Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, Beijing, China, 2-4 Nov. 2007

ADAPTIVE THRESHOLD EDGE DETECTION WITH NOISE IMMUNITY BY MULTI-SCALE ANALYSIS

SI-CONG YUE, RONG-CHUN ZHAO, JIANG-BIN ZHENG

School of Computer, Northwestern Polytechnical University, Xi'an 710072, China
E-MAIL:typhoonadam@gmail.com

Abstract

An adaptive threshold edge detection algorithm based on dyadic wavelet transform is presented in this paper. At first a multi-scale edge response function (MERF) is defined as the multiple scales point-wise products of the dyadic wavelet transform to enhance significant image structures and suppress noise. Thereafter, an adaptive threshold is calculated and imposed on the MERF to identify edges as the local maxima of the MERF gradient map without synthesizing the edge maps at several scales together, which was employed in many multi-scale techniques. Experiments on synthetic benchmark and natural images showed that the proposed adaptive threshold multi-scale edge detection algorithm achieves better detection results than that for a single scale, especially on the localization performance; and edge and noise can be better distinguished by MERF comparing with the Mallat wavelet-based multi-scale algorithm and Canny edge detector.

Keywords: Edge detection; wavelet transform; adaptive threshold; MERF

1. Introduction

Edge detection plays an important role in image analysis and computer vision for edges characterize the object boundary and carry important information of an image. Edge detection concerns the localization of significant variations of the grey level image and the identification of the boundaries of homogeneous regions in an image based on properties such as intensity. This information is very useful for applications in 3D reconstruction, recognition, image enhancement, registration, compression, and so on.

In early years, many edge detection algorithms had been developed such as Roberts, Sobel, Prewitt and Laplacian [1], in which image derivatives are computed. However differentiation of an image is sensitive to noise. Recently some other novel edge detection techniques have also been proposed [2-8]. The common approach to this problem is based on smoothed gradient estimation, also called filtered derivative methods. Canny [2] presented the well-known three criteria of edge detector: good detection, good localization and low spurious response and he showed that the optimal detector for a step edge should be the first derivative of Gaussian, which estimates the gradient after smoothing with a Gaussian function.

However there are undesirable effects associated with smoothing, i.e., loss of information and displacement of prominent structures in the image plane. Attempting to achieve simultaneous good detection and good localization results in a tradeoff between the level of smoothing and the variance of the estimated step location. Consequently, besides the shape of the detector, another important problem is to set a proper scale of detection filter. Fine-scaled filters are sensitive to edge signal but also prone to noise, whereas coarse-scaled filters are robust to noise but could filter out fine details. So it is of interest to exploit multiple levels of smoothing simultaneously to describe the variety of the edge structure.

It is well known that wavelet transform is naturally a multi-scale analysis and it can provide a multi-resolution representation for an image with flexible localization in both time and frequency domains[5,6]. In literature [5], Mallat illustrated mathematically that signals and noise have different singularities in wavelet domain and edge structures present observable magnitudes along the scale, while noise decreases rapidly. In this regard, Mallat and Zhong [6] developed a discrete wavelet transform based on smoothed gradient estimation, where the level of smoothing varies with dyadic scale. Thereafter the applications of wavelet transform in image processing have received significant attention and some very efficient wavelet-based multi-scale edge detection algorithms had been proposed [9-12].

In this paper we presented a simple but efficient multi-scale algorithm based on forming multiple scales point-wise products of smoothed gradient estimators (MERF), which is intended to enhance multi-scale peaks due to edges, while suppressing noise. This idea was first put forward by Rosenfeld [13] before the advent of wavelets and it is shown that multi-scale products can improve the edge localization [9]. Unlike many multi-scale edge detectors, where the edge maps were formed at several scales and then synthesized together, our scheme determined edges as the local maxima in MERF after an adaptive

multi-scale thresholding scheme, which merged the merits of wavelet interscale dependencies and the thresholding technique. Through MERF and the adaptive threshold, feature-level synthesis of multiple scales will be achieved and an integrated edge map will be formed efficiently while avoiding the ill-posed edge synthesis process.

The remainder of this paper is organized as follows. In Section 2, we discuss the multi-scale analysis and MERF, achieving an effective multi-scale wavelet-based feature synthesis. Section 3 introduces the adaptive thresholding scheme. The experimental results and comparison with other edge detection algorithms are given in Section 4, and, finally, conclusions are drawn in Section 5.

