

# Basis functions for early vision

Markus Michaelis

GSF-Medis, Neuherberg, D-85764 Oberschleißheim, Germany

Gerald Sommer

Institut für Informatik, Christian-Albrechts-Universität, Preußerstr.1-9  
D-24105 Kiel, Germany

Neuherberg, August 20, 1994

**Institut für Informatik und Praktische Mathematik  
Christian-Albrechts-Universität Kiel**

**Bericht Nr. 9413**

**Abstract:** It is commonly agreed on that the first step in early vision consists of projections of the image to a set of basis functions. Usually the spatial distribution of the basis functions is homogeneous and the projection is a convolution but in general this will not be the case. In the literature there is a wealth of different basis functions, each of them optimal with respect to certain criteria. On the other hand, there seems to be a convergence towards derivatives of Gaussians or harmonic modulations of Gaussians (Gabor functions). In this report we discuss the principles and analysing methods underlying the choice of these functions. One of these methods that recently became of exceptional importance is the energy/phase representation. We investigate in detail the quality of successive orders of derivatives of Gaussians as odd/even pairs for the energy/phase concept. In addition we work out to which extent derivatives of Gaussians can be approximated by Gabor functions.

---

<sup>1</sup>Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft unter den Aktenzeichen So 320/1-1 und Ei 322/1-1 gefördert.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Basic principles in early vision</b>	<b>2</b>
2.1	Uncertainty principle . . . . .	2
2.1.1	Basic form . . . . .	2
2.1.2	Some extensions . . . . .	3
2.1.3	Information theoretical point of view . . . . .	3
2.1.4	What is where? . . . . .	6
2.2	Hilbert transform . . . . .	8
2.3	Scale-space . . . . .	10
2.4	Regularization and noise . . . . .	11
2.5	Discrete case . . . . .	11
2.6	Wavelets . . . . .	11
2.7	Biological vision systems . . . . .	12
2.8	Applications . . . . .	12
<b>3</b>	<b>Basis functions</b>	<b>13</b>
3.1	Gabor functions . . . . .	13
3.2	Derivatives of Gaussians . . . . .	13
3.3	Hermite functions . . . . .	13
3.4	Other functions . . . . .	14
<b>4</b>	<b>Derivatives of Gaussians</b>	<b>15</b>
4.1	Definition . . . . .	15
4.2	Fourier transform . . . . .	16
4.3	Normalization . . . . .	16
<b>5</b>	<b>Optimal odd/even pairs with Gaussian derivatives</b>	<b>18</b>
5.1	Same $L^2$ or $L^1$ norm . . . . .	19
5.2	Minimal Fourier energy at negative frequencies . . . . .	19
5.3	Monomodal energy . . . . .	21
5.4	Fit of the energy to a Gaussian . . . . .	23
5.5	Linear phase . . . . .	25
5.6	Conclusion . . . . .	27
<b>6</b>	<b>Relationship between derivatives of Gaussians and Gabor functions</b>	<b>28</b>
6.1	Definition and normalization of Gabor functions . . . . .	28
6.2	Asymptotic behavior of the derivatives of Gaussians . . . . .	29
6.3	Fit of Gabor functions to derivatives of Gaussians . . . . .	30
6.4	Conclusion . . . . .	33
<b>7</b>	<b>Summary</b>	<b>33</b>
	<b>Appendix A: Scale-space and the diffusion equation</b>	<b>34</b>
	<b>Appendix B: Hermite transform</b>	<b>35</b>
	<b>Appendix C: Integrals of the Gaussian</b>	<b>36</b>
	<b>Appendix D: Hermite polynomials</b>	<b>36</b>

# 1 Introduction

In many computational early vision methods the first processing step can be modeled as the projection of the image to a set of basis functions and it is believed that biological visual systems can be modeled in the same manner. In technical systems this step usually is performed by convolution, i.e. by one function that is translated on a homogeneous grid. In general, this restriction is not desirable. Therefore, we use the term 'projection'. The term 'basis function' is chosen because the projection can be seen as a change of the representation of the image from a Dirac (pixel) basis to another basis in the sense of linear algebra. In our case the new basis can be overcomplete or even be incomplete. An overcomplete basis will be necessary in general if the representation has to be invariant with respect to a certain transformation (spatial shift, rotation, scaling). An incomplete basis can be necessary for complexity reasons or to get rid of irrelevant information by projecting the image to a subspace.

However, before answering questions about the optimal sampling of the parameters for translating, rotating, and scaling a function, the type of function (its shape) that is optimal for a given task must be found. There are many different tasks in computer and biological vision and even for one task (e.g. edge detection) there are many different optimization criteria, each of which results in a different optimal function. Accordingly, in the computer vision literature a vast number of 'optimal' functions can be found. On the other hand, the receptive field profiles of simple cells in the visual cortex are frequently matched to Gabor functions or derivatives of Gaussians in the neuroscience literature.

Motivated by biological vision systems and independently by theoretical insights there has been a growing interest in these functions in computer vision. It seems that they are optimal with respect to the most generic criteria, i.e. for systems that are not committed to special tasks or input signals. Whether this is the case or not is hard to decide because the answer depends on the following processing steps. But the design of complete vision systems that are optimal for general vision tasks is unknown. Nevertheless, there are basic principles that are highly distinguished by theoretical considerations as for example the uncertainty principle or the energy/phase concept.

The content of this report is (1) a presentation of the relevant literature, (2) a discussion of the principles underlying the design of basis functions, and (3) an investigation of the relevant basis functions. In this report we concentrate on 1D functions or on 1D cross sections respectively. Most of the literature is not discussed in detail. The purpose is not a complete review of the articles but to show the background of previous work that has to be considered to realize the importance of the mentioned functions in computer vision and the knowledge about them. In section 2 the most relevant principles underlying the design of basis functions are discussed. In some papers only one of these principles (uncertainty, scale-space, energy/phase etc.) is particularly emphasised and the basis functions have to be chosen accordingly. In contrast, our point of view is that **all the principles are important but there is not one special principle that has to be satisfied under all circumstances**. As a result of this discussion we point out the importance of derivatives of Gaussians and Gabor functions.

The problem is that for weakly oscillating Gabor functions and lower derivatives of Gaussians not all principles can be satisfied at the same time. To understand the limitations we investigate in detail some properties of the derivatives of Gaussians and Gabor functions. In particular we emphasize the question of odd/even (quadrature) pairs and the comparison of the different functions. Even though we deal with lower percentage effects, we think that it is worth the effort because of the exceptional importance of the investigated functions and principles. For weakly oscillating functions there are a num-

ber of associated functions that satisfy some of the required principles, e.g. true Hilbert transforms, DC-, and ramp-corrected versions of the Gabor functions.

## 2 Basic principles in early vision

In this section we discuss in some detail the uncertainty principle, the Hilbert transform which is closely related to the energy/phase concept, scale-space, regularization and noise, the discrete case, wavelets, and biological vision systems. Besides these items there are other points that have to be considered for the design of basis functions:

- Separability: The analysis in this report is essentially 1D. In 2D there are new degrees of freedom. The functions are usually cartesian or polar separable in position or frequency space [3, 34, 62, 35].
- For theoretical analysis, functions that have simple analytical expressions are preferable.
- A small spatial extent is preferable to save computation time. This restriction might not be valid if hardware is cheap as in biological vision systems.
- Another important question concerns the sampling scheme that is necessary for a given function. This is closely connected to steerability, a topic that we will investigate in a following report.
- A main criterion is the optimality for following processing steps. But these steps depend on the application. A step towards evaluating them with respect to the application is to investigate the responses to certain signals, e.g. single event, multiple event, and texture responses. In this report, however, we concentrate on the functions themselves.

### 2.1 Uncertainty principle

The uncertainty principle in its most cited form states that a function cannot be localized both in the spatial and in the frequency domain. The mathematical foundations for this principle are well understood. On the other hand, there are more general forms of the principle and the deeper meaning behind it and its relevance and consequences in signal and image processing are still vague. The uncertainty principle became more recognized in recent years because it is closely connected to the popular bivariate representations [65] and the question of appropriate basis functions.

#### 2.1.1 Basic form

The uncertainty principle first appeared in quantum mechanics in the 20th and was applied to the field of signal processing by Denis Gabor in his famous article in 1946 [25]. Its mathematical form is given by:

$$\Delta x \Delta \omega \geq 1/2 \quad (1)$$

with the variances (the metric) defined by:

$$(\Delta x)^2 = \int_{-\infty}^{\infty} (x - \bar{x})^2 f(x) f^*(x) dx \quad (\Delta \omega)^2 = \int_{-\infty}^{\infty} (\omega - \bar{\omega})^2 \hat{f}(\omega) \hat{f}^*(\omega) d\omega \quad (2)$$

$\hat{f}(\omega) = (2\pi)^{-1/2} \int f(x)e^{i\omega x} dx$  is the Fourier transform,  $\bar{x}$  and  $\bar{\omega}$  are the expectation values. The complex function that minimizes the inequality (1) is the Gabor function:

$$G_{\sigma,c}(x) = e^{-\frac{x^2}{2\sigma^2}} e^{ic\frac{x}{\sigma}} \quad (3)$$

The relevance of the uncertainty principle for vision was pointed out by Marcelja and Daugman [47, 17]. Daugman extended the uncertainty principle to two dimensions. The optimal functions in 2D are Gaussians with any aspect ratio and a complex harmonic modulation in any orientation. As in the 1D case the cross-sections parallel to the modulation are given by (3). In practice the modulation is always chosen parallel to one of the axes of the Gaussian. Depending on the parameter  $\sigma$ , the Gabor functions are better localized in position or frequency space but they always exhibit the best possible joint localization. For the majority of the papers that use Gabor functions the above is the justification for using them. But the subject has some more facets that we will discuss in the following.

### 2.1.2 Some extensions

The uncertainty principle is not restricted to position and spatial frequency. In quantum mechanics it is stated for all eigenbases of hermitian operators that do not commute. Another point is the definition of the variance as the 2nd moment in (2). In quantum mechanics this choice is motivated by the probability interpretation of the wave function but in general there are other metrics possible. Some of them might even be preferable in computer vision. Other metrics are discussed by Stork et al. [66]. This paper has been rightly criticized by Yang [76], Daugman [19], and Klein and Beutter [31] because it contains some errors concerning the metrics and the optimal functions for real instead of complex function spaces. Nevertheless, the criticism of Stork et al. of the unreflected use of the 2nd moment as the variance is justified. However, Klein and Beutter point out that for reasonable metrics the optimal functions will be more or less Gabor like shaped. The eigenfunctions of the harmonic oscillator (Hermite functions) that were erroneously suggested (and confused with the derivatives of Gaussians) by Stork et al. maximize rather than minimize the uncertainty.

### 2.1.3 Information theoretical point of view

The basic question in this subsection is: **How much information is contained in a function  $f$ ?** The answer depends of course on the way the information is coded in the function. We assume that the coding is **linear**, i.e. the information is coded in the coefficients  $a_k$  with respect to some basis functions  $g_k$  ( $\langle \cdot | \cdot \rangle$  is the scalar product):

$$f = \sum_k a_k g_k \quad (4)$$

$$a_k = \langle f | \tilde{g}_k \rangle \quad \text{with} \quad \langle g_k | \tilde{g}_l \rangle = \delta_{kl}$$

The  $\tilde{g}_k$  constitute the dual basis. This equation implies that a physically possible sensor exists that corresponds to  $\tilde{g}_k$ . The measurement of the coefficients should be independent of each other what means that the basis functions are more or less orthogonal. It is helpful to associate to  $f$  the abstract function  $|f\rangle$  as in quantum mechanics.  $f(x) = \langle x | f \rangle$  or  $f(\omega) = \langle \omega | f \rangle$  are just different representations of the same function with  $\langle x | x_0 \rangle = \delta_{x_0}$ ,  $\langle x | \omega_0 \rangle = e^{i\omega_0 x}$ ,  $\langle \omega | \omega_0 \rangle = \delta_{\omega_0}$ . Hence we leave the symbol  $f$  unchanged.

If the resolution for the amplitude of the coefficients is finite, the amount of information depends on the number of coefficients. **How many basis functions contribute to  $f$ ?** How big is the alphabet? For example the information content of one letter, say 'o',

depends on the alphabet. If it is a member of the latin alphabet it is about 5 to 7 Bit. If it is a 600 dpi PostScript bitmap it is several kBit. If in our example the basis consists only of  $f$  itself, just one coefficient is needed. On the other hand, if there are **no restrictions** to the sensors (basis functions) the number of basis functions clearly can be infinite. Hence, without fixing the restrictions of the sensors there is no sensible answer to the posed question.

From the theory of signal processing we know an answer. For a signal of length  $\Delta x = x_{\max} - x_{\min}$  and sensors with a bandwidth of  $\Delta\omega = \omega_{\max} - \omega_{\min}$  the number of coefficients that can be transmitted is  $2\Delta x\Delta\omega$ . We now want to derive a deeper understanding of this formula. The most obvious physical restriction is with respect to the position (respective time). The signal is only measurable in a period  $\Delta x$  and there will be a maximal spatial resolution  $\delta x$ . Hence, we could assume  $\Delta x/\delta x$  basis functions, each of which has e.g. a rectangular profile (Fig.1).

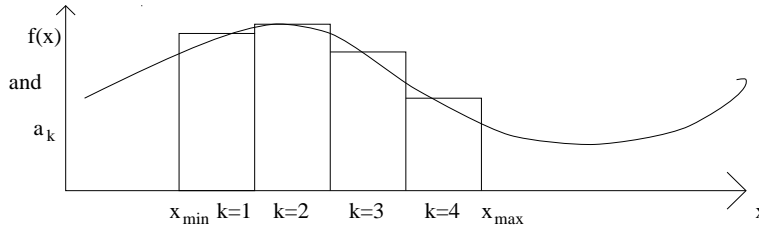


Figure 1: Rectangular basis functions

An example for such a sensor is a CCD-camera where the number of pixels gives the number of basis functions that sample the original signal (illuminance distribution). In the next step we try to improve the sensor to obtain more information from the signal. Therefore, in Fig.2 the rectangular function and the Gaussian function are compared as sensors.

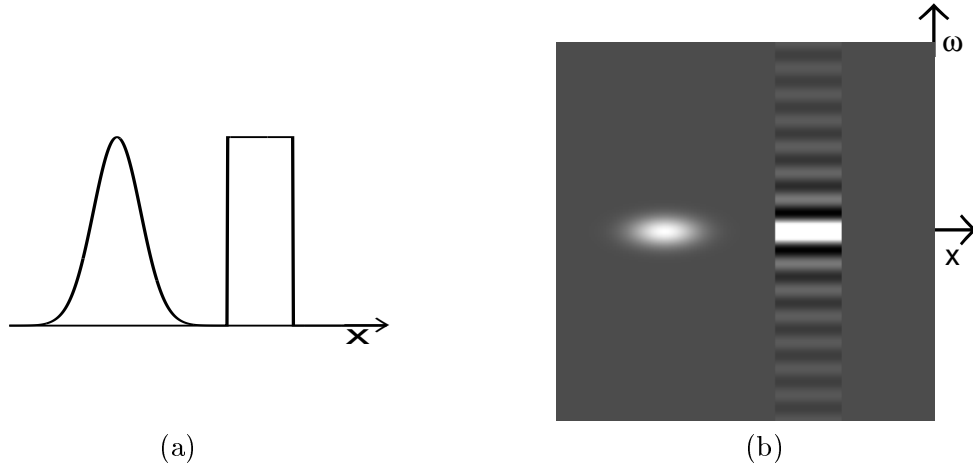


Figure 2: Comparison of rectangular and Gaussian function as sensor

Both functions have approximately the same spatial width, i.e. they have to be shifted by the same amount for the next basis function to be orthogonal. Clearly the Gaussian function will never be exactly orthogonal to shifted copies. But a small dependence between the coefficients is tolerable, at least if it is below the resolution of the amplitude of the coefficients. We will now establish the connection to the uncertainty principle: In practice the resolution often is not given by a spatial shift  $\delta x$  but by the bandwidth  $\Delta\omega$  of the sensor. Fig.2b shows that the spread of the rectangle in frequency space is much

larger than for the Gaussian. Because the Fourier basis is orthogonal, shifting the Gaussian in the frequency space by the width of its support results in an orthogonal function. Hence, for the Gaussian more (quasi) orthogonal basis functions are possible. It can be proofed that for the metric (2) the function with the best joint localization is the Gaussian. Shifting the Gaussian in frequency space leads to complex Gabor functions. The shift leaves the variances unchanged and hence the Gabor functions have the same optimal joint localization as the Gaussian.

Our question can now be answered as follows. If the restrictions are given by sensors that respond in a certain time or spatial interval and a certain frequency band and no other possibilities or restrictions exist the largest vector space in which the signal can be embedded is given by the span of Gabor functions paving the time-frequency area of the sensor. This implies that Gabor functions are the finest probes that are possible to investigate the signal or, in other words, to maximally decorrelate the information in the signal.

The improved performance of Gabor functions with respect to rectangular functions is achieved by opening the frequency channel. With this new degree of freedom the construction of orthogonal basis functions is not only possible by spatial shifts but also by frequency shifts.

