Abstract

In most previous studies involving the ideal observer, the task considered has been that of simple detection where it is assumed that there is complete a priori knowledge of the background and of the possible object's shape, amplitude, and position. It is shown that redefining the detection task to include the possibility of an unknown, slowly varying background reduces the importance of the low-frequency components in the image for the ideal observer. More complicated tasks than object detection are also considered, such as determination of an object's position and width and the resolution of two objects. These higher-order tasks further enhance the importance of the high-frequency information content of the image.

Introduction

There is growing interest in the use of the ideal observer to determine the best task performance possible in a given imaging situation. The definition of such a standard is useful to quantitatively evaluate image quality or to determine the absolute efficiency with which human observers can perform specified imaging tasks. The ideal observer is not a real entity, but rather an algorithm that is designed to accomplish the task at hand. Normally the ideal observer itself is not implemented. Instead, only its measure of performance is calculated to determine how well it would have done if it had been implemented. Since the ideal observer typically is developed using a maximum likelihood estimation procedure, it will provide optimal performance in the absence of prior information.

In recent years the ideal observer approach has been applied to the simple detection task to obtain a measure of image quality for various imaging modalities. In the simple detection task, it is assumed that there is complete a priori knowledge of the background and of the possible object's shape, amplitude, and position. The task is to decide whether the object is present or not. Although the simple detection task represents a drastically oversimplified situation, tacitly it is hoped that the performance of this simple task is closely related to the tremendously complex task of diagnosis. For example, the use of the Landolt C has been proposed as a measure of image quality. The task is to identify which of the four (or eight) possible orientations of the hole in the annulus is correct for a specific C. It is fairly obvious that this task is closely related to a multiple application of the simple detection task.

In an effort to further simplify the expression for the ideal observer's performance of the simple detection task, Wagner, et al., introduced the concept of the effective sampling aperture to characterize the size of the object to be detected as well as the spatial resolution and noise properties of an imaging system. This approach is based on the fact that the simple formula, which results for detectability under the assumption that the object shape, the system MTF, and noise power spectrum are all Gaussians, appears to be applicable to a variety of other functional shapes provided the effective area of each shape is used. Detectability based upon the simple detection task using either the effective sampling aperture or the detection of a simple object, such as a disc, has been employed to calculate the absolute efficiency of x-ray imaging and to optimize various imaging systems. This approach has also been employed to characterize human observer performance of detection of various sorts. The absolute efficiency of the human observer with respect to the ideal observer for the simple detection of objects in noiseless images has been measured through carefully conducted experiments. The references cited above are only a sampling of the many articles written on these topics and are not meant to represent a complete survey of the available literature.

The reason the simple detection task has received so much attention is that it is so simple. The relaxation of any of the assumptions made about the task would complicate the theoretical results and add additional degrees of freedom to the already difficult experimental procedures. However, it will eventually be necessary to consider the effect of removing all of the simplifying assumptions in order to make the connection between the simple detection task and even the simplest practical diagnostic task. For example, the absolute calibration of actual imaging systems can rarely be relied upon so a priori knowledge of the
value of the background level is likely to be of little benefit. Thus, the background level must be determined from each image. This eliminates the usefulness of the zero-frequency component in the image for the detection task. It will be argued below that this reasoning may be extended to include more complicated background distributions and that the result is to diminish the significance of the low-frequency contribution to the detection task.

In the hierarchy of imaging tasks, detection is the least complex, with recognition and identification each representing increasingly more difficult tasks. In many imaging situations the distinction between these various tasks becomes blurred. It may not be possible to truly say a feature in an image is detected unless there is some ability to recognize the class to which that feature belongs. This becomes even more important when the feature is superimposed upon a complicated background. For example, the "detection" of lesions in the lung or breast involves much more the recognition of lesions, as distinguished from normal structures, than the mere detection of a region of increased density. Recognition rests upon the synthesis of information about various parts of an object, the study of which is the topic of pattern recognition, a field that has not yet reached maturity. The objective of the present work is to refocus attention on these higher order tasks, which might be more closely related to realistic imaging problems. It will be shown that these higher order tasks place increased emphasis on the high-frequency components in the image.

