A Methodology for Quantitative Performance Evaluation of Detection Algorithms

Tapas Kanungo, *Senior Member, IEEE,* M. Y. Jaisimha, John Palmer, and Robert M. Haralick, *Fellow, IEEE*

***Abstract*— We present a methodology for the quantitative performance evaluation of detection algorithms in computer vision. A common method is to generate a variety of input images by varying the image parameters and evaluate the performance of the algorithm, as algorithm parameters vary. Operating curves that relate the probability of misdetection and false alarm are generated for each parameter setting. Such an analysis does not integrate the performance of the numerous operating curves. In this paper, we outline a methodology for summarizing many operating curves into a few performance curves. This methodology is adapted from the human psychophysics literature and is general to any detection algorithm. The central concept is to measure the effect of variables in terms of the equivalent effect of a critical signal variable, which in turn facilitates the determination of the breakdown point of the algorithm. We demonstrate the methodology by comparing the performance of two-line detection algorithms.**

I. Introduction

QUANTITATIVE performance evaluation of computer vision algorithms is important in order to compare the performance of dissimilar algorithms on a common quantifiable basis. The usual method is to vary parameters of the input images or parameters of the algorithms and then construct operating curves that relate the probability of misdetection and false alarm for each parameter setting. Such an analysis does not integrate the performance of the numerous operating curves. In this paper we outline a methodology for summarizing many operating curves into a few performance curves. This methodology is adapted from the human psychophysics literature and is general to any detection algorithm.

How exactly does one define performance? Issues that need to be addressed are: i) What image population is relevant? ii) Is the performance evaluated independent of the algorithm? iii) How are differences in performance measured? An area with previous work on quantitative performance evaluation is in edge detection and thresholding [1]–[7]. Most of the papers present an analysis that is specific to edge detection. Furthermore, the performance is finally a number, e.g., percentage of edge points detected, etc. There is little further analysis of the sensitivity of performance to relevant factors, such as the context of the edge.

Manuscript received March 9, 1993; revised March 7, 1995. The associate editor coordinating the review of this paper and approving it for publication was Prof. Rama Chellappa.

T. Kanungo, M. Y. Jaisimha, and R. M. Haralick are with the Department of Electrical Engineering, University of Washington, Seattle, WA 98195 USA.

J. Palmer is with the Department of Psychology, University of Washington, Seattle, WA 98195 USA.

IEEE Log Number 9415096.

In this paper, we present a methodology for designing experiments to characterize detection algorithms. We adopt an established methodology that has been used and tested in psychophysics. The central concept is to measure the effect on performance of variables in terms of the equivalent effect of a critical signal variable. For example, psychophysicists study the performance of humans in the task of edge and grating detection by measuring the contrast necessary for detection under a variety of conditions [8]–[10]. The effect of various conditions is measured by the equivalent effect of contrast as quantified by the contrast threshold. We demonstrate the use of this methodology for detection algorithm performance evaluation by comparing the performance of two line detection algorithms. The task was to detect the presence or absence of a vertical edge in the middle of an image containing a grating mask and additive Gaussian noise. We compare two line detection algorithms: i) The facet edge detector [11] followed by the Burns line finder [12], and ii) the facet edge detector followed by the Hough transform.

The methodology is defined in a general enough way that it can be readily applied to other computer vision algorithms that can be specified as a *detection task*. Typical problems where this methodology can be applied are—the detection of cracks, the detection of curves, the detection of objects, pose error estimation, machine part inspection, etc. Some of the results presented here were presented earlier in [13], [14] and this methodology has been partially applied to object recognition [15] and machine part inspection [16].

In Section II, we describe the general performance evaluation methodology. The example experiment we perform to demonstrate the methodology is described in Section III. Here, we discuss the detection tasks, two algorithms for detecting our targets, and describe the population of images the algorithms and the experimental protocol. In Section IV, we summarize the results. The benefits of our methodology and its application to other detection problems is discussed in Section V.

II. Data Analysis Methodology

Consider a typical detection task wherein a system is required to report the presence or absence of a target in an input image. Typically, the system first computes a number that gives a measure of the evidence of the presence of the target, and then reports that a target is present if the evidence strength is greater than a particular value. In any detection task there are some variables that affect the signal to noise ratio S/N, in the image, which, in turn, affect the performance

