# Comparing edge detection algorithm performance under degrading signal to noise ratio conditions

Kelce S. Wilson
Air Force Research Laboratory, Wright-Patterson AFB, OH 45433

## ABSTRACT

A metric is developed for evaluating performance degradation of edge detection algorithms as a function of signal to noise ratio (SNR). The metric combines both missed detections and false alarms to form a composite score. This provides a basis for objectively comparing the performance of different techniques and quantifies relative noise tolerance. It is applied to various popular algorithms, Sobel, Roberts, Prewitt, and Laplacian of Gaussian, but is described in sufficient detail to facilitate easy application to other edge detection methods. Results shown allow selection of the most optimum method for application to images with known SNR levels.

Keywords: edge detection, image processing, feature extraction, metric, SNR

## 1. INTRODUCTION

Traditional edge detectors have generally required a high enough SNR and fine enough image clarity that the transition from one region to another can be found with little uncertainty. Each of the methods examined in this effort, Sobel, Prewitt, Roberts, and Laplacain of Gaussian, have their own advantages, however they may have different tolerances to noise. All four methods were tested against identical image sets in order to compare the relative performance degradation in increasing amounts of noise and also tested using a highly blurred image.

The methods tested here are generally not well-suited for application to high-noise environments such as synthetic aperture radar (SAR) imagery[1]; those classes of images requires specialized noise-tolerant techniques[2]. Since edge detection is a form of high-pass filtering, and spatially uncorrelated speckle has a definite high frequency component, many algorithms rely on low pass filtering of the image in order to make the process more band-pass. No noise-mitigation was tested, since the comparison metrics introduced here can be implemented easily and applied to any desired method. The scores shown for the traditional methods serve as a basis for quantifying improvements in custom algorithms, or allow selection of the optimum method from those shown, once the image SNR can be calculated or estimated.

Perhaps the best way to stress edge detectors in a controlled manner is to add spatially uncorrelated Gausian noise with a calibrated power level to the pixel grayscale intensities. This was done with a constructed image whose edges were known exactly so that reliable truth models could be used as the scoring standard. Both probability of detection ($P_D$) and probability of false alarm ($P_{FA}$) were found for each method at a series on SNR levels. These two metrics were also combined into a third that includes both values in a single number. In addition, a highly blurred image was processed in order to create a visual contrast between the performance of each technique. Since the blur test did not use a constructed image, the exact edges could not be known with certainty. This prevented a quantitative comparison. A study similar to the increasing noise tolerance comparisons, but for increasing amounts of blur, will be addressed in a future effort.

## 2. NOISE TOLERANCE

Figure 1 shows a constructed image with varying levels of SNR. The image was constructed by creating two overlapping circles with values of 1.0 on a background with values of 0.0. Total size is 101 by 141 pixels. The region of overlap has pixel values of 2.0, resulting in a simple addition of the two circles. All four edge detection methods are applied to the same image, and the results are shown in separate columns. The first, left-hand, column contains the image, followed by the Sobel method, then Roberts, Prewitt, and finally Laplacian of Gaussian in the right-hand column. The first row contains the results for 20 dB SNR. Here SNR is calculated by comparing the mean of the pixel values (approximately 1.16) with the mean of the additive Gaussian noise. The noise is multiplied by the proper scale so that it has a mean value of 0.016 for the 20 dB case. At this SNR level, all methods return acceptable results.


Optical Pattern Recognition XII, David P. Casasent, Tien-Hsin Chao, Editors,
Proceedings of SPIE Vol. 4387 (2001) © 2001 SPIE · 0277-786X/01/$15.00

The second row shows a 15 dB SNR image followed by the respective edge detection results. The Laplacian of Gaussian method appears to be the most sensitive to even low levels of noise, while the other methods appear to be barely perturbed. In fact, even though the other methods appear to be returning nearly perfect results, the $P_D$ is hovering only barely above 0.9. Most of the detected edges are moved slightly when compared with the perfectly noise-free version of the image.

