National Taiwan University

Time Frequency Analysis and Wavelet Transforms
(Autumn 2009)

Term Paper
Wavelet for Edge Detection

Student Name	**Feng-Ju,Chang 張鳳洳**
Student ID	**R98942063**
Class	**GICE 1st grade**
Instructor	**Jian-Jiun Ding**

Contents

***Abstract*—** There are many edge detection methods nowadays such as Sobel, Roberts, Laplacian etc. However, these gradient-based and zero-crossing finding algorithms are very sensitive to noise. They may misjudge the noise points as the part of real edges and miss some real edges because of the noises' interference. Furthermore, the masks' sizes are fixed and cannot be dilated for the need of varieties of problem domains. So, a better edge detection method, Canny edge detector occurs. It improves the resistance of the noise and the scale of the Gaussian filter is able to be adjusted. However, the Canny operator still suffers from some practical limitations. The wavelet transform which has the good locality and multi-scale identity is recognized as an efficient way to detect edges currently. It satisfies the need of edge detection in multi-scales so that the noise can be avoided, more real edges will be kept, and the practical limitation of Canny can be overcome. Therefore, the wavelet-based edge detection scheme is considered better than the traditional gradient-based, zero-crossing finding as well as Canny operators. Although the wavelet transform is able to be applied alone in the edge detection, it is not necessary for every kind of images. So some auxiliary methods such as Lipschitz regularity, adaptive thresholding, scale multiplication are needed or using the deformation of the wavelet like the directional wavelets and the complex wavelets to make the performance of the edge detection better and better.

Keywords: Gradient-based, Canny, The dyadic wavelet transform, scale multiplication, the directional wavelet, the complex wavelet

1. Introduction

1.1 What are Edges and Edge Detection?

Edges are curves that follow a path of rapid change in intensity. In image processing, an edge is often interpreted as one class of singularities and they can be characterized easily as discontinuities where the gradient approaches infinity. However, the images we encounter with now are all discrete, so edges in an image are defined as the local maxima of the gradient.

Because edge is the most basic feature of images and it is the main tool in pattern recognition, image segmentation, and scene analysis, we need to do the edge detection. An edge detector is basically a high pass filter that can be applied to extract the edge points in an image. It significantly reduces the amount of data and filters out useless information, while preserving the important structural properties in an image [20].

1.2 Conventional Edge detectors

Conventional edge detection mechanisms examine the pixels where the first derivative of the intensity is larger in magnitude than some threshold, or finding places where the second derivative of the intensity has a zero crossing. **Traditional edge-detection algorithms** usually belong to one of the following classes: **gradient-based edge detectors**, **Laplacian of Gaussian (LOG)**, **zero crossing**, and **Canny edge detector** [6]. The gradient-based method is an early technique to detect edges. Basically, this method uses some specifically designed masks to traverse the image and detects matching edge patterns by locating the maxima of the gradient magnitude of the image. The whole process is realized by the discrete convolution with a set of directional derivative masks. The well known edge detection operators based on the gradient calculation are Roberts, Prewitt, Sobel, Frei-Chen operators, and etc. Another kind of measures, the Laplacian of Gaussian combines the Gaussian filtering with a search for zero crossings in the second derivative of the image. Like LoG, the zero crossing method searches for the zero crossing in the second derivative of the image but without the Gaussian smoothing. Overall, they are efficient and easy to apply due to their forms of a 3x3 pattern grid. In certain but least situations where the edges are highly directional, some edge detector can work especially well in that the designed pattern fit the real edges better. Unfortunately, when there are noises in images, the gradient based method is very sensitive to noises and may detect noise points as the part of edges; what's worse, some real edges are missed simultaneously due to the corruption of the noise.

The last classification of the traditional edge detector is Canny edge detector. It can resist the noise because the Gaussian function is used to smooth the image. This method is optimal for step edges corrupted by white noise [17]. John F. Canny followed a list of criteria in order to improve the gradient-based edge detectors. The first is **good detection (low error rate)**. It is important that the edges which occur in an image should not be missed and that there should be no spurious responses. So the algorithm needs to mark as many real edges in the image as possible. The second criterion is **good localization**, it means the distance between the actual and located position of the edge should be minimal. The third is **minimal response**, a given edge in the image should only be marked once, and image noise should not create false edges. The third criterion is implemented because the first two criteria were not substantially enough to completely eliminate the possibility of multiple responses to an edge [17].

After reducing the noise through the Gaussian function, Canny edge detector finds the image gradient to highlight regions with high spatial derivatives. And then it tracks along these regions, suppresses any pixel that is not at the maximum **(non-maximum suppression)**. The gradient array is now further reduced by **hysteresis thresholding**. Hysteresis thresholding is used to track along the remaining pixels that have not been suppressed. It uses two thresholds (low and high thresholds) and the low threshold is usually 0.5 multiplied by the high threshold (0.4 in MATLAB). If the magnitude is below the low threshold, it is set to zero (made it a nonedge). If the magnitude is above the high threshold, it is made an edge. And if the magnitude is between the high and low thresholds, then it is set to zero unless there is a path from this pixel to a pixel with a gradient above high threshold [17].

Although Canny's method is known to many as the better edge detector than the gradient-based, LoG, and zero-crossing methods, it still suffers from some practical limitations. Firstly, close edges may affect each other in the process especially when the standard deviation of the Gaussian function is too large which results in inaccurate edge locations and some edge losses. Secondly, the hysteresis thresholding requires not only the trial and error adjustment of two thresholds to produce a satisfactory edge result for each different input image but also the control of the imaging environment to assure the validity of the pre-adjusted thresholds [3].

1.3 Edge Detectors With Wavelets

To overcome the disadvantages of the edge detectors mentioned in 1.2. More edge detecting or generally singularity investigation techniques are developed. The Fourier transform is one of them. It was viewed as the main mathematical tool for analyzing singularities. However, the Fourier transform is too global to be well adapted on local singularities. It is hard to find the location and spatial distribution of singularities using Fourier transform. The wavelet transform, on the other hand, is a local analysis. It is more suitable than the Fourier transform for the time-frequency analysis, which is essential for singularity detection. This was a major motivation for the study of the wavelet transform in mathematics and in applied domains [2][16].

For several decades, the wavelet transforms have been found to be remarkable mathematical tools to analyze singularities including the edges, corners, and etc. Its idea is actually similar to that of Canny approach which can be viewed as a multi-scale version of the gradient-based approaches for the standard deviation of the Gaussian is adjustable. This idea is extended to the wavelet maxima representation in two dimensions [2]. **Mallat, Hwang, and Zhong [1] have proved that the maxima of the wavelet transform modulus is certainly be used to detect the location of the irregular structures.** Further, these maxima are used to calculate Lipschitz exponents of edges. **In mathematics, the sharpness of an edge can be described with Lipschitz exponent [2].** The locations with lower Lipschitz regularities are more likely to be details and noise.

The main reason and advantage for applying the wavelet transform to the detection of edges in an image is the possibility of choosing the size of the details that will be detected. How many edges we want to get is set by the wavelet scale. In the case of the discrete wavelet transform, the choice of the scale is performed by multiple signal passage through the wavelet filter [9]. **When processing a 2-D image, the wavelet analysis is performed separately for the horizontal and the vertical directions. Thus, the vertical and the horizontal edges are detected separately.**

The 2D discrete wavelet transform (DWT) decomposes the images into sub-images, 3 details and 1 approximation. The approximation looks similar to the input image but only 1/4 of original size. The 2-D DWT is an extension of the 1-D DWT in both the horizontal and the vertical direction. We label the resulting sub-images from an octave (a single iteration of the DWT) as LL (the approximation or we say the smoothing image of the original image which contains the most information of the original image), LH (preserves the horizontal edge details), HL (preserves the vertical edge details), and HH (preserves the diagonal details which are influenced by noise greatly), according to the filters used to generate the sub-image [4][6]. For example, HL means that we used a high pass filter along the rows, and a low pass filter along the columns. This process can repeat continuously by putting the first octave's LL sub-image through another set of low pass and high pass filters. These iterative procedures construct the multi-resolution analysis. The diagram is shown in Fig. 1.

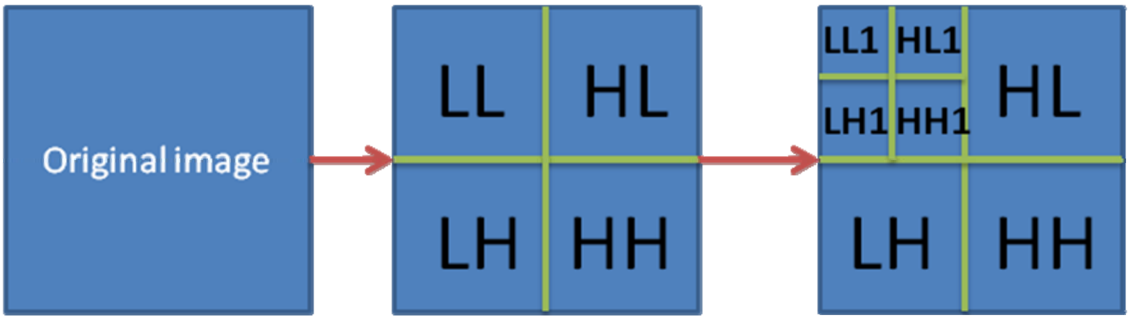


Fig. 1: Image decomposition based on the wavelet transform

1.4 The Organization of This Tutorial

In this tutorial, I will introduce the dyadic wavelet transform in the next section. The section 3 will characterize the edges with Lipschitz regularity and incorporate it into the multi-scale wavelet model for edge detection. For section 4~8 , the wavelet transform combined with other methods or the deformation of itself are introduced to make the edge detection results much better than that of the traditional wavelet based method. In section 9, two performance analysis for edge detection will be introduced.

2. The Construction of Wavelets for Edge Detection

2.1 The Continuous Wavelet transform

There are numerous types of wavelet transforms. The first is the continuous wavelet transform (CWT).

$$W_\psi f(a,b) = \frac{1}{\sqrt{b}} \int_{-\infty}^{\infty} x(t)\psi\left(\frac{t-a}{b}\right)dt \tag{1}$$

where $x(t)$ is the input, $\psi(t)$ is the mother wavelet, a is the location, b is the scaling. The value of a is any real number and b is any positive real number. The resolution of the wavelet transform is invariant along a (location axis) but variant along b (scaling axis). (See Fig. 2)

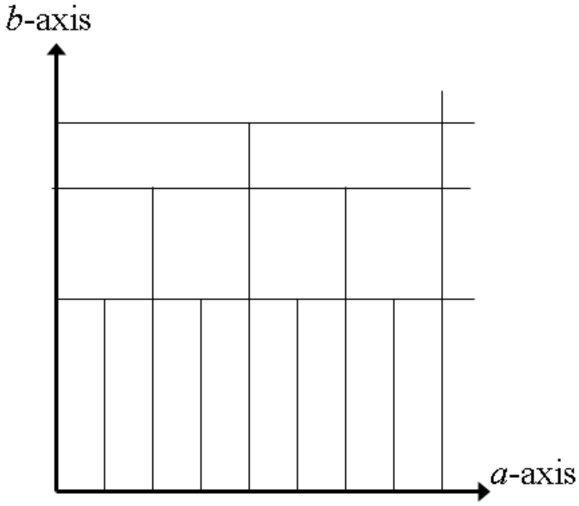


Fig. 2(This figure is extracted from the slides of the TFW lecture)

In Fig. 2, we can see that when b is small, the width of the mother wavelet $\psi(t)$ is narrow, which means bad time resolution and good frequency resolution. On the other hand, if b is large, the width of the mother wavelet $\psi(t)$ is wide, which means good time resolution and bad frequency resolution. There are many ways to choose the mother wavelet, such as Haar basis and the Mexican hat function,

$$\psi(t) = \frac{2^{5/4}}{\sqrt{3}}\left(1 - 2\pi t^2\right)e^{-\pi t^2} \tag{2}$$

Actually, the Mexican hat is the second order derivative of the Gaussian function (a smoothing function). Therefore, we can choose derivatives of the Gaussian function to be the mother wavelet,

$$\psi(t) = \frac{d^p}{dt^p}e^{-\pi t^2} \tag{3}$$

and its vanish moment is directly p. In fact, the wavelet transform is a high pass filter in that it's vanish moment is more than 1. So if we want to preserve or reconstruct the low frequency information, the scaling function (a low pass filter) is needed.

Although the continuous WT is good in mathematics, it is hard to implement and finding the scaling function is difficult. Hence, in practical, the discrete WT and the continuous WT with discrete coefficients are more useful.

2.2 The Muti-scale Edge Detection With The Dyadic Wavelet Transform

The continuous WT with discrete coefficient which uses the dyadic scale $2^j, j \in Z, j \in (-\infty, \infty)$ to be the scaling is the dyadic wavelet transform. Many papers view this type of transform to be the discrete wavelet transform directly. Nevertheless, it is just the special type of the continuous WT.

