**Article:**
Rockett, Peter (2003) Performance Assessment of Feature Detection Algorithms: A Methodology and Case Study on Corner Detectors. IEEE TRANSACTIONS ON IMAGE PROCESSING, 12 (12). pp. 1668-1676. ISSN: 1057-7149

https://doi.org/10.1109/TIP.2003.818041

# Performance Assessment of Feature Detection Algorithms: A Methodology and Case Study on Corner Detectors

Peter I. Rockett

***Abstract*—In this paper, we describe a generic methodology for evaluating the labeling performance of feature detectors. We describe a method for generating a test set and apply the methodology to the performance assessment of three well-known corner detectors: the Kitchen-Rosenfeld, Paler *et al.*, and Harris-Stephens corner detectors. The labeling deficiencies of each of these detectors is related to their discrimination ability between corners and various of the features which comprise the class of noncorners.**

***Index Terms*—Corner detection, feature detection, performance evaluation, ROC curves.**

## I. Introduction

THE requirement for robust and reliable corner detection is well-rehearsed in the literature [1] and in an attempt to meet this, many corner detection algorithms have been proposed over the years [2]. Unfortunately, assessment of the labeling performance of these detectors has been largely subjective: typically detector results on real images are shown – usually with few details of the imaging conditions or detector parameters employed – and a subjective appraisal is made of the labeling performance. In recent years there has been an increasing emphasis on *quantitative* performance evaluation in computer vision (which we briefly review in Section II). In this paper, we present a generic methodology for assessing the performance of feature detectors. To illustrate this methodology, we describe its application to corner detectors, an area where there has been comparatively little work on performance evaluation. We return to the generic nature of the proposed methodology in Section V but in the following sections we focus on the issue of corner detector evaluation. We employ the well-known receiver operating characteristic (ROC) paradigm [3] to assess the performance on the two-class labeling problem (corners versus noncorners). ROCs have the advantage of mapping-out the whole range of detector operating points independent of the priors on a given labeling problem; the slope of the ROC can be related to the posterior probability of class membership [4] and the area under the curve (AUC) is a concise measure of performance [5] although Adams and Hand [6] have pointed-out the need for care when comparing two detectors whose ROCs cross. Central to the generation of an ROC are two labeled datasets: a true feature dataset (here, corners) and a dataset of counter-examples (noncorners) and a key challenge is to obtain two labeled sets of sufficient size to obtain statistically meaningful results. In Section II, we review relevant work on the performance evaluation of feature detectors. In the present work we employ a detailed model of the image formation process to obtain datasets of the necessary cardinality; we describe the generation of the model data in Section III.

In Section IV, we illustrate the application of the methodology presented where we examine the performance and operation of three well-known corner detectors: the Kitchen & Rosenfeld detector [7], the Paler detector [8] and the Harris & Stephens detector [9] also known as the Plessey corner detector. One advantage of the present approach is that it enables us to probe the key issues in the functioning of each of the detectors as well as assessing labeling performance in operation.

We discuss the implications and extensions of this work in Section V before offering concluding remarks in Section VI.

## II. Feature Detector Performance Evaluation

The need for performance evaluation protocols in computer vision has greater gained acceptance in recent years—see Bowyer and Phillips [10] and Courtney *et al.* ([11] and references therein) for a partial survey. Most of the work done to date on evaluating the performance of feature detectors has concerned edge detection algorithms, an area initiated by the seminal work of Abdou and Pratt [12].

Heath *et al.* [13] have reviewed a number of methods for assessing edge detector performance and presented an approach of using manually labeled edges from real images to generate a dataset labeled with the ground truth from which ROC plots were produced. An ingenious aspect of the method of Heath *et al.* was to introduce a "don't know" class which was applied to image regions where it was not possible for a human observer to unambiguously label edges. It is, however, a philosophical issue whether judging a feature detector by its concordance with the opinions of a group of human observers is legitimate—the issue would seem to hinge on the ultimate purpose of the edge detection process. In addition, whereas an edge detector would employ a small image patch (maybe 3–7 pixels square) a human observer would utilize much larger-scale contextual cues.

As far as the quantitative assessment of corner detectors is concerned there has been little published work. Courtney *et al.* [11] describe a Monte Carlo procedure for constructing probability density functions for the Harris and Stephens detector

Manuscript received July 17, 2002; revised March 11, 2003. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Arnaud Jacquin.

The author is with the Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield S1 3JD, U.K. (e-mail: p.rockett@ shef.ac.uk).

Digital Object Identifier 10.1109/TIP.2003.818041

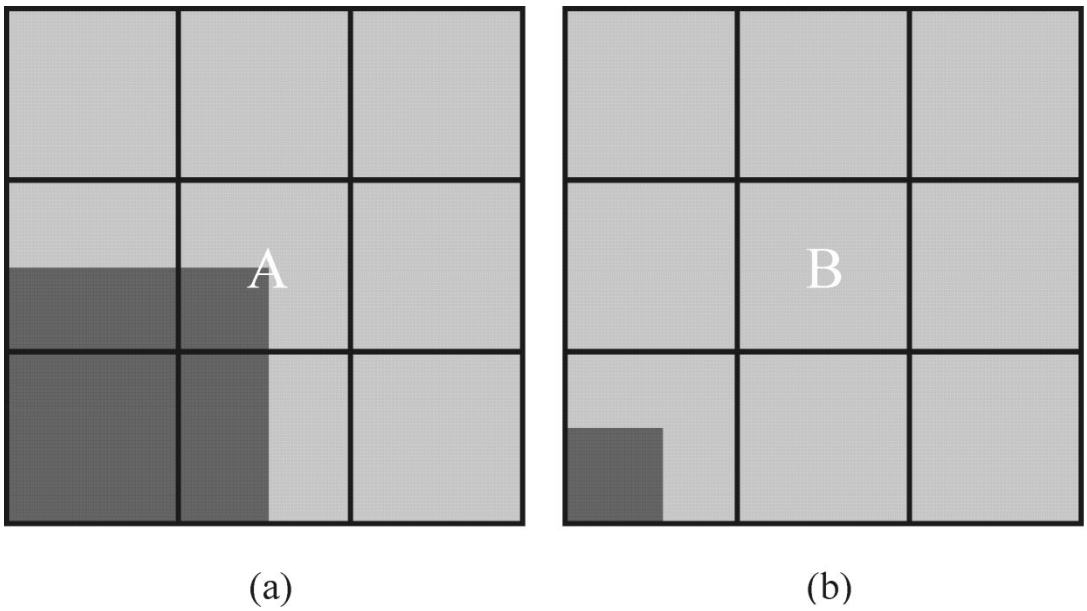


Fig. 1. Image patch in the vicinity of a corner feature. In (a), pixel "A" should be labeled as a corner whereas in (b) pixel "B" should be labeled as a (nonobvious) noncorner.

