Canny Edge Detection Enhancement by Scale Multiplication

Paul Bao, Lei Zhang, and Xiaolin Wu

Abstract—The technique of scale multiplication is analyzed in the framework of Canny edge detection. A scale multiplication function is defined as the product of the responses of the detection filter at two scales. Edge maps are constructed as the local maxima by thresholding the scale multiplication results. The *detection* and *localization* criteria of the scale multiplication are derived. At a small loss in the *detection* criterion, the *localization* criterion can be much improved by scale multiplication. The product of the two criteria for scale multiplication is greater than that for a single scale, which leads to better edge detection performance. Experimental results are presented.

Index Terms—Edge detection, scale multiplication, multiscale analysis.

───────────── ✦ ─────────────

1 INTRODUCTION

EDGES carry important information of an image. Numerous edge detection techniques have been proposed [1], [2], [3], [4], [5], [6], [7], [8], [9]. The common approach is to apply the first (or second) derivative to the smoothed image and then find the local maxima (or zero-crossings). Canny [2] first presented the well-known three criteria of edge detectors: *good detection*, *good localization*, and *low spurious response* and showed that the optimal detector for an isolated step edge should be the *first derivative of Gaussian* (FDOG). An error in Canny's *localization* criterion was fixed by Tagare and de Figueiredo [5] and by Koplowitz and Greco [6]. Recently, Demigny [4] analyzed in detail the optimal detection filters in discrete domain.

An important issue in edge detection is the scale of detection filter. Small-scaled filters are sensitive to edge signals but also prone to noise, whereas large-scaled filters are robust to noise but could filter out fine details. As suggested by Marr and Hildreth [1], multiple scales could be employed to describe and synthesize the varieties of edge structures. The idea of scale multiplication was first exploited by Rosenfeld in [3], where it is shown that the scale products can improve the edge localization. Mallat [10] illustrated mathematically that signals and noise have different singularities and edge structures present observable magnitudes along the scales, while noise decreases rapidly. With this observation, Xu et al. [11] proposed a wavelet-based spatially selective filtering technique by multiplying the adjacent scales. Sadler and Swami [12] applied the wavelet-multiscale-products to step detection and estimation and Bao and Zhang [13] presented a denoising scheme by thresholding the multiscale products.

Inspired by the work of Canny [2] and others [5], [6] on criteria of edge detection filters, we define the *detection* and *localization* criteria of the scale multiplication. We show that through scale multiplication the *localization* accuracy can be significantly improved with only a small loss in the *detection* criterion and the product of the two criteria for the scale multiplication is greater than that for a single scale, leading to superior edge detection results. A simple but efficient edge detector by scale multiplication is then proposed.

─────────────

• *P. Bao is with the Nanyang Technological University, Singapore, 639798. E-mail: aspbao@ntu.edu.sg.*
• *L. Zhang and X. Wu are with McMaster University, Hamilton, ON, Canada, L8G 4K1. E-mail: {johnray, xwu}@mail.ece.mcmaster.ca.*

2 CRITERIA OF OPTIMAL EDGE DETECTION FILTERS

Suppose that an FIR filter $f(x)$ supported in $[-T, T]$ is used as the detector. The signal to be detected is a single step edge corrupted by noise: $W(x) = G(x) + n(x)$, where $n(x) \sim N(0, \sigma^2)$ is *Gaussian* white noise and $G(x)$ is a step edge whose magnitude is A when $x \geq 0$. The response of $G(x)$ to $f(x)$ is $H_G(x) = \int_{-T}^{T} G(x-t)f(t)dt$ and that of $n(x)$ is $H_n(x) = \int_{-T}^{T} n(x-t)f(t)dt$ and then the response of $W(x)$ to $f(x)$ is $H_W(x) = H_G(x) + H_n(x)$. Considering the edges as the local maxima of $H_W(x)$, Canny [2] generalized three criteria to evaluate an edge detector:

1. *Good detection.* It is expected that at abscissa $x = 0$, the signal-to-noise ratio (SNR) should be as high as possible. The criterion is:

$$SNR = A \left| \int_{-T}^{0} f(x)dx \right| \Big/ \left(\sigma \sqrt{\int_{-T}^{T} f^2(x)dx} \right). \quad (2.1)$$

2. *Good localization.* The detected edge should be as close to location $x = 0$ as possible. Denote by y the local maximum position in $H_W(x)$. Canny approximated $y \approx H_n'(y)/H_G''(0)$ and defined the criterion as the reciprocal of the standard deviation of y. He computed $E[H_n'^2(y)] = \sigma^2 \int_{-T}^{T} f'^2(x)dx$. However, Tagare and Figueiredo [5] pointed out this expression is correct only if $H_n'(x)$ is sampled at the same position for all realizations of $H_n(x)$. But, the location y is not constant. In [6], Koplowitz and Greco gave another approximation of y:

