Purdue University Purdue e-Pubs

ECE Technical Reports

Electrical and Computer Engineering

8-1-1996

Automatic Gradient Threshold Determination for Edge Detection Using a Statistical Model A Description of the Model and Comparison of Algorithms

Peter V. Henstock, Purdue University School of Electrical and Computer Engineering

David M. Chelberg Purdue University School of Electrical and Computer Engineering

Follow this and additional works at: http://docs.lib.purdue.edu/ecetr

Henstock, Peter V. and Chelberg, David M., "Automatic Gradient Threshold Determination for Edge Detection Using a Statistical Model A Description of the Model and Comparison of Algorithms" (1996). *ECE Technical Reports.* Paper 95. http://docs.lib.purdue.edu/ecetr/95

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Automatic Gradient Threshold Determination For Edge Detection Using a Statistical Model: A Description of the Model and Comparison of Algorithms

PETER V. HENSTOCK DAVID M. CHELBERG

TR-ECE 96-13 August 1996



School of Electrical and Computer Engineering Purdue University West Lafayette, Indiana 47907-1285 Automatic Gradient Threshold Determination for Edge Detection Using a Statistical Model A Description of the Model and Comparison of Algorithms

Peter V. Henstock, David M. Chelberg

School of Electrical Engineering Purdue University West Lafayette, Indiana 47907-1285

This work was supported in part by NSF Grant No. IRI-9011421.

Contents

Ι	Intr	roduction	1
11	Hist A B	t ogram Modeling Background	3 3 4
II	Des	cription of the Different Methods	6
	А	Global Descent	7
	В	Two Step Estimation	7
		B.1 Alpha-Beta Estimation	8
		B.2 Percentage Estimation	10
		B.3 Threshold Determination	11
IV	' Exp	periments	12
	A	Data Preparation	12
	В	Running the Algorithms	14
	С	Testing Other Methods	15
	D	Computed and Subjective Thresholds	15
	Е	Robustness to Noise	21
\mathbf{V}	Con	nclusion	23

ii

List of Tables

1	The mean and variance of the average distance (area) between the computed			
	and subjective thresholds, averaged over the 16 images	18		
2	A list of the algorithms presented in Table 1	18		
3	The mean and variance of the area between the subjective and computed			
	thresholds of the 5 different alpha-beta estimation methods	19		
4	The mean and variance of the area between the subjective and computed			
	thresholds of the 4 different percentage estimation methods	19		
5	The mean and variance of the area between the subjective and computed			
	thresholds of the 3 different threshold estimation methods	20		

List of Figures

1	A bimodal histogram which can be easily thresholded	5
2	A typical histogram for edge thresholding with the estimated densities of the non-edge and edge pixels	5
3	With some histograms, noise is required to spread the peak near zero gradient	0
	units	12
4	A battery of 16 images of varying scenes. Moving left to right starting at the	
	top, the images are: a house scene, an xray skull image, bethl, contact, crowd,	
	dilts, a synthetic dragon cartoon, flir images of a truck, man (gih), girl2, jo,	
	john, a satellite image of the earth, Madonna quantized to a few gray levels,	
	a text scan, and a still life $(1m16)$.	16
5	The thresholded images using the best algorithm.	17
6	Images thresholded at 27, 39, 57, and 128 gradient units for subjective deci-	
	sions. The computed threshold is the second image	19
7	The percent of non-edge pixels in each of the 16 images as determined by	
	the average of the 5 researchers (solid lines) and the average of the 5 best	
	algorithms (dotted lines).	20
8	Synthetic images with added white Gaussian zero-mean noise of variances 1.0,	
	128.0, 1024.0, and 4096.0.	22
9	Type I, Type II, and overall misclassification errors for different noise levels	
	using the best algorithm.	22
10	Histograms for the synthetic image under three different added noise variances.	26
8 9	Synthetic images with added white Gaussian zero-mean noise of variances 1.0, 128.0, 1024.0, and 4096.0. Type I, Type II, and overall misclassification errors for different noise levels using the best algorithm	
10	Histograms for the synthetic image under three different added noise variances.	26

iv

Abstract

In the enhancement/thresholding method of edge detection, the gradient values of pixels exceeding a certain threshold are designated as edge pixels. However, selecting a threshold has commonly been performed through ad hoc measures. This paper describes a method for automatically selecting a threshold using a 5-parameter model. The model is based on the weighted sum of two gamma density functions corresponding to edge and non-edge pixels. A variety of statistical and fitting methods for finding the model parameters were evaluated by comparing their computed thresholds to perceptual thresholds determined by subjects for 16 different images. The performance of the model was also analyzed under different noise levels.

Index Terms: Automatic thresholding, edge detection, gradient, Sobel operator, statistical classification.

