves the highest correlations in the case of JPEG compression artifacts when it is tuned to be sensitive to severe changes in the high frequency components at highly salient locations.

It is surprising that the extremely simple computational model of orientation and spatial frequency selectivity used in the HaarPSI suffices to obtain comparatively high correlations with human opinion scores. Additionally, these correlations are stable with respect to a wide range of parameters C and α (cf. Figure 2.4). This could indicate that the computational structure of the HaarPSI succeeds at reproducing the *functional essence* of at least some parts of the human visual system. It is, however, quite likely that the HaarPSI owes some of its experimental success to the limitations of the used benchmark databases, which only consider a limited number of reference images and specific types of distortions. Certainly, orientation selectivity in the primary visual cortex is not restricted to horizontal and vertical edges.

Another computational principle that plays an important role in natural neural systems and that was recently successfully applied in the context of perceptual image similarity measurement is *divisive normalization* [111]. While the similarity measure $S(a, b, C)$ introduces some kind of normalization, divisive normalization is not included in any of the computational stages of the HaarPSI. It remains an open question if and how the HaarPSI could be further improved by incorporating divisive normalization in a similar fashion as the concepts of orientation selectivity and spatial frequency selectivity.

Many practical applications demand image similarity metrics to yield values that are easy to interpret. Ideally, an image similarity of 0.9 would in fact indicate that the average human would also assess a similarity of 90 % between two images or that a decrease in similarity to 0.8 corresponds to a 10 % decrease in perceived quality for a human viewer. Due to the generality and difficulty of this task, computational models of image similarity typically only aim at establishing a monotonic relationship with human mean opinion scores, which is also reflected in the choice of the SROCC as a measure of consistency. In the case of the HaarPSI, applying $l_\alpha^{-1}(\cdot)^2$ to the final similarity score significantly linearizes its relationship with human opinion scores, thereby leading to the strong linear correlations depicted in the scatter plots in Figure 2.5. While $l_\alpha^{-1}(\cdot)^2$ is monotonically increasing on $[\frac{1}{2}, 1)$ and therefore not affecting the SROCC, we hope that this improves the readability and applicability of the HaarPSI. To also provide an objective measure of linear correlation,

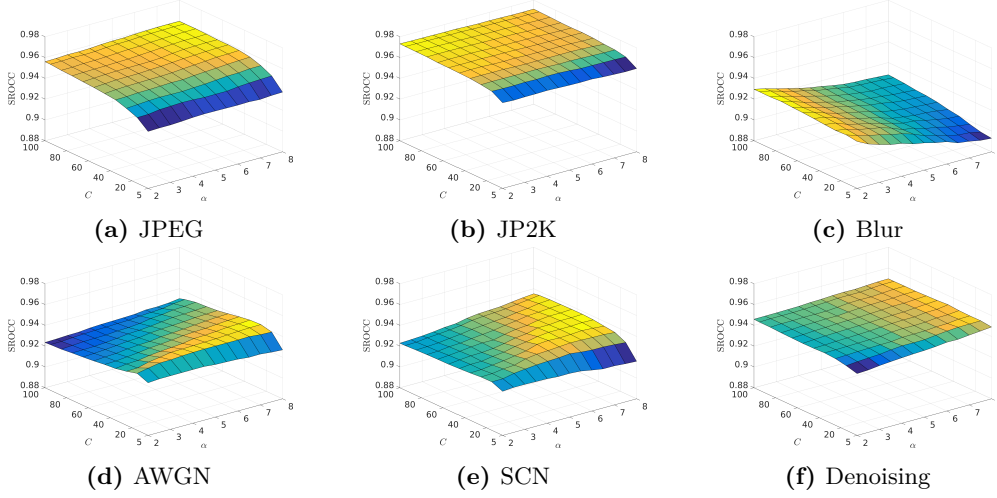


Figure 2.6: Spearman rank order correlations as functions of the parameters C and α for images affected by (a) JPEG compression, (b) JP2K compression, (c) Gaussian Blur, (d) additive Gaussian white noise, (e) spatially correlated noise white noise, and (f) denoising. All correlations are with respect to TID 2013.

we repeated the numerical evaluation from Section 2.3 with the Pearson product-moment correlation instead of the SROCC (see Tables A.1 and A.2 in the appendix). The results of this analysis indicate that even without additional nonlinear regression, the HaarPSI has a highly linear relationship with human mean opinion scores from different databases and across varying types of distortion.

The HaarPSI can conceptually be understood as a simplified version of the FSIM. Both metrics rely on the construction of two maps, where one map measures local similarities between a reference image and a distorted image and the other map assesses the relative importance of image areas. However, in the HaarPSI, these maps are defined only in terms of a single Haar wavelet filterbank, while the FSIM utilizes an implementation of the phase congruency measure that requires the input images to be convolved with 16 complex-valued filters and contains several non-trivial computational steps, like adaptive thresholding. Another difference is that the FSIM uses the phase congruency measure both as a weight function in (2.26) and as a part of the local similarity measure (2.25). In the HaarPSI, the weight function (2.39) and the local similarity measure (2.38) are strictly separated in the sense that they are based on distinct bands of the frequency spectrum.

These conceptual simplifications lead to a significant decrease in execution time (see Table 2.2) and enable a better understanding of how single elements of the measure and properties of the input images contribute to the final similarity score. In the case of the HaarPSI, it is clear that the local similarity measure is based on high-frequency information, while the weight map, which provides a crude measure of visual saliency, is using filters

Table 2.4: SROCC on benchmark databases for different choices of wavelet filters.

	Daub2PSI	Daub4PSI	Sym4PSI	CDFPSI	Coif1PSI	HaarPSI
<i>Grayscale images</i>						
LIVE	0.9620	0.9530	0.9552	0.9604	0.9603	0.9690
TID 2008	0.8971	0.8796	0.8915	0.8836	0.8965	0.9043
TID 2013	0.8064	0.7982	0.8022	0.7965	0.8055	0.8094
CSIQ	0.9492	0.9442	0.9454	0.9404	0.9485	0.9546
<i>Color images</i>						
LIVE	0.9659	0.9610	0.9630	0.9675	0.9644	0.9683
TID 2008	0.8992	0.8804	0.8950	0.8932	0.8986	0.9097
TID 2013	0.8724	0.8643	0.8696	0.8633	0.8716	0.8732
CSIQ	0.9603	0.9577	0.9592	0.9596	0.9593	0.9604

Lower and higher correlations than HaarPSI (statistically significant with $p < 0.05$). The highest correlation in each row is written in **boldface**.

that are tuned to lower frequencies. We suspect that a similar principle plays an important role in the FSIM, where additional high-frequency filters are applied to obtain the gradient map used in the local similarity measure (2.25). However, for the FSIM, it is difficult to verify this, as filters that are tuned to lower frequencies are only implicitly used in the computation of the phase congruency measure, which is in turn part of both the local similarity measure and the weight map.

We do not have a straightforward explanation as to why the HaarPSI outperforms the FSIM with respect to correlations with human opinion scores (see Table 2.1). After all, both measures have a similar overall structure and implement similar principles such as frequency and orientation selectivity. We assume that the reduced complexity of the HaarPSI also limits uncontrollable side effects when accentuating different aspects of the input images by varying the parameters C and α . This could improve the chance of successfully fitting subsets of benchmark databases when only considering two free parameters, but also decrease the generalization error. Furthermore, the principle of orientation selectivity is implemented differently in the HaarPSI in the sense that measurements regarding horizontal and vertical structures are only combined at the very end, that is, when taking the weighted average. It is well known that orientation selectivity is a strong organization principle in the primary visual cortex, where neurons that are tuned to similar orientations are grouped together in so-called orientation columns [82]. It thus seems reasonable that a consistent separation of the information yielded by vertical and horizontal filters has a positive effect on the correlations with human opinion scores. From a computational point of view, it is very beneficial to apply discrete Haar wavelet filters instead of other wavelet filters. However, by changing \mathbf{h}_1^{1D} and \mathbf{g}_1^{1D} in (2.32) to the respective filters, the measure given in

(2.40) can easily be defined for other wavelets. Table 2.4 depicts the performance of such measures based on selected Daubechies wavelets [43], symlets [45], coiflets [44] and the Cohen-Daubechies-Feauveau wavelet [36] with respect to the four databases considered in Section 2.3. It is interesting to see that Haar filters not only seem to be the computationally most efficient but also the qualitatively best choice for the measure (2.40).

CHAPTER 3

Visual Servoing

The contents of this chapter are based on:

- DUFLOT, L.-A., REISENHOFER, R., TAMADAZTE, B., ANDREFF, N., and KRUPA, A.: “Wavelet and shearlet-based image representations for visual servoing”. To appear in *Int. J. Robot. Res.*
-

Prelude Wavelet- and Shearlet-Based Transforms

The theory of wavelets is certainly one of the biggest success stories in the field of applied harmonic analysis. Broadly speaking, a wavelet is an oscillating but localized function whose shifts and dilates can be used to construct systems of functions that are well suited to represent and analyze arbitrary functions in $L^2(\mathbb{R})$. One of the most significant advantages of wavelet-based transforms with respect to classical Fourier analysis is that, due to being localized in the time domain, wavelets not only yield information about which frequencies are present in a signal but can also identify when they are occurring in the time domain. What sets wavelets apart from other approaches in time-frequency analysis, such as the short-time Fourier transform, is that instead of considering fixed windows that provide localization in the time domain and using the modulation operator as a means of changing the frequency of a generating function, the construction of systems of wavelets is based on dilation. By squeezing or stretching a function, the dilation operator changes the frequencies but also the localization in the time domain (see, e.g., Figure 2.1(a)). As a result, wavelet-based systems are adaptive in the sense that high-frequency wavelets have a higher degree of localization in the time domain than their low-frequency counterparts.

The first definition of a transform that can be considered a wavelet transform was already given in 1910 by Alfred Haar [76] (see prelude to Chapter 2). However, it was the groundbreaking work of pioneers like Ingrid Daubechies [43], Stéphane Mallat [120] and Yves Meyer [128] in the late 1980’s and early 1990’s, that made discrete wavelet transforms applicable in many areas of engineering and applied mathematics and led to the development of a new generation of multiscale signal and image representations. A

comprehensive introduction to the theory of wavelets can be found in textbooks such as [45] or [121], which have also been the main source for the first part of this prelude.

Formally, a function $\psi \in L^2(\mathbb{R})$ is called a *wavelet* if it satisfies the wavelet admissibility condition

$$\int_{\mathbb{R}} \left| \widehat{\psi}(\xi) \right|^2 |\xi|^{-1} d\xi < \infty, \quad (3.1)$$

and we denote shifts and dilates of a generating wavelet ψ by

$$\psi_{a,b} = |a|^{-\frac{1}{2}} \psi \left(\frac{\cdot - b}{a} \right), \quad a, b \in \mathbb{R}, \quad a \neq 0, \quad (3.2)$$

where a is a *dilation parameter* and b is a *translation parameter*. The wavelet admissibility condition (3.1) is a necessary condition for the existence of an inversion formula of the continuous wavelet transform of a function $f \in L^2(\mathbb{R})$ associated with ψ that is defined as

$$(C\mathcal{W}_\psi f)(a, b) = \langle f, \psi_{a,b} \rangle_{L^2} = |a|^{-\frac{1}{2}} \int_{\mathbb{R}} f(x) \overline{\psi \left(\frac{x-b}{a} \right)} dx, \quad a, b \in \mathbb{R}, \quad a \neq 0. \quad (3.3)$$

Note that if $\psi \in L^1(\mathbb{R})$, which implies that $\widehat{\psi}$ is continuous, (3.1) can only be satisfied if ψ has a vanishing mean value, that is if

$$\widehat{\psi}(0) = \int_{\mathbb{R}} \psi(x) dx = 0. \quad (3.4)$$

The admissibility condition hence enforces wavelets to exhibit some kind of oscillatory behavior, which can also be observed in the examples of wavelets plotted in Figures 2.1(a), 3.1(a) and 3.1(c). A first example of a function that satisfies (3.1) and (3.4), the Haar wavelet ψ^{haar} , was already discussed in the prelude to Chapter 2. In particular, it was shown in Theorem 2.0.2 that the set Ψ^{haar} , which consists of shifts and dilates of ψ^{haar} , is an ONB for $L^2(\mathbb{R})$. In the proof of Theorem 2.0.2, we explicitly constructed linear combinations of Haar wavelets that can approximate any given function in $L^2(\mathbb{R})$ arbitrarily well by representing the difference of increasingly coarse step functions. This approach implicitly defined a nested structure of approximation spaces which were defined as the closure of the linear span of the shifts of dilated indicator functions $\mathbb{1}_{[0,1)}(2^{-j}\cdot)$. This specific structure is in fact an instance of a so-called multiresolution analysis (MRA) [120], which is a powerful framework that enables us to define a wide class of wavelet-based ONBs for $L^2(\mathbb{R})$.

Definition 3.0.1 (MRA, as in [45]). *A sequence $\{V_j\}_{j \in \mathbb{Z}}$ of closed subspaces in $L^2(\mathbb{R})$ that*

are nested in the sense that

$$\cdots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \cdots, \quad (3.5)$$

constitutes a multiresolution analysis (MRA) if the following properties are satisfied:

$$\forall j \in \mathbb{Z}: \quad f \in V_j \Leftrightarrow f(2^j \cdot) \in V_0, \quad (3.6)$$

$$\forall m \in \mathbb{Z}: \quad f \in V_0 \Leftrightarrow f(\cdot - m) \in V_0, \quad (3.7)$$

$$\overline{\bigcup_{j=-\infty}^{\infty} V_j} = L^2(\mathbb{R}), \quad (3.8)$$

$$\bigcap_{j=-\infty}^{\infty} V_j = \{0\}, \quad (3.9)$$

and if there exists a scaling function $\varphi \in L^2(\mathbb{R})$ such that the sequence $\{\varphi(\cdot - m)\}_{m \in \mathbb{Z}}$ is an ONB of V_0 .

In the case of Haar wavelets, the approximation space V_0 is defined as the closure of the linear span of the integer shifts of the indicator function $\mathbb{1}_{[0,1)}$, which thus assumes the role of the scaling function φ in Definition 3.0.1. In the proof of Theorem 2.0.2, we already saw that the Haar wavelets $\psi_{j,m}^{\text{haar}}$ can be used to encode the difference between two consecutive approximation spaces V_j and V_{j-1} . To construct an ONB for $L^2(\mathbb{R})$ from an arbitrary MRA, we now consider the so-called *wavelet spaces* W_j , which are defined for $j \in \mathbb{Z}$ as the orthogonal complement of V_j in V_{j-1} , that is,

$$W_j = V_j^\perp \subset V_{j-1}, \quad j \in \mathbb{Z}. \quad (3.10)$$

Due to $V_{j-1} = W_j \oplus V_j$ for all $j \in \mathbb{Z}$ and the nested structure of the approximation spaces (see (3.5)), this implies that

$$W_{j_1} \perp W_{j_2}, \quad \text{if } j_1 \neq j_2, \quad (3.11)$$

for all $j_1, j_2 \in \mathbb{Z}$. Furthermore, the MRA properties (3.8) and (3.9) imply that

$$\bigoplus_{j \in \mathbb{Z}} W_j = L^2(\mathbb{R}). \quad (3.12)$$

In summary, the sequence of wavelet spaces $\{W_j\}_{j \in \mathbb{Z}}$ consists of mutually orthogonal spaces whose infinite direct sum equals $L^2(\mathbb{R})$. The following theorem shows that it is indeed possible to define a generating wavelet ψ based on the scaling function of an MRA that

can be used to construct orthonormal bases for the spaces W_j , as well as $L^2(\mathbb{R})$.

Theorem 3.0.2 (Wavelet ONB for $L^2(\mathbb{R})$, [45, 121]). *Let $\{V_j\}_{j \in \mathbb{Z}}$ be an MRA (see Definition 3.0.1) and $\varphi \in L^2(\mathbb{R})$ be the associated scaling function. Furthermore, let us consider the sequence $\mathbf{h} = \{h_n = \sqrt{2} \langle \varphi(2 \cdot -n), \varphi \rangle_{L^2}\}_{n \in \mathbb{Z}}$, that is,*

$$\varphi = \sqrt{2} \sum_{n \in \mathbb{Z}} h_n \varphi(2 \cdot -n), \quad (3.13)$$

and the trigonometric polynomial

$$m_0(\xi) = 2^{-\frac{1}{2}} \sum_{n \in \mathbb{Z}} h_n e^{-2\pi i n \xi}. \quad (3.14)$$

Then, the function

$$\widehat{\psi}(\xi) = e^{\pi i \xi} \overline{m_0(\xi/2 + 1/2)} \widehat{\varphi}(\xi/2) \quad (3.15)$$

is the Fourier transform of a wavelet ψ such that for each $j \in \mathbb{Z}$, the sequence $\{\psi_{j,m} = 2^{-\frac{j}{2}} \psi(2^{-j} \cdot -m)\}_{m \in \mathbb{Z}}$ is an ONB for the space W_j (cf. (3.10)), and the family $\{\psi_{j,m}\}_{j,m \in \mathbb{Z}}$ is an ONB for $L^2(\mathbb{R})$.

It should be noted that not every wavelet ONB for $L^2(\mathbb{R})$ necessarily constitutes an MRA. For instance, the so-called Journé wavelet (see, e.g., [45]), is a somewhat pathological example of a mother wavelet ψ , for which the set $\{\psi_{j,m}\}_{j,m \in \mathbb{Z}}$ is an ONB but which is not associated with an MRA.

Equation (3.13) is also known as the *scaling relation* and illustrates the special connection between the scaling function φ and the *scaling filter* \mathbf{h} . Note that by inverting the Fourier transform, the definition of the *mother wavelet* ψ in (3.15) can also be written as

$$\psi = \sqrt{2} \sum_{n \in \mathbb{Z}} g_n \varphi(2 \cdot -n), \quad (3.16)$$

where the so-called *wavelet filter* is defined as the sequence $\mathbf{g} = \{g_n = (-1)^{n-1} \overline{h_{-n-1}}\}_{n \in \mathbb{Z}}$. Two examples of wavelets and scaling functions that are associated with an MRA are plotted along with their respective filters \mathbf{g} and \mathbf{h} in Figure 3.1. While Equation (3.15) is not the only way of defining orthonormal wavelet bases for the spaces W_j , the filter \mathbf{g} can always be obtained by considering the sequence $\{g_n = \sqrt{2} \langle \varphi(2 \cdot -n), \psi \rangle_{L^2}\}_{n \in \mathbb{Z}}$ and we denote the associated trigonometric polynomial with

$$m_1(\xi) = 2^{-\frac{1}{2}} \sum_{n \in \mathbb{Z}} g_n e^{-2\pi i n \xi}. \quad (3.17)$$

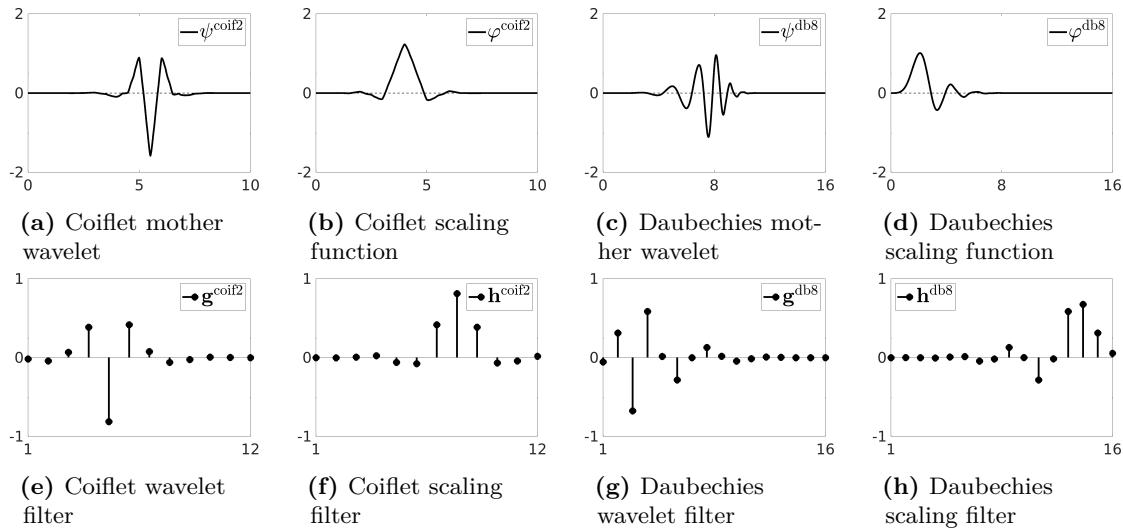


Figure 3.1: Examples of mother wavelets, scaling functions and their associated discrete filters.

The question remains, however, how a scaling function φ that defines an MRA in the sense of Definition 3.0.1 can be obtained. Applying the Fourier transform to both sides of the scaling relation (3.13) yields

$$\widehat{\varphi}(\xi) = m_0(\xi/2)\widehat{\varphi}(\xi/2) = m_0(\xi/2)m_0(\xi/4)\widehat{\varphi}(\xi/4) = \cdots, \quad (3.18)$$

and similarly in the case of (3.16),

$$\widehat{\psi}(\xi) = m_1(\xi/2)\widehat{\varphi}(\xi/2). \quad (3.19)$$

Assuming that $\widehat{\varphi}$ is continuous in 0 and ψ normalized in the sense that $\widehat{\psi}(0) = \int_{\mathbb{R}} \psi(x) dx = 1$, it follows from (3.18) that

$$\widehat{\varphi}(\xi) = \prod_{j=1}^{\infty} m_0(2^{-j}\xi). \quad (3.20)$$

This means that, under certain conditions, the trigonometric polynomial m_0 defined by a sequence \mathbf{h} via (3.14) can in fact be used to construct a scaling function φ that in turn defines an MRA. Sufficient conditions for the sequence \mathbf{h} such that the scaling function (3.20) defines an MRA can, for instance, be found in [45] or [121].

It was already observed in the prelude to Chapter 2 that by interpreting the entries of a discrete-time signal in $\ell^2(\mathbb{Z})$ as the coefficients of a linear combination of shifts of a dilate of the indicator function $\mathbb{1}_{[0,1]}$, the corresponding Haar wavelet coefficients can directly be computed via discrete convolutions with the filters \mathbf{h} and \mathbf{g} (cf. Theorem 2.0.2).

This relationship is not exclusive to Haar wavelets but can be defined for general MRAs. Analogous to Equation (2.2) in the prelude to Chapter 2, let us now consider dyadic dilates and shifts of a mother wavelet ψ of an MRA, that is,

$$\psi_{j,m} = 2^{-j/2} \psi(2^{-j} \cdot -m), \quad j, m \in \mathbb{Z}, \quad (3.21)$$

and the associated discrete wavelet transform of a function $f \in L^2(\mathbb{R})$, defined as

$$(\mathcal{W}_\psi f)(j, m) = \langle f, \psi_{j,m} \rangle_{L^2} = 2^{-j/2} \int_{\mathbb{R}} f(x) \overline{\psi(2^{-j} \cdot -m)} dx, \quad j, m \in \mathbb{Z}. \quad (3.22)$$

Note that here, the term *discrete* refers to the discrete sampling of the scale parameter j and the shift parameter m . In this setting, the family $\{\psi_{j,m}\}_{j,m \in \mathbb{Z}}$ also constitutes an ONB for $L^2(\mathbb{R})$ (cf. Theorem 3.0.2). Using Equation (3.16), it follows that

$$(\mathcal{W}_\psi f)(1, m) = \langle f, \psi_{1,m} \rangle_{L^2} = \left\langle f, \sum_{n \in \mathbb{Z}} g_n \varphi(\cdot - 2m - n) \right\rangle_{L^2} = \sum_{n \in \mathbb{Z}} \overline{g_{n-2m}} \langle f, \varphi_{0,n} \rangle_{L^2}, \quad (3.23)$$

or, more general,

$$(\mathcal{W}_\psi f)(j, m) = \sum_{n \in \mathbb{Z}} \overline{g_{n-2m}} \langle f, \varphi_{j-1,n} \rangle_{L^2}, \quad j, m \in \mathbb{Z}. \quad (3.24)$$

Similarly, it holds for the scaling function φ that

$$\langle f, \varphi_{j,m} \rangle_{L^2} = \sum_{n \in \mathbb{Z}} \overline{h_{n-2m}} \langle f, \varphi_{j-1,n} \rangle_{L^2}, \quad j, m \in \mathbb{Z}. \quad (3.25)$$

Analogous to the proof of Theorem 2.0.2, we now assume that for an arbitrarily large positive integer $J \in \mathbb{N}$ and a finite sequence $\mathbf{c}_{-J} \in \ell^2(\mathbb{Z})$, a given function $f \in L^2(\mathbb{R})$ can be expressed as a linear combination of shifts of the dilated scaling function $\varphi_{-J,0}$:

$$f = \sum_{n \in \mathbb{Z}} c_{-J,n} \varphi_{-J,n} = \sum_{n \in \mathbb{Z}} \langle f, \varphi_{-J,n} \rangle_{L^2} \varphi_{-J,n}. \quad (3.26)$$

Here, the sequence \mathbf{c}_{-J} can be interpreted as a finite and discrete-time representation of the function f . The coefficients of the discrete wavelet transform (3.22) of f can now directly be computed for all scales $j > -J$ by iteratively convolving the sequence \mathbf{c}_{-J} with the associated scaling filter \mathbf{h} and the wavelet filter \mathbf{g} , yielding a faithful transition between

the continuous setting and the discrete-time world:

$$\mathbf{d}_j = (\mathbf{c}_{j-1} * \tilde{\mathbf{g}})_{\downarrow 2}, \quad \text{and} \quad \mathbf{c}_j = (\mathbf{c}_{j-1} * \tilde{\mathbf{h}})_{\downarrow 2}, \quad (3.27)$$

where $\tilde{g}_n = \overline{g_{-n}}$ and $\tilde{h}_n = \overline{h_{-n}}$. As in the prelude to Chapter 2, the sequences \mathbf{d}_j contain the so-called detail coefficients while the entries of the sequences \mathbf{c}_j are denoted as approximation coefficients. Note that omitting the dyadic subsampling operator $\downarrow 2$ in Equation 3.27 corresponds to a wavelet transform where the translation parameters are sampled with a higher density than in (3.22). Such transforms are often denoted as stationary or non-subsampled wavelet transforms. This type of transform was already used in the definition of the HaarPSI in Chapter 2 (see Equation 2.40) and will also play an important role in the definition of wavelet- and shearlet-based visual servoing schemes later in this chapter.

So far in this prelude, we have only considered wavelet transforms of one-dimensional signals. All of the three applications discussed in this thesis, however, are concerned with the analysis of images and thus require a two-dimensional approach. A straightforward way of defining two-dimensional wavelet transforms is to consider two-dimensional wavelet generators that are based on tensor products of a one-dimensional mother wavelet and the associated one-dimensional scaling functions. For a scaling function $\varphi \in L^2(\mathbb{R})$ and a mother wavelet $\psi \in L^2(\mathbb{R})$, we consider separable two-dimensional wavelet generators

$$\psi^{(1)}(\mathbf{x}) = (\varphi \otimes \psi)(\mathbf{x}) = \varphi(x_1)\psi(x_2), \quad (3.28)$$

$$\psi^{(2)}(\mathbf{x}) = (\psi \otimes \varphi)(\mathbf{x}) = \psi(x_1)\varphi(x_2), \quad (3.29)$$

$$\psi^{(3)}(\mathbf{x}) = (\psi \otimes \psi)(\mathbf{x}) = \psi(x_1)\psi(x_2), \quad (3.30)$$

and the two-dimensional scaling function

$$\varphi^{(1)}(\mathbf{x}) = (\varphi \otimes \varphi)(\mathbf{x}) = \varphi(x_1)\varphi(x_2). \quad (3.31)$$

Analogous to the one-dimensional setting, we apply the dyadic scaling matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad (3.32)$$

and define shifts and L^2 -normalized dilates of a two-dimensional wavelet generator by

$$\psi_{j,\mathbf{m}} = 2^j \psi(\mathbf{A}^j \cdot -\mathbf{m}), \quad j \in \mathbb{Z}, \mathbf{m} \in \mathbb{Z}^2. \quad (3.33)$$

Note that by writing \mathbf{A}^j instead of \mathbf{A}^{-j} , the usage of the dilation parameter j is slightly different than in the one-dimensional case (cf. Equation 3.21). Here, the high-frequency wavelets are associated with parameters $j \gg 0$, while low-frequency wavelets are obtained by choosing $j \ll 0$. This convention will be maintained throughout the remainder of this thesis and was also used in the definition of the HaarPSI in Chapter 2. For a function $f \in L^2(\mathbb{R}^2)$, the discrete wavelet transform with respect to a wavelet generator $\psi \in L^2(\mathbb{R}^2)$ can now be defined as

$$(\mathcal{W}_\psi f)(j, \mathbf{m}) = \langle f, \psi_{j, \mathbf{m}} \rangle_{L^2} = 2^j \int_{\mathbb{R}^2} f(\mathbf{x}) \overline{\psi(\mathbf{A}^j \mathbf{x} - \mathbf{m})} \, d\mathbf{x}, \quad j \in \mathbb{Z}, \mathbf{m} \in \mathbb{Z}^2. \quad (3.34)$$

Analogous to (3.26), by interpreting a given discrete-time image $\mathbf{I} \in \ell^2(\mathbb{Z}^2)$ as an approximation of a function $f \in L^2(\mathbb{R}^2)$ in the sense that

$$f \approx 2^J \sum_{\mathbf{m} \in \mathbb{Z}^2} I[\mathbf{m}] \varphi^{(1)}(\mathbf{A}^J \cdot -\mathbf{m}), \quad (3.35)$$

for a fixed scale $J \in \mathbb{N}$, the MRA framework facilitates the faithful computation of wavelet coefficients $(\mathcal{W}_\psi f)(j, \mathbf{m})$ in terms of convolutions of \mathbf{I} with the associated scaling and wavelet filters. Figure 3.2 depicts the first and second stage of a two-dimensional wavelet transform of a digital image. Both representations shown in Figure 3.2 contain a coarse approximation of the original image associated with the scaling function $\varphi^{(1)}$ as well as several levels of detail coefficients, associated with the wavelets $\psi^{(1)}$, $\psi^{(2)}$, and $\psi^{(3)}$.

Just as in the one-dimensional case, the two-dimensional mother wavelets defined in Equations (3.28) to (3.30) can be used to define an ONB for $L^2(\mathbb{R}^2)$, as illustrated by the following theorem.

Theorem 3.0.3 (Wavelet ONB for $L^2(\mathbb{R}^2)$, e.g. [121]). *Let $\varphi, \psi \in L^2(\mathbb{R})$ be the scaling function respectively the mother wavelet of an MRA for $L^2(\mathbb{R})$. Then, the two-dimensional wavelets $\psi^{(1)}$, $\psi^{(2)}$, and $\psi^{(3)}$ (see (3.28) to (3.30)) are the generators of an ONB for $L^2(\mathbb{R}^2)$ in the sense that the set $\left\{ \psi_{j, \mathbf{m}}^{(l)} : j \in \mathbb{Z}, \mathbf{m} \in \mathbb{Z}^2, l \in \{1, 2, 3\} \right\}$ is an ONB for $L^2(\mathbb{R}^2)$.*

By virtue of being localized in the time domain, wavelets are well suited for characterizing transient features such as singularities in the one-dimensional setting. To a certain extent, this property is carried over to the two-dimensional world. For instance, the detail coefficients associated with the wavelet generators $\psi^{(1)}$, $\psi^{(2)}$, and $\psi^{(3)}$ in the wavelet-based image representations in Figure 3.2 are clearly highlighting the edges of the image depicted in Figure 3.2(a). However, it can be shown that systems of two-dimensional wavelets are in a sense not optimal for representing the curvilinear features typically occurring

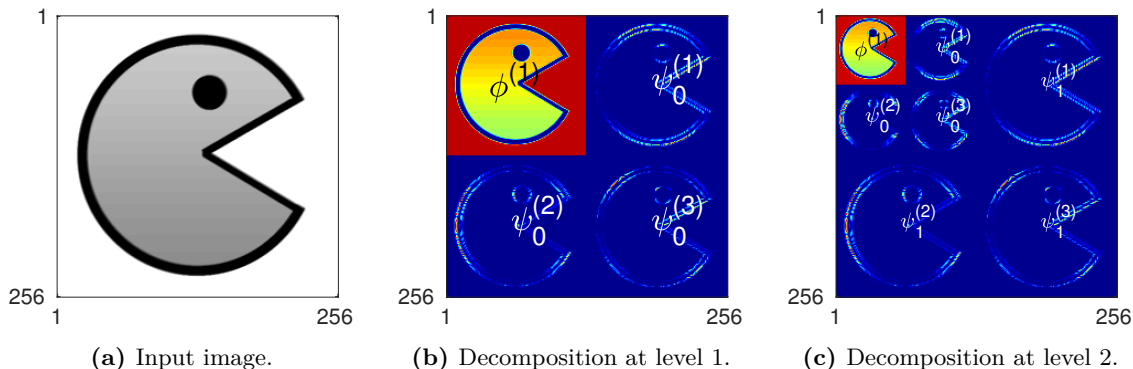


Figure 3.2: Wavelet decompositions of a digital image at level 1 and 2. Each pixel in 3.2(b) and 3.2(c) corresponds to the inner product of the image 3.2(a) with a discrete wavelet filter. The coefficients associated with the low-pass scaling function $\varphi^{(1)}$ define a coarse approximation of the original image. The coefficients associated with the high-pass wavelet functions $\psi^{(1)}$, $\psi^{(2)}$, and $\psi^{(3)}$ highlight horizontal, vertical and diagonal edges, respectively. Wavelet transforms often yield sparse representations of images, which is indicated by the large number of zeros in 3.2(b) and 3.2(c).

in a two-dimensional image [50]. The main reason behind this that such features are typically associated with a direction (e.g. the tangent direction of an edge), that is, they are *anisotropic*. The scaling matrix \mathbf{A} defined in (3.32), however, is *isotropic* in the sense that both dimensions are either squeezed or stretched by the same amount. To optimally characterize transient anisotropic two-dimensional features, it would be preferable to use systems in which the high-frequency elements are also directionally sensitive. This motivated the investigation of multiscale transforms for two- and higher-dimensional signals that are based on anisotropic scaling such as curvelets [25], contourlets [49], and shearlets [104, 110]. All of these approaches have since found numerous applications in image processing tasks such as image denoising [58], inpainting [95], or edge detection [108, 144, 171]. In this chapter, we will focus on shearlet-based image representations. However, we will see in the prelude to Chapter 4 that many of the theoretical results regarding approximation properties can indeed be transferred between different constructions of multiscale representation systems.

In most cases, systems that are based on anisotropic scaling contain redundancies in the sense that their elements are not linearly independent. In particular, this means that such systems do not constitute an ONB and hence do not provide uniquely defined expansions. At first glance, this might seem like a drawback. However, when applying a transform that is based on a redundant system, it is often possible to find representations that are sparser and thus, in a sense, more efficient, than when only considering ONBs. Furthermore, for the specific applications considered in this thesis, the most important question is not if

and how an analyzed image can be reconstructed after applying a certain transform but whether the transform succeeds at reliably extracting the features that are most significant for the respective task. In such a setting, redundant systems can have huge advantages over ONBs.

Before introducing shearlet-based transforms, we will thus take a short detour and briefly discuss the definition of a *frame*, which provides a useful generalization of the concept of an orthonormal basis and is one of the most important tools for the investigation of redundant representation systems. For a countable index set Λ , a family $\Phi = \{\varphi_\lambda\}_{\lambda \in \Lambda} \subset L^2(\mathbb{R}^2)$ is called a *frame* for $L^2(\mathbb{R}^2)$ if there exist constants $0 < A \leq B$ such that

$$A \|f\|_2^2 \leq \sum_{\lambda \in \Lambda} |\langle f, \varphi_\lambda \rangle_{L^2}|^2 \leq B \|f\|_2^2, \quad \forall f \in L^2(\mathbb{R}^2). \quad (3.36)$$

This condition is a relaxation of Parseval's identity $\sum_{\lambda \in \Lambda} |\langle f, \varphi_\lambda \rangle_{L^2}|^2 = \|f\|_2^2$, which holds for orthonormal systems if and only if they also constitute a basis. If $A = B$, the corresponding frame is called a *tight frame*. If furthermore $A = B = 1$, (3.36) reduces to Parseval's identity and the associated frame is called a *Parseval frame*. Note that even Parseval frames do not necessarily define an ONB if they do not constitute an orthonormal system. The transform of a function $f \in L^2(\mathbb{R}^2)$ with respect to a frame is defined via the so-called *analysis operator*

$$\mathcal{T}_\Phi: L^2(\mathbb{R}^2) \rightarrow \ell^2(\Lambda): f \mapsto \{\langle f, \varphi_\lambda \rangle_{L^2}\}_{\lambda \in \Lambda}, \quad (3.37)$$

whose adjoint

$$\mathcal{T}_\Phi^*: \ell^2(\Lambda) \rightarrow L^2(\mathbb{R}^2): \{c_\lambda\}_{\lambda \in \Lambda} \mapsto \sum_{\lambda \in \Lambda} c_\lambda \varphi_\lambda, \quad (3.38)$$

is often denoted as the *synthesis operator*. Due to \mathcal{T}_Φ^* being bounded, a series of the form $\sum_{\lambda \in \Lambda} c_\lambda \varphi_\lambda$, where \mathbf{c} is in $\ell^2(\Lambda)$, and $\{\varphi_\lambda\}_{\lambda \in \Lambda}$ is a frame, is associated with a notion of *unconditional convergence* (see, e.g., [69]). In particular, this means that certain manipulations such as rearrangements, or interchanging the summation with a linear operator, are permitted. Due to the fact that the so-called *frame operator* $\mathcal{S}_\Phi = \mathcal{T}_\Phi^* \mathcal{T}_\Phi: L^2(\mathbb{R}^2) \rightarrow L^2(\mathbb{R}^2)$ is invertible (see, e.g., [33]), unconditional convergence can also be applied to obtain the following reconstruction formula:

$$f = \mathcal{S}_\Phi^{-1} \mathcal{S}_\Phi f = \mathcal{S}_\Phi^{-1} \sum_{\lambda \in \Lambda} \langle f, \varphi_\lambda \rangle_{L^2} \varphi_\lambda = \sum_{\lambda \in \Lambda} \langle f, \varphi_\lambda \rangle_{L^2} \mathcal{S}_\Phi^{-1} \varphi_\lambda, \quad \forall f \in L^2(\mathbb{R}^2), \quad (3.39)$$

This means that, analogous to the case of an orthonormal basis, it is always possible to reconstruct a function f from its *analysis coefficients* $\mathcal{T}_\Phi f$. Note that in the case of tight frames, $\mathcal{S}_\Phi = \frac{1}{A}$, and for Parseval frames, the frame operator is just the identity. A

comprehensive introduction to the theory of frames can be found in [33].

Let us now turn to shearlet-based transforms for functions in $L^2(\mathbb{R}^2)$. The main idea behind the construction of shearlet systems is to replace the isotropic scaling matrix \mathbf{A} (see (3.32)) with an anisotropic scaling matrix

$$\mathbf{A}_{1/2} = \begin{bmatrix} 2 & 0 \\ 0 & 2^{1/2} \end{bmatrix}, \quad (3.40)$$

which treats both dimensions differently. Broadly speaking, applying the anisotropic scaling matrix $\mathbf{A}_{1/2}$ to the argument of a generating function will cause the function to be more severely squeezed in the first dimension than in the second one. This means that consecutive applications of the scaling matrix $\mathbf{A}_{1/2}$ will introduce an increasing directional bias to the high-frequency elements of a shearlet-based system. The difference between isotropic scaling and anisotropic scaling is illustrated in Figure 3.3 in the case of digital two-dimensional wavelet and shearlet filters. Note that the exponent $1/2$ defines a so-called *parabolic* scaling matrix. There exist generalizations in which $1/2$ can be replaced with an exponent $\alpha \in [0, 1]$ [71], which will be discussed in more detail in Chapter 4, and generalizations which are not restricted to the dyadic scaling factor 2 [66]. In the wavelet case, due to their localization in the time domain, it is necessary to also consider shifts of dilated mother wavelet in order to define an ONB for $L^2(\mathbb{R})$ or $L^2(\mathbb{R}^2)$. Analogously, in the case of systems of directionally sensitive functions, we require a means of changing their preferred orientation in order to have a chance of obtaining frames or even bases for $L^2(\mathbb{R}^2)$. In the shearlet setting, this is done by applying the shear matrices

$$\mathbf{S}_k = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}, \quad k \in \mathbb{Z} \quad (3.41)$$

to the argument of a shearlet generator. Other constructions that are based on anisotropic scaling, such as curvelets [25], use rotations rather than shears. However, one of the most significant advantages of the shear operator is that for shear parameters $k \in \mathbb{Z}$, the integer grid is invariant under the action of the shear matrix \mathbf{S}_k . In particular, this facilitates the faithful digitization of shearlet-based transforms defined in $L^2(\mathbb{R})$.

Similar to the wavelet case, a function $\psi \in L^2(\mathbb{R}^2)$ is said to be an *admissible shearlet* if it satisfies

$$\iint_{\mathbb{R}^2} \left| \widehat{\psi}(\xi_1, \xi_2) \right|^2 \xi_1^{-2} d\xi_2 d\xi_1 < \infty, \quad (3.42)$$

which implies the existence of an inverse of the continuous shearlet transform (see, e.g., [105]), and for a shearlet generator $\psi \in L^2(\mathbb{R}^2)$, we will consider anisotropically dilated, sheared,

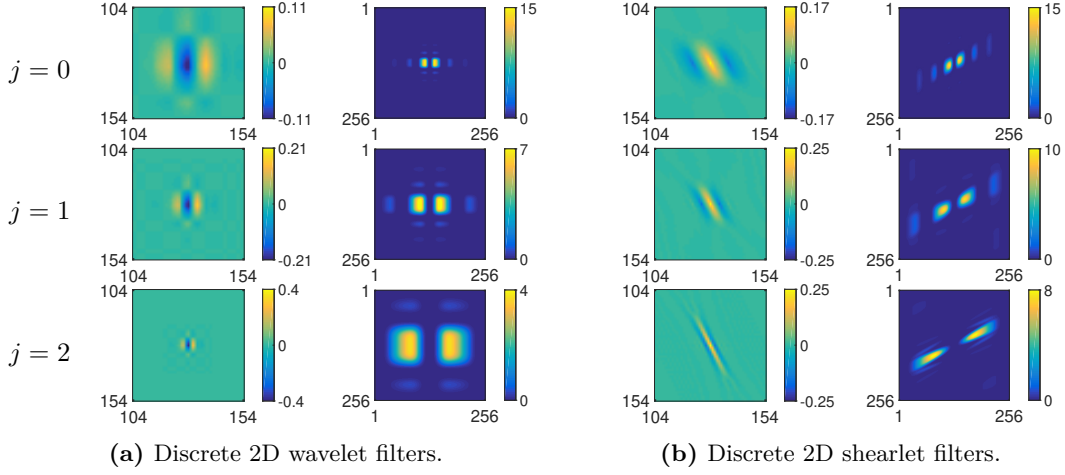


Figure 3.3: Comparison of discrete 2D wavelet and shearlet filters at different scales. Due to the anisotropic scaling, the high-frequency shearlet filters in the last row of 3.3(b) are elongated and better adapted to fit edges in images than the high-frequency wavelet filters in the last row of 3.3(a). The frequency spectrum of each filter is depicted in the right columns of 3.3(a) and 3.3(b).

and shifted L^2 -normalized functions of the form

$$\psi_{j,k,\mathbf{m}} = 2^{\frac{3j}{4}} \psi(\mathbf{S}_k \mathbf{A}_{1/2}^j \cdot -\mathbf{m}), \quad j \in \mathbb{Z}, k \in \mathbb{Z}, \mathbf{m} \in \mathbb{Z}^2. \quad (3.43)$$

Like in the wavelet case, the discrete shearlet-based transform $\mathcal{SH}_\psi f$ of a function $f \in L^2(\mathbb{R}^2)$ with respect to a generating shearlet $\psi \in L^2(\mathbb{R}^2)$ is given for a scale parameter $j \in \mathbb{Z}$, a shear parameter $k \in \mathbb{Z}$, and a translation parameter $\mathbf{m} \in \mathbb{Z}^2$ by

$$(\mathcal{SH}_\psi f)(j, k, \mathbf{m}) = \langle f, \psi_{j,k,\mathbf{m}} \rangle_{L^2} = 2^{\frac{3j}{4}} \int_{\mathbb{R}^2} f(\mathbf{x}) \overline{\psi(\mathbf{S}_k \mathbf{A}_{1/2}^j \mathbf{x} - \mathbf{m})} \, d\mathbf{x}. \quad (3.44)$$

One of the problems of using shear matrices of the form (3.41) to construct shearlet systems is that they have an intrinsic directional bias in the sense that changing the preferred orientation of a shearlet generator by 90° is only possible in the limit $k \rightarrow \infty$. Furthermore, due to the effect of the shear operator, shearlets $\psi_{j,k,\mathbf{m}}$ with $|k| \gg 0$ will be unnaturally elongated. A straightforward solution to this problem is given by considering so-called *cone-adapted* shearlet systems. Such systems are based on two shearlet generators denoted as ψ and $\tilde{\psi}$, where the essential support of $\hat{\psi}$ is contained in the horizontal frequency cones and the essential support of $\hat{\tilde{\psi}}$ is contained in the vertical frequency cones. By considering the matrices

$$\tilde{\mathbf{A}}_{1/2} = \begin{bmatrix} 2^{1/2} & 0 \\ 0 & 2 \end{bmatrix}, \quad \text{and} \quad \mathbf{S}_k^\top = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}, \quad k \in \mathbb{Z}, \quad (3.45)$$

and with a two-dimensional scaling function $\varphi \in L^2(\mathbb{R}^2)$, a discrete cone-adapted shearlet system is defined as the union

$$SH(\varphi, \psi, \tilde{\psi}) = \Phi(\varphi) \cup \Psi(\psi) \cup \tilde{\Psi}(\tilde{\psi}), \quad (3.46)$$

where

$$\Phi(\varphi) = \{\varphi_{\mathbf{m}} = \varphi(\cdot - \mathbf{m}) : \mathbf{m} \in \mathbb{Z}^2\}, \quad (3.47)$$

$$\Psi(\psi) = \left\{ \psi_{j,k,\mathbf{m}} = 2^{j\frac{3}{4}} \psi(\mathbf{S}_k \mathbf{A}_j \cdot -\mathbf{m}) : j \in \mathbb{N}_0, |k| < \left\lceil 2^{\frac{j}{2}} \right\rceil, \mathbf{m} \in \mathbb{Z}^2 \right\}, \quad (3.48)$$

$$\tilde{\Psi}(\tilde{\psi}) = \left\{ \tilde{\psi}_{j,k,\mathbf{m}} = 2^{j\frac{3}{4}} \tilde{\psi}(\mathbf{S}_k^\top \tilde{\mathbf{A}}_j \cdot -\mathbf{m}) : j \in \mathbb{N}_0, |k| < \left\lceil 2^{\frac{j}{2}} \right\rceil, \mathbf{m} \in \mathbb{Z}^2 \right\}. \quad (3.49)$$

Here, the condition $|k| < \left\lceil 2^{\frac{j}{2}} \right\rceil$ prohibits shears of a generating shearlet that would change the preferred orientation by more than 45° . In particular, this forces the essential support of a shearlet in the frequency domain to remain within the cones in which the corresponding generating shearlet is located (hence the name *cone-adapted*). In turn, this means that for a cone-adapted shearlet system to cover the whole frequency plane, it is important to choose the generating shearlets ψ and $\tilde{\psi}$ such that they are supported in different frequency cones. In particular, this can be ensured by defining the functions φ, ψ and $\tilde{\psi}$ analogous to Equations (3.28) to (3.31) as tensor products of one-dimensional scaling and wavelet functions. The tiling of the frequency domain induced by the system (3.46) is schematically depicted in Figure 3.4(b), where it can also be compared to the tiling induced by a discrete wavelet system, shown in Figure 3.4(a). Note that in the case of cone-adapted systems, the definition of the shearlet transform $(\mathcal{SH}_\psi f)$ (cf. (3.44)) needs to be adapted for $\tilde{\psi}$ in the sense that the matrices \mathbf{S}_k and $\mathbf{A}_{1/2}$ have to be replaced with \mathbf{S}_k^\top and $\tilde{\mathbf{A}}_{1/2}$, respectively.

It seems evident that systems based on anisotropic scaling should be better suited for characterizing the curvilinear features that often govern the geometry of two-dimensional images than systems based on isotropic scaling. For a special class of images, so-called *cartoon-like images*, which were first considered in this context in [52], it can indeed be shown that shearlet-based systems of the form (3.46) yield in a sense optimal approximations.

Definition 3.0.4 (Cartoon-like image function, [52], as in [104]). *Let $f \in L^2(\mathbb{R}^2)$ be compactly supported in $[0, 1]^2$. If f can be written as*

$$f = f_0 + \mathbb{1}_B \cdot f_1, \quad (3.50)$$

$B \subset [0, 1]^2$ is a measurable set whose boundary is a closed C^2 -curve with bounded curvature, $f_0, f_1 \in C^2(\mathbb{R}^2)$ are compactly supported in $[0, 1]^2$, and $\|f_0\|_{C^2}, \|f_1\|_{C^2} \leq 1$, then f is a

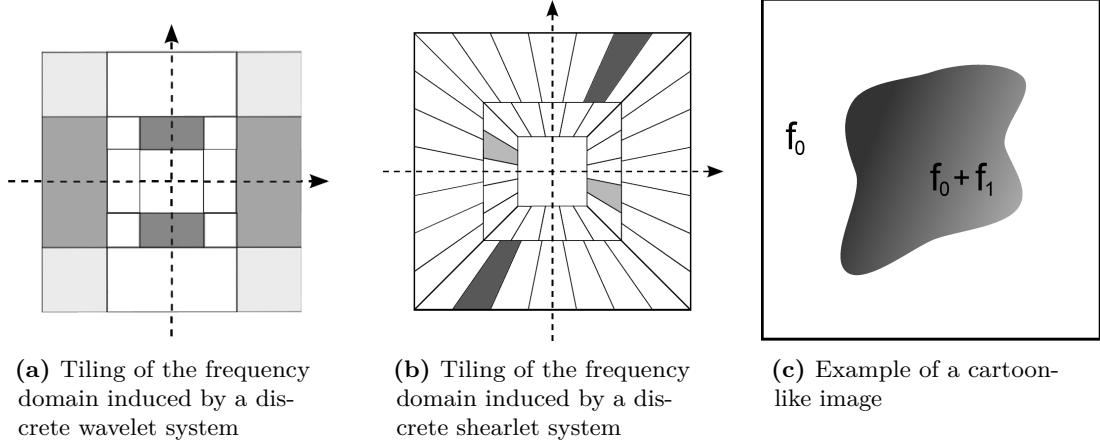


Figure 3.4: Schematic depiction of partitions of the Fourier domain induced by (a): wavelets and (b): shearlets. In each image, the areas drawn in the same shade of gray indicate the frequency support of an element of a discrete wavelet or shearlet system associated with the same generator and the same scaling and rotation parameters.

cartoon-like image function.

An example of a cartoon-like image is shown in Figure 3.4(c). Before stating the associated optimality result, we need to briefly discuss the notion of N -term approximations, which will provide us with a meaningful way of comparing the approximation properties of different systems of analyzing functions with respect to a given class of signals. Let Λ be a countable index set and $\Phi = \{\varphi_\lambda\}_{\lambda \in \Lambda}$ denote a normalized frame for $L^2(\mathbb{R}^2)$, that is, $\|\varphi_\lambda\|_2 = 1$ for all $\lambda \in \Lambda$. Furthermore, let $f \in L^2(\mathbb{R}^2)$, $N \in \mathbb{N}$, and $\Lambda_N \subset \Lambda$ denote the subset of Λ that contains the N largest coefficients of $\mathcal{T}_\Phi f$ in magnitude, that is,

$$|\Lambda_N| = N, \text{ and } \forall \lambda \in \Lambda_N, \mu \in \Lambda \setminus \Lambda_N: |\langle f, \varphi_\lambda \rangle_{L^2}| \geq |\langle f, \varphi_\mu \rangle_{L^2}|. \quad (3.51)$$

Then, the N -term approximation of f associated with the N largest coefficients is defined as

$$f_N = \sum_{\lambda \in \Lambda_N} \langle f, \varphi_\lambda \rangle_{L^2} \mathcal{S}_\Phi^{-1} \varphi_\lambda, \quad (3.52)$$

where \mathcal{S}_Φ denotes the frame operator associated with the frame Φ (cf. (3.39)). It was shown in [52] that when only considering sets Λ_N that are obtained by a practically feasible selection process, it is in fact possible to characterize the optimal approximation behavior of a frame with respect to cartoon-like image functions.

Definition 3.0.5 (Optimally sparse approximations of cartoon-like images, [52], as in [106]). *Let $\Phi = \{\varphi_\lambda\}_{\lambda \in \Lambda}$ be a normalized frame for $L^2(\mathbb{R}^2)$. Then, Φ is said to yield optimally sparse approximations of cartoon-like image functions if the N -term approximations associated*

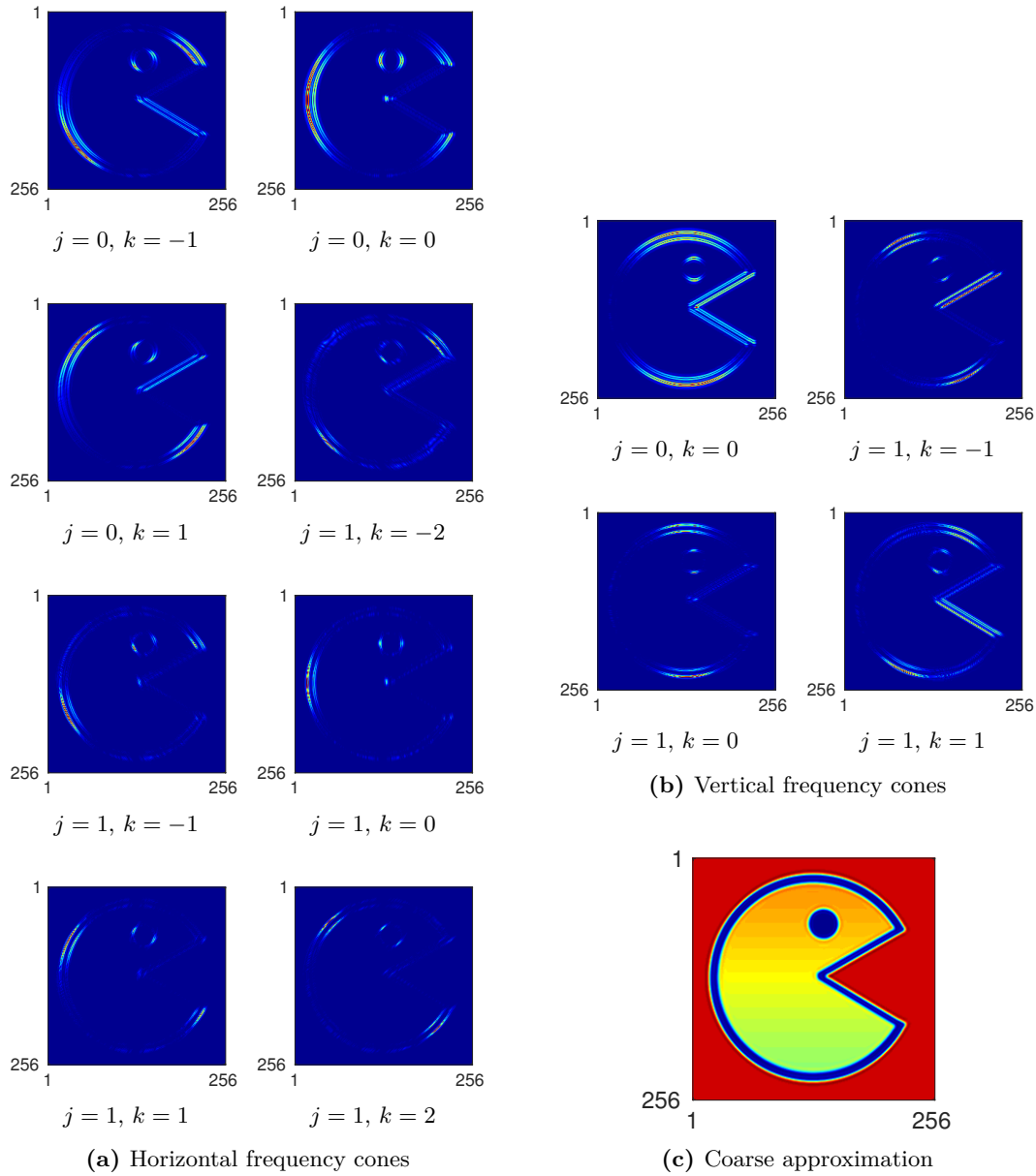


Figure 3.5: Non-subsampled shearlet decomposition at level 2 of the 256×256 digital image shown in Figure 3.2(a). The parameter j denotes the scale of the corresponding shearlet filter while the shear parameter k controls its orientation. All coefficient matrices have the same size as the input image. The depicted decomposition is thus highly redundant with a redundancy factor of 13.

with the N largest coefficients of the sequence $\mathcal{T}_\Phi f$ in magnitude $f_N \in L^2(\mathbb{R}^2)$ satisfy

$$\|f - f_N\|_2 \lesssim N^{-1}, \quad (3.53)$$

for all cartoon-like image functions $f \in L^2(\mathbb{R}^2)$.