2.2.1 1-D Edge Detection

I have mentioned in 2.1 that we usually use the smoothing function to derive the mother wavelet. So let's do it. Assume that the smoothing function is twice differentiable. A smoothing function can be viewed as the impulse response of a low pass filter. Let $\theta_s(x) = (1/s)\theta(x/s)$, which is the s-dilation of $\theta(x)$ and let the 1-D signal $f(x)$ be a real function in $L^2(R)$. $L^2(R)$ denotes the Hilbert space of measurable functions, such that [1][2]

$$\int_{-\infty}^{\infty}|f(x)|^2 dx < \infty \tag{4}$$

Edges at the scale s are defined as local sharp variation points of $f(x)$ smoothed by $\theta_s(x)$. Let

$$\psi^1(x) = \frac{d\theta(x)}{dx} \quad \text{and} \quad \psi^2(x) = \frac{d^2\theta(x)}{dx^2} \tag{5}$$

The functions $\psi^1(x)$ and $\psi^2(x)$ can be considered to be wavelets because their integral is equal to 0, i.e. their vanishing moments are at least 1.

$$\int_{-\infty}^{\infty}\psi^1(x)dx = \int_{-\infty}^{\infty}\psi^2(x)dx = 0 \tag{6}$$

The wavelet transform is computed by convolving the signal with a dilated wavelet. So, the wavelet transform of $f(x)$ at the scale $s = 2^j, j \in Z, j \in (-\infty, \infty)$ and position x with respect to each of these wavelets in (5) are given by [1]

$$W_s^1 f(x) = f * \psi_s^1(x) \quad \text{and} \quad W_s^2 f(x) = f * \psi_s^2(x) \tag{7}$$

$$W_s^1 f(x) = f * \left(s \frac{d\theta_s}{dx} \right)(x) = s \frac{d}{dx} (f * \theta_s)(x) \tag{8}$$

and

$$W_s^2 f(x) = f * \left(s^2 \frac{d^2\theta_s}{dx^2} \right)(x) = s^2 \frac{d^2}{dx^2} (f * \theta_s)(x) \tag{9}$$

From equation (8) and (9), we can see that the wavelet transforms $W_s^1 f(x)$ and $W_s^2 f(x)$ are proportional to the first and second derivative of $f(x)$ smoothed by $\theta_s(x)$ respectively. For a fixed scale s, the local extrema of $W_s^1 f(x)$ along the x variable, correspond to the zero-crossings of $W_s^2 f(x)$ and to the inflection points of $(f * \theta_s)(x)$ (see Fig. 3) [1]. When the smoothing function $\theta(x)$ is a Gaussian function, the determination of the local extrema of $W_s^1 f(x)$ is equivalent to the Canny edge detection.

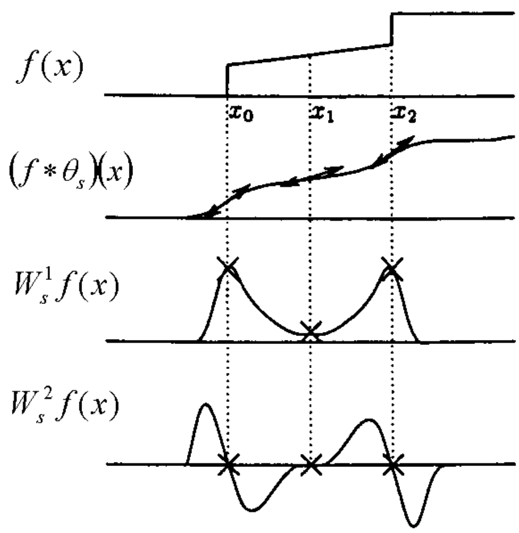


Fig. 3 : Extrema of $W_s^1 f(x)$ and the zero-crossings of $W_s^2 f(x)$ are the inflection points of $(f * \theta_s)(x)$. Points of abscissa x_0 and x_2 are sharp variations of $(f * \theta_s)(x)$ and are local maxima of $|W_s^1 f(x)|$. The local minimum of $|W_s^1 f(x)|$ at x_1 is also an inflection point but it is a slow variation point. (This figure is extracted from p.622 of [1] : http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=119727&isnumber=3425)

For a 1-D signal, there are usually four kinds of edges, which can be indicated in one dimension as in Fig. 4 (a). 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. 4 (b) shows the wavelet transform of this signal on different edges in dyadic scales. Fig. 4 (c) shows the modulus maxima which indicates the location of the edges. It is clear that by examining the wavelet transform, we can extract a lot of information about edges. For example, we can see whether it is a gradual change or leap, or whether it is a giant cliff, or a momentary spike, by looking only at the wavelet representation.

2.2.2 2-D Edge Detection

Now consider the dyadic wavelet transform in the 2-D image. Let the 2-D smoothing function to be any function whose double integral is nonzero. We define two wavelets that are , respectively, the partial derivatives along the *x* and *y* of a 2-D smoothing function $\theta(x,y)$:

$$\psi^1(x,y)=\frac{\partial\theta(x,y)}{\partial x} \text{ and } \psi^2(x,y)=\frac{\partial\theta(x,y)}{\partial y} \tag{10}$$

Let $\psi_s^1(x,y)=(1/s)^2\psi^1(x/s,y/s)$, $\psi_s^2(x,y)=(1/s)^2\psi^2(x/s,y/s)$ which are the s-dilation of $\psi^1(x,y)$ and $\psi^2(x,y)$ respectively. In addition, $s=2^j, j\in Z, j\in(-\infty,\infty)$. Let the 2-D signal $f(x,y)\in L^2(R^2)$. $L^2(R^2)$ denotes the Hilbert space of 2-D square-integrable functions, such that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|f(x,y)|^2\,dxdy<\infty \tag{11}$$

Then, the wavelet transform defined with respect to $\psi^1(x,y)$ and $\psi^2(x,y)$ has two components:

$$W_s^1 f(x,y)=f*\psi_s^1(x,y) \text{ and } W_s^2 f(x,y)=f*\psi_s^2(x,y) \tag{12}$$

Similar to equation (7)~(9), we can easily prove that [1]

$$\begin{pmatrix} W_s^1 f(x,y) \\ W_s^2 f(x,y) \end{pmatrix} = s\begin{pmatrix} \frac{\partial}{\partial x}(f*\theta_s)(x,y) \\ \frac{\partial}{\partial y}(f*\theta_s)(x,y) \end{pmatrix} = s\nabla(f*\theta_s)(x,y) \tag{13}$$

If we want to locate the positions of rapid variation (edges) of an image $f(x,y)$, we should consider the local maxima of the gradient magnitude at various scales which is given by

$$M_s f(x,y)=\left\|\begin{pmatrix} W_s^1 f(x,y) \\ W_s^2 f(x,y) \end{pmatrix}\right\| = \|s\nabla(f*\theta_s)(x,y)\| = \sqrt{(W_s^1 f(x,y))^2+(W_s^2 f(x,y))^2}\ , \tag{14}$$

$s=2^j, j\in Z, j\in(-\infty,\infty)$. The function $M_s f(x,y)$ is also called t**he modulus of the wavelet transform at the scale** s [1].

A point (x, y) is a multi-scale edge point at scale s if the magnitude of the gradient $M_s f(x, y)$ attains a local maximum there along the gradient direction $A_s f(x, y)$, defined by

$$A_s f(x,y) = \tan^{-1}\left(\frac{W_s^2 f(x,y)}{W_s^1 f(x,y)}\right) \tag{15}$$

The angle in (15) can also be used to recover the two components $W_s^1 f(x, y)$ and $W_s^2 f(x, y)$ from the modulus $M_s f(x, y)$. Here, we choose the Gaussian function as the smoothing function for Mallat's wavelet, i.e. [1]

$$\theta(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{16}$$

An example of the multi-scale edge detection is represented in Fig. 5.

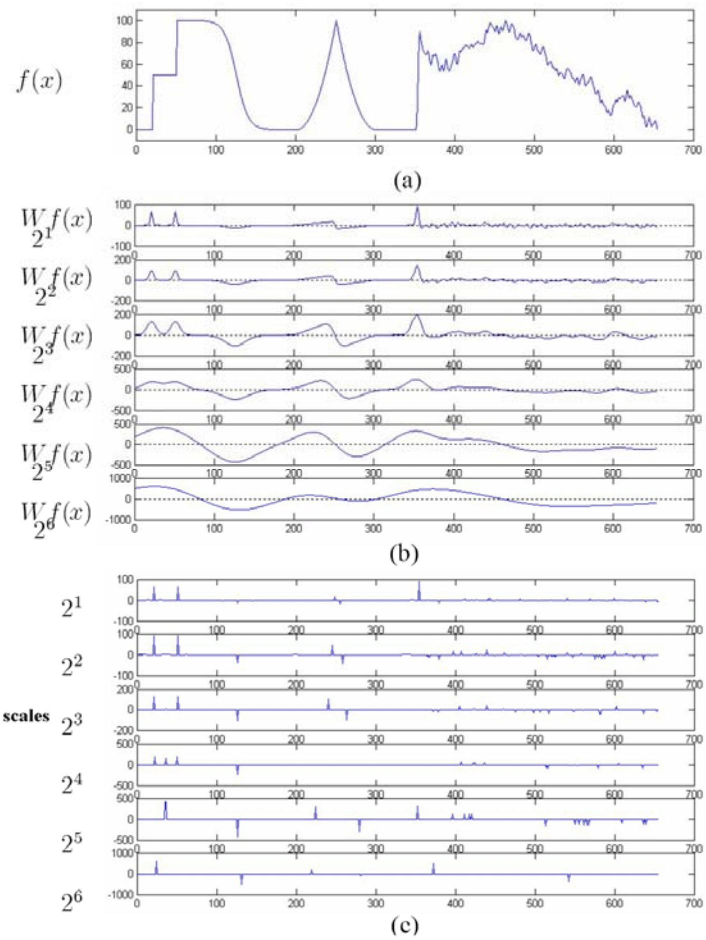
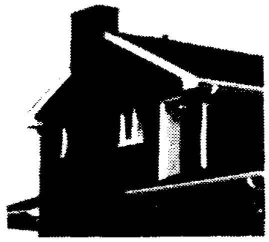
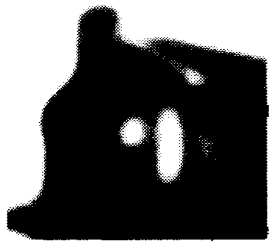
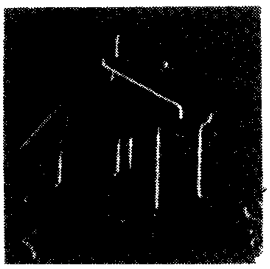
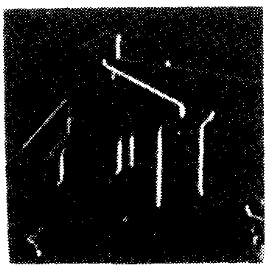


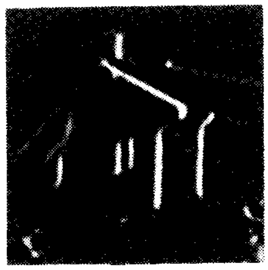
Fig. 4: Wavelet transform of different types of edge throughout scales: (a) Original image (b) Wavelet Transform computed up to the scale 2^6. (c) At each scale, each Dirac indicates the position of a modulus location. (This figure extracted from p.29 of [2] : http://eref.uqu.edu.sa/files/Others/Image%20Processing/A%20wavelet%20approach%20to%20edge%20detection%20-%20Thesis.pdf)

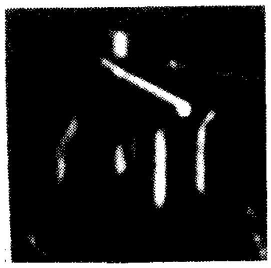

(a)

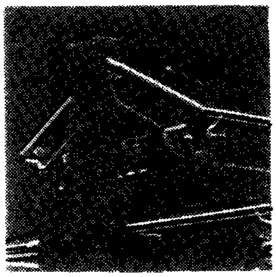

(b)

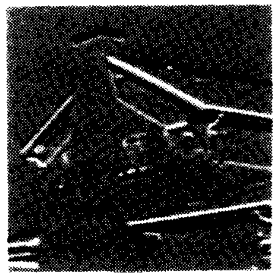

(c)

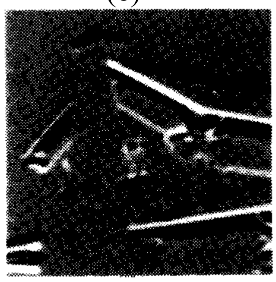

(d)


(e)


(f)

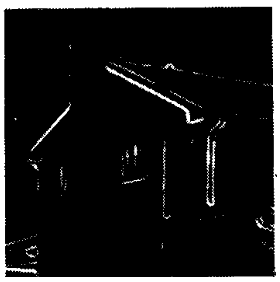

(g)

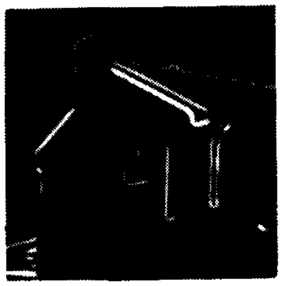

(h)


(i)

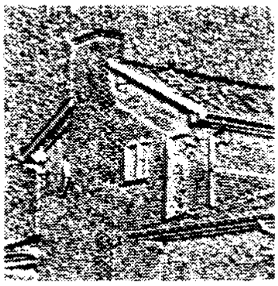

(j)

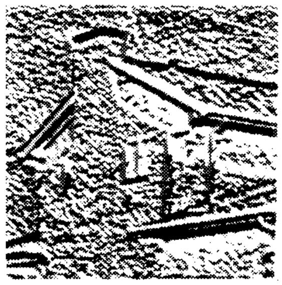

(k)


(l)


(m)


(n)


(o)


(p)


(q)


(r)