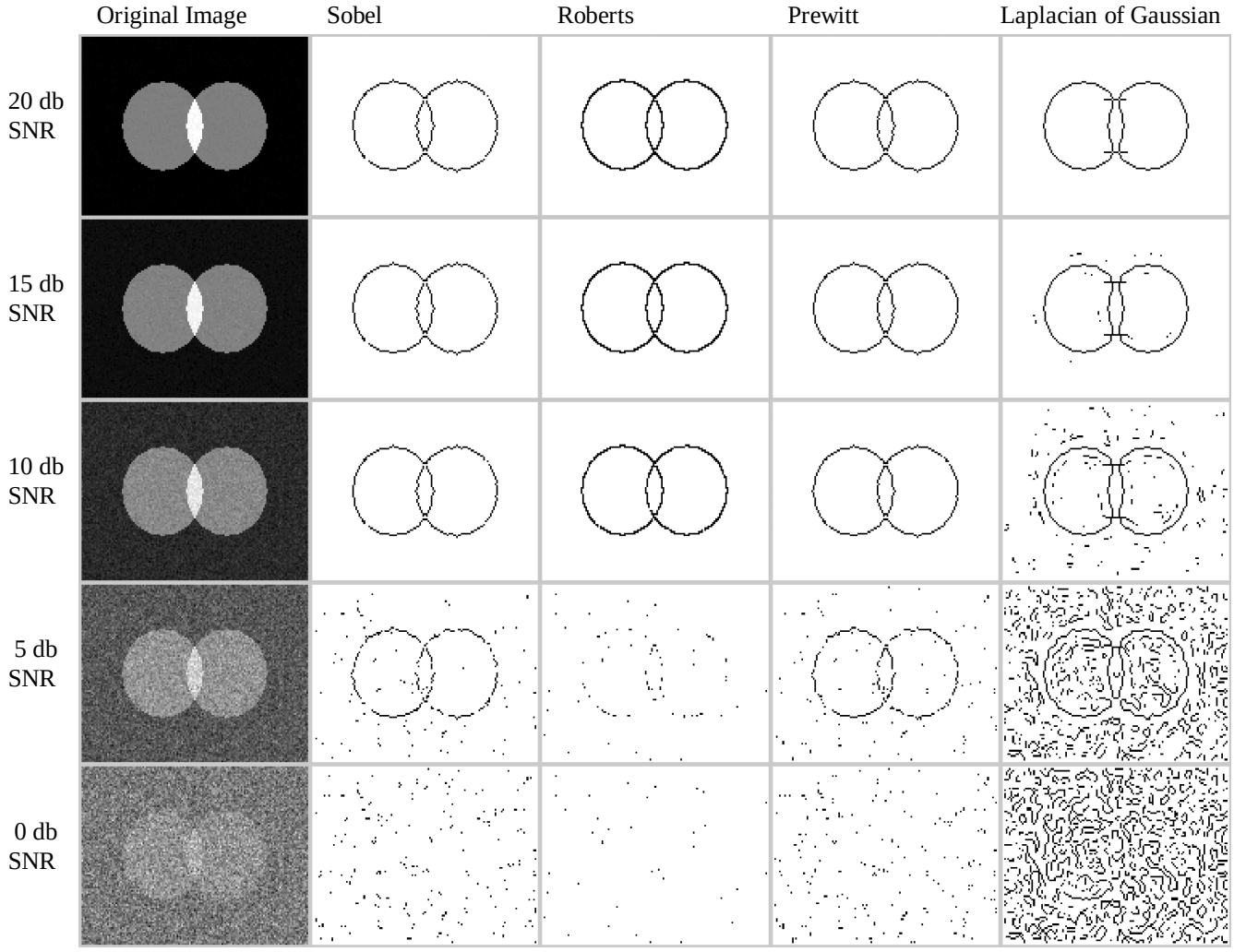


Figure 1: Edge detection in the presence of calibrated noise.

By the time the SNR drops to about 5 dB, the Roberts and Laplacian of Gaussian methods are nearly useless. Sobel and Prewitt perform for a few more dB loss in SNR until they, too produce meaningless results. Section 4 addresses the issue of quantitative scoring and the calculation of $P_D$ and $P_{FA}$.

