Detection and discrimination of known signals in inhomogeneous, random backgrounds

H. H. Barrett and J. P. Rolland
Department of Radiology and Optical Sciences Center
University of Arizona, Tucson, Arizona 85724

and

R. F. Wagner and K. J. Myers
Center for Devices and Radiological Health, FDA
12720 Twinbrook Parkway, Rockville, Maryland 20857

ABSTRACT

Two studies of the effect of background inhomogeneity on observer performance in radionuclide emission imaging are presented. In the first, the task is detection of a Gaussian blob, and the imaging aperture is a pinhole of Gaussian profile. In the second, a simple discrimination task called the Rayleigh task is considered, and the aperture has a rectangular profile. In both cases performance of a suboptimal linear observer is calculated; in the first study the observer is one derived in a classic paper by Harold Hotelling, while in the second study the observer is a simple non-prewhitening matched filter. In both studies an important variable is the aperture size, and a key question is whether a small aperture or compact point spread function is advantageous. The main result is that a large aperture may perform very well or even optimally with a spatially uniform background but fail badly when the background is non-uniform. Thus predictions of image quality based on stylized tasks with uniform background must be viewed with caution.

1. INTRODUCTION

Image quality, though extremely important in all areas of medical imaging, is very difficult to assess or even define. Ultimately, image quality must be defined in terms of clinical efficacy; the best image is the one that leads to the best outcome for the patient. This definition is, however, of little operational value in the design and optimization of imaging systems, since studies of clinical efficacy are very tedious and imprecise. It would be virtually impossible to determine whether some small variation of an imaging system had any effect on patient care.

A less ambitious approach is to try to determine the performance of some model observer on a well specified model task. For example, there is a considerable literature on the ideal Bayesian observer, defined as one who has full knowledge of all relevant statistical properties of the images and of the task at hand, and who uses that information in such a way as to minimize a suitably defined risk.[1,2] For simple detection and discrimination tasks in which the signals are exactly specified (the so-called signal-known-exactly or SKE tasks), the performance of the ideal observer is readily calculable and can indeed be used as a basis for system assessment and optimization. It is implicitly assumed in this approach that a system optimized for a model observer and a model task will also be optimal for a range of real clinical tasks. It is therefore of considerable practical importance to verify the validity of this assumption.

An interesting model problem for this discussion was proposed by Wagner, Brown and Metz[3] in 1981. They considered the so-called Rayleigh task where the observer must decide whether a scene contains one Gaussian blob or two. The blobs were superimposed on a weak, uniform background, and the ideal observer was found to be a simple linear filter (though not exactly a matched filter). The imaging systems studied, though somewhat idealized, were intended to be representative of imaging sytems used in nuclear medicine. Specifically, they considered a simple pinhole, a uniformly redundant array (URA) coded aperture, and a large open aperture without any internal structure. We shall dub the latter aperture the great gaping aperture or GGA.

The findings of this study were rather surprising. Wagner et al. found that both the URA and the GGA were far superior to the pinhole aperture for the task studied. Furthermore, the GGA was superior to the URA except where the Gaussian blobs were almost point sources. The conclusion seemed to be that high spatial resolution (or a compact point spread function) was not advantageous for this task.

Since this conclusion did not accord with clinical experience in nuclear medicine, an extensive series of studies was undertaken to determine where the problem lay. In the previous paper in this volume, Myers et al.[4] looked at a number of variations on the theme set by Wagner et al. They studied the effects of object orientation and background strength and also the decision strategy of the observer, allowing both a prewhitening matched filter and a simple non-prewhitening matched filter in addition to the Bayesian observer. Still the conclusion remained--spatial resolution was of little value.

A common feature of all the studies reported by both Wagner et al. and Myers et al. was that both the signal and the background were always nonrandom and exactly known to the observer. In this paper we consider tasks where the signal is still known exactly but the background is allowed to vary randomly. Our goal is to study whether the background variability imposes further restrictions on the design of imaging systems. If so, the implications for choice of task in image assessment must be carefully weighed.