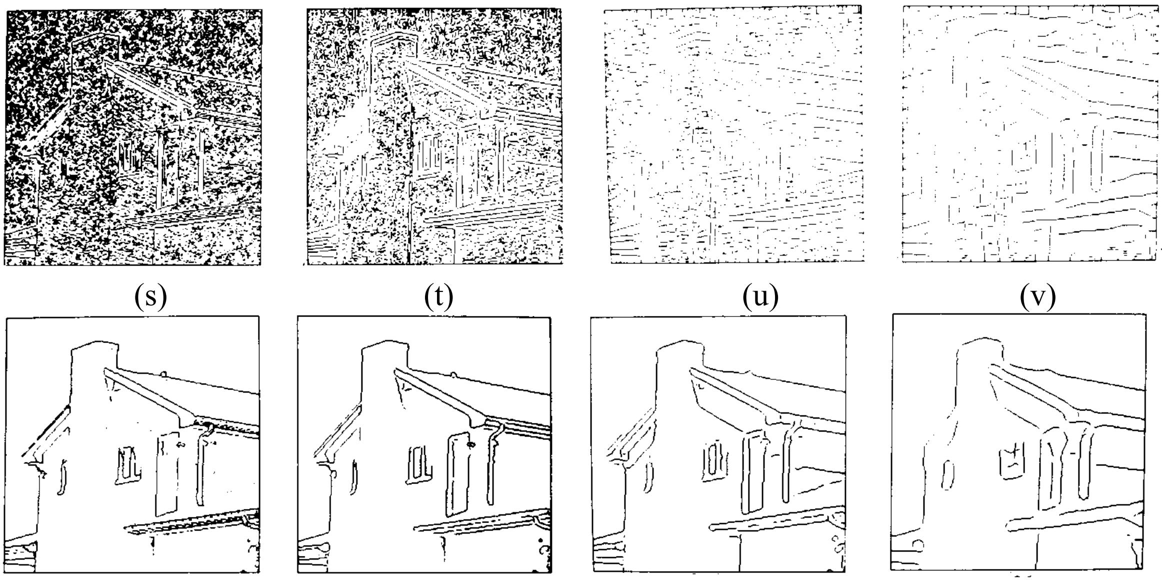


Fig. 5: (a) Original image. Low frequencies whose scales below the scale 2^4 are carried by the image (Scaling function) $S_{2^4} f(x,y)$ that is shown in (b). (c)~(f) give the images $W_s^1 f(x,y) \mid s = 2^j, 1 \le j \le 4$, and the scale increases from left to right. (g)~(j) give the images $W_s^2 f(x,y) \mid s = 2^j, 1 \le j \le 4$. Black, grey, and white pixels indicate, respectively, negative, zero, and positive sample values. (k)~(n) display the modulus image $M_s f(x,y) \mid s = 2^j, 1 \le j \le 4$. Black pixels indicate zero values and white ones correspond to the highest values. (o)~(r) gives the angle images $A_s f(x,y) \mid s = 2^j, 1 \le j \le 4$. Angle values range from 0(black) to 2π (white). (s)~(v) display in black the position of the points that are local maxima of $M_s f(x,y)$, in the direction given by the angle images $A_s f(x,y)$. (w)~(z) show the modulus maxima, where the modulus value $M_s f(x,y)$ is larger than a given threshold. (This figure is extracted from p.635 of [1]: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=119727&isnumber=3425)

Like the non-maximum suppression in a Canny edge detection, we detect the points where the modulus of $\nabla(f * \theta_s)(x,y)$ is locally maximum. At each scale s, the modulus maxima of the wavelet transform are thus defined as points (x,y) where the modulus image $M_s f(x,y)$ is locally maximum, along the gradient direction given by $A_s f(x,y)$. These modulus maxima are inflection points of $(f * \theta_s)(x,y)$. We record the position of each modulus maximum and the values of $M_s f(x,y)$ and $A_s f(x,y)$ at the corresponding location [1].

After obtaining the local modulus maxima and the angle information, we now can compute the edge image at scale s by thresholding the magnitude function as:

$$E_s(x,y) = \begin{cases} 0, & \|W_s f(x,y)\| < T \\ 1, & \|W_s f(x,y)\| \geq T \end{cases} \quad (17)$$

where T is the threshold value that separates the image into the edges and the background, and can be determined using Otsu's method [7].

From Fig. 5 (s)~(v), we can see at fine scales, i.e. the width of the mother wavelet is narrow, there are many modulus maxima created by the image noise and light textures. After thresholding in Fig. 5 (w)~(z), the modulus maxima created by the light noise and textures have disappeared and only the important edges remain.

After many discussions, we are able to see that the dyadic wavelet transform is very useful for the edge detections in 1-D or 2-D. Due to its dyadic scales and the recursive relationship between the wavelet and the scaling function, we can use the fast algorithm suggested by Mallat et al. to make it research large scales more 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. In the section 3, we combine the Lipschitz regularity with the multi-scale wavelet transform in order to 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.

3. A Multi-scale Wavelet Model with the Lipschitz Regularity for Edge Detection

3.1 Characterization of Edges with Lipschitz Reguarity

In an image, all edges are created with different magnitude and varying direction. Some are more significant, and others are blurred and insignificant, or even maybe the noise. **The edges of more significance are usually more essential 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. Besides, the local Lipschitz regularity can be efficiently measured by wavelet transforms.

I first introduce the definition of Lipschitz exponent in 1-D. [1]

<u>**Definition 1**</u>

- 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 $h < h_0$,

$$|f(x_0+h) - P_n(h)| \leq A|h|^\alpha \quad (18)$$

- 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 (18) is satisfied if $x_0 + h \in (a,b)$.

- We call the superior bound of all values α at x_0 which satisfy equation (18) as Lipschitz regularity.
- 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. [1]

Definition 2

- 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:

$$\left|f(x+h,y+k)-f(x,y)\right| \le A_\varepsilon (h^2+k^2)^{\frac{\alpha}{2}} \tag{19}$$

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 $\left|W_s^1 f(x,y)\right|$ and $\left|W_s^2 f(x,y)\right|$. The value of each of these components is bounded by equation (14).**

To measure the Lipschitz regularity of the edges, we can use the following theorems which are proved by the definition of $\psi_s^1(x,y), \psi_s^2(x,y)$ and equation (12). [1]

Theorem 1

Let $f(x,y)\in L^2(R^2)$ and $0\le\alpha\le1$. 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 $s>0$,

$$\left|M_s f(x,y)\right| \le A_\varepsilon s^\alpha \tag{20}$$

Theorem 2

Let n be a positive integer and $\alpha\le n$. Let $f(x,y)\in L^2(R^2)$. 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,

$$\left|M_s f(x,y)\right| \le A(s^\alpha + \left|x-x_0\right|^\alpha) \tag{21}$$

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.

Condition 1: *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,*

$$\left|M_s f(x,y)\right| \le A s^\varepsilon \tag{22}$$

Condition 2: *There exists a constant B such that for all points (x,y) in a neighborhood of (x_0,y_0)*

and any scale s,

$$|M_s f(x,y)| \le B\left(s^{\alpha} + \frac{\left((x-x_0)^2+(y-y_0)^2\right)^{\frac{\alpha}{2}}}{\frac{1}{2}\left|\log\left((x-x_0)^2+(y-y_0)^2\right)\right|} \right) \tag{23}$$

Theorem 1 and Theorem 2 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. [2]**

3.2 Scales of Edges and Effect of Noise on Edge Detection

The scale of the mother wavelet controls the significance of edges to be shown. Edges of higher significance are more likely to be kept by the wavelet transform across scales. On the other hand, edges of lower significance are more likely to disappear when the scale increases. **The significance of an edge can be mathematically measured by its Lipschitz exponent using Theorem 1 and Theorem 2.**

In most of the observed images, the noise is unavoidable. It is broadly defined as an additive (or possibly multiplicative) contamination of an image [2]. Due to its unpredictability, we often use the term "random noise" to describe an unknown contamination added to an image. Assume the image without noise is f, and the noise is n, then the contaminated image g can be represented as

$$g = f + n \tag{24}$$

We can measure the effect of the noise by the signal –to-noise ratio (SNR).[2]

$$SNR = \frac{\|f\|^2}{\|n\|^2} = \frac{\int\int_{\Omega} f^2(x,y)dxdy}{\int\int_{\Omega} n^2(x,y)dxdy} \tag{25}$$

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

$$H_f = \|f * L\|^2 \tag{26}$$

If the noise variance is n_0^2, the squared norm of the noise output is

$$H_n = \|n * L\|^2 \tag{27}$$

And the new signal-to-noise ratio is:

$$SNR = \frac{H_f}{H_n} \tag{28}$$

It is easy to prove that

$$\frac{H_f}{H_n} \geq \frac{\|f\|^2}{\|n\|^2} \tag{29}$$

From equation (29), 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 [1][2][18]. J. Canny [18] 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. [1] 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.**

3.3 Multi-scale Edge Detection with Lipschitz Regularity

For the traditional wavelet-based edge detection scheme in the section 2, we use the different scales in different time and get the edge maps with distinct scales, so **different locations in the same image use the same scale meanwhile**. Now I would like to introduce another dissimilar " multi-scale" method. Here, "multi-scale" means **different locations in the same image use the unlike scale at the same time.** To do this, we first use the coefficients of the wavelet transform across scales to measure the local Lipschitz regularity by Theorem 1 and 2. If the coefficients of the wavelet transform are likely to increase then the Lipschitz regularity is positive, if they are likely to decrease then the Lipschitz regularity is negative (e.g. white noise). Locations with lower Lipschitz regularities are more likely to be insignificat details or noise.

Take Fig. 4 for example, the step edge is more likely to be the real edge, and the Dirac as well as the fractal edges are more likely to be the interference for the image. Hence, the coefficients of the wavelet transform across increasing scales for step edges, but decrease for another two kind of edges. **After getting the Lipschitz regularity of the whole image, we can use a larger-scale wavelet at positions where Lipschitz regularity is slower to remove the effect of noise, while using a smaller-scale wavelet at positions where the Lipschitz regularity is larger to preserve the precise position of the edges [2].**

Fig. 6 shows the noiseless Lena image, we can see the result of new multi-scale edge detection with Lipschitz regularity is better than Canny's edge detection for it preserves more real edges. The performance of the new method is more obvious in the images corrupted by the white noise which are shown in Fig. 7~8. **It is apparent to see the improved method gives more continuous and precise edges.**

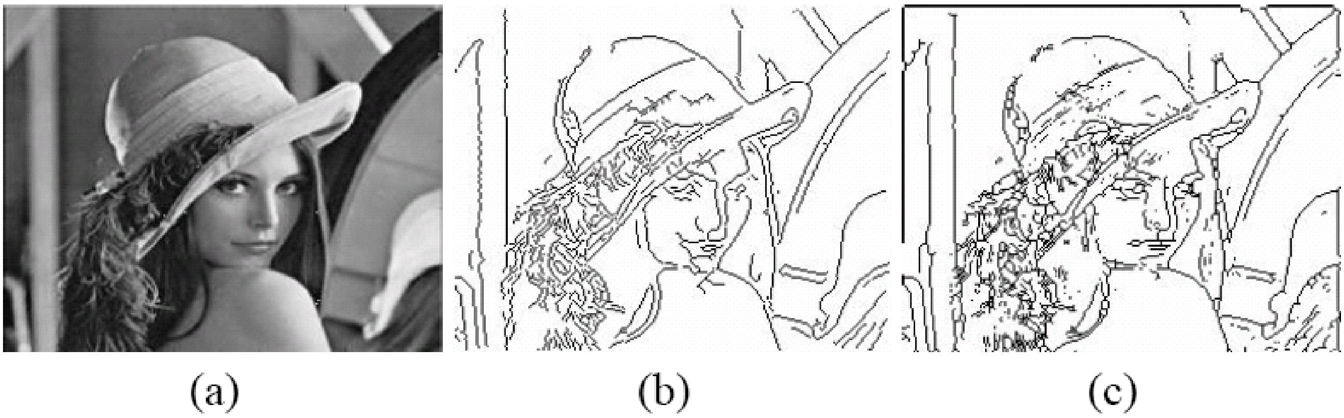


Fig. 6: Edge detection for Lena image. (a) The Lena image. (b) Edges by the Canny edge detector. (c) Edges by the multi-scale edge detection using wavelet transform and Lipschitz regularity.

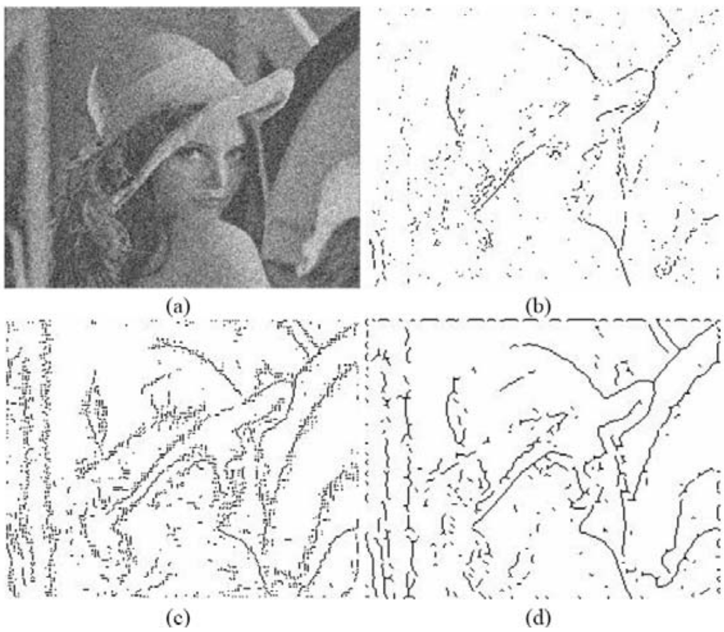


Fig. 7 : Edge detection for a Lena image with white noise: (a) Lena image (SNR=30dB). (b) Edges by the Sobel edge detector. (c) Edges by Canny edge detection with adjusted variance. (d) Edges by multi-level edge detection using wavelet transform and Lipschitz regularity.