The following theorem illustrates that, apart from a logarithmic factor, the optimal approximation rate (3.53) can in fact be achieved for cartoon-like images when considering cone-adapted shearlet systems.

Theorem 3.0.6 (Shearlet-based approximations of cartoon-like functions are optimally sparse, e.g., [106]). *Let $\psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ be aptly chosen shearlet generators and $\varphi \in L^2(\mathbb{R}^2)$ be an aptly chosen scaling function. Then, apart from a log-factor, the cone-adapted shearlet system $SH(\varphi, \psi, \tilde{\psi})$ provides optimally sparse approximations of cartoon-like image functions in the sense that*

$$\|f - f_N\|_2 \lesssim N^{-1}(\log_2 N)^{3/2}, \quad (3.54)$$

for all cartoon-like image functions $f \in L^2(\mathbb{R}^2)$, where f_N denotes the N -term approximation associated with the N largest shearlet coefficients in magnitude.

During the past decade, several libraries implementing different types of shearlet transforms for finite and discrete data have been developed. The fast finite shearlet transform (FFST) [78] implements non-subsampled transforms based on band-limited shearlet generators. ShearLab [109] implements subsampled transforms that are based on separable and compactly supported shearlet generators while ShearLab 3D [107] contains both two- and three-dimensional non-subsampled shearlet transforms which are based on non-separable and compactly supported shearlets.

Figure 3.5 shows the coefficients of a non-subsampled shearlet-based transform of a digital image after two stages of decomposition that was obtained with ShearLab 3D. Comparing Figure 3.2 and Figure 3.5 also illustrates the difference between subsampled and non-subsampled decompositions. The main advantage of subsampled shearlet- and wavelet-based transforms is that the total number of coefficients is almost equal to the original number of pixel values, while non-subsampled transforms often introduce a significant amount of redundancy. An exhaustive overview of the theory and applications of shearlets can be found in [104].

3.1 Introduction

Visual servoing is a technique in which visual features provided by one or multiple vision sensors are used to control the motion of a robot in a closed-loop scheme [31, 83]. The past thirty years have seen rapid advances in this field, with applications in industrial manipulation [115], medical robotics [3, 103] or, more recently, drone navigation [9, 126]. The general goal of vision-based control laws is to succeed at a positioning task by minimizing the difference between a set of desired visual features \mathbf{s}^* and a set of currently observed features $\mathbf{s}(t)$. The historically first type of visual features considered in visual servoing were geometric features such as points [30], lines [146], and image moments [160], and most visual servoing approaches strongly depend on the ability to detect and track visual features over time during the control process.

Recently, new methods emerged that avoid challenging visual tracking tasks by directly using readily available information such as image intensities [151, 161], mutual information [42], the sum of conditional variance [147], the Fourier transform [124, 125], or Gaussian mixtures [38]. Such techniques, which are typically referred to as direct visual servoing schemes, are often considered to be more accurate and robust than visual servoing approaches that are based on the extraction of geometric features.

The work described in this chapter deals with the development of new direct visual servoing approaches where the input control signals are chosen as the coefficients of wavelet- and shearlet-based multiscale image representations. Multiscale representations of an image are typically obtained by repeatedly smoothing and subsampling the image while also storing the detail information lost at each stage of this process. Such decompositions are not only useful for defining increasingly coarse approximations of images, but also to obtain sparse representations in the sense that most coefficients describing the detail information lost in each transformation step are typically close to zero. A widely used multiscale image representation scheme is the so-called Laplacian pyramid, introduced in 1983 by Burt and Adelson [20], which is based on repeated applications of a Gaussian blurring kernel. The Laplacian pyramid was later extended by Simoncelli et al. [152] to yield improved representations of oriented features such as curves and edges. Another way of defining multiscale image representations can be found by considering the two-dimensional wavelet- and shearlet-based transforms that were already discussed in the prelude to this chapter.

Despite their prominent role in digital image processing, the application of wavelet- or shearlet-based image representations as signal control inputs in a vision-based control scheme has only recently been considered. A six degrees of freedom (DOF) visual servoing approach based on the low-pass approximation of an image obtained from a discrete wavelet transform was proposed in [132, 133]. Shearlet-based transforms were considered in a

six DOF ultrasound-based visual servoing scheme that was proposed for automatically positioning an ultrasound probe mounted on a robotic arm [54, 55]. These preliminary investigations already indicate the feasibility and potential benefits of considering wavelet- or shearlet-based multiscale image representations in a visual servoing scheme when dealing with unfavorable conditions that include noise, partial image occlusions, or changes in illumination. In particular, using multiscale image decompositions could provide an interesting compromise between purely geometric feature-based approaches and image-intensity-based visual servoing techniques. The coarse approximation of an image yielded by a multiscale representation scheme is a smoothed and subsampled version of the original image and thus robust to noise. In addition, the detail coefficients obtained at the different decomposition levels are highlighting basic image features, such as edges or curves, and can thus be seen as implicit visual feature detectors.

3.1.1 Contributions

Up to now, one of the main obstacles in the application of wavelet- and shearlet-based multiscale image representations in the field of visual servoing was the issue of obtaining an analytical formulation of the interaction model, typically denoted as the *interaction matrix*, that links the variation of the wavelet or shearlet coefficients to the spatial velocity of the visual sensor. For instance, in [54, 55], the interaction matrix was obtained via a numerical approximation. In this chapter, we analytically derive interaction matrices for control laws based on both wavelet and shearlet coefficients. That is, the variation of the wavelet or shearlet coefficients in the visual feature vector over time is analytically linked to the velocity of the camera which can be controlled by a robot with six degrees of freedom.

Moreover, the proposed wavelet- and shearlet-based control laws are tested in both simulation and experimental scenarios using a 6 DOF Cartesian robot with an eye-in-hand configuration. The goal of these experiments is to provide a thorough qualitative and quantitative analysis of their respective strengths and weaknesses. In particular, numerous experiments under favorable (nominal) conditions and unfavorable (partial occlusions, unstable illumination) working conditions were carried out. It can be highlighted that wavelet as well as shearlet-based visual servoing approaches provide good performances with regards to accuracy and robustness to external disturbances.

Section 3.1.2 provides a short review of a classical direct visual servoing approach, so-called photometric visual servoing, in which image intensities are used as the only visual features. It is then shown in Sections 3.2 and 3.3 how the concept of photometric visual servoing scheme can be extended to obtain wavelet- and shearlet-based control laws and how the associated interaction matrices can be derived analytically. Section 3.4 finally

presents the results obtained from several simulations and experimental validations under different working conditions and a comparison of wavelet- and shearlet-based schemes with the classical photometric approach.

3.1.2 Photometric Visual Servoing

Before discussing wavelet- and shearlet-based visual servoing control laws, we will briefly revisit the well-known photometric visual servoing approach in order to recall some basic concepts and notation. The objective of a visual servoing scheme is to control the motion of a robot such that a set of $N \in \mathbb{N}$ visual features $\mathbf{s} \in \mathbb{R}^N$ depending on the robot's pose $\mathbf{r}(t) \in SE(3)$ at time t matches a set of desired features $\mathbf{s}^* \in \mathbb{R}^N$, obtained at the desired pose \mathbf{r}^* . While the feature vector \mathbf{s} should technically be viewed as a function of the robot's pose, to simplify our notation, we will treat the set of visual features as a function of time by setting $\mathbf{s}(t) := \mathbf{s}(\mathbf{r}(t))$.

For the remainder of this chapter, we will use $I(x, y, t) \in \mathbb{R}_{\geq 0}$ to denote the image intensities at the coordinates (x, y) observed at time t . Furthermore, we write $I_t = I(\cdot, \cdot, t)$ to denote the two-dimensional image observed at time t . This notation will be helpful in unambiguously stating expressions that contain operations which are only defined for two-dimensional functions such as two-dimensional wavelet and shearlet transforms. The control laws derived in this chapter will eventually be based on the spatial gradient ∇I_t as well as wavelet and shearlet transforms of the two-dimensional image I_t , that is, the coefficients $(\mathcal{W}_\psi I_t)(j, \mathbf{m})$ and $(\mathcal{SH}_\psi I_t)(j, k, \mathbf{m})$ (cf. (3.34) and (3.44)). Note that formally, these objects are only defined in the continuum. Consequently, the corresponding visual servoing schemes are also derived in the continuous realm. In practice, however, we cannot observe the complete continuous-time image I_t but only a discretized approximation \mathbf{I}_t that is typically obtained by spatial averaging and sampling. This means that in any implementation of the proposed control laws, the spatial gradient as well as the wavelet- and shearlet-based transforms of I_t need to be approximated in the discrete realm. This can be achieved by convolving \mathbf{I}_t with gradient filters and by computing the discrete-time wavelet and shearlet transforms of \mathbf{I}_t , respectively.

In the case of photometric visual servoing, introduced in [37], the visual feature vector at time t is a finite set of image intensities sampled at a sequence of points $\{(x_n, y_n)\}_{n \leq N} \subset \mathbb{R}^2$:

$$\mathbf{s}_{\text{ph}}(t) = [I(x_1, y_1, t), I(x_2, y_2, t), \dots, I(x_N, y_N, t)]^\top. \quad (3.55)$$

To reach the desired pose \mathbf{r}^* , a control law is applied to minimize the visual error given by

$$\mathbf{e}_{\text{ph}}(t) = \mathbf{s}_{\text{ph}}(t) - \mathbf{s}_{\text{ph}}^* \quad (3.56)$$

over time by accordingly moving the robotic system. In order to obtain an error-decreasing six-dimensional camera velocity vector $\mathbf{v} = [v_x, v_y, v_z, \omega_x, \omega_y, \omega_z]^\top$, where v_x , v_y , and v_z denote linear translational, and ω_x , ω_y , and ω_z denote angular velocities, it is necessary to establish a link between the variation of the visual features \mathbf{s}_{ph} , defined in the image plane, and the movement of the camera. In [37], the authors formalize this relationship by linearly describing the time derivative of the image intensities in terms of the spatial velocities of the camera via a so-called interaction matrix \mathbf{L}_{sph} :

$$\frac{d\mathbf{s}_{\text{ph}}(t)}{dt} = \mathbf{L}_{\text{sph}}(t)\mathbf{v}(t). \quad (3.57)$$

By assuming temporal luminance constancy and applying the optical flow constraint equation (OFCE) introduced in [81], we can write

$$\frac{\partial I(x, y, t)}{\partial t} = -\left(\nabla I_t(x, y)\right)^\top \left(\frac{dx}{dt}, \frac{dy}{dt}\right)^\top. \quad (3.58)$$

Using the velocity vector $\mathbf{v}(t)$, (3.58) can be written as

$$\frac{\partial I(x, y, t)}{\partial t} = -\left(\nabla I_t(x, y)\right)^\top \mathbf{L}_p(x, y)\mathbf{v}(t), \quad (3.59)$$

where $\mathbf{L}_p(x, y)$ denotes the interaction matrix of the point $(x, y) \in \mathbb{R}^2$ corresponding to the projection of a three-dimensional point in the scene into the image plane, as proposed in [31]. \mathbf{L}_p is defined for a point (x, y) in the image plane as

$$\mathbf{L}_p(x, y) = \begin{bmatrix} -\frac{1}{Z} & 0 & \frac{x}{Z} & xy & -(1+x^2) & y \\ 0 & -\frac{1}{Z} & \frac{y}{Z} & 1+y^2 & -xy & -x \end{bmatrix}, \quad (3.60)$$

where $Z \in \mathbb{R}$ is the depth of the considered three-dimensional point in the Cartesian camera frame. Here, the depth Z is assumed to remain constant over time.

Applying (3.57) and (3.59), we can explicitly compute the interaction matrix \mathbf{L}_{sph} at a point in time t as follows:

$$\mathbf{L}_{\text{sph}}(t) = - \begin{bmatrix} \left(\nabla I_t(x_1, y_1)\right)^\top \mathbf{L}_p(x_1, y_1) \\ \vdots \\ \left(\nabla I_t(x_N, y_N)\right)^\top \mathbf{L}_p(x_N, y_N) \end{bmatrix}. \quad (3.61)$$

Finally, the velocity vector of the camera at a time t is obtained by

$$\mathbf{v}(t) = -\lambda \left(\mathbf{L}_{\text{sph}}(t)\right)^+ \mathbf{e}_{\text{ph}}(t), \quad (3.62)$$

where $\lambda > 0$ is a gain parameter, and $(\mathbf{L}_{s_{ph}}(t))^+$ denotes the Moore-Penrose pseudo-inverse of the interaction matrix $\mathbf{L}_{s_{ph}}(t)$.

To increase the domain of convergence, we consider the Levenberg-Marquardt method, which is a generalization of the traditional Gauss-Newton optimization approach (3.62). With a damping parameter $\mu > 0$, the corresponding control law can then be written as

$$\mathbf{v}(t) = -\lambda \left(\mathbf{H}(t) + \mu \text{diag}(\mathbf{H}(t)) \right)^{-1} \left(\mathbf{L}_{s_{ph}}(t) \right)^\top \mathbf{e}_{ph}(t), \quad (3.63)$$

where $\mathbf{H}(t) = (\mathbf{L}_{s_{ph}}(t))^\top \mathbf{L}_{s_{ph}}(t)$, and $\text{diag}(\mathbf{H}(t))$ denotes the diagonal matrix given by the diagonal entries of $\mathbf{H}(t)$.

3.2 Wavelet-Based Visual Servoing

Instead of the pixel intensities used in (3.55), we now consider the design of a visual servoing scheme in which the visual features are defined as the coefficients of a wavelet-based multiscale representation of an image.

Let us consider a finite set of $L \in \mathbb{N}$ generating wavelets $\{\psi^{(l)}\}_{l \leq L} \subset L^2(\mathbb{R}^2)$, which typically consists of a total of four separable generators constructed from one-dimensional wavelet and scaling functions, as defined in Equations (3.28) to (3.31). The wavelet-based feature vector for a set of $N \in \mathbb{N}$ triples $\{(l_n, j_n, \mathbf{m}_n)\}_{n \leq N} \subset \{1, \dots, L\} \times \mathbb{N} \times \mathbb{Z}^2$, where l_n specifies a wavelet generator, j_n defines the scale and \mathbf{m}_n the translation parameter is given at a time t as

$$\mathbf{s}_w(t) = \left[(\mathcal{W}_{\psi^{(l_1)}} I_t)(j_1, \mathbf{m}_1), (\mathcal{W}_{\psi^{(l_2)}} I_t)(j_2, \mathbf{m}_2), \dots, (\mathcal{W}_{\psi^{(l_N)}} I_t)(j_N, \mathbf{m}_N) \right]^\top. \quad (3.64)$$

The goal is now to derive an interaction matrix $\mathbf{L}_{s_w} \in \mathbb{R}^{N \times 6}$ that relates the camera motion defined by the six-dimensional velocity vector \mathbf{v} to the time derivative of the wavelet-based feature vector \mathbf{s}_w , such that the following linearization holds:

$$\frac{d\mathbf{s}_w(t)}{dt} = \mathbf{L}_{s_w}(t) \mathbf{v}(t). \quad (3.65)$$

Let us assume that all wavelet generators as well as the image intensities I are continuously differentiable. With the two-dimensional generalization of the Leibniz integral rule, also known as the Reynolds transport theorem (see for example [60]), we can express the time

derivative of a single entry of the feature vector \mathbf{s}_w at a fixed point (l, j, \mathbf{m}) by

$$\frac{d(\mathcal{W}_{\psi^{(l)}} I_t)(j, \mathbf{m})}{dt} = \frac{d \langle I_t, \psi_{j, \mathbf{m}}^{(l)} \rangle_{L^2}}{dt} = \iint_{\mathbb{R}^2} \frac{\partial I(x, y, t)}{\partial t} \psi_{j, \mathbf{m}}^{(l)}(x, y) \, dx \, dy. \quad (3.66)$$

Then, by applying (3.59), we get

$$\frac{d \langle I_t, \psi_{j, \mathbf{m}}^{(l)} \rangle_{L^2}}{dt} = - \iint_{\mathbb{R}^2} (\nabla I_t(x, y))^T \mathbf{L}_p(x, y) \mathbf{v}(t) \psi_{j, \mathbf{m}}^{(l)}(x, y) \, dx \, dy. \quad (3.67)$$

Let $\mathbf{e}_i \in \mathbb{R}^6$ denote the i -th canonical unit vector in \mathbb{R}^6 . Then we can simplify our notation by setting

$$I_t^{(i)} = (\nabla I_t(x, y))^T \mathbf{L}_p(x, y) \mathbf{e}_i, \quad (3.68)$$

for $i \in \{1, \dots, 6\}$. Note that in (3.68), $\mathbf{L}_p(x, y) \mathbf{e}_i$ selects the i -th column of the matrix $\mathbf{L}_p(x, y)$. In other words, $I_t^{(i)}$ encodes the correlation of the image gradient with the movement contributed by the i -th degree of freedom. Using (3.68), the interaction matrix with respect to the wavelet-based feature vector \mathbf{s}_w can be written as:

$$\mathbf{L}_{\mathbf{s}_w}(t) = - \begin{bmatrix} (\mathcal{W}_{\psi^{(l_1)}} I_t^{(1)})(j_1, \mathbf{m}_1) & \cdots & (\mathcal{W}_{\psi^{(l_1)}} I_t^{(6)})(j_1, \mathbf{m}_1) \\ \vdots & \ddots & \vdots \\ (\mathcal{W}_{\psi^{(l_N)}} I_t^{(1)})(j_N, \mathbf{m}_N) & \cdots & (\mathcal{W}_{\psi^{(l_N)}} I_t^{(6)})(j_N, \mathbf{m}_N) \end{bmatrix} \in \mathbb{R}^{N \times 6}. \quad (3.69)$$

Each column of $\mathbf{L}_{\mathbf{s}_w}$ represents the wavelet transform of a two-dimensional image $I_t^{(i)}$ with respect to the system of wavelets defined by the wavelet generators $\{\psi^{(l)}\}_{l \leq L}$, and the sequence $\{(l_n, j_n, \mathbf{m}_n)\}_{n \leq N}$. Constructing the interaction matrix $\mathbf{L}_{\mathbf{s}_w}$ thus corresponds to computing the image gradient of I_t and performing six discrete wavelet transforms. While the fast wavelet transform (FWT) is indeed a linear time algorithm, this might still be too much in time-critical applications. One possibility to limit the computational effort is to reduce the number of considered scales j . Another approach, which reduces the number of necessary discrete wavelet transforms from six to two, is given by only considering an approximation of $\mathbf{L}_{\mathbf{s}_w}$.

Let us assume that all generator functions $\psi^{(l)}$ are compactly supported in the spatial domain with their support being centered around the origin. It was already noted in [13] that in such a setting, the velocities $\left(\frac{dx}{dt}, \frac{dy}{dt}\right)$ in the optical flow constraint equation (3.58) remain approximately constant on the support set of a wavelet $\psi_{j, \mathbf{m}}^{(l)}$. Formally, this assumption is equivalent with replacing $\mathbf{L}_p(x, y)$ inside the integral in (3.67) with a matrix

$\mathbf{L}_{\text{pw}}(j, \mathbf{m})$, that only depends on the center of the support set of the wavelet $\psi_{j, \mathbf{m}}^{(l)}$ and is thus fixed for any pair of parameters $(j, \mathbf{m}) \in \mathbb{N} \times \mathbb{Z}^2$. Therefore, $\mathbf{L}_{\text{pw}}(j, \mathbf{m})$ can simply be computed by evaluating \mathbf{L}_{p} at the center of the associated wavelet:

$$\mathbf{L}_{\text{pw}}(j, \mathbf{m}) = \mathbf{L}_{\text{p}}(2^{-j}m_1, 2^{-j}m_2). \quad (3.70)$$

By using (3.70), an approximation $\tilde{\mathbf{L}}_{\text{sw}}(t) \approx \mathbf{L}_{\text{sw}}(t)$ of the wavelet-based interaction matrix at a time t that, instead of six, only requires the computation of two discrete wavelet transforms is thus given by:

$$\tilde{\mathbf{L}}_{\text{sw}}(t) = - \begin{bmatrix} \left[(\mathcal{W}_{\psi^{(l_1)}} \frac{\partial I_t}{\partial x})(j_1, \mathbf{m}_1), (\mathcal{W}_{\psi^{(l_1)}} \frac{\partial I_t}{\partial y})(j_1, \mathbf{m}_1) \right] \cdot \mathbf{L}_{\text{pw}}(j_1, \mathbf{m}_1) \\ \vdots \\ \left[(\mathcal{W}_{\psi^{(l_N)}} \frac{\partial I_t}{\partial x})(j_N, \mathbf{m}_N), (\mathcal{W}_{\psi^{(l_N)}} \frac{\partial I_t}{\partial y})(j_N, \mathbf{m}_N) \right] \cdot \mathbf{L}_{\text{pw}}(j_N, \mathbf{m}_N) \end{bmatrix} \in \mathbb{R}^{N \times 6}. \quad (3.71)$$

Let us now denote the visual error with respect to the wavelet feature vector by

$$\mathbf{e}_{\text{w}}(t) = \mathbf{s}_{\text{w}}(t) - \mathbf{s}_{\text{w}}^*. \quad (3.72)$$

In order to minimize \mathbf{e}_{w} , we can again apply the Levenberg-Marquardt approach

$$\mathbf{v}(t) = -\lambda \left(\mathbf{H}(t) + \mu \text{diag}(\mathbf{H}(t)) \right)^{-1} \left(\mathbf{L}_{\text{sw}}(t) \right)^{\top} \mathbf{e}_{\text{w}}(t), \quad (3.73)$$

where $\lambda > 0$ is a controller gain parameter, $\mu > 0$ a damping factor, and $\mathbf{H}(t) = \left(\mathbf{L}_{\text{sw}}(t) \right)^{\top} \mathbf{L}_{\text{sw}}(t)$.

The interaction matrix $\tilde{\mathbf{L}}_{\text{sw}}$ is based on wavelet transforms of the partial derivatives of the two-dimensional image I_t . However, by applying integration by parts, $\tilde{\mathbf{L}}_{\text{sw}}$ could also be computed by considering the partial derivatives of the wavelets $\psi_{j, \mathbf{m}}^{(l)}$. While this approach is of limited practical value, it has some interesting theoretical aspects and is thus briefly described in the upcoming Section 3.2.1.

3.2.1 Computing the Interaction Matrix via Derivative Wavelets

Using integration by parts, we can move the partial derivatives in (3.71) from the two-dimensional image I_t to the respective wavelet and write:

$$\left\langle \frac{\partial I_t}{\partial x}, \psi_{j, \mathbf{m}}^{(l)} \right\rangle_{L^2} = \left\langle I_t, \frac{\partial \psi_{j, \mathbf{m}}^{(l)}}{\partial x} \right\rangle_{L^2}, \quad \text{and} \quad \left\langle \frac{\partial I_t}{\partial y}, \psi_{j, \mathbf{m}}^{(l)} \right\rangle_{L^2} = \left\langle I_t, \frac{\partial \psi_{j, \mathbf{m}}^{(l)}}{\partial y} \right\rangle_{L^2}. \quad (3.74)$$

This means that instead of computing the wavelet transform of the image gradient ∇I_t , one can compute the wavelet transform of I_t associated to the partial derivatives of the wavelet generators $\psi^{(l)}$. In particular, for separable two-dimensional wavelet generators, the partial derivatives can easily be computed by considering the derivatives of the corresponding one-dimensional wavelet and scaling functions (cf. Equations (3.28) to (3.31)).

This approach of obtaining wavelet transforms that are based on the derivatives of wavelets was already investigated in [13] with the goal of defining an efficient multiscale transform-based framework for computing optical flows. In particular, it was shown that for a sufficiently smooth one-dimensional scaling function $\varphi \in L^2(\mathbb{R})$, and an associated one-dimensional wavelet $\psi \in L^2(\mathbb{R})$, that are defined by a pair of finite impulse response filters $\mathbf{h}, \mathbf{g} \subset \ell^2(\mathbb{Z})$ (cf. Equations (3.13) and (3.16)) and associated trigonometric polynomials m_0 and m_1 (cf. Equations (3.14) and (3.17)), an MRA can be constructed that is based on a wavelet $\psi^{\text{der}} \in L^2(\mathbb{R})$, and a scaling function $\varphi^{\text{der}} \in L^2(\mathbb{R})$ that satisfy the following relationships:

$$\psi^{\text{der}} = \psi', \quad (3.75)$$

$$\varphi^{\text{der}}(\cdot + 1) - \varphi^{\text{der}} = \varphi'. \quad (3.76)$$

In [13], this was achieved by computing the derivatives of ψ and φ in the frequency domain via

$$\widehat{f}'(\xi) = 2\pi i \xi \widehat{f}(\xi), \quad (3.77)$$

where f is differentiable and $f', f \in L^1(\mathbb{R})$, and considering the identity

$$2\pi i \xi = (e^{2\pi i \xi} - 1) \left(\prod_{j=1}^{\infty} \frac{e^{2^{-j+1}\pi i \xi} + 1}{2} \right)^{-1}, \quad \xi \in \mathbb{R}. \quad (3.78)$$

Eventually, it was shown that the wavelet ψ^{der} and the scaling function φ^{der} can be defined by the trigonometric polynomials

$$m_0^{\text{der}}(\xi) = \frac{2m_0(\xi)}{e^{2\pi i \xi} + 1}, \quad (3.79)$$

$$m_1^{\text{der}}(\xi) = \frac{e^{2\pi i \xi} - 1}{2} m_1(\xi), \quad (3.80)$$

in the sense that

$$\widehat{\varphi}^{\text{der}}(\xi) = \prod_{j=1}^{\infty} m_0^{\text{der}}(2^{-j}\xi), \quad (3.81)$$

$$\widehat{\psi}^{\text{der}}(\xi) = m_1^{\text{der}}(2^{-1}\xi)\widehat{\varphi}^{\text{der}}(2^{-1}\xi). \quad (3.82)$$

The following theorem, which is not part of the original work presented in [13], provides a simple rule for constructing the filters \mathbf{h}^{der} and \mathbf{g}^{der} associated with the trigonometric polynomials defined in Equations (3.79) and (3.80).

Theorem 3.2.1. *Let ψ be a wavelet and φ be a scaling function defined by a low-pass filter \mathbf{h} , and a high-pass filter \mathbf{g} via Equations (3.19), (3.20), (3.14), and (3.17). Then, the derivative wavelet ψ^{der} and the scaling function φ^{der} satisfying (3.75) and (3.76) are defined by finite impulse response filters that satisfy*

$$h_n^{\text{der}} = 2h_n - h_{n+1}^{\text{der}}, \quad (3.83)$$

$$g_n^{\text{der}} = \frac{g_{n+1} - g_n}{2}, \quad (3.84)$$

and the associated trigonometric polynomials

$$m_0^{\text{der}}(\xi) = 2^{-\frac{1}{2}} \sum_{n \in \mathbb{Z}} h_n^{\text{der}} e^{-2\pi i n \xi}, \quad (3.85)$$

$$m_1^{\text{der}}(\xi) = 2^{-\frac{1}{2}} \sum_{n \in \mathbb{Z}} g_n^{\text{der}} e^{-2\pi i n \xi}, \quad (3.86)$$

via Equations (3.81) and (3.82).

Proof. We use the definitions of m_0 and m_0^{der} as well as (3.79) and compute

$$m_0(\xi) = 2^{-\frac{1}{2}} \sum_{n \in \mathbb{Z}} h_n e^{-2\pi i n \xi} = \frac{1}{2} \left(e^{2\pi i \xi} + 1 \right) m_0^{\text{der}}(\xi) \quad (3.87)$$

$$= 2^{-\frac{3}{2}} \left(\sum_{n \in \mathbb{Z}} h_n^{\text{der}} e^{-2\pi i (n-1)\xi} + \sum_{n \in \mathbb{Z}} h_n^{\text{der}} e^{-2\pi i n \xi} \right) \quad (3.88)$$

$$= 2^{-\frac{1}{2}} \sum_{n \in \mathbb{Z}} \frac{h_{n+1}^{\text{der}} + h_n^{\text{der}}}{2} e^{-2\pi i n \xi}. \quad (3.89)$$

For \mathbf{g}^{der} , the definitions of the trigonometric polynomials m_1 and m_1^{der} , combined with (3.80) yield

$$m_1^{\text{der}}(\xi) = 2^{-\frac{1}{2}} \sum_{n \in \mathbb{Z}} g_n^{\text{der}} e^{-2\pi i n \xi} = \frac{1}{2} \left(e^{2\pi i \xi} - 1 \right) m_1(\xi) \quad (3.90)$$

$$= 2^{-\frac{3}{2}} \left(\sum_{n \in \mathbb{Z}} g_n e^{-2\pi i (n-1)\xi} - \sum_{n \in \mathbb{Z}} g_n e^{-2\pi i n \xi} \right) \quad (3.91)$$

$$= 2^{-\frac{1}{2}} \sum_{n \in \mathbb{Z}} \frac{g_{n+1} - g_n}{2} e^{-2\pi i n \xi}. \quad (3.92)$$

□

Theorem 3.2.1 provides an analytically interesting perspective on the entries of the matrix $\tilde{\mathbf{L}}_{s_w}$ as coefficients of an MRA-based representation of the image I_t . However, it does unfortunately not yield a straightforward method to compute $\tilde{\mathbf{L}}_{s_w}$ by simply applying the filters \mathbf{h}^{der} and \mathbf{g}^{der} . While ψ' is in fact equal to the derivative wavelet ψ^{der} , the derivative of the scaling function needs to be computed via $\varphi^{\text{der}}(\cdot + 1) - \varphi^{\text{der}} = \varphi'$ (cf. Equations (3.75) and (3.76)). Furthermore, the usual assumption when representing a digital image with an MRA-based transform is that the observed pixel values correspond to coefficients defining a coarse approximation with respect to the scaling function φ (cf. Equation (3.35)). However, when computing the coefficients of the derivative wavelet-based MRA, we would already assume that the pixel values correspond to a coarse approximation with respect to φ^{der} .

3.3 Shearlet-Based Visual Servoing

Shearlet-based image representations can be used in visual servoing tasks analogously to the wavelet-based approach described in Section 3.2. For a finite set of $L \in \mathbb{N}$ shearlet generators $\{\psi^{(l)}\}_{l \leq L} \subset L^2(\mathbb{R}^2)$, the shearlet-based feature vector for a sequence of $N \in \mathbb{N}$ quadruples $\{(l_n, j_n, k_n, \mathbf{m}_n)\}_{n \leq N} \subset \{1, \dots, L\} \times \mathbb{N} \times \mathbb{Z} \times \mathbb{Z}^2$, where l_n specifies a shearlet generator, j_n defines the scale, k_n the shear operator, and \mathbf{m}_n the translation parameter, is given as a function of t by

$$\mathbf{s}_{\text{sh}}(t) = \left[(\mathcal{SH}_{\psi^{(l_1)}} I_t)(j_1, k_1, \mathbf{m}_1), \dots, (\mathcal{SH}_{\psi^{(l_N)}} I_t)(j_N, k_N, \mathbf{m}_N) \right]^\top, \quad (3.93)$$

where the shearlet transform \mathcal{SH} is defined as in (3.44).

Retaining the notation from Section 3.2 and after repeating the computations (3.66) to (3.68) for the shearlet case, the interaction matrix associated with the shearlet-based feature vector \mathbf{s}_{sh} at a time t can be written as:

$$\mathbf{L}_{\mathbf{s}_{\text{sh}}}(t) = - \begin{bmatrix} (\mathcal{SH}_{\psi^{(l_1)}} I_t^{(1)})(j_1, k_1, \mathbf{m}_1) & \cdots & (\mathcal{SH}_{\psi^{(l_1)}} I_t^{(6)})(j_1, k_1, \mathbf{m}_1) \\ \vdots & \ddots & \vdots \\ (\mathcal{SH}_{\psi^{(l_N)}} I_t^{(1)})(j_N, k_N, \mathbf{m}_N) & \cdots & (\mathcal{SH}_{\psi^{(l_N)}} I_t^{(6)})(j_N, k_N, \mathbf{m}_N) \end{bmatrix} \in \mathbb{R}^{N \times 6}. \quad (3.94)$$

The main formal difference between shearlet- and wavelet-based visual servoing control laws can be found in the definition of the approximated interaction matrix $\tilde{\mathbf{L}}_{s_{\text{sh}}}$. This is

Table 3.1: Overview of the four different wavelet- and shearlet-based control laws considered in the numerical and experimental validation.

	<i>s-wavelet</i>	<i>s-shearlet</i>	<i>ns-wavelet</i>	<i>ns-wavelet</i>
Length of feature vector $\mathbf{s}_w/\mathbf{s}_{sh}$	20059	24576	262144	327680
Number of scales in the transform	3	4	4	4
Scales j used for $\mathbf{s}_w/\mathbf{s}_{sh}$	$j \in \{0, 1\}$	$j \in \{0, 1\}$	only $j = 0$	only $j = 0$
Duration of one iteration	100 ms	140 ms	700 ms	800 ms

due to the fact that the center of a shearlet is not only determined by the scale parameter j , but also depends on the shear parameter k . Under the assumption that the shearlet generators are compactly supported in the time-domain with their supports being centered around the origin, we can write

$$\mathbf{L}_{psh}(j, k, \mathbf{m}) = \mathbf{L}_p\left(2^{-j}(m_1 - km_2), 2^{-j/2}m_2\right), \quad (3.95)$$

if the corresponding generator belongs to the horizontal frequency cones respectively

$$\mathbf{L}_{psh}(j, k, \mathbf{m}) = \mathbf{L}_p\left(2^{-j/2}m_1, 2^{-j}(m_2 - km_1)\right), \quad (3.96)$$

if the generator belongs to the vertical frequency cones. The approximated shearlet-based interaction matrix $\tilde{\mathbf{L}}_{s_{sh}}(t) \approx \mathbf{L}_{s_{sh}}(t)$ is now given by

$$\tilde{\mathbf{L}}_{s_{sh}}(t) = - \begin{bmatrix} \left[(\mathcal{SH}_{\psi(t_1)} \frac{\partial I_t}{\partial x})(j_1, k_1, \mathbf{m}_1), (\mathcal{SH}_{\psi(t_1)} \frac{\partial I_t}{\partial y})(j_1, k_1, \mathbf{m}_1) \right] \cdot \mathbf{L}_{psh}(j_1, k_1, \mathbf{m}_1) \\ \vdots \\ \left[(\mathcal{SH}_{\psi(t_N)} \frac{\partial I_t}{\partial x})(j_N, k_N, \mathbf{m}_N), (\mathcal{SH}_{\psi(t_N)} \frac{\partial I_t}{\partial y})(j_N, k_N, \mathbf{m}_N) \right] \cdot \mathbf{L}_{psh}(j_N, k_N, \mathbf{m}_N) \end{bmatrix} \in \mathbb{R}^{N \times 6}. \quad (3.97)$$

The visual error with respect to the shearlet-based feature vector

$$\mathbf{e}_{sh}(t) = \mathbf{s}_{sh}(t) - \mathbf{s}_{sh}^*, \quad (3.98)$$

can again be minimized by applying a Levenberg-Marquardt optimization

$$\mathbf{v}(t) = -\lambda \left(\mathbf{H}(t) + \mu \text{diag}(\mathbf{H}(t)) \right)^{-1} \left(\mathbf{L}_{s_{sh}}(t) \right)^\top \mathbf{e}_{sh}(t), \quad (3.99)$$

where $\lambda > 0$ is a gain parameter, $\mu > 0$ a damping factor and $\mathbf{H}(t) = \left(\mathbf{L}_{s_{sh}}(t) \right)^\top \mathbf{L}_{s_{sh}}(t)$.

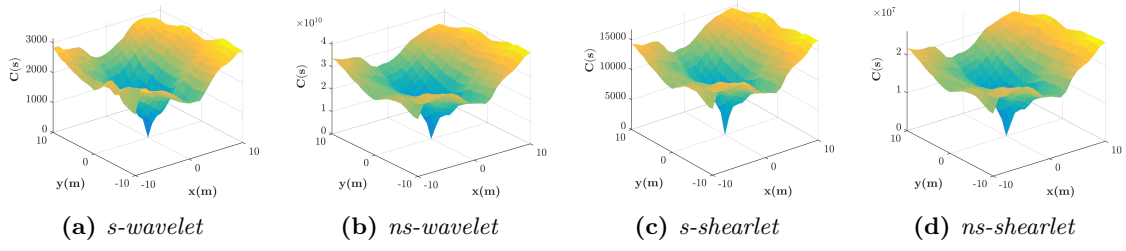


Figure 3.6: Visualization of the cost functions associated with wavelet- and shearlet-based visual servoing control laws. The value $\|\mathbf{s} - \mathbf{s}^*\|_2$ is plotted against linear translations in x - and y -direction.

3.4 Numerical and Experimental Validation

To investigate the applicability of the proposed wavelet- and shearlet-based control laws in both favorable and unfavorable conditions, we performed a series of experiments and numerical simulations. The author of this thesis contributed initial MATLAB scripts that implement the wavelet- and shearlet-based optimization procedures (3.73) and (3.99) but the experiments discussed in this section were primarily designed and conducted by Lesley-Ann Dufлот.

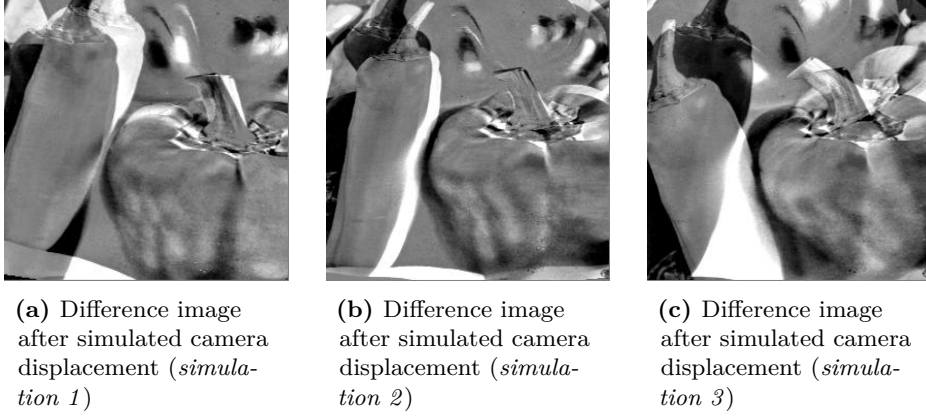
For the numerical and experimental validation, we consider wavelet- and shearlet-based features vectors \mathbf{s}_w and \mathbf{s}_{sh} (cf. Equations (3.64) and (3.93)) that are based on both subsampled and non-subsampled discrete-time transforms. The respective wavelet transforms are computed for two-dimensional digital images via the MATLAB Wavelet Toolbox¹. For non-subsampled shearlet-based transforms, we use ShearLab 3D² [107], and for subsampled shearlet transforms ShearLab 1.1² [109]. The complete visual servoing framework was finally implemented within the Visual Servoing Platform (ViSP)³ C++ library, which also facilitated a simple means of comparing the newly proposed control laws with the classical photometric approach. For the remainder of this section, we use *s-wavelet* and *ns-wavelet* to denote the implementation of the control law (3.73) that is based on a subsampled, respectively non-subsampled, wavelet transform and *s-shearlet*, respectively *ns-shearlet*, to denote the implementation of the control law (3.99) based on subsampled and non-subsampled shearlet transforms.

Table 3.1 provides an overview of the main properties of the interaction matrices associated with the four considered control laws. In all cases, we consider interaction matrices of the form \mathbf{L}_{s_w} and $\mathbf{L}_{s_{sh}}$ (cf. Equations (3.69) and (3.94)) rather than the

¹ <https://en.mathworks.com/products/wavelet.html>

² <http://www.shearlab.org/software>

³ <https://visp.inria.fr/>



	Linear displacement			Angular displacement		
	x	y	z	ω_x	ω_y	ω_z
<i>simulation 1</i>	10 mm	−10 mm	100 mm	5 °	−5 °	10 °
<i>simulation 2</i>	5 mm	−5 mm	100 mm	−20 °	10 °	−5 °
<i>simulation 3</i>	−2 mm	2 mm	−50 mm	5 °	10 °	−30 °

The pixel length is ≈ 1.2 mm.

(d) Initial camera pose displacements in the simulated experiments.

Figure 3.7: Overview of the three different camera displacements considered in our numerical simulations.

approximations $\tilde{\mathbf{L}}_{s_w}$ and $\tilde{\mathbf{L}}_{s_{sh}}$ (cf. Equations (3.71) and (3.97)). Note that due to the high redundancy of the non-subsampled transforms, besides the associated coarse approximation, we only consider the detail coefficients from a single scale for the feature vectors in the case of *ns-wavelet* and *ns-shearlet*. Examples of the feature vectors used in the implemented control laws *s-wavelet*, *ns-wavelet*, *s-shearlet*, and *ns-shearlet* can be found in Appendix B in Figures B.1(b), B.3, B.2, and B.4, respectively.

For all simulations and experiments, the depth Z in the definition of the interaction matrix \mathbf{L}_p (cf. Equation 3.60) was estimated manually and assumed to remain constant during the visual servoing process. To increase the domain of convergence, the parameters of the Levenberg-Marquardt method λ and μ (cf. Equation (3.73)) were chosen adaptively as functions of the ℓ^2 -norm of the currently observed error \mathbf{e}_w , respectively \mathbf{e}_{sh} . Examples of the behavior of the associated cost functions $\|\mathbf{s}_w - \mathbf{s}_w^*\|_2$, and $\|\mathbf{s}_{sh} - \mathbf{s}_{sh}^*\|_2$ with respect to linear translations in the image plane are plotted for all of the four considered control laws in Figure 3.6.

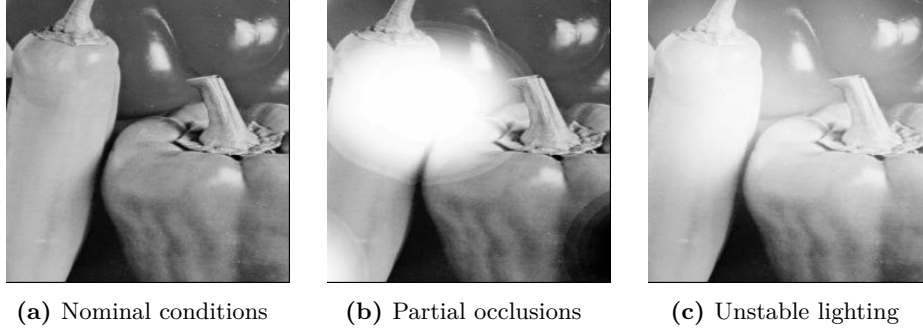


Figure 3.8: The three different conditions considered in the numerical simulations.

3.4.1 Numerical Simulations

We first consider numerical simulations of positioning tasks in which the simulated camera is displaced by an element in $SE(3)$ after saving a set of desired visual features \mathbf{s}_{ph}^* , \mathbf{s}_w^* , or \mathbf{s}_{sh}^* , respectively, that are associated with the initial camera pose. The respective control laws are then applied with the goal of recovering the original pose. A main advantage of working with a simulated environment is that it provides reliable testing conditions in the sense that it is independent of factors such as the calibration of the visual sensor, mechanical properties of the robot, or possible measurement errors. In all simulated experiments, the newly proposed wavelet- and shearlet-based control laws are also compared to the classical photometric approach. Note that the pixel length in the images used in our simulations is considered to be approximately 1.2 mm. In total, we use three different camera displacements, which will subsequently be denoted as *simulation 1*, *simulation 2*, and *simulation 3* (see Figure 3.7). To further test for stability in unfavorable conditions, we also consider the case of small occlusions and variations in illumination (see Figure 3.8).

Table 3.2 reports the final positioning errors of the simulated camera in nominal conditions (i.e., without external influences such as occlusions) for all three considered displacement vectors. In this setting, all of the five evaluated control laws successfully converge at the initial camera poses with only small deviations.

The final positioning errors when the visual servoing process is impeded by partial occlusions are compiled in Table 3.3. In this setting, roughly a fifth of the initially observed image remains occluded for the duration of the positioning process (cf. Figure 3.8(b)). In the case of *simulation 1*, all five control laws also seem to successfully converge towards the desired pose with respect to the two-dimensional plane parallel to the observed image. However, all control laws also seem to have difficulties with exactly recovering the correct position with respect to the z -axis and the angular displacement around the x - and y -axes.

Table 3.2: Final camera pose displacements yielded by the five considered control laws in simulated positioning tasks in nominal conditions.

	x	y	z	ω_x	ω_y	ω_z
<i>simulation 1</i>						
initial displacement	10 mm	−10 mm	100 mm	5 °	−5 °	10 °
photometry	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	−10 ^{−4}	10 ^{−4}
<i>s-wavelet</i>	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	−10 ^{−4}	10 ^{−4}
<i>s-shearlet</i>	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	−10 ^{−4}	10 ^{−4}
<i>ns-wavelet</i>	10 ^{−4}	10 ^{−4}	10 ^{−3}	10 ^{−4}	−10 ^{−3}	10 ^{−4}
<i>ns-wavelet</i>	10 ^{−4}	10 ^{−4}	10 ^{−3}	10 ^{−3}	−10 ^{−4}	10 ^{−4}
<i>simulation 2</i>						
initial displacement	5 mm	−5 mm	100 mm	−20 °	10 °	−5 °
photometry	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	−10 ^{−4}	10 ^{−4}
<i>s-wavelet</i>	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}
<i>s-shearlet</i>	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}
<i>ns-wavelet</i>	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−3}	10 ^{−4}	−10 ^{−4}
<i>ns-wavelet</i>	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}
<i>simulation 3</i>						
initial displacement	−2 mm	2 mm	−50 mm	5 °	10 °	−30 °
photometry	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}
<i>s-wavelet</i>	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}
<i>s-shearlet</i>	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}
<i>ns-wavelet</i>	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}
<i>ns-wavelet</i>	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}	10 ^{−4}

x, y, z : Translational displacement along the respective axis.

$\omega_x, \omega_y, \omega_z$: Angular displacement around the respective axis.

A similar behavior can be observed in the case of *simulation 2*, although it should be noted that both in *simulation 1* as well as in *simulation 2*, the *s-wavelet* control law is significantly better at recovering the displacement along the z -axis. When considering *simulation 3*, which is characterized by a rather severe angular displacement around the z -axis, all of the five control laws completely fail to perform the image registration task in the presence of a partial occlusion.

The results of the simulated positioning task in the case of unstable lighting conditions can be found in Table 3.4. Here, the initial feature vectors \mathbf{s}_{ph}^* , \mathbf{s}_{w}^* , and \mathbf{s}_{sh}^* , are obtained in the presence of an additional light source (cf. Figure 3.8(c)), while the visual servoing process after displacing the simulated camera is carried out in nominal conditions (cf. Figure 3.8(a)). The results are similar to the case of partial occlusions in the sense that for *simulation 1*, and *simulation 2*, all control laws are mostly successful in recovering

Table 3.3: Final camera pose displacements yielded by the five considered control laws in simulated positioning tasks impeded by partial occlusions.

	x	y	z	ω_x	ω_y	ω_z
<i>simulation 1</i>						
initial displacement	10 mm	−10 mm	100 mm	5 °	−5 °	10 °
photometry	0.69	0.01	−4.61	4.30	−5.64	0.05
<i>s-wavelet</i>	0.56	−0.16	0.52	4.43	−5.60	0.17
<i>s-shearlet</i>	0.56	−0.16	−1.94	4.31	−5.51	0.08
<i>ns-wavelet</i>	0.78	−0.14	−2.79	4.30	−5.60	0.15
<i>ns-wavelet</i>	0.74	−0.14	−2.74	4.33	−5.54	0.13
<i>simulation 2</i>						
initial displacement	5 mm	−5 mm	100 mm	−20 °	10 °	−5 °
photometry	1.18	−0.24	−42.84	−20.29	9.14	−1.50
<i>s-wavelet</i>	0.55	2.06	−11.06	−20.30	9.17	−1.27
<i>s-shearlet</i>	0.52	1.18	−28.98	−20.40	9.18	−1.50
<i>ns-wavelet</i>	1.10	0.10	−31.24	−20.37	9.18	−1.43
<i>ns-wavelet</i>	1.04	0.16	−31.04	−20.36	9.22	1.47
<i>simulation 3</i>						
initial displacement	−2 mm	2 mm	−50 mm	5 °	10 °	−30 °
photometry	✗	✗	✗	✗	✗	✗
<i>s-wavelet</i>	✗	✗	✗	✗	✗	✗
<i>s-shearlet</i>	✗	✗	✗	✗	✗	✗
<i>ns-wavelet</i>	✗	✗	✗	✗	✗	✗
<i>ns-wavelet</i>	✗	✗	✗	✗	✗	✗

x, y, z : Translational displacement along the respective axis.

$\omega_x, \omega_y, \omega_z$: Angular displacement around the respective axis.

✗ = fail.

the correct position with respect to the x - and y -axes as well as the angular displacement around the z -axis, with the exception of the photometric control law, which completely fails in the case of *simulation 2*. A possible explanation for the fact that in this specific case, the classical photometric approach is outperformed by the newly proposed wavelet- and shearlet-based control laws is that the detail coefficients obtained in the respective transforms are in fact invariant to changes in the mean (cf. Equation 3.4) and should thus more stable in this setting. However, in the case of *simulation 3*, again all of the five control laws fail to successfully converge to the initial camera pose in the presence of a large angular displacement around the z -axis.

Table 3.4: Final camera pose displacements yielded by the five considered control laws in simulated positioning tasks impeded by unstable lighting conditions.

	x	y	z	ω_x	ω_y	ω_z
<i>simulation 1</i>						
initial displacement	10 mm	−10 mm	100 mm	5 °	−5 °	10 °
photometry	0.84	0.57	−31.49	4.27	−5.61	−0.26
<i>s-wavelet</i>	0.10	0.69	−23.48	4.43	−5.65	−0.10
<i>s-shearlet</i>	0.52	0.49	−13.57	4.26	−5.60	−0.07
<i>ns-wavelet</i>	0.83	0.60	−23.66	4.26	−5.65	−0.22
<i>ns-wavelet</i>	0.78	0.59	−22.76	4.27	−5.62	−0.21
<i>simulation 2</i>						
initial displacement	5 mm	−5 mm	100 mm	−20 °	10 °	−5 °
photometry	✗	✗	✗	✗	✗	✗
<i>s-wavelet</i>	0.34	0.66	−59.68	−20.27	9.25	−2.40
<i>s-shearlet</i>	0.80	0.89	−52.28	−20.31	9.28	−2.12
<i>ns-wavelet</i>	1.34	0.456	−60.99	−20.32	9.33	−2.20
<i>ns-wavelet</i>	1.20	0.52	−59.19	−20.28	9.35	−2.22
<i>simulation 3</i>						
initial displacement	−2 mm	2 mm	−50 mm	5 °	10 °	−30 °
photometry	✗	✗	✗	✗	✗	✗
<i>s-wavelet</i>	✗	✗	✗	✗	✗	✗
<i>s-shearlet</i>	✗	✗	✗	✗	✗	✗
<i>ns-wavelet</i>	✗	✗	✗	✗	✗	✗
<i>ns-wavelet</i>	✗	✗	✗	✗	✗	✗

x, y, z : Translational displacement along the respective axis.

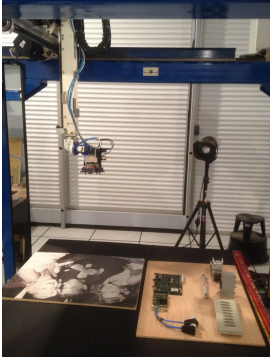
$\omega_x, \omega_y, \omega_z$: Angular displacement around the respective axis.

✗ = fail.

3.4.2 Experimental Validation

For the experimental validation of the newly proposed wavelet- and shearlet-based control laws, we used a CCD camera which was mounted on a robotic system with six degrees of freedom. The corresponding experimental setup is shown in Figure 3.9(a). The camera yielded images with a resolution of 450×450 pixels, which were resized to a resolution of 256×256 pixels to decrease the computational cost of the considered control laws, at a frame rate of 25 frames per second. A personal computer clocked at 2.4 GHz was used to execute the implementations of the wavelet- and shearlet-based control laws and to communicate with the robotic system and the mounted CCD camera.

As in the numerical simulations discussed in Section 3.4.1, the experimental evaluation was carried out by using the considered visual servoing control laws to recover the initial



(a) Experimental setup

	Linear displacement			Ang. displacement		
	x	y	z	ω_x	ω_y	ω_z
<i>experiment 1</i>	50 mm	50 mm	100 mm	5°	-5°	-4°
<i>experiment 2</i>	-20 mm	-30 mm	-20 mm	10°	2°	2°

(b) Initial camera pose displacements considered in the experimental validation.

Figure 3.9: Overview of the two different camera displacements considered in the experimental evaluation of the wavelet- and shearlet-based control laws.

pose of the displaced camera by considering reference feature vectors \mathbf{s}_w^* and \mathbf{s}_{sh}^* that were obtained at the original position. The two considered displacements, which can be described by elements in $SE(3)$ and will subsequently be denoted as *experiment 1*, and *experiment 2*, can be found in Table 3.9(b). In order to keep this section concise, only the results with respect to *experiment 1* are directly reported here. The detailed results of the respective wavelet- and shearlet-based control laws when considering the initial displacement defined by *experiment 2* can be found in Appendix C. Again, the positioning tasks were carried out in both favorable and unfavorable conditions, where in the latter case, the visual servoing process was impeded by either partial occlusions or unstable lighting conditions.

Due to the high computational demand associated with the large feature vectors yielded by non-subsampled wavelet- and shearlet-based transforms (cf. Table 3.1), we only considered the control laws *s-wavelet* and *s-shearlet*, which are based on subsampled transforms. Throughout all experiments, the depth parameter Z (cf. Equation (3.60)) was assumed to remain constant and defined as $Z = 0.8$ m, that is, approximately equal to the distance between the camera and the observed planar scene.

Figures 3.10 and C.1 illustrate the behavior of the control laws *s-wavelet* and *s-shearlet* in the considered positioning tasks in nominal conditions. The development of the translational and angular displacement of the camera over time for both control laws is depicted in Figures 3.10(e), 3.10(f), C.1(e), and C.1(f), respectively. The associated difference images obtained at the respective final iteration are shown in Figures 3.10(c), 3.10(d), C.1(c), and C.1(d). For both of the two considered displacements, both control laws succeed at the positioning task in the sense that after 400 to 700 iterations, the errors with respect to the initial camera pose are vanishingly small.