**Why not open more channels?** In a strict sense the new channel is no new degree of freedom because the Dirac (position) base is already complete for all reasonable (and unreasonable) functions. This is expressed in the uncertainty principle. It is not possible to pave the  $x, \omega$ -plane with Diracs. If  $x$  is already sampled with Diracs there is no more freedom through the  $\omega$ -channel. Opening the new channel is just a possibility to fully exploit the capabilities of the sensor. That means if there exists a sensor for another orthogonal basis this channel can be opened too and will result in more information that can be retrieved. There will be also a corresponding uncertainty principle. It simply means that we have finer sensors not by improved sampling in position but with respect to some other basis. On the other hand, in practice signals are almost always investigated in position and frequency space. The exceptional role of the Fourier base results from the following facts, that are closely related to each other:

- Frequency and position are best suited to describe the restrictions for common sensors.
- The Fourier base is complementary to the position base, i.e.  $|\langle x|\omega\rangle| = 1$  with linear phase.  $e^{i\omega x}$  is maximally spread with respect to  $x$  and contains no position information.
- $e^{i\omega x}$  are the eigenfunctions for the spatial shift operator  $i\partial_x$ .
- $e^{i\omega x}$  are the eigenfunctions of linear shift invariant operators, i.e. spatially non adaptive (constant) operators that do not single out one position.

Now that we have a deeper understanding about the what and why of Gabor and Fourier bases we conclude the following:

If the base (sensors, features, alphabet) is given by the problem in hand it should be used. The uncertainty principle then reveals the consequences for the inference process as it is explained in section 2.1.4. If no alphabet is given, Gabor functions are for natural restrictions (position and bandwidth) the finest possible generic sensors.

For example in applications like texture segmentation or compression the methods are often motivated by maximal information decorrelation and not by the detection of features. On the other hand, in many recognition tasks the question is not to pack as much information as possible in a signal or to decorrelate it maximally but to recognize the natural alphabet (e.g. edges, lines etc.). The performance of Gabor functions in this case is not too bad as well [12, 46, 51]. Some more comments about Gabor functions can be found in section 3.1.

**Gabors 'proof':** Gabor proofs in his paper [25] that the Gabor functions have the best joint localization with respect to the metric of equation (2). But concerning the more general question of how much information can be coded in a signal it is rather a motivation than a strict proof. The metric is given ad hoc from an example. It is the same as in quantum mechanics but there we have different physical foundations and other questions in mind. Following a comment in [31] the influence of other reasonable metrics should not be too severe and the optimal functions are more or less Gabor shaped but to our knowledge this topic has not been investigated in depth. Another point is that Gabor functions are not orthogonal. Hence, the coefficients are correlated and there is less information in their amplitudes than for orthogonal functions. The exact consequences of this fact have not been investigated too. Orthogonal bases have been developed for subband coders as quadrature mirror filters or orthogonal wavelets. But for these functions the uncertainty principle is ignored even though their task is to decorrelate the information in the signal. Finally, we want to remind that in Gabor's paper as well as in this section only linear coding schemes as in equation (4) were considered.

#### 2.1.4 What is where?

One of the basic tasks in image analysis is to localize and classify the objects in an image. This is expressed by the question 'what is where'. There are many approaches to this question. In this subsection we comment on one approach that relates this question closely to the quantum mechanical inference process.

Wilson together with Granlund [72] and Knutsson [73] drew up another consequence of the uncertainty principle for the inference from pixels to symbols (features)(see also Sommer [65]). We give a very concise summary of their inference mechanism. According to them the basic features 's' are characterized by invariances and are represented in images by the invariant subspaces of corresponding operators (corresponding to the invariances). The subspaces are spanned by the eigenfunctions  $\Phi_k$  of the operators. E.g. in the case of shifts the operator is  $i\partial_x$  and the eigenfunctions are  $e^{i\omega x}$ . The inference to the features is done by probabilities. The probability  $P$  of the feature  $s$  that is related to the eigenfunction  $\Phi_k$  (denoted  $s(\Phi_k)$ ) to be present in the signal  $f$

$$f(x) = \sum_k a_k \Phi_k \quad (5)$$

is given by

$$P(s(\Phi_k)|f) = \frac{|a_k|^2}{\sum_l |a_l|^2} \quad (6)$$

Also serial inferences are possible by introducing in addition to the  $\Phi_k$  a second eigenbasis  $\Psi_l$  to another operator. The possibility  $P(s(\Phi_k)|f)$  of the symbol  $s$  to be present in the signal  $f$  can be estimated by:



$$P(s(\Phi_k)|f) = \sum_l P(s(\Psi_l)|f) |\langle \Psi_l | \Phi_k \rangle|^2 \quad (7)$$

For an illustration one should think of inferring to textures directly as in (6) or by first inferring to edges via (6) and then to textures by (7).

This 'subspace structure of inference' as it is called by the authors is a one-to-one translation from quantum mechanics to image processing. The well known effects of the uncertainty principle, e.g. interferences and the non commutativity are derived. For non commuting operators the features cannot be simultaneously sharp. 'Sharp' means that  $f$  consists of only one eigenfunction of the corresponding operator. For serial inferences the order in which the bases are applied changes the result. The most popular example is again that between position (operator  $x$ , eigenvector  $\delta_x$ ) and spatial frequency (operator  $i\partial_x$ , eigenvector  $e^{i\omega x}$ ) but it is true for all other features too.

To our opinion the usefulness of the details of the inference process are doubtful mainly because real symbols/features (even edges) are too complex to be linked directly to the subspaces. The linear structure of the subspace concept is not rich enough to render the complexity of the subdivision of a signal into features. The projection to a set of basis functions will only be the first of several layers in an approach to feature detection.

Some of the effects (interferences) are shared by the energy concept instead of probabilities. The effect whether one or two events are seen is more related to scale ([55]) or context than to the inference process even though the uncertainty principle is necessary for the understanding. Anyway, a linear modul is necessary in all methods and the point concerning the incompatibility of different features is undoubtedly true.

We will not go further into the details because our concern is: **What can we learn about basis functions from this formalism?**

The goal is to get from pixels to symbols. According to Wilson et al. the symbols are connected directly to the basis functions and the (invariant) subspace structure. Especially three bases are mentioned: The Dirac base  $\delta_x$ , the Fourier base  $e^{i\omega x}$ , and the Gabor base. The motivation for these bases in [73] is as follows. The Fourier base is invariant to shifts and therefore spans the invariant subspaces for the object classification. The Dirac base is invariant to the Fourier base (classification) and is therefore related to position. The Gabor functions are the best trade off between both domains and allow a desired degree of invariance in one of them.

This motivation is not wrong but misleading. Of course Fourier coefficients are used for position invariant object recognition. But beside the position invariance the connection to the objects is as close or as far and as obvious (or not obvious) as for the Dirac base or other bases. Accordingly, the role of the Gabor base in object classification is not cleared up. We already mentioned that real features are not related in a simple way to the invariant subspaces. On the other hand, the Dirac base and the Fourier base are in general used to analyse all other functions and the Gabor base is known to have many applications in computer vision. Some motivation for these functions was given in section 2.1.

The final conclusion is the following: The uncertainty principle does not indicate one special basis that is preferable for object recognition. If the features and appropriate basis functions are given, the uncertainty principle shows the consequences of this basis for the inference process. The formalism of the inference process of [72, 73] should not be overemphasized because in general, features are not given by simple projections but in following (nonlinear) processing layers.

## 2.2 Hilbert transform

During the last decade complex quadrature filters, with the real and imaginary part connected by a Hilbert transform, became an important tool in image processing. A good introduction to the Hilbert transform can be found in the textbook of Bracewell [8]. In the following paragraphs we distinguish three lines leading to the use of quadrature filters.

### Gabor functions and analytical signals

In Gabor's paper [25] the signal is extended to an analytical signal to allow an elegant mathematical formalism. Gabor's 'elementary signals' or Gabor functions as they were called later are therefore complex analytical signals. One benefit compared to real signals concerns the metric of equation (2) that is used in the paper. The spread of the Gabor function in Fourier space  $\Delta\omega$  corresponds to the shift that is necessary to obtain an (quasi) orthogonal function only in the case of the complex function but not for the real one (see Fig.3), because the latter has two lobes.

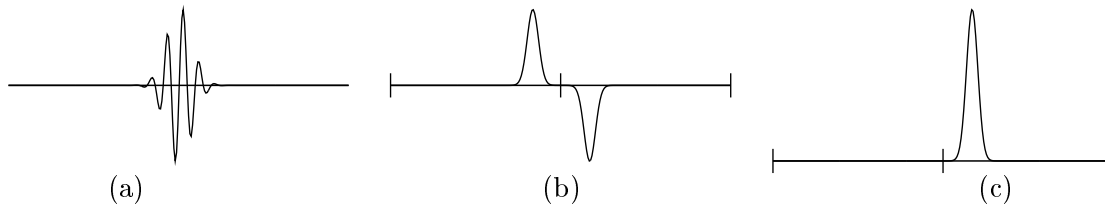


Figure 3: Fourier transform for the real and analytical signal. (a) signal, (b) Im of FFT of (a) (Re is zero), (c) Re of FFT of analytical signal of (a) (Im is zero). The FFT in (b) has two lobes. Hence, the metric from equation (2) is not very meaningful to describe the width of the signal (b).

### Feature detection

To detect features independently of their profile (line, edge, mixed) quadrature pairs of filters are used [3, 26, 29, 36, 53, 55, 59, 62]. The most profound analysis is the paper of Perona and Malik [59]. Usually the quadrature pair consists of a real even part and an imaginary odd part. The response is transformed to an energy/phase representation where the energy detects the features and the phase recognizes their symmetry type.

### Australian feature detection

An Australian group (Owens, Venkatesh, Morrone, Burr ..) [56, 57, 58, 70, 71] founded a new school of feature detection. Based on comprehensive psychophysical experiments and some mathematics they state that features that are perceived by humans are connected to the energy maxima of the analytic signal of the image. The method is in principle the same as in the last paragraph. The difference is that the former starts by defining features and then designs optimal filters for them whereas the latter (in its basic form) uses always broad band filters and defines as features what is detected in the energy.

### Comments

The sine and cosine Gabor functions are not true Hilbert pairs. This is easy to see as the cosine functions do not integrate to zero. Therefore, in some applications modified versions are used. E.g in Lades et al. [40], Ronse [61], and Michaelis et al. [53] the following DC corrected cosine function:

$$G_{\sigma,c}(x) = e^{-\frac{x^2}{2\sigma^2}} (e^{ic\frac{x}{\sigma}} - e^{-\frac{1}{2}c^2}) \quad (8)$$

For large  $c$  the correction is neglectible, and therefore one usually chooses  $c \geq 5$ . But large  $c$  are not always possible (e.g. [53]). Some authors even require an imaginary

component that is insensitive to linear ramps [59, 61]. A ramp-corrected complex Gabor function is:

$$G_{\sigma,c}(x) = e^{-\frac{x^2}{2\sigma^2}} (e^{ic\frac{x}{\sigma}} - e^{-\frac{1}{2}c^2} - i\frac{cx}{\sigma}e^{\frac{1}{2}c^2}) \quad (9)$$

Other corrected Gabor functions are also used, e.g. in [29]. Figure 4 compares the Gabor cosine function with some related functions. For demonstration purposes the oscillation parameter is  $c = 0.1$ , much smaller than any reasonable value. The imaginary part in fig.4 is always a Gabor sine function.

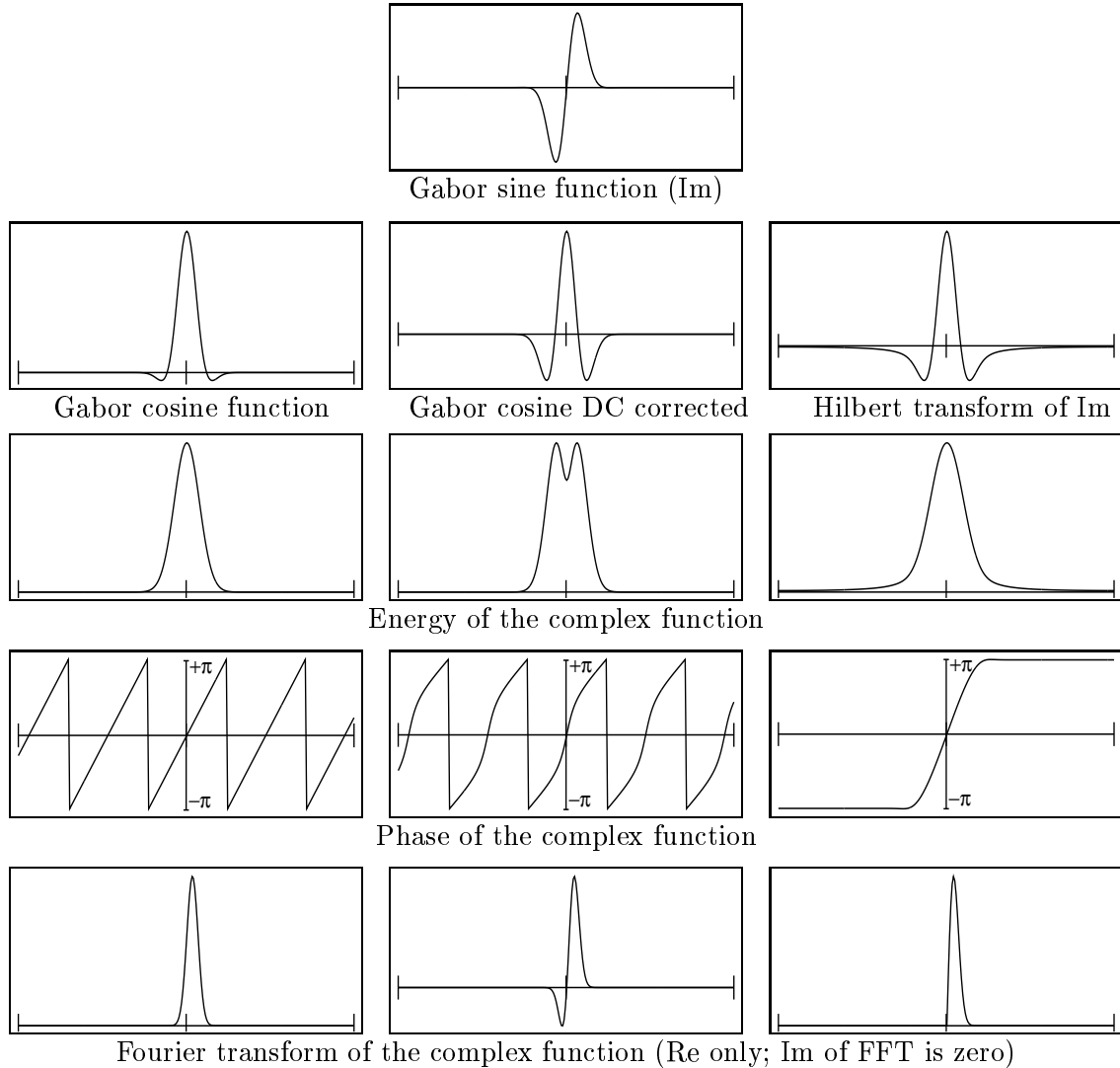


Figure 4: Comparison of different odd/even pairs for Gabor functions. All three even (Re) variants have the same odd (Im) partner that is depicted in the first row. For further explanations see text.

- The first column is the Gabor cosine function that obviously does not integrate to zero. The energy in position space (i.e. the response for a line) is the unmodulated Gaussian of the Gabor function. The phase is perfectly linear. The Fourier transform is Gaussian shaped too.
- The second column is the DC corrected version from equation (8). The energy is not single peaked and the Fourier transform is not single lobed (but remember that

$c$  is unreasonable small). The phase is not linear. Especially there's not much shift between the phases  $-\pi/2$ ,  $0$ , and  $\pi/2$  and hence, the phase estimation is not very robust.

- The last column is the true Hilbert transform of the Gabor sine function. The DC component is zero and the phase is approximately linear. The drawback of this function is its width. The slow decay results from a convolution of the Gabor sine function with  $1/x$  by the Hilbert transform. As a consequence the energy is not the unmodulated Gaussian but a broad function with a support much larger than that of the Gabor sine function. In practice this means that the response is much larger than the feature and might be influenced by other features in the neighborhood (interferences). The Hilbert transforms of the Gabor sine functions and many other functions of interest do not have simple analytical forms.

In this extreme case none of the functions is very satisfactory and for larger  $c$  all become the same. In between the trade offs have to be considered carefully. Especially we want to make the point that the energy of the Hilbert transform pair is not what might be expected naively as the envelope of the Gabor sine function. The details of the trade offs of the Hilbert transform and other odd/even pairs should be better investigated.

These trade offs exist for others than Gabor functions too. The Hilbert transform is a very good candidate to design odd/even pairs but there are also alternative possibilities. For example Perona and Malik [59] use a Gaussian derivative and its true Hilbert transform whereas in Michaelis and Sommer [54, 55] a pseudo quadrature pair of two Gaussian derivatives is used and the merits and drawbacks are discussed.

## 2.3 Scale-space

In a narrow sense scale-space means the continuous scale LoG analysis of Witkin [75] that was a seminal work for hundreds of following papers. This and all other work related to the LoG underlines the importance of the derivatives of Gaussians. Babaud et. al [4] proofed that the Gaussian is the unique kernel for smoothing 1D signals without producing new zero-crossings. Yuille and Poggio [78] and Koenderink [32] investigated the 2D case. In 2D the zero-crossings can split and merge but nevertheless the authors show the exceptional role of the Gaussian even in the 2D case (zero-crossings never vanish at finer). More details about scale-space behavior can be found in appendix A.

But the importance of scale-space violations depend on the information that is retrieved from the image. E.g. Kube and Perona [39] point to vanishing zero crossings in the case of energy filters for edge detection. On the other hand, Ronse [61] shows that for the Morrone/Owens edge detection scheme [56, 57] the only requirement for a smoothing filter not to change the edge information is zero Fourier phase. Hence, the scale-space behavior should be treated in the same way as the uncertainty principle or the true quadrature pairs: **important but not undisputable**.

