A Wavelet Approach to Edge Detection

A thesis presented

by

Jun Li

to

The Department of Mathematics and Statistics

in partial fulfillment of the requirements

for the degree of

Master of Science

in the subject of

Mathematics

Sam Houston State University

Huntsville, Texas

August, 2003

A Wavelet Approach to Edge Detection

by

Jun Li

APPROVED:

_____

Jianzhong Wang
Thesis Director

_____

Charles D. Luning

_____

Harry P. Konen

Approved:

_____

Brian Chapman, Dean
College of Arts and Sciences

DEDICATION

This thesis is dedicated to my parents,
Dianwen Li and Yongsheng Xu.

Abstract

Jun Li, <u>A Wavelet Approach to Edge Detection</u>. Master of Science (Mathematics), August 2003, Sam Houston State University. Huntsville, Texas.

This thesis presents a new approach to edge detection using wavelet transforms. First, we briefly introduce the development of wavelet analysis. Then, some major classical edge detectors are reviewed and interpreted with continuous wavelet transforms. The classical edge detectors work fine with high-quality pictures, but often are not good enough for noisy pictures because they cannot distinguish edges of different significance. The proposed wavelet based edge detection algorithm combines the coefficients of wavelet transforms on a series of scales and significantly improves the results. Finally, a cascade algorithm is developed to implement the wavelet based edge detector.

Approved:

_____

Jianzhong Wang

Thesis Director

Acknowledgments

I would like to express my appreciation to Dr. Jianzhong Wang for the great amount of time, effort, support and patient guidance throughout the two semesters that I have been working on this thesis. I also thank Dr. Charles D. Luning and Dr. Harry P. Konen for reading and correcting my thesis.

Contents

List   of   Figures

List   of   Tables

Preface

When I first started writing this thesis, I was quite in doubt whether such a topic was too applied to be of any value in mathematics research. However, with further study of wavelet analysis, I realized that this particular field of mathematics would not have been developing so quickly were it not for the vast applications that required such a mathematical tool.

In St. Louis, Victor Wickerhauser was using this same mathematics to help the FBI store fingerprints more economically; while at Yale, Ronald Coifman used it to coax a battered recording of Brahms playing the piano into yielding its secrets. In France, Yves Meyer of the University of Paris-Dauphina found himself talking to astronomers about how these new techniques could be used to study the large-scale structure of the universe.

The technique of edge detection has long been studied by researchers in electrical engineering and in mathematics. It was not until the work of Mallat and his colleagues that wavelet was introduced into such studies. Mallat and his colleagues put their emphasis on decomposing and reconstructing digital signals from their edge information. Here, we make improvements based on some of the previous results, and develop a wavelet approach for the problem of edge detection.

Chapter 1
Introduction

1.1   What is edge and what is an edge detector?

An edge in an image is a contour across which the brightness of the image changes abruptly. In image processing, an edge is often interpreted as one class of singularities. In a function, singularities can be characterized easily as discontinuities where the gradient approaches infinity. However, image data is discrete, so edges in an image often are defined as the local maxima of the gradient. This is the definition we will use here.

Edge detection is an important task in image processing. It is a main tool in pattern recognition, image segmentation, and scene analysis. An edge detector is basically a high-pass filter that can be applied to extract the edge points in an image. This topic has attracted many researchers and many achievements have been made [14]-[20]. In this paper, we will explain the mechanism of edge detectors from the point of view of wavelets and develop a way to construct edge detection filters using wavelet transforms.

1.2   Introduction of the classical edge detectors

Many classical edge detectors have been developed over time. They are based on the principle of matching local image segments with specific edge patterns. The edge detection is realized by the convolution with a set of directional derivative masks [21]. The popular

edge detection operators are Roberts, Sobel, Prewitt, Frei-Chen, and Laplacian operators ([17], [18], [21], [22]). They are all defined on a 3 by 3 pattern grid, so they are efficient and easy to apply. In certain situations where the edges are highly directional, some edge detector works especially well because their patterns fit the edges better.

1.3 Noise and its influence on edge detection

However, classical edge detectors usually fail to handle imanges with strong noise, as shown in Fig. 1.1. Noise is unpredictable contamination on the original image. It is usually introduced by the transmission or compression of the image.

(a) Lena image   (b) Edges using Canny detector
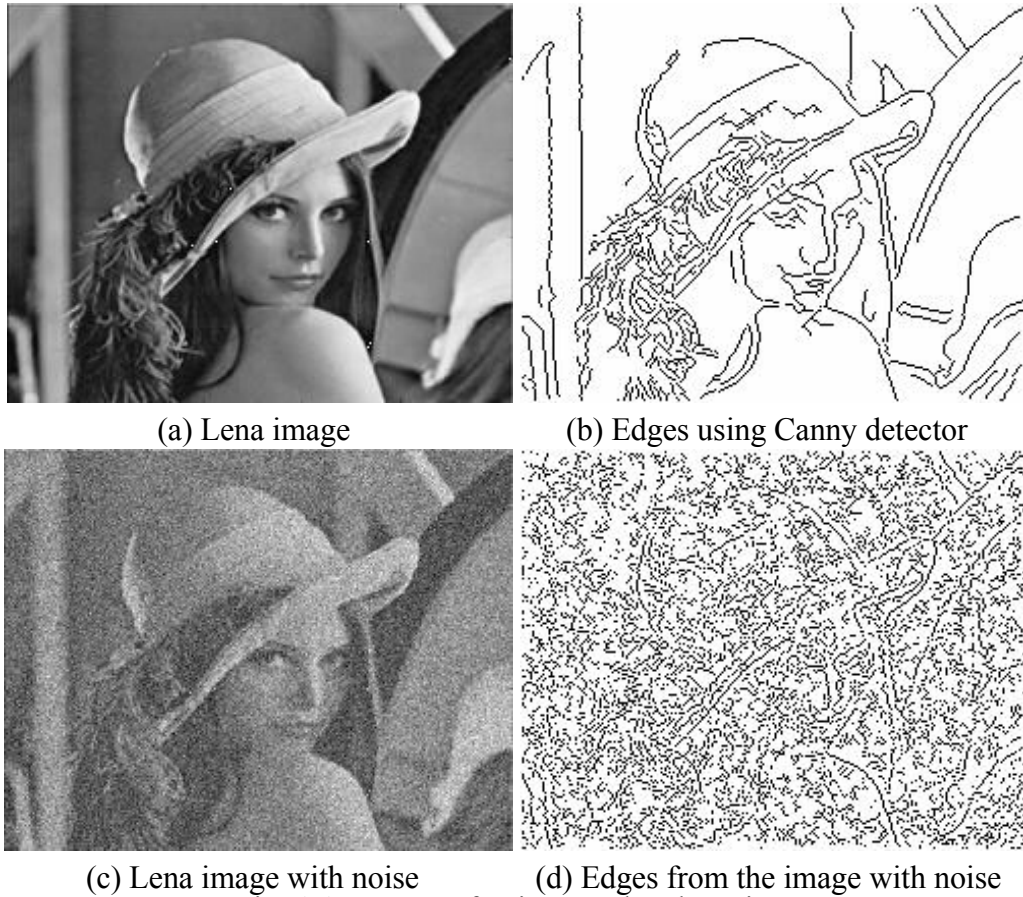
(c) Lena image with noise   (d) Edges from the image with noise
Fig. 1.1: Impact of noise on edge detection

There are various kinds of noise, but the most widely studied two kinds are white noise and *"salt and pepper"* noise. Fig. 1.1 shows the dramatic difference between the result of edge detection from two similar images, with the later one affected by some white noise.

1.4   Edge detector with pre-smoothing

To reduce the influence of noise, two techniques were developed from 1979 to 1984. Marr, et al. suggested filtering the images with the Gaussian before edge detection ([28]-[30]); Hueckel [22] and Haralick [33] proposed to approximate the image with a smooth function.

The weakness of the above approaches is that the optimal result may not be obtained by using a fixed operator. Canny [4] developed a computational approach to edge detection in 1991. He derived that an optimal detector can be approximated by the first derivative of a Gaussian.

1.5   Development of wavelet analysis

The concept of wavelet analysis has been developed since the late 1980's.   However, its idea can be traced back to the Littlewood-Paley technique and Calderón-Zygmund theory [25] in harmonic analysis. Wavelet analysis is a powerful tool for time-frequency analysis. Fourier analysis is also a good tool for frequency analysis, but it can only provide global frequency information, which is independent of time.   Hence, with Fourier analysis, it is impossible to describe the local properties of functions in terms of their spectral properties, which can be viewed as an expression of the Heisenberg uncertainty principle [13].

In many applied areas like digital signal processing, time-frequency analysis is critical. That is, we want to know the frequency properties of a function in a local time interval. Engineers and mathematicians developed analytic methods that were adapted to these problems, therefore avoiding the inherent difficulties in classical Fourier analysis.

For this purpose, Dennis Gabor [34] introduced a "sliding-window" technique. He used a Gaussian function g as a "window" function, and then calculated the Fourier transform of a function in the "sliding window". The analyzing function is

$$g_{a,b}(x) = e^{iax}g(x-b), \quad a,b \in \mathbb{R}.$$

The Gabor transform is useful for time-frequency analysis.   The Gabor transform was later generalized to the windowed Fourier transform in which g is replaced by a "time-local" function called the "window" function. However, this analyzing function has the disadvantage that the spatial resolution is limited by the fixed size of the Gaussian envelope [13].

In 1985, Yves Meyer ([23], [24]) discovered that one could obtain orthonormal bases for $L^2(\mathbb{R})$ of the type

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), \qquad j, k \in \mathbb{Z},$$

and that the expression

$$f = \sum_{j,k\in\mathbb{Z}} < f, \psi_{j,k} > \psi_{j,k},$$

for decomposing a function into these orthonormal wavelets converged in many function spaces.

The most preeminent books on wavelets are those of Meyer ([23], [24]) and Daubechies [8]. Meyer focuses on mathematical applications of wavelet theory in harmonic analysis; Daubechies gives a thorough presentation of techniques for constructing wavelet bases with desired properties, along with a variety of methods for mathematical signal analysis [14].

A particular example of an orthonormal wavelet system was introduced by Alfred Haar [35]. However, the Haar wavelets are discontinuous and therefore poorly localized in frequency.

Stéphane Mallat ([36], [37]) made a decisive step in the theory of wavelets in 1987 when he proposed a fast algorithm for the computation of wavelet coefficients. He proposed the pyramidal schemes that decompose signals into subbands. These techniques can be traced back to the 1970s when they were developed to reduce quantization noise [39]. The framework that unifies these algorithms and the theory of wavelets is the concept of a multi-resolution analysis (MRA). An MRA is an increasing sequence of closed, nested subspaces $\{V_j\}_{j\in\mathbb{Z}}$ that tends to $L^2(\mathbb{R})$ as j increases. V_j is obtained from V_{j+1} by a dilation of factor 2. V_0 is spanned by a function ϕ that satisfies

$$\phi(x) = 2 \sum h_n \phi(2x - n). \tag{1.1}$$

Equation (1.1) is called the *"two-scale equation"*, and it plays an essential role in the theory of wavelet bases. In chapter three, Mallat's wavelet theories are discussed in more detail.

In 1997, Chui and Wang [11] further discussed the asymptotically optimal time-frequency localization by scaling functions and wavelets. In their paper they proved the convergence of the time-frequency window sizes of cardinal polynomial B-wavelets, which are used in Mallat's algorithm and are important in many other wavelet applications.

1.6   Edge detector using wavelets

Now that we have talked briefly about the development of edge detection techniques and wavelet theories, we next discuss how they are related.

Edges in images can be mathematically defined as local singularities. Until recently, the Fourier transforms was the main mathematical tool for analyzing singularities. However, the Fourier transform is global and not well adapted to local singularities. It is hard to find the location and spatial distribution of singularities with Fourier transforms. Wavelet analysis is a local analysis, it is especially suitable for time-frequency analysis [1], which is essential for singularity detection. This was a major motivation for the study of the wavelet transform in mathematics and in applied domains.

With the growth of wavelet theory, the wavelet transforms have been found to be remarkable mathematical tools to analyze the singularities including the edges, and further, to detect them effectively. This idea is similar to that of John Canny [4]. The Canny

approach selects a Gaussian function as a smoothing function θ; while the wavelet-based approach chooses a wavelet function to be θ'.   Mallat, Hwang, and Zhong ([5], [6]) proved that the maxima of the wavelet transform modulus can detect the location of the irregular structures. Further, a numerical procedure to calculate their Lipschitz exponents has been provided.   One and two-dimensional signals can be reconstructed, with a good approximation, from the local maxima of their wavelet transform modulus.

The wavelet transform characterizes the local regularity of signals by decomposing signals into elementary building blocks that are well localized both in space and frequency. This not only explains the underlying mechanism of classical edge detectors, but also indicates a way of constructing optimal edge detectors under specific working conditions.

1.7   The purpose of this paper and its organization

The purpose of this paper is to explain the working mechanism of edge detectors from the point of view of wavelet transforms and to develop a new wavelet-based multi-level edge detection method.

Chapter two is an introduction to wavelet transforms.   The relationship between wavelet transforms and edge detection is discussed in detail in chapter three.   In light of Mallat's wavelet theory, we are able to construct a universal wavelet model for edge detection. After a brief review of some classical edge detectors in chapter four, we go on in chapter five to interpret those edge detectors from the point of view of wavelet transforms. Finally, in chapter six, we developed a multiscale wavelet model for edge detection.

The fast algorithm suggested by Mallat, et al. ([5], [6]) is based on a dyadic scale, which makes it reach large scales quickly. However, when dealing with noisy images, the noise level (SNR) is sensitive to the change of scales. A dyadic sequence of scales cannot always optimally adapt to the effect of noise. With the cascade algorithm developed in chapter four, one can quickly reach a scale of the wavelet that adapts to the scale of signals and the noise level, which minimizes the effect of noise on edge detection.

1.8   Some notation

Before we begin the discussion of wavelet transform, we introduce some notation. $L^2(\mathbb{R})$ denotes the Hilbert space of measurable functions, such that

$$\int_{-\infty}^{+\infty} |f(x)|^2 \, dx < +\infty. \tag{1.2}$$

For $f \in L^2(\mathbb{R})$, the norm, $||f||$, is given by

$$||f||^2 = \int_{-\infty}^{+\infty} |f(x)|^2 \, dx. \tag{1.3}$$

The convolution of two functions $f(x) \in L^2(\mathbb{R})$ and $g(x) \in L^2(\mathbb{R})$ is denoted by

$$f * g(x) = \int_{-\infty}^{+\infty} f(u)g(x-u)du. \tag{1.4}$$

The Fourier transform of a function $f(x)$ is defined by

$$\widehat{f}(\omega) = \lim_{N \to \infty} \int_{-N}^{N} f(x)e^{-i\omega x}dx, \tag{1.5}$$

and the s-dilation of $f(x)$ is defined by

$$f_s(x) = \frac{1}{s}f\left(\frac{x}{s}\right), \ s > 0. \tag{1.6}$$

Similarly, let $L^2\left(\mathbb{R}^2\right)$ be the Hilbert space of two-dimensional square-integrable functions.
The norm of $f \in L^2\left(\mathbb{R}^2\right)$ is given by

$$||f||^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |f(x,y)|^2 \, dxdy, \tag{1.7}$$

and the Fourier transform of $f \in L^2\left(\mathbb{R}^2\right)$ is defined by

$$\widehat{f}(\omega_x, \omega_y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) \, e^{-i(\omega_x x + \omega_y y)} \, dxdy. \tag{1.8}$$

For any function $f \in L^2\left(\mathbb{R}^2\right)$, $f_s(x,y)$ denotes the s-dilation of $f(x,y)$

$$f_s(x,y) = \frac{1}{s^2} f\left(\frac{x}{s}, \frac{y}{s}\right), \; s > 0. \tag{1.9}$$

Chapter 2
Introduction to Wavelet Transforms

2.1 Continuous wavelet transforms

We know that the Fourier transform of a signal $f(t)$ is defined as

$$\widehat{f}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt.$$

Since $e^{-i\omega t}$ is globally supported and not in L^2, Fourier analysis is global in nature, which makes it not suitable for detecting local singularities.