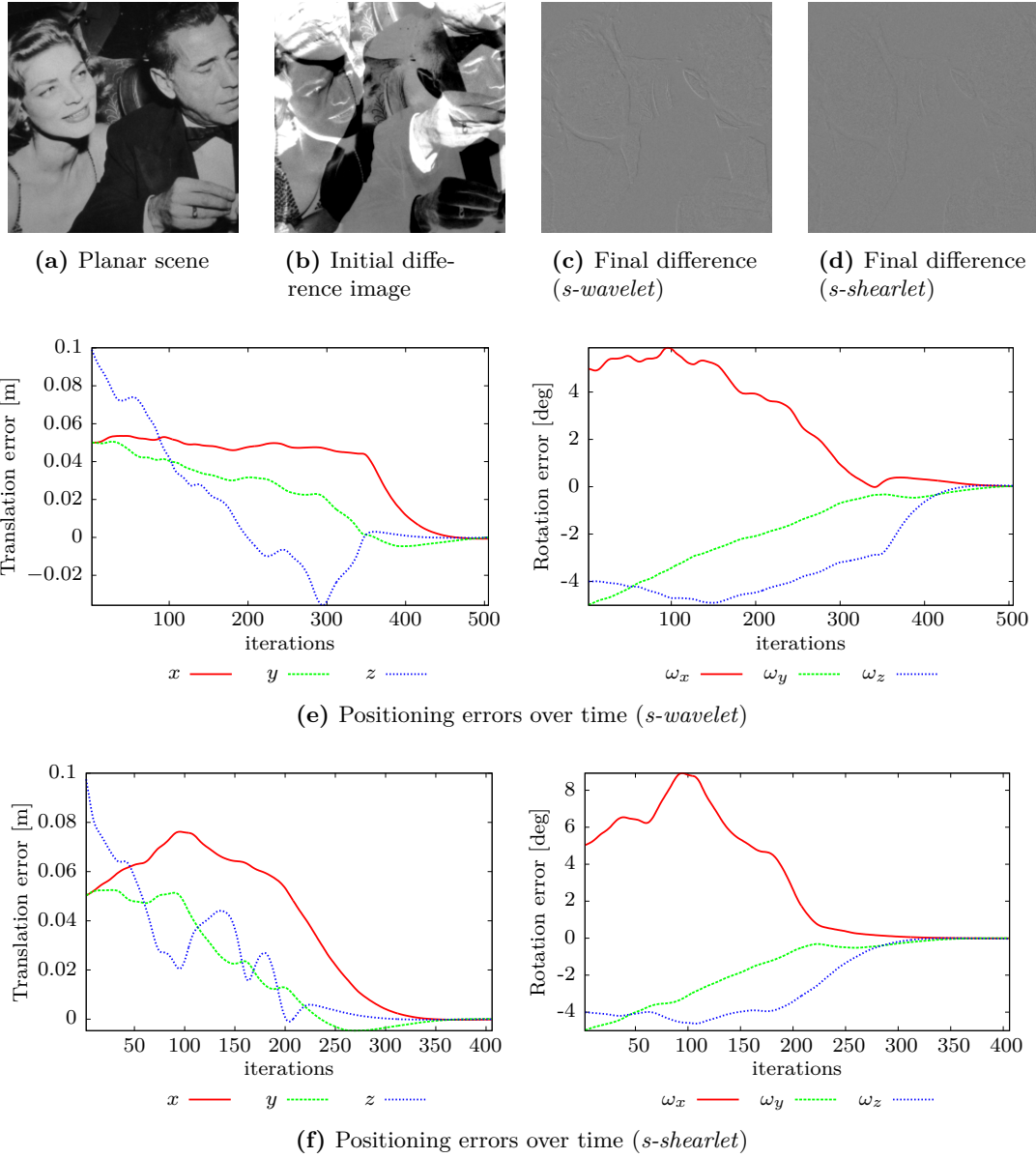


Figure 3.10: Wavelet- and shearlet-based visual servoing in nominal conditions with a planar scene. The initial displacement was chosen according to *experiment 1* (cf. Table 3.9(b)).

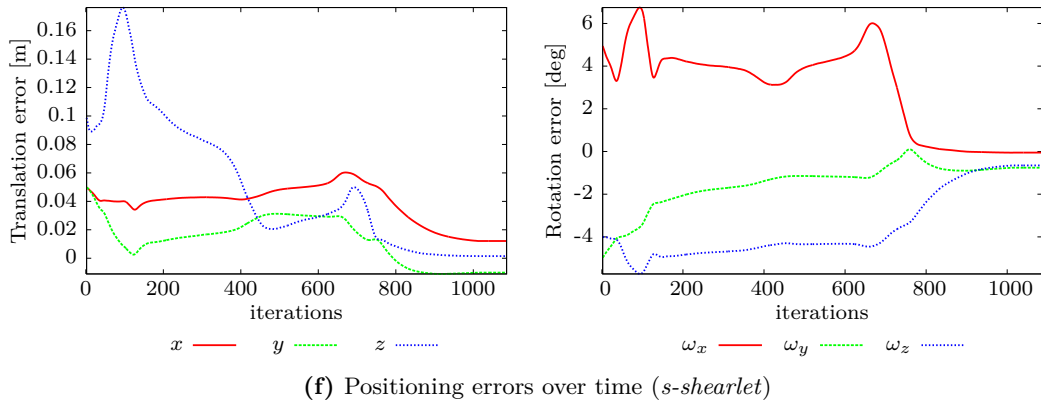
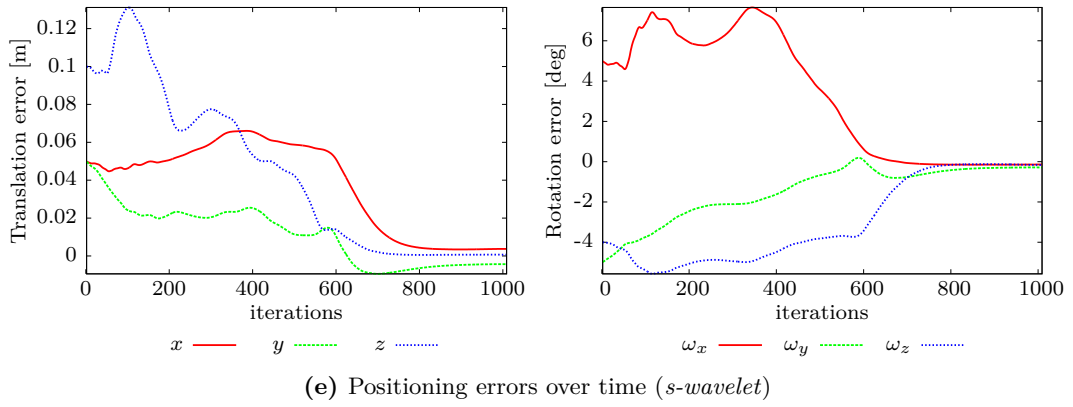
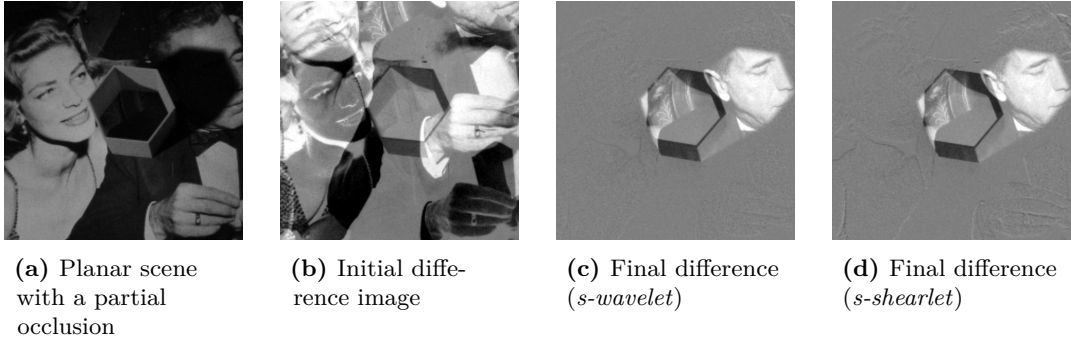


Figure 3.11: Wavelet- and shearlet-based visual servoing with a planar scene that is partially occluded by an object. The initial displacement was chosen according to *experiment 1* (cf. Table 3.9(b)).

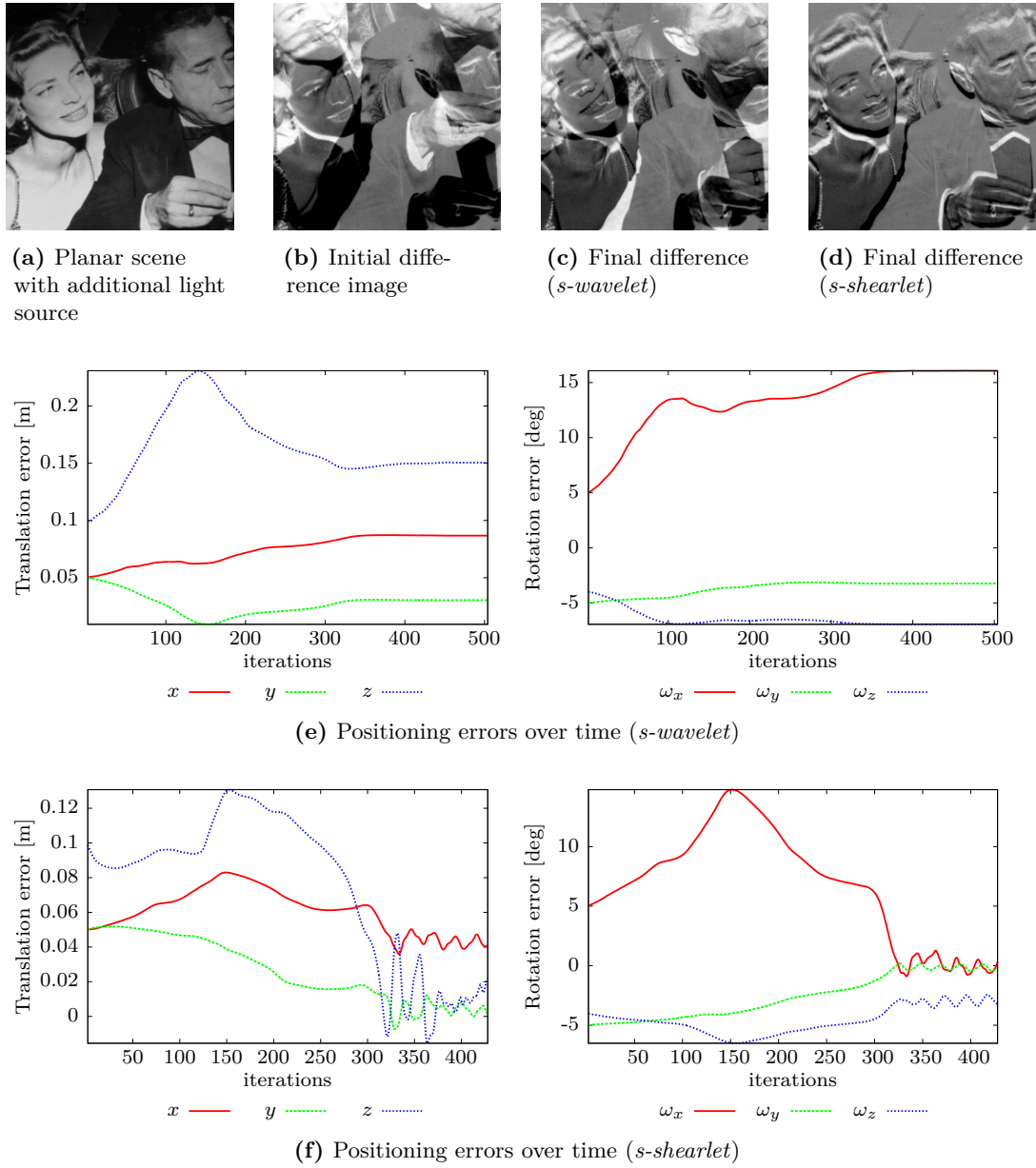


Figure 3.12: Wavelet- and shearlet-based visual servoing with a planar scene and unstable lighting conditions. The initial displacement was chosen according to *experiment 1* (cf. Table 3.9(b)).

Figure 3.11 contains the results of the wavelet- and shearlet-based control laws with respect to *experiment 1* when the visual servoing process is impeded by a partial occlusion that is caused by a three-dimensional object (cf. Figure 3.11). In this scenario, both control laws are again mostly successful at recovering the initial camera pose. It is, however, not entirely clear from the experimental results whether the local minimum of the respective cost function towards the control laws seem to converge exactly coincides with the original camera pose. Furthermore, in the case of *experiment 2*, the shearlet-based control law completely fails at correctly identifying the initial displacement, while the control law based on a subsampled wavelet transform still succeeds (see Figure C.2). A possible explanation for this might be found in the anisotropic nature of shearlet-based transforms. It was already discussed in the prelude to this chapter that, due to the application of an anisotropic scaling matrix (cf. Equation 3.40), shearlet-based transforms are better suited for representing curvilinear features such as edges, than wavelet-based transforms. This also suggests that a shearlet-based feature vector \mathbf{s}_{sh} should indeed be more sensitive to the highly anisotropic structures introduced by the occluding object considered in this scenario, than the wavelet-based feature vector \mathbf{s}_{w} .

Analogous to the simulated experiments, we also tested the control laws *s-wavelet* and *s-shearlet* in an unfavorable environment characterized by unstable lighting conditions. The evolution of the translational and angular positioning errors of the camera in this setting, as well as the eventually obtained difference images, are compiled in Figures 3.12 and C.3. In the simulated experiments, we were able to observe an improvement with respect to the classical photometric approach in this setting (cf. Table 3.4) and conjectured that this could be connected to the insensitivity of the detail coefficients of a wavelet- or shearlet-based transform to changes in the mean. Unfortunately, we were not able to fully reproduce this behavior in the non-simulated experiments. While the final difference images shown in Figures 3.12(d) and C.3(d) indicate that for both *experiment 1* and *experiment 2*, the shearlet-based control law *s-shearlet* is successful at undoing some of the initial displacement, the plots in Figures 3.12(f) and C.3(f) reveal that the camera poses at the respective last iterations still significantly differ from the original position. Furthermore, in both these cases, the positioning error graphs show a highly undesirable oscillatory behavior which indicates that the control law is not successfully converging to a local minimum. The wavelet-based control law *s-wavelet* exhibits a similar behavior in the case of *experiment 2*. In the case of *experiment 1*, on the other hand, *s-wavelet* seems to converge but also fails to recover the initial camera pose. In particular with respect to the observed oscillatory behavior, it is not entirely clear whether these negative results are due to intrinsic properties of the proposed methods or if they could be remedied by appropriately adapting parts of the implementation, such as the choices for the damping

parameter μ and the gain parameter λ in (3.73), respectively (3.99).

Further details regarding the numerical and experimental validation can be found in our original publication [56], where we also evaluated the proposed control laws in a setting where a three-dimensional object was observed rather than a purely planar scene. The online version of [56] furthermore contains video recordings of some of the simulations and experiments described in this section¹.

3.5 Conclusion

The main contribution of this chapter was the analytical derivation of interaction matrices that linearly describe the relationship between the time derivative of wavelet- and shearlet-based feature vectors and the translational and angular velocities of a robotic system with six degrees of freedom, which facilitates the application of wavelet- and shearlet-based control laws in visual servoing tasks. In particular, these control laws can be seen as a best of both worlds approach in the sense that the approximation coefficients of a wavelet- or shearlet-based multiscale image representation can be treated analogous to the case of classical photometric visual servoing, while the detail coefficients are sensitive to geometrical features such as edges and thereby more related to visual servoing schemes that are based on detecting and tracking specific features.

To qualitatively evaluate the accuracy and robustness of the newly proposed visual servoing schemes in both favorable and unfavorable conditions, we performed a series of simulations and experiments involving positioning tasks in which the control laws were applied to recover the original pose of a displaced camera. In nominal conditions without external disturbances, the wavelet- and shearlet-based control laws demonstrated a high accuracy in the simulations (cf. Table 3.2) as well as during the experimental validation (cf. Figures 3.10 and C.1), which indicates their general applicability in visual servoing tasks.

The robustness of the newly proposed control laws to partial occlusions of the observed scene was comparable to the robustness of the classical photometric approach in our simulated experiments (cf. Table 3.3). This is most likely due to the fact that all wavelet- and shearlet-based feature vectors also contained a coarse approximation of the original image (cf. Figures B.1(b), B.2, B.3, and B.4), which is comparable to the feature vector considered in the photometric case and thus equipped with similar stability properties with respect to partial occlusions. In particular, a reasonable degree of robustness to partial occlusions could also be observed in the experiments with a robotic system (see Figures 3.11 and C.2).

¹ <https://doi.org/10.1177/0278364918769739>

When considering unstable lighting conditions, our simulations indicate a slight improvement of the newly proposed control laws over the photometric approach (cf. Table 3.4). We were, however, unsuccessful in reproducing a convincing degree of robustness to unstable lighting in our subsequently conducted real-world experiments (cf. Figures 3.12 and C.3). To understand this behavior, it is important to observe that, due to the vanishing mean property (3.4) of the respective analyzing functions, the detail coefficients of the considered wavelet-, or shearlet-based transforms are invariant to changes in the mean and thus also robust to changes in the overall luminosity of a scene. While this is certainly not the case for the associated approximation coefficients, one would still expect that a visual servoing control law that utilizes wavelet- or shearlet-based feature vectors would thus exhibit a certain degree of robustness to unstable lighting conditions. However, the detail coefficients of the wavelet- or shearlet-based representation of a digital image are typically characterized by a high degree of sparsity (see, e.g., Figures 3.5 and B.1(b)), which could also lead to problems when applying the associated control laws. Broadly speaking, when considering a large displacement between the desired and the current pose, the cost function associated with the detail coefficients of a wavelet- and shearlet-based feature vector will locally be governed by various shallow local minimums and thus not yield a conclusive direction during optimization. This could also partly explain the oscillatory behavior exhibited by the wavelet- and shearlet-based control laws during the experimental validation (cf. Figures 3.12(f), C.3(e), and C.3(f)). One way of taking this into consideration would be to restrict the experimental analysis with respect to unstable lighting to a setting in which only small displacements are being considered, as opposed to the rather severe displacements used in the simulations and experiments discussed in this chapter. Another approach that could lead to improved wavelet- and shearlet-based visual servoing schemes that are truly robust to variations in lighting, even when dealing with large displacements, is to consider transforms that are based on complex-valued generalizations of wavelets and shearlets [74, 97, 142, 157]. The magnitude response of complex-valued wavelet and shearlet transforms is known to exhibit a certain degree of shift invariance, which should lead to smoother cost functions and could help increase the robustness of the proposed methods, especially with respect to the detail coefficients.

Another drawback of the newly proposed control schemes is their comparatively high computational demand, especially when considering non-subsampled transforms. However, it should be noted that for the simulations and experiments considered in this chapter, the computation of wavelet- and shearlet-based coefficients was performed by external MATLAB libraries. We expect that the runtime can be significantly decreased by considering a pure C++ implementation and by using GPU hardware for parallelized computations and that it should eventually be possible to design wavelet- and shearlet-based visual servoing

controllers that can also be considered for real-time applications.

CHAPTER 4

Edge, Ridge, and Blob Detection

The contents of this chapter are based on:

- REISENHOFER, R. and KING, E. J.: “Edge, ridge, and blob detection with symmetric molecules”. In preparation.
 - REISENHOFER, R., KIEFER, J., and KING, E. J.: “Shearlet-based detection of flame fronts”. *Exp. Fluid.* (Feb. 2016), vol. 57(3): p. 41
 - KING, E. J., REISENHOFER, R., KIEFER, J., LIM, W.-Q., LI, Z., and HEYGSTER, G.: “Shearlet-based edge detection: flame fronts and tidal flats”. *Applications of Digital Image Processing XXXVIII*. ed. by G. TESCHER, A. G. Vol. 9599. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. 2015
-

Prelude α -Molecules

We have already discussed the construction principles behind wavelet- and shearlet-based systems of functions and some of the properties of their associated transforms in the preludes to Chapters 2 and 3. As we recall, the most significant difference between the definition of systems of two-dimensional wavelets and systems of shearlets is that the former are constructed by considering an isotropic scaling matrix (cf. Equation (3.32)), while the latter are defined by applying a so-called parabolic scaling matrix (cf. Equation (3.40)), that yields anisotropically dilated elements, which are better suited for representing two-dimensional curvilinear singularities. While systems of shearlets indeed exhibit an approximation behavior with respect to a certain model class of two-dimensional images that is provably in a sense optimal (cf. Theorem 3.0.6), they are not the only multiscale representation system that has been proposed to address the weaknesses of wavelets when dealing with anisotropic features in two-dimensional signals. Other prominent examples include ridgelets [26, 27], which apply a maximally anisotropic scaling matrix in the sense that dilation only occurs in a single direction, curvelets [131], which use a rotation matrix instead of a shear matrix to change the preferred orientation of a generator, as well as contourlets [49].

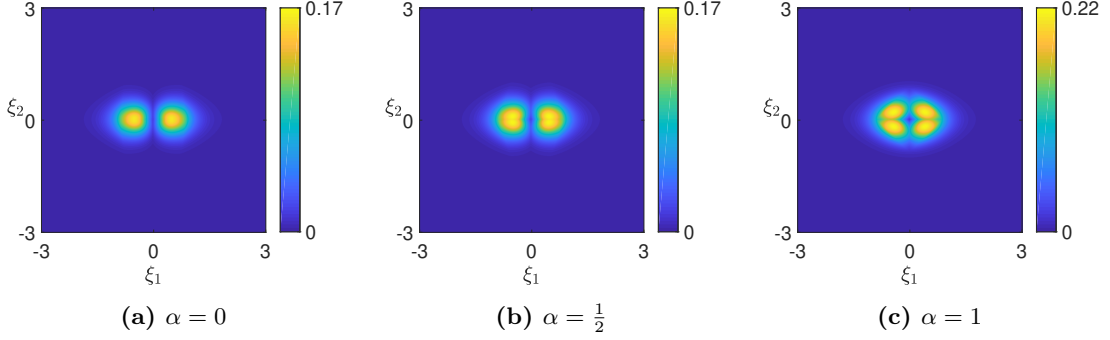


Figure 4.1: The weight function on the right hand side of (4.5) for parameters $M = 2$, $N_1 = N_2 = 8$, $s_\lambda = 8$, and $\alpha \in \{0, \frac{1}{2}, 1\}$.

In spite of various differences in their construction and their approximation behavior for different classes of signals, many of the aforementioned function systems are similar in the sense that all of them are constructed by applying translation, scaling, rotation, or shear operators to a possibly infinite number of so-called generator functions, which typically need to satisfy a certain admissibility condition. In a series of papers [71, 72, 73], so-called α -molecules were introduced to provide a unifying framework that includes different constructions of multiscale representations in order to enable the transfer of results from one theory to another, to allow for a categorization of multiscale representations with respect to the approximation behavior, and to facilitate the construction of novel systems.

Systems of α -molecules are constructed by translating, scaling, and rotating members of a possibly infinite set of generator functions, where the latter two operations are performed by applying scaling matrices

$$\mathbf{A}_{s,\alpha} = \begin{bmatrix} s & 0 \\ 0 & s^\alpha \end{bmatrix}, \quad s > 0, \alpha \in [0, 1], \quad (4.1)$$

and rotation matrices

$$\mathbf{R}_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}, \quad \theta \in \mathbb{T}, \quad (4.2)$$

to the argument of a generator function, where $\mathbb{T} = [-\frac{\pi}{2}, \frac{\pi}{2})$. Note that the parameter α in (4.1) controls the degree of anisotropy and interpolates between isotropic scaling ($\alpha = 1$), which is used in the definition of wavelet-based systems, and fully anisotropic scaling ($\alpha = 0$), as applied in the definition of ridgelets.

For a fixed parameter $\alpha \in [0, 1]$, a single α -molecule is thus defined by its corresponding generator and a point (s, θ, \mathbf{y}) in the parameter space

$$\mathbb{P} := \mathbb{R}_+ \times \mathbb{T} \times \mathbb{R}^2, \quad (4.3)$$

where s is a scaling, θ a rotation, and \mathbf{y} a translation parameter.

Let Λ be an index set and $\alpha \in [0, 1]$ be fixed. A family of functions $\{m_\lambda\}_{\lambda \in \Lambda} \subset L^2(\mathbb{R}^2)$ is called a system of α -molecules of order (L, M, N_1, N_2) , with control parameters $L, M, N_1, N_2 \in \mathbb{N}_0 \cup \{\infty\}$, if each molecule can be written as

$$m_\lambda = s_\lambda^{(1+\alpha)/2} g^{(\lambda)}(\mathbf{A}_{s_\lambda, \alpha} \mathbf{R}_{\theta_\lambda}(\cdot - \mathbf{y}_\lambda)), \quad (4.4)$$

where $(s_\lambda, \theta_\lambda, \mathbf{y}_\lambda) \in \mathbb{P}$ is a point in the α -molecule parameter space and the corresponding generator $g^{(\lambda)} \in L^2(\mathbb{R}^2)$ satisfies

$$\left| \partial^\rho \widehat{g^{(\lambda)}}(\boldsymbol{\xi}) \right| \lesssim \min \left\{ 1, s_\lambda^{-1} + |\xi_1| + s_\lambda^{-(1-\alpha)} |\xi_2| \right\}^M \cdot (1 + \|\boldsymbol{\xi}\|_2^2)^{\frac{-N_1}{2}} \cdot (1 + (\xi_2)^2)^{\frac{-N_2}{2}}, \quad (4.5)$$

with $\boldsymbol{\xi} \in \mathbb{R}^2$ for all $|\boldsymbol{\rho}| \leq L$, such that the implicit constants associated with the definition of \lesssim are uniform over $\lambda \in \Lambda$. Note that $\boldsymbol{\rho} \in \mathbb{N}_0^2$ denotes a multi-index with $|\boldsymbol{\rho}| = \rho_1 + \rho_2$. Further note that if one or more of the parameters L, M, N_1 , and N_2 equal ∞ , the respective values in (4.5) can be arbitrarily large. Three examples of the weight function (4.5) for different choices of α are visualized in Figure 4.1.

Equation (4.5) provides a maximum amount of flexibility by allowing each molecule of an α -molecule system to be based on a different generator. In practice, however, systems of α -molecules are often constructed from only a small number of generators. For instance, wavelet orthonormal bases of $L^2(\mathbb{R}^2)$ are typically based on four generators (cf. Equations (3.28) to (3.31)), while so-called cone-adapted shearlet systems only require three different generators (cf. Equations (3.47) to (3.49)). Note that in the latter case, due to the application of the shear matrix, the elements of a cone-adapted shearlet system can not directly be expressed in the form (4.4) when just considering the two shearlet generators from Equations (3.48) and (3.49). However, it can be shown that the cone-adapted discrete shearlet system (3.46) is equivalent to an α -molecule system that is based on an infinite number of sheared generators, all of which satisfy Condition (4.5) [71].

Analogous to the wavelet- and shearlet-based transforms discussed in the prelude to Chapter 3, the transform of a function $f \in L^2(\mathbb{R}^2)$ with respect to an α -molecule system $\{m_\lambda\}_{\lambda \in \Lambda} \subset L^2(\mathbb{R}^2)$ can be defined in terms of the analysis operator:

$$(\mathcal{T}f)(g^{(\lambda)}, s_\lambda, \theta_\lambda, \mathbf{y}_\lambda) = \langle f, m_\lambda \rangle_{L^2} = s_\lambda^{(1+\alpha)/2} \int_{\mathbb{R}^2} f(\mathbf{x}) \overline{g^{(\lambda)}(\mathbf{A}_{s_\lambda, \alpha} \mathbf{R}_{\theta_\lambda}(\mathbf{x} - \mathbf{y}_\lambda))} \, d\mathbf{x}. \quad (4.6)$$

Note that the decay condition (4.5) requires the generators of an α -molecule system to satisfy certain time-frequency localization properties, where the parameter L describes the degree of spatial localization, M is the number of directional (almost) vanishing moments

and N_1 and N_2 describe the smoothness of a generator. At first glance, Condition (4.5) can seem a bit unintuitive. Let us thus briefly elaborate on the relationship between the single control parameters L , M , N_1 , and N_2 , and the respective properties of the α -molecule system. We have already observed in the prelude to Chapter 3, that for integrable wavelet generators, the wavelet admissibility condition implies that the generator has a vanishing mean (cf. Equation 3.4). This notion can be generalized in the sense that a one-dimensional function f is said to have $N \in \mathbb{N}$ *vanishing moments*, if

$$\int_{\mathbb{R}} f(x) x^n dx = 0, \quad \forall n \in \{0, \dots, N-1\}. \quad (4.7)$$

By considering the Taylor expansion, it can in fact be shown that the number of vanishing moments of a wavelet generator also determines the decay rate of the associated wavelet coefficients at locations at which an analyzed signal is smooth. To illustrate the link between the control parameter M in Condition 4.5 and the vanishing moments of an α -molecule generator, let us consider the following lemma.

Lemma 4.0.1 (as in [73]). *Let $f: \mathbb{R} \rightarrow \mathbb{C}$ be a continuous and compactly supported function that possesses N vanishing moments. Then, it holds that*

$$|\widehat{f}(\xi)| \lesssim \min \{1, |\xi|\}^N. \quad (4.8)$$

Proof. Due to

$$\frac{d^n}{d\xi^n} \widehat{f}(0) = (-2\pi i)^n \int_{\mathbb{R}} f(x) x^n dx, \quad n \in \mathbb{N}, \quad (4.9)$$

the vanishing moments of f imply that $\frac{d^n}{d\xi^n} \widehat{f}(0) = 0$ for all $n \in \{0, \dots, N-1\}$. As f is compactly supported, \widehat{f} is analytic, and considering the respective power series expansion at 0 yields

$$|\widehat{f}(\xi)| \lesssim |\xi|^N. \quad (4.10)$$

Furthermore, since $f \in L^1(\mathbb{R})$, \widehat{f} is bounded and we can write

$$|\widehat{f}(\xi)| \lesssim \min \{1, |\xi|\}^N. \quad (4.11)$$

□

Lemma (4.8) illustrates that the first factor in (4.5), which requires the generators of an α -molecule systems to decay in the frequency domain near the origin, is in fact a necessary condition for the existence of vanishing moments.

As for the connection between the control parameters L , N_1 , and N_2 , and the degree of spatial localization, respectively the smoothness of a generator, one can consider a consequence of the Riemann-Lebesgue lemma, which states that for $f \in L^1(\mathbb{R})$, the Fourier transform goes to zero at infinity, that is, $\lim_{|\xi| \rightarrow \infty} \widehat{f}(\xi) = 0$. Let an integrable function f be N -times continuously differentiable with integrable derivatives, then $(2\pi i \xi)^N \widehat{f}(\xi) = \widehat{\frac{d^N}{dx^N} f}(\xi)$ (cf. Equation (3.77)), and the Riemann-Lebesgue lemma implies that $\lim_{|\xi| \rightarrow \infty} \frac{d^N}{dx^N} f(\xi) = 0$, and, in turn, that $\widehat{f}(\xi) \lesssim \xi^{-N}$ for $\xi \rightarrow \infty$. Broadly speaking, this means that regularity in the time domain implies fast decay in the frequency domain, and, due to the interchangeable nature of the Fourier transform and its inverse, vice versa, which is reflected in Condition (4.5) by the control parameters, L , N_1 , and N_2 .

For the final part of this prelude, we will briefly sketch how the framework of α -molecules can be applied to transfer results about the approximation properties of one type of multiscale representation system to another. We have already seen that shearlet-based frames can yield optimally sparse N -term approximations of cartoon-like image functions (cf. Theorem 3.0.6), which is mostly owed to the fast decay of shearlet coefficients at locations where the respective elements of the shearlet frame are not spatially *and* directionally aligned with a curvilinear singularity. In general, there is a close relationship between the decay rate of the N -term approximation error of a frame and the sparsity, that is, the decay rate, of the respective sequence of analysis coefficients. An analytically tractable way of measuring the sparsity of a sequence is by considering a possibly non-convex ℓ^p -(quasi)-norm, where $p \in (0, 1]$. It can in fact be shown that if, for a given function or class of functions, the respective sequences of analysis coefficients of two distinct frames have a similar decay behavior, these frames also yield the same N -term approximation rate (see, e.g., [71]). This observation motivated the following definition of *sparsity equivalence* for frames, which was first proposed in [73], and which is based on the cross-Gramian, that is, the matrix whose entries are all pairwise inner products of elements from the two considered frames.

Definition 4.0.2 (Sparsity equivalence, [73]). *Let $p \in (0, 1]$, and let $\{\psi_\lambda\}_{\lambda \in \Lambda}$, and $\{\tilde{\psi}_{\tilde{\lambda}}\}_{\tilde{\lambda} \in \tilde{\Lambda}}$ denote frames for $L^2(\mathbb{R}^2)$ with discrete index sets Λ , and $\tilde{\Lambda}$. The frames are said to be sparsity equivalent in $\ell^p(\mathbb{Z})$, if the cross-Gramian*

$$\mathbf{G}_{\lambda, \tilde{\lambda}} = \left\langle \psi_\lambda, \tilde{\psi}_{\tilde{\lambda}} \right\rangle_{L^2}, \quad \lambda \in \Lambda, \tilde{\lambda} \in \tilde{\Lambda}, \quad (4.12)$$

satisfies

$$\|\mathbf{G}\|_{\ell^p \rightarrow \ell^p} < \infty, \quad (4.13)$$

where $\|\cdot\|_{\ell^p \rightarrow \ell^p}$ denotes the operator norm of a matrix on $\ell^p(\mathbb{Z})$.

One of the main contributions of the theory of α -molecules thus far is the derivation of sufficient conditions for the sparsity equivalence of two distinct systems of α -molecules that are frames for $L^2(\mathbb{R}^2)$. As one would expect, these conditions partly depend on the order of an α -molecule system, which imposes restrictions on the regularity, the spatial localization, and the number of vanishing moments of the associated generators. However, when considering systems of α -molecules, the cross-Gramian (4.12) also strongly depends on the associated sets of dilation, rotation, and translation parameters. The following notion of (α, k) -consistency was first proposed in [73] and yields a practical means of evaluating whether two different parameter sets in \mathbb{P} are suitable for obtaining sparsity equivalent frames. Note that the precise definition of the so-called α -scaled index distance function $\omega_\alpha(\cdot, \cdot)$, which provides a distance measure for points in the parameter space \mathbb{P} , is omitted here for brevity, but can be found in [71].

Definition 4.0.3 ((α, k) -consistency, as in [71]). *Let $\alpha \in [0, 1]$, $k > 0$, and Λ , and $\tilde{\Lambda}$ denote two discrete index sets. Two α -molecule parameter sets $\{\mathbf{q}_\lambda = (s_\lambda, \theta_\lambda, \mathbf{y}_\lambda)\}_{\lambda \in \Lambda} \subset \mathbb{P}$, and $\{\tilde{\mathbf{q}}_{\tilde{\lambda}} = (\tilde{s}_{\tilde{\lambda}}, \tilde{\theta}_{\tilde{\lambda}}, \tilde{\mathbf{y}}_{\tilde{\lambda}})\}_{\tilde{\lambda} \in \tilde{\Lambda}} \subset \mathbb{P}$, are called (α, k) -consistent, if*

$$\sup_{\lambda \in \Lambda} \sum_{\tilde{\lambda} \in \tilde{\Lambda}} \omega_\alpha(\mathbf{q}_\lambda, \tilde{\mathbf{q}}_{\tilde{\lambda}})^{-k} < \infty, \quad \text{and} \quad \sup_{\tilde{\lambda} \in \tilde{\Lambda}} \sum_{\lambda \in \Lambda} \omega_\alpha(\mathbf{q}_\lambda, \tilde{\mathbf{q}}_{\tilde{\lambda}})^{-k} < \infty, \quad (4.14)$$

where $\omega_\alpha(\cdot, \cdot): \mathbb{P} \times \mathbb{P} \rightarrow [1, \infty)$ denotes the so-called α -scaled index distance (see Definition 4.1. in [71]).

As demonstrated by the following theorem, the notion of (α, k) -consistency and the order of an α -molecule system can be used to obtain a sufficient condition for the sparsity equivalence of frames for $L^2(\mathbb{R}^2)$.

Theorem 4.0.4 ([71]). *Let $\alpha \in [0, 1]$, $k > 0$, $p \in (0, 1]$, and let $\{m_\lambda\}_{\lambda \in \Lambda}$, and $\{\tilde{m}_{\tilde{\lambda}}\}_{\tilde{\lambda} \in \tilde{\Lambda}}$, where Λ and $\tilde{\Lambda}$ are discrete index sets, denote systems of α -molecules of order (L, M, N_1, N_2) that are frames for $L^2(\mathbb{R}^2)$. If the associated parameter sets are (α, k) -consistent and it furthermore holds that*

$$L \geq 2\frac{k}{p}, \quad M > 3\frac{k}{p} - \frac{3-\alpha}{2}, \quad N_1 \geq \frac{k}{q} + \frac{1+\alpha}{2}, \quad \text{and} \quad N_2 \geq 2\frac{k}{p}, \quad (4.15)$$

the systems of α -molecules are sparsity equivalent in $\ell^p(\mathbb{Z})$.

Theorem (4.0.4) can now be used to obtain results about the approximation behavior of a given α -molecule frame by relating it to another system of α -molecules for which these results have already been established. To give an example, let us consider the following

generalization of the definition of cartoon-like image functions (cf. Definition 3.0.4). Note that for the remainder of this prelude, we will use C^β with $\beta \in (1, 2]$ to denote the space of continuously differentiable functions whose derivatives are Hölder continuous with exponent $\beta - 1$.

Definition 4.0.5 (C^β -cartoon-like image functions, as in [71]). *Let $\beta \in (1, 2]$ and $f \in L^2(\mathbb{R}^2)$ be compactly supported in $[0, 1]^2$. Then f is called a C^β -cartoon-like image function if it can be written as*

$$f = f_0 + \mathbb{1}_B \cdot f_1, \quad (4.16)$$

where $B \subset [0, 1]^2$ is a measurable set whose boundary is a regular closed piecewise C^β -curve, and $f_0, f_1 \in C^\beta([0, 1]^2)$ are compactly supported.

It was shown in [70] that a tight frame of so-called α -curvelets provides, up to a log-factor, optimally sparse approximations of C^β -cartoon-like image functions in the sense that for a fixed anisotropy parameter $\alpha \in [\frac{1}{2}, 1)$, and $\beta = \alpha^{-1}$, there exists a constant $C > 0$ such that for every C^β -cartoon-like image function f , it holds that

$$\|f - f_N\|_2 \leq CN^{-\frac{\beta}{2}} \cdot (\log_2 N)^{\frac{\beta+1}{2}}, \quad \text{for } N \rightarrow \infty, \quad (4.17)$$

where f_N denotes the N -term approximation of f obtained by choosing the N largest coefficients yielded by the analysis operator of the considered α -curvelet frame (cf. Equation (3.52)). Note that in the parabolic case of $\alpha = \frac{1}{2}$, (4.17) is equivalent to the decay rate of shearlet-based N -term approximations discussed in the prelude to Chapter 3 (see Equation (3.54)). In particular, Equation (4.17) generalizes the result from Chapter 3 by showing that a higher regularity of the boundary curves and the smooth regions of a cartoon-like image allows for better approximation rates that can be achieved by increasing the degree of anisotropy of the scaling matrix $\mathbf{A}_{s,\alpha}$ (cf. (4.1)). It was furthermore shown in [71] that the system of α -curvelets that yields (4.17) also constitutes a system of α -molecules of order $(\infty, \infty, \infty, \infty)$. The following theorem combines both results and provides a powerful tool for obtaining a large class of multiscale representation systems that yield an (almost) optimal approximation behavior with respect to cartoon-like image functions.

Theorem 4.0.6 (Optimally sparse approximations of C^β -cartoon-like images, [71]). *Let $\alpha \in [1/2, 1)$, and $\beta = \alpha^{-1}$. Furthermore, let $\{m_\lambda\}_{\lambda \in \Lambda}$ denote a system of α -molecules of order (L, M, N_1, N_2) , that is a tight frame for $L^2(\mathbb{R}^2)$, where Λ is a discrete index set, and $\mathbf{Q}_\Lambda \subset \mathbb{P}$ denotes the associated set of parameters. Then, if for some fixed $k > 0$, the parametrization \mathbf{Q}_Λ is (α, k) -consistent with the parametrization of a specific tight frame*

of α -curvelets (see Definition 2.2. in [71]), and it holds that

$$L \geq k(1 + \beta), \quad M \geq \frac{3k}{2}(1 + \beta), \quad N_1 \geq \frac{k}{2}(1 + \beta) + \frac{1 + \alpha}{2}, \quad \text{and} \quad N_2 \geq k(1 + \beta), \quad (4.18)$$

the $\{m_\lambda\}_{\lambda \in \Lambda}$ -based N -term approximations of C^β -cartoon-like image functions are (almost) optimally sparse, that is,

$$\|f - f_N\|_2 \lesssim N^{-\frac{\beta}{2} + \varepsilon}, \quad (4.19)$$

where f is a C^β -cartoon-like image function, f_N denotes the N -term approximation obtained from the N largest coefficients in magnitude (cf. Equation 3.52), and $\varepsilon > 0$ can be chosen arbitrarily small.

4.1 Introduction

The correct localization of the significant structures in an image as well as the precise characterization of their geometry are two of the most eminent tasks in digital image processing with an overwhelming number of applications. Since the advent of digital image processing, a large body of research has been devoted to the development and analysis of algorithms for the detection and characterization of features such as edges, ridges, or well-defined two-dimensional shapes. Surprisingly, in many practical situations, even highly optimized implementations of popular methods such as the Canny edge detector [28], approaches that are based on directionally sensitive filters [88], or even multiscale systems of anisotropic functions [171] are not always capable of reliably identifying the features in question. This is often the case for images which are heavily distorted by noise, in which different features are strongly overlapping, or where the geometry of the considered features is characterized by a high variation and irregularities such as corner points. It is also worth noting that feature detection in general is a task where, despite recent advances, computers are still often outperformed by humans.

A main difficulty researchers face when developing methods for feature detection is to identify universal and computationally tractable properties that are characteristic of or even unique to points at which a certain feature is localized. In the case of edge detection, the first and most basic observation was that edges are typically associated with changes in contrast and can thus be identified by considering points with large image gradients. This approach led to the development of early edge enhancement filters such as the Roberts filter [148], the Prewitt filter [140], and the Sobel filter [153], which are still widely applied today. However, any procedure merely considering the magnitude of gradient filters as an indicator is highly sensitive to noise and variations in image illumination. Two famous and influential descriptors which already made a big step towards capturing the structural nature of edges

were proposed in the 1980's by Marr and Hildreth [123], and Canny [28]. Marr and Hildreth observed that edges coincide with zero-crossings in the second directional-derivative of the image intensities while Canny developed an algorithm for edge detection which identifies points at which the image gradient reaches a local maximum. The Canny edge detector is also based on a gradient filter, and the Marr-Hildreth-operator applies a discrete Laplace filter. However, both methods are not just evaluating whether a magnitude exceeds a certain threshold but consider local structural properties by testing for zero-crossings and local maxima.

A fully dimensionless approach to identifying features such as edges and ridges can be found in the so-called local energy model, which was proposed in the late 1980's and postulates that the location of features coincides with points in an image for which the Fourier components are maximally *in phase* [129, 130, 163]. However, to precisely calculate the degree of phase congruency, it is necessary to optimize a certain value over inputs from a continuous interval for every point in an image. This computationally highly inefficient procedure was later significantly simplified by Kovess, who showed that an equivalent measure can be formulated in terms of convolutions with differently scaled and oriented complex-valued wavelet filters [99, 100]. The resulting phase congruency measure yields a contrast invariant feature detector that simultaneously detects edges and ridges and can be efficiently implemented to process digital images. It has furthermore strongly inspired the measures for edge, ridge, and blob detection presented here and will therefore be introduced in greater detail later in this work.

In this chapter, we derive measures for the detection of edges, ridges, and blobs that are inspired by the local energy model and the notion of phase congruency but realized by making full use of the flexibility of modern construction principles for anisotropic analyzing functions provided within the framework of α -molecules. Saying that the Fourier components are in phase at the location of an edge basically means that locally the considered function is purely defined by odd-symmetric sine components while all even-symmetric cosine components vanish. In the first step, we will derive a novel edge measure that utilizes this observation by testing for the local symmetry properties of a function by considering differently scaled and oriented odd- and even-symmetric analyzing functions constructed in the α -molecule framework. This measure can then be generalized to detect ridges by interchanging the roles of the odd- and even-symmetric molecules. We further show how, with a few modifications regarding the measure and the construction of analyzing functions, similar principles can be applied to the case of blob detection. Each analyzing function in a multiscale system considered in this work is associated with a specific pair of scaling and orientation parameters and thus conveys not only information about the location of features but also about their geometry in terms of feature width and orientation.

Eventually, we demonstrate how this information can be utilized to obtain measures for local tangent directions, widths, and heights, that yield detailed information about the geometry of an image.

4.1.1 Contributions

Variants of the edge measure derived in Section 4.3.2 and the ridge measure derived in Section 4.3.3 that are based on complex-valued shearlets have already been published previously by the author of the present thesis and co-authors [96, 142, 144]. The main contribution of the work presented in this chapter is to provide a comprehensive derivation and description of the proposed measures within the framework of α -molecules, which provides enough flexibility to allow for a unified treatment of the cases of edge, ridge, and blob detection. In particular, we demonstrate in Theorem 4.3.4 that by considering tensor products of derivatives of the one-dimensional Gaussian and their Hilbert transforms, one can obtain a large class of well-behaved α -molecules that also provide the necessary symmetry properties. We furthermore show in Section 4.3.4 how the concept of considering local symmetry properties can also be generalized to yield a blob measure and specifically derive functions that estimate the local tangent orientations of edge contours and ridge centerlines, local diameters of ridges and blobs, and the local contrast of a feature.

To evaluate the strengths and weaknesses of the proposed measures in comparison with other state of the art methods for feature detection, we performed extensive numerical experiments with respect to synthetic images with reliable ground truths, which can be found in Section 4.5. To furthermore demonstrate the applicability of the proposed framework, we consider the application of the ridge measure in the context of retinal image analysis in Section 4.6.1, flame front characterization in Section 4.6.3, and show how the proposed blob measure can be applied for automatically counting the number of cell colonies in a Petri dish in Section 4.6.2.

We also developed a MATLAB toolbox that implements all of the proposed measures for two-dimensional digital images. The implementation strategies are briefly explained in Section 4.4 and the toolbox can be downloaded from <http://www.math.uni-bremen.de/cda/software.html>.

4.2 Related Work

We begin by giving a short review of the most important tools and concepts that will later be used to derive the respective edge, ridge, and blob measures.

4.2.1 Derivative of Gaussian Wavelets

In order to detect features by testing for local symmetry properties, we require even- and odd-symmetric real-valued functions with good localization in both the time and the frequency domain. In the remainder of this work, we will consider two simple tools to construct one-dimensional even- and odd-symmetric generators with these properties, namely taking derivatives of the Gaussian as well as applying the so-called Hilbert transform. A comprehensive discussion of the methods reviewed in this section can be found in [12]. For $k \in \mathbb{N}_0$, let us denote the k -th derivative of the unnormalized Gaussian by

$$G_k(x) = (-1)^k H_k(x) e^{-x^2}, \quad (4.20)$$

where H_k is the k -th Hermite polynomial. The Hermite polynomials are orthogonal with respect to the weight function e^{-x^2} and can be defined by $H_0(x) = 1$ and the recurrence relation

$$H_{k+1}(x) = 2xH_k(x) - 2kH_{k-1}(x), \quad (4.21)$$

where $H_{-1}(x) = 0$. For more details on orthogonal polynomials and Hermite polynomials in particular, see [159]. It is furthermore well known that for a differentiable function f with $f, f' \in L^1(\mathbb{R})$, its derivative can be expressed algebraically in the frequency domain (cf. Equation 3.77), and that the Gaussian $G_0(\sqrt{\pi}x) = e^{-\pi x^2}$ is a fixed point of the Fourier transform. G_k can thus be expressed in the Fourier domain by

$$\widehat{G}_k(\xi) = (2\pi i \xi)^k \sqrt{\pi} e^{-\pi^2 \xi^2}. \quad (4.22)$$

Note that if k is odd, \widehat{G}_k is purely imaginary, while if k is even, \widehat{G}_k is purely real. This implies that G_k is odd-symmetric if k is odd and even-symmetric if k is even. Furthermore, the Gaussian lies in the Schwartz space of rapidly decreasing functions. That is, the Gaussian is infinitely differentiable and all of its derivatives decay faster than the inverse of any polynomial. With $\widehat{G}_k(0) = 0$, this implies that G_k satisfies the wavelet admissibility condition (cf. Equation (3.1)) for any integer $k > 0$. The Hilbert transform is a bounded linear operator on $L^2(\mathbb{R})$ and defined for functions $f \in L^2(\mathbb{R})$ by

$$(\mathcal{H}f)(x) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{f(t)}{t-x} dt, \quad (4.23)$$

or equivalently in the Fourier domain via

$$\widehat{\mathcal{H}f}(\xi) = -i \operatorname{sgn}(\xi) \hat{f}(\xi), \quad (4.24)$$

where sgn denotes the sign function.

It is easy to see from (4.24) that if \hat{f} is purely real, $\widehat{\mathcal{H}f}$ is purely imaginary and vice versa. This implies that for any even-symmetric function f , its Hilbert transform $\mathcal{H}f$ is odd-symmetric, while any odd-symmetric function will become even-symmetric under the Hilbert transform. The Fourier-based definition (4.23) also illustrates that the Hilbert transform leaves the magnitude of the Fourier transform invariant. In particular, this implies that the wavelet admissibility condition (3.1) is preserved by the Hilbert transform and that for any function $f \in L^2(\mathbb{R})$, it holds that $\|f\|_2 = \|\mathcal{H}f\|_2$. Note that the latter does not hold for the L^1 -norm. Two examples of even- and odd-symmetric wavelets based on derivatives of the Gaussian are plotted along with their respective Hilbert transforms in Figure 4.2.

4.2.2 Feature Detection via Phase Congruency

By taking the Fourier series expansion, any real-valued function $f \in L^2(-\pi, \pi)$ can be written as

$$f(x) = \frac{1}{2\pi} \sum_{n \in \mathbb{Z}} |c_n| \cos(nx + \varphi_n) = \frac{|c_0|}{2\pi} + \frac{1}{\pi} \sum_{n \in \mathbb{N}^+} |c_n| \cos(nx + \varphi_n), \quad (4.25)$$

for almost every $x \in (-\pi, \pi)$, where $c_n = \langle f, e^{in\cdot} \rangle_{L^2}$, and φ_n denotes the phase angle of c_n , i.e., $c_n = |c_n| e^{i\varphi_n}$. It was already observed in the mid-80s [129, 130] that at the location of features like jump-discontinuities, the angles $nx + \varphi_n$ are in congruency for different choices of $n \in \mathbb{N}$. This led to the definition of a so-called phase congruency measure (e.g. [99]), given by

$$\text{PC}(f, x) = \max_{\mu \in [0, 2\pi]} \frac{\sum_{n \in \mathbb{N}} |c_n| \cos(nx + \varphi_n - \mu)}{\sum_{n \in \mathbb{N}} |c_n|}, \quad (4.26)$$

where c_n and φ_n are defined as above. This measure can be computed for any non-zero square integrable function and every point in its domain. $\text{PC}(f, x)$ takes 1 if the phases of the frequency components of f are locally in perfect congruency, indicating the presence of a significant feature at x , and goes towards 0 if the converse is the case. However, maximizing the weighted mean in (4.26) over all possible phases for each point in the domain of a signal is a computationally demanding task. To simplify matters, Morrone and Owens suggested considering local energy functions, whose maxima coincide with the local maxima of the phase congruency measure [129, 163]. These local energies were obtained by first convolving a signal with an even-symmetric filter and its odd-symmetric Hilbert transform and then summing the squares of the respective outputs (see also [5]).

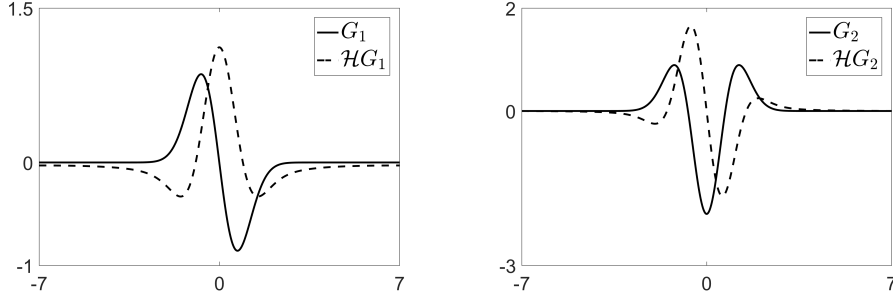


Figure 4.2: The first and the second derivative of the Gaussian plotted alongside their respective Hilbert transforms.

The apparent issue behind computing (4.26) is that phase congruency is a local property but based on the coefficients of the discrete Fourier transform, which – when viewed independently – only provide information about *which* frequency components are present in a signal but not *where* they occur. A possible solution was given by Kovess, who proposed extracting the local phase congruency by using sets of differently scaled complex-valued wavelets [99]. Similar to the use of sines and cosines in the local energy approach, the real and imaginary parts of these wavelets formed Hilbert transform pairs, with the real part being even- and the imaginary part being odd-symmetric. However, by considering multiple scalings, Kovess was able to combine phase information from different parts of the frequency domain and to eventually derive a dimensionless and easy-to-compute complex wavelet-based phase congruency measure, namely

$$\text{PC}_{\psi^c}(f, x) = \frac{\left| \sum_{a \in A} \langle f, \psi_{a,x}^c \rangle_{L^2} \right|}{\sum_{a \in A} \left| \langle f, \psi_{a,x}^c \rangle_{L^2} \right| + \varepsilon}, \quad (4.27)$$

where $\psi^c \in L^2(\mathbb{R})$ is a complex-valued wavelet, $A \subset \mathbb{R}_+$ a set of scaling parameters, $\psi_{a,x}^c = a^{-\frac{1}{2}} \psi\left(\frac{\cdot - x}{a}\right)$ (i.e., ψ^c centered at x and scaled by a) and $\varepsilon > 0$ prevents division by zero. Analogous to (4.26), this measure is defined for any square integrable function f and each point x in its domain. It goes towards 1 if the phases of the local frequency components are in congruency and towards 0 if they are widely spread. More details about the generalization of this measure to two dimensions and its implementation can be found here [99, 100, 101].

The measure (4.27) can effortlessly be computed and is by construction contrast invariant. Features such as edges and ridges are picked up solely by the traces their structure leaves in the phases of the frequency representation of a signal rather than the local magnitude of contrast. Furthermore, as Kovess points out in his original work, it

is an interesting property of the phase congruency measure that from the perspective of (4.26) or (4.27), there is no difference between edges and ridges.

While contrast invariance is indeed a highly desirable feature in many edge and ridge detection tasks, localizing a property of the frequency components in the time representation of a signal comes at a price. Indeed, it was already observed in [157] and [142] that replacing isotropic wavelets with anisotropically scaled analyzing functions in (4.27) deteriorates the detection of features in images rather than improving it. This should come as a surprise, as one would assume that anisotropically scaled elements would be better suited for detecting typically anisotropic features such as edges or ridges than their isotropically scaled counterparts. Finding a way of bringing together the intuitions behind the phase congruency measure $\text{PC}_{\psi^c}(f, x)$ and modern constructions of anisotropic analyzing functions was in fact one of the main factors motivating the present work.

4.2.3 Maximum Point Estimation From Discrete Samples

Each two-dimensional analyzing function within a system of α -molecules is associated with a certain scaling and rotation parameter (cf. (4.4)). These parameters can be used to obtain first estimates of the local tangent direction and the width of a feature by considering the most significant analyzing function (i.e., the member of a system of functions yielding the largest coefficient). To obtain more precise measurements, we consider a refinement procedure which allows us to estimate the maximum point of a function defined in the continuum from only a few discrete samples.