## 3. BLUR TOLERANCE

All four methods perform rather poorly in the presence of significant blur. When applied to photographs, either direct digital pictures, or digitized film, blur is not a very significant issue. However, some non-photographic image collection systems, such as SAR, have inherent blur due to basic requirements of the image formation process. The baseline methods examined here cannot be expected to do well in such applications.

Most practical images that pose challenges to edge detectors contain both blur and noise. So even though a quantification of blur tolerance awaits further investigation, selection of a preferred method should not be made without at least giving a cursory examination to the blur issue. This is done in Figure 2.

The image shown is a digitized photograph of the shadow cast by a grid that is illuminated in the near field. Although the collection and digitization processes were relatively sharp and noise-free, the image does not lend itself to trivial edge detection. The blur may not be visually apparent at the scale in which the figure is rendered. The original image size is 467 by 588 pixels.

Both Sobel and Prewitt return a significant number of false edges, suggesting that in the blur study, the $P_{FA}$ will be rather high. Roberts falis to detect much of anything, while the Laplacian of Gausian appears to do best. Such a claim awaits a quantitive validation, but it currently appears that in a high-blur and high-noise image, a significant trade-off exists between using either Sobel/Prewitt or Laplacian of Gaussian.

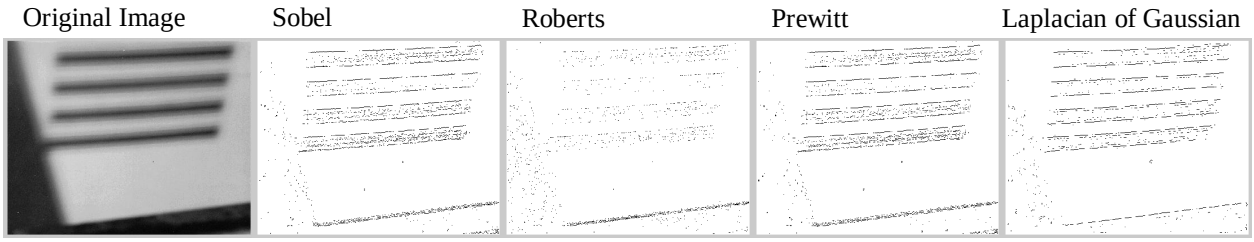


Figure 2: Edge detection in a highly blurred image.

## 4. PERFORMANCE METRIC CALCULATION

Calculation of both and $P_D$ and $P_{FA}$ for each method required some creativity in defining terms. Typical $P_D$ and $P_{FA}$ estimation is based on comparing two binary states: detect or non-detect versus target present or not present. This allows four cases as shown below:

| | **Detect** | **Non-Detect** |
|---|---|---|
| **Target present** | Increases $P_D$ | Reduces $P_D$ |
| **Target not present** | Increases $P_{FA}$ | Reduces $P_{FA}$ |

Unfortunately, such a simple scheme does not allow for spatially variant results. It is useless to define the edges of the constructed image as belonging to a single group of pixels if the methods consistently find the edges at different locations. For example, the Sobel and Prewitt methods consistently returned detected edges one pixel above and one pixel to the left of the Roberts and Laplacian of Gaussian algorithms. Additionally, the Roberts results tended to be wider than the others. This necessitated finding a method-specific truth model.

The truth models for $P_D$ for each method and a single $P_{FA}$ standard for all methods are shown in Figure 3. The $P_D$ images were created by running the edge detectors against a perfectly noise-free version of the overlapping circles. The $P_{FA}$ truth model was created separately and was intended to tolerate some noise-induced shift of the detected edges.

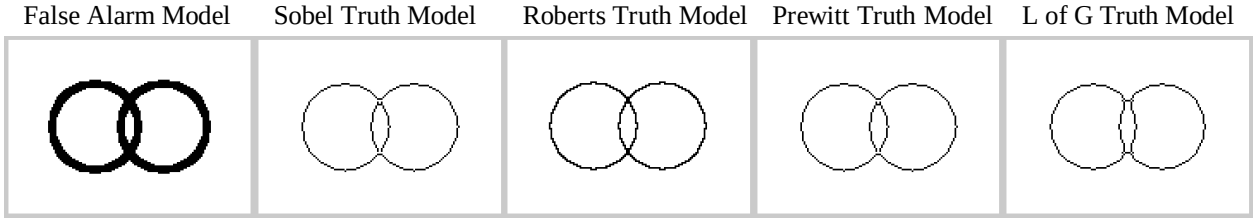


Figure 3: $P_{FA}$ truth model and $P_D$ truth models for each method.

