ELSEVIER

Pattern Recognition Letters 23 (2002) 1771–1784

Pattern Recognition Letters

www.elsevier.com/locate/patrec

Edge detection by scale multiplication in wavelet domain

Lei Zhang *, Paul Bao

Department of Computing, The Hong Kong Polytechnic University, Hung Hum, Kowloon, Hong Kong

Received 21 September 2001; received in revised form 17 January 2002

Abstract

This paper proposes a wavelet based edge detection scheme by scale multiplication. The dyadic wavelet transforms at two adjacent scales are multiplied as a product function to magnify the edge structures and suppress the noise. Unlike many multiscale techniques that first form the edge maps at several scales and then synthesize them together, we determined the edges as the local maxima directly in the scale product after 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. Experiments on natural images are compared with the *Laplacian of Gaussian* and *Canny edge detection* algorithms.

Keywords: Edge detection; Scale multiplication; Wavelet transforms

1. Introduction

Edge detection is an essential process in image analysis and many techniques have been proposed. Some edge detection filters were developed with optimality (Canny, 1986; Shen and Castan, 1992). Canny (1986) evaluated the detectors by three criteria: *good detection*, *good localization* and *low spurious response*, and he showed that the optimal detector for an *isolated step* edge should be the *first derivative of Gaussian*.

The optimal *Canny edge detector* for *ramp* edges was proposed by Petrou and Kittler (1991). Canny restricted the detector as a finite impulse response (FIR) filter. Sarkar and Boyer (1991) extended it to infinite impulse response (IIR) filter. Besides the shape of the detector, another important problem is to set a proper detection scale. As suggested by Marr and Hildreth (1980), multiple scales should be employed to describe the variety of the edge structures. Then these multiscale descriptions will be synthesized to form an edge map.

Canny (1986) used a fine-to-coarse feature synthesis strategy to mingle the multiscale edge information based on a set of predefined rules. Bergholm (1987) combined the multiscale edges in a coarse-to-fine tracking manner. The RRES (reasoning about edges in scale space) scheme of Lu and Jain (1992) tends to be more complex with so many knowledge rules and continuous scale space.

Considering that the synthesis of the multiscale edges is intricate and itself an ill-posed problem,

* Corresponding author.
E-mail address: cslei@inet.polyu.edu.hk (L. Zhang).

0167-8655/02/$ - see front matter
PII: S0167-8655(02)00151-4

Jeong and Kim (1992) selected an optimal scale adaptively for each of the pixels by minimizing an objective function, but the results suffered from the complicated shape of the function and the sensitivity to the initial scale. Ziou and Tabbone (1993) ran a subpixel *Laplacian* operator at two scales and recovered the edges with four-step edge models. Park et al. (1995) divided an image into several regions based on a discontinuity measure calculated over a window and then selected different resolution (i.e. scale) for each of the regions.

This paper wants to find a simple but efficient multiscale scheme. Wavelet transform (WT) is naturally a multiscale analysis and some WT-based edge detection techniques were proposed (Mallat and Zhong, 1992; Sadler and Swami, 1999; Aydn et al., 1996). The dyadic wavelet constructed by Mallat and Zhong (1992) is a *quadratic spline*, which approximates *the first derivative of Gaussian*. The corresponding dyadic wavelet transform (DWT) is equivalent to the *Canny edge detection*. Mallat calculated the local maxima of DWT at each scale and formed a multiscale edge representation of an image.

Signals and noise have different singularities mathematically (Mallat and Hwang, 1992). In DWT domain, it is represented that the edge structures present observably at each subband while noise decreases rapidly along the scales. It has been observed that multiplying the adjacent scales could sharpen edges while diluting noise (Xu et al., 1994; Sadler and Swami, 1999).

In this paper, we presented a scale-multiplication-based edge detection scheme. Two DWT subbands are multiplied as a product function. Unlike many multiscale edge detectors, where the edge maps were formed at several scales and then synthesized together, our scheme determines edges as the local maxima in the product function after a thresholding. The scale multiplication enhances image structures and suppresses noise. An integrated edge map will be formed efficiently while avoiding the ill-posed edge synthesis process. It will be shown that much improvement is obtained on the localization accuracy and the detection results are better than using either of the two scales only.

It is often required to increase the filter's scale to better smooth noise. An edge would disappear or be dislocated if there is another edge curve at its neighborhood. In this paper, it is also found that the scale multiplication will reduce the interference of neighboring edges a lot.

The paper is organized as follows. Section 2 introduces DWT. In Section 3, the scale-multiplication-based edge detection scheme is presented and analyzed by synthetic images. In Section 4, experiments are performed on natural images in comparison with the *Laplacian of Gaussian* (LOG) and Canny edge detection algorithms. The conclusion is given in Section 5.

2. The dyadic wavelet transform

A function $\psi(x)$ is called a wavelet if its average is equal to 0. The DWT of $f(x)$ at dyadic scale 2^j and position x is

$$W_j f(x) = f * \psi_j(x) \tag{1}$$

where $*$ denotes convolution operation and ζ_j denotes the dyadic dilation of function ζ

$$\zeta_j(x) = 2^{-j}\zeta(2^{-j}x) \tag{2}$$

For the details please refer to Daubechies (1992) and Mallat and Zhong (1992).

