On Higher Dimensional Multiscale Edge Extraction

Ji-Young Lim

Academic advisor:
Prof. Dr.-Ing. H. Siegfried Stiehl

Universität Hamburg
Fachbereich Informatik
Arbeitsbereich Kognitive Systeme

March, 2002

Abstract

Edge extraction is one of the key issues in image analysis and computer vision. In particular, multiscale approaches to edge extraction have proven to reveal important information about edges in an image. In this report, we first review up to recently published multiscale edge extraction approaches as well as nonlinear diffusion and wavelet approaches. As a result from our in-depth literature study on edge extraction, we identify as major open problems in higher dimensional edge extraction i) the development of appropriate higher dimensional edge models, ii) the effect of curvature as related to scale, and iii) optimal scale selection. We then propose a theoretical framework for optimal scale selection in higher dimensional edge extraction based on higher dimensional edge models.

Zusammenfassung

Die Extraktion von Kanten aus digitalen Bildern kann als eines der zentralen Forschungsprobleme in den Gebieten Bildanalyse und Computer Vision betrachtet werden. Insbesondere liefern dabei sog. Multiskalen-Verfahren zur Kantenextraktion wichtige Informationen über Kanten in einem Bild. In diesem Bericht werden zunächst jüngere Arbeiten zur Multiskalen-Kantenextraktion sowie Ansätze basierend auf nichtlinearer Diffusion und Wavelets aufgearbeitet und bewertet. Als offene Probleme wurden dabei identifiziert: die Entwicklung mehrdimensionaler Kantenmodelle, die Untersuchung des Einflusses der Kantenkrümmung im Skalenraum sowie die Entwicklung von Verfahren zur Bestimmung der jeweils optimalen Skala. Anschließend wird ein neuartiger Ansatz zur Bestimmung der optimalen Skala für mehrdimensionale Kantenmodelle entwickelt.

Contents

1 Introduction

Edge extraction is one of the key issues in image analysis and computer vision. The goal of edge extraction is to obtain a complete and meaningful description from an image by characterizing intensity changes which take place at various spatial scales depending on their physical origin. A great effort has been devoted to edge extraction, and various approaches have been reported in the extensive literature over a few decades.

In multiscale edge extraction approaches based on the *linear scale-space theory*, the behavior of edges across scales is analyzed, which can reveal precious information about the nature of the underlying physical process that gives rise to edges, or, synonymously, intensity variations in the image ([2], [3], [7], [11], [12], [14], [16], [17], [21], [27]). Though smoothing greatly reduces the effect of random noise, it can also smooth across edges, which may be unwanted for the purpose of extracting edges. Hence, *nonlinear multiscale approaches* have been developed in order to preserve (or even enhance) edges ([8], [31], [39]). Besides, the *wavelet representation* as another type of multiscale representation is worth noting ([23], [26]), since the linear scale-space representation can be considered as a special case of the continuous wavelet representation. These three approaches, i.e. linear, nonlinear, and wavelet approaches deal with the matter of extracting edges based on a multiscale analysis. Therefore, it is worthwhile to take a glance at these approaches in order to get an idea of the main characteristics of each approach with respect to improving key problems of edge extraction in various ways.

As a result from our in-depth literature study on edge extraction, we can identify as major open problems in higher dimensional edge extraction i) the development of appropriate higher dimensional edge models, ii) the effect of curvature as related to scale, and iii) optimal scale selection.

This report is organized as follows: We first survey in Section 2 up to recently published multiscale edge extraction approaches. Moreover, in Section 3 we give a review of both the nonlinear diffusion approach and the wavelet approach. Then in Section 4, we establish higher dimensional edge models for higher dimensional multiscale edge extraction, and based on them we propose a theoretical framework for optimal scale selection in higher dimensional edge extraction in Section 5. Finally, we summarize and present our future work in Section 6.

2 Linear Multiscale Approaches to Edge Extraction

The fundamentals of edge extraction as well as seminal approaches to edge extraction were surveyed in [18]. Moreover, we reviewed in [18] the principles of the linear scale-space theory as well as key multiscale approaches to edge extraction. In this section, we chronologically look over recently published important multiscale edge extraction approaches (particularly as related to optimal scale selection) in order to gain an overview of the development of multiscale edge extraction. Note that for the sake of brevity, details will be omitted but main ideas of each work are highlighted and shortcomings are analyzed.

Jeong and Kim [16] suggested a regularization method for edge extraction, where they estimated a unique scale adaptively for each point in the image. Assuming $\sigma(x, y)$ to be the scale parameter of the Gaussian, the authors introduced an energy function $E(\sigma)$ defined as a functional developed over continuous scale-space, and considered the process of selecting a set of optimal scales for edge extraction as the minimization of the energy function. Consequently, the edge extraction problem reduces to finding $\sigma(x, y)$ such that $E(\sigma)$ is minimized. Jeong and Kim used a parallel relaxation algorithm in order to solve the resulting nonconvex optimization problem. In their approach, however, the result suffered from the complications associated with the objective function and the selected scales were sensitive to the initial guess.

Gökmen and Jain [14] introduced an image and surface representation which samples an image in both the scale space and the continuity space for the purpose of decomposing the image/surface into a larger number of descriptions. Based on this representation, they developed "a generalized edge detector" by specifying the values of two parameters, one of which controls the shape of the filter (τ) whereas the other controls the scale of the filter (λ). As a result of their approach, an edge representation is generated in the scale-continuity ($\lambda\tau$) space. In their work, however, the determination of the optimal values of parameters λ and τ for a given image and for specific tasks remained unsolved.

Lindeberg [20], [21] proposed an automatic method for selecting locally appropriate scales for edge extraction based on the normalized derivative operator; maxima in the convolution of the signal with the Gaussian over scales of normalized derivatives reflect the scales over which spatial variations take place in the signal. The normalized derivative operator was introduced in order i) to compensate for the generally observed decrease in amplitude of image intensity caused by linear scale-space smoothing, and ii) to ensure that image structures are processed by the vision system in such a way that the processing results are not critically

dependent upon how large the image structures actually are. Since the scale of a structure selected by this approach is often too small to provide reliable estimates of the derivatives, Lindeberg suggested a more global post-extraction stage, where a measure of edge significance is defined and integrated along connected chains of pixels in order to distinguish real edges from artifacts. Only edge chains with a significant measure above a given threshold are then considered as important edges. However, the determination of this threshold is unspecific, and this post-extraction process is relatively tentative and heuristic.

Elder [11] and Elder and Zucker [12] presented a method of local scale adaptation based on the statistical reliability for detecting and localizing local edges in images regardless of their physical origin (e.g., occlusion, shadows, or textures). The authors did not restrict the goal of the local computation to the extraction of a specific type of edge (e.g., occlusion edges) only, because different types of edges are locally indistinguishable by local computation. Thus the goal of the local computation was set to detect, localize, and characterize *all* edges over broad scales, regardless of the underlying structures from which they project. They showed that knowledge of sensor properties (i.e. the second moment of the sensor noise) and operator norms can be exploited to define a unique, locally computable "minimum reliable scale" for local estimation at each point in the image, which led to a method for estimating the local blur of image contours. However, two main problems in their approach can be identified. First, the authors used the second-order derivative for avoiding the main defect of the gradient operator (e.g. multiple separable responses to a single edge). As a matter of fact, however, it is not guaranteed in practice that zero-crossings of the second-order derivative are well localized, and for the worse the higher-order derivatives (e.g., the second- and third-order derivative) are generally more sensitive to noise and computationally more expensive than the gradient operator. Second, the accuracy of their approach is dependent on two parameters, i.e., one is the standard deviation value of the sensor noise and the other is the value of the overall significance level for a reliability criterion. However, it is hard to measure correctly the sensor noise in real images, and the overall significance level parameter is assumed by the authors without any plausible grounding.

3 Other Multiscale Approaches to Edge Extraction

In this section, we review other approaches, i.e., nonlinear diffusion approaches and wavelet approaches, related to multiscale edge extraction.

3.1 Nonlinear Diffusion Approach

Linear scale-space approaches are based on blurring with the Gaussian kernel, and the related theory provides a well founded formalism for early visual computation and applications (e.g. for feature extraction). However, there exist some limitations in using an rotationally symmetric Gaussian kernel for a vision system; for example, smoothing across object boundaries may result in an undesirable effect on the shape of the edge profile and on the accuracy in localization of edges in edge extraction. The main idea of nonlinear diffusion approaches to edge extraction is to preserve and even enhance edges while smoothing out spurious detail off edges. In Weickert [37], one can find a nice tutorial-like introduction to nonlinear diffusion. To make this chapter self-contained, we briefly recapitulate the main developments and applications of nonlinear diffusion approaches.