The Gaussian has two other fundamental and unique benefits as a smoothing kernel. First it has minimal uncertainty and second it is **isotropic and cartesian separable**. Therefore, we would like to derive scaled signals by Gaussian smoothing. To satisfy the wavelet like scaling ( $f(x) \mapsto f(sx)$ ) as well as the Gaussian scaling, functions are preferred that satisfy ( $D$  is a Gaussian):

$$D_\sigma(x) * f(x) \propto f(sx) \quad s = s(\sigma) \quad (10)$$

This type of scaling is investigated in depth by Koenderink and van Doorn [34] and is referred to as *relative invariance*. If  $f$  has free parameters, denoted with  $\tau$ , a weaker form

is possible where the smoothed function has changed parameters  $\tau'$ :

$$D_\sigma(x) * f_\tau(x) \propto f_{\tau'}(sx) \quad s = s(\sigma) \quad (11)$$

This behavior is exhibited by Gabor functions (equation (3)) that have the free oscillation parameter  $c$ .

## 2.4 Regularization and noise

Torre and Poggio [68] show the ill-posed nature of edge detection due to noise. In their paper, edge detection is seen as differentiation. The authors show that an optimal regularization can be done by filtering with a spline that is close to a Gaussian. Canny [14] asked for the optimal filter that maximizes the product of a localization measure and the SNR. He obtained a function close to the first derivative of a Gaussian. Mehrotra et al. [51] calculated the same criterion for Gabor sine functions. These functions do not perform too bad if  $c$  (see equation (3)) is chosen to be approximately 1 (but see the comment in section 6.3). This is not too surprisingly as the Gabor sine function resembles very much the first derivative of a Gaussian (section 6.3). Larger  $c$  result in multiple responses, whereas for smaller  $c$  the detection becomes hampered because of noise maxima. Manjunath and Chellappa [46] discuss the same for complex Gabor filters used as energy feature detectors. For some reason they only use  $c = \pi$ , a value that is not very reasonable for edge detection.

## 2.5 Discrete case

For theoretical considerations, the basis functions most often are supposed to be continuous. We assume that future vision systems need spatial resolutions such that this continuous point of view is justified. On the other hand, there always will be some discretization and in past systems discretization often played a dominant role. Therefore, we hint to some papers that take into account the effect of discretization.

Usually for discretization the unmodified continuous basis functions are sampled. It can be calculated which sampling rate is necessary to limit the discretization effects. This is done for example in [7] for the Gabor functions. If the sampling of the basis functions has to be sparse because of speed requirements or a sparse sampling of the signal, it is not possible just to sample the continuous basis functions. Discrete modifications have to be used instead. Examples can be found in Davies [20], Wilson and Bhalerao [74], and Hashimoto and Sklansky [27] for the Gaussian, in Lindeberg [41] for the scale space analysis, in Martens [49] for the Hermite transform that includes derivatives of Gaussians and Hermite functions. For discrete derivatives see also Lindeberg [42] and Hummel and Lowe [30].

## 2.6 Wavelets

Wavelets are bases that consist of translated and scaled copies of one mother wavelet:

$$f_{\tau,s}(x) = \sqrt{s}f(sx - \tau) \quad (12)$$

Among the most popular wavelets are Gabor functions [37] even though for small  $c$  (eq. (3)) Gabor cosine functions do not integrate to zero and hence are not admissible. In the case of Gabor functions the term 'wavelets' refers to the scaling (12) and to the analysis of the response in a combined position scale space. The theory of wavelets brought new insights about sampling schemes for complete bases, regularity and other things.

There is a class of wavelets, the orthogonal wavelets, that is of special interest. We have already mentioned that Gabor functions are not orthogonal. In the case of orthogonal wavelets, where specific mother wavelets and sampling schemes have to be used, all scaled and translated functions are orthogonal to each other. The basis is therefore minimal and complete and the coefficients are uncorrelated. With these wavelets orthogonal, discrete multiresolution pyramids can be built where the information in the different scales is uncorrelated [44, 45]. On the other hand, the uncertainty principle is ignored in the design of orthogonal wavelets and they are not always plausible as feature detectors, e.g. the orthogonal Gabors of Daubechies et al. [15]. If the goal is the analysis and not the coding of the data the orthogonal sampling schemes with its minimal number of coefficients are unattractive anyway [52, 64].

## 2.7 Biological vision systems

Strong motivations for the design of early vision methods stem from findings in biological vision systems. In the early 80th the famous LoG filter had its counterpart in the ganglion cells of the retina. However, today it is believed that the first interesting processing steps take place only in the visual cortex. Especially the interpretation of the first layer (V1) is usually based on the work of Hubel and Wiesel. According to them the neurons in V1 act as feature detectors and can be classified in simple cells, complex cells, and hypercomplex cells. The simple cells are believed to be the first processing units and they are modeled by linear projections to basis functions. Complex cells might be modeled by energy feature detectors as in [60]. Energy feature detectors are widely used in computer vision, signal processing, and simulations of biological vision systems [29, 38, 46, 53, 55, 59, 70]. The simple cells are mostly matched to Gabor functions [16, 17, 47, 60]. But derivatives of Gaussians are also suggested by other authors [77, 33, 34, 49, 66].

## 2.8 Applications

Of course the final justification for using projections to basis functions in general and Gabor functions, derivatives of Gaussians, and related functions in particular is the usefulness for the problems in early vision. We list the most important early vision tasks and cite papers that use basis functions in general and Gabor functions and derivatives of Gaussians in particular to underline their relevance.

- **Compression and Coding**

LoG coding by Burt and Adelson [13] and Hermite transform coding by Martens [50]. In case of the non orthogonal Gabor functions the calculation of the coefficients is a non trivial problem. Many solutions have been suggested, e.g. by Bastiaans [6] or by Daugman [18] who uses neural networks for the computation. Coders that use orthogonal wavelets have less coefficients [44, 45]. On the other hand, this might not be the decisive argument if other bases are more adapted to the visual relevant features or if they are 'finer probes' because of optimal uncertainty (may be less entropy and better vector quantization possible).

- **Texture segmentation**

Frequently Gabor functions are used where the energy, phase or both informations are used for texture segmentation [10, 11, 7, 69].

- **Edge and line detection**

First derivative [14] or second derivative [48] edge detectors (see also [68]). Energy detectors for edges and lines [38, 46, 59, 70, 56].

- **Junction detection and classification**  
Andersson [3], Rosenthaler et al. [62], Michaelis and Sommer [55].
- **Stereo**  
Sanger [63].
- **Optical flow**  
Adelson and Bergen [2], Fleet [21], Heeger [28].
- **Shape from X**  
Freeman and Adelson [24].
- **Restoration and Enhancement**  
Knutsson et al. [36].

Besides these applications there are methods as the jets or differential geometry that cover a wide range of applications and that are based on derivatives of Gaussians. Jets were introduced by Koenderink and van Doorn [33, 34]. Applications of differential geometry can be found e.g. in papers of Barth et al. [5] and Florack [23].

### 3 Basis functions

In this section we list the basis functions that are of interest for early vision tasks. In section 2.5 we mentioned functions for the discrete case. These functions and all other functions that are purely technical optimizations for special tasks without any generic value will not be considered. The derivatives of Gaussians and the Gabor functions are investigated in detail in sections 4, 5 and 6.

#### 3.1 Gabor functions

Gabor functions are optimal with respect to the uncertainty principle and hence they are the finest possible sensors. In addition they prove to be good feature detectors [46, 51] such that du Buf [12] states that the justification for the Gabor functions in the visual cortex might not be the uncertainty principle but their optimal preprocessing for the following layers. A drawback of the Gabor functions is that the real part does not integrate to zero, especially for small  $c$  that are used for edge detection. Different solutions are possible as explained in section 2.2.

#### 3.2 Derivatives of Gaussians

Derivatives of Gaussians are motivated by their optimal feature detection properties [14, 48, 59], their connection to scale-space [32, 75], and their close connection to differential geometrical methods [23] and Taylor expansions [33, 34]. They have a simple analytical structure what makes them easy to handle and they need no DC correction even though the spatial extent has to be quite large for even orders. Odd/even pairs for the derivatives of Gaussians can be constructed by taking the true Hilbert transform [59] or successive derivatives [33, 54, 55].

#### 3.3 Hermite functions

The importance of the Hermite functions in image processing is due to their connection to the derivatives of Gaussians. With the latter they are often confused because of their similar analytical form:

$$\begin{aligned}
P_n(x) &= N e^{-\frac{x^2}{2}} H_n(x) \\
D_n(x) &= N e^{-x^2} H_n(x)
\end{aligned}
\tag{13}$$

$H_n$  denotes the  $n$ 'th order Hermite polynomial,  $N$  is a normalization factor.  $P_n$  and  $D_n$  are the Hermite function and the derivative of Gaussian of  $n$ 'th order. The difference between both is the scale of the Gaussian in relation to the scale of the Hermite polynomial. In case of the Hermite functions the Hermite polynomials grow as fast as the exponential decays and hence the maxima of the Hermite functions have all about the same height (Fig.5), giving them the shape of a truncated sine/cosine wave.

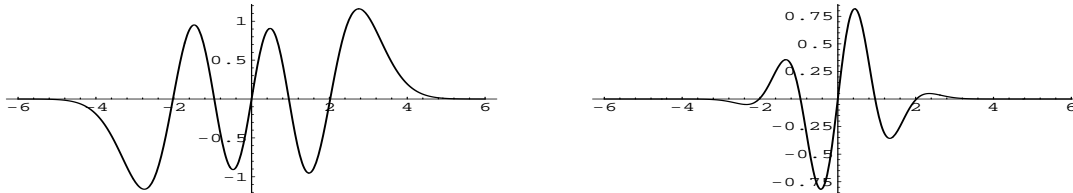


Figure 5:  $P_5(x)$  (left) and  $D_5(x)$  (right): Hermite function and derivative of Gaussian of order 5

The connections of the Hermite functions to the derivatives of Gaussians is as follows: In case of the Hermite transform (Martens [49]) the filter functions (analysing) are the derivatives of Gaussians, whereas the pattern functions (reconstruction) are the Hermite functions. This unsymmetry is necessary because the derivatives of Gaussians are not orthogonal. Instead the Hermite transform can be viewed as the transform of a Gaussian windowed image. With respect to the Gaussian weighted window the derivatives are orthogonal. Some more details about the Hermite transform can be found in appendix B.

A second point of view connecting both functions is that of Koenderink and van Doorn [34]. They call the derivatives of Gaussians neighborhood operators and the Hermite functions ripples. The basic question for them is a taxonomy of all functions satisfying a *relative invariance* scaling equation ((10), section 2.3). This question leads to the diffusion equation. The derivatives of Gaussians are solutions to this equation. They can be decomposed into the product of an invariant part (the ripple) and a scaling window (a Gaussian). The ripples (Hermite functions) themselves satisfy the quantum mechanical equation of the harmonic oscillator. Hence, they are orthogonal.

The Hermite functions have two interesting properties. First they **maximize** the uncertainty principle (with respect to a certain class of functions, [31]) and second, as for the Gaussian, their Fourier transform has the same functional form as the function itself.

### 3.4 Other functions

Koenderink suggests also solutions of the polar separable harmonic oscillator (times Gaussian window) as basis functions [34]. This basis can easily be converted to the cartesian separable derivatives of Gaussians.

The Linköping school suggests functions that are polar separable in Fourier space [35, 36, 3, 26, 62]. The angular component is usually given by a power of the cosine function whereas the radial component is based on Gabor functions. One benefit of the polar separability is that the response to an edge separates in a term that depends on the profile of the edge and a term that depends on the orientation.

Finally, the orthogonal wavelets are another interesting class even though they are not always optimal if focus is on the analysis of the image content. The restrictions for the



design of the functions (and the sampling scheme) hamper their use for feature detection.

## 4 Derivatives of Gaussians

In this section we collect some definitions, conventions, and calculations that are necessary for the following and that are often a source of trouble because of different sign conventions and constants.

### 4.1 Definition

The Gaussian has the general form:

$$D_0(x) = N(a) e^{-ax^2} \quad (14)$$

$N$  is a normalization constant that depends on  $a$  but not on  $x$ . There are several possibilities for the coefficient  $a$  in the exponent: (1) The standard form is  $a = (2\sigma^2)^{-1}$  where  $\sigma$  is the standard deviation. (2) In the Hermite transform a Gaussian window as well as the squared window are used (see appendix B) and hence  $a$  can be  $a = \sigma^{-2}$ . (3) In scale-space theory the natural scale parameter  $s$  is given by  $a = (4s)^{-1}$  because this Gaussian is the Green's function of the simplest diffusion equation (see appendix A). In general we assume the Gaussian to have the standard form. If not otherwise stated, expressions containing  $\sigma$  refer to this form.

The normalization constant can be such that  $L^1(D_0) = 1$  ( $N_1$ ) or  $L^2(D_0) = 1$  ( $N_2$ ).  $L^2$  is especially interesting if a reconstruction is required as in the Hermite transform (appendix B). But we will also consider the  $L^1$  norm because sometimes it is preferable for the interpretation of the responses. Using the integrals of appendix C we can calculate the normalization constants:

$$\begin{aligned} (L^2(D_0))^2 &= 2 \int_0^\infty N_2^2 e^{-2ax^2} dx = N_2^2 \frac{\sqrt{\pi}}{\sqrt{2a}} = N_2^2 \sigma \sqrt{\pi} \\ &\Rightarrow N_2 = (2a)^{\frac{1}{4}} \pi^{-\frac{1}{4}} = \sigma^{-\frac{1}{2}} \pi^{-\frac{1}{4}} \end{aligned} \quad (15)$$

$$\begin{aligned} L^1(D_0) &= 2 \int_0^\infty N_1 e^{-ax^2} dx = N_1 \frac{\sqrt{\pi}}{\sqrt{a}} = N_1 \sigma \sqrt{2\pi} \\ &\Rightarrow N_1 = \frac{\sqrt{a}}{\sqrt{\pi}} = (\sqrt{2\pi}\sigma)^{-1} \end{aligned} \quad (16)$$

The derivatives of the Gaussian are ( $H_n$  are the Hermite polynomials, appendix D):

$$\begin{aligned} D_0 &= N e^{-ax^2} \\ D_{1x} &= -N 2ax e^{-ax^2} \\ D_{2x} &= N 2a(2ax^2 - 1) e^{-ax^2} \\ D_{3x} &= -N 4a^2(2ax^3 - 3x) e^{-ax^2} \\ D_{4x} &= N 4a^2(4a^2x^4 - 12ax^2 + 3) e^{-ax^2} \\ D_{nx} &= \frac{d^n D_0}{dx^n} = (-1)^n N a^{\frac{n}{2}} H_n(\sqrt{a}x) e^{-ax^2} \end{aligned} \quad (17)$$

The formula for the general expression  $D_{nx}$  is derived with (56) of appendix D. Fig.6 shows the first four derivatives.

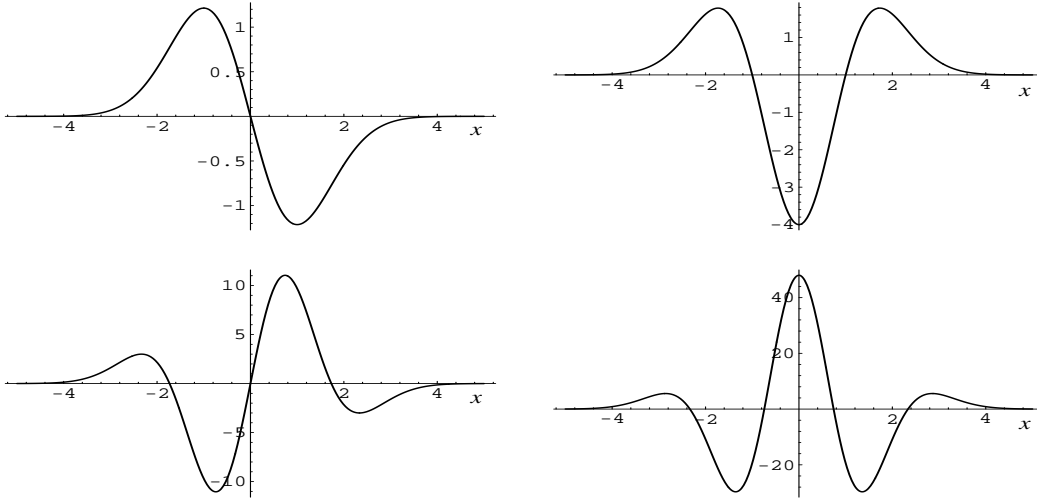


Figure 6: First four derivatives of a Gaussian ( $N = \sigma = 1$ ).

## 4.2 Fourier transform

We use the following convention for the Fourier transform and its inverse (it is the same as in [9, 43] and Mathematica):

$$\begin{aligned} \hat{f}(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{i\omega x} dx \\ f(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{-i\omega x} d\omega \end{aligned} \quad (18)$$

To calculate the Fourier transform of the derivatives of Gaussians we use the derivative theorem:

$$\int_{-\infty}^{\infty} \frac{d^n f(x)}{dx^n} e^{i\omega x} dx = (-i\omega)^n \int_{-\infty}^{\infty} f(x) e^{i\omega x} dx \quad (19)$$

Using this theorem we get:

$$\hat{D}_{nx}(\omega) = (-i\omega)^n \hat{D}_0 = N (-i\omega)^n \frac{1}{\sqrt{2a}} e^{-\frac{\omega^2}{4a}} = N (-i\omega)^n \sigma e^{-\frac{(\sigma\omega)^2}{2}} \quad (20)$$

The Fourier transform of  $D_0$  is from [9]. Fig.7 shows the Fourier transform of the Gaussian and its first three derivatives.