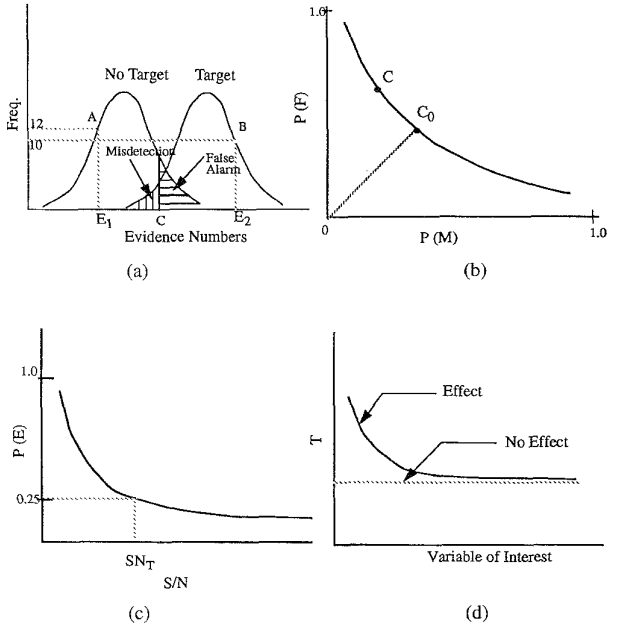


Fig. 1. Performance curves for the typical experiment described in the methodology. (a) Step 1—frequency counts for target and nontarger case; (b) Step 2—operating characteristics relating P(M) and P(F); (c) Step 3—equal cost probability of error as a function of signal to noise ratio; and (d) Step 4—threshold as a function of the variable of interest.

of the detection system. One example of such a variable is the edge contrast. Increasing the edge contrast of the signal results in an increased S/N ratio and an improvement in the system's ability to detect edge targets.

The methodology is applied in four steps. The first two are standard decision analysis [2], the last two are inspired by psychophysical methods [10].

Step 1: First, two noise-free images are created: one with the target and one without. Then, a large number of images are created from these base images by adding different realizations of noise. Each image is then provided as input to the system whose performance is to be evaluated. The output of the system is a number that is a measure of the evidence of the presence of a target. This number is referred to as the *evidence strength*. Each evidence strength has an associated frequency (i.e., the number of times it appears at the output of the system in the course of the experiment). The frequency count is plotted versus the evidence strength to obtain the graph shown in Fig. 1(a). Two frequency distributions are obtained, one for the case of target present and one for the case of no target present. In this figure, the point A represents the fact that the system outputs E_1 as the evidence strength for twelve input images that do not contain the targets. Similarly, the point B represents the fact that the system outputs E_2 as the evidence strength for ten input images that contain the targets.

Step 2: To decide whether or not the system has indicated the absence or presence of a target, an evidence criterion C has to be applied to the evidence strengths output by the system. If the evidence strength corresponding to an image is greater than the evidence criterion, then the target is declared to be present in the image. To study the performance of the system, the evidence criterion is varied through a set of values. See Fig. 1(a). If the system claims that the input image does not contain the target when in fact it does, the decision is defined as a *misdetection*. Similarly, if the system claims that the input image contains the target when in fact it does not, the decision is defined as a false alarm. The probability of misdetection, $P(misdetection)$ is defined as $P(no\ target|target)$ = number of misdetections/total number of input images with target. The probability of false alarm, $P(false\ alarm)$, is defined as $P(target|no\ target)$ = number of false alarms/total number of input images with no target. For each value of the evidence criterion there are corresponding values of $P(misdetection)$ and $P(false\ alarm)$. The plot of $P(misdetection)$ versus $P(false\ alarm)$ as the evidence criterion is varied is called the *operating characteristic*, see Fig. 1(b).

For the equal bias case, choose the operating criterion, C_0 as the evidence strength for which $P(misdetection) = P(false\ alarm)$. The equal bias probability of error, $P(E)$ is then defined as $P(E) = (P(misdetection) + P(false\ alarm)/2$ for equal probability of target and no-target.

Step 3: For different values of the signal to noise ratio, repeat steps 1 and 2. Each value of S/N results in an operating characteristic from which the equal cost probability of error, $P(E)$, can be determined. $P(E)$ is plotted versus S/N in Fig. 1 (c). Choose C_T, the value of S/N for which $P(E) = 0.25$, as the *contrast threshold*. The value of $P(E)$ corresponding to the C_T is chosen half way between pure chance ($P(E) = 0.5$) and perfect ($P(E) = 0.0$). If the particular application so demands, the value of $P(E)$ that defines the contrast threshold C_T can be chosen differently.

Step 4: For different values of the variable of interest V, repeat Steps 1, 2, and 3. A typical plot of contrast threshold C_T versus the variable of interest V is shown in Fig. 1(d). This curve now characterizes the effect of the variable of interest in terms of the effect of contrast. The effect of any variable can be measured by its effect on the contrast threshold.

In summary, Steps 1 and 2 result in a measure of performance using standard decision analysis. Step 3 provides a measure in terms of a signal variable at a fixed S/N ratio. Step 4 measures the effect of any other variable by the common currency of the signal variable.

III. The Experiment

The Detection Task

We illustrate the general methodology with an analysis of a particular issue in edge and line detection. The question raised is—How selective is an edge detection algorithm to clutter (irrelevant) edges? For example, can the algorithm detect an edge in the context of a texture containing oriented edges? A similar problem has been analyzed for human detection performance in [8] (a typical real-world example of such a detection task in the image processing domain is in the inspection of composite materials and metals slabs for cracks).