Probability of detection was calculated in this manner: The method-specific truth model formed a list of *must detect* pixels. Any pixel from this group that was absent in the results from a later high-noise test was declared a missed detection and counted against $P_D$. Additional comments are warranted for the Laplacian of Gaussian method. The long horizontal lines appearing at the top and bottom corners of the circle overlap region are artificial edges. They are more apparent in the top

right corner of Figure 1. Since these points are not really desired, much of the lines were excluded from the *must detect* group.

Probability of false alarm was calculated differently than with the typical binary model. Rather than each detection falling into one of only two classes: valid (increases $P_D$) or invalid (increases $P_{FA}$), a third option was allowed: shifted detection. This third option had no effect on either $P_D$ or $P_{FA}$, since the validity could not be determined in an efficient manner. The $P_{FA}$ truth model forms a *may detect* group which fully contains all $P_D$ truth models. The six possible cases are shown below:

| | **Detect** | **Non-Detect** |
|---|---|---|
| **Pixel inside $P_D$ Truth Model** | Increases $P_D$ | Reduces $P_D$ |
| **Pixel in $P_{FA}$ but not $P_D$ Truth Model** | No effect | No effect |
| **Pixel outside $P_{FA}$ Truth Model** | Increases $P_{FA}$ | Reduces $P_{FA}$ |

Combining the two values into a single metric was done simply: the composite metric is the geometric mean of $(1 - P_{FA})$ and $P_D$. This gives a value that goes to 1.0 only when both $P_D$ and $P_{FA}$ are at their optimum values of 1.0 and 0.0, respectively. If either of $P_D$ goes to 0.0, or $P_{FA}$ goes to 1.0, the combined metric goes to 0.0 no matter what the other value may be. Notional $P_D$ and $P_{FA}$ curves are shown in Figure 4 to provide a graphical example of the behavior of the metric under various conditions. The top row shows some imaginary $P_D$ curves; the second row contains the $P_{FA}$ values that are to be combined with the data directly above. Corresponding composite score curves are shown in the bottom row of each column. Plots shown in Figure 4 are notional only and are not representative of the edge detection scores.

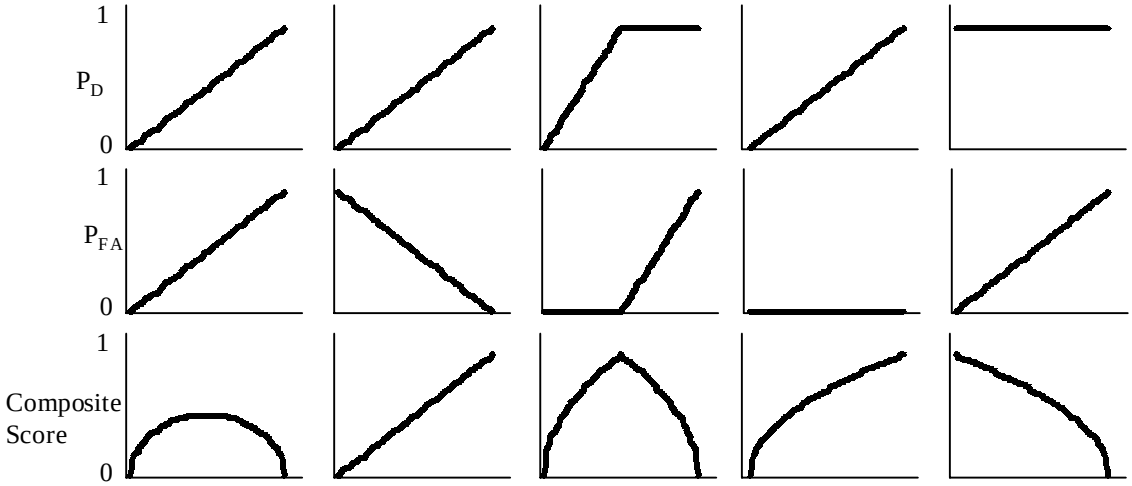


Figure 4: Relationship between notional $P_D$, $P_{FA}$, and composite scores.