The DWT can be designed as a multiscale edge detector that is equivalent to Canny edge detector. Suppose that $\theta(x)$ is a differentiable smooth function whose integral is 1 and converges to 0 at infinity. Let wavelet $\psi(x)$ be the first order derivative of $\theta(x)$

$$\psi(x) = \mathrm{d}\theta(x)/\mathrm{d}x \tag{3}$$

Then

$$W_j f(x) = f * \psi_j(x) = f * \left(2^j \frac{\mathrm{d}\theta_j}{\mathrm{d}x}\right)(x) = 2^j \frac{\mathrm{d}}{\mathrm{d}x}(f * \theta_j)(x) \tag{4}$$

The $W_j f(x)$ is proportional to the first derivative of $f(x)$ smoothed by a smooth function $\theta_j(x)$. When $\theta(x)$ is *Gaussian*, the determination of the local extrema of $W_j f(x)$ is equivalent to the Canny edge detection.

The wavelet used in this paper is the Mallat wavelet (Mallat and Zhong, 1992). The corresponding $\theta(x)$ is a cubic spline, and thus $\psi(x)$ is a quadratic spline.

$$\theta(x)=\begin{cases}0 & |x|\geqslant 1\\ \theta(-x) & 0\leqslant x\leqslant 1\\ -8x^3-8x^2+4/3 & -0.5\leqslant x\leqslant 0\\ 8(x+1)^3/3 & -1\leqslant x\leqslant -0.5\end{cases} \tag{5}$$

$$\psi(x)=\begin{cases}0 & |x|\geqslant 1\\ -\psi(-x) & 0\leqslant x\leqslant 1\\ -24x^2-16x & -0.5\leqslant x\leqslant 0\\ 8(x+1)^2 & -1\leqslant x\leqslant -0.5\end{cases} \tag{6}$$

In Fig. 1, $\theta(x)$ and $\psi(x)$ are plotted as well as a Gaussian function and its first derivative. It can be seen that $\theta(x)$ approximates the Gaussian function closely and the DWT is equivalent to Canny edge detection.

In the case of images, two wavelets $\psi^1(x,y)$ and $\psi^2(x,y)$ should be utilized. Suppose $\theta(x,y)$ is a 2-D differentiable smooth function whose integral is 1 and converges to 0 at infinity. The two wavelets are:

$$\psi^1(x,y)=\frac{\partial\theta(x,y)}{\partial x},\qquad \psi^2(x,y)=\frac{\partial\theta(x,y)}{\partial y} \tag{7}$$

Denote

$$\zeta_j(x,y)=2^{-2j}\zeta(2^{-j}x,2^{-j}y) \tag{8}$$

the dilation of $\zeta(x,y)$ by 2^j, the WT of $f(x,y)$ at scale 2^j and position (x,y) has two components

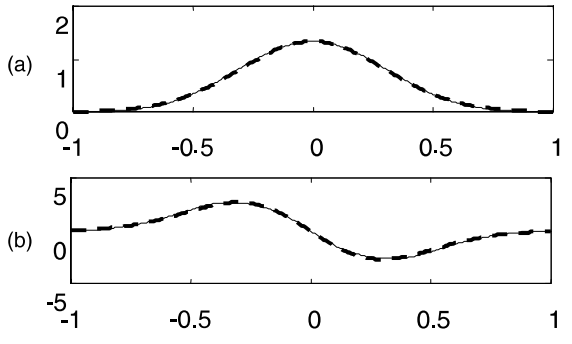


Fig. 1. (a) The smooth function $\theta(x)$ (—) and a Gaussian function (- - -). (b) Wavelet $\psi(x)$ (—) and the first derivative of the Gaussian in (a) (- - -).

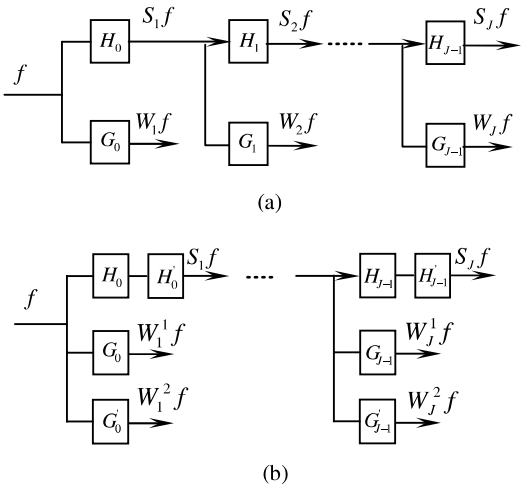


Fig. 2. (a) 1-D DWT, (b) 2-D DWT. Here filter $H_j(G_j)$ is the 2^j dilation of $H_0(G_0)$ (putting 2^j-1 zeros between each of coefficients of $H_0(G_0)$) and $H_j'(G_j')$ is the transition of $H_j(G_j)$.

$$W_j^1f(x,y)=f*\psi_j^1(x,y)$$

and

$$W_j^2f(x,y)=f*\psi_j^2(x,y) \tag{9}$$

The 1-D and 2-D DWT are illustrated in Fig. 2. Filter $H_j(G_j)$ is the 2^j dilation of $H_0(G_0)$ (i.e. placing 2^j-1 zeros between each of the coefficients of $H_0(G_0)$) and $H_j'(G_j')$ is the transition of $H_j(G_j)$.