In the experiment, the input is either a target image or a nontarget image. The target image consists of a vertical edge at a known column position, irrelevant square wave grating at various orientations, and Gaussian noise. The nontarget image consists of only the grating and the Gaussian noise—no vertical edge. The task is to detect the presence or absence of the edge of interest without being "confused" by the grating. The signal to noise ratio is manipulated by varying the noise standard deviation. The variable of interest is the orientation of the grating. We measure the extent of an algorithm's sensitivity to the orientation of the irrelevant grating.

The Algorithms

In this section, we describe the two algorithms used to detect the presence or absence of a vertical edge at the center of the image.

Algorithm 1: This algorithm has two stages: an edge detection stage followed by a line detection stage. The edge detection stage labels individual pixels as edge or nonedge. The line detection stage identifies groups of edge pixels that might represent a line. For edge detection we use the second directional derivative facet edge detector [11]. For the second stage, the line detection is done by the Burns line finder [12].

The Burns line finder takes the output of the edge detector and labels all the connected lines it finds. As a measure of the confidence of the algorithm in the presence (or absence) of a vertical line at the location of interest (the column of pixels at or "near" the center of the image), we count the number of pixels in the middle three columns that were classified by the Burns line finder as line pixels. This count is referred to as the *evidence strength.*

Algorithm 2: Like Algorithm 1, this algorithm also has two stages: An edge detection stage followed by a line detection stage. For edge detection, as in Algorithm 1, we use the second directional derivative facet edge detector. The line detection is performed using the Hough transform technique.

The line detection stage takes the output of the edge detector and maps the edge pixels to the distance-angle Hough space [11]. An edge with no noise or clutter grating is first passed through the edge detector. The perfect edge image obtained is then mapped onto the Hough space. Since this is a perfect image, all the edge pixels map on to one bin in the Hough space. As the perfect step is progressively distorted the grating and noise, the number of pixels that fall into the the same bin decreases. Since, we know in advance the exact location of the edge, only the pixels in the vicinity of the edge need to vote for the edge. Hence, the Hough space is computed for only a small region around the location of the edge. The bin count is used as the evidence strength for the experiment.

Experimental Protocol

The images used were 513×513 pixels. They were defined as one of two types: 1) A vertical step edge superimposed with a masking grating and noise, 2) just the masking grating with noise. The grating is characterized by the amplitude, the frequency, the orientation relative to the vertical direction, the fraction of the period when it is high/low (i.e., the *duty cycle*) and its phase angle relative to the center column of the image. We use a grating that is a square wave with a 50% duty cycle and a half period W of 16 pixels. The phase ϕ of the grating is the offset of the rising edge of the grating from the rising edge of the vertical edge of interest. When the offset is zero, the phase is 0 degrees and the grating has a constructive interference. When the offset is equal to W, the phase is 180 degrees, and the grating has a destructive interference on the vertical edge. For the experiment the phase is fixed at 180 degrees, i.e., the offset was W. The orientation θ of the grating with respect to the edge is varied through a set of possible angles. The grating is rotated about the center pixel of the image. The contrast C_e of the edge and the grating C_g (which is specified in terms of fractions of the mean gray value L_0) are also varied. Similarly the standard deviation, σ_η, of the Gaussian noise also takes values that are fractions of L_0. The mean, μ_η, of the noise is assumed to be zero. The values of the parameters of the experiment are as follows: $L_0 = 100; C_m =$ 10% of $L_0; W =$ 16 pixels; $\sigma_\eta =$ 20% of $L_0; \mu_\eta =$ $0; \phi =$ 180 degrees (destructive phase) ; $C_e =$ 2% of $L_0, \ldots,$ 26% of $L_0; \theta = 0, 1, 3, 5, 45, 90$ degrees.

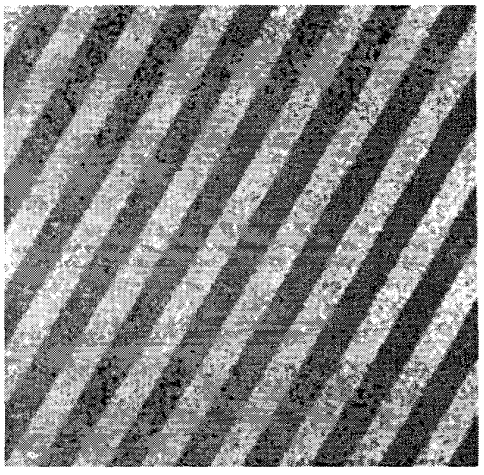
To create a no-edge image the following procedure is used: Create a grating image with high and low values G_h and G_l given by $G_h = L_0 + \frac{L_0 \times C_m}{2}$ and $G_l = L_0 - \frac{L_0 \times C_m}{2}$, and orientation is θ degrees, with respect to the vertical edge. The phase of the grating is assumed to be π at the center pixel (r_c, c_c). The gray value $f(r, c)$ is computed as follows. First, the following function is evaluated at each pixel (r, c).

$$y(r,c) = \sin\left[((r - r_c)\cos(\pi\theta/180) + (c - c_c)\sin(\pi\theta/180) + W\phi/180)\pi/W)\right] \quad (1)$$