I Introduction

The goal of most object recognition systems is to identify the major features of an image and match them against given models of known objects. Shape is commonly used for identification since it is one of the most distinguishing characteristics of an object. The shape of an object is generally considered to be a high level property that can be computed from a hierarchy of various lower order structures. At the lowest level, edgels or edge pixels represent the edges in the image that are grouped into line segments and then into various curves, lines, rectangles, etc. Making accurate decisions about edgels at the lowest level will facilitate more rapid and better performance at higher levels of the system. Consequently, the edge detection process is of fundamental importance in object recognition. This paper describes a model-based method for accurately and automatically determining a threshold that separates edge pixels from non-edge pixels in intensity images.

Edges are defined as sharp intensity changes over a small area of an image. Since this general definition does not lend itself to a specific mathematical formula, many edge detection algorithms have been used. Edge detectors fall into three major classes: edge fitting operators, zero crossings of second derivative operators, and enhancement/thresholding operators [1]. The first category attempts to orient a mathematical model to the edge using a best-fit approximation. The second method attempts to find the inflection points of the edge that are the zero crossings of the second derivative operator, by definition. This paper focuses on the third category of the enhancement/thresholding methods since they are the most common in practice [2, 3].

Enhancement/thresholding methods apply differential or gradient-based operators to an image. The operator is typically a 3×3 or 5×5 matrix which is convolved with the image. The Sobel or Prewitt methods use two operators which measure the local differences in contrast in the horizontal and vertical directions and combine the results into gradient units. Areas of high contrast have a high gradient value; areas of low contrast, a low value.

A threshold of the gradient values must be used to classify each pixel of an image as either

an edge or a non-edge pixel. Consequently, even in this class of operators, the presence of an edge is still imprecise. Many differential operators have been studied in depth, yet the determination of a threshold is still very difficult since it may depend upon the application, the source of the image, and the subjective perception of the viewer.

Our approach to identify the threshold dividing edge and non-edge pixels is based upon a statistical model of the gradient values. By creating a histogram of the gradient values, we can fit a statistical density function to both the edge pixel values and non-edge pixel values. Once the densities have been found, the problem of finding the threshold is equivalent to a statistical classification problem. This paper analyzes the effectiveness of several approaches in finding the density functions and the corresponding threshold.

Studies in the evaluation of different edge detectors have contributed indirectly to work in automatic threshold determination. In [1, 4, 5], several edge characteristics were identified for comparing edge operators:

- 1. the number of correctly identified edge points
- 2. the number of incorrectly identified edge points (false positives)
- 3. the number of missed edge points (false negatives)
- 4. the number of multiple detections for a single edge, which is related to edge thickness
- 5. the distance between the identified edge and the true edge
- 6. edge continuity

After applying different operators to test images, the above characteristics were quantitatively assessed and combined into a weighted error sum, allowing for comparison. In order to compare the operators' performance, a threshold had to be selected. Venkatesh and Kitch [1] used a least-failure-measure (lmf) threshold which minimized their error vector. While qualitative measurement of an error sum has desirable characteristics, it is not clear that the minimization of the error sum will, in fact, lead to an accurate determination of the threshold. McLean and Jernigan [6] selected thresholds according to the "visual quality of the resulting edge map, not according to any objective measure of the actual operator performance". Most of the work in edge evaluation has focused on synthetic images with known edges or compared the results of the operators against themselves, rather than against an absolute standard, or a model-based criteria [5]. As a result, these methods cannot be used to select an edge threshold, but serve to validate our use of both synthetic images and subjective thresholds to evaluate the performance of the edge thresholding.

Previous work in gradient threshold selection includes the work of Haddon [7] and Zuniga and Haralick [8]. After forming an estimate of the noise within the image, Haddon derived a probability density of the edge strength from which he could select a threshold. With an accurate measure of the noise, a global threshold could be selected to be independent of the strength and number of edge pixels in the image. However, the approach assumes uniform noise and only one image demonstrates its performance.

Zuniga and Haralkick [8] approached threshold selection as a Bayesian decision problem involving two densities. Rather than fit densities to the edge and non-edge pixels, they derived two conditional densities based on the gradient histogram by using a facet model in conjunction with a hypothesis test of the gradient values. For synthetic images, they computed thresholds as a function of the noise that were better than subjective decisions, although no real images were tested.

In Section II, we survey automatic classification methods and describe our model in detail. Section III introduces the different techniques of statistical estimation used to predict the parameters of our model. Section IV describes and presents the results of our experiments comparing the statistically calculated thresholds with the subjective perception of edge thresholds, and evaluating the performance of the calculated thresholds under different noise conditions.

II Histogram Modeling

A Background

The decision between edge and non-edge pixels based upon a gradient histogram is very similar to the automatic segmentation/thresholding problem. Using grayscale or color as the distinguishing criteria, pixels can be classified automatically. In determining the threshold

for different classes, these methods rely upon the modality, shape, or moments of the classes in the histogram [9]. For the bilevel case in particular, these algorithms generally define the location of the threshold to be at the minimum value of the valley between two peaks of the histogram, which can be clearly seen in Figure 1. As Kapur, Sahoo, and Wong [10] point out, different algorithms try to improve the histograms by making the threshold more visible through statistical or other enhancement techniques.