3. Edge detection by scale multiplication

3.1. The discrimination of singularity and noise by DWT

The behavior of signal across scales in wavelet domain depends on the local regularity that can be measured by *Lipschitz exponents* mathematically. A function $f(x)$ is *Lipschitz* α at x_0 $(0\leqslant\alpha\leqslant 1)$ if and only if there exists a constant K_1 such that in the neighborhood of x_0:

$$|f(x)-f(x_0)|\leqslant K_1|x-x_0|^{\alpha} \tag{10}$$

We call the superior bound of all α satisfying Eq. (10) as *Lipschitz regularity*. And the relations between α and wavelet amplitude is described by (Mallat and Hwang, 1992):

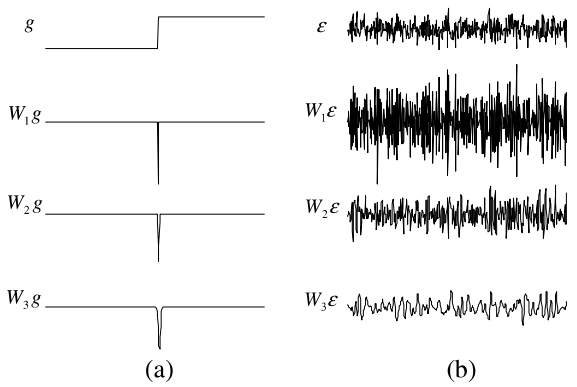


Fig. 3. (a) The DWT of step g at the first three scales, (b) the DWT of noise ε at the first three scales.

$$|W_j f(x)| \leqslant K_2 (2^j)^{\alpha} \tag{11}$$

where K_2 is a constant.

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 $-1/2$.

It can be seen from Eq. (11) that for singularities of signals, whose Lipschitz regularities are $\alpha \geqslant 0$, 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, the transform amplitudes will decrease rapidly along the scales. In Fig. 3 the DWT of a step function and noise are illustrated. It can be observed that the DWT amplitudes of the step are large across scales, but those of noise decay rapidly.

3.2. The scale multiplication

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

$$P_j^f(x) = W_j f(x) W_{j+1} f(x) \tag{12}$$

Subscript j means that the correlation is computed at scales 2^j and 2^{j+1}.

As Fig. 3 has shown, the peaks due to edges tend to propagate across scales. The production function will enhance the edge structures. But if $f(x)$ is *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 (Mallat and Hwang, 1992). Directly multiplying the DWT at adjacent scales will dilute the noise.

With scale varying along dyadic sequence (2^j), $j \in Z$, the support of wavelet base $\psi_j(x)$ will increase rapidly. This is also to say $W_j f(x)$ will become smoother rapidly along scales. If three or more adjacent scales were incorporated in the multiplication, edges would not be sharpened more but much edge dislocation would occur. So it is appropriate to analyze the multiplication using two scales.

In Fig. 4(a), a block signal g and its noisy version $f = g + \varepsilon$ are illustrated, where ε is Gaussian white noise. Their DWT at the first three scales are given in Fig. 4(b) and (c). It is shown that at the finest scale the wavelet coefficients $W_1 f$ are almost dominated by noise. At the second and third scales, the noise diluted rapidly. It can also be seen that at the small scales the positions of the step edges are better localized. But some noise may be falsely considered as edges. At the large scales, the SNR is improved and edges can be detected more correctly but with the decreasing of the accuracy of the edge location. In Fig. 4(d), the product P_j^f, $j = 1$–3, are illustrated. Apparently the step edges are more observable in P_j^f than in $W_j f$.

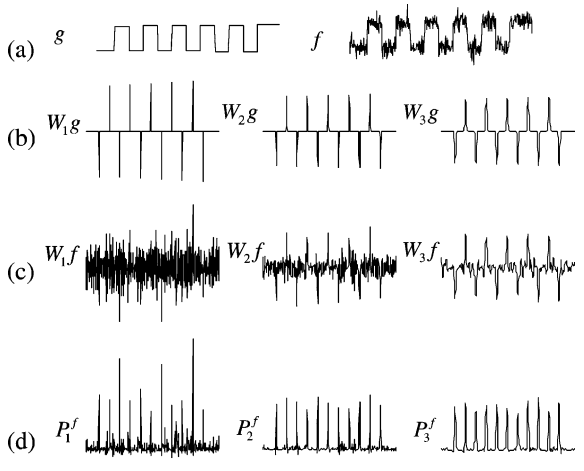


Fig. 4. (a) Blocks g and its noisy version f, (b) the DWT of g at the first three scales, (c) the DWT of noisy f at the first three scales, (d) the product function P_j^f with $j = 1$–3.

3.3. The thresholding

In the first-derivative-based edge detection schemes, the gradient image should be thresholded to eliminate false edges produced by noise. With a single threshold t, there will still be some false edges if t is small, and portions of a contour may be missed if t is too large. In the Canny edge detection, a double thresholding algorithm was employed. After non-maxima suppression, a low threshold t_{l} and a high threshold $t_{\mathrm{h}} \approx 2t_{\mathrm{l}}$ are applied to obtain double thresholded edge images, I_{l} and I_{h}. The algorithm selects edges in I_{l} that link to the edges in I_{h}.

The double thresholding algorithm can also be applied in our scheme. Notice that edges and noise can be better distinguished in the scale product than in a single scale, a properly chosen threshold could suppress the noise maxima effectively. Due to this observation, 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 $P_j^f(x)$ is less than zero, the point will be considered as noise and filtered out.

Refer to Fig. 2(a), suppose that the input is Gaussian white noise $\varepsilon \sim N(0, \sigma^2)$. The DWT of ε on scale 2^j is $W_j\varepsilon(x) = \varepsilon * \psi_j(x)$. For expression convenience, we denote $X_j(x) = W_j\varepsilon(x)$ and the product function is rewritten as $Y_j(x) = P_j^{\varepsilon}(x) = X_j(x)X_{j+1}(x)$.

Obviously, X_j is *Gaussian colored* noise and $X_j \sim N(0, \sigma_j^2)$, where $\sigma_j = \|\psi_j\|\sigma$ and

$$\|\psi_j\| = \sqrt{\int \psi_j^2(x)\,\mathrm{d}x} \tag{13}$$

X_j and X_{j+1} are highly correlated due to the similarities between wavelet bases ψ_j and ψ_{j+1}.

