LINEAR DISCRIMINANTS AND IMAGE QUALITY

H H Barrett[1,2,3], T Gooley[3], K Girodias[2],
J Rolland[1,2,4], T White[1,2] and J Yao[1,2]

[1] Optical Sciences Center, University of Arizona, Tucson, AZ 85721
[2] Department of Radiology, University of Arizona, Tucson AZ, 85724
[3] Program in Applied Mathematics, University of Arizona, Tucson, AZ 85721
[4] Dept. of Computer Science, University of North Carolina, Chapel Hill, NC

Abstract

The use of linear discriminant functions, and particularly a discriminant function derived from the work of Harold Hotelling, as a means of assessing image quality is reviewed. The relevant theory of ideal or Bayesian observers is briefly reviewed, and the circumstances under which this observer reduces to a linear discriminant are discussed. The Hotelling oberver is suggested as a linear discriminant in more general circumstances where the ideal observer is nonlinear and ususally very difficult to calculate. Methods of calculation of the Hotelling discriminant and the associated figure of merit, the Hotelling trace, are discussed. Psychophysical studies carried out at the University of Arizona to test the predictive value of the Hotelling observer are reviewed, and it is concluded that the Hotelling model is quite useful as a predictive tool unless there are high-pass noise correlations introduced by post-processing of the images. In that case, we suggest that the Hotelling observer be modified to include spatial-frequency-selective channels analogous to those in the visual system.

Keywords

Image quality; medical imaging; linear discriminant functions; ideal observer; Hotelling trace.

1. INTRODUCTION

A general definition of image quality has proven to be an elusive goal. Indeed, in the image-processing literature, image assessment is most often purely subjective, and no objective definition of quality is even attempted. The radiology literature is somewhat more sophisticated in this respect; image quality is usually defined there in terms of how well some observer can perform some task of diagnostic interest. The difficulty in that case is in choosing a task and an observer.

By far the most common observer of real radiographic images is the physician, though there is also considerable interest in automated or machine observers. For the human observer, task performance can be measured by psychophysical studies. If the task is binary (i.e., the observer has only two possible choices), the results of such studies can be analyzed by use of ROC (receiver operating characteristic) curves. A common figure of merit for image quality is thus the area under the ROC curve

(AUC) or the associated detectability index d' or d_a.

Though psychophysical studies and ROC analysis satisfy our requirement for a rigorous definition of image quality, there are still many problems in practice. The studies are time consuming and expensive, especially if the observers are physicians or if real clinical images are used. Moreover, the results are too specific to answer many questions of practical importance. An ROC study can give a definitive comparison of two imaging systems for one particular disease entity and one set of engineering parameters for each system, but it says nothing about how either system would perform with other parameters or for other diseases.

For these reasons, there is considerable interest in the use of model observers for which the performance indices such as AUC can be calculated rather than measured. If we had a model observer whose performance correlated well with that of the human, we could use it to study the effects of variation of task or system parameters. Such a tool would be extremely valuable for optimizing and effectively using radiographic imaging systems.

The most widely investigated model observer is the ideal or Bayesian observer, defined as one who has full statistical knowledge of the task and who makes best use of that knowledge to minimize a suitably defined risk. The strategy of the ideal observer for a binary task is to calculate a test statistic called the likelihood ratio and to compare it to a threshold in order to decide between the two alternatives; this strategy maximizes the AUC. The performance of the ideal observer sets an upper limit to the performance obtainable by any observer, including the human, and it might be hoped that a system optimized for the ideal observer would also be optimized for the human.

Though this approach seems reasonable, significant problems are encountered in practice. Most importantly, the likelihood ratio is only rarely calculable. Indeed, almost all investigations of the ideal observer have concentrated on detection of an exactly specified signal (or perhaps discrimination of two exactly known signals) superimposed on an exactly known background. We refer to such situations as SKE/BKE (signal known exactly, background known exactly). The SKE/BKE paradigm is obviously quite different from clinical radiology where, even for simple lesion-detection tasks, the background is cluttered with normal anatomic structures and the lesion to be detected is highly variable in size, location, shape and contrast.

The reason for the concentration on SKE/BKE tasks is that the likelihood ratio in that case can be calculated by simple linear filtering. For detection of a known signal on a flat background, where the only randomness is measurement noise that can be modeled as a stationary, white, Gaussian random process, the likelihood ratio is the output of a matched filter. If the noise is stationary and Gaussian but not white, the likelihood ratio is calculated by a so-called prewhitening matched filter.

Even in the SKE/BKE case, the performance of an ideal observer can be very different from that of a human observer. For example, Myers et al. (1985) found that human performance relative to the ideal was dramatically degraded by certain kinds of noise correlations. One interpretation of this result, and of similar results by other authors, is that the human observer is incapable of performing the prewhitening operation. This interpretation has led to the suggestion that the correct model for predicting human performance is the quasi-ideal or non-prewhitening (NPW) ideal observer who uses a simple matched filter, even in the presence of colored noise, to derive a test statistic. Though this test statistic is inferior to the optimum test statistic (the likelihood ratio), it does have the virtue of correctly predicting human performance in a range of SKE/BKE tasks.

Unfortunately, as we shall see in Section IV, the NPW model can yield very poor correlation with the human if there is inherent randomness in the task. Furthermore, the ideal observer is usually not an option except for SKE/BKE since the likelihood ratio is impossible to calculate. We must therefore look for other observer models that remain calculable for a wide variety of realistic tasks yet correlate well