[20] R.L. Kirby, "A Product Rule Relaxation Method," Technical Report 772, Univ. Maryland, 1979.

[21] J. Kittler, W.J. Christmas, and M. Petrou, "Probabilistic Relaxation for Matching Problems in Computer Vision," *Proc. Fourth Int'l Conf. Computer Vision,* pp. 666-673, 1993.

[22] J. Kittler and J. Illingworth, "Relaxation Labeling Algorithms—A Review," *Image and Vision Computing,* vol. 3, no. 4, pp. 206-216, 1985.

[23] G. Lohmann, "An Evidential Reasoning Approach to the Classification of Satellite Images," Technical Report DLR-FB 91-29, Deutsche Forschungsanstalt für Luft- und Raumfahrt, 1991.

[24] D.G. Lowe, *Perceptual Organization and Visual Recognition.* Dortrecht-Boston-London: Kluwer Academic Publishers, 1985.

[25] D.G. Lowe, "Three-Dimensional Object Recognition from Single Two-Dimensional Images," *Artificial Intelligence,* vol. 31, pp. 355-395, 1987.

[26] H. Ogawa, "A Fuzzy Relaxation Technique for Partial Shape Matching," *Pattern Recognition Letters,* vol. 15, no. 6, pp. 349-355, 1994.

[27] K.E. Price, "Relaxation Matching Techniques—A Comparison," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 7, no. 5, pp. 617-623, 1985.

[28] A. Rosenfeld, R.A. Hummel, and S.W. Zucker, "Scene Labeling by Relaxation Operations," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 6, no. 6, pp. 420-433, 1976.

[29] P.L. Rosin and G.A.W. West, "Nonparametric Segmentation of Curves into Various Representations," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 12, pp. 1140-1153, Dec. 1995.

[30] G. Shafer, *A Mathematical Theory of Evidence.* Princeton Univ. Press, 1976.

[31] G. Shafer, "Perspectives on the Theory and Practice of Belief Functions," *Int'l J. Approximate Reasoning,* vol. 4, pp. 323-362, 1990.

[32] P. Smets, "The Combination of Evidence in the Transferable Belief Model," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 12. no. 5, pp. 447-458, May 1990.

[33] V.W. Soo and K. Huang, "On Evidential Relaxation Labeling—A Scheme Toward Knowledge-Based Vision," *J. Information Science and Eng.,* vol. 9, no. 2, pp. 153-175, 1993.

[34] R.V. Southwell, "Stress-Calculation in Frameworks by the Method of Systematic Relaxation of Constraints, I and II," *Proc. Royal Soc. Edinburgh,* vol. 151, no. 872, pp. 56-95, 1935.

[35] A.J. Stoddart, M. Petrou, and J.V. Kittler, "On the Foundations of Probabilistic Relaxation with Product Support," *J. Math. Imaging and Vision,* vol. 9, no. 1, pp. 29-48, July 1998.

[36] G. Temple, "The General Theory of Relaxation Methods Applied to Linear Systems," *Proc. Royal Soc. Edinburgh,* vol. 169, no. 939, pp. 476-500, 1939.

[37] P.H.S. Torr, "Geometric Motion Segmentation and Model Selection," *Phil. Trans. Royal Soc. A,* pp. 1321-1340, 1998.

[38] G. Vosselmann, *Relational Matching.* Berlin-Heidelberg-New York: Springer, 1992.

[39] Z. Wang and G. Klir, *Fuzzy Measure Theory.* Plenum Press, 1992.

[40] G. Winkler, *Image Analysis, Random Fields, and Dynamic Monte Carlo Methods.* Berlin-Heidelberg-New York: Springer, 1995.

[41] Q.X. Wu, "A Correlation-Relaxation-Labeling Framework for Computing Optical Flow-Template Matching from a New Perspective," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 9, pp. 843-853, Sept. 1995.

[42] D.M. Wuenscher and K.L. Boyer, "Robust Contour Decomposition Using a Constant Curvature Criterion," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 13, no. 1, pp. 41-51, 1991.

[43] L.A. Zadeh, "Fuzzy Logic = Computing with Words," *IEEE Trans. Fuzzy Systems,* vol. 4, no. 2, pp. 103-111, 1996.

[44] S.W. Zucker, E.V. Krishnamurthy, and R.L. Haar, "Relaxation Processes for Scene Labeling: Convergence, Speed, and Stability," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 8, no. 1, pp. 41-48, 1978.

[45] S.W. Zucker, Y.G. Leclerc, and J.L. Mohammed, "Continuous Relaxation and Local Maxima Selection: Conditions for Equivalence," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 3, no. 2, pp. 117-127, 1981.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** http://computer.org/publications/dlib.

A Method for Objective Edge Detection Evaluation and Detector Parameter Selection

Yitzhak Yitzhaky, *Member*, *IEEE*, and Eli Peli

Abstract—Subjective evaluation by human observers is usually used to analyze and select an edge detector parametric setup when real-world images are considered. In this paper, we propose a statistical objective performance analysis and detector parameter selection, using detection results produced by different detector parameters. Using the correspondence between the different detection results, an estimated best edge map, utilized as an estimated ground truth (EGT), is obtained. This is done using both a receiver operating characteristics (ROC) analysis and a Chi-square test, and considers the trade off between information and noisiness in the detection results. The best edge detector parameter set (PS) is then selected by the same statistical approach, using the EGT. Results are demonstrated for several edge detection techniques, and compared to published subjective evaluation results. The method developed here suggests a general tool to assist in practical implementations of parametric edge detectors where an automatic process is required.

Index Terms—Edge detection evaluation, detector parameters, receiver operating characteristics.

