8-8-2003

Imagery chain assessment for feature extraction

Rufus H. Cofer

Samuel Peter Kozaitis

PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Imagery chain assessment for feature extraction

Rufus H. Cofer
Samuel Peter Kozaitis

SPIE.

Image chain assessment for feature extraction

R. H. Cofer and Samuel P. Kozaitis
Florida Institute of Technology
Department of Electrical and Computer Engineering
150 W. University Blvd.
Melbourne, FL 32901

ABSTRACT

It is shown that the image chain has important effects upon the quality of feature extraction. Exact analytic ROC results are given for the case where arbitrary multivariate normal imagery is passed to a Bayesian feature detector designed for multivariate normal imagery with a diagonal covariance matrix. Plots are provided to allow direct visual inspection of many of the more readily apparent effects. Also shown is an analytic tradeoff that says doubling background contrast is equal to halving sensor to scene distance or sensor noise. It is also shown that the results provide a lower bound to the ROC of a Bayesian feature detector designed for arbitrary multivariate normal distributions.

.

Keywords: feature extraction, Bayesian detector, image chain

1. INTRODUCTION

The utility of image sensors is that they can digest orders of magnitude more imagery than a human. However the quality of the automated results is not always apparent, especially when imagery with different statistics than designed for is to be processed or when specific effects of the imagery chain are not compensated for. In particular, it is can be exceedingly difficult to determine the quality of automated feature extraction as the ideal operating point of maximum probability of detection and minimum false alarm is approached.

At their core, a great many feature detection problems can be usefully modeled using arbitrary normal distributions. This assumption is justified in the cases where most feature and background dependencies can be effectively dealt by statistical conditioning. Section 2 below gives a terse review of relevant background needed to understand the analytic development of Section 3. Section 4 ties the analytic development to insights into the effects of mismatch of the feature detector as designed and then used, including many of the prominent possible variabilities in the image chain.

2. THEORETICAL BACKGROUND

We will make use of the following well-known facts from probability and pattern recognition theory:[1-3]

Multivariate normal distribution of a r.v. vector, X

$$p(X) = (2\pi)^{-n/2}\left|\Sigma\right|^{-1/2}\exp\left[-1/2(X-\mu)^T\Sigma^{-1}(X-\mu)\right] \triangleright N(\mu,\Sigma) \tag{1}$$

Distribution of a transformation of the above vector

$$\text{If } Y = A^T X + B \text{ then } p(Y) \triangleright N\left(A^T\mu + B, A^T\Sigma A\right) \tag{2}$$

Visual Information Processing XII, Zia-ur Rahman, Robert A. Schowengerdt, Stephen E. Reichenbach, Editors, Proceedings of SPIE Vol. 5108 (2003) © 2003 SPIE · 0277-786X/03/$15.00

The Bayesian detector

$$\lambda(X) \stackrel{\Delta}{=} \frac{p(X|H_1)}{p(X|H_0)} \geq \eta \longrightarrow D = H_1, \leq \eta \longrightarrow D = H_0 \tag{3}$$

where:

$\lambda(X)$ is called the likelihood ratio of X,
D is the output detection and
η is the Bayesian threshold.

The probability of correct detection, P_D, of the above detector

$$P_D = \int_{X_1} p(X|H_1)dX \tag{4}$$

where:

$$X_1 = \{all\ X |\ D = 1\}$$

The probability of false alarm, P_F, of the above detector

$$P_F = \int_{X_1} p(X|H_0)dX \tag{5}$$

where:

$$X_1 = \{all\ X |\ D = 1\}$$

The receiver operating curve, ROC, of the above detector

The nature of the receiver operating curve, ROC, of the above detector is shown in Fig. 1.

Properties of the ROC of the above detector

- The point, A, of Fig. 1 is the ideal, but usually unattainable, operating point.
- The ROC of the above detector is optimal in the sense that no other detector can be found with operation in the darkly shaded region of Fig. 1.
- In general, it is extremely difficult to impossible to analytically determine the ROC for the above Bayesian detector given arbitrary $p(X|H_0)$ and $p(X|H_1)$.
- In general, it is increasingly difficult to empirically determine the ROC for the above Bayesian detector as it approaches the ideal operating point, A, of Fig. 1.