Due to the nature of the gradient thresholding problem, most of these bilevel classification techniques cannot be used. Histograms of the gradient image normally do not have two visible peaks and a valley as shown in Figure 2. This means that using the modality, shape, or moments to classify the peaks will be futile since the two peaks and the valley separating them cannot be identified.

Other classifications such as the p-tile [9] assume that the percentage of pixels in one class is known a priori. The percentage of edge pixels is usually estimated to be 10 to 20% of the overall pixels of an image [3, 11, 12]. However since these percentages depend strongly upon the images and the signal to noise ratio (SNR) present, such techniques will only provide a rough estimate of the threshold and not an accurate measure.

B Our Model

Our approach employs a statistical classification based upon a 5-parameter model. As mentioned before, the gradient value reflects changes in grayscale over a local region of an image. This change is generally the result of both edges (contrast changes) and noise. We assume that we can statistically model the data by probability density functions. Given a 512 x 512 image with over 250,000 points, we have a large sample of data points which can be used to find density functions. We model edge pixels and non-edge pixels by using two density functions which sum to the original histogram.

The density functions which we use to model the edge and non-edge pixels are gamma density functions. We assume that the edge and non-edge contrasts, as defined by the horizontal



Figure 1: A bimodal histogram which can be easily thresholded.



Figure 2: A typical histogram for edge thresholding with the estimated densities of the non-edge and edge pixels.

L

and vertical gradients ∇_x and ∇_y , can be modeled by the Gaussian distribution based on the central limit theorem. The gradient squared values given by $\nabla^2 = \nabla_x^2 + \nabla_y^2$ for each density have a non-central chi-squared distribution with two degrees of freedom. However, since this distribution is difficult to compute, we instead model the gradient values given by $\nabla = \sqrt{\nabla_x^2 + \nabla_y^2}$ that was also used by [7, 8]. Based on synthetic data, the gamma density given by the equation

$$f(x \mid \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}.$$

provided a good approximation over the range of gradient values within the images $\chi^2 : p < .01$.

Two parameters model each of the two distributions and the remaining parameter, p_0 , describes the ratio between the edge and non-edge pixels. Thus the overall 5-parameter model, which is fit to the initial histogram, is the sum of a gamma density representing the edge pixels and a gamma density representing the non-edge pixels weighted by the ratio factor p_0 which can be formally written:

$$f(x \mid \alpha_0, \beta_0)p_0 + f(x \mid \alpha_1, \beta_1)(1-p_0)$$

and is shown in Figure 2. The threshold of most statistical classification systems is chosen to minimize the probability of error using either a MAP (maximum *a posteriori*) or ML (maximum likelihood) criteria [4]. This threshold can be determined from our model without requiring any particular modality, concavity, or shape and without assuming knowledge of the initial ratio of the densities, which overcomes the limitations of other automatic classification techniques. Thus, by accurately estimating the five parameters of our model, we can statistically determine a threshold to distinguish between edges and non-edges.

III Description of the Different Methods

Using our model described in the previous section, there are several different methods of estimating the five parameters. The first method, a global descent, tries to estimate all five parameters simultaneously. The other methods divide the problem of calculating the model's five parameters into two algorithm steps. The first step, which we will refer to as "alphabeta estimation" attempts to find the α and β parameters of both the edge and non-edge

density functions. The second step, "percentage estimation", computes the ratio to edge and non-edge pixels in the image. This section describes these estimation techniques in detail.

A Global Descent

For the global descent method, the goal is to find the combination of five parameters that best fits the gradient histogram data. We are searching in a 5 dimensional space for a set of parameters that most closely approximates the histogram. We use a Powell algorithm [13], a multidimensional minimization based upon successive line minimizations that do not require a gradient calculation. It utilizes conjugate directions in forming line minimizations which converge quadratically to the minimum value of the function.

In order to use this search strategy, we had to define a "best fit" of the model to the data. We investigated three distance measures for the entire experiment, each of which provides a measure of how well a model with given parameters fits the actual data points. The distance measures identify the difference in quantized curves x and y.

- d1: the absolute distance error given by $\sum_i |y_i x_i|$
- d2: the squared error given by $\sum_i (y_i x_i)^2$
- d3: the area between the curves using a trapezoidal approximation

While the measures are strongly correlated, they represent different accuracies and computation speeds, and vary with the amount of noise present. For the global descent method, only the d1 measure was used since the method is already quite computationally expensive.

B Two Step Estimation

Breaking the estimation process into two steps, the alpha-beta estimation and the percentage estimation are computed alternately until a convergence is reached. This: strategy is the same as the EM (expectation-maximization) algorithm formulation [14, 15] and significantly reduces the search space. In alpha-beta estimation, the two parameters For each density function are estimated (four parameters total), given p_0^k

 $lpha_0^k,eta_0^k,lpha_1^k,eta_1^k\mid p_0^k$

where p_0^k is the relative probability of density0 verses density1 at the kth iteration. With these estimates, a new p_0^k parameter is formed in the percentage estimation step

$$p_0^{k+1} \mid lpha_0^k, eta_0^k, lpha_1^k, eta_1^k$$

These two estimation steps are repeated until the estimated values converge or until no progress is being made.