2. Multi-scale Analysis and MERF

2.1. Dyadic Wavelet Transform as a Multi-scale Edge Detector

A wavelet transform represents a signal as a linear combination of elementary atoms that appear at different resolutions. It is computed by convoluting the input signal with dilated wavelet filters recursively.

We denote by $\xi_s(x)$ the dilation of a function $\xi(x)$ by a scale factor s

$$\xi_s(x) = \frac{1}{s}\xi(\frac{x}{s}) \tag{1}$$

Suppose that $\phi(x)$ is a differentiable smooth function whose integral is equal to 1 and that it converges to 0 at infinity. Let's define $\psi(x)$ as the first-order derivative of $\phi(x)$. The dyadic wavelet transform of signal $f(x)$ at scale 2^j and position x is defined as

$$\begin{aligned} W_j f(x) &= f * \psi_j(x) \\ &= f * (2^j \frac{d\phi_j}{dx})(x) = 2^j \frac{d}{dx}(f * \phi_j)(x) \end{aligned} \tag{2}$$

It can be seen that the dyadic wavelet transform $W_j f(x)$ is the first derivative of $f(x)$ smoothed by $\phi_j(x)$. In particular, when $\phi(x)$ is a Gaussian function, the local extrema determination in $W_j f(x)$ is equivalent to the well-known Canny edge detector [2].

The dyadic wavelet transform can also be extended to two-dimensional (2-D) images. Suppose $\phi(x,y)$ is a 2-D differentiable smooth function whose integral is equal to 1 and converges to 0 at infinity. For example $\phi(x,y)$ could be the tensor product of one-dimensional (1-D) smooth functions: $\phi(x,y) = \phi(x)\phi(y)$. We define the two wavelets $\psi^x(x,y)$ and $\psi^y(x,y)$ at horizontal and vertical directions as

$$\psi^x(x,y) = \frac{\partial\phi(x,y)}{\partial x}, \quad \psi^y(x,y) = \frac{\partial\phi(x,y)}{\partial y} \tag{3}$$

The dilation of any 2-D function $\xi(x,y)$ by scale s can be, therefore, denoted by

$$\xi_s(x,y) = \frac{1}{s^2}\xi(\frac{x}{s},\frac{y}{s}) \tag{4}$$

So the dyadic wavelet transform of 2-D function $f(x,y)$ at scale 2^j and position (x,y) has two components. Similarly to (2) these two components can be written as

$$\begin{pmatrix} W_j^x f(x,y) \\ W_j^y f(x,y) \end{pmatrix} = 2^j \begin{pmatrix} \frac{\partial}{\partial x}(f * \phi_j)(x,y) \\ \frac{\partial}{\partial y}(f * \phi_j)(x,y) \end{pmatrix} = 2^j \vec{\nabla}(f * \phi_j)(x,y) \tag{5}$$

In the case when $\phi(x,y)$ is a Gaussian function, detecting the local extrema form $\begin{pmatrix} W_j^x f(x,y) \\ W_j^y f(x,y) \end{pmatrix}$ is equivalent to the Canny edge detection.

The wavelet used in this paper is the Mallat wavelet constructed by Mallat and Zhong [6]. The associated smooth function $\phi(x)$ is a cubic spline and the wavelet $\psi(x)$ is a quadratic spline. Details about derivation of the Mallat wavelet can be found in [6].

2.2. Multi-scale Edge Response Function

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 $0 \le \alpha \le 1$ if and only if there exists a constant K_1 such that in the neighborhood of x_0:

$$|f(x) - f(x_0)| \le K_1 |x - x_0|^\alpha \tag{6}$$

The superior bound of all α satisfying Equation (6) is called as Lipschitz regularity. And the relations between α and wavelet amplitude is described by

$$|W_j f(x)| \le K(2^j)^\alpha \tag{7}$$

Using the mathematical concept of the Lipschitz regularity, Mallat et al. [5], [6] analyzed the evolution of singularities and noise across wavelet scales. Signals and noise behave very differently in the wavelet transform domain. Singularities are more regular than noise and have higher Lipschitz regularities. For example, the Lipschitz regularity of a step edge is 0. If a structure is smoother than the step, it will have positive Lipschitz regularity. Otherwise the definition can be extended to negative value for singularities worst than discontinuities, such as white noise. White noise is almost singular everywhere and has a uniform Lipschitz regularity that is equal to -1/2.

