Decision Strategies Maximizing the Area Under the LROC Curve

Parmeshwar Khurd and Gene Gindi

Medical Image Processing Lab,
Depts. of Electrical & Computer Engineering and Radiology,
SUNY at Stony Brook, NY 11794-2350, USA

ABSTRACT

For the 2-class detection problem (signal absent/present), the likelihood ratio is an ideal observer in that it minimizes Bayes risk for arbitrary costs and it maximizes AUC, the area under the ROC curve. The AUC-optimizing property makes it a valuable tool in imaging system optimization. If one considered a different task, namely, joint detection and localization of the signal, then it would be similarly valuable to have a decision strategy that optimized a relevant scalar figure of merit. We are interested in quantifying performance on decision tasks involving location uncertainty using the LROC methodology. We derive decision strategies that maximize the area under the LROC curve, A_{LROC}. We show that these decision strategies minimize Bayes risk under certain reasonable cost constraints. We model the detection-localization task as a decision problem in three increasingly realistic ways. In the first two models, we treat location as a discrete parameter having finitely many values resulting in an (L+1) class classification problem. In our first simple model, we do not include search tolerance effects and in the second, more general, model, we do. In the third and most general model, we treat location as a continuous parameter and also include search tolerance effects. In all cases, the essential proof that the observer maximizes A_{LROC} is obtained with a modified version of the Neyman-Pearson lemma using Lagrange multiplier methods. A separate form of proof is used to show that in all three cases, the decision strategy minimizes the Bayes risk under certain reasonable cost constraints.

Keywords: LROC, ideal observers, Bayes risk

1. INTRODUCTION

In medical imaging, scalar figures of merit summarizing task performance can be used to objectively optimize and compare imaging systems or data processing methods. An important task is detection of a lesion or defect, heretofore referred to as a "signal". For such detection problems, it is useful to derive optimal decision rules. For the familiar 2-class detection problem (signal absent or present), the likelihood ratio is an ideal observer in several senses. In particular, it is ideal (optimal) in the sense that it minimizes Bayes risk for arbitrary costs and it also maximizes AUC, the area under the ROC curve, or any partial area under the ROC curve.[1] The AUC-optimizing property of this ideal decision strategy makes it an extremely valuable tool in imaging system optimization and other applications. If one considered a different task, namely, joint detection and *localization* of the signal, then it may be similarly valuable to have a decision strategy that optimized a relevant scalar figure of merit.

Here, we are interested in quantifying performance on decision tasks involving location uncertainty using the LROC methodology.[2] Recall that the LROC curve plots, as an operating threshold is varied, the probability of correct localization for a single signal versus the probability of false alarm. (We shall define the LROC curve more precisely below.) The LROC methodology has been used for task-based image quality evaluation in SPECT,[3,4] PET,[5] mammography[6] and chest radiography.[2] Previous studies used human observers[2–6] or model observers[4,7] that were not optimal in any sense. We derive decision strategies that maximize a scalar figure of merit, the area under the LROC curve. We also analyze these optimal decision strategies from a Bayes risk perspective and show that these decision strategies minimize Bayes risk under certain reasonable cost constraints.

Further author information: (Send correspondence to Gene Gindi)
Gene Gindi: E-mail: gindi@clio.mil.sunysb.edu, Telephone: 631-444-7803

Medical Imaging 2005: Image Perception, Observer Performance, and Technology Assessment,
edited by Miguel P. Eckstein, Yulei Jiang, Proceedings of SPIE Vol. 5749
(SPIE, Bellingham, WA, 2005) · 1605-7422/05/$15 · doi: 10.1117/12.595915

We model the detection-localization task as a decision problem in three increasingly realistic ways. In the first two models, we treat location as a discrete parameter having finitely many values. In our first simple model, we do not include search tolerance effects and this results in an $L+1$ class classification problem. In the second, more general, model, we do include search tolerance effects. In the third and most general model, we treat location as a continuous parameter and also include search tolerance effects. We present our results for the first, second and third models in Sections 2, 3 and 4, respectively, and conclude with a discussion in Sec. 5. In this paper, we focus on the formalism and not on any specific medical applications.