Perona and Malik [31] first proposed a nonlinear edge-preserving diffusion method. Perona and Malik (the PM approach) considered an anisotropic diffusion equation by varying the conductance parameter over space and time

$$\begin{aligned}\frac{\partial L}{\partial t} &= \text{div}\ (c(x,y,t)\nabla L) \\ &= c(x,y,t)\nabla^2 L + \nabla c \cdot \nabla L,\end{aligned}$$

where $L(x,y;t)$ is obtained by convolving the original image with a Gaussian kernel, and $c(x,y,t)$ corresponds to the diffusion coefficient function. It is noticeable that if $c(x,y,t)$ remains constant the equation above reduces to the isotropic heat equation, since $\nabla c(\cdot) = 0$. The diffusion coefficient function in the PM approach is chosen locally as a function g of the magnitude of the gradient, i.e.,

$$c(x,y,t) = g\left(\|\nabla L\|\right). \tag{3.1}$$

The conductance $g(\cdot)$ is a positive, decreasing, and nonlinear function, and plays a role in limiting blurring near edges (i.e., preserving as well as sharpening edges). The authors suggested two kinds of the conductance $g(\|\nabla L\|)$:

$$g(\|\nabla L\|) = e^{-\left(\frac{\|\nabla L\|}{\lambda}\right)^2}$$

and

$$g(\|\nabla L\|) = \frac{1}{1 + \left(\frac{\|\nabla L\|}{\lambda}\right)^2} \qquad (\lambda > 0),$$

where the constant λ is fixed for controlling the effect of a given gradient value. The first conductance function emphasizes high contrast edges over low contrast ones, while the second gives a privilege to wide regions over smaller ones. The experimental results of the PM approach showed that edges remained stable over a very long diffusion time. However, the PM approach has a few problems (see for detail e.g. [8], [37], or [38]). First, in case of strongly noisy images, the gradient tends to oscillate (it is well known that the derivative operators are not robust against noise), and these oscillations are kept throughout the nonlinear diffusion process. Therefore, they proposed to perform a low-pass filtering prior to the diffusion process in order to reduce these oscillations, which in turn caused a loss of accuracy in edge localization. Second, the PM approach explicitly intended the forward-backward diffusion for the desirable result of blurring small fluctuations and sharpening edges. However, it is well-known that backward diffusion is an ill-posed process, where the solution, if any, is highly sensitive even to the slightest perturbations of the initial data.

Catté *et al.* [8] proposed to replace $g(\|\nabla L\|)$ of Eq. 3.1 by $g(\|\nabla L_\sigma\|)$ given by

$$\frac{\partial L}{\partial t} = \text{div } (g(\|\nabla L_\sigma\|)\nabla L),$$

where $L_\sigma = G_\sigma * L$ and G_σ denotes the linear Gaussian blurring at each step of the non-uniform diffusion process for the purpose of a reliable gradient estimate at each time step t. In the approach by Catté *et al.*, smoothing the image before the diffusion process is not necessary, and the Gaussian kernel guarantees that the edge-controlled diffusion is a monotonically decreasing process. On the other hand, Whitaker and Pizer [40] suggested a multiscale diffusion technique which makes the scale parameter σ for the gradient measurement time-dependent, i.e.,

$$\frac{\partial L}{\partial t} = \text{div } (g(\|\nabla L_{\sigma(t)}\|)\nabla L).$$

From a practical point of view, the approaches by Catté *et al.* and by Whitaker and Pizer offer the advantage of making the filter rather insensitive to noise. This aspect eliminates the drawback of the PM approach which misinterprets strong oscillations due to noise as edges that should be preserved or even enhanced.

Weickert [39] introduced an image restoration technique based on a multiscale method in which a nonlinear diffusion filter is steered by the so-called interest operator (e.g., second-order moment matrix, structure tensor) for enhancing line-like structures. In relation to coherence-enhancing anisotropic diffusion filtering, a new algorithm which uses the first-order derivative filters optimized with respect to the best gradient direction estimation was introduced in [32].

In general, the nonlinear diffusion approach takes higher order differential invariant properties of images into account, leading to the preservation and even enhancement of edges. However, the differential approaches are often only well defined in two dimensions and are of considerably higher computational cost.

3.2 Wavelet Approach

Wavelets are a mathematical tool for hierarchically decomposing functions, and they have shown great potential and applicability in many fields. Wavelets are functions that satisfy certain requirements ([36]): They should integrate to zero, which implies that the function has to be well localized, and also both the direct and inverse wavelet transform should exist. As a mathematical expression, a wavelet is a function $\psi(x) \in L^2$ such that

$$\int_{-\infty}^{+\infty} \psi(x) = 0.$$

The dilation of $\psi(x)$ by a factor s, which controls the wavelet scale, is given by

$$\psi_s(x) = \frac{1}{s}\psi\left(\frac{x}{s}\right).$$

The wavelet transformation with scale s of a function $f(x)$ at position x is given by

$$W_s f(x) = f * \psi_s(x).$$

As can be seen, the wavelet transform is a linear operation convolving the signal with a dilated filter. Such a decomposition has been extensively studied in signal processing and computer vision (see for detail e.g. [24], [34], or [36]).

It is noticeable that, in general, wavelet analysis differs from Fourier analysis in the fact that Fourier basis functions are localized in the frequency domain but not in the spatial domain, and vice versa (for example, even small changes in the Fourier domain will produce changes everywhere in the spatial domain after having applied the inverse Fourier transform), whereas wavelets are local both in the frequency/scale (via dilation) domain and in the spatial

(via translation) domain. On account of this localization property, it is possible to analyze data at different scales or resolutions much better than by the Fourier transform. We skip here the mathematical fundamentals of the wavelet theory, since one can find a detailed introduction in [24].

In the context of computer vision, wavelets have been used with enormous success in data compression and image noise suppression. The simplest wavelet basis is the Haar wavelet, where the basis function is the characteristic function of the interval $[0, 1)$ (see for details e.g., [36]). An obvious disadvantage of the Haar wavelet is that the Haar wavelet transform is not continuous, and therefore the choice of the Haar basis for representing fine-scaled images is not appropriate. As a more general framework, Mallat's multiresolution analysis (MRA) [22] can be cited. The MRA is a tool for a constructive description of different wavelet bases, where the basis function is chosen in order to satisfy some continuity, smoothness, and tail requirements. Additionally, the family of the basis functions forms an orthonormal basis for the reference space. It was shown in [22] that the difference of information between the approximation of a signal at scales 2^{j+1} and 2^j can be extracted by decomposing the signal on a wavelet orthonormal basis of $L^2(\mathbb{R}^n)$. In $L^2(\mathbb{R})$, a wavelet orthonormal basis is a family of functions $\left(\sqrt{2^j}\psi(2^j x - n)\right)_{(j,n)\in\mathbb{Z}^2}$ that is built by dilating and translating a unique function ψ. This decomposition defines an orthonormal multiresolution representation, which is called a "wavelet representation". A wavelet representation of a function consists of a coarse overall approximation together with detail-related coefficients that influence the function at various scales, and it can be applied, for example, to data compression in image coding.

Mallat proposed to apply a wavelet transform to edge extraction. He used the Laplacian of the Gaussian as a wavelet for the wavelet transform. The edge positions were assumed to be zero-crossings of a wavelet transform, and it was shown that 1-D signals can be well reconstructed from a stabilized zero-crossing representation. The completeness and stability of the zero-crossing representation was shown by the result that the reconstruction was independent of the choice of the initial point. However, this result was only indicated by numerical experiments, not by a mathematical proof. Also, the experiment was restricted only to 1-D cases.

As an extension, Mallat and Hwang [25] showed that the local maxima of the wavelet transform modulus detect the locations of irregular structures and can be used for numerical

procedures with their Lipschitz exponents[1] by which singularities are generally characterized in mathematics. Additionally, it was shown numerically that 1-D and 2-D signals can be reconstructed with a good approximation from the local maxima of their transform modulus. On the basis of the proof that Lipschitz exponents could be computed from the evolution across scales of the wavelet transform modulus maxima as shown in [25], Mallat and Zhong [26] investigated the properties of multiscale edges through the wavelet theory. Based on the assumption that a multiscale Canny edge extraction is equivalent to finding the local maxima of a wavelet transform, they defined edges as the points for which the modulus of the gradient vector is maximal in the direction towards which the gradient vector points in an image, and therefore they used the first-order derivatives of the 2-D Gaussian as wavelets for a 2-D wavelet transform. Also, they described an algorithm for reconstructing a close approximation of 1-D and 2-D signals from their multiscale edges.

On the basis of the wavelet theory proposed in [22], [25], and [26], Schröder and Wörz [33] had investigated to which degree the continuous theory, e.g. edge classification based on Lipschitz exponents, can be related to the practical case of digital images. Their experimental results showed that as yet the multiscale approach to edge extraction based on the wavelet theory is not satisfactory in practice, and thus differs from the verification result derived by Mallat and Zhong [26]. As a consequence, future work has to be devised on the key issue of how to discretize correctly the continuous theory with respect to edge classification based on the wavelet transformation.

[1]For example, the larger the uniform Lipschitz exponent α_0 is, the "more regular" the corresponding singularity at the given point x_0 will be. If $f(x)$ is discontinuous but bounded in the neighborhood of x_0, its uniform Lipschitz exponent in the neighborhood of x_0 is 0.

4 Higher Dimensional Edge Extraction

Up to here, we went through the key literature on multiscale edge extraction approaches. In this section, after giving a synopsis of several prevailing trends with respect to multiscale edge extraction and marking the subjects that have to be considered critically, we establish higher dimensional edge models for higher dimensional multiscale edge extraction.

