Performance Characterization of Edge Detectors

Visvanathan Ramesh
Robert M. Haralick
Department of Electrical Engineering, FT-10
University of Washington
Seattle WA 98195

Abstract

Edge detection is the most fundamental step in vision algorithms. A number of edge detectors have been discussed in the computer vision literature. Examples of classic edge detectors include: the Marr-Hildreth edge operator [11], Facet edge operator [6] and the Canny edge operator [2]. Edge detection using morphological techniques are attractive because they can be efficiently implemented in near real time machine vision systems that have special hardware support [10]. However, little performance characterization of edge detectors has been done. In general, performance characterization of edge detectors has been done mainly by plotting empirical curves of performance. Quantitative performance evaluation of edge detectors was first performed by Abdou and Pratt [1]. It is the goal of this paper to perform a theoretical comparison of gradient based edge detectors and morphological edge detectors. By assuming that an ideal edge is corrupted with additive noise we derive theoretical expressions for the probability of misdetection (the probability of labelling of a true edge pixel as a non-edge pixel in the output). Further, we derive theoretical expressions for the probability of false alarm (the probability of labelling of a non-edge pixel as an output edge pixel) by assuming that the input to the operator is a region of flat graytone intensity corrupted with additive Gaussian noise of zero mean and variance σ^2. Even though the blurring step in the morphological operator introduces correlation in the additive noise, we make an approximation that the output samples after blurring are i.i.d. Gaussian random variables with zero mean and variance σ^2/M where M is the window size of the blurring kernel. The false alarm probabilities obtained by using this approximation can be shown to be upperbounds of the false alarm probabilities computed without the approximation. The theory indicates that the blur-min operator is clearly superior when a 3 by 3 window size is used. Since we only have an upperbound for the false alarm probability the theory is inadequate to confirm the superiority of the blur-min operator. Empirical evaluation of the performance indicates that the blur-min operator is superior to the gradient based operator. Evaluation of the edge detectors on real images also indicate superiority of the blur-min operator. Application of hysteresis linking, after edge detection, significantly reduces the misdetection rate, but increases the false alarm rate.

1 Introduction

Edge detection is the most fundamental step in vision algorithms. Ideal intensity edges are defined by relatively steep intensity changes between two regions in an image. These intensity changes are often modelled as step, ramp or roof functions. Examples of classic edge detectors include: the Marr-Hildreth edge detector [11], Facet edge detector [6] and the Canny edge detector [2]. Edge detection using morphological techniques are attractive because they can be efficiently implemented in near real time machine vision systems that have special hardware support [10]. A number of edge detectors have been discussed in the computer vision literature. However, little performance characterization of edge detectors has been done. In general, performance characterization of edge detectors has been done mainly by plotting empirical curves of performance. Quantitative performance evaluation of edge detectors was first performed by Abdou and Pratt [1]. It is the goal of this paper to perform a theoretical comparison of gradient based edge detectors and morphological

edge detectors. By assuming that an ideal edge is corrupted with additive noise we derive theoretical expressions for the probability of misdetection (the probability of labelling of a true edge pixel as a non-edge pixel in the output). Further, we derive theoretical expressions for the probability of false alarm (the probability of labelling of a non-edge pixel as an output edge pixel) by assuming that the input to the operator is a region of flat graytone intensity corrupted with additive noise.

2 Analysis of Gradient based Edge finding methods

2.1 Edge detection – Noise model

The input image gray values are assumed to be corrupted with noise which may be modelled as a Gaussian distribution with zero mean and standard variance σ. That is:

$$I(r,c) = I_t(r,c) + \eta(r,c) \tag{1}$$

where, $I(r,c)$ is the observed image gray value, $I_t(r,c)$ is the true gray value and $\eta(r,c)$ is the noise component. Often $\eta(r,c)$ is assumed to be zero mean Gaussian with a standard deviation of σ. In order to obtain misdetection characteristics, we assume that the edge is a ramp edge with true gradient magnitude G_t and we analyze the effects of the noise in the estimated gradient magnitude. In order to analyze the false alarm characteristics of an operator, we assume that $I_t(r,c)$ is a constant value in the edge operator neighborhood.

2.2 Probability of Misdetection of a Gradient Edge

Here we analyze edge detectors that use the gradient across a pixel to label a particular pixel as edge or non-edge pixel. We assume that the gradient at a particular pixel is estimated by computing a least squares to the gray levels in the pixel's neighborhood. It is also assumed that the input image is corrupted with additive Gaussian noise with zero mean and variance σ^2. If we approximate the image graytone values in the pixel's neighborhood by a plane $\alpha r + \beta c + \gamma$, then the gradient value $g = \sqrt{\alpha^2 + \beta^2}$. To estimate α and β we use a least squares criterion. On the basis of these estimate, we can derive the density function for the estimated gradient.

Under our assumptions about the noise model in the input image it can be shown that the fitted parameters $\hat{\alpha}$ and $\hat{\beta}$ are Gaussian random variables with means μ_α, μ_β and variances $\sigma_\alpha^2, \sigma_\beta^2$ respectively. We use the notation U_i to denote unit normal random variables with zero mean and unit variance. Under this notation we can rewrite the expressions for α and β as: $\alpha = \mu_\alpha + \sigma_\alpha U_1$ and $\beta = \mu_\beta + \sigma_\beta U_2$. Note that: ${\sigma_\alpha}^2 = {\sigma_\beta}^2$ when a square neighborhood is used in the fit and they are related to the input noise variance σ^2 by the expression:

$${\sigma_\alpha}^2 = \frac{\sigma^2}{\Sigma_r \Sigma_c r^2}. \tag{2}$$

Here $\Sigma_r \Sigma_c r^2$ is the summation of the squared row index values over the neighborhood used in the least squares fit.