The above equation implies that the wavelet transform magnitudes increase for positive α with increasing scales. Contrarily, wavelet transform magnitudes decrease for negative Lipschitz regularities with increasing scales. In Fig. 1, the DWT at first three scales of a noisy test signal $f = g + \varepsilon$ are illustrated. Notice that the signal singularities evolve across scales with observable peaks while noise decays rapidly along scales. When ε 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 scale 2^j [5].

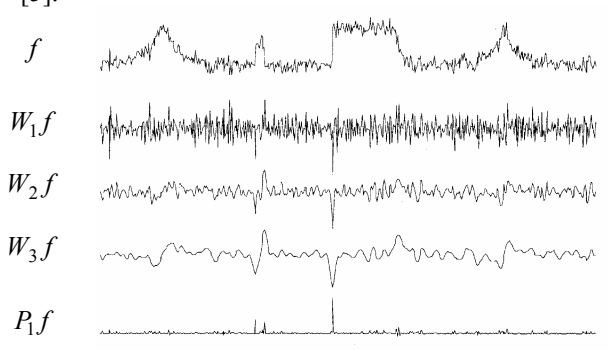


Figure 1. The DWT at first three scales and MERF of a noisy test signal

With the observation of Fig. 1, we can imagine that the product of the DWT at adjacent scales would amplify edge structures and dilute noise. So the multi-scale edge response function (MERF) is defined as:

$$P_j f(x) = W_j f(x) W_{j+1} f(x) \tag{8}$$

Similarly for 2-D images, in x and y directions the MERF have two components

$$\begin{aligned} P_j^x f(x,y) &= W_j^x f(x,y) W_j^x f(x,y) \\ P_j^y f(x,y) &= W_j^y f(x,y) W_j^y f(x,y) \end{aligned} \tag{9}$$

For an edge point (x,y), $W_j^d f(x,y)$ and $W_{j+1}^d f(x,y)$, $d = x, y$ should have the same sign. So both $P_j^x f(x,y)$ and $P_j^y f(x,y)$ will be nonnegative. Setting the points with $P_j^d f(x,y) < 0$, $d = x, y$ to 0, the modulus and angle of point (x,y) are defined as

$$M_j f(x,y) = \sqrt{P_j^x f(x,y) + P_j^y f(x,y)} \tag{10}$$

$$A_j f(x,y) = \arctan\left(\frac{\operatorname{sgn}(W_j^y f(x,y)) \sqrt{\left| P_j^y f(x,y) \right|}}{\operatorname{sgn}(W_j^x f(x,y)) \sqrt{\left| P_j^x f(x,y) \right|}} \right) \tag{11}$$

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 $A_j f(x,y)$.

3. Adaptive Thresholding for Gradient Module

3.1. The Thresholding Scheme

Naturally, in gradient module local maximum estimation, the gradient image should be thresholded to eliminate false edges produced by noise. So the determination of the threshold value is extremely critical to the threshold-based algorithms. We denote by $f = g + \varepsilon$ the measurements of image g corrupted by Gaussian white noise $\varepsilon \sim N(0, \sigma^2)$. From Fig. 1, it can be noticed that, at finer scales, if the threshold t applied to $W_j f$ is relatively high, some edge structures may be suppressed as noise. Otherwise, if it is relatively low, many noisy pixels would be undesirably preserved. However, in the MERF $P_j f$, it can be seen that the significant structures are strengthened while the noise is weakened. MERF $P_j f$ results in a more effective discrimination between edges and noise than $W_j f$. With such observations we propose an adaptive thresholding scheme for MERF gradient module, to merge the merits of the thresholding technique and wavelet interscale dependencies, unlike other wavelet-base multi-scale edge detection where the thresholding schemes do not exploit the dependencies that exist between adjacent wavelet scales. The algorithm is summarized as follows.

1) Compute the MERF $P_j f$ of input image f up to J scales.

2) Estimate the standard deviation of noise σ [14] and Calculate the thresholds $t(j)$.

3) Threshold the gradient modulus $M_j f(x,y)$ by $t(j)$ with local maximum in the direction of $A_j f(x,y)$.

3.2. Adaptive Threshold

Suppose that the input image is Gaussian white noise ε. For the convenience of expression, we denote the DWT of ε by

$$U_j^d(x,y) = W_j^d \varepsilon(x,y) = \varepsilon * \psi_j^d(x,y), \quad d = x, y \tag{12}$$

U_j^d is a Gaussian colored noise and its standard deviation is

$$\sigma_j = \left\| \psi_j \right\| \sigma \tag{13}$$

Where $\left\| \psi_j \right\| = \sqrt{\iint \psi_j^2(x,y) dx dy}$.