$$y \approx -H_n'(0) \Big/ \left(H_G''(0) + H_n''(0) \right). \quad (2.2)$$

Let $\varsigma = -H_n'(0)$ and $\xi = H_G''(0) + H_n''(0)$. ς and ξ are independent Gaussian variables. Denote by μ_ς and μ_ξ and σ_ς^2 and σ_ξ^2 the means and variances of them, respectively. The pdf of y is [6]:

$$\Pr(y) = \frac{1}{2\pi\sigma_\varsigma\sigma_\xi} \cdot \frac{e^{-c}}{a} \left(1 + b\sqrt{\pi/a} \cdot e^{b^2/a} \cdot erf(b/\sqrt{a}) \right), \quad (2.3)$$

where $erf(x)$ is the error function and $a = \frac{y^2}{2\sigma_\varsigma^2} + \frac{1}{2\sigma_\xi^2}$, $b = -\frac{\mu_\varsigma y}{2\sigma_\varsigma^2} - \frac{\mu_\xi}{2\sigma_\xi^2}$, and $c = \frac{\mu_\varsigma^2}{2\sigma_\varsigma^2} + \frac{\mu_\xi^2}{2\sigma_\xi^2}$. Note that $\int y^2 \Pr(y)dy$ is infinite. Limiting the edge location between interval $[-T, T]$, we compute the variance of y as $E[y^2] = \int_{-T}^{T} y^2 \Pr(y)dy$. Thus, the *good localization* criterion is defined as

$$L = 1 \Big/ \sqrt{\int_{-T}^{T} y^2 \Pr(y)dy}. \quad (2.4)$$

3. *Low spurious response.* Since the input is a single step, the detector should not produce multiple maxima. Canny defined by

$$x_{\max}(f) = 2\pi \sqrt{\int_{-T}^{T} f'^2(x)dx} \Big/ \int_{-T}^{T} f''2(x)dx$$

the mean distance between two noise maxima and expected $x_{\max}(f)$ to be as big as possible.

In Canny's view, the optimal edge detection filter should maximize the product of *detection* and *localization* criteria with the constraint in *low spurious response*. However, Demigny [4] pointed out that the criterion of *low spurious response* is not essential and he further proved that the influence of this criterion is greatly reduced by thresholding operation and a good filter can be determined by only optimizing the *detection-localization* product.
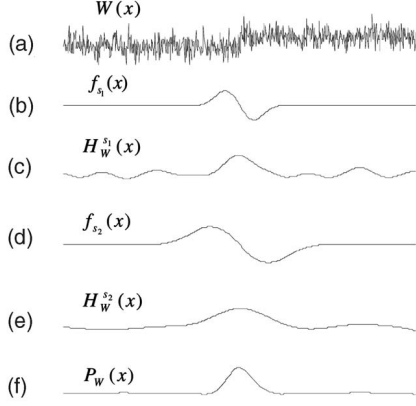
Fig. 1. (a) Noisy step edge $W(x)$. (b) The FDOG at scale s_1. (c) The response of $f_{s_1}(x)$ to $W(x)$. (d) The FDOG at scale s_2. (e) The response of $f_{s_2}(x)$ to $W(x)$. (f) The responses product of scales s_1 and s_2.

3   ANALYSIS OF SCALE MULTIPLICATION IN EDGE DETECTION

3.1   Scale Multiplication

Denote by $f_s(x) = f(x/s)/s$ the dilation of $f(x)$ by scale s. The support of $f_s(x)$ is $[-T_s, T_s]$, where $T_s = sT$. We denote the response of $W(x)$ to $f_s(x)$ as $H_W^s(x) = H_G^s(x) + H_n^s(x)$, where $H_G^s(x)$ and $H_n^s(x)$ are the responses of $G(x)$ and $n(x)$, respectively. In this paper, we take the FDOG $f(x) = -x \cdot e^{-x^2/2}$ as the edge detection filter and then $f_s(x) = -xe^{-x^2/(2s^2)}/s^2$. We use a small scale s_1 and a large scale s_2 to detect the step edge. The responses at the two scales are $H_W^{s_1}(x)$ and $H_W^{s_2}(x)$. The scale multiplication is defined as the product of $H_W^{s_1}(x)$ and $H_W^{s_2}(x)$

$$P_W(x) = H_W^{s_1}(x) \cdot H_W^{s_2}(x) \tag{3.1}$$

Fig. 1 shows a noisy step edge $W(x)$, the edge detection filters $f_{s_1}(x)$ and $f_{s_2}(x)$ $(s_2 = 2s_1)$, the responses at the two scales, and their products $P_W(x)$. We see that with scale s_1, the step edge is more accurately localized but some false local maxima are produced; with scale s_2, fewer false edges are detected but traded off with a decreased accuracy in edge location. $P_W(x)$ combines the advantages of the two scales. In $P_W(x)$, the step edge is much sharper compared with $H_W^{s_2}(x)$, while noise is better suppressed compared with $H_W^{s_1}(x)$. Intuitively, more robust detection results can be obtained if edges are determined as the local maxima in $P_W(x)$ after thresholding.