Let $f \in C^2(\mathbb{R})$, $\{x_n\}_{n=1}^N \subset \mathbb{R}$ be a strictly increasing sequence of $N \in \mathbb{N}$ sampling points (i.e. $x_n < x_{n+1}$ for all $1 \leq n \leq N-1$) and denote

$$n^* = \operatorname{argmax}_{n \leq N} f(x_n). \quad (4.28)$$

If $1 < n^* < N$, the function f has at least one local maximum point in the interval $[x_{n^*-1}, x_{n^*+1}]$, which is in real-world applications often taken as a best guess for the global maximum point of f . One method of refining the estimate x_{n^*} for a local maximum point is to assume that f can be approximated on $[x_{n^*-1}, x_{n^*+1}]$ by a parabola fit through the points $(x_{n^*-1}, f(x_{n^*-1}))$, $(x_{n^*}, f(x_{n^*}))$, and $(x_{n^*+1}, f(x_{n^*+1}))$, that is

$$f(x) \approx c_2(x - x_{n^*})^2 + c_1(x - x_{n^*}) + c_0 \quad (4.29)$$

for $x \in [x_{n^*-1}, x_{n^*+1}]$, where the parameters $c_2, c_1, c_0 \in \mathbb{R}$ are chosen such that

$$f(x_{n^*}) = c_0, \quad (4.30)$$

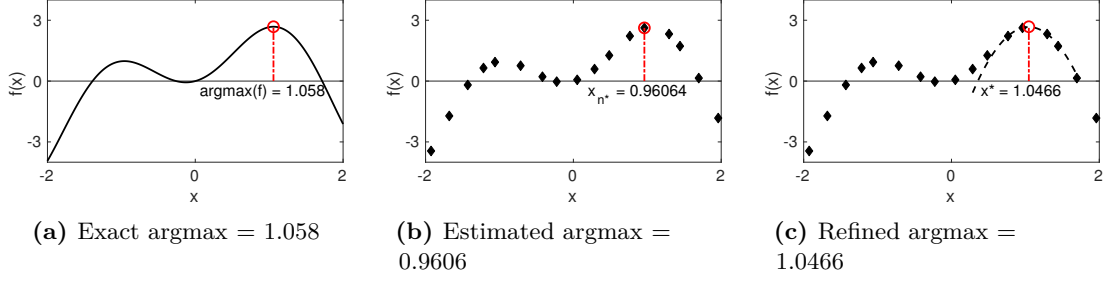


Figure 4.3: Example of a refined maximum point estimation based on discrete samples. (a): A C^2 -function and its maximum point in the interval $[-2, 2]$. (b): An estimate x_{n^*} of the maximum point obtained from sampling the function at 17 different points. (c): A refined estimate x^* of the maximum point obtained from fitting a parabola through x_{n^*} and its nearest neighbors.

$$f(x_{n^*-1}) = c_2(x_{n^*-1} - x_{n^*})^2 + c_1(x_{n^*-1} - x_{n^*}) + c_0, \quad \text{and} \quad (4.31)$$

$$f(x_{n^*+1}) = c_2(x_{n^*+1} - x_{n^*})^2 + c_1(x_{n^*+1} - x_{n^*}) + c_0. \quad (4.32)$$

Setting $d_- = x_{n^*-1} - x_{n^*}$ and $d_+ = x_{n^*+1} - x_{n^*}$, the parameters c_2 and c_1 are given by

$$c_2 = \frac{d_+(f(x_{n^*-1}) - f(x_{n^*})) - d_-(f(x_{n^*+1}) - f(x_{n^*}))}{d_+((d_-)^2 - d_+d_-)}, \quad (4.33)$$

$$c_1 = \frac{(d_-)^2(f(x_{n^*+1}) - f(x_{n^*})) - (d_+)^2(f(x_{n^*-1}) - f(x_{n^*}))}{d_-(d_+d_- - (d_+)^2)} \quad (4.34)$$

and the maximum point of the parabola (4.29), denoted as x^* , can be computed by

$$x^* = x_{n^*} - \frac{c_1}{2c_2}. \quad (4.35)$$

Note that in the case $d_- = -1$ and $d_+ = 1$, equations (4.33) and (4.34) simplify significantly. A short review of this and other methods for estimating maximal points from discrete samples in the context of edge and line detection can be found in [40].

Figure 4.3 illustrates how (4.35) can be used to yield refined estimates of local maxima from a finite number of samples. Here, we will use (4.35) to improve on estimates of local tangent orientations and widths of features from finite samples of coefficients associated with differently scaled and oriented α -molecules.

4.3 Symmetric Molecule-Based Feature Detection

In this section, we will evaluate the likelihood of a certain feature being centered at a point $\mathbf{y} \in \mathbb{R}^2$ in the image domain for a given image $f \in L^2(\mathbb{R}^2)$ by considering the transform of f with respect to systems of even- and odd-symmetric α -molecules and demonstrate how

additional information about the detected features, namely the local tangent orientations of edges and ridges as well as local widths (diameters) of ridges and blobs and heights (contrasts) can be extracted. Formally, we will define three mappings that serve as feature detectors:

$$\text{an edge measure} \quad E(f, \mathbf{y}): L^2(\mathbb{R}^2) \times \mathbb{R}^2 \rightarrow [0, 1], \quad (4.36)$$

$$\text{a ridge measure} \quad R(f, \mathbf{y}): L^2(\mathbb{R}^2) \times \mathbb{R}^2 \rightarrow [0, 1], \quad (4.37)$$

$$\text{and a blob measure} \quad B(f, \mathbf{y}): L^2(\mathbb{R}^2) \times \mathbb{R}^2 \rightarrow [0, 1]. \quad (4.38)$$

Furthermore, we will define functions that extract local properties of a feature centered at a given point in the image domain, namely

$$\text{tangent orientation measures} \quad O_x(f, \mathbf{y}): L^2(\mathbb{R}^2) \times \mathbb{R}^2 \rightarrow \left[-\frac{\pi}{2}, \frac{\pi}{2}\right), \quad x \in \{E, R\}, \quad (4.39)$$

$$\text{local width measures} \quad W_x(f, \mathbf{y}): L^2(\mathbb{R}^2) \times \mathbb{R}^2 \rightarrow \mathbb{R}_+, \quad x \in \{R, B\}, \quad (4.40)$$

$$\text{and local height measures} \quad H_x(f, \mathbf{y}): L^2(\mathbb{R}^2) \times \mathbb{R}^2 \rightarrow \mathbb{R}, \quad x \in \{E, R, B\}. \quad (4.41)$$

In particular, $E(f, \mathbf{y})$, and $H_E(f, \mathbf{y})$ will be defined in Section 4.3.2, $R(f, \mathbf{y})$, $W_R(f, \mathbf{y})$, and $H_R(f, \mathbf{y})$ in Section 4.3.3, $B(f, \mathbf{y})$, $W_B(f, \mathbf{y})$, and $H_B(f, \mathbf{y})$ in Section 4.3.4, and $O_E(f, \mathbf{y})$, and $O_R(f, \mathbf{y})$ in Section 4.3.5.

Before turning to the derivation of the measures in (4.36) to (4.41), we would like to explain the basic principle of symmetric molecule-based feature detection in the one-dimensional setting. Figure 4.4(a) depicts what we consider an ideal one-dimensional edge, namely a step discontinuity preceded and followed by two distinct constant functions. An ideal ridge, which could be described as a short constant function interrupting a constant baseline, is shown in Figure 4.4(b). The features plotted in Figures 4.4(a) and 4.4(b) are both centered at the point $1/2$ and the main question we aim to answer is how this point could be discriminated from its neighborhood and, in general, from other points not lying in the center of a ridge, edge, or another kind of feature. Note that the notions of ridges and blobs coincide in the one-dimensional setting.

One way of looking at this problem is to observe that the function plotted in Figure 4.4(a) is odd-symmetric around $1/2$, while the function shown in Figure 4.4(b) is even-symmetric around $1/2$. This is consistent with the phase congruency-based approach to feature detection, in which edges are characterized by phase congruency with respect to the angle

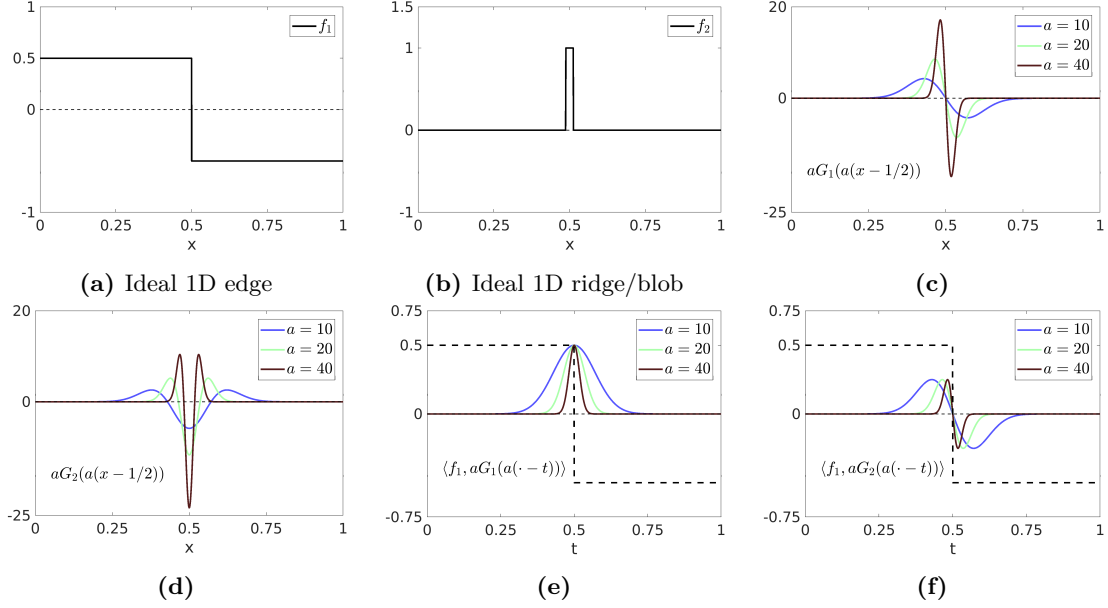


Figure 4.4: Behavior of even- and odd-symmetric coefficients at the location of ideal features. (a): An idealized one-dimensional edge. (b): An idealized one-dimensional ridge/blob. (c): Three different L^1 -normalized dilates of the first derivative of the Gaussian centered at $1/2$. (d): Three different L^1 -normalized dilates of the second derivative of the Gaussian centered at $1/2$. (e): Inner products of the dilates of the first derivative of the Gaussian with an ideal edge plotted as a function of the shift parameter t . (f): Inner products of the dilates of the second derivative of the Gaussian with the ideal edge.

$\pi/2$ and ridges are characterized by phase congruency with respect to the angle 0 (cf. (4.27)). In the phase congruency framework, this simply means that in the vicinity of an edge, a function can purely be represented by odd-symmetric sine components, while in the vicinity of a ridge, a function solely consists of even-symmetric cosine components. For edges, we can additionally observe a certain kind of self-similarity. When restricting the function plotted in Figure 4.4(a) to any neighborhood around the point $1/2$ and then scaling it back to the full interval $[0, 1]$, the function will remain unchanged. Ridges on the other hand are not defined by a single step-discontinuity but by two step-discontinuities which are in close proximity of each other. Any ridge can thus be associated with a width (i.e. the distance between its step-discontinuities) and is not invariant under the restricting and rescaling operation described above. In the one-dimensional setting, we will only deal with the case of edge detection but this issue will be taken into consideration when defining the two-dimensional ridge measure in Section 4.3.3.

To summarize: In order to detect edges, we aim to test each point in the domain of a function for the scale-invariant symmetry and self-similarity properties described in the preceding paragraph. If a point is at the location of an edge, restrictions of the analyzed

function to reasonably small neighborhoods around this point should look roughly the same and, in particular, be odd-symmetric when centered at the analyzed point. A simple and efficient way of testing for these properties is to consider the coefficients associated with differently scaled even- and odd-symmetric wavelets that are centered around the point of interest. Again, this approach is consistent with (4.27), where complex-valued wavelets whose real part is even- and whose imaginary part is odd-symmetric are applied to approximate the phase congruency measure (4.26).

Figures 4.4(c) and 4.4(d) show three differently scaled versions of the odd-symmetric first and the even-symmetric second derivative of the Gaussian (cf. (4.20) and (4.22)), all of which are L^1 -normalized to one. Figures 4.4(e) and 4.4(f) depict the corresponding inner products with the idealized one-dimensional edge as a function of the translation parameter t . For the remainder of this chapter, we shall call inner products of a function f with an even-symmetric wavelet or α -molecule the *even-symmetric coefficients* of f and the inner products with an odd-symmetric wavelet or α -molecule the *odd-symmetric coefficients*. Independently of scale, the even-symmetric coefficients shown in Figure 4.4(f) go to zero at the location of the step-discontinuity, as the even-symmetric wavelets are orthogonal to the locally odd-symmetric function centered at the point $1/2$. Furthermore, the odd-symmetric coefficients all peak in the point $1/2$ with a maximum value of approximately $1/2$, which is due to the fact that the jump size of the singularity is one and because

$$\frac{1}{\|G_1\|_1} \int_{-\infty}^0 G_1(x) \, dx = \frac{G_0(0)}{2G_0(0)} = \frac{1}{2}. \quad (4.42)$$

In summary, Figure 4.4 illustrates that at the location of an ideal edge, L^1 -normalized symmetric wavelets show scale invariant behavior in the sense that odd-symmetric coefficients remain constant under scaling, while even-symmetric coefficients remain fixed at zero. In particular, this behavior is also invariant of the jump size of the singularity and thus unaffected by changes in contrast.

To translate the above observations into an edge measure for functions in $L^2(\mathbb{R})$, we first use the derivatives of the Gaussian and the Hilbert transform defined in Section 4.2.1 to define sets of odd- and even-symmetric L^1 -normalized wavelets:

$$\Psi^e = \left\{ \frac{G_{2k}}{\|G_{2k}\|_1} : k \in \mathbb{N} \right\} \cup \left\{ \frac{\mathcal{H}G_{2k-1}}{\|\mathcal{H}G_{2k-1}\|_1} : k \in \mathbb{N} \right\} \subset L^2(\mathbb{R}) \cap L^1(\mathbb{R}), \quad (4.43)$$

$$\Psi^o = \left\{ \frac{G_{2k-1}}{\|G_{2k-1}\|_1} : k \in \mathbb{N} \right\} \cup \left\{ \frac{\mathcal{H}G_{2k}}{\|\mathcal{H}G_{2k}\|_1} : k \in \mathbb{N} \right\} \subset L^2(\mathbb{R}) \cap L^1(\mathbb{R}). \quad (4.44)$$

When applying generators from the sets Ψ^e and Ψ^o in practice, it will often be sufficient

to numerically approximate the L^1 -norm of the corresponding derivative of the Gaussian, respectively its Hilbert transform. Assuming that the zeros of the associated Hermite polynomial are known, the following lemma shows how the L^1 -norms of derivatives of the Gaussian can also be computed explicitly.

Lemma 4.3.1. *Let G_k denote the k -th derivative of the Gaussian and $z = \{z_n\}_{n=1}^k$ be the strictly increasing sequence of zeros of the associated Hermite polynomial H_k (cf. (4.20)), that is, $H_k(z_n) = 0$ for all $1 \leq n \leq k$ and $z_n < z_{n+1}$ for all $1 \leq n < k-1$. The L^1 -norm of G_k is then given by*

$$\|G_k\|_1 = 2 \left((-1)^{\lfloor k/2 \rfloor} G_{k-1}(0) + 2 \sum_{n=1}^{\lfloor k/2 \rfloor} (-1)^{n-1} G_{k-1}(z_n) \right). \quad (4.45)$$

Proof. As it was already noted, the Hermite polynomials are orthogonal with respect to the measure e^{-x^2} . Their zeros are therefore real and unique (a proof can be found in [159]). It is easy to see from the recursion (4.21), that H_k is of degree k and thus has exactly k real-valued zeros. This means that (4.45) is well defined for all $k \in \mathbb{N}$. Due to $e^{-x^2} > 0$ for all $x \in \mathbb{R}$, the zeros of H_k coincide with the zeros of $G_k(x) = (-1)^k H_k(x) e^{-x^2}$ (cf. (4.20)). Let us denote the number of negative zeros with $\tilde{k} = \lfloor k/2 \rfloor$ and consider the strictly increasing sequence $\tilde{z} = (-\infty, z_1, \dots, z_{\tilde{k}}, 0)$, which partitions \mathbb{R}^- into intervals on which G_k yields either strictly positive or strictly negative values. By exploiting the symmetry of G_k and with the fundamental theorem of calculus, it follows that

$$\frac{\|G_k\|_1}{2} = \int_{-\infty}^0 |G_k(x)| \, dx = \sum_{n=1}^{\tilde{k}+1} \int_{\tilde{z}_n}^{\tilde{z}_{n+1}} |G_k(x)| \, dx = \sum_{n=1}^{\tilde{k}+1} (-1)^{n-1} \int_{\tilde{z}_n}^{\tilde{z}_{n+1}} G_k(x) \, dx \quad (4.46)$$

$$= G_{k-1}(\tilde{z}_2) - \lim_{a \rightarrow -\infty} G_{k-1}(a) + \sum_{n=2}^{\tilde{k}+1} (-1)^{n-1} (G_{k-1}(\tilde{z}_{n+1}) - G_{k-1}(\tilde{z}_n)) \quad (4.47)$$

$$= G_{k-1}(z_1) + (-1)^{\tilde{k}} (G_{k-1}(0) - G_{k-1}(z_{\tilde{k}})) + \sum_{n=1}^{\tilde{k}-1} (-1)^n (G_{k-1}(z_{n+1}) - G_{k-1}(z_n)) \quad (4.48)$$

$$= (-1)^{\tilde{k}} G_{k-1}(0) + 2 \sum_{n=1}^{\tilde{k}} (-1)^{n-1} G_{k-1}(z_n). \quad (4.49)$$

□

Obtaining the exact set of zeros of a given Hermite polynomial is in general not a trivial task. However, at least for the first five Hermite polynomials, the zeros can easily be found using elementary methods, as illustrated by the following lemma.

Lemma 4.3.2. *The sets of zeros for G_1 , G_2 , G_3 , G_4 , and G_5 are*

$$z^{(1)} = \{0\}, \quad (4.50)$$

$$z^{(2)} = \left\{ -\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right\}, \quad (4.51)$$

$$z^{(3)} = \left\{ -\sqrt{\frac{3}{2}}, 0, \sqrt{\frac{3}{2}} \right\}, \quad (4.52)$$

$$z^{(4)} = \left\{ -\sqrt{\frac{3}{2} + \sqrt{\frac{3}{2}}}, -\sqrt{\frac{3}{2} - \sqrt{\frac{3}{2}}}, \sqrt{\frac{3}{2} - \sqrt{\frac{3}{2}}}, \sqrt{\frac{3}{2} + \sqrt{\frac{3}{2}}} \right\}, \quad (4.53)$$

$$z^{(5)} = \left\{ -\sqrt{\frac{5}{2} + \sqrt{\frac{5}{2}}}, -\sqrt{\frac{5}{2} - \sqrt{\frac{5}{2}}}, 0, \sqrt{\frac{5}{2} - \sqrt{\frac{5}{2}}}, \sqrt{\frac{5}{2} + \sqrt{\frac{5}{2}}} \right\}, \quad (4.54)$$

respectively.

Proof. The zeros of $G_k(x) = (-1)^k H_k(x) e^{-x^2}$ coincide with the zeros of the k -th Hermite polynomial H_k . For $H_1(x) = 2x$, the only zero is 0. In the case of $H_2(x) = 4x^2 - 2 = (2x - \sqrt{2})(2x + \sqrt{2})$, the zeros are $\left\{ -\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right\}$. We can factorize

$$H_3(x) = 8x^3 - 12x = (8x^2 - 12)x = \left(\frac{4}{\sqrt{2}}x - \frac{6}{\sqrt{3}} \right) \left(\frac{4}{\sqrt{2}}x + \frac{6}{\sqrt{3}} \right) x, \quad (4.55)$$

which yields the zeros $\left\{ -\sqrt{\frac{3}{2}}, 0, \sqrt{\frac{3}{2}} \right\}$. In the case of $H_4(x) = 16x^4 - 48x^2 + 12$, we consider the normalized polynomial

$$x^2 - 3x + \frac{3}{4} = \left(x - \frac{3}{2} \right)^2 - \frac{3}{2} = \left(x - \sqrt{\frac{3}{2}} + \frac{3}{2} \right) \left(x + \sqrt{\frac{3}{2}} - \frac{3}{2} \right), \quad (4.56)$$

which yields the set of zeros $z^{(4)}$. The complete set of zeros $z^{(5)}$ in the case of $H_5(x) = 32x^5 - 160x^3 + 120x$ is obtained by considering

$$x^2 - 5x + \frac{15}{4} = \left(x - \frac{5}{2} \right)^2 - \frac{5}{2} = \left(x - \sqrt{\frac{5}{2}} + \frac{5}{2} \right) \left(x + \sqrt{\frac{5}{2}} - \frac{5}{2} \right). \quad (4.57)$$

□

Remark 4.3.3. *Using Lemmas 4.3.1 and 4.3.2, the L^1 -norms of the first five derivatives*

of the Gaussian can be computed explicitly:

$$\|G_1\|_1 = 2, \quad (4.58)$$

$$\|G_2\|_1 = 4G_1\left(-\frac{\sqrt{2}}{2}\right) \approx 3.43105554, \quad (4.59)$$

$$\|G_3\|_1 = 4G_2\left(-\sqrt{\frac{3}{2}}\right) - 2G_2(0) \approx 7.57008256, \quad (4.60)$$

$$\|G_4\|_1 = 4G_3\left(-\sqrt{\frac{3}{2} + \sqrt{\frac{3}{2}}}\right) - 4G_3\left(-\sqrt{\frac{3}{2} - \sqrt{\frac{3}{2}}}\right) \approx 19.85573915, \quad (4.61)$$

$$\|G_5\|_1 = 2G_4(0) + 4G_4\left(-\sqrt{\frac{5}{2} + \sqrt{\frac{5}{2}}}\right) - 4G_4\left(-\sqrt{\frac{5}{2} - \sqrt{\frac{5}{2}}}\right) \approx 59.25755290. \quad (4.62)$$

Let us now turn to the definition of a one-dimensional edge measure that is based on even- and odd-symmetric generators from the sets Ψ^e and Ψ^o . Let $\psi^o \in \Psi^o$, $\psi^e \in \Psi^e$, $a > 0$ be fixed, $\mathbf{J} = \{j_n\}_{n=1}^{N_J} \subset \mathbb{Z}$ be an increasing sequence of $N_J \in \mathbb{N}$ scaling parameters and denote L^1 -normalized dilates and shifts of the generating wavelets ψ^o and ψ^e with

$$\psi_{j,y}^o = a^j \psi^o(a^j(\cdot - y)), \quad (4.63)$$

$$\psi_{j,y}^e = a^j \psi^e(a^j(\cdot - y)), \quad (4.64)$$

for scaling parameters $j \in \mathbf{J}$ and translation parameters $y \in \mathbb{R}$. Note that $f \mapsto a^j f(a^j \cdot)$ defines an isometric mapping of L^1 . As already discussed earlier in the case of the first derivative of the Gaussian, the odd-symmetric coefficient at the location of an ideal edge is, independently of scaling, fully determined by the jump size of the discontinuity and the integral of the generating wavelet ψ^o over \mathbb{R}_- . To be precise, the value

$$K_{\psi^o} = \int_{-\infty}^0 \psi^o(x) \, dx, \quad (4.65)$$

is equal to the odd-symmetric coefficients at the location of an ideal edge with jump size one. With a soft-thresholding parameter $\beta > 0$ that corresponds to the minimal jump size that is required for the detection of an edge and a fixed scaling offset for the even-symmetric wavelets $j^e \in \mathbb{R}_{\geq 0}$, a one-dimensional edge measure for a function $f \in L^2(\mathbb{R})$ and $y \in \mathbb{R}$ is

given by

$$\check{E}^{1D}(f, y) = \frac{\left| \sum_{j \in J} \langle f, \psi_{j,y}^o \rangle_{L^2} \right| - \sum_{j \in J} \left| \langle f, \psi_{j-j^e,y}^e \rangle_{L^2} \right| - \beta N_J K_{\psi^o}}{N_J \max_{j \in J} \left\{ \left| \langle f, \psi_{j,y}^o \rangle_{L^2} \right| \right\} + \varepsilon}, \quad (4.66)$$

where $\varepsilon > 0$ prevents division by zero. The measure \check{E}^{1D} is bounded above by one but to ensure a mapping to $[0, 1]$, we finally set

$$E^{1D}(f, y) = \max\{0, \check{E}^{1D}(f, y)\}. \quad (4.67)$$

For example, if a function g has an ideal edge at a point y^* with jump size s , then all odd-symmetric coefficients will virtually¹ be equal to K_{ψ^o} times the jump size s , and all even-symmetric coefficients will be zero. That is, when disregarding ε , the measure $\check{E}^{1D}(g, y^*)$ reduces to

$$\check{E}^{1D}(g, y^*) = \frac{N_J s K_{\psi^o} - \beta N_J K_{\psi^o}}{N_J s K_{\psi^o}} = 1 - \frac{\beta}{s}, \quad (4.68)$$

illustrating that the jump size of edges that can be detected by the measure E^{1D} is indeed bounded below by the parameter β .

The term βK_{ψ^o} in (4.66) can also be seen as a soft-thresholding parameter that is applied to all odd-symmetric coefficients and thereby implicitly denoises the analyzed signal f . In practice, the parameter β is typically chosen as a function of the expected level of noise present in a given signal or image.

Besides the lower bound on s introduced by β , the measure E^{1D} is by construction contrast invariant. Due to normalization with respect to the odd-symmetric coefficient with the greatest magnitude, E^{1D} is independent of the jump size of the discontinuity associated with an edge. In particular, regardless of the jump size, E^{1D} will be close to one at the location of an ideal edge and only diminish at a point y if the analyzed functions have even-symmetric components around y , or if the odd-symmetric coefficients at y have a strong variance with respect to scaling. E^{1D} thus provides a continuous measure of the structural similarity of the neighborhood of a given point with an ideal edge, which sets it categorically apart from most gradient-based approaches to edge detection, that are

¹ Note that the derivatives of the Gaussian are not compactly supported in the time domain (cf. Figure 4.2) and that an ideal edge that separates two different constant functions covering the whole real line is not in $L^2(\mathbb{R})$.

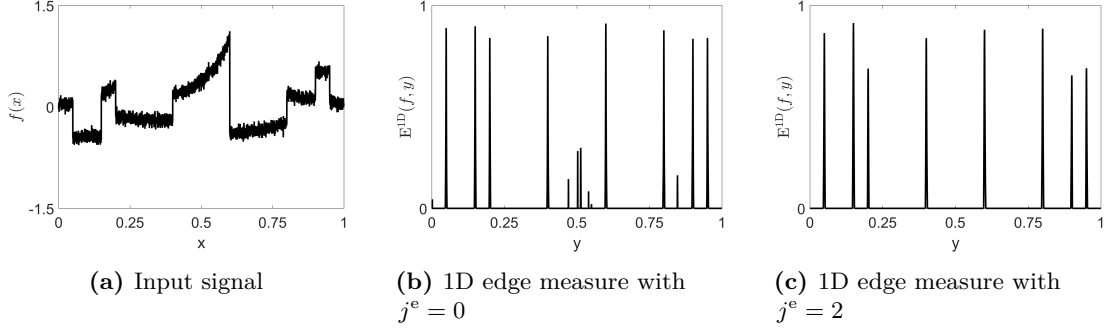


Figure 4.5: Detection of one-dimensional edges. (a): A piecewise polynomial function perturbed with additive Gaussian white noise (computed using [51]). (b): The one-dimensional edge measure (4.67) with parameters $\psi^o = \frac{G_1}{\|G_1\|_1}$, $\psi^e = \frac{\mathcal{H}G_1}{\|\mathcal{H}G_1\|_1}$, $a = 2$, $N_J = 4$, $\beta = 0.03$ and $j^e = 0$. (c): The one-dimensional edge measure (4.67) with a scaling offset $j^e = 2$ (other parameters are the same as in 4.5(b)).

measuring jump sizes.

Broadly speaking, the odd-symmetric coefficients in (4.66) serve as evidence in favor of – while the even-symmetric coefficients serve as evidence against – the presence of an edge at a given location. To suppress the emergence of false positives in the extended neighborhood of an actual edge, it has proven useful to consider even-symmetric wavelets that have lower frequencies than their odd-symmetric counterparts, which can be achieved by setting the scale offset parameter j^e to a real value greater than 0. The effect of the scale offset parameter is illustrated in Figure 4.5, which visualizes the one-dimensional edge measure E^{1D} computed on a piecewise polynomial signal with additive Gaussian noise. The odd- and even-symmetric wavelets were chosen to be the first derivative of the Gaussian and its Hilbert transform, that is $\psi^o = \frac{G_1}{\|G_1\|_1}$ and $\psi^e = \frac{\mathcal{H}G_1}{\|\mathcal{H}G_1\|_1}$. The other parameters were selected as follows: $a = 2$, $N_J = 4$, $\beta = 0.03$, $j^e = 0$ for the plot in Figure 4.5(b) and $j^e = 2$ for the plot in Figure 4.5(c).

The main difference between edges and ridges is that edges are locally odd-symmetric features while ridges are locally even-symmetric features. Thus, to obtain a one-dimensional ridge measure, it almost suffices to simply switch the roles of the even- and odd-symmetric wavelets in (4.66). However, due to the fact that ridges, in contrast to edges, are associated with widths, even-symmetric coefficients are not scaling invariant at the location of an ideal ridge as it is the case for odd-symmetric coefficients at the location of an ideal edge. As the one-dimensional edge measure was mainly derived to illustrate the general concept behind this work, a ridge measure will only be rigorously defined for the two-dimensional case in Section 4.3.3.

4.3.1 Symmetric α -Molecules for Edge, Ridge, and Blob Detection

We now generalize the one-dimensional edge measure E^{1D} to edges, ridges and blobs in the two-dimensional setting. With small adjustments, the symmetry and self-similarity properties that characterize edges and ridges in the one-dimensional setting can also be utilized to extract features from two-dimensional signals. A function describing a two-dimensional image at a point lying on an edge is locally odd-symmetric with respect to the tangent of the edge contour. On the other hand, the tangents of the contour of a ridge define axes of symmetry associated with even-symmetry while a function is locally point symmetric at the center of a blob. In order to test for two-dimensional symmetry properties, we will consider generating functions in $L^2(\mathbb{R}^2) \cap L^1(\mathbb{R}^2)$ that are odd- or even-symmetric wavelets in one dimension or defined by even-symmetric wavelets in both dimensions, where the latter will only be applied in the case of blob detection. A simple and efficient way of obtaining such generators is to consider tensor product constructions based on the unnormalized one-dimensional Gaussian G_0 as well as even- and odd-symmetric one-dimensional wavelets from the sets Ψ^e and Ψ^o . As mentioned in Section 4.2.1, we will use derivatives of the Gaussian and their Hilbert transforms as the generating one-dimensional odd- and even-symmetric L^1 -normalized wavelets to detect features by analyzing the local symmetry properties of a function. To this end, we define the sets

$$\Psi_2^e = \left\{ c_1 c_2 \pi^{-\frac{1}{2}} \psi^e(c_1 \cdot) G_0(c_2 \cdot) : \psi^e \in \Psi^e, c_1, c_2 \in \mathbb{R}_+ \right\} \subset L^2(\mathbb{R}^2) \cap L^1(\mathbb{R}^2), \quad (4.69)$$

$$\Psi_2^o = \left\{ c_1 c_2 \pi^{-\frac{1}{2}} \psi^o(c_1 \cdot) G_0(c_2 \cdot) : \psi^o \in \Psi^o, c_1, c_2 \in \mathbb{R}_+ \right\} \subset L^2(\mathbb{R}^2) \cap L^1(\mathbb{R}^2), \quad (4.70)$$

that contain two-dimensional wavelet generators that are based on even- and odd-symmetric one-dimensional wavelets and the Gaussian G_0 , while the generators in

$$\widetilde{\Psi}_2^e = \{ c_1 c_2 \psi^e(c_1 \cdot) \psi^e(c_2 \cdot) : \psi^e \in \Psi^e, c_1, c_2 \in \mathbb{R}_+ \} \subset L^2(\mathbb{R}^2) \cap L^1(\mathbb{R}^2) \quad (4.71)$$

are even-symmetric and have vanishing moments in both dimensions. The parameters c_1 and c_2 can be used to adjust the scale of the even- and odd-symmetric generators as well as their aspect ratio. Examples of two-dimensional symmetric generators contained in Ψ_2^e , Ψ_2^o , and $\widetilde{\Psi}_2^e$ are depicted in Figure 4.6.

Analogous to the one-dimensional case, we require sets of differently scaled analyzing functions to test for scale-invariant local symmetry properties. However, the local symmetry at the location of a two-dimensional feature is typically only given with respect to a specific symmetry axes. Thus, the analysis of two-dimensional signals cannot solely be based on differently scaled wavelet-like functions but also requires a means of changing their

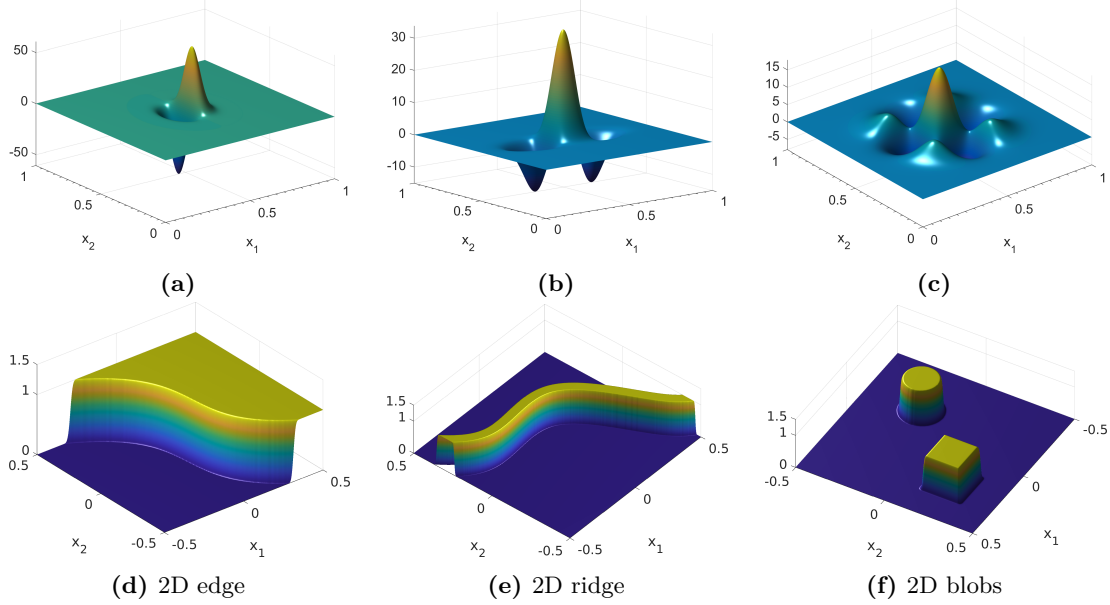


Figure 4.6: Examples of L^1 -normalized two-dimensional symmetric generators and ideal two-dimensional features. (a): Odd-symmetric generator given by a separable product of G_1 and G_0 . (b): Even-symmetric generator based on G_2 and G_0 . (c): Even-symmetric generator given by the tensor product of G_2 with itself. (d): Example of an ideal edge. (e): Example of an ideal ridge. (f): Examples of ideal blobs.

preferred orientation. A natural way of obtaining systems of dilations, rotations and shifts of a single two-dimensional generator is to consider the α -molecule framework [71] discussed in the prelude to this chapter. Let $\alpha \in [0, 1]$ be fixed, $g \in \Psi_2^e \cup \Psi_2^o \cup \widetilde{\Psi}_2^e$ be a symmetric generator, $a > 0$, $J \subset \mathbb{Z}$ be a set of scaling parameters and $\Theta \subset [-\frac{\pi}{2}, \frac{\pi}{2})$ be a set of scaling parameters, then the corresponding system of L^1 -normalized symmetric molecules is given by

$$SM(g, \alpha, a, J, \Theta) = \left\{ m_{j, \theta, \mathbf{y}} = a^{j(1+\alpha)} g(\mathbf{A}_{a^j, \alpha} \mathbf{R}_\theta(\cdot - \mathbf{y})) : j \in J, \theta \in \Theta, \mathbf{y} \in \mathbb{R}^2 \right\}. \quad (4.72)$$

The parameter α describes the degree of anisotropy introduced to a system by the scaling matrix $\mathbf{A}_{a^j, \alpha}$ and should be chosen as a function of the expected level of noise and the regularity of the structures that are to be detected. If a signal was perturbed with severe noise but the contours of its edges or ridges are smooth, introducing a high degree of anisotropic scaling by choosing α close to zero can significantly improve the detection performance. However, setting α close to zero can lead to problems at locations where the contour of a feature is irregular and can thus not be well approximated by a linear function. Note that the elements in (4.72) are L^1 -normalized, while classical systems of α -molecules are normalized with respect to the L^2 -norm (cf. (4.4)).

The following theorem specifies the order of symmetric molecule systems obtained from generators defined by (4.70), (4.69), and (4.71) in terms of a localization parameter L , a vanishing moments parameter M and smoothness parameters N_1 and N_2 (cf. (4.5)).

Theorem 4.3.4. *Let $k \in \mathbb{N}$ and denote four types of generators by*

$$g_k^{(1)}(\mathbf{x}) = G_k(x_1)G_0(x_2), \quad g_k^{(2)}(\mathbf{x}) = \mathcal{H}G_k(x_1)G_0(x_2), \quad (4.73)$$

$$g_k^{(3)}(\mathbf{x}) = G_k(x_1)G_k(x_2), \quad g_k^{(4)}(\mathbf{x}) = \mathcal{H}G_k(x_1)\mathcal{H}G_k(x_2). \quad (4.74)$$

Let $\alpha \in [0, 1]$, $a > 0$, $J \subset \mathbb{Z}$ be a set of scaling parameters and $\Theta \subset [-\frac{\pi}{2}, \frac{\pi}{2})$ a set of rotation parameters then it holds for all $\boldsymbol{\rho} \in \mathbb{N}_0^2$, $M \leq k$ and $N_1, N_2 \in \mathbb{N}_0$ that

$$\left| \partial^{\boldsymbol{\rho}} \widehat{g_k^{(1)}}(\boldsymbol{\xi}) \right| \lesssim \min \left\{ 1, a^{-j} + |\xi_1| + a^{-j(1-\alpha)} |\xi_2| \right\}^M \cdot (1 + \|\boldsymbol{\xi}\|_2^2)^{-\frac{N_1}{2}} \cdot (1 + (\xi_2)^2)^{-\frac{N_2}{2}}, \quad (4.75)$$

where the implicit constants are independent of the choice of parameters $j \in J$ and we call the associated set $SM(\|g_k^{(1)}\|_1^{-1}g_k^{(1)}, \alpha, a, J, \Theta)$ a system of L^1 -normalized symmetric α -molecules of order $(\infty, k, \infty, \infty)$. Furthermore, with α , a , J and Θ as above, the set $SM(\|g_k^{(3)}\|_1^{-1}g_k^{(3)}, \alpha, a, J, \Theta)$ is also a system of L^1 -normalized α -molecules of order $(\infty, k, \infty, \infty)$, while the systems $SM(\|g_k^{(2)}\|_1^{-1}g_k^{(2)}, \alpha, a, J, \Theta)$ and $SM(\|g_k^{(4)}\|_1^{-1}g_k^{(4)}, \alpha, a, J, \Theta)$ are of order $(k-1, k, \infty, \infty)$.

Proof. Let $\boldsymbol{\rho} \in \mathbb{N}_0^2$ and $N_1, N_2 \in \mathbb{N}_0$, then we can use that $e^{-x^2} \lesssim (1 + |x|)^{-N}$ for all $N \in \mathbb{N}_0$, and write

$$\left| \partial^{\boldsymbol{\rho}} \widehat{g_k^{(1)}}(\boldsymbol{\xi}) \right| = \left| \partial^{\boldsymbol{\rho}} \left((2\pi i \xi_1)^k \pi e^{-\pi^2(\xi_1)^2} e^{-\pi^2(\xi_2)^2} \right) \right| \quad (4.76)$$

$$= \left| (2i)^k \pi^{k+1} \left(\partial_1^{\rho_1} (\xi_1)^k e^{-\pi^2(\xi_1)^2} \right) \left(\partial_2^{\rho_2} e^{-\pi^2(\xi_2)^2} \right) \right| \quad (4.77)$$

$$\asymp |\xi_1|^{k+\rho_1} e^{-\pi^2(\xi_1)^2} \cdot |\xi_2|^{\rho_2} e^{-\pi^2(\xi_2)^2} \quad (4.78)$$

$$\lesssim |\xi_1|^{k+\rho_1} |\xi_2|^{\rho_2} (1 + |\xi_1|)^{-(N_1+k+\rho_1)} (1 + |\xi_2|)^{-(N_1+N_2+\rho_2)} \quad (4.79)$$

$$= \frac{|\xi_1|^{k+\rho_1}}{(1 + |\xi_1|)^{k+\rho_1}} \frac{|\xi_2|^{\rho_2}}{(1 + |\xi_2|)^{\rho_2}} (1 + |\xi_1|)^{-N_1} (1 + |\xi_2|)^{-(N_1+N_2)} \quad (4.80)$$

$$\leq \min \left\{ 1, |\xi_1|^{k+\rho_1} |\xi_2|^{\rho_2} \right\} (1 + |\xi_1|)^{-N_1} (1 + |\xi_2|)^{-(N_1+N_2)} \quad (4.81)$$

$$\lesssim \min \left\{ 1, a^{-j} + |\xi_1| + a^{-j(1-\alpha)} |\xi_2| \right\}^k (1 + \|\boldsymbol{\xi}\|_2^2)^{-\frac{N_1}{2}} \cdot (1 + (\xi_2)^2)^{-\frac{N_2}{2}}, \quad (4.82)$$

independently of the choice of parameters $a > 0$, $j \in \mathbb{Z}$, and $\alpha \in [0, 1]$. This shows (4.75) and a similar argument leads to the same properties for the generator $g_k^{(3)}$.

For $g_k^{(2)}$, we consider $|\boldsymbol{\rho}| = \rho_1 + \rho_2 < k$, $N_1, N_2 \in \mathbb{N}_0$ and write

$$\left| \partial^\rho \widehat{g_k^{(2)}}(\boldsymbol{\xi}) \right| = \left| \partial^\rho \left(-i \operatorname{sgn}(\xi_1) (2\pi i \xi_1)^k \pi e^{-\pi^2(\xi_1)^2} e^{-\pi^2(\xi_2)^2} \right) \right| \quad (4.83)$$

$$= \left| (2i)^k \pi^{k+1} \left(\partial_1^{\rho_1} |\xi_1| (\xi_1)^{k-1} e^{-\pi^2(\xi_1)^2} \right) \left(\partial_2^{\rho_2} e^{-\pi^2(\xi_2)^2} \right) \right|, \quad (4.84)$$

where $\partial_1^{\rho_1} |\xi_1| (\xi_1)^{k-1} e^{-\pi^2(\xi_1)^2}$ only exists for $\rho_1 < k$. By repeating the same steps as in the case of $\widehat{g_k^{(1)}}$, we can conclude that

$$\left| \partial^\rho \widehat{g_k^{(2)}}(\boldsymbol{\xi}) \right| \lesssim \min \left\{ 1, a^{-j} + |\xi_1| + a^{-j(1-\alpha)} |\xi_2| \right\}^k (1 + \|\boldsymbol{\xi}\|_2^2)^{-\frac{N_1}{2}} \cdot (1 + (\xi_2)^2)^{-\frac{N_2}{2}}, \quad (4.85)$$

independently of the choice of parameters $a > 0$, $j \in \mathbb{Z}$ and $\alpha \in [0, 1]$. The same argument can be used to show that $SM(\|g_k^{(4)}\|_1^{-1} g_k^{(4)}, \alpha, a, J, \Theta)$ is of order $(k-1, k, \infty, \infty)$. \square

Theorem 4.3.4 illustrates that symmetric molecule systems based on derivatives of the Gaussian are well localized and smooth in the time domain while the number of vanishing moments increases with the number of derivatives. However, it also shows that changing the symmetry properties of a generator by applying the Hilbert transform reduces smoothness in the Fourier domain and thus localization in the time domain.

4.3.2 Edge Detection

In the neighborhood of a point \mathbf{y} that lies on an ideal edge (cf. Figure 4.6(d)), the image is locally odd-symmetric with respect to a symmetry axis defined by the tangent of the edge contour. The measure E^{1D} can thus be generalized to yield a two-dimensional edge measure by restricting the analysis to symmetric molecules whose direction of vanishing moments is orthogonal to the tangent of the edge contour going through \mathbf{y} . Let $g^o \in \Psi_2^o$, $g^e \in \Psi_2^e$ be a pair of odd- and even-symmetric generators, $\alpha \in [0, 1]$, $a > 0$, $\mathbf{J} = \{j_n\}_{n=1}^{N_J} \subset \mathbb{Z}$ be an increasing sequence of $N_J \in \mathbb{N}$ scaling parameters, and $\Theta \subset \mathbb{T}$ be a set of orientation parameters. For the remainder of this section, we will denote elements from the symmetric molecule system $SM(g^o, \alpha, a, \mathbf{J}, \Theta)$ with $m_{j,\theta,\mathbf{y}}^o$ and elements from the system $SM(g^e, \alpha, a, \mathbf{J}, \Theta)$ with $m_{j,\theta,\mathbf{y}}^e$ (cf. (4.72)). For an image defined by a two-dimensional function $f \in L^2(\mathbb{R}^2)$, we first denote the most significant scaling and orientation parameters at a point $\mathbf{y} \in \mathbb{R}^2$ by

$$(j^*(\mathbf{y}), \theta^*(\mathbf{y})) = \operatorname{argmax}_{(j,\theta) \in \mathbf{J} \times \Theta} \left| \langle f, m_{j,\theta,\mathbf{y}}^o \rangle_{L^2} \right|. \quad (4.86)$$

Then, with a soft-thresholding parameter $\beta > 0$ and a scaling offset parameter $j^e \in \mathbb{R}_{\geq 0}$, a two-dimensional edge measure is given by

$$\check{E}(f, \mathbf{y}) = \frac{\left| \sum_{j \in \mathbf{J}} \left\langle f, m_{j, \theta^*(\mathbf{y}), \mathbf{y}}^o \right\rangle_{L^2} \right| - \sum_{j \in \mathbf{J}} \left| \left\langle f, m_{j-j^e, \theta^*(\mathbf{y}), \mathbf{y}}^e \right\rangle_{L^2} \right| - \beta N_J K_{\psi^o}}{N_J \left| \left\langle f, m_{j^*, \theta^*(\mathbf{y}), \mathbf{y}}^o \right\rangle_{L^2} \right| + \varepsilon}, \quad (4.87)$$

where $\varepsilon > 0$ prevents division by zero and K_{ψ^o} (as defined in (4.65)) corresponds to the odd-symmetric coefficient at the location of an ideal edge with jump-size one. As in the one-dimensional case, we finally set

$$E(f, \mathbf{y}) = \max\{0, \check{E}(f, \mathbf{y})\}, \quad (4.88)$$

to ensure that the measure maps to $[0, 1]$.

As we already pointed out during the derivation of the one-dimensional edge measure $E^{1D}(f, y)$ earlier in this section (cf. (4.66)), the measure (4.87) is partly based on the observation that the coefficients obtained from an odd-symmetric and L^1 -normalized analyzing function are invariant to scaling if it is spatially and directionally aligned with an ideal edge. In this regard, we would like to emphasize an interesting connection to other shearlet-based approaches to edge detection, which utilize the special decay behavior of L^2 -normalized shearlets. It was shown in [75] that the coefficients of band-limited shearlets at a point that lies on a smooth boundary curve, and whose local tangent direction is aligned with the considered shearlet, decay with $O(s^{-3/4})$ for $s \rightarrow \infty$, where s denotes the scaling parameter (cf. (4.4)). Similar results have also been shown in the case of compactly supported shearlet generators [108]. This is related to the measures proposed here in the sense that, when considering a two-dimensional generator $\psi \in L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$ and parabolic scaling (i.e., $\alpha = 1/2$), it holds that

$$\frac{\|\psi(\mathbf{A}_{s, 1/2} \cdot)\|_2}{\|\psi(\mathbf{A}_{s, 1/2} \cdot)\|_1} = \frac{s^{-3/4} \|\psi\|_2}{s^{-3/2} \|\psi\|_1} = s^{3/4} \frac{\|\psi\|_2}{\|\psi\|_1}, \quad s > 0. \quad (4.89)$$

In other words, considering L^1 - instead of L^2 -normalization precisely neutralizes the decay of shearlet coefficients at points that lie on smooth boundary curves.

Further note that, due to $\|m^o\|_1 = 1$ for all $m^o \in SM(g^o, \alpha, a, \mathbf{J}, \Theta)$, the most significant scale parameter $j^*(\mathbf{y})$ at a point y can directly be related to the contrast (i.e., the height) of an edge going through \mathbf{y} . By considering K_{ψ^o} (cf. (4.65)), which denotes the value of an odd-symmetric coefficient at the location of an ideal edge with jump size one, we define

the local height measure for edges as

$$H_E(f, \mathbf{y}) = \frac{\left\langle f, m_{j^*, \theta^*(\mathbf{y}), \mathbf{y}}^o \right\rangle_{L^2}}{K_{\psi^o}}. \quad (4.90)$$

4.3.3 Ridge Detection

For ridge detection, we can exploit the fact that in the neighborhood of a point that lies on the centerline of an ideal ridge (cf. Figure 4.6(e)), the image is locally even-symmetric with respect to a symmetry axis defined by the tangent of the centerline. A two-dimensional ridge measure can in principle be obtained by simply interchanging the roles of the odd- and even-symmetric molecules m^o and m^e in the definition of the edge measure $E(f, \mathbf{y})$. However, to fully retain contrast invariance, it is necessary to also take into account the width of the ridge, that is, the distance between the two jump singularities that define a ridge in an idealized setting. Contrary to the case of edge detection, even-symmetric coefficients of L^1 -normalized molecules that are centered around the location of a ridge are not invariant to scaling but depend on the scaling parameter as well as the width of the respective ridge. This behavior is illustrated with respect to ridges of three different widths in Figure 4.7. For an even-symmetric generator $g^e \in \Psi_2^e$, let us denote the coefficient at the location of an ideal ridge whose tangent direction in \mathbf{y} agrees with the orientation of g^e and that has width $2r$ and height 1 by

$$K_{g^e}(r) = K_{\psi^e}(r) = \int_{-r}^r \psi^e(x) \, dx, \quad (4.91)$$

where ψ^e is the one-dimensional even-symmetric wavelet associated with the two-dimensional generator g^e (cf. (4.70)). We furthermore define the radius of g^e as half of the width of the ideal ridge which yields the most significant coefficient, that is,

$$r_{g^e} = \operatorname{argmax}_{r \in \mathbb{R}} |K_{g^e}(r)|, \quad (4.92)$$

and extend this notion to arbitrarily shifted and rotated molecules by setting $r_{m_{0,\theta,\mathbf{y}}^e} = r_{m_{0,0,0}^e} = r_{g^e}$. The radius of a dilated molecule can directly be computed as a function of the scale parameter and the underlying generator, that is, $r_{m_{j,\theta,\mathbf{y}}^e} = a^{-j} r_{m_{0,0,0}^e} = a^{-j} r_{g^e}$. Note that in most practical cases, the radius r_{g^e} is equivalent to the distance between the origin and the first zero crossing of the associated one-dimensional wavelet ψ^e (cf. Figure 4.7(a)). It is, for instance, easy to show this property in the case of the even-symmetric generators $\frac{G_2}{\|G_2\|_1}$ and $\frac{G_4}{\|G_4\|_1}$, for which we have already obtained the exact sets of zeros in Lemma 4.3.2.

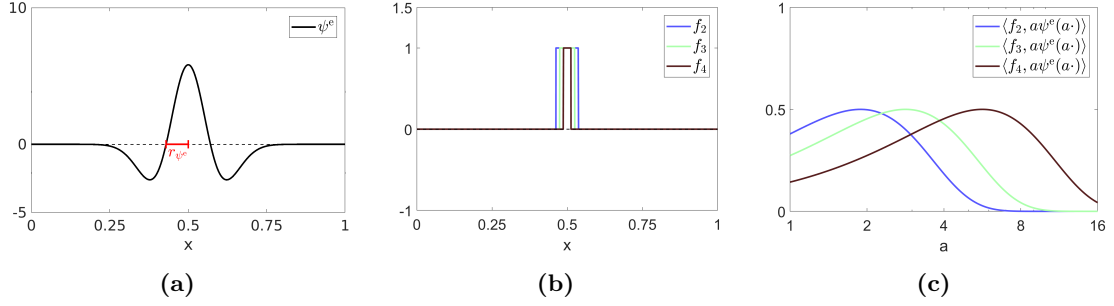


Figure 4.7: Behavior of even-symmetric coefficients at the location of an ideal ridge. (a): The radius r_{ψ^e} of a one-dimensional even-symmetric wavelet, as defined in (4.92). (b): Three ideal one-dimensional ridges with different widths. (c): The inner products of ψ^e with the ideal ridges plotted as a function of the scaling parameter a .

Theorem 4.3.5. Let $z^{(2)}$ and $z^{(4)}$ denote the zeros of G_2 and G_4 , respectively, as in Lemma 4.3.2. Then,

$$\operatorname{argmax}_{r \in \mathbb{R}} |K_{G_2}(r)| = z_2^{(2)} = \frac{\sqrt{2}}{2}, \quad (4.93)$$

$$\operatorname{argmax}_{r \in \mathbb{R}} |K_{G_4}(r)| = z_3^{(4)} = \sqrt{\frac{3}{2}} - \sqrt{\frac{3}{2}}. \quad (4.94)$$

Proof. The second derivative of the Gaussian G_2 is strictly positive on $(-\infty, -z_2^{(2)}) \cup (z_2^{(2)}, \infty)$ and strictly negative on $(-z_2^{(2)}, z_2^{(2)})$. The unique argument of the maximum of $r \mapsto |K_{G_2}(r)|$ is therefore $z_2^{(2)} = \frac{\sqrt{2}}{2}$. G_4 is strictly positive on $(-\infty, -z_4^{(4)}) \cup (-z_3^{(4)}, z_3^{(4)}) \cup (z_4^{(4)}, \infty)$ and strictly negative on $(-z_4^{(4)}, -z_3^{(4)}) \cup (z_3^{(4)}, z_4^{(4)})$. Due to

$$G_3(-z_3^{(4)}) - G_3(-z_4^{(4)}) + G_3(-z_4^{(4)}) \approx -3.9 < 0, \quad (4.95)$$

$z_3^{(4)} = \sqrt{\frac{3}{2}} - \sqrt{\frac{3}{2}}$ is the unique argument of the maximum of $r \mapsto |K_{G_4}(r)|$. \square

Let $g^e \in \Psi_2^e$, $g^o \in \Psi_2^o$ be a pair of odd-symmetric and even-symmetric generators, $\alpha \in [0, 1]$, $a > 0$, $\mathbf{J} = \{j_n\}_{n=1}^{N_J} \subset \mathbb{Z}$ be an increasing sequence of $N_J \in \mathbb{N}$ scaling parameters, and $\Theta \subset \mathbb{T}$ be a set of orientation parameters. Furthermore let $f \in L^2(\mathbb{R}^2)$ be a two-dimensional image. Analogous to the edge detection case, we denote the scaling and rotation parameters associated with the most significant even-symmetric molecule at a point $\mathbf{y} \in \mathbb{R}^2$ in the image plane by

$$(j^*(\mathbf{y}), \theta^*(\mathbf{y})) = \operatorname{argmax}_{(j, \theta) \in \mathbf{J} \times \Theta} \left| \langle f, m_{j, \theta, \mathbf{y}}^e \rangle_{L^2} \right|. \quad (4.96)$$

For each point \mathbf{y} , we first estimate the width of an assumed ideal ridge whose centerline passes through \mathbf{y} and whose tangent direction at \mathbf{y} agrees with the most significant orientation parameter $\theta^*(\mathbf{y})$. We use $j^-(\mathbf{y})$ and $j^+(\mathbf{y})$ to denote the scaling parameters in the strictly increasing sequence \mathbf{J} that precede, respectively succeed, the most significant scale $j^*(\mathbf{y})$. The width measure is obtained by computing the maximum point of a parabola fit through the points $\left(a^{-j^-(\mathbf{y})}r_{g^e}, \left\langle f, m_{j^-(\mathbf{y}), \theta^*(\mathbf{y}), \mathbf{y}}^e \right\rangle_{L^2}\right)$, $\left(a^{-j^*(\mathbf{y})}r_{g^e}, \left\langle f, m_{j^*(\mathbf{y}), \theta^*(\mathbf{y}), \mathbf{y}}^e \right\rangle_{L^2}\right)$, and $\left(a^{-j^+(\mathbf{y})}r_{g^e}, \left\langle f, m_{j^+(\mathbf{y}), \theta^*(\mathbf{y}), \mathbf{y}}^e \right\rangle_{L^2}\right)$. Following (4.35), we define the local width measure as

$$W_R(f, \mathbf{y}) = 2 \left(a^{\frac{c_1}{2c_2} - j^*(\mathbf{y})} r_{g^e} \right), \quad (4.97)$$

where the values c_1 and c_2 are chosen according to (4.33) and (4.34).

