A Wavelet-based Edge Detection Method by Scale Multiplication

Lei Zhang and Paul Bao
Dept. of Computing, The Hong Kong Polytechnic University
Hung Hum, Kowloon, Hong Kong
E-mail: csbao@comp.polyu.edu.hk

Abstract

A wavelet-based multiscale edge detection scheme is presented in this paper. By multiplying the wavelet coefficients at two adjacent scales to magnify significant structures and suppress noise, we determined edges as the local maxima directly in the scale product after an efficient thresholding instead of first forming the edge maps at several scales and then synthesizing them together, which was employed in many multiscale techniques. It is shown that the scale multiplication achieves better results than either of the two scales, especially on the localization performance. Experiments on natural images are compared with the Laplacian of Gaussian (LOG) and Canny edge detection algorithms.

1. Introduction

Edge detection is an essential process in image analysis and many techniques have been proposed. Canny [3] 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*. Besides the shape of the detector, another important problem is to set a proper detection scale. As suggested by Marr and Hildreth [2], multiple scales should be employed to describe the variety of the edge structures.

Canny [3] used a fine-to-coarse feature synthesis strategy to mingle the multiscale edge information based on a set of predefined rules. Bergholm [4] combined the multiscale edges in a coarse-to-fine tracking manner. The RRES scheme of Lu and Jain [5] 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, Jeong and Kim [9] 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 [6] ran a subpixel *Laplacian* operator at two scales and then recovered the edges with four-step edge models. Park *et al.* [7] 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.

Wavelet analysis is naturally a multiscale and multiresolution analysis. Mallat and Zhong [11] constructed a dyadic wavelet and the corresponding dyadic wavelet transform (DWT) is equivalent to the *Canny edge detection*. In wavelet domain, the edge structures present observably at each subband while noise decreases rapidly along the scales. With this observation, Xu *et al.* [12] proposed a spatially selected filtering technique by multiplying the adjacent DWT scales to enhance the significant structures. Sadler and Swami [13] analyzed the multiscale product of DWT and applied it to the step detection and estimation.

In this paper, we presented a scale-multiplication-based edge detection scheme with the DWT. Two adjacent 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. 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 with either of the two scales only.

2. The dynamic wavelet transform

A function $\psi(x)$ is called a wavelet if its average is equal to 0. Let $\zeta_j(x) = 2^{-j}\zeta(2^{-j}x)$ be the dilation of function $\zeta(x)$ by dyadic scale 2^j. The DWT of $f(x)$ at scale 2^j and position x is

$$W_j f(x) = f * \psi_j(x) \qquad (1)$$

Where $*$ denotes convolution operation. For the details please refer to Daubechies [10] and Mallat [11]. Suppose $\theta(x)$ is a differentiable smooth function whose integral is 1 and converges to 0 at infinity. Let $\psi(x) = d\theta(x)/dx$, then

1051-4651/02 $17.00 © 2002 IEEE

$$W_j f(x) = f * \psi_j(x) = 2^j \frac{d}{dx}(f * \theta_j)(x) \qquad (2)$$

$W_j f(x)$ is proportional to the first derivative of $f(x)$ smoothed by $\theta_j(x)$. The wavelet used in this paper is the Mallat wavelet [11], whose $\theta(x)$ is a cubic spline that approximates the *Gaussian* function closely. And then the DWT is equivalent to the *Canny edge detection*.

In 2-D case, two wavelets should be utilized:

$$\psi^1(x,y) = \frac{\partial \theta(x,y)}{\partial x}, \quad \psi^2(x,y) = \frac{\partial \theta(x,y)}{\partial y} \qquad (3)$$

Denote $\zeta_j(x,y) = 2^{-2j}\zeta(2^{-j}x, 2^{-j}y)$ the dilation of $\zeta(x,y)$ by 2^j, the wavelet transform of $f(x,y)$ at scale 2^j and position (x,y) has two components

$$W_j^1 f(x,y) = f * \psi_j^1(x,y), \quad W_j^2 f(x,y) = f * \psi_j^2(x,y) \qquad (4)$$

3. The algorithm

3.1. Scale multiplication

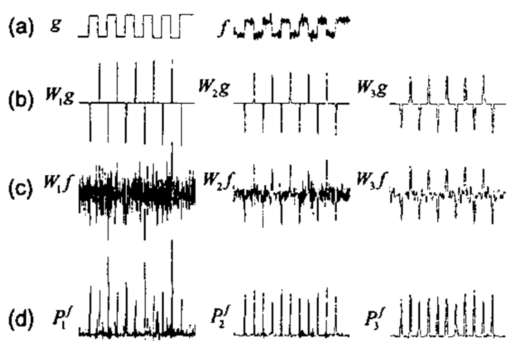

Fig. 1. (a) Blocks g and its noisy version f. (b) The DWT of g. (c) The DWT of noisy f. (d) The product function P_j^f with $j = 1 \sim 3$.