over a very limited range of corner configurations. Mohanna and Mokhtarian [14] have employed a methodology similar to the one used by Heath *et al.* for edges by asking a panel of human observers to hand label corners in a set of test images; these authors took a feature to be a corner if more than 7 out of the 10 panel members labeled it thus and took the position of a given corner to be the average position indicated by their observers. Whatever the merits of assessing edge detectors using hand-labeled datasets, clearly the approach is problematic in the case corners since these are fairly rare and a very large number of images would be needed to produce a statistically meaningful test set (see Section III-B). Mohanna and Mokhtarian increased the numbers of corners in their test set by affine and similarity transformation although as a consequence, not all of their test features can have been statistically independent. In addition, it is unclear whether the image location labeled by a human observer necessarily corresponds to the correct pixel of *projection* of a corner in object space.

In the following section, we describe the development of a ground truth dataset based on an accurate imaging model which overcomes the drawbacks of the approaches set-out in this section.

## III. Data Model

In formulating our approach, we take the objective of a corner detection process to be: *To label the pixel site which corresponds to the optical projection of the corner feature in object space*. This introduces a subtlety in that as a corner detector of finite (invariably odd) region of support is swept over the digital image, a corner projecting in an *adjacent* pixel site to the location currently under consideration will produce a pattern rather similar to a corner. Nonetheless the site adjacent to the true corner should be labeled as a *noncorner*. This situation is illustrated in Fig. 1.

In Fig. 1(a), the pixel labeled "A" should be labeled as a corner whereas in Fig. 1(b), pixel "B"—which would present a pattern closely resembling a corner—should nonetheless be labeled as a noncorner. In this paper we term the feature at pixel "B" a *nonobvious noncorner* to distinguish it from features which are obviously noncorners, such as edges and uniform patches. Most corner detectors give a number of high responses in the *vicinity* of a corner, a situation which is usually mitigated by nonmaximal suppression; in pattern space, corners and nonobvious noncorners thus seem to be rather close and are often confused. For this reason we explicitly consider nonobvious noncorners.

### A. Image Patch Generation

There is a sentiment in the image processing community, expressed, for example, in [13] that synthetic image data is somehow not representative of *real* images. We would argue that the data generated by the model used here is *highly representative* of *idealized* features. The value of approximate yet tractable models is beyond doubt in most fields of physics and engineering. In addition, a point that has been overlooked by most critics of synthetic data is that the design methodology of almost all detectors is the recognition of *idealized* features. For example, ideal step edges in the case of the Canny edge detector [15] and points of rapid, two-dimensional intensity change for the Harris-Stephens corner detector [9]. We are aware of no conventional feature detector whose development has been *explicitly* influenced by the need to reject clutter, texture and so on. Consequently, detector responses to the data used here are at very least, a measure of self-consistency and in practice, indicative of performance across a far wider domain. We are not aware of anybody who regards the Canny detector as of no worth because it is developed from an "unrealistic" —although eminently sensible—model of an edge.

Further, in Section II we argued that generating a test set by hand labeling corners is impractical. In this work we have extended and enhanced previous synthetic-based approaches by

accurately modeling the *whole optical imaging process*. (We return in Section V to yet further extensions.)

In the present work we assume we are imaging a "knife-edge" right-angled ("L"-) corner located on the optical axis. These "knife-edge" features are then projected onto the CCD image plane to a perfect focus through a diffraction-limited lens which in practice involves convolving the object space with an Airy function [16]. The image quality of most reasonable lenses is usually limited by diffraction through the optical aperture when stopped-down sufficiently and in this paper we have assumed an aperture of f8 as a typical figure. (Interestingly, omitting this stage – equivalent to assuming a lens of infinite diameter aperture – produces the most significant effect on the (mis)labeling of edges rather than corners. The influence of the optical system on the performance of feature detectors will be reported elsewhere.)

We then generate a set of corner patterns by applying randomly determined affine transformations of the optical field image projected onto the CCD. First the image is sheared to produce a corner with an random opening angle uniformly distributed in the range $\in \left[45^0 \ldots 135^0\right]$. Then, the image is rotated by a random angle $\in \left[0^0 \ldots 180^0\right]$ and translated by uniformly distributed displacements

$$\Delta x, \Delta y \in \left[-1.5\, pixels \ldots + 1.5\, pixels\right].$$

If $|\Delta x| < 0.5\, pixel \wedge |\Delta y| < 0.5\, pixel$ then the generated feature is a corner, otherwise it is a nonobvious noncorner. The gray levels either side of the corner were then linearly mapped to the randomly selected ranges, $I_{high}, I_{low} \in [0 \ldots 255]$. The generated data thus spans the whole pattern space of corners.

The diffracted and transformed optical signal was then integrated over the square CCD pixel sites (which, without loss of generality, we assume to exactly abut). Finally, we add Gaussian-distributed noise to the pixel values and quantize to one of 255 levels to simulate an 8-bit analog-to-digital conversion process in a framegrabber. Up to this final quantization stage, all calculations were performed with floating-point arithmetic. The process flow is illustrated in Fig. 2; note there is a change of scale between the first three images and the last two.