——————————— ✦ ———————————

1 INTRODUCTION

ALTHOUGH implementations of most edge detectors usually involve a prior necessary step of parameter selection, no automatic parameter selection process exists, mainly due to the strong dependency of the optimal parameter set (PS) of an edge detector on the input image [1], [2], and the difficulty in assessing edge detection results for real-world images.

1.1 Edge Detection Evaluation Approaches

Edge detectors are usually evaluated subjectively by observers [1], [2], [3]. Most of the objective evaluation methods assume knowledge of specific features such as known object boundaries in simple synthetic images. In such cases, the edge detection can be quantitatively measured, based on the known ideal detection considered to be the ground truth (GT) [4], [5], [6]. In real images, manual specification of the edges was applied to form a GT [7], [8]. A method called *local edge coherence* measures local properties of continuation and thinness by examining local neighborhoods surrounding the detected edge points [9] and was later generalized to arbitrary surrounding neighborhoods size [10], and to deal with turning and intersecting edge points [11]. However, an assumption of the same specific local properties may not be desired by all edge detectors and edge detection applications. For example, in Peli's detector [12], the brighter and darker sides of the edge are taken into account, and a two-line (bright and dark) edge outline is desired to represent the two sides.

The method we propose here performs automatic statistical analysis of the correspondence of detection results produced using different detector parameters. Statistical measures are used to estimate the GT considering the trade off between true and false edges, and to extract the best detector's PS. The process does not enforce any criteria regarding the output format of the detector and avoids both the use of human ratings and manual specification of

———————————

● *The authors are with the Schepens Eye Research Institute, Harvard Medical School, 20 Stanford Street, Boston, MA 02114.*
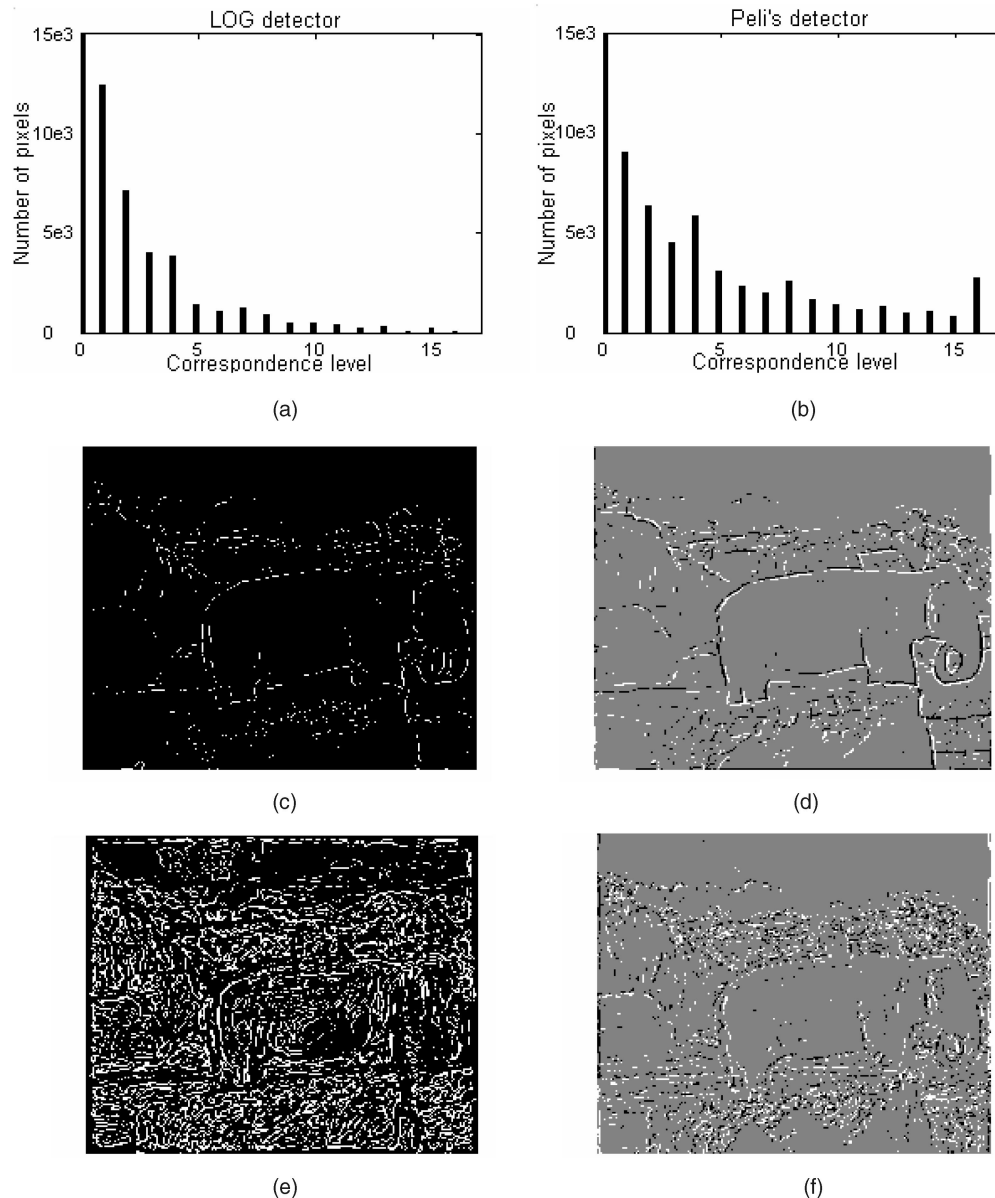*E-mail: {eli, itzik}@vision.eri.harvard.edu.*

Fig. 1. Different correspondence results for the LOG (leftward graphs) and Peli's (rightward graphs) edge detectors applied to the "Elephant" image. (a) and (b) are the distributions of edge locations correspondences obtained from the sum of 16 implementations of the detectors with different PSs. (c) and (d) are locations with at least 50 percent correspondence (between 9 and 16). (e) and (f) are locations with less than 50 percent correspondence.

the GT and, thus, it provides a general technique for automatic parameters selection.