3.2   The Detection Criterion

The detection criterion proposed by Canny [2] is based on the linear property of convolution. In $P_W(x)$, however, nonlinearity is introduced by the multiplication operation. Similar to Canny's definition, the detection criterion is considered as the SNR of $P_W(x)$ at $x = 0$, where the step edge occurs. If the system input is a noiseless signal $G(x)$, $P_W(0)$ would be $H_G^{s_1}(0) \cdot H_G^{s_2}(0)$ and if the input is noise $n(x)$, the output would be $H_n^{s_1}(0) \cdot H_n^{s_2}(0)$. We define the detection criterion as

$$SNR_P = \frac{\sqrt{\left| H_G^{s_1}(0) \cdot H_G^{s_2}(0) \right|}}{\sqrt{E[\left| H_n^{s_1}(0) \cdot H_n^{s_2}(0) \right|]}}.$$

Note that, if $s_1 = s_2$ (i.e., $P_W(x)$ is the square of $H_W^{s_1}(x)$), SNR_P is the same as the SNR in (2.1) with $f(x) = f_{s_1}(x)$. Since $H_G^s(0)$ is invariant with scale s and $H_G^s(0) = A$, we have $SNR_P = A/\sqrt{E[|H_n^{s_1}(0) \cdot H_n^{s_2}(0)|]}$. $H_n^{s_1}(0)$ and $H_n^{s_2}(0)$ are linear filtering outputs

of Gaussian noise $n(x)$ and, thus, zero mean and *jointly Gaussian* distributed:

$$\Pr(z_1, z_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} \exp\left\{ -\frac{1}{2(1 - \rho^2)} \left(\frac{z_1^2}{\sigma_1^2} - \frac{2\rho z_1 z_2}{\sigma_1\sigma_2} + \frac{z_2^2}{\sigma_2^2} \right) \right\},$$

where

$$\sigma_1 = \sigma \cdot \sqrt{\sqrt{\int f_{s_1}^2(t)dt}} = \sqrt[4]{\pi/(4s_1^2)} \cdot \sigma \text{ and}$$

$$\sigma_2 = \sigma \cdot \sqrt{\sqrt{\int f_{s_2}^2(t)dt}} = \sqrt[4]{\pi/(4s_2^2)} \cdot \sigma \tag{3.2}$$

and ρ is the correlation coefficient of $H_n^{s_1}(0)$ and $H_n^{s_2}(0)$:

$$\rho = \frac{\int f_{s_1}(t)f_{s_2}(t)dt}{\sqrt{\int f_{s_1}^2(t)dt} \cdot \sqrt{\int f_{s_2}^2(t)dt}} = \sqrt{\frac{2^3 s_1^3 s_2^3}{(s_1^2 + s_2^2)^3}}. \tag{3.3}$$

Note that ρ is invariant with the ratio $r = s_2/s_1$. The expectation of $\left| H_n^{s_1}(0) \cdot H_n^{s_2}(0) \right|$ is

$$E\left[\left| H_n^{s_1}(0) \cdot H_n^{s_2}(0) \right|\right] = 2\sigma_1\sigma_2(\cos\alpha + \alpha\sin\alpha)/\pi, \tag{3.4}$$

where α is given by $\sin\alpha = \rho, -\pi/2 < \alpha \leq \pi/2$. Then, we can compute SNR_P as

$$SNR_P = \frac{A}{\sigma} \cdot \frac{\sqrt[4]{\pi s_1 s_2}}{\sqrt{\cos\alpha + \alpha\sin\alpha}}. \tag{3.5}$$

When ratio $r = s_2/s_1$ is fixed, ρ is a constant and α is fixed. If A/σ is also fixed, we see that SNR_P is proportional to $\sqrt{s_1}$. This observation is similar tp that on the detection criterion SNR in (2.1). Substituting $f_s(x) = -xe^{-x^2/s^2}/s^2$ for $f(x)$ in (2.1), we can compute that $SNR = \sqrt{s} \cdot (A/\sigma) \cdot \sqrt[4]{4/\pi}$. Apparently, SNR is also proportional to $\sqrt{s}$.

3.3   The Localization Criterion

Denote by z the local maximum position of $P_W(x)$ near $x = 0$. The derivative of $P_W(x)$ at z is $P_W'(z) = H_W'^{s_1}(z) \cdot H_W^{s_2}(z) + H_W^{s_1}(z) \cdot H_W'^{s_2}(z) = 0$. By Taylor expansion of the derivatives of $H_W^s(z)$ and $H_W'^s(z)$ about $x = 0$ and neglecting the $o(z^2)$ terms, we have

$$z \approx \frac{-H_n'^{s_1}(0)H_W^{s_2}(0) - H_n'^{s_2}(0)H_W^{s_1}(0)}{H_W''^{s_1}(0)H_W^{s_2}(0) + H_W''s_2(0)H_W^{s_1}(0) + 2H_n'^{s_1}(0)H_n'^{s_2}(0)}. \tag{3.6}$$

We used $H_G'^s(0) = 0$ in deriving the above equation. For expression convenience, we denote $\varsigma_i = H_W^{s_i}(0)$, $\xi_i = H_n'^{s_i}(0)$, and $\eta_i = H_W''^{s_i}(0)$, $i = 1, 2$ and then $z \approx \frac{-\xi_1\varsigma_2 - \xi_2\varsigma_1}{\eta_1\varsigma_2 + \eta_2\varsigma_1 + 2\xi_1\xi_2}$.