By definition $\sum_i^v U_i^2$, where U_i's are i.i.d. unit normal variables, is distributed as a chi-square distribution with v degrees of freedom. Also, $\sum_i^v (U_i + \delta_i)^2$ is distributed as a non-central chi-square distribution with non-centrality parameter $\sum_{i=1}^v \delta_i^2$. Now:

$$\begin{aligned} \hat{G}^2 = \hat{\alpha}^2 + \hat{\beta}^2 &= \sigma_\alpha^2 (U_1 + \delta_\alpha)^2 + \sigma_\beta^2 (U_2 + \delta_\beta)^2 \\ &= \sigma_\alpha^2 \cdot {\chi_1}^2(\delta_\alpha^2) + \sigma_\beta^2 \cdot {\chi_1}^2(\delta_\beta^2) \end{aligned} \tag{3}$$

where: $\delta_\alpha = \frac{\mu_\alpha}{\sigma_\alpha}$ and $\delta_\beta = \frac{\mu_\beta}{\sigma_\beta}$. Now, $(U_1 + \delta_\alpha)^2$ and $(U_2 + \delta_\beta)^2$ are non-central chi-square distributed with noncentrality parameters δ_α^2 and δ_β^2 and 1 degree of freedom. The distribution of the sum of two non-central

chi-square distributed random variables is also non-central chi-square distributed. Press [13] has shown that the distribution of linear functions of independent non-central chi-square variates with positive coefficients can be expressed as mixtures of distributions of central chi-square's. In addition, the non-central chi-square distribution is a special case of a type I Bessel function distribution and the more general form of the distribution of a linear function of Bessel independent random variables is derived in Springer [14].

We have shown that the distribution for the gradient can be derived from first principles. In the situation where the input noise is additive zero mean Gaussian noise we have shown that the ratio $G^2/{\sigma_\alpha}^2$ is a non-central chi-square distribution. with 2 degrees of freedom and non-centrality parameter $C = ({\mu_\alpha}^2+{\mu_\beta}^2)/{\sigma_\alpha}^2$.

Now, the gradient edge detection process labels a pixel as an edge if $\hat{G}^2 > T^2$. Hence a pixel is labeled as an edge pixel if $\hat{G}^2/{\sigma_\alpha}^2 > T^2/{\sigma_\alpha}^2$. But ${\sigma_\alpha}^2 = \sigma^2/\Sigma_r\Sigma_c r^2$ and $\hat{G}^2/{\sigma_\alpha}^2$ is distributed like a ${\chi_2}^2(C)$. Therefore the probability of detecting the edge can be given by

$$P(\hat{G}^2 > T^2) = Prob\left({\chi_2}^2(C) > \frac{T^2\Sigma_r\Sigma_c r^2}{\sigma^2}\right) \tag{4}$$

and the probability of midsdetecting the edge is given by

$$P_{misdetection} = Prob\left({\chi_2}^2(C) < \frac{T^2\Sigma_r\Sigma_c r^2}{\sigma^2}\right) \tag{5}$$

.

2.3 Probability of False alarm at the edge detector output

We assume that the input data at the edge detection step is a region of constant gray tone values with additive Gaussian noise. Since a pixel is labelled an edge pixel if the estimated gradient value, G, is greater than a specified threshold, T, the probability of false detection is $Prob(G > T)$. The coefficients α and β of the facet model described in chapter 2 are normally distributed with zero mean. If the input noise variance is σ^2 then the variance of α, ${\sigma_\alpha}^2$ is equal to:

$$\sigma^2/\sum_r\sum_c r^2. \tag{6}$$

The variance of β, ${\sigma_\beta}^2$, is equal to:

$$\sigma^2/\sum_r\sum_c c^2. \tag{7}$$

Note that the summations are done over the index set for r and c. Since $G^2 = \alpha^2 + \beta^2$, if we assume a square neighborhood then the $G^2/{\sigma_\alpha}^2$ is chi-square distributed with 2 degrees of freedom. So the probability of labelling of a noise pixel as an edge pixel can be computed once we know the variance for the parameter α. Specifically, the probability of false alarm is given by:

$$P_{falsealarm} = Prob\left({\chi_2}^2 > \frac{T^2\sum_r\sum_c r^2}{\sigma^2}\right) \tag{8}$$

Note that only when the operator uses a square neighborhood the estimates of the variances for α and β are equal. The above simplification is possible only under this condition. On the other hand when a rectangular neighborhood is used the only difference is that G^2 is distributed as a linear combination of two chi-square distributed random variables.

2.4 Analysis of edge operator with hysteresis thresholds

In this section we show how the above analysis can be used to derive expressions for the false alarm and misdetection probabilities when the hysteresis linking idea of Canny [2] is used. Canny uses two thresholds:

- a high gradient threshold, T_1, to mark potential edge candidates, and
- a low gradient threshold, T_2, that assigns edge label to pixels if there exists at least one pixel in the pixel's neighborhood that has gradient magnitude greater than T_1.

More formally:

$$\begin{aligned} O(r,c) &= 1 \text{ if } G(r,c) > T_1 \text{ or} \\ &\quad \text{if } G(r,c) > T_2 \text{ and } \exists (R,C) \in N_{r,c} \ni G(R,C) > T_1. \\ &= 0 \text{ elsewhere} \end{aligned} \tag{9}$$

Let F_g denote the cumulative distribution function for the gradient magnitude. Let W denote the number of pixels in the neighborhood. Then the probability of labelling a pixel as an edge pixel is given by :

$$P(edge) = 1 - F_g(T_1) + ((F_g(T_1) - F_g(T_2))(1 - \{F_g(T_1)\}^{W-1})) \tag{10}$$

The term $1 - F_g(T_1)$ is the probability that the gradient magnitude is greater than T_1. The rest of the term is the probability that the current pixel being examined has a gradient magnitude between T_2 and T_1 and there exists at least another pixel with gradient value greater than T_1. Here we assume that the candidate pixels considered in the window have similar orientation estimates. That is, their edge orientation estimates are close to each other. One can relax this assumption and include the effects of the noise on the orientation estimate. The cumulative distribution would then be on two variables, the orientation and gradient magnitude.

2.4.1 Probability of misdetection

Using equation (10) we can write the expression for the probability of misdetection (when hysteresis linking is used) as:

$$P_{misdetection} = q_h = F_g(T_2) + (F_g(T_1) - F_g(T_2))F_g(T_1)^{W-1} \tag{11}$$

A glance at the above expression indicates that this probability is going be smaller than the misdetection probability for an edge operator with a single gradient threshold. The probability of misdetection, when hysteresis linking is not used, is given by $F_g(T_1)$. Since T_2 is much less than T_1, $F_g(T_2)$ is less than $F_g(T_1)$. The second term in the above expression can be atmost equal to $F_g(T_1) - F_g(T_2)$. Hence $q_h <= F_g(T_1)$.