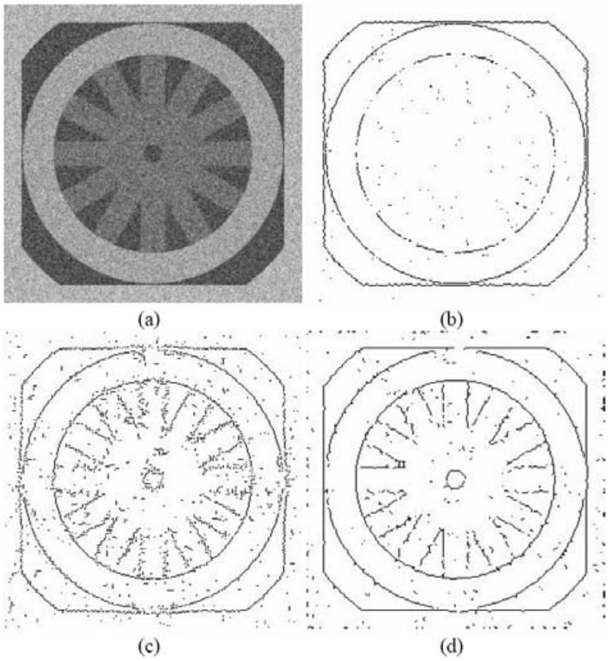


Fig. 8 : Edge detection for a wheel image with white noise. (a) Wheel image (SNR=10dB). (b) Edges by the Sobel edge detector. (c) Edges by Canny edge detection with adjusted variance. (d) Edges by multi-level edge detection using wavelet transform and Lipschitz regularity.
(Fig.6~8 are extracted from p.58~59, p.62 of [2] :
http://eref.uqu.edu.sa/files/Others/Image%20Processing/A%20wavelet%20approach%20to%20edge%20detection%20-%20Thesis.pdf))

4. A novel Thresholding Method for Wavelet Multi-scale Edge Detection

From the section 2 and 3, we have known that the advantage of the multi-scale measurement of the dyadic wavelet transform makes it very useful for edge detection. After considering the Lipschitz regularity of the edges and using varied scales in different places of the image at the same time, we can accurately preserve more real edges and filter out more noise in the noisy edges. Another essential issue in the edge detection is how to thresholding. The examples for the wavelet transform mentioned above only use one threshold for the whole transformed images (equation (17)). The only one threshold results in the local maxima of magnitude from weak edges will be filtered out together with magnitude extrema from intensity irregularities and noises [3]. So, an adaptive threshold method is proposed by Sun Wenchang et al., in the hope of being effective on edge detection of images with both strong and weak edges. Furthermore, they propose a procedure to synthesize multi-scale edges of the image in order to obtain only one precise edge map.

4.1 Adaptive Thresholding Method

First, we calculate the modulus of the wavelet transform at the scale s, $M_s f(x,y), s=2^j, j \in Z, j \in (-\infty,\infty)$ and angle images at the scale s, $A_s f(x,y)$ through the dyadic wavelet transform of original image by equation (14) and (15). Second, using the similar scheme like the non-maximum suppression to get the probable edge image $P_s(x,y)$. Then, scan this probable edge image with a $n \times n$ window, and the threshold is calculated from the wavelet transform coefficients within the window, [3]

$$T = T_0 + \alpha_0 \times \sum_{i,j} C_{i,j} \tag{30}$$

where T is the threshold, T_0 is initial value of T, $C_{i,j}$ is the corresponding wavelet transform coefficient in the window, α_0 is a scale coefficient to determine how much T depends on $C_{i,j}$. The size of the window needed to be chosen very carefully but there is no standard to tell us how to choose is better. If the window is too small, noises and intensity irregularities will affect the threshold more, and false alarm rate increases. While the window is too large, weak edges may be filtered. According to the experiences, the window size of 24×24 is proper.

4.2 An Iterative Way to Synthesize the Multi-scale Edges

I have mentioned in 3.2 that the scale of the wavelet deeply affects the edges to be shown. If the scale is small, edge details are abundant, and the precision of edge location is relatively high, but noises cause disturbances easily. And if the scale is large, the edges are more stable, with good noise resistance but worse precision of edge location [1]. To solve this problem, we are able to use the following iterative procedures to get the better edge map : [3]

Step 1: Denote $C_{2^j}(x,y)$ as the edge image of scale 2^j. Starting from a proper large scale to reduce the most of the noise, and then decrease the value of *j* by 1. When the edge shift between edge images of adjacent scales is within 1 pixel, for each existed edge pixel in $C_{2^j}(x,y)$, we search edge pixels in $C_{2^{j-1}}(x,y)$ within a 3×3 window centered at the pixel corresponding to the one in $C_{2^j}(x,y)$. If we can find the corresponding pixel in the $C_{2^{j-1}}(x,y)$, label it to be one of the candidate edge pixels, otherwise set it to be zero.

Step 2: After finding all candidate edge pixels in step 1. Connect the pixels with similar gradient magnitudes and arguments in $C_{2^{j-1}}(x, y)$ to form a chain. Delete isolated chains if the length of the chains is too short. Make the rest chains form an edge image with single pixel width, and denoted them by $E_{2^{j-1}}(x, y)$.

Step 3: Set $j = j - 1$. If $j > 1$, go to step (1); if $j = 1$, then edge image $E_{2^0}(x, y)$ is the final synthesis of multi-scale edges.

In Fig. 9, we can see the wavelet transform with the adaptive thresholding and new synthesis way has better performance than other edge detectors no matter in the noiseless or noisy images. The Prewitt detector (Fig. 9 (b)) has too thick edges and Canny operator (Fig. 9 (c) and (d)) lose some real edges. In noisy condition (Fig. 9 (f)), the Prewitt and Canny detectors (Fig. 9(g) and (h)) contain too much noise edges. Therefore, we can conclude that the edges obtained by the proposed method are adequate and precisely located. Besides, it has better noise resistance.

(a) Original Lena

(b) Prewitt detector

(c) Canny (0.00004, 0.1)

(d) Canny (0.4, 0.4)

(e) Proposed method

(f) Lena with noise

(g) Prewitt detector

(h) Canny detector

(i) Proposed method

Fig. 9 : (a) is the original image of Lena. (b)~(e) are result images of the edge detection for (a), for the bracket(,) of the Canny edge detector, the left number means the low threshold and the right number means the higher threshold. (f) is Lena with pepper-and-salt noises (noise density is 0.05). (g)~(i) are results for (f).

(This figure is extracted from [3] : http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5301959&isnumber=5300799)

5. Combining Wavelet Transform and Canny Operator for Edge Detection

In the section 1, we have told about that the discrete wavelet transform (DWT) decomposes the original image into LL, LH, HL, and HH four parts iteratively (Fig. 1). In the end, the LH and HL parts contain the horizontal and vertical details and may be some noises. For edge detection, the HH areas are always affected by the noise seriously. Therefore, we usually preserve only the LH and HL parts and discard the HH section and ignore the LL part. However, the four areas obtaind from 2-D DWT in fact all contain the useful information of the original image, so these areas should be used completely when detecting the image edges. Hence, Jianjia Pan combine the one octave of the discrete wavelet transform, Canny operator , the dyadic wavelet transform (the continuous WT with discrete coefficients) to do the edge detection.

5.1 Using Canny Operator on the Edge Detection of LL part

The LL area is the smoothed version of the original image which contains the most information of the original image. Because the Canny operator is a edge detecting operator based on an optimal algorithm, which has the most stringent criterions of edge detection which are mentioned in 1.2 and it outperforms on processing the image contaminated by additive Gaussian noise. Therefore, the Canny operator is adopted to detect the edges of the low frequency sub-image (LL part). The realization process has been discussed in 1.2. In the step of smoothing image, the 5×5 Gaussian function is used. For the hysteresis thresholding, Jianjia Pan let the high threshold to be about 1.5 to 2 times of the low threshold.

Through adopting Canny operator to detect the low frequency sub-image, we can obtain a clear edge image which will miss some real edges and exist some sham edges [4]. Thus the edges of the high frequency sub-images (LH, HL, HH parts) should be fused.

5.2 Using The Wavelet Transform on the Edge detection of the LH, HL, and HH Parts

After an octave, a single iteration of the DWT, we get the high frequency parts: LH, HL, HH in addition to the low frequency part : LL. Due to the high frequency property of the noise, the noises need be reduced before extracting the edges from the high frequency sub-images.

5.2.1 The Denoising Algorithm

We first get the wavelet coefficients (the modulus of the wavelet transform) and angle information by equation (14) and (15). Then, the wavelet coefficients are multiplied by a denoising factor which is relative to their own coefficients' value. This denoising factor is less than 1 and will decrease if the absolute value of the wavelet coefficients increases. This algorithm considers energy distribution property of the wavelet decomposition of the image globally, and obtains a better denoising result, which support the foundation for extracting the edge of the high-frequency sub-images [4]. The process function is as follows.

$$F(x,y) = \begin{cases} w(x,y), & |w(x,y)| \ge |3\sigma| \\ 0, & |w(x,y)| \le |aver| \\ w(x,y) \times k, & else \end{cases} \tag{31}$$

where $w(x,y)$ denotes the high frequency coefficient, $F(x,y)$ denotes the high frequency coefficient gained after denoising, σ 、 $aver$ indicate the variance and the mean of the high frequency coefficients in different wavelet decomposition levels.

On the basis of the statistic property of the wavelet coefficients, k is a function which is relative to the coefficient:

$$k = e^{-aw(x,y)+b} - 1 \tag{32}$$

When $w(x,y)$ is greater than 3σ, it means $w(x,y)$ consisting most of the information of the original image. So, $k = 1$ and

$$e^{-3\sigma a+b} - 1 = 1 \tag{33}$$

When $w(x,y)$ is smaller than $aver$, $w(x,y)$ will approaches 0 and

$$e^{-a \times aver+b} - 1 = 0 \tag{34}$$

From equation (33) and (34), we can solve the values of a and b,

$$a = \frac{-\ln 2}{3\sigma - aver} \quad , \quad b = a \times aver \tag{35}$$

This algorithm is aiming for the wavelet coefficients multiplied by different denoising factors which can reduce the image noise and keep useful details.

5.2.2 Edge Detection of LH, HL, and HH Parts Based on Wavelet Transform

After eliminating noises, the edges of the high-frequency sub-images can be detected using **wavelet modulus maxima algorithm** which has been mentioned in 2.2.2. Because this algorithm can describe multi-scale edges of the target in the image which has translation, scale and rotation-invariant performance, it is effective to detect the edges [4].

After using the wavelet modulus maxima algorithm, we are able to get the gradient module images in scale s, $M_s f(x,y)$ of the high frequency sub-images LH、HL、HH. And then, we find the local modulus maxima of three sub-images using non-maximum suppression to get the edge images G_{LH}, G_{HL}, G_{HH}. Finally, we combine these three edge images by the fusion rule which is described as follows: [4]

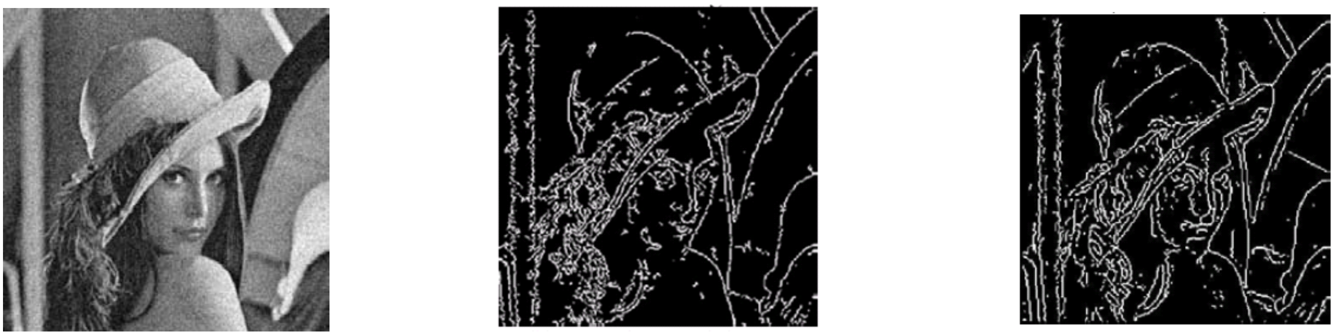
$$D_H(i,j) = r_{LH} \times D_{LH}(i,j) + r_{HL} \times D_{HL}(i,j) + r_{HH} \times D_{HH}(i,j) \tag{36}$$

where $D_{LH}(i,j)$、$D_{LH}(i,j)$、$D_{LH}(i,j)$ indicate the wavelet coefficients corresponding to the edge images G_{LH}, G_{HL}, G_{HH}. $D_H(i,j)$ denotes the wavelet coefficient after fusion. r_{LH}、r_{HL} and r_{HH} mean the corresponding weights, the sum of the three weights is 1.

From all above processes, we can get the edges of LL, LH, HL, and HH parts. The final edge image is obtained through wavelet composition from the fusion edge sub-images [19].

Fig. 10 shows the Lena results of the traditional wavelet transform and the proposed algorithm. From the outcomes, we are capable of concluding that the proposed algorithm which combined wavelet transform and the Canny operator not only get rid of the noise effectively but also enhance the image edge's details and locate the edge accurately.