4.1 The Issues

Edges in the image are locations where the intensity significantly changes from one level to a different one, i.e., transitions from a bright region to a dark region, or vice versa. There roughly exist two ways in defining edges: One is to define edges as zero-crossings of the Laplacian ([4], [23], [27], [29]), while the other is to define edges as local maxima of the gradient magnitude ([2], [7], [17], [31]). A simple example of edge detection by derivative operators is illustrated after [15] in Fig. 1; (a) an image with a light stripe on a dark background, (b) the intensity profile of a horizontal line, (c) the result of the first-order (e.g. the gradient) derivative of the intensity profile, and (d) the result of the second-order (e.g. the Laplacian) derivative of the intensity profile. Mathematically, positions of the maxima of the gradient magnitude correspond to zero-crossings of the Laplacian. One can see from Fig. 1 that a zero-crossing (Fig. 1-(d)) corresponds to the location of each edge. However, the local maxima approach has some important advantages in contrast to the zero-crossing approach. The problem of zero-crossing edge detection algorithms was closely analyzed in Clark [9]. In edge detection, a zero-crossing of the Laplacian can either be a maximum or a minimum of the gradient magnitude. In Fig. 1-(d), zero-crossings also appear between the edges, where no intensity variation occurs. In other words, the maxima of the gradient magnitude correspond to points of sharp variation, whereas the minima correspond to the points of the least sharp variation. It is difficult to distinguish two types of zero-crossing only with the second-order derivative operator, and thus the third-order derivative operator is needed to check the sign of the second-order derivative. However, the third-order derivatives are generally computationally more expensive and more sensitive to noise. Moreover, zero-crossings alone give only position information which does not contain any information of edge strength. This means that, with zero-crossing information only, a small amplitude fluctuation cannot be discriminated from a visually salient discontinuity. From a practical point of view, therefore, the local maxima approach using the gradient operator (the so-called non-maximum suppression

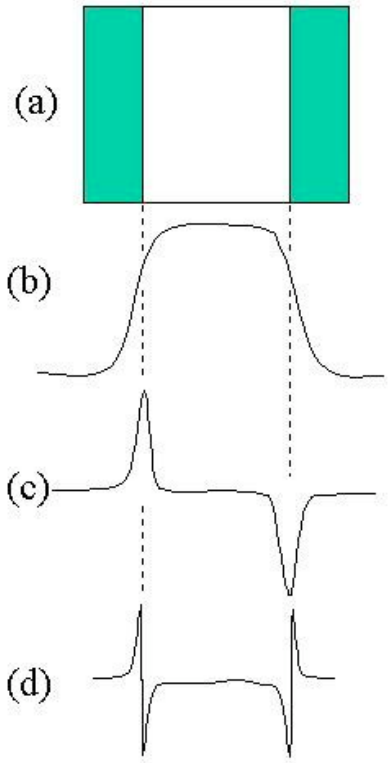


Figure 1: Edge detection by derivative operators (after Gonzalez and Woods [15])

proposed first by Canny [7]) has been gaining ground for edge extraction. In the case of non-maximum suppression, a maximum of the gradient magnitude in the gradient direction is defined as an edge point.

Canny's edge detector [7] which employed non-maximum suppression based on the step edge model is prevailing in edge detection. However, its main underlying problem is that Canny's approach did not show clearly how to combine edges from different scales. In the context of scale selection in edge extraction, the edge focusing method (i.e., the so-called coarse-to-fine tracking proposed by Bergholm [3]) has been often referred to. Aside from the computational complexity of this approach, there exist some critical problems in applying it to edge extraction. First, it is difficult to select the coarse scale at which good extraction performance can be expected (it is noticeable that as a consequence the approach in [2] favored a fine-to-coarse procedure based on first-order derivatives of the Gaussian). Second, there is no guarantee for the assumption that optimal localization can be attained at the finest scale (if noise in an image is high, localization accuracy becomes very poor at finer scales). Furthermore, care must be taken in the choice of the step size of the scale parameter,

as edge elements should not move more than one spatial pixel per focusing step in order to allow for a stable edge following algorithm.

Let alone the typical critical issues of edge extraction described above, it is now necessary to consider the matter of how to approach higher dimensional edge extraction. Compared to 1-D edge extraction using e.g. the 1-D sigmoid edge model as a favored model, higher dimensional edge extraction is not as simple as the 1-D case, since the additional edge properties, e.g. orientation and curvature for a given edge point on higher dimensional edge contour, have to be taken into account. Most approaches to higher dimensional edge extraction have used the 1-D step or, respectively, the sigmoid edge model. However, since the 1-D sigmoid edge model represents only ideally smooth intensity changes (or discontinuities), higher dimensional edge properties cannot be expressed sufficiently by it. As an exception, the extraction of higher dimensional straight edges (i.e., non-curved edges) can be reduced to the 1-D case using the direction of the gradient, meaning that the 1-D sigmoid edge model can be aligned to the direction of the gradient of higher dimensional straight edges. From a practical point of view, on the other hand, it is not difficult to envisage some problems or limits of applying the 1-D sigmoid edge model to extracting curved edges which usually appear in images. With the 1-D sigmoid edge model, it is impossible to reveal the aspect of scale as related to edge curvature. A typical example of high curvature contours is a corner, and arbitrarily smoothing its curve results in destroying its properties. Or, a sharp corner smoothed with large scale may result in a non-corner.

Besides, most approaches to multiscale higher dimensional edge extraction are inclined to extend the 2-D case to the 3-D case, the latter of which contains the same problem with respect to scale as related to curvature as mentioned above. It is noticeable that Breil and Sonka [6] recently proposed a directional 3-D edge detector designed for anisotropic image data, where their edge detector was based on interpolating the image intensity function in a small neighborhood of every voxel by a tri-cubic polynomial. However, their approach did not deal with scale as related to edge curvature.

As a consequence, it is necessary to establish higher dimensional edge models in order to reveal the aspect of scale as related to edge curvature in higher dimensional edge extraction.

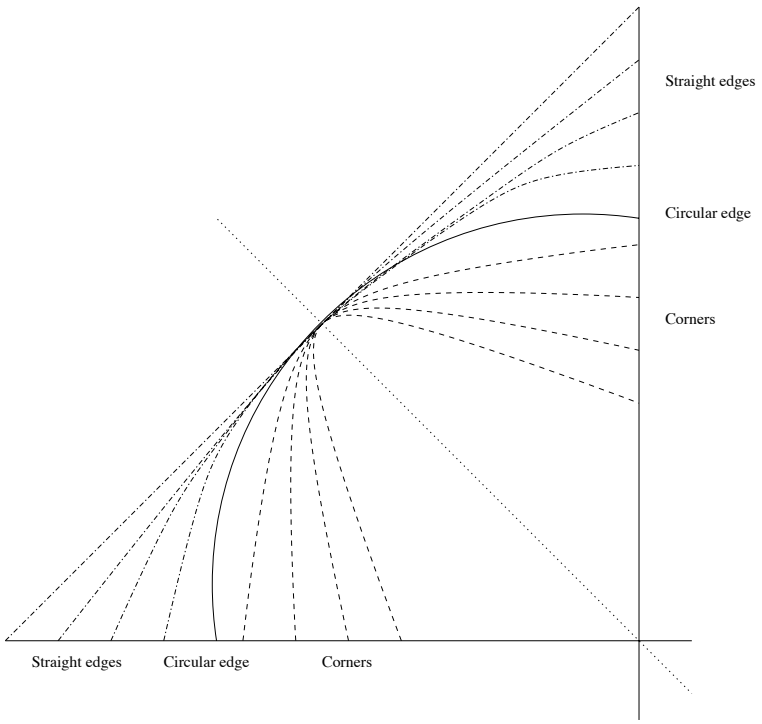


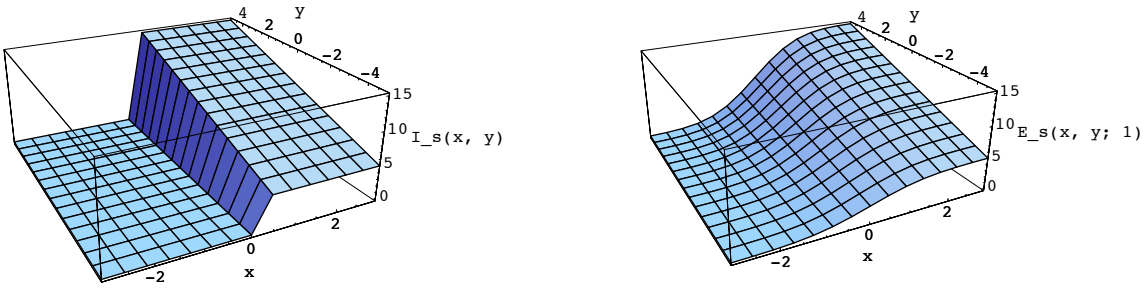
Figure 2: Higher dimensional edge classification

4.2 Higher Dimensional Edge Model

In order to establish higher dimensional edge models for the purpose of analyzing the effect of curvature related to scale for a given higher dimensional edge, we classify higher dimensional edges according to their curvature into three types, i.e. *straight edges*, *circular edges*, and *corners*, while assuming either step- or sigmoid-like profile. In more concrete terms as related to higher dimensional edge classification, for a given radius R of a circular edge its corresponding curvature K is given by the reciprocal of the radius (i.e., $K = 1/|R|$), and the curvature stays constant along circular edges. If the curvature does not stay constant (i.e. either ascending or descending along a given edge curve), an edge on the edge curve of descending curvature is assumed to form a straight edge[2], and contrarily, an edge on the edge curve of ascending curvature is assumed to represent a corner (or corner-like structure). Fig. 2 depicts the classification.

As a matter of fact, one can differently classify higher dimensional edges according to other alternative criteria. Our classification for establishing higher dimensional edge models

[2]Strictly speaking, the curvature of *straight edges* must be zero. However, we here consider as "straight-like edges" also, and for simple terminology, we use here *straight edges*.