To obtain the approximate frequency contents of a signal $f(t)$ in the neighborhood of some desired location $t = b$, we can window the function using an appropriate window function $\phi(t)$ to produce the windowed function $f_b(t) = f(t)\phi(t - b)$ and then take the Fourier transform of $f_b(t)$. This is called the short-time Fourier transform (STFT). It can be defined at the location (b, ξ) in the time-frequency plane as

$$G_\phi f(b, \xi) = \int_{-\infty}^{\infty} f(t)\overline{\phi_{b,\xi}(t)}dt \tag{2.10}$$

where

$$\phi_{b,\xi}(t) = \phi(t - b)e^{i\xi t}. \tag{2.11}$$

One can recover the time function $f(t)$ by taking the inverse Fourier transform of $G_\phi f(b, \xi)$, then multiply by $\overline{\phi(t - b)}$ and integrate with respect to b:

$$f(t) = \frac{1}{2\pi \, ||\phi||^2} \int_{-\infty}^{\infty} d\xi e^{i\xi t} \int_{-\infty}^{\infty} G_\phi f(b, \xi)\overline{\phi(t - b)}db \tag{2.12}$$

11

If we use the Gaussian function as the window function, i.e., $\phi(t) = \frac{1}{2\pi\alpha}e^{-\frac{t^2}{4\alpha}}$ ($\alpha > 0$), the transform is called the Gabor transform [34]. If we use a dilated and translated wavelet function as $\phi_{b,\xi}(t)$, i.e., $\phi_{b,\xi}(t) = \psi_{b,\xi}(t)$, the transform is a continuous wavelet transform.

Definition 1 The continuous wavelet transform(CWT) of a function $f(t) \in L^2$ with respect to some analyzing wavelet ψ is defined as

$$W_\psi f(b, a) = \int_{-\infty}^{\infty} f(t)\psi_{b,a}(t)dt,$$

where

$$\psi_{b,a}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \qquad a > 0, \tag{2.13}$$

and $\psi(t)$ satisfies the following condition:

$$\widehat{\psi}(0) = \int_{-\infty}^{\infty} \psi(t)dt = 0. \tag{2.14}$$

In addition to satisfying (2.14), a wavelet is sometimes constructed such that it has vanishing moments of higher order. A wavelet is said to have vanishing moments of order m if

$$\int_{-\infty}^{\infty} t^p\psi(t)dt = 0, \qquad p = 0, ..., m - 1. \tag{2.15}$$

We often derive $\psi(t)$ from the derivative of a smoothing function $\phi(t) \in L^2(\mathbb{R})$,

$$\int_{-\infty}^{\infty} \phi(t)dt = 1$$

$$\psi(t) = \frac{d\phi}{dt}.$$

Theorem 1   Let $\phi(t) \in L^2(\mathbb{R})$ be a smoothing function that is n times differentiable. Let $\psi_p(t)$ be the pth derivative of $\phi(t)$, i.e.

$$\psi_p(t) = \frac{d^p \phi(t)}{dt^p}, \quad p < n, \tag{2.16}$$

then $\psi_p(t)$ has vanishing moments of order p.

Proof.   Since

$$\int_{-\infty}^{\infty} \psi(t)dt = \lim_{t \to \infty} \phi(t) - \lim_{t \to -\infty} \phi(t) = 0 \tag{2.17}$$

According to (2.14) and (2.15), $\psi(t)$ is a wavelet with vanishing moments of order 1.

Assume that $\psi_{p-1}(t)$ has vanishing moments of order $p - 1$, i.e.

$$\int_{-\infty}^{\infty} t^{p-2} \psi_{p-1}(t)dt = 0, \tag{2.18}$$

Then

$$\begin{aligned}
\int_{-\infty}^{\infty} t^{p-1} \psi_p(t)dt &= \int_{-\infty}^{\infty} t^{p-1} d\psi_{p-1}(t) \tag{2.19} \\
&= t^{p-1} \psi_{p-1}(t)\big|_{-\infty}^{\infty} - (p-1) \int_{-\infty}^{\infty} t^{p-2} \psi_{p-1}(t)dt \\
&= 0.
\end{aligned}$$

Thus by induction $\psi_p(t)$ has vanishing moments of order p.

■

Similar to STFT, the inverse wavelet transform involves a two-dimensional integration over the scale parameter a and the translation parameter b :

$$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} db \int_{-\infty}^{\infty} \frac{1}{a^2} \left(W_\psi f\,(b,a) \right) \psi_{b,a}\,(t)\,da, \tag{2.20}$$

where C_ψ is a constant given by

$$C_\psi = \int_{-\infty}^{\infty} \frac{\left|\widehat{\psi}\left(\omega\right)\right|^2}{|\omega|} d\omega < \infty. \tag{2.21}$$

The condition (2.21), known as the admissibility condition, restricts the class of functions that can be wavelets. It actually implies (2.14) [1].

2.2   Multi-resolution analysis

A careful choice of the mother wavelet ϕ is crucial to the Fast Wavelet Transform, (FWT). The mother wavelet should in a certain way "reproduce itself" when it is subject to scaling. More specifically, the mother wavelet ψ satisfies a linear identity having the following structure:

$$D_2\phi(t) = \sum_{k=0}^{n} c_k\phi(t-k), \tag{2.22}$$

where D_2 is a scaling operation such that $D_2\phi(t) = \phi(\frac{t}{2})$. In another words, $\phi_j(t)$ forms an MRA.

Definition 2 A multi-resolution analysis, abbreviated MRA, is constituted by the following ingredients (a)-(c).

(a) A bilateral sequence $(V_j|j \in \mathbb{Z})$ of closed subspaces of L^2. These V_j are ordered by inclusion,

$$... \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset ... \subset V_j \subset V_{j-1} \subset ... \subset L^2 \tag{2.23}$$

(smaller values of j correspond to larger spaces V_j), with

$$\cap_j V_j \;=\; \{0\} \quad \text{(seperation axiom)}, \tag{2.24}$$

$$\overline{\cup_j V_j} \;=\; L^2 \quad \text{(completeness axiom)}. \tag{2.25}$$

(b) The V_j are connected to each other by a rigid scaling property:

$$V_{j+1} = D_2 \left(V_j \right) \qquad \forall j \in \mathbb{Z} \tag{2.26}$$

(c) V_0 contains one basis vector per base step 1, i.e., there is a function $\phi \in L^2 \cap L^1$ such that its translates $(\phi\left(\cdot - k\right) | k \in \mathbb{Z})$ form an orthonormal basis of V_0. The function ϕ is commonly called the scaling function of the MRA and it determines V_0 :

$$V_0 = \left\{ f \in L^2 | f(t) = \sum_k c_k \phi(t - k), \; \sum_k |c_k|^2 < \infty \right\} \tag{2.27}$$

2.3   Scale of a wavelet function

First, we consider some special characteristics of Gaussian filters. A Gaussian function through a Gaussian filter is still a Gaussian function, but with a dilated variance. Let $g_\sigma(x, y)$ be a normalized Gaussian function with variance σ^2, then

$$
(g_{\sigma_1} * g_{\sigma_2})(x, y)
$$

$$
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_{\sigma_1}(x - h, y - k) g_{\sigma_2}(h, k) \, dh \, dk
$$

$$
= \frac{1}{4\pi^2 \sigma_1^2 \sigma_2^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{(x-h)^2 + (y-k)^2}{2\sigma_1^2}} e^{-\frac{h^2 + k^2}{2\sigma_2^2}} \, dh \, dk
$$

$$
= \frac{1}{4\pi^2 \sigma_1^2 \sigma_2^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{\left(\sigma_1^2 + \sigma_2^2\right)h^2 - 2\sigma_2^2 hx + \sigma_2^2 x^2 + \left(\sigma_1^2 + \sigma_2^2\right)k^2 - 2\sigma_2^2 ky + \sigma_2^2 y^2}{2\sigma_1^2 \sigma_2^2}} \, dh \, dk
$$

$$
= \frac{1}{4\pi^2 \sigma_1^2 \sigma_2^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{\left(h - \frac{\sigma_2^2 x}{\sigma_1^2 + \sigma_2^2}\right)^2 + \left(k - \frac{\sigma_2^2 x}{\sigma_1^2 + \sigma_2^2}\right)^2 + \frac{\sigma_1^2 \sigma_2^2 \left(x^2 + y^2\right)}{2\left(\sigma_1^2 + \sigma_2^2\right)^2}}{2\frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}}} \, dh \, dk
$$

$$
= \frac{1}{4\pi^2 \sigma_1^2 \sigma_2^2} e^{-\frac{x^2 + y^2}{2\left(\sigma_1^2 + \sigma_2^2\right)^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{\left(h - \frac{\sigma_2^2 x}{\sigma_1^2 + \sigma_2^2}\right)^2 + \left(k - \frac{\sigma_2^2 y}{\sigma_1^2 + \sigma_2^2}\right)^2}{2\frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}}} \, dh \, dk
$$

$$
= \frac{1}{2\pi(\sigma_1^2 + \sigma_2^2)} e^{-\frac{x^2 + y^2}{2\left(\sigma_1^2 + \sigma_2^2\right)^2}}
$$

$$
= g_{\sqrt{\sigma_1^2 + \sigma_2^2}}(x, y) \tag{2.28}
$$

Now, we consider a special class of wavelets, which can be defined as the derivative of a smoothing function, i.e.,

$$
\psi(x, y) = \frac{\partial}{\partial x} \phi(x, y),
$$

where $\int \int_{\mathbb{R}^2} \phi(x, y) = 1$. If the Gaussian is used as the smoothing function, this class of wavelets is called a Gaussian wavelet. From (2.28), we can develop a cascade algorithm

to find the wavelet transform of $f(x, y)$ on a series of scales.

$$
\begin{aligned}
W^1 f(ns, x, y) &= \psi_{ns} * f(x, y) \\
&= \frac{\partial \phi_{ns}}{\partial x} * f(x, y) \\
&= \frac{\partial(\phi_{(n-1)s} * \phi_s)}{\partial x} * f(x, y) \\
&= \frac{\partial(\phi_{(n-1)s} * f)}{\partial x} * \phi_s(x, y) \\
&= W^1 f((n-1)s, x, y) * \phi_s(x, y) \quad\quad (2.29)
\end{aligned}
$$

Thus, for any integer $n > 1$,

$$
W^1 f(ns, x, y) = \underbrace{\phi_s * \ldots * \phi_s}_{n-1} * W^1 f(s, x, y) \quad\quad (2.30)
$$

Similarly,

$$
W^2 f(ns, x, y) = \underbrace{\phi_s * \ldots * \phi_s}_{n-1} * W^2 f(s, x, y) \quad\quad (2.31)
$$

2.4   Approximation of the Gaussian filter with B-splines

Next, we use cardinal B-splines to approximate the Gaussian filter.

Definition 3   The cardinal B-spline of order m, denoted by $N_m(x)$, is inductively defined by the multi-convolution of the box function:

$$
\begin{aligned}
N_1(x) &= B(x) = \begin{cases} 1, & 0 \le x \le 1; \\ 0, & \text{otherwise} \end{cases}, \\
N_m(x) &= N_{m-1} * N_1(x)
\end{aligned}
$$

i.e.

$$N_m(x) \;=\; \overbrace{N_1 * N_1 * \cdots * N_1}^{m}(x) \qquad (2.32)$$

Although the cardinal B-spline of order 1 is not an accurate approximation of a Gaussian function, it has been shown in [38] that the cardinal B-spline of order n asymptotically converges to a Gaussian function $\phi_n(x)$ as n goes to infinity.

$$\phi_n(x) \;\approx\; N_n(x),$$

$$\;=\; \overbrace{N_1 * \ldots * N_1}^{n-2}(x) \qquad (2.33)$$

where $N_n(x)$ is a B-spline with support n.

We now introduce a cascade algorithm that detects edges continuously on all integer scales.

Let $H(x)$ be the Haar wavelet

$$H(x) \;=\; \begin{cases} -1, & 0 \le x < \frac{1}{2}; \\ 1, & \frac{1}{2} \le x \le 1; \\ 0, & \text{otherwise.} \end{cases}$$

$$\psi_1(x) \;\approx\; H(x),$$

where $\psi_1(x)$ is the Gaussian wavelet of scale 1. Thus the Gaussian wavelet at integer scale $n > 2$ can be approximated as

$$\psi_n(x) \;\approx\; \overbrace{N_1 * \ldots * N_1}^{n-2} *H(x)$$

$$\;=\; N_{n-2} * H(x).$$

We can easily expand this method to calculate the wavelet transform of a discrete 2-D

image $f(x, y)$, by defining $B(x, y)$, $H^1(x, y)$ and $H^2(x, y)$ as follows:

$$N_1(x, y) = B(x, y) = \begin{cases} 1, 0 \leq x \leq 1, 0 \leq y \leq 1; \\ 0, \text{ otherwise.} \end{cases}$$

$$H^1(x, y) = \begin{cases} -1, 0 \leq x < \frac{1}{2}, 0 \leq y \leq 1; \\ 1, \frac{1}{2} \leq x \leq 1, 0 \leq y \leq 1; \\ 0, \text{ otherwise.} \end{cases}$$

$$H^2(x, y) = \begin{cases} -1, 0 \leq x \leq 1, 0 \leq y < \frac{1}{2}; \\ 1, 0 \leq x \leq 1, \frac{1}{2} \leq y \leq 1; \\ 0, \text{ otherwise.} \end{cases}$$

And thereby,

$$\overbrace{}^{n-2} \tag{2.34}$$

$$W^1 f(n, x, y) \approx \overbrace{N_1 * ... * N_1}^{n-2} * H^1 * f(x, y);$$

$$\overbrace{}^{n-2}$$

$$W^2 f(n, x, y) \approx \overbrace{N_1 * ... * N_1}^{n-2} * H^2 * f(x, y).$$

We use (2.34) to calculate the horizental and vertical wavelet transform at scale n.

In practice, we sometimes need the wavelet transform with a rotated direction.   To

do this, we define the approximation for a rotated wavelet:

$$H^3(x, y) = \begin{cases} -1, 0 \leq x < y \leq 1; \\ 1, 0 \leq y \leq x \leq 1; \\ 0, \text{ otherwise.} \end{cases}$$

$$H^4(x, y) = \begin{cases} -1, 0 \leq x < -y \leq 1; \\ 1, 0 \leq -y \leq x \leq 1; \\ 0, \text{ otherwise.} \end{cases}$$

Similarly,

$$W^3 f(n, x, y) \approx \overbrace{N_1 * \dots * N_1}^{n-2} * H^3 * f(x, y); \qquad (2.35)$$

$$W^4 f(n, x, y) \approx \overbrace{N_1 * \dots * N_1}^{n-2} * H^4 * f(x, y). \qquad (2.36)$$

Fig. 2.1 shows how the Gaussian wavelet is approximated by B-spline. Each

(a)   (b)

(c)   (d)

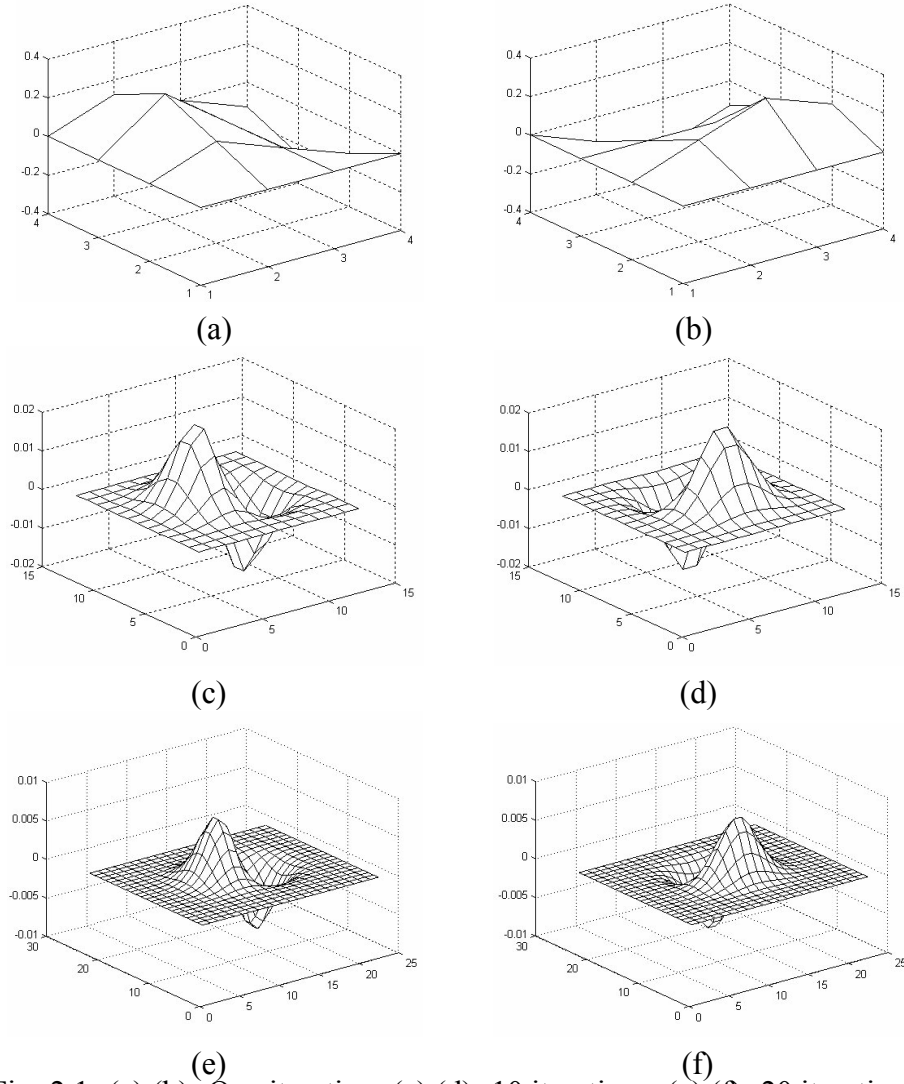(e)   (f)

Fig. 2.1: (a),(b): One iteration; (c),(d): 10 iterations; (e),(f): 20 iterations

iteration of convolution with N_1 raise the order of the B-spline by 1.