All variables ς_i, ξ_i, and η_i are linear filtering outputs of noise $n(x)$, so they are *jointly Gaussian*. Let $\vec{X} = col\{\varsigma_1, \varsigma_2, \xi_1, \xi_2, \eta_1, \eta_2\}$. The mean vector and covariance matrix of $\vec{X}$ are $\vec{M} = E[\vec{X}]$ and $\boldsymbol{\Omega} = E[(\vec{X} - \vec{M})(\vec{X} - \vec{M})^T]$. The pdf of $\vec{X}$ is [14]

$$\Pr(\vec{X}) = \frac{1}{(2\pi)^3\sqrt{|\boldsymbol{\Omega}|}} \exp\left(-\frac{1}{2}\vec{X} - \vec{M})^T \boldsymbol{\Omega}^{-1}\vec{X} - \vec{M}) \right), \tag{3.7}$$

where $|\boldsymbol{\Omega}|$ is the determinant of matrix $\boldsymbol{\Omega}$. The elements of $\vec{M} = E[\vec{X}]$ are computed as $E[\varsigma_i] = H_G^{s_i}(0) = A$, $E[\xi_i] = 0$, and $E[\eta_i] = H_G''s_i(0) = -A/s_i^2$. The elements of $\boldsymbol{\Omega}$ are

$$cov(\varsigma_i, \varsigma_j) = \sigma^2 \int f_{s_i}(x)f_{s_j}(x)dx$$
$$= \sigma^2 \cdot \sqrt{2\pi}s_is_j/(s_i^2 + s_j^2)^{1.5}, i, j = 1, 2 \tag{3.8a}$$

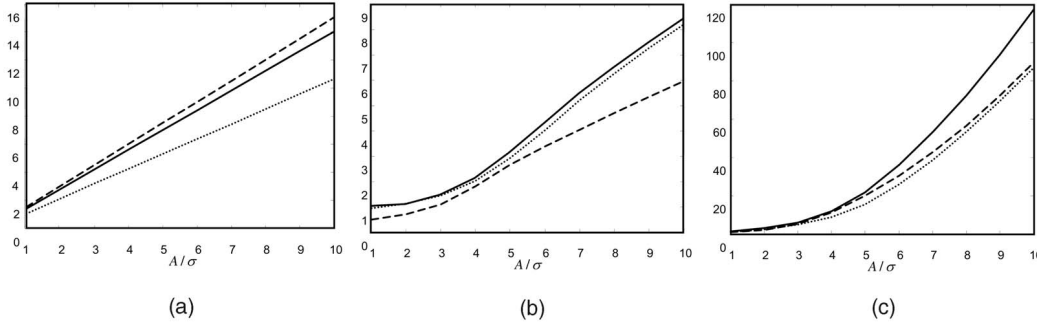Fig. 2. When $s_1 = 2^0$ and $s_2 = 2^1$. (a) The curves Σ_1 (dotted), Σ_2 (dashed), and Σ_P (solid). (b) The curves Λ_1 (dotted), Λ_2 (dashed), and Λ_P (solid). (c) The curves $\Sigma_1 \times \Lambda_1$ (dotted), $\Sigma_2 \times \Lambda_2$ (dashed), and $\Sigma_P \times \Lambda_P$ (solid).

Fig. 3. When $s_1 = 2^1$ and $s_2 = 2^2$. (a) The curves Σ_1 (dotted), Σ_2 (dashed), and Σ_P (solid). (b) The curves Λ_1 (dotted), Λ_2 (dashed), and Λ_P (solid). (c) The curves $\Sigma_1 \times \Lambda_1$ (dotted), $\Sigma_2 \times \Lambda_2$ (dashed), and $\Sigma_P \times \Lambda_P$ (solid).
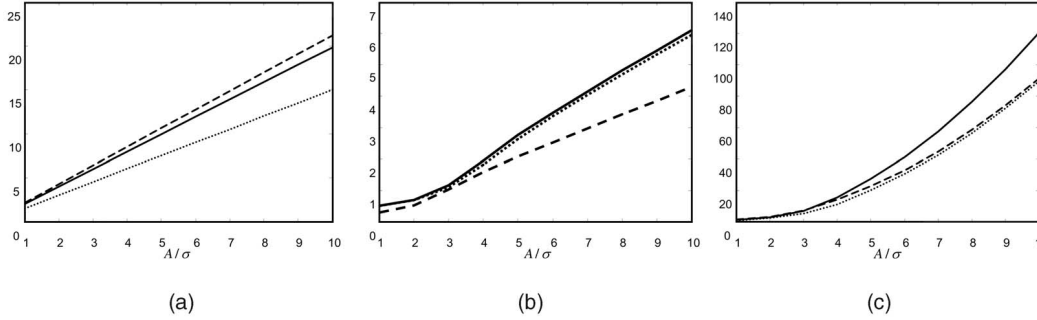
$$\text{cov}(\xi_i, \xi_j) = \sigma^2 \int f'_{s_i}(x) f'_{s_j}(x) dx = \sigma^2 \cdot 3\sqrt{2\pi} s_i s_j / (s_i^2 + s_j^2)^{2.5}, i,j = 1,2,$$
(3.8b)