Denote $t_{\mathrm{sc}}(j)$ the threshold applied to $Y_j(x)$. For the de-noising purpose, it is expected that $t_{\mathrm{sc}}(j)$ could suppress almost all the values in $Y_j(x)$, i.e. $P(y_j < t_{\mathrm{sc}}(j)) \to 1$. We give a form of threshold $t_{\mathrm{sc}}(j)$ as follows.

Normalize X_j and X_{j+1} as:

$$\overline{X}_j = X_j/\sigma_j, \qquad \overline{X}_{j+1} = X_{j+1}/\sigma_{j+1} \tag{14}$$

So $\overline{X}_j$, $\overline{X}_{j+1} \sim N(0,1)$. Define

$$\overline{Y}_j(x) = \overline{X}_j(x) \cdot \overline{X}_{j+1}(x) \tag{15}$$

and then

$$Y_j(x) = \|\psi_j\| \cdot \|\psi_{j+1}\|\sigma^2 \cdot \overline{Y}_j(x) \tag{16}$$

Let

$$\overline{Y}_{j,+}(x) = (\overline{X}_j(x) + \overline{X}_{j+1}(x))/2$$

and

$$\overline{Y}_{j,-}(x) = (\overline{X}_j(x) - \overline{X}_{j+1}(x))/2 \tag{17}$$

Then

$$\overline{Y}_j(x) = \overline{Y}_{j,+}^2(x) - \overline{Y}_{j,-}^2(x) \tag{18}$$

and $\overline{Y}_{j,+}(x)$, $\overline{Y}_{j,-}(x)$ are Gaussian distributed: $\overline{Y}_{j,+} \sim N(0, \sigma_{j,+}^2)$ and $\overline{Y}_{j,-} \sim N(0, \sigma_{j,-}^2)$, where

$$\sigma_{j,+} = \frac{1}{2}\sqrt{\int \left(\psi_j(x)/\|\psi_j\| + \psi_{j+1}(x)/\|\psi_{j+1}\|\right)^2 \mathrm{d}x},$$

$$\sigma_{j,-} = \frac{1}{2}\sqrt{\int \left(\psi_j(x)/\|\psi_j\| - \psi_{j+1}(x)/\|\psi_{j+1}\|\right)^2 \mathrm{d}x} \tag{19}$$

Since there is a strong correlation between $\psi_j(x)$ and $\psi_{j+1}(x)$, so that $\sigma_{j,+}^2$ is much more than $\sigma_{j,-}^2$. Take the wavelet used in this paper as an example, for any scale parameter j the ratio of $\sigma_{j,+}^2$ to $\sigma_{j,-}^2$ is about 5.44. (In discrete implementations, the ratio is of some distortion, at the first three scales the ratios are: $\sigma_{j,+}^2/\sigma_{j,-}^2 = 3.30, 4.62, 5.22$. When $j > 3$ the ratio approximates 5.44.)

Let $\bar{t}_{\mathrm{sc}}(j) = t_{\mathrm{sc}}(j)/(\|\psi_j\| \cdot \|\psi_{j+1}\|\sigma^2)$, we have

$$\begin{aligned} P(y_j < t_{\mathrm{sc}}(j)) &= P(\bar{y}_j < \bar{t}_{\mathrm{sc}}(j)) \\ &= P(\bar{y}_{j,+}^2 < \bar{t}_{\mathrm{sc}}(j) + \bar{y}_{j,-}^2) \\ &\geqslant P(\bar{y}_{j,+}^2 < \bar{t}_{\mathrm{sc}}(j)) = P(|\bar{y}_{j,+}| < \sqrt{\bar{t}_{\mathrm{sc}}(j)}) \end{aligned} \tag{20}$$

$\overline{Y}_{j,+}$ is Gaussian distributed, $\sqrt{\bar{t}_{\mathrm{sc}}(j)} \geqslant 4\sigma_{j,+}$ will lead to

$$P(|\bar{y}_{j,+}| < \sqrt{\bar{t}_{\mathrm{sc}}(j)} \,|\, \sqrt{\bar{t}_{\mathrm{sc}}(j)} \geqslant 4\sigma_{j,+}) > 0.9999$$

Therefore

$$\begin{aligned} P(y_j < t_{\mathrm{sc}}(j) \,|\, t_{\mathrm{sc}}(j) &\geqslant 16\|\psi_j\| \cdot \|\psi_{j+1}\|\sigma^2\sigma_{j,+}^2) \\ &\geqslant P(|\bar{y}_{j,+}| < \sqrt{\bar{t}_{\mathrm{sc}}(j)} \,|\, \sqrt{\bar{t}_{\mathrm{sc}}(j)} \geqslant 4\sigma_{j,+}) \to 1 \end{aligned}$$

In real applications, the input is $f = g + \varepsilon$ where g is the original image and f is a measurement. Due to the linearity of WT, the DWT of f can be written as $W_j f = W_j g + W_j \varepsilon$. At fine scales, $W_j \varepsilon$ will be predominant in $W_j f$ except for some significant edge structures to be detected. Since the contrast of image singularities and noise is greatly amplified in P_j^f, threshold $t_{sc}(j)$ will be much effective in discriminating edges from noise. In our experiments a setting of

$$t_{sc}(j) = c \cdot \|\psi_j\| \cdot \|\psi_{j+1}\| \sigma^2 \sigma_{j,+}^2 \tag{21}$$

with $c \approx 20$ yields impressive results.

It is well known that the scale, i.e. the width, of the detection filter should be determined to give a better trade-off between the detection and the localization efficiency. Small scale leads to higher resolution while large scale leading to lower false edge rate. Canny (1986) pointed out that small operator width should be used whenever they have sufficient SNR. This is natural for that as long as the SNR is high enough to dilute the effects of noise, the filter width should be set as small as possible to improve the localization and reduce the interference of neighboring edges.