## 5. RESULTS

Actual scoring of the different methods required averaging results from several iterations of the noise tolerance test process. Since the additive noise was random, the score calculated for each manifestation of the noise was slightly different. After averaging 10 results, the $P_D$, $P_{FA}$, and composite scores converged. The process was as follows:

1. The image was formed in Matlab and the average power in the signal pixels was found to be 1.16.
2. A Gaussian noise vector, *N*, was generated with Matlab's default noise generator.
3. The average power for the elements of *N* was found. Ideally it would have been 1.0, but instead, it was slightly off.
4. *N* was scaled to achieve a 20 dB average power difference between the circles and the noise.
5. The scaled version of *N* was added to the pristine circle image.
6. All four edge detection methods were run, and the metrics were calculated.
7. *N* was scaled to achieve a 0.5 dB lower SNR than just implemented, and steps 5 and 6 were repeated until the SNR was 0.
8. A new version of *N* was found and steps 2 through 7 were repeated a total of 10 times.
9. The metric curves generated from all 10 iterations were then averaged.

Results for $P_D$, $P_{FA}$, and composite scores are shown in Figures 5, 6, and 7, respectively. The relatively rapid drop-off of $P_D$ for the Roberts method with degrading SNR conditions is quite apparent in Figure 5. Also apparent in Figure 6 is the rather high $P_{FA}$ for the Laplacian of Gaussian method. According the Figure 7, the Roberts method appears to offer superior results for SNR values above 9 dB, at which point Sobel and Prewitt degrade the most gracefully. Below about a 4 dB SNR, none of the methods have much practical usefulness. These figures all support the earlier visual interpretation of Figure 1.