$$\text{cov}(\eta_i, \eta_j) = \sigma^2 \int f''_{s_i}(x) f''_{s_j}(x) dx = \sigma^2 \cdot 15\sqrt{2\pi} s_i s_j / (s_i^2 + s_j^2)^{3.5}, i,j = 1,2,$$
(3.8c)

$$\text{cov}(\varsigma_i, \xi_j) = 0, \text{cov}(\varsigma_i, \eta_j) = 0, \text{ and } \text{cov}(\eta_i, \xi_j) = 0, i,j = 1,2. \quad (3.8d)$$

In fact, we can write $\Pr(\vec{X}) = \Pr(\varsigma_1, \varsigma_2) \cdot \Pr(\xi_1, \xi_2) \cdot \Pr(\eta_1, \eta_2)$ because ς_i, ξ_i, and η_i are mutually independent, where $\Pr(\bullet, \bullet)$ is the bivariate Gaussian distribution.

We limit the local maximum of $H_W^{s_i}(x)$ between $[-T_{s_i}, T_{s_i}]$. Suppose $s_2 > s_1$, then $T_{s_2} > T_{s_1}$. Referring to Fig. 1, the edge is enhanced in $P_W(x)$ and we limit y, the local maximum of $P_W(x)$, between $[-T_{s_1}, T_{s_1}]$. The variance of z can be computed by $E[z^2] = \int_{-T_{s_1}}^{T_{s_1}} z^2 \Pr(z) dz$ if we know $\Pr(z)$, the pdf of z. However, z is a complex function of ς_i, ξ_i, and η_i and it is hard to obtain the analytic form of $\Pr(z)$. Instead, we compute numerically the variance of z by using $\Pr(X)$:

$$E[z^2] = \iiint\iiint\limits_{|z| \leq T_{s_1}} z^2 \Pr(\vec{X}) d\varsigma_1 d\varsigma_2 d\xi_1 d\xi_2 d\eta_1 d\eta_2. \quad (3.9)$$

Finally, similar to (2.4), we define the localization criterion of z as

$$L_P = 1/\sqrt{E[z^2]}. \quad (3.10)$$

3.4 Numerical Comparison of SNR_P, L_P with SNR, L

SNR_P is computed with (3.5) and SNR defined in (2.1) is computed by $SNR = \sqrt{s} \cdot (A/\sigma) \cdot \sqrt[4]{4/\pi}$. Localization criteria L_P in (3.10) and L in (2.4) are computed numerically. Since the Gaussian function decays rapidly, we set the support of $f(x) = -x \cdot e^{-x^2/2}$ to $[-4,4]$. Thus, the support of $f_s(x) = -xe^{-x^2/s^2}/s^2$ is

$[-4s, 4s]$, i.e., $T_s = 4s$. Parameters μ_ς, μ_ξ, σ_ς, and σ_ξ in probability function $\Pr(y)$ (2.3) are computed as

$$\mu_\varsigma = E[H'_n(0)] = 0,$$

$$\sigma_\varsigma = \sqrt{E[H'^2_n(0)]} = \sigma\sqrt{\int_{-T_s}^{T_s} f'^2_s(x) dx} = \sigma\sqrt{3\sqrt{\pi}/(4s^3)},$$

$$\mu_\xi = E[H''_G(0)] = Af'_s(0) = -A/s^2,$$

$$\sigma_\xi = \sqrt{E[H''^2_n(0)]} = \sigma\sqrt{\int_{-T_s}^{T_s} f''^2_s(x) dx} = \sigma\sqrt{15\sqrt{\pi}/(8s^5)}.$$

Denote by Σ_1 and Σ_2 the SNR values of $f_{s_1}(x)$ and $f_{s_2}(x)$ and by Λ_1 and Λ_2 the L values of them. The corresponding values of criteria SNR_P and L_P are denoted by Σ_P, Λ_P. In the sequel, we set $s_2 = 2 \cdot s_1$ and compute Σ_1, Σ_2, Λ_1, Λ_2, Σ_P, and Λ_P versus A/σ.