2.5   Wavelets with vanishing moments of higher order

Up to this point, only the first-degree derivative of the smoothing function is used as a wavelet. This is usually sufficient for edge detection of normal images. But in some special cases where we wish to detect discontinuities that are less noticeable, wavelets defined by higher degree derivatives may be used. In such cases, we can simply repeat the convolution with H. Each time we do the convolution with H, we increase the order of the derivative by one, and thus increase the order of vanishing moments by 1. This can easily be shown as follows:

$$
\begin{aligned}
\psi_{(k)} &= \frac{d^k \phi_n}{dx^k} \\
&= \frac{d^k}{dx^k}\left(\phi_{n-1} * \phi_1\right) \\
&= \frac{d^{k-1}\phi_{n-1}}{dx^{k-1}} * \frac{d\phi_1}{dx}, \quad (k < n).
\end{aligned}
\tag{2.37}
$$

Thus, by induction and from (2.32), (2.33),

$$
\psi_{(k)} = \phi_{n-k} * \overbrace{\frac{d\phi_1}{dx} * \ldots * \frac{d\phi_1}{dx}}^{k}
$$

In 2-D,

$$
\begin{aligned}
W_{(k)}f(n, x, y) &= \psi_{(k)} * f(x, y) \\
&\approx \overbrace{N_1 * \ldots * N_1}^{n-k-1} * \overbrace{H * \ldots * H}^{k} * f(x).
\end{aligned}
\tag{2.38}
$$

Fig. 2.2 shows an approximation of a wavelet with vanishing moments of order two.

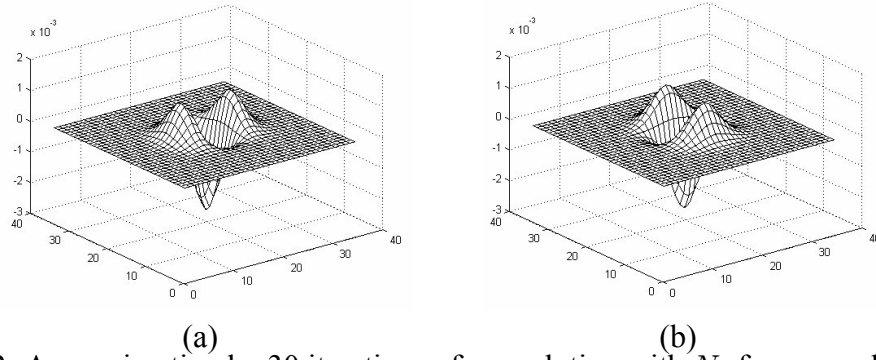(a)   (b)

Fig. 2.2: Approximation by 30 iterations of convolution with N_1 for a wavelet with vanishing moments of order two

Chapter 3
Edge Detection Using Wavelets

A remarkable property of the wavelet transform is its ability to characterize the local regularity of functions. For an image $f(x, y)$, its edges correspond to singularities of $f(x, y)$, and thus are related to the local maxima of the wavelet transform modulus. Therefore, the wavelet transform is an effective method for edge detection.

3.1 General rules of edge detection

Edge detectors are modified gradient operators. Since an edge is characterized by having a gradient of large magnitude, edge detectors are approximations of gradient operators. Because noise influences the accuracy of the computation of gradients, usually an edge detector is a combination of a smoothing filter and a gradient operator. An image is first smoothed by the smoothing filter and then its gradient is computed by the gradient operator.

The 2-norm of the gradient is $\|\nabla f\|_2 = \sqrt{f_x^2 + f_y^2}$. In order to simplify computation, in image processing we often use the 1-norm $\|\nabla f\|_1 = |f_x| + |f_y|$ instead. Their discrete forms are

$$\|\nabla f\|_2 = \sqrt{(f(i-1, j) - f(i+1, j))^2 + (f(i, j-1) - f(i+1, j))^2} \qquad (3.39)$$

and

$$\|\nabla f\|_1 = \frac{1}{2}|f(i-1, j) - f(i+1, j)| + \frac{1}{2}|f(i, j-1) - f(i, j+1)| \qquad (3.40)$$

respectively.

In image processing, the following two ways are used to define edges.

(1) Local maxima definition. Let $f(x,y) \in H^2(\Omega)$. A point $(\overline{x}, \overline{y}) \in \Omega^0$ is called an edge point of the image $f(x,y)$, if $\|\nabla f\|$ has a (strictly) local maximum at $(\overline{x}, \overline{y})$, i.e., $\|\nabla f(\overline{x}, \overline{y})\| > \|\nabla f(x,y)\|$ in a neighborhood of $(\overline{x}, \overline{y})$. An edge curve in an image is a continuous curve on which all points are edge points. The set of all edge points of $f(x,y)$ is called an edge image of $f(x,y)$.

(2) Threshold definition. Let $f(x,y) \in H^2(\Omega)$. Assume that

$$\max_{(x,y)\in\Omega} \|\nabla f(x,y)\| = M. \tag{3.41}$$

Choose K, $0 < K < M$, as the edge threshold. If $\|\nabla f(\overline{x}, \overline{y})\| \geq K$, then $(\overline{x}, \overline{y})$ is called an edge point of $f(x,y)$.

3.2 Mallat's wavelet

We first define a smoothing function $\phi(x)$, then we build the wavelet $\psi(x)$, which is associated with $\phi(x)$.

We know that the Fourier transform of the smoothing function $\phi(x)$ can be written as an infinite product

$$\phi(\omega) = e^{-i\varpi\omega} \prod_{p=1}^{+\infty} H(2^{-p}\omega) \tag{3.42}$$

where $H(\omega)$ is a 2π periodic differentiable function such that

$$|H(\omega)|^2 + |H(\omega + \pi)|^2 \leq 1 \text{ and } |H(0)| = 1 \tag{3.43}$$

We define a wavelet $\psi(x)$ whose Fourier transform $\widehat{\psi}(\omega)$ is given by

$$\widehat{\psi}(2\omega) = e^{-i\varpi\omega}G(\omega)\widehat{\phi}(\omega), \tag{3.44}$$

where $G(\omega)$ is a 2π periodic function.

Let the smoothing function $\theta(x)$ be such that $\psi(x)$ is the first-order derivative of $\theta(x)$.

A family of 2π periodic functions $H(\omega)$ and $G(\omega)$, that satisfy these constrains, is given

by

$$H(\omega) = e^{i\omega/2}(\cos(\omega/2))^{2n+1}, \tag{3.45}$$

$$G(\omega) = 4ie^{i\omega/2}\sin(\omega/2), \tag{3.46}$$

From (3.45), we derive

$$\widehat{\phi}(\omega) = \left(\frac{\sin(\omega/2)}{\omega/2}\right)^{2n+1}, \tag{3.47}$$

$$\widehat{\psi}(\omega) = iw\left(\frac{\sin(\omega/4)}{\omega/4}\right)^{2n+2}. \tag{3.48}$$

The Fourier transform $\widehat{\theta}(\omega)$ of the primitive is therefore

$$\widehat{\theta}(\omega) = \left(\frac{\sin(\omega/4)}{\omega/4}\right)^{2n+2} \tag{3.49}$$

Equations (3.47), (3.49) prove that $\psi(x)$ is a quadratic spline with compact support,

whereas $\theta(x)$ is a cubic spline whose integral is equal to 1, as shown in Fig. 3.1.
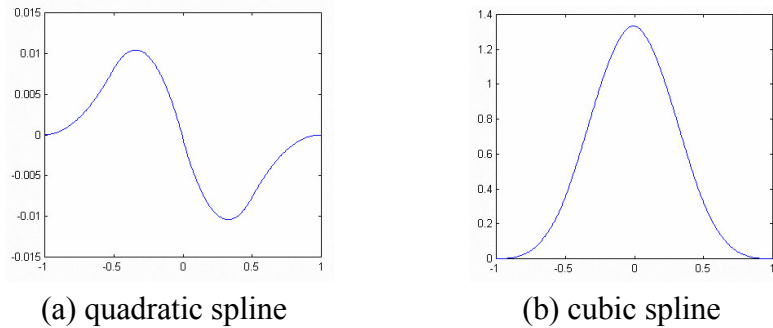

(a) quadratic spline   (b) cubic spline
Fig. 3.1: B-splines defined in Mallat's wavelet.

Now the wavelet is expanded to 2-D. Define

$$\psi_1(x,y) = 2\psi(x)\phi(2y)\text{and } \psi_2(x,y) = 2\phi(2x)\psi(y). \tag{3.50}$$

Since $\psi(x) = \frac{d\theta(x)}{dx}$, these two wavelets can be rewritten as

$$\psi_1(x,y) = \frac{\partial\theta_1(x,y)}{\partial x}\text{and } \psi_2(x,y) = \frac{\partial\theta_2(x,y)}{\partial y}, \text{with} \tag{3.51}$$

$$\theta_1(x,y) = 2\theta(x)\phi(2y)\text{and } \theta_2(x,y) = 2\phi(2x)\theta(y).$$

The functions $\theta_1(x,y)$ and $\theta_2(x,y)$ are numerically close enough so that they can be considered to be equal to a single function $\theta(x,y)$ in a first approximation.

3.3 Construction of wavelets for edge detection

We assume the smoothing function $\theta(x)$ is twice differentiable and define, respectively, $\psi^a(x)$ and $\psi^b(x)$ as the first- and second-order derivatives of $\theta(x)$

$$\psi^a(x) = \frac{d\theta(x)}{dx}\text{and } \psi^b(x) = \frac{d^2\theta(x)}{dx^2}. \tag{3.52}$$

By the definition in (2.14), the functions $\psi^a(x)$ and $\psi^b(x)$ can be considered to be wavelets because their integral is equal to 0.

$$\int_{-\infty}^{+\infty} \psi^a(x)dx = 0\text{and } \int_{-\infty}^{+\infty} \psi^b(x)dx = 0$$

A wavelet transform is computed by convolving the signal with a dilated wavelet. The wavelet transform of $f(x)$ at the scale s and position x, computed with respect to the wavelet $\psi^a(x)$, is defined by

$$W_s^a f(x) = f * \psi_s^a(x). \tag{3.53}$$

The wavelet transform of $f(x)$ with respect to $\psi^b(x)$ is

$$W_s^b f(x) = f * \psi_s^b(x). \tag{3.54}$$

It follows that

$$W_s^a f(x) = f * (s\frac{d\theta_s}{dx})(x) = s\frac{d}{dx}(f * \theta_s)(x), \text{ and} \tag{3.55}$$

$$W_s^b f(x) = f * (s^2\frac{d^2\theta_s}{dx^2})(x) = s^2\frac{d^2}{dx^2}(f * \theta_s)(x). \tag{3.56}$$

Therefore, the wavelet transforms $W_s^a f(x)$ and $W_s^b f(x)$ are, respectively, the first and second derivative of the signal smoothed at the scale s. The local extrema of $W_s^a f(x)$ thus correspond to the zeros of $W_s^b f(x)$ and to the inflection points of $f * \theta_s(x)$.

Let $\psi_s^1(x, y) = \frac{1}{s^2}\psi_1(\frac{x}{s}, \frac{y}{s})$ and $\psi_s^2(x, y) = \frac{1}{s^2}\psi_2(\frac{x}{s}, \frac{y}{s})$. Let $f(x, y) \in L^2(\mathbb{R}^2)$. The wavelet transform of $f(x, y)$ at the scale s has two components defined by

$$W_s^1 f(x, y) = f * \psi_s^1(x, y) \text{ and } W_s^2 f(x, y) = f * \psi_s^2(x, y). \tag{3.57}$$

It is straight forward that

$$\begin{pmatrix} W_s^1 f(x, y) \\ W_s^2 f(x, y) \end{pmatrix} = s \begin{pmatrix} \frac{\partial}{\partial x}(f * \theta_s)(x, y) \\ \frac{\partial}{\partial y}(f * \theta_s)(x, y) \end{pmatrix} = s\nabla(f * \theta_s)(x, y) \tag{3.58}$$

Hence, edge points can be located from the two components, $W_s^1 f(x, y)$ and $W_s^2 f(x, y)$, of the wavelet transform.

We choose the Gaussian as the smoothing function for Mallat's wavelet. i.e.

$$\theta(x, y) = \frac{1}{2\pi\sigma^2}e^{-\frac{(x^2+y^2)}{2\sigma^2}}. \tag{3.59}$$

Thus, the corresponding wavelet $\psi^1(x, y)$ and $\psi^2(x, y)$ is given by

$$\psi^1(x, y) = \frac{\partial\theta}{\partial x} = -\frac{x}{2\pi\sigma^4}e^{-\frac{(x^2+y^2)}{2\sigma^2}};$$

$$\psi^2(x,y) \quad = \quad \frac{\partial \theta}{\partial y} = -\frac{y}{2\pi\sigma^4}e^{-\frac{(x^2+y^2)}{2\sigma^2}}. \qquad (3.60)$$

Higher degree differentiations of θ also comply with the definition of wavelets in (2.14).

$$\frac{\partial^2 \theta}{\partial x \partial y} \quad = \quad \frac{xy}{2\pi\sigma^6}e^{-\frac{(x^2+y^2)}{2\sigma^2}}; \qquad (3.61)$$

$$\frac{\partial^4 \theta}{\partial^2 x \partial^2 y} \quad = \quad \frac{\left(\sigma^2 - x^2\right)\left(\sigma^2 - y^2\right)}{2\pi\sigma^{10}}e^{-\frac{(x^2+y^2)}{2\sigma^2}}. \qquad (3.62)$$

3.4   Edge detection with wavelet maxima

In a signal, there are usually four kinds of edges, which can be indicated in one dimension as in Fig. 3.2(a).