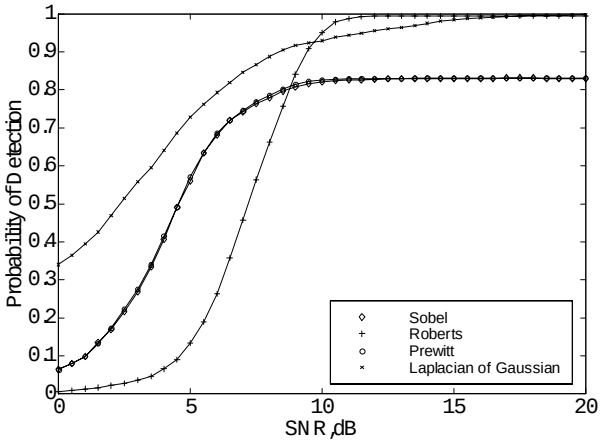


Figure 5: $P_D$ for each method

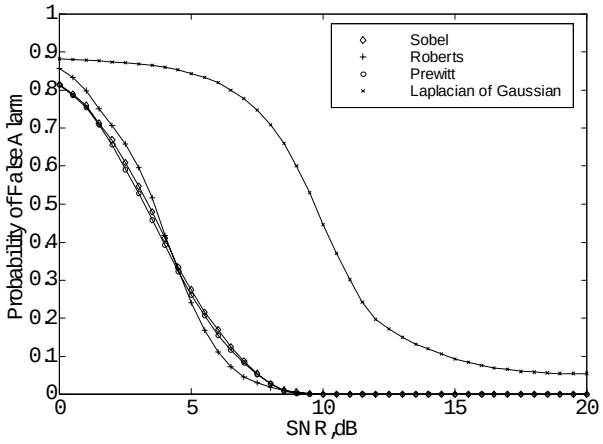


Figure 6: $P_{FA}$ each method

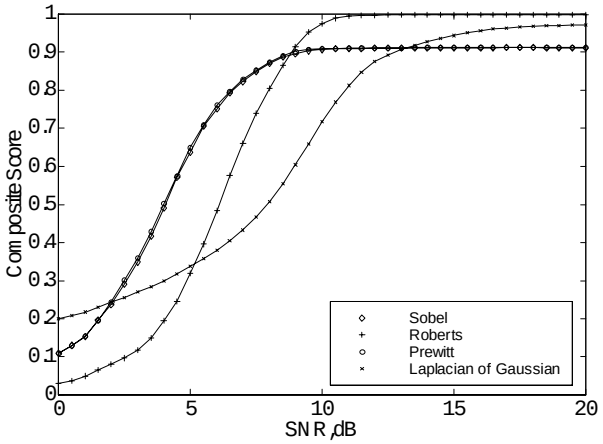


Figure 7: Composite Score for each method

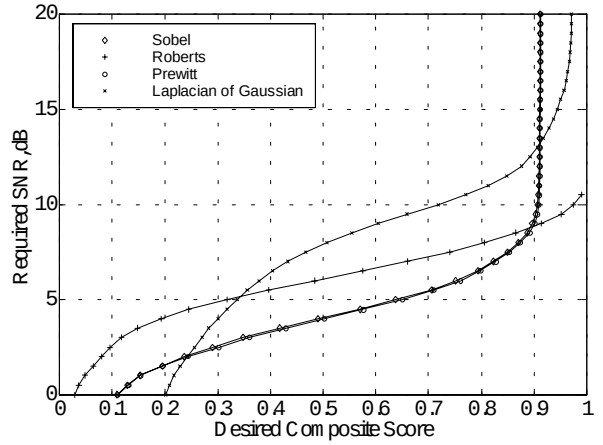


Figure 8: SNR required by method

In order to create a more readily understood comparison of the noise rejection advantage, Figure 8 shows the SNR required by each method in order to achieve a specific composite score. This was formed essentially by flipping and rotating the data for Figure 7. SNR is on the vertical axis, while desired score is on the horizontal. At first this may appear a little confusing, but after understanding how to read the figure, the interpretation becomes easy. The curves trace the following set of points: In order for one of the methods to achieve a minimum composite score of *x*, it requires the image have an SNR of at least *y*. This allows a quantitative comparison of noise rejection.

For example, in order to achieve practical results (a useful score of at least 0.7), the Laplacian of Gaussian method requires an SNR of at least 10 dB. Achieving a score above 0.9 appears to be impractical using Sobel/Prewitt, but it can be accomplished by Roberts or Laplacian of Gaussian. Comparing Figures 1 and 3 indicates that most of the missed detections by Sobel and Prewitt above 10 dB come from *shifts* in the positions of the detections. So if minor positional errors are unimportant, those two methods appear to work the best.

## 6. CONCLUSION

A metric was developed for evaluating performance degradation of edge detection algorithms as a function of SNR. The metric combined both missed detections and false alarms to form a composite score. This provided a basis for objectively comparing the performance of four popular methods: Sobel, Roberts, Prewitt, and Laplacian of Gaussian. From the results shown, it appears that the Sobel and Prewitt methods have little practical differences, and in most circumstances are the optimum methods out of those examined. The Roberts method was shown to miss a significant amount of detections for SNR levels much lower then about 10 dB, while Laplacian of Gaussian introduced serious false alarms even while the other methods were still performing well. The Laplacian of Gaussian, however, did appear to perfom the best in a single non-quantitative blur comparison.

The comparison method is simple enough and described in sufficient detail to facilitate easy application to other edge detection algorithms or noise-suppression pre-processing techniques applied prior to one of the tested methods.

## REFERENCES


1. Wilson, Kelce S. and Power, Gregory J., "Region Growing Shadow Segmentation in Synthetic Aperture Radar Images", Proceedings of the International Conference on Imaging Science, Systems, and Technology, CISST 2000, Vol. 2, pp. 31-42, Las Vegas, June 2000
2. Montesinos, P. and Chouk, S., "Edge Detection by Anisotropic Diffusion", Proceedings of the International Conference on Imaging Science, Systems, and Technology, CISST 2000, Vol. 2, pp. 605-609, Las Vegas, June 2000
3. Sperl, Daniel E., "Post-Processing Resolution Enhancement of Open Skies Photographic Imagery", Master's Thesis, US Air Force Institute of Technology, Wright-Patterson Air Force Base, Mar 2000