Fig. 2a plots the curves of Σ_1, Σ_2, and Σ_P versus A/σ when $s_1 = 2^0$. Fig. 2b plots the corresponding curves of Λ_1, Λ_2, and Λ_P, and Fig. 2c plots the products $\Sigma_1 \times \Lambda_1$, $\Sigma_2 \times \Lambda_2$, and $\Sigma_P \times \Lambda_P$. Figs. 3a, 3b, and 3c plot these curves for $s_1 = 2^1$. Similar observations are made on other values of s_1. We see that Σ_P is much greater than Σ_1 and is only slightly less than Σ_2. At the same time, Λ_P is not only much higher than Λ_2, but also slightly higher than Λ_1. Finally, $\Sigma_P \times \Lambda_P$ is much higher than $\Sigma_1 \times \Lambda_1$ and $\Sigma_2 \times \Lambda_2$, meaning that scale multiplication significantly improves edge detection results.

3.5 Thresholding

In first-derivative-based edge detection, the gradient image should be thresholded to eliminate false edges produced by noise. With a single threshold t, some false edges may appear if t is too small and some true edges may be missed if t is too large. In [2], Canny proposed a double thresholding algorithm. After nonmaxima suppression, a low threshold t_l and a high threshold $t_h \approx 2t_l$ are applied to obtain double thresholded edge maps, I_l and I_h. The algorithm selects edges in I_l that link to the edges in I_h. The double
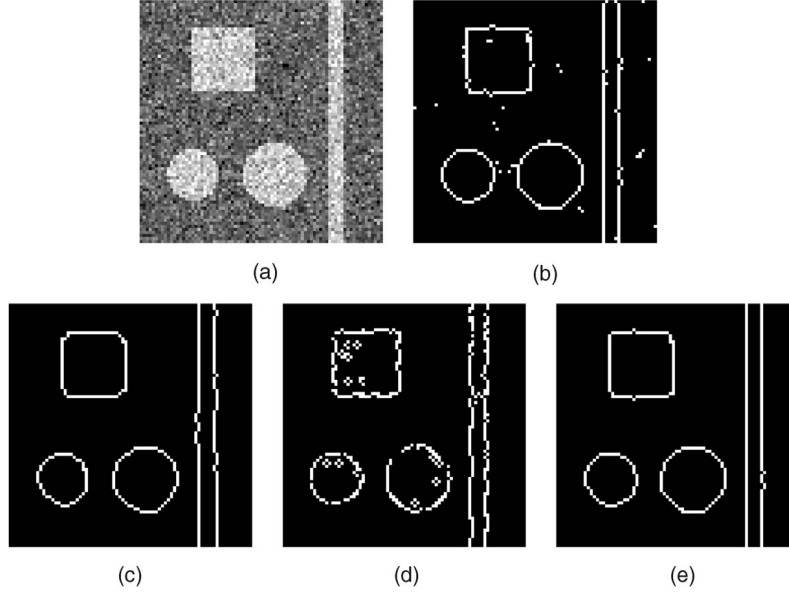
Fig. 4. (a) Noisy synthetic image. (b) Edge map by CED at a small scale s_1. (c) Edge map by CED at a large scale s_2. (d) Edge map by ADED. (e) Edge map by the proposed SMED with s_1 and s_2.

thresholding algorithm can also be applied to our scheme. Considering that edges and noise can be better distinguished in the scale products than in a single scale, however, we adopt the simple single thresholding strategy.

The edges are detected as the local maxima in $P_W(x)$. A significant edge at abscissa x_0 will occur at both the scales and the signs of $H_W^{s_1}(x_0)$ and $H_W^{s_2}(x_0)$ will be the same, so that $P_W(x_0)$ will be nonnegative. If $P_W(x) < 0$, the point is considered as noise and filtered out. Suppose the input is pure noise $n(x)$ and then the products function is $P_n(x) = H_n^{s_1}(x) \cdot H_n^{s_2}(x)$. Variables $H_n^{s_1}(x)$ and $H_n^{s_2}(x)$ are *jointly Gaussian* with standard deviations σ_1 and σ_2, respectively, and correlation coefficient ρ. The pdf of their product, $P_n(x)$, is [14, p. 42]

$$\Pr(z) = \frac{1}{\pi\Gamma(1/2)\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(\frac{\rho z}{(1-\rho^2)\sigma_1\sigma_2}\right)K_0\left(\frac{|z|}{(1-\rho^2)\sigma_1\sigma_2}\right),$$
(3.11)

where $\Gamma(t) = \int_0^\infty e^{-u}u^{t-1}du$ is the Gamma function and K_0 is the modified Bessel function of the second kind with order zero. The standard deviation of $P_n(x)$ is

$$\kappa = \sqrt{E[z^2]} = \sqrt{E[u^2v^2]} = \sqrt{1+2\rho^2}\cdot\sigma_1\sigma_2. \qquad (3.12)$$

A threshold $t = c \cdot \kappa$ can suppress most of the $P_n(x)$ values by increasing c. In real applications, the input is the combination of noise and signal. The noise will be dominant in the filter response except for the significant edge structures. We find that setting $t_P = c \cdot \kappa$, $c > 5$ leads to satisfactory results, i.e., suppressing most of the noise while well preserving the edges.