For signal structures, the DWT amplitudes would increase or keep invariant when increasing the scale 2^j. On the contrary, those of noise will decay rapidly along the scales. Directly multiplying the DWT at two adjacent scales will amplify the edge structures and dilute the noise. The scale product function of $f(x)$ is defined as the correlation of the DWT of $f(x)$ at two adjacent scales

$$P_j^f(x) = W_j f(x) \cdot W_{j+1} f(x) \qquad (5)$$

In Fig. 1 (a), a block signal g and its noisy version f are illustrated. Their DWT are given in Fig. 1 (b) and (c). At the finest scale the wavelet coefficients are almost dominated by noise. It can also be seen that at small scales the step edges are better localized but some noise may be falsely considered as edges. At the large scales, edges can be detected more correctly but with the decreasing of the accuracy of the edge location. In Fig. 1 (d), the product P_j^f, $j = 1 \sim 3$, are illustrated. Apparently the step edges are more observable in P_j^f than in $W_j f$.

3.2. Thresholding

We assert the edges as the local maxima in P_j^f. A significant edge at x_0 will occur on both the adjacent scales with the same sign, so that $P_j^f(x_0)$ should be non-negative. P_j^f should be thresholded to filter noise.

Let noise $\varepsilon \sim N(0,\sigma^2)$ and $W_j\varepsilon(x) = \varepsilon * \psi_j(x)$. For expression convenience, we denote $X_j(x) = W_j\varepsilon(x)$ and $Y_j(x) = P_j^\varepsilon(x) = X_j(x) \cdot X_{j+1}(x)$. $X_j \sim N(0, \sigma_j^2)$, where $\sigma_j = \|\psi_j\|\sigma$ and $\|\psi_j\| = \sqrt{\int \psi_j^2(x)dx}$.

Denote $t_{sc}(j)$ the threshold applied to $Y_j(x)$. It is expected that $t_{sc}(j)$ could suppress most of the noise, i.e., $P(y_j < t_{sc}(j)) \to 1$. Normalize X_j and X_{j+1} as:

$$\bar{X}_j = X_j / \sigma_j, \quad \bar{X}_{j+1} = X_{j+1} / \sigma_{j+1} \qquad (6)$$

Define $\bar{Y}_j(x) = \bar{X}_j(x) \cdot \bar{X}_{j+1}(x)$ and then

$$Y_j(x) = \|\psi_j\| \cdot \|\psi_{j+1}\| \sigma^2 \cdot \bar{Y}_j(x) \qquad (7)$$

Let

$$\bar{Y}_{j,+}(x) = (\bar{X}_j(x) + \bar{X}_{j+1}(x))/2 \text{ and}$$
$$\bar{Y}_{j,-}(x) = (\bar{X}_j(x) - \bar{X}_{j+1}(x))/2 \qquad (8)$$

Thus

$$\bar{Y}_j(x) = \bar{Y}_{j,+}^2(x) - \bar{Y}_{j,-}^2(x) \qquad (9)$$

and $\bar{Y}_{j,+} \sim N(0, \sigma_{j,+}^2)$ and $\bar{Y}_{j,-} \sim N(0, \sigma_{j,-}^2)$ with

$$\sigma_{j,+} = \frac{1}{2}\sqrt{\int (\psi_j(x)/\|\psi_j\| + \psi_{j+1}(x)/\|\psi_{j+1}\|)^2 dx},$$
$$\sigma_{j,-} = \frac{1}{2}\sqrt{\int (\psi_j(x)/\|\psi_j\| - \psi_{j+1}(x)/\|\psi_{j+1}\|)^2 dx} \qquad (10)$$

Since there is a strong correlation between $\psi_j(x)$ and $\psi_{j+1}(x)$, so $\sigma_{j,+}^2$ is much larger than $\sigma_{j,-}^2$. Let $\bar{t}_{sc}(j) = t_{sc}(j)/(\|\psi_j\| \cdot \|\psi_{j+1}\|\sigma^2)$, we have

$$P(y_j < t_{sc}(j)) \geq P(\bar{y}_{j,+}^2 < \bar{t}_{sc}(j)) = P(|\bar{y}_{j,+}| < \sqrt{\bar{t}_{sc}(j)})$$