2. TASKS

Two separate studies are reported in this paper. In the first, the task is detection of a known signal in a random background. In particular, we consider a pinhole gamma-ray camera viewing a random radioactive field in which the spatial autocorrelation function is a Gaussian. Though this statistical description of the background is still simplified, it represents a step towards real radiological objects. The signal to be detected is a weak, nonrandom blob with a Gaussian profile, and the aperture is also assumed to have a Gaussian profile. The second study considers the Rayleigh task as in the work of Wagner et al. and Myers et al. Here the aperture is a sharp-edged square pinhole of variable size. In both studies, performance of the task is limited by both the inhomogeneities of the background and by Poisson noise.

Because of the background nonuniformity, the probability density function of the noise in these studies is not simple, and the ideal observer cannot be implemented by a simple linear filter. Since it is not clear whether human observers can perform nonlinear operations[5], and we do not know how to determine the form of the optimum nonlinear operations in any case, we shall consider only linear observers in this paper. For study I we consider the Hotelling observer.[6-9] Though this observer is generally inferior to the Bayesian, its performance is easier to calculate, and it has been found to correlate well with the performance of human observers in detection problems where both the signal and the background have considerable variability.[9]

For study II we use a simple non-prewhitening matched filter, which is simply a template for the expected difference signal. This observer is inferior to the Hotelling observer, but it has been widely used in other studies of image quality, including the paper of Myers et al. with which we wish to make contact. Since this approach is well known,[10] we shall not describe it in detail here. The Hotelling approach, on the other hand, is less familiar, so it is briefly reviewed in the next section.

3. THE HOTELLING OBSERVER

Ultimately, the task of a radiographic system is pattern recognition--the clinician must recognize the characteristic patterns of disease--so it is natural to turn to the literature on pattern recognition[11] for guidance in assessing imaging systems. Pattern recognition is usually divided into three steps: segmentation, feature extraction and classification. Segmentation, which refers to preprocessing operations necessary to isolate a region of interest in the image, will not be discussed further here. Feature extraction involves calculation of a set of L numbers ("features") from the image data. These numbers are usually arranged in an Lx1 column vector called a feature vector. Equivalently, this vector may be regarded as a point in an L-dimensional feature space. Finally, classification is the process of dividing this feature space into K regions corresponding to the K classes into which we would like to classify the object.

There are several things we need to know about this process: How do we choose the features in the first place, how do we partition the space, and how well does our classifier perform? Answers to all of these questions may be found in a classic paper by Harold Hotelling.[6] Translated into more modern terminology,[11] Hotelling's approach was based on two "scatter matrices" S_1 and S_2. The intraclass scatter matrix S_2 is simply the average covariance matrix of the features, averaged over the K object classes and over all sources of variability in the data. The interclass scatter matrix S_1 measures how far the class means for the features deviate from the grand mean.

Since there are L features, both S_1 and S_2 are LxL matrices, but it may help in visualizing these matrices to consider the special case $L = 1$ and $K = 2$, so that there is just one feature and both matrices reduce to scalars. In that case, S_1 is the squared difference in the mean value of the feature under the two classes, while S_2 is its average variance. It would then be natural to define the separability of the two classes as S_1/S_2. A similar definition holds in the general case, but there we must use the matrix $S_2^{-1}S_1$ rather than S_1/S_2. Furthermore, it is very desirable to have a scalar measure of class separability, and one way to form a scalar invariant from a matrix is to take its trace (sum of its diagonal elements). The Hotelling trace is therefore defined as

$$J = \mathrm{tr}\,[S_2^{-1}S_1\,]\,. \tag{1}$$

Hotelling also showed that an optimum set of features (i.e. ones that maximize J) could be formed from an arbitrary data set by using feature-extraction operators derived from an eigenanalysis of $S_2^{-1}S_1$. Using these feature operators, we have been able to show that J fixes an upper bound on the probability of error in diagnosis.[9] Finally, in several relatively realistic problems in nuclear medicine, the performance of the Hotelling classifier was found to correlate well with that of humans.[9,12]

In summary, then, the Hotelling trace appears to be a useful metric for image quality for several reasons:

1. It is a scalar, invariant to rotations and translations in the feature space;
2. It is intuitively appealing, increasing with increased interclass spread (S_1) or decreasing class variability (S_2);
3. It is calculable even when the ideal observer is not;
4. It has been found to correlate well with human performance;
5. It is related to an upper bound on probability of error in classification.