3.4. Two dimensions

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

$$P_j^{f,1}(x,y) = W_j^1 f(x,y) \cdot W_{j+1}^1 f(x,y)$$

and

$$P_j^{f,2}(x,y) = W_j^2 f(x,y) \cdot W_{j+1}^2 f(x,y) \tag{22}$$

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 $P_j^{f,1}(x_0, y_0)$ and $P_j^{f,2}(x_0, y_0)$ will be non-negative and the orientation information of the gradient is lost, which should be recovered from $W_j^1 f(x_0, y_0)$ and $W_j^2 f(x_0, y_0)$.

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

$$M_j f(x,y) = \sqrt{P_j^{f,1}(x,y) + P_j^{f,2}(x,y)} \tag{23}$$

$$A_j f(x,y) = \arctan\left(\frac{\operatorname{sgn}(W_j^2 f(x,y)) \cdot \sqrt{P_j^{f,2}(x,y)}}{\operatorname{sgn}(W_j^1 f(x,y)) \cdot \sqrt{P_j^{f,1}(x,y)}} \right) \tag{24}$$

As in the Canny edge detection algorithm, an edge point is asserted wherever $M_j f(x,y)$ 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 Section 3.3, a proper threshold $t_{sc}^i(j)$ that could be applied to $P_j^{f,i}(x,y)$, $i = 1, 2$, is

$$t_{sc}^i(j) = c \cdot \|\psi_j^i\| \cdot \|\psi_{j+1}^i\| \cdot \sigma^2 \cdot (\sigma_{j,+}^i)^2 \tag{25}$$

where c is a constant and

$$\|\psi_j^i\| = \sqrt{\int\int (\psi_j^i(x,y))^2 \, dx \, dy} \tag{26}$$

$$\sigma_{j,+}^i = \frac{1}{2} \sqrt{\int\int (\psi_j^i(x,y)/\|\psi_j^i\| + \psi_j^i(x,y)/\|\psi_j^i\|)^2 \, dx \, dy} \tag{27}$$

c can be chosen around 20. By experimental experience, setting the threshold applied to $M_j f(x,y)$ as

$$t_{sc}(j) = 0.8 \ * \ \sqrt{t_{sc}^1(j) + t_{sc}^2(j)} \tag{28}$$

could achieve satisfying results.

3.5. Performance analysis

In this section, we take two synthetic images as examples to show that the scale multiplication will improve the detection performance (especially on the localization accuracy) and reduce the interference of neighboring edges.

3.5.1. Isolated step edge

First we take the isolated step edge as an example. By DWT, we find edges at two adjacent scales respectively using the Canny edge detection algorithm. Then we detect edges in the product of the two scales with our scheme. The *figure of merit*

F of Pratt (1991) is used to evaluate the performance

$$F = \frac{1}{\max\{N_I, N_A\}} \sum_{k=1}^{N_A} \frac{1}{1 + \alpha d^2(k)} \tag{29}$$

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.

Fig. 5(a) shows a 256×256 isolated step edge corrupted with Gaussian white noise. Fig. 5(c) and (d) are the edge detection results by the small scale 2^3 and large scale 2^4. Fig. 5(b) is the detected edge map by the scale multiplication scheme. Denote by F_P the figure of merit value of Fig. 5(b) and F_1, F_2 those of (c) and (d). These values are shown in Table 1. As expected, F_P is the greatest. F_1 is less than F_2 for that some false edges are caused by noise.

Next we focus on the *localization* accuracy of the three edge images. If the distance $d(k)$ is not greater than 4 pixels, this edge is considered as a true edge. Denote by N the total number of true edges that are detected, we define the *mean square distance* as

$$D = \sqrt{\frac{1}{N} \sum_{i=1}^{N} d^2(i)} \tag{30}$$

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

Denote by D_P the mean square distance value of Fig. 5(b) and D_1, D_2 those of (c) and (d). It can be seen from Table 2 that not only D_P is less than D_2 but also it is less than D_1. This implies that the scale multiplication improves the localization accuracy significantly while keeping high detection efficiency.

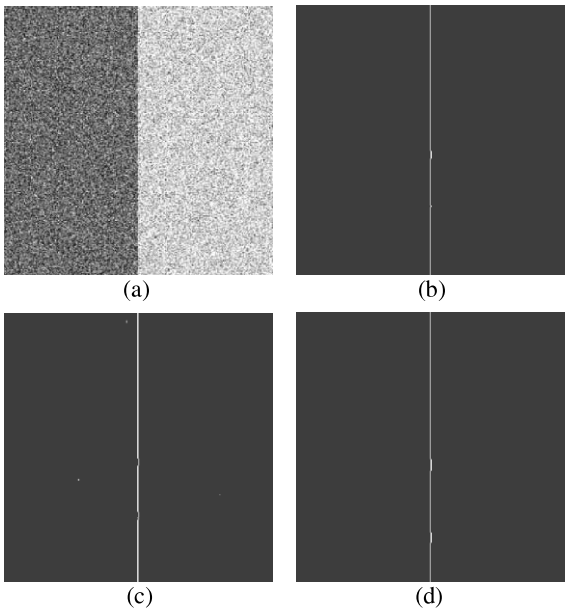


Fig. 5. Noisy step and edge maps: (a) noisy step edge, (b) by scale multiplication, (c) by scale 2^3, (d) by scale 2^4.