2. DISCRETE-LOCATION MODEL WITHOUT SEARCH TOLERANCE

Initially, we will focus on certain $(L+1)$-class classification (a.k.a. multiple hypothesis testing) problems where the classifier needs to decide if an observed N-dimensional data vector $\mathbf{x}$ contains a signal and if so, the classifier needs to decide to which of L signal classes the signal belongs. The L classes correspond to L disjoint regions where the signal could occur. We will refer to the signal-absent hypothesis as H_0 and the L signal-present hypotheses as $H_i,\ i=1,\cdots,L$. Let the prior probabilities (prevalences) of each class be $p(H_i), i=0,\cdots,L$ and let $p(\mathbf{x}|H_i),\ i=0,\cdots,L$ denote the $(L+1)$ data pdf's conditioned on the various hypotheses. Note that the data pdf $p(\mathbf{x}|H_0)$ corresponds to the presence of a background only, while $p(\mathbf{x}|H_i),\ i\in\{1,\cdots,L\}$ corresponds to a background plus a signal in region i. In our general treatment, there is no presumption of a fixed background, and signal characteristics can also vary apart from simple location variability.

Any classifier for this problem can be specified with the tuple of disjoint regions $(R_0,R_1,\cdots,R_L)$ in the data vector space, where we decide H_i if the data vector $\mathbf{x}$ belongs to region R_i. For simplicity, we will assume that the range of $\mathbf{x}$ is $\Re^N$, where $\Re^N$ is the N-dimensional Cartesian product space of the real numbers. However, our theory is applicable for the range being any subset of $\Re^N$. Then $R_0=\Re^N-\cup_{i=1}^{L}R_i$.

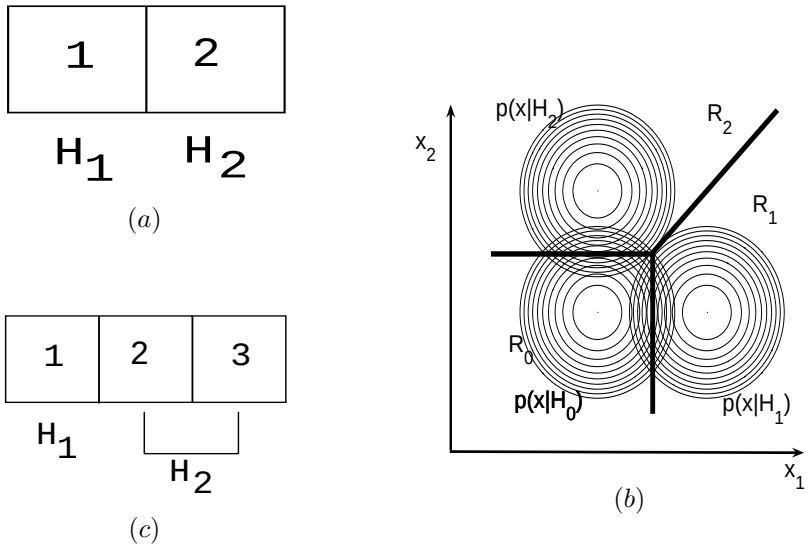

Figure 1. Hypotheses and Classifier Regions

For example, consider the 2-pixel object $\mathbf{x}=(x_1,x_2)$ in Fig. 1(a), where H_1,H_2 correspond to the signal present in pixel $1,2$ respectively. In this case, Fig. 1(b) shows a plausible contour plot of the pdf's and partitioning into regions R_0,R_1 and R_2. Note that L need not equal N. For example, the 3-pixel object in Fig. 1(c) has H_2 corresponding to the signal in pixel 2 or 3. Note that the signal can extend across many pixels in the data vector $\mathbf{x}$. In fact, $\mathbf{x}$ could be more abstract. It could be the result of an indirect imaging system such as tomography, where the signal is physically localized in an underlying space and reveals itself as a distributed signal in the sinogram (here $\mathbf{x}$).