Setting $\sqrt{\bar{t}_{sc}(j)} \geq 4\sigma_{j,+}$ will lead to

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

and then

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

In real applications, the input is $f = g + \varepsilon$ where g is the original image and then $W_j f = W_j g + W_j \varepsilon$. 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 $t_{sc}(j)=c\cdot\|\psi_j\|\cdot\|\psi_{j+1}\|\sigma^2\sigma_{j,+}^2$ with $c\approx 20$ yields impressive results.

3.3. Two dimensions

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

$$P_j^{f,1}(x,y)=W_j^1 f(x,y)\cdot W_{j+1}^1 f(x,y)$$
$$P_j^{f,2}(x,y)=W_j^2 f(x,y)\cdot W_{j+1}^2 f(x,y) \qquad (11)$$

Similar to 1-D case, for an edge point (x_0,y_0), both $P_j^{f,1}(x_0,y_0)$ and $P_j^{f,2}(x_0,y_0)$ should be non-negative but the orientation information of the gradient is lost, which could be recovered from $W_j^1 f(x_0,y_0)$ and $W_j^2 f(x_0,y_0)$. The modulus and angle of point (x,y) are defined as

$$M_j f(x,y)=\sqrt{P_j^{f,1}(x,y)+P_j^{f,2}(x,y)} \qquad (12)$$

$$A_j f(x,y)=\arctan\left(\frac{\operatorname{sgn}(W_j^2 f(x,y))\cdot\sqrt{P_j^{f,2}(x,y)}}{\operatorname{sgn}(W_j^1 f(x,y))\cdot\sqrt{P_j^{f,1}(x,y)}}\right) \qquad (13)$$

An edge point is asserted wherever $M_j f(x,y)$ has a local maximum in the direction of the gradient given by $A_j f(x,y)$. Similar to section 3.2, let $t_{sc}^i(j)$, $i=1,2$, is

$$t_{sc}^i(j)=c\cdot\|\psi_j^i\|\cdot\|\psi_{j+1}^i\|\cdot\sigma^2\cdot(\sigma_{j,+}^i)^2 \qquad (14)$$

Where c is a constant and

$$\|\psi_j^i\|=\sqrt{\iint(\psi_j^i(x,y))^2\,dxdy} \qquad (15)$$

$$\sigma_{j,+}^i=\frac{1}{2}\sqrt{\iint\left(\psi_j^i(x,y)/\|\psi_j^i\|+\psi_j^i(x,y)/\|\psi_j^i\|\right)^2 dxdy} \qquad (16)$$

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

$$t_{sc}(j)=0.8*\sqrt{t_{sc}^1(j)+t_{sc}^2(j)} \qquad (17)$$

could achieve satisfying results.

4. Performance analysis

Fig. 2 (a) is a 256×256 isolated noisy step edge. We find edges first at two adjacent scales by *Canny edge detection* and then in the product of the two scales with our scheme. The *figure of merit* F of Pratt [1] is used to evaluate the performance

$$F=\frac{1}{\max\{N_I,N_A\}}\sum_{k=1}^{N_A}\frac{1}{1+\alpha d^2(k)} \qquad (18)$$

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.

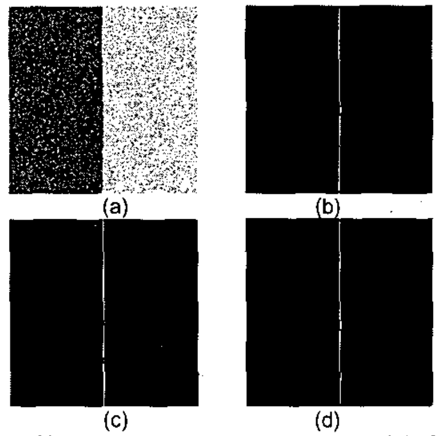

Fig. 2. Noisy step and its edge maps. (a) Noisy step edge. (b) By scale multiplication. (c) By scale 2^3. (d) By scale 2^4.

Fig. 2 (c) and (d) are the edge maps by small scale 2^3 and large scale 2^4. Fig. 2 (b) is the edge map by our scheme. Denote by F_P the *figure of merit* value of Fig. 2 (b) and F_1, F_2 those of (c) and (d). These values are shown in Table I. As expected, F_P is the greatest, which means the best performance. F_1 is less than F_2 because some false edges are caused by noise.