4. METHODS: STUDY I

The problem of interest in this paper is pinhole imaging of a spatially inhomogeneous distribution of radioactivity. For study I image quality is assessed on the basis of a simple detection task. A number of mathematical simplifications make the problem analytically tractable without losing essential features. We assume that:

1. The object is planar and infinite in lateral extent;

2. The aperture is infinitely thin and located equidistantly between the object and detector;
3. The pinhole transmittance is described by a Gaussian function of width σ_p as projected to the detector plane;
4. The object background is described by a constant term plus a random term, the latter taken as a stationary random process with a Gaussian autocorrelation function of width σ_b;
5. The signal to be detected is a Gaussian blob of width σ_s and known amplitude and location.

In spite of all the Gaussians that appear in this problem, the noise is not "Gaussian", since the probability density function (pdf) of the noise is not Gaussian. If there were no object variability, the noise probability law would be Poisson, which might be reasonably approximated by a Gaussian, but in general the noise pdf is rather difficult to calculate. This fact prohibits the use of the ideal observer for this problem. For the Hotelling observer, on the other hand, the pdf is not needed. Rather, we need only calculate the first- and second-order moments of the detected image, from which we find S_1 and S_2. Full details of this calculation will be presented elsewhere; here we simply sketch the procedure, give the results and discuss their significance.

Basically, the procedure is to image one realization of the lumpy background, with or without the signal, through the pinhole in a deterministic fashion, and then to recognize that this deterministic function is the mean of a Poisson random process.[13] The moments of the Poisson random process are, of course, readily calculated, but then the result must be averaged over all realizations of the background random process. The key assumption that makes this problem tractable is that the background is stationary, which means that the covariance matrices that go into S_2 are diagonalized by a discrete Fourier transform. The matrix inverse may then be expressed as an integral over the Fourier domain, and we obtain:

$$J = \frac{1}{4}\int_{\infty} d^2\rho \, \frac{|S(\rho)P(\rho)|^2}{[\,2\pi\sigma_p{}^2B + |P(\rho)|^2 S_b(\rho)]}\,, \tag{2}$$

where $S(\rho)$ is the Fourier transform of the signal to be detected, $P(\rho)$ is the Fourier transform of the aperture transmission, B is the mean background level, and $S_b(\rho)$ is the power spectral density of the background fluctuations in the object. The integral in Eq. (2) must be performed numerically.

The integrand in Eq. (2) may be regarded as a generalization of the frequency-dependent noise-equivalent quanta (NEQ), a concept introduced by Shaw[14,15] and used extensively by Wagner and others for detection problems. Our generalized NEQ is given by

$$NEQ(\rho) = \frac{|P(\rho)|^2}{[\,2\pi\sigma_p{}^2B + |P(\rho)|^2 S_b(\rho)]}\,. \tag{3}$$

The usual expression for NEQ is recovered from this form by setting S_b to zero and recognizing that $2\pi\sigma_p{}^2B$ is the Poisson noise power spectral density for a uniform background.

5. METHODS: STUDY II

Study II is based on the Rayleigh task in which the null hypothesis H_0 is that the scene contains a single Gaussian blob of width σ_s and amplitude A, while the alternative hypothesis H_1 is that it contains two blobs, each of amplitude A/2, with separation 2d. Under either hypothesis, the signal is superimposed on a random background as in study I.

The observer in this study is the non-prewhitening filter, or simply a template for the expected difference signal. This filter takes no account of the statistics of either the Poisson process or the background fluctuations. We denote the output of this filter by λ and define its signal-to-noise ratio as

$$[SNR_{npw}]^2 \equiv \frac{[E\{\lambda|H_1\}-E\{\lambda|H_0\}]^2}{var(\lambda)}. \quad (4)$$

The expectations in this equation must, of course, include both the Poisson and background randomness. The variance of λ in the denominator may be assumed to be independent of the hypothesis if the signals are weak. It is straightforward to relate both the numerator and the denominator of this equation to the scatter matrices used in the Hotelling approach, and we find

$$[SNR_{npw}]^2 = \frac{tr[S_1]^2}{tr[S_1 S_2]}. \quad (5)$$

Once again, because of the assumed stationary statistics in our problem, the actual evaluation of this general matrix expression reduces to integrals in the frequency domain:

$$[SNR_{npw}]^2 = \frac{\left[\int_{\infty} d^2\rho \, |S(\rho)P(\rho)|^2\right]^2}{\left[\int_{\infty} d^2\rho \, |S(\rho)P(\rho)|^2 \{ 2\pi\sigma_p{}^2 B + |P(\rho)|^2 S_b(\rho) \}\right]}. \quad (6)$$