(a) $I_s(x,y)$ (b) $E_s(x,y;t_E) : t_E = 1$

Figure 3: Straight edge models : $a = 1$, $b = 10$

is the first attempt for approaching higher dimensional edge extraction theoretically in order to analyze the effects of curvature as related to scale in multiscale edge extraction. Although our higher dimensional edge models based on this classification may not be sufficiently general to represent all edge types in real images, it is realistic and reliable to use them as a first step in order to analyze the effect of curvature as related to scale in higher dimensional edge extraction theoretically.

As a consequence, and as a first step towards higher dimensional multiscale edge extraction, we consider two classes of higher dimensional edges (i.e. straight and circular edges) for which we will derive higher dimensional edge models in the remainder of this section.

4.2.1 Straight Edge Model

A 2-D straight edge can be described by I_s given by the spatially separable function

$$I_s(x,y) = (ay + b)\mathcal{H}(x),$$

where a and b are constants and $\mathcal{H}(x)$ is the Heaviside function. $I_s(x,y)$ represents an ideal straight step edge at $x = 0$ with a linear intensity variation along the y axis. As a special case of our straight edge model, when $a = 0$ (i.e., non-variation of intensity along the straight edge line), it is equal to the typical 1-D ideal step edge model which most existing approaches employed.

A sigmoid straight edge with edge width t_E, denoted as $E_s(x,y;t_E)$, is represented by the

convolution of $I_s(x,y)$ with the Gaussian with variance t_E

$$\begin{aligned} E_s(x,y;t_E) &= I_s(x,y) * G(x,y;t_E) \\ &= (ay+b)\Phi(x;t_E), \end{aligned} \tag{4.2}$$

where $G(x,y;t_E)$ is the normalized 2-D Gaussian and $\Phi(x;t_E)$ is the integral function of the Gaussian (i.e. the normalized integral error curve) (see Appendix 1 for the detailed derivation). For sigmoid edge models, we employ $\Phi(x;t_E)$ since it has some advantages for representing an intensity discontinuity: First, it is possible to model sigmoid edges with arbitrary edge width t_E. Also, $\Phi(x;t_E)$ is smooth, and thus it is differentiable along its curve. Furthermore, with $\Phi(x;t_E)$ one can theoretically approach the problem of optimal scale selection for the underlying edge (see also [2], [3], [4], [7], [12], [17]).

It is noticeable that the linear intensity variation of $I_s(x,y)$ is not affected by smoothing with the Gaussian as can be seen from $E_s(x,y;t_E)$ in Eq. 4.2. The step and the sigmoid straight edge models for the case of $a=1$ and $b=10$ are shown in Fig. 3.

4.2.2 Circular Edge Model

A unit circular edge with radius R is represented by I_c

$$I_c(x,y) = \mathcal{H}(R^2 - x^2 - y^2).$$

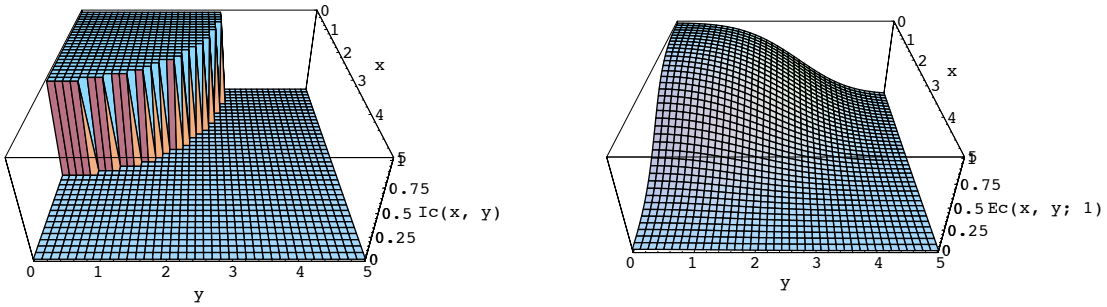
The circular edge model described by the Heaviside function has been used for modeling curved edges with constant curvature e.g. in [4], [10], [28], [29], and [35].

The sigmoid unit circular edge with edge width t_E, denoted as $E_c(x,y;t_E)$, is represented by the convolution of $I_c(x,y)$ with the Gaussian with variance t_E (see Appendix 2 for the detailed derivation) given by

$$\begin{aligned} &E_c(x,y;t_E) = I_c(x,y) * G(x,y;t_E) \\ &= R\int_{-1}^{1} G(x - R\cdot\gamma;t_E)\left[\Phi\left(y + R\sqrt{1-\gamma^2};t_E\right) - \Phi\left(y - R\sqrt{1-\gamma^2};t_E\right)\right] d\gamma. \end{aligned} \tag{4.3}$$

With the help of Mathematica, the step and the sigmoid circular edge models (when $R=3$) can be computed and displayed, as shown in Fig. 4.

It is remarkable that the analytical solution of $\int G\cdot\Phi$ is unknown ([10]). This means that we can hardly find the analytical solution of $E_c(x,y;t_E)$ in Eq. 4.3. Fortunately, without loss of generality, one can transform $E_c(x,y;t_E)$ represented in Cartesian coordinates into



(a) $I_c(x,y)$ (b) $E_c(x,y;t_E) : t_E = 1$

Figure 4: Circular edge models : $R = 3$

$E_c(r;t_E)$ in polar coordinates, where $r^2 = x^2 + y^2$ (see e.g. [28]). The details with respect to the sigmoid circular edge model in polar coordinates are described in Section 5.2.2.

5 Optimal Scale Selection

We are interested in the principal way of how to select an optimal scale for higher dimensional edge extraction. Using the scale-space representation of the higher dimensional edge models established in the previous section, we intend to analyze the behavior of higher dimensional edges over scales for the purpose of selecting their optimal scale values.

Optimal scale selection in our approach will be carried out on the basis of the so-called fine-to-coarse multiscale analysis. We are convinced that the fine-to-coarse multiscale analysis is more suitable for edge extraction than the coarse-to-fine approach (its problems were briefly stated in Section 4.1).

In this section, we briefly review the scale invariance theory in order to understand the role of scale with respect to the linear scale-space theory, and we look over the normalized derivative operator for scale selection. After that, we describe the theoretical framework for optimal scale selection based on higher dimensional edge models.

5.1 Related Theory

5.1.1 Scale Invariance

As an introduction to the scale invariance theory (e.g. [13], [20], or [30]), we here mostly refer to Pauwels *et.al.* [30] in order to figure out the principle of scale invariance (note that we will not recapitulate their close derivations and proofs).

Scale invariance implies the absence of a preferred scale in a vision system ([13], [30]). That is: When there is no a priori information as to what structure in an image we are looking for, there should be no preferred scale with respect to computational processes. In terms of dimensional analysis, a function relating physical observables must be independent of the choice of dimensional units; they do not change under the given scalings, and hence they are called *dimensionless* ([13]). In concrete terms with respect to scale-space filters, there should exist *a fixed kernel function* (or, synonymously, "*parent-kernel*") ϕ such that at different levels of the scale parameter t kernel k_t is a simple rescaling of this parent-kernel by means of a rescaling-function $\Psi : \mathbb{R}^+ \to \mathbb{R}^+$,

$$k_t(x) = k(x,t) = \frac{1}{\Psi(t)} \phi\left(\frac{x}{\Psi(t)}\right),$$

where t is the scale parameter and k_t is a scale-space kernel([30]). Note the striking similarity to wavelets (see [24], [36])

It was proven in [30] that the corresponding rescaling-function for 1-D signals (i.e. $x \in \mathbb{R}$) is given by

$$\Psi(t) = t^{\frac{1}{\alpha}},$$

where $\alpha > 0$. The parent-kernel of the Gaussian kernel is then obtained by taking $\alpha = 2$:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}},$$

where $\Psi(t) = \sqrt{t}$.

Also in [30], the rescaling-function for 2-D signals (i.e. $\vec{x} \in \mathbb{R}^2$) was derived. Similarly to the 1-D case, the rescaling condition that each filter is the scaled version of some "unscaled" (rotationally invariant) parents-filter ϕ yields,

$$k_t(\vec{x}) = \frac{1}{\Psi^2(t)} \phi\left(\frac{\vec{x}}{\Psi(t)}\right),$$

where the corresponding rescaling function is given by $\Psi(t) = t^{\frac{1}{\alpha}}$ $(\alpha > 0)$. Once again choosing $\alpha = 2$ yields the Gaussian kernel.

Although in [30] the result was not further extended to the N-D cases, we may easily generalize their 2-D result as follows: For N-D signals (i.e. $\vec{x} \in \mathbb{R}^N$), it must hold that

$$k_t(\vec{x}) = \frac{1}{\Psi^N(t)} \phi\left(\frac{\vec{x}}{\Psi(t)}\right),$$

where the corresponding rescaling-function is given by $\Psi(t) = t^{\frac{1}{\alpha}}$ $(\alpha > 0)$. Furthermore, the N-D Gaussian kernel is obtained when $\alpha = 2$.