Since it is difficult to accurately estimate the parameters for two densities simultaneously from their sum as in the first step, we divided all the points in the histogram into two non-overlapping groups using the prior estimates for the density parameters and the p_0 value. Again using the EM algorithm formulation, we compute a ratio of the densities weighted by p_0 at each point of the histogram, defined as

$$ratio(i) = \frac{p_0 \times density0(i)}{p_0 \times density0(i) + (1 - p_0) \times density1(i)}$$

where density 0 and density 1 are the reconstructions of the gamma densities given the estimates of the a and β parameters from the most recent estimation. It should be noted that this ratio function is actually not a density function, but only a ratio of densities with values ranging from 0 to 1. From this ratio function we form two separate density estimates

$$density0(i) = histogram(i) \times ratio(i)$$

$$density1(i) = histogram(i) \times (1 - ratio(i))$$

for which new a and β parameters can be more easily estimated. By truncating the histogram where possible, as described later, the computation of this step can be reduced since it is directly proportional to the length of the histogram.

B.1 Alpha-Beta Estimation

We compared five different methods for estimating the \boldsymbol{a} and $\boldsymbol{\beta}$ parameters of the two densities. Two of these employed an iterative estimation and the remainder employed a two-parameter Powell descent algorithm using different distance measurements.

1) Method of Moments This method generates point estimators by equating the first two moments of the data with the first two moments of the parameter estimates given by

1

the following equations:

$$Data_moment1 = \frac{1}{n} \sum_{i=1}^{n} x_i$$
$$Data_moment2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2$$
$$Estimatemoment1 = \alpha\beta$$
$$Estimate_moment2 = \alpha\beta^2 + (\alpha\beta)^2.$$

Solving these for a and β , we get

$$\alpha = \frac{1}{\beta n} \sum_{i=1}^{n} x_i$$
$$\beta = \frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i} - \frac{1}{n} \sum_{i=1}^{n} x_i$$

2) Maximum Likelihood Estimation As the name implies, this approach is the formal ML estimation using the derivative of the gamma density function which finds the maximum of the product of densities using the derivative's zero crossings:

$$\prod_{i=1}^{n} f(x_i \mid \alpha, \beta) = \frac{\prod_{j=1}^{n} x_j^{(\alpha-1)} e^{-\sum_{\beta \in \mathcal{X}_i} x_j}}{\beta^{\alpha n} \Gamma(\alpha)^n}$$

Since the log is a monotonic function, we can take the log of both sides

$$\log(\prod f(x_i \mid \theta)) = (a - 1) \sum \log x_i - \frac{\sum x_i}{\beta} - a n \log \beta - n \log \Gamma(\alpha).$$

Taking the derivative of this with respect to α and β and setting the resulting equations to 0, we obtain the MLE of a and β , respectively:

$$\frac{\partial}{\partial \alpha} \to \sum \log(x_i) - n \log 4 - n \frac{\partial}{\partial \alpha} \log[\Gamma(\alpha)] = 0$$

where

$$\frac{\partial}{\partial \alpha} \log[\Gamma(\alpha)] = \frac{\int_0^\infty t^{(\alpha-1)} \log t e^{-t} dt}{\int_0^\infty t^{\alpha-1} e^{-t} dt}$$

and

$$\frac{\partial}{\partial\beta} \to \beta = \sum_{i} \frac{x_i}{n\alpha}$$

It should be noted that the $\partial/\partial \alpha$ of the log of gamma function requires significant computation with the integral. While it may be possible to perform a table look-up of precomputed values for similar images, this is impractical over a wide set of images due to the high precision required and large range of possible alpha values.

3–5) The Powell Descent Method This method is the same overall algorithm as the global descent method described earlier in this section. The difference is that the Powell method for alpha-beta estimation is searching for two parameters for each density, and each density is handled separately. The search space is significantly reduced from the five dimensions of the global descent method to only two dimensions. The Powell method minimizes the distance between the model of the density and the density computed from the histogram ratio function. That is,

distance(density0-from-ratio, density0(α_0, β_0)) distance(density1-from-ratio, density1(α_1, β_1)).

We used three different Powell methods, for alpha-beta algorithms 3-5, with respective distance functions d1, d2, and d3 as defined in Subsection A.

B.2 Percentage Estimation

Four different algorithms were compared for the percentage estimation of non-edges. The percentage of non-edge pixels or p_0 is a scaling ratio parameter for the two densities. It determines the relative weights of the two densities which sum to the gradient histogram. Three of the algorithms use the Golden Section descent method with the distance measures d1 through d3. The fourth employs the EM algorithm [14, 15] on a Bernoulli estimate of the ratio.

The Golden Section [13] descent algorithm is a one-dimensional minimization method based upon bracketing the minimum value, that is quite efficient in searching for a minimum value. Given the two densities described by the a and β parameters, p_0 was defined as the value which minimized the distance between the histogram and the sum of the two ;gammadensities weighted by p_0 . The equation is:

$$\sum_{i} distance\{histogram(i), [density0(i) \ x \ p_0 + density1(i) \ x \ (1 - p_0)]\}$$

where the distance is one of the aforementioned distance functions d1 through d3.