2.4.2 Probability of false alarm

Using equation (10) we can write the expression for the probability of false alarm as:

$$P_{falsealarm} = 1 - F_{gf}(T_1) + ((F_{gf}(T_1) - F_{gf}(T_2))(1 - \{F_{gf}(T_1)\}^{W-1})) \tag{12}$$

where F_{gf} is the cumulative distribution function for the gradient magnitude when the input data is a flat graytone surface with additive noise. The above expression indicates that the probability of false alarm is higher than when a single gradient threshold of T_1 is used.

3 Analysis of Morphological edge detectors

In the previous section we derived the expressions for the probability of misdetection and probability of false alarm for conventional gradient based edge detectors. We turn our attention to morphological edge detectors. Specifically, we derive the expression for the probability of false alarm and the probability of misdetection for the blur-min edge detector given in Lee et al [10].

3.1 Blur-min edge detector – Description

We first provide the definitions of grayscale dilation and grayscale erosion and give the details of the blur-min edge detector. The dilation of a gray-scale image $f(r,c)$ by a grayscale structuring element b is denoted by d and is defined by:

$$d(r,c) = max_{i,j}(f(r-i,c-j)+b(i,j)) \tag{13}$$

where the maximum is taken over all (i,j) in the domain of b such that $(r-i,c-j)$ is in the domain of f. The domain of the result is the dilation of the domains of f and b.

The erosion of a gray-scale image f by a structuring element b is denoted by e and is given by:

$$e(r,c) = min_{i,j}(f(r+i,c+j)-b(i,j)) \tag{14}$$

where the minimum is taken over all (i,j) in the domain of b. The domain of e is the domain of f eroded by the domain of b.

In the Blur-minimum morphological edge detector a pixel is assigned an edge label if the edge strength computed is above a given threshold T. The edge strength I_e is given by the equation:

$$I_e = min\{I_1 - erosion(I), dilation(I_1) - I_1\}. \tag{15}$$

The operation of the blur-minimum edge detector can be illustrated by a simple example. Consider an one-dimensional step edge sequence and let the neighborhood size for the blur, dilation and erosion operations be 3. That is, the erosion and dilation use a flat structuring element with domain $\{-1,0,1\}$. We use the following notation in the example: o - original data, b - blurred original, e - erosion of original, and d - dilation of original.

o	0	0	0	E	E	E
b	0	0	$\frac{E}{3}$	$\frac{2E}{3}$	E	E
e(b)	0	0	0	$\frac{E}{3}$	$\frac{2E}{3}$	E
d(b)	0	$\frac{E}{3}$	$\frac{2E}{3}$	E	E	E
b-e(b)	0	0	$\frac{E}{3}$	$\frac{E}{3}$	$\frac{E}{3}$	0
d(b)-b	0	$\frac{E}{3}$	$\frac{E}{3}$	$\frac{E}{3}$	0	0
min(b-e(b),d(b)-b)	0	0	$\frac{E}{3}$	$\frac{E}{3}$	0	0

It can be seen that if the threshold chosen is $\frac{E}{3}$ then the third and fourth pixels in the one-dimensional sequence will get labelled as edge pixels.

3.2 Distributions of Grayvalues in Dilation and Erosion Residues

We now derive the cdf for the grayvalues in the dilation-residue and the erosion residue. Let M be the number of pixels in the neighborhood. Let Z_i denote the grayvalue at the i'th pixel location before the blurring step. Let X_i denote the grayvalue at the i'th pixel location after the blurring step. The blurring step in the blur-minimum edge detector introduces correlation between neighboring pixels' grayvalues. In the derivations that follow the false alarm and misdetection characteristics for the blur-minimum edge detector is obtained by ignoring the effects of correlation on the additive noise. If Z_i's were i.i.d Gaussian random variables with zero mean and variance σ_z^2, then the X_i's are Gaussian distributed with zero mean and variance σ_z^2/M. However there exists correlation between neighboring X_i's. The probability of false alarm obtained by assuming that the X_i's are independent with zero mean and variance σ_z^2/M is actually an upper bound for the probability of false alarm obtained by taking correlation into account. We do not rigorously prove this claim, but provide an explanation for this in the appendix.

3.2.1 Dilation, Erosion Residues and Edge Strength

The dilation residue at the jth location is given by:

$$max_i(X_i) - X_j = max\{X_1 - X_j, X_2 - X_j, \ldots, X_M - X_j\} \tag{16}$$

Note that the *max* operation is over all elements in the edge detector's neighborhood. We can see that each term that is an argument to the *max* operator is distributed as the difference of two Gaussian random variables. Let: $Y_i = X_i - X_j$. Then the dilation residue at pixel j, $Y_d(j)$, can be written as:

$$Y_d(j) = max_{i,i\neq j}\{max\{Y_i\}, 0\}. \tag{17}$$

Similarly the erosion residue can be written as:

$$Y_e(j) = -min_{i,i\neq j}\{min\{Y_i\}, 0\}. \tag{18}$$

The output edge-strength O is given by:

$$O(j) = min(Y_d(j), Y_e(j)). \tag{19}$$

In order to obtain the cumulative distribution function for O, we first derive the expressions for the distributions of maxima and minima of Y_i's and then obtain expressions for the distributions of Y_d and Y_e. In these derivations we assume that the Y_i's are independent random variables, but they need not be identically distributed.