Simple detection

The optimum method for deciding which of two alternative two-dimensional functions is present in a noisy image was developed by Harris for uncorrelated noise and was later extended by Wagner to include possible noise correlation. It was assumed that the noise in the image is normally distributed, additive, stationary, and independent of the signal. Both derivations were based upon the maximum likelihood approach, which is known to yield optimum results when there is a lack of a priori information about the relative frequency of occurrence of the alternative functions. Let the two possible functions be \( f_1(x,y) \) and \( f_2(x,y) \). The derivations consisted of constructing a decision function \( d(x,y) \) that maximizes the log-likelihood ratio of the ratio of the likelihood that the given image is due to \( f_1 \) to that due to \( f_2 \). The square of the signal-to-noise ratio for this binary decision, defined as the square of the occurrence of the alternative functions. Let the two possible functions be \( f_1(x,y) \) and \( f_2(x,y) \). The derivations consisted of constructing a decision function, namely, the derivative of the logarithm of the ratio of the likelihood that the given image is due to \( f_1 \) to that due to \( f_2 \). The scale of the signal-to-noise ratio for this binary decision, defined as the square of the difference between the mean decision function values for the \( f_1 \) and \( f_2 \) divided by the variance in the decision-function for either alternative, was found by Wagner to be

\[
\text{SNR}^2 = \iint |H|^2 \left( \frac{|F_1 - F_2|^2}{S} \right) \, du \, dv,
\]

where \( F_1(u,v) \) and \( F_2(u,v) \) are the Fourier transforms of \( f_1 \) and \( f_2 \), \( H(u,v) \) is the contrast transfer function, and \( S(u,v) \) is the noise power spectrum. The integration is over the orthogonal spatial-frequency variables, \( u \) and \( v \), and is to include all relevant regions in that domain. The contrast transfer function \( H \) is the modulation transfer function scaled to allow the units of the image to differ from those of the original functions, \( f_1 \) and \( f_2 \). The ideal observer can only achieve this optimized performance if the correlations in the noise, as characterized by \( S \), are known and taken into account. For example, in the case of simple detection of Gaussian-shaped objects in CT reconstructions the optimum SNR is 25% larger than that achieved if the noise were assumed to be uncorrelated.

The integrand in Eq. 1 may be viewed as the density of \( \text{SNR}^2 \) for the discrimination between \( f_1 \) and \( f_2 \) as represented in the frequency domain whose integral over all frequencies is the total \( \text{SNR}^2 \). The fundamental quantity that affects the contribution of the available difference signal power, \( |F_1 - F_2|^2 \), to the \( \text{SNR}^2 \) is the density of noise-equivalent quanta

\[
\text{NEQ}(u,v) = \frac{|H(u,v)|^2}{S(u,v)}.
\]

The NEQ is obviously a property solely of the imaging system. It summarizes the relationship between the attenuated power of the signal and the noise power. Roughly speaking, NEQ is related to the information density transmitted by the imaging system per unit signal power. The precise relation between the information capacity of an image and NEQ may be found in Ref. 5.
In the simple detection task it is assumed that the background is completely known beforehand. If $f_1$ is the background function and $f_2$ is the sum of the background and the known object function, the difference signal is just the object function $f(x,y)$. The SNR for the simple (binary) detection problem, alternatively known as the detection sensitivity index $d'$, therefore is given by

$$SNR^2 = d'^2 = \int \int_{\text{NEQ}} |F|^2 dudv,$$

(3)

where $F(u,v)$ is the Fourier transform of the object $f(x,y)$. It is seen that for the simple detection problem all of the frequency components of the object, after being weighted by NEQ, contribute equally to $SNR^2$. The changes in this expression that result from either unknown background or alteration in the type of task will be explored below.