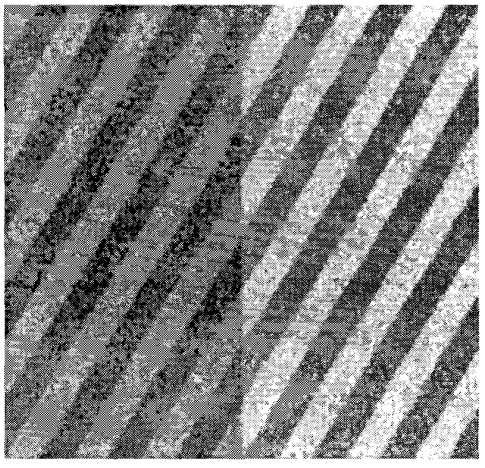
Next, if $y(r, c) \geq 0.0$, $f(r, c) = G_h$; or else, $f(r, c) = G_l$. Next, convolve the image with a 2×2 mask with all entries equal to unity to get the no-noise, no-edge image. Finally, add zero mean Gaussian noise with standard deviation σ_η to each pixel value to get the final noisy no-edge image.

To create an edge image the following procedure was followed: Create a grating image with high and low values G_h and G_l given by $G_h = L_0 + \frac{L_0 \times C_m}{2}$ and $G_l = L_0 - \frac{L_0 \times C_m}{2}$, and orientation θ as before. Next, create a vertical step edge image with a step edge one column to the right of the center column of the image. The high and low values, S_h and S_l, for the step are given by the equations: $S_h = \frac{L_0 \times C_e}{2}$ and $S_l = -\frac{L_0 \times C_e}{2}$. Add the rotated grating image to the edge image to get a edge image with grating. Convolve the image with a 2×2 mask with all entries equal to unity to get the basic masked edge image. Finally, add zero mean Gaussian noise with standard deviation σ_η to each pixel value to get the final noisy no-edge image.

Two sample images used in the experiment are shown in Fig. 2. The image (a) has only the grating at an orientation of 45 degrees. The image (b) has the grating at the same orientation as the image on the right with an additional vertical edge.



(a)



(b)

Fig. 2. Sample images for the experiment. The image (a) has only the grating at an orientation of 45 degrees. The image (b) has the grating at the same orientation as the left as well as the vertical edge.

IV. Results

In this section, we report the results of our analysis of the two line detection algorithms outlined in Section III, using the methodology outlined in Section II. We first discuss the evaluation of Algorithm 1 (Facet + Burns), and finally, compare its performance with that of Algorithm 2 (Facet + Hough).

The first step is to measure the frequency distribution of the evidence strengths for a vertical edge (the target) given images with or without an edge. For Algorithm 1, the evidence strength is the count of vertical edge pixels detected by the Burns line finder in the middle column of the input image. These frequency histograms are shown for the case with $\sigma_\eta = 20$, $\theta = 45°$ in Fig. 3. As expected, the edge images result in a distribution with higher evidence strength values, and this appears to the right of the graph.

The second step is to measure the operating characteristic from these frequency distributions. Fig. 4 shows the probability of false alarms as a function of the probability of misdetection for the full range of possible evidence criteria. This results in a monotonically decreasing function. Notice that as the value of the edge detection criterion is lowered, the probability of false alarm decreases but the probability of misdetection increases. The operating criterion is chosen as the point for which $P(\text{misdetection}) = P(\text{false alarm})$ and is graphically given by the intersection of the 45 degree line and the operating characteristic. The probability of error $P(E) = (P(\text{misdetection}) + P(\text{false alarm}))/2$. For example, the probability of error is 0.29, when $C_e = 6\%L_0$ in Fig. 4.

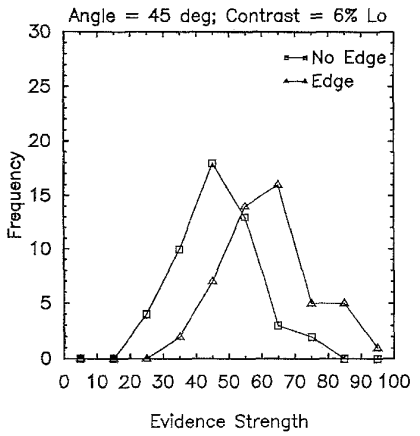


Fig. 3. Step 1—Histogram for evidence strengths for images with and without the vertical edge. In this case the grating angle was 45 degrees, the edge contrast was 6% of mean gray value. The mean gray value was kept constant in all experiments at 100. It can be seen that the evidence strength corresponding to images with vertical edge on the average have higher values than those images without the vertical edge. This histogram corresponds to 100 trials—50 with vertical-edge images and 50 without vertical-edge images.