$f(x)$

(a)

$Wf(2^1, x)$

$Wf(2^2, x)$

$Wf(2^3, x)$

$Wf(2^4, x)$

$Wf(2^5, x)$

$Wf(2^6, x)$

(b)

2^1
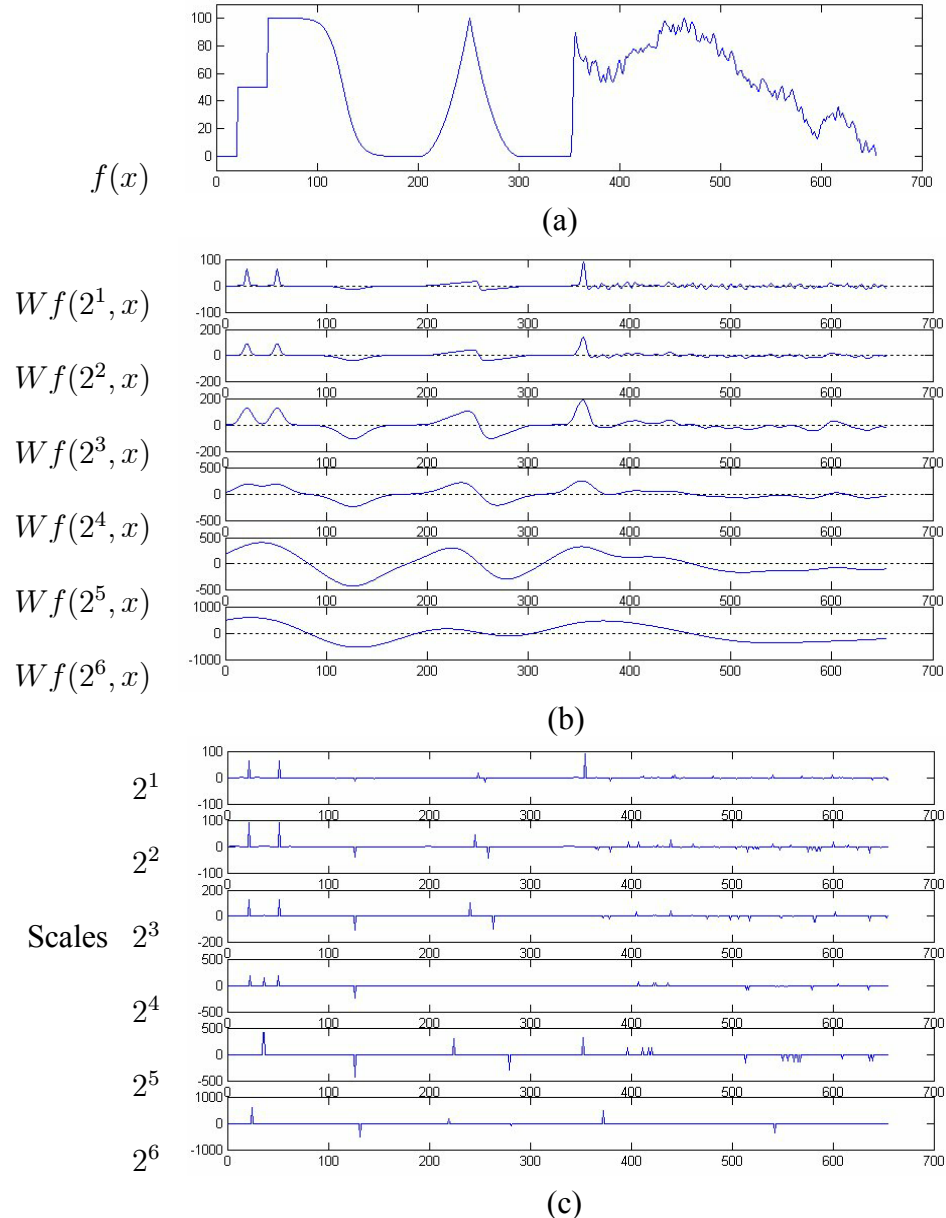
2^2

Scales 2^3

2^4

2^5

2^6

(c)

Fig. 3.2: Wavelet transform of different types of edges throughout scales : (a) Original signal; (b) wavelet transform computed up to the scale 2^6; (c)At each scale, each Dirac indicates the position of a modulus location.

The edges between 0 and 100 are step edges; The edge between 100 and 200 is a smoothed step edge; The edge between 200 and 300 is a Dirac edge; and the edges between 350 and 650 are fractal edges. Fig. 3.2(b) shows the wavelet transform of this signal on different edges in dyadic scales. Fig. 3.2(c) shows the modulus maxima which indicates the location of the edges. By taking the wavelet transform, we can locate all four kinds of edges. Here, we will explain the reason by presenting some theorems.

In an image, all edges are not created equal. Some are more significant, and some are blurred and insignificant. The edges of more significance are usually more important and more likely to be kept intact by wavelet transforms. The insignificant edges are sometimes introduced by noise and preferably removed by wavelet transforms. In mathematics, the sharpness of an edge can be described with Lipschitz exponent. We will later show that local Lipschitz regularity can be efficiently measured by wavelet transforms.

We first introduce the definition of Lipschitz exponent in 1-D.

Definition 4

- Let n be a positive integer and $n \leq \alpha \leq n + 1$. A function $f(x)$ is said to be Lipschitz α, at x_0, if and only if there exists two constants A and $h_0 > 0$, and a polynomial of order n, $P_n(x)$, such that for $h < h_0$

$$|f(x_0 + h) - P_n(h)| \leq A |h|^{\alpha}. \tag{3.63}$$

- the function $f(x)$ is uniformly Lipschitz α over the interval (a, b), if and only if there exists a constant A and for any $x_0 \in (a, b)$, there exists a polynomial of order n, $P_n(h)$, such that equation (3.63) is satisfied if $x_0 + h \in (a, b)$.

- We call Lipschitz regularity of $f(x)$ at x_0, the superior bound of all values α at x_0.

- We say that a function is singular at x_0, if it is not Lipschitz 1 at x_0.

The definition is now extended to 2-D.

Definition 5   Let $0 \le \alpha \le 1$. For any $\varepsilon > 0$, the function $f(x, y)$ is called uniformly Lipschitz α on the interval of $(a + \varepsilon, b - \varepsilon) \times (c + \varepsilon, d - \varepsilon)$, if there exists a constant A_ε, such that, for $(x, y) \in (a + \varepsilon, b - \varepsilon) \times (c + \varepsilon, d - \varepsilon)$, $h^2 + k^2 < \varepsilon^2$, the following holds:

$$|f(x + h, y + k) - f(x, y)| \le A_\varepsilon (h^2 + k^2)^{\frac{\alpha}{2}}. \tag{3.64}$$

We refer to the Lipschitz uniform regularity of $f(x, y)$ as the upper bound α_0 of all α such that $f(x, y)$ is uniformly Lipschitz α.

The local Lipschitz regularity of a function $f(x, y)$ is estimated from the evolution across scales of both $|W_s^1 f(x, y)|$ and $|W_s^2 f(x, y)|$. The value of each of these components is bounded by

$$Mf(s, x, y) = \sqrt{|W_s^1 f(x, y)|^2 + |W_s^2 f(x, y)|^2}, \tag{3.65}$$

where the function $Mf(s, x, y)$ is called he modulus of the wavelet transform at the scale 2^j. It is proportional to the modulus of the gradient vector $\nabla (f * \theta_{2^j}) (x, y)$.

Theorem 2   Let $f(x, y) \in L^2(\mathbb{R}^2)$ and $0 < \alpha < 1$. For any $\varepsilon > 0$, $f(x, y)$ is uniformly Lipschitz α over $(a + \varepsilon, b - \varepsilon) \times (c + \varepsilon, d - \varepsilon)$, if and only if for any $\varepsilon > 0$ there exists a constant A_ε such that for $(x, y) \in (a + \varepsilon, b - \varepsilon) \times (c + \varepsilon, d - \varepsilon)$ and any scale $s > 0$

$$|Mf(s, x, y)| \le A_\varepsilon s^\alpha. \tag{3.66}$$

Proof.   For the "only if" part, since $f(x, y) \in L^2 \cap L^1(\mathbb{R}^2)$ is uniformly Lipschitz α over $(a + \varepsilon, b - \varepsilon) \times (c + \varepsilon, d - \varepsilon)$, there exists an $A_1 > 0$ such that, given any scale $s > 0$, for $(x, y) \in (a + \varepsilon, b - \varepsilon) \times (c + \varepsilon, d - \varepsilon)$ and any scale $h > 0, k > 0$, $h^2 + k^2 < \min(\varepsilon^2, s^2)$

$$|f(x + h, y + k) - f(x, y)| \leq A_1 (h^2 + k^2)^{\frac{\alpha}{2}}. \tag{3.67}$$

According to the definition of $\psi_s^1(x, y)$, for any given (x, y),

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\psi_s^1(h, k)dhdk = 0 \tag{3.68}$$

$$\psi_s^1(x, y) = \frac{1}{s^2}\psi^1(\frac{x}{s}, \frac{y}{s}) < \frac{C}{s^2\left(1 + (\frac{x^2+y^2}{s^2})^{1+\frac{\alpha}{2}}\right)} \tag{3.69}$$

Thus with (3.68), we have

$$
\begin{aligned}
\left|W^1 f(s, x, y)\right| &= \left|(\psi_s^1 * f)(x, y)\right| \\
&= \left|\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x + h, y + k)\overline{\psi_s^1(h, k)}dhdk\right| \\
&= \left|\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (f(x + h, y + k) - f(x, y))\,\overline{\psi_s^1(h, k)}dhdk\right| \\
&\leq \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |(f(x + h, y + k) - f(x, y))|\,\left|\psi_s^1(h, k)\right| dhdk
\end{aligned}
$$

$$\tag{3.70}$$

Let $\Omega = \{(h, k) \mid h^2 + k^2 < s^2\}$, thus, we apply (3.67) inside the neighborhood of Ω,

$$
\begin{aligned}
&\int\int_{\Omega} |(f(x + h, y + k) - f(x, y))|\,\left|\psi_s^1(h, k)\right| dhdk \\
&\leq \int\int_{\Omega} A_1(h^2 + k^2)^{\frac{\alpha}{2}}\,\left|\psi_s^1(h, k)\right| dhdk \\
&\leq A_1\left\|\psi_s^1\right\| s^\alpha = A_1 s^\alpha
\end{aligned}
\tag{3.71}
$$

and we use the decay condition (3.69) of the wavelet outside Ω,

$$
\int\int_{\mathbb{R}^2\backslash\Omega} |(f(x+h,y+k)-f(x,y))|\left|\psi_s^1(h,k)\right| dhdk
$$

$$
\leq \int\int_{\mathbb{R}^2\backslash\Omega} C_1 \frac{1}{s^2\left(1+(\frac{h^2+k^2}{s^2})^{1+\frac{\alpha}{2}}\right)} dhdk
$$

$$
= s^\alpha \int\int_{\mathbb{R}^2\backslash\Omega} C_1 \frac{1}{s^\alpha+(h^2+k^2)^{1+\frac{\alpha}{2}}} dhdk
$$

$$
\leq s^\alpha C_2. \tag{3.72}
$$

From (3.70),(3.71) and (3.72), we have

$$
\left|W^1 f(s,x,y)\right| \leq \frac{A_\varepsilon}{\sqrt{2}} s^\alpha,
$$

$$
\text{where } A_\varepsilon = \sqrt{2}(A_1+C_2).
$$

Similarly,

$$
\left|W_s^2 f(x,y)\right| \leq \frac{A_\varepsilon}{\sqrt{2}} s^\alpha.
$$

Thus,

$$
|Mf(s,x,y)| = \sqrt{|W_s^1 f(x,y)|^2 + |W_s^2 f(x,y)|^2}
$$

$$
\leq A_\varepsilon s^\alpha.
$$

Therefore, the "only if" part is proved.

For the "if" part, since $|Mf(s,x,y)| \leq A_\varepsilon s^\alpha$,

$$
\left|W_s^1 f(x,y)\right| \leq A_\varepsilon s^\alpha,
$$

$$
\left|W_s^2 f(x,y)\right| \leq A_\varepsilon s^\alpha. \tag{3.73}
$$

Moreover, according to the definition of the wavelet ψ_s^1,

$$f(x,y) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} dh_1 \int_{-\infty}^{+\infty} dk_1 \int_0^{+\infty} \left(W_{s_1}^1 f(h_1, k_1) \right) \psi_{s_1}^1(x + h_1, y + k_1) ds_1, \quad (3.74)$$

where C_ψ is a constant and is given by

$$C_\psi = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\widehat{\psi_s^1}(\omega_x, \omega_y)}{(\omega_x^2 + \omega_y^2)^{\frac{1}{2}}} d\omega_x d\omega_y < \infty.$$

Therefore, from (3.73) and (3.74), we have

$$|f(x + h, y + k) - f(x, y)|$$

$$\leq \frac{1}{C_\psi} \int_{-\infty}^{+\infty} dh_1 \int_{-\infty}^{+\infty} dk_1$$

$$\int_0^{+\infty} A_\varepsilon s^\alpha \left| \psi_{s_1}^1(x + h + h_1, y + k + k_1) - \psi_{s_1}^1(x + h_1, y + k_1) \right| ds_1$$

Since s is arbitrary, without loss of generality, we choose s such that $s^2 = h^2 + k^2$.

Using the mean value theorem,

$$\int_{-\infty}^{+\infty} dh_1 \int_{-\infty}^{+\infty} dk_1$$

$$\int_s^{+\infty} A_\varepsilon s^\alpha \left| \psi_{s_1}^1(x + h + h_1, y + k + k_1) - \psi_{s_1}^1(x + h_1, y + k_1) \right| ds_1$$

$$\leq A_\varepsilon s^\alpha \int_s^{+\infty} \frac{(h^2 + k^2)^{\frac{1}{2}} C}{s_1^3} ds_1 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty}$$

$$\sup \left(\frac{1}{1 + \left(\left(\frac{h_1 + h}{s_1} \right)^2 + \left(\frac{k_1 + k}{s_1} \right)^2 \right)^{1+\alpha}}, \frac{1}{1 + \left(\frac{h_1^2 + k_1^2}{s_1^2} \right)^{1+\alpha}} \right) dh_1 dk_1$$

$$\leq A_\varepsilon (h^2 + k^2)^{\frac{\alpha}{2}} \int_s^{+\infty} \frac{C_1}{s_1^3} ds_1 \leq A_\varepsilon C_2 (h^2 + k^2)^{\frac{\alpha}{2}}$$

Furthermore,

$$\int_{-\infty}^{+\infty} dh_1 \int_{-\infty}^{+\infty} dk_1$$

$$\int_0^s A_\varepsilon s^\alpha \left| \psi_{s_1}^1(x+h+h_1, y+k+k_1) - \psi_{s_1}^1(x+h_1, y+k_1) \right| ds_1$$

$$\leq \int_{-\infty}^{+\infty} dh_1 \int_{-\infty}^{+\infty} dk_1$$

$$\int_0^s A_\varepsilon s^\alpha \left(|\psi_{s_1}^1(x+h+h_1, y+k+k_1)| + |\psi_{s_1}^1(x+h_1, y+k_1)| \right) ds_1$$

$$\leq A_\varepsilon s^\alpha \int_0^s \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{2C s_1^2}{1+(h_1^2+k_1^2)^{1+\frac{\alpha}{2}}} dh_1 dk_1 ds_1$$

$$\leq A_\varepsilon C_3 (h^2+k^2)^{\frac{\alpha}{2}}$$

Therefore,

$$|f(x+h, y+k) - f(x,y)| \quad \leq \quad A_1(h^2+k^2)^{\frac{\alpha}{2}},$$

$$\text{where } A_1 \quad = \quad A_\varepsilon(C_2+C_3)$$

∎

Theorem 3   Let n be a positive integer and $\alpha \leq n$. Let $f(x,y) \in L^2(\mathbb{R}^2)$. If $f(x,y)$ is Lipschitz α at (x_0, y_0), then there exists a constant $A > 0$ such that for all point (x,y) in a neighborhood of (x_0, y_0) and any scale s

$$|Mf(s,x,y)| \leq A(s^\alpha + |x-x_0|^\alpha) \tag{3.75}$$

Conversely, let $\alpha < n$ be a noninteger value. The function $f(x,y)$ is Lipschitz α at (x_0, y_0), if the following two conditions hold.

- *There exists some $\varepsilon > 0$ and a constant A such that for all points (x, y) in a neighborhood of (x_0, y_0) and any scale s*

$$|Mf(s, x, y)| \le As^{\varepsilon}. \tag{3.76}$$

- *There exists a constant B such that for all points (x, y) in a neighborhood of (x_0, y_0) and any scale s*

$$|Mf(s, x, y)| \le B\left(s^{\alpha} + \frac{\left((x - x_0)^2 + (y - y_0)^2\right)^{\frac{\alpha}{2}}}{\frac{1}{2}\left|\log\left((x - x_0)^2 + (y - y_0)^2\right)\right|}\right) \tag{3.77}$$

Theorem 3 is an extension of Theorem 2 and the proof is similar.   Theorem 2 and Theorem 3 prove that the wavelet transform is particularly well adapted to estimating the local regularity of functions.   Because edges are singularities characterized by Lipschitz exponent, it is appropriate to locate edges by finding local singularities with the wavelet transform.   Fig. 3.3 shows how the coefficients of the wavelet transform change across scales.

3.5   2-D discrete convolution

The convolution of two bi-variate functions $f(x, y)$ and $g(x, y)$ is defined by

$$(f * g)(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x - u, y - v)g(u, v)dudv,$$

provided that the integral on the right hand side exists.

Data from digital images are discrete.   So to deal with such image data, we need discrete filters, which can be applied by discrete convolutions.

**Definition 6**   The 2-D discrete convolution of two sequences **g** and **x** is a 2-D sequence

$\mathbf{g} * \mathbf{x}$ given by

$$(\mathbf{g} * \mathbf{x})(i,j) = \sum_{k,l} g(k,l)x(i-k,j-l), \;\; i \in \mathbb{Z}, \; j \in \mathbb{Z} \tag{3.78}$$

provided the series in (3.78) is convergent for each $i, j \in \mathbb{Z}$.