3.2.2 Distribution of maxima and minima of Y_i's

The cumulative distribution function of the minimum of Y_i's (the first order statistic $Y_{(1)}$.) is given by:

$$\begin{aligned} Prob(min(Y_i, i = 1, \ldots, M-1) \leq x) &= Prob(Y_{(1)} \leq x) \\ &= 1 - Prob(Y_i > x, i = 1, \ldots M-1) \end{aligned}$$

If the Y_i's are independent and identically distributed with cdf F_y, we have:

$$\begin{aligned} F_{y_{(1)}}(x) &= Prob(min(Y_i, i = 1, \ldots, M-1) \leq x) \\ &= 1 - \{Prob(Y_i > x)\}^{M-1} \\ &= 1 - \{1 - F_y(x)\}^{M-1} \end{aligned} \tag{20}$$

The cumulative distribution function of the maximum of i.i.d. Y_i's (the highest order statistic $Y_{(M-1)}$) is given by:

$$\begin{aligned} F_{y_{(m-1)}}(x) &= Prob(max(Y_i, i = 1, \ldots, M-1) \leq x) \\ &= Prob(Y_{(M-1)} \leq x) \\ &= Prob(Y_i \leq x, i = 1, \ldots M-1) \\ &= F_y(x)^{M-1} \end{aligned} \tag{21}$$

When the Y_i's are independent, but not identically distributed, then the cdf of the minimum value of Y_i's is given by:

$$F_{y_{(1)}}(x) = 1 - \prod_{i=1}^{M-1}(1 - F_{y_i}) \tag{22}$$

where F_{y_i} is the cdf for Y_i.
Similarly, the cdf of the maximum of Y_i's is given by:

$$F_{y_{(M-1)}}(x) = \prod_{i=1}^{M-1} F_{y_i}. \tag{23}$$

3.2.3 Distribution of Y_d and Y_e

We know that: $Y_e = -min\{Y_{(1)}, 0\}$. Hence Y_e can be written as:

$$\begin{aligned} Y_e &= -Y_{(1)} \qquad \text{if } Y_{(1)} \leq 0 \\ &= 0 \qquad \text{elsewhere} \end{aligned} \tag{24}$$

The cdf of Y_e is given by the following expression when the Y_i's are i.i.d random variables:

$$\begin{aligned} F_{Y_e}(x) &= (1 - F_y(-x))^{M-1} \qquad x > 0 \\ &= 1 - (1 - F_y(0))^{M-1} \qquad x = 0 \end{aligned} \tag{25}$$

When the Y_i's are independent but not identically distributed: the cdf of Y_e is given by:

$$\begin{aligned} F_{Y_e}(x) &= \prod_{i=1}^{M-1} (1 - F_{y_i}(-x)) \qquad x > 0 \\ &= 1 - \prod_{i=1}^{M-1} (1 - F_{y_i}(0)) \qquad x = 0 \end{aligned} \tag{26}$$

Similarly, $Y_d = max\{Y_{(M-1)}, 0\}$. Hence Y_d can be written as:

$$\begin{aligned} Y_d &= Y_{(M-1)} \qquad \text{if } Y_{(M-1)} > 0 \\ &= 0 \qquad \text{elsewhere} \end{aligned} \tag{27}$$

Then the cdf of Y_d, for i.i.d Y_i's, is given by:

$$F_{Y_d}(x) = F_y(x)^{M-1} \qquad x \geq 0 \tag{28}$$

When Y_i's are independent, but not identically distributed, F_{Y_d} is given by:

$$F_{Y_d}(x) = \prod_{i=1}^{M-1} F_{y_i}(x) \qquad x \geq 0 \tag{29}$$

3.3 Distribution of Edge Strength

The output edge strength is given by $O = min(Y_d, Y_e)$. The cdf for O can be easily obtained if Y_d and Y_e were independent. In order to derive the expression for the cdf of O, we rewrite the expression for O as:

$$O = max(min(-Y_{(1)}, Y_{(M-1)}), 0). \tag{30}$$

This is done in order to bring out the fact that the distribution of edge strength is dependent on the joint distribution of the min and max of the samples in the detector window. The joint density of the minimum and maximum of the Y_i's can be written (for i.i.d Y_i's) as:

$$f(x_1, x_2) = (M-1)(M-2) f_y(x_1) f_y(x_2) [F_y(x_2) - F_y(x_1)]^{M-3} \qquad x_1 \leq x_2 \tag{31}$$

The cdf for O (for iid Y_i's) is then given by the following expression:

$$F_O(x) = 1 - \int_{-\infty}^{-x} \int_{x}^{\infty} f(x_1, x_2) dx_1 dx_2 \tag{32}$$

$$= (1 - F_y(-x))^{M-1} + ({F_y}^{M-1}(x) - (F_y(x) - F_y(-x))^{M-1}) \qquad x > 0 \tag{33}$$

The cdf for O for independent, but not identical Y_i's, is given by the expression (32) with $f(x_1, x_2)$ equal to:

$$f(x_1, x_2) = \sum_{(i,j), i \neq j} f_{y_i}(x_1) f_{y_j}(x_2) \prod_{k=1, k \neq i, k \neq j}^{M-1} F_{y_k}(x_2) - F_{y_k}(x_1) \tag{34}$$

3.4 Probability of False alarm

In this section, we use the results of the previous section(s) to derive the expression for the probability that a non-edge pixel gets labelled as an edge pixel. We assume that the input to the detector is a sequence of i.i.d Gaussian samples with zero mean and variance σ_z^2.

The probability of falsely labelling a noise pixel as an edge pixel is given by: $p = 1 - F_O(T)$, where T is the edge strength threshold used, where $F_O(T)$ is the cumulative distribution function for the edge strength obtained when the Y_i's are Gaussian with zero mean and variance $2\sigma^2$, where $\sigma^2 = \sigma_z^2/M$.

3.5 Probability of Misdetection

Using the results obtained in previous sections, we can derive the expression for the probability of misdetection of the blur-min edge detector. Let the true gradient value (the slope of the ramp) be G_t. Let the neighborhood size be M pixels. Assuming that the slope spans the entire neighborhood the gray values in the ramp can be written as: $I_1 + ((G_t)i), i = 0, \ldots, M-1$, where I_1 is the gray value in the left most pixel in the window. We assume that the image values are corrupted with additive Gaussian noise with zero mean and variance σ_z^2. We can see that the X_i's in section (3.2) are nothing but Gaussian random variables with mean $I_1 + (G_t i)$ and variance σ^2. The differences of X_i's, Y_i's, are also Gaussian random variables with means, μ_i, and variance $2\sigma^2$. Here the μ_i's are specified by the sequence: $G_t(M-1)/2, G_t(M-2)/2, \ldots, G_t, -G_t, \ldots, -G_t(M-2)/2, -G_t(M-1)/2$. Note that there are only $M-1$ Y_i's, because the difference from the center pixel is zero. Since the Y_i's are independent, we can use equations (22) and (23) to obtain the cdf's for the minimum and maximum of Y_i's. The derivation for the cdf of the output edge strength is analogous to the derivation in section (3.2). If F_O is the cdf for the output edge strength, the probability of misdetection, when a threshold T is used, is given by:

$$P_{misdetection} = F_O(T) \tag{35}$$

As was seen in section (3.2) we need to numerically integrate $f(x_1, x_2)$, the joint pdf of the minimum and maximum of the random variables, in order to compute the above probability. When the noise standard deviation σ is less than $G_t/4$, we can use the following approximation for computing the distribution of the edge strength:

$$Prob(O \leq x) = 1 - Prob(Z_1 > x, Z_2 > x) \tag{36}$$

where Z_1 and Z_2 are independent random variables with cdf's given $F_Z(x)$ by:

$$Prob(Z \leq x) = 1 - \prod_{i=1}^{(M-1)/2} F_{y_i}(x) \tag{37}$$

Here Z_1 and Z_2 are the dilation and erosion residues. Hence the edge strength distribution is given by:

$$Prob(O \leq x) = 1 - (1 - F_z(x))^2 \tag{38}$$

The above simplification is possible because modes of the density functions for the Mth sample and the first sample are well separated and the *min* and the *max* of the samples may be considered to be independent of each other.

When σ is comparable to G_t then equation (32) will have to be used and the probability has to be computed by numerical integration.

4 Theoretical Plots and Comparison

In this section we plot the theoretical and empirical false alarm and misdetection curves for the edge detectors evaluated. We plot $P_{misdetection}$ and $P_{falsealarm}$ against gradient or edge strength threshold T, for various

noise variances σ_z^2. We also plot the theoretical operating curve $P_{misdetection}$ vs $P_{falsealarm}$ for these edge detectors. The theoretical plots were obtained by varying/setting the values for the parameters in the ranges specified below:

- Ideal edge gradient values – 0, 3, 3.5, 4, 8, 10, 100
- Noise variance values – 1, 10, 25, 100
- Gradient threshold values – 0.1 to 50.0
- Edge operator window size – 5 x 5, 3 x 3
- hysteresis threshold value – $T_2 = 0.5 * T_1$.

From the theoretical analysis, it is clear that gradient edge detection with hysteresis linking is superior to gradient edge detection without hysteresis linking. As has been pointed out by Hancock and Kittler [4], the output from the hysteresis linking algorithm may consist of short segments obtained due to correlated noise. For a given threshold T_1 the probability of false alarm with hysteresis linking is higher. This is expected since we are admitting more pixels to be edge pixels based on contextual information. However, if T_1 is sufficiently large, the false alarm obtained with hysteresis linking is comparable to that obtained without hysteresis linking. The plots obtained confirm the above points.

Figure 1(a) gives the theoretical false alarm vs misdetection plot for a conventional gradient based edge detector. The graylevel noise variance was set at 25, and a 5 by 5 window size was used. The true edge slope was varied from 2 to 5. In a normal image only a fraction of the pixels are true edge pixels and hence the absolute count of the number of pixels falsely labelled as edge pixels would be quite high. Figure 1(b) gives the theoretical false alarm vs misdetection plot for the blur-minimum edge operator. Figure 1(c) gives the plots comparing the detectors for an edge slope of 4.0, noise variance of 25 and window size of 5 by 5. It can be seen that when the edge slope is equal to 4, the false alarm rate of the gradient based edge detector, for a misdetection rate of 10% is approximately 2 percent. It can be seen that for a misdetection rate of 10%, the corresponding false alarm rate for the blur-minimum edge detector when correlation effects (due to blurring) are not considered is as much as 18%. But this value can be shown to be an upperbound for a range of edge strength threshold values (Please see appendix A). Hence no quantitative statement about the false alarm rate for the blur-minimum edge operator can be made. Figure 1(d) shows the comparison of the edge detectors when a window size of 3 by 3 is used. It can be seen that the gradient based detector performs the worst. An intuitive explanation for this is as follows: the blurmin operator uses "min" and "max" to estimate the edge strength and as the sample size grows (i.e. the window size is bigger) the estimates for the min and max will tend towards $-\infty$ and ∞. Thus the estimated edge strength will be large, even though the noise variance is not as high. On the other hand with the gradient based scheme, a larger the window size implies better fit and the lesser variance for the estimates. Hence one would expect the false alarm probability to decrease with increasing window size for the gradient based operator, whereas the probability would increase (if one ignores effects of correlation) with increasing window size for the morphological operator.

In order to verify our theory we generated test ramp images with varying levels of additive noise and plotted the false alarm vs misdetection characteristics. One such plot obtained, when the true edge slope was 4.0 and the input noise variance was 100, is given in figure 2(a). Figure 2(a) shows that the performance of the blur-minimum operator is the best, followed by the gradient based edge operator. The performance of the blur-minimum operator (when noise is not correlated) is poor as predicted by theory.

From the expressions for the false alarm and misdetection probabilities for the morphological edge detector, it is not easy to infer whether it is superior to the other detectors. Since the theoretical false alarm characteristics only gives an overestimate on the probability of false alarm for the blur-minimum operator, one cannot say anything about its superiority over the gradient based edge detection scheme. We are in the process of computing the edge strength distribution, by taking correlation effects (introduced during blurring) into account. Currently, our theoretical model assumes that the output after blurring is a ramp edge with additive Gaussian noise. The added noise at each pixel is assumed to be i.i.d Gaussian samples. We will address the effects of correlation between neighboring pixels in a subsequent paper.

Figure 3 illustrates the results obtained by applying the edge detectors on the Dr.Einstein image and the brain image. The operator window size was set at 3 by 3 and the threshold varied. One can view both the output from both edge detectors as estimates of the edge strength and the fundamental difference between the two detectors is the way in which the edge gradient estimation is done. The morphological estimates the edge gradient by a non-linear technique whereas the conventional least squares fitting method uses a linear filter. In order to evaluate the detection schemes with real images we fix a particular threshold, T, for the blur-minimum edge operator and then vary, T_g, the threshold for the gradient based edge operator until the same degree of false alarm (due to texture in the data) was obtained. It can be seen from figure 2 that the blur-minimum edge operator output captures more of the structure in the brain image. Also, the edges obtained by using the blur-min operator are thinner. A thorough evaluation of the performance of the operators on real images can be done if the images were obtained by controlled experiments and if ground truth information were available.