(a) The Original image　　(b) Detecting result by the traditional wavelet transform　　(c) The proposed algorithm



(d) The original noisy image (SNR = 15.8 dB)　(e) Detecting result by the traditional wavelet transform　(e) The proposed algorithm

Fig. 10 (This Figure is extracted from [4] : http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5207422&isnumber=5207404)

6. Edge Detection by Scale Multiplication in Wavelet Domain

Many multi-scale techniques mentioned above are all first form the edge maps at several scales and then synthesize them together, a novel wavelet-based edge detection scheme by scale multiplication is proposed. The dyadic wavelet transforms at two adjacent scales are multiplied as a product function to emphasize the edge structures and suppress the noise. They determined the edges as the local maxima directly in the scale product before an efficient thrsholding. It is shown that the scale multiplication achieves better results than either of the two scales, especially on the localization performance. The dislocation of neighboring edges is also improved when the width of detection filter is set large to smooth noise [5]. In the above subsections, the dyadic wavelet transform is abbreviated to DWT for simplicity.

6.1 The Discrimination of Singularity and Noise by DWT

As we have mentioned in 3.1, the behavior of a signal across scales in wavelet domain depends on the local regularity that can be measured by Lipschitz exponents mathematically.

The Lipschitz regularity of a step edge is 0 and the definition can be extended to negative values for singularities worst than discontinuities, such as white noise. White noise is almost singular everywhere and has a uniform Lipschitz regularity equaling to -1/2 [5]. It can be seen from equation (22), for singularities of signals whose Lipschitz regularities are $\alpha \geq 0$, the DWT amplitudes would increase or keep invariant when increasing the scale 2^j . On the contrary, the Lipschitz regularity of white noise is less than 0, so the transform amplitudes will decrease rapidly along the scales. **In Fig. 11, we can observe that the DWT amplitudes of the step edge are large across scales, but those of noise decay rapidly.**

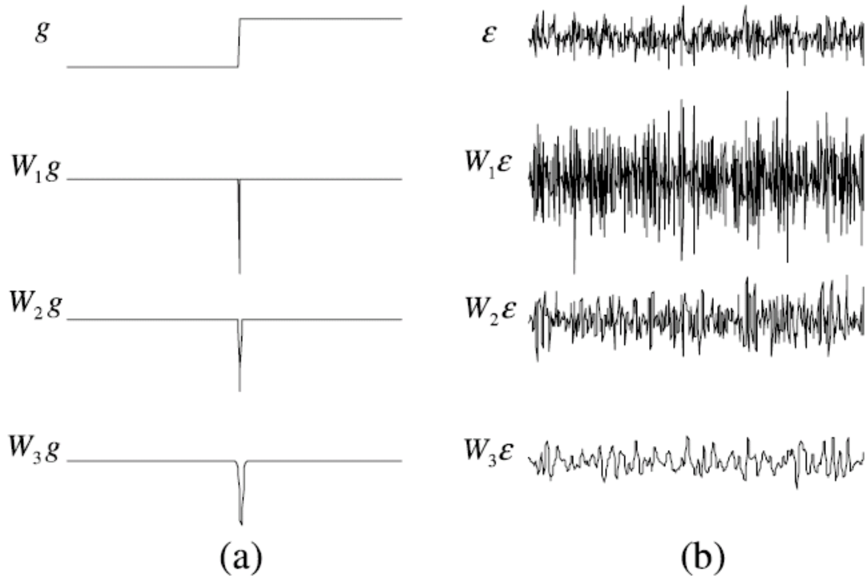


Fig. 11: (a) The DWT of sep edge g at the first three scales ($j = 1 \sim 3$). (b) The DWT of the noise ε at the first three scales($j = 1 \sim 3$). (This figure is extracted from [5] : http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V15-45XT9HX-5&_user=7761193&_rdoc=1&_fmt=&_orig=search&_sort=d&_docanchor=&view=c&_searchStrId=1174726732&_rerunOrigin=google&_acct=C000051951&_version=1&_urlVersion=0&_userid=7761193&md5=06ba66b85df7405a30dd2110c140c753)

6.2 The Scale Multiplication

The scale product function of $f(x)$ is defined as the correlation of two adjacent DWT scales,

$$P_j^f(x) = W_j f(x) W_{j+1} f(x) \tag{37}$$

Subscript j means that the product is computed at scales 2^j and 2^{j+1}. As Fig. 11 has shown, the peaks of real edges tend to propagate across scales. The production function will enhance the edge structures. But when $f(x)$ is a Gaussian white noise, it could be proved that the average number of local maxima at scale 2^{j+1} is half of that at scales 2^j [1]. Hence, if we multiply the DWT at adjacent scales then the noise will be diluted.

Why we not multiply three or more scales but only consider two adjacent scales? That's because the width of the mother wavelet $\psi_j(x)$ increases rapidly along dyadic sequence ($2^j, j \in Z$) and $W_j f(x)$ will become smoothier rapidly too. Therefore, as three or more adjacent scales were incorporated in the multiplication, edges would not be sharpened more and much edge dislocation would occur. So it is appropriate to analyze the multiplication using adjacent two scales.

In Fig. 12(a), a block signal g and its noisy version $f = g + \varepsilon$ are illustrated, where ε is the Gaussian white noise. It is shown that the noise is almost dominated in the wavelet coefficients $W_1 f$ at the finest scale. When the scale increases, the noise becomes smaller very quickly. It can also be seen that the positions of the step edges are better localized at the small scale. But some noise may be falsely considered as edges. At the large scales, on the other hand, the SNR is improved and edges can be detected more correctly but the accuracy of the edge location is decreasing. Finally, it is obviously to see in Fig. 12 (d) that the location of the step edges are more clear in P_j^f than in $W_j f$.

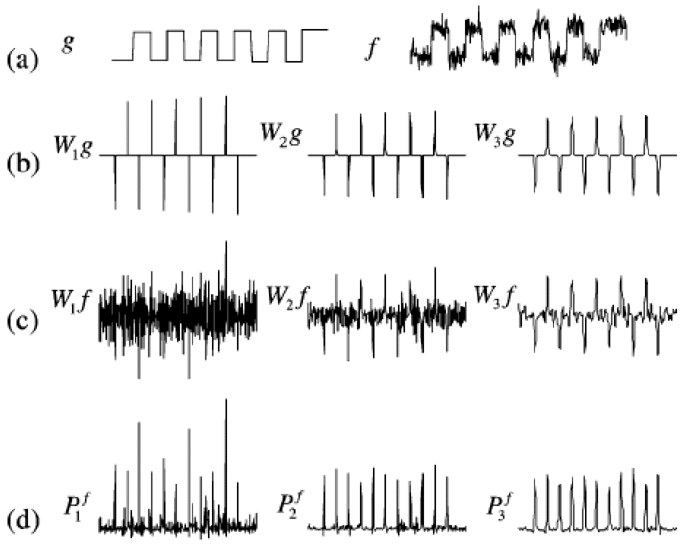


Fig. 12 : (a) The block signal g and its noisy version f. (b) The DWT of g at the first three scales ($j = 1 \sim 3$). (c) The DWT of the noisy signal f at the first three scales ($j = 1 \sim 3$). (d) The product function $P_j^f, j = 1 \sim 3$. (This figure is extracted from [5] : http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V15-45XT9HX-5&_user=7761193&_rdoc=1&_fmt=&_orig=search&_sort=d&_docanchor=&view=c&_searchStrId=1174726732&_rerunOrigin=google&_acct=C000051951&_version=1&_urlVersion=0&_userid=7761193&md5=06ba66b85df7405a30dd2110c140c753)

6.3 The Thresholding

In the classical edge detectors, we generally use only a single threshold t to obtain the edge map. However, if t is too small, there will be some false edges. On the contrary, if t is too large, portions of the contour may be missed. In the Canny operator, two thresholds are employed. After non-maximum suppression, a low threshold t_L and a high threshold $t_H \approx 2t_L$ are applied to obtain double thresholded edge images, I_L and I_H. Canny selects edges in I_L that link to the edges in I_H. In the wavelet multi-scale edge detection, we often adopt only one threshold for the whole transformed image. So the weak edges are always missed. To solve this problem, an adaptive threshold which is calculated by the wavelet coefficients is used. The more details have been discussed in the section 4.

Notice that edges and noise can be better distinguished in the scale product than in a single scale, so a properly chosen threshold could suppress the noise maxima effectively. Therefore, the single threshold is preferred for the simplicity.

We assert the edges as the local maxima in P_j^f. A significant edge at abscissa x_0 will occur on both the adjacent scales and the signs of $W_j f(x_0)$ and $W_{j+1} f(x_0)$ will be the same, so that $P_j^f(x_0)$ should be non-negative. If $P_j^f(x_0)$ is less than zero, then the point will be considered as noise and filtered out. The threshold in the scale multiplication scheme is set to be [5]

$$t_{sc}(j) = c \cdot \|\psi_j\| \cdot \|\psi_{j+1}\| \sigma^2 \sigma_{j,+}^2 \tag{38}$$

where ψ_j is the Mallat wavelet [1] at scale 2^j which is the first derivative of the Gaussian function and

$$\|\psi_j\| = \sqrt{\int \psi_j^2(x)dx} \tag{39}$$

σ is the standard deviation of the input signal which is assumed to be the Gaussian white noise $\varepsilon \sim N(0,\sigma^2)$ and

$$\sigma_{j,+} = \frac{1}{2}\sqrt{\int \left(\frac{\psi_j(x)}{\|\psi_j\|} + \frac{\psi_{j+1}(x)}{\|\psi_{j+1}\|} \right)^2 dx} \tag{40}$$

The constant $c \approx 20$ yields impressive results. To see more details of the induction of this threshold, please refer to [5].

In two dimensions, two product functions should be defined in x and y directions.

$$P_j^{f,1}(x,y) = W_j^1 f(x,y) W_{j+1}^1 f(x,y) \text{ and } P_j^{f,2}(x,y) = W_j^2 f(x,y) W_{j+1}^2 f(x,y) \tag{41}$$

For an edge point (x_0, y_0), $W_j^i f(x_0, y_0)$ and $W_{j+1}^i f(x_0, y_0)$, $i = 1, 2$ should have the same sign. So both $P_j^1 f(x_0, y_0)$ and $P_j^2 f(x_0, y_0)$ will be nonnegative and the orientation in formation of the gradient is lost, which should be recovered from $W_j^1 f(x_0, y_0)$ and $W_j^2 f(x_0, y_0)$ [5].

Setting the points with $P_j^{f,1}(x,y) < 0 \;\; or \;\; P_j^{f,2}(x,y) < 0$ to 0, the modulus and angle of point (x,y) are defined as

$$M_j f(x,y) = \sqrt{P_j^{f,1}(x,y) + P_j^{f,2}(x,y)} \tag{42}$$

$$A_j f(x,y) = \tan^{-1}\left(\frac{\mathrm{sgn}\left(W_j^2 f(x,y)\right)\cdot\sqrt{P_j^{f,2}(x,y)}}{\mathrm{sgn}\left(W_j^1 f(x,y)\right)\cdot\sqrt{P_j^{f,1}(x,y)}} \right) \tag{43}$$

As in the Canny edge detection algorithm, an edge point is asserted to be where $M_j f(x,y)$ has a local maximum in the direction of the gradient given by $A_j f(x,y)$. The modulus map $M_j f(x,y)$ should be thresholded to remove noise. Similar to 6.1.3, a proper threshold $t_{sc}^i(j)$ that could be applied to $P_j^{f,i}(x,y)$, $i = 1$, 2, is

$$t_{sc}^i(j) = c\cdot\left\|\psi_j^i\right\|\cdot\left\|\psi_{j+1}^i\right\|\sigma^2(\sigma_{j,+}^i)^2 \tag{44}$$

Where c is a constant and

$$\left\|\psi_j^i\right\| = \sqrt{\iint(\psi_j^i(x,y))^2\,dxdy} \tag{45}$$

$$\sigma_{j,+}^i = \frac{1}{2}\sqrt{\iint\left(\psi_j^i(x)\Big/\left\|\psi_j^i\right\| + \psi_{j+1}^i(x)\Big/\left\|\psi_{j+1}^i\right\|\right)^2 dxdy} \tag{46}$$

c can be chosen around 20 and by the experimental experience,

$$t_{sc}(j) = 0.8\times\sqrt{t_{sc}^1(j) + t_{sc}^2(j)} \tag{47}$$

could achieve satisfying results.

In Fig. 13~17, the Canny edge detection and LOG algorithms are employed for comparison with the scale multiplication scheme. There are two parameters in the Canny operator. One is the high threshold t_H $(\approx 2t_L)$. The other is σ_g, the standard deviation of the Gaussian function that is used to adjust the width of the detection filter. LOG method also has two adjustable parameters: the standard deviation σ_g of the Gaussian function and the threshold t to suppress false edges.

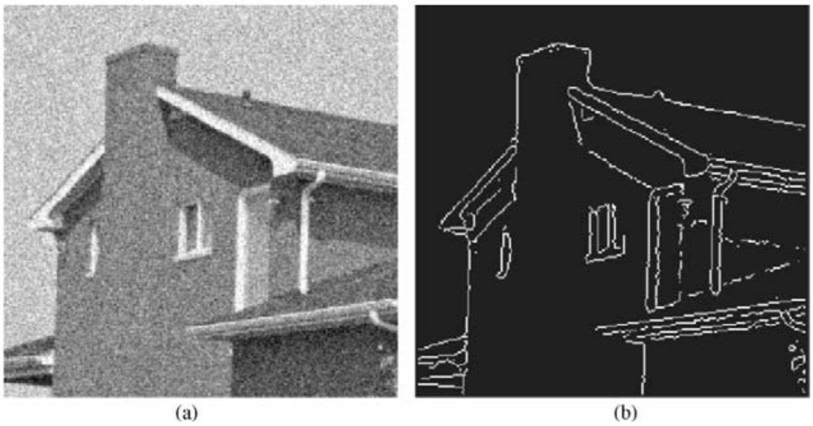


Fig. 13 : (a) Noisy house (SNR 16.52 dB). (b) Edge map by scale multiplication scheme (at scales 2^2 *and* 2^3).