Т

L

The fourth estimation algorithm is based on a Bernoulli estimate of the ratio p_0 and uses the EM algorithm. The equation is:

$$Po = \frac{k \, s \, u \, m}{n}$$

where

$$k \, s \, u \, m = \sum_{i} histo(i) \frac{p_0 \sum_{j} density0(j)}{p_0 \sum_{j} (density0(j)) + (1 - p_0) \sum_{j} \overline{density1(j)}}$$

and n is the total number of points in the histogram.

B.3 Threshold Determination

Given two overlapping densities, there are several methods of determining a threshold described in pattern recognition and signal processing texts. Most of these methods are designed for normal densities, rather than the gamma densities of our approach. Although some researchers such as Canny prefer dual thresholds for edges, this paper will focus on single threshold estimation. The results can be generalized for dual thresholds with our model of the edges with an appropriate threshold criteria given the edge model. We considered the MAP and ML decision thresholds, which are the most common in practice, and also tested a direct p_0 threshold. These three methods of computing the threshold were compared in the experiments for their accuracy using our model.

The MAP threshold decides that an edge is present when

 $p(x \mid edge) \times p(edge) > p(x \mid non_edge) \times p(non_edge)$

and a non-edge otherwise. In the notation of this section, we can rewrite this as

density
$$l x (1 - p_0) > density 0 x p_0$$
.

The maximum likelihood threshold decides that an edge occurs where

$$p(x \mid edge) > p(x \mid non-edge)$$
 or $densityl > density0$.

This differs from the MAP decision in that it ignores the prior probabilities of edge and non-edges.

The p-tile method of selecting edges forms a threshold at a given percentage, ignoring all other information. Since p_0 is a parameter of the model, we tested this approach, even

though this threshold is based directly on one parameter rather than the **3** parameters of the MAP threshold and the 2 parameters of the maximum likelihood threshold.

IV Experiments

Two different techniques were used to analyze the effectiveness of our model. In the first experiment, 16 diverse images were processed with the different algorithms and the computed thresholds were compared with the subjective edge threshold decisions made by 5 researchers. The second experimented evaluates the robustness of the model to different noise levels applied to a synthetic image. We will discuss the image preparation and the procedures in computing the thresholds, followed by the two evaluation methods.

A Data Preparation

All the methods described in Section III operate on the gradient histogram prepared from the images. To standardize the testing, all the images were prepared using the following steps: adding noise, smoothing, applying the Sobel operator, and truncating the histogram which will be explained in more detail.



Figure 3: With some histograms, noise is required to spread the peak near zero gradient units.

Additive white Gaussian noise N(O,1) (zero mean and unit variance) was added to each

L

image and rounded to the **255** grayscale values of the original images to spread the sharp spikes which occurred in some of the histograms. For images with very low noise and little texture which often occur in synthetic images, the histogram takes the form of a delta function at gradient value 0 (no edge), and a small peak at a gradient value proportional to the contrast of the edge shown in Figure 3. Recalling that the gamma density at x = 0 is zero, it is clear that the pixels of the non-edge density will not accurately fit the data. By adding Gaussian noise N(0, 1) to the image, the gradient values are shifted away from the zero value, spreading the delta function. Since the resultant thresholds are robust to the addition of low variance noise, this noise was added to all the images.

The image was then smoothed using a 3 x 3 Gaussian filter with unit variance. A Gaussian filter is commonly applied before edge detection to reduce the effects of noise by smoothing, and as a form of regularization as described by Poggio [16]. The smoothing had virtually no effect on the overall shape of the histogram.

Although our algorithm can be used by any differential or gradient based operator obtaining similar results, only the 3 x 3 Sobel operator was used in this experiment. This operator was chosen since it is effective in the presence of noise and is widely used [3, 2]. After convolving the Sobel operator with the image, the output gradient values were placed in a histogram with a bin size of 1 gradient unit.

The largest 0.01% of the gradient values of the histogram were truncated which served several purposes. First, by reducing the length of the histogram, computation was saved since the statistics describing the data fit need only be computed for pertinent values. Second, by eliminating these outlying values, the accuracy was improved since such values tended to bias both the statistical-based methods and the fitting methods. Using different percentages for truncation did not change the resulting thresholds as long as the percentages were much less than the percentage of edge pixels in the image. The histogram was then normalized to unit area for purposes of comparing data sets.

B Running the Algorithms

The goal of this portion of the experiment was to find and compare thresholds from the different algorithms under the constraints of our model. We applied each method to the prepared histograms described earlier in this section. An initial seed of **88%** non-edge pixels was used for all the images since they commonly compromise between **85%** and 90% of image pixels, though this is strongly dependent upon the image and the definition of edges. For all cases except the global Powell descent involving simultaneous estimation of all the parameters, the first iteration of the parameter estimation step utilized the maximum likelihood estimation (MLE) method. This step provided a standard, quick, and reasonable initialization using the updated percentage information for all the algorithms and did not appear to bias any of the algorithms.