5.1.2 The Normalized Derivative Operator

Lindeberg [20] [21] presented a scale selection method by considering the behavior of the scale under rescaling of the image pattern. Lindeberg considered, as a simple example, a 1-D sinusoidal input signal and introduced a γ-*normalized derivative operator* defined by

$$\partial_{\xi,\gamma-\text{norm}} = t^{\frac{\gamma}{2}} \partial_x,$$

which corresponds to the change of variables

$$\xi = \frac{x}{t^{\frac{\gamma}{2}}},$$

where t is the scale parameter. In case that $\gamma = 1$, ξ-coordinates and their associated normalized derivative operator are *dimensionless* (see for this terminology the previous section), and the perfect scale invariance is guaranteed. Even when $\gamma \neq 1$, it was proven that sufficient scale invariance can be achieved. Also, it was shown in [20] for a sinusoidal signal that the scale at which a normalized derivative gains its maximum over scales is proportional to the wavelength of the signal. Consequently, Lindeberg pointed out that maxima over scales of normalized derivatives reflect the scale over which spatial variations take place in the signal.

It is remarkable that already in 1988 Korn [17] suggested to introduce a normalizing factor for the gradient of the Gaussian in multiscale edge extraction. Korn's normalizing factor for the gradient of the Gaussian was given by $\sqrt{2\pi}\sigma$ ($\sigma^2 = t$), which is equivalent to the normalized derivative operator by Lindeberg when $\gamma = 1$, leaving aside the coefficient $\sqrt{2\pi}$. However, the operator response using the normalizing factor $\sqrt{2\pi}\sigma$ of the gradient of the Gaussian is monotonically increasing along the scale axis; i.e., no explicit maximum that can be selected as the optimal scale value exists somewhere along the scale axis. In [17], therefore, Korn used a heuristics for optimal scale selection. Successively, the approach in [2] drew upon Korn's normalizing factor, and came up with an indirect scale selection scheme (though not via a maximum) through relating the derivation to an underlying sigmoid edge model (which Korn did not).

5.2 Scale Selection in Higher Dimensional Edge Extraction

Let us look into the details of scale selection in higher dimensional edge extraction using our higher dimensional edge models introduced in Section 4.2.

5.2.1 The Straight Edge Case

The scale-space representation of a straight edge $E_s(x, y; t_E)$ in Eq. 4.2 is given by the convolution with the Gaussian kernel

$$\begin{aligned} L_{E_s}(x, y; t) &= E_s(x, y; t_E) * G(x, y; t) \\ &= E_s(x, y; t_E + t) \\ &= (ay + b)\Phi(x; t_E + t), \end{aligned}$$

where t is the scale parameter. The gradient of $L_{E_s}(x, y; t)$ is given by

$$\begin{aligned}\nabla L_{E_s}(x, y; t) &= \begin{pmatrix} \frac{\partial L_{E_s}}{\partial x} & \frac{\partial L_{E_s}}{\partial y} \end{pmatrix}^T \\ &= \left((ay + b)G(x; t_E + t) \quad a\Phi(x; t_E + t)\right)^T.\end{aligned}$$

By the definition that the position of an edge corresponds to that of the maximum of the gradient magnitude, the magnitude of the gradient of an edge (e.g., at $(0, k)$, $k \in \mathbb{R}$) is derived as

$$\begin{aligned}\|\nabla L_{E_s}(0, k; t)\| &= \sqrt{((ak + b)G(0; t_E + t))^2 + (a\Phi(0; t_E + t))^2} \\ &= \sqrt{\frac{(ak + b)^2}{2\pi(t_E + t)} + \frac{a^2}{4}},\end{aligned} \tag{5.4}$$

which corresponds to the response of the multiscale straight edge operator.

Apart from the constants (i.e. a, b, and k), the response function $\|\nabla L_{E_s}(0, k; t)\|$ in Eq. 5.4 is dependent on the term $1/\sqrt{2\pi(t_E + t)}$ which we denote as $M(t)$

$$M(t) = \frac{1}{\sqrt{2\pi(t_E + t)}}.$$

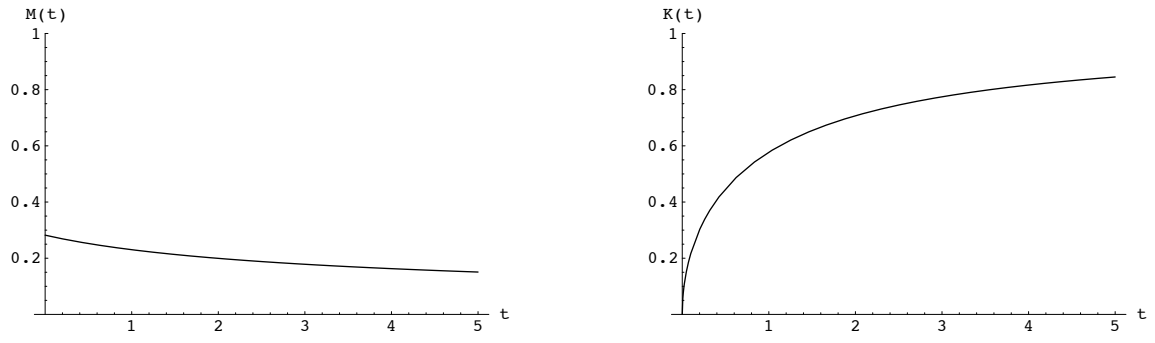
$M(t)$ is monotonously decreasing with respect to the scale parameter t, which means that the edge response $\|\nabla L_{E_s}\|$ is getting weaker as the scale parameter increases, and hence it does not give any maximum along the t-axis.

Let us consider another response function, denoted as $K(t)$, using Korn's normalized derivative operator ($k(\sigma) = \sqrt{2\pi}\sigma$ according to Korn's notation or, respectively, adapted to our notation $k(t) = \sqrt{2\pi t}$). $K(t)$ is given by multiplying $k(t)$ with $M(t)$, i.e.,

$$K(t) = k(t) \cdot M(t) = \frac{\sqrt{t}}{\sqrt{(t_E + t)}}.$$

Contrarily to $M(t)$, $K(t)$ is monotonously increasing along the scale axis, which implies that the edge operator response is getting stronger as the scale parameter increases. Fig. 5 depicts $M(t)$ and $K(t)$ in case of $t_E = 2$. Since both $M(t)$ and $K(t)$ do not give any maximum value along the scale, it can be said that they are not suitable to be used as multiscale edge response functions for directly selecting optimal scale values.

On the other hand, let $S(t)$ be the response function of $M(t)$ multiplied with a coefficient function $\eta(t) = t^\alpha$ giving the maximum at $t = t_E$. We set the limit of α as $0 < \alpha < \frac{1}{2}$; when $\alpha = 0$, $S(t)$ is equal to $M(t)$, and when $\alpha = \frac{1}{2}$, $S(t)$ is almost equivalent with $K(t)$. $S(t)$ can



(a) $M(t)$ (b) $K(t)$

Figure 5: Monotonous responses of straight edge operator: $t_E = 2$

be written as

$$S(t) = \eta(t) \cdot M(t) = \frac{t^\alpha}{\sqrt{2\pi(t_E + t)}} \qquad \left(0 < \alpha < \frac{1}{2}\right),$$

where the value of α $(0 < \alpha < \frac{1}{2})$ with which $S(t)$ gives a maximum at $t = t_E$ must be found. If $S(t)$ has a maximum at $t = t_E$, then the derivative of $S(t)$ with respect to t must be zero at $t = t_E$. That is,

$$\begin{aligned}\frac{dS(t)}{dt} &= \frac{\alpha t^{\alpha-1}}{\sqrt{2\pi(t_E + t)}} - \frac{t^\alpha}{\sqrt{2\pi(t_E + t)}2(t_E + t)} \\ &= \frac{t^\alpha}{\sqrt{2\pi(t_E + t)}}\left(\frac{\alpha}{t} - \frac{1}{2(t_E + t)}\right).\end{aligned}$$

$dS(t)/dt$ has to be 0 at $t = t_E$ $(t_E \neq 0)$

$$\left[\frac{\alpha}{t} - \frac{1}{2(t_E + t)}\right]_{t=t_E} = 0,$$

which leads to

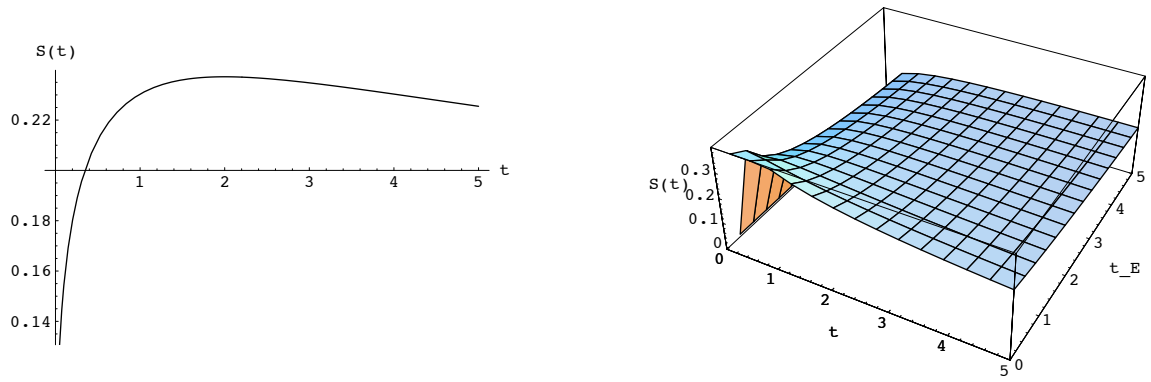
$$\frac{\alpha}{t_E} = \frac{1}{4t_E} \quad \rightarrow \quad \alpha = \frac{1}{4}.$$

Substituting $\frac{1}{4}$ for α, $S(t)$ is finally given by

$$S(t) = \frac{t^{\frac{1}{4}}}{\sqrt{2\pi(t_E + t)}}.$$

Fig. 6 shows $S(t)$; (a) in case $t_E = 2$, the maximum position of $S(t)$ corresponds to $t = 2$, (b) $S(t)$ gives maxima at $t = t_E$ as t_E varies.