Continuous wavelet models can be converted into discrete edge detectors by discretizing the wavelet with appropriate scales. Suppose the support of a wavelet $\psi(x,y)$ is within an n by n grid ($n = 2k+1, k \in \mathbb{Z}^+$), then the discrete edge detector defined by the wavelet is

$$g(i,j) = \int_{j-\frac{1}{2}}^{j+\frac{1}{2}} \int_{i-\frac{1}{2}}^{i+\frac{1}{2}} \psi(x,y)dxdy, \;\; i = -k..k, j = -k..k. \tag{3.79}$$
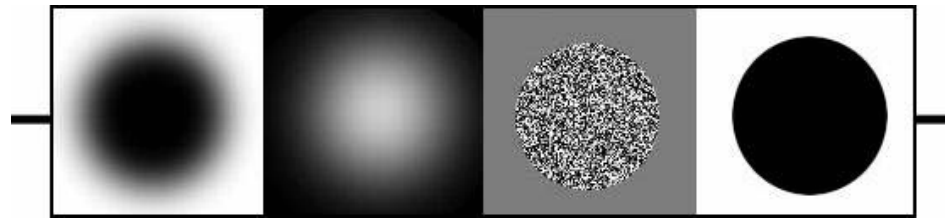
(a)

(b)

Fig. 3.3: (a) An image with different type of edges. A sample of pixels is taken across the indicated line in the middle; (b) Modulus of the continuous wavelet transform from scale 1 to 2^6(brighter color indicates edges).

Chapter 4
Review of Classical Edge Detectors

Classical edge detectors use a pre-defined group of edge patterns to match each image segments of a fixed size. 2-D discrete convolutions are used here to find the correlations between the pre-defined edge patterns and the sampled image segment.

$$(f * m)(x, y) = \sum_i \sum_j f(i, j) m(x - i, y - j), \tag{4.80}$$

where f is the image and m is the edge pattern defined by

$$m \quad = \quad \begin{array}{|c|c|c|} \hline m(-1, -1) & m(-1, 0) & m(-1, 1) \\ \hline m(0, -1) & m(0, 0) & m(0, 1) \\ \hline m(1, -1) & m(1, 0) & m(1, 1) \\ \hline \end{array},$$

$$m(i, j) \quad = \quad 0, \text{ if } (i, j) \text{ is not in the defined grid.}$$

These patterns are represented as filters, which are vectors (1-D) or matrices (2-D). For fast performance, usually the dimension of these filters are 1×3 (1-D) or 3×3 (2-D).

From the point of view of functions, filters are discrete operators of directional derivatives. Instead of finding the local maxima of the gradient, we set a threshold and consider those points with gradient above the threshold as edge points. Given the source image $f(x, y)$, the edge image $E(x, y)$ is given by

$$E = \sqrt{(f * s)^2 + (f * t)^2}, \tag{4.81}$$

where s and t are two filters of different directions.

4.1   Roberts edge detector

$$s = \begin{array}{|c|c|c|} \hline 0 & 0 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & -1 \\ \hline \end{array} \qquad t = \begin{array}{|c|c|c|} \hline 0 & 0 & 0 \\ \hline 0 & 1 & 0 \\ \hline -1 & 0 & 0 \\ \hline \end{array}$$

The edge patterns are shown in Fig. 4.1.
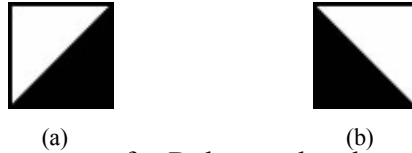

(a)   (b)

Fig. 4.1. Edge patterns for Roberts edge detector: (a) s; (b) t

These filters have the shortest support, thus the position of the edges is more accurate. On the other hand, the short support of the filters make it very vulnerable to noise. The edge pattern of this edge detector makes it especially sensitive to edges with a slope around $\frac{\pi}{4}$. Some computer vision programs use the Roberts edge detector to recognize edges of roads.

4.2   Prewitt edge detector

$$s = \begin{array}{|c|c|c|} \hline -1 & -1 & -1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 1 & 1 \\ \hline \end{array}, \qquad t = \begin{array}{|c|c|c|} \hline -1 & 0 & 1 \\ \hline -1 & 0 & 1 \\ \hline -1 & 0 & 1 \\ \hline \end{array}$$

The edge patterns are shown in Fig. 4.2.

(a)   (b)

Fig. 4.2: Edge patterns for Prewitt and Sobel edge detectors: (a)s; (b)t

These filters have longer support. They differentiate in one direction and average in

the other direction. So the edge detector is less vulnerable to noise. However, the position

of the edges might be altered due to the average operation.

4.3   Sobel edge detector

$$s = \begin{array}{|c|c|c|} \hline -1 & -2 & -1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}, \qquad t = \begin{array}{|c|c|c|} \hline -1 & 0 & 1 \\ \hline -2 & 0 & 2 \\ \hline -1 & 0 & 1 \\ \hline \end{array}$$

The edge patterns are similar to those of the Prewitt edge detector as shown in Fig.

4.2. These filters are similar to the Prewitt edge detector, but the average operator is more

like a Gaussian, which makes it better for removing some white noise.

4.4   Frei-Chen edge detector

A 3×3 subimage b of an image f may be thought of as a vector in $\mathbb{R}^9$. For example,

$$b = \begin{array}{|c|c|c|} \hline b_4 & b_3 & b_2 \\ \hline b_5 & b_0 & b_1 \\ \hline b_6 & b_7 & b_8 \\ \hline \end{array} \qquad \longrightarrow \qquad b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_8 \end{bmatrix}$$

Let V denote the vector space of 3×3 subimages. B_V, an orthogonal basis for V,

is used for the Frei-Chen method. The subspace E of V that is spanned by the subimages

v_1, v_2, v_3 and v_4 is called the edge subspace of V. The Frei-Chen edge detection method

bases its determination of edge points on the size of the angle between the subimage b and

its projection on the edge subspace.

$$\begin{array}{|c|c|c|} \hline 1 & \sqrt{2} & 1 \\ \hline 0 & 0 & 0 \\ \hline -1 & -\sqrt{2} & -1 \\ \hline \end{array} \quad \begin{array}{|c|c|c|} \hline 1 & 0 & -1 \\ \hline \sqrt{2} & 0 & -\sqrt{2} \\ \hline 1 & 0 & -1 \\ \hline \end{array} \quad \begin{array}{|c|c|c|} \hline 0 & -1 & \sqrt{2} \\ \hline 1 & 0 & -1 \\ \hline -\sqrt{2} & 1 & 0 \\ \hline \end{array} \quad \begin{array}{|c|c|c|} \hline \sqrt{2} & -1 & 0 \\ \hline -1 & 0 & 1 \\ \hline 0 & 1 & -\sqrt{2} \\ \hline \end{array}$$
$$\qquad v_1 \qquad\qquad\qquad v_2 \qquad\qquad\qquad v_3 \qquad\qquad\qquad v_4$$

0	1	0
-1	0	-1
0	1	0

v_5

-1	0	1
0	0	0
1	0	-1

v_6

1	-2	1
-2	4	-2
1	-2	1

v_7

-2	1	-2
1	4	1
-2	1	-2

v_8

1	1	1
1	1	1
1	1	1

v_9

$$\theta_E = \cos^{-1}\left(\frac{\sum_{i=1}^{4}(v_i \cdot b)^2}{\sum_{i=1}^{9}(v_i \cdot b)^2}\right)^{\frac{1}{2}} \tag{4.82}$$

The edge patterns are shown in Fig. 4.3.

(a)   (b)   (c)

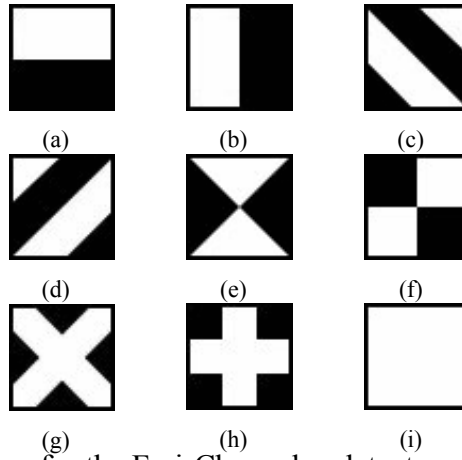(d)   (e)   (f)

(g)   (h)   (i)

Fig. 4.3: Edge Patterns for the Frei-Chen edge detector: (a) v1; (b) v2; (c) v3; (d) v4; (e) v5; (f) v6; (g) v7; (h) v8; (i) v9.

As you can see from the patterns, the subimages in the edge space are typical edge patterns with different directions; the other subimages resemble lines and blank space. Therefore, the angle θ_E is small when the subimage contains edge-like elements, and θ_E is large otherwise.

4.5 Canny edge detection

Canny edge detection [4] is an important step towards mathematically solving edge detection problems. This edge detection method is optimal for step edges corrupted by white noise. Canny used three criteria to design his edge detector. The first requirement is reliable

detection of edges with low probability of missing true edges, and a low probability of detecting false edges. Second, the detected edges should be close to the true location of the edge. Lastly, there should be only one response to a single edge.

To quantify these criteria, the following functions are defined:

$$SNR(f) \quad = \quad \frac{A}{n_0} \cdot \frac{\left| \int_{-\infty}^{0} f(x)dx \right|}{\left[\int_{-\infty}^{\infty} f^2(x)\, dx \right]^{\frac{1}{2}}},$$

$$Loc(f) \quad = \quad \frac{A}{n_0} \cdot \frac{|f'(0)|}{\left[\int_{-\infty}^{\infty} f'^2(x)dx \right]^{\frac{1}{2}}},$$

where A is the amplitude of the signal and n_0^2 is the variance of noise. $SNR(f)$ defines the signal-to-noise ratio and $Loc(f)$ defines the localization of the filter $f(x)$.

Now, by scaling f to f_s, we get the following "uncertainty principle":

$$SNR(f_s) \quad = \quad \sqrt{s} SNR(f),$$

$$Loc(f_s) \quad = \quad \frac{1}{\sqrt{s}} Loc(f).$$

That is, increasing the filter size increases the signal-to-noise ratio but also decreases the localization by the same factor. This suggests maximizing the product of the two. So the object function is defined as:

$$J(f) = \frac{\left| \int_{-\infty}^{0} f(x)dx \right|}{\left[\int_{-\infty}^{\infty} f^2(x)dx \right]^{\frac{1}{2}}} \cdot \frac{|f'(0)|}{\left[\int_{-\infty}^{\infty} f'^2(x)dx \right]^{\frac{1}{2}}}, \qquad (4.83)$$

where $f(x)$ is the filter for edge detection. The optimal filter that is derived from these requirements can be approximated with the first derivative of the Gaussian filter,

$$f(x) = -\frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}.$$

The choice of the standard deviation for the Gaussian filter, σ, depends on the size,

or scale, of the objects contained in the image.   For images with multiple size objects, or unknown size one approach is to use Canny detectors with different σ values.   The outputs of the different Canny filters are combined to form the final edge image.

Chapter 5
The Wavelet Point of View of Some Classical Edge Detectors

From the definition of the wavelet transform (2.14), we see that the edge detectors are actually discretized wavelet functions. The convolution with edge detectors gives the wavelet transform of the image at a certain scale. Here,we approximate the continuous wavelet models, which will result in the classical edge detectors when discretized. Once we have a continuous model, we can then change the scales and detect edges on different scale levels.

5.1   Roberts edge detector

The Roberts edge detector finds edges using the Roberts approximation to the derivative. It directly discretizes the gradient operator as follows.

$$R(i,j) =$$
$$\sqrt{\frac{1}{8}[(f(i-1,j-1) - f(i+1,j+1))^2 + (f(i-1,j+1) - f(i+1,j-1))^2]}.$$

$$(5.84)$$

The Roberts edge detector can be described as a discretized model of the following two orthogonal wavelets:

$$\text{Roberts-s} \quad : \quad \psi^1(x,y) = (y-x)e^{-\frac{x^2+y^2}{2}},$$

$$\text{Roberts-t} \quad : \quad \psi^2(x,y) = (x+y)e^{-\frac{x^2+y^2}{2}}.$$

Fig. 5.1 shows a 3-D plot of the wavelets.

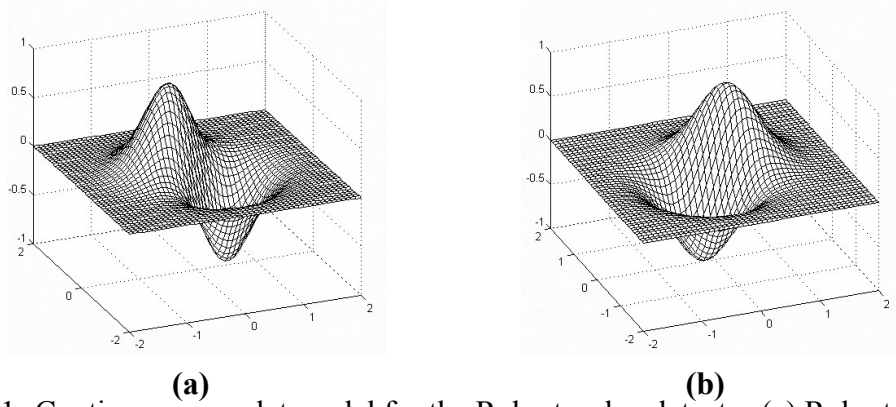**(a)**   **(b)**
Fig. 5.1: Continuous wavelet model for the Roberts edge detector (a) Roberts-s; (b) Roberts-t.

5.2   Prewitt edge detector

The Prewitte edge detector finds edges using the Prewitt approximation to the derivative.

Let $B(x)$ be the box function: $B(x) = \chi_{[-1,1]}(x)$ and $B(x,y) = B(x)B(y)$. The Prewitt detector P is constructed as follows:

$$P_f = ||\nabla(B * f)||$$

$$
\begin{aligned}
P_f(i,j) \;=\; & \frac{1}{6}|[f(i-1,j-1) + f(i-1,j) + f(i-1,j+1)] - \\
& [f(i+1,j-1) + f(i+1,j) + f(i+1,j+1)]| + \\
& \frac{1}{6}[f(i-1,j-1) + f(i,j-1) + f(i+1,j-1)] - \\
& [f(i-1,j+1) + f(i,j+1) + f(i+1,j+1)]|
\end{aligned}
\tag{5.85}
$$

The Prewitt edge detector can be described as a discretized model of the following two orthogonal Haar wavelets:

$$\psi^1(x,y) \;=\; \begin{cases} -1, & -1.5 \le x < 0; \; -1.5 \le y \le 1.5, \\ 1, & 0 \le x < 1.5; \; -1.5 \le y \le 1.5, \\ 0, & \text{otherwise} \end{cases}$$

$$\psi^2(x,y) \;=\; \begin{cases} -1, & -1.5 \le y < 0; \; -1.5 \le x \le 1.5 \\ 1, & 0 \le y < 1.5; \; -1.5 \le x \le 1.5 \\ 0, & \text{otherwise} \end{cases} .$$

Fig. 5.2 shows a 3-D plot of the wavelets.

(a)   (b)
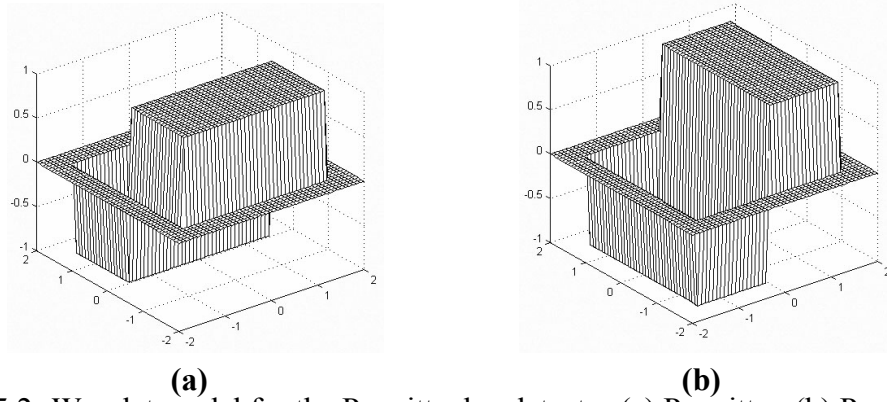
Fig. 5.2: Wavelet model for the Prewitt edge detector (a) Prewitt-s; (b) Prewitt-t.