1.2 Parametric Edge Detection Methods

Typically, edge detection is performed by enhancing the image edges with a linear filter that approximates a first or a second derivative, followed by a decision stage in which a threshold is applied. An earlier smoothing step is frequently applied using a Gaussian mask [13]. While the smoothing operation reduces noise and minor edges, it may cause a distortion of edge locations. The severity of that distortion depends mainly on the size of the smoothing operator, and on the spatial shapes of the edges [14], [15], [16]. Edge enhancement operators that combine smoothing and derivation can then be a first derivative of Gaussian (Canny [17]) and a Laplacian of the Gaussian (LOG) [18]). Peli [12] developed a visual model-based edge detection method, where the visual receptive field like filtering channels are used at the enhancement stage, and the threshold is the contrast sensitivity of the human eye. Information from different frequency bands (spatial scales) in the image is combined to produce the final edge detection. The parameters in Peli's detector are the center frequency and the width of its edge enhancing band pass filters.

2 ANALYSIS AND EVALUATION OF EDGE DETECTION RESULTS

Here, a statistical analysis is performed to both produce an EGT and then use it to select the detector parameter values for an image. The edge detection techniques used to demonstrate the proposed analysis are the classic single-band LOG and Canny, and Peli's technique. The proposed method does not compare between detection approaches, but it performs a general automatic self-evaluation and parameter selection within a range of examined detector parameters. It assumes that the best detection of a certain edge detector in a given image is that which is most consistent with the verity of detection outputs that can be produced by the detection algorithm when different parameters are used. As a result of the edge drifting scale space affects [15], this consistency may be

TABLE 1
Definition of the Outcome Probabilities: (a) In the GT Estimation Process and (b) for the Best PS Selection Process,
According to Statistical Decision Theory Terminology

Statistical term	Probability definition	Probability description	Notation
True Positive (Hit)	$P(E_{PGT_i} / E_{D_j})$	Points decided as 'edges' in the PGT_i, and coincide with edges in D_j.	TP_{PGT_i, D_j}
False Positive (False alarm)	$P(E_{PGT_i} / NE_{D_j})$	Points decided as 'edges' in the PGT_i, but coincide with non-edges in D_j.	FP_{PGT_i, D_j}
True Negative (Correct rejection)	$P(NE_{PGT_i} / NE_{D_j})$	Points that were decided as 'non-edges' in the PGT_i, and coincide with non-edges points in D_j.	TN_{PGT_i, D_j}
False Negative (Miss)	$P(NE_{PGT_i} / E_{D_j})$	Points that were decided as 'non-edges' in the PGT_i, and coincide with edges points in D_j.	FN_{PGT_i, D_j}

(a)

True Positive (Hit)	$P(E_{D_j} / E_{EGT})$	Edge points in D_j, which coincide with 'edge' points in the EGT.	$TP_{D_j EGT}$
False Positive (False alarm)	$P(E_{D_j} / NE_{EGT})$	Edge points in D_j, which do not coincide with 'edge' points in the EGT.	$FP_{D_j EGT}$
True Negative (Correct rejection)	$P(NE_{D_j} / NE_{EGT})$	Non-edge points in D_j, which coincide with the 'non-edge' points in the EGT.	$TN_{D_j EGT}$
False Negative (Miss)	$P(NE_{D_j} / E_{EGT})$	Non-edge points in D_j, which coincide with the 'edge' points in the EGT.	$FN_{D_j EGT}$

(b)

"E" and "NE" denote "edges" and "nonedges," respectively.

compromised in some locations in the image when different scales are produced by different values of a detector smoothing parameter.

First, an EGT is automatically constructed, given a range of detection results obtained from different detector PSs. Then, the single PS that provides the best match to the EGT is identified. Since the best PS depends on the image [1], [2], the results are appropriate to the specific image and possibly to similar images, but the process should be carried out again for dissimilar images. The range of parameters to be tested should be large enough to cover a wide range of detection results from noisy to sparse. Here, we implemented 16 PS combinations, which appear to reasonably cover such a range; however, higher number of PSs may be used when edge detections with very fine differences between them have to be evaluated. The actual parameter values depend on the actual implementation setup such as the image intensity range and size. The standard deviation of the Gaussian in the Canny detector here is from 0.7 to 1.3 in steps of 0.2, the high threshold value is from 0.05 to 0.35 in steps of 0.1, and the low threshold is set to one third of the high. In the LOG detector, the standard deviation is from 2.5 to 4 in steps of 0.5 and the threshold is from 0.1 to 0.85 in steps of 0.25. The thresholding here is performed on the absolute value of the difference between the values

of the LOG from opposite sides of the zero-crossing. In Peli's detector, the filters' center frequencies and bandwidths are from their original values (1 octave wide and 1 octave apart) to 30 percent higher, in steps of 10 percent. The LOG and Peli's detectors are used for demonstration and analysis, while Canny detector (used also in [1]) is utilized for comparison to subjective results in Section 3. The images used here ("Elephant," "Grater," "Turtle," and "Airplane") were taken from the Web site [19], and are also shown in [1].