Equivalent statistics

Any function:

$$Y = f(x) \tag{6}$$

is called an equivalent statistic to the $\lambda(X)$ of Eq. (3) when the decision rule:

$$Y \geq \eta' \longrightarrow D = H_1, \leq \eta' \longrightarrow D = H_0 \tag{7}$$

produces the same ROC as that of Eq. (3). The statistics of Y are:

$$P_D = \int_{Y_1} p(Y|H_1)dY \tag{8}$$

$$P_F = \int_{Y_1} p(Y|H_0)dY \tag{9}$$

where:

$$Y_1 = \{all\ Y |\ D = 1\}$$

The ROC of a scalar normal Bayesian detector

Given the scalar variable, x, with the conditional probabilities $p(x|H_0) \triangleright N(0,1)$ and $p(x|H_1) \triangleright N(\mu, \sigma^2)$, the ROC of the corresponding Bayesian detector is analytically given as:

$$P_F = erfc(\eta)/2 \tag{10}$$

$$P_D = erfc(\frac{\eta - \mu}{\sigma})/2 \tag{11}$$

3. THEORETICAL DEVELOPMENT

In the Bayesian detector of Eq. (3), let $p(X|H_0) \triangleright N(\mu_0, \Sigma_0)$ and $p(X|H_1) \triangleright N(\mu_1, \Sigma_1)$ via Eq. (1). Then $\lambda(X)$ of Eq. (3) is:

$$\lambda(X) = \exp\left[-1/2(X-\mu_1)^T \Sigma^{-1}(X-\mu_1) - 1/2(X-\mu_0)^T \Sigma^{-1}(X-\mu_0)\right] \tag{12}$$

Letting $A = \Phi$ and $b = -\mu_0$ in Eq. (2), we obtain the equivalent statistic,Y:

$$Y = \exp\left[-1/2(X-\Delta\mu)^T \Sigma_1^{-1}(X-\Delta\mu) - 1/2 X^T \Sigma_0^{-1} X\right] \tag{13}$$

where:

$$\Delta\mu = \mu_1 - \mu_0 \tag{14}$$

The statistics of Y are $p(Y|H_0) \triangleright N(0, \Sigma_0)$ and $p(Y|H_1) \triangleright N(\Delta\mu, \Sigma_1)$.

A somewhat simpler equivalent statistic, Y', can be had by taking the natural log of both sides of Eq. (13) and simplifying:

$$Y' = X^T \Sigma_0^{-1} X - (X-\Delta\mu)^T \Sigma_1^{-1}(X-\Delta\mu) \tag{15}$$

In general as stated above, it is most difficult to analytically determine the ROC of Y'. However, this is not the case for $\Sigma_0 = \Sigma_1 \stackrel{\Delta}{=} \Sigma$ diagonal. Manipulation of Eq. (15) under this condition will result in the equivalent statistic, Y'':

$$Y'' = \Delta\mu^T \Sigma^{-1} X \tag{16}$$

Now assume an X_D distributed as:

$$p(X_D|H_i) \triangleright N(\mu_{Di}, \Sigma_{Di}),\ i = 0,1 \tag{17}$$

is passed as X through Eq. (19) to obtain a Y" value of Y_D. Using Eq. (2), the statistics of Y_D are:

$$p(Y_D|H_i) \triangleright N(\Delta\mu^T \Sigma^{-1} \mu_{Di}, \Delta\mu^T \Sigma^{-1} \Sigma_{Di} \Sigma^{-1} \Delta\mu),\ i = 0,1 \tag{18}$$