This data model thus accurately represents an *idealized* imaging situation and the performance of a corner detector on this dataset represents an *upper bound* on labeling performance which factors such as texture and clutter may well reduce; we would argue that knowing an upper bound on the performance of a detector is extremely valuable. Nonetheless, the basic methodology described here is extendable to cover factors such as finite depth-of-field (see Section V) and image clutter [17].

The particular results reported in this paper have been generated for a representative imaging setup with a lens of 25 mm focal length and an object distance of 1 meter. (In fact, the only role of the object distance is to set the optical magnification.) The CCD pixel sensors have been taken as 7.5 $\mu$m on a side (15 $\times$ the mean optical wavelength of 500 nm) and we have used $40 \times 40$ samples in the optical field to correspond to one CCD pixel, this figure having been determined as sufficient to produce accurate results. The Gaussian noise had a variance of 4, being the figure observed in a number of analog camera/framegrabber combinations examined in this laboratory.

### B. Size and Composition of Datasets

The labels produced on a two-class classification problem can be described by a binomial distribution with probability of success, $P$ (where $P$ is a function of the classifier operating point). We estimate $P$ over a finite sample and we would like to have confidence intervals on this measure. Confidence intervals on a binomial-distributed variate were considered by Clopper and Pearson [18] for rather small sample sizes but more generally, we can adjust the sample size to obtain a given confidence interval for some observed $P$. Unfortunately, direct evaluation of the factorial quantities in the binomial distribution is problematic for the sample sizes of interest here but we can use an approximation to the normal distribution [19] to obtain confidence intervals. In this way we conclude that for probability estimates, $P \geq 0.15$, 10 000 samples are sufficient to give a confidence interval of $\pm 5\%$ at a 95% confidence level; this range of probabilities is of the greatest practical interest for ROC plots.

In order to construct ROC plots we require a test set of corners and another test set of noncorners. Our corner dataset comprises 10 000 examples generated in the manner shown in Fig. 2. The set of noncorners comprises the subclasses of: nonobvious noncorners (see Fig. 1), edges and uniform patches and we can identify two possible measures of performance. Firstly, we can evaluate performance in the *context* of a whole image for which the composition of the noncorner set needs to reflect the proportions of the various noncorner subclasses present. Here we proceed by assuming that the prior probability of corners is $\sim 0.1\%$ and therefore the proportion of pixels which are nonobvious noncorners (NONCs) is $\sim 0.8\%$ – every corner has eight nearest neighbors which are NONCs. We further assume that around 5% of all pixel sites are edges and the remainder are uniform patches. Taking the number of NONCs as 1000, the composition of the whole image *noncorner* set is determined accordingly. We refer to this whole image noncorner set as dataset "A;" clearly the vast majority of this dataset is uniform patches. In a sense, evaluation on dataset "A" reflects the expectation of any pixel randomly selected from an image being labeled correctly.

In addition to studying the typical labeling performance one might expect over a given image, it is highly instructive to study the *operation* of a detector, in particular its ability to differentiate corners from the various noncorner subclasses. For this reason, we have generated a second dataset which we designate "B" which comprises 10 000 of each of the noncorner subclasses from which can obtain reliable statistics on discrimination ability between, for example, corners and NONCs.

All of the noncorner subclasses were generated in a very similar manner to the corner features. The generation of nonobvious noncorners have been described in Section III-A. Edge generation is identical except that the starting corner image is replaced with a semi-infinite half plane. To produce uniform patches it is sufficient to randomly select a nominal patch intensity, $I_{patch} \in [0 \ldots 255]$ and then corrupt each pixel with Gaussian-distributed noise.