Remark 4.3.6. Let $\mathbf{x} \in \mathbb{R}^2$ be fixed and assume that $g^e \in C^2(\mathbb{R}^2)$, then the mapping

$$s \mapsto s^{1+\alpha} g^e(\mathbf{A}_{s,\alpha} \mathbf{R}_{\theta^*(\mathbf{y})}(\mathbf{x} - \mathbf{y})) f(\mathbf{x}) \quad (4.98)$$

is also C^2 on \mathbb{R}_+ . This implies that the mapping

$$s \mapsto \left\langle f, m_{a, \theta^*(\mathbf{y}), \mathbf{y}}^e \right\rangle_{L^2} = s^{1+\alpha} \int_{\mathbb{R}^2} g^e(\mathbf{A}_{s,\alpha} \mathbf{R}_{\theta^*(\mathbf{y})}(\mathbf{x} - \mathbf{y})) f(\mathbf{x}) \, d\mathbf{x} \quad (4.99)$$

is also C^2 on closed intervals in \mathbb{R}_+ , which is important for the feasibility of the refinement procedure for $W_R(f, \mathbf{y})$ described above.

Under the assumption that an ideal ridge with width $W_R(f, \mathbf{y})$, height 1, and a local tangent direction that agrees with $\theta^*(\mathbf{y})$ passes through \mathbf{y} , we can directly compute the associated even-symmetric coefficients of differently dilated molecules with rotation parameter $\theta^*(\mathbf{y})$ and denote

$$K_{g^e}(f, j, \mathbf{y}) = K_{g^e} \left(a^j W_R(f, \mathbf{y})/2 \right) = \int_{-a^j W_R(f, \mathbf{y})/2}^{a^j W_R(f, \mathbf{y})/2} \psi^e(x) \, dx. \quad (4.100)$$

Based on the estimated width $W_R(f, \mathbf{y})$, we can further define a measure that computes the height of an assumed ideal ridge whose centerline passes through \mathbf{y} as the ratio between the most significant coefficient and the value $K_{g^e}(f, j^*(\mathbf{y}), \mathbf{y})$, that is,

$$H_R(f, \mathbf{y}) = \frac{\left\langle f, m_{j^*(\mathbf{y}), \theta^*(\mathbf{y}), \mathbf{y}}^e \right\rangle_{L^2}}{K_{g^e}(f, j^*(\mathbf{y}), \mathbf{y})}. \quad (4.101)$$

A two-dimensional ridge measure can now be defined by carefully modifying the edge measure $E(f, \mathbf{y})$. To detect locally even-symmetric features instead of odd-symmetric ones, the roles of the even-symmetric molecules m^e and odd-symmetric molecules m^o need to be interchanged. While the odd-symmetric coefficients at the location of an ideal edge are invariant to scaling, this is not the case for the even-symmetric coefficients at the location of an ideal ridge. This can be taken into account by adjusting the normalizing factor for every scaling parameter $j \in \mathbf{J}$ using the values $H_R(f, \mathbf{y})K_{g^e}(f, j, \mathbf{y})$, which correspond to the even-symmetric coefficients at the location of an ideal ridge with width $W_R(f, \mathbf{y})$ and height $H_R(f, \mathbf{y})$. Finally, when using strongly oscillating generators, even the sign of the even-symmetric coefficients at the location of an ideal ridge is not invariant to scaling. This can be addressed by considering the expected sign $\text{sgn}(H_R(f, \mathbf{y})K_{g^e}(f, j, \mathbf{y}))$ for each scale parameter when taking the sum of the even-symmetric coefficients. To simplify our notation, we write $h(f, j, \mathbf{y}) = H_R(f, \mathbf{y})K_{g^e}(f, j, \mathbf{y})$ and a two-dimensional ridge measure is now given by

$$\check{R}(f, \mathbf{y}) = \frac{\sum_{j \in \mathbf{J}} \text{sgn}(h(f, j, \mathbf{y})) \langle f, m_{j, \theta^*(\mathbf{y}), \mathbf{y}}^e \rangle_{L^2} - \sum_{j \in \mathbf{J}} |\langle f, m_{j-j^o, \theta^*(\mathbf{y}), \mathbf{y}}^o \rangle_{L^2}| - \beta N_J |K_{g^e}(f, j^*(\mathbf{y}), \mathbf{y})|}{\sum_{j \in \mathbf{J}} \max\{|\langle f, m_{j, \theta^*(\mathbf{y}), \mathbf{y}}^e \rangle_{L^2}|, |h(f, j, \mathbf{y})|\} + \varepsilon}, \quad (4.102)$$

where $\varepsilon > 0$ prevents division by zero, $\beta > 0$ is a soft-thresholding parameter, and $j^o \in \mathbb{R}_{\geq 0}$ denotes the scaling offset between even- and odd-symmetric molecules. As in the case of edge detection, we finally set

$$R(f, \mathbf{y}) = \{0, \check{R}(f, \mathbf{y})\}. \quad (4.103)$$

4.3.4 Blob Detection

The measures $E(f, \mathbf{y})$ and $R(f, \mathbf{y})$ detect features that exhibit symmetry properties that are locally defined by a singular symmetry axis, namely the normal of the tangent of an edge contour or the centerline of a ridge. A similar approach can be used to detect features in two-dimensional images that have less anisotropic symmetry properties in the sense that locally, the center of such features is a point of symmetry for more than one or even all possible directions. Let us consider an idealized blob as a filled square, which has four axes of even symmetry, or a filled circle, for which each line that passes through its center is an axis of even symmetry (cf. Figure 4.6(f)).

To detect such blobs, we will use a generator $\tilde{g}^e \in \widetilde{\Psi}_2^e$ that is defined by the tensor product of an even-symmetric wavelet with itself (cf. (4.71)). Like a square, such a generator has by construction four axes of symmetry and is invariant to rotations of $\frac{\pi}{2}$. Let

us denote the coefficient at the location of an ideal square-shaped blob with side length (width) $2r$ and height 1 by

$$\tilde{K}_{\tilde{g}^e}(r) = \int_{-r}^r \int_{-r}^r \psi^e(x_1) \psi^e(x_2) \, dx_2 \, dx_1 = K_{\tilde{g}^e}(r)^2, \quad (4.104)$$

and retain the definition of the radius of \tilde{g}^e as half of the side length of the square-shaped blob which maximizes its response, that is,

$$r_{\tilde{g}^e} = \operatorname{argmax}_{r \in \mathbb{R}} K_{\tilde{g}^e}(r)^2 = \operatorname{argmax}_{r \in \mathbb{R}} |K_{\tilde{g}^e}(r)|. \quad (4.105)$$

Let $g^o \in \Psi_2^o$ be an odd-symmetric generator, $a > 0$, $\mathbf{J} = \{j_n\}_{n=1}^{N_J} \subset \mathbb{Z}$ be an increasing sequence of $N_J \in \mathbb{N}$ scaling parameters, and $\Theta \subset \mathbb{T}$ be a set of orientation parameters. Due to the isotropic nature of the features we aim to detect, and contrary to the cases of edge and ridge detection, we set $\alpha = 1$ and denote elements from the molecule system $SM(\tilde{g}^e, 1, a, \mathbf{J}, \Theta)$ with $\tilde{m}_{j,\theta,\mathbf{y}}^e$ and elements from the system $SM(g^o, 1, a, \mathbf{J}, \Theta)$ with $m_{j,\theta,\mathbf{y}}^o$. For a two-dimensional image $f \in L^2(\mathbb{R}^2)$, we again denote the most significant pair of scaling and rotation parameters with respect to the even-symmetric generator \tilde{g}^e at a point $\mathbf{y} \in \mathbb{R}^2$ with

$$(j^*, \theta^*) = \operatorname{argmax}_{(j,\theta) \in \mathbf{J} \times \Theta} \left| \left\langle f, \tilde{m}_{j,\theta,\mathbf{y}}^e \right\rangle_{L^2} \right|. \quad (4.106)$$

As in the case of ridge detection, the even-symmetric coefficients at the center of a blob are not invariant to scaling but depend on both the scaling parameter j and the width of the respective blob. For each point \mathbf{y} , we again estimate the width of an assumed ideal blob by fitting a parabola through the points $\left(a^{-j^-(\mathbf{y})} r_{\tilde{g}^e}, \left\langle f, \tilde{m}_{j^-(\mathbf{y}),\theta^*(\mathbf{y}),\mathbf{y}}^e \right\rangle_{L^2} \right)$, $\left(a^{-j^*(\mathbf{y})} r_{\tilde{g}^e}, \left\langle f, \tilde{m}_{j^*(\mathbf{y}),\theta^*(\mathbf{y}),\mathbf{y}}^e \right\rangle_{L^2} \right)$, and $\left(a^{-j^+(\mathbf{y})} r_{\tilde{g}^e}, \left\langle f, \tilde{m}_{j^+(\mathbf{y}),\theta^*(\mathbf{y}),\mathbf{y}}^e \right\rangle_{L^2} \right)$, where $j^-(\mathbf{y})$ denotes the predecessor and $j^+(\mathbf{y})$ the successor of $j^*(\mathbf{y})$ in the sequence \mathbf{J} and compute

$$W_B(f, \mathbf{y}) = 2 \left(a^{\frac{c_1}{2c_2} - j^*(\mathbf{y})} r_{\tilde{g}^e} \right), \quad (4.107)$$

where the values c_1 and c_2 are chosen according to (4.33) and (4.34). Under the assumption that an ideal square-shaped blob with width $W_B(f, \mathbf{y})$ and height 1 is centered on \mathbf{y} , we can also directly compute the associated even-symmetric coefficients of differently dilated molecules and denote

$$\tilde{K}_{\tilde{g}^e}(f, j, \mathbf{y}) = \tilde{K}_{\tilde{g}^e} (a^j W_B(f, \mathbf{y})/2) = K_{\tilde{g}^e} (a^j W_B(f, \mathbf{y})/2)^2. \quad (4.108)$$

The height of the blob can then be computed by

$$H_B(f, \mathbf{y}) = \frac{\left\langle f, \widetilde{m}_{j^*(\mathbf{y}), \theta^*(\mathbf{y}), \mathbf{y}}^e \right\rangle_{L^2}}{K_{\widetilde{g}^e}(f, j^*(\mathbf{y}), \mathbf{y})^2}. \quad (4.109)$$

We are now ready to define a measure for blob detection that can be understood as a generalization of the ridge measure $R(f, \mathbf{y})$ in the sense that we aim to detect features that are associated with more than one axis of even symmetry. In particular, we will take into account that axes of even symmetry induce rotation invariance. A perfect square, for instance, has four symmetry axes and is invariant to rotations by $\frac{\pi}{2}$, while a pentagon has five axes of even symmetry and is invariant to rotations by $\frac{2\pi}{5}$. Let for each point in the image plane $\Theta^*(\mathbf{y}) \subset \Theta$ denote the subset of rotation parameters that correspond to the expected rotation invariance properties of the blob centered on \mathbf{y} . The subset $\Theta^*(\mathbf{y})$ thus depends on the most significant rotation parameter $\theta^*(\mathbf{y})$ and on the symmetry and rotation invariance properties of the specific shape we aim to detect. Note that in the case of circles, $\Theta^*(\mathbf{y}) = \Theta$ independently of the most significant scale parameter $j^*(\mathbf{y})$ for all points \mathbf{y} . With a soft-thresholding parameter $\beta > 0$ and a scaling offset $j^o \in \mathbb{R}_{\geq 0}$, the blob measure is given by

$$\check{B}(f, \mathbf{y}) = \frac{\sum_{j \in \mathbf{J}} \min_{\theta \in \Theta^*(\mathbf{y})} \left\{ \text{sgn}(H_B(f, \mathbf{y})) \left\langle f, \widetilde{m}_{j, \theta, \mathbf{y}}^e \right\rangle_{L^2} \right\} - \sum_{j \in \mathbf{J}} \max_{\theta \in \Theta^*(\mathbf{y})} \left\{ \left| \left\langle f, m_{j-j^o, \theta, \mathbf{y}}^o \right\rangle_{L^2} \right| \right\} - \beta N_J K_{\widetilde{g}^e}(f, j^*(\mathbf{y}), \mathbf{y})^2}{\sum_{j \in \mathbf{J}} \max \left\{ \left| \left\langle f, \widetilde{m}_{j, \theta^*(\mathbf{y}), \mathbf{y}}^e \right\rangle_{L^2} \right|, |H_B(f, \mathbf{y})| K_{\widetilde{g}^e}(f, j, \mathbf{y})^2 \right\}}, \quad (4.110)$$

where $\varepsilon > 0$ prevents division by zero. Eventually, we define

$$B(f, \mathbf{y}) = \{0, \check{B}(f, \mathbf{y})\}. \quad (4.111)$$

4.3.5 Estimation of Local Tangent Directions for Edges and Ridges

The width estimates $W_R(f, \mathbf{y})$ and $W_B(f, \mathbf{y})$ exploit the fact that even-symmetric generators from the sets Ψ_2^e and $\widetilde{\Psi}_2^e$ are themselves associated with a width which in turn depends on the scale parameter j . By considering the orientation parameter θ , a similar approach can be used to estimate local tangent directions of anisotropic features such as edges or ridges. Let us first consider the case of edges and a system $SM(g^o, \alpha, a, \mathbf{J}, \Theta)$ of odd-symmetric molecules with $g^o \in \Psi_2^o$, $\alpha \in [0, 1]$, $a > 0$, an increasing sequence $\mathbf{J} = \{j_n\}_{n=1}^{N_J} \subset \mathbb{Z}$ of $N_J \in \mathbb{N}$ scaling parameters, and a clockwise (or counter-clockwise) ordered sequence $\Theta = \{\theta_n\}_{n=1}^{N_\Theta} \subset \mathbb{T}$ of $N_\Theta \in \mathbb{N}$ orientation parameters on the torus. Each odd-symmetric molecule $m_{j, \theta, \mathbf{y}}^o \in SM(g^o, \alpha, a, \mathbf{J}, \Theta)$ (cf. (4.72)) has a preferred orientation defined by the rotation parameter θ and the preferred orientation of the generator g^o . At the location of

an ideal edge, the odd-symmetric coefficient is maximized with respect to rotation if the direction of the one-dimensional wavelet ψ^o used in the construction of the generator g^o is orthogonal to the local tangent of the edge. For an image $f \in L^2(\mathbb{R}^2)$ and a point $\mathbf{y} \in \mathbb{R}^2$, let us again denote the most significant scaling and orientation parameters with

$$(j^*(\mathbf{y}), \theta^*(\mathbf{y})) = \underset{(j, \theta) \in \mathbf{J} \times \Theta}{\operatorname{argmax}} \left| \langle f, m_{j, \theta, \mathbf{y}}^o \rangle_{L^2} \right|. \quad (4.112)$$

Analogous to the width measures $W_R(f, \mathbf{y})$ and $W_B(f, \mathbf{y})$, we can use $\theta^*(\mathbf{y})$ as a first approximation of the tangent direction of a potential edge going through the point \mathbf{y} . Again, this approximation can be refined by computing the maximum point of a parabola fit through the points $\left(\theta^-(\mathbf{y}), \langle f, m_{j^*(\mathbf{y}), \theta^-(\mathbf{y}), \mathbf{y}}^e \rangle_{L^2} \right)$, $\left(\theta^*(\mathbf{y}), \langle f, m_{j^*(\mathbf{y}), \theta^*(\mathbf{y}), \mathbf{y}}^e \rangle_{L^2} \right)$, and $\left(\theta^+(\mathbf{y}), \langle f, m_{j^*(\mathbf{y}), \theta^+(\mathbf{y}), \mathbf{y}}^e \rangle_{L^2} \right)$, where $\theta^-(\mathbf{y})$ denotes the preceding and $\theta^+(\mathbf{y})$ the succeeding orientation parameter of $\theta^*(\mathbf{y})$ in the sequence Θ . The local tangent orientation can eventually be estimated by applying (4.35), that is,

$$O_E(f, \mathbf{y}) = \theta^*(\mathbf{y}) - \frac{c_1}{2c_2}, \quad (4.113)$$

where the values c_1 and c_2 are chosen as in (4.33) and (4.34). To obtain a function $O_R(f, \mathbf{y})$ that estimates local tangent orientation of ridges, it suffices to replace the odd-symmetric generator g^o with an even-symmetric generator $g^e \in \Psi_2^e$.

4.4 Implementation for Digital Images

A MATLAB toolbox named *Symmetric Molecule-based Feature Detector* (SymFD) that implements the measures defined in Section 4.3 for digital two-dimensional grayscale images can be downloaded from <http://www.math.uni-bremen.de/cda/software.html>. SymFD uses cyclic two-dimensional convolutions with digital symmetric molecule filters to obtain the required even- and odd-symmetric coefficients. The convolutions are carried out in the frequency domain via the fast Fourier transform (FFT). The computational complexity of the evaluation of any of the measures implemented in SymFD is thus of order $O(MN \log(N))$, where N denotes the number of pixels of the input image and M the number of considered symmetric molecule filters.

A main difficulty of developing digital implementations of the proposed feature detectors is to find a parametrization that makes it easy to configure SymFD for different types of applications and inputs while retaining the flexibility of the original definitions. In this section, we briefly summarize how systems of digital symmetric molecule filters can be defined and how the different parameters of the proposed edge, ridge, and blob measures

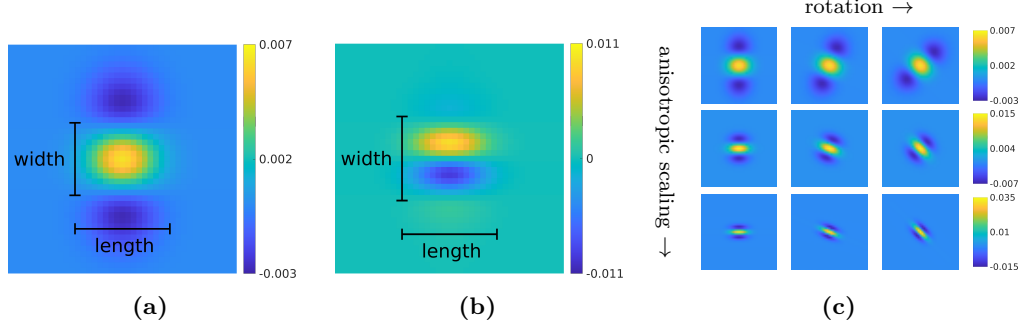


Figure 4.8: Digital symmetric molecules filters. (a): Example of an even-symmetric molecule filter. (b): Example of an odd-symmetric molecule filter. (c): Digital symmetric molecule filters that were obtained by rotating and anisotropically dilating the even-symmetric generator depicted in (a) ($\alpha = 0.2$).

are represented in SymFD.

In the continuum, a symmetric molecule system $SM(g, \alpha, a, J, \Theta) \subset L^2(\mathbb{R}^2)$ is defined by an even- or odd-symmetric generator $g \in L^2(\mathbb{R}^2)$, an anisotropy parameter $\alpha \in [0, 1]$, a constant scaling factor $a > 0$, a set of scaling parameters $J \subset \mathbb{Z}$, and a set of rotation parameters $\Theta \subset \mathbb{T}$ (see (4.72)). In SymFD, digital symmetric molecule filters are constructed by sampling the frequency domain representations $\widehat{m_{j,\theta,y}}(\xi)$ of symmetric molecules on a uniformly spaced grid of the size of the given input image. Note that all even- and odd-symmetric generators in the set $\Psi_2^e \cup \widetilde{\Psi}_2^e \cup \Psi_2^o$ are based on tensor products of derivatives of the one-dimensional Gaussian and their Hilbert transforms (see (4.69), (4.70), and (4.71)), for which we can use the explicit formulations in the Fourier domain (4.22) and (4.24). Each function in the set $\Psi_2^e \cup \widetilde{\Psi}_2^e \cup \Psi_2^o$ of two-dimensional symmetric generators is associated with two scaling parameters $c_1, c_2 \in \mathbb{R}_+$ that determine the shape of their effective support (cf. (4.72)). In SymFD, the parameters `maxFeatureWidth` and `maxFeatureLength` can be used to directly specify the width and length of a generating symmetric molecule filter in terms of pixels. Here, the width is defined as the distance between the two zero crossings of the associated one-dimensional wavelet that are closest to the origin while the length is defined as the size of the interval centered on the origin which contains 95% of the energy of the dilated one-dimensional Gaussian. Figures 4.8(a) and 4.8(b) depict examples of an even-symmetric and an odd-symmetric digital filter, respectively, and their corresponding lengths and widths. Note that in the case of even-symmetric molecules, this notion of width agrees with the definition of the radius given in (4.92) and (4.105).

The anisotropy parameter α is equivalent to the parameter `alpha` in SymFD. The constant scaling factor a is parameterized by the value `scalesPerOctave` which determines

the number of scales within each dyadic scaling step, that is,

$$a = 2^{1/\text{scalesPerOctave}}. \quad (4.114)$$

The set of scaling parameters \mathbf{J} can then be specified via the parameter `minFeatureWidth` which defines the width of the symmetric molecule filters on the scale associated with the highest frequencies. That is, $\mathbf{J} = \{0, \dots, j_{\max}\}$ with

$$j_{\max} = \max \{j \in \mathbb{N}: a^{-j} \text{maxFeatureWidth} \geq \text{minFeatureWidth}\}. \quad (4.115)$$

In SymFD, the set of rotation parameters Θ is defined by the integer `nOrientations` as a sequence of uniformly spaced samples from the interval $[-\frac{\pi}{2}, \frac{\pi}{2})$, that is,

$$\Theta = \left\{ \theta_n = \frac{n\pi}{\text{nOrientations}} - \frac{\pi}{2} \right\}_{n=0}^{\text{nOrientations}-1}. \quad (4.116)$$

Examples of differently dilated and rotated digital symmetric molecule filters are plotted in Figure 4.8(c). SymFD furthermore supports the use of both rotation matrices \mathbf{R}_θ (see (4.2)) and shear matrices

$$\mathbf{S}_h = \begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{S}_h^\top = \begin{bmatrix} 1 & 0 \\ h & 1 \end{bmatrix}, \quad h \in \mathbb{R}, \quad (4.117)$$

to change to preferred orientation of a symmetric generator. In practice, however, we recommend that one applies the rotation matrix \mathbf{R}_θ . As discussed in the prelude to Chapter 3, a main advantage of the shear operator is that it preserves the integer grid for shear parameters $s \in \mathbb{Z}$ which facilitates the construction of discrete shearlet transforms that provide a faithful transition between the continuous and the digital realm. However, such transforms are typically only defined for special combinations of generators, scaling parameters, and orientation parameters and thus substantially lacking in flexibility. For this reason, all digital filters used in SymFD are constructed by sampling their analogs from the continuum in the Fourier domain. This approach yields maximal freedom but also renders the most significant advantage of applying shears instead of rotations moot.

The parameters introduced so far fully define a system of symmetric molecule filters that is based on a generator g which matches the symmetry of the feature that is to be detected (i.e., even symmetry in the cases of ridges and blobs and odd symmetry in the case of edges). In SymFD, the generator of opposite symmetry is by default constructed by applying the Hilbert transform to the one-dimensional wavelet used in the definition of g . Furthermore, a parameter `evenOddScaleOffset` can be used to define j^e in (4.87) and

j° in (4.102) and (4.110).

The soft-thresholding parameter β used in all three proposed feature measures is equivalent to the parameter `minContrast` in SymFD. Furthermore, SymFD implements two steps of post-processing that can be applied if necessary. The parameter `thinningThreshold` defines a threshold that is used to obtain a binary feature map which is then further processed with morphological thinning using MATLAB's `bwmorph` function. The parameter `minComponentLength` can be used to remove all connected components in the thinned binary feature map that do not have the specified minimal size.

Note that there also exists a Python implementation developed by Jonas Wloka, which is based on a previous version of the SymFD toolbox called *Complex Shearlet-Based Edge and Ridge Measure* (CoShREM). Both toolboxes can also be downloaded from <http://www.math.uni-bremen.de/cda/software.html>.

4.5 Performance Evaluation on Synthetic Images

To evaluate the performance of the proposed measures, we first consider sets of synthetic digital grayscale images for which the locations and properties of features are unambiguously defined. For each type of feature, we created two test images of varying complexity that contain exemplary instances of the respective feature. To test the stability of the proposed measures in adverse conditions, we furthermore applied a combination of Gaussian and Poisson (shot) noise to each test image with three different degrees of severity (no noise, medium noise, severe noise). Visualizations of all six synthetic images, their distorted versions as well as the corresponding ground truths are compiled in Appendix D.

In all cases, both the synthetic image and the ground truth are based on a geometry that was defined semi-randomly on the unit square and then discretized on a grid of

Table 4.1: Overview of algorithms considered in the evaluation with synthetic images.

	Edges	Ridges	Blobs	Tangent dir.	Widths	Parameters	Avg. time
SymFD	✓	✓	✓	✓	✓	12	18.56 s
Shearlet [171]	✓	✗	✗	✓	✗	2	2.51 s
PhaseCong [99]	✓	✓	✗	✓	✗	11	1.53 s
Canny [28]	✓	✗	✗	✗	✗	3	0.39 s
Sobel [153]	✓	✗	✗	✗	✗	1	0.05 s
Steger [156]	✗	✓	✗	✗	✓	4	1.64 s
Frangi [63]	✗	✓	✗	✓	✗	8	0.43 s
Circular Hough [7, 46]	✗	✗	✓	✗	✓	4	0.36 s

768 × 768 pixels. Hence, the ground truth is not directly tied to the discrete digital image and therefore less likely to be biased towards any of the different algorithms we consider in our numerical experiments. For all three features, the feature location ground truth is stored as a binary image where white pixels denote the contours of edges, centerlines of ridges, and centers of blobs, respectively. In the case of edges and ridges, the ground truth also contains the orientation of the tangent for each pixel that lies on the contour of an edge or the centerline of a ridge. The same holds for local feature widths in the case of ridges and blobs. Due to the fact that the height measures $H_R(f, \mathbf{y})$ and $H_B(f, \mathbf{y})$ directly depend on the corresponding width measures (cf. (4.101) and (4.109)), they are not explicitly included in the numerical evaluation.

We consider different metrics to evaluate the results obtained by a feature detection algorithm with respect to the corresponding ground truth. For a given image, let N_{gt} denote the number of pixels that are marked as feature points in the ground truth and N_{det} the number of points that were detected as feature points by the algorithm under evaluation. We denote the set of feature points in the ground truth with $P_{\text{gt}} = \{\mathbf{p}_n\}_{n=1}^{N_{\text{gt}}} \subset \mathbb{N}^2$ and the set of detected feature points with $P_{\text{det}} = \{\mathbf{p}_n\}_{n=1}^{N_{\text{det}}} \subset \mathbb{N}^2$. Furthermore, let $D_{\text{det}} = \{\min_{\mathbf{p}_2 \in P_{\text{gt}}} \|\mathbf{p}_1 - \mathbf{p}_2\|_2 : \mathbf{p}_1 \in P_{\text{det}}\}$ denote the set of all Euclidean distances from pixel coordinates in P_{det} to the set P_{gt} , and $D_{\text{gt}} = \{\min_{\mathbf{p}_2 \in P_{\text{det}}} \|\mathbf{p}_1 - \mathbf{p}_2\|_2 : \mathbf{p}_1 \in P_{\text{gt}}\}$ denote the set of all distances from points in P_{gt} to the set P_{det} . A figure of merit that evaluates how successful an algorithm was at correctly detecting the location of features can then be defined by

$$\text{FOM} = \frac{1}{\max\{N_{\text{gt}}, N_{\text{det}}\}} \sum_{n=1}^{N_{\text{gt}}} \frac{1}{1 + \gamma d_n^2}, \quad (4.118)$$

where $d_n \in D = D_{\text{gt}} \cup D_{\text{det}}$ denotes the n -th largest element in the union of both sets of distances, and $\gamma > 0$. All FOM-values reported in this section were computed with $\gamma = \frac{1}{9}$. Note that for $D = D_{\text{det}}$ and accordingly adjusted summation bounds, (4.118) yields exactly the definition of Pratt's figure of merit (PFOM) [2], which is defined as

$$\text{PFOM} = \frac{1}{\max\{N_{\text{gt}}, N_{\text{det}}\}} \sum_{d \in D_{\text{det}}} \frac{1}{1 + \gamma d^2}, \quad (4.119)$$

where $\gamma > 0$ is typically chosen as $\gamma = \frac{1}{9}$. The main reason for using a slightly adjusted version here is that PFOM tends to be too forgiving of false positives in the case that $N_{\text{det}} < N_{\text{gt}}$.

For a fixed radius $r > 0$, we define the set of successfully detected points in the ground truth (true positives) as $P_{\text{tp}} = \{\mathbf{p}_1 \in P_{\text{gt}} : \min_{\mathbf{p}_2 \in P_{\text{det}}} \|\mathbf{p}_1 - \mathbf{p}_2\|_2 \leq r\}$ and let $N_{\text{tp}} \in \mathbb{N}$

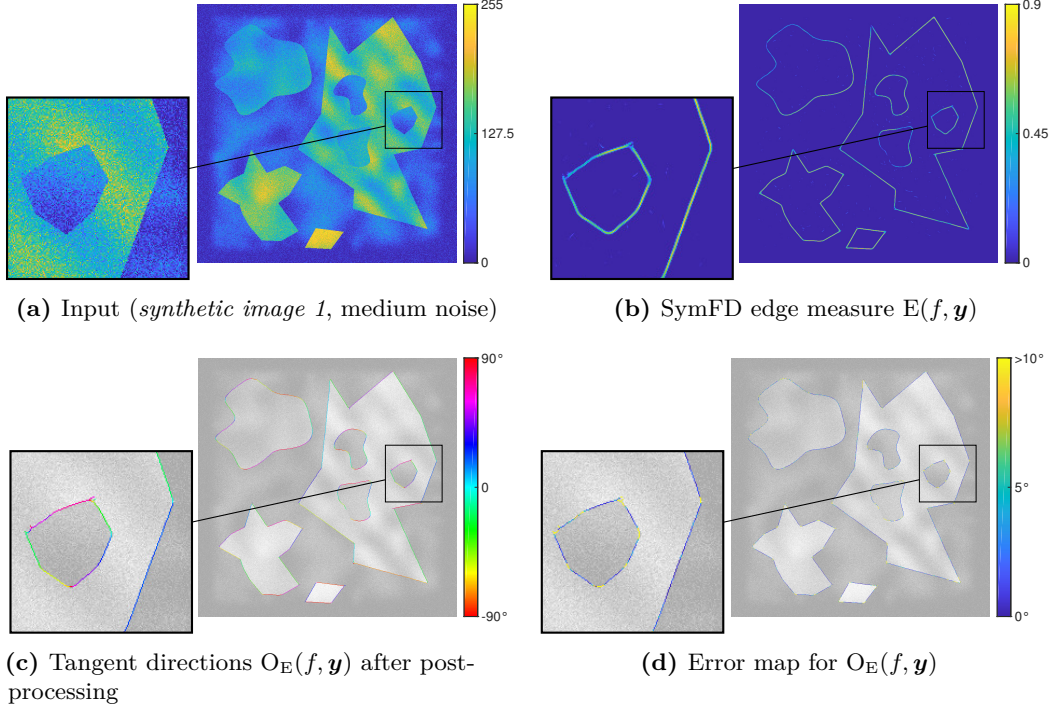


Figure 4.9: Detection of edges and local tangent orientations in a synthetic image using SymFD with parameters $\psi^o = \frac{\mathcal{H}G_2}{\|\mathcal{H}G_2\|_1}$, $\psi^e = \frac{G_2}{\|G_2\|_1}$, $\text{maxFeatureWidth} = 16$, $\text{maxFeatureLength} = 16$, $\alpha = \frac{1}{2}$, $\text{minFeatureWidth} = 4$, $\text{scalesPerOctave} = 2$, $\text{nOrientations} = 16$, $j^e = 1$, and $\beta = 15$. The input image is of size 768×768 and was distorted by a combination of Gaussian and Poisson noise.

denote the number of points in P_{tp} . To evaluate width and tangent orientation measures, we compute the mean of the set of absolute errors obtained by comparing the corresponding ground truth values at points in P_{tp} with the values yielded by an algorithm at the nearest neighbors of the ground truth points in the set P_{det} . Formally, we define the mean absolute error (MAE) of an arbitrary measure $M(f, \mathbf{y})$ computed on a digital image f as

$$\text{MAE}_M = \frac{1}{N_{\text{tp}}} \sum_{n=1}^{N_{\text{tp}}} \left| M_{\text{gt}}(f, \mathbf{p}_n) - M \left(f, \underset{\mathbf{p} \in P_{\text{det}}}{\text{argmin}} \|\mathbf{p}_n - \mathbf{p}\|_2 \right) \right|, \quad (4.120)$$

where $\mathbf{p}_n \in P_{\text{tp}}$ and $M_{\text{gt}}(f, \mathbf{p}_n)$ denotes the ground truth of the estimated property at the pixel \mathbf{p}_n . Note that in the case of tangent orientation measures, the error is computed on the torus. Finally, we define the number of true positives (TP) and the ratio of successfully detected feature points (TPR) as

$$\text{TP} = N_{\text{tp}}, \text{ and } \text{TPR} = \frac{N_{\text{tp}}}{N_{\text{gt}}}, \quad (4.121)$$

Table 4.2: Evaluation of five edge detection methods on synthetic images.

	No noise			Medium noise			Severe noise		
	FOM	MAE _{O_E}	TPR	FOM	MAE _{O_E}	TPR	FOM	MAE _{O_E}	TPR
<i>Synthetic image 1</i>									
SymFD	0.91	1.79 °	100 %	0.89	3.19 °	100 %	0.72	4.46 °	83 %
Shearlet [171]	0.85	3.78 °	100 %	0.82	5.78 °	98 %	0.64	8.71 °	84 %
PhaseCong [99]	0.91	6.13 °	100 %	0.85	8.48 °	95 %	0.58	15.59 °	72 %
Canny [28]	0.92	n/a	100 %	0.90	n/a	100 %	0.72	n/a	86 %
Sobel [153]	0.93	n/a	100 %	0.73	n/a	83 %	0.11	n/a	17 %
<i>Synthetic image 2</i>									
SymFD	0.90	4.55 °	100 %	0.85	7.43 °	96 %	0.64	13.28 °	76 %
Shearlet [171]	0.85	6.58 °	100 %	0.79	9.67 °	96 %	0.55	13.32 °	73 %
PhaseCong [99]	0.87	9.46 °	97 %	0.72	12.11 °	82 %	0.42	17.01 °	52 %
Canny [28]	0.91	n/a	100 %	0.87	n/a	99 %	0.65	n/a	79 %
Sobel [153]	0.93	n/a	100 %	0.65	n/a	76 %	0.16	n/a	22 %

respectively.

Throughout this section, the results of SymFD are not only compared to the ground truth but also to the results of other well-known feature detection algorithms. An overview of all considered algorithms is given in Table 4.1, which lists for each method the types of features it can detect, the number of parameters of the algorithm and the average time it took to process one synthetic digital image. The average processing time was measured on an Intel Core i7-4790 CPU clocked at 3.60 GHz.

The parameters of each method were separately optimized for each test image and each noise configuration. In the cases of edge and ridge detection, the optimization was carried out in terms of the FOM-value. For blob detection, the goal was to maximize the number of correctly detected blobs while keeping the number of false positives close to zero.

All MATLAB scripts and configurations that were used to obtain the numerical results presented in this section can be downloaded from <http://www.math.uni-bremen.de/cda/software.html>.

4.5.1 Edges

The synthetic images for the evaluation the edge measure $E(f, \mathbf{y})$ and the tangent orientation measure $O_E(f, \mathbf{y})$ are based on shapes that are defined by closed smooth spline curves and polygons. The main advantage of considering splines and polygons is that in both cases, with the exception of corner points in a polygon, the tangent can easily be computed for every point on the respective curve. The binary images that represent the feature location ground truth were obtained by marking each pixel which intersects with any of the defined curves. Each polygon and each spline curve is constructed from a finite sequence

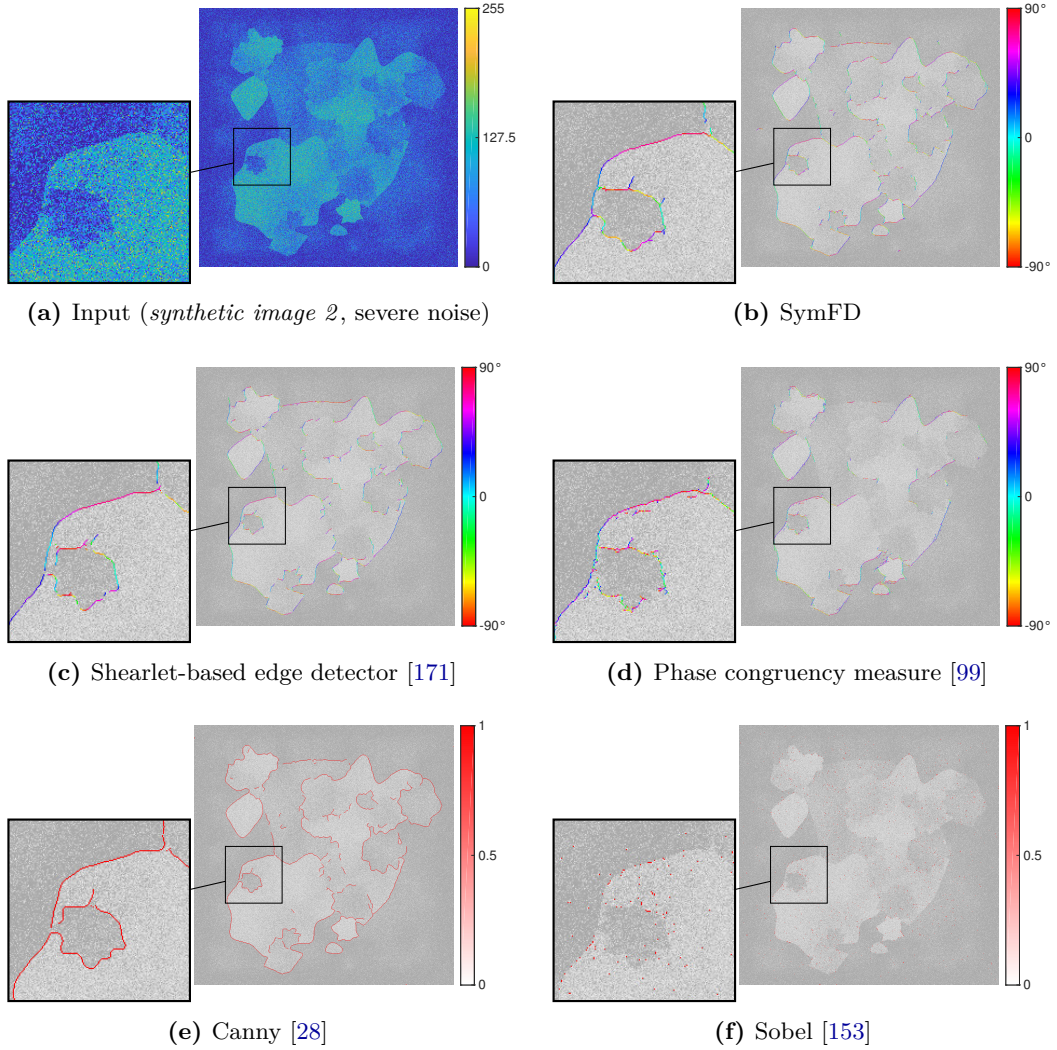


Figure 4.10: Visual comparison of the detection results yielded by different edge detection methods. Where applicable, the detected edge contour is shown in combination with estimates of the local tangent directions.

of points. These point sequences were predefined in the case of *synthetic image 1* (see Figure D.1(a)) and semi-randomly selected in *synthetic image 2* (see Figure D.1(d)). The shapes in both images were filled using a range of constant values and different types of gradients to ensure a certain amount of contrast variation. Furthermore, both images were overlaid with a smooth surface to test the capability of an algorithm to properly distinguish between sharp and smooth transitions. The synthetic images, their distorted versions, and the corresponding ground truths are compiled in Figure D.1 in Appendix D.

The edge measure $E(f, \mathbf{y})$ and the tangent orientation measure $O_E(f, \mathbf{y})$ after post-

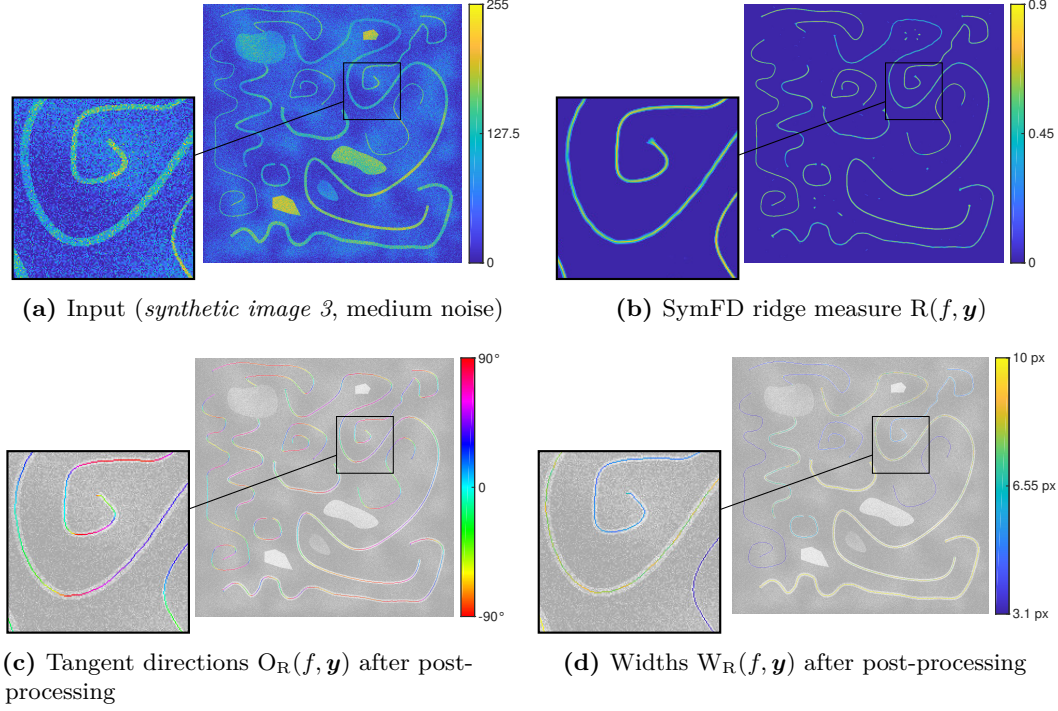


Figure 4.11: Detection of ridges, local tangent orientations, and ridge widths in a synthetic image. The results were yielded by SymFD with parameters $\psi^e = \frac{G_2}{\|G_2\|_1}$, $\psi^o = \frac{\mathcal{H}G_2}{\|\mathcal{H}G_2\|_1}$, $\text{maxFeatureWidth} = 10$, $\text{maxFeatureLength} = 15$, $\alpha = 0.2$, $\text{minFeatureWidth} = 3$, $\text{scalesPerOctave} = 6$, $\text{nOrientations} = 16$, $j^o = 1$, and $\beta = 20$. All ridges in the input have positive contrast. The analysis is therefore restricted to locations y where the height measure is positive, i.e., $H_R(f, y) \geq 0$. The input image is of size 768×768 and was distorted by a combination of Gaussian and Poisson noise.

processing are plotted in Figure 4.9 in the case of *synthetic image 1* distorted with medium noise. We also compare the performance of SymFD with results obtained from the Canny edge detector [28], the Sobel edge filter [153], the phase congruency measure [99], and a shearlet-based edge detector [171], which can be seen as a shearlet-based generalization of Canny. In the case of the Canny and Sobel edge detectors, we applied the implementation provided by MATLAB's `edge` function. The phase congruency measure was computed with the `phasecong3` function [101] and an implementation of the shearlet-based edge detector was kindly provided by the author of [171].

The detection performance of the considered algorithms for both synthetic images and three different levels of noise are summarized in Table 4.2. The edge detection performance was measured in terms of the FOM. Estimates of the local tangent directions are compared to the ground truth via the MAE_{O_E} . Note that the TPR is equal to the percentage of ground truth feature points that were used in the computation of the MAE_{O_E} (cf. (4.120) and (4.121)). The radius defining the set of true positives P_{tp} was set to three pixels. Note

Table 4.3: Evaluation of four ridge detection methods on synthetic images.

	No noise				Medium noise				Severe noise			
	FOM	MAE _{W_R}	MAE _{O_R}	TPR	FOM	MAE _{W_R}	MAE _{O_R}	TPR	FOM	MAE _{W_R}	MAE _{O_R}	TPR
<i>Synthetic image 3</i>												
SymFD	0.90	0.12 px	1.20 °	100 %	0.89	0.26 px	2.15 °	100 %	0.81	0.84 px	4.80 °	94 %
Steger [156]	0.90	0.71 px	n/a	100 %	0.85	1.31 px	n/a	100 %	0.81	2.56 px	n/a	98 %
Frangi [63]	0.80	n/a	1.24 °	99 %	0.80	n/a	2.18 °	99 %	0.20	n/a	3.26 °	23 %
PhaseCong [99]	0.89	n/a	15.67 °	99 %	0.81	n/a	31.39 °	96 %	0.48	n/a	36.07 °	52 %
<i>Synthetic image 4</i>												
SymFD	0.88	0.55 px	4.18 °	99 %	0.87	0.59 px	4.28 °	98 %	0.84	0.77 px	5.12 °	97 %
Steger [156]	0.88	1.12 px	n/a	98 %	0.87	1.80 px	n/a	99 %	0.83	1.85 px	n/a	95 %
Frangi [63]	0.85	n/a	4.30 °	98 %	0.85	n/a	4.55 °	97 %	0.82	n/a	6.94 °	97 %
PhaseCong [99]	0.81	n/a	25.71 °	97 %	0.82	n/a	29.18 °	96 %	0.64	n/a	33.55 °	74 %

that out of the algorithms included in the evaluation, only SymFD, the phase congruency measure, and the shearlet-based edge detector provide estimates of tangent orientations.

A visual comparison of the results yielded by the different methods when processing *synthetic image 2* with severe distortions can be found in Figure 4.10.

4.5.2 Ridges

Similar to the case of edge detection, we generated synthetic digital images that contain exemplary ridges whose centerlines are based on smooth spline curves. The sequences of points defining the spline curves were again predefined in *synthetic image 3* (see Figure D.2(a)) and chosen in a semi-random fashion in *synthetic image 4* (see Figure D.2(d)). To obtain ridges, the spline curves were thickened by a value ranging from three to ten pixels. In the case of *synthetic image 3*, the ridge-shapes were filled using different constant values and different types of gradients, while in *synthetic image 4*, all ridge-shapes are filled with the same value. To test the capability of an algorithm to correctly differentiate between edges and ridges, both synthetic images also contain filled shapes that are based on closed curves. Furthermore, both images were overlaid with a smooth surface. Visualizations of the noise-free and distorted versions of *synthetic image 3* and *synthetic image 4* as well as the corresponding ground truths can be found in Figure D.2 in Appendix D.

Detection results for a noisy version of *synthetic image 3* yielded by the ridge measure $R(f, \mathbf{y})$, the tangent orientation measure $O_R(f, \mathbf{y})$, and the width measure $W_R(f, \mathbf{y})$ are depicted in Figure 4.11. The ridge detection performance of SymFD is furthermore compared with results obtained from the phase congruency measure [99], the vessel enhancement filter proposed by Frangi et al. [63], and a widely used scale space-based ridge detection method proposed by Steger [156]. For the phase congruency measure, we

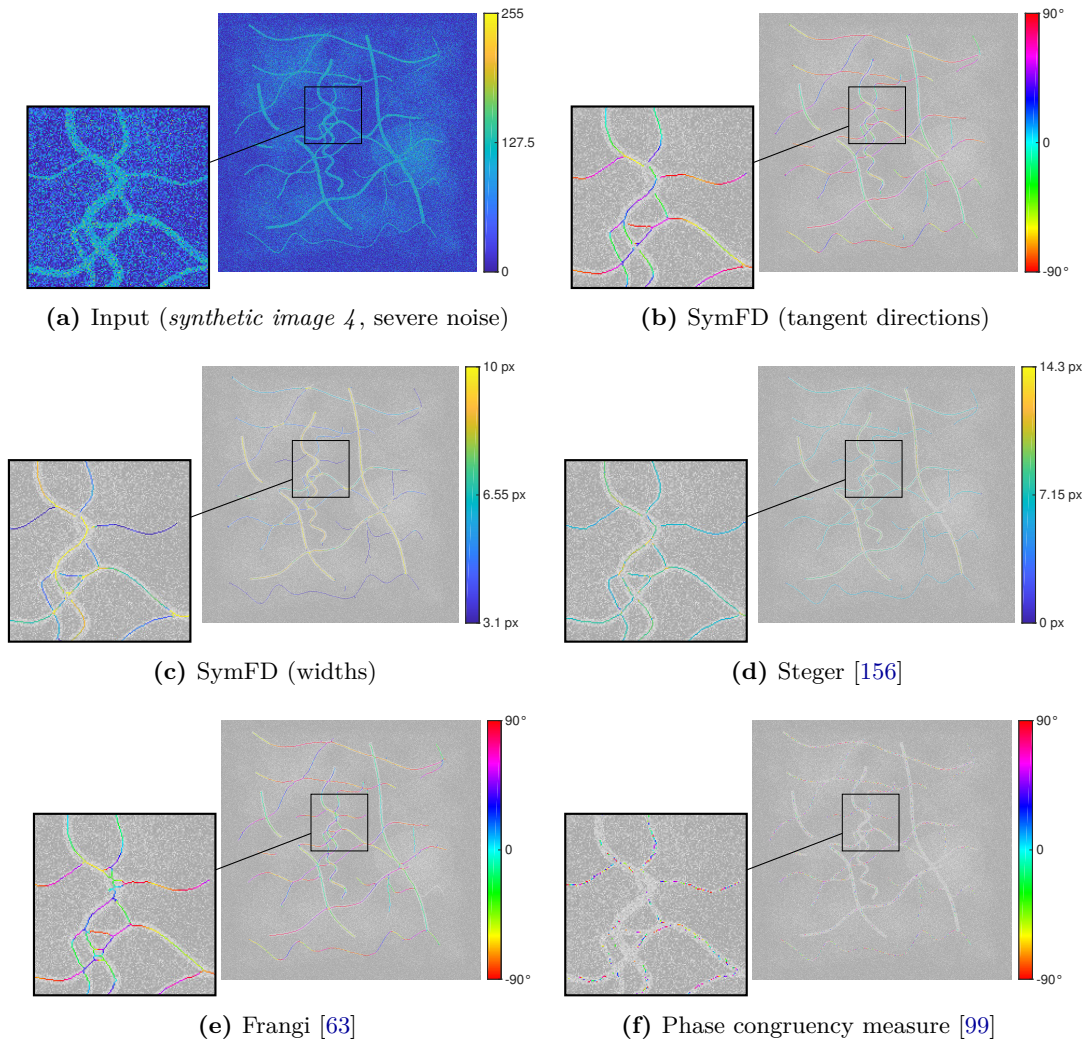


Figure 4.12: Visual comparison of the detection results yielded by different ridge detection methods. Where applicable, the detected centerline of a ridge is shown in combination with estimates of the local tangent direction and/or the width of the ridge.

applied the same implementation already used during the evaluation of different edge detection methods. An implementation of the Frangi vesselness measure can be found on the MATLAB Central File Exchange [102]. To obtain binary detection results, the feature maps computed by the Frangi vesselness measure were further processed with the same thresholding and thinning methods that are implemented in SymFD. For the Steger ridge detection method, we used an implementation provided by a freely available ImageJ plugin [164].

The ridge detection performances of SymFD, the phase congruency measure, the Frangi vessel enhancement filter, and the method proposed by Steger are summarized in Table 4.3.

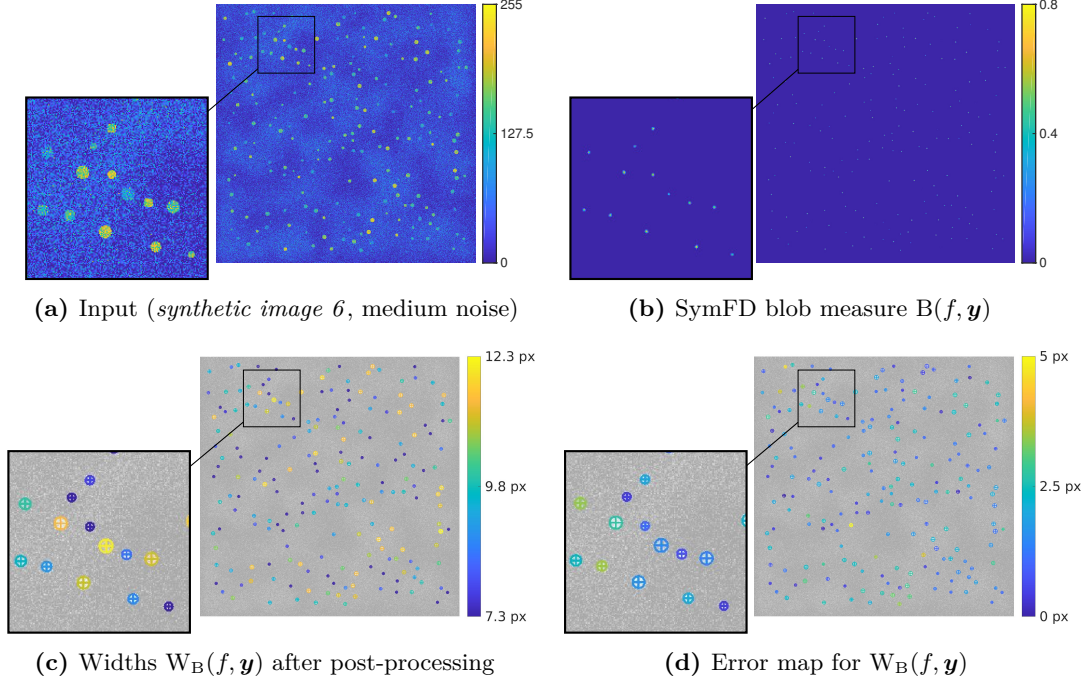


Figure 4.13: Detection of blobs and blob widths (diameters) in a synthetic image. The results were yielded by SymFD with parameters $\psi^e = \frac{\mathcal{H}G_1}{\|\mathcal{H}G_1\|_1}$, $\text{maxFeatureWidth} = 16$, $\text{maxFeatureLength} = 16$, $\alpha = 1$, $\text{minFeatureWidth} = 6$, $\text{scalesPerOctave} = 8$, $\text{nOrientations} = 16$, $j^\circ = 0$, and $\beta = 5$. All blobs in the input have positive contrast. The analysis is therefore restricted to locations \mathbf{y} where the height measure is positive, i.e., $H_B(f, \mathbf{y}) \geq 0$. The input image is of size 768×768 and was distorted by a combination of Gaussian and Poisson noise.

The similarity between the ridge location ground truth and the map of feature points detected by a single algorithm is again measured in terms of the FOM. Estimates of the local tangent directions and ridge widths are compared to the ground truth via the MAE_{OR} and the MAE_{WR} , respectively. The radius defining the set of true positives P_{tp} was set to three pixels. Similar to the evaluation of edge detection methods, errors in width and tangent direction measurements are only computed for feature points in the set P_{tp} . Out of the algorithms included in the evaluation, SymFD, the phase congruency measure and the implementation applied to obtain the Frangi vesselness measure also provide estimates of local tangent directions. Furthermore, only SymFD and the Steger ridge detector yield estimates of local ridge widths.

A visual comparison of the results yielded by the different methods when processing *synthetic image 4* with severe distortions can be found in Figure 4.12.

Table 4.4: Evaluation of two blob detection methods on synthetic images.

	No noise				Medium noise				Severe noise			
	TP	FP	MAE _{ctr}	MAE _{W_B}	TP	FP	MAE _{ctr}	MAE _{W_B}	TP	FP	MAE _{ctr}	MAE _{W_B}
<i>Synthetic image 5</i> (31 blobs)												
SymFD	31	0	2.16 px	2.62 px	31	0	2.16 px	2.95 px	31	0	2.09 px	3.13 px
Circular Hough [7, 46]	31	0	0.99 px	0.61 px	30	0	1.82 px	2.33 px	20	0	2.03 px	8.15 px
<i>Synthetic image 6</i> (200 blobs)												
SymFD	200	0	1.10 px	0.87 px	198	0	1.04 px	0.97 px	155	1	1.30 px	1.19 px
Circular Hough [7, 46]	200	0	0.88 px	0.56 px	194	3	0.99 px	1.04 px	47	1	1.10 px	1.66 px

4.5.3 Blobs

To evaluate the proposed blob measure $B(f, \mathbf{y})$ and width estimates $W_B(f, \mathbf{y})$, we consider two synthetic images that show aggregations of filled circles with varying diameters and contrasts. *Synthetic image 5* (see Figure D.3(a)) consists of 31 large circles with diameters ranging from 30 to 50 pixels while *synthetic image 6* (see Figure D.3(d)) contains a total of 200 small blobs whose diameters are between 7 and 13 pixels. In both cases, the center points of the circles were semi-randomly selected in the sense that two center points have at least a distance of 20 pixels in *synthetic image 6* and 100 pixels in *synthetic image 5*. Noisy versions of *synthetic image 5* and *synthetic image 6* and the corresponding ground truths are visualized in Figure D.1 in Appendix D.