An alternate equivalent statistic is:

$$Y_D^{'} = Y_D - \mu_0 \tag{20}$$

and the distributions:

$$p(Y_D' \mid H_0) \triangleright N\left(0, \Delta\mu^T \Sigma^{-1} \Sigma_{D0} \Sigma^{-1} \Delta\mu\right) \tag{21}$$

$$p(Y_D' \mid H_1) \triangleright N\left(\Delta\mu^T \Sigma^{-1} \Delta\mu_D, \Delta\mu^T \Sigma^{-1} \Sigma_{D1} \Sigma^{-1} \Delta\mu\right) \tag{22}$$

where:

$$\Delta\mu_D = \mu_{D1} - \mu_{D0} \tag{23}$$

Letting:

$$Y_D'' = Y_D' / \sigma \tag{24}$$

where:

$$\sigma = sqrt\left(\Delta\mu^T \Sigma^{-1} \Sigma_{D0} \Sigma^{-1} \Delta\mu\right) \tag{25}$$

then via Eq. (2), the equivalent statistic of interest here, Y_D'', has the distributions:

$$p(Y_D'' \mid H_0) \triangleright N(0,1) \tag{26}$$

$$p(Y_D'' \mid H_1) \triangleright N\left(\Delta\mu^T \Sigma^{-1} \Delta\mu_D / \sigma, \Delta\mu^T \Sigma^{-1} \Sigma_{D1} \Sigma^{-1} \Delta\mu / \sigma\right) \tag{27}$$

Finally, the ROC of the Bayesian detector of Eq. (16) when supplied with a signal having statistics given by Eq. (18) is given by:

$$P_F = erfc(\eta_D'') / 2 \tag{29}$$

$$P_D = erfc\left(\frac{\eta'' - \Delta\mu^T \Sigma^{-1} \Delta\mu_D / \sigma}{\Delta\mu^T \Sigma^{-1} \Sigma_{D1} \Sigma^{-1} \Delta\mu / \sigma}\right) / 2 \tag{30}$$

4. INTERPRETATION AND RESULTS

A typical recognition strategy is to search an image for features by a pixel-by-pixel local neighborhood process. It is desired to recognize those locations possessing features with high probability, the P_D of Eq. (4) or (8). It is also desired to have low false alarm in regions of no feature, the P_F of Eq. (5) or (9). Regions near the feature are "don't care," Fig. 2, in the sense that a simple morphological operation can reduce multiple local detections to just one.

In recognition hardware, conditional probability conditioning, piecewise linear processing such as nearest neighbor and/or a desire for cost-effective hardware can reduce the final feature recognition decision to that of Eq. (16). The question then is how variations within the image chain affect the accuracy of the automated recognition.

The first question that will be answered is the detection accuracy under well-matched conditions, i.e. in Eq. (18):

$$\Sigma_{D0} = \Sigma_{D1} \overset{\Delta}{=} \Sigma \tag{31}$$

This results in a $Y_D^{"}$ of Eq. (24) having the distributions:

$$p(Y_D'' \mid H_0) \triangleright N(0,1) \tag{32}$$

$$p(Y_D'' \mid H_1) \triangleright N\left(sqrt(\Delta\mu^T \Sigma^{-1} \Delta\mu), 1\right) \tag{33}$$

Thus the best ROC results that is obtainable from this specific Bayesian detector is given by:

$$P_F = erfc(\eta_D'') / 2 \tag{34}$$

$$P_D = erfc(\eta''_D - l)/2 \tag{35}$$

where:

$$l = sqrt(\Delta\mu^T \Sigma^{-1} \Delta\mu) \tag{36}$$

It is now possible to provide a family of ROCs as a function of l as shown in Fig. 3. However as l contains all of the information concerning the quality of detection and as the region of highest interest is near the ideal operating point, A, of Fig. 1, the alternate plot of Fig. 4 is typically more useful. Here given a problem specific value of l from Eq. (36), one can pick a level of P_F of interest and quickly see the attainable P_D. The higher utility of Fig. 4 becomes apparent when one realizes that it better covers the range of very low P_F's typically sought in many feature detection scenarios.