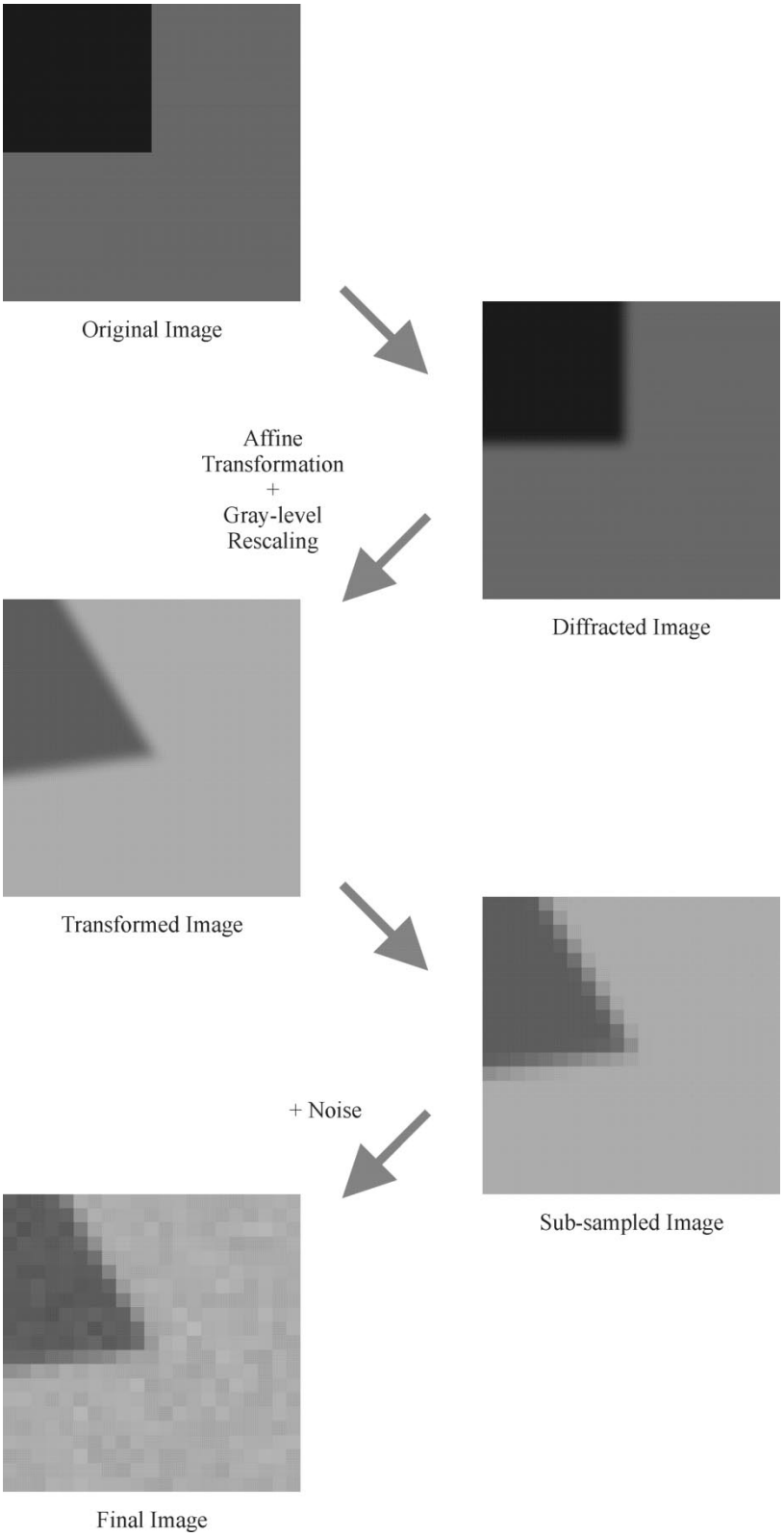


Fig. 2. Illustration of the corner generation process. Note: There is a change of scale of a factor of 40 between the first group of three images and the last two – see text for full details.

To allow other workers to assess the performance of additional corner detectors, the datasets generated here are available at http:// www.shef.ac.uk/eee/staff/pir/Datasets/CornerDatasets_f8_40.zip.

## IV. Application to Corner Detectors

We have applied the present methodology to evaluate the performance of three well-known corner detectors. The Kitchen

and Rosenfeld (KR) detector [7] is a seminal contribution to the field and forms a measure of the corner-like properties of an image point ("corner-ness") by taking the product of gray-level curvature and the gradient magnitude. Subsequently, other authors have pointed-out that many other significant corner detectors effectively fall into this curvature-times-gradient-magnitude category although exact implementations may well produce significant differences in performance. The KR detector here has been implemented over a $3 \times 3$ neighborhood.

We have also considered the detector of Paler *et al.* [8] which appears to operate on a very different principle from those detectors based on image calculus in that a median filtered version of the image is subtracted from the original and a cornerness measure formed by multiplying the gray-level differences with the contrast over a window. Davies [20], however, has shown that subject to certain assumptions, this detector also conforms to the curvature-times-gradient-magnitude format.

The corner detector of Harris and Stephens (H-S) [9] is based on an autocorrelation measure and uses only first-derivatives to form a cornerness quantity as opposed to the second-derivatives used by the KR detector which are well-known to be noisy. Noble [1], however, has shown that the H-S detector again fits the curvature-times-gradient-magnitude model.

Corner detectors usually produce significant false responses in the *vicinity* of corners and conventionally, nonmaximal suppression (NMS) is used to select the peak local response. Despite its practical utility, NMS is ultimately a "repair" stage to ameliorate detector deficiencies – in this work we focus only on the *quality* of the underlying corner measure on the grounds that a detector based on a good corner measure can perhaps be improved by NMS but a detector based on a poor measure cannot be brought up to the same quality by NMS. The minor exception to not treating NMS is with the KR detector where, arguably, an NMS stage constitutes an intrinsic part of calculating one of the possible corner measures [7].

In a number of the ROC plots which follow, the false positive fraction does not reach a value of unity even for zero threshold. (To prevent misleading discontinuities in the ROC plots we take a positive label to be the condition where $\text{corner measure} > \text{threshold}$; note the *strict* inequality.) Where the $\text{false positive fraction} < 1$ we modify the AUC measure to be a *fill-factor*, AUC', where we normalize the AUC by the maximum false positive fraction. This gives an indication of the degree to which the characteristic is filling the reachable portion of the plot; an AUC' measure of unity is ideal.

### A. Kitchen and Rosenfeld Detector [7]

The ROC plot for the basic KR detector generated with dataset "A" (reflecting whole image labeling) but *without* nonmaximally suppressing gradient values is shown in the upper curve of Fig. 3. (The gaps in the plot are caused by calculating the cornerness measure over quantized gray-level values rather than insufficient data.)

Clearly the basic KR detector is rather poor and we can investigate the reasons for this from the response to each subclass (dataset "B") plotted against threshold in Fig. 4 where the thresholds span the range of practically useful values.

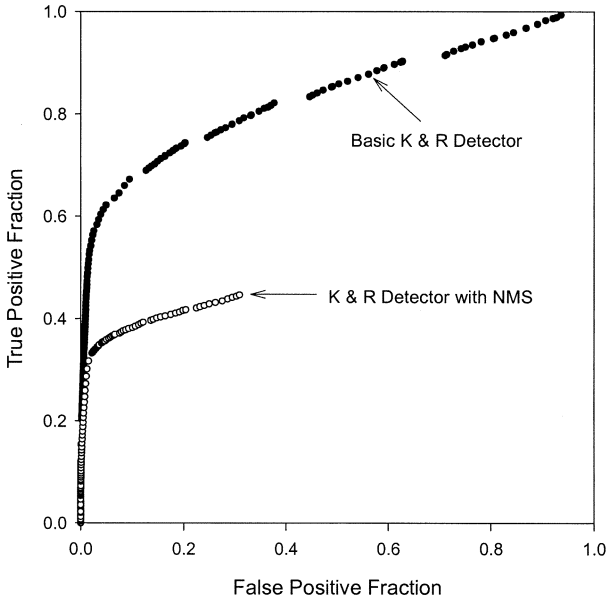


Fig. 3. ROC plots for the basic and nonmaximally suppressed KR detector.