We are particularly interested in classifiers that have a convenient alternative parametrization in the form of a 3-tuple $(t(\mathbf{x}),i(\mathbf{x}),\tau)$. Such a classifier works by first computing a scalar test statistic $t(\mathbf{x})$ from the data.

It decides signal present if $t(\mathbf{x}) > \tau$, where τ is some threshold (which determines the operating point) and it also computes a location $i(\mathbf{x})$, $i(\mathbf{x}) \in \{1, \cdots, L\}$, where the signal may have occurred. The regions in the data vector space corresponding to the 3-tuple $(t(\mathbf{x}), i(\mathbf{x}), \tau)$ are given by: $R_0 = \{\mathbf{x} : t(\mathbf{x}) < \tau\}$, $R_j = \big\{\mathbf{x} : t(\mathbf{x}) > \tau$ and $i(\mathbf{x}) = j\big\}$, $j \in \{1, \cdots, L\}$.

Classification performance on detection-localization tasks is often measured by plotting the LROC curve. The abscissa of the LROC curve is the false-positive rate given by

$$P_{FP} = \int_{\Re^N - R_0} p(\mathbf{x}|H_0) d\mathbf{x} = \sum_{i=1}^{L} \int_{R_i} p(\mathbf{x}|H_0) d\mathbf{x}. \tag{1}$$

The ordinate of the LROC curve is the probability of correct localization, i.e. the probability that the signal is correctly detected and localized, given by

$$P_{CL} = \frac{1}{1 - p(H_0)} \sum_{i=1}^{L} \int_{R_i} p(\mathbf{x}|H_i) p(H_i) d\mathbf{x}. \tag{2}$$

(Note that the prior probability that the data contains a signal at location j, $j \in \{1, \cdots, L\}$ given that it contains a signal is $\frac{p(H_j)}{1-p(H_0)}$). Under the 3-tuple classifier scheme, the false-positive rate and the probability of correct localization are given by $P_{FP}(\tau) = \Pr\big((t(\mathbf{x})|H_0) > \tau\big)$ and $P_{CL}(\tau) = \frac{1}{1-p(H_0)} \sum_{j=1}^{L} \Pr\Big((t(\mathbf{x})|H_j > \tau)$ and $i(\mathbf{x}) = j\Big) p(H_j)$, respectively. The classical ROC curve plots the true-positive rate, given by

$$P_{TP} = \frac{1}{1 - p(H_0)} \int_{\Re^N - R_0} \sum_{i=1}^{L} p(\mathbf{x}|H_i) p(H_i) d\mathbf{x}. \tag{3}$$

against the false-positive rate. Fig. 2 shows an example of the LROC and the ROC curves.

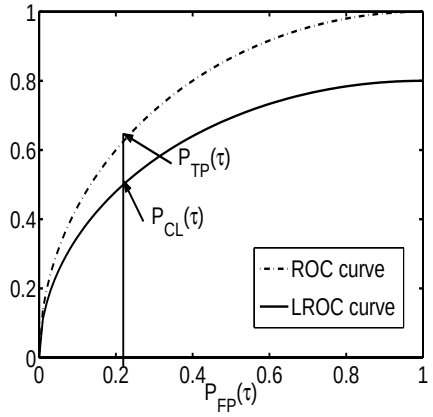

Figure 2. The ROC and LROC curves

The area under the LROC curve, A_{LROC}, is a commonly used figure of merit to summarize performance on detection-localization tasks.[2] We will now present an optimal decision strategy, referred to as the modified GLRT (generalized likelihood ratio test). It will be shown in Sec. 2.1 that the modified GLRT maximizes the area under the LROC curve (and also any partial area under the LROC curve). The intuitively pleasing optimal decision strategy is:

$$t(\mathbf{x}) \quad = \quad \max_{j \in \{1, \cdots, L\}} \frac{p(\mathbf{x}|H_j) p(H_j)}{p(\mathbf{x}|H_0)}$$

$$i(\mathbf{x}) = \arg \max_{j \in \{1,\cdots,L\}} \frac{p(\mathbf{x}|H_j)p(H_j)}{p(\mathbf{x}|H_0)}$$
$$\text{Decide } H_{i(\mathbf{x})} \text{ if } t(\mathbf{x}) > \tau, \text{ else decide } H_0 \quad (4)$$