SymFD blob detection results for a distorted version of *synthetic image 6* are shown in Figure 4.13. The circle detection performance of SymFD is furthermore compared to the circular Hough transform [7, 46], which can be computed in MATLAB via the function `imfindcircles`. To compare detection results of an algorithm with the ground truth, we consider the number TP of true positives with respect to a radius of six pixels, as defined in (4.121), and the number FP of false positives. Here, FP is equal to the number of detected blob center points which are not in the vicinity of an actual center of a blob. That is, FP is equal to the number of elements in the set $P_{fp} = \{p_1 \in P_{det} : \min_{p_2 \in P_{gt}} \|p_1 - p_2\| > 6\}$. We further consider the mean absolute errors in terms of center point location and blob width denoted by MAE_{ctr} and MAE_{W_B}, respectively. The results of the numerical evaluation of the blob detection performance of SymFD and the circular Hough transform are summarized in Table 4.4.

A visual comparison of the results yielded by SymFD and the circular Hough transform in the cases of *synthetic image 5* and *synthetic image 6* distorted by severe Gaussian and Poisson noise is presented in Figure 4.14.

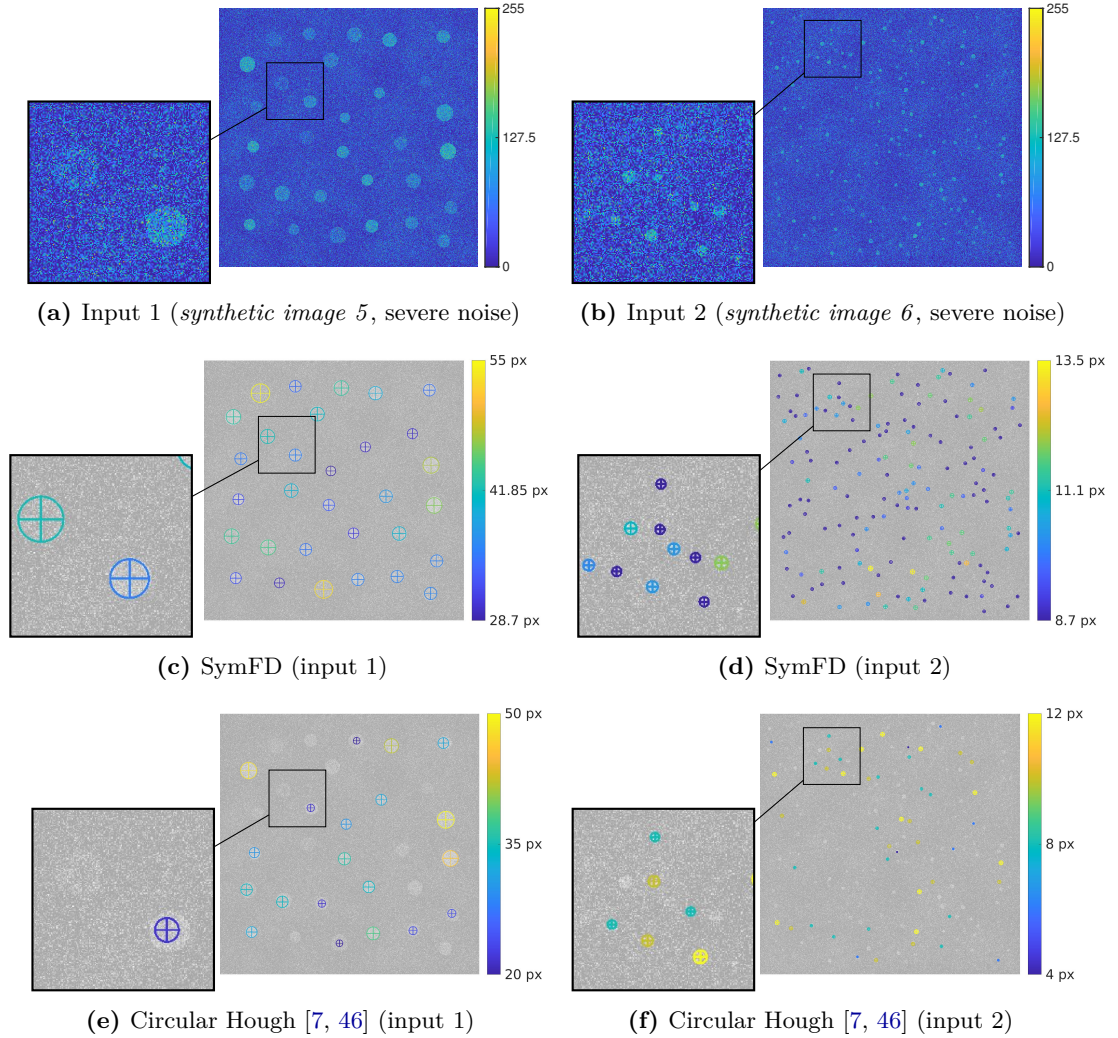
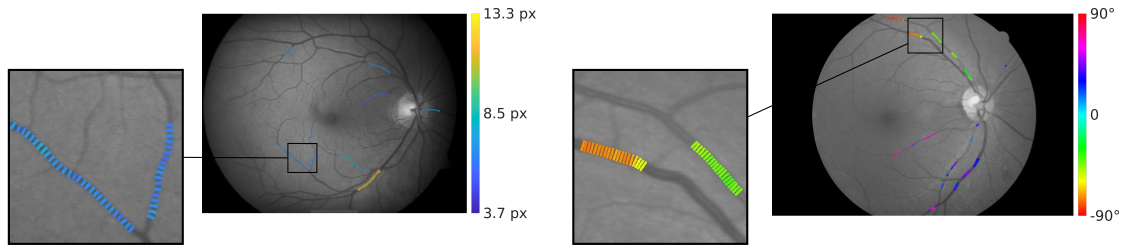


Figure 4.14: Visual comparison of the detection results yielded by the measures $B(f, \mathbf{y})$ and $W_B(f, \mathbf{y})$ implemented in SymFD and the circular Hough transform.

4.6 Applications

Previous versions of $E(f, \mathbf{y})$ and $R(f, \mathbf{y})$ that are based on complex-valued shearlets [96, 142, 144] were already successfully applied in different feature extraction tasks such as the detection and characterization of flame fronts [144] (see also Section 4.6.3), the detection of borders of tidal flats in the Wadden Sea [96], the extraction of fracture-traces in rockmasses [14], the detection of channel boundaries in seismic data [94], and the automated detection of the boundaries of touching cells in scanning electron (SEM) images [127].

To further evaluate the applicability of the proposed measures, we consider two tasks which require the correct localization and characterization of certain morphological com-



(a) Image 7 of the REVIEW VDIS dataset with manually annotated vessel profiles.

(b) Image 2 of the REVIEW CLRIS dataset with manually annotated vessel profiles.

Figure 4.15: Examples of digital images in the REVIEW retinal vessel reference dataset [48] with manually annotated ground truth vessel profiles.

ponents in biomedical images. In Section 4.6.1, we demonstrate how the width measure $W_R(f, \mathbf{y})$ and estimates of local tangent directions yielded by $O_R(f, \mathbf{y})$ can be used in combination with the ridge measure $R(f, \mathbf{y})$ to obtain characterizations of the geometry of retinal blood vessels. In particular, the blood vessel profiles yielded by SymFD are validated and compared to other state of the art algorithms on the publicly available REVIEW database [48]. In Section 4.6.2, we investigate the applicability of the blob measure $B(f, \mathbf{y})$ for automatically counting the number of grown cell colonies in a Petri dish. Finally, in Section 4.6.3, we evaluate the applicability of the proposed edge and ridge measures in the context of flame front characterization.

4.6.1 Determining Widths and Orientations of Retinal Blood Vessels

Retinal imaging is one of only a few existing techniques which provides non-invasive observations of the human circulatory system. Variations in the geometry of the retinal vasculature can furthermore be linked to a wide range of ocular diseases and systemic diseases with a vascular component such as diabetes [11, 32], Alzheimer's disease [168], or high blood pressure [84]. In recent years, many methods have been proposed for automatically extracting and analyzing the geometry of blood vessels in digital retinal images [4, 135].

To give an example of how SymFD can be applied in the context of retinal image analysis, we consider the REVIEW retinal vessel reference dataset [48]. For each image in the REVIEW database, three independent experts were asked to manually define vessel edge profiles on a number of preselected segments. Each profile consists of a pair of opposite edge points that indicate the boundaries of a blood vessel. A single profile thus not only contains information about the location of a retinal blood vessel but provides a complete characterization in terms of the local width and the local normal direction of the centerline. Two examples of images in the REVIEW database are shown along with manually marked

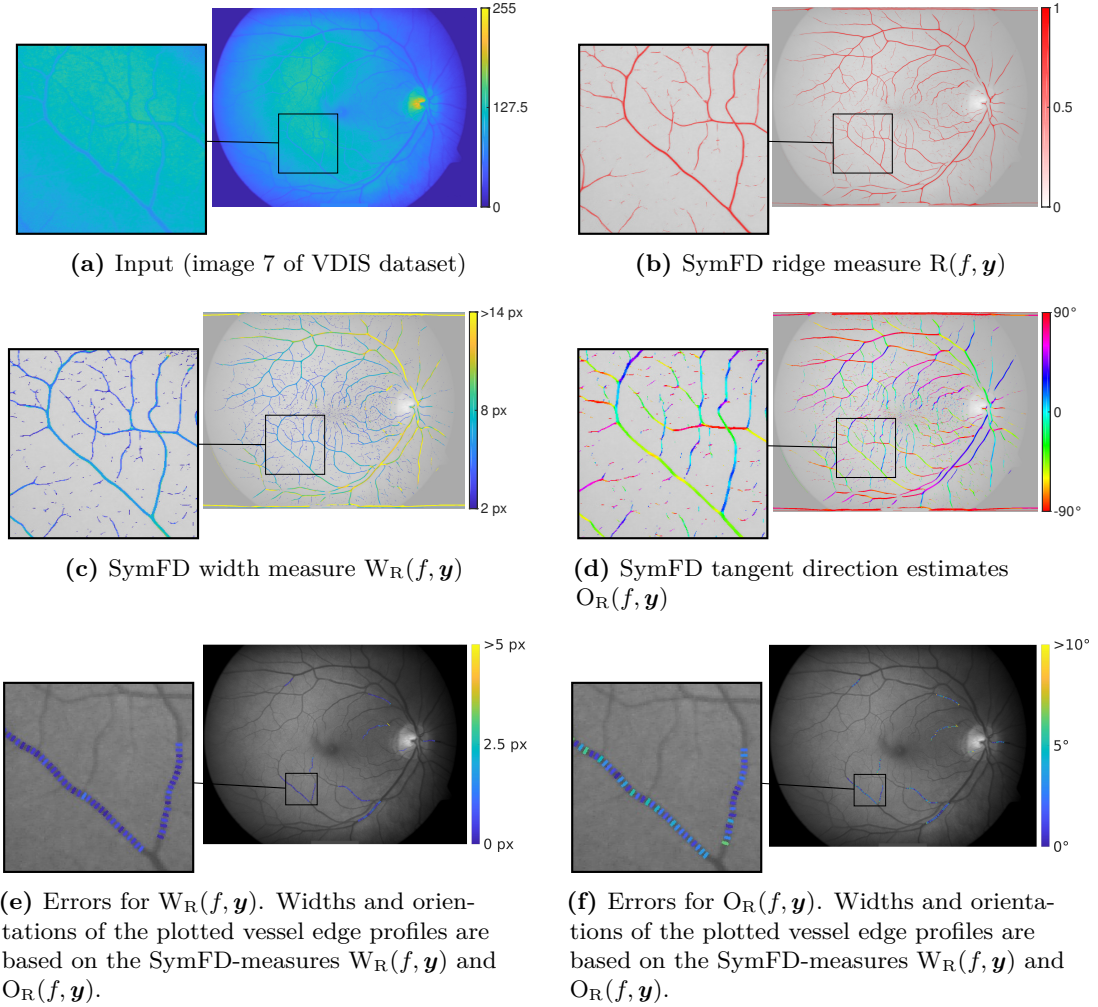


Figure 4.16: Width and orientation measurements of retinal blood vessels yielded by SymFD. The input image is part of the VDIS dataset from the REVIEW retinal vessel reference database.

vessel profiles in Figure 4.15.

The REVIEW database consists of four subsets of digital images that aim to provide meaningful benchmarks for different types of challenges in retinal image processing. The *high resolution image set* (HRIS) contains four images that show severe cases of diabetic retinopathy. Each image in the HRIS data set has a resolution of $896 \times 609 \times 3$ pixels but has previously been sub-sampled by a factor of four such that, disregarding human error, the vessel widths are known to a precision of a quarter of a pixel. The *vascular disease image set* (VDIS) contains eight retinal images of size $1360 \times 1024 \times 3$ that were randomly selected from a database of images of patients attending a retinopathy clinic. The *central light reflex image set* (CLRIS) consists of only two images of a resolution of $2160 \times 1440 \times 3$

Table 4.5: Evaluation of SymFD orientation measurements on the REVIEW database.

	VDIS		KPIS		HRIS		CLRIS	
	MAE _{O_R}	SR	MAE _{O_R}	SR	MAE _{O_R}	SR	MAE _{O_R}	SR
<i>Human Observers</i>								
O1	2.69 °	100 %	2.40 °	100 %	2.82 °	100 %	1.62 °	100 %
O2	2.97 °	100 %	2.52 °	100 %	2.41 °	100 %	1.61 °	100 %
O3	3.47 °	100 %	3.18 °	100 %	2.70 °	100 %	1.51 °	100 %
<i>Algorithm</i>								
SymFD	3.30 °	99 %	2.74 °	100 %	2.46 °	100 %	2.93 °	95 %

pixels that show early atherosclerotic changes in retinal vessels. Finally, the *kick point image set* (KPIS) contains two sub-sampled images where the normal direction of the centerline of a given location was computationally determined from the neighboring pixels on the centerline and the human observers where only asked to identify the kick-points determining the width. In total, the REVIEW database contains 5066 manually marked vessel profiles.

The diameters of blood vessels that are visible in the images of the database roughly range from 2 to 20 pixels. To obtain a precise detection and characterization of both thin and wide blood vessels, we use not a single pair but two pairs of systems of even- and odd-symmetric molecules that cover different but overlapping regions in the frequency domain. Both pairs of systems are based on the even-symmetric one-dimensional wavelet $\psi^e = \frac{\mathcal{H}G_1}{\|\mathcal{H}G_1\|_1}$ and the odd-symmetric wavelet $\psi^o = \frac{G_1}{\|G_1\|_1}$ and cover two octaves with `scalesPerOctave` = 4, $\alpha = 1$, and `nOrientations` = 16. The first pair of systems aims at the detection of thin vessel and is defined by the parameters `maxFeatureWidth`₁ = 8, `maxFeatureLength`₁ = 24, and `minFeatureWidth`₁ = 2. The second pair is capable of detecting vessels with a diameter of up to 24 pixels and is defined by the parameters `maxFeatureWidth`₂ = 24, `maxFeatureLength`₂ = 72, and `minFeatureWidth`₂ = 6. Each image in the database is processed by computing the ridge measure $R(f, \mathbf{y})$, the width measure $W_R(f, \mathbf{y})$ and the tangent direction measure $O_R(f, \mathbf{y})$ for both pairs of systems of even- and odd-symmetric molecules with parameters $j^o = 0$, and $\beta = 2$. In each case, the computations are carried out solely on the green channel of the processed RGB image, in which the blood vessels typically have the most significant contrast. For every pixel location y , we keep the values yielded by the pair of systems associated with the larger ridge measure $R(f, \mathbf{y})$. We furthermore restrict the analysis to ridges with negative contrast, that is, to locations where $H_R(f, \mathbf{y}) \leq 0$. As the width measure $W_R(f, \mathbf{y})$ is not only based on the even-symmetric coefficient on the most significant scale but also on the coefficients with respect to the preceding and succeeding scaling parameters, measurements where the

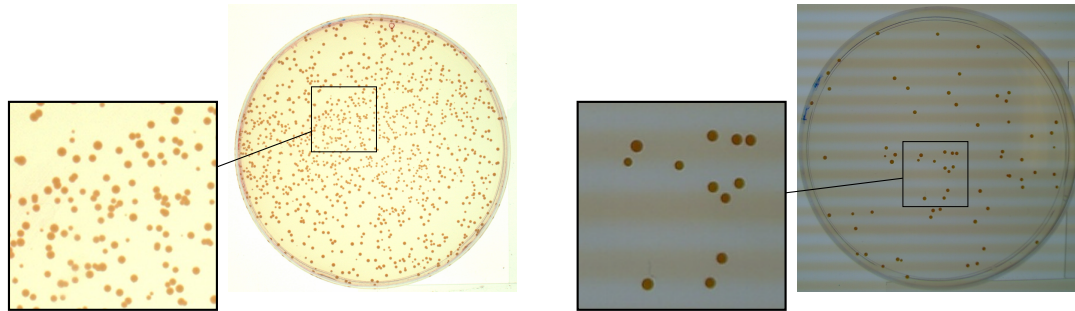
Table 4.6: Evaluation of width measurements obtained by nine different algorithms on the REVIEW Database

	VDIS		KPIS		HRIS		CLRIS	
	SD_{W_R}	SR	SD_{W_R}	SR	SD_{W_R}	SR	SD_{W_R}	SR
<i>Human Observers</i>								
O1	0.54 px	100 %	0.23 px	100 %	0.29 px	100 %	0.57 px	100 %
O2	0.62 px	100 %	0.21 px	100 %	0.26 px	100 %	0.70 px	100 %
O3	0.67 px	100 %	0.23 px	100 %	0.28 px	100 %	0.57 px	100 %
<i>Algorithms</i>								
SymFD	1.20 px	99 %	0.41 px	100 %	0.60 px	100 %	1.44 px	95 %
ARIA [8]	0.95 px	99 %	0.29 px	100 %	0.32 px	100 %	0.95 px	100 %
ETOS [10]	0.80 px	100 %	0.36 px	100 %	0.45 px	100 %	0.53 px	100 %
ESP [47]	0.77 px	100 %	0.33 px	100 %	0.42 px	100 %	1.47 px	93 %
Graph [169]	1.43 px	96 %	0.67 px	99 %	0.57 px	100 %	1.78 px	94 %
Gregson [68]	1.49 px	100 %	0.60 px	100 %	2.84 px	100 %	2.84 px	100 %
HHFW [18]	0.88 px	78 %	0.39 px	96 %	0.93 px	88 %	n/a	0 %
1DG [176]	2.11 px	100 %	0.40 px	100 %	4.14 px	100 %	4.14 px	99 %
2DG [117]	1.33 px	77 %	0.34 px	100 %	6.02 px	99 %	6.02 px	27 %

first or last scales are most significant are not considered. The obtained feature maps in the case of image 7 from the VDIS dataset are plotted in Figures 4.16(a) and 4.16(d).

To validate the obtained measurements, the width and tangent direction maps yielded by SymFD are compared to the widths and orientations of the manually marked profiles. For each vessel edge profile in the ground truth, we consider the difference with respect to the width measure $W_R(f, \mathbf{y})$ and the orientation measure $O_R(f, \mathbf{y})$ at the nearest location y in a two-pixel radius for which the ridge measure $R(f, \mathbf{y})$ is greater than zero. If no such pixel exists, the corresponding vessel edge profile counts as undetected. Visualizations of the tangent direction error and the width error for measurements obtained from SymFD are visualized in the case of image 7 from the VDIS dataset in Figures 4.16(e) and 4.16(f). For the tangent orientation estimates $O_R(f, \mathbf{y})$, we compute the mean absolute error MAE_{O_R} on the torus with respect to all manually marked profiles for each of the four subsets of images. The respective values along with the *success rate* SR, which denotes the percentage of successfully matched vessel profiles, are reported in Table 4.5.

For the validation of width measurements $W_R(f, \mathbf{y})$, we follow the procedure proposed in [8, 47, 169] and consider the standard deviation SD_{W_R} of differences between the measure $W_R(f, \mathbf{y})$ and the ground truth vessel diameters instead of the mean absolute error. This way, the comparison is independent of possibly different implicit definitions of vessel widths. We furthermore compare SymFD to a number of classical and state of the art methods for the extraction of vessel diameters from digital retinal images. In [176], vessel widths are obtained by fitting a *one-dimensional Gaussian* (1DG) model to the



(a) The shown image was acquired with a high-definition camera with a resolution of 1538×1536 pixels. Median human count of cell colonies: 1090.

(b) The shown image was obtained from a low-cost webcam with horizontal stripe artifacts and acquired at a resolution of 1000×1000 pixels. Median human count of cell colonies: 69.

Figure 4.17: Two pictures showing grown cell colonies in a Petri dish. The dataset containing the displayed images was originally developed for the evaluation of the OpenCFU software package [65].

vessel intensity profiles. A similar approach using a *two-dimensional Gaussian* (2DG) model was proposed in [117]. The *half height full width* (HHFW) model defines the width of a vessel as the distance between the points at which the vessel intensity profile reaches half of the maximum intensity left and right of an assumed center point [18]. Gregson et al. proposed to obtain vessel widths by fitting a rectangle to the intensity profile such that the area under the rectangle is equal to the area under the intensity profile [68]. A graph-based edge segmentation technique for measuring vessel widths was proposed in [169], while the *extraction of segment profiles* (ESP) algorithm proposed in [47] is based on an active contour model. The *automated retinal image analyzer* (ARIA) employs a wavelet-based approach for vessel segmentation which is followed by a refinement of edge locations [8]. Finally, *edge tracking based on orientation scores* (ETOS) is an algorithm that yields state of the art vessel width measurements by simultaneously tracking both edges of a vessel in the coupled space of positions and orientations [10]. The accuracy of vessel width measurements on images from the REVIEW database obtained by SymFD and all aforementioned methods in terms of the standard deviation SD_{W_R} and success rate SR are compiled in Table 4.6.

4.6.2 Automated Cell Colony Counting

Counting the number of cell colonies growing on an agar disk is a common but time-consuming task in many areas of biomedical research. To evaluate the applicability of the blob measure $B(f, \mathbf{y})$ for automatically obtaining the number of colonies from digital images, we consider a set of pictures which was originally developed for the evaluation of

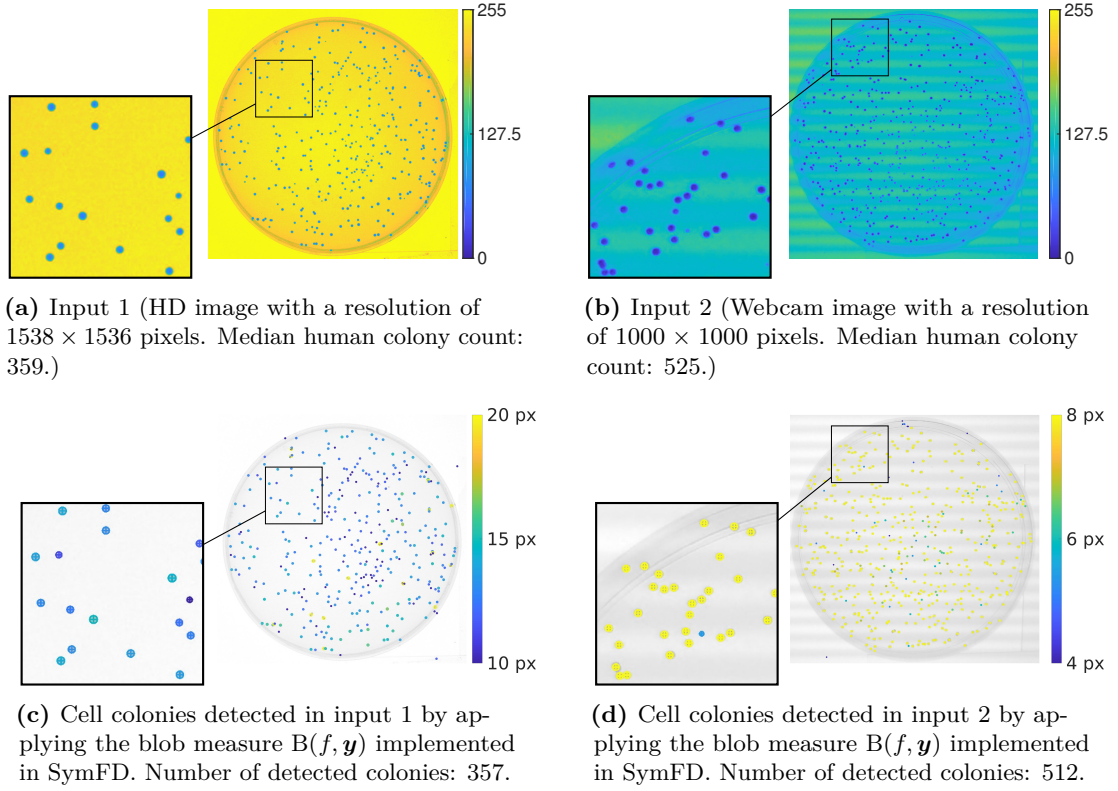


Figure 4.18: SymFD-based detection of cell colonies in digital photographs of Petri dishes.

the OpenCFU software package [65]. The set consists of digital images of 19 different plates containing between 10 and 1000 *Staphylococcus aureus* colonies and can be downloaded from <http://opencfu.sourceforge.net/>. Each plate was photographed once with a high-definition camera and once with a low-cost webcam. To provide a reference for the evaluation of automated counting methods, the number of colonies on each plate was independently counted by seven trained humans. For a single plate, the ground truth was then defined as the median of the numbers reported by the human experts. Examples of the considered images are displayed in Figure 4.17.

To automatically count the number of cell colonies in the high-definition images, all 19 images were processed with SymFD by computing the blob measure $B(f, \mathbf{y})$ on the blue color channels using the parameters $\psi^e = \frac{\mathcal{H}G_1}{\|\mathcal{H}G_1\|_1}$, $\text{maxFeatureWidth}_{\text{hd}} = 20$, $\text{maxFeatureLength}_{\text{hd}} = 20$, $\text{minFeatureWidth}_{\text{hd}} = 10$, $\alpha = 1$, $\text{scalesPerOctave} = 3$, $\text{nOrientations} = 16$, $j^o = -1$, and $\beta_{\text{hd}} = 15$. The obtained feature maps were then transformed into binary images by applying the threshold 0.03. Centers of cell colonies were eventually obtained by computing the centroids of all connected components in the binary image via MATLAB's `regionprops` function. For processing the webcam images, the

Table 4.7: Evaluation of the accuracy of automatically obtained cell colony counts.

	HD		Webcam	
	With Mask	Without Mask	With Mask	Without Mask
SymFD	1.95 %	3.57 %	2.23 %	5.95 %
OpenCFU [65]	1.93 %	3.85 %	2.79 %	n/a
ImageJ [22]	6.64 %	80.93 %	11.11 %	n/a
NICE [35]	9.91 %	20.56 %	13.04 %	n/a

The accuracy is reported in terms of the median relative deviation from counts provided by human experts.

parameters specifying the effective support of the applied symmetric molecules had to be adapted to accommodate for the lower resolution and were set to $\text{maxFeatureWidth}_{\text{wc}} = 8$, $\text{maxFeatureLength}_{\text{wc}} = 8$, and $\text{minFeatureWidth}_{\text{wc}} = 4$. Furthermore, mostly due to horizontal stripe artifacts (cf. Figure 4.17(b)), the contrasts of the cell colonies are slightly lower in the images captured by a webcam than in their high-definition counterparts. Thus, for processing the webcam images, the parameter specifying the minimal contrast was chosen as $\beta_{\text{wc}} = 6$. Examples of cell colony detection results obtained by SymFD for one high-definition image and one webcam image are visualized in Figure 4.18.

To validate the accuracy of the cell colony counts obtained by SymFD, we follow the procedure proposed in [65] and consider the median relative deviation from the counts provided by the human experts across all 19 images. We furthermore compare the accuracy of SymFD with three other methods for the automated detection and counting of cell colonies. The *NIST's Integrated Colony Enumerator* (NICE) [35] is a freely available software package based on a combination of extended-minima transforms and thresholding operations. We also consider an ImageJ plugin developed by Cai et al. [22] that is based on adaptive thresholding, the application of a watershed algorithm, and particle filtering. The results obtained by SymFD are compared to OpenCFU which is a highly versatile and efficiently implemented open source library for the detection of cell colonies [65]. The median relative deviations from human counts for SymFD and the aforementioned methods with respect to high-definition and webcam images of all 19 plates are compiled in Table 4.7. The cell colony counts for OpenCFU, NICE, and the ImageJ plugin proposed by Cai et al. were kindly provided by the author of [65]. In particular NICE, and the ImageJ macro are sensitive to artifacts in the background and at the border of the Petri dish. To facilitate a fair comparison, Table 4.7 also reports the median relative errors when using an additional foreground mask. To illustrate the robustness of the considered methods with respect to images of agar plates with both small and large numbers of cell colonies, semilog graphs

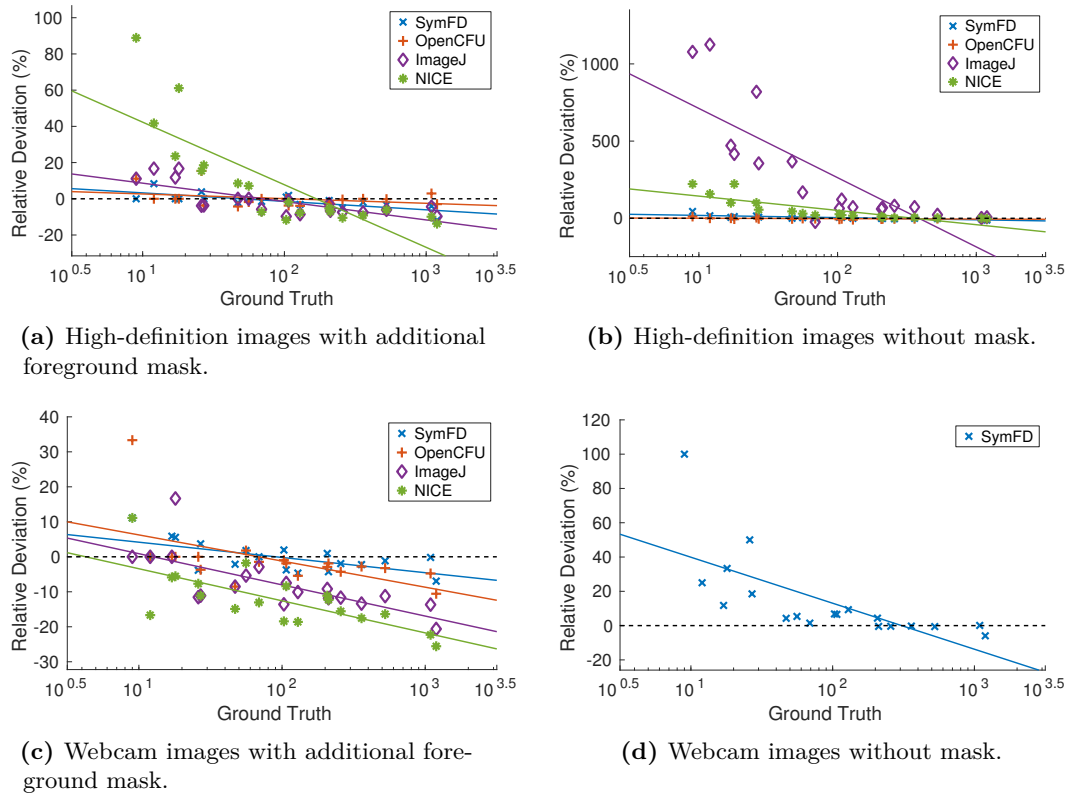


Figure 4.19: Semilog graphs and trendlines for four automated cell counting methods, plotting the relative deviation for a single plate against the number of cell colonies in the corresponding plate.

plotting for each algorithm the relative deviation for a single plate against the number of cell colonies in the corresponding plate are shown Figure 4.19.

4.6.3 Detection and Characterization of Flame Fronts

The development of laser combustion diagnostics employing planar imaging techniques in the 1970s and 80s has transformed combustion research [6, 39, 86, 162]. Using a light sheet to illuminate an entire two-dimensional cross-section of a flame and imaging the laser-induced emission onto a camera provides spatially correlated information in contrast to pointwise scanning. In particular, the use of short-pulse laser sources and gated cameras enables imaging on time scales that are shorter than flow and diffusion phenomena, and hence a true snapshot of a flame can be taken. Consequently, studying transient phenomena is possible by capturing flame structures under turbulent conditions. However, processing and evaluating such images is a challenge. Appropriate methods must be reproducible, accurate, and quantitative. In addition, the information desired should be available within

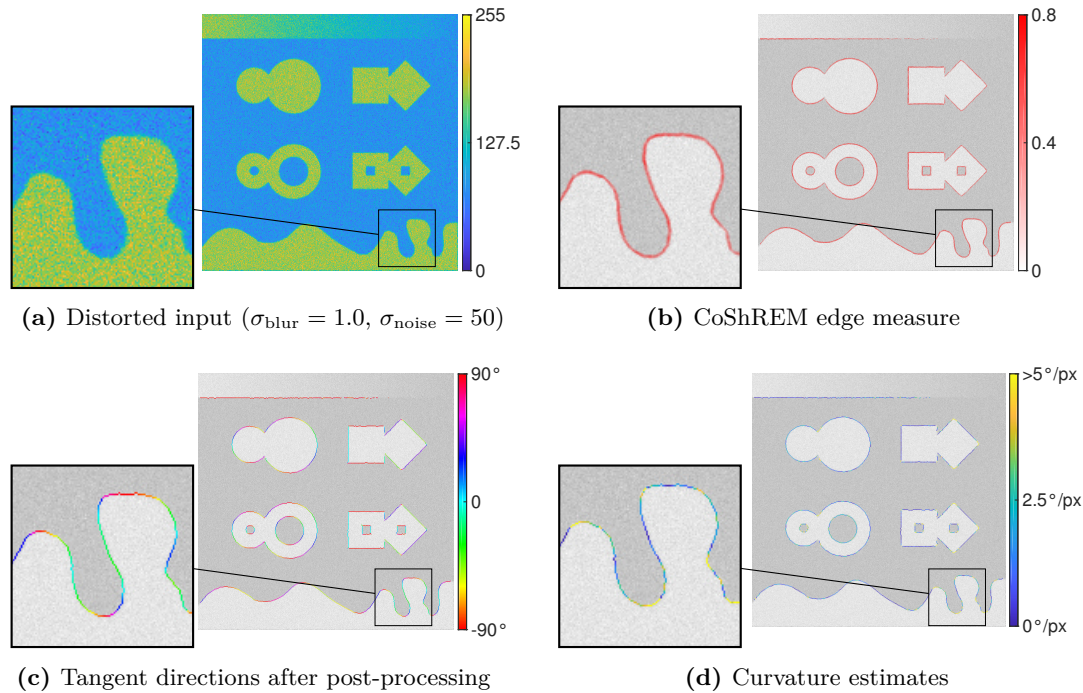


Figure 4.20: Detection and characterization of edges in a noisy mock image with CoShREM. The input is perturbed with Gaussian blur ($\sigma_{\text{blur}} = 1.0$) and additive Gaussian white noise ($\sigma_{\text{noise}} = 50$).

a reasonable period of time. This is particularly important when large data sets need to be processed.

The most common task is to identify and to characterize the flame front in an image recorded by planar laser-induced fluorescence (PLIF) [158], laser Rayleigh scattering (LRS) [136] or particle imaging velocimetry (PIV) [136]. Regarding data processing, this task comes down to an edge detection problem. Needless to say, the edge detection step is crucial, as a slightly differently detected edge may suggest significantly different flame parameters, e.g. in terms of the flame front curvature.

The majority of existing approaches for detecting the flame front in an image are based either on direct binarization [87, 119] or on local intensity gradients [21, 57]. When direct binarization is applied, an intensity threshold filter is used delivering a binary image containing areas of zeroes and ones, representing unburnt and burnt regions. The boundary between the two is the flame front, from which further information can be derived. In gradient-based methods, the first step is to convert the initial image by computing an approximation of the gradient pixel-by-pixel. Then a local or global threshold is applied in order to discriminate between the steep gradients typical for the flame front and less

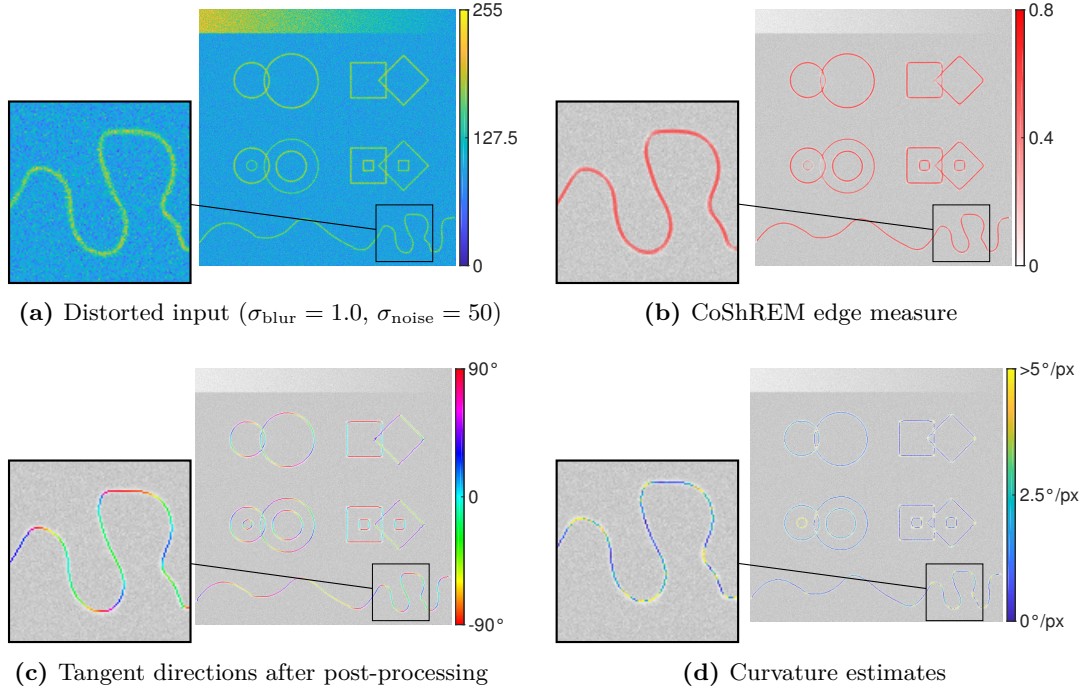


Figure 4.21: Detection and analysis of ridges in a noisy mock image with CoShREM. The input is perturbed with Gaussian blur ($\sigma_{\text{blur}} = 1.0$) and additive Gaussian white noise ($\sigma_{\text{noise}} = 50$).

pronounced structures. This may be a simple threshold or a sophisticated combination of multiple thresholds via hysteresis [28]. Subsequently, the remaining flame front data points can be fitted with a mathematical function, from which parameters such as curvature and flame surface density can be derived eventually. Pre-processing the original images with filters for noise reduction and contour enhancement may be required in order to improve the clearness and robustness of the flame front detection [21, 122, 158]. However, when the signal-to-noise ratio is low or the edge to be detected is not sufficiently steep, the common data processing algorithms may reach their limits. Further, the typical results must be further processed (e.g., by fitting a cubic spline to the detected edge) in order to obtain geometric information like curvature [136]. Finally, traditional edge-detection algorithms are not capable of detecting ridges as coherent structures, which is problematic when analyzing images of short-lived radicals like CH and HCO. Instead, completely different ridge-detection methods have to be applied, which often are based on approximating local optima [114, 154] or matching ridges to set shapes like circles [141].

In this section, we investigate the potential of the newly proposed edge and ridge measures for evaluating data from planar laser diagnostics. Note that the numerical

Table 4.8: Numerical comparison of CoShREM with five other edge detectors. The table shows PFOM values for all considered algorithms and a total of 20 differently distorted versions of the mock image shown in Figure 4.20, where 1.0 would indicate a perfect reproduction of the ground-truth. For each algorithm, parameters remained fixed for all test images but were carefully optimized such that the maximal error was minimized across all levels of noise. The binary ground-truth was drawn from hand and consisted of minimally connected lines (i.e. with the exception of intersections, each pixel with value 1 has at most two neighbors with value 1). To ensure a fair comparison, a thinning operation was applied to the binary outcome of each method. For a visual comparison of the results in the noisiest case ($\sigma_{\text{blur}} = 1.5$, $\sigma_{\text{noise}} = 100$), see Figure E.1.

	$\sigma_{\text{blur}} = 0.0$					$\sigma_{\text{blur}} = 0.5$				
	$\sigma_{\text{noise}} = 0$	20	50	80	100	$\sigma_{\text{noise}} = 0$	20	50	80	100
CoShREM	0.97	0.96	0.93	0.92	0.91	0.97	0.96	0.93	0.92	0.91
Canny [28]	0.90	0.90	0.88	0.88	0.88	0.90	0.90	0.88	0.89	0.88
Sobel [153]	0.93	0.92	0.89	0.70	0.44	0.93	0.91	0.89	0.65	0.43
PhaseCong [99]	0.95	0.94	0.90	0.78	0.72	0.95	0.94	0.88	0.75	0.62
Shearlet [171]	0.88	0.88	0.88	0.87	0.88	0.88	0.89	0.88	0.88	0.87
Canny (std. param.)	0.92	0.11	0.10	0.11	0.11	0.92	0.11	0.10	0.11	0.11

	$\sigma_{\text{blur}} = 1.0$					$\sigma_{\text{blur}} = 1.5$				
	$\sigma_{\text{noise}} = 0$	20	50	80	100	$\sigma_{\text{noise}} = 0$	20	50	80	100
CoShREM	0.96	0.95	0.94	0.92	0.91	0.95	0.94	0.93	0.90	0.89
Canny [28]	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.88	0.87	0.87
Sobel [153]	0.91	0.90	0.76	0.44	0.33	0.89	0.93	0.47	0.32	0.27
PhaseCong [99]	0.94	0.92	0.83	0.60	0.39	0.89	0.80	0.72	0.33	0.16
Shearlet [171]	0.88	0.88	0.88	0.87	0.85	0.87	0.88	0.87	0.86	0.86
Canny (std. param.)	0.92	0.10	0.10	0.10	0.11	0.91	0.10	0.10	0.10	0.11

evaluation presented in this section was performed prior to the development of the SymFD toolbox using the complex shearlet-based edge and ridge measure (CoShREM). CoShREM is a MATLAB toolbox¹ that implements essentially the same edge, ridge, and tangent direction measures that were defined in Section 4.3, but only considers complex-valued shearlet transforms instead of the more general α -molecule framework.

4.6.3.1 Synthetic Inputs

We first consider synthetic images with associated ground truths that were specifically designed by Johannes Kiefer to represent the special characteristics of experimentally obtained flame data. To allow for a systematic assessment of the stability of the evaluated methods, the images were corrupted by varying degrees of Gaussian blur, Gaussian noise, and Poisson noise. One set of images exhibits thin ridges, which are commonly observed

¹ CoShREM can also be downloaded from <http://www.math.uni-bremen.de/cda/software.html>.

Table 4.9: Numerical comparison of the stability under additional Poisson noise. The table shows PFOM values for all six considered edge detectors and a total of five differently distorted versions of the mock image shown in Figure 4.20(a), where 1.0 would indicate a perfect reproduction of the ground-truth. The test images were first perturbed with five different levels of additive Gaussian noise. Then, each pixel was resampled from a Poisson distribution with an expectancy of one tenth of the original pixel value. Finally, the values of the thereby obtained grayscale image were rescaled by a factor of 10. To test the stability with respect to this kind of shot noise, the same parameters as in Table 4.8 were used for all algorithms.

	$\sigma_{\text{noise}} = 0$	20	50	80	100
CoShREM	0.95	0.94	0.91	0.89	0.83
Canny [28]	0.90	0.89	0.87	0.50	0.28
Sobel [153]	0.65	0.49	0.36	0.22	0.24
PhaseCong [99]	0.93	0.90	0.70	0.24	0.00
Shearlet [171]	0.89	0.88	0.86	0.78	0.65
Canny (std. param.)	0.18	0.11	0.11	0.11	0.12

when short-lived radicals such as CH and HCO are visualized using PLIF. The other data set shows broader areas (edge detection), which are characteristic of LRS and PIV data, as well as PLIF images of long-lived radicals like OH.

Figures 4.20 and 4.21 visualize the results of the CoShREM edge, respectively ridge, detector in the case of two synthetic images corrupted by Gaussian noise and Gaussian blur. In particular, Figures 4.20(d) and 4.21(d) depict estimated curvatures, which were obtained for each pixel lying on a detected edge by computing the central difference with respect to the tangent direction estimates shown in Figures 4.20(c) and 4.21(c).

We further compare the CoShREM implementation of the newly proposed measures to a number of established approaches, such as the Canny edge detector [28], the Sobel edge detector [153], and the phase congruency measure [99, 100], as well as the shearlet-based edge detector developed by Yi et al. in 2009 [171]. In the case of the Canny and the Sobel edge detectors, the implementations provided by the MATLAB Image Processing Toolbox (version 9.2) were applied. The software used to compute the phase congruency measure can be downloaded from Peter Kovess's homepage [101], while an implementation of the other shearlet-based edge detector was kindly provided by the authors of [171]. For the latter, pre-processing with a Gaussian smoothing filter was added such that the algorithm could also handle more severe levels of noise.

The comparison is carried out visually but also numerically by computing Pratt's figure of merit (PFOM) (cf. Equation 4.119) on the obtained results¹. The images analyzed in

¹ While the corresponding ground-truth is often created applying one of the edge detectors on the noiseless image, in this case, it was handmade by the authors to prevent favoring a particular method.

Table 4.10: Numerical comparison of CoShREM and the phase congruency-based ridge detector. The table shows again PFOM values, where 1.0 would indicate a perfect reproduction of the ground-truth. Both methods were applied to a total of 20 differently distorted versions of the mock image shown in Figure 4.21. For both algorithms, parameters remained fixed for all test images but were carefully optimized such that the maximal error was minimized across all levels of noise. The binary ground-truth was drawn from hand and consisted of minimally connected lines (i.e. with the exception of intersections, each pixel with value 1 has at most two neighbors with value 1). To ensure a fair comparison, a thinning operation was applied to the binary outcome of each method. For a visual comparison for three differently distorted images, see Figure E.3.

	$\sigma_{\text{blur}} = 0.0$					$\sigma_{\text{blur}} = 0.5$				
	$\sigma_{\text{noise}} = 0$	20	50	80	100	$\sigma_{\text{noise}} = 0$	20	50	80	100
CoShREM	0.94	0.95	0.95	0.93	0.93	0.93	0.93	0.92	0.89	0.89
PhaseCong [99]	0.90	0.94	0.94	0.93	0.93	0.88	0.92	0.93	0.92	0.88
	$\sigma_{\text{blur}} = 1.0$					$\sigma_{\text{blur}} = 1.5$				
	$\sigma_{\text{noise}} = 0$	20	50	80	100	$\sigma_{\text{noise}} = 0$	20	50	80	100
CoShREM	0.94	0.92	0.92	0.90	0.86	0.93	0.92	0.91	0.88	0.88
PhaseCong [99]	0.88	0.92	0.92	0.88	0.81	0.88	0.91	0.89	0.74	0.37

Figures 4.20 and 4.21 were perturbed by four levels of Gaussian blur (i.e. convolution with a Gaussian filter kernel) and five levels of additive Gaussian white noise, yielding a total of 20 different levels of corruption, before applying CoShREM and other edge detection algorithms. To put an emphasis on testing the stability of the respective methods, the parameters configuring the various algorithms remained fixed throughout the evaluation. For each method, the associated set of parameters was chosen such that both visually and with respect to the PFOM metric, the maximal error was minimized across all levels of noise. To highlight the importance of carefully choosing fitting parameters for each edge detector, the Canny edge detector applied with its default parameters in MATLAB was also included in the comparison. The numerical results of this comparison are summarized in Table 4.8 in the case of edge detection, and in Table 4.10 for the considered ridge detectors.

Visual comparisons of the results yielded by the respective edge and ridge detection methods can be found in Figures E.1 and E.3 in Appendix E. Furthermore, a visual comparison of tangent direction and curvature estimates obtained from different algorithms can be found in Figures E.2 and E.4. Note that the computation of the local tangent orientations differs from method to method, however, all curvature estimates were obtained by computing the central difference with respect to the local tangent orientations yielded by the respective algorithm.

While there are a number of different sources of noise in combustion diagnostics [89, 91, 92, 137], the assumption of an overall Gaussian noise distribution is reasonable. However,

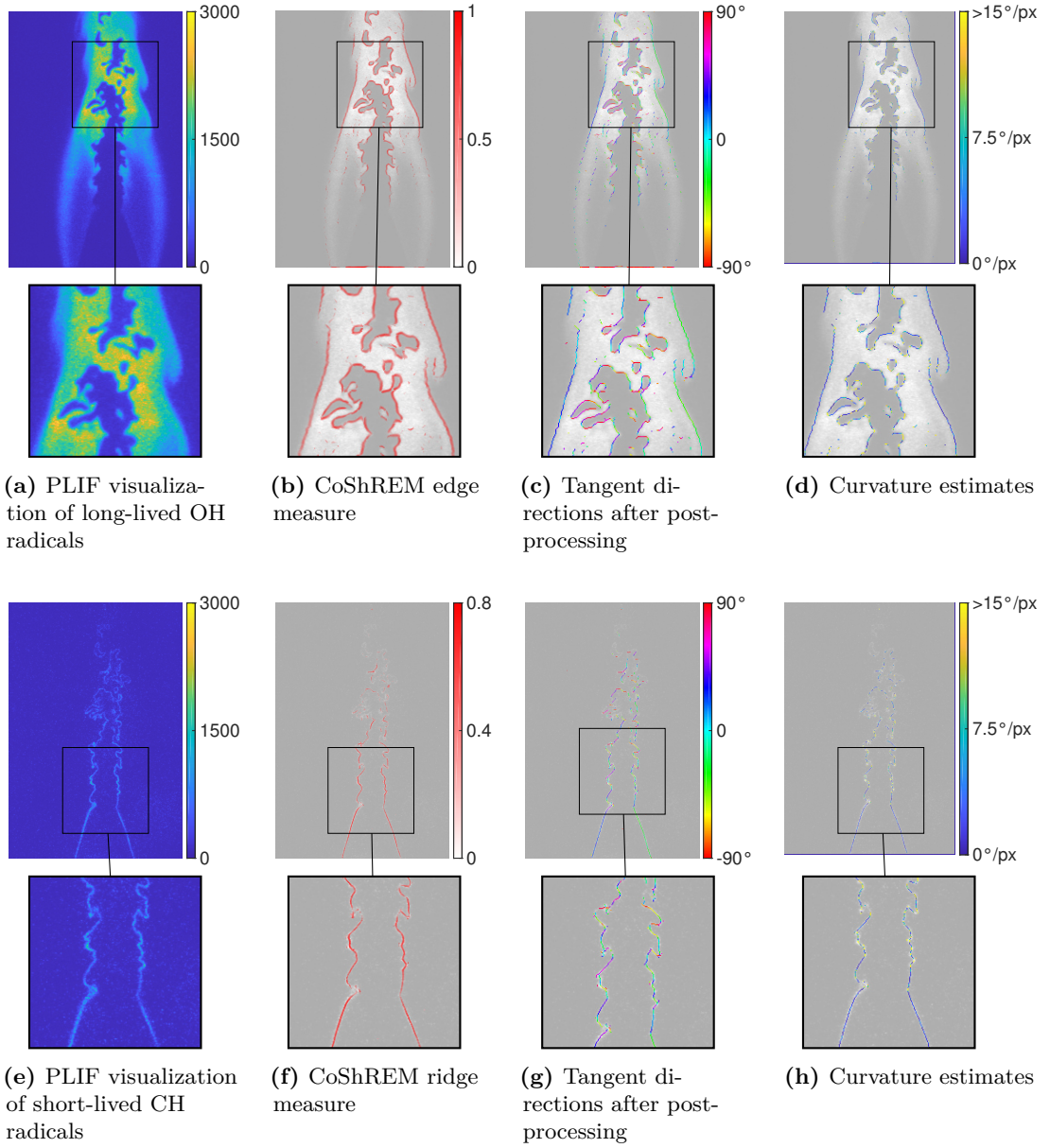


Figure 4.22: Detection and analysis of flame fronts with the edge measure $E(f, \mathbf{y})$ and the ridge measure $R(f, \mathbf{y})$, as implemented in CoShREM.

for completeness, we have also conducted a comparison of the edge detection algorithms in a setting where the input image was corrupted by five different levels of additive Gaussian noise and additional Poisson/shot noise. The respective results are compiled in Table 4.9.

4.6.3.2 Edge and Ridge Detection on PLIF Images

We conclude Section 4.6.3 by applying CoShREM to CH and OH PLIF images of a turbulent jet flame to demonstrate its performance in the analysis of experimental data. For this purpose, we reuse images recorded in simultaneous single-shot CH/OH PLIF experiments of a turbulent jet flame. A description of the experimental conditions as well as an analysis of the images can be found in a previous article [87], where the burner and the diagnostic setup as well as the two images are discussed from a combustion point of view in detail. In Figures 4.22(a) to 4.22(d), detected flame front locations, local tangent orientations, and local curvature in a PLIF recording of long-lived OH radicals are shown. Figures 4.22(e) to 4.22(h) depict detected flame front locations, local tangent orientations, and local curvature in a PLIF recording of short-lived CH radicals. It can be seen that the flame front is picked up reliably in both cases.

4.7 Discussion of Numerical Results

The results of the numerical evaluation on synthetic images summarized in Tables 4.2, 4.3, and 4.4 indicate that the detection accuracy of the proposed measures $E(f, \mathbf{y})$, $R(f, \mathbf{y})$, and $B(f, \mathbf{y})$ at least matches the performance of some of the most popular state of the art algorithms for edge, ridge, and blob detection. In particular, all three measures show a high robustness in the presence of noise. Maybe the most significant instance of this can be found in the case of blob detection, where the circular Hough transform clearly fails to reliably detect filled circles in severely distorted images, while SymFD still yields a perfect detection result in the case of *synthetic image 5* and correctly identifies three quarters of the blobs in *synthetic image 6* (see Table 4.4 and Figure 4.14). However, the experiments also show that regarding the mere detection of features, SymFD does not provide a significant advantage over some of the already established methods for the considered test inputs. In the case of edge detection, the well-known Canny edge detector [28] matches or even slightly outperforms the detection rate of SymFD for all of the six test images (see Table 4.2) while the same is true in the case of ridge detection for the Steger algorithm [156] (see Table 4.3).

One of the most significant benefits of the proposed framework is that it not only allows for the detection of different types of features but also yields a comprehensive characterization of their geometry in terms of local tangent directions and diameters. The

respective mean absolute errors reported in Tables 4.2 and 4.3 indicate a comparatively high accuracy of the tangent orientation measures $O_E(f, \mathbf{y})$ and $O_R(f, \mathbf{y})$, and the width measure $W_R(f, \mathbf{y})$. However, the experiments also show that for all considered test images, the width and orientation measures are sensitive to noise, where the most significant decrease in accuracy can be found in the case of $O_E(f, \mathbf{y})$ and *synthetic image 2* (see Table 4.2). The visualization of the local error of the tangent direction estimates shown in Figure 4.9(d) furthermore reveals that the accuracy of the measure $O_E(f, \mathbf{y})$ deteriorates in the proximity of points on which the edge curve is not smooth (e.g. corner points). In the case of blob detection, the errors of the width measure $W_B(f, \mathbf{y})$ summarized in Table 4.4 and visualized in Figure 4.13(d) are relatively large, even in comparison with the results yielded by the circular Hough transform. This can in part be explained by the fact that the two-dimensional generators from the set $\widetilde{\Psi}_2^e$ are based on tensor products of one-dimensional even-symmetric wavelets and thus designed to detect square-shaped rather than circular blobs, which is also reflected in the definition of the function $\widetilde{K}_{\tilde{g}^e}(r)$ in Section 4.3.4 (see (4.104)). We suspect that the accuracy of the measure $W_B(f, \mathbf{y})$ with respect to filled circles would significantly be improved by using circular two-dimensional generators and adjusting the definition of $\widetilde{K}_{\tilde{g}^e}(r)$ accordingly. A first choice for such a generator could be the two-dimensional Mexican hat wavelet. We furthermore suspect that using a generator that actually matches the symmetry properties and shape of the objects that are to be detected would also have a positive effect on the overall accuracy of the measure $B(f, \mathbf{y})$.