## 4.3 Normalization

If the convolution with derivatives of Gaussians is meant to be the true derivative of the smoothed signal, the constant  $N$  in (17) is  $N_1$  of  $D_0$  from equation (16). However, if we are interested in these functions as feature detectors and especially if we want to combine successive orders to odd/even pairs this choice is not convenient. From equations (17) and (54) in appendix C we see that  $L^1(D_{nx}) \propto \sigma^{-n}$  (with  $N_1$ ) and likewise  $L^2(D_{nx}) \propto \sigma^{-n}$  (with  $N_2$ ) or  $L^2(D_{nx}) \propto \sigma^{-(2n+1)/2}$  (with  $N_1$ ).

Hence, the amplitude of the responses for higher derivatives decreases fast with increasing  $\sigma$  and increasing order  $n$ . In order to use the Gaussian derivatives as feature detectors,  $N$  has to be chosen such that all orders  $n$  are  $L^1$  or  $L^2$  normalized to unity.

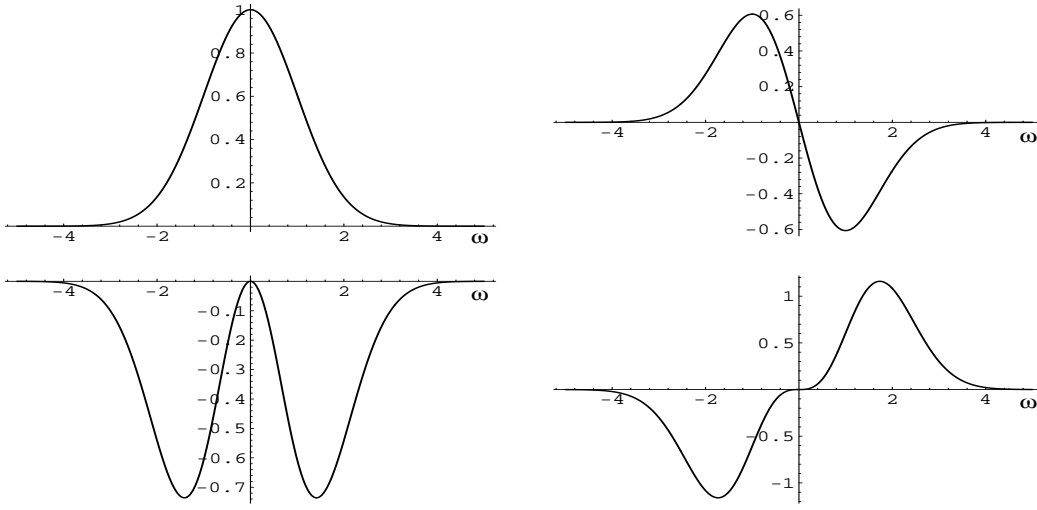


Figure 7: Fourier transform of the Gaussian and its first three derivatives ( $N = \sigma = 1$ ). The zero's order is identical in both domains (for  $\sigma = 1$ ). The first order is identical up to an amplitude factor. The Fourier transforms are purely real for even orders and imaginary for odd orders.

### $L^2$ normalization

The calculation of the  $L^2$  norm is most easily done in Fourier space. With the Fourier transform convention of (18), the frequency basis is normalized and the  $L^2$  norm is preserved. Using (53) from appendix C we calculate:

$$\begin{aligned}
 L^2(D_{nx}) &= N \left( \int_{-\infty}^{\infty} \omega^{2n} \frac{1}{2a} e^{-\frac{\omega^2}{2a}} d\omega \right)^{\frac{1}{2}} = N a^{-\frac{1}{2}} \left( \int_0^{\infty} \omega^{2n} e^{-\frac{\omega^2}{2a}} d\omega \right)^{\frac{1}{2}} = \\
 &= N \sqrt{\frac{1 \cdot 3 \dots (2n-1) \sqrt{\pi}}{\sqrt{2}}} a^{n-\frac{1}{2}} = N \sqrt{\frac{1 \cdot 3 \dots (2n-1) \sqrt{\pi}}{2^n \sigma^{2n-1}}}
 \end{aligned} \tag{21}$$

### $L^1$ normalization

An analytical expression for the  $L^1$  norm is difficult to derive. Even though there is an analytical expression for the integrals of derivatives of Gaussians [43, section 5.6.2]

$$\int_0^x e^{-z^2} H_n(z) dz = H_{n-1}(0) - e^{-x^2} H_{n-1}(x) \tag{22}$$

there is no general expression for the zeros of Hermite polynomials [67]. Hence, it is possible to calculate the norm analytically order by order but not to derive a general expression for any order. For this reason we prefer to do the calculation numerically (except for zero and first order).

$$L^1(D_0) = N 2 \int_0^{\infty} e^{-ax^2} dx = N \frac{\sqrt{\pi}}{\sqrt{a}} = N \sqrt{2\pi\sigma} \tag{23}$$

$$L^1(D_1) = N 2 \int_0^{\infty} 2axe^{-ax^2} dx = N 2 \tag{24}$$

$$\begin{aligned}
L^1(D_{nx}) &= N a^{\frac{n}{2}} \int_{-\infty}^{\infty} |H_n(\sqrt{a}x)| e^{-ax^2} dx = \\
&= N a^{\frac{n-1}{2}} \int_{-\infty}^{\infty} |H_n(z)| e^{-z^2} dz = N \sigma^{n-1} 2^{\frac{1-n}{2}} \int_{-\infty}^{\infty} |H_n(z)| e^{-z^2} dz
\end{aligned} \tag{25}$$

The integral  $I(n) := \int_{-\infty}^{\infty} |H_n(z)| e^{-z^2} dz$  is calculated numerically:

$n$	0	1	2	3	4	5
$I(n)$	1.77245	2	3.43106	7.57009	19.8557	59.2576
$n$	6	7	8	9	10	
$I(n)$	195.901	704.821	2725.68	11225.8	48890.4	

Table 1: Numerically calculated Integrals of the  $L^1$  norm of the derivatives of Gaussians

## 5 Optimal odd/even pairs with Gaussian derivatives

There is growing evidence that feature detection has to be done by energy detectors that consist of complex filters of an even real part and an odd imaginary part. In section 2.1 we discussed the Hilbert transform as a natural method to construct a partner for a given odd or even function. We have also mentioned that for weakly oscillating functions the true Hilbert transform is not always the best choice because it results in functions with a large support. Moreover, considering a framework that already uses all lower derivatives of Gaussians (e.g. jets or differential geometry) we would like to save the computational burden to calculate the projections to the Hilbert transforms. In addition the Hilbert transforms have the drawback that they are not given analytically.

We would like to build complex odd/even filters  $F$  from derivatives of Gaussians alone. From (21) and (25) we see that the dependence on the order  $n$  of the  $L^1$  norm as well as the  $L^2$  norm and the amplitude is given by  $\sigma^{-n}$ . Hence, to make the relative strength of the odd and even function independent of their scale the lower order function has to be multiplied by a factor  $\sigma^{-1}$ . **With  $k$  we introduce a free parameter that controls the relative weight of the odd and even part** of the complex filter  $F$ .

$$F := D_{2mx} \mp i k \sigma^{\pm 1} D_{(2m\pm 1)x} \tag{26}$$

With  $D_{nx}$  we denote the derivatives from equation (17) and we assume both functions to have the **same** normalization constant  $N$ . **Note that  $k$  is always at the imaginary, odd part.** This apparently changes the formulas for  $k$  for the two cases that the odd part has the higher or lower order. The relative sign of the odd and even part in (26) is such that the Fourier energy at negative frequencies compensates as can be seen from fig.7. For the opposite Fourier transform sign convention the sign of the imaginary part changes.

The free parameter  $k$  raises the question: **What is the best relative amplitude of the two functions?** In this section we discuss several possible answers. As we will see, for the lower derivatives there is no choice for  $k$  that optimizes all requirements. Depending on the task we have to accept more or less satisfactory compromises. The optimization criteria that are discussed in the following subsections are (1) same  $L^2$  or  $L^1$  norm of the odd and even part, (2) minimal Fourier energy at negative frequencies of the complex filter, (3) monomodal energy, (4) fit of the energy to a Gaussian, and (5) linear phase of the complex filter.

In principle the comparison of the different choices for the relative weight  $k$  should include an analysis of the responses to signals and noise. But this is beyond the scope of this report, and even when considering signals and noise, a final decision is difficult because there is no single generic signal or noise model. Hence, we restrict in this report to the investigation of the filters themselves. Therefore, the implications of our results are of a qualitative nature. Considering for example the monomodal energy criterion we know that side maxima lead to false detections or that nonlinear phase leads to a worse performance in classifying odd and even signals. The quantitative investigation of these things needs signal and noise models. But even without this we get a qualitative feeling for the performance of complex Gaussian derivative filters and we know qualitatively how this performance is influenced by increasing or decreasing the relative weight  $k$ .

### 5.1 Same $L^2$ or $L^1$ norm

The  $L^2$  norm and the  $L^1$  norm of  $D_{nx}$  are given by the equations (21) and (25). According to (25) there is no simple analytical expression for the  $L^1$  norm and therefore we show numerical results for  $k_{L^1}$  in table 2.  $n$  denotes the **higher** order of the odd and even part.

$$\frac{L^2(D_{nx})}{L^2(D_{(n-1)x})} = \sqrt{n - \frac{1}{2}} \sigma^{-1} \implies k_{L^2} = \begin{cases} \sqrt{n - \frac{1}{2}} & n \text{ even} \\ 1/\sqrt{n - \frac{1}{2}} & n \text{ odd} \end{cases} \quad (27)$$

$$\frac{L^1(D_{nx})}{L^1(D_{(n-1)x})} = \frac{I(n)}{\sqrt{2} I(n-1)} \sigma^{-1} \implies k_{L^1} = \text{see table 2} \quad (28)$$

The integrals  $I(n)$  are given by table 1.

$n$	1	2	3	4	5
$k_{L^1}$	1.25332	1.21306	0.64097	1.85469	0.47386
$n$	6	7	8	9	10
$k_{L^1}$	2.3376	0.3931	2.7345	0.34338	3.0796

Table 2: Numerically calculated weight factor  $k_{L^1}$  for the odd and even part to have the same  $L^1$  norm

The following table shows that  $k$  is approximately the same for both norms, i.e.  $k_{L^1} \approx k_{L^2}$  ( $n = 1$  is not very interesting). To be precise, we listed  $k_{L^2}/k_{L^1}$  for  $n$  even and  $k_{L^1}/k_{L^2}$  for  $n$  odd in the table to better demonstrate the convergence.

$n$	1	2	3	4	5
$k_{L^2}/k_{L^1}$	0.886	1.0096	1.0135	1.0087	1.0052
$n$	6	7	8	9	10
$k_{L^2}/k_{L^1}$	1.0032	1.0021	1.0015	1.0011	1.0008

Table 3: The table shows the relative magnitude of  $k_{L^2}$  and  $k_{L^1}$ . The difference between both weights is small and vanishes with increasing order  $n$ .

### 5.2 Minimal Fourier energy at negative frequencies

A complex function with the odd and even part in quadrature has vanishing Fourier energy at negative frequencies. Therefore, we calculate the optimal relative amplitude  $k$

that minimizes the relative amount of Fourier energy at negative frequencies. By  $A(k)$  we denote the energy at positive frequencies divided by the energy at negative frequencies. This expression has to be maximized.

$$\begin{aligned}
A(k) &:= \frac{\int_0^\infty (\omega^{2m} + k\sigma^{\pm 1}\omega^{2m\pm 1})^2 e^{-\sigma^2\omega^2} d\omega}{\int_{-\infty}^0 (\omega^{2m} + k\sigma^{\pm 1}\omega^{2m\pm 1})^2 e^{-\sigma^2\omega^2} d\omega} = \frac{\int_0^\infty (\omega^{2m} + k\sigma^{\pm 1}\omega^{2m\pm 1})^2 e^{-\sigma^2\omega^2} d\omega}{\int_0^\infty (\omega^{2m} - k\sigma^{\pm 1}\omega^{2m\pm 1})^2 e^{-\sigma^2\omega^2} d\omega} \\
&= \frac{\int_0^\infty (\omega^{4m} + 2k\sigma^{\pm 1}\omega^{4m\pm 1} + k^2\sigma^{\pm 2}\omega^{4m\pm 2}) e^{-\sigma^2\omega^2} d\omega}{\int_0^\infty (\omega^{4m} - 2k\sigma^{\pm 1}\omega^{4m\pm 1} + k^2\sigma^{\pm 2}\omega^{4m\pm 2}) e^{-\sigma^2\omega^2} d\omega}
\end{aligned}$$

We now split into the two cases that the odd function has higher or lower order than the even function. Using formula (53) yields ( $\{2m-1\}$  denotes the product  $1\cdot 3\cdots(2m-1)$ ):

**Case  $2m-1$ :**

$$\begin{aligned}
A(k) &= \frac{\frac{\{4m-1\}\sqrt{\pi}}{2^{2m+1}\sigma^{4m+1}} + 2k\frac{(2m-1)!}{2\sigma^{4m+1}} + k^2\frac{\{4m-3\}\sqrt{\pi}}{2^{2m}\sigma^{4m+1}}}{\frac{\{4m-1\}\sqrt{\pi}}{2^{2m+1}\sigma^{4m+1}} - 2k\frac{(2m-1)!}{2\sigma^{4m+1}} + k^2\frac{\{4m-3\}\sqrt{\pi}}{2^{2m}\sigma^{4m+1}}} = \\
&= \frac{\{4m-1\}\sqrt{\pi} + k2^{2m+1}(2m-1)! + k^2\{4m-3\}2\sqrt{\pi}}{\{4m-1\}\sqrt{\pi} - k2^{2m+1}(2m-1)! + k^2\{4m-3\}2\sqrt{\pi}} =: \frac{u(k)}{v(k)}
\end{aligned}$$

The maxima of  $A(k)$  are found by taking the derivative with respect to  $k$ . If the numerator of  $A(k)$  is called  $u$  and the denominator  $v$  we have:

$$\begin{aligned}
\frac{\partial A}{\partial k} &\stackrel{(!)}{=} 0 \implies u'v - v'u = 0 = \\
&= 2^{2m+2}\sqrt{\pi}\{4m-3\}(2m-1)! (2k_F - 4m + 1) = 0 \\
&\implies k_F = \sqrt{2m - \frac{1}{2}}
\end{aligned} \tag{29}$$

**Case  $2m+1$ :**

$$\begin{aligned}
A(k) &= \frac{\frac{\{4m-1\}\sqrt{\pi}}{2^{2m+1}\sigma^{4m+1}} + 2k\frac{(2m)!}{2\sigma^{4m+1}} + k^2\frac{\{4m+1\}\sqrt{\pi}}{2^{2m+2}\sigma^{4m+1}}}{\frac{\{4m-1\}\sqrt{\pi}}{2^{2m+1}\sigma^{4m+1}} - 2k\frac{(2m)!}{2\sigma^{4m+1}} + k^2\frac{\{4m+1\}\sqrt{\pi}}{2^{2m+2}\sigma^{4m+1}}} = \\
&= \frac{\{4m-1\}2\sqrt{\pi} + k2^{2m+2}(2m)! + k^2\{4m+1\}\sqrt{\pi}}{\{4m-1\}2\sqrt{\pi} - k2^{2m+2}(2m)! + k^2\{4m+1\}\sqrt{\pi}} =: \frac{u(k)}{v(k)}
\end{aligned}$$

$$u'v - v'u \stackrel{(!)}{=} 0 = 2^{2m+3}(2m)!\{4m-1\}\sqrt{\pi} (k_F^2(4m+1) - 2)$$

$$\implies k_F = \sqrt{\frac{2}{4m+1}} \quad (30)$$

Taking into account that  $k_F$  in this case weights the higher order derivative it is the same result as in the case  $2m-1$ . Changing the factor  $k_F\sigma$  from the odd to the lower order even function gives  $(k_F\sigma)^{-1}$  and denoting the higher order by  $n$  (i.e.  $2m+1 =: n$ ) gives  $k_F^{-1} = \sqrt{(4m+1)/2} = \sqrt{(2m+1) - \frac{1}{2}}$ . This is the same expression as in the case  $2m-1$  where  $n = 2m$ . It turns out that:

$$k_F = k_L$$

### Quality of the odd/even pair

Table 4 shows the ratio of the energy at negative frequencies to the total energy ( $v/(u+v)$ ) in the notation of (29),(30).

order	2/1	2/3	4/3	4/5	6/5	6/7
$\frac{u}{u+v}$	0.0393411	0.0242336	0.0174848	0.0136703	0.0112203	0.0095142

Table 4: Energy at negative frequencies divided by the total energy for derivatives of Gaussian odd/even pairs

Figure 8 gives a qualitative impression of the performance of the complex filter using the relative weight  $k_F$  of the imaginary part. For  $D_{2x} + ik_F\sigma^{-1}D_{1x}$  the result is quite reasonable. In contrast, the energy for the functions where the odd part has the higher order ( $D_{2x} - ik_F\sigma D_{3x}$  and  $D_{4x} - ik_F\sigma D_{5x}$ ) is double peaked. This very likely causes false detections and localizations. We will investigate this in more detail in the next section. The phase of all functions is slightly dominated by the real part (flat slope at  $0, \pm\pi$ ). The phase is investigated in section 5.5.