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

3.5.2. Neighboring step edges

If two or more edges occur in a neighborhood, they may interference each other with the increasing of the width of the filter. Lu and Jain (1989, 1992) discussed the behavior of the edges in scale space and pointed out that with a large scale the dislocation of an edge will occur if there is another edge in the neighborhood. If we select a small scale parameter, the detection result would noise sensitive. 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. First, in comparison with the small scale, edges are enhanced and noise is diluted. The noise sensitivity will be reduced. Second, in comparison with the large scale, the dislocation of the edges will be

Table 2
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

significantly improved, as illustrated in the following example.

Fig. 6(a) is the 256×256 noisy neighboring step edges. One edge is located at $x = 124$ and jumps from low level to high level. Another is located at $x = 134$ and jumps from high level to low level. (If the two edges are sufficiently close, they can be viewed as a pulse.) Fig. 6(c) and (d) are the edge figures formed by small scale 2^3 and large scale 2^4. Fig. 6(b) is the result by the scale multiplication. The figure of merit value F_P, F_1 and F_2 of Fig. 6(b)–(d) are listed in Table 3. In Fig. 6(d), since the width of detection filter is large, the two edges are with some dislocation and their distance is greater than 10 pixels. Thus although some false edges occur in Fig. 6(c), F_1 is still greater than F_2. Fig. 6(b) possesses both the advantages of Fig. 6(c) and (d) with few false edges and little edge dislocation. This is consistent with that F_P is greater than both F_1 and F_2.

The mean square distances D_P, D_1 and D_2 of Fig. 6(b)–(d) are given in Table 4. D_P is slightly greater D_1 and both of them are much smaller than D_2, implying that much dislocation of the edges occur in Fig. 6(d).

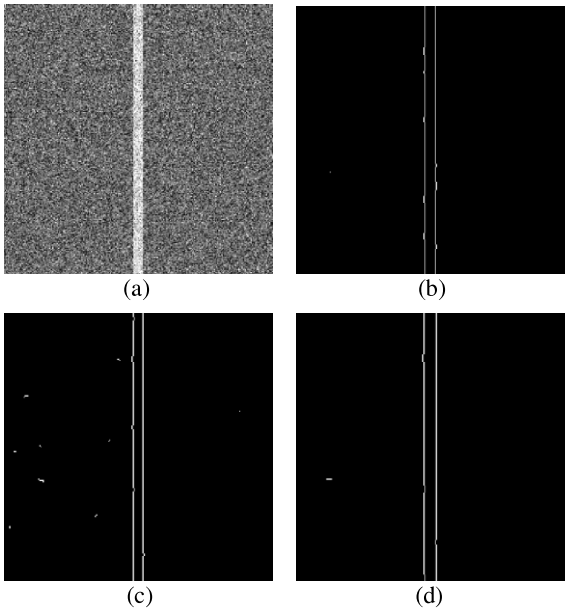


Fig. 6. Noisy two neighboring steps and edge maps: (a) noisy neighboring step edges, (b) by scale multiplication, (c) by scale 2^3, (d) by scale 2^4.

Table 3
The figure of merit values of the two scales and their multiplication for neighboring step edges

F_P	F_1	F_2
0.9899	0.9477	0.8914

Table 4
The mean square distance values of the two scales and their multiplication for neighboring step edges

D_P	D_1	D_2
0.2852	0.1837	1.0049

4. Experiments

Experiments on some benchmark images are used to validate the proposed scheme. The Canny edge detection and LOG algorithms are employed for comparison. There are two parameters in Canny edge detection. One is the high threshold t_h. The other is σ_g, the standard deviation of the Gaussian function that is used to adjust the width of the detection filter. Also two parameters are taken in LOG scheme, the standard deviation σ_g of the Gaussian function and the threshold t to suppress false edges. In the proposed scale multiplication based scheme by DWT, the filter width will be enlarged rapidly with the increasing of the dyadic scale parameter 2^j. Setting the small scale as 2^2 or 2^3 will be sufficient in general. In the following experiments, we fixed the small scale as 2^2 and then the large scale is 2^3.

Fig. 7(a) is a 256×256 noisy *House* image (SNR = 16.52 dB). Fig. 7(b) is the edge image detected by the scale multiplication scheme. Fig. 8(a)–(c) are the edge maps generated by Canny edge detection with $\sigma_g = 1$ and threshold $t_h = 0.25$, 0.32, 0.39 respectively. Fig. 9(a)–(c) show the Canny edge detection results with $\sigma_g = 2$ and $t_h = 0.21$, 0.26, 0.31. The edge maps by LOG operator are illustrated in Fig. 10(a)–(c) with $\sigma_g = 2$, $t = 1.9$, 2.4, 2.9 and Fig. 11(a)–(c) with $\sigma_g = 2.8$, $t = 0.5$, 0.7, 0.9.

(a) (b)

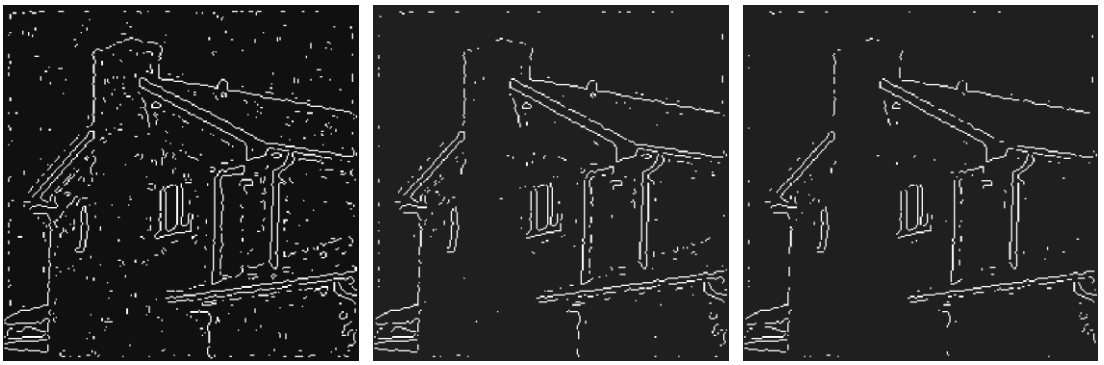
Fig. 7. (a) Noisy house (SNR = 16.52 dB), (b) edge map by scale multiplication scheme.