2.1 Ground Truth Estimation

N edge detection results $D_j (j = 1, \ldots, N)$ are computed using N combinations of the detector's parameters. The results are then tested for correspondence. A pixel location identified as an edge by all N detector setups will have the highest correspondence (N), and a location identified as an edge by only one detector setup will have the lowest. A histogram of the correspondence for $N = 16$ is presented in Figs. 1a and 1b for the LOG and Peli methods. In Peli's detector, correspondence is accounted only if a bright edge matches a bright edge or a dark edge matches a dark edge. Usually, points with higher correspondence belong to more distinct luminance edges and are considered to be more related to boundaries of main objects in the
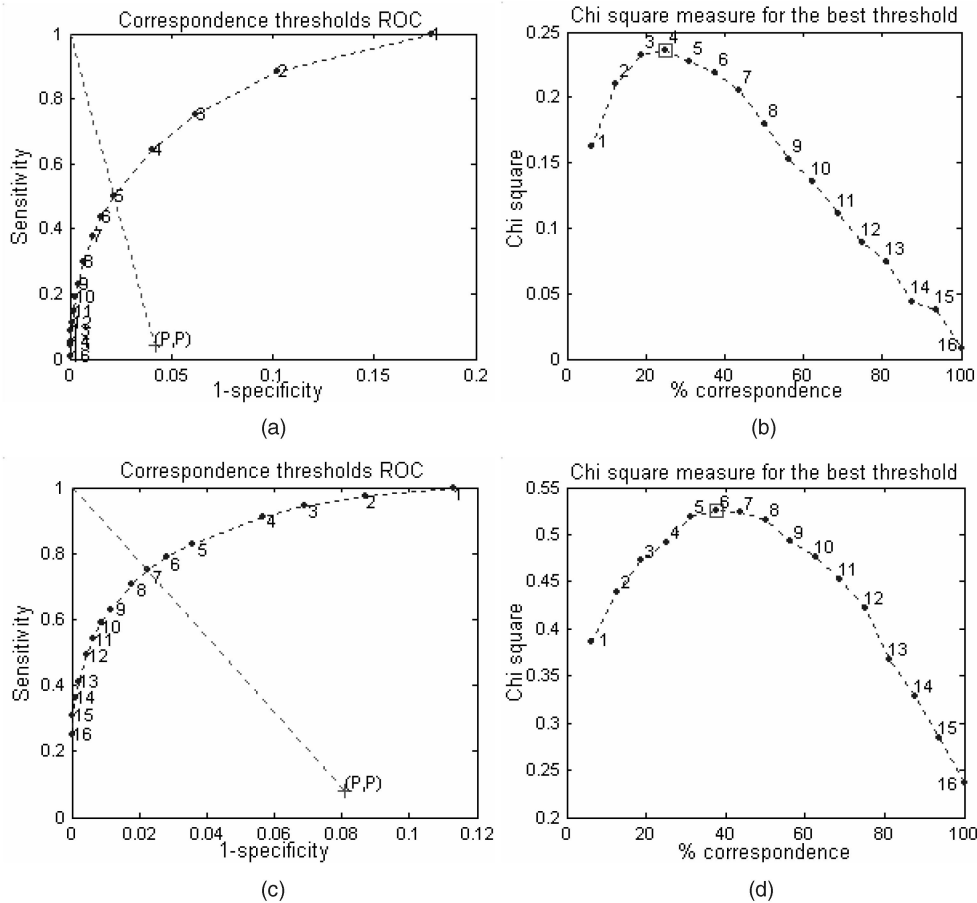
Fig. 2. CT selection measures for the LOG (upper graphs) and Peli's (lower graphs) detectors implemented on the Elephant image. (a) LOG: The average ROC curve (across the 16 PSs) evaluated from 16 CT levels at the 16 correspondence levels; the best CT level according to the intersection between the *diagnosis line* and the ROC curve is 5 (31 percent correspondence). (b) LOG: The Chi-square measure for the best CT level shows a maximum at CT level 4 (25 percent correspondence). (c) Same as (a), but for Peli's detector. According to the CTROC the best CT level is 7 (38 percent correspondence). (d) Same as (b), but for Peli's detector. According to the Chi-square measure, the best CT level is 6 (31 percent correspondence).

image rather than noise or minor features that may appear disturbing to the viewer. Examples of this behavior are presented in Figs. 1c, 1d, 1e, and 1f, where locations with at least 50 percent correspondence are presented in Figs. 1c and 1d, and locations with less then 50 percent correspondence are presented in Figs. 1e and 1f. These examples illustrate that statistical correspondence between the different detections can be used as a basis for a distinction between true and false identified edges. N detection results produce N possible correspondence levels, thus, N possible *correspondence threshold* (CT) values can be applied to distinguish between points with correspondence higher, equal or lower than that CT.

2.1.1   Construction of a CT ROC (CTROC) Curve

ROC analysis [20], [21] is applied here to find the best CT that forms the EGT by considering the tradeoff between increasing the information and decreasing the noise in the detection result.

To form the CTROC curve, a CT is applied at each of the N correspondence levels. In each CT level i, points with correspondence above or equal to the CT will be considered as "edges" and the other points will be considered as "nonedges," forming a *potential* GT (PGT_i) for that CT level i. Examples of such PGTs are presented in Figs. 1c and 1d for CT levels $i = N/2 + 1$ (50 percent correspondence). Each PGT map associated with each CT level (PGT_i) is then examined against the entire N separate detection results D_j generating four probabilities for each individual match. Table 1a presents these outcome probabilities in the GT estimation process according to statistical decision theory terminology. For each PGT_i,

the average of all the probabilities resulting from its match with each of the individual detection results j is calculated. Defining an "edge" pixel as "1" and a "nonedge" pixel as "0", these averages are:

$$\overline{\mathrm{TP}}_{\mathrm{PGT}_i} = \frac{1}{N}\sum_{j=1}^{N}\mathrm{TP}_{\mathrm{PGT}_i,\mathrm{D}_j}$$