Following the initial estimation step, subsequent parameter estimation followed by the percentage estimation was performed and the process repeated as shown:

- 1. initial $p_0^k = 88\%$
- 2. initial alpha-beta estimation (p_0^k) returns a, ", $\beta_0^k, \alpha_1^k, \beta_1^k$
- 3. overall estimation loop { percentage estimation $(\alpha_0^k, \beta_0^k, \alpha_1^k, \beta_1^k)$ updates p_0^{k+1} alpha-beta estimation $(\alpha_0^k, \beta_0^k, \alpha_1^k, \beta_1^k, p_0^{k+1})$ returns $\alpha_0^{k+1}, \beta_0^{k+1}, \alpha_1^{k+1}, \beta_1^{k+1}$

$$k = k + 1$$

where the k denotes the iteration number.

The complete iterative process of parameter estimation and percentage estimation, shown in the overall estimation loop was repeated a maximum of 12 times or until the threshold converged within 6 gradient units of the previous iteration for all the algorithms. The latter condition required the minimum precision of the threshold to be approximately **0.5%** (6 out of a maximum 1200 gradient units) between on successive iterations.

The parameter estimation algorithms iterated for a maximum of 50 iterations or terminated when the both the β_0 and β_1 parameters converged to within .3% of the previous iteration. The a parameters were not used as a termination condition since they were less sensitive

than the β parameters and also tended to vary proportionally to the β parameters. The percentage estimation algorithms terminated when consecutive estimates converged within 0.3%.

For all the methods, the range of valid percentage of non-edge pixels was restricted to [1%,99%] of the truncated histogram. According to the 5-parameter model, the histogram consisted of the sum of two non-zero gamma densities representing the edge and non-edge pixels. Bounding the size of the non-edge histogram away from 0% and 100% barred the elimination of one of the edge densities.

C Testing Other Methods

Before analyzing the results of our 5-parameter model, we tested several other techniques described by Sahoo, Soltani, and Wong in their overview of thresholding methods, to verify that these methods would not, in general, be effective for different gradient images. We prepared the data using the same process of adding noise, filtering, and applying the Sobel operator as before. As expected, the p-tile, node, and concavity methods were not effective due to the varying shapes and edge to non-edge ratios of the different image histograms. The p-tile method always selected the same percentage of pixels as the threshold, but the percentage of edges in the battery of 16 images varied widely. The centroid and other methods including the Otsu method could not find a threshold with the absence of two peaks in most of the gradient histograms.

D Computed and Subjective Thresholds

The purpose of the subjective analysis was to find how well the mathematical threshold compared with the subjective perceptual determination of edges. We used the human visual system as the standard of edge detection since humans are able to rapidly and accurately perceive global structures in images such as edges through a complex process, often referred to as perceptual organization [17, 18, 19, 20].

For the experiment, the 16 images shown in Figure 4 were prepared as described in Sub-



Figure 4: A battery of 16 images of varying scenes. Moving left to right starting at the top, the images are: a house scene, an xray skull image, bethl, contact, crowd, dilts, a synthetic dragon cartoon, flir images af a truck, man (gih), girl2, jo, john, a satellite image of the earth, Madonna quantized to a few gray levells, a text scan, and a still life (iml6).

section \mathbf{A} to form a histogram. Since the full range of thresholds is not necessary, we first restricted the valid range of thresholds and then divided it into 256 equally spaced possible thresholds, for greater accuracy.

Five subjects with experience in edge detection and computer vision were shown two black and white 5.4" x 5.4" images on the computer screen–the original image and the thresholded image. For each of the 16 images, the subjects were asked to raise or lower the threshold using the mouse buttons so that the thresholded image matched their idea of a "best" edge for the original image.



Figure 5: The thresholded images using the best algorithm.

Each of the algorithms of Section III were run on the histograms, giving a set of computed thresholds. We used the area under the normalized histogram between each computed threshold and the average subjected threshold as the metric for comparing the performance of the different algorithms. Table 1 shows the mean and variance of the area differences for the

Mean	Var.	$\alpha - \beta$	% Est.	Thr.	Mean	Var.	$a - \beta$	% Est.	Thr.
0.0611	0.0020	4	1	2	0.1381	0.0064	1	3	1
0.0618	0.0019	2	4	2	0.1387	0.0064	1	1	1
0.0633	0.0010	4	4	2	0.1387	0.0068	2	1	1
0.0741	0.0037	4	3	2	0.1399	0.0067	2	3	1
0.0836	0.0020	1	4	2	0.1520	0.0067	1	3	2
0.0963	0.0036	2	1	2	0.1522	0.0067	1	1	2
0.0967	0.0038	2	4	3	0.1522	0.0076	2	2	2
0.1016	0.0051	2	3	2	0.1562	0.0167	4	3	3
0.1084	0.0044	4	4	3	0.1577	0.0131	3	1	1
0.1166	0.0061	4	1	1	0.1577	0.0131	3	3	1
0.1219	0.0024	1	4	1	0.1606	0.0081	5	2	1
0.1229	0.0045	5	2	2	0.1618	0.0368	2	3	3
0.1234	0.0161	4	1	3	0.1620	0.0370	2	1	3
0.1324	0.0063	4	4	1	0.1621	0.0042	1	2	2
0.1362	0.0066	2	4	1	0.1637	0.0124	3	2	1
0.1365	0.0074	4	3	1	0.1651	0.0073	global	global	global

Table 1: The mean and variance of the average distance (area) between the computed and subjective thresholds, averaged over the 16 images.