In short, the theoretical result derived above implies that an optimal scale value in straight edge extraction based on the fine-to-coarse multiscale tracking can be directly selected by multiplying the response function $\|\nabla L_{E_s}(0, k; t)\|$ in Eq. 5.4 with $t^{\frac{1}{4}}$.



(a) $t_E = 2$ (b) along t_E and t

Figure 6: $S(t)$ of straight edges

5.2.2 The Circular Edge Case

The scale-space representation of a circular edge described in Eq. 4.3, $L_{E_c}(x, y; t)$, is given by the convolution with the Gaussian kernel. For simplicity of the following derivation, $L_{E_c}(x, y; t)$ in Cartesian coordinates can be transformed into $L_{E_c}(r; t)$ in polar coordinates ($r^2 = x^2 + y^2$):

$$L_{E_c}(r; t) = \mathcal{H}(R - r) * G(r; t_E + t),$$

where t_E and t respectively correspond to the edge width of a circular edge and the scale parameter, and

$$G(r; t_E + t) = \frac{1}{2\pi(t_E + t)} e^{-\frac{r^2}{2(t_E+t)}}.$$

Considering the polar coordinates (r, θ), for any point $P = (x, y)$ in Cartesian coordinates we have

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases} \quad \text{and} \quad \begin{cases} r = \sqrt{x^2 + y^2} \\ \theta = \tan^{-1}\left(\frac{y}{x}\right) \end{cases},$$

and also

$$\frac{\partial r}{\partial x} = \frac{x}{\sqrt{x^2 + y^2}} = \frac{r\cos\theta}{r} = \cos\theta$$
$$\frac{\partial r}{\partial y} = \frac{y}{\sqrt{x^2 + y^2}} = \frac{r\sin\theta}{r} = \sin\theta.$$

Then, the gradient of $L_{E_c}(r;t)$ is given by

$$\begin{aligned}\nabla L_{E_c}(r;t) &= \left(\frac{\partial L_{E_c}(r;t)}{\partial x} \quad \frac{\partial L_{E_c}(r;t)}{\partial y}\right)^T \\ &= \left(\frac{\partial L_{E_c}(r;t)}{\partial r}\frac{\partial r}{\partial x} \quad \frac{\partial L_{E_c}(r;t)}{\partial r}\frac{\partial r}{\partial y}\right)^T \\ &= \left(\frac{\partial L_{E_c}(r;t)}{\partial r}\cdot\cos\theta \quad \frac{\partial L_{E_c}(r;t)}{\partial r}\cdot\sin\theta\right)^T,\end{aligned} \tag{5.5}$$

and the magnitude of the gradient of $L_{E_c}(r;t)$ is given by

$$\begin{aligned}\|\nabla L_{E_c}(r;t)\| &= \sqrt{\left(\frac{\partial L_{E_c}(r;t)}{\partial x}\right)^2 + \left(\frac{\partial L_{E_c}(r;t)}{\partial y}\right)^2} \\ &= \sqrt{\left(\frac{\partial L_{E_c}(r;t)}{\partial r}\right)^2\cdot\cos^2\theta + \left(\frac{\partial L_{E_c}(r;t)}{\partial y}\right)^2\cdot\sin^2\theta} \\ &= \left|\frac{\partial L_{E_c}(r;t)}{\partial r}\right|,\end{aligned}$$

which is derived (see Appendix 3 for the detailed derivation) as

$$|\nabla L_{E_c}(r;t)| = \frac{R}{t_E+t}e^{-\frac{r^2+R^2}{2(t_E+t)}}I_1\left(\frac{R\cdot r}{t_E+t}\right),$$

where $I_1(\cdot)$ is the modified Bessel function of integer order 1 [3], which is depicted in Fig. 7. From Fig. 8 one can see that $|\nabla L_{E_c}(r;t)|$ gives maxima along $r = R$. This illustrates the

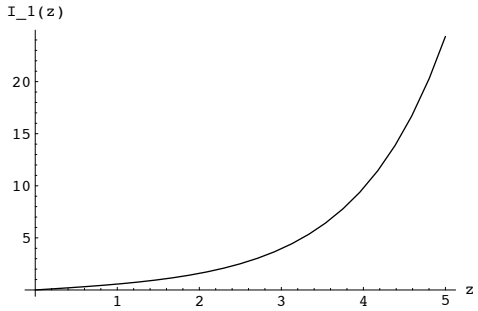


Figure 7: The modified Bessel function of integer order 1 : $I_1(z)$

[3]The modified Bessel function of integer order n is defined ([1]) by

$$I_n(z) = \frac{1}{2\pi}\int_0^{2\pi}\cos(n\theta)e^{z\cos\theta}d\theta.$$

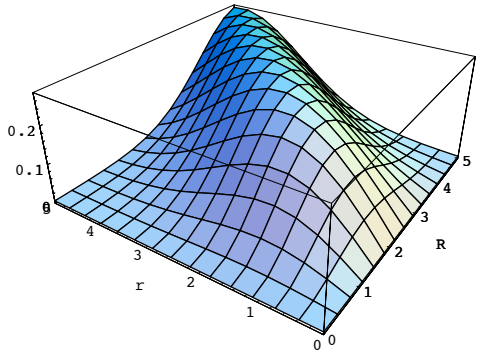


Figure 8: $|\nabla L_{E_c}(r;t)|$

definition that the position of the edge (i.e. $r = R$ or, respectively, $x^2 + y^2 = R^2$) corresponds to that of the maximum of the gradient magnitude.

Let us assume $|\nabla L_{E_c}|$ at edges (i.e. $r = R$) to be a response function of t denoted as

$$M(t) = |\nabla L_{E_c}(R;t)| = \frac{R}{t_E + t} e^{-\frac{R^2}{t_E+t}} I_1\left(\frac{R^2}{t_E + t}\right). \tag{5.6}$$

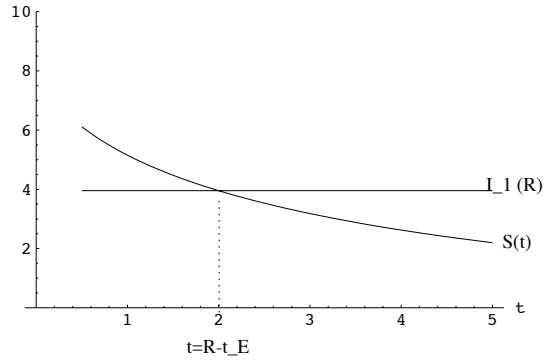
$M(t)$ is the monotonically decreasing function with respect to t, which implies that the response $|\nabla L_{E_c}|$ at edges is getting weaker as the scale increases, and thus it does not give any maximum. Therefore, the response of $M(t)$ is not proper for direct optimal scale selection. On the other hand, let $S(t)$ be a response function by multiplying $M(t)$ with e^R as given by

$$\begin{aligned} S(t) &= e^R \cdot M(t) \\ &= \frac{R}{t_E + t} e^{\frac{-R^2+R(t_E+t)}{t_E+t}} I_1\left(\frac{R^2}{t_E + t}\right), \end{aligned}$$

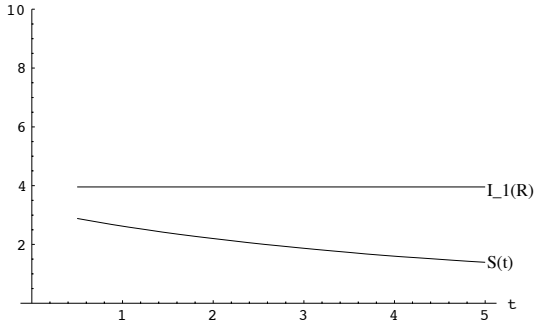
where we observe the response of $S(t)$ when $t_E + t = R$. If $t_E + t = R$, $S(t)$ is simplified as

$$S(t)|_{t_E+t=R} = \frac{R}{R} e^{\frac{-R^2+R^2}{R}} I_1(R) = I_1(R). \tag{5.7}$$

Eq. 5.7 implies that $S(t)$ gives the response $I_1(R)$ when $t_E + t = R$. In other words, the scale value satisfying $S(t) = I_1(R)$ corresponds to $t = R - t_E$, which means that the value of t_E is selected from the response of $S(t)$ uniquely since $S(t)$ is a monotonically decreasing function of t. For a given R, $I_1(R)$ is known (see Fig. 7), from which one can uniquely obtain t satisfying $S(t)$=$I_1(R)$. The derived t, in turn, can be used to derive the optimal scale value t_E (i.e., $t_E = R - t$) according to Eq. 5.7. Here it is noticeable that there does not exist any t from the response of $S(t)$ satisfying $S(t)$=$I_1(R)$ where $R \leq t_E$; as a matter of fact, it



(a) $R = 3, t_E = 1$: $t = 2$ uniquely satisfies $S(t) = I_1(R)$



(b) $R = 3, t_E = 4$: $S(t)$ cannot meet $I_1(R)$ along t

Figure 9: t satisfying $S(t) = I_1(R)$

	R									
$t = R - t_E$	1	2	3	4	5	6	7	8	9	10
$t_E = 1$	-	1	2	3	4	5	6	7	8	9
$t_E = 2$	-	-	1	2	3	4	5	6	7	8
$t_E = 3$	-	-	-	1	2	3	4	5	6	7

Table 1: The values of t at which $S(t) = I_1(R)$: When $t = R - t_E$, the value of $S(t)$ is equivalent with that of $I_1(R)$.

is meaningless both from a theoretical and a practical perspective to consider the case for which the radius of a given circular edge is smaller than its edge width, and our theoretical derivation in Eq. 5.7 shows that $R \leq t_E$ corresponds to $t \leq 0$. We denote the relationship $R \leq t_E$ the *curvature-scale constraint*. Fig. 9 illustrates that in (a) $S(t = 2)$ meets $I_1(R = 3)$ exactly at $t = 2$ in case $t_E = 1$ and in (b) $S(t)$ cannot meet $I_1(R)$ at any t since $R = 3$ is smaller than $t_E = 4$. Additionally, Table 1 represents the values of t satisfying $S(t) = I_1(R)$ with respect to increasing R and t_E.