5.3   Sobel edge detector

Let $L(x)$ be the central cardinal linear B-spline, $L(c,y) = L(x)L(y)$. The Sobel edge detector S is constructed as

$$S_f = \|\nabla(L * f)\| .$$

If the 1-norm is chosen, then its discrete form is

$$S_f(i,j) \;=\; \frac{1}{8} |[f(i-1,j-1) + 2f(i-1,j) + f(i-1,j+1)] - \\ [f(i+1,j-1) + 2f(i+1,j) + f(i+1,j+1)]| +$$

$$\frac{1}{8}|[f(i-1,j-1)+2f(i,j-1)+f(i+1,j-1)]-$$
$$[f(i-1,j+1)+2f(i,j+1)+f(i+1,j+1)]|. \qquad (5.86)$$

The Sobel edge detector can be described as a discretized model of the two orthogonal wavelets:

$$\psi^1(x,y) = -xe^{-\frac{x^2+y^2}{2}}, \qquad (5.87)$$
$$\psi^2(x,y) = -ye^{-\frac{x^2+y^2}{2}}.$$

Fig. 5.3 shows a 3-D plot of the wavelets.

(a)   **(b)**
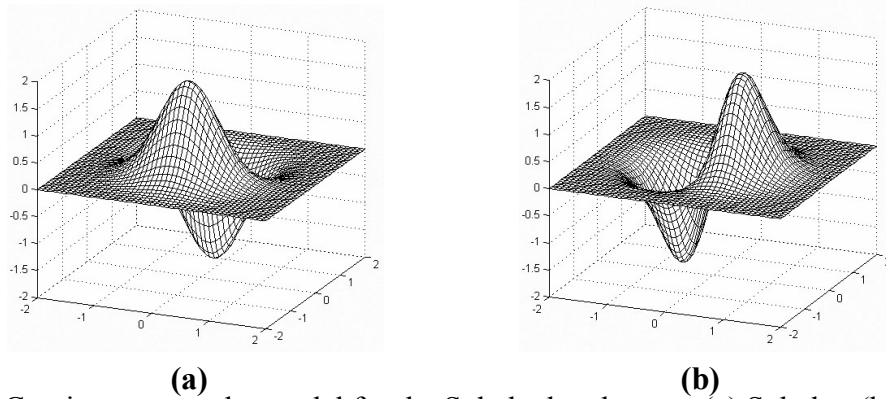
Fig. 5.3: Continuous wavelet model for the Sobel edge detector (a) Sobel-s; (b) Sobel-t.
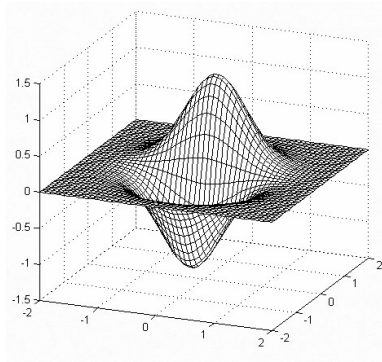
5.4   Frei-Chen edge detector

The Frei-Chen edge detector actually uses a combination of 4 pairs of wavelets.   Two pairs are the first degree partial derivatives of the Gaussian smoothing function, which are sensitive to step edges. The other two pairs are second and fourth mixed partial derivatives of the Gaussian smoothing function, which are sensitive to Dirac edges.   They can be
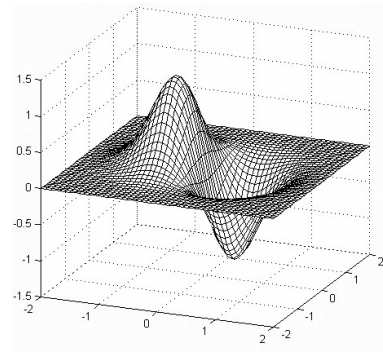
described with wavelets as follows:

$$
v_1 \quad : \quad \psi^1(x,y) = \frac{\partial \phi}{\partial y} = -\frac{1}{2\pi} y e^{-\frac{x^2+y^2}{2}},
$$

$$
v_2 \quad : \quad \psi^2(x,y) = \frac{\partial \phi}{\partial x} = -\frac{1}{2\pi} x e^{-\frac{x^2+y^2}{2}},
$$

$$
v_3 \quad : \quad \psi^3(x,y) = \frac{\partial^3 \phi}{\partial y'^3} = \frac{1}{4\sqrt{2\pi}} (6(x+y) - (x+y)^3) e^{-\frac{x^2+y^2}{2}}, \text{where } y' = \frac{x+y}{\sqrt{2}}
$$

$$
v_4 \quad : \quad \psi^4(x,y) = \frac{\partial^3 \phi}{\partial x'^3} = \frac{1}{4\sqrt{2\pi}} (6(y-x) - (y-x)^3) e^{-\frac{x^2+y^2}{2}}, \text{where } x' = \frac{y-x}{\sqrt{2}}
$$

$$
v_5 \quad : \quad \psi^5(x,y) = \frac{\partial^2 \phi}{\partial x' \partial y'} = \frac{(x^2 - y^2)}{4\pi} e^{-\frac{x^2+y^2}{2}},
$$

$$
v_6 \quad : \quad \psi^6(x,y) = \frac{\partial^2 \phi}{\partial x \partial y} = \frac{1}{2\pi} xy e^{-\frac{x^2+y^2}{2}},
$$

$$
v_7 \quad : \quad \psi^7(x,y) = \frac{\partial^4 \phi}{\partial^2 x \partial^2 y} = \frac{1}{2\pi} (1 - x^2)(1 - y^2) e^{-\frac{x^2+y^2}{2}},
$$

$$
v_8 \quad : \quad \psi^8(x,y) = \frac{\partial^4 \phi}{\partial^2 x' \partial^2 y'}
$$

$$
= \frac{1}{4\pi} (x^4 + y^4 - 2x^2 y^2 - 4x^2 - 4y^2 + 4) e^{-\frac{x^2+y^2}{2}},
$$

$$
v_9 \quad : \quad \phi(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}.
$$

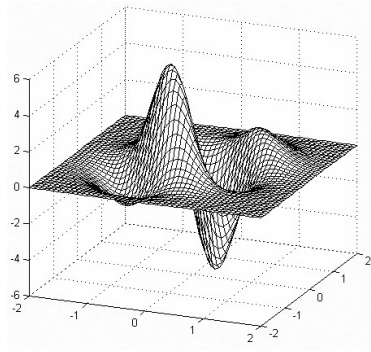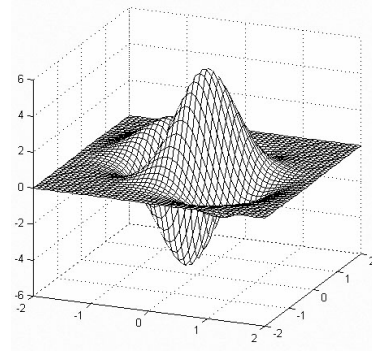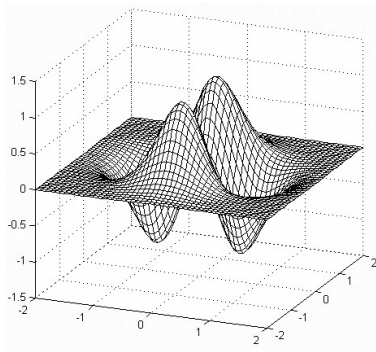Fig. 5.4 shows a 3-D plot of the wavelets and the smoothing function.

(a) v1   (b) v2

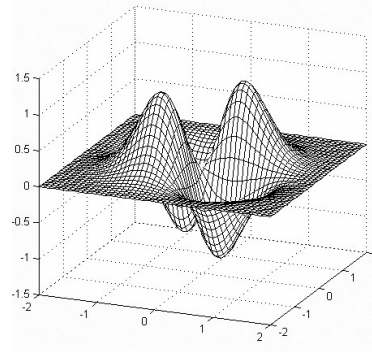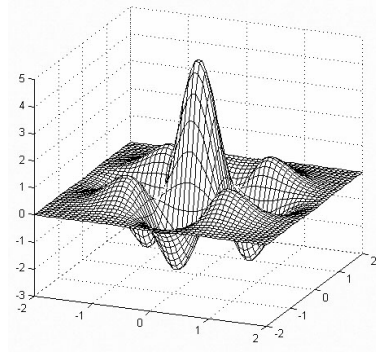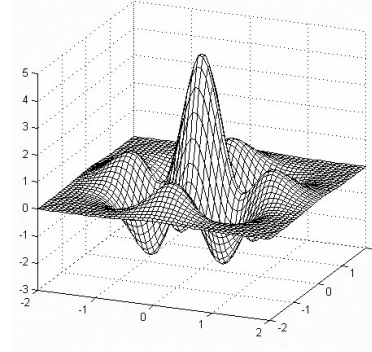(c) v3

(d) v4

(e) v5

(f) v6

(g) v7

(h) v8
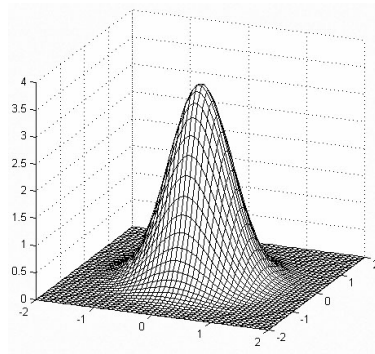
(i)v9

Fig. 5.4: Continuous wavelet model for the Frei-Chen edge detector.

5.5   Canny edge detection

Since Canny edge detection uses the first derivative of a Gaussian as as its filter, it is very close to the method of wavelet transforms.   However, from the point of view of wavelet transforms, we can use a more effective algorithm to adjust the scale of the filters.   The function in (5.87) is also an explicit form of the wavelet model for Canny edge detection.

Chapter 6
A Multiscale Wavelet Model for Edge Detection

The continuous wavelet model we constructed here not only explains the working mechanism of most classical edge detectors, but also has several significant advantages in practical applications. The scale of the wavelet used in our model can be adjusted to detect edges of different levels of scale. Also, the smoothing function used in the construction of a wavelet reduces the effect of noise. Thus, the smoothing step and edge detection step are combined together to achieve the optimal result.

6.1   Scales of edges

The resolution of an image is directly related to the proper scale for edge detection. High resolution and small scale will result in noisy and discontinuous edges; low resolution and large scale will result in undetected edges. The scale is not adjustable with classical edge detectors, but with the wavelet model, we can construct our own edge detectors with proper scales.

Because image data is always discrete, the practical scale in images is usually integer. With the cascade algorithm in chapter two and the wavelet-based edge detection method in chapter three, we can detect edges of a series of integer scales in an image. This can be useful when the image is noisy, or when edges of certain detail or texture are to be neglected.

The scale controls the significance of edges to be shown. Edges of higher significance are more likely to be kept by the wavelet transform across scales. Edges of lower significance are more likely to disappear when the scale increases. Fig. 6.1 shows edges of different significance on a dyadic sequence of scales. The significance of an edge can be mathematically measured by its Lipschitz exponent using Theorem 2 and Theorem 3.

scale: 10

scale: 20

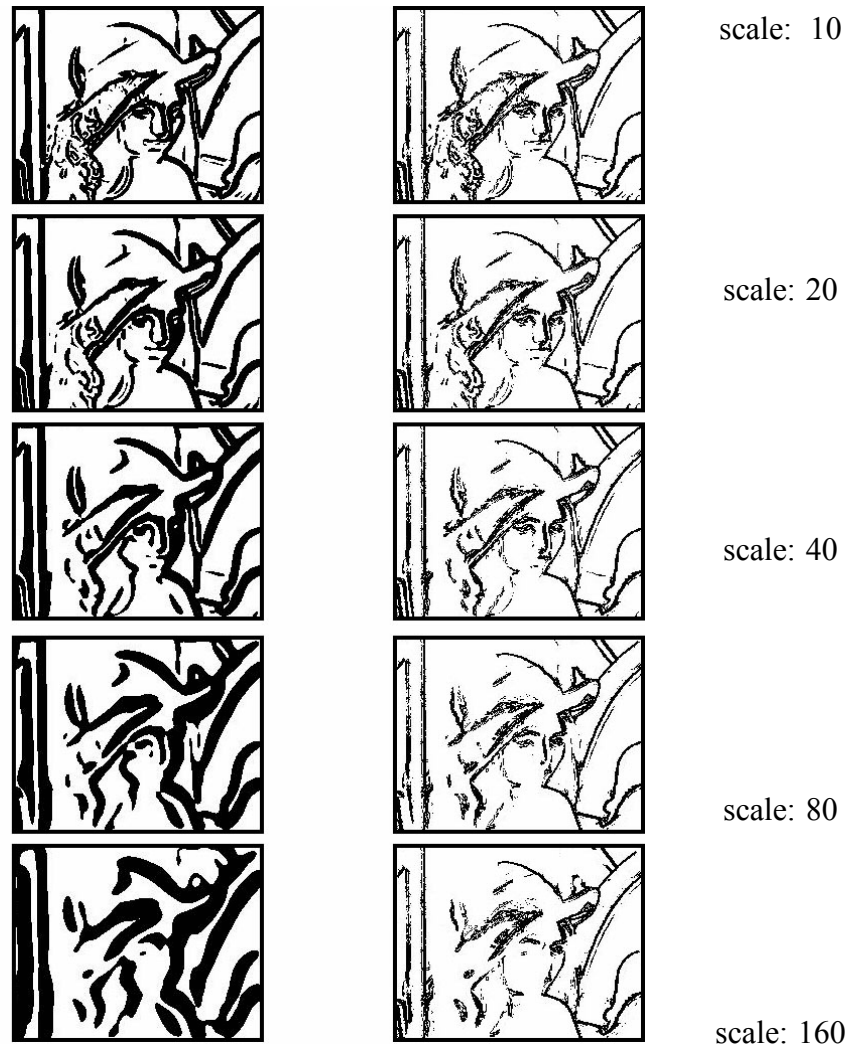scale: 40

scale: 80

scale: 160

Fig. 6.1: The left column is the thresholded edge of different scales, and the right column is the corresponding edges from high resolution; The first row indicates the result from classical edge detectors, and the following rows show the result when scales of edges are increased in the continuous model.

6.2   Effect of noise on edge detection

Noise is broadly defined as an additive (or possibly multiplicative) contamination of an image. Noise is not predictable. Hence we often use the term *"random noise"* to describe an unknown contamination added to an image. Assume the image without noise is u, and the noise is n, then the contaminated image u_0 can be represented as

$$u_0 = u + n.$$

We can measure the effect of noise by the signal-to-noise ratio (SNR).

$$SNR = \frac{||u||^2}{||n||^2} = \frac{\int \int_\Omega u^2(x,y)dxdy}{\int \int_\Omega n^2(x,y)dxdy}$$

If the noise is a Gaussian white noise, and the normalized low-pass filter is $f(x,y)$, the filter output given image $u(x,y)$ is

$$H_u = ||u * f||^2$$

If the noise variance is n_0^2, the noise output is

$$H_n = ||n * f||^2, \tag{6.88}$$

and signal-to-noise ratio is:

$$SNR = \frac{H_u}{H_n}. \tag{6.89}$$

It is reasonable to assume that the image signal has a power spectrum that concentrates on the lower frequencies, and the noise has a bounded uniform distribution in its power spectrum, i.e.

$$\int_{-\omega_0}^{\omega_0} \int_{-\omega_0}^{\omega_0} |\widehat{u}(\omega_x, \omega_y)|^2 \, d\omega_x d\omega_y \quad > \quad \frac{||\widehat{u}||^2}{4} \tag{6.90}$$

$$E\left(|\widehat{n}(\omega_x, \omega_y)|^2\right) = \begin{cases} \frac{||\widehat{n}||^2}{16\omega_0^2}, & -2\omega_0 < \omega_x < 2\omega_0 \\ & -2\omega_0 < \omega_y < 2\omega_0 \\ 0, \text{otherwise.} \end{cases} \tag{6.91}$$

where ω_0 is a parameter for an ideal low-pass filter such that

$$\left|\widehat{f}(\omega_x, \omega_y)\right|^2 = \begin{cases} \frac{1}{4\omega_0^2}, & -\omega_0 < \omega_y < \omega_0 \\ & -\omega_0 < \omega_y < \omega_0 \\ 0, \text{otherwise.} \end{cases}$$

Thus,

$$\begin{aligned} H_u &= ||u * f||^2 = \left|\left|\widehat{u} \cdot \widehat{f}\right|\right|^2 \\ &= \int_{-\omega_0}^{\omega_0} \int_{-\omega_0}^{\omega_0} \frac{1}{4\omega_0^2} |\widehat{u}(\omega_x, \omega_y)|^2 \, d\omega_x d\omega_y \\ &\geq \frac{1}{16\omega_0^2} ||\widehat{u}||^2 . \end{aligned} \tag{6.92}$$

$$\begin{aligned} H_n &= ||n * f||^2 = \left|\left|\widehat{n} \cdot \widehat{f}\right|\right|^2 \\ &= \int_{-\omega_0}^{\omega_0} \int_{-\omega_0}^{\omega_0} \left|\widehat{n}(\omega_x, \omega_y)\widehat{f}(\omega_x, \omega_y)\right|^2 \, d\omega_x d\omega_y \\ &= \frac{||\widehat{n}||^2}{16\omega_0^2} \int_{-\omega_0}^{\omega_0} \int_{-\omega_0}^{\omega_0} \left|\widehat{f}(\omega_x, \omega_y)\right|^2 \, d\omega_x d\omega_y \\ &\leq \frac{||\widehat{n}||^2}{16\omega_0^2} \end{aligned} \tag{6.93}$$

From (6.92) and (6.94), we have

$$\frac{H_u}{H_n} \geq \frac{\frac{1}{16\omega_0^2} ||\widehat{u}||^2}{\frac{1}{16\omega_0^2} ||\widehat{n}||^2} = \frac{||u||^2}{||n||^2} \tag{6.94}$$

From (6.94), we see that SNR is increased by applying the low-pass filter.