### 5.3 Monomodal energy

The energy for odd/even pairs of derivatives of Gaussians tends to be **non-monomodal**. The energies that are depicted in figs 8 and 9 are the responses for a Dirac signal. Feature detection is done by evaluating the maxima in the energy of the response to a signal. Maxima that are caused by the function and not by the signal lead to false detections and localizations. In practice things are not as critical as they might seem from the figures because real signals usually are smooth and not Diracs and hence, the responses are smoothed too. Nevertheless, the detection and localization performance of a monomodal energy filter will be better than for a non-monomodal filter (e.g. with noise that consists of Dirac peaks). To investigate this quality criterion we calculate the derivative of the energy  $E$ .

$$E := \sqrt{D_{nx}^2 + k^2\sigma^{-2}D_{(n-1)x}^2}$$

We are interested in the extrema of  $E$  that are indicated by the zeros of the first derivative of  $E$  with respect to  $x$ . We can examine  $E^2$  as well, because  $(E^2)' = 2EE'$  and  $E$  has no zeros or poles (the prime denotes the derivative). We choose  $N = 1$  and  $a = 1, \sigma = 1/\sqrt{2}$  in (17) for convenience.

$$E^2 = e^{-2x^2} (H_n^2(x) + k^2 2H_{n-1}^2(x))$$

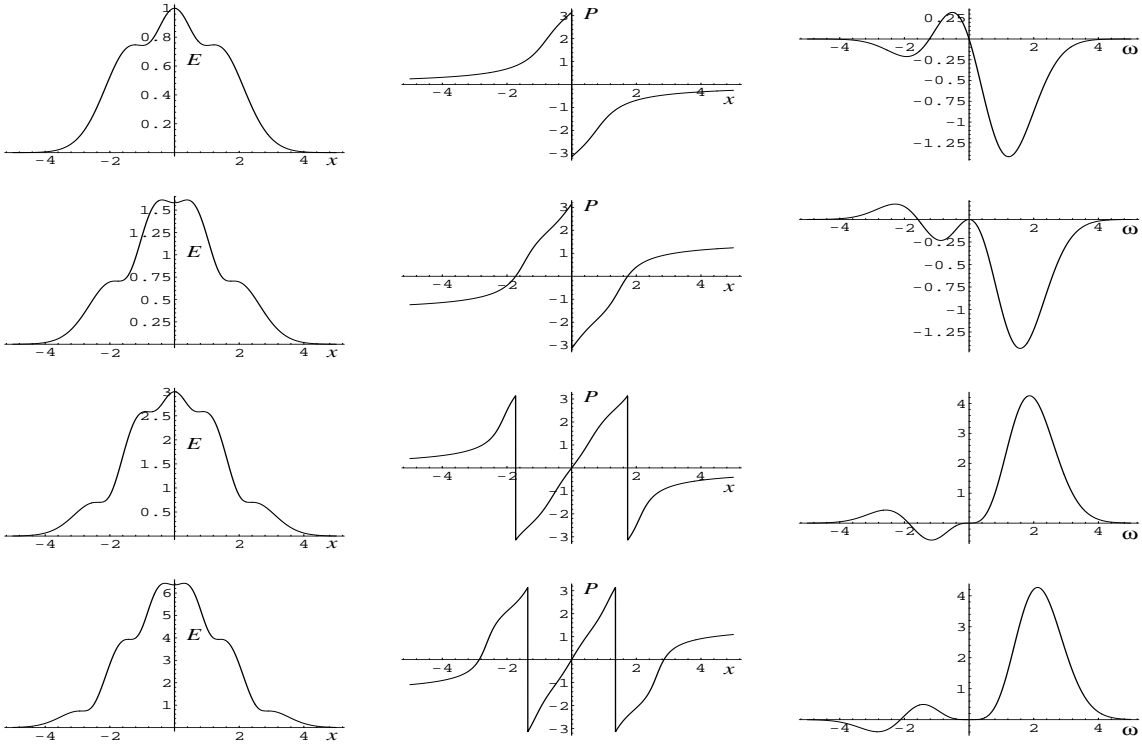


Figure 8: From top to bottom:  $D_{2x} + ik_F \sigma^{-1} D_{1x}$ ,  $D_{2x} - ik_F \sigma D_{3x}$ ,  $D_{4x} + ik_F \sigma^{-1} D_{3x}$ ,  $D_{4x} - ik_F \sigma D_{5x}$ .  $D_{nx}$  is from (4) with  $N = \sigma = 1$ . From left to right: energy (E) and phase (P) in position space, and energy in Fourier space.

$$\frac{dE^2}{dx} = e^{-2x^2} (-4xH_n^2 - 8xk^2H_{n-1}^2 + 2H_nH_n' + 4k^2H_{n-1}H_{n-1}')$$

Using the recurrence relations (58) for  $H_n'$  and (59) for  $H_{n-1}'$  we obtain

$$\frac{dE^2}{dx} = e^{-2x^2} (-4xH_n^2 + 4H_nH_{n-1}(n - k^2)) \quad (31)$$

For  $k^2 = n$  the second term vanishes and the derivative of the energy has only double zeros and for  $n$  odd a triple zero at  $x = 0$ . Note that according to our definition in (26) we actually dealt with  $k^{-1}$  in the case that  $n - 1$  is even. Having this in mind the optimal choice is:

$$k_E = \begin{cases} \sqrt{n} & n \text{ even} \\ 1/\sqrt{n} & n \text{ odd} \end{cases} \quad (32)$$

Then, except for  $x = 0$ , the energy has only saddle points but no extrema. From (31) we see that at  $x = 0$  we have for  $n$  even a first order maximum and for  $n$  odd a third order maximum. The third order maximum implies, that the peak is broad as can be seen in the middle of the second and fourth row of fig.9. For  $k^2 \neq n$  there are only maxima and minima and no saddle points (remember that  $H_{n-1}$  has maxima or minima at the zeros of  $H_n$ ). But the extrema at the slope of the energy are weak if  $k$  is not very different from  $k_E$ . The relative weight  $k = k_E$  is also suggested by Koenderink and van Doorn [33] but without giving a motivation for this choice.



Looking at fig.9, that shows the energy for different weights  $k$ , we derive the following preliminary conclusion: Giving the even part less weight than for  $k_E$  results in very poor energy shapes that have local minima at the center instead of maxima. This is the case if the higher order is odd and  $k_F = k_{L^2}$  is taken as the weight. **It is even preferable to give more weight to the even part than for  $k_E$ .** This pronounces the middle peak. If the higher order is odd the third order central maximum becomes a sharper first order maximum. The trade off is that the more the even part is emphasized the more the saddle points at the slope of the energy split in increasingly pronounced maxima and minima. If the higher order is even,  $k_{L^2}$  might be a good compromise. In the other case the even part should be more pronounced or at least equal than for  $k_E$ . Figure 9 shows the energy for different weights of the odd part.

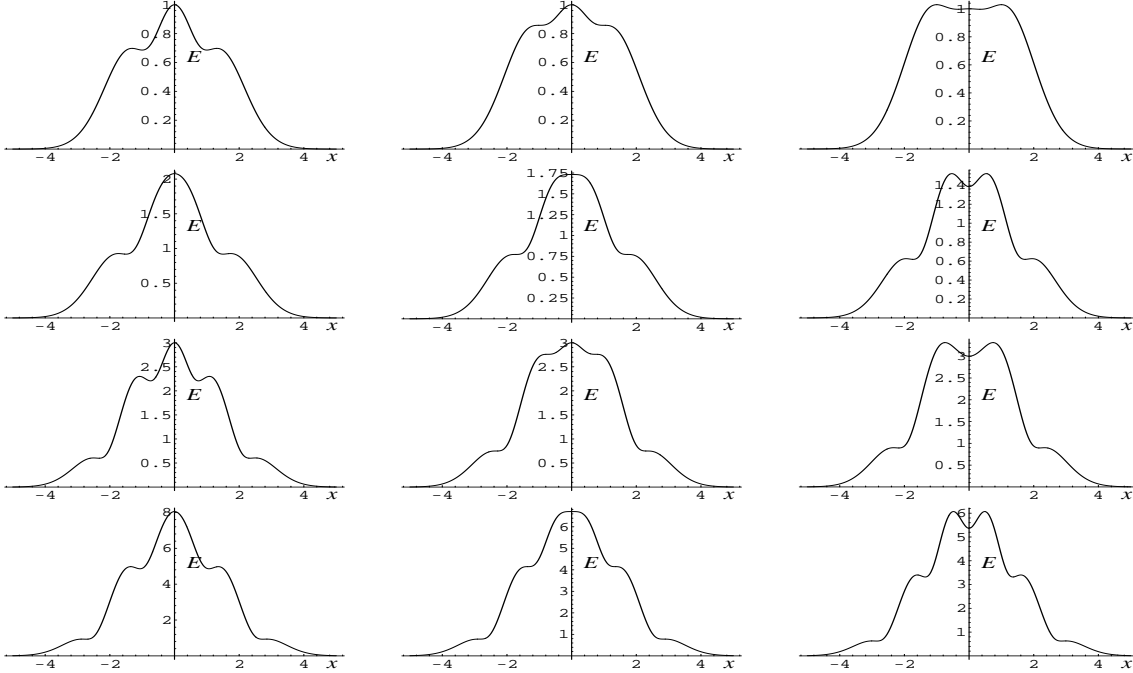


Figure 9: From top to bottom: energy for  $D_{2x} + ikD_{1x}$ ,  $D_{2x} - ikD_{3x}$ ,  $D_{4x} + ikD_{3x}$ ,  $D_{4x} - ikD_{5x}$ .  $D_{nx}$  is from (4) with  $N = \sigma = 1$ . From left to right:  $k = 0.8k_E$ ,  $k = k_E$ ,  $k = 1.2k_E$  if the odd part has lower order (upper and third row) and reverse otherwise.

#### 5.4 Fit of the energy to a Gaussian

The Gaussian function has optimal properties with respect to regularization and uncertainty as we explained earlier in sections 2.1 and 2.4. It therefore seems to be a natural requirement to fit the energy of the complex filter  $F$  from equation (26) as well as possible to a Gaussian. The energy is defined in (33) where we fixed the scale of the Gaussian derivatives to  $\sigma = 1/\sqrt{2}$ .

$$E := N \sqrt{D_{nx}^2 + 2k^2 D_{(n-1)x}^2} \quad (33)$$

$$N := \left( \sqrt{\pi/2} \{2n-1\} + \sqrt{2\pi} k^2 \{2n-3\} \right)^{-\frac{1}{2}}$$

$\{2n-1\}$  denotes the product  $1 \cdot 3 \dots (2n-1)$ . The derivatives  $D_{nx}, D_{(n-1)x}$  are taken from equation (17). Both functions have the same normalization constant  $N$ , such that the energy is  $L^2$  normalized to unity. We restrict the fit to normalized Gaussians  $L^2(D_0) = 1$

but the results are pretty much the same if the norm of the Gaussian is varied too. The normalized Gaussians and the  $L^2$ -error are given by:

$$D_0(x) := \sigma^{-\frac{1}{2}} \pi^{-\frac{1}{4}} e^{-\frac{x^2}{2\sigma^2}} \quad (34)$$

$$\text{error}(k, \sigma) := L^2(E - D_0) = \left( 2 - 2 \int ED_0 dx \right)^{\frac{1}{2}}$$

The error is minimized if the projection of  $E$  to  $D_0$  is maximized. For every order  $n$  we calculate  $\text{error}(k, \sigma)$  and search for the maximum. It turns out that this criterion is not very decisive because the quality of the fit is almost independent of  $k$ . However, the best fit is achieved for weights  $k$  that are approximately equal to  $k_{L^2}$ . In the case that the even function has the higher order this choice is quite reasonable. On the other hand, if the odd function has the higher order the energy function is double peaked with a minimum at the center what does not seem to be an appropriate choice. Table 5 and figure 10 show the results. With  $k_G$  we denote the weight that optimizes the fit to a Gaussian. For the case that the odd function has the higher order we also show the results of the best fit for  $k = k_{P_1}$ .  $k_{P_1}$  will be introduced in the next section and avoids the double peaked energy. The results show that the  $L^2$  error is not very different for this choice, even though the two weights are quite different. Notice that the scale of the optimal fitting Gaussian converges to  $\sigma = 1$ , what is the scale of the approximating Gabor functions for higher order Gaussian derivatives (see section 6.3).

order	$k$	$\sigma$	$L^2$ error in %
2/1	$k_G = 1.27$	$\sigma = 1.13$	11.4%
2/3	$k_G = 0.63$	$\sigma = 1.05$	10.9%
2/3	$k_{P_1} = 0.53$	$\sigma = 1.06$	12.4%
4/3	$k_G = 1.87$	$\sigma = 1.03$	9.3%
4/5	$k_G = 0.47$	$\sigma = 1.02$	8.2%
4/5	$k_{P_1} = 0.43$	$\sigma = 1.02$	8.7%

Table 5: Results for optimal Gaussian fits to the energy. The weight  $k$  with the best fit is denoted  $k_G$ .

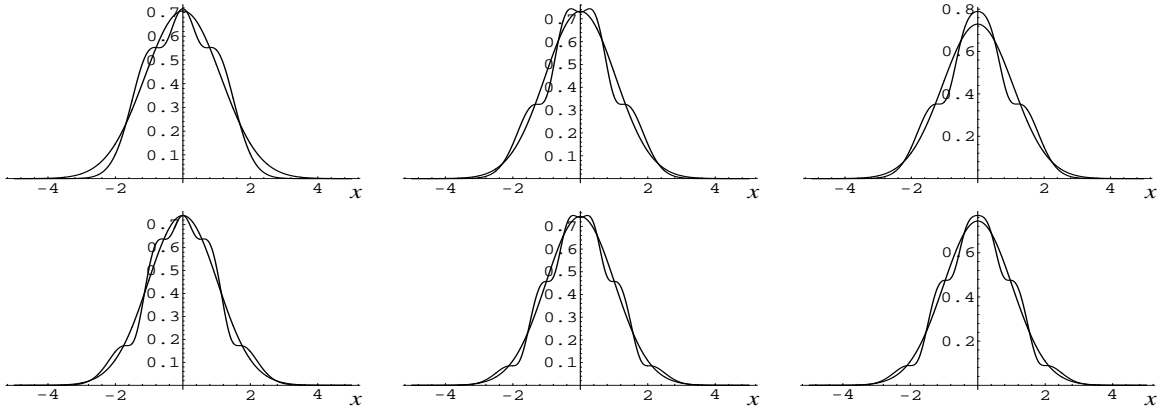


Figure 10: Optimal Gaussian fits to the energy of complex derivative of Gaussian filters. Depicted are the energy and the optimal fitting Gaussian for the following functions. The upper row depicts from left to right:  $D_{2x} + ik_G D_{1x}$ ,  $D_{2x} - ik_G D_{3x}$ ,  $D_{2x} - ik_{P_1} D_{3x}$ . In the lower row are depicted:  $D_{4x} + ik_G D_{3x}$ ,  $D_{4x} - ik_G D_{5x}$ ,  $D_{4x} - ik_{P_1} D_{5x}$ .

## 5.5 Linear phase

The phase in position space should be approximately linear at the center of the function. The phase is used for classifying the shape of the local signal structure (line, edge, etc.). The motivation for linear phase is robustness against noise, shifts and small deformations of the signal. It is defined by:

$$P := \mp \arctan\left(\frac{k\sigma^{\pm 1}D_{2m\pm 1}}{D_{2m}}\right) \quad (35)$$

For linear phase, the phase is a direct measure of the shift between a delta-event and the center of the basis function. In the extreme case nonlinear phase means flat plateaus with jumps between the plateaus. This is depicted in Fig.11 that shows examples with dominating real or imaginary part. A jumping phase contains almost no information, because the slightest deformation of a signal causes it to be classified according to the phase value at the plateaus.. In this section the phase is handled without wrap around at  $\pm\pi$  because otherwise the derivative of the phase is meaningless at these points.

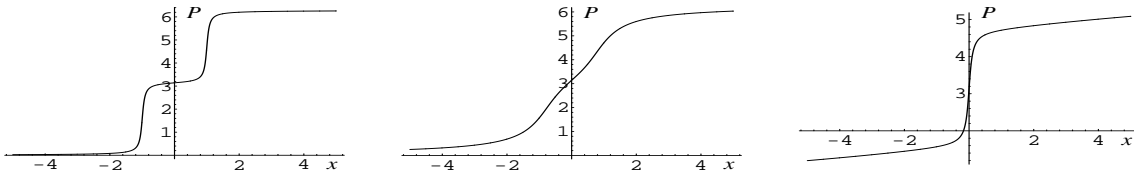


Figure 11: Phase of  $D_{2x} + ikD_{1x}$  with  $k = 0.1k_F$ ,  $k = k_F$ , and  $k = 10k_F$  (from left to right),  $\sigma = 1$ . For small or large  $k$  there's no robust classification possible because the phase 'jumps' between flat plateaus.

The theoretically optimal linear phase can be defined as follows: The phase difference between the first zero  $x_1$  of the even part and  $x = 0$  is  $\pi/2$  and the total phase difference is  $n\pi$  if  $n$  is the higher order of the odd and even part. Hence, the optimal slope  $S_{\text{opt}}$  of the phase is  $\pi/(2x_1)$  between  $-nx_1$  and  $+nx_1$ , and 0 outside. This is depicted in Fig.12. This definition could be made also with the last zero instead of the first or some mean value but we are interested first of all in an optimal phase near the center.

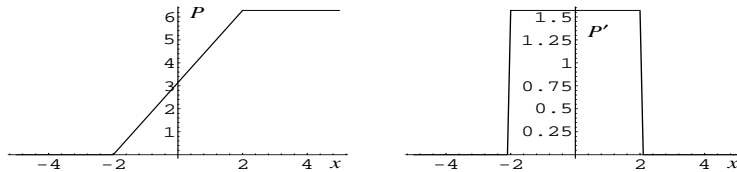


Figure 12: Theoretically optimal phase for  $D_{2x} + ikD_{1x}$ ,  $\sigma = 1$  (left) and its derivative (right).

The figures 13-16 show the derivative of the phase of the lower order odd/even pairs for some values of  $k$ . With increasing order the influence of the various normalizations decreases rapidly. On the other hand, the main differences are at the center of the function, where they are the most decisive.