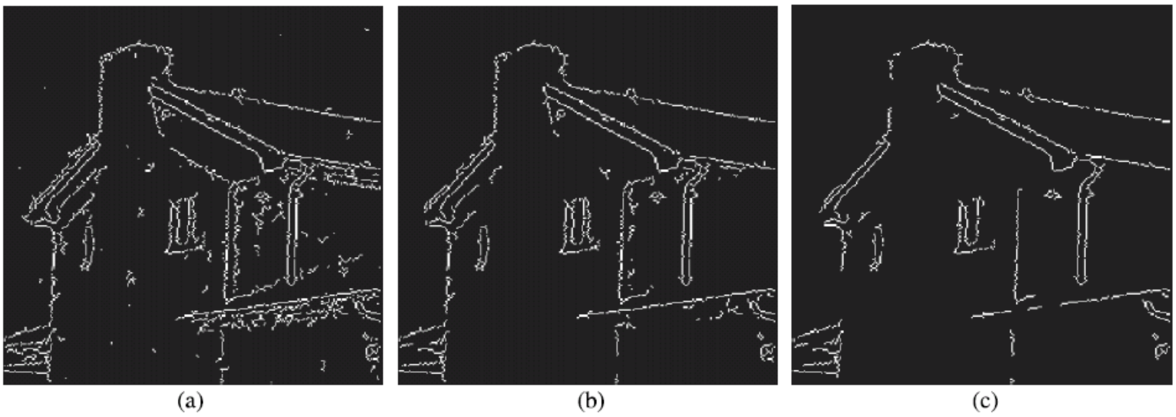


Fig. 14 : Edge maps by Canny with $\sigma_g = 1$: (a) $t_H = 0.25$. (b) $t_H = 0.32$. (c) $t_H = 0.39$.

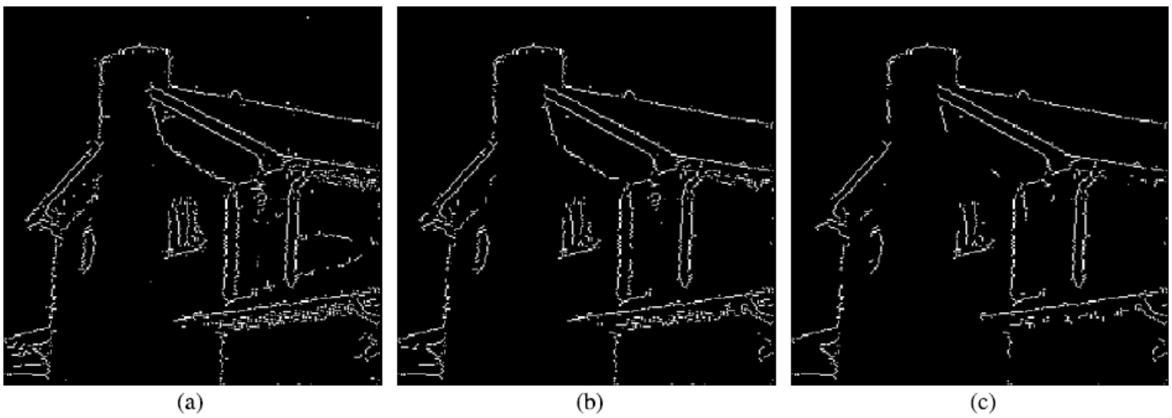


Fig. 15 : Edge maps by Canny with $\sigma_g = 2$: (a) $t_H = 0.21$. (b) $t_H = 0.26$. (c) $t_H = 0.31$.

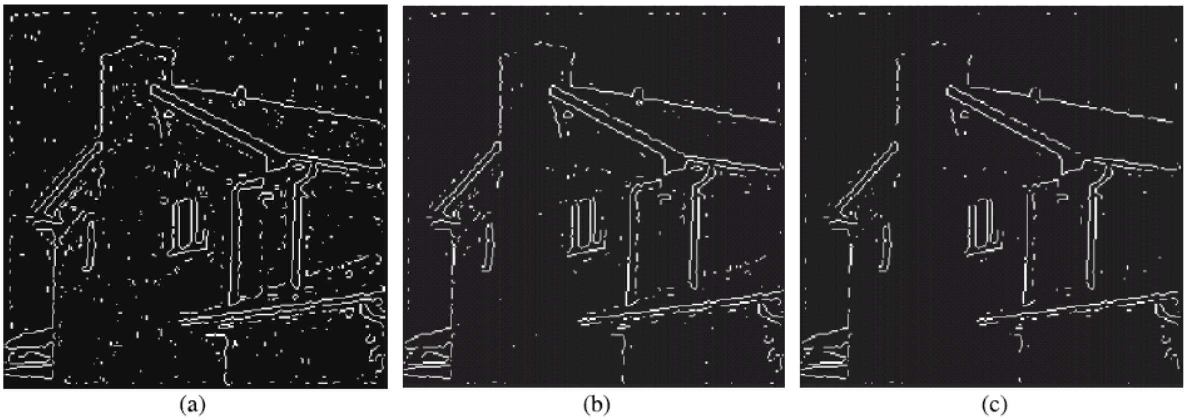


Fig. 16 : Edge maps by LOG with $\sigma_g = 2$: (a) $t = 1.9$. (b) $t = 2.4$. (c) $t = 2.9$.

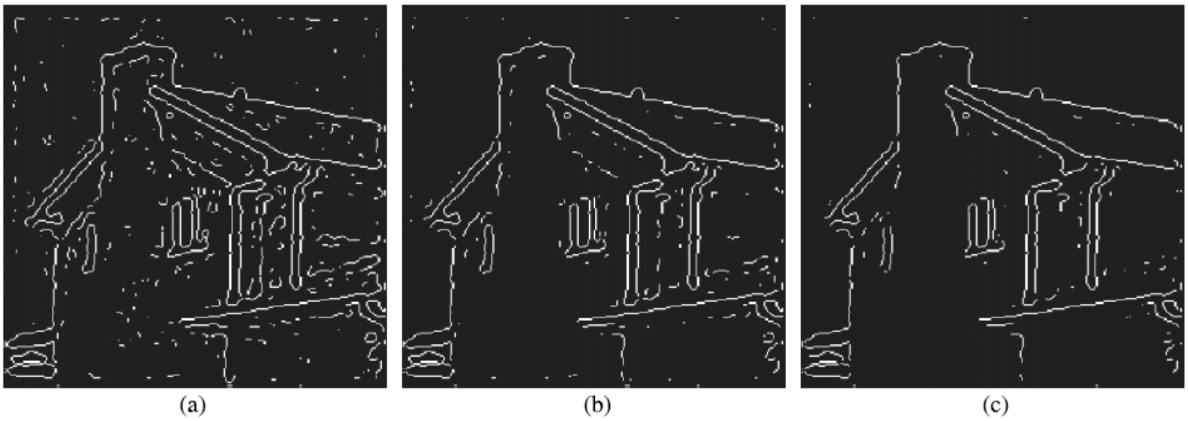


Fig. 17 : Edge maps by LOG with $\sigma_g = 2.8$: (a) $t = 1.5$. (b) $t = 0.7$. (c) $t = 0.9$.

From Fig. 14~17, we can see in Canny edge detection and LOG algorithms, the edge maps are sensitive to noise at the small scale. When the threshold is increased to suppress noise, some of the true edges disappear. At large scale, the location accuracy is decreased. It can be seen from Fig. 13 (b) that with the proposed scheme, much better results are obtained. False edges are eliminated and the edges on the house are clearly visible. Many relatively faint edges undetected by LOG and Canny edge detection are enhanced by the scale multiplication and found in the final edge map.

7. Wavelet-Based Edge Detection Using Gabor and Cauchy Directional Wavelets

The most applications of the edge detection require the following criteria: [1][6][7]

(1) **Multiresolution** : The edge detection method should allow the resolution of the edges to be variable. The classical edge detector cannot do it, but the wavelet transform can..

(2) **Localization** : The basis elements used in the edge detection should be well concentrated in both the spatial and the frequency domains. The Fourier transform is lack of this property.

(3) **Directionality** : The edge detection technique should contain the basis functions oriented at the variety of directions, not just horizontal and vertical directions.

(4) **Anisotropy** : To capture the smooth contours in the images, the edge detection technique should contain the basis functions with variety of shapes, in particular with different aspect ratios.

The first two criteria are successfully provided by the dyadic wavelet transform which is separable and isotropic, but it is lack of the properties (3) and (4). So, if we encounter with an edge which is not in the horizontal or vertical direction and it is very smooth, then the traditional wavelet-based scheme will have poor performance. Therefore, non-separable directional wavelet transform such as the Gabor wavelet transform and the Cauchy wavelet transform are needed to solve this problem.

7.1 Gabor Wavelet Transform

A directional mother wavelet is a function $\psi(x,y) \in L^2(R^2)$ where its Fourier transform $\hat{\psi}(\omega_x, \omega_y)$ has support in a convex cone in the spatial frequency space, with apex at the origin. An example of this type is the Gabor or Morlet wavelet. A Gabor mother wavelet is a Gaussian modulated by a sinusoid. It is a non-orthogonal wavelet with the modulated sinusoidal frequency of ω_0 and the standard deviations σ_x and σ_y which is expressed as : [7]

$$\psi(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + j\omega_0 x \right] \tag{48}$$

Through this representation, we can write the directional wavelet transform of 2-D signal (image) $f(x,y)$ at scale s and orientation θ as :

$$W_{s,\theta} f(x,y) = (f * \psi_{s,\theta})(x,y) \tag{49}$$

where the Gabor wavelet $\psi_{s,\theta}(x,y)$, obtained by dilation and rotation of the mother wavelet function $\psi(x,y)$. Thus

$$\psi_{s,\theta}(x,y) = \frac{1}{s}\psi\left(\frac{x^{'}}{s}, \frac{y^{'}}{s}\right) \tag{50}$$

$$x^{'} = x\cos\theta + y\sin\theta \tag{51}$$

$$y^{'} = -y\sin\theta + y\cos\theta \tag{52}$$

For the edge detection, we compute the wavelet transform of the image at scale $s = a^m$, $for\ m = 0,1,...,M-1$ and orientations $\theta_n = \frac{n\pi}{N}$, $for\ n = 0,1,...,N-1$. M and N denote the total number of scales and the orientations. Therefore the Gabor wavelet functions $\psi_{m,n}(x,y)$ are designed so to ensure that the half-peak magnitude supports of the filter responses in the frequency spectrum touch one another [7], as shown in Fig. 18. By doing this, it can be ensured that the filters will capture the maximum information with minimum redundancy.

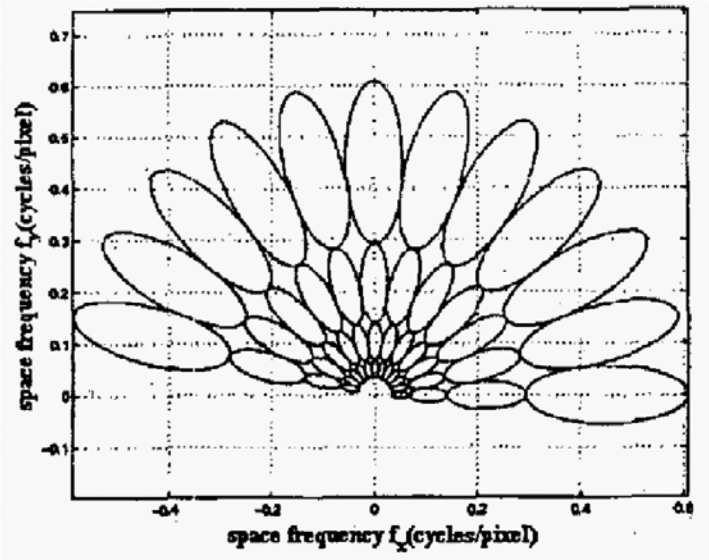


Fig. 18 : Magnitude support of Gabor wavelets in the frequency domain with the highest frequency $\omega_H = 0.9\pi$, $a = 2$, $N = 12$, $M = 4$. (This figure is extracted from [7] : http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1461748&isnumber=31419)

After getting $W_{s,\theta} f(x,y)$, we choose the maximum value of the transform through non-maximum suppresion:

$$W_m f(x,y) = Max_n (W_{m,n} f(x,y)) \tag{53}$$

If this maximum value is greater than the threshold then we determine it as an edge point, else it is ignored.

7.2 Cauchy Wavelet Transform

The 1-D Cauchy wavelet of order k, in space and frequency domains, is defined as [7] :

$$\psi_k(x) = \left(\frac{j}{j+x} \right)^k \tag{54}$$

In 1-D, the positive half-line is a convex cone. Thus a natural generalization to 2-D will be a wavelet whose support in spatial-frequency space is contained in a convex cone with apex at the origin; therefore the defined Cauchy wavelet is a directional wavelet. Let $C(0,\delta)$ be the convex cone generated at the direction of x with angle width equal to $\delta < \pi$. We define the 2-D Cauchy wavelet with spatial-frequency support in this cone and zero elsewhere as : [7]

$$\psi_k(x,y) = (-1)^{k+1}\left(\frac{(\sin\delta)^{k+1}}{\cos^2\left(\frac{\delta}{2}\right)(x+j) - \sin^2\left(\frac{\delta}{2}\right)y}\right)^{k+1} \tag{55}$$

Fig. 19 shows the 3-D view of magnitude of the Cauchy wavelet with $k = 4$.