5 Conclusion

In this paper we gave a theoretical evaluation of the performance of gradient based edge detectors and morphological edge detectors. The performance analysis was done by assuming an ideal edge model and a noise model and by deriving expressions for probability of false alarm and probability of misdetection of edge pixels. Under the Gaussian noise model assumption, the theory indicates that the blur-min edge detector is superior to gradient based edge detectors when a 3 by 3 window is used. The empirical plots indicate that the blur-min edge operator is superior when a 5 by 5 window is used, however the theoretical plots do not confirm this because the theory just provide an upperbound. We also see that hysteresis linking significantly improves the misdetection rate, but the false alarm rate increases for a given gradient threshold. In our analysis of hysteresis linking we assumed that the candidate pixels for the linking operator were grouped based on similar orientation. The derivations for the probability distribution of the orientation estimate is given in [9]. We can use those results in conjunction with our results to study the effects of the grouping threshold and the gradient thresholds T_1 and T_2..

References

[1] I.E.Abdou and W.K.Pratt, "Quantitative Design and Evaluation of Enhancement/ Thresholding Edge detectors," Proceedings of the IEEE, Vol. 67, pp. 753 - 763, May 1979.

[2] F.J.Canny, "Finding edges and lines in images," Tech.Rep. 720, MIT AI Lab, June 1983.

[3] H.A.David, *Order Statistics*, John Wiley & Sons, New York, 1970.

[4] E.R.Hancock and J.Kittler, "Adaptive Estimation of Hysteresis Thresholds," Proceedings of CVPR, 1991, pp. 196 - 201.

[5] Haralick, R.M., *"Edge and Region Analysis for Digital Image Data,"* Computer Graphics and Image Processing, Vol.12, 1980, pp. 60-73.

[6] R.M.Haralick, *"Digital step edges from zero crossing of second directional derivatives,"* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, pp. 58-68, Jan. 1984.

[7] R.M.Haralick, "Second directional derivative zero-crossing detector using the cubic facet model," Image Analysis, Vol.1, Proceedings, Norway, June 1985.

[8] Haralick, R.M., and L.G.Shapiro, *Computer and Robot Vision*, Vol.1, Reading, MA: Addison-Wesley, 1991.

[9] R.M.Haralick, V.Ramesh, and K.Thornton, "Experimental Protocol in Computer Vision," Unpublished manuscript.

[10] J.S.Lee, R.M.Haralick, and L.G.Shapiro, "Morphologic Edge Detection," IEEE Journal of Robotics and Automation, Vol. RA-3, No.2, April 1987.

[11] D.Marr and E.Hildreth, *"Theory of edge detection,"* Proceedings of Royal Society of London, B, Vol. 207, pp. 187-217, 1980.

[12] V.S.Nalwa and T.O.Binford, "On Detecting Edges," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No. 6, pp. 699 - 714, Nov. 1986.

[13] Press, S.J., *"Linear combinations of non-central chi-square variates,"* Annals of Mathematical Statistics, Vol. 37, pp. 480-487.

[14] Springer, M.D., *The Algebra of Random Variables*, New York: John Wiley and Sons, 1979.

A Upper bound for the Probability of false alarm of Blur-min operator

In our derivation of the cdf of the edge strength, we stated that we ignored the effects of noise correlation introduced by blurring. Assuming that the X_i's, inputs to the dilation and erosion operations, are independent random variables, we derived the probability of false alarm of the morphological operator. Here we show how the probability of false alarm obtained is an upper bound of the probability that will be obtained when the correlation effects are considered.

Rather than giving a general proof we take a one-dimensional case and illustrate why this is true. Let the number of pixels in the operator's window be M (assume 2L + 1 = M). Let Z_i's be independent Gaussian random variables with zero mean and variance σ_z^2 The X_i's may then be written as:

$$X_i = \frac{1}{M} \sum_{m=-L+i}^{L+i} Z_m \tag{39}$$

Then $X_{i+k} - X_i$, $0 \leq k \leq L$ is given by:

$$X_{i+k} - X_i = \frac{1}{M}\{ \sum_{m=L+i+1}^{L+i+k} Z_m - \sum_{l=-L+i}^{-L+i+k-1} Z_l\} \tag{40}$$

When the edge operator window is centered on pixel locations i and $i+k$ the windows overlap. The above expression just states that the difference between grayvalues at two pixels k distance apart is given by the difference between the sums of the Z_i's in the right and left non-overlapping areas. It can be seen that the differences,$(X_i - X_j)$'s (in the operator's neighborhood) are given by:

$$0, \quad Y_1, \quad Y_1 + Y_2, \ldots, \quad Y_1 + Y_2 + \ldots + Y_L, \quad Y_{-1}, \quad Y_{-1} + Y_{-2}, \quad Y_{-1} + Y_{-2} + \ldots Y_{-L} \tag{41}$$

where the Y_i's are independent Gaussian random variables with zero mean and variance $2\sigma_z^2/M^2$ (denote density by $f_y(x)$). To illustrate how these terms were obtained, let us consider Y_1. In this case, the operator windows centered on pixel locations $i+1$ and i have $M-1$ pixels in common and the difference $\frac{Z_{L+i+1} - Z_{-L+i}}{M}$ is equal to Y_1. In general we set:

$$Y_j = \frac{Z_{L+i+j} - Z_{-L+i+j-1}}{M} \tag{42}$$

and the differences $X_{i+k} - X_i$ are equal to:

$$\sum_{j=1}^{k} Y_j \qquad k > 0$$
$$\sum_{j=-k}^{-1} Y_j \qquad k < 0 \tag{43}$$

The cdf for the edge strength can then be obtained by using the joint distribution of the *min* and *max* of the random variables given in the equation (41).

Let us consider a window that is 5 pixels wide. Then the differences $X_{i+k} - X_i$ for $k = -2, -1, 0, 1, 2$ are given by:

$$Y_{-1}, \quad Y_{-1} + Y_{-2}, \quad 0, \quad Y_1, \quad Y_1 + Y_2 \tag{44}$$

The random variables with the negative subscripts are independent of the ones with the positive subscripts. Hence we can think of computing the edge strength in two stages. First compute the edge strength O_1 that corresponds to $min(-min(Y_{-1}, Y_{-1} + Y_{-2}, 0), max(Y_{-1}, Y_{-1} + Y_{-2}, 0))$ and then compute edge strength O_2 that corresponds to $min(-min(Y_1, Y_1 + Y_2, 0)$,$max(Y_1, Y_1 + Y_2, 0))$. We can then take the minimum of the two results to obtain O.