U_j^d and U_{j+1}^d are jointly Gaussian distributed with probability density function (pdf) [15]

$$p(u_j,u_{j+1})=\frac{1}{2\pi\sigma_j\sigma_{j+1}\sqrt{1-\rho_{j,j+1}^2}}\cdot\exp\left\{-1/2(1-\rho_{j,j+1}^2)\left[u_j^2/\sigma_j^2-(2\rho_{j,j+1}u_ju_{j+1}/\sigma_j\sigma_{j+1})+u_{j+1}^2/\sigma_{j+1}^2\right]\right\} \tag{14}$$

Where the correlation coefficient $\rho_{j,j+1}$ of U_j^d and U_{j+1}^d is

$$\rho_{j,j+1}=\frac{\iint\psi_j(x,y)\psi_{j+1}(x,y)dxdy}{\sqrt{\iint\psi_j^2(x,y)dxdy\iint\psi_{j+1}^2(x,y)dxdy}} \tag{15}$$

We denote MERF by $V_j^d=U_j^dU_{j+1}^d$. Then the standard deviation of V_j^d is [15]:

$$\begin{aligned}\kappa_j&=\sqrt{E[v_j^2]}=\sqrt{E[u_j^2u_{j+1}^2]}\\&=\sqrt{1+2\rho_{j,j+1}^2}\,\sigma_j\sigma_{j+1}\end{aligned} \tag{16}$$

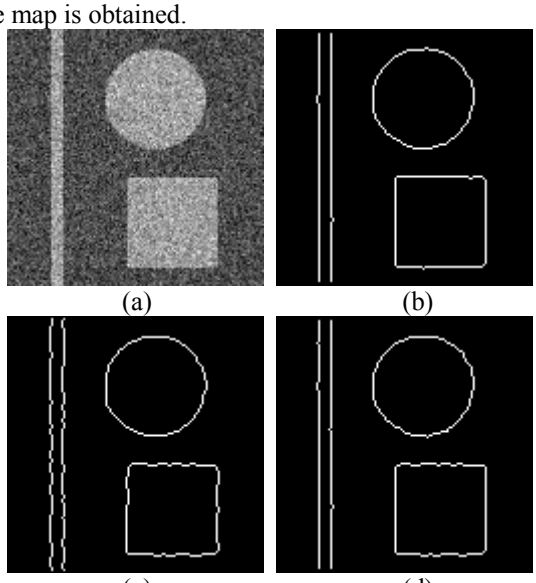
For the value of probability $P\{v_j\le c\cdot\kappa_j\}$, when $c\ge 5$, the probability will be close to 1, which implying that a threshold $t^d(j)=c\cdot\kappa_j$ will suppress most of the data in V_j^d by increasing c. In real applications, the input is the combination of noise and signal. At fine scales, the noise will be predominant in the filter response except for some significant edge structures to be detected. Since the contrast of image singularities and noise is greatly amplified in MERF, the thresholds $t(j)$ applied to the MERF gradient modulus map $M_jf(x,y)$ are set as

$$t(j)=\alpha\cdot\sqrt{t^x(j)+t^y(j)} \tag{17}$$

Significant structure edge points are identified if its corresponding module of MERF gradient is greater than the adaptive threshold $t(j)$. By experiment, setting $\alpha\approx 0.8$ could achieve satisfying results, i.e., suppressing most of noise while preserving the edges.

4. Experiments

In this section, we took a synthetic image, in which the true edges are known, as benchmark image to test the detection performance by some quantitative measurement and we experimented on several natural images to validate the proposed adaptive threshold multi-scale scheme (we denote it as AMED). The canny edge detector (CED) and Mallat wavelet based edge detector (WED) are employed for comparison. In CED we fixed the low threshold to be half of the high one, and consequently there are two parameters in CED. One is the high threshold t_h to suppress false edges. The other is σ_g, the standard deviation of the Gaussian function that is used to adjust the width of the detection filter. For MERF, we choose the larger scale as 2^3 to keep the false edge rate low, while achieving high edge location accuracy by multiplying a small scale 2^2. In the following experiment results of CED, and WED, we adjusted the parameters until a visually best edge map is obtained.



(a) (b) (c) (d)

Figure 2. Noisy synthetic image and edge maps: (a) noisy benchmark, (b) edge map by the proposed algorithm, (c) edge map by Canny, (d) edge map by Mallat wavelet.