The numerical experiments conducted in Section 4.6.1 show that the measures $R(f, \mathbf{y})$, $O_R(f, \mathbf{y})$, and $W_R(f, \mathbf{y})$ can be used to reliably describe the geometry of blood vessels in different types of retinal images. In particular, the tangent direction estimates yielded by $O_R(f, \mathbf{y})$ are highly accurate in the sense that their mean deviation from the ground truth is only slightly above the variation within the group of human experts (see Table 4.5). While the width measure $W_R(f, \mathbf{y})$ yields more accurate estimates of local vessel diameters than many of the classical approaches it is also clearly outperformed by some of the more recent algorithms that were specifically developed for processing retinal images, such as ARIA [8] and ETOS [10] (see Table 4.6).

The applicability of the blob measure $B(f, \mathbf{y})$ for automatically counting the number of cell colonies on an agar plate was investigated in Section 4.6.2. It has been suggested that in practice, an average relative deviation of less than 20 % will often be negligible in comparison to noise introduced by other experimental factors [65]. When using an additional foreground mask, SymFD achieves a median relative deviation from the ground truth of about 2 % on a set of 19 high-definition images but also on a set of 19 low-quality webcam images (cf. Table 4.7). This accuracy is comparable to the accuracy of the widely

used OpenCFU software package [65] and suggests a high reliability of the results obtained by SymFD when applied for cell colony detection. The values reported in Table 4.7 also show only a slight decrease in accuracy for SymFD when omitting the foreground mask, which indicates that the measure $B(f, \mathbf{y})$ is not sensitive to structures in the image background or at the border of a Petri dish. Similar to other methods, SymFD tends to slightly overestimate the number of cell colonies on sparse plates while underestimating the cell colony count for high-density plates (cf. Figure 4.19). This is mostly due to the high significance of false positives when only a few colonies are present and the difficulty of correctly identifying single colonies in dense clusters which often occur on high-density plates.

Section 4.6.3 demonstrates that the measures $E(f, \mathbf{y})$, $R(f, \mathbf{y})$, $O_E(f, \mathbf{y})$, and $O_R(f, \mathbf{y})$ can be reliably applied over a wide range of noise levels to characterize the geometry of flame fronts. When compared to other state of the art algorithms, the CoShREM implementation of the newly proposed edge and ridge measures provides a consistently high accuracy in a setting where the parameters of the considered methods remain fixed across different types and levels of noise (cf. Tables 4.8, 4.9, and 4.10). At least visually, these results can also be confirmed when considering experimentally obtained data of a turbulent jet flame (cf. Figures 4.22). In particular, the results obtained in Section 4.6.3 suggest that considering finite differences computed on the post-processed results of the tangent direction estimates $O_E(f, \mathbf{y})$, and $O_R(f, \mathbf{y})$ yields a simple but effective way of obtaining curvature estimates.

The numerical experiments described in Sections 4.5 and 4.6 suggest that the proposed measures indeed define a highly versatile and powerful framework that can be reliably applied in a wide range of different feature detection and characterization tasks. There is, however, a trade-off for this high degree of flexibility, namely the relatively large number of parameters that are required to be selected when configuring SymFD for a specific task (cf. Table 4.1). Furthermore, evaluating the proposed edge, ridge, and blob measures on a digital image often requires the computation of more than 100 convolutions with digital symmetric molecule filters. While an average execution time of slightly less than 20 seconds (cf. Table 4.1) is still acceptable in most tasks, SymFD is currently far from being applicable in real-time applications.

4.8 Conclusion

In the present work, we propose a novel framework for the detection and characterization of features such as edges, ridges, and blobs in two-dimensional images. Drawing inspiration from investigations of the phase congruency property [99, 129, 130], the developed measures

exploit the special symmetry properties of directionally sensitive analyzing functions that are based on tensor products of L^1 -normalized derivatives of the one-dimensional Gaussian and their Hilbert transforms. Differently scaled and oriented two-dimensional analyzing functions are then constructed within the recently introduced α -molecule framework [71], which provides a maximum degree of flexibility. Combining both old and new ideas from applied harmonic analysis and computational vision, the developed measures are by construction contrast invariant and furthermore capable of fully characterizing the geometry of features in terms of local tangent directions and local widths.

We have demonstrated that the proposed measures can yield state of the art detection performances when considering sets of both clean and distorted artificial images that are associated with reliable ground truths. Furthermore, we have given examples of how the developed feature detectors can be applied in tasks like the detection of flame fronts, the characterization and detection of blood vessels in retinal images, or the automated counting of cell colonies in a Petri dish. The obtained experimental results are promising and suggest a high applicability of the proposed measures in a wide range of diverse applications. However, they also reveal a number of open issues that leave ample room for future improvements.

It was already discussed in Section 4.7 that the width measure $W_B(f, \mathbf{y})$ seems to be inherently biased when used for the characterization of blobs with a circular shape, because the considered tensor product constructions implicitly define two-dimensional generators that are associated with square shapes. A possible solution to this problem would be to further generalize the proposed framework to non-separable two-dimensional generators that are not based on tensor products such as the two-dimensional Mexican hat wavelet. It could furthermore be interesting to investigate which types of generators are ideal when using the blob measure $B(f, \mathbf{y})$ to detect other regular polygons or ellipsoids and whether an orientation measure similar to $O_E(f, \mathbf{y})$ and $O_R(f, \mathbf{y})$ would be successful at recovering the exact rotation of such objects.

The tangent direction and width measures $O_E(f, \mathbf{y})$, $O_R(f, \mathbf{y})$, $W_R(f, \mathbf{y})$, and $W_B(f, \mathbf{y})$ utilize a refinement procedure in which a parabola is fit through a finite number of points in order to locally approximate the behavior of even- and odd-symmetric coefficients when considered as functions of the scaling parameter a^j or the rotation parameter θ . A parabola fit requires only three discrete sample points and yields a uniquely defined argument of the maximum that can be computed by a simple closed formula. These properties make the parabola a practical but maybe not ideal candidate to provide a model for the considered functions. It is not entirely clear how well a parabola can approximate the respective functions and how stable the obtained fits are in the presence of noise. A further mathematical analysis of the expected behavior of the even- and odd-symmetric

coefficients when considering specific two-dimensional generators could improve both our understanding and the precision of the proposed width and orientation measures.

Derivatives of the Gaussian and their Hilbert transforms constitute a large classes of even- and odd-symmetric generators with explicit formulations in the frequency domain. A main advantage of using such functions in the proposed framework is that this allows us to obtain arbitrarily scaled and rotated two-dimensional digital filters by directly sampling the corresponding symmetric molecules in the frequency domain. While derivatives of the Gaussian belong to the Schwartz space of rapidly decreasing functions and are thus equipped with a desirable decay behavior in the time domain, none of the generators considered in this work are actually compactly supported in the time domain. It would be interesting to investigate how the proposed measures could be defined and implemented using compactly supported generators such as, for example, splines. Considering that features such as edges, ridges, or blobs are defined by structures that are highly localized in the time domain, this could lead to significant improvements with respect to detection rates and accuracy.

By considering pairs of analyzing functions that only differ with respect to 90° phase shifts (i.e. define Hilbert transform pairs) on different scales and with respect to different orientations, the measures $E(f, \mathbf{y})$, $R(f, \mathbf{y})$, and $B(f, \mathbf{y})$ are strongly inspired by functional properties that are known to be exhibited by neurons in the early human visual system. Another property that has widely been observed in neural populations is that the response of single neurons is often normalized in the sense that it is divided by a weighted sum of the responses of neighboring neurons [15, 29, 79]. This principle is often termed *divisive normalization* and closely related to the normalization procedure which ensures contrast invariance in the proposed measures (cf. (4.66), (4.87), (4.102), and (4.110)). The implementations of the developed edge, ridge, and blob measures for digital images could furthermore also be modeled as artificial neural networks, where the applied even- and odd-symmetric digital filters define a convolutional layer, the absolute value is used as a non-linearity and max-pooling as well as divisive normalization are applied to obtain the final normalized values. It would be interesting to see whether a neural network architecture that is based on similar design principles would yield computations that are related to the formulas proposed in this work when trained for the detection of edges, ridges, or blobs. Such an approach might also lead to new ideas how the proposed measures could be improved to better handle difficulties such as corner points, intersections of ridges, or densely grouped blobs.

The present work only considers two-dimensional images. However, similar measures and construction principles of even- and odd-symmetric generators could also be applied in a three-dimensional setting to detect and characterize surfaces, planes, or filled three-

dimensional objects. The definition as well as the mathematical and experimental analysis of such measures could also be a fruitful topic of future research.

4.9 Acknowledgments

We thank J.-L. Starck for suggesting to extend the framework to the case of blob detection and M. Schäfer for insightful discussions on the subject of α -molecules. We are grateful to Q. Geissmann for kindly sharing a dataset of images of cell colonies and the corresponding ground truth counts. We furthermore thank D. Labate for sharing the MATLAB implementation of his shearlet-based edge detector.

Bibliography

1. ABBOTT, B. P., ABBOTT, R., ABBOTT, T., ABERNATHY, M., ACERNESE, F., ACKLEY, K., ADAMS, C., ADAMS, T., ADDESSO, P., ADHIKARI, R., et al.: “Observation of gravitational waves from a binary black hole merger”. *Phys. Rev. Lett.* (2016), vol. 116(6): p. 061102 (cit. on p. 1).
2. ABDOU, I. E. and PRATT, W. K.: “Quantitative design and evaluation of enhancement/thresholding edge detectors”. *Proc. IEEE* (May 1979), vol. 67(5): pp. 753–763 (cit. on p. 117).
3. ABOLMAESUMI, P., SALCUDEAN, S. E., ZHU, W.-H., SIROUSPOUR, M. R., and DI MAIO, S. P.: “Image-guided control of a robot for medical ultrasound”. *IEEE Trans. Robot. Autom.* (Feb. 2002), vol. 18(1): pp. 11–23 (cit. on p. 53).
4. ABRÀMOFF, M. D., GARVIN, M. K., and SONKA, M.: “Retinal imaging and image analysis”. *IEEE Rev. Biomed. Eng.* (2010), vol. 3: pp. 169–208 (cit. on p. 127).
5. ADELSON, E. H. and BERGEN, J. R.: “Spatiotemporal energy model for the perception of motion”. *J. Opt. Soc. Am. A* (1985), vol. 2(2): pp. 284–299 (cit. on p. 90).
6. ALDÉN, M., BOOD, J., LI, Z., and RICHTER, M.: “Visualization and understanding of combustion processes using spatially and temporally resolved laser diagnostic techniques”. *Proc. Combust. Inst.* (2011), vol. 33: pp. 69–97 (cit. on p. 134).
7. ATHERTON, T. J. and KERBYSON, D. J.: “Size invariant circle detection”. *Image Vis. Comput.* (1999), vol. 17(11): pp. 795–803 (cit. on pp. 116, 125, 126).
8. BANKHEAD, P., SCHOLFIELD, C. N., MCGEOWN, J. G., and CURTIS, T. M.: “Fast retinal vessel detection and measurement using wavelets and edge location refinement”. *PLoS One* (Mar. 2012), vol. 7(3): pp. 1–12 (cit. on pp. 130, 131, 142).
9. BARAJAS, M., DÁVALOS-VIVEROS, J. P., GARCIA-LUMBRERAS, S., and GORDILLO, J. L.: “Visual servoing of UAV using cuboid model with simultaneous tracking of multiple planar faces”. *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Nov. 2013: pp. 596–601 (cit. on p. 53).

10. BEKKERS, E., DUITTS, R., BERENDSCHOT, T., and TER HAAR ROMENY, B.: “A multi-orientation analysis approach to retinal vessel tracking”. *J. Math. Imag. Vis.* (July 2014), vol. 49(3): pp. 583–610 (cit. on pp. [130](#), [131](#), [142](#)).
11. BEKKERS, E., ZHANG, J., DUITTS, R., and TER HAAR ROMENY, B.: “Curvature based biomarkers for diabetic retinopathy via exponential curve fits in $SE(2)$ ”. (2015), vol. Ed. by X, C., MK, G., JJ, L., E, T., and Y, X.: pp. 113–120 (cit. on p. [127](#)).
12. BENEDETTO, J. J.: *Harmonic analysis and applications*. Vol. 23. CRC Press, 1996 (cit. on p. [89](#)).
13. BERNARD, C. P.: “Discrete wavelet analysis for fast optic flow computation”. *Appl. Comput. Harmon. Anal.* (2001), vol. 11(1): pp. 32–63 (cit. on pp. [58](#), [60](#), [61](#)).
14. BOLKAS, D., VAZAIOS, I., PEIDOU, A., and VLACHOPOULOS, N.: “Detection of rock discontinuity traces using terrestrial LiDAR data and space-frequency transforms”. *Geotech. Geol. Eng.* (Dec. 2017), vol. (cit. on p. [126](#)).
15. BONDS, A. B.: “Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex”. *Vis. Neurosci.* (1989), vol. 2(1): pp. 41–55 (cit. on p. [145](#)).
16. BOSSE, S., MANIRY, D., MÜLLER, K.-R., WIEGAND, T., and SAMEK, W.: “Full-reference image quality assessment using neural networks”. *Int. Work. Qual. Multi-med. Exp.* 2016 (cit. on p. [18](#)).
17. BOSSE, S., MANIRY, D., WIEGAND, T., and SAMEK, W.: “A deep neural network for image quality assessment”. *2016 IEEE International Conference on Image Processing (ICIP)*. 2016 (cit. on p. [18](#)).
18. BRINCHMANN-HANSEN, O. and HEIER, H.: “Theoretical relations between light streak characteristics and optical properties of retinal vessels”. *Acta Ophthalmologica* (1986), vol. 64(S179): pp. 33–37 (cit. on pp. [130](#), [131](#)).
19. BRUNA, J. and MALLAT, S.: “Invariant scattering convolution networks”. *IEEE Trans. Pattern Anal. Mach. Intell.* (Aug. 2013), vol. 35(8): pp. 1872–1886 (cit. on p. [1](#)).
20. BURT, P. and ADELSON, E.: “The Laplacian pyramid as a compact image code”. *IEEE Trans. Comm.* (Apr. 1983), vol. 31(4): pp. 532–540 (cit. on p. [53](#)).
21. C. D. SLABAUGH, A. C. PRATT, and R. P. LUCHT: “Simultaneous 5 kHz OH-PLIF/PIV for the study of turbulent combustion at engine conditions”. *Appl. Phys. B Laser Optic.* (2015), vol. 118: pp. 109–130 (cit. on pp. [135](#), [136](#)).

22. CAI, Z., CHATTOPADHYAY, N., LIU, W. J., CHAN, C., PIGNOL, J.-P., and REILLY, R. M.: “Optimized digital counting colonies of clonogenic assays using ImageJ software and customized macros: comparison with manual counting”. *Int. J. Radiat. Biol.* (2011), vol. 87(11): pp. 1135–1146 (cit. on p. [133](#)).
23. CALDERÓN, A.: “Intermediate spaces and interpolation, the complex method”. *Stud. Math.* (1964), vol. 24(2): pp. 113–190 (cit. on p. [2](#)).
24. CANDÈS, E. J., ROMBERG, J., and TAO, T.: “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information”. *IEEE Trans. Inform. Theor.* (Feb. 2006), vol. 52(2): pp. 489–509 (cit. on p. [4](#)).
25. CANDÈS, E. J. and DONOHO, D. L.: *Curvelets: a surprisingly effective nonadaptive representation for objects with edges*. Tech. rep. Stanford Univ. Ca. Dept. of Statistics, 2000 (cit. on pp. [3](#), [45](#), [47](#)).
26. CANDÈS, E. J. and DONOHO, D. L.: “Ridgelets: a key to higher-dimensional intermittency?” *Phil. Trans. Roy. Soc. Lond.* (1999), vol. 357(1760): pp. 2495–2509 (cit. on pp. [3](#), [79](#)).
27. CANDÈS, E. J.: “Ridgelets: theory and applications”. PhD Thesis. Stanford University, 1998 (cit. on pp. [3](#), [79](#)).
28. CANNY, J.: “A computational approach to edge detection”. *IEEE Trans. Pattern Anal. Mach. Intell.* (1986), vol. 8(6): pp. 679–698 (cit. on pp. [86](#), [87](#), [116](#), [119–121](#), [136–138](#), [141](#), [179](#)).
29. CARANDINI, M., HEEGER, D. J., and MOVSHON, J. A.: “Linearity and normalization in simple cells of the macaque primary visual cortex”. *J. Neurosci.* (1997), vol. 17(21): pp. 8621–8644 (cit. on p. [145](#)).
30. CHAUMETTE, F. and BOUKIR, S.: “Structure from motion using an active vision paradigm”. *Proceedings of the 11th IAPR International Conference on Pattern Recognition*. Aug. 1992: pp. 41–44 (cit. on p. [53](#)).
31. CHAUMETTE, F. and HUTCHINSON, S.: “Visual servo control. I. Basic approaches”. *IEEE Robot. Autom. Mag.* (Dec. 2006), vol. 13(4): pp. 82–90 (cit. on pp. [53](#), [56](#)).
32. CHEUNG, C. Y.-L., LAMOUREUX, E., IKRAM, M. K., SASONGKO, M. B., DING, J., ZHENG, Y., MITCHELL, P., WANG, J. J., and WONG, T. Y.: “Retinal vascular geometry in asian persons with diabetes and retinopathy”. *J. Diabetes Sci. Technol.* (2012), vol. 6(3): pp. 595–605 (cit. on p. [127](#)).
33. CHRISTENSEN, O. et al.: *An introduction to frames and Riesz bases*. Springer, 2016 (cit. on pp. [46](#), [47](#)).

34. CISCO: “Cisco visual networking index: forecast and Methodology, 2015–2020”. *White paper* (2016), vol. (cit. on p. 16).
35. CLARKE, M. L., BURTON, R. L., HILL, A. N., LITORJA, M., NAHM, M. H., and HWANG, J.: “Low-cost, high-throughput, automated counting of bacterial colonies”. *Cytometry* (2010), vol. 77A(8): pp. 790–797 (cit. on p. 133).
36. COHEN, A., DAUBECHIES, I., and FEAUVEAU, J.-C.: “Biorthogonal bases of compactly supported wavelets”. *Comm. Pure Appl. Math.* (1992), vol. 45(5): pp. 485–560 (cit. on p. 35).
37. COLLEWET, C. and MARCHAND, E.: “Photometric visual servoing”. *IEEE Trans. Robot.* (Aug. 2011), vol. 27(4): pp. 828–834 (cit. on pp. 55, 56).
38. CROMBEZ, N., CARON, G., and MOUADDIB, E. M.: “Photometric Gaussian mixtures based visual servoing”. *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Sept. 2015: pp. 5486–5491 (cit. on p. 53).
39. D. C. FOURGUETTE, R. M. ZURN, and M. B. LONG: “Two-dimensional Rayleigh scattering thermometry in a turbulent nonpremixed methane-hydrogen flame”. *Combust. Sci. Tech.* (1986), vol. 44: pp. 307–317 (cit. on p. 134).
40. D. G. BAILEY: “Sub-pixel estimation of local extrema”. *Proceedings of Image and Vision Computing New Zealand*. Nov. 2003: pp. 414–419 (cit. on p. 93).
41. DALY, S. J.: “Application of a noise-adaptive contrast sensitivity function to image data compression”. *Opt. Eng.* (1990), vol. 29(8): pp. 977–987 (cit. on p. 17).
42. DAME, A. and MARCHAND, E.: “Mutual information-based visual servoing”. *IEEE Trans. Robot.* (Oct. 2011), vol. 27(5): pp. 958–969 (cit. on p. 53).
43. DAUBECHIES, I.: “Orthonormal bases of compactly supported wavelets”. *Comm. Pure Appl. Math.* (1988), vol. 41(7): pp. 909–996 (cit. on pp. 2, 12, 35, 37).
44. DAUBECHIES, I.: “Orthonormal bases of compactly supported wavelets II: variations on a theme”. *SIAM J. Math. Anal.* (1993), vol. 24(2): pp. 499–519 (cit. on p. 35).
45. DAUBECHIES, I.: *Ten lectures on wavelets*. Vol. 61. CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 1992 (cit. on pp. 2, 35, 38, 40, 41).
46. DAVIES, E.: “Chapter 10 – circle detection”. *Machine Vision*. Ed. by DAVIES, E. Third Edition. Signal Processing and its Applications. Burlington: Morgan Kaufmann, 2005: pp. 283–313 (cit. on pp. 116, 125, 126).

47. AL-DIRI, B., HUNTER, A., and STEEL, D.: "An active contour model for segmenting and measuring retinal vessels". *IEEE Trans. Med. Imag.* (Sept. 2009), vol. 28(9): pp. 1488–1497 (cit. on pp. 130, 131).
48. AL-DIRI, B., HUNTER, A., STEEL, D., HABIB, M., HUDAIB, T., and BERRY, S.: "REVIEW – A reference data set for retinal vessel profiles". *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Aug. 2008: pp. 2262–2265 (cit. on p. 127).
49. DO, M. N. and VETTERLI, M.: "Contourlets". *Studies in Computational Mathematics*. Vol. 10. Elsevier, 2003: pp. 83–105 (cit. on pp. 3, 45, 79).
50. DONOHO, D. L.: "Sparse components of images and optimal atomic decompositions". *Constr. Approx.* (Jan. 2001), vol. 17(3): pp. 353–382 (cit. on p. 45).
51. DONOHO, D., MALEKI, A., and SHAHRAM, M.: *WaveLab 850: software toolkit for time-frequency analysis*. Available from <https://statweb.stanford.edu/~wavelab/>. 2006 (cit. on p. 101).
52. DONOHO, D. L.: "Sparse components of images and optimal atomic decomposition". *Constr. Approx.* (2001), vol. 17: pp. 353–382 (cit. on pp. 49, 50).
53. DOSSELMANN, R. and YANG, X. D.: "A comprehensive assessment of the structural similarity index". *Signal, Image and Video Processing* (2011), vol. 5(1): pp. 81–91 (cit. on p. 17).
54. DUFLOT, L.-A., KRUPA, A., TAMADAZTE, B., and ANDREFF, N.: "Shearlet-based vs. photometric-based visual servoing for robot-assisted medical applications". *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2016: pp. 4099–4104 (cit. on p. 54).
55. DUFLOT, L.-A., KRUPA, A., TAMADAZTE, B., and ANDREFF, N.: "Towards ultrasound-based visual servoing using shearlet coefficients". *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2016: pp. 3420–3425 (cit. on p. 54).
56. DUFLOT, L.-A., REISENHOFER, R., TAMADAZTE, B., ANDREFF, N., and KRUPA, A.: "Wavelet and shearlet-based image representations for visual servoing". To appear in *Int. J. Robot. Res.* (cit. on pp. 37, 75).
57. E. BAYLEY, A. E., HARDALUPAS, Y., and M. K. P. TAYLOR, A. M. K. P.: "Local curvature measurements of a lean, partially premixed swirl-stabilised flame". *Exp. Fluid.* (2012), vol. 52: pp. 963–983 (cit. on p. 135).

58. EASLEY, G. R., LABATE, D., and COLONNA, F.: “Shearlet-based total variation diffusion for denoising”. *IEEE Trans. Image Process.* (Feb. 2009), vol. 18(2): pp. 260–268 (cit. on p. 45).
59. FIELLER, E. C., HARTLEY, H. O., and PEARSON, E. S.: “Tests for rank correlation coefficients. I”. *Biometrika* (1957), vol. 44(3/4): pp. 470–481 (cit. on p. 27).
60. FLANDERS, H.: “Differentiation under the integral sign”. *Am. Math. Mon.* (1973), vol. 80(6): pp. 615–627 (cit. on p. 57).
61. FOUCART, S. and RAUHUT, H.: *A mathematical introduction to compressive sensing*. Vol. 1. 3. Birkhäuser Basel, 2013 (cit. on p. 4).
62. FOURIER, J.: *Theorie analytique de la chaleur*. Chez Firmin Didot, père et fils, 1822 (cit. on p. 1).
63. FRANGI, A. F., NIESSEN, W. J., VINCKEN, K. L., and VIERGEVER, M. A.: “Multiscale vessel enhancement filtering”. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 1998: pp. 130–137 (cit. on pp. 116, 122, 123).
64. GABOR, D.: “Theory of communication. Part 1: the analysis of information”. *Journal of the Institution of Electrical Engineers – Part III: Radio and Communication Engineering* (26 Nov. 1946), vol. 93: 429–441(12) (cit. on p. 2).
65. GEISSMANN, Q.: “OpenCFU, a new free and open-source software to count cell colonies and other circular objects”. *PLoS One* (Feb. 2013), vol. 8(2): pp. 1–10 (cit. on pp. 131–133, 142, 143).
66. GENZEL, M. and KUTYNIOK, G.: “Asymptotic analysis of inpainting via universal shearlet systems”. *SIAM J. Imag. Sci.* (2014), vol. 7(4): pp. 2301–2339 (cit. on p. 47).
67. GIROD, B.: “What’s wrong with mean-squared error?” *Digital Images and Human Vision*. 1993: pp. 207–220 (cit. on p. 17).
68. GREGSON, P., SHEN, Z., SCOTT, R., and KOZOUSEK, V.: “Automated grading of venous beading”. *Comput. Biomed. Res.* (1995), vol. 28(4): pp. 291–304 (cit. on pp. 130, 131).
69. GRÖCHENIG, K.: *Foundations of time-frequency analysis*. Springer Science & Business Media, 2013 (cit. on p. 46).
70. GROHS, P., KEIPER, S., KUTYNIOK, G., and SCHÄFER, M.: “Cartoon approximation with α -curvelets”. *J. Fourier Anal. Appl.* (Dec. 2016), vol. 22(6): pp. 1235–1293 (cit. on p. 85).

71. GROHS, P., KEIPER, S., KUTYNIOK, G., and SCHÄFER, M.: “ α -molecules”. *Appl. Comput. Harmon. Anal.* (2016), vol. 41(1): pp. 297–336 (cit. on pp. [6](#), [47](#), [80](#), [81](#), [83–86](#), [103](#), [144](#)).
72. GROHS, P., KEIPER, S., KUTYNIOK, G., and SCHÄFER, M.: “Parabolic molecules: curvelets, shearlets, and beyond”. *Approximation Theory XIV: San Antonio 2013*. Ed. by FASSHAUER, G. E. and SCHUMAKER, L. L. Springer International Publishing, 2014: pp. 141–172 (cit. on p. [80](#)).
73. GROHS, P. and KUTYNIOK, G.: “Parabolic molecules”. *Found. Comput. Math.* (Apr. 2014), vol. 14(2): pp. 299–337 (cit. on pp. [6](#), [80](#), [82–84](#)).
74. GROSSMANN, A. and MORLET, J.: “Decomposition of Hardy functions into square integrable wavelets of constant shape”. *SIAM J. Math. Anal.* (1984), vol. 15(4): pp. 723–736 (cit. on p. [76](#)).
75. GUO, K. and LABATE, D.: “Characterization and analysis of edges Using the continuous shearlet transform”. *SIAM J. Imag. Sci.* (2009), vol. 2(3): pp. 959–986 (cit. on p. [106](#)).
76. HAAR, A.: “Zur Theorie der orthogonalen Funktionensysteme”. *Math. Ann.* (1910), vol. 69(3): pp. 331–371 (cit. on pp. [2](#), [11](#), [12](#), [37](#)).
77. HAIRER, M.: “A theory of regularity structures”. *Invent. Math.* (Nov. 2014), vol. 198(2): pp. 269–504 (cit. on p. [1](#)).
78. HÄUSER, S. and STEIDL, G.: “Fast finite shearlet transform”. *arXiv preprint arXiv:1202.1773* (2012), vol. (cit. on p. [52](#)).
79. HEEGER, D. J.: “Normalization of cell responses in cat striate cortex”. *Vis. Neurosci.* (1992), vol. 9(2): pp. 181–197 (cit. on p. [145](#)).
80. HEIL, C.: *A basis theory primer: expanded edition*. Springer Science & Business Media, 2010 (cit. on p. [2](#)).
81. HORN, B. K. and SCHUNCK, B. G.: “Determining optical flow”. *Artif. Intell.* (1981), vol. 17(1): pp. 185–203 (cit. on p. [56](#)).
82. HUBEL, D. H. and WIESEL, T. N.: “Sequence regularity and geometry of orientation columns in the monkey striate cortex”. *J. Comp. Neurol.* (1974), vol. 158(3): pp. 267–293 (cit. on p. [34](#)).
83. HUTCHINSON, S., HAGER, G. D., and CORKE, P. I.: “A tutorial on visual servo control”. *IEEE Trans. Robot. Autom.* (Oct. 1996), vol. 12(5): pp. 651–670 (cit. on p. [53](#)).

84. IKRAM, M. K., WITTEMAN, J. C., VINGERLING, J. R., BRETELER, M. M., HOFMAN, A., and JONG, P. T. de: "Retinal vessel diameters and risk of hypertension". *Hypertension* (2006), vol. 47(2): pp. 189–194 (cit. on p. 127).
85. INTERNATIONAL TELECOMMUNICATION UNION: "ITU-T P.1401, methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models". (2012), vol. (cit. on p. 27).
86. J. DYER, M. J. and R. CROSLLEY, D. R.: "Two-dimensional imaging of OH laser-induced fluorescence in a flame". *Optic. Lett.* (1982), vol. 7: pp. 382–384 (cit. on p. 134).
87. J. KIEFER, Z. S. LI, J. ZETTERBERG, X. S. BAI, and M. ALDEN: "Investigation of local flame structures and statistics in partially premixed turbulent jet flames using simultaneous single-shot CH and OH planar laser-induced fluorescence imaging". *Combust. Flame* (2008), vol. 154: pp. 802–818 (cit. on pp. 135, 141).
88. JACOB, M. and UNSER, M.: "Design of steerable filters for feature detection using Canny-like criteria". *IEEE Trans. Pattern Anal. Mach. Intell.* (2004), vol. 26(8): pp. 1007–1019 (cit. on p. 86).
89. J.H. FRANK, S.A. KAISER, and M.B. LONG: "Multiscalar imaging in partially premixed jet flames with argon dilution". *Combust. Flame* (2005), vol. 143: pp. 507–523 (cit. on p. 139).
90. JIA, Y., LIN, W., and KASSIM, A. A.: "Estimating just-noticeable distortion for video". *IEEE Trans. Circ. Syst. Video Tech.* (2006), vol. 16(7): pp. 820–829 (cit. on p. 17).
91. J.M. SEITZMAN, R.K. HANSON, P.A. DEBARBER, and C.F. HESS: "Application of quantitative two-line OH planar laser-induced fluorescence for temporally resolved planar thermometry in reacting flows". *Appl. Optic.* (June 1994), vol. 33(18): pp. 4000–4012 (cit. on p. 139).
92. K. MCMILLIN, B. K., L. PALMER, J. L., and K. HANSON, R. K.: "Temporally resolved, two-line fluorescence imaging of NO temperature in a transverse jet in a supersonic cross flow". *Appl. Optic.* (Dec. 1993), vol. 32(36): pp. 7534–7545 (cit. on p. 139).
93. KANG, L., YE, P., LI, Y., and DOERMANN, D.: "Convolutional neural networks for no-reference image quality assessment". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2014: pp. 1733–1740 (cit. on p. 18).

94. KARBALAALI, H., JAVAHERIAN, A., DAHLKE, S., and TORABI, S.: "Channel edge detection using 2D complex shearlet transform: a case study from the South Caspian Sea". *Explor. Geophys.* (2017), vol. (cit. on p. 126).
95. KING, E. J., KUTYNIOK, G., and LIM, W.-Q.: "Image inpainting: theoretical analysis and comparison of algorithms". *Wavelets and Sparsity XV*. Vol. 8858. Sept. 2013: p. 885802 (cit. on pp. 4, 45).
96. KING, E. J., REISENHOFER, R., KIEFER, J., LIM, W.-Q., LI, Z., and HEYGSTER, G.: "Shearlet-based edge detection: flame fronts and tidal flats". *Applications of Digital Image Processing XXXVIII*. Ed. by G. TESCHER, A. G. Vol. 9599. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. 2015 (cit. on pp. 79, 88, 126).
97. KINGSBURY, N.: "Image processing with complex wavelets". *Phil. Trans. Roy. Soc. Lond.* (1999), vol. 357(1760): pp. 2543–2560 (cit. on p. 76).
98. KÖRNER, T. W.: *Fourier analysis*. Cambridge university press, 1989 (cit. on p. 2).
99. KOVESI, P.: "Image features from phase congruency". *Videre: A Journal of Computer Vision Research* (1999), vol. 1(3): pp. 1–26 (cit. on pp. 87, 90, 91, 116, 119–123, 137–139, 143, 179–182).
100. KOVESI, P.: "Phase congruency: a low-level image invariant". *Psychol. Res.* (2000), vol. 64: pp. 136–148 (cit. on pp. 19, 87, 91, 138).
101. KOVESI, P. D.: *MATLAB and Octave functions for computer vision and image processing*. Centre for Exploration Targeting, School of Earth and Environment, The University of Western Australia. Available from <http://www.csse.uwa.edu.au/~pk/research/matlabfns/> (cit. on pp. 20, 91, 121, 138).
102. KROON, D.-J.: *Hessian based Frangi vesselness filter*. Available from <https://de.mathworks.com/matlabcentral/fileexchange/24409-hessian-based-frangi-vesselness-filter> (cit. on p. 123).
103. KRUPA, A., GANGLOFF, J., DOIGNON, C., MATHELIN, M. F. de, MOREL, G., LEROY, J., SOLER, L., and MARESCAUX, J.: "Autonomous 3-D positioning of surgical instruments in robotized laparoscopic surgery using visual servoing". *IEEE Trans. Robot. Autom.* (Oct. 2003), vol. 19(5): pp. 842–853 (cit. on p. 53).
104. KUTYNIOK, G. and LABATE, D.: *Shearlets: multiscale analysis for multivariate data*. Springer Science & Business Media, 2012 (cit. on pp. 3, 45, 49, 52).

105. KUTYNIOK, G. and LABATE, D.: “Shearlets – multiscale analysis for multivariate data”. Ed. by KUTYNIOK, G. and LABATE, D. Birkhäuser, 2012. Chap. Introduction to shearlets: pp. 1–38 (cit. on p. 47).
106. KUTYNIOK, G., LEMVIG, J., and LIM, W.-Q.: “Shearlets – multiscale analysis for multivariate data”. Ed. by KUTYNIOK, G. and LABATE, D. Birkhäuser, 2012. Chap. Shearlets and optimally sparse approximations: pp. 145–199 (cit. on pp. 50, 52).
107. KUTYNIOK, G., LIM, W.-Q., and REISENHOFER, R.: “ShearLab 3D: faithful digital shearlet transforms based on compactly supported shearlets”. *ACM Trans. Math. Software* (Jan. 2016), vol. 42(1): 5:1–5:42 (cit. on pp. 3, 52, 64).
108. KUTYNIOK, G. and PETERSEN, P.: “Classification of edges using compactly supported shearlets”. *Appl. Comput. Harmon. Anal.* (2017), vol. 42(2): pp. 245–293 (cit. on pp. 45, 106).
109. KUTYNIOK, G., SHAHRAM, M., and ZHUANG, X.: “ShearLab: a rational design of a digital parabolic scaling algorithm”. *SIAM J. Imag. Sci.* (2012), vol. 5(4): pp. 1291–1332 (cit. on pp. 52, 64).
110. LABATE, D., LIM, W.-Q., KUTYNIOK, G., and WEISS, G.: “Sparse multidimensional representation using shearlets”. Vol. 5914. 2005 (cit. on pp. 3, 45).
111. LAPARRA, V., BERARDINO, A., BALLÉ, J., and SIMONCELLI, E. P.: “Perceptually optimized image rendering”. arXiv preprint arXiv:1701.06641. 2017 (cit. on p. 32).
112. LARSON, E. C. and CHANDLER, D. M.: “Most apparent distortion: full-reference image quality assessment and the role of strategy”. *J. Electron. Imag.* (2010), vol. 19(1) (cit. on pp. 17, 24–26).
113. LIN, W. and KUO, C.-C. J.: “Perceptual visual quality metrics: a survey”. *J. Vis. Comm. Image Represent* (2011), vol. 22(4): pp. 297–312 (cit. on p. 17).
114. LINDBERG, T.: “Edge detection and ridge detection with automatic scale selection”. *Int. J. Comput. Vis.* (1998), vol. 30(2): pp. 117–154 (cit. on p. 136).
115. LIPPIELLO, V., SICILIANO, B., and VILLANI, L.: “Position-based visual servoing in industrial multirobot cells using a hybrid camera configuration”. *IEEE Trans. Robot.* (Feb. 2007), vol. 23(1): pp. 73–86 (cit. on p. 53).
116. LIU, A., LIN, W., and NARWARIA, M.: “Image quality assessment based on gradient similarity”. *IEEE Trans. Image Process.* (Apr. 2012), vol. 21(4): pp. 1500–1512 (cit. on pp. 17, 18).

117. LOWELL, J., HUNTER, A., STEEL, D., BASU, A., RYDER, R., and KENNEDY, R. L.: “Measurement of retinal vessel widths from fundus images based on 2-D modeling”. *IEEE Trans. Med. Imag.* (Oct. 2004), vol. 23(10): pp. 1196–1204 (cit. on pp. [130](#), [131](#)).
118. LUBIN, J.: “A human vision system model for objective picture quality measurements”. *International Broadcasting Convention* (1997), vol.: pp. 498–503 (cit. on p. [17](#)).
119. M. Z. HAQ, SHEPPARD, C. G. W., R. WOOLEY, D. A. GREENHALGH, and R. D. LOCKETT: “Wrinkling and curvature of laminar and turbulent premixed flames”. *Combust. Flame* (2002), vol. 131: pp. 1–15 (cit. on p. [135](#)).
120. MALLAT, S.: “A theory for multiresolution signal decomposition: the wavelet representation”. *IEEE Trans. Pattern Anal. Mach. Intell.* (July 1989), vol. 11(7): pp. 674–693 (cit. on pp. [2](#), [3](#), [12](#), [37](#), [38](#)).
121. MALLAT, S.: *A wavelet tour of signal processing*. Ed. by MALLAT, S. 3rd. Academic Press, 2009 (cit. on pp. [38](#), [40](#), [41](#), [44](#)).
122. MALM, H., SPARR, G., HULT, J., and F. KAMINSKI, C. F.: “Nonlinear diffusion filtering of images obtained by planar laser-induced fluorescence spectroscopy”. *J. Opt. Soc. Am.* (2000), vol. 17: pp. 2148–2156 (cit. on p. [136](#)).
123. MARR, D. and HILDRETH, E.: “Theory of edge detection”. *Proc. Biol. Sci.* (1980), vol. 207(1167): pp. 187–217 (cit. on p. [87](#)).
124. MARTURI, N., TAMADAZTE, B., DEMBÉLÉ, S., and PIAT, N.: “Image-guided nanopositioning scheme for SEM”. *IEEE Trans. Autom. Sci. Eng.* (Jan. 2018), vol. 15(1): pp. 45–56 (cit. on p. [53](#)).
125. MARTURI, N., TAMADAZTE, B., DEMBÉLÉ, S., and PIAT, N.: “Visual servoing schemes for automatic nanopositioning under scanning electron microscope”. *2014 IEEE International Conference on Robotics and Automation (ICRA)*. May 2014: pp. 981–986 (cit. on p. [53](#)).
126. MÁTHÉ, K., BUŞONIU, L., BARABÁS, L., IUGA, C. I., MICLEA, L., and BRABAND, J.: “Vision-based control of a quadrotor for an object inspection scenario”. *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*. June 2016: pp. 849–857 (cit. on p. [53](#)).

127. MEMARIANI, A., NIKOU, C., ENDRES, B. T., BASSÈRES, E., GAREY, K. W., and KAKADIARIS, I. A.: “DETCIC: detection of elongated touching cells with inhomogeneous illumination using a stack of conditional random fields”. *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP*. INSTICC. SciTePress, 2018: pp. 574–580 (cit. on p. [126](#)).
128. MEYER, Y.: *Wavelets and operators*. Vol. 1. Cambridge university press, 1992 (cit. on pp. [2](#), [37](#)).
129. MORRONE, M. C. and OWENS, R. A.: “Feature detection from local energy”. *Pattern Recogn. Lett.* (1987), vol. 6: pp. 303–313 (cit. on pp. [87](#), [90](#), [143](#)).
130. MORRONE, M. C., ROSS, J. R., BURR, D. C., and OWENS, R. A.: “Mach bands are phase dependent”. *Nature* (1986), vol. 324(6094): pp. 250–253 (cit. on pp. [20](#), [87](#), [90](#), [143](#)).
131. “New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities”. *Comm. Pure Appl. Math.* (), vol. 57(2): pp. 219–266 (cit. on p. [79](#)).
132. OURAK, M., TAMADAZTE, B., and ANDREFF, N.: “Partitioned camera-OCT based 6 DOF visual servoing for automatic repetitive optical biopsies”. *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Oct. 2016: pp. 2337–2342 (cit. on p. [53](#)).
133. OURAK, M., TAMADAZTE, B., LEHMANN, O., and ANDREFF, N.: “Wavelets-based 6 DOF visual servoing”. *2016 IEEE International Conference on Robotics and Automation (ICRA)*. May 2016: pp. 3414–3419 (cit. on p. [53](#)).
134. PALMER, S. E.: *Vision science: photons to phenomenology*. Vol. 1. MIT press Cambridge, MA, 1999 (cit. on p. [30](#)).
135. PATTON, N., ASLAM, T. M., MACGILLIVRAY, T., DEARY, I. J., DHILLON, B., EIKELBOOM, R. H., YOGESAN, K., and CONSTABLE, I. J.: “Retinal image analysis: concepts, applications and potential”. *Progr. Retin. Eye. Res.* (2006), vol. 25(1): pp. 99–127 (cit. on p. [127](#)).
136. PFADLER, S., BEYRAU, F., and LEIPERTZ, A.: “Flame front detection using conditioned particle image velocimetry (CPIV)”. *Optic. Express* (2007), vol. 15: pp. 15444–15456 (cit. on pp. [135](#), [136](#)).
137. PFADLER, S., BEYRAU, F., LÖFFLER, M., and LEIPERTZ, A.: “Application of a beam homogenizer to planar laser diagnostics”. *Optic. Express* (Oct. 2006), vol. 14(22): pp. 10171–10180 (cit. on p. [139](#)).

138. PONOMARENKO, N., LUKIN, V., ZELENSKY, A., EGIAZARIAN, K., CARLI, M., and BATTISTI, F.: “TID2008 – a database for evaluation of full-reference visual quality assessment metrics”. *Advances of Modern Radioelectronics* (2009), vol. 10(4): pp. 30–45 (cit. on pp. [25](#), [26](#)).
139. PONOMARENKO, N., JIN, L., IEREMEIEV, O., LUKIN, V., EGIAZARIAN, K., ASTOLA, J., VOZEL, B., CHEHDI, K., CARLI, M., BATTISTI, F., and KUO, C.-C. J.: “Image database TID2013: peculiarities, results and perspectives”. *Signal Process. Image Comm.* (2015), vol. 30: pp. 57–77 (cit. on pp. [25](#), [26](#)).
140. PREWITT, J. M.: “Object enhancement and extraction”. *Picture processing and Psychopictorics* (1970), vol. 10(1): pp. 15–19 (cit. on p. [86](#)).
141. R. O. DUDA and P. E. HART: “Use of the Hough transformation to detect lines and curves in pictures”. *Comm. ACM* (Jan. 1972), vol. 15: pp. 11–15 (cit. on p. [136](#)).
142. REISENHOFER, R.: “The complex shearlet transform and applications to image quality assessment”. Matser’s Thesis. Technische Universität Berlin, 2014 (cit. on pp. [76](#), [88](#), [92](#), [126](#)).
143. REISENHOFER, R., BOSSE, S., KUTYNIOK, G., and WIEGAND, T.: “A Haar wavelet-based perceptual similarity index for image quality assessment”. *Signal Process. Image Comm.* (2018), vol. 61: pp. 33–43 (cit. on p. [11](#)).
144. REISENHOFER, R., KIEFER, J., and KING, E. J.: “Shearlet-based detection of flame fronts”. *Exp. Fluid.* (Feb. 2016), vol. 57(3): p. 41 (cit. on pp. [45](#), [79](#), [88](#), [126](#)).
145. REISENHOFER, R. and KING, E. J.: “Edge, ridge, and blob detection with symmetric molecules”. In preparation. (cit. on p. [79](#)).
146. RENAUD, P., ANDREFF, N., DHOME, M., and MARTINET, P.: “Experimental evaluation of a vision-based measuring device for parallel machine-tool calibration”. *2002 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Vol. 2. 2002: 1868–1873 vol.2 (cit. on p. [53](#)).
147. RICHA, R., SZNITMAN, R., TAYLOR, R., and HAGER, G.: “Visual tracking using the sum of conditional variance”. *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Sept. 2011: pp. 2953–2958 (cit. on p. [53](#)).
148. ROBERTS, L. G.: “Machine perception of three-dimensional solids”. PhD thesis. Massachusetts Institute of Technology, 1963 (cit. on p. [86](#)).
149. SHEIKH, H. R. and BOVIK, A. C.: “Image information and visual quality”. *IEEE Trans. Image Process.* (2006), vol. 15: pp. 430–444 (cit. on p. [17](#)).

150. SHEIKH, H. R., WANG, Z., CORMACK, L., and BOVIK, A. C.: *LIVE image quality assessment database release 2*. Available from <http://live.ece.utexas.edu/research/quality> (cit. on pp. 25, 26).
151. SILVEIRA, G. and MALIS, E.: “Direct visual servoing: vision-based estimation and control using only nonmetric information”. *IEEE Trans. Robot.* (Aug. 2012), vol. 28(4): pp. 974–980 (cit. on p. 53).
152. SIMONCELLI, E. P. and FREEMAN, W. T.: “The steerable pyramid: a flexible architecture for multi-scale derivative computation”. *Proceedings of the International Conference on Image Processing*. Vol. 3. Oct. 1995: 444–447 vol.3 (cit. on p. 53).
153. SOBEL, I. and FELDMAN, G.: “A 3×3 isotropic gradient operator for image processing”. Presentation at the Stanford Artificial Intelligence Project. 1968 (cit. on pp. 86, 116, 119–121, 137, 138, 179).
154. STAAL, J., D. ABRÀMOFF, M. D., NIEMEIJER, M., A. VIERGEVER, M. A., and GINNEKEN, B. van: “Ridge-based vessel segmentation in color images of the retina”. *IEEE Trans. Med. Imag.* (Apr. 2004), vol. 23(4): pp. 501–509 (cit. on p. 136).
155. STARCK, J.-L., MURTAGH, F., and FADILI, J. M.: *Sparse image and signal processing: wavelets, curvelets, morphological diversity*. Cambridge university press, 2010 (cit. on p. 4).
156. STEGER, C.: “An unbiased detector of curvilinear structures”. *IEEE Trans. Pattern Anal. Mach. Intell.* (1998), vol. 20(2): pp. 113–125 (cit. on pp. 116, 122, 123, 141).
157. STORATH, M.: “Amplitude and sign decompositions by complex wavelets – theory and applications to image analysis”. Dissertation. München: Technische Universität München, 2013 (cit. on pp. 76, 92).
158. SWEENEY, M. and HOCHGREB, S.: “Autonomous extraction of optimal flame fronts in OH planar laser-induced fluorescence images”. *Appl. Optic.* (2009), vol. 48: pp. 3866–3877 (cit. on pp. 135, 136).
159. SZEGÖ, G.: *Orthogonal polynomials*. Vol. 23. American Mathematical Soc., 1939 (cit. on pp. 89, 97).
160. TAHRI, O., TAMTSIA, A. Y., MEZOUAR, Y., and DEMONCEAUX, C.: “Visual servoing based on shifted moments”. *IEEE Trans. Robot.* (June 2015), vol. 31(3): pp. 798–804 (cit. on p. 53).
161. TAMADAZTE, B., PIAT, N. L. F., and MARCHAND, E.: “A direct visual servoing scheme for automatic nanopositioning”. *IEEE ASME Trans. Mechatron.* (Aug. 2012), vol. 17(4): pp. 728–736 (cit. on p. 53).

162. THUROW, B., JIANG, N., and LEMPERT, W.: “Review of ultra-high repetition rate laser diagnostics for fluid dynamic measurements”. *Meas. Sci. Tech.* (2013), vol. 24: p. 012002 (cit. on p. [134](#)).
163. VENKATESH, S. and OWENS, R.: “On the classification of image features”. *Pattern Recogn. Lett.* (1990), vol. 11(5): pp. 339–349 (cit. on pp. [87](#), [90](#)).
164. WAGNER, T., HINER, M., and XRAYNAUD: *thorstenwagner/ij-ridgedetection: ridge detection 1.4.0*. Aug. 2017 (cit. on p. [123](#)).
165. WANG, Z., BOVIK, A. C., SHEIKH, H. R., and SIMONCELLI, E. P.: “Image quality assessment: from error visibility to structural similarity”. *IEEE Trans. Image Process.* (2004), vol. 13(4): pp. 600–612 (cit. on pp. [17–20](#)).
166. WANG, Z., SIMONCELLI, E. P., and BOVIK, A. C.: “Multi-scale structural similarity for image quality assessment”. *Proceedings of 37th IEEE Asilomar Conference on Signals, Systems and Computers*. 2003 (cit. on p. [17](#)).
167. WATSON, A., BORTHWICK, R., and TAYLOR, M.: “Image quality and entropy masking”. *SPIE Proceedings*. Vol. 3016. 1997: pp. 1–11 (cit. on p. [17](#)).
168. WILLIAMS, M. A., MCGOWAN, A. J., CARDWELL, C. R., CHEUNG, C. Y., CRAIG, D., PASSMORE, P., SILVESTRI, G., MAXWELL, A. P., and MCKAY, G. J.: “Retinal microvascular network attenuation in Alzheimer’s disease”. *Alzheimers Dement.* (2015), vol. 1(2): pp. 229–235 (cit. on p. [127](#)).
169. XU, X., NIEMEIJER, M., SONG, Q., SONKA, M., GARVIN, M. K., REINHARDT, J. M., and ABRAMOFF, M. D.: “Vessel boundary delineation on fundus images using graph-based approach”. *IEEE Trans. Med. Imag.* (June 2011), vol. 30(6): pp. 1184–1191 (cit. on pp. [130](#), [131](#)).
170. YE, P. and DOERMANN, D.: “No-reference image quality assessment using visual codebooks”. *IEEE Trans. Image Process.* (2012), vol. 21(7): pp. 3129–3138 (cit. on p. [18](#)).
171. YI, S., LABATE, D., EASLEY, G. R., and KRIM, H.: “A shearlet approach to edge analysis and detection”. *IEEE Trans. Image Process.* (2009), vol. 18(5): pp. 929–941 (cit. on pp. [45](#), [86](#), [116](#), [119–121](#), [137](#), [138](#), [179](#), [180](#)).
172. ZHANG, L. and LI, H.: “SR-SIM: a fast and high performance IQA index based on spectral residual”. *19th IEEE International Conference on Image Processing*. Sept. 2012: pp. 1473–1476 (cit. on pp. [17](#), [18](#)).

173. ZHANG, L., SHEN, Y., and LI, H.: “VSI: a visual saliency-induced index for perceptual image quality assessment”. *IEEE Trans. Image Process.* (Oct. 2014), vol. 23(10): pp. 4270–4281 (cit. on pp. [17–19](#)).
174. ZHANG, L., ZHANG, L., MOU, X., and ZHANG, D.: “FSIM: a feature similarity index for image quality assessment”. *IEEE Trans. Image Process.* (Aug. 2011), vol. 20(8): pp. 2378–2386 (cit. on pp. [5](#), [18](#), [19](#), [21](#)).
175. ZHANG, P., ZHOU, W., WU, L., and LI, H.: “SOM: semantic obviousness metric for image quality assessment”. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), vol.: pp. 2394–2402 (cit. on p. [18](#)).
176. ZHOU, L., RZESZOTARSKI, M. S., SINGERMAN, L. J., and CHOKREFF, J. M.: “The detection and quantification of retinopathy using digital angiograms”. *IEEE Trans. Med. Imag.* (Dec. 1994), vol. 13(4): pp. 619–626 (cit. on p. [130](#)).

List of Figures

1.1	Basic pipeline for applied harmonic analysis-based image analysis	4
2.1	Haar wavelets and the associated scaling function	15
2.2	Main components of the HaarPSI	23
2.3	Examples for HaarPSI similarity and weight maps	24
2.4	HaarPSI parameter optimization	25
2.5	HaarPSI scatter plots	30
2.6	Influence of parameters C and α on distortion-specific correlations	33
3.1	Examples of mother wavelets, scaling functions and their associated discrete filters	41
3.2	Wavelet decomposition of a digital image	45
3.3	Comparison of digital 2D wavelet and shearlet filters	48
3.4	Frequency domain partitions induced by wavelet- and shearlet-based systems	50
3.5	Non-subsampled shearlet decomposition of a digital image	51
3.6	Visualizations of the cost functions associated with wavelet- and shearlet-based visual servoing control laws	64
3.7	Overview of the three different camera displacements considered in our numerical simulations	65
3.8	Three different conditions for simulated experiments	66
3.9	Overview of the two different camera displacements considered in the experimental evaluation of the wavelet- and shearlet-based control laws	70
3.10	Wavelet- and shearlet-based visual servoing in nominal conditions with a planar scene (<i>experiment 1</i>)	71
3.11	Wavelet- and shearlet-based visual servoing with a planar scene that is partially occluded by an object (<i>experiment 1</i>)	72
3.12	Wavelet- and shearlet-based visual servoing with a planar scene and unstable lighting conditions (<i>experiment 1</i>)	73
4.1	Examples of the α -molecules weight functions	80
4.2	First and the second derivative of the Gaussian and their respective Hilbert transforms	91

4.3	Example of a refined maximum point estimation based on discrete samples	93
4.4	Behavior of even- and odd-symmetric coefficients at the location of ideal features.	95
4.5	Detection of one-dimensional edges	101
4.6	Examples of L^1 -normalized 2D symmetric generators and ideal 2D features	103
4.7	Behavior of even-symmetric coefficients at the location of an ideal ridge . .	108
4.8	Digital symmetric molecules filters	114
4.9	Detection of edges and local tangent orientations in a synthetic image using SymFD	118
4.10	Visual comparison of the detection results yielded by different edge detection methods	120
4.11	Detection of ridges, local tangent orientations, and ridge widths in a synthetic image via SymFD	121
4.12	Visual comparison of the detection results yielded by different ridge detection methods	123
4.13	Detection of blobs and blob widths in a synthetic image using SymFD . . .	124
4.14	Visual comparison of SymFD and the circular Hough transform in the case of blob detection	126
4.15	Examples of digital images in the REVIEW database	127
4.16	Width and orientation measurements of retinal blood vessels yielded by SymFD	128
4.17	Two pictures showing grown cell colonies in a Petri dish	131
4.18	SymFD-based detection of cell colonies in digital photographs of Petri dishes	132
4.19	Semilog graphs and trendlines for four automated cell colony counting methods	134
4.20	Detection and characterization of edges in a noisy mock image with CoShREM	135
4.21	Detection and analysis of ridges in a noisy mock image with CoShREM . .	136
4.22	Detection and geometric characterization of flame fronts with CoShREM . .	140
B.1	Example of a wavelet-based feature vector \mathbf{s}_w	171
B.2	Example of shearlet-based feature vector \mathbf{s}_{sh}	171
B.3	Example of non-subsampled wavelet-based feature vector \mathbf{s}_w	172
B.4	Example of non-subsampled shearlet-based feature vector \mathbf{s}_{sh}	172
C.1	Wavelet- and shearlet-based visual servoing in nominal conditions with a planar scene (<i>experiment 2</i>)	173
C.2	Wavelet- and shearlet-based visual servoing with a planar scene that is partially occluded by an object (<i>experiment 2</i>)	174

C.3	Wavelet- and shearlet-based visual servoing with a planar scene and unstable lighting conditions (<i>experiment 2</i>)	175
D.1	The synthetic images used for the evaluation of different edge detection methods.	176
D.2	The synthetic images used for the evaluation of different ridge detection methods.	177
D.3	The synthetic images used for the evaluation of different blob detection methods.	178
E.1	Visual comparison of different edge detection algorithms	179
E.2	Visual comparison of tangent direction and curvature estimates obtained from CoShREM and two other methods	180
E.3	Visual comparison of the robustness of CoShREM and the phase congruency-based ridge detector	181
E.4	Visual comparison of tangent direction and curvature estimates obtained from CoShREM and the phase congruency measure	182

List of Tables

1.1	List of commonly used symbols	7
2.1	SROCC of IQA metrics on benchmark databases	28
2.2	Mean SROCC and average execution time	28
2.3	Distortion-specific SROCC for different IQA metrics	29
2.4	SROCC on benchmark databases for different choices of wavelet filters . . .	34
3.1	Overview of the control laws considered in the numerical and experimental validation	63
3.2	Final camera pose displacements yielded by the five considered control laws in simulated positioning tasks in nominal conditions	67
3.3	Final camera pose displacements yielded by the five considered control laws in simulated positioning tasks impeded by partial occlusions	68
3.4	Final camera pose displacements yielded by the five considered control laws in simulated positioning tasks impeded by unstable lighting conditions . . .	69
4.1	Overview of algorithms considered in the evaluation with synthetic images .	116
4.2	Evaluation of five edge detection methods on synthetic images	119
4.3	Evaluation of four ridge detection methods on synthetic images	122
4.4	Evaluation of two blob detection methods on synthetic images	125
4.5	Evaluation of SymFD orientation measurements on the REVIEW database	129
4.6	Evaluation of width measurements obtained by nine different algorithms on the REVIEW Database	130
4.7	Evaluation of the accuracy of automatically obtained cell colony counts . .	133
4.8	Numerical comparison of CoShREM with five other edge detectors	137
4.9	Numerical comparison of the stability under additional Poisson noise	138
4.10	Numerical comparison of CoShREM and the phase congruency-based ridge detector	139
A.1	Pearson correlations for IQA metrics on benchmark databases	169
A.2	Distortion-specific Pearson correlations for different IQA metrics	170

A Pearson Correlations for Different Image Quality Assessment Metrics

Table A.1: Pearson correlations for IQA metrics on benchmark databases.