The third step is to measure the probability of error as a function of the signal to noise ratio. For this experiment, the signal to noise ratio was manipulated by varying the vertical edge contrast, C_e. The results for many orientations, θ, are shown in Fig. 5. This curve falls from a maximum expected error of 0.5 to no errors as the edge contrast increases. If the distributions found with Step 1 are equal variance Gaussian, these functions will be cumulative Gaussians. Such a function is roughly linear in its middle range and this approximation is used to extract the contrast threshold, C_T, at performance equal to 0.25 error. Contrast threshold C_T is the edge contrast required to get a 25% error rate or equivalently 75% detection rate. More precise results can be based on [17]. From this threshold and a template of these functions, one can estimate the probability of error for any contrast.

The fourth step is to measure the effect of the orientation of the irrelevant grating. Fig. 6 shows the contrast threshold for several orientations. It can be seen that for grating orientations greater than 5° and less than 90°, you need a vertical edge contrast of 6% of L_0 to have a 75% detection rate. The performance is remains relatively constant over this range of grating orientations. There is a sudden deterioration of performance as the grating orientations becomes smaller that 5°. In fact, when the grating orientation is 0°, i.e., the grating

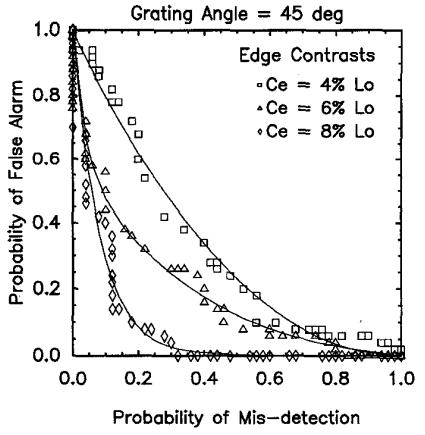


Fig. 4. Step 2—Operating curves for grating angle of 45 degrees and vertical edge contrasts of 4%, 6%, and 8% of mean gray value, L_0. A data point on a curve corresponds to a fixed criterion (threshold that was used in the histogram). When this criterion is varied over the range of evidence strength values and the associated probabilities of misdetection and false alarm are plotted. The equal bias probability of error, $P(E)$, for each contrast is the point of intersection of the 45 degree line (P(misdetection) = P(false alarm)) and the operating curve for that contrast.

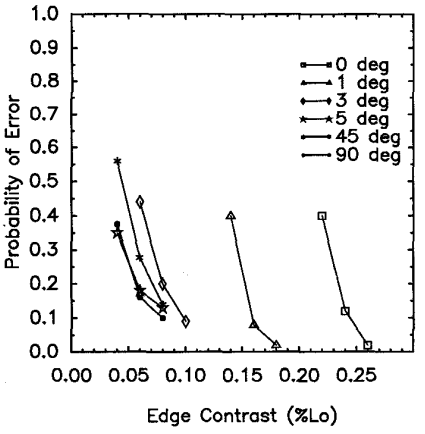


Fig. 5. Step 3—The effect of contrast on the equal bias probability of error. For each orientation it can be seen that as the contrast of the vertical edge increases, the error rate decreases. The contrast threshold, C_T, is the contrast required for a 25% error rate. We can see that when the grating orientation is 0 degrees, an edge contrast of 23% is required in order to get a 75% detection rate.

is in destructive phase, an edge contrast of 23% is required in order to get a 75% detection rate.

The same analysis was done for Algorithm 2. The results are plotted along with the results of Algorithm 1 in Fig. 7. We can see that both algorithms have similar worst case performance (grating orientation of 0°). But, Algorithm 1 has a better asymptotic performance. Furthermore, the performance of both algorithms start deteriorating at around 5°.

Without doing the analysis outlined above, it would have been difficult to hypothesize the worst and asymptotic performance of the two algorithms. In addition, it would have been difficult to predict the location of the knee, i.e., break down point, of the curve. We are currently developing statistical measures for each step in this analysis [11].

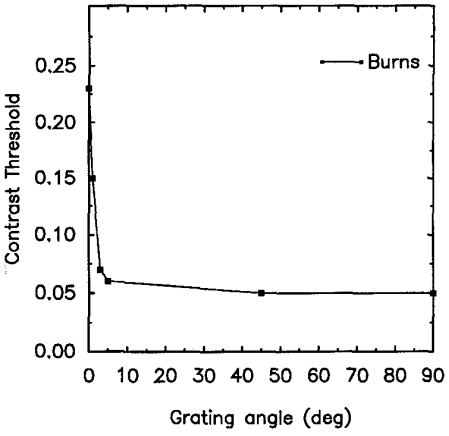


Fig. 6. Step 4—Contrast threshold, the edge contrast required for 75% detection rate, as a function of grating orientation for Algorithm 1 (Facet + Burns). It can be seen that the performance does not deteriorate further when the grating orientation increases from 5°–90°. The performance drops rapidly, i.e., the algorithm breaks down, as the grating orientation decreases from 5°–0°. We can see that when the grating angle is 45°, an edge contrast of 6% of L_0 is required for 75% detection rate, whereas for 0 degrees a 23% contrast is require for the same performance.