Fig. 2(a) is a 128×128 noisy synthetic benchmark image with a circle, a square and two neighboring step edges in it. The added noise level is $\sigma=22$ (SNR=11.93dB). Fig. 2(b) shows the MAED results. Fig. 2(c) is the detected edge map by CED with $\sigma_g=1.0$ and $t_h=0.4$. Fig. 2(d) illustrates the edge map generated by WED. To objectively compare the edge detection results for benchmark image, we use the measurement figure of merit F [16] to evaluate the performance quantitatively:

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

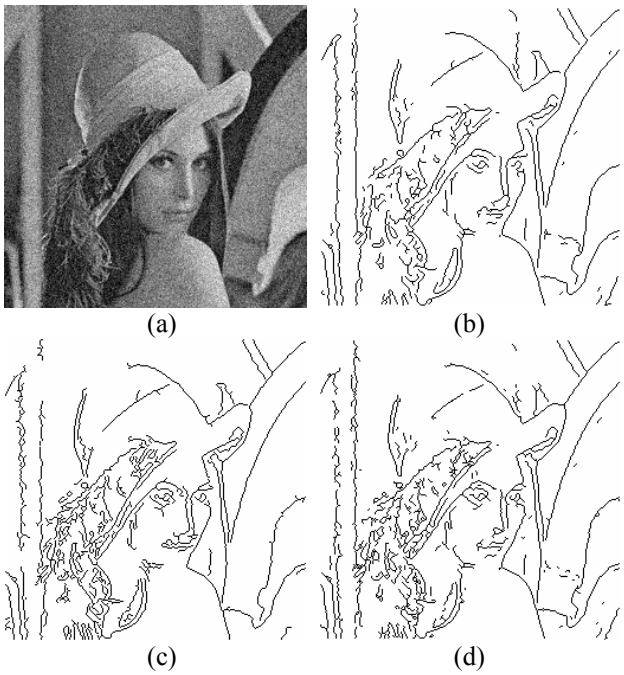
where N_I is the number of the actual edges and N_A is the number of the detected edges. $d(k)$ denotes the distance from the kth actual edge to the corresponding detected edge. α is a scaling constant set to 1/9 as in Pratt's work. The greater the F, the better the detection results. In Table 1, the values of F for the edge maps in Figs. 2 are listed.

Table 1. The figure of merit values of edge maps by CED, WED and MAED

Method	CED	WED	AMED
F	0.9597	0.9898	0.9932

Fig. 3(a) shows the 256×256 noisy Lena image (SNR=15.53dB). Fig. 3(b)-(d) is the edge maps by MAED,

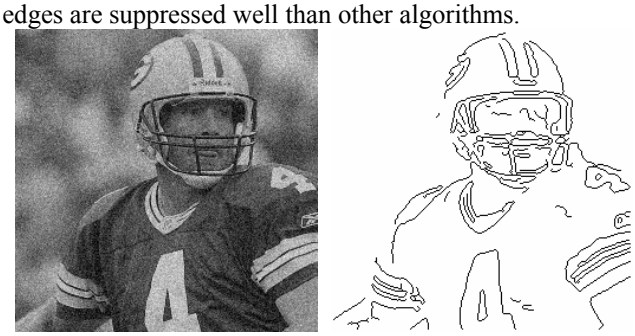
CED and WES respectively. The parameters threshold t and the standard deviation of the Gaussian function σ_g in CED are $\sigma_g = 0.8$, $t_h = 0.3$.



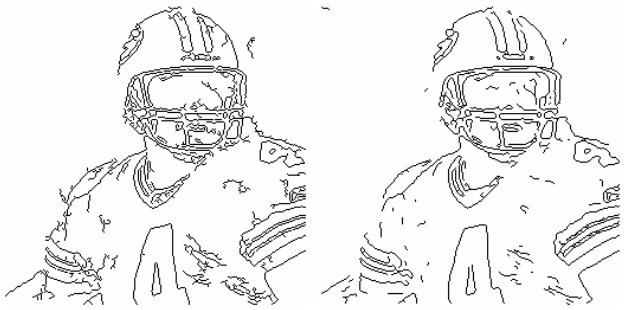
(a) (b) (c) (d)

Figure 3. Noisy image Lena and edge maps: (a) noisy image Lena, (b) edge map by the proposed algorithm, (c) edge map by Canny, (d) edge map by Mallat wavelet.