Detection in unknown background

Some information about the structure of the background in an image must be known a priori or be derivable from the image itself in order to perform any detection task. This is painlessly evident when an attempt is made to incorporate unknown background in a detection algorithm such as the ideal observer. When human observers view familiar images they typically do not realize the importance of the background because they recognize nearly every feature presented for the observer. Sometimes a visual inspection of a dozen or more people, the need to identify an individual from the set of similar objects can readily make evident the importance of the background (the other faces). In numerous types of diagnostic procedures the distinction between normal and abnormal structure is subtle, requiring the radiologist to consciously consider the background (normal) structure.

It is impossible to develop a universal model for background structure and to draw general conclusions about the effects of background upon detection because background structure is so variable. The simplest assumption that can be made if the background is not completely known beforehand is that it is constant. In most imaging situations this assumption fails. However, in a large class of images it is very reasonable to assume that the background is slowly varying. For example, it is often assumed that the background has a linear or quadratic dependence upon the spatial variables $x$ and $y$. In the implementation of a detection algorithm, the coefficients in the low-order polynomial expansion of the background are considered as variables to be determined from the image itself. An equivalent way to express a slowly varying background would be in terms of a low-order sine-cosine expansion. A maximum-likelihood approach is used to determine the coefficient of each term in the expansion together with the amplitude of the known function that represents the object to be detected. The low-frequency components of the data contain contributions from the background, the object to be detected, and the noise. If the amplitudes of the background can have arbitrary values and are not interrelated, they can only be determined from the data through inference using the known object function whose amplitude is determined from the high-frequency data components. The low-frequency components of the data clearly do not help determine the amplitude of the object function in this situation. Therefore, the detection sensitivity index given in Eq. 4 must be modified to include only those frequencies for which the arbitrary background components do not exist. One way to accomplish this is to multiply the integrand by a weighting function that is zero below the maximum background frequency and unity above it.

In some situations there may be a priori knowledge of the general appearance of the background. This type of knowledge can be thought of in terms of a randomized ensemble of possible backgrounds. Then the maximum likelihood approach must be replaced by one of maximum a posteriori (MAP) to obtain optimum performance. It may be possible to model such an ensemble in terms of a mean value and a covariance matrix for either the spatial or Fourier domain representations or perhaps for the coefficients in a functional series expansion. This may simplify the resulting MAP equation. While the details of the technique become important for determining its location as well as its amplitude (detection). The effect on detectability of not knowing the possible object's position is not well understood and deserves more attention.
Object specification

The detection of an object is simply related to the estimation of the amplitude of the object with a specified waveform. As mentioned in the introduction, there are other parameters of interest that further specify an object and their estimation may be more closely related to complex diagnosis than is the detection task. Suppose it is desired to estimate the position of an object with a known waveform that is superimposed upon a completely known background. The accuracy with which an individual parameter can be estimated may be derived from the expression for the SNR² for the binary decision problem, Eq. 1. To address the accuracy of position determination let the binary decision be whether the waveform \( f_1(x) \) or the waveform \( f_2(x + \Delta) \) is present. Then, using Eq. 1, it is easy to show that the SNR² for this task is

\[
\text{SNR}_\Delta^2 = 4 \int \text{NEQ} |F|^2 \sin^2(\pi \alpha \Delta) \, du \, dv .
\]

Recall that this SNR² refers to properties of the decision function, which for this problem must depend solely upon the only variable, the displacement \( \Delta \). For small displacements, the "signal" in the signal-to-noise ratio must be proportional to \( \Delta \). Therefore, in the limit of small \( \Delta \) it is possible to unambiguously identify the SNR² as the ratio of \( \Delta^2 \) to the variance of \( \Delta^2 \). The formula for this is

\[
\frac{\Delta^2}{\sigma_{\Delta}^2} = \lim_{\Delta \to 0} \text{SNR}_\Delta^2 \quad \text{as} \quad \sigma_{\Delta}^2 = 4 \pi^2 \Delta^2 \int \text{NEQ} |F|^2 u^2 \, du \, dv .
\]