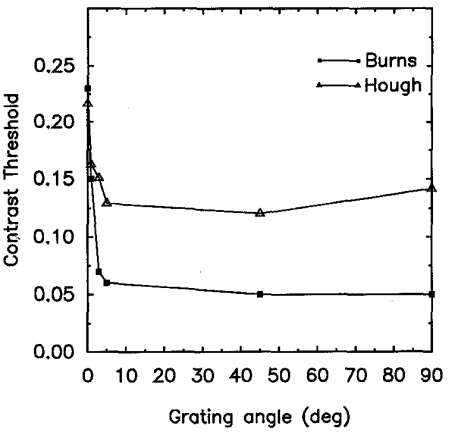


Fig. 7. Contrast threshold, the contrast necessary for a 75% detection rate, as a function of grating angle for Algorithm 1 (Facet + Burns) and Algorithm 2 (Facet + Hough). It can be seen that Algorithm 1 has a better performance than Algorithm 2. Both have equally bad performance when the grating orientation is 0°—both need approximately 23% contrast for 75% detection rate. But when the orientation of the grating is greater than 5°, the breakdown point for both the algorithms, Algorithm 1 performs much better—Algorithm 1 needs about 6% edge contrast whereas Algorithm 2 needs about 13% edge contrast. This behavior would not have been obvious from few operating curves (output of Step 2).

V. Discussion

Advantages of using thresholds:

This methodology builds on previous efforts to formalize detection problems. It follows others in using decision analysis to combine the two kinds of errors into a single error probability given a decision criterion. The current analysis extends this by manipulating a signal-to-noise variable to measure a threshold as is common in psychophysics. Thresholds have several advantages as performance measures:

- Thresholds are defined independent of the algorithm. In our case we used a contrast threshold that gave us a 75% detection rate.
- By defining threshold at a fixed performance rate, one can compare the effect of other variables at a known and comparable S/N ratio.
- Thresholds vary over as large a range as the signal-to-noise ratio. Thus, thresholds will have a dynamic range that is not as restricted as the probability of error measure. The probability of error is only defined over the initial range of perfect (0 error) to chance (0.5 error). This is a problem because very low error probabilities cannot be measured in a practical experiment.
- Thresholds can be measured without factorial experiments by adaptively choosing signal-to-noise ratios appropriate for each set of conditions.. For example, if we need the threshold for which the $P(E)$ is 0.25, there is no need to run the experiment with parameters that give S/N for which the $P(E)$ is far from 0.25. This can substantially reduce the size of experiments. Furthermore, as can be seen in our experiments, we did not have to run the experiment for many orientations between 5° and 90° since the performance does not change much within this range. One can characterize the "knee", i.e., the break down point, of the curve in greater detail and than other parts where there is not much change in performance.
- Thresholds provide a way to measure differences in performance over large ranges. For example, is one algorithm worse than another by a 10% difference in the threshold, a 100% difference, or a 1000% difference? Thresholds give a way of distinguishing between large and small effects.

The use of thresholds is not new in signal processing. The most common example is the the notion of bandwidth. The bandwidth is usually defined as the frequency range within which amplitude response of a filter is is greater than 3 db. In this case, the threshold used is the 3 db amplitude response.

Analyzing algorithm behavior and design of better algorithms:

Since different algorithms can be compared using this methodology, it can be used a a tool for understanding the behavior of different algorithms. For example, why should one algorithm have better asymptotic performance than other? What parameters in the algorithm control the location of the break down point of the curve? Can they be modified to suit our requirements? Furthermore, these performance curves can be used to design better algorithms using the good features of various algorithms.

Summarizing performance curves:

An important point is that this methodology allows us to summarize many operating curve by a few performance curves. In our experiment, each histogram and corresponding operating curve resulted from 100 trials. Each curve in Step 3 of the methodology represents three operating curves or 300 trials. The final curve in Step 4 of the methodology represents $3 \times 6 = 18$ operating curves, or $3 \times 6 \times 100 = 1800$ trials. In contrast, most methodologies existing in the literature today provide only the operating curves (Step 2 of our methodology). Thus, our methodology allows the researcher to convey more information in a meaningful way.

Analytic performance evaluation:

In case an analytic model is available, it is not necessary to run the experiments to compute the operating curve. In fact, the probabilities can be computed form the analytic expression for the probabilities of mis-detection and false alarm. But we still encounter the problem of summarizing the numerous operating curves. Our methodology can be applied, without modification to the analytic results, just as it is applied to the empirical results. Thus, the analytic results can be summarized just as we summarized the empirical results. Furthermore, the there are cases when either the analytic model of an algorithm is not numerically tractable or is not known. In such cases, it is possible to approach the performance evaluation problem in a quantitative fashion. For more details on analytic performance evaluation of vision algorithms, please see [18].

Applications:

A strength of the methodology is that it can be applied to any detection problem. The line detection example developed in this paper was for demonstrating the application of this methodology. This methodology has been partially adapted for performance evaluation of object recognition algorithms [15] and machine inspection algorithms [16]. The key steps to applying this methodology to any algorithm are i) converting the algorithm into a detection algorithm, and ii) choosing the appropriate signal variable to use as the threshold.