(a) (b) (c)

Fig. 8. Edge maps by Canny with $\sigma_g = 1$: (a) $t_h = 0.25$, (b) $t_h = 0.32$, (c) $t_h = 0.39$.

(a) (b) (c)

Fig. 9. Edge maps by Canny with $\sigma_g = 2$: (a) $t_h = 0.21$, (b) $t_h = 0.26$, (c) $t_h = 0.31$.



(a) (b) (c)

Fig. 10. Edge maps by LOG with $\sigma_g = 2$: (a) $t = 1.9$, (b) $t = 2.4$, (c) $t_h = 2.9$.

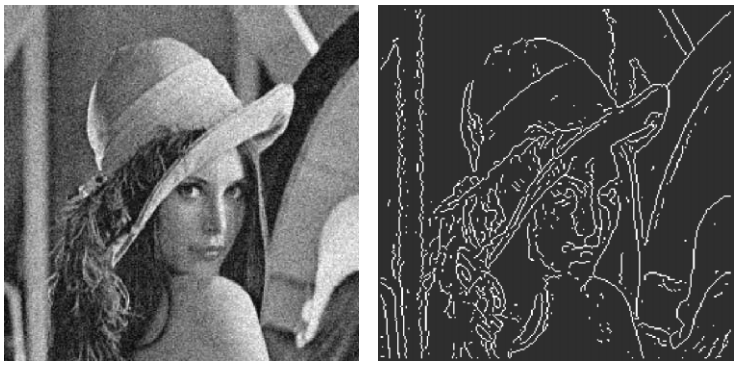
(a) (b) (c)

Fig. 11. Edge maps by LOG with $\sigma_g = 2.8$: (a) $t = 0.5$, (b) $t = 0.7$, (c) $t = 0.9$.

In Canny edge detection and LOG algorithms, with small scale the edge maps are sensitive to noise. When the threshold is increased to suppress noise, some of the true edges disappear. With large scale, the location accuracy is decreased. It can be seen from Fig. 7(b) that with our 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.

Fig. 12(a) is the 256×256 noisy *Lenna* (SNR = 16.34 dB). Fig. 12(b) shows the edge map by the scale multiplication scheme. Fig. 13(a)–(c) show the edge maps generated by Canny edge detection with $\sigma_g = 1.0$, 1.4, 1.8 respectively. Fig. 14(a)–(c) are the results by LOG with $\sigma_g = 1.6$, 2.0, 2.4. From Figs. 13 and 14 it can be seen that finer edges are detected as well as many false edges at fine scales and noise are suppressed but some edges are also missed or dislocated (such as the face and hair of Lenna) at large scales. In Fig. 12(b) through the scale multiplication more fine edges are asserted while false edges being removed really well.

Figs. 15–20 are the experiments on *Peppers* and *Cameraman*. The presented scheme also achieves satisfying performances.



(a) (b)

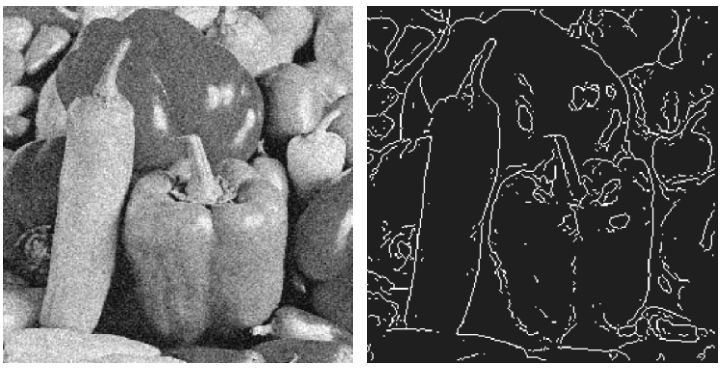
Fig. 12. (a) Noisy Lenna (SNR = 16.34 dB), (b) edge map by scale multiplication scheme.

(a) (b) (c)

Fig. 13. Edge maps by Canny: (a) $\sigma_g = 1$, (b) $\sigma_g = 1.4$, (c) $\sigma_g = 1.8$.

(a) (b) (c)

Fig. 14. Edge maps by LOG: (a) $\sigma_g = 1.6$, (b) $\sigma_g = 2.0$, (c) $\sigma_g = 2.4$.

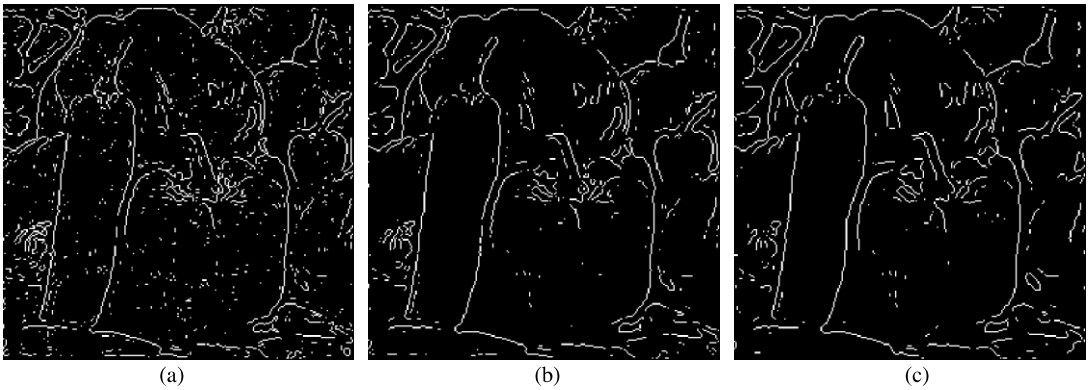


(a) (b)

Fig. 15. (a) Noisy Peppers (SNR = 16.36 dB), (b) edge map by scale multiplication scheme.