The probabilities of both Type I error (missed detection) and Type II error (false alarm) in edge detection are minimized when the signal-to-noise ratio of the detection filter is maximized. J. Canny [4] showed that the optimal filter can be approximated by the derivative of Gaussian functions. Some of the classical edge detectors do not work well with noisy images, because their SNR is not maximized through the detection filter.

Mallat, et. al. [6] have provided a method of signal denoising based on wavelet maxima. If we assume the presence of a white noise, it creates singularities with negative Lipschitz regularity. Thus, by looking at the amplitude of modulus maxima across scales, we can keep edges that propagate to a certain coarser scale and remove other edges which are dominated by noise.

scale: 2
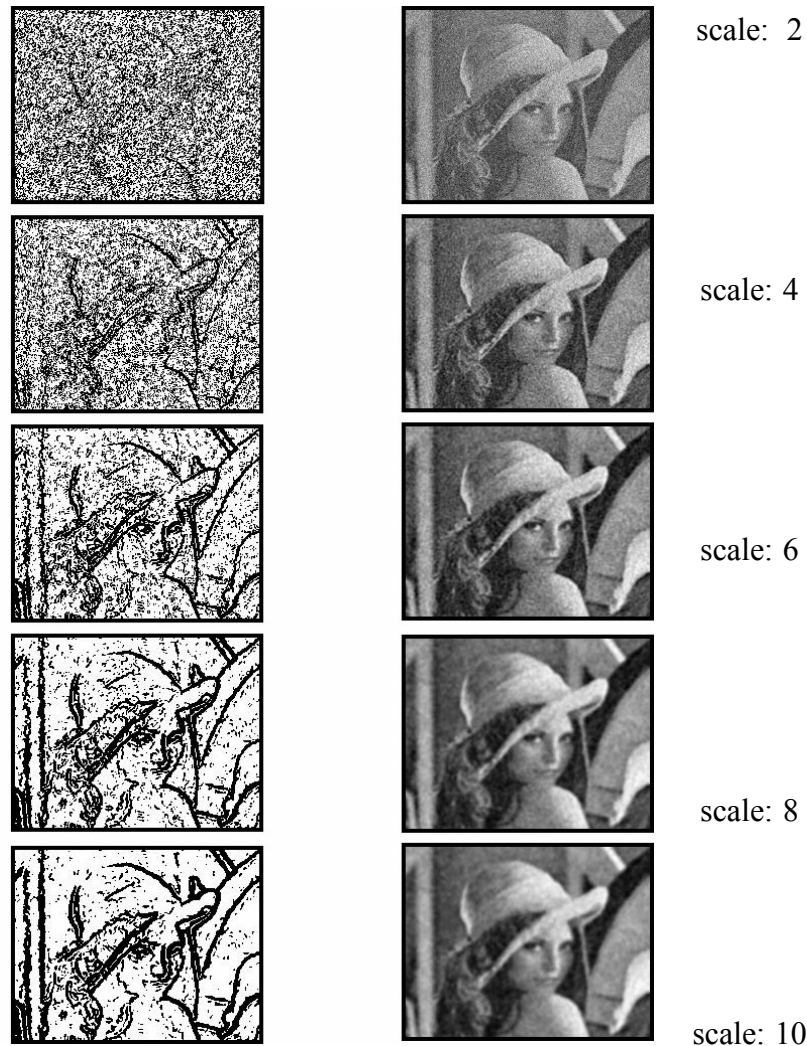
scale: 4

scale: 6

scale: 8

scale: 10

Fig. 6.2: The left column shows the results of edge detection with different scales for a noisy image; The right column shows the denoised images under corresponding edge scales; The first row is the result from classical edge detectors, and the following rows shows the results when scales of edges are increased by the continuous algorithm

6.3   Multiscale edge detection

We have explained how to obtain edges of different scales and how the effect of noise can be reduced by adjusting the scale of the edge detection wavelets. We now give a method of edge detection on a series of scales to obtain satisfactory results on an image contaminated by white noise.

Wavelet filters of large scales are more effective for removing noise, but at the same time increase the uncertainty of the location of edges. Wavelet filters of small scales preserve the exact location of edges, but cannot distinguish between noise and real edges.

From Theorem 2 and Theorem 3, we can use the coefficients of the wavelet transform across scales to measure the local Lipschitz regularity. That is, when the scale increases, the coefficients of the wavelet transform are likely to increase where the Lipschitz regularity is positive, but they are likely to decrease where the Lipschitz regularity is negative. We know that locations with lower Lipschitz regularities are more likely to be details and noise. As we can see from Fig.3.3, when the scale increases, the coefficients of the wavelet transform increase for step edges, but decrease for Dirac and fractal edges. So we can use a larger-scale wavelet at positions where the wavelet transform decreases rapidly across scales to remove the effect of noise, while using a smaller-scale wavelet at positions where the wavelet transform decreases slowly across scale to preserve the precise position of the edges. Using the cascade algorithm in chapter two, we can observe the change of the wavelet transform coefficient between each adjacent scales, and distinguish different kind

of edges. Then we can keep the scales small for locations with positive Lipschitz regularity and increase the scales for locations with negative Lipschitz regularity.

Fig. 6.3 shows that for a image without noise, the result of our method is similar to that of Canny's edge detection. For images with white noise in Fig. 6.4 - 6.8, our method gives more continuous and precise edges.   Table 6.1 shows that the SNR of the edges obtained by the multiscale wavelet transform is significantly higher than others.

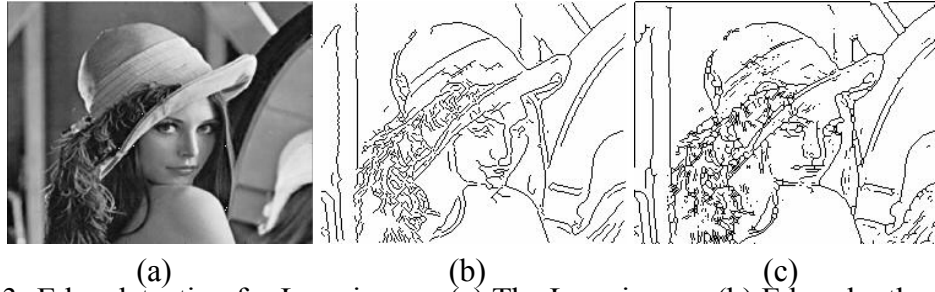(a)   (b)   (c)

Fig. 6.3: Edge detection for Lena image: (a) The Lena image; (b) Edges by the Canny edge detector; (c) Edges by the multiscale edge detection using wavelet transform

(a)

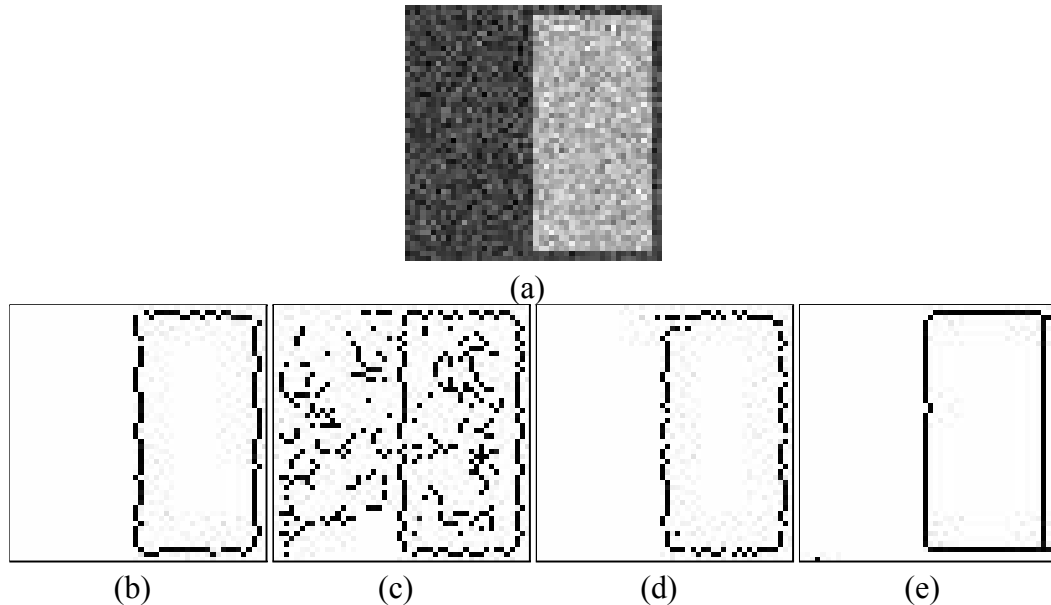(b)   (c)   (d)   (e)

Fig. 6.4: Edge detection for a block image with noise: (a) A block image (SNR=10db); (b) Edges by the Sobel edge detector; (c) Edges by Canny edge detection with default variance; (d) edges by Canny edge detection with adjusted variance; (e) Edges by the multiscale edge detection using wavelet transform
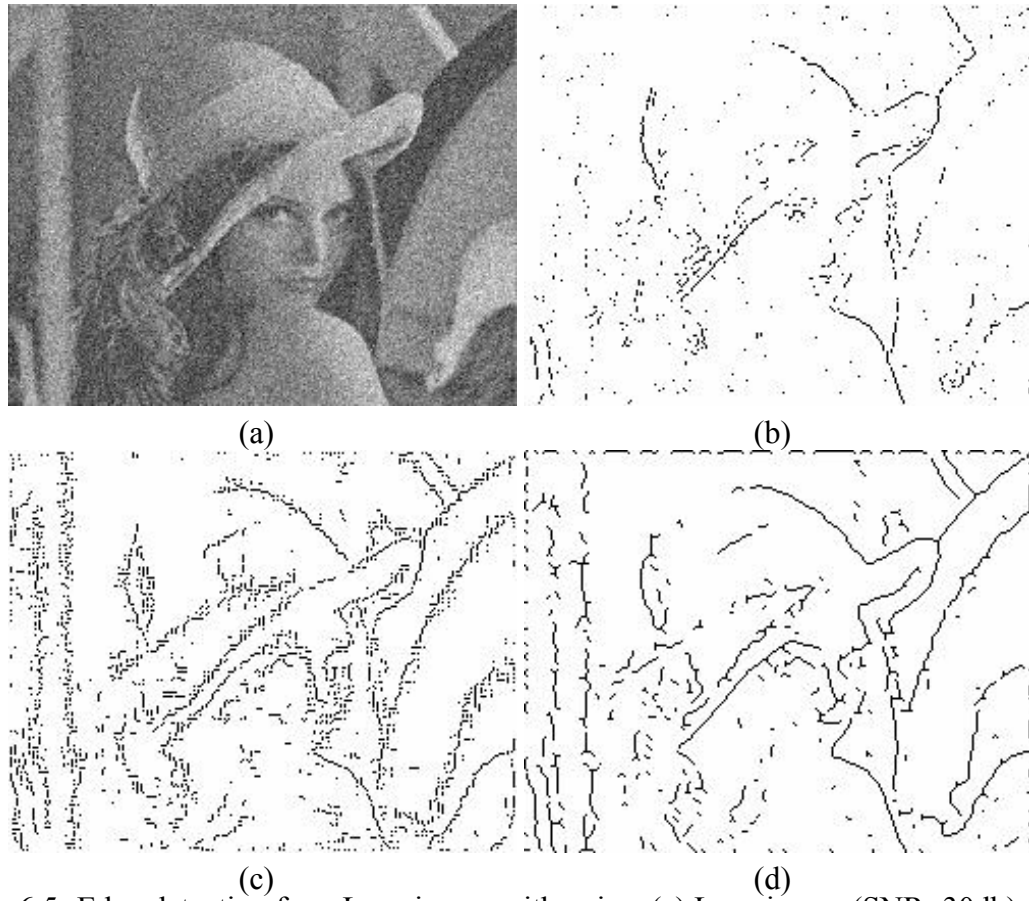
(a)   (b)

(c)   (d)

Fig. 6.5: Edge detection for a Lena image with noise: (a) Lena image (SNR=30db); (b) Edges by the Sobel edge detector; (c) Edges by Canny edge detection with adjusted variance; (d) Edges by multi-level edge detection using wavelets

Table 6.1: False rate of the detected edges

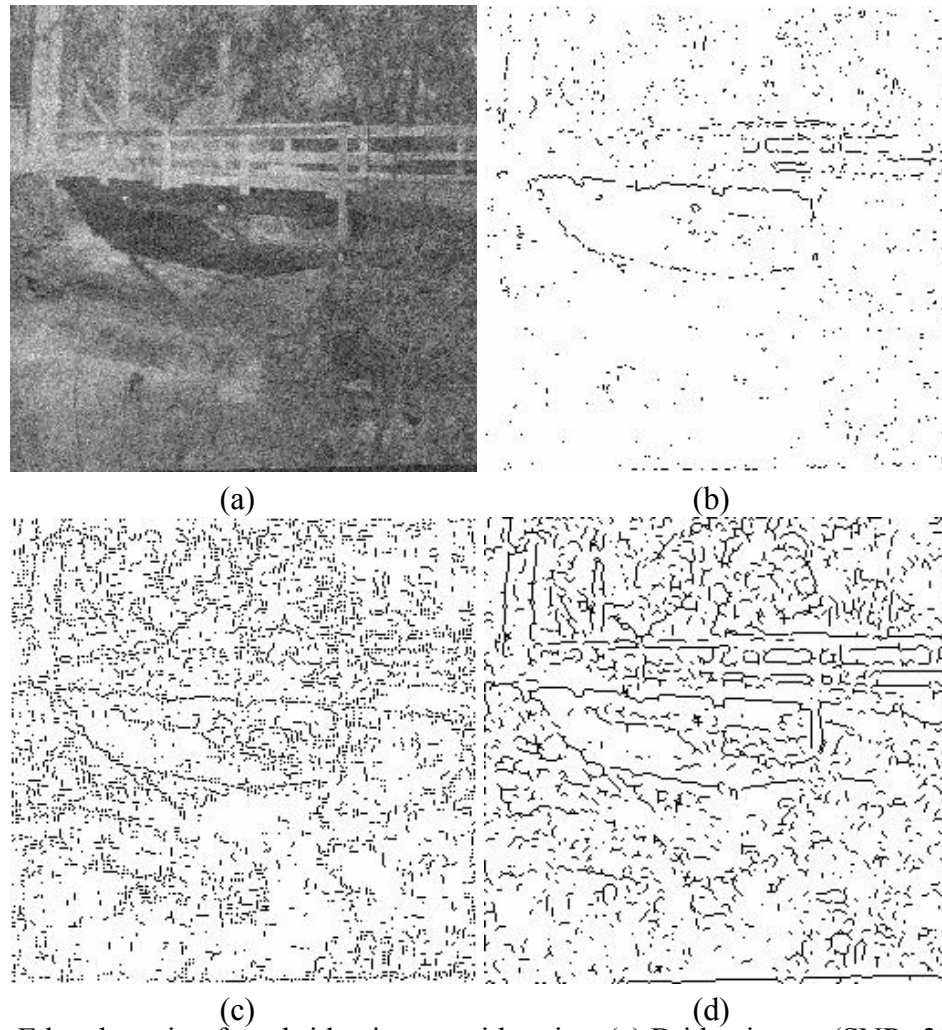| Image (SNR=10db)	Canny (default variance)	Canny (adjusted variance)	Sobel	Multiscale
Lena Image	114.3%	57.5%	60.8%	43.9%
Bridge Image	78.2%	72.3%	86.5%	68.8%
Pepper Image	188.3%	71.7%	69.2%	68.0%
Wheel Image	307.3%	63.8%	43.4%	36.7%
Block Image	197.8%	17.9%	6.0%	0.8%