6. RESULTS AND DISCUSSION

Representative results for study I, obtained by numerical integration of Eq. (2), are given in Fig. 1, where we plot J as a function of pinhole diameter σ_p for various values of the power spectral density of the background in the object at zero spatial frequency. The upper curve in Fig. 1 corresponds to a uniform background, in which case the Hotelling and ideal observers are identical. Note that in this case the only limitation is counting statistics, so the pinhole should be as large as possible to collect as many photons as possible; spatial resolution per se is unimportant. On the other hand, as soon as the background is significantly nonuniform, there is a clear optimum size for the pinhole. Not surprisingly, the optimum situation is for the pinhole to be about the same size as the signal.

Similar results for study II, obtained by numerical integration of Eq. (6), are shown in Fig. 2. In this case the optimum aperture size is comparable to the size of the signal even without background inhomogeneity. The large apertures, however, are much more strongly influenced by background inhomogeneity than are the smaller ones. (Note that the scale is logarithmic.) Thus a large aperture may perform satisfactorily, though less than optimally, for a uniform background, but fail badly for a nonuniform background. In this respect, the two studies yield similar results: the large apertures simply do not encode sufficient information to allow reliable discrimination between signal and inhomogeneous background.

It is interesting to compare Fig. 2 to the results obtained by Myers et al.[4] for the high-contrast Rayleigh task, where the SNR increased monotonically with aperture size. We see that this dependence does not hold for the low-contrast Rayleigh problem considered in this paper; further investigation is needed to account for this difference.

7. CONCLUSIONS

The present problem is interesting in that relatively little numerical computation is needed. The scatter matrices S_1 and S_2 can be derived analytically, and S_2 can be analytically inverted in the Fourier domain. Numerical computation is needed only to perform the final trace operation in J, which can be expressed as an integral in the Fourier domain. In addition, the SNR for the non-prewhitening filter can also be expressed in terms of the scatter matrices. The problem is thus a useful way to gain insight into the use of the Hotelling formalism and its relation to the non-prewhitening filter.

The results show that the choice of task is very important in assessing and optimizing imaging systems. If the task is detection of a known signal on a uniform background, the pinhole size should be as large as possible. Similarly, both Wagner et al. and Myers et al. found that the high-contrast Rayleigh task was best performed by a very large square aperture. Obviously, these conclusions are not in accord with clinical experience, where some degree of spatial resolution is needed to perform realistic clinical tasks. In that respect, the stationary nonuniform background is more realistic and leads to the intuitively appealing conclusion that the aperture should be matched to the signals to be detected or discriminated. If the aperture is substantially smaller than the signal, photon collection suffers unnecessarily, while if it is much larger, the spatial resolution is inadequate to allow reliable discrimination between signal and background.

In the extreme case where background noise dominates over Poisson noise, the collection efficiency per se of the aperture is unimportant, and the deterministic point spread function completely determines the ability of any observer to distinguish signal and background. In this limit, therefore, two systems designed for the same point spread function will have the same performance on any (background-limited) task.

To summarize, this work has graphically demonstrated the hazards of too narrowly specifying the task in model calculations of the performance of imaging systems. Systems that work well, or even optimally, on stylized tasks where the signal and background are both known exactly may fail badly with slightly more realistic tasks that include background variablity. Thus the conclusions of the earlier paper by Wagner et al., while correct for the simple problem considered there, should not be extrapolated to more complicated imaging situations.

8. ACKNOWLEDGMENTS

The authors have benefitted greatly from discussions with David Brown. This work was supported in part by the National Institutes of Health under grant no. PO1 CA23417.

9. REFERENCES

1. A.D. Whalen, Detection of Signals in Noise, Academic Press, New York (1971)

2. R.F. Wagner, C.D. Metz, and D.G. Brown, "Signal detection theory and medical image assessment," Recent Developments in Digital Imaging, 1984 AAPM Annual Summer School, New York:AIP (1985).

3. R.F. Wagner, G.D. Brown, and C.E. Metz, "On the multiplex advantage of coded source/aperture photon imaging," Proc. of the SPIE 314, 72-76 (1981).