3.6 Two Dimensions

In 2D images, two scale product functions are needed. Denote by $f_s^x(x,y)$ and $f_s^y(x,y)$ the two detection filters in x and y directions. Their responses to image I are $H_x^s(x,y)$ and $H_y^s(x,y)$. The product functions are defined as $P_I^x(x,y) = H_x^{s_1}(x,y) \cdot H_x^{s_2}(x,y)$ and $P_I^y(x,y) = H_y^{s_1}(x,y) \cdot H_y^{s_2}(x,y)$. For an edge point (x_0,y_0), $H_x^s(x_0,y_0)$, or $H_y^s(x_0,y_0)$ will have the same sign at adjacent scales s_1 and s_2 so both $P_I^x(x_0,y_0)$ and $P_I^y(x_0,y_0)$ will be nonnegative and the orientation information of the edge is lost, which should be recovered from $H_x^{s_1}(x_0,y_0)$ and $H_y^{s_1}(x_0,y_0)$. Setting the points with $P_I^x(x,y) < 0$ (or $P_I^y(x,y) < 0$) to 0, the modulus and angle of point (x,y) are defined as

$$M_I(x,y) = \sqrt{P_I^x(x,y) + P_I^y(x,y)} \text{ and}$$

$$A(x,y) = \arctan\left(\frac{\text{sgn}\left(H_y^{s_1}(x,y)\right)\cdot\sqrt{P_I^y(x,y)}}{\text{sgn}(H_x^{s_1}(x,y))\cdot\sqrt{P_I^x(x,y)}}\right). \qquad (3.13)$$

As in the Canny edge detector, an edge point is asserted wherever $M_I(x,y)$ has a local maximum in the direction given by $A_I(x,y)$. The modulus map $M_I(x,y)$ should be thresholded to remove noise. Referring to Section 3.5, the threshold applied to $P_I^x(x,y)$ is $t_P^x = c \cdot \kappa_x$, $c > 5$. The threshold t_P^y applied to $P_I^y(x,y)$ is obtained similarly. We set the threshold to $M_I(x,y)$ as $t_P = \sqrt{t_P^x + t_P^y}$.

4 EXPERIMENTAL RESULTS

We experimented the proposed technique on a synthetic image and two natural images. The synthetic image, in which the true edges are known, is used to test the edge detection performance by some quantitative measurement. With the scale multiplication strategy, we choose the larger scale big enough to keep the false edge rate low, while achieving high edge location accuracy by multiplying a small scale. Let $s_2 = r \cdot s_1$ with $r > 1$. Too small an r would not incorporate sufficient information of two different scales, whereas too large an r would reduce the correlation of the two scales. $r = 2$ is a good choice and it is convenient to discrete implementation. Suppose at scale s_1 the discrete filter length is l_1, then at scale $s_2 = 2s_1$ the filter length is $l_2 = 2l_1 + 2$. At the finest scale, $s = 2^0$, we discretize the FDOG filter to be $f_1 = [-1, 1]$, and at the second scale $s = 2^1$, the filter is $f_2 = [-1, -3, -2, 2, 3, 1]/4$, etc. The proposed edge detector (we denote it as SMED) is compared with the *Canny edge detector* (CED) and the recently proposed *anisotropic diffusion edge detector* (ADED) by Black et al. [8]. The source code of method ADED is obtained from the Web site ftp://figment.csee.usf.edu/pub/Edge_Comparison/source_code/anisotropic.tar.gz. In ADED, there is one parameter, threshold t. We set scale $s_1 = 2^2$ and scale $s_2 = 2^3$ in implementing the SMED and CED in the following experiments. In CED, we fix the low threshold to be half of the high one and set the high threshold as $t_j = \lambda\sigma_j$ at scale s_j, where σ_j (referring to (3.2)) is the noise level at that scale and constant λ is used to control the threshold. In the following experimental results of CED and ADED, we adjusted the threshold until a visually best edge map is obtained.

TABLE 1
The Figure of Merit Values of the Edge Maps in Fig. 4

| Edge Maps	Fig. 4 (b)	Fig. 4 (c)	Fig. 4 (d)	Fig. 4 (e)
F	0.8908	0.8413	0.8022	0.9756

Fig. 4a shows a noisy synthetic image with one square, two circles, and two neighboring step edges in it. The added noise level is $\sigma = 35$. Figs. 4b and 4c illustrate the edge maps generated by CED at the two scales. The constant λ is set to 2.4 and 2.0, respectively, at the two scales. We see that, at the small scale there are many false edges caused by noise, while at the large scale the edge localization accuracy is decreased as a tradeoff of low false-edge-rate. Note that the distance between the two neighboring step edges in Fig. 4c is mistakenly amplified due to the interference of adjacent edge structures when detector scale is increased. Fig. 4d shows the ADED results. The threshold is $t = 1$. It is seen that ADED does not suppress noise well. Fig. 4e is the edge map detected by SMED with $c = 6$. We see that the results remove most of the false edges and achieve very high edge localization performance. The shapes of the objects are precisely detected.
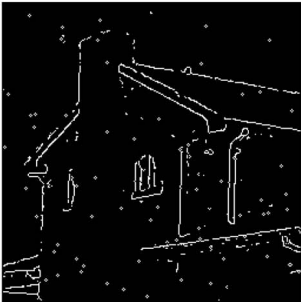