	Alpha-Beta Estimation		Percentage Estimation		Threshold
	$\alpha - \beta$		% Est.		Thr
1	Method of Moments	1	Golden Section (d1)	1	MLE
2	Max Likelihood Est.	2	Golden Section (d2)	2	MAP
3	Powell Descent (d1)	3	Golden Section (d3)	3	p-tile
4	Powell Descent (d2)	4	Bernoulli EM		
5	Powell Descent (d3)				

Table 2: A list of the algorithms presented in Table 1



Figure 6: Images thresholded at 27, 39, 57, and 128 gradient units for subjective decisions. The computed threshold is the second image.

best algorithms. Converting the thresholds into percentages of non-edge pixels, the average of the 5 best algorithms and average subjective edge decisions are graphed in Figure 7. As an example of the edges, three of the researchers chose the thresholds of 27, 57, and 128 for the im16 image shown in Figure 6. The best algorithm selected the second image of the sequence with a threshold of 39. The thresholds for all the images with the best algorithm is shown in Figure 5.

Al	pha-Beta Estimation Method	Mean	Variance
1	Method of Moments	0.210	0.025
2	Maximum Likelihood Est.	0.142	0.013
3	Powell Descent with d1	0.257	0.026
4	Powell Descent with d2	0.262	0.084
5	Powell Descent with d3	0.313	0.042

Table 3: The mean and variance of the area between the subjective and computed thresholds of the 5 different alpha-beta estimation methods.

Pe	ercent Estimation Method	Mean	Variance
1	Golden Section d1	$0.\overline{224}$	0.036
2	Golden Section d2	0.311	0.058
3	Golden Section d3	0.228	0.034
4	Bernoulli EM	0.185	0.029

Table 4: The mean and variance of the area between the subjective and computed thresholds of the 4 different percentage estimation methods.

For comparing the algorithms, Table 1 demonstrates that many of the algorithms have very similar performance over the different images. To better view the data, we have grouped the data by averaging the values for every alpha-beta estimation, percent estimation, and threshold technique in Tables 3, 4, and 5, respectively.



Figure 7: The percent of non-edge pixels in each of the 16 images as determined by the average of the 5 researchers (solid lines) and the average of the 5 best algorithms (dotted lines).

	Threshold	Mean	Variance
1	MLE	0.221	0.037
2	MAP	0.187	0.029
3	p-tile	0.302	0.051

Table 5: The mean and variance of the area between the subjective and computed thresholds of the 3 different threshold estimation methods.

The first two alpha-beta algorithms of Table 3 employ statistical methods to determine the parameters, but the last three use the Powell Descent fitting technique. Although the results in terms of accuracy and robustness are fairly similar, the statistical methods are many times faster than their counterparts. Consequently, the maximum likelihood estimation technique is the preferred method in terms of both speed and accuracy.

In Table 4, the Bernoulli EM algorithm was the most accurate method for percent estimation and was also much faster than the Golden Section counterparts. The reasoning is similar to the statistical versus the fitting methods of the alpha-beta algorithms.

For Table 5, the percentage estimation step for the thresholding algorithms was not very accurate as expected, since it does not use all the model information to find an optimized threshold, but rather only one parameter of the model. The MAP threshold proved the best method, but the MLE was also effective.

E Robustness to Noise

The previous section demonstrated that the 5-parameter model and corresponding algorithms were robust over a wide range of images. This experiment tests the robustness of the algorithms over different noise levels applied to a synthetic image. Using the top algorithms from Table 1, we added N(0, a) noise to ten copies of a synthetic image with a^2 ranging from 1.0 to 4096.0, in multiples of 2.0. Four examples of the noisy images are shown in Figure 8. We then calculated the threshold for each image at each noise level.

To analyze the data, we wanted to calculate the misclassification errors associated with the average threshold choices at each noise level. Since the gradient values increase with the noise level (Figure 10), the thresholds must also increase accordingly for the model to work. For this evaluation, we assumed the $a^2 = 1.0$ case corresponded to the perfect classification. We then calculated the Type I, Type II, and total errors as a function of the noise variance which are plotted in Figures 9, respectively. As can be seen from the graph, the total classification error is low (< .1) until $\sigma^2 = 512$ which corresponds to a SNR as low as 3.0. The reason

.....



Figure 8: Synthetic images with added white Gaussian zero-mean noise of variances 1.0, 128.0, 1024.0, and 4096.0.