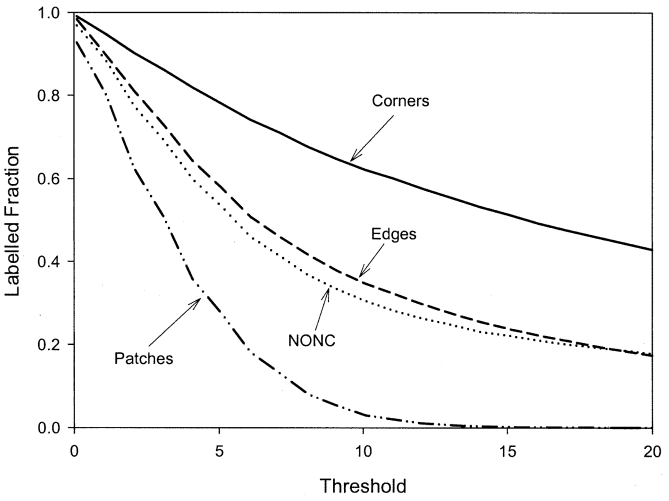


Fig. 4. Labeling response of the basic Kitchen and Rosenfeld detector.

From Fig. 4 we can see that while the detector maintains a reasonable discrimination over NONCs, it responds strongly to edges and also to uniform patches to a lesser degree. These false responses are qualitatively well-known [21] leading Kitchen and Rosenfeld [7] to modify their corner measure to use the nonmaximally suppressed gradient magnitude. The ROC plot for the KR detector with NMS is also shown in Fig. 3. It is clear that the performance is very poor and the reasons for this degradation are apparent from comparing Fig. 5 (the labeling response for the NMS-KR detector) with Fig. 4. While the response to all the noncorner subclasses (false positives) has been effectively suppressed by NMS, unfortunately the response to corners has also been reduced significantly. Recall that here a corner is considered correctly labeled if the pixel into which it projects from object space is labeled. Clearly the KR detector is exhibiting poor localization, an observation which is not new [21] but it is gratifying that the present methodology gives

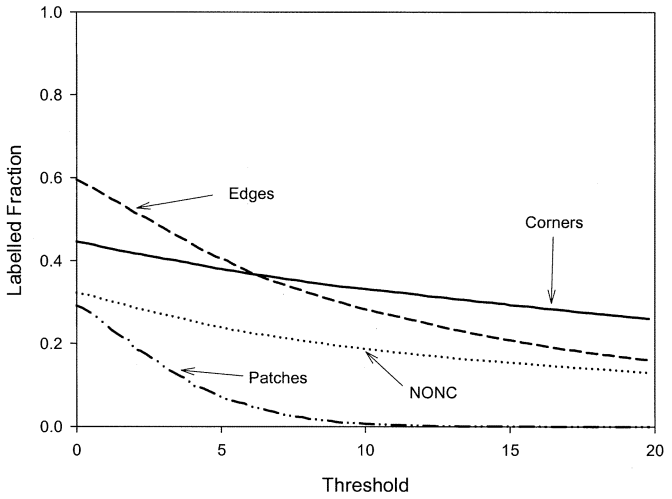


Fig. 5. Labeling response of the Kitchen and Rosenfeld detector with nonmaximal suppression.

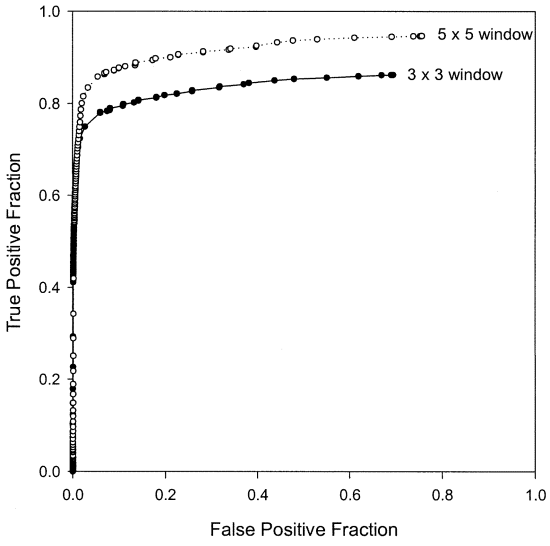


Fig. 6. ROC plots for the Paler detector for $3 \times 3$ and $5 \times 5$ windows.

results consistent with previous subjective evaluation as well as providing a means to *quantify* the effect.

### *B. Paler, Föglein, Illingworth, Kittler Detector [8]*

The Paler detector is based on the observation that median filtering does not greatly distort edges but does affect corners. Here the size of the median filter window is a parameter. The ROC plots (dataset A) for both $3 \times 3$ and $5 \times 5$ windows are shown in Fig. 6; using a $7 \times 7$ window does not produce a significantly better ROC plot than the $5 \times 5$ detector. The conclusion from Fig. 6 is in agreement with Davies [20] that a $5 \times 5$ window gives somewhat better results.

The labeling responses (over dataset B) for both the $3 \times 3$ and $5 \times 5$ windows are shown in Figs. 7 and 8, respectively. It is noteworthy that although the $5 \times 5$ detector has an improved selectivity for corners, the response for all of the noncorner classes, particularly NONCs, is also increased. This would tend to suggest that the $5 \times 5$ detector would suffer from an increased rate of false responses in the vicinity of a corner but if these could be filtered by nonmaximal suppression then the larger window size may be preferable. Although not shown here, the trend of increasing window size reducing the corner/NONC discrimination is further worsened for the $7 \times 7$ detector.