(a)
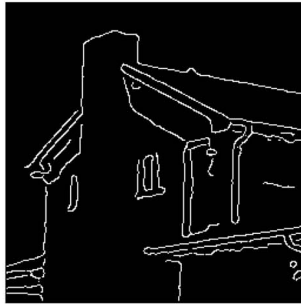
(b)   (c)

(d)   (e)

Fig. 5. (a) Noisy image *House*. (b) Edge map by CED at a small scale s_1. (c) Edge map by CED at a large scale s_2. (d) Edge map by ADED. (e) Edge map by the proposed SMED with s_1 and s_2.

(a)

(b)   (c)

(d)   (e)

Fig. 6. (a) Image *Flower*. (b) Edge map by CED at a small scale s_1. (c) Edge map by CED at a large scale s_2. (d) Edge map by ADED. (e) Edge map by the proposed SMED with s_1 and s_2.

To objectively compare the edge detection results for this synthetic image, the measurement *figure of merit* F proposed by Pratt [15] is used to evaluate the performance quantitatively:

$$F = \frac{1}{\max\{N_I, N_A\}} \sum_{k=1}^{N_A} \frac{1}{1 + \alpha d^2(k)}, \qquad (4.1)$$

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. 4b, 4c, 4d, and 4e are listed.

In Fig. 5, the experimental results on a 256×256 noisy *House* image (added noise level is $\sigma = 25$) by the three schemes are shown. The constant λ in CED is 2.6 and 2.0, respectively, at the two scales. The threshold in ADED is $t = 0.8$. The threshold constant in SMED is $c = 6$. The last experiment is on a naturally noisy image. The image *Flower* taken by a digital camera is corrupted by background noise in the acquisition process. The noise level is estimated as $\sigma = 11$. Fig. 6 showed the edge maps. The constant λ in CED is 2.5 and 2.1, respectively. The threshold in ADED is $t = 1.8$. The constant in SMED is set to $c = 7$. It can be seen that the proposed scheme achieved very good results with few false edges and high localization accuracies.

5 CONCLUSION

We developed a scale multiplication-based scheme to improve the performance of traditional Canny edge detector. Taking the advantage of similarities in the filter's responses at adjacent scales, the new scheme multiplies the responses to enhance edge structures while diluting noise and detect the edges as the local maxima in the scale products. Our theoretical analyses show that scale multiplication can improve the edge localization accuracy and then yield better edge detection results. Experiments on synthetic and natural images were made to test the proposed method.

REFERENCES

[1]   D. Marr and E. Hildreth, "Theory of Edge Detection," *Proc. Royal Soc. London,* vol. 207, pp. 187-217, 1980.
[2]   J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 8, pp. 679-698, 1986.
[3]   A. Rosenfeld, "A Nonlinear Edge Detection Technique," *Proc. IEEE,* pp. 814-816, May 1970.
[4]   D. Demigny, "On Optimal Linear Filtering for Edge Detection," *IEEE Trans. Image Processing,* vol. 11, pp. 728-1220, July 2002.
[5]   H.D. Tagare and R.J.P. de Figueiredo, "On the Localization Performance Measure and Optimal Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 12, no. 12, pp. 1186-1190, Dec. 1990.
[6]   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, Dec. 1994.
[7]   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—Part B: Cybernetics,* vol. 31, pp. 589-601, Aug. 2001.
[8]   M. Black, G. Sapiro, D. Marimont, and D. Heeger, "Robust Anisotropic Diffusion," *IEEE Trans. Image Processing,* vol. 7, pp. 421-432, Mar. 1998.
[9]   H. Jeong and C.I. Kim, "Adaptive Determination of Filter Scales for Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, no. 5, pp. 579-585, May 1992.
[10] S. Mallat, *A Wavelet Tour of Signal Processing.* Academic Press, 1998.
[11] Y. Xu et al., "Wavelet Transform Domain Filters: A Spatially Selective Noise Filtration Technique," *IEEE Trans. Image Processing,* vol. 3, pp. 747-758, Nov. 1994.
[12] B.M. Sadler and A. Swami, "Analysis of Multiscale Products for Step Detection and Estimation," *IEEE Trans. Information Theory,* vol. 45, pp. 1043-1051, Apr. 1999.
[13] P. Bao and L. Zhang, "Noise Reduction for Magnetic Resonance Image via Adaptive Multiscale Products Thresholding," *IEEE Trans. Medical Imaging,* vol. 22, pp. 1089-1099, Sept. 2003.
[14] K.S. Miller, *Multidimensional Gaussian Distributions.* New York: Wiley, 1964.
[15] W.K. Pratt, *Digital Image Processing,* second ed. John Wiley & Sons, Apr. 1991.