If all of the pixels on the target have equal variance, σ^2, then it is seen that:

$$l = \sqrt{n}\Delta\mu/\sigma \tag{37}$$

which provides access to several image chain engineering design tradeoffs. This for instance could be an important in-flight consideration if the sensor's ISA or EV value can be readjusted in real time. First if the σ^2 results from internal image sensor noise at each pixel, then the σ^2 of a pixel is fixed regardless of the sensor to scene viewing distance, r. Thus since n, the number of pixels on the feature is governed by:

$$n = k/r^2 \tag{39}$$

where k is a constant of proportionality, one sees that if the sensor is placed at twice the range the sensor noise's standard deviation must be halved to maintain constant performance. Alternately if the system design calls for twice the sensor noise, then the sensor to scene spacing will need to be halved. Alternately, yet again, if the sensor can be flown into an scene reflectance region where the feature to local background gray-level difference is increased by a factor of 4 then both the sensor to scene distance and the sensor noise could be increased by a factor of 2 for constant performance. Fig. 4 can be used in all cases to determine the exact values of resulting P_F and P_D.

A different kind of problem that image sensors may have is 'hot' and 'cold' pixels, e. g. CCD arrays. Automatic pixel-to-pixel gain mechanisms can restore the average gray-level responses, but will now cause variabilities in individual pixel noise levels. Letting each pixel, i, noise level be given by σ_i^2, solution of Eq. (36) will now give:

$$l = sqrt\left(\sum_{i=1}^{n} \frac{\Delta\mu_i^2}{\sigma_i^2}\right) \tag{40}$$

allowing design investigation of the exact engineering tradeoffs required to overcome any specific configuration of 'hot' or 'cold' pixels. Such action could be useful in reducing the sensor costs.

It should be noted that many image chains can have non-linear gray-level response occurring after the source of the σ^2 or σ_i^2 s considered here. Unless compensated for, this non-linear response will also reduce performance. Variations of the above theory can be relatively easily developed to handle the exact performance effects.

Many more complex image-chain and scene phenomology effects can be handled by Eqs. (18), (29) and (30). Being able to handle simultaneous variations in feature and background means and covariances permit analytic investigations of effects where the feature and background variances are different, effects of sensor defocus, target rotations, effects of shadowing and others. Examples of such effects are not pursued here but can be easily evaluated in MatLab or Mathematica.

Finally, it should be noted that it can be quite difficult to obtain the ROC of Bayesian feature detectors designed for arbitrary statistics. This includes obtaining the ROC of the Bayesian feature detector for the distributions of Eq. (17). However, it is easily possible to obtain lower bounds in this latter case, via Eqs. (18), (29) and (30). This could be quite important in early rapid-prototyping of sensor/feature detector designs.

5. CONCLUSIONS

It is shown that the image-chain has important effects upon the quality of feature detection. Exact analytic ROC results are given for the case where arbitrary multivariate normal imagery is passed to a Bayesian feature detector designed for multivariate normal imagery with a diagonal covariance matrix. Plots are provided to allow direct visual inspection of many of the more readily apparent effects. Also shown is an analytic tradeoff that says doubling background contrast is equal to halving sensor to scene distance or is equal to halving sensor noise. It is also shown that the results provide a lower bound to the ROC of a Bayesian feature detector designed for arbitrary multivariate normal distributions.

REFERENCES

1. R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification, 2nd Ed.*, John Wiley & Sons, 2001
2. R.T. Cox, *Algebra of Probable Inference,* John Hopkins University Press, 2001
3. Y. Viniotis, *Probability and Random Processes for Electrical Engineers,* McGraw Hill, 1998