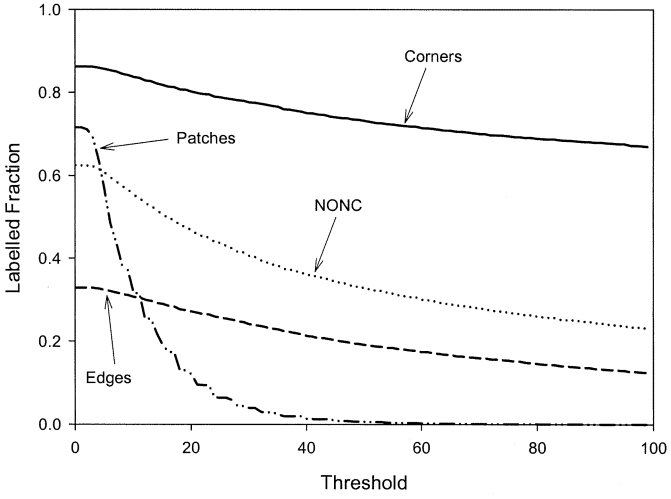


Fig. 7. Labeling response for $3 \times 3$ Paler detector.

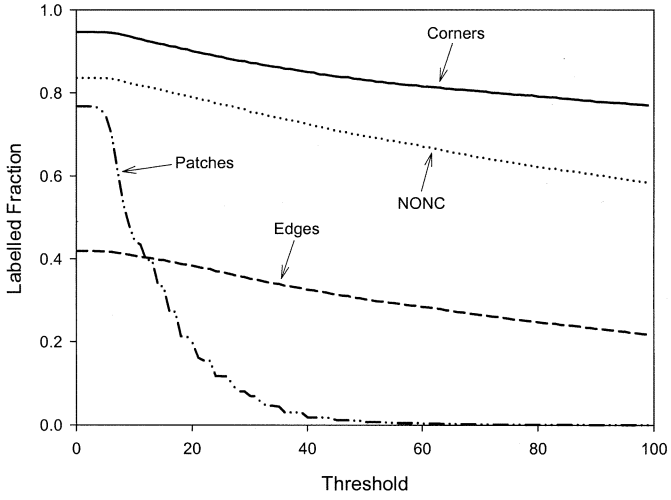


Fig. 8. Labeling response for $5 \times 5$ Paler detector.

### *C. Harris & Stephens Detector [9]*

The Harris and Stephens detector proceeds by forming a set of partial first derivatives, $I_{x,y} = \partial I / \partial x, y$ whence

$$\left.\begin{aligned} A &= G(0,\sigma) \otimes I_x^2 \\ B &= G(0,\sigma) \otimes I_y^2 \\ C &= G(0,\sigma) \otimes I_x I_y \end{aligned}\right\} \tag{1}$$

where $G(0,\sigma)$ is a zero-mean Gaussian smoothing kernel of variance, $\sigma^2$ and $\otimes$ is the convolution operator. Forming the symmetric matrix:

$$\mathbf{M} = \begin{bmatrix} A & C \\ C & B \end{bmatrix} \tag{2}$$

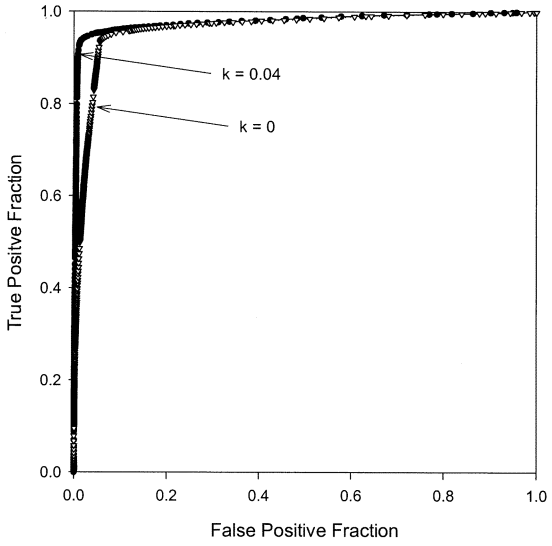


Fig. 9. ROC plots for the Harris-Stephens detector. Filled circles are for $k = 0.04$ and open triangles for $k = 0$ – see text for explanation.

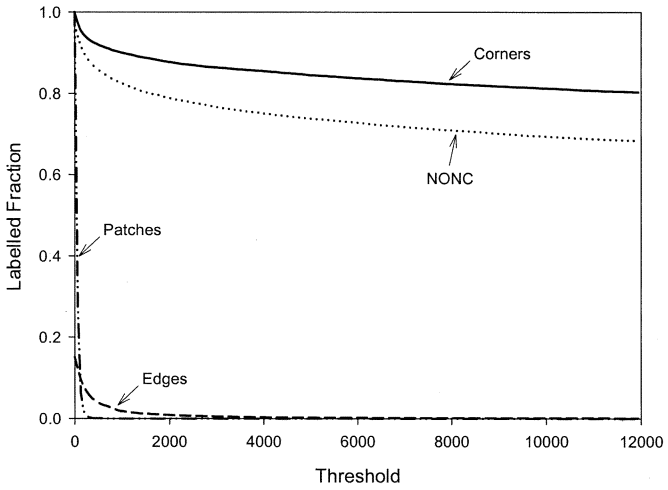


Fig. 10. Labeling response for Harris-Stephens detector for $k = 0.04$.

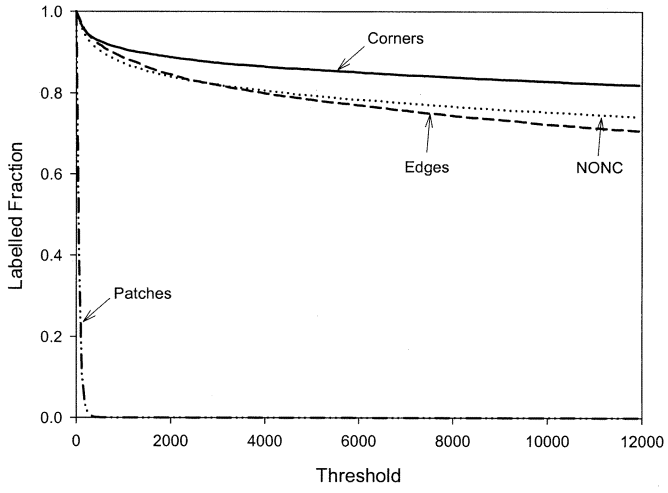


Fig. 11. Labeling response for Harris-Stephens detector for $k = 0$. Note the greatly increased response to edges over Fig. 10.