$$= \frac{1}{N}\sum_{j=1}^{N}\left(\frac{1}{\mathrm{K}\cdot\mathrm{L}}\sum_{k=1}^{\mathrm{K}}\sum_{l=1}^{\mathrm{L}}\mathrm{PGT}_{i_1}\cap\mathrm{D}_{j_1}\right) \qquad (1)$$

$$\overline{\mathrm{FP}}_{\mathrm{PGT}_i} = \frac{1}{N}\sum_{j=1}^{N}\mathrm{FP}_{\mathrm{PGT}_i,\mathrm{D}_j}$$

$$= \frac{1}{N}\sum_{j=1}^{N}\left(\frac{1}{\mathrm{K}\cdot\mathrm{L}}\sum_{k=1}^{\mathrm{K}}\sum_{l=1}^{\mathrm{L}}\mathrm{PGT}_{i_1}\cap\mathrm{D}_{j_0}\right) \qquad (2)$$

$$\overline{\mathrm{TN}}_{\mathrm{PGT}_i} = \frac{1}{N}\sum_{j=1}^{N}\mathrm{TN}_{\mathrm{PGT}_i,\mathrm{D}_j}$$

$$= \frac{1}{N}\sum_{j=1}^{N}\left(\frac{1}{\mathrm{K}\cdot\mathrm{L}}\sum_{k=1}^{\mathrm{K}}\sum_{l=1}^{\mathrm{L}}\mathrm{PGT}_{i_0}\cap\mathrm{D}_{j_0}\right) \qquad (3)$$

$$\overline{\mathrm{FN}}_{\mathrm{PGT}_i} = \frac{1}{N}\sum_{j=1}^{N}\mathrm{FN}_{\mathrm{PGT}_i,\mathrm{D}_j}$$

$$= \frac{1}{N}\sum_{j=1}^{N}\left(\frac{1}{\mathrm{K}\cdot\mathrm{L}}\sum_{k=1}^{\mathrm{K}}\sum_{l=1}^{\mathrm{L}}\mathrm{PGT}_{i_0}\cap\mathrm{D}_{j_1}\right), \qquad (4)$$
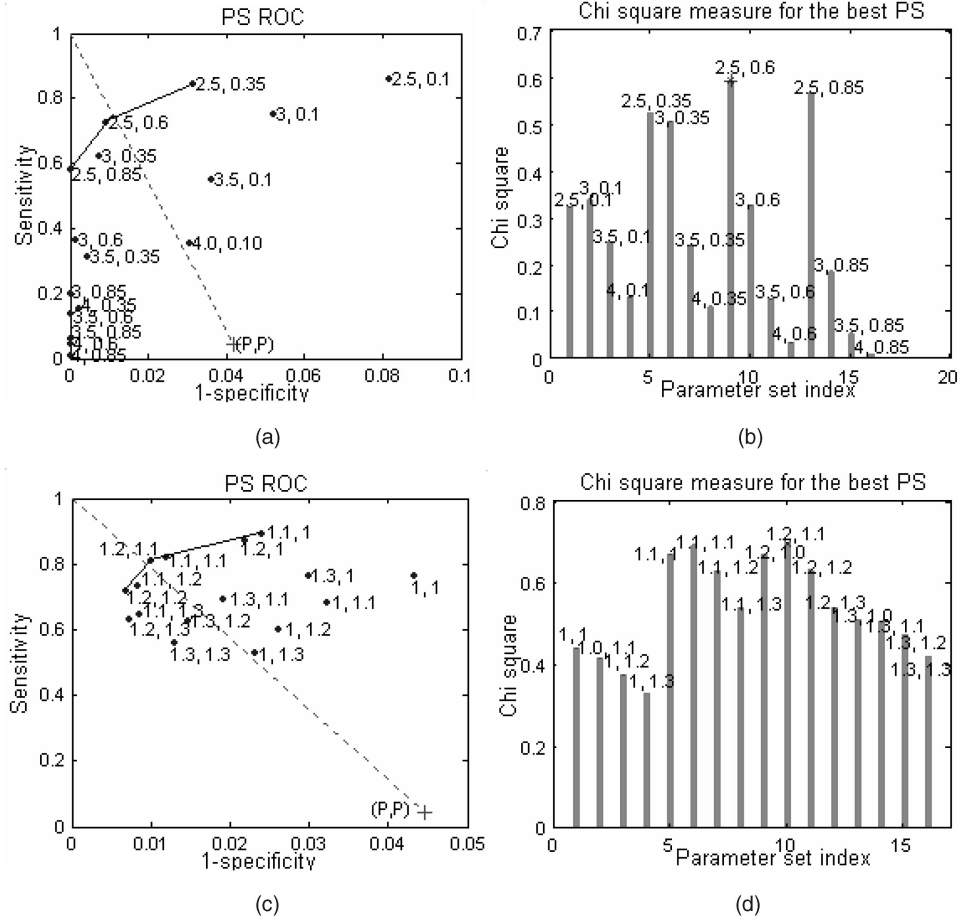
Fig. 3. PS selection measures for the LOG (upper graphs) and Peli's (lower graphs) detectors. (a) The match points in the ROC plane for all the examined PSs. The numbers near the points are the standard deviation and the threshold parameters. The potentially best PSs are connected forming the PSROC curve, and the best set is (2.5, 0.6). (b) Chi-square measure for the match results between the different detections and the EGT shows a maximum for the parameters (2.5, 0.6). (c) Same as (a), but for Peli's detector. The numbers near the points are the numbers with which the detector's center frequency and the frequency bandwidth of the enhancement filters were multiplied. The best set is (1.2, 1.1). (d) Same as (b), but for Peli's detector. The best PS according to the Chi-square measure was (1.2, 1.1).