Another appropriate example where our methodology could be used is the detection of corners and junctions [19], [20]. To analyze corner, consider an image that can contain a curve with or without a single corner. The corner detection algorithm outputs an evidence strength that indicates whether there is a corner in an image. We can manipulate the angle of the corner to find the signal-to-noise threshold. This performance measure can be used to study the effect of other variables such as the length of the lines making up the corner. A similar approach can be used to analyze junction detectors.

To analyze automatic target detection algorithms, input images would either contain image of the target or no target. The algorithm would first have to detect the presence or absence of the target in the image. Now, one could fix all the variables (distance, shape of target, etc.) except the contrast of the signal. The contrast could be used to control the signal-to-noise ratio and the variable of interest could be the specular reflectance of the target. One could study how the performance deteriorates as the specular reflectance increases.

In the case of inspection of machined parts, the vision algorithm algorithm decides whether a machine part is satisfactory ("within spec") or not ("out of spec"). This is a detection task. The errors are either misdetection errors or false alarm errors. In this case, the degree of defect in the machined part could be used as the signal variable. For more details see [16]

The task of pose error estimation can be converted into a detection task by asking the question: Is the estimated pose of the object within specific error bounds or not? A measure of error could be computed as follows. After the pose of an object is estimated, the average distance between the vertices of the original object and the back-projected object could be

used as an error measure. A threshold on this error makes converts the problem into a detection task. For more details please refer to [15].

VI. Conclusion

We describe a methodology for characterizing and summarizing the performance of any detection algorithm. It extends the previous applications of decision analysis by the addition of threshold measures inspired from psychophysics. This is a general methodology that can be applied to any detection algorithm.

Note: In a recent paper [21] we have discussed how to choose optimal operating points when we are given an ROC curve, the prior probability of target and no-target, and the the costs associated with each decision. In order to compute the optimal operating points, one has to fit smooth curves through empirical TOC curves and histograms. This has been addressed in [22].

Acknowledgment

The authors would like to thank H. Barlow, H. Barrett, and C. Metz for their comments, and R. Vishvanathan for providing us with an implementation of the Burns line finder.

References

[1] E. S. Deutsch and J. R. Fram, "A quantitative study of the orientational bias of some edge detector schemes," *IEEE Trans. Comput.*, vol. C-27, no. 3, pp. 205–213, Mar. 1978.

[2] I. E. Abdou and W. K. Pratt, "Qualitative design and evaluation of enhancement/thresholding edge detector," *Proc. IEEE*, vol. 67, no. 5, pp. 753–763, 1979.

[3] T. Peli and D. Malah, "A study of edge detection algorithms," *Comput. Graphics Image Processing*, vol. 20, pp. 1–21, 1982.

[4] L. Kitchen and A. Rozenfeld, "Edge evaluation using local edge coherence," *IEEE Trans. Sys., Man Cybern.*, vol. SMC-11, no. 9, pp. 597–605, 1981.

[5] R. M. Haralick and J. S. J. Lee, "Context dependent edge detection and evaluation," *Pattern Recognition*, vol. 23, no. 1/2, pp. 1–19, 1990.

[6] H. L. Tan, S. B. Gelfand, and E. J. Delp, "A comparative cost function approach to edge detection," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, pp. 1337–1349, Dec. 1989.

[7] P. K. Sahoo, "A survey of thresholding techniques," *Computer Vision, Graphics, and Image Processing*, vol. 41, no. 2, pp. 233–260, 1988.

[8] F. W. Campbell and J. J. Kulikowski, "Orientation selectivity of the human visual system," *J. Physiology*, vol. 187, pp. 437–445, 1966.

[9] D. M. Green and J. A. Swets, *Signal Detection Theory and Phsychophysics*. New York: Robert E. Krieger Pub., second ed., 1974.

[10] G. A. Gescheider, *Psychophysics: Methods, Theory, and Application*. Hillsdale, NJ: Erlbaum, 1985.

[11] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision, Vol. 1, 2)*. Reading, MA: Addison-Wesley, 1992.

[12] J. B. Burns, A. R. Hanson, and E. M. Riseman, "Extracting straight lines," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PMI-8, pp. 425–455, July 1986.

[13] T. Kanungo, M. Y. Jaisimha, R. M. Haralick, and J. Palmer, "An experimental methodology for performance characterization of a a line detection algorithm," in *Proc. of SPIE Symp. Advances Intell. Syst.*, Boston, MA, Nov 1990, vol. 1385, pp. 104–112.

[14] ——, "A quantitative methodology for analyzing the performance of detection algorithms," in *Proc. IEEE Int. Conf. Comput. Vision*, Berlin, May 1993, pp. 247–252.

[15] O. I. Camps, L. G. Shapiro, and R. M. Haralick, "Recognition using prediction and probabilistic matching," in *Proc. IEEE/RSJ Fifth Int. Conf. Intell. Robots*, Raleigh, NC, July 1992, pp. 1044–1052.