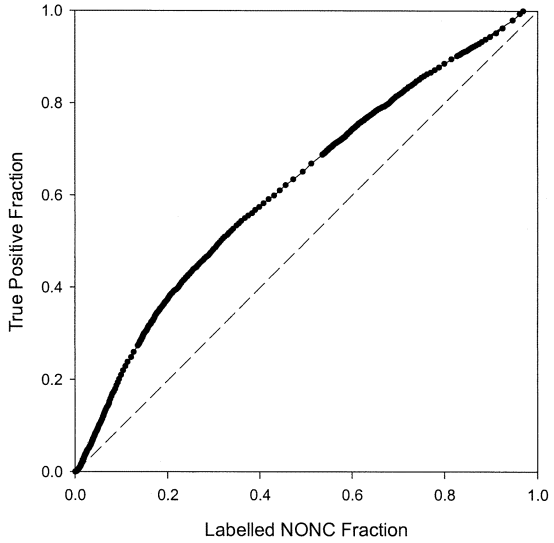


Fig. 12. ROC plot for Harris-Stephens detector for $\sigma = 1$ and $k = 0.04$. Here the noncorner class comprises only nonobvious noncorners (NONCs). The dashed line is the ROC characteristic for random guessing.

leads to the "inspired formulation" [9] of the cornerness measure

$$C = \det(\mathbf{M}) - k \times \mathrm{trace}^2(\mathbf{M}). \tag{3}$$

It is interesting to note that Gaussian convolutions in (1) were justified by Harris and Stephens as needed to suppress noise in the partial derivatives, $I_{x,y}$ but it is clear from the above (1)–(3) that either omitting this stage or smoothing $I_{x,y}$ *before* calculating $A$, $B$, $C$ results in a cornerness measure which is identically zero. Thus, rather than merely suppressing noise, the Gaussian convolution would appear to be fundamental to the operation of the detector in that it isotropically modifies the frequency spectra of the various quantities – in this sense it functions similarly to the median-filtering stage in the Paler detector. This pivotal role of the Gaussian filter does not hitherto seem to have been *explicitly* noted in the literature.

The ROC plot for the H-S detector for $\sigma = 1$; $k = 0.04$ is shown in Fig. 9 along with the plot for the case where $k = 0$. It is clear that the detector with $k = 0.04$ is superior and achieves a lower false positive score for a given true positive rate. From the labeling performance shown in Figs. 10 and 11 it is clear that the principal function of the trace-squared contribution to the corner measure (3) is to suppress the response of the detector to edges.

One matter for concern with the H-S detector is the poor discriminability of corners over NONCs that is apparent from Fig. 10. A modified ROC characteristic is shown in Fig. 12 where the noncorner class comprises solely NONC features; this plot illustrates corner/NONC discrimination for the H-S detector and it is worth noting that the dashed line in this figure is the ROC characteristic we would obtain for purely random assignment of class labels. It is clear that the H-S detector characteristic is uncomfortably close to random guessing. To quantify this, the AUC' measure for Fig. 12 is 0.6085 (whereas the comparable statistic for even the basic KR detector is 0.6636.) These quantitative observations are consistent with the widely observed defect of the H-S detector of producing large numbers of false responses around a true corner.

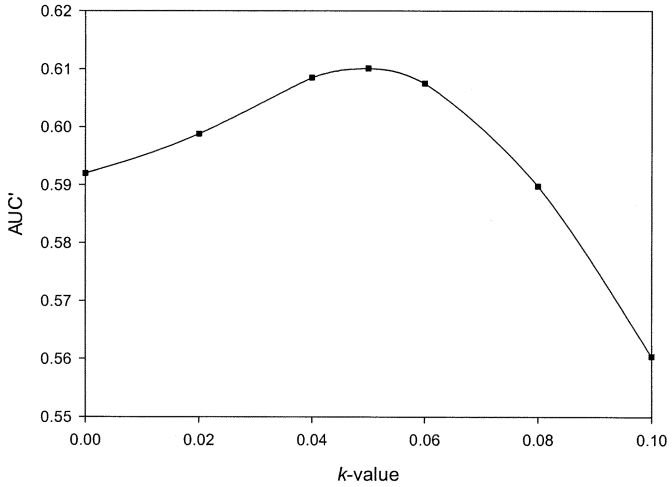


Fig. 13. AUC' measure versus $k$-value for the Harris-Stephens detector. Note the suppressed zero on the ordinate.

As a final example of the evaluation methodology presented here, we examine the optimal setting of one of the parameters in the H-S detector. Harris and Stephens [9] mentioned no value for $k$ in (3) but 0.04 appears to be widely used. We have applied the present methodology to determining the $k$-value which optimizes the AUC' measure for the family of corner versus NONC ROC plots such as Fig. 12. A plot of AUC' versus $k$-value (for $\sigma = 1$) is shown if Fig. 13 which suggests that a value of 0.05 is optimal, at least for the present imaging conditions. The improvement in performance is, however, very small and thus the figure of $k = 0.04$ seems well-founded. Nonetheless, Fig. 13 illustrates an important application of the present approach to optimization of detector parameters with respect quantitative labeling performance.

## V. Discussion and Future Work

Clearly the present methodology generates *idealized* image examples which can be viewed as yielding an upper bound on performance; if a feature detector cannot perform well on such "clean" examples then it unlikely to perform well on real-world examples contaminated by clutter and texture. Although the illustrations used here relate to corner detection, the methodology is generic in the sense that it can be used to produce test data for any arbitrary feature. All that is required is a high-resolution image of the idealized feature to replace the first image in Fig. 2. The construction of a suitable dataset proceeds analogously.