	PSNR	VIF	SSIM	MS-SSIM	GSM	MAD	SRSIM	FSIM	VSI	HaarPSI
LIVE	0.8585	0.9411	0.8290	0.7670	0.7799	0.9559	0.7758	0.8595	0.7647	0.9592
TID 2008	0.5190	0.7769	0.7401	0.7897	0.7779	0.8290	0.8242	0.8341	0.8107	0.9032
TID 2013	0.4785	0.7335	0.7596	0.7773	0.7966	0.8074	0.7984	0.8322	0.8373	0.8904
CSIQ	0.7512	0.9219	0.7916	0.7720	0.7471	0.9500	0.7520	0.8208	0.8392	0.9463

Lower and higher correlations than HaarPSI (statistically significant with $p < 0.05$).
The highest correlation in each row is written in **boldface**.
All correlations obtained **without nonlinear regression**.

Table A.2: Distortion-specific Pearson correlations for different IQA metrics.

	PSNR	VIF	SSIM	MS-SSIM	GSM	MAD	SR-SIM	FSIM	VSI	HaarPSI
<i>LIVE</i>										
jpg2k	0.8747	0.9476	0.8925	0.8697	0.8564	0.9725	0.8800	0.9036	0.8662	0.9673
jpg	0.8650	0.9600	0.9279	0.9184	0.9131	0.9742	0.9028	0.9117	0.9037	0.9779
gwn	0.9792	0.9632	0.9583	0.9181	0.8904	0.9764	0.8684	0.9263	0.9171	0.9791
gblur	0.7744	0.9575	0.8881	0.8450	0.8565	0.9486	0.8411	0.9086	0.8544	0.9576
ff	0.8753	0.9560	0.8619	0.8113	0.7925	0.9461	0.7837	0.8515	0.8151	0.9444
<i>TID 2008</i>										
gwn	0.9336	0.8657	0.7494	0.7433	0.8078	0.8165	0.8284	0.8076	0.8719	0.9029
gwnc	0.9208	0.8928	0.7758	0.7772	0.7833	0.8267	0.8625	0.8671	0.9045	0.9131
scn	0.9526	0.8578	0.7678	0.7583	0.8422	0.8598	0.8492	0.8217	0.8862	0.9283
mn	0.8627	0.8900	0.7496	0.7849	0.5512	0.7566	0.7345	0.8106	0.6114	0.7480
hfn	0.9680	0.9441	0.8228	0.8176	0.8452	0.8931	0.8657	0.8597	0.8934	0.9393
in	0.8566	0.8146	0.6202	0.6220	0.6218	0.0417	0.6912	0.7044	0.7651	0.8077
qn	0.8729	0.7442	0.7239	0.7602	0.8090	0.7981	0.7586	0.7986	0.8077	0.8602
gblr	0.8439	0.9388	0.8936	0.8745	0.8761	0.9227	0.9078	0.9078	0.8731	0.8934
den	0.9428	0.8968	0.9208	0.9156	0.9052	0.9612	0.9133	0.9344	0.9162	0.9739
jpg	0.8597	0.9327	0.9319	0.9279	0.9546	0.9487	0.9444	0.9299	0.9566	0.9647
jpg2k	0.8629	0.9169	0.9492	0.9365	0.9564	0.9733	0.8965	0.9566	0.9632	0.9856
jpgt	0.6258	0.8720	0.8375	0.8150	0.8441	0.8556	0.8573	0.8446	0.8705	0.8882
jpg2kt	0.8528	0.8307	0.8252	0.7970	0.7958	0.8295	0.7932	0.7883	0.8142	0.8688
pn	0.5831	0.7366	0.6685	0.6637	0.7013	0.8242	0.7381	0.7297	0.7314	0.7936
bdist	0.6277	0.8340	0.8659	0.7861	0.8822	0.8007	0.7864	0.8410	0.6198	0.8069
ms	0.6845	0.5896	0.6834	0.6735	0.7431	0.5709	0.6098	0.6700	0.6420	0.5358
ctrst	0.5819	0.8816	0.5158	0.7686	0.7068	0.2573	0.6978	0.7275	0.6995	0.6446
<i>TID 2013</i>										
gwn	0.9519	0.9010	0.7954	0.7891	0.8500	0.8732	0.8569	0.8435	0.8928	0.9248
gwnc	0.8948	0.8641	0.7615	0.7629	0.8216	0.8297	0.8603	0.8543	0.8975	0.8998
scn	0.9513	0.8783	0.7840	0.7681	0.8420	0.8804	0.8371	0.8240	0.8714	0.9261
mn	0.8447	0.8772	0.7569	0.7929	0.5934	0.7804	0.7615	0.8214	0.6585	0.7737
hfn	0.9607	0.9454	0.8342	0.8307	0.8575	0.9098	0.8702	0.8669	0.8939	0.9415
in	0.8856	0.8489	0.6625	0.6541	0.6602	0.2741	0.7183	0.7216	0.7776	0.8325
qn	0.8855	0.7805	0.7514	0.7752	0.8199	0.8365	0.7677	0.8096	0.8119	0.8643
gblr	0.8952	0.9530	0.8832	0.8616	0.8565	0.9336	0.8893	0.8922	0.8548	0.9030
den	0.9572	0.8914	0.9199	0.9110	0.9116	0.9602	0.9114	0.9304	0.9187	0.9690
jpg	0.8972	0.9332	0.9278	0.9207	0.9470	0.9510	0.9343	0.9242	0.9479	0.9750
jpg2k	0.9078	0.9184	0.9424	0.9183	0.9462	0.9663	0.8772	0.9360	0.9494	0.9787
jpgt	0.6410	0.9000	0.8721	0.8476	0.8697	0.8537	0.8772	0.8761	0.8972	0.9177
jpg2kt	0.8834	0.8692	0.8260	0.7929	0.7960	0.8648	0.7914	0.8010	0.8179	0.8913
pn	0.6702	0.7686	0.7481	0.7376	0.7718	0.8513	0.8034	0.7957	0.7971	0.8376
bdist	0.1448	0.5027	0.5589	0.4608	0.5939	0.3184	0.4436	0.5237	0.1356	0.4441
ms	0.7482	0.6829	0.7309	0.6823	0.8153	0.6654	0.6364	0.7103	0.7367	0.6365
ctrst	0.4812	0.8730	0.4941	0.7268	0.6701	0.2601	0.6520	0.6838	0.6595	0.5916
ccs	0.1378	0.3404	0.4349	0.4237	0.3739	0.0351	0.2491	0.6069	0.6852	0.6003
mgn	0.9187	0.8559	0.7358	0.7301	0.7903	0.8422	0.8049	0.8008	0.8505	0.8786
cn	0.8548	0.8992	0.8459	0.8105	0.9286	0.9280	0.9260	0.9214	0.9301	0.9571
lcni	0.9372	0.9034	0.9058	0.8917	0.9472	0.9520	0.9439	0.9364	0.9463	0.9686
icqd	0.9227	0.8582	0.8083	0.7767	0.8240	0.8626	0.7574	0.8053	0.8083	0.8826
cha	0.8569	0.9441	0.9519	0.9071	0.9563	0.9560	0.8819	0.9478	0.9498	0.9549
ssr	0.9167	0.9067	0.9528	0.9197	0.9601	0.9658	0.9135	0.9412	0.9449	0.9791
<i>CSIQ</i>										
gwn	0.9437	0.9590	0.8043	0.8254	0.8517	0.9486	0.8669	0.7959	0.8875	0.9433
jpeg	0.7898	0.9590	0.9165	0.9064	0.8964	0.9696	0.8731	0.9077	0.8833	0.9780
jpg2k	0.9270	0.9360	0.8967	0.8843	0.8793	0.9808	0.8428	0.9106	0.9008	0.9853
gpn	0.9527	0.9552	0.7844	0.7790	0.8293	0.9548	0.7777	0.8160	0.8698	0.9470
gblr	0.9081	0.9627	0.8692	0.8670	0.8575	0.9713	0.8675	0.8843	0.8761	0.9623
ctrst	0.8888	0.9294	0.7666	0.9003	0.8656	0.9306	0.8878	0.8765	0.8686	0.9229

Lower and higher correlations than HaarPSI (statistically significant with $p < 0.05$).

The highest correlation in each row is written in **boldface**.

All correlations obtained **without nonlinear regression**.

B Examples of Visual Feature Vectors Used for Visual Servoing

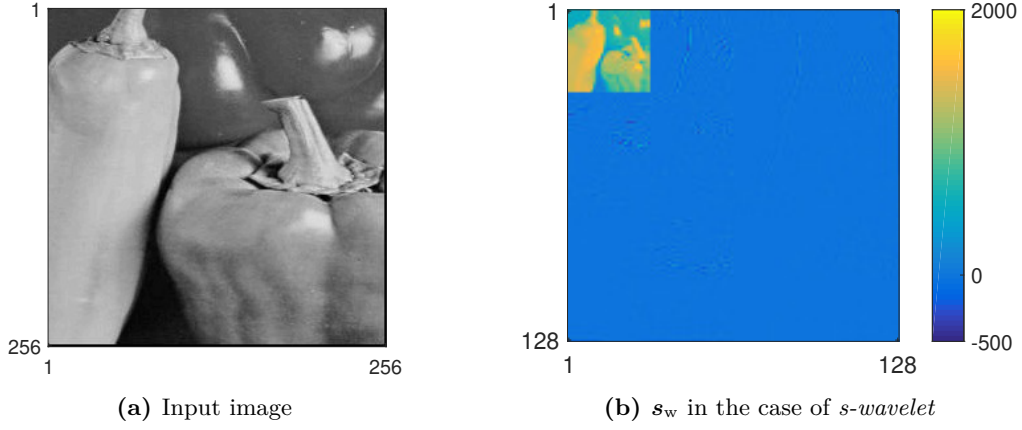


Figure B.1: The feature vector s_w was obtained by performing a discrete wavelet transform at level 3. The upper left corner in Figure B.1(b) contains a coarse approximation of the original image of size 32×32 . The remaining pixels correspond to detail coefficients at the second and third stage of decomposition. Note that the high-frequency detail coefficients from the first stage of decomposition were omitted.

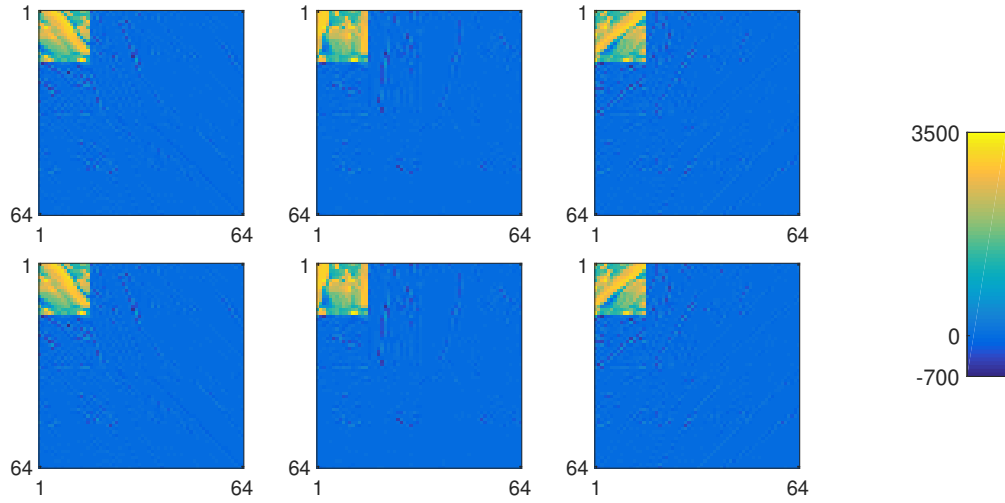


Figure B.2: s_{sh} in case of s -shearlet. Concatenating the six displayed coefficient matrices yields the feature vector s_{sh} . The matrices were obtained by performing a discrete subsampled shearlet transform at level 4. Each upper left corner contains a coarse approximation of the original image of size 16×16 along one of the six considered directions. The remaining pixels in each matrix correspond to detail coefficients at the third and fourth stage of decomposition. Note that the high-frequency detail coefficients from the first and second stage of decomposition were omitted.

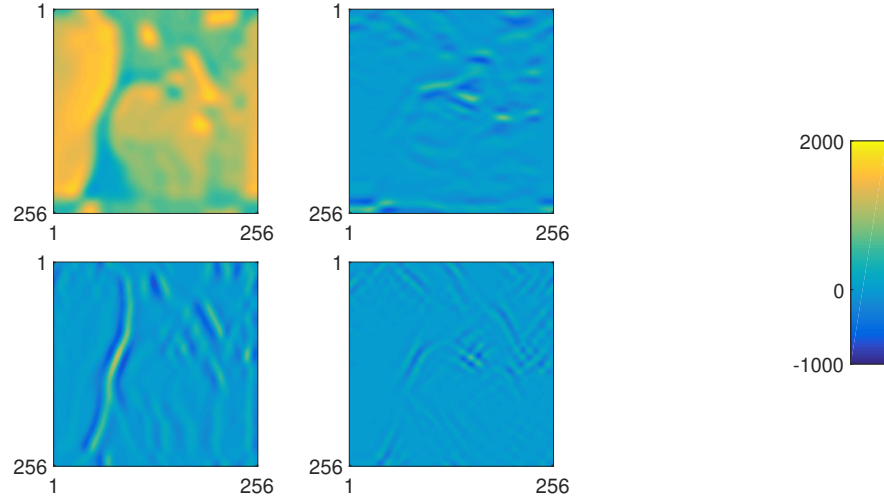


Figure B.3: s_w in case of ns -wavelet. Concatenating the four displayed coefficient matrices yields the feature vector s_w . The matrices were obtained by performing a non-subsampled wavelet transform at level 4. The first coefficient matrix defines a coarse approximation of the original image. The three remaining matrices show detail coefficients at the fourth level of decomposition. Similar to the s -wavelet case, the high-frequency coefficients yielded by the first, second and third stage of decomposition were omitted. Note that all matrices are of the same size as the original image.

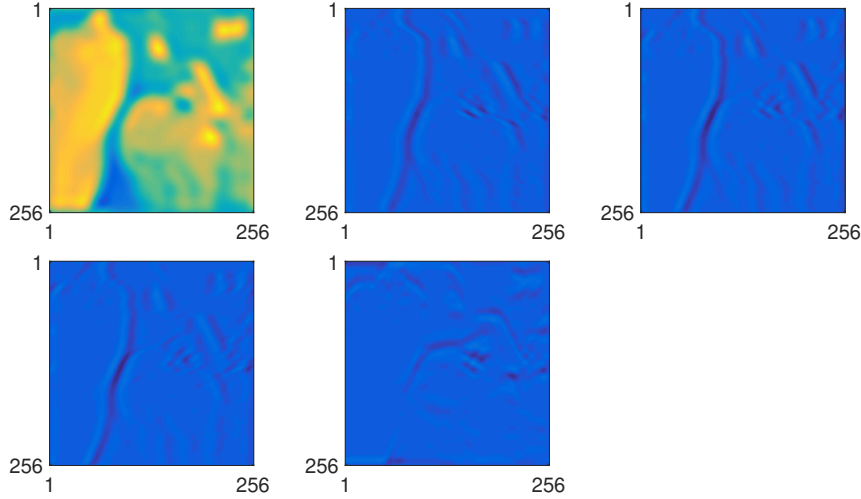


Figure B.4: s_{sh} in case of ns -shearlet. Concatenating the five displayed coefficient matrices yields the feature vector s_{sh} . The matrices were obtained by performing a non-subsampled shearlet transform at level 4. The first coefficient matrix defines a coarse approximation of the original image. Each of the other matrices was obtained by convolving the input image with a discrete shearlet filter associated with a different orientation. Note that the high-frequency shearlet coefficients from the first, second and third stage of decomposition were omitted and that each matrix has the same size as the original image.

C Additional Visual Servoing Experiments

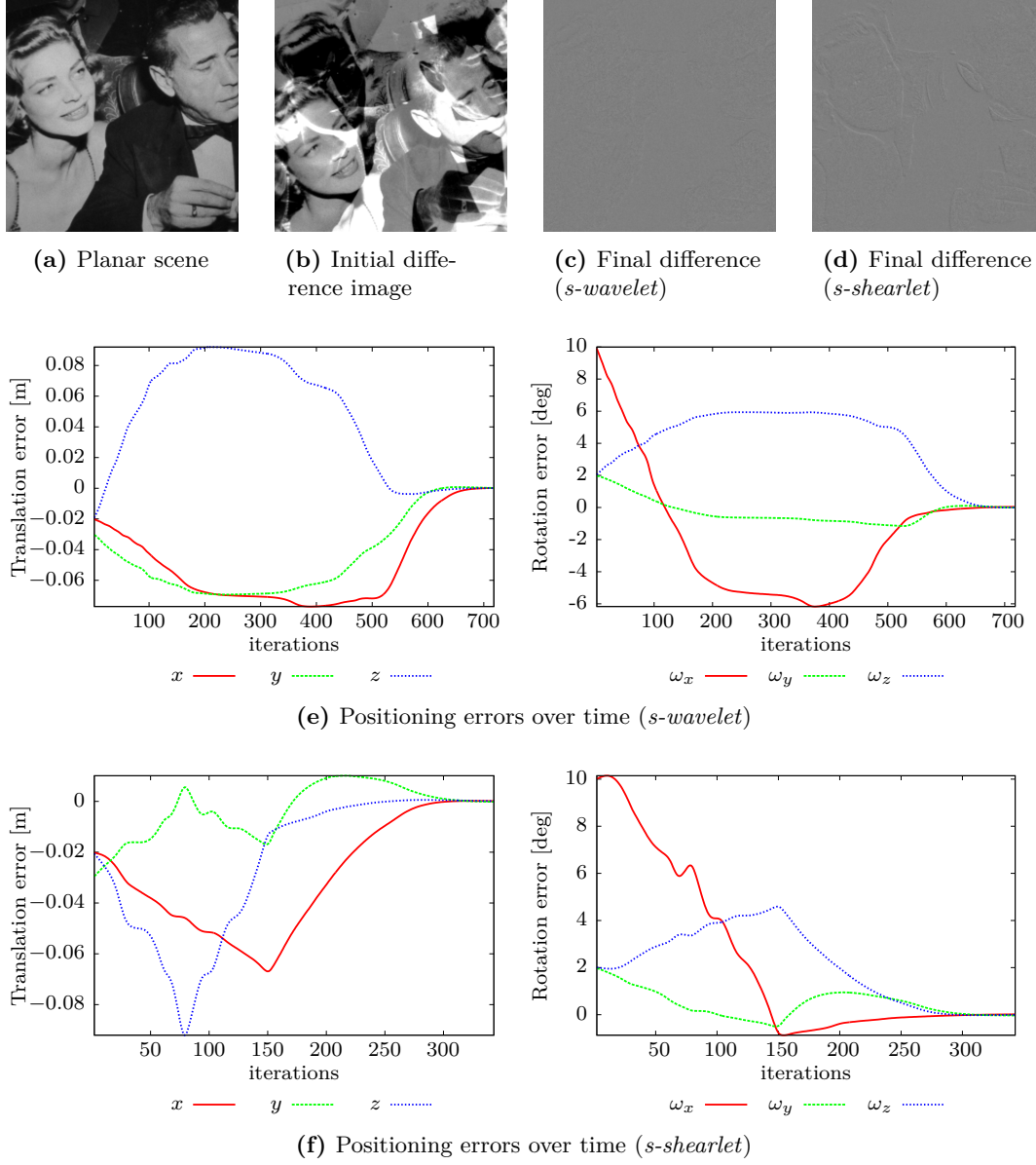


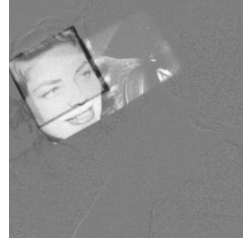
Figure C.1: Wavelet- and shearlet-based visual servoing in nominal conditions with a planar scene. The initial displacement was chosen according to *experiment 2* (cf. Table 3.9(b)).



(a) Planar scene with a partial occlusion



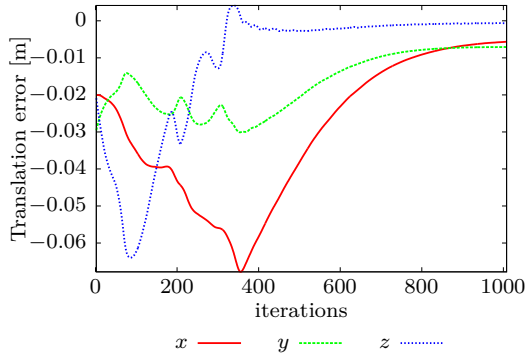
(b) Initial difference image



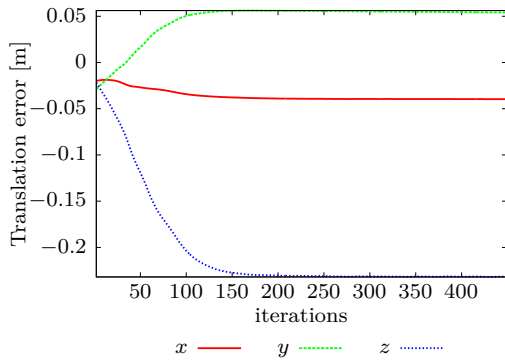
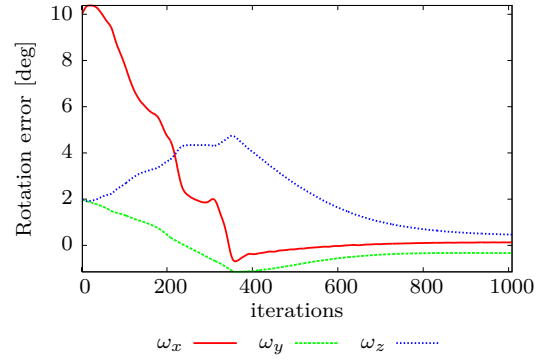
(c) Final difference (*s-wavelet*)



(d) Final difference (*s-shearlet*)



(e) Positioning errors over time (*s-wavelet*)



(f) Positioning errors over time (*s-shearlet*)

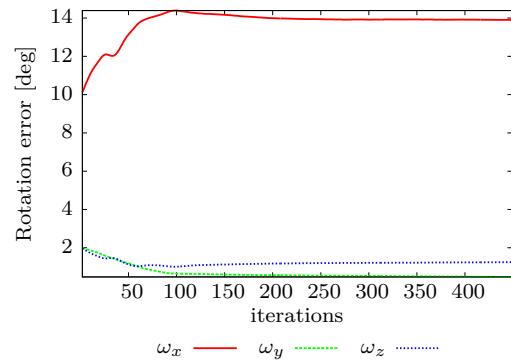


Figure C.2: Wavelet- and shearlet-based visual servoing with a planar scene that is partially occluded by an object. The initial displacement was chosen according to *experiment 2* (cf. Table 3.9(b)).



(a) Planar scene with additional light source



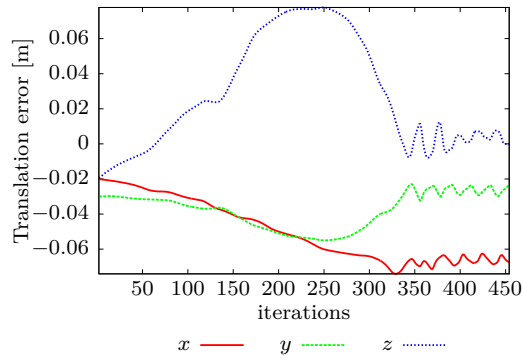
(b) Initial difference image



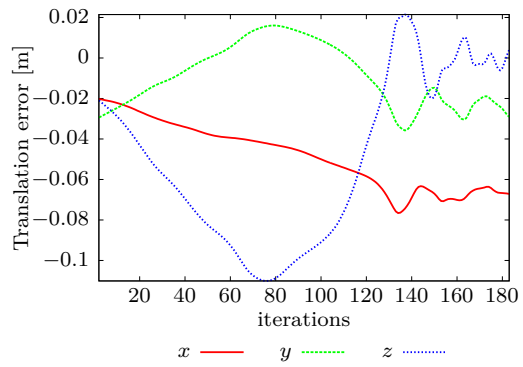
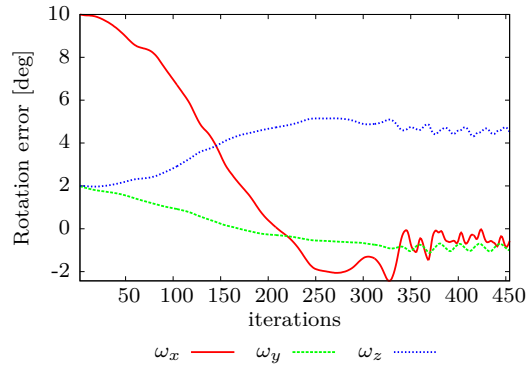
(c) Final difference (*s-wavelet*)



(d) Final difference (*s-shearlet*)



(e) Positioning errors over time (*s-wavelet*)



(f) Positioning errors over time (*s-shearlet*)

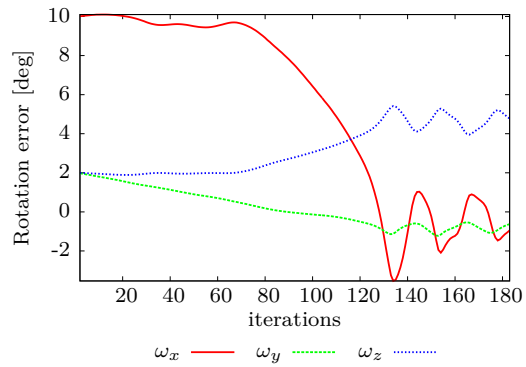


Figure C.3: Wavelet- and shearlet-based visual servoing with a planar scene and unstable lighting conditions. The initial displacement was chosen according to *experiment 2* (cf. Table 3.9(b)).

D Synthetic Images Used for the Evaluation of SymFD

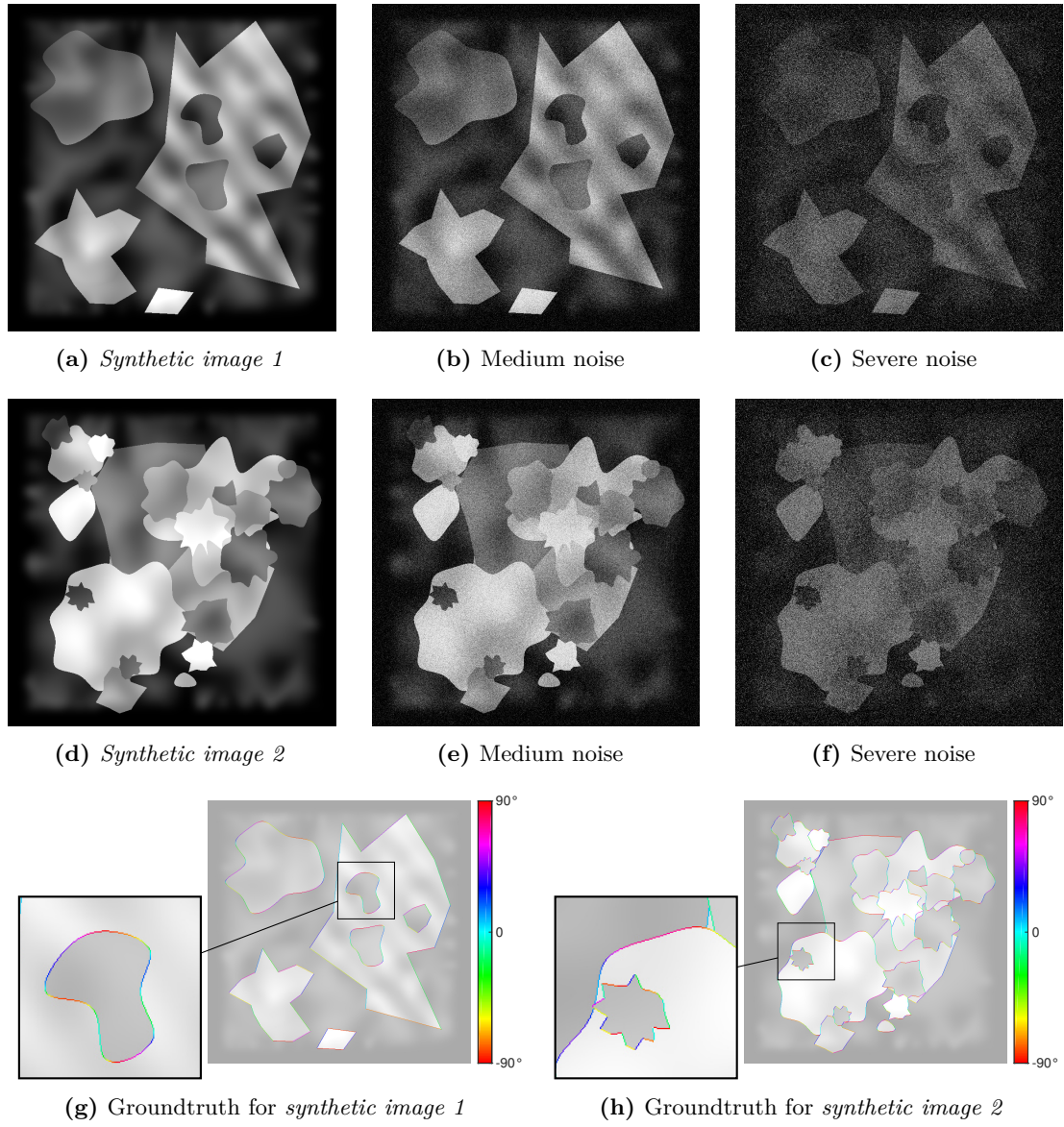


Figure D.1: The synthetic images used for the evaluation of different edge detection methods.

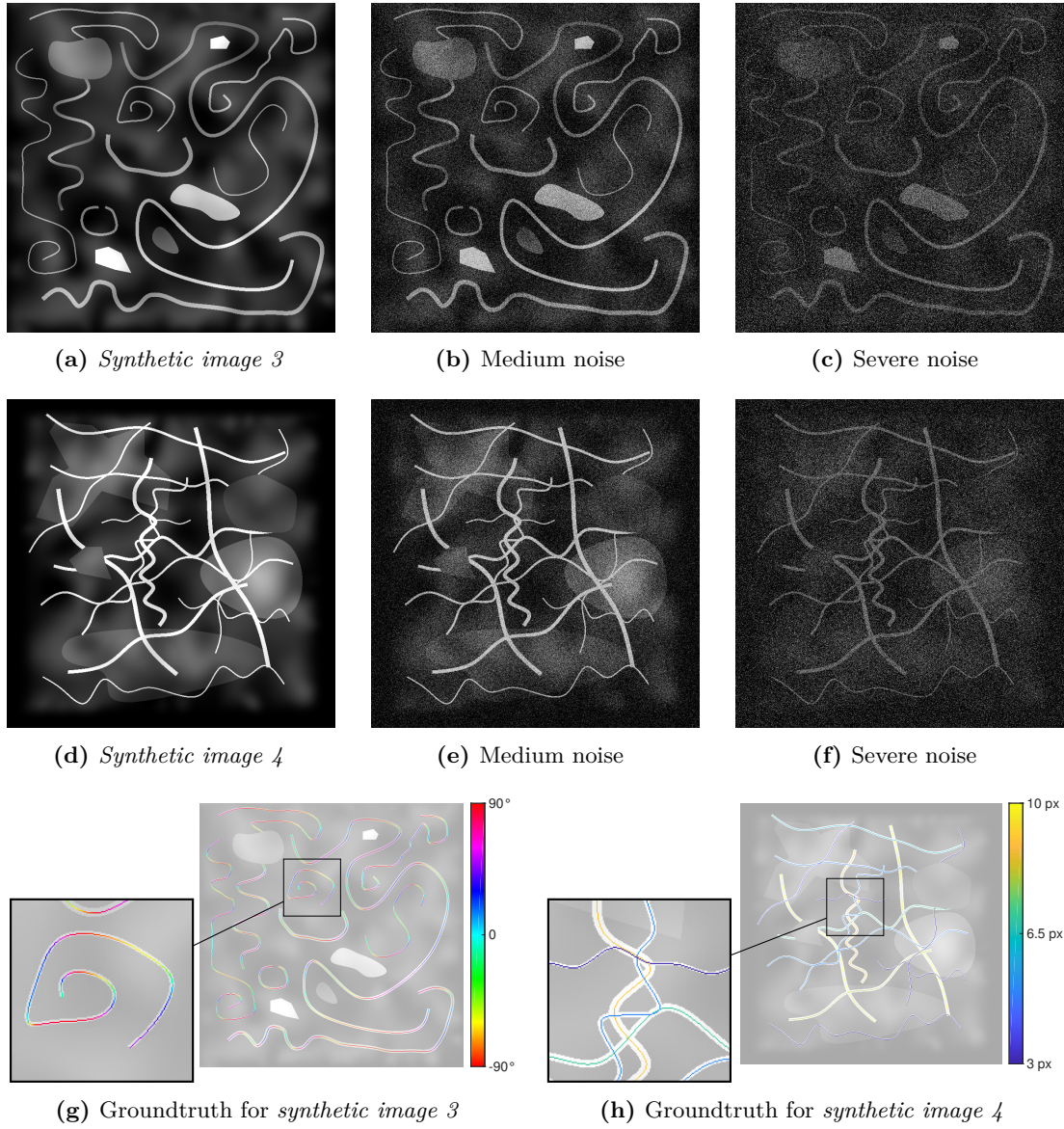


Figure D.2: The synthetic images used for the evaluation of different ridge detection methods.

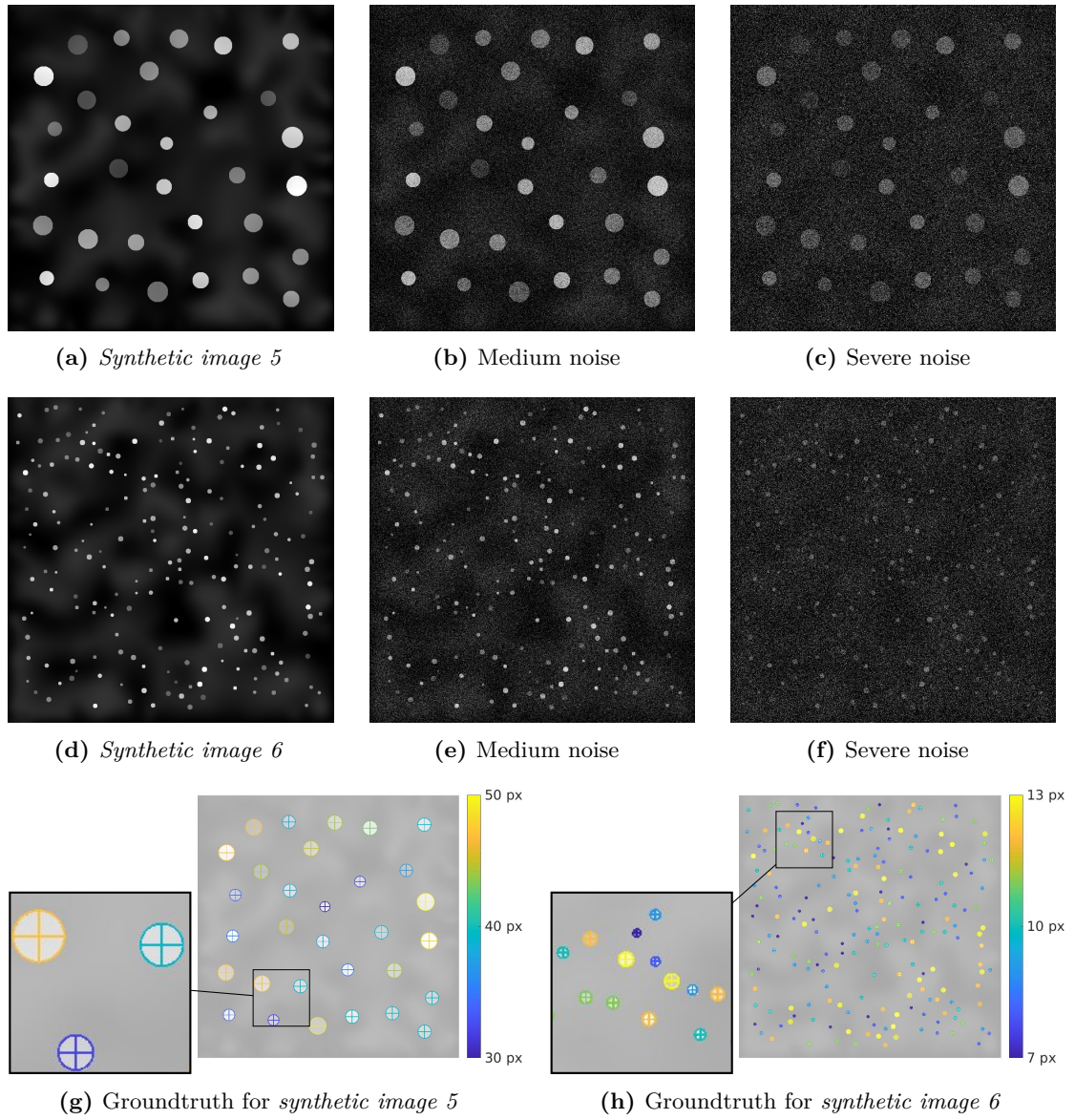


Figure D.3: The synthetic images used for the evaluation of different blob detection methods.

E Visual Comparison of Feature Detectors on Synthetic Flame Data

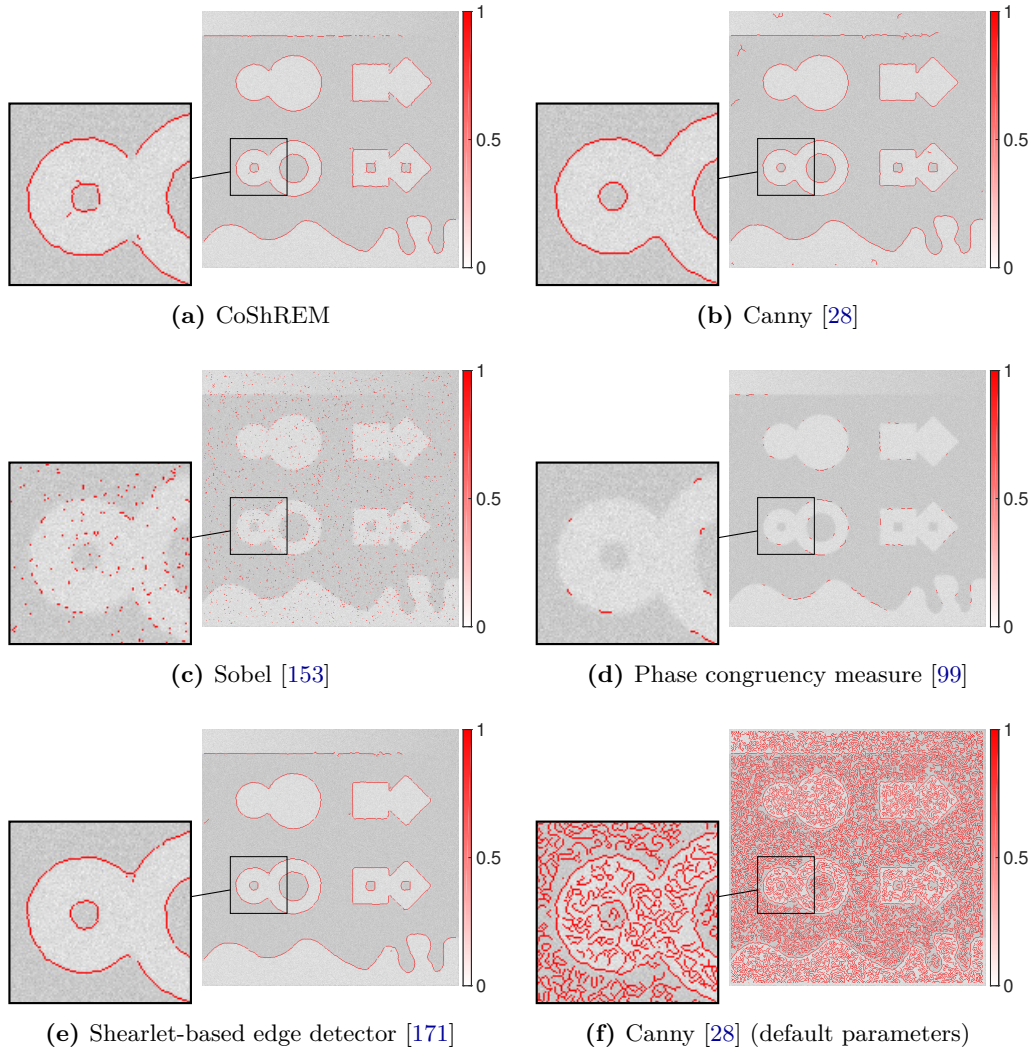


Figure E.1: Visual comparison of different edge detection algorithms. The processed image was perturbed with Gaussian blur ($\sigma_{\text{blur}} = 1.5$) and additive Gaussian white noise ($\sigma_{\text{noise}} = 100$). The PFOM values corresponding to the results shown here can be found Table 4.8.

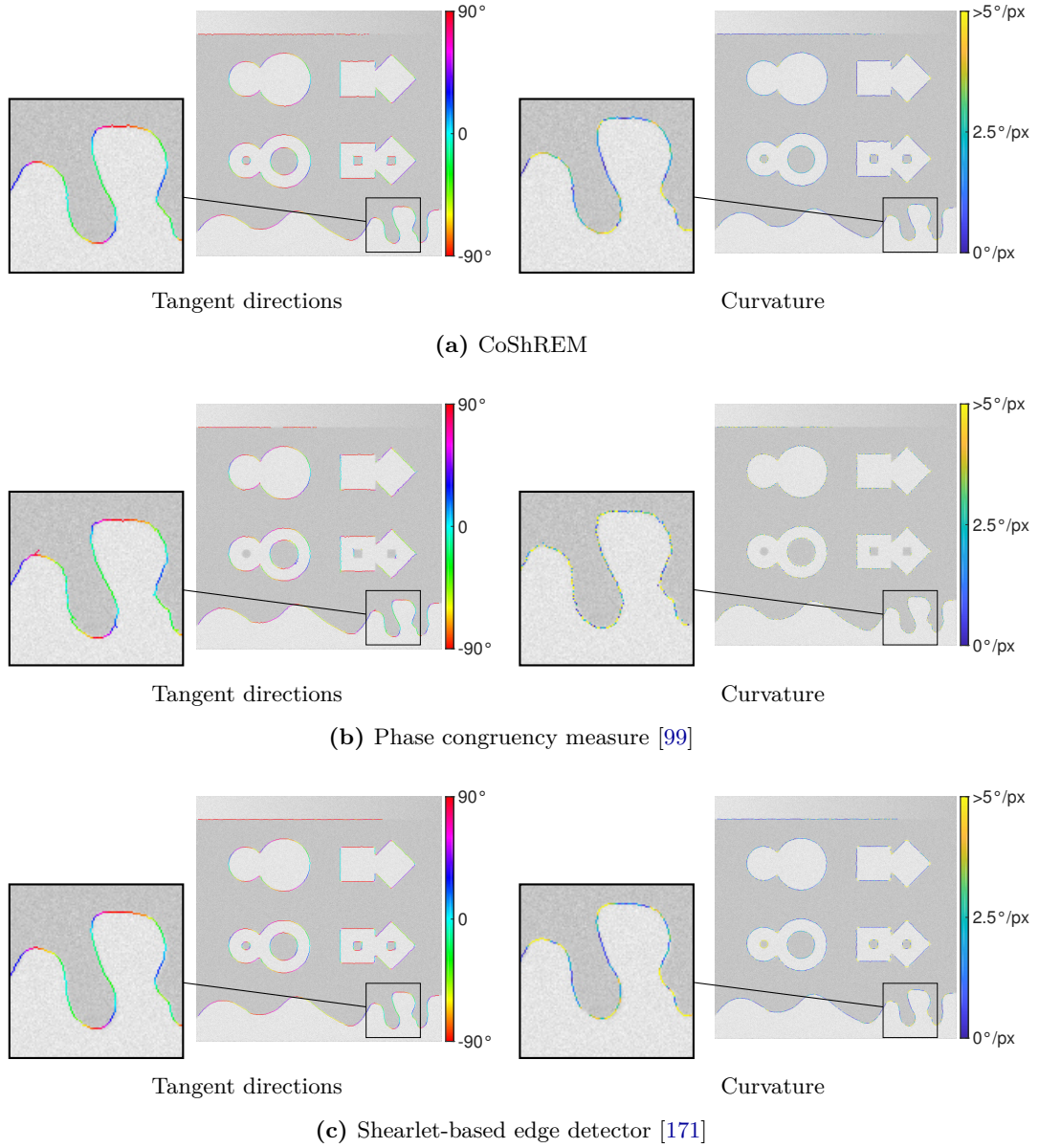


Figure E.2: Visual comparison of tangent direction and curvature estimates obtained from CoShREM, the phase congruency measure [99] and the shearlet-based edge detector [171]. The processed image was perturbed with Gaussian blur ($\sigma_{\text{blur}} = 1.0$) and additive Gaussian white noise ($\sigma_{\text{noise}} = 50$) (cf. Figure 4.20(a)). The PFOM values corresponding to the results shown here can be found in Table 4.8.

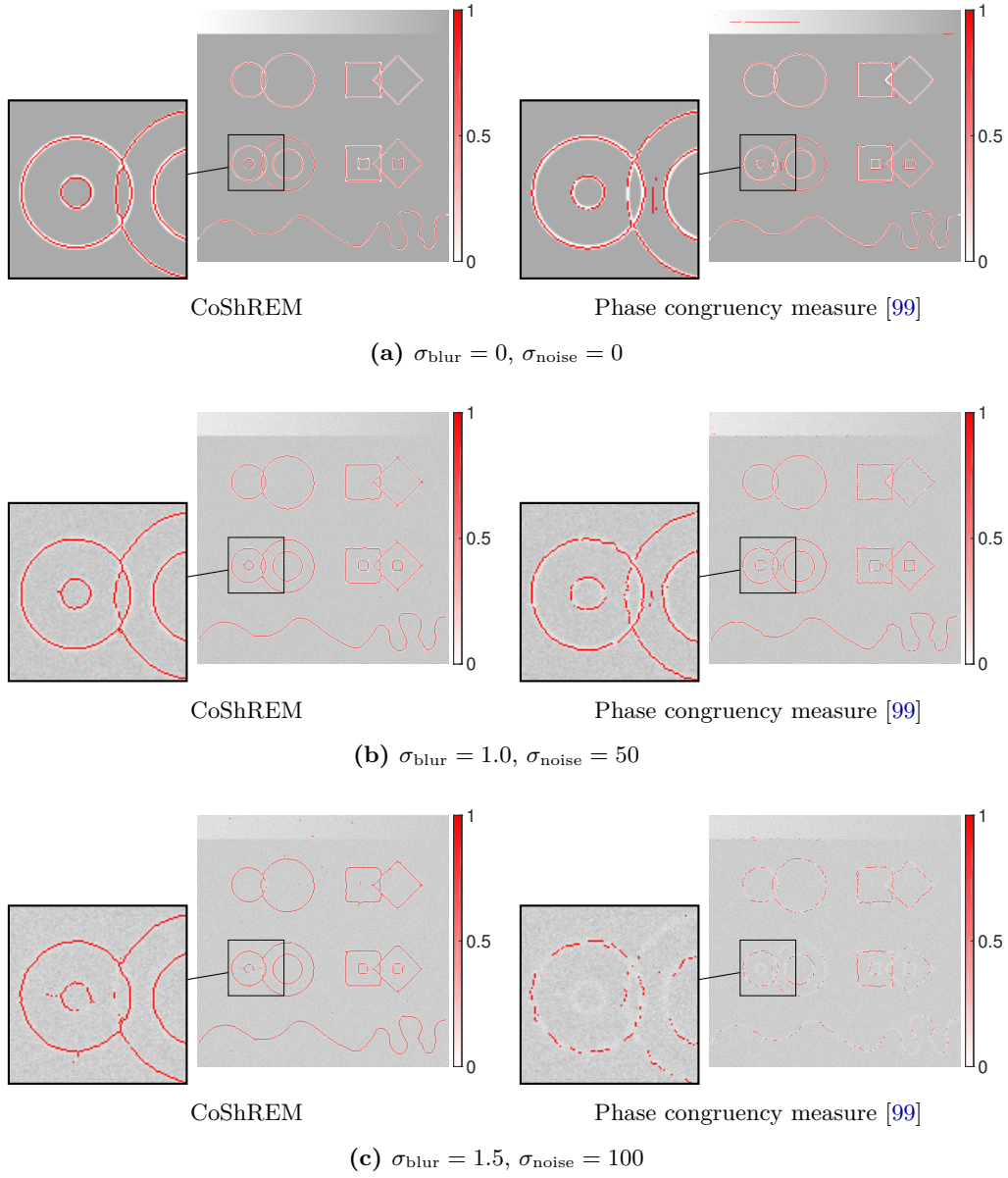


Figure E.3: Visual comparison of the robustness of CoShREM and the phase congruency-based ridge detector with respect to Gaussian blur and Gaussian noise. The PFOM values corresponding to the results shown here can be found in Table 4.10.

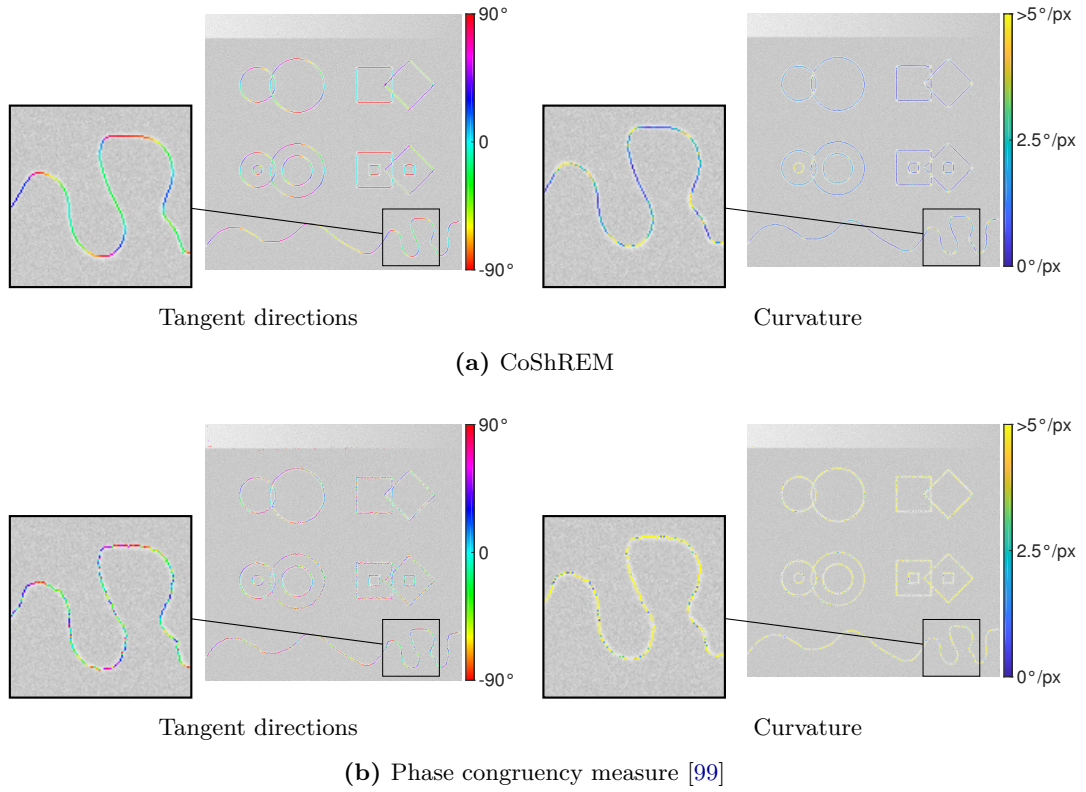


Figure E.4: Visual comparison of tangent direction and curvature estimates obtained from CoShREM and the phase congruency measure. The processed image is the same as in figure Figure 4.21(a) ($\sigma_{\text{blur}} = 1.0$, $\sigma_{\text{noise}} = 50$).

Acronyms

Notation	Description
1DG	one-dimensional Gaussian model
2DG	two-dimensional Gaussian model
ARIA	automated retinal image analyzer
CH	methylidyne radical
CLRIS	central light reflex image set
CoShREM	complex shearlet-based edge and ridge measure
CSIQ	Laboratory of Computational and Subjective Image Quality image quality database
DMOS	difference mean opinion score
DOF	degrees of freedom
ESP	extraction of segment profiles
ETOS	edge tracking based on orientation scores
FOM	figure of merit
FP	false positives
FR IQA	full reference image quality assessment
FSIM	feature similarity index
FSIMC	color-sensitive feature similarity index
GSM	gradient similarity measure
HaarPSI	Haar wavelet-based perceptual similarity index
HaarPSIC	color-sensitive Haar wavelet-based perceptual similarity index
HCO	formyl radical

Notation	Description
HHFW	half height full width
HRIS	high resolution image set
HVS	human visual system
IQA	image quality assessment
KPIS	kick point image set
LIVE	Laboratory for Image & Video Engineering image quality assessment database
LRS	laser Rayleigh scattering
MAD	most apparent distortion
MAE	mean absolute error
MOS	mean opinion score
MRA	multiresolution analysis
MS-SSIM	multi-scale structural similarity index
MSE	mean squared error
NICE	NIST's integrated colony enumerator
NR IQA	no reference image quality assessment
OH	hydroxyl radical
ONB	orthonormal basis
PFOM	Pratt's figure of merit
PIV	particle imaging velocimetry
PLIF	planar laser-induced fluorescence
PSNR	peak signal-to-noise ratio

Notation	Description
RR IQA	reduced reference image quality assessment
SD	standard deviation
SR	success rate
SR-SIM	spectral residual-based similarity
SROCC	Spearman's rank-order correlation coefficient
SSIM	structural similarity index
SymFD	symmetric molecule-based feature detector
TID	Tampere image database
TP	true positives
TPR	true positive rate
VDIS	vascular disease image set
VIF	visual information fidelity
ViSP	visual Servoing Platform
VSI	visual saliency-induced index