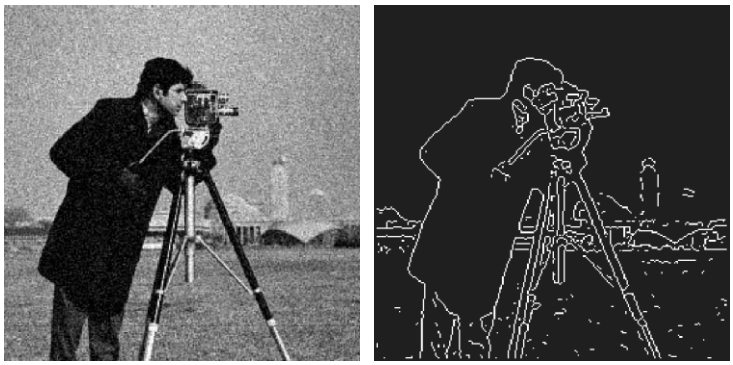
(a) (b) (c)

Fig. 16. Edge maps by Canny: (a) $\sigma_g = 1$, (b) $\sigma_g = 1.4$, (c) $\sigma_g = 1.8$.



(a) (b) (c)

Fig. 17. Edge maps by LOG: (a) $\sigma_g = 1.6$, (b) $\sigma_g = 2.0$, (c) $\sigma_g = 2.4$.



(a) (b)

Fig. 18. (a) Noisy Cameraman (SNR = 16.52 dB), (b) edge map by scale multiplication scheme.

(a) (b) (c)

Fig. 19. Edge maps by Canny: (a) $\sigma_g = 1$, (b) $\sigma_g = 1.4$, (c) $\sigma_g = 1.8$.

(a) (b) (c)

Fig. 20. Edge maps by LOG: (a) $\sigma_g = 1.6$, (b) $\sigma_g = 2.0$, (c) $\sigma_g = 2.4$.

5. Conclusion

In this paper, we proposed a DWT based edge detection scheme by scale multiplication. Since there exist high spatial similarities in wavelet subbands, we defined a scale product function as the multiplication of two adjacent scales of wavelet coefficients to amplify edge structures while diluting noise. Then the edges are determined as the local maxima of the product to avoid the ill-posed edge synthesis process in most multiscale detection schemes. Noting that edges and noise can be better distinguished in the scale product, we only use a properly determined threshold to suppress the noise maxima instead of the double thresholding utilized by Canny. Through synthetic images, it is shown that the scale multiplication achieves really 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. At last the experiments on natural images are compared with the LOG and Canny edge detection algorithms.

References

Aydn, T. et al., 1996. Multidirectional and multiscale edge detection via M-band wavelet transform. IEEE Trans. Image Processing 5, 1370–1377.

Bergholm, F., 1987. Edge focusing. IEEE Trans. PAMI PAMI-9, 726–741.

Canny, J., 1986. A computational approach to edge detection. IEEE Trans. PAMI PAMI-8, 679–698.

Daubechies, I., 1992. Ten Lectures on Wavelets. SIAM, Philadelphia, PA.

Jeong, H., Kim, C.I., 1992. Adaptive determination of filter scales for edge detection. IEEE Trans. PAMI 14, 579–585.

Lu, Y., Jain, R.C., 1989. Behavior of edges in scale space. IEEE Trans. PAMI 11, 337–356.

Lu, Y., Jain, R.C., 1992. Reasoning about edges in scale space. IEEE Trans. PAMI 14, 450–468.

Mallat, S., Hwang, W.L., 1992. Singularity detection and processing with wavelets. IEEE Trans. Information Theory 32, 617–643.

Mallat, S., Zhong, S., 1992. Characterization of signals from multiscale edges. IEEE Trans. PAMI 14, 710–732.

Marr, D., Hildreth, E., 1980. Theory of edge detection. Proc. Royal Soc. London 207, 187–217.

Park, D.J., Nam, K.N., Park, R.H., 1995. Multiresolution edge detection techniques. Pattern Recogn. 28, 211–229.

Petrou, M., Kittler, J., 1991. Optimal edge detectors for ramp edges. IEEE Trans. PAMI 13, 483–491.

Pratt, W.K., 1991. Digital Image Processing, second ed. John Wiley & Sons, New York.

Sadler, B.M., Swami, A., 1999. Analysis of multiscale products for step detection and estimation. IEEE Trans. Information Theory 45, 1043–1051.

Sarkar, S., Boyer, K.L., 1991. On optimal infinite impulse response edge detection filters. IEEE Trans. PAMI 13, 1154–1171.

Shen, J., Castan, S., 1992. An optimal linear operator for step edge detection. CGVIP: Graphical Models Image Process. 54, 112–133.

Xu, Y. et al., 1994. Wavelet transform domain filters: a spatially selective noise filtration technique. IEEE Trans. Image Process. 3, 747–758.

Ziou, D., Tabbone, S., 1993. A multi-scale edge detector. Pattern Recogn. 26, 1305–1314.