In the present paper, in using the Airy function for the point spread function we have implicitly assumed quasimonochromatic incoherent illumination. Extension to broadband illumination for a source of arbitrary spectrum is straightforward; further extension to fully or partially coherent illumination is also possible [16] although somewhat more involved. Here, for simplicity we have also assumed that the image features are in perfect focus but a finite depth of optical field can be modeled via the Lommel functions [16]. More generally, the influence of imaging conditions on the performance of feature detectors is an area that has not received much attention [22] – this too is an avenue of future research.

Most of the previous analyses of corner detectors have assumed a real-valued, continuous intensity function, for example [1], [2], [20]. Clearly images are discrete and quantized and therefore it is conceivable that even different implementations of the same algorithm may produce different results. Despite each of the three corner detectors considered having been shown to broadly conform to the curvature-times-gradient-magnitude form, it is interesting that each has its own particular shortcoming. A detailed comparison of corner detectors will be the subject of future publication.

In this paper we have approached the issue of localization *implicitly* by taking a true positive to be a label in the pixel in which the object-space corner projects – if a neighboring pixel is labeled this is counted as a false positive. Clearly localization and its interaction with nonmaximal suppression is complex and again, will be the subject of further research.

The present use of a standardized methodology allows a quantitative comparison or the objective evaluation of novel detectors – we see this as the key contribution of the present paper. Nonetheless, it is reassuring that the present results agree broadly with previous subjective observations.

## VI. Conclusions

In this paper, we have described a methodology for the performance evaluation of feature detectors. The evaluation framework is generic but to illustrate the procedure, we have analyzed three well-known corner detectors. In particular, we have been able to pinpoint the deficiencies of each detector which we have related to its ability to discriminate between corners and various of the noncorner classes.

## References

[1] J. A. Noble, "Finding corners," *Image Vis. Comput.*, vol. 6, no. 2, pp. 121–128, May 1988.
[2] K. Rohr, "Localization properties of direct corner detectors," *Int. J. Math. Imag. Vis.*, vol. 4, no. 2, pp. 139–150, 1994.
[3] M. H. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine," *Clin. Chem.*, vol. 39, no. 4, pp. 561–577, 1993.
[4] T. Kanungo and R. M. Haralick, "Receiver operating characteristic curves and optimal bayesian operating points," *Proc. IEEE Int. Conf. Image Processing*, pp. 256–259, Oct. 1995.
[5] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, 1997.
[6] N. M. Adams and D. J. Hand, "Comparing classifiers when misallocation costs are uncertain," *Pattern Recognit.*, vol. 32, no. 7, pp. 1139–1147, July 1999.
[7] L. Kitchen and A. Rosenfeld, "Gray-level corner detection," *Pattern Recognit. Lett.*, vol. 1, no. 2, pp. 95–102, Dec. 1982.
[8] K. Paler, J. Föglein, J. Illingworth, and J. Kittler, "Local ordered grey-levels as an aid to corner detection," *Pattern Recognit.*, vol. 17, no. 5, pp. 535–543, 1984.
[9] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. 4th Alvey Vision Conf.*, Manchester, U.K., 1988, pp. 147–151.
[10] K. W. Bowyer and P. J. Phillips, "Overview of work in empirical evaluation of computer vision algorithms," in *Empirical Evaluation Techniques in Computer Vision*, K. W. Bowyer and P. J. Phillips, Eds. New York: IEEE Computer Soc. Press, 1998.
[11] P. Courtney, N. A. Thacker, and A. F. Clark, "Algorithmic modeling for performance evaluation," *Mach. Vis. Applicat.*, vol. 9, no. 5-6, pp. 219–228, 1997.
[12] I. E. Abdou and W. K. Pratt, "Quantitative design and evaluation of enhancement/thresholding edge detectors," *Proc. IEEE*, vol. 67, no. 5, pp. 753–763, May 1979.

[13] M. D. 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.
[14] F. Mohanna and F. Mokhtarian, "Performance evaluation of corner detection algorithms under similarity and affine transforms," in *Proc. British Machine Vision Conf. (BMVC2001)*, Manchester, U.K., Sept. 2001, pp. 353–362.
[15] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, pp. 679–698, Nov. 1986.
[16] M. Born and E. Wolf, *Principles of Optics*, 7th ed. Cambridge, U.K.: Cambridge Univ. Press, 1999.
[17] T. Kanungo, M. Y. Jaisihma, J. Palmer, and R. M. Haralick, "A methodology for quantitative performance evaluation of edge detection algorithms," *IEEE Trans. Image Processing*, vol. 4, pp. 1667–1674, Dec. 1995.
[18] C. J. Clopper and E. S. Pearson, "The use of confidence or fiducial limits illustrated in the case of the binomial," *Biometrika*, vol. 26, no. 3-4, pp. 404–413, 1934.
[19] D. A. Berry and B. W. Lindgren, *Statistics: Theory and Methods*, 2nd ed. Belmont, CA: Wadsworth, 1996, pp. 135–136.
[20] E. R. Davies, "Median-based methods of corner detection," in *Proc. 4th Int. Conf. Pattern Recognit.*, Cambridge, U.K., Mar. 1988, pp. 360–369.
[21] Z. Zheng, H. Wang, and E. K. Teoh, "Analysis of gray level corner detection," *Pattern Recognit. Lett.*, vol. 20, no. 2, pp. 149–162, Feb. 1999.
[22] R. G. White and R. A. Schowengerdt, "Effect of point-spread functions on precision edge measurements," *J. Opt. Soc. Amer.*, vol. 11, no. 10, pp. 2593–2603, Oct. 1994.

**Peter I. Rockett** was born in London, U.K., in 1954. He received the B.Sc. degree in electronic engineering in 1976, the M.Sc. degree in solid-state electronics in 1977, and the Ph.D. degree in semiconductor physics in 1980.

After holding various fellowships and research positions, he was appointed to a tenured faculty position in electronic systems in the Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield, U.K., in 1990. His current research interests are in image processing, feature detection, and statistical pattern recognition.