There is no canonical measure for the linearity of the phase. Therefore, we investigated several different measures to clear up their influence on the results. In the following we show two of these measures to demonstrate the robustness of the qualitative results against the different ad hoc criteria. The optimal relative weights  $k_{P_1}$  and  $k_{P_2}$  minimize the following expressions  $I_1(k)$  and  $I_2(k)$ :

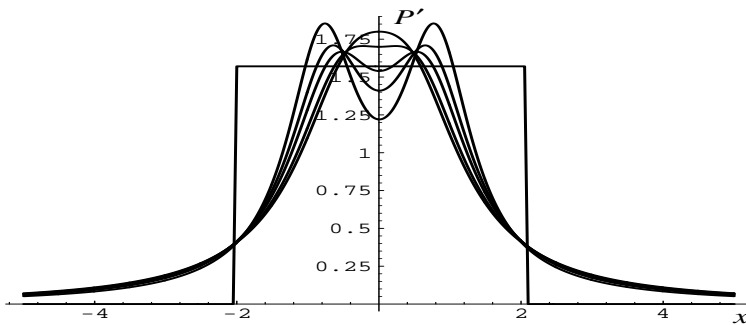


Figure 13: Derivative of the phase  $P'$  of  $D_{2x} + ikD_{1x}$  for  $k_F = 1.22, k_E = 1.41, k_{P_1} = 1.54, k_{P_3} = 1.7, k = 1.8$  and the theoretically optimal phase for comparison (smaller  $k$  have lower central minima).

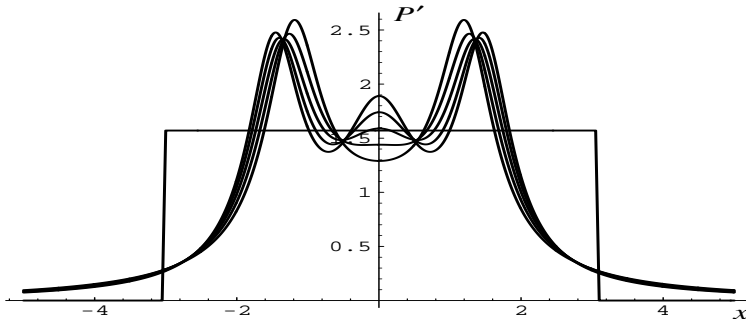


Figure 14: Derivative of the phase  $P'$  of  $D_{2x} - ikD_{3x}$  for  $k_F = 0.63, k_E = 0.58, k_{P_1} = 0.53, k_{P_3} = 0.48, k = 0.43$  and the theoretically optimal phase for comparison (larger  $k$  have larger central maxima).

$$I_1(k) = \int_{-\infty}^{\infty} \left| \arctan\left(\frac{\partial P}{\partial x}\right) - \arctan(S_{\text{opt}}) \right| e^{-\frac{x^2}{\sigma^2}} dx \quad (36)$$

$$I_2(k) = \int_{-x_1}^{x_1} \left| \frac{\partial P}{\partial x} - \bar{S} \right|, \quad \text{with } \bar{S} := \frac{1}{2x_1} \int_{-x_1}^{x_1} \frac{\partial P}{\partial x} dx \quad (37)$$

$I_1(k)$  tries to fit the phase to the optimal phase  $S_{\text{opt}}$  whereas  $I_2(k)$  tries to keep the derivative of the phase as flat as possible in the region between the first zeros  $x_1$  of the even part.  $\bar{S}$  is the mean derivative in this interval. The arctan function is introduced not to over emphasize large deviations of  $P'$  from  $S_{\text{opt}}$ , i.e. the averaging is with respect to the difference in angle and not in gradient because the latter is singular at  $90^\circ$ . The exponential in  $I_1$  weights the phase values with respect to their importance according to the energy.

We want to mention also a third criterion that might seem reasonable at the first sight. From figs. 13-16 we see that for a certain  $k$  the minimum of  $P'$  at the centre turns into a maximum. We choose  $k_{P_3}$  to be the one where  $P''$  has a third order zero at the minimum and hence  $P'$  is maximally flat. But it turns out that this linearity measure will be disadvantageous for most purposes for the following two reasons: (1) The region where the phase is linear is small and outside this region the linearity is worse than for  $k_{P_1}$  and  $k_{P_2}$ . (2) The energy is unreasonable for this choice.

The linearity measures are rather ad hoc but it turns out that the exact choice of the measure is not very critical. This gives more weight to the results than for the case that they depend on the accidental choice of the measure. Especially, for all three linearity

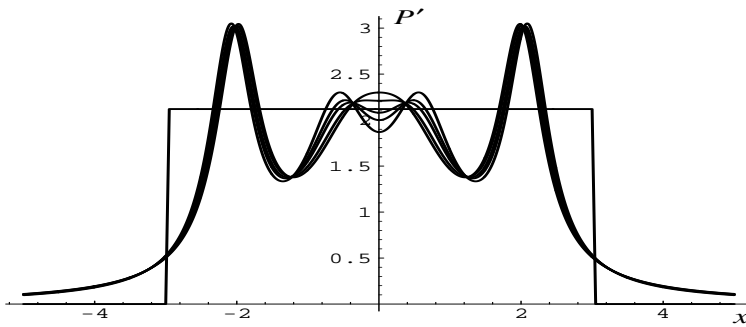


Figure 15: Derivative of the phase  $P'$  of  $D_{4x} + ikD_{3x}$  for  $k_F = 1.87, k_E = 2.0, k_{P_1} = 2.08, k_{P_3} = 2.21, k = 2.3$  and the theoretically optimal phase for comparison (smaller  $k$  have lower central minima).

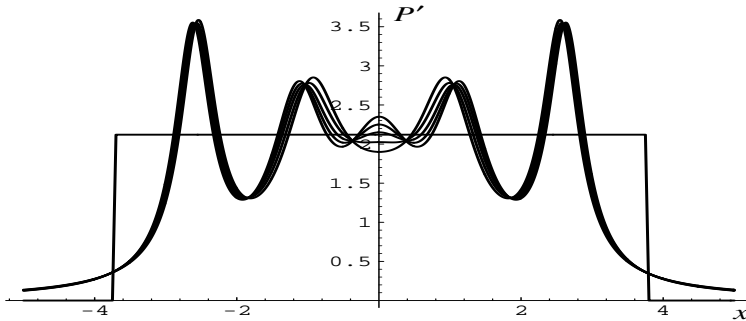


Figure 16: Derivative of the phase  $P'$  of  $D_{4x} - ikD_{5x}$  for  $k_F = 0.47, k_E = 0.45, k_{P_1} = 0.43, k_{P_3} = 0.405, k = 0.38$  and the theoretically optimal phase for comparison (larger  $k$  have larger central maxima).

measures it **confirms that optimal linear phase is derived with more weight on the higher order part than for the previous criteria.** Table 6 in the following conclusion shows the optimal  $k_P$ 's derived from (36) and (37).

## 5.6 Conclusion

In summary we achieved the following results: The weights from the  $L^1$  normalization do not differ much from those of the  $L^2$  normalization. The criterion of minimal Fourier energy at negative frequencies gives the same weights as the  $L^2$  normalization. Hence, there are essentially 3 different weights  $k_F, k_E,$  and  $k_{P_1}$ , where  $k_E$  always is between the other two. If the even function has the higher order  $k_F$  gives more weight and  $k_{P_1}$  gives less weight to the even part than  $k_E$ . If the odd function has the higher order it is vice versa. An overview of the different weights is given in table 6.

Now, what is altogether the best  $k$ ? Remember that we did no investigations including signals and noise and surely the answer is task and signal dependent. But for a common detection, localization, and classification task we would give the following advice: Mostly the detection and localization is more important and moreover the quality of the phase is relatively robust compared to the quality of the energy. Therefore, we exhaust the possible detection and localization performance by optimizing the energy. For this we give more or at least as much weight to the even part as for  $k_E$ . If the higher order is even,  $k_F$  might be a good compromise at the cost of a poorer quality of the phase. If the higher order is odd we choose  $k_{P_1}$  at the cost of more Fourier energy at negative frequencies and at the cost of different  $L^1$  and  $L^2$  normalizations of the odd and even part. But in any case the

order	2/1	2/3	4/3	4/5
$k_{L^1}$	1.21	0.64	1.85	0.47
$k_F = k_{L^2}$	1.22	0.63	1.87	0.47
$k_G$	1.27	0.63	1.87	0.47
$k_E$	1.41	0.58	2.0	0.45
$k_{P_1}$	1.52	0.53	2.06	0.43
$k_{P_2}$	1.52	0.53	2.08	0.43
$k_{P_3}$	1.7	0.48	2.21	0.405

Table 6: Optimal  $k$  according to the different criteria.  $k_{P_2}$  is included to demonstrate the small difference between the linearity measures.

deviation from  $k_E$  can't be very large because otherwise the energy gets undesired strong side extrema.

## 6 Relationship between derivatives of Gaussians and Gabor functions

### 6.1 Definition and normalization of Gabor functions

The Gabor functions are defined as:

$$\begin{aligned} G_c(x) &:= N_c e^{-\frac{x^2}{2\sigma^2}} \cos\left(\frac{cx}{\sigma}\right) \\ G_s(x) &:= N_s e^{-\frac{x^2}{2\sigma^2}} \sin\left(\frac{cx}{\sigma}\right) \end{aligned} \tag{38}$$

$N_c$  and  $N_s$  are normalization constants. The Fourier transform is:

$$\begin{aligned} \hat{G}_c(\omega) &= N_c \frac{\sigma}{2} \left( e^{-\frac{(\omega\sigma-c)^2}{2}} + e^{-\frac{(\omega\sigma+c)^2}{2}} \right) \\ \hat{G}_s(\omega) &= N_s \frac{\sigma}{2} \left( e^{-\frac{(\omega\sigma-c)^2}{2}} - e^{-\frac{(\omega\sigma+c)^2}{2}} \right) \end{aligned} \tag{39}$$

Usually the Gabor functions are applied without normalization or with a normalized Gaussian. However, we are interested in  $L^1$  or  $L^2$  normalized Gabor functions.

#### $L^2$ normalization

The  $L^2$  norm is calculated in Fourier space. Using (39) and the integrals from (54) we obtain:

$$\begin{aligned} L^2(G_c) &= \left( \int_{-\infty}^{\infty} \hat{G}_c^2(\omega) d\omega \right)^{\frac{1}{2}} = N_c \sqrt{\frac{1}{2} \sigma \sqrt{\pi} (1 + e^{-c^2})} \\ L^2(G_s) &= \left( \int_{-\infty}^{\infty} \hat{G}_s^2(\omega) d\omega \right)^{\frac{1}{2}} = N_s \sqrt{\frac{1}{2} \sigma \sqrt{\pi} (1 - e^{-c^2})} \end{aligned} \tag{40}$$

#### $L^1$ normalization

The  $L^1$  norm is calculated numerically:

$$\begin{aligned}
L^1(G_c) &= N_c \sigma \int_{-\infty}^{\infty} |e^{-\frac{z^2}{2}} \cos(cz)| dz =: N_c \sigma I_c(c) \\
L^1(G_s) &= N_s \sigma \int_{-\infty}^{\infty} |e^{-\frac{z^2}{2}} \sin(cz)| dz =: N_s \sigma I_s(c)
\end{aligned} \tag{41}$$

Figure 17 shows  $I_c(c)$ ,  $I_s(c)$ , and their quotients with the  $L^2$  norm from (40) for  $\sigma = 1$  ( $\sqrt{2}I_{c/s}(c)/\pi^{\frac{1}{4}}(1 \pm e^{-c^2})^{\frac{1}{2}}$ ). The quotient reveals the difference between the two normalizations. It turns out that both norms are approximately independent of  $c$  above  $c \approx 2$ . Papers that are interested in local frequency analysis or texture classification usually don't use small  $c$ . In contrast we are interested also in values of  $c$  below 2.

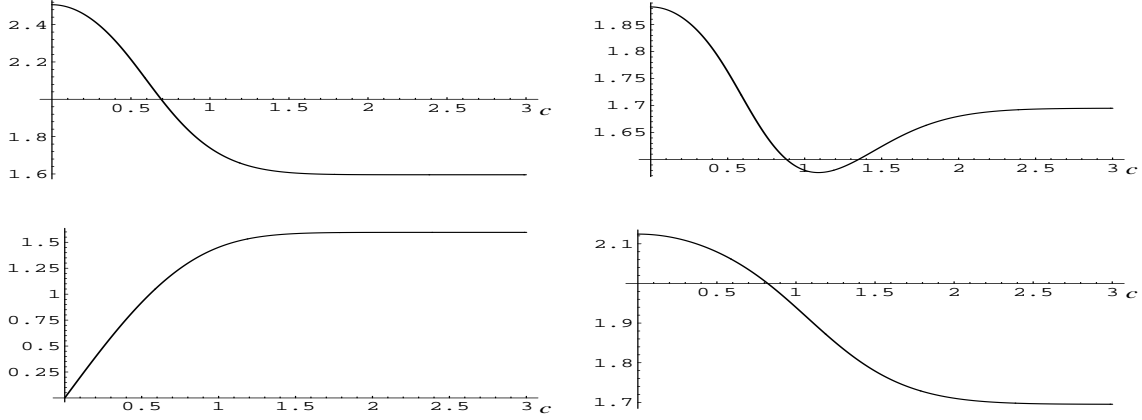


Figure 17:  $I_c(c)$  (upper left) and  $I_c(c)$  divided by the  $L^2$  norm (see text) (upper right).  $I_s(c)$  (lower left) and  $I_s(c)$  divided by the  $L^2$  norm (lower right). Beside a constant factor of about 1.695 the two norms differ by at most 7% for reasonable values of  $c$  above 1.

## 6.2 Asymptotic behavior of the derivatives of Gaussians

The asymptotic behavior of the derivatives of Gaussians  $D_{nx}$  with growing order  $n$  is given by the following equation. It is derived from the textbook of Magnus et al. [43, 5.6.3] by changing from  $He_n$  to  $H_n$  by  $He_{2n}(x\sqrt{2}) = 2^{-n}H_{2n}(x)$ .

$$\begin{aligned}
H_{2n}(x) &= (-1)^n 2^{2n} (2n-1)! e^{\frac{x^2}{2}} \left[ \cos(\sqrt{4n+1}x) + O(n^{-\frac{1}{4}}) \right] \\
H_{2n+1}(x) &= (-1)^n 2^{2n+1} (2n-1)! (2n+1) e^{\frac{x^2}{2}} \left[ \sin(\sqrt{4n+3}x) + O(n^{-\frac{1}{4}}) \right]
\end{aligned} \tag{42}$$

From (17) we see that:

$$\begin{aligned}
D_{nx}(x) &= N (-1)^n 2^{-\frac{n}{2}} \sigma^{-n} H_n\left(\frac{x}{\sigma\sqrt{2}}\right) e^{-\frac{x^2}{2\sigma^2}} \\
&\rightarrow_{n \rightarrow \infty} \begin{cases} N (-1)^m 2^m (n-1)! \sigma^{-n} e^{-\frac{x^2}{4\sigma^2}} \cos(\sqrt{2n+1} \frac{x}{\sqrt{2}\sigma}) , & n = 2m \\ N (-1)^{m+1} 2^{\frac{n}{2}} n (n-2)! \sigma^{-n} e^{-\frac{x^2}{4\sigma^2}} \sin(\sqrt{2n+1} \frac{x}{\sqrt{2}\sigma}) , & n = 2m+1 \end{cases}
\end{aligned} \tag{43}$$

This means that for larger orders  $n$  the derivatives of Gaussians look like Gabor functions. For  $n$  odd they look like Gabor sine functions, for  $n$  even they look like Gabor

cosine functions. If the derivatives of Gaussians have the scale  $\sigma$ , the Gabor functions have the scale  $\sqrt{2}\sigma$  and an oscillation parameter  $c = \sqrt{2n+1}$ . The error decreases rapidly with  $O(n^{-1/4})$ . Therefore, only for the lower orders we expect deviations between Gabor functions and derivatives of Gaussians worth mentioning. In the following subsection we investigate the quality of the fit for the orders  $n = 1, 2, 3$ .

### 6.3 Fit of Gabor functions to derivatives of Gaussians

In this section we investigate the optimal  $L^2$  fit for the lower orders, where the approximation of (43) does not hold.