Consequently, the optimal scale value for circular edges can be uniquely selected using the response function $S(t)$ described in Eq. 5.7 such that the optimal scale value of a given circular edge with the radius R is given by $t_E = R - t$, where t satisfies $S(t) = I_1(R)$.

6 Discussion

Based on the extensive review on edge extraction in Section 2 and Section 3 as well as in [18], some prevailing tendencies of edge extraction can be found: First, regardless of the number of dimension of the given image, edge extraction is generally considered as a one dimensional problem, and thus the 1-D step edge (or sigmoid) model is often used. Most 2-D approaches are simply extended from the 1-D approach. Second, multiscale approaches to edge extraction have attracted a substantial amount of interest due to the importance of scale in edge extraction. Therefore, optimal scale selection is an important issue in multiscale edge extraction. Third, in most approaches, one can hardly find a concrete reliable account of how to apply the continuous theory to digital images. As a consequence, it is clear that there still exist many drawbacks to be improved in edge extraction in the context of higher dimensions, of optimal scale selection, and of discrete implementation of a continuous theory.

In this report, we proposed an theoretical framework for optimal scale selection in higher dimensional edge extraction. In Section 4, we established higher dimensional edge models based on our higher dimensional edge classification, i.e. straight edges, circular edges, and corners. Using this models, we aimed to find an optimal scale for higher dimensional edges based on the theoretical analysis of the effects of curvature as related to scale in Section 5. According to the result described in Section 5.2, an optimal scale value in straight edge extraction can be selected from the response function $\|\nabla L_{E_s}(0, k; t)\|$ in Eq. 5.4 multiplied with $t^{\frac{1}{4}}$, and the optimal scale value in circular edge extraction can be selected uniquely from the response function $S(t)$ described in Eq. 5.7 such that the optimal scale value of a given circular edge with radius R is given by $t_E = R - t$, where t satisfies $S(t) = I_1(R)$.

Our future work can be sketched as follows: First, the result of our theoretical framework for optimal scale selection must be validated and evaluated by applying it to synthetic higher dimensional images, and can be further used for optimal scale selection for real images. Second, in our approach, straight and circular edges are considered only up to now, while corners are not. In the future, it is worth taking into account of how to apply the developed framework of optimal scale selection for circular edge extraction to the case of corners. Third, the radius of a given circular edge plays an important role for selecting an optimal scale value in our approach, and its value is a priori given. However, strictly speaking, the curvature of curved edges (i.e., $1/|R|$) in real images is not prior known. Therefore, curvature detection of curved edges is supposed to be ahead of optimal scale selection.

A Appendix

Appendix 1 (from page 15)

$$\begin{aligned}
E_s(x,y) &= I_s(x,y) * G(x,y;t_E) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_s(x-\alpha, y-\beta)\cdot G(\alpha,\beta;t_E)d\alpha d\beta \\
&= \int_{-\infty}^{\infty} \mathcal{H}(x-\alpha)G(\alpha;t_E)d\alpha \int_{-\infty}^{\infty}(a(y-\beta)+b)G(\beta;t_E)d\beta \\
&= \int_{-\infty}^{x} G(\alpha;t_E)d\alpha \cdot \left[(ay+b)\int_{-\infty}^{\infty} G(\beta;t_E)d\beta - a\int_{-\infty}^{\infty}\beta G(\beta;t_E)d\beta\right] \\
&= \Phi(x;t_E)\cdot[(ay+b) - a\cdot 0] \\
&= (ay+b)\Phi(x;t_E),
\end{aligned}$$

Appendix 2 (from page 15)

$$\begin{aligned}
E_c(x,y;t_E) &= I_c(x,y) * G(x,y;t_E) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{H}(R^2-\alpha^2-\beta^2)G(x-\alpha, y-\beta;t_E)d\beta d\alpha,
\end{aligned}$$

where $\mathcal{H}(R^2-\alpha^2-\beta^2) = 1$ when $|\beta| \le \sqrt{R^2-\alpha^2}$ and $-R \le \alpha \le R$. This leads to

$$E_c(x,y;t_E) = \int_{-R}^{R}\int_{-\sqrt{R^2-\alpha^2}}^{\sqrt{R^2-\alpha^2}} G(x-\alpha;t_E)G(y-\beta;t_E)d\beta d\alpha,$$

(assuming $\gamma = \alpha/R$, which leads to $\alpha = R\cdot\gamma$ and $d\alpha = R\cdot d\gamma$)

$$= \int_{-1}^{1}\int_{-\sqrt{R^2-R^2\cdot\gamma^2}}^{\sqrt{R^2-R^2\cdot\gamma^2}} G(x-R\cdot\gamma;t_E)G(y-\beta;t_E)d\beta R\cdot d\gamma$$

(we replace ξ for $y-\beta$, leading to $\xi \to y \mp R\sqrt{1-\gamma^2}$ when $\beta \to \pm R\sqrt{1-\gamma^2}$)

$$\begin{aligned}
&= R\int_{-1}^{1}\int_{y+R\sqrt{1-\gamma^2}}^{y-R\sqrt{1-\gamma^2}} G(x-R\cdot\gamma;t_E)\cdot[-G(\xi;t_E)]\,d\xi d\gamma \\
&= R\int_{-1}^{1} G(x-R\cdot\gamma;t_E)\left[\Phi\left(y+R\sqrt{1-\gamma^2};t_E\right) - \Phi\left(y-R\sqrt{1-\gamma^2};t_E\right)\right]d\gamma.
\end{aligned}$$

Appendix 3 (from page 23)

Provided that $f(r)$ and $g(r)$ are both rotationally symmetric, the convolution of $f(r)$ with $g(r)$ in polar coordinates is defined as ([5, p. 339])

$$\begin{aligned}
f(r) * g(r) &= \int_0^{\infty}\int_0^{2\pi} f(r')g(s)r'dr'd\theta' \\
&\qquad (s^2 = r^2 + r'^2 - 2rr'\cos\theta').
\end{aligned}$$

Then, $L(r;t) = \mathcal{H}(R-r) * G(r; t_E + t)$ is given by

$$\begin{aligned} L(r;t) &= \int_0^{\infty} \int_0^{2\pi} r' \mathcal{H}(R-r') \frac{1}{2\pi(t_E+t)} e^{-\frac{r'^2+r^2-2rr'\cos\theta'}{2(t_E+t)}} dr' d\theta' \\ &= \frac{1}{t_E+t} \int_0^R r' e^{-\frac{r'^2+r^2}{2(t_E+t)}} \int_0^{2\pi} \frac{1}{2\pi} e^{\frac{rr'\cos\theta'}{t_E+t}} d\theta' dr' \\ &= \frac{1}{t_E+t} \int_0^R r' e^{-\frac{r'^2+r^2}{2(t_E+t)}} I_0\left(\frac{rr'}{t_E+t}\right) dr', \end{aligned}$$

where $I_0(\cdot)$ is the modified Bessel function of integer order 0 (see for its definition Footnote 3).