Fig. 6.6: Edge detection for a bridge image with noise: (a) Bridge image (SNR=30db); (b) Edges by the Sobel edge detector; (c) Edges by Canny edge detection with adjusted variance; (d) Edges by multi-level edge detection using wavelet
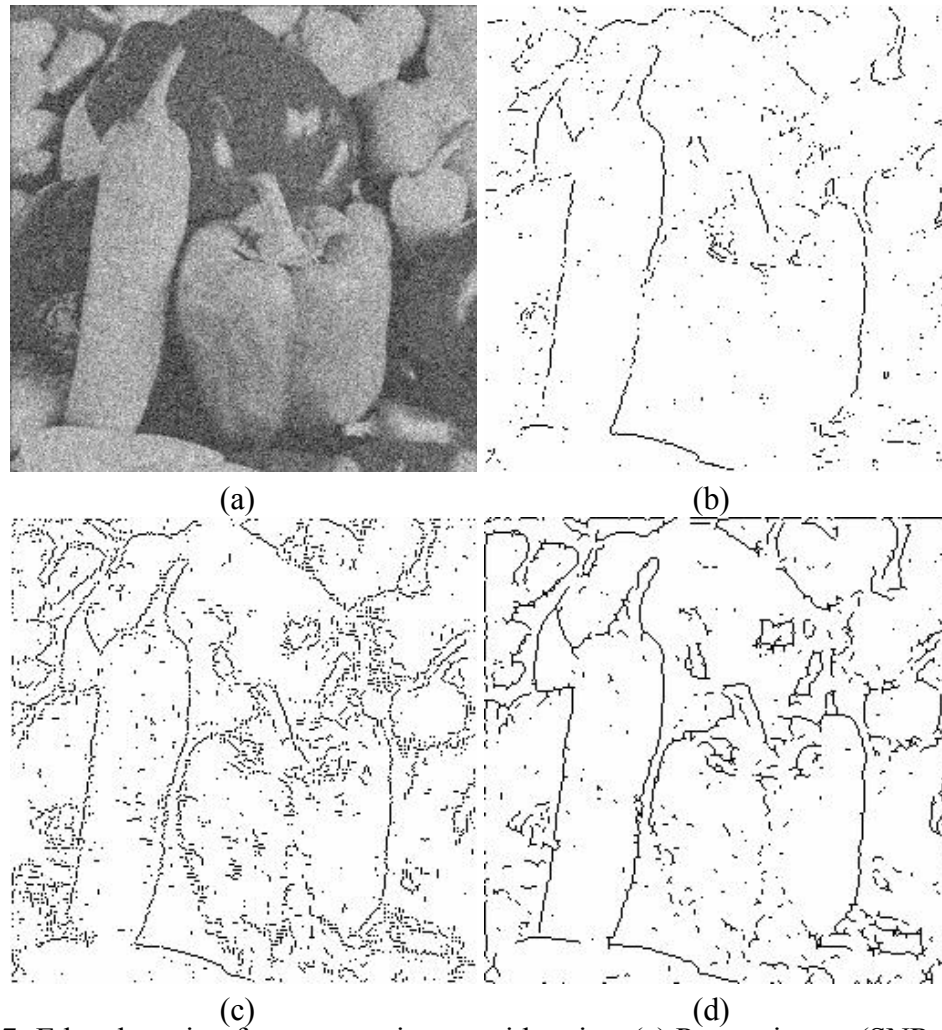
Fig. 6.7: Edge detection for a pepper image with noise: (a) Pepper image (SNR=10db); (b) Edges by the Sobel edge detector; (c) Edges by Canny edge detection with adjusted variance; (d) Edges by multi-level edge detection using wavelet

(a)   (b)

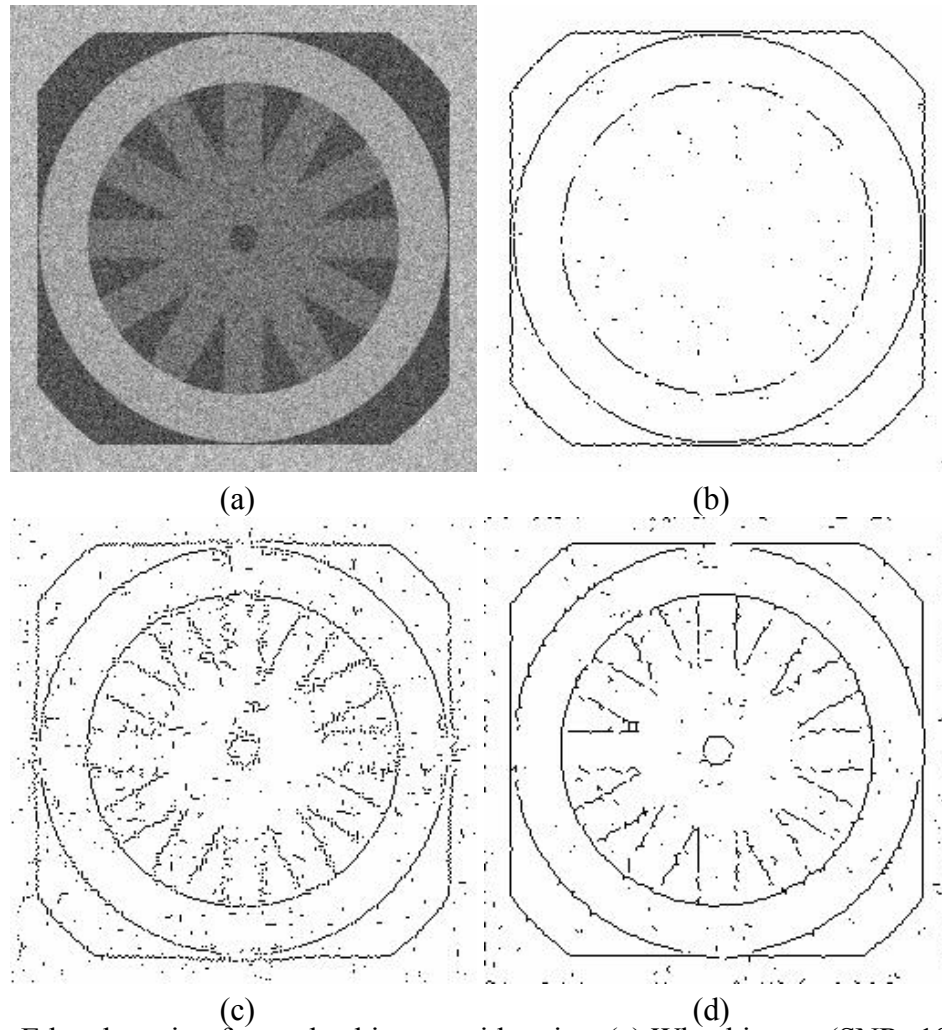(c)   (d)

Fig. 6.8: Edge detection for a wheel image with noise: (a) Wheel image (SNR=10db); (b) Edges by the Sobel edge detector; (c) Edges by Canny edge detection with adjusted variance; (d) Edges by multi-level edge detection using wavelet

References

[1] J. C. Goswami, A. K. Chan, 1999, *"Fundamentals of wavelets: theory, algorithms, and applications,"* John Wiley & Sons, Inc.

[2] G. X. Ritter, J. N. Wilson, 1996, *"Handbook of computer vision algorithms in image algebra,"* CRC Press, Inc.

[3] R. C. Gonzalez, R. E. Woods, 1993, *"Digital image processing,"* Addison-Wesley Publishing Company, Inc.

[4] J. Canny, 1986, *"A computational approach to edge detection,"* *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 679-698.

[5] S. Mallat, S. Zhong, 1992, *"Characterization of signals from multiscale edges,"* *IEEE Trans. Pattern Anal. Machine Intell.*, vol.14, no.7, pp. 710-732.

[6] S. Mallat, W. L. Hwang, 1992, *"Singularity detection and processing with wavelets,"* *IEEE Trans. Inform. Theory*, vol.38, no.2, pp. 617-643.

[7] Y. Y. Tang, L. Yang, J. Liu, 2000, *"Characterization of Dirac-Structure Edges with Wavelet Transform,"* *IEEE Trans. Sys. Man Cybernetics-Part B: Cybernetics,* vol.30, no.1, pp. 93-109.

[8] I. Daubechies, 1992, *"Ten lectures on Wavelets,"* CBMS-NSF Series in Appl. Math., #61, SIAM, Philadelphia.

[9] M. Unser, A. Aldroubi, M. Eden, 1992, *"On the asymptotic convergence of B-spline wavelets to Gabor functions,"* *IEEE Trans. Inform.* pp. 864-871.

[10] C.K. Chui, 1992, *"An Introduction to Wavelets,"* Academic Press, Boston.

[11] C.K. Chui, J. Z. Wang, 1997, *" A study of asymptotically optimal time-frequency localization by scaling functions and wavelets,"* *Annals of Numerical Mathematics* 4(1997), pp. 193-216.

[12] C. K. Chui, J. Z. Wang, 1993, *"An analysis of cardinal spline-wavelets,"* *J. Approx.* pp. 54-68.

[13] A. Cohen, R. D. Ryan, 1995, *" Wavelets and Multiscale Signal Processing,"* Chapman & Hall.

[14] J. J. Benedetto, M. W. Frazier, 1994, *"Wavelets-Mathematics and Applications,"* CRC Press, Inc.

[15] R. J. Beattie, 1984, *"Edge detection for semantically based early visual processing,"* dissertation, Univ. Edinburgh, Edinburgh, U.K..

[16] B. K. P. Horn, 1971, *"The Binford-Horn line-finder,"* Artificial Intell. Lab., Mass. Inst. Technol., Cambridge, AI Memo 285.

[17] L. Mero, 1975, *"A simplified and fast version of the Hueckel operator for finding optimal edges in pictures,"* *Pric. IJCAI*, pp. 650-655.

[18] R. Nevatia, 1977, *"Evaluation of simplified Hueckel edge-line detector,"* *Comput., Graph., Image Process.*, vol. 6, no. 6, pp. 582-588.

[19] Y. Y. Tang, L.H. Yang, L. Feng, 1998, *"Characterization and detection of edges by Lipschitz exponent and MASW wavelet transform,"* *Proc. 14th Int. Conf. Pattern Recognit.*, Brisbane, Australia, pp. 1572-1574.

[20] K. A. Stevens, 1980, *"Surface perception from local analysis of texture and contour,"* Artificial Intell. Lab., Mass. Instr. Technol., Cambridge, Tech. Rep. AI-TR-512.

[21] K. R. Castleman, 1996, *"Digital Image Processing,"* Englewood Cliffs, NJ: Prentice-Hall.

[22] M. Hueckel, 1971, *"An operator which locates edges in digital pictures,"* *J. ACM*, vol. 18, no. 1, pp. 113-125.

[23] Y. Meyer, 1987, *"Ondelettes, fonctions splines et analyses graduées,"* Rapport Ceremade 8703.

[24] Y. Meyer, 1987, *"Ondelettes et Opérateurs"*, Paris: Hermann.

[25] A. Zygmund, 1968, *"Trigonometric Series,"* 2nd ed., Cambridge: Cambridge Univ. Press.

[26] G. Beylkin, R. Coifman, V. Rokhlin, 1991, *" Fast wavelet transform and numerical algorithms: I,"* *Commun. Pure and Appl. Math.* 44, pp. 141-183.

[27] M. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer, L. Raphael, 1992, *"Wavelets and Their Applications,"* Boston, MA: Jones and Bartlett.

[28] D. Marr, E. C. Hildreth, 1980, *"Theory of edge detection," Proc. R. Soc.,* B 207, pp. 187-217, London, U.K..

[29] D. Marr, S. Ullman, 1981, *"Directional selectivity and its use in early visual processing," Proc. R. Soc.* Lond. B, vol. 208.

[30] D. Marr, T. Poggio, 1979, *"A theory of human stereo vision," Proc. R. Soc.*, London, U.K., B 204, pp. 301-328.

[31] E. C. Hildreth, 1980, *"Implementation of a theory of edge detection,"* M.I.T. Artificial Intell. Lab., Cambridge, MA, Tech. Rep. AI-TR-579.

[32] M. A. Shah, R. Jain, 1984, *"Detecting time-varing corners," Proc. 7th IJCPR*, pp. 2-4.

[33] R. M. Haralick, 1984, *"Digital step edges from zero crossing of second directional derivatives," IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, no. 1, pp. 58-68.

[34] D. Gabor, 1946, *"Theory of communication," JIEE (London)* 93, pp. 429-457.

[35] A. Haar, 1910, *"Zur theorie der orthogonalen Funktionensysteme,"* Math. Ann. 69: pp. 331-371.

[36] S. Mallat, 1989, *" Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$," Trans. Amer. Math. Soc.* 315: pp. 69-87.

[37] S. Mallat, 1998, *"A Wavelet Tour of Signal Processing,"* New York: Academic.

[38] C. K. Chui, J. Z. Wang, 1991, *"A cardinal spline approach to wavelets," Proc. Amer. Math. Soc.*, 113: pp. 785-793.

[39] D. Esteban, C. Galand, 1977, *" Applications of quadrature mirror filters to split-band voice coding schemes," Int. Conf. Acoust. Speech Signal Process.*, Hartford, CT, pp. 191-195.

[40] C. K. Chui, 1992, *"Wavelets: A tutorial in Theory and Applications,"* Academic Press.

Appendix A
Matlab code for the cascade algorithm for the wavelet model

```
%y=psi(axis,a,h)

%axis:  1 for x, 2 for y,

%       3 for (x+y), 4 for (x-y)

%a:      variance

%h:      order of derivative

%       support = (a+h+1)x(a+h+1)

%e.g.

%       y=psi(1,10,0) gives you a Gaussian

%        with support of 11 by 11

%       y=psi(3,6,4) give you a 4th derivative

%        of Gaussian rotated pi/4.

function y=psi(axis,a,h)

   v=[1 1;1 1]/4;

   d1=[0 0 0; 0 1 -1; 0 0 0]/2;

   d2=[0 0 0; 0 1 0; 0 -1 0]/2;

   d3=[0 0 0; 0 1 0; 0 0 -1]/2;

   d4=[0 0 0; 0 1 0; -1 0 0]/2;

   d=eval(['d' '0'+axis]);
```

```
if h>0

    y=d;

    for i=1:h-1

        y=conv2(y,d);

    end

else

    y=v;

    a=a-1;

end

for i=1:a

    y=conv2(y,v);

end
```

Appendix B
Matlab code for multiscale edge detection

```
%y=multiedge(a,s)

%a:      original image

%s:      maximum level of wavelet scale

function y=multiedge(a,s);

h1=[1 0 -1; 2 0 -2; 1 0 -1];

h2=[1 2 1;0 0 0;-1 -2 -1];

v=[1 2 1;2 4 2;1 2 1];

a1=filter2(h1,a);

a2=filter2(h2,a);

[m n]=size(a);

ns=h1;

for ii=1:s

   nsh=conv2(ns,v);

   p=max(max(ns))/max(max(nsh));

   % Increase scale of the wavelet

   ah1=filter2(v,a1)*p;

   ah2=filter2(v,a2)*p;

   ns=nsh;

   % Distinguish noise and real edges
```

```
    a1=a1.*(abs(a1)<=abs(ah1))+ah1.*(abs(a1)>abs(ah1));

    a2=a2.*(abs(a2)<=abs(ah2))+ah2.*(abs(a2)>abs(ah2));

end

e=(a1.^2+a2.^2);

% Threshold

y=(e>(filter2(ones(10))/100,e)).*(e>mean2(e));
```

Vita

Jun Li was born in Shanghai, China, on July 31st, 1977. In 1994, he attended Tongji University in Shanghai, China, and received a Bachelor of Engineering Degree in Urban Planning in 1999. In August 2001, he entered Sam Houston State University and in August 2003 received his Master of Science Degree in Mathematics.