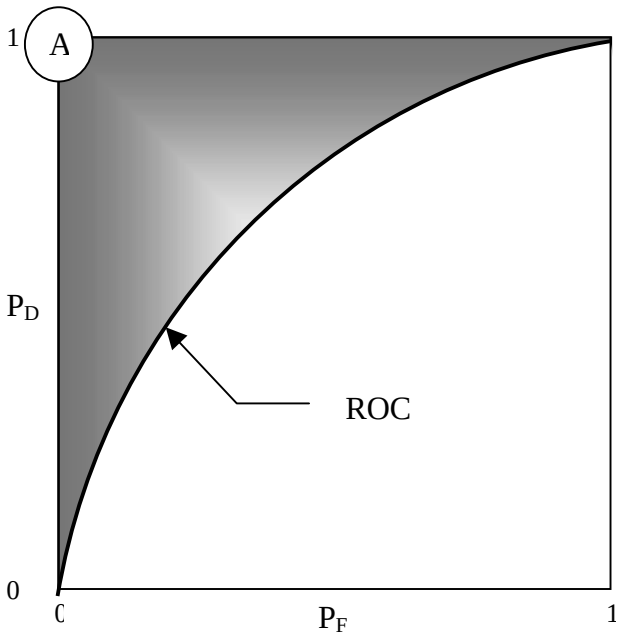

Figure 1. The Receiver Operating Curve of a Feature Detector

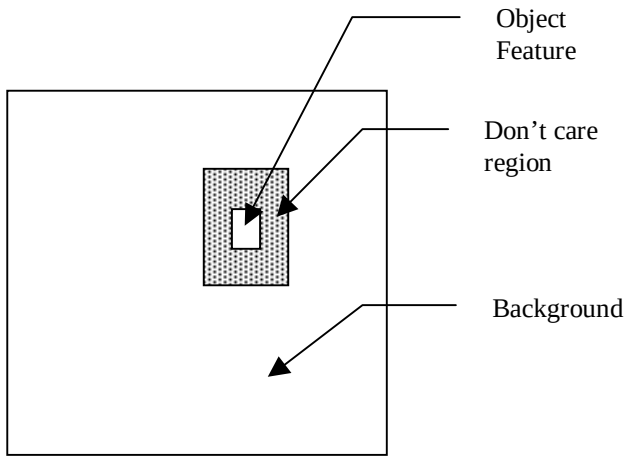

Figure 2. Typical Search for an Object Feature

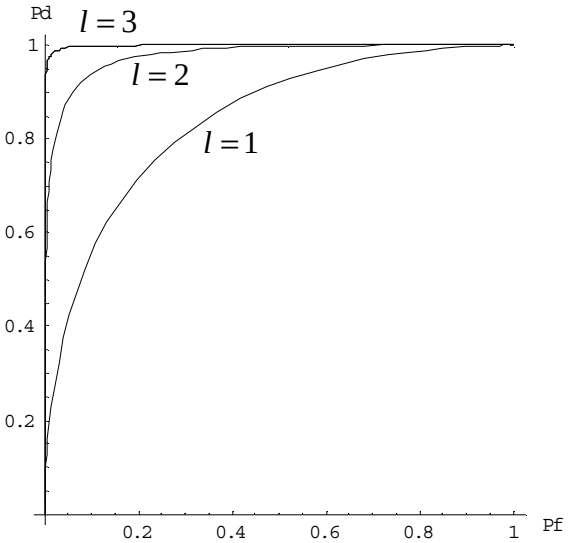

Figure 3. Normalized ROC

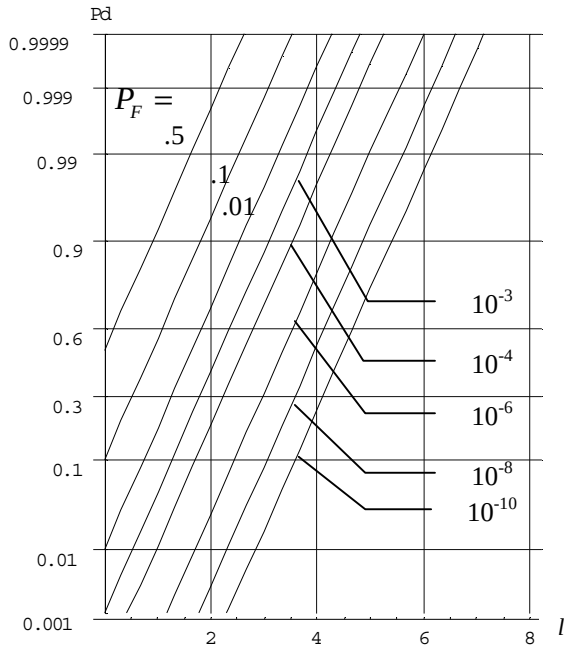

Figure 4. Normalized Detection Probability