The joint density of the random variables Y_1 and $Y_2 + Y_1$ is given by a multivariate normal distribution, $g(x_1, x_2)$, with zero mean and covariance matrix $\sum$. The diagonal entries are equal to $2\sigma_z^2/M^2$ and $4\sigma_z/M^2$ respectively and the non-diagonal entries are equal to the covariances, $2\sigma_z^2/M^2$. The $Prob(O = 0)$, the probability that the estimate of the edge strength is equal to zero, is given by the integral of $g(x_1, x_2)$ with limits for x_1 and x_2 going from 0 to ∞. This integral can be shown to be greater than 0.5. When correlation effects are ignored, the probability is 0.5. In addition, it can also be seen that the joint density function (when correlation is used) for the minimum and maximum of the random variables Y_1 and $Y_1 + Y_2$ is given by:

$$f(x_1, x_2) = \frac{M^2}{4\pi\sigma_z^2} exp\left(\frac{-(x_2 - x_1)^2 M^2}{4\sigma_z^2}\right)\left[exp\left(\frac{-(x_1)^2 M^2}{4\sigma_z^2}\right) + exp\left(\frac{-(x_2)^2 M^2}{4\sigma_z^2}\right)\right] \tag{45}$$

This function decreases rapidly, compared to the joint density function for uncorrelated case, as x_1 and x_2 increase. Hence the cdf: $Prob(O <= T)$ for a given threshold is greater than, $Prob(O_u <= T)$, the probability obtained by making the uncorrelated assumption. The probability, $P(O > T)$, is given by: $P(O_1 > T)P(O_2 > T)$. Since we have upper bounds for each of these probabilities, we have an upper bound for the final probability desired. This final probability is nothing but the probability of false alarm. This illustrates how the theoretical expression for the false alarm probability obtained by ignoring correlation effects is an upper bound of the actual false alarm probability.

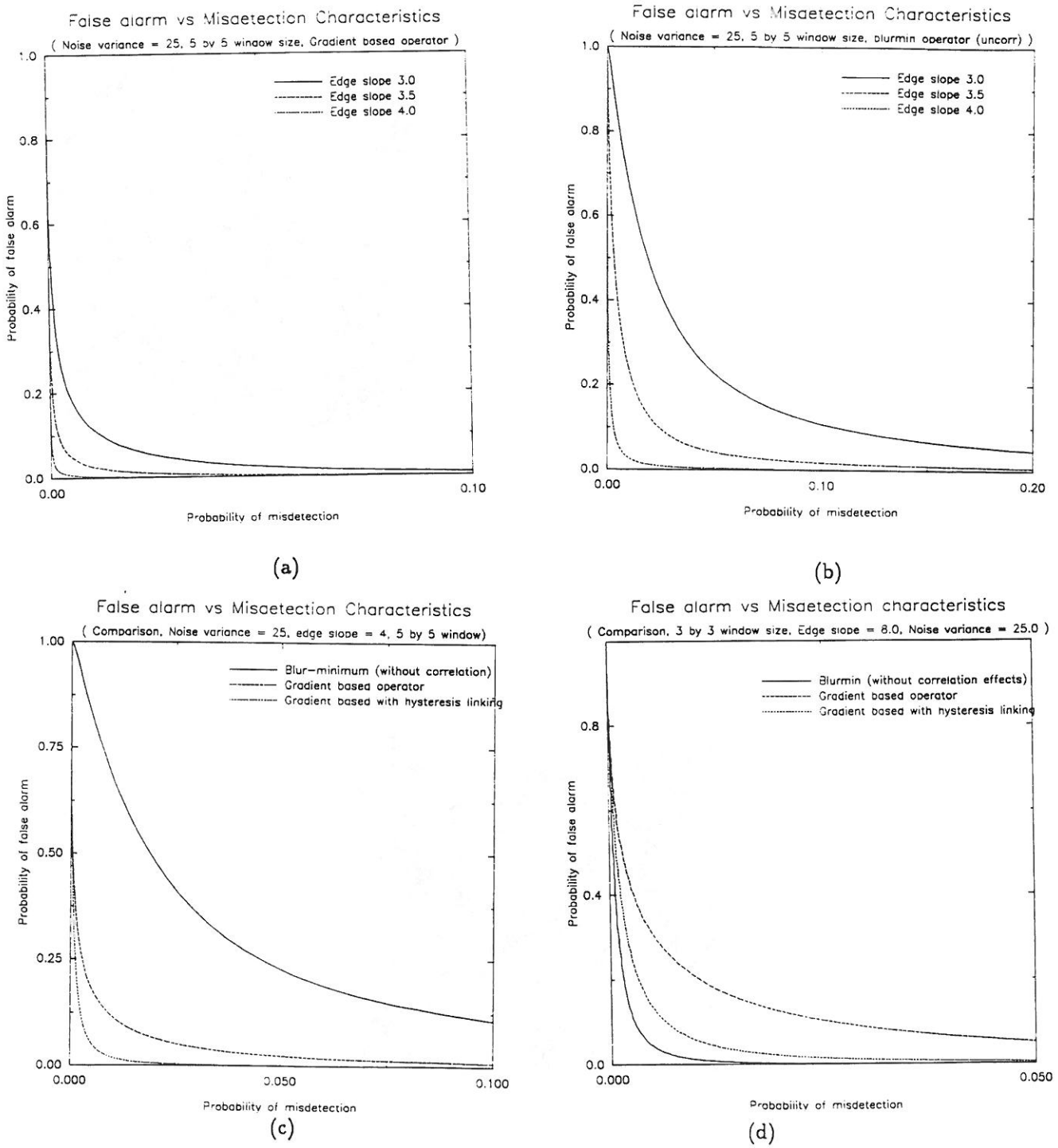
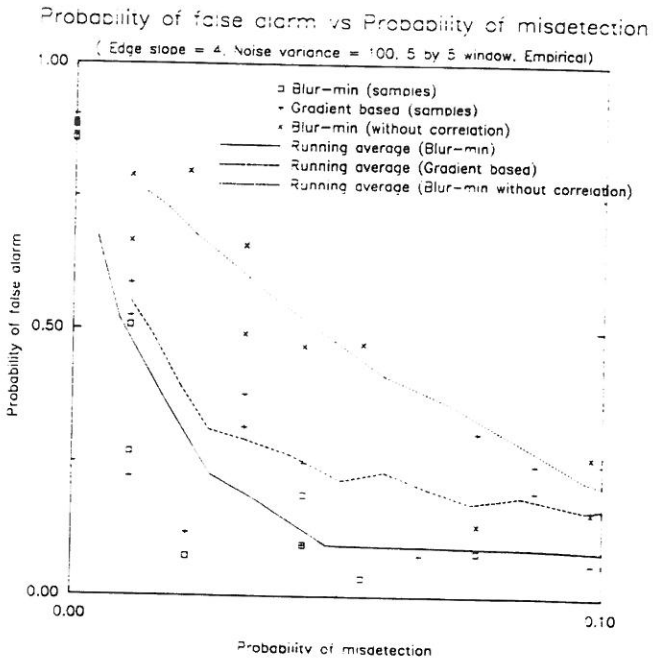
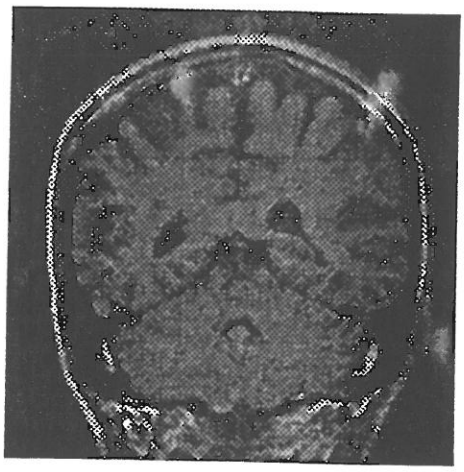