where K and L are the dimensions of the image, $\text{PGT}_{i_1} \text{PGT}_{i_0}$ are the pixels in the PGT_i decided as "edges" and "nonedges," respectively, and D_{j_1} and D_{j_0} are, respectively, pixels detected as edges and nonedges in the detection j. The expressions inside the parentheses are described in Table 1a. Each point in the resulting CTROC curve shows the average match between all the detection results and the PGT_i. The coordinates of this point are the average TP rate (TPR) and the average FP rate (FPR). The TPR, known also as *sensitivity*, is defined as:

$$\text{TPR}_{\text{PGT}_i} = \frac{\overline{\text{TP}}_{\text{PGT}_i}}{\text{P}}, \qquad (5)$$

where the *prevalence* P is average relative number of positive detections (edges) in the detection results [20], $\text{P} = \overline{\text{TP}}_{\text{PGT}_i} + \overline{\text{FN}}_{\text{PGT}_i}$, $\forall i$. The FPR, known also as $(1 - \textit{specificity})$, is defined as:

$$\text{FPR}_{\text{PGT}_i} = \frac{\overline{\text{FP}}_{\text{PGT}_i}}{1 - \text{P}}, \qquad (6)$$

where $1 - \text{P} = \overline{\text{FP}}_{\text{PGT}_i} + \overline{\text{TN}}_{\text{PGT}_i}$. The N points in the CTROC plane will then be: $(sensitivity, 1 - specificity) = (\text{TPR}_{\text{PGT}_i}, \text{FPR}_{\text{PGT}_i})$, forming the CTROC curve. As the CT level i decreases, TP_{PGT_i} and FP_{PGT_i} increase and, thus, the *sensitivity* and $(1 - specificity)$ both increase. Therefore, the CTROC curve is monotonically increasing. Figs. 2a and 2c show the resulting CTROC curves

constructed by implementing a CT at each of the correspondence levels that appear in Figs. 1a and 1b, respectively.

2.1.2 Extraction of the Estimated Ground Truth

We define the best CT as the one that forms a PGT that gives the best match to the entire edge detection results. This PGT will be our EGT. Using the calculated statistical data, we will find the best CT by implementing two measures: the CTROC curve using a *diagnosis line* and the *Chi-square* test [20]. Two such measures are usually used in areas such as detection theory [21] and psychophysics [20], and may give a more reliable evaluation than only one measure.

Measure 1: An ROC curve with a *diagnosis line*. The optimal CT forms a point in the CTROC plane that is the closest to the ideal point (0, 1). First, we define a *diagnosis line* by connecting the points (P, P) and (0, 1) on the CTROC plane. Then, the best CT will be at the intersection of the *diagnosis line* and the ROC curve [20], and practically the closest point to the intersection. The *diagnosis line* is the line on the ROC plane where FP = FN. When P = 0.5, this line is the diagonal line (1, 0), (0, 1), but, in the general case, it is the line (P, P), (0, 1). Implementations of this are shown in Figs. 2a and 2c.

Measure 2: Chi-square. The Chi-square test of the optimal CT can be written as [20]:

$$\chi^2_{\text{PGT}_i} = \frac{(sensitivity - Q_{\text{PGT}_i})}{(1 - Q_{\text{PGT}_i})} \cdot \frac{(specificity - (1 - Q_{\text{PGT}_i}))}{Q_{\text{PGT}_i}}, \qquad (7)$$