The derivatives of Gaussians are normalized to  $L^2(D_{nx}) = 1$  according to (21) with a scale  $\sigma = 1/\sqrt{2}$ :

$$D_{nx}(x) = \frac{(-1)^n 2^{\frac{1}{4}}}{\sqrt{1 \cdot 3 \dots (2n-1) \sqrt{\pi}}} H_n(x) e^{-x^2} \quad (44)$$

We restrict the fits to the case that the Gabor functions are normalized to  $L^2(G) = 1$  as well:

$$\begin{aligned} G_c &= \sqrt{\frac{2}{\sigma \sqrt{\pi}(1+e^{-c^2})}} e^{-\frac{x^2}{2\sigma^2}} \cos\left(\frac{cx}{\sigma}\right) \\ G_s &= \sqrt{\frac{2}{\sigma \sqrt{\pi}(1-e^{-c^2})}} e^{-\frac{x^2}{2\sigma^2}} \sin\left(\frac{cx}{\sigma}\right) \end{aligned} \quad (45)$$

Now we are looking for the parameters  $c, \sigma$  that minimize  $L^2(D_{nx} \pm G_{c/s})$ , where  $G_s$  is used for  $n$  odd and  $G_c$  for  $n$  even. The sign  $\pm$  is  $+-+ \dots$  for the orders  $1, 2, 3, \dots$

$$\begin{aligned} L^2(D_{nx} \pm G) &= \sqrt{(L^2(D_{nx}))^2 + (L^2(G))^2 \pm 2 \int D_{nx} G dx} = \\ &= \sqrt{2 \pm 2 \int D_{nx} G dx} =: \sqrt{2 \pm 2A_n(c, \sigma)} \end{aligned}$$

$$A_n(c, \sigma) = \begin{cases} \frac{2^{\frac{3}{4}}}{\sqrt{\pi\{2n-1\}\sigma(1+e^{-c^2})}} \int e^{-(1+\frac{1}{2\sigma^2})x^2} H_n(x) \cos\left(\frac{cx}{\sigma}\right) dx, & n \text{ even} \\ -\frac{2^{\frac{3}{4}}}{\sqrt{\pi\{2n-1\}\sigma(1-e^{-c^2})}} \int e^{-(1+\frac{1}{2\sigma^2})x^2} H_n(x) \sin\left(\frac{cx}{\sigma}\right) dx, & n \text{ odd} \end{cases} \quad (46)$$

$\{2n-1\}$  denotes the product  $1 \cdot 3 \dots (2n-1)$ . The  $L^2$  error is  $\sqrt{2 \pm 2A_n}$  and for a perfect fit  $A_n$  would be 1. Using Mathematica we obtain:

**Case  $n = 1$**

$$A_1(c, \sigma) = 2^{\frac{9}{4}} c \sigma^{\frac{3}{2}} (1 - e^{-c^2})^{-\frac{1}{2}} e^{-\frac{c^2}{2+4\sigma^2}} (1 + 2\sigma^2)^{-\frac{3}{2}}$$

$$\frac{\partial A_1}{\partial \sigma} \stackrel{(!)}{=} 0 \implies \sigma = \frac{1}{\sqrt{6}} \sqrt{c^2 + \sqrt{c^4 + 9}} \quad (47)$$



According to (47)  $\sigma$  depends approximately linear on  $c$  if  $c$  is not too small. Figure 18a shows  $L^2(D_{1x} + G_s(c, \sigma))$  with  $\sigma = \sigma(c)$  according to (47). The optimal fit is for  $c \rightarrow 0$  and  $\sigma \rightarrow 1/\sqrt{2}$ , where  $A_1$  approaches 1. This can be understood by considering the Taylor expansion of  $\sin(cx/\sigma)$ . For very small  $c$  the first term in the Taylor expansion  $cx/\sigma$  suffices and it has the same functional form than the first order Hermite polynomial. Hence, for small  $c$  the Gabor function perfectly fits  $D_{1x}$ . On the other hand, for small  $c$  the amplitude of the sine function goes to zero whereas the normalization constant in (45) goes to infinity. This is quite unnatural and Fig. 18a shows that the fit is not much worse for larger  $c$ . For  $c = 1$  and  $\sigma = 0.833$  the  $L^2$  error is about 1.5%. Figure 18b shows the remaining error of the fit.

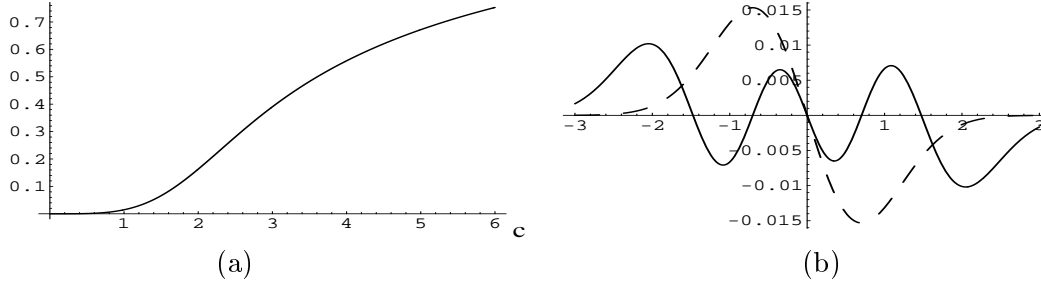


Figure 18: (a) shows  $L^2(D_{1x} + G_s(c, \sigma(c)))$  in the range of  $c \in [0, 6]$ . For  $c \rightarrow 0$   $D_{1x}$  and  $G_s$  become identical. The dependence on  $c$  is not very strong. (b) shows  $D_{1x} + G_s$  for  $D_{1x}$ :  $\sigma = 0.707$  and for  $G_s$ :  $\sigma = 0.833, c = 1$ . The amplitude of  $D_{1x} + G_s$  is just 1.3% of the amplitude of  $D_{1x}$ . Hence, there is no visible difference between  $D_{1x}$  and  $G_s$ . The function  $D_{1x}$  in (b) (dashed) is depicted only to illustrate the spatial scale. **It has not the true amplitude.**

### Comment to the paper of Mehrotra et al.

Mehrotra et al. [51] investigate the edge detection performance of Gabor sine functions according to Canny's criteria. The authors denote the Gabor function by  $G(x) = \exp(-x^2/(2\sigma^2) + i\omega x)$ , what is not very convenient. Choosing the parameter in the oscillatory part  $c/\sigma$  instead of  $\omega$  makes  $\sigma$  a true scale parameter and  $c$  a true shape parameter, whereas  $\omega$  mixes both things. Hence, the finding in [51] that  $\sigma\omega$  is the appropriate tuning parameter is evident, because  $\sigma\omega = c$ .

The authors then motivated as an optimal choice  $c \approx 1$ . The above results provide an easy understanding of the fact that the performance is close to  $D_{1x}$ . On the other hand, the low values of the quality parameter  $C$  in [51] (spacing between noise maxima) for small  $c$  are hard to believe, because for an appropriate choice of  $\sigma$  and  $c$  the results should become the same as for  $D_{1x}$ . We did not check the calculations in [51] in detail.

### Case $n = 2$

$$A_2(c, \sigma) = -2^{\frac{9}{4}} 3^{-\frac{1}{2}} (1 + e^{-c^2})^{-\frac{1}{2}} e^{-\frac{c^2}{2+4\sigma^2}} (1 + 2\sigma^2)^{-\frac{5}{2}} \sigma^{\frac{1}{2}} (1 + 2\sigma^2 + 2c^2\sigma^2)$$

$$\frac{\partial A_2}{\partial \sigma} \stackrel{!}{=} 0 \implies c^2 = \frac{20\sigma^4 - 4\sigma^2 - 7 + \sqrt{400\sigma^8 + 160\sigma^6 + 24\sigma^4 + 104\sigma^2 + 41}}{8\sigma^2} \quad (48)$$

This complex formula can be fitted very well by the simple linear dependence  $c \approx -0.26 + 2.37\sigma$  (see Fig. 19c). Figure 19a shows  $L^2(D_{2x} + G_c(c, \sigma))$  with  $c = c(\sigma)$  according

to (48). The optimal fit is for  $c = 2.29$  and  $\sigma = 1.073$ , where the  $L^2$  error is 10%. The Taylor expansion of the cosine function is  $\cos(cx/\sigma) \approx 1 - c^2 x^2 / (2\sigma^2)$  what is the 2nd order Hermite polynomial if  $c = 2\sigma$  (up to a factor 2). For  $\sigma = 1/\sqrt{2}$ , the scale we choose in (44), this results in  $c = \sqrt{2}$  what is indeed the solution of (48). But in contrast to the case  $n = 1$  this is not the overall optimal solution.

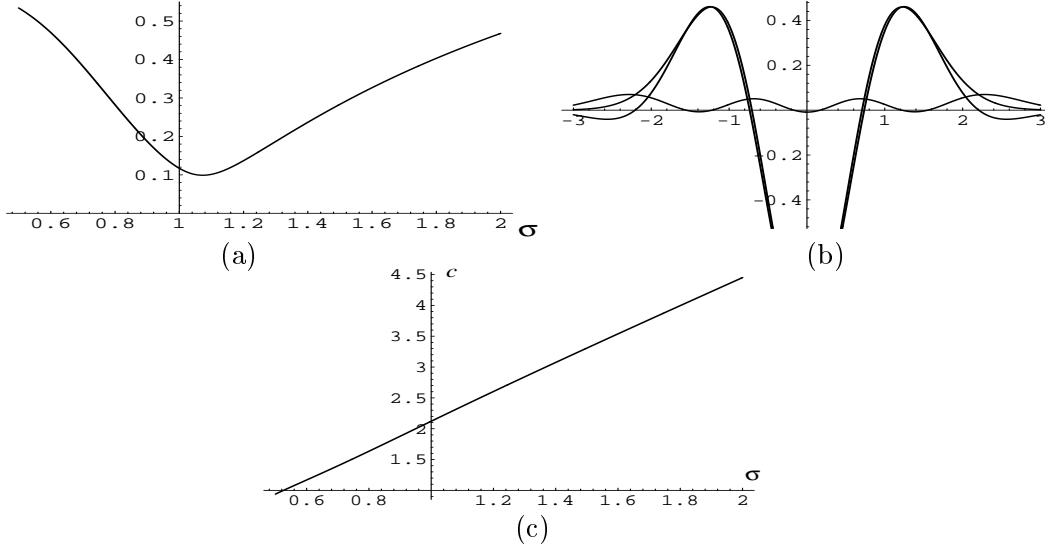


Figure 19: (a) shows  $L^2(D_{2x} + G_c(c(\sigma), \sigma))$  in the range of  $\sigma \in [0.5, 2.0]$ . (b) shows  $D_{2x}$  ( $\sigma = 0.707$ ),  $-G_c$  ( $\sigma = 1.073, c = 2.29$ ), and  $D_{2x} + G_c$ . (c) shows the approximate linearity of  $c(\sigma)$ .

### Case $n = 3$

The analytical calculations become rapidly more involved with higher orders. Hence, we present only numerical results in Fig. 20. Figure 20a shows  $L^2(D_{3x} - G_s(c, \sigma))$ . First we see that the dependence between  $c$  and  $\sigma$  is approximately linear. Second, the figure shows that the quality of the fit does not change much for variations of  $c$  and  $\sigma$  along the optimal curve, in contrast to variations orthogonal to this curve. The optimal fit is depicted in fig. 20b. The values for the optimal fit are:  $\sigma = 1.03, c = 2.62$  and the  $L^2$  error is about 6%.

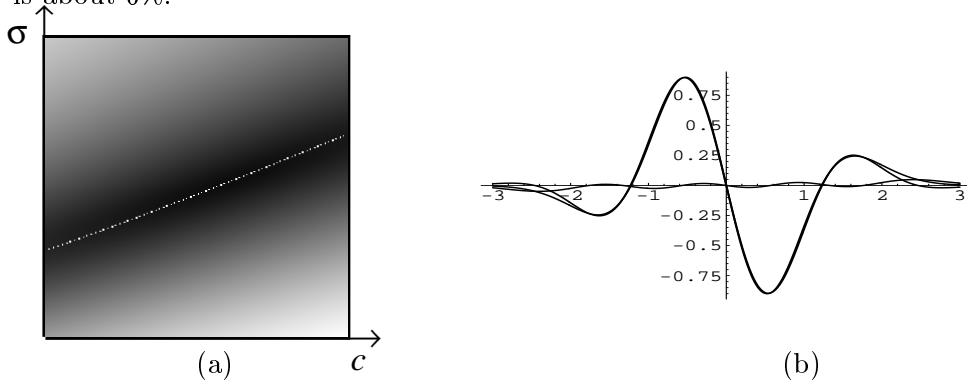


Figure 20: (a) shows  $L^2(D_{3x} - G_s(c, \sigma))$  in the range of  $c \in [2, 3]$  (small  $c$  left),  $\sigma \in [0.5, 1.5]$  (small  $\sigma$  bottom). The dotted line marks the ridge that corresponds to optimal  $\sigma$  for given  $c$ . (b) shows  $D_{3x}$  ( $\sigma = 0.707$ ),  $G_s$  ( $\sigma = 1.03, c = 2.62$ ), and  $D_{3x} - G_s$ .

### Case $n \geq 4$

For larger orders  $n$  the asymptotic behavior ( $\sigma = 1$ ,  $c = \sqrt{2n+1}$ ) gives a very good approximation of the optimal fit.

## 6.4 Conclusion

For higher orders the derivatives of Gaussians can be approximated almost perfectly by Gabor functions. From the lower order functions  $D_{1x}$  can be approximated perfectly too. The worst fit is for  $D_{2x}$  where the  $L^2$  error is 10%. For  $D_{3x}$  the error is 6%. When high precision is required as in some differential geometric methods the error for the lower orders is too large to use the Gabor functions instead of the derivatives of Gaussians. But for other more qualitative applications the fit is quite well. Note that we did the fit with the true Gabor functions. In practice a DC-corrected version has to be used for the lower even orders instead (section 2.2).

## 7 Summary

We have provided strong motivations that the first processing step for all common early vision tasks is adequately done by projecting the image to a set of basis functions. We discussed the basic principles that determine the choice of the basis functions. But in contrast to some papers that overemphasize one particular principle we also pointed to their limitations. We derived some interesting results on the construction of optimal odd/even pairs from successive orders of derivatives of Gaussians and on the relation to the Gabor functions. It turns out that higher derivatives (and the first derivative as well) can be approximated almost perfectly by Gabor functions and there is no difference to the true Hilbert transforms or DC-, and ramp-corrected functions. In this case all principles (uncertainty, scale-space behavior, energy/phase etc.) can be satisfied simultaneously and no compromises are necessary. On the other hand, for weakly oscillating functions there exists no optimal solution that satisfies all requirements. This is especially true for the second and third derivatives that are important for edge and line detection. Hence, we investigated this case in detail to understand the limitations and to provide the criteria to find the best compromise for a given task.

## Appendix A: Scale-space and the diffusion equation

All fundamental basis functions are based more or less on the Gaussian. This is not too surprisingly as the Gaussian plays an exceptional role in many parts of mathematics. One strong justification for the use of the Gaussian in computer vision is its relation to the concept of scale. This relation has been investigated by Babaud et al. [4], Yuille and Poggio [78], and Koenderink [32]. All three papers have slightly different premises and statements but essentially they present the same results.

### Babaud et al.

The paper treats the 1D case and defines scaling by the convolution of a signal  $f(x)$  with a kernel  $g_s(x) := sg(sx)$ , where  $g$  satisfies some natural conditions. The convolution is denoted by  $\Phi(x, s) := f(x) * g_s(x)$ . The central proposition states that if for **first order** extrema (i.e.  $\partial_x \Phi = 0, \partial_{xx} \Phi \neq 0$ ) the *monotonicity condition*  $\partial_s \Phi \partial_{xx} \Phi < 0$  is valid the kernel must be a Gaussian. This implies that local maxima only decrease towards larger scales and local minima only increase.

The connection to the diffusion equation is made in the sufficiency part of the proof. The monotonicity condition is valid for all  $\Phi$  that are solutions of the diffusion equation

$$\frac{\partial^2 \Phi}{\partial x^2} = \frac{\partial \Phi}{\partial s} \quad (49)$$

The Gaussian is the Green's function of the diffusion equation and hence the solutions  $\Phi$  are convolutions of the source term (the signal) by a Gaussian. For the diffusion equation to have the form (49) the Gaussian is denoted:

$$g_s(x) = (4\pi s)^{-\frac{1}{2}} e^{-\frac{x^2}{4s}} \quad (50)$$

With the usual form of the Gaussian, where the denominator of the exponent is  $2\sigma^2$  instead of  $4s$  the diffusion equation looks slightly different, but the results are the same. The two forms are connected by a transformation of the scale parameter  $2s = \sigma^2$ .

The authors point out that the monotonicity condition is not possible for higher order extrema. This is one hint that the importance of this scale-space behavior is sometimes over-emphasized. The argument usually is that signals can be analysed by their extrema (or level crossings, see below) only if the extrema can be traced back to finest scales to give the connection to the signal. But there always will be extrema for which this is not the case and moreover the methods to analyse the signal and the interpretation of extrema and other events are by no means unique.

From the monotonicity condition follows a result for zero-crossings too. We are interested in the zero-crossings of an arbitrary linear differential operator. Such operators commute with the Gaussian convolution and hence it suffices to investigate zero-crossings of  $\Phi$  itself. The result is that zero-crossing contours in scale-space are always locally parabolic and concave up, i.e. zero-crossings never disappear towards finer scales.

As an outlook the 'dumbbell' example is given that shows that in 2D zero-crossings are free to split and merge with increasing scale even though they are never created from nothing. A simple tree-structure of the events as in the 1D case is not possible.

### Yuille and Poggio

Yuille and Poggio generalize the results of Babaud et al. to two dimensions. As in the paper of Babaud et al. signals are investigated that are scaled by convolution with a kernel. The basic theorem is stated not for extrema but for zero-crossings of the Laplacian and shows that the Gaussian (isotropic or elongated) is the only kernel that does not create new zero-crossings. The result is then generalized for level crossings of any differential operator

$L$  that commutes with the diffusion equation, i.e. events of the type  $L(g * f) = \text{const}$  are never created towards larger scales. As already mentioned, in 2D a split and merge is possible.

It turns out that the heat equation is the only one with the required scaling behavior of its solutions. Other filters are expected to have this scaling behavior only for scales where they are approximate solutions of the heat equation. E.g. the difference of Gaussians and the Gabor functions for scales that are not too small.

### Koenderink

Koenderink starts with a function  $\Phi(x, y, s)$  with the boundary condition that for scale 0 it is the original image  $\Phi(x, y, 0) = f(x, y)$ . It is not required (in the beginning) that  $\Phi$  results from a convolution. Koenderink requires that level crossings  $\Phi(x, y, s) = \text{const}$  can always be traced back to finer scales. Then,  $\Phi$  has to be a solution of the 2D diffusion equation.