In Fig. 4, the experiment results on a 256×256 noisy Footballer image (SNR=17.13dB) by the three schemes are shown. Fig. 4(b) shows the edge map by MAED. Fig. 4(c) is the results by CED with $\sigma_g = 0.8$ and $t_h = 0.3$. Fig. 2(d) illustrates the edge map generated by WED. From Fig. 3 and Fig. 4 it can be seen that the proposed scheme MAED achieved very good results with few false edges and high localization accuracies. On one hand a rich of class of edges, are detected with better localization through MAED. On the other hand, the edge map is "clear", which means that false edges are suppressed well than other algorithms.



(a) (b)



(c) (d)

Figure 4. Noisy image Footballer and edge maps: (a) noisy footballer, (b) edge map by the proposed algorithm, (c) edge map by Canny, (d) edge map by Mallat wavelet.

5. Conclusions

We proposed an adaptive threshold multi-scale edge detection scheme with noise immunity by MERF. Taking the advantage of similarities in the wavelet filter's responses at adjacent scales, we defined a MERF function as the point-wise product of the responses of wavelet in adjacent scales to amplify edge structures while diluting noise; and detect the edges as the local maxima in the MERF gradient using a properly designed adaptive threshold scheme. Through MERF we can achieve feature-level synthesis process of multiple scales and at the same time avoid the ill-posed edge results synthesis process in most multi-scale detection schemes. The theoretical analyses on MERF and adaptive threshold show that the new scheme can improve the edge localization accuracy and yield better edge detection results. At last experiments on synthetic and natural images were made to test the proposed method, in which edges and noise can be better distinguished in comparison with the Mallat wavelet-based edge detector and Canny edge detector.

Acknowledgments

The work in this paper was supported by the National Natural Science Foundation of China (No.60403008) and the National High Technology Development 863 Program of China (No.2006AA01Z324).

References

[1] R. C. Gonzalez, R. E. Woods and S. L. Eddins, Digital Image Processing Using MATLAB. Publishing House of Electronics Industry, Beijing, China, 2004.

[2] J. Canny, "A computational approach to edge detection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, no. 2, pp.679-698, 1986.

[3] J. Koplowitz and V. Greco, "On the Edge Location Error for Local Maximum and Zero-crossing Edge Detectors," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 16, no. 12, pp.1207–1212, 1994.

[4] M.C. Shin, D.B. Goldgof, K.W. Bowyer, and S. Nikiforou, "Comparison of Edge Detection Algorithms Using a Structure from Motion Task," IEEE Trans. System, Man, and Cybernetics, vol. 31, no. 4, pp. 589-601, 2001.

[5] S. Mallat and W. L. Hwang, "Singularity Detection and Processing with Wavelets," IEEE. Trans. Information Theory, vol. 38, no. 2, pp.617-643, 1992.

[6] S. Mallat and S. Zhong, "Characterization of Signals from Multiscale Edges," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 14, no. 7, pp.710-732, 1992.

[7] P. Bao , L. Zhang and X. Wu, "Canny Edge Detection Enhancement by Scale Multiplication", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 9, pp.1485-1490, 2005.

[8] D. Demigny, "On Optimal Linear Filtering for Edge Detection," IEEE Trans. Image Processing, vol. 11, no. 7, pp.728-737, 2002.

[9] B.M. Sadler and A. Swami, "Analysis of Multiscale Products for Step Detection and Estimation," IEEE Trans. Information Theory, vol. 45, pp.1043-1051, 1999.

[10] H. Wei and L. Shen, "Multiscale Edge Detection by Using Anti-symmetrical Biorthogonal Wavelets," Acta Electronica Sinica, vol. 30, no. 3, pp.313-316, 2002.

[11] H. Yang, W. Zhang, "Research on Image Edge Detection Based on Multiscale Wavelet Transform and Fuzzy Clustering," Computer Science, vol. 33, no. 1, pp. 174-176, 2006

[12] G. Zhang and Q. Liu, "Robust Edge Detection Based on Stationary Wavelet Transform," Journal of Southeast University, vol.22, no. 2, pp.218-221, 2006.

[13] A. Rosenfeld and M. Thurston, "Edge and Curve Detection for Visual Scene Analysis," IEEE Trans. Computers, vol. 20, no. 5, pp.562–569, 1971.

[14] D.L. Donoho and I.M. Johnstone, "Ideal Spatial Adaptation via Wavelet Shrinkage," Biometrika, vol. 81, no. 3, pp.425–455, 1994.

[15] J. K. Patel and C. B. Read, Handbook of The Normal Distribution. Marcel Dekker, New York, USA, 1982.

[16] W.K. Pratt, Digital Image Processing. John Wiley & Sons, New York, USA, 2001.