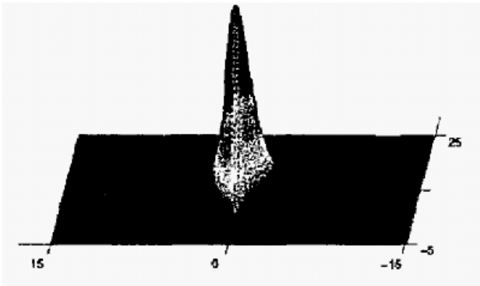


Fig. 19 : A 3-D view of magnitude of the Cauchy wavelet with $k = 4$.(This figure is extracted from [7] : http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1461748&isnumber=31419)

Through this representation, we can write the directional wavelet transform of 2-D signal (image) $f(x,y)$ at scale s and orientation θ as :

$$W_{s,\theta} f(x,y) = (f * \psi_{s,\theta})(x,y) \tag{56}$$

where the Cauchy wavelet $\psi_{s,\theta}(x,y)$, obtained by dilation and rotation of the mother wavelet function $\psi(x,y)$. Thus

$$\psi_{s,\theta}(x,y) = \frac{1}{s}\psi\left(\frac{x'}{s},\frac{y'}{s}\right) \tag{57}$$

$$x' = x\cos\theta + y\sin\theta \tag{58}$$

$$y' = -y\sin\theta + y\cos\theta \tag{59}$$

For the edge detection, we compute the wavelet transform of the image at scale $s = a^m$, $for\ m = 0,1,...,M-1$ and orientations $\theta_n = \frac{n\pi}{N}$, $for\ n = 0,1,...,N-1$. M and N denote the total number of scales and the orientations. Therefore the Cauchy wavelets $\psi_{m,n}(x,y)$, like the Gabor wavelet functions, are designed so to ensure that the magnitude supports of the filter responses in the frequency spectrum touch one another. By doing this, it can be ensured that the filters will capture the maximum information with minimum redundancy.

After getting $W_{s,\theta} f(x,y)$, we choose the maximum value of the transform:

$$W_m f(x,y) = Max_n(W_{m,n} f(x,y)) \tag{60}$$

If this maximum value is greater than the threshold then it is an edge point, else it is ignored.

It is obvious that, both Gabor and Cauchy wavelet require more computations than the dyadic wavelet transform explained in the section 2 that requires only two wavelet transform computations at the x and y directions.

7.3 2-step Directional Wavelet-Based Edge Detection

To combine the superior performance of non-separable directional wavelet-based techniques and less computations of the separable isotropic wavelet-based techniques, a two-step wavelet-based edge detection using isotropic and the directional wavelets is proposed. [7]

Step 1: The ordinary isotropic 2-D wavelet transform is applied by computing the gradient of the image which has been smoothed by the Gaussian function at the different scales.

Step 2: Double threshoIding the wavelet transform coefficients using two threshold levels T_L and T_U derived from the extended Otsu's method. If the magnitude of the wavelet transform coefficient of a pixel is less than then T_L, then the pixel is ignored but if the magnitude of the wavelet transform coefficient is greater than or equal to the upper threshold level T_U, then the pixel is marked as edge points.

Step 3: If the pixels with the magnitude of the wavelet transform coefficients are greater than or equal to the lower threshold level T_L and less than the upper threshold level T_U, we consider them as candidate edge points. For these points, we compute the phase of the wavelet transform coefficient $A_s f(x, y)$ that represents the angle perpendicular to the edge direction.

Step 4: We compute the Gabor and Cauchy wavelet transform of the image at those points with θ equal to the computed phase of the wavelet transform coefficient in the Step 3.

Step 5: Since the associated filter to the Gabor or Cauchy wavelet transform has the maximum sensitivity at that direction, then the magnitude of the directional wavelet transform coefficients $W_{s,\theta} f(x, y)$ can be thresholded using Otsu's single thresholding method, to decide if that point is an edge point or not.

In Fig. 20, the noisy step image with SNR = 16.32 dB is experimented. Let the total number of scales $M = 2$ and the total number of orientations $N = 12$ .Angular resolution is $\frac{\pi}{12}$. J. M. Niya et al. implement the ordinary wavelet-based edge detection technique in the section 2 and use Otsu’s double thresholding method (Fig. 20 (c) and (d)). Using directional wavelet transform for candidate edge points as described in 7.3, we obtain superior edge images (Fig. 20 (e) for Gabor wavelet and Fig. 20 (f) for Cauchy wavelet).

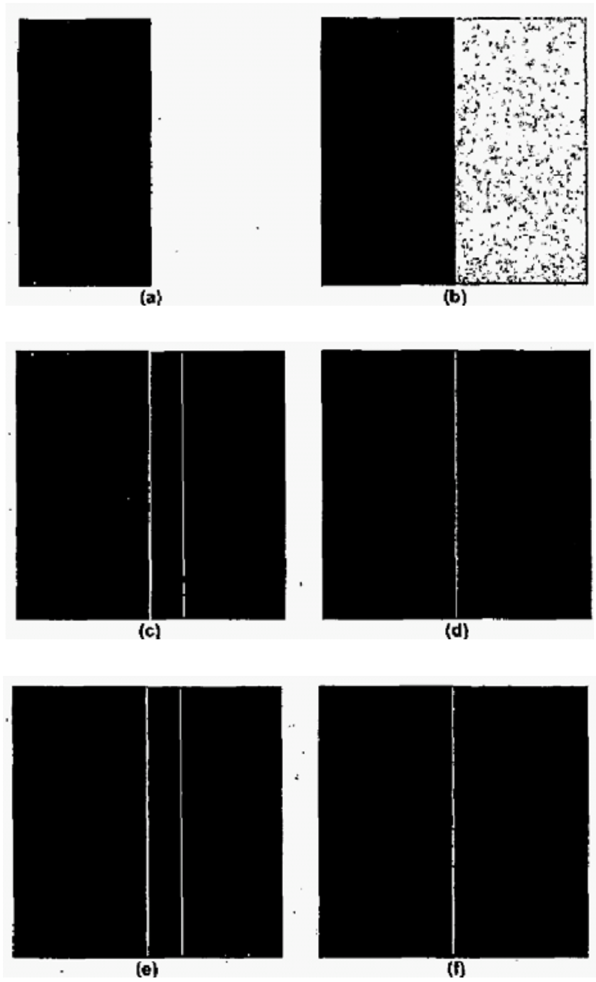


Fig. 20 : Noisy step image and its edge pictures: (a) Step image. (b) Noisy image. (c) Lower thresholded edge picture. (d) Upper thresholded edge picture. (e) Edge picture via Gabor directional wavelet transform. (f) Edge picture via Cauchy directional wavelet transform rising the algorithm proposed. (This figure is extracted from [7] : http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1461748&isnumber=31419)

8. The Rough Introduction of The Dual-Tree Complex Wavelet Transform

From section 3~7, we can see many application of the wavelet-based edge detection method are all based on the dyadic wavelet transform (DWT) which has many benefits, such as its efficient computational algorithm and sparse representation. Nevertheless, it still suffers from four fundamental shortcomings which make their performance of the edge detection be not very well. [8]

Problem 1 : Oscillations

Because wavelets are in fact the bandpass functions, the wavelet coefficients tend to oscillate positive and negative around singularities [8]. This condition considerably complicates wavelet-based processing, making singularity extraction and signal modeling very challenging.

Problem 2 : Shift Variance

A small shift of the signal greatly perturbs the wavelet coefficient oscillation pattern around singularities. This drawback complicates wavelet-domain processing. Algorithms must be made capable of coping with the wide range of possible wavelet coefficient patterns caused by shifted singularities.

Problem 3 : Aliasing

While the two channels (approximation and detail) in an octave recombine at the end to perfectly reproduce the signal seen at the input, the anti-aliasing is somewhat fragile [6].

Problem 4 : Lack of directionality

The multi-dimensional wavelet transform spreads out information, like ripples from a drop splashing in a pool of water. This lack of directionality greatly complicates modeling and processing of geometrical image features like ridges and edges. One solution is using the directional wavelets which is introduced in the section 7, the other is the complex wavelet.

To solve above four problems of DWT, Selesnick et al. discuss the complex wavelet transform which is also abbreviated to CWT (not to be confused with the continuous wavelet transform(CWT)). They claim that image processing, especifically edge detection, can be enhanced with their technique. They propose using different filters at each stage of transform and the transform is directionally selective, meaning that it does not spread out the edges like the way that a separable 2-D DWT does.

The dual-tree complex wavelet transform (CWT) is a valuable enhancement of the traditional real wavelet transform (DWT) that is nearly shift invariant and directionally selective in two and higher dimensions. The multidimensional (M-D) dual-tree CWT is non-separable but is based on a computationally efficient, separable filter bank (FB) [8]. Since the real and imaginary parts of the dual-tree CWT are, in fact, conventional real wavelet transforms, the CWT benefits from the vast theoretical, practical, and computational resources that have been developed for the standard DWT. For example, software and hardware developed for implementation of the real DWT can be used directly for the CWT. But, in addition, the magnitude and phase of CWT coefficients can be exploited to develop new effective wavelet-based algorithms, especially for applications that the DWT is not suitable or has bad performance.

An Example of the dual-tree complex wavelet for edge detection is illustrated in Fig. 21. When the test image is reconstructed from one level of the DWT coefficients, ringing and aliasing effects are apparent (Fig. 21 (b)). However, the reconstruction of the image from one level of the CWT does not exhibit these phenomena.

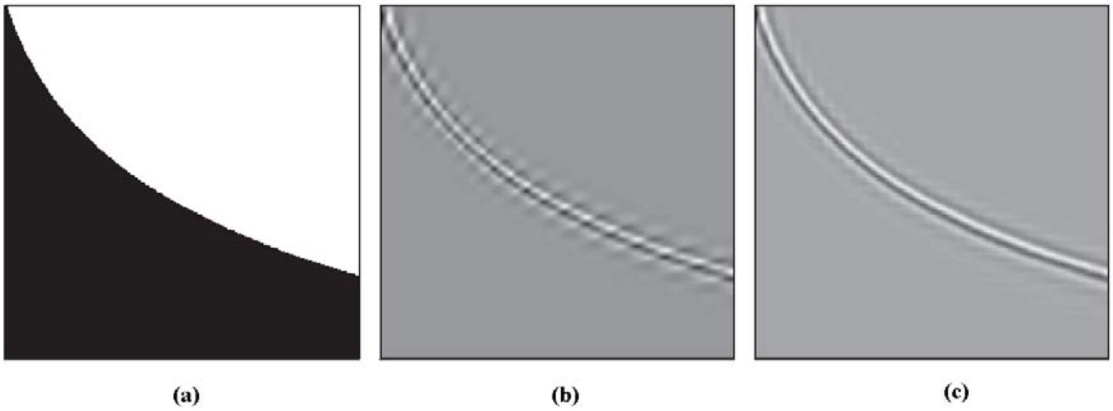


Fig. 21 : (a) Test image with sharp edge on hyperbolic trajectory. (b) DWT. (c) CWT.
(This figure is extracted from [8])

9. Performance Analysis for Edge Detection

To determine the effect of the noise in the image, we can use SNR to judge it. The figure of merit and the mean square distance of the localization accuracy are the methods to evaluate the performance of the edge detection in addition to the missed detection rate (MDR) and the false alarm rate.

The *figure of merit* F of Pratt [5] which is defined as

$$F = \frac{1}{\max\{N_I, N_A\}} \sum_{k=1}^{N_A} \frac{1}{1+\alpha\, d^2(k)} \tag{61}$$

where N_I is the number of the actual edges and N_A is the number of the detected edges. $d(k)$ denotes the distance from the kth actual edge to the corresponding detected edge. α is a scaling constant set to 1/9 as in Pratt's work. The greater the F, the better the detection results.

Take the scale multiplication scheme and the directional wavelets for example.

Example 1: Scale Multiplication

- Isolated Step Edge

Fig. 22 (a) shows a 256 x 256 isolated step edge corrupted by the Gaussian white noise. Fig. 22 (b) is the detection result of the scale multiplication. Fig. 22 (c) and (d) are the detecting results of the small scale 2^3 and large scale 2^4.

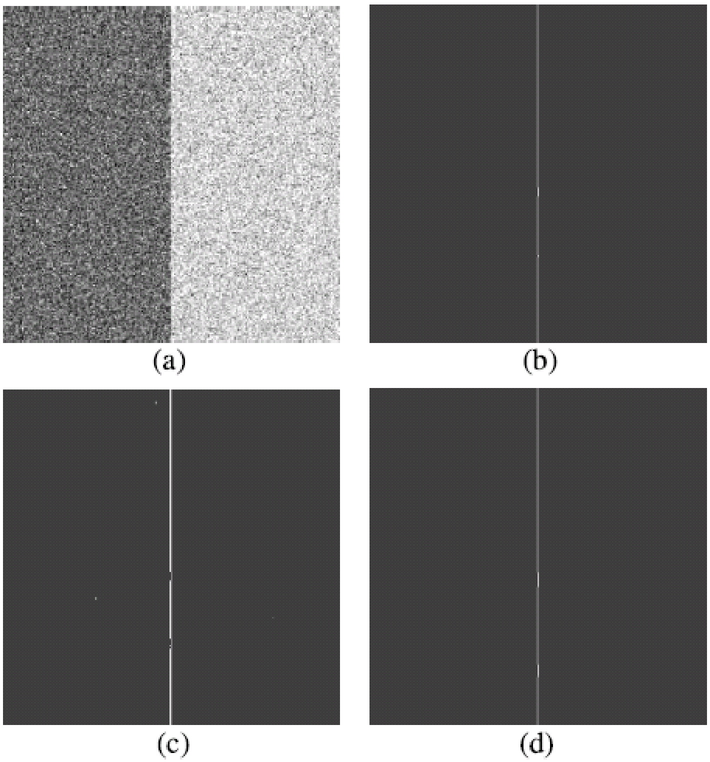


Fig. 22 : Noisy step and edge maps. (a) Noisy step edge. (b) by scale multiplication scheme. (c) by scale 2^3. (d) by scale 2^4.