[16] B. R. Modayur, L. G. Shapiro, and R. M. Haralick, "Visual inspection of machined parts," in *Proc. IEEE Conf. Comput. Vision Pattern Recognition*, Champaign, IL, June 1992, pp. 393–398.

[17] D. J. Finney, *Probit Analysis*. Cambridge, Cambridge Univ. Press, 1971.

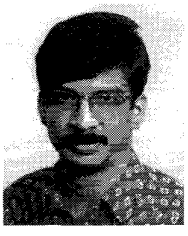
[18] V. Ramesh and R. M. Haralick, "Random perturbation models and performance characterization in computer vision," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, Champaign, IL, June 1992, pp. 521–527.

[19] C. H. Teh and R. T. Chin, "On the detection of dominant points on digital curves," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 859–872, 1989.

[20] J. Matas and J. Kittler, "Contextual junction finder," in *Proc. British Machine Vision Conf.*, Leeds, Sept. 1992, pp. 119–128.

[21] T. Kanungo and R. M. Haralick, "Receiver operating characteristic curves and optimal Bayesian operating points," in *Proc. IEEE Int. Conf. Image Processing*, Washington, D.C., Oct. 1995, vol. 3, pp. 256–259.

[22] T. Kanungo, D. M. Gay, and R. M. Haralick, "Constrained monotone regression of ROC curves and histograms using splines and polynomials," in *Proc. IEEE Int. Conf. Image Processing*, Washington, D.C., Oct. 1995, vol. 2, pp. 292–295.



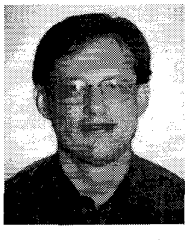
Tapas Kanungo (SM'83), was born on May 5, 1964, in Varanasi, India. He received the B.E. degree in electronics and communication engineering from the Regional Engineering College, Tiruchirapalli, India, in 1986. He received the M.S. in electrical engineering from the University of Washington, Seattle, in 1990, and is currently working toward the Ph.D. degree in electrical engineering at the University of Washington, Seattle.

From 1986–1988, he was with the Computer Science Group at the Tata Institute of Fundamental Research, Bombay, India. He worked at AT&T Bell Laboratories, Murray Hill, NJ, during the summer of 1994, and at the IBM Almaden Research Center, San Jose, CA, during the summer of 1993. From 1988 until the present, he has been a research assistant at the Intelligent Systems Laboratory, University of Washington. His research interests include reconstruction, document understanding, performance evaluation, image databases, morphology, shape decomposition, and human vision.

Mr. Kanungo received the second prize at the Annual Industrial Affiliate's Poster Competition, in 1992, and in 1990, he was a recipient of the Watamull Scholarship.

M. Y. Jaisimha received the B.E. in electronics and communications engineering from the University of Roorkee, India, in 1987, and the M.S. degree in electrical engineering, in 1989, from Tulane University, New Orleans, LA. He is currently a Ph.D. candidate at the Department of Electrical Engineering, University of Washington, Seattle.

Presently, he is a research engineer at the StatSci Division of MathSoft, Inc., Seattle. His research interests are in image analysis, image databases, vector quantization, and pattern recognition.



John Palmer received the B.S. in psychology at the University of Washington. in 1976. and the Ph.D. in psychology at the University of Michigan, in 1984.

He joined the faculty of the Department of Psychology as an assistant professor, from 1984–1992. He is a research scientist in the Department of Psychology at the University of Washington. His research focuses on attention in humans, particularly in the context of visual search and visual memory. He is also interested in the analogous phenomena of selective processing within artificial visual systems, and applications using visual performance models of either humans or artificial systems.

Robert M. Haralick (S'62–S'67–M'69–SM'76–F'84) received the B.A. in mathematics from the University of Kansas, in 1964, the B.S. degree in electrical engineering, in 1966, and the M.S. degree in electrical engineering, in 1967. He completed the Ph.D. degree at the University of Kansas, in 1969.

He joined the faculty of the Electrical Engineering Department at the University of Kansas, where he last served as a professor, from 1975–1978. In 1979, he joined the Electrical Engineering Department at Virginia Polytechnic Institute and State University, where he was a professor and director of the Spatial Data Analysis Laboratory. He is the Boeing Clairmont Egtvedt Professor in electrical engineering at the University of Washington. His recent work is in shape analysis and extraction, using the techniques of mathematical morphology, robust pose estimation, techniques for making geometric inferences from perspective projection information, propagation of random perturbations through image analysis algorithms, and document analysis.

Dr. Haralick served as Vice President of Research at Machine Vision International, Ann Arbor, MI, from 1984–1986. He is a Fellow of the IAPR for his contributions in image processing, computer vision, and mathematical morphology. He serves on the editorial board of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, as well as the editorial board of *Real Time Imaging*. He is an associate editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and an associate editor for *Journal of Electronic Imaging*.