Figure 9: Type I, Type II, and overall misclassification errors for different noise levels using the best algorithm.

that the total error is less than the Type II error is that the total error represents a weighted sum of the errors. The relatively small error over the low SNR indicates that the algorithm is effective at identifying a threshold. At high noise levels, the 3×3 Sobel is much less effective as an edge detection operator which degrades the performance of any thresholding technique. As an interesting note, the Type II error was significantly higher than the Type I error. This implies that the researchers preferred more false positives than false negatives in estimating their thresholds.

V Conclusion

In object recognition systems, choosing a threshold to accurately find edge pixels in an image at the lowest level can lead to significant computational savings at higher levels. Since most automatic thresholding techniques do not apply to the specific problem of edge detection, heuristic approaches are commonly used in research. We have developed a model of edges that strongly agrees with the subjective perception of edges consisting of the weighted sum of two gamma densities to represent edge and non-edge pixels. The model proved effective over a wide range of images and performed well in the presence of noise with a SNR as low as **3.0**.

After testing a number of the algorithms to calculate the thresholds based on the model, we recommend both the Powell descent with d2 distance measure and the maximum likelihood estimate in conjunction with the Bernoulli EM and MAP threshold. 'These algorithms performed the best in terms of accuracy and computational speed for our set of images.

This paper has discussed the use of automatic threshold determination over the entire image. Many applications however require local rather than global thresholds. Our method can be applied directly to such applications without modifying the algorithm by simply taking the gradient histogram data from selected regions of the image instead of the whole image. As mentioned earlier, the model is also generalizable to operators other than the Sobel.

References

[1] S. Venkatesh and L. J. Kitchen, "Edge evaluation using necessary components," CVGIP:

Graphical Models and Image Processing, vol. 54, pp. 23-30, 1992.

- [2] B. K. P. Horn, Robot Vision. New York: McGraw-Hill Book Company, Inc., 1985.
- [3] D. H. Ballard and C. M. Brown, Computer Vision. New Jersey: Prentice-Hall, Inc., 1982.
- [4] I. E. Abdou and W. K. Pratt, "Quantitative design and evaluation of enhancement/thresholding edge detectors," Proceedings of the IEEE, vol. 67, no. 5, pp. 753-763, May 1979.
- [5] L. Kitchen and A. Rosenfeld, "Edge evaluation using local edge coherence," IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-11, no. 9, pp. 597-605, September 1981.
- [6] G. F. McLean and M. E. Jernigan, "Hierarchial edge detection," *Computer* Vision, Graphics, and Image Processing, vol. 44, pp. 350-366, 1988.
- [7] J. F. Haddon, "Generalised threshold selection for edge detection," *Pattern Recognition*, vol. 21, no. 3, pp. 195–203, 1988.
- [8] O. A. Zuniga and R. M. Haralick, "Gradient threshold selection using the facet model," Pattern Recognition, vol. 21, no. 5, pp. 493-503, 1988.
- [9] P. K. Sahoo, S. Soltani, and A. K. C. Wong, "A survey of thresholdling techniques," Computer Vision, Graphics, and Image Processing, vol. 41, pp. 233–260, 1988.
- [10] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," Computer Vision, Graphics, and Image Processing, vol. 29, pp. 273-285, 1985.
- [11] L. S. Davis, "A survey of edge detection techniques," Computer *Graphics* and Image Processing, vol. 4, pp. 248–270, 1975.
- [12] M. Tanaka and T. Katayama, "Edge detection and restoration of noisy images by the expectation-maximization algorithm," *Signal* Processing, vol. 67, no. 5, pp. 753–763, May 1989.
- [13] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, Numerical Recipes in C. Cambridge: Cambridge University Press, 1992.
- [14] M. Tanaka and T. Katayama, "Edge detection and restoration of noisy images by the expectatons-maximization algorithm," Signal Processing, vol. 17, pp. 213-226, 1989.
- [15] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood for incomplete data via the EM algorithm," J. Royal Stat. Society, vol. B-39, pp. 1–38, 1977.
- [16] V. Torre and T. A. Poggio, "On edge detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, no. 2, pp. 147–163, March 1986.

- [17] A. Sha'ashua and S. Ullman, "Structural saliency: The detection of globally salient structures using a locally connected network," *Proceedings of the Second International Conference on Computer Vision*, 1988, pp. 321–327.
- [18] A. Rosenfeld, "Pyramid algorithms for efficient vision," in *Vision: Coding and Efficiency* (C. Blakemore, ed.), pp. 423–430, Cambridge: Cambridge University Press, 1990.
- [19] J. R. Pomerantz and M. Kubovy, "Theoretical approaches to perceptual organization: Simplicity and likelihood principles," in *Handbook of Perception ana! Human Performance* (K. R. Boff, L. Kaufman, and J. P. Thomas, eds.), vol. 2, pp. 36.1–36.46, New York: Wiley, 1986.
- [20] A. Rosenfeld, "Theoretical techniques: Pyramid algorithms for perceptual organization," *Behavior Research Methods, Instruments, and Computers*, vol. 18, no. 6, pp. 595– 600, 1986.



Figure 10: Histograms for the synthetic image under three different added noise variances.