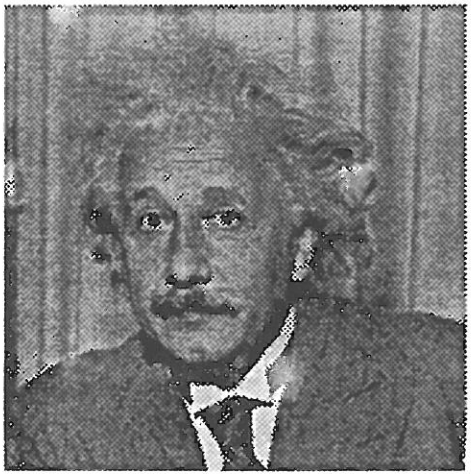
Figure 1: **False alarm vs Misdetection plots:** (a) gradient based operator, (b) blur-min operator (without correlation effects), (c) comparison and (d) comparison (for 3 by 3 window size)



(a)

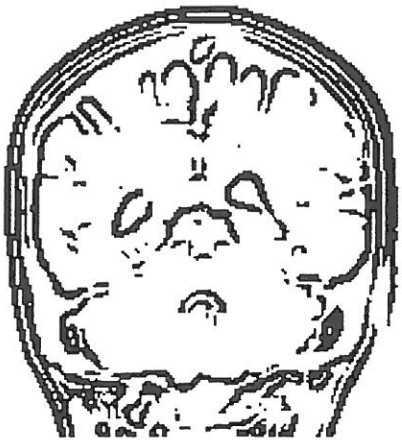


(b)

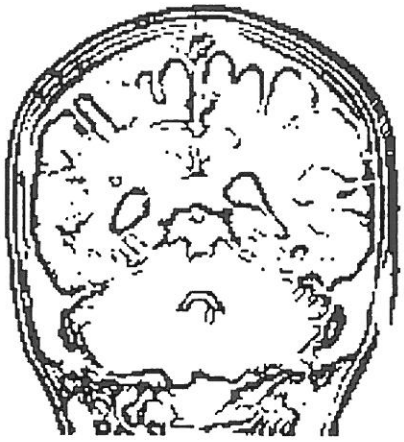


(c)

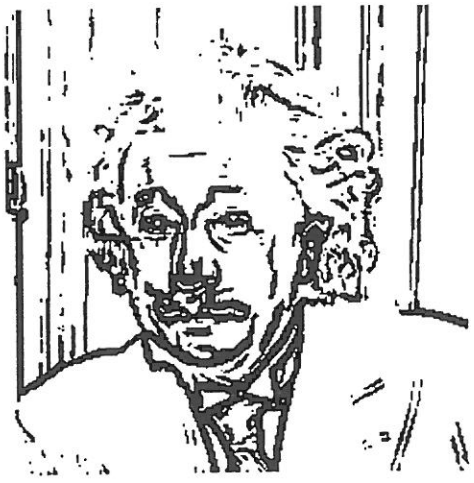
Figure 2: Empirical results on synthetic data and example real images used: (a) Empirical results, (b) Image 1 and (c) Image 2



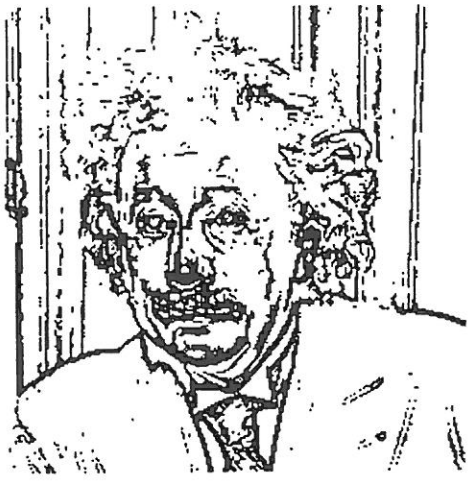
(a)



(b)



(c)



(d)

Figure 3: Results with real images: (a) gradient based operator, 3 by 3 window, T = 7, Image 1 (b) blur-min operator (3 by 3 window, T=7), Image 1 , (c) Image 2 – gradient based operator (3 by 3 window, T=7), and (d) Image 2 – blurmin operator (3 by 3 window, T=7)