Denote the figure of merit of Fig. 22(b)~(d) as F_p, F_1, F_2 respectively. The Table 1 shows that F_p is greatest, so the performance of the scale multiplication scheme is better than individual scales. F_1 is less than F_2 for the some false edges are caused by noise.

Table 1: The figure of merit values of the two scales and their multiplication for the isolated step edge

F_P	F_1	F_2
0.9929	0.9496	0.9877

(Fig. 22 and Table 1 are extracted from [5] : http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V15-45XT9HX-5&_user=7761193&_rdoc=1&_fmt=&_orig=search&_sort=d&_docanchor=&view=c&_searchStrId=1174726732&_rerunOrigin=google&_acct=C000051951&_version=1&_urlVersion=0&_userid=7761193&md5=06ba66b85df7405a30dd2110c140c753)

- <u>Neighboring Step Edges</u>

If two or more edges occur in a neighborhood, they may interfere each other with the increasing of the width of the wavelet. Fig. 23 (a) shows a 256 x 256 neighboring step edges corrupted by the Gaussian white noise. Fig. 23 (b) is the detection result of the scale multiplication. Fig. 23 (c)(d) are the detecting results of the small scale 2^3 and large scale 2^4.

Generally, with a single scale, it is intricate to properly balance the edge dislocation and the noise sensitivity. With the scale multiplication, this problem can be largely resolved.

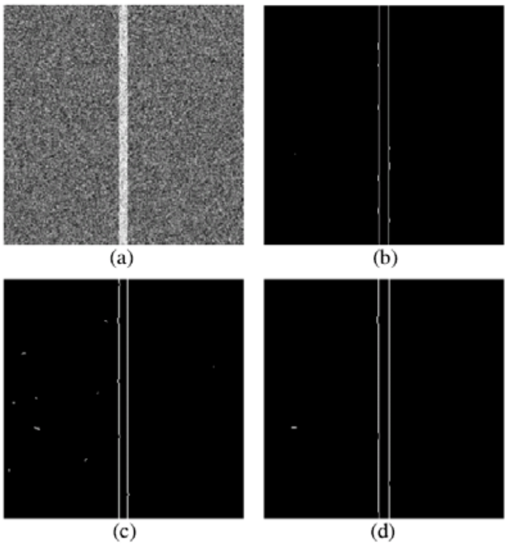


Fig. 23 : Noisy two neighboring steps and edge maps. (a) Noisy neighboring step edges. (b) by scale multiplication scheme. (c) by scale 2^3. (d) by scale 2^4.

Table 2: The figure of merit values of the two scales and their multiplication for the neighboring step edges

F_P	F_1	F_2
0.9899	0.9477	0.8914

(Fig. 23 and Table 2 are extracted from [5] : http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V15-45XT9HX-5&_user=7761193&_rdoc=1&_fmt=&_orig=search&_sort=d&_docanchor=&view=c&_searchStrId=1174726732&_rerunOrigin=google&_acct=C000051951&_version=1&_urlVersion=0&_userid=7761193&md5=06ba66b85df7405a30dd2110c140c753)

Observe Fig. 23, in comparison with the small scale (2^3), edges are enhanced and noise is diluted in the scale multiplication method. Moreover, the dislocation of the edges will be significantly improved in comparison with the large scale (2^4). From Table 2, F_p is still the greatest and F_1 is also less than F_2 for the some false edges are caused by noise. Therefore, Fig. 23 (b) possesses both the advantage of Fig. 23 (c) and (d) with few false edges and little edge dislocation.

Example 2: The directional wavelets

Denote the figure of merit of Fig. 20(c)~(f) as F_1, F_2, F_3, F_4 respectively. The Table 3 shows that the two-step edge detection based on directional wavelets detects edges better than the ordinary wavelet-based algorithm. And also the Cauchy wavelet shows superior performance compared to the Gabor wavelet.

Table 3 : The figure of merit values of 4 edge images described above

F_1	F_2	F_3	F_4
0.9277	0.9333	0.9833	0.9841

(Table 3 is extracted from [7] :
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1461748&isnumber=31419)

The second measurement of the performance for edge detection is the location accuracy of the edge images. If the distance $d(k)$ is not greater than 4 pixels, this edge is considered as a true edge. Denote N As the total number of true edges that are detected, L. Zhang and P. Bao [5] define *the mean square distance* as

$$D = \sqrt{\frac{1}{N}\sum_{i=1}^{N} d^2(i)} \tag{62}$$

The smaller the D, the better the localization accuracy will be achieved.

Take the scale multiplication scheme and the directional wavelets for example.

Table 4: The mean square distance values of the two scales and their multiplication **for the isolated step edge**

D_P	D_1	D_2
0.1782	0.2271	0.2887

Table 5: The mean square distance values of the two scales and their multiplication **for the neighboring step edges**

D_P	D_1	D_2
0.2852	0.1837	1.0049

(Table 4 and Table 5 are extracted from [5] :
http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V15-45XT9HX-5&_user=7761193&_rdoc=1&_fmt=&_orig=search&_sort=d&_docanchor=&view=c&_searchStrId=1174726732&_rerunOrigin=google&_acct=C000051951&_version=1&_urlVersion=0&_userid=7761193&md5=06ba66b85df7405a30dd2110c140c753)

Denote the mean square distance of Fig. 22 and 23(b)~(d) as D_p, D_1, D_2 respectively. It can been from Table 4 that D_p is less than not only D_1 but also D_2. This implies that the scale multiplication improves the localization accuracy significantly while keeping high detection efficiency. For the neighboring step edges, Table 5 shows that D_p is slightly larger than D_1 and both of them are much smaller than D_2, implying that much dislocation of the edges occur in Fig. 23(d).

10. Conclusion

To get the accurate edge map, we need multi-resolution to obtain the strong and weak real edges simultaneously. The scales of the classical edge detectors such as Sobel, Prewitt, etc., however are not adjustable. Although the scale of the Gaussian function in Canny edge detection can be modulated, it still suffers from some practical limitations. So, the wavelet transform is a better choice to do the edge detection in that we can construct our own edge detectors with proper scales. The dyadic wavelet transform which is in fact the continuous WT with discrete coefficients is always be viewed as the discrete wavelet transform (DWT) in many papers. The wavelet transform is able to characterize the local regularity of signals by decomposing the signals into low frequency and high frequency blocks that are well localized both in space and frequency. 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. However, if we now deal with the noisy images, the noise level (SNR) is very sensitive to the change of scales and the dyadic sequence of scales cannot always optimally adapt to the effect of noise. Hence, we have to consider the Lipschitz regularity which is computed by the wavelet coefficients. According the magnitude and the polarity of the Lipschitz regularity, we can set proper scales in different locations of the same image meanwhile in order to eliminate more noise and keep the real edges accurately.

Another important issue for the edge detection is thresholding. We usually use only one threshold in the wavelet muti-scale edge detection which will miss the weak edges. So an adaptive thresholding method with the wavelet coefficients is introduced in the section 4, and a novel synthesis method of the multi-scale edges is also illustrated.

We know that the discrete wavelet transform decompose the signal into the LL, LH, HL, and HH parts. For the edge detection, we usually only preserve the LH and HL areas and ignore other blocks. This action makes the performance of the edge detection poor and the noise still exist. So a new edge detection scheme which combine the Canny operator and the wavelet transform is introduced in the section 5. The low frequency edges are detected by Canny and the high frequency edges are detected by the wavelet modulus maxima algorithm to restore the edges after denoising by the wavelet. Finally, the low and high frequency sub-images are integrated by a proposed fusion rule.

In the traditional multi-scale techniques, we always form the edge map at several scales first and then synthesize them together to obtain the final result. The dyadic transform at two adjacent scales are multiplied as a product function in the scale multiplication scheme. This proposed method determine the edge points as the local maxima of directly on the product function which avoid the ill-posed synthesis process in most multi-scale detection schemes. Moreover, this method outperforms on the accurate edge location and improve the dislocation problem of the neighboring edges when the width of the mother wavelet is set large to smooth the noise.

The dyadic wavelet transform has the good property of multi-resolution in edge detection and it localizes the edges properly. However, it only detects the edges in the horizontal and vertical directions. The edges in practical have a variety of directions. Also, if the edges to be detected or extracted have a specified direction (straight lines, oriented textures, and etc.), the separable isotropic wavelet filters fail. So we need the directional wavelets like the Gabor and the Cauchy wavelets to solve above problems. Because of their complex computations, we use these two directional wavelets in the edge-like points which magnitude is between the low and high threshold. For the surely edge and nonedge points, we use the ordinary isotropic 2-D wavelet transform.

Finally, I introduce the dual-tree complex transform for its specialty. It improves the four shortcomings of the DWT : oscillation, shift variance, aliasing, and lack of directionality and make the results of edge detection better. To understand about it more, please refer to [8].

By the way, two performance measurements: *the figure of merit* and *the localization accuracy with the mean square distance* are also be mentioned to investigate the performance of different edge maps with varying methods.

All in all, it is apparent that the DWT is a well accepted technique to detect edges currently. However, the scoring method or the performance measurements are not fully and clearly explained in some of the papers. Furthermore, few images are experimented in most of the papers. For these two reasons, it is difficult to convince the readers that the proposed method is really better than other schemes. Besides, the noise in the noisy image is always assumed to be the white noise. In practice, it is almost random and not white. So, the assumption is too narrow to make the outcomes convincible. The above comments are not suggest that which methods are not valid useful, but we all know that there are still areas remaining that can be refined, advanced and further tested.

11. References

[1] S. Mallat, W. L. Hwang, “Singularity detection and processing with wavelet,” *IEEE Tran. Inform. Theory*, Vol. 38, No. 2, March, 1992.

[2] Jun Li, “A wavelet approach to edge detection,” *Sam Houston State University*, August, 2003.

[3] S. Wenchang, S. Jianshe, Z. Lin, “Wavelet Multi-scale Edge Detection Using Adaptive threshold,” *IEEE*, 2009.

[4] J. Pan, “Edge detection combining wavelet transform and Canny operator based on fusion rules,” *IEEE Proceedings of the 2009 international conference on wavelet analysis and pattern recognition*, July, 2009.

[5] L. Zhang, P. Bao, “Edge detection by scale multiplication in wavelet domain,” *Elsevier Science B.V. Pattern Recognition letters 23*, pp. 1771-1784, 2002.

[6] E. Brannock, M. Weeks, “A synopsis of recent work in edge detection using the DWT,” *IEEE*, 2008.

[7] J. M. Niya, A. Aghagolzadeh, M. A. Tinati, S. Feizi, “2-step wavelet-based edge detection using Gabor and Cauchy directional wavelets,” *IEEE Trans. Acoustic, Speech, Signal processing*, vol. 37, pp. 2091-2110, Dec., 1989.

[8] I. W. Selesnick, R. G. Baraniuk, N. G. Kingsbury, “The dual-tree complex wavelet transform, ” *IEEE Signal Processing Magazine*, pp.123-151, Nov. 2005.

[9] V. Skorpil, J. Stastny, “Wavelet transform for image analysis,” *IEEE*, 2003.

[10] O. Bezvesilniy, V. Vinogradov, D. Vavriv, K. Schunemann, “Wavelet-based image processing: edge detection and noise reduction,” *ICECom*, Oct., 2003.

[11] A. Sur, N. Patra, S. Chakraborty, I. Saha, “A new wavelet based edge detection technique for iris imagery,” *IEEE IACC*, 2009.

[12] C. Q. Zhan, L. J. Karam, “Wavelet-based adaptive image denoising with edge preservation,” *IEEE*, 2003.

[13] E. Brannock, M. Weeks, “Edge detection using wavelets,” *ACM SE*, March, 2006.

[14] X. Wang, “image edge detection based on lifting wavelet,” *IEEE International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2009.

[15] M. Chabert, J.Y. Tourneret, G. Mesnager, “Edge detection in speckled SAR images using the continuous wavelet transform,” *IEEE*, 1996.

[16] M. Sonka, V. Hlavac, R. Boyle, *Image processing, analysis, and machine vision*, Thomson Learning, 2008.

[17] J. Canny, “A computational approach to edge detection,” *IEEE Trans. Patt. Anal. Machine Intell.*, Vol. PAMI-8, pp.679-698, 1986.

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

[19] G. Pajares, “A wavelet-based image fusion tutorial,” *Pattern recognition*, Vol.37 (9): pp.1855-1872, 2004.

[20] Bill Green, "Canny Edge Detection Tutorial," 2002, available at http://www.dspguide.com/ch24/7.htm
[21] Wikipedia. "Lipschitz Continuity," http://en.wikipedia.org/wiki/Lipschitz_continuity.