4. K.J. Myers, R.F. Wagner, D.G. Brown, and H.H. Barrett, "Efficient utilization of aperture and detector by optimal coding," to be published in the Proc. of the SPIE 1090 (1989).

5. R.F. Wagner, K.J. Myers, A.E. Burgess, "Higher-order tasks: human vs. machine performance," to be published in the Proc. of the SPIE 1090 (1989).

6. H. Hotelling", The generalization of Student's ratio," Ann. Mth. Stat. 2, 360 (1931).

7. H.H. Barrett, W.E. Smith, K.J. Myers, T.D. Milster, and R.D. Fiete, "Quantifying the performance of imaging systems," Proc. SPIE, 535, 65-69,(1985); H.H. Barrett, K.J. Myers, and R.F. Wagner, "Beyond Signal-detection theory," Proc. SPIE 626, 231-239 (1986).

8. W.E. Smith and H.H. Barrett, "Hotelling trace criterion as a figure of merit for the optimization of imaging systems," J. Opt. Soc. Am. A, 3, 717-725 (1986).

9. R.D. Fiete, H.H. Barrett, W.E. Smith, and K.J. Myers, "The Hotelling trace criterion and its correlation with human observer performance," J. Opt. Soc. Am. A., 4, 945-953 (1987).

10. R.F. Wagner and D.G. Brown, "Unified SNR analysis of medical imaging systems," Phys. Med. Biol. 30, 489-518 (1985).

11. K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, New York (1972).

12. R.D. Fiete, H.H. Barrett, E.B. Cargill, K.J. Myers, W.E. Smith, "Psychophysical validation of the Hotelling trace criterion as a metric for system performance," SPIE 767, 298-305 (1987).

13. H.H. Barrett and W. Swindell, Radiological Imaging: The Theory of Image Formation, Detection, and Processing, Academic Press, New York (1981).

14. R. Shaw, "Evaluating the efficiency of imaging processes," Rep. Prog. Phys. 41, 1103-1155 (U.K., 1978).

15. J.C. Dainty and R. Shaw, Image Science, Academic Press, London (1974).

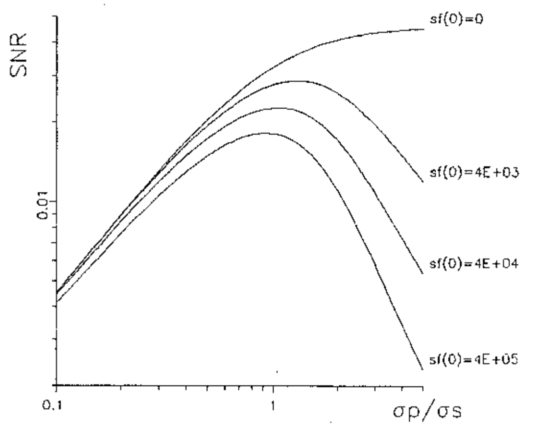

Fig. 1 Plot of the signal-to-noise ratio (the Hotelling trace) for study I as a function of the ratio of the aperture width to the signal width. The parameter Sf(0) is the power spectral density of the object background at zero spatial frequency. The signal width is 10 units and the width of the autocorrelation function of the background is 42 units. (All widths are expressed as standard deviations.)

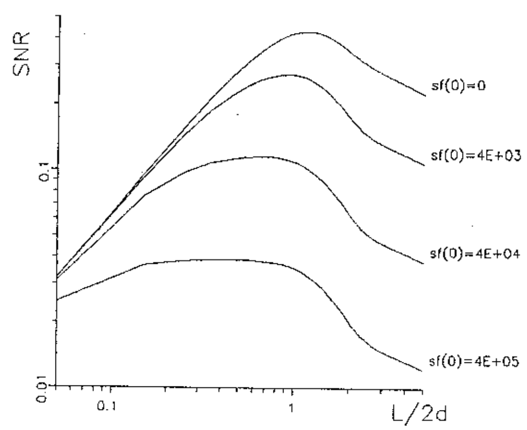

Fig. 2 Plot similar to Fig. 1 but for study II. The abscissa is the ratio of the aperture width L to the separation of the two sources, and the ordinate is the SNR for a non-prewhitening matched filter. Parameters as in Fig. 1.