where $Q_{\text{PGT}_i} = \overline{\text{TP}}_{\text{PGT}_i} + \overline{\text{FP}}_{\text{PGT}_i}$. A higher $\chi^2_{\text{PGT}_i}$ indicates a better CT. The best CT will be the value of i that maximizes $\chi^2_{\text{PGT}_i}$. Figs. 2b
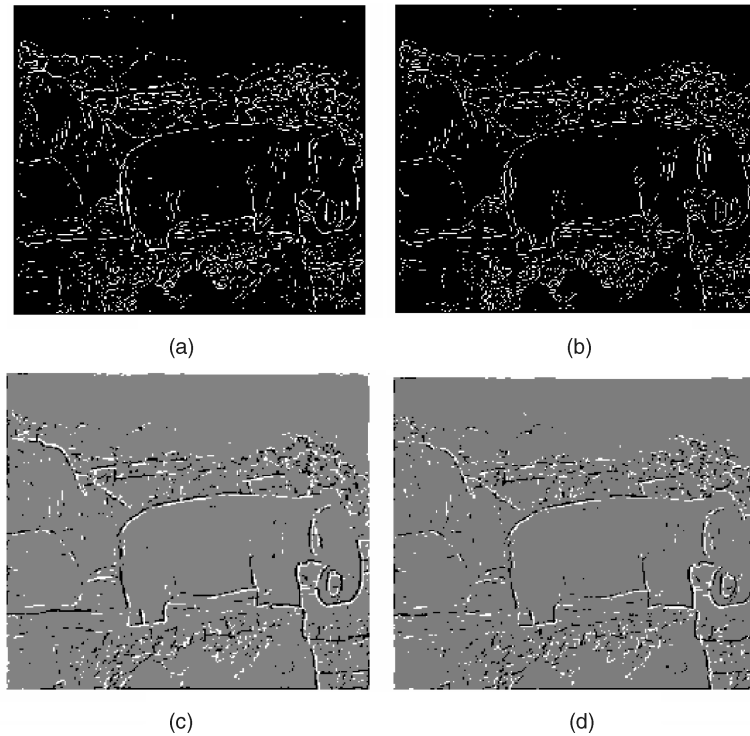
Fig. 4. Resulting edge maps for the LOG (upper row) and Peli's (lower row) detectors. (a) and (c) The EGT using the selected CT according to the Chi-square measure. (b) and (d) The edge detection using the selected PS.
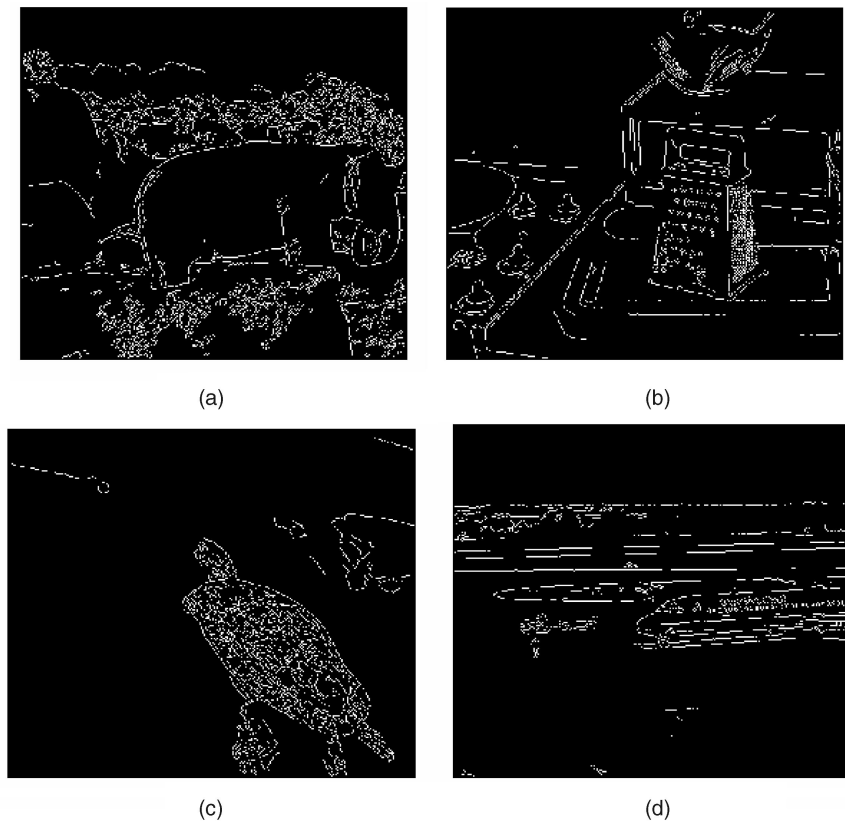

Fig. 5. Detection results using the Canny detector with the automatically selected PSs. The images "Elephant," "Grater," "Turtle," and "Airplane" were taken from [19], and are also shown in [1]. These results can be compared to results presented in the Web site [23] with the PSs selected in [1].

and 2d show the values of the Chi-square measure for different CT levels for the LOG and Peli's detectors, respectively. The results can be compared to those of the CTROC in Figs. 2a and 2c. It can be seen that, in these examples, a difference of one level occurred. The EGTs formed by the CTs selected according to the Chi-square test are presented in Figs. 4a and 4c. Implementations of the two measures with many other images usually gave close results (mostly zero or one level difference). The EGT can be used as the final detection.

However, it is often required to implement an edge detection operation with many similar images. In such cases, the best detector's PS can be extracted from one image and then be used for the others.

2.2   Extraction of the Best PS

The best PS of the detector is the one that forms the detection that gives the best match to the EGT. The best match is decided here also according to ROC and Chi-square measures. The resulting matches between each detection and the EGT are presented in the PS ROC (PSROC) plane. The formulations of the TPR and the FPR here are similar to (5) and (6), respectively, but with the probabilities $\mathrm{TP_{D_jEGT}}$, $\mathrm{FP_{D_jEGT}}$, $\mathrm{TN_{D_jEGT}}$, and $\mathrm{FN_{D_jEGT}}$, described in Table 1b. The potentially best points here are those points that no other point has both better *sensitivity* and *specificity* [20]. These points can be connected forming an outer bound with regard to the other points. We will term this line: the "PSROC curve." The best PS point in this case is the point on the PSROC curve that is the closest to the *diagnosis line*. Examples for that are shown in Figs. 3a and 3c. A similar concept is the *Pareto front* [22], which is the outer bound that connects the points that represent the best parameters for segmentation algorithms in a fitness-cost space.

The Chi-square measure in this case, $\chi^2_{\mathrm{D_jEGT}}$, will be similar to (7), but with $\mathrm{Q_{D_jEGT_i}}$ instead of $\mathrm{Q_{PGT}}$, where $\mathrm{Q_{D_jEGT_i}} = \mathrm{TP_{D_jEGT}} + \mathrm{FP_{D_jEGT}}$. The match results according to the Chi-square criterion appear in Figs. 3b and 3d, where the highest bars indicate the best PSs. Each of these sets is one of the potentially best sets according to the PSROC measure. The edge detection results formed using the selected parameters are shown in Figs. 4b and 4d.

3   COMPARISON TO SUBJECTIVE PARAMETER SELECTION RESULTS

Results of the proposed objective technique are compared here to subjective parameter selection results [1], using the Canny detector employed in both works. From the 28 images used in [1], we have found two pairs ("Elephant," "Grater") and ("Turtle," "Airplane") that fitted two criteria according to the results in [1]: 1) for each pair the selected PSs were the same, but were different between the two pairs; 2) the differences between the two PSs had to be effective in terms of the change in the "noisiness" of the detection. A noneffective difference would be, for example, a higher standard deviation (less noisiness) and a lower threshold (more noisiness). Although the actual parameter values varies due to different implementations of the algorithm, the automatic technique selected the same set for the first two images and another same set for the second two (according to both ROC and Chi-square measures), similar to the results of the subjective evaluation. Furthermore, in both the subjective and objective determinations, the difference between the two selection results was from a noisier-bound set for the first pair of images to a quieter-bound one for the second pair. The detection results using the selected PSs presented in Fig. 5, show a high similarity to the results preferred by the viewers presented in [23].