## Appendix B: Hermite transform

The Hermite functions themselves are not very important in computer vision. The interest in these functions stems from their close connection to the derivatives of Gaussians. If the projection of a signal to the latter is viewed as the transformation of a Gaussian windowed signal the Hermite functions are the dual functions for the reconstruction of the original signal. It has been shown that the derivatives of Gaussians are not too different from the Gabor functions but with the advantage that basis functions of different orders are mutually orthogonal for the same window. Though, in contrast to orthogonal wavelet (subband) coders, basis functions for different windows have some overlap.

The interplay of the derivatives of Gaussians and the Hermite functions in the Hermite transform is explained in the following equations. All equations are modulo some normalizing coefficients. More details can be found in [49]. The Hermite transform is a windowed transform, i.e. we distinguish a window function in addition to the analysing function and the reconstruction function.

Let  $f(x)$  be the signal and  $V(x) = \sigma^{-\frac{1}{2}} e^{-\frac{x^2}{2\sigma^2}}$  the Gaussian window that selects a weighted region of interest of the signal. Then the Hermite decomposition and reconstruction is given by the following two formulas:

$$\begin{aligned} f(x) &= \sum_n \sum_k c_n(k) P_n(x - k\Delta x) W^{-1}(x) \\ c_n(k) &= \int f(x) D_n(x - k\Delta x) dx \end{aligned} \quad (51)$$

$W(x) := \sum_k V(x - k\Delta x)$  is the weighting function.  $D_n(-x) = H_n(x/\sigma) V^2(x) = \partial_x^n V^2(x)$  are the analysis functions which is the  $n$ 'th order derivative of a Gaussian.  $P_n(x) = H_n(x/\sigma) V(x)$  are the reconstruction functions, the  $n$ 'th order Hermite functions.  $H_n(x)$  is the  $n$ 'th order Hermite polynomial. In the above notation  $D_n$  is the derivative of the square of the Gaussian in its usual form but this results only in some different constants and a rescaling of  $\sigma$ . The reconstruction functions are orthogonal:

$$\int P_m(x) P_n(x) dx = \delta_{mn} \quad (52)$$

The sum over  $k$  in equation (51) is the sum over all shifted non-orthogonal Gaussian windows. Within each window the sum over  $n$  is an orthogonal expansion. The expansion coefficients  $c_n$  can therefore easily be obtained by a projection to the analysing functions  $D_n$ . This step is painful for the non-orthogonal Gabor functions.

## Appendix C: Integrals of the Gaussian

Throughout the text integrals of polynomials times Gaussians are needed. For the non-definite integrals there is no simple analytical expression but mostly the following definite integrals are sufficient [9] ( $a > 0$ ):

$$\int_0^{\infty} x^n e^{-ax^2} dx = \begin{cases} \frac{1 \cdot 3 \dots (2k-1)\sqrt{\pi}}{2^{k+1}a^{k+\frac{1}{2}}}, & n = 2k \\ \frac{k!}{2a^{k+1}}, & n = 2k+1 \end{cases} \quad (53)$$

For the first orders and  $a = (2\sigma^2)^{-1}$  the explicit forms are:

$$\begin{aligned} \int_0^{\infty} e^{-ax^2} dx &= \frac{\sqrt{\pi}}{2\sqrt{a}} = \frac{\sqrt{\pi}}{\sqrt{2}} \sigma \\ \int_0^{\infty} xe^{-ax^2} dx &= \frac{1}{2a} = \sigma^2 \\ \int_0^{\infty} x^2 e^{-ax^2} dx &= \frac{\sqrt{\pi}}{4a^{\frac{3}{2}}} = \frac{\sqrt{\pi}}{\sqrt{2}} \sigma^3 \\ \int_0^{\infty} x^3 e^{-ax^2} dx &= \frac{1}{2a^2} = 2\sigma^4 \\ \int_0^{\infty} x^4 e^{-ax^2} dx &= \frac{3\sqrt{\pi}}{8a^{\frac{5}{2}}} = \frac{3\sqrt{\pi}}{\sqrt{2}} \sigma^5 \end{aligned} \quad (54)$$

## Appendix D: Hermite polynomials

The Hermite polynomials are defined to be the orthogonal polynomials with respect to a Gaussian weighted scalar product [43]:

$$\int_{-\infty}^{\infty} H_m(x)H_n(x)e^{-x^2} dx = \begin{cases} 0 & , \quad m \neq n \\ \sqrt{\pi} 2^n n! & , \quad m = n \end{cases} \quad (55)$$

To calculate these polynomials Rodrigues' formula can be used:

$$H_n(z) = (-1)^n e^{z^2} \frac{d^n e^{-z^2}}{dz^n} \quad (56)$$

The explicit form of the first orders and the general form is:

$$\begin{aligned} H_0(z) &= 1 \\ H_1(z) &= 2z \\ H_2(z) &= 4z^2 - 2 \\ H_3(z) &= 8z^3 - 12z \\ H_4(z) &= 16z^4 - 48z^2 + 12 \\ H_5(z) &= 32z^5 - 160z^3 + 120z \\ H_n(z) &= n! \sum_{m=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^m}{m!(n-2m)!} (2z)^{n-2m} \end{aligned} \quad (57)$$

We also need the recurrence relations (the prime denotes the derivative):

$$H'_n(z) = 2nH_{n-1}(z) \tag{58}$$

$$H_{n+1}(z) = 2zH_n(z) - H'_n(z) \tag{59}$$

$$2zH_n(z) = H_{n+1}(z) + 2nH_{n-1}(z) \tag{60}$$

## References

- [1] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, Dover Publications, New York, 1972.
- [2] E.H. Adelson and J.R. Bergen, *Spatiotemporal energy models for the perception of motion*, JOSA A, vol.2, no. 2, 284-299, 1985.
- [3] M.T. Andersson, *Controllable multidimensional filters and models in low level computer vision*, PhD Thesis, Linköping University, S-58183 Linköping, Sweden, Diss. no. 282, ISBN 91-7870-981-4, 1992.
- [4] J. Babaud, A.P. Witkin, M. Baudin and R.O. Duda, *Uniqueness of the Gaussian kernel for scale-space filtering*, IEEE PAMI-8, no.1, 26-33, 1986.
- [5] E. Barth, T. Caelli and C. Zetsche, *Image encoding, labeling, and reconstruction from differential geometry*, CVGIP GMIP, vol.55, no.6, 428-446, 1993.
- [6] M.J. Bastiaans, *Gabor's expansion of a signal into Gaussian elementary signals*, Proc. IEEE, vol.68, No.4, 538-539, 1980.
- [7] A.C. Bovik, M. Clark and W.S. Geisler, *Multichannel texture analysis using localized spatial filters*, IEEE PAMI-12, no.1, 55-73, 1990.
- [8] R. Bracewell, *The Fourier Transform and Its Applications*, McGraw-Hill, New York, 1965.
- [9] I.N. Bronstein and K.A. Semendjajew, *Taschenbuch der Mathematik*, Harri Deutsch, Thun, 1985.
- [10] J.M.H. du Buf, *Gabor phase in texture discrimination*, Signal Processing, vol.21, 221-240, 1990.
- [11] J.M.H. du Buf and P. Heitkaemper, *Texture features based on Gabor phase*, Signal Processing, vol.23, 227-244, 1991.
- [12] J.M.H. du Buf, *Responses of simple cells: events, interferences, and ambiguities*, Biol. Cybernetics, vol.68, 321-333, 1993.
- [13] P.J. Burt and E.H. Adelson, *The Laplacian pyramid as a compact image code*, IEEE COM-31, no.4, 532-540, 1983.
- [14] J. Canny, *A computational approach to edge detection*, IEEE PAMI-8, no.6, 679-698, 1986.
- [15] I. Daubechies, S. Jaffard, J.-L. Journe, *A simple wilson orthonormal basis with exponential decay*, SIAM J.Math.Anal., vol.22, no.5, 554-572, 1991.
- [16] J.G. Daugman, *Two-dimensional spectral analysis of cortical receptive field profiles*, Vision Research, vol.20, 847-856, 1980.
- [17] J.G. Daugman, *Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two-dimensional visual cortical filters*, JOSA A, vol.2, no.7, 1160-1169, 1985.
- [18] J.G. Daugman, *Complete discrete 2D Gabor transform by neural networks for image analysis and compression*, IEEE ASSP-36, no.7, 1169-1179, 1988.



- [19] J.G. Daugman, *Quadrature-phase simple-cell pairs are appropriately described in complex analytic form*, JOSA A, vol.10, no.2, 375-377, 1993.
- [20] E.R. Davies, *Design of optimal Gaussian operators in small neighbourhoods*, Image and Vision Computing, vol.5, no.3, 199-205, 1987.
- [21] D.J. Fleet, *Measurement of image velocity*, Kluwer Academic Press Publishers, Boston, 1992.
- [22] D.J. Fleet and A.D. Jepson, *Stability of phase information*, IEEE PAMI-15, no.12, 1253-1268, 1994.
- [23] L. Florack, *The syntactical structure of scalar images*, PhD thesis, University of Utrecht, Netherlands, 1993.
- [24] W.T. Freeman and E.H. Adelson, *The design and use of steerable filters for image analysis*, IEEE PAMI-13, no.9, 891-906, 1991.
- [25] D. Gabor, *Theory of communication*, IEE Part III, vol.93, no.26, 429-457, 1946.
- [26] L. Haglund, *Adaptive multidimensional filtering*, PhD Thesis, Linköping University, S-58183 Linköping, Sweden, Diss. no. 284, 1992.
- [27] M. Hashimoto and J. Sklansky, *Multiple-order derivatives for detecting local image characteristics*, CVGIP, vol.39, 28-55, 1987.
- [28] D.J. Heeger, *Optical flow using spatiotemporal filters*, Int. J. of Computer Vision, 279-302, 1988.
- [29] F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans, and O. Kübler, *Simulation of neural contour mechanisms: from simple to end-stopped cells*, Vision Research, vol.32, no.5, 963-981, 1992.
- [30] R. Hummel and D. Lowe, *Computational considerations in convolution and feature-extraction in images*, in J.C. Simon (ed.) From Pixels to Features, Elsevier Science Publishers L B.V.(North Holland), 91-102, 1989.
- [31] S.A. Klein and B. Beutner, *Minimizing and maximizing the joint space-spatial frequency uncertainty of Gabor-like functions: comment*, JOSA, vol.9, no.2, 337-340, 1992.
- [32] J.J. Koenderink, *The structure of images*, Biol. Cybernetics, vol.50, 363-370, 1984.
- [33] J.J. Koenderink and A.J. van Doorn, *Representation of local geometry in the visual system*, Biol. Cybernetics, vol.55, 367-375, 1987.
- [34] J.J. Koenderink and A.J. van Doorn, *Generic neighborhood operators*, IEEE PAMI-14, no.6, 597-605, 1992.
- [35] H. Knutsson, *Filtering and reconstruction in image processing*, PhD thesis, Univ. of Linköping, Sweden, 1982.
- [36] H. Knutsson, R. Wilson, and G.H. Granlund, *Anisotropic nonstationary image estimation and its applications: Part I - restoration of noisy images*, IEEE COM-31, no.3, 388-397, 1983.

- [37] R. Kronland-Martinet, J. Morlet and A. Grossmann, *Analysis of sound patterns through wavelet transforms*, Int. J. of Pat. Rec. and Art. Intelligence, vol.1, no.2, 273-302, 1987.
- [38] P. Kube, *Properties of energy edge detectors*, CVPR 92, 586-591, 1992.
- [39] P. Kube and P. Perona, *Scale-space properties of quadratic edge detectors*, ECCV 94, Vol.I, 115-122, 1994.
- [40] M. Lades et al., *Distortion invariant object recognition in the dynamic link architecture*, IEEE Computers, Vol.42, No.3, 300-311, 1993.
- [41] T. Lindeberg, *Scale-space for discrete signals*, IEEE PAMI-12, no.3, 234-254, 1990.
- [42] T. Lindeberg, *Discrete derivative approximations with scale-space properties: A basis for low-level feature extraction*, J. of Math. Imaging and Vision, vol.3, 349-376, 1993.
- [43] W. Magnus, F. Oberhettinger, and R.P. Soni, *Formulas and Theorems for the Special Functions of Mathematics and Physics*, Springer, 1966.
- [44] S.G. Mallat, *A Theory for multiresolution signal decomposition: The wavelet representation*, IEEE PAMI-11, no.7, 674-693, 1989.
- [45] S.G. Mallat, *Multifrequency channel decompositions of images and wavelet models*, IEEE ASSP-37, no.12, 2091-2110, 1989.
- [46] B.S. Manjunath and R. Chellappa, *A unified approach to boundary perception: edges, textures, and illusory contours*, IEEE NN-4, no.1, 96-107, 1993.
- [47] S. Marcelja, *Mathematical descriptions of the responses of cortical cells*, JOSA A, vol.70, 1297-1300, 1980.
- [48] D. Marr and E. Hildreth, *Theory of edge detection*, Proc. Royal Society London B 207, 187-217, 1980.
- [49] J.-B. Martens, *The Hermite transform - theory*, IEEE ASSP-38, no.9, 1595-1606, 1990.
- [50] J.-B. Martens, *The Hermite transform - applications*, IEEE ASSP-38, no.9, 1607-1618, 1990.
- [51] R. Mehrotra, K.R. Namuduri and N. Ranganathan, *Gabor filter-based edge detection*, Pattern Recognition, vol.25, no.12, 1479-1494, 1992.
- [52] M. Michaelis and G. Sommer, *Using wavelets for feature detection*, in: Theoretical Foundations of Computer Vision, R. Klette and W. Kropatsch (Eds.), 109-118, Mathematical Research vol.69, Akademie Verlag, Berlin, 1993.
- [53] M. Michaelis, S. Perz, C. Black and G. Sommer, *Detection and classification of P waves using Gabor wavelets*, IEEE Computers in Cardiology, London, 531-534, 1993.
- [54] M. Michaelis and G. Sommer, *Keypoint characterization in images*, Proceedings of the SPIE, vol.2093, 450-461, 1994.
- [55] M. Michaelis and G. Sommer, *Junction classification by multiple orientation detection*, ECCV'94, Stockholm, Vol.I, 101-108, 1994 (extended version available).

- [56] M.C. Morrone and R.A. Owens, *Feature detection from local energy*, Pattern Recognition Letters, vol.6, no.5, 303-313, 1987.
- [57] M.C. Morrone and D.C. Burr, *Feature detection in human vision: a phase-dependent energy model*, Proc. R. Soc. London B, vol.235, 221-245, 1988.
- [58] R.A. Owens, *Feature-free images*, Pattern Recognition Letters, vol.15, 35-44, 1994.
- [59] P. Perona and J. Malik, *Detecting and localizing edges composed of steps, peaks and roofs*, UC Berkeley, Tech. Rep. UCB/CSD 90/590, 1990.
- [60] D.A. Pollen and S.F. Ronner, *Visual cortical neurons as localized spatial frequency filters*, IEEE SMC-13, no.5, 907-916, 1983.
- [61] C. Ronse, *On idempotence and related requirements in edge detection*, IEEE PAMI-15, no.5, 484-491, 1993.
- [62] L. Rosenthaler, F. Heitger, O. Kübler and R. von der Heydt, *Detection of general edges and keypoints*, ECCV 92, 78-86, 1992.
- [63] T.D. Sanger, *Stereo disparity computation using Gabor filters*, Biol. Cybernetics, vol.59, 405-418, 1988.
- [64] E. Simoncelli, W.T. Freeman, E.H. Adelson, D.J. Heeger, *Shiftable multi-scale transforms*, MIT Tech. Rep. no.161, 1991.
- [65] G. Sommer, *Bivariate local signal representations*, in: Theoretical Foundations of Computer Vision, R. Klette and W. Kropatsch (Eds.), 119-130, Mathematical Research vol.69, Akademie Verlag, Berlin, 1993.
- [66] D.G. Stork and H.R. Wilson, *Do Gabor functions provide appropriate descriptions of visual receptive fields?*, JOSA A, vol.7, no.8, 1362-1373, 1990.
- [67] G. Szegö, *Orthogonal Polynomials*, American Mathematical Society, 1959.
- [68] V. Torre and T.A. Poggio, *On edge detection*, IEEE PAMI-8, no.2, 147-163, 1986.
- [69] M.R. Turner, *Texture discrimination by Gabor functions*, Biol. Cybernetics, vol.55, 71-82, 1986.
- [70] S. Venkatesh, J. Cooper and B. White, *Local energy, the pre-envelope, and filter resolution*, ICPR 92, vol.III, 13-17, 1992.
- [71] S. Venkatesh S. and R. Owens, *On the classification of image features*, Pattern Recognition Letters, vol.11, 339-349, 1990.
- [72] R. Wilson and G. Granlund, *Uncertainty principle in image processing*, IEEE PAMI-6, no.6, 758-767, 1984.
- [73] R. Wilson and H. Knutsson, *Uncertainty and inference in the visual system*, IEEE SMC-18, no.2, 305-312, 1988.
- [74] R. Wilson and A.H. Bhalerao, *Kernel design for efficient multiresolution edge detection and orientation estimation*, IEEE PAMI-14, no.3, 384-390, 1992.
- [75] A.P. Witkin, *Scale-space filtering*, 8th Int. Joint Conf. on Art. Intelligence, 1019-1022, 1983.

- [76] J. Yang, *Do Gabor functions provide appropriate descriptions of visual cortical receptive fields?: comment*, JOSA A, vol.9, no.2, 334-336, 1992.
- [77] R.A. Young, *The Gaussian derivative theory of spatial vision: Analysis of cortical cell receptive field line-weighting profiles*, General Motors Research Publication GMR-4920, 1985.
- [78] A.L. Yuille and T.A. Poggio, *Scaling theorems for zero crossings*, IEEE PAMI-8, no.1, 15-25, 1986.