Cancelling the \( \Delta^2 \) factors on both sides of the equation yields the final result

\[
\frac{\Delta^2}{\sigma_{\Delta}^2} = \int \text{NEQ} |F|^2 u^2 \, du \, dv .
\]

The variable \( u \) is the spatial frequency associated with the spatial variable \( x \) in which direction the object position is to be measured. The position accuracy in the \( y \)-direction is found by replacing \( u^2 \) by \( v^2 \). When NEQ is a constant, the integrand of Eq. 7 is seen to be the square of the Fourier transform of the partial derivative of \( f(x,y) \) with respect to \( x \). The integral in this case is the same as the integral over \( x \) and \( y \) of the square of the same partial derivative. This is a familiar result for uncorrelated noise and no blurring. See, for example, Ref. 35 in which the noise was assumed to be signal dependent. It says that information concerning the position of an object exists only where the object function changes with position, which is obvious.

Figure 1a shows a one-dimensional example of the use of the binary decision approach to determine the accuracy with which the ideal observer can estimate the position of an object. The Gaussian waveform is shown in two alternative positions. Decision theory dictates that the useful signal for deciding between these two positions is the difference signal shown as the dotted line. The SNR for the decision is given by Eq. 1, so the relevant contribution from the object is the square of the Fourier transform of the difference signal, that is the power spectrum of the difference signal. This is shown in Fig. 1b as the dotted line. It can be seen that the amplitudes of the object that are important for position estimation occur at higher frequencies than the power spectrum of the object itself, shown as the solid line. Equation 7 gives the precise mathematical statement of this. Figures 1c-f show the same conclusion is reached for the tasks of object width and binary-object separation estimation. In the latter task the objective is to determine the separation between two similar objects, as for example in the measurement of the separation between binary stars.

Table 1 summarizes the accuracies attainable in the various object specifications discussed above. The square of the performance indices presented is equal to the integral over all spatial frequencies of the respective integrands. It is observed that for the higher order tasks, higher frequencies are more important than for the lower order task of amplitude estimation, which is equivalent to object detection. For width determination, the \( u^2 \) weighting is reinforced by the partial derivative of \( F \) with respect to \( u \), which is typically small at low frequencies. For the estimation of the separation of binary objects, the high-frequency contributions are enhanced by the factor of \( u^4 \).

The relevance of the above results may be demonstrated by consideration of a typical screen/film combination, Hi-Plus/XRP. Figure 2a displays the MTF², noise power spectrum, and resulting NEQ spectrum for this screen/film system at a net diffuse optical density of unity. Assume the object and the NEQ spectrum to be circularly symmetric. Then with a change of the integration variables in Eq. 7 to polar coordinates, the integral over the polar angle may trivially be done resulting in a one-dimensional integral.

\[
\int \text{NEQ} |F|^2 u^2 \, du \, dv = 4 \pi \int \text{NEQ} |F|^2 v^2 \, dv .
\]

Table 1:

<table>
<thead>
<tr>
<th>Object Specification</th>
<th>Accuracy</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object width</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Object separation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall estimation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 1. One-dimensional examples of binary-decision tasks related to the measurement of object position, object width, and the separation between two similar objects. In a, c, and e the solid and dashed lines show the two alternative waveforms and the dotted lines show the difference of these signals. In b, d, and f the corresponding power spectra of these curves are displayed. These power spectra, which determine the relative contributions to the SNR², demonstrate the increased significance of the high-frequency components in the signals for these measurement tasks.
Fig. 2. (a) The MTF², noise power spectrum, and NEQ spectrum, for Hi-Plus/XRP. (b) Various weightings of the NEQ spectrum that contribute to the SNR² for object detection (f NEQ) and object position measurement (f³ NEQ) for point-like objects. These show that much higher frequencies are important for localization than for detection.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Performance Index</th>
<th>Integrand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplitude, A</td>
<td>A σ⁻¹</td>
<td>NEQ</td>
</tr>
<tr>
<td>Width, W</td>
<td>W σ⁻¹</td>
<td>NEQ</td>
</tr>
<tr>
<td>Position, Δ</td>
<td>σ⁻¹</td>
<td>4π²NEQ</td>
</tr>
<tr>
<td>Binary Separation, d</td>
<td>σ⁻¹</td>
<td>π⁴d²NEQ</td>
</tr>
</tbody>
</table>