Next we focus on the *localization* accuracy of the three edge maps. 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

$$D=\sqrt{\frac{1}{N}\sum_{i=1}^{N}d^2(i)} \qquad (19)$$

Table I. The *figure of merit* values of the two scales and their multiplication.

F_P	F_1	F_2
0.9929	0.9496	0.9877

Table II. The *mean square distance* values of the two scales and their multiplication.

D_P	D_1	D_2
0.1782	0.2271	0.2887

The smaller the D, the better the *localization* accuracy will be achieved. Denote by D_P the *mean* *square distance* of Fig. 2 (b) and D_1, D_2 those of (c) and (d). It can be seen from Table II that not only D_P is less than D_2 but also it is less than D_1. Scale multiplication improves the *localization* accuracy significantly while keeping high *detection* efficiency.

(a) (b)

Fig. 4. (a) Noisy Lenna (SNR=16.34dB). (b) Edge map by scale multiplication scheme.

(a) (b)

Fig. 5. By *Canny*. (a) $\sigma_g = 0.8$. **(b)** $\sigma_g = 1.6$.

(a) (b)

Fig. 6. By *LOG*. (a) $\sigma_g = 1.6$. **(b)** $\sigma_g = 2.4$.

Next the Lenna image is used to validate the proposed scheme. The *Canny edge detection* and *LOG* algorithms are employed for comparison. In the two methods, the standard deviation of the Gaussian function, σ_g, is used to adjust the width of the detection filter. In the proposed scale multiplication based scheme by DWT, we toke the small scale as 2^2 and then the large scale is 2^3.

Fig. 4 (a) is the 256×256 noisy *Lenna* (SNR=16.34dB). Fig. 4 (b) shows the edge map by the scale multiplication scheme. Fig. 5 (a)~(b) show the edge maps generated by *Canny edge detection* with $\sigma_g = 0.8, 1.6$ respectively. Fig. 6 (a)~(b) are the results by *LOG* with $\sigma_g = 1.6, 2.4$. From Fig. 5 and Fig. 6 it can be seen that when scale is small, finer edges are detected as well as many false edges. If the scale is enlarged, noise are suppressed but some edges are also missed or dislocated (such as the face and hair of *Lenna*). In Fig. 4 (b) much better result is achieved. On the one hand, a rich class of edges, even some fine ones missed by the other two schemes at fine scale, are detected with better localization through the scale multiplication. On the other hand, the edge map is "clear", which means that false edges are suppressed well. The experiments on other benchmark images export the similar results.

References

[1] William K. Pratt, *Digital Image Processing*, John Wiley & Sons, 2nd edition, 1991.

[2] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Royal Soc*, London, vol. 207, pp. 187-217, 1980.

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

[4] Fredrik Bergholm, "Edge focusing, " *IEEE Trans. PAMI*, vol. PAMI-9, pp. 726-741, 1987.

[5] Yi Lu and R. C. Jain, "Reasoning about edges in scale space, " *IEEE Trans. PAMI*, vol. 14, pp. 450-468, 1992.

[6] D. Ziou and S. Tabbone, "A multi-scale edge detector, " *Pattern Recognition*, vol. 26, pp. 1305-1314, 1993..

[7] D. J. Park, K. N. Nam and R. H. Park, "Multiresolution edge detection techniques, " *Pattern Recognition*, vol. 28, pp. 211-229, 1995.

[8] B. Chanda, M. K. Kund and Y. V. Padmaja, "A multiscale morphologic edge detector, " *Pattern Recognition*, vol. 31, pp. 1469-1478, 1998.

[9] Hong Jeong and C. I. Kim, "Adaptive determination of filter scales for edge detection, " *IEEE Trans. PAMI*, vol. 14, pp. 579-585, 1992.

[10] I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia, PA: SIAM, 1992.

[11] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges, " *IEEE Trans. PAMI*, vol. 14, pp. 710-732, 1992.

[12] Y. Xu *et al*, "Wavelet transform domain filters: a spatially selective noise filtration technique, " *IEEE Trans. Image Processing*, vol. 3, pp. 747-758,1994.

[13] Brain M. Sadler and A. Swami, "Analysis of multiscale products for step detection and estimation, " *IEEE Trans. Information Theory*, vol. 45, pp. 1043-1051, 1999.