The derivative of $L(r;t)$ ($T = t_E + t$) with respect to r is derived as

$$\begin{aligned} \frac{dL(r;t)}{dr} &= \frac{1}{T} \int_0^R r' \frac{d\left(e^{-\frac{r'^2+r^2}{2T}} I_0\left(\frac{rr'}{T}\right)\right)}{dr} dr' \qquad \left(\text{since } I_0'(z) = I_1(z)\right) \\ &= \frac{1}{T} \int_0^R \frac{r'^2}{T} e^{-\frac{r'^2+r^2}{2T}} I_1\left(\frac{rr'}{T}\right) - \frac{rr'}{T} e^{-\frac{r'^2+r^2}{2T}} I_0\left(\frac{rr'}{T}\right) dr' \\ &= \frac{1}{T} \int_0^R -r' \frac{d\left(e^{-\frac{r'^2+r^2}{2T}}\right)}{dr'} I_1\left(\frac{rr'}{T}\right) - e^{-\frac{r'^2+r^2}{2T}} \underbrace{\frac{rr'}{T} I_0\left(\frac{rr'}{T}\right)}_{**} dr', \end{aligned} \tag{A.8}$$

where the term denoted as '**' is derived further using the recurrence relations[4] of the Bessel functions ([1])

$$\begin{aligned} \frac{rr'}{T} I_0\left(\frac{rr'}{T}\right) &= \frac{rr'}{2T} I_0\left(\frac{rr'}{T}\right) + \frac{rr'}{2T} I_0\left(\frac{rr'}{T}\right) \\ &= \frac{rr'}{2T} I_0\left(\frac{rr'}{T}\right) + I_1\left(\frac{rr'}{T}\right) + \frac{rr'}{2T} I_2\left(\frac{rr'}{T}\right) \\ &= \frac{I_0\left(\frac{rr'}{T}\right) + I_2\left(\frac{rr'}{T}\right)}{2} \frac{rr'}{T} + I_1\left(\frac{rr'}{T}\right) \\ &= r' \frac{d\left(I_1\left(\frac{rr'}{T}\right)\right)}{dr'} + I_1\left(\frac{rr'}{T}\right) \frac{dr'}{dr'}. \end{aligned}$$

[4] By definition,

$$\begin{aligned} I_n'(z) &= \frac{I_{n-1}(z) + I_{n+1}(z)}{2}, \\ I_n(z) &= \frac{z}{2n} I_{n-1}(z) - \frac{z}{2n} I_{n+1}(z). \end{aligned}$$

As a consequence, Eq A.8 is

$$\begin{aligned}
&\frac{dL(r;t)}{dr} \\
&= \frac{1}{T}\int_0^R -r' \frac{d\left(e^{-\frac{r'^2+r^2}{2T}}\right)}{dr'} I_1\left(\frac{rr'}{T}\right) - e^{-\frac{r'^2+r^2}{2T}}\left(r' \frac{d\left(I_1\left(\frac{rr'}{T}\right)\right)}{dr'} + I_1\left(\frac{rr'}{T}\right)\frac{dr'}{dr'}\right) dr' \\
&= \frac{1}{T}\int_0^R -\frac{d\left(r' e^{-\frac{r'^2+r^2}{2T}} I_1\left(\frac{rr'}{T}\right)\right)}{dr'} dr' \\
&= -\frac{R}{T} e^{-\frac{R^2+r^2}{2T}} I_1\left(\frac{Rr}{T}\right),
\end{aligned}$$

and accordingly,

$$\begin{aligned}
\frac{dL(r;t)}{dr} &= -\frac{R}{t_E+t} e^{-\frac{R^2+r^2}{2(t_E+t)}} I_1\left(\frac{Rr}{t_E+t}\right), \\
\left|\frac{dL(r;t)}{dr}\right| &= \frac{R}{t_E+t} e^{-\frac{R^2+r^2}{2(t_E+t)}} I_1\left(\frac{Rr}{t_E+t}\right).
\end{aligned}$$

References

[1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical functions;9th printing.* Dover Publisher, New York, 1972.

[2] S. Back, H. Neumann, and H. S. Stiehl. On scale-space edge detection in computed tomograms. In *Proc. 11. Mustererkennung DAGM-Symposium, 1989, Springer-Verlag, Berlin*, pages 216–223, 1989.

[3] F. Bergholm. Edge focusing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 9(6):726–741, 1987.

[4] V. Berzins. Accuracy of Laplacian edge detectors. *Computer Vision, Graphics, and Image Processing*, 27:195–210, 1984.

[5] R. N. Bracewell. *The Fourier Transform and Its Applications;third edition.* McGraw-Hill, 2000.

[6] M. Brejl and M. Sonka. Directional 3D Edge Detection in Anisotropic Data: Detector Design and Performance Assessment. *Computer Vision and Image Understanding*, 77:84–110, 2000.

[7] J. F. Canny. A computational approach to edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[8] F. Catté, P. L. Lions, J. M. Morel, and T. Coll. Image Selective Smoothing and Edge Detection by Nonlinear Diffusion. *SIAM Journal on Numerical Analysis*, 29(1):182–193, 1992.

[9] J. J. Clark. Authenticating Edges Produced by Zero-Crossing Algorithms. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(1):43–57, 1989.

[10] C. Drewniok. *Objektlokalisation durch Adaption parametrischer Grauwertmodelle und ihre Anwendung in der Luftbildauswertung.* Dissertation, Fachbereich Informatik, Universität Hamburg, Germany, 1999.

[11] J. H. Elder. Are Edges Incomplete? *Internat. Journal of Computer Vision*, 34(2/3):97–122, 1999.

[12] J. H. Elder and S. W. Zucker. Local Scale Control for Edge Detection and Blur Estimation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(7):699–716, 1998.

[13] L. M. J. Florack, B. M. ter H. Romeny, J. J. Koenderink, and M. A. Viergever. Scale and the differential structure of images. *Image and Vision Computing*, 10(6):376–388, 1992.

[14] M. Gökmen and A. K. Jain. $\lambda\tau$-Space Representation of Images and Generalized Edge Detector. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):545–563, 1997.

[15] R. C. Gonzalez and R. E. Woods. *Digital Image Processing.* Addison-Wesley, Reading, MA, 1993.

[16] H. Jeong and C. I. Kim. Adaptive Determination of Filter Scales for Edge Detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(5):579–585, 1992.

[17] A. F. Korn. Toward a symbolic representation of intensity changes in images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1093(5):610–625, 1988.

[18] J. Y. Lim. On the Role of the Gaussian Kernel in Edge Detection and Scale-Space Methods. Technical Report FBI-HH-B-230/01, Fachbereich Informatik, Universität Hamburg, Germany, 2001.

[19] J. Y. Lim. On the Discrete Scale-Space Formulation. Technical Report FBI-HH-B-231/01, Fachbereich Informatik, Universität Hamburg, Germany, 2001.

[20] T. Lindeberg. Feature Detection with Automatic Scale-Selection. *Internat. Journal of Computer Vision*, 30(2):77–116, 1998.

[21] T. Lindeberg. Edge detection and ridge detection with automatic scale selection. *Internat. Journal of Computer Vision*, 30(2):117–154, 1998.

[22] S. Mallat. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.

[23] S. Mallat. Zero-Crossings of a Wavelet Transform. *IEEE Trans. on Information Theory*, 37(4):1019–1033, 1991.

[24] S. Mallat. *A wavelet tour of signal processing.* Academic Press, San Diego, 1998.

[25] S. Mallat and W. L. Hwang. Singularity Detection and Processing with Wavelets. *IEEE Trans. on Information Theory*, 38(2):617–643, 1992.

[26] S. Mallat and S. Zhong. Characterization of Signals from Multiscale Edges. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(7):710–732, 1992.

[27] D. Marr and E. Hildreth. Theory of edge detection. In *Proc. Royal Society of London B 207*, pages 187–217, 1980.

[28] H. Neumann, K. Ottenberg, and H. S. Stiehl. Accuracy of Regularized Differential Operators for Discontinuity Localization in 1-D and 2-D Intensity Functions. In *Proc. 1. Int. IEEE Workshop on Robust Computer Vision*, pages 214–260. R. M. Haralic, W. Förstner (Eds.) Academic Press, 1992.

[29] K. Ottenberg. *Model-Based Extraction of Geometric Structure from Digital Images.* Dissertation, Faculteit der Geneeskunde, Universiteit Utrecht, Netherlands, 1993.

[30] E. J. Pauwels, L. J. Van Gool, P. Fiddelaers, and T. Moons. An Extended Class of Scale-Invariant and Recursive Scale Space Filters. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(7):691–701, 1995.

[31] P. Perona and J. Malik. Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.

[32] H. Scharr and J. Weickert. An Anisotropic Diffusion Algorithm with Optimized Rotation Invariance. In *Proc. 22. Mustererkennung DAGM-Symposium, Kiel*, pages 460–467. Springer, Berlin, 2000.

[33] F. Schröder and S. Wörz. *Theoretische und experimentelle Untersuchungen zur Anwendung der Wavelettransformation auf Analyse digitaler Signalstrukturen.* Diplomarbeit, Fachbereich Informatik, Universität Hamburg, Germany, 2000.

[34] E. J. Stollnitz, T. D. DeRose, and D. H. Salesin. *Wavelets for Computer Graphics: Theory and Applications.* Morgan Kaufmann Publishers, Inc., San Francisco, California, 1996.

[35] P. W. Verbeek and L. J. van Vliet. On the Location Error of Curved Edges in Low-Pass Filtered 2-D and 3-D Images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(7):726–733, 1994.

[36] B. Vidaković and P. Müller. Wavelets for Kids: A Tutorial Introduction. Technical Report AMS Subject Classification: 42A06, 41A05, 65D05, Duke University, 1991.

[37] J. Weickert. A Review of Nonlinear Diffusion Filtering. In B. N. ter Haar Romeny et al., editor, *Scale-Space Theory in Computer Vision*, pages 3–27, Berlin, 1997. Springer.

[38] J. Weickert. A Semidiscrete Nonlinear Scale-Space Theory and Its Relation to The Perona-Malik Paradox. In F. Solina et al., editor, *Advances in Computer Vision*, pages 1–10, Wien, 1997. Springer.

[39] J. Weickert. Coherence-Enhancing Diffusion Filtering. *Internat. Journal of Computer Vision*, 31:111–127, 1999.

[40] R. T. Whitaker and S. M. Pizer. A Multi-Scale Approach to Nonuniform Diffusion. *Computer Vision, Graphics, and Image Processing*, 57:99–110, 1993.

[41] A. P. Witkin. Scale-Space Filtering. In *Proc. of 8th Int. Joint Conf. Artificial Intelligence, Karlsruhe*, pages 1019–1021, 1983.