Summary of the integrands appropriate for determining the ultimate accuracy with which the specified tasks can be performed. The higher order tasks place increased emphasis on higher frequencies through powers of u, the spatial frequency corresponding to the direction in which the measurements are to be performed.
\[ o^{-2} = 4\pi^3 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{NEQ} |F|^2 w^3 \text{dw} \]  

where \( w = (u^2 + v^2)^{1/2} \) is the radial frequency. Figure 2b shows this integrand for Hi-Plus/XRP after factoring out the object power spectrum, together with the corresponding integrand for object detection (Eq. 3) under the assumption of circular symmetry. The square of the performance indices for position estimation and detection are the respective areas under these two curves after multiplication by the power spectrum of the object under consideration. If the object is point-like (less than 50 \( \mu \)m wide) its power spectrum will be flat over the displayed range of frequencies. Then the important frequencies for detection are between 0.5 and 2.5 mm\(^{-1}\). However, for position measurement the important frequencies are from 2 to beyond 6 mm\(^{-1}\). The increased importance of the high-frequency components of the image for position determination are clearly evident. If the NEQ spectrum of the imaging system had dropped precipitously after a frequency of 3 mm\(^{-1}\), it would have very little effect upon the ability to detect small objects. On the other hand the accuracy with which such objects could be located would be severly impaired. It should be pointed out that the MTF of this system is about 0.1 at 3 mm\(^{-1}\). Ordinarily, this might be considered to be the cutoff frequency of this system. However, in this example, the ideal observer derives significant information for MTF values down to 0.02 for the purpose of position determination.

This observation has profound consequences for a large range of procedures in imaging science from the optimization of imaging systems to the choice of sampling rates to digitize images. It can be seen from the above exercise that a system optimized for object detection may not be optimized for object localization. Furthermore, from the observation that neither the \( \text{MTF}^2 \) nor the NEQ spectrum in Fig. 2a can be approximated by a Gaussian (appearing as an inverted parabola on this graph), it can be concluded that the simple effective-aperture approach applicable to the detection task\(^9, 10\) will not adequately approximate the integrals indicated in Table 1 for the higher order tasks.

Discussion

The extension of the simple detection problem to more complicated tasks has been considered. In any imaging task, some assumption must be made about the possible background. The effect on the ideal observer of the inclusion of an unknown background in the detection task is to reduce the significance of image amplitudes in the frequency intervals assumed to be spanned by the unknown background. The measurement of object width, object position, and separation between two similar objects have also been addressed. These higher order tasks, which all revolve about localization of object edges, place more significance on the high-frequency components present in the image than does simple detection. Thus, system studies that are based upon the simple detection task may be misleading when the the system is to be used to perform more complicated tasks. The highly complex task of radiographic diagnosis is likely to be more closely related to the higher order tasks considered here than to the detection task. This may partially help explain the preference of radiologists for images with superior high-frequency response, even when the size of the objects involved in the diagnostic task would not seem to require it. The results presented here indicate that the seemingly subtle high-frequency response of imaging systems must be carefully considered when designing or intercomparing such systems. Even computing such simple task of the-envelope type of calculations that are afforded for simple detection through the use of effective-aperture approximations may not adequately predict system performance of higher order tasks.

Acknowledgements

The author gratefully acknowledges many stimulating and productive conversations with Robert F. Wagner, C. Carl Jaffe, and Arthur E. Burgess. This work was supported by the U. S. Department of Energy under contract number W-7405-ENG-36.

References


66