4   DISCUSSION

The EGT and the best detector's PS depend on the range of the parameters initially used to produce the various detections. The range of the parameters should therefore extend from forming very noisy detections to forming very sparse ones. This requirement is a drawback since the method is thereby not fully automatic, however, this requirement is less manually demanding than finding the best PS as the proper range can be preselected easily for most images. The purpose of the method developed here is to improve the practical use of parametric edge detectors and to assist with the uncertainty associated with detector parameter selection. The method may be used to evaluate detections from

different edge detectors only if these detectors aim to the same output format. Otherwise, noncorrespondence between detections of different detectors may be a result of the dissimilarity between the desired edge detection formats. Additional information and results of this work may be found on our Web site [24].

REFERENCES

[1]   M. Heath, S. Sarkar, T. Sanocki, and K.W. Bowyer, "A Robust Visual Method for Assessing the Relative Performance of Edge Detection Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 12, pp. 1338-1359, Dec. 1997.

[2]   M. Heath, S. Sarkar, T. Sanocki, and K.W. Bowyer, "Comparison of Edge Detectors: A Methodology and Initial Study," *Computer Vision and Image Understanding,* vol. 69, no. 1, pp. 38-54 Jan. 1998.

[3]   M.C. Shin, D. Goldgof, and K.W. Bowyer, "Comparison of Edge Detector Performance through Use in an Object Recognition Task," *Computer Vision and Image Understanding,* vol. 84, no. 1, pp. 160-178, Oct. 2001.

[4]   T. Peli and D. Malah, "A Study of Edge Detection Algorithms," *Computer Graphics and Image Processing,* vol. 20, pp. 1-21, 1982.

[5]   J.R. Farm and E.W. Deutsch, "On The Quantitative Evaluation of Edge Detection Schemes and Their Comparison with Human Performance," *IEEE Trans. Computer,* vol. 24, no. 6, pp. 616-628, June 1975.

[6]   T. Kanungo, M.Y. Jaisimha, J. Palmer, and R.M. Haralick, "A Methodology for Quantitative Performance Evaluation of Detection Algorithms," *IEEE Trans. Image Processing,* vol. 4, no. 12, pp. 1667-1673, Dec. 1995.

[7]   K. Bowyer, C. Kranenburg, and S. Dougherty, "Edge Detector Evaluation Using Empirical ROC Curves," *Computer Vision and Image Understanding,* vol. 84, no. 1, pp. 77-103, Oct. 2001.

[8]   M.C. Shin, D. Goldgof, and K.W. Bowyer, "An Objective Comparison Methodology of Edge Detection Algorithms Using a Structure from Motion Task, " *Empirical Evaluation Techniques in Computer Vision,* IEEE CS, pp. 235-254, 1998.

[9]   L. Kitchen and A. Rosenfeld, "Edge Evaluation Using Local Edge Coherence," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 11, no. 9, pp. 597-605, Sept. 1981.

[10]   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.

[11]   Q. Zhu, "Efficient Evaluations of Edge Connectivity and Width Uniformity," *Image and Vision Computing,* vol. 14, pp. 21-34, 1996.

[12]   E. Peli, "Feature Detection Algorithm Based on a Visual System Model," *Proc. IEEE,* vol. 90, pp. 78-93, 2002.

[13]   D. Ziouand, S. Tabbone, "Edge Detection Techniques- An Overview," Technical Report no. 195, Dept. of Math. and Informatique, Universite de Sherbrooke, 1997.

[14]   T. Lindeberg, *Scale-Space Theory in Computer Vision.* Dordrecht, The Netherlands: Kluwer Academic, 1994.

[15]   V. Berzins, "Accuracy of Laplacian Edge Detectors," *Computer Vision, Graphics, and Image Processing,* vol. 27, pp. 195-210, 1984.

[16]   M. Shahand, A. Sood, and R. Jain, "Pulse and Staircase Edge Models," *Computer Vision, Graphics, and Image Processing* vol. 34, pp. 321-343, 1986.

[17]   J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Analysis and Machine intelligence,* vol. 8, no. 6, pp. 679-698, Nov. 1986.

[18]   D. Marr and E.C. Hildreth, "Theory of Edge Detection," *Proc. Royal Soc., London B,* vol. 207, pp. 187-217, 1980.

[19]   The Computer Vision/Image Analysis Research Laboratory, Univ. of South Florida, http://figment.csee.usf.edu/~kranenbu/roc.html, Jan. 2002.

[20]   H.C. Kraemer, *Evaluating Medical Tests: Objective and Quantitative Guidelines.* Newbury Park, Calif.: Sage Publications, 1992.

[21]   N.A. Macmillan and C.D. Creelman, *Detection Theory: A User's Guide.* Cambridge: Cambridge Univ. Press, 1991.

[22]   M.R. Everingham, H. Muller, and B.T. Thomas, "Evaluating Image Segmentation Algorithms Using the Pareto Front," *Proc. Seventh European Conf. Computer Vision,* pp. IV:34-48, May 2002.

[23]   The Computer Vision/Image Analysis Research Laboratory, Univ. of South Florida, http://marathon.csee.usf.edu/edge/edge_detection.html, Dec. 2002.

[24]   The Vision Rehabilitation Laboratory, Schepens Eye Research Inst., Boston Mass., http://www.eri.harvard.edu/faculty/peli/papers/Appendix_EdgeDetectionEval.pdf, Dec. 2002.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** http://computer.org/publications/dlib.