Our strategy differs from a conventional GLRT[8] often used in the signal-processing literature, in that it uses the knowledge of the prior probabilities $p(H_j)$. In Sec. 2.2, we will present costs under which the modified GLRT minimizes the Bayes risk. We may note that for $L = 1$, our detection-localization problem reduces to a standard 2-class detection problem and as expected, the decision strategy in (4) reduces to the 2-class ideal observer (aside from an irrelevant scaling factor) that computes the likelihood ratio $\Lambda(\mathbf{x}) = \frac{p(\mathbf{x}|H_1)}{p(\mathbf{x}|H_0)}$ and decides signal present if $\Lambda(\mathbf{x}) > \tau$.

2.1. Modified GLRT Maximizes A_{LROC}

We will show that the modified GLRT in (4) maximizes A_{LROC} and any partial area under the LROC curve parametrized by two thresholds We do this by proving the following lemma (which we called the modified Neyman-Pearson lemma), analogous to the 2-class Neyman-Pearson lemma in Fukunaga,[9] Kay,[8] Duda and Hart[10] and Casella and Berger[11]:

Lemma For all classifiers satisfying $P_{FP} = \alpha$, the modified GLRT in (4) maximizes P_{CL}.

Proof 1

We will first give a proof for the modified Neyman-Pearson lemma analogous to the proof of the 2-class Neyman-Pearson lemma in Fukunaga[9] and Kay.[8]

We need to maximize

$$E = P_{CL} + \lambda(P_{FP} - \alpha) \quad (5)$$

where λ denotes a Lagrange multiplier. Using (2) and (1), we get:

$$E = \frac{1}{1 - p(H_0)} \sum_{i=1}^{L} \int_{R_i} \Big(p(\mathbf{x}|H_i)p(H_i) + (1 - p(H_0))\lambda p(\mathbf{x}|H_0) \Big) d\mathbf{x} \quad - \lambda\alpha \quad (6)$$

We need to find regions R_i so that E is maximized. Clearly, the maximum is achieved when the term corresponding to each data point is positive and adds maximally to the total integral, i.e. the region $R_i,\ i \in \{1, \cdots, L\}$ corresponds to $i = \arg\max_{j \in \{1,\cdots,L\}} \big(p(\mathbf{x}|H_j)p(H_j) + \lambda(1 - p(H_0))p(\mathbf{x}|H_0) \big)$ and $\big(p(\mathbf{x}|H_i)p(H_i) + \lambda(1 - p(H_0))p(\mathbf{x}|H_0) \big) > 0$. This decision strategy is identical to the modified GLRT with $\tau = -(1 - p(H_0))\lambda$.

Proof 2

A more elementary proof of the 2-class Neyman-Pearson lemma is given in Casella and Berger[11] that does not use the mechanism of Lagrange multipliers. We provide a second proof of our modified lemma which also does not use Lagrange multipliers.

Let $(R_0^*, R_1^*, \cdots, R_L^*)$ denote the regions for our modified GLRT, i.e. the modified GLRT assigns $\mathbf{x}$ to $R_i^*,\ i = 1, \cdots, L$ iff. $p(\mathbf{x}|H_i)p(H_i) > \tau^* p(\mathbf{x}|H_0)$ and $p(\mathbf{x}|H_i)p(H_i) > p(\mathbf{x}|H_j)p(H_j),\ \forall j \in \{1, \cdots, L\}$. Note that $R_0^* = \Re^N - \cup_{i=1}^L R_i$. Let (P_{CL}^*, P_{FP}^*) denote the corresponding correct localization and false positive probabilities for the modified GLRT. Let $(R_0', R_1', \cdots, R_L')$ specify any other classifier with the same false positive probability $P_{FP}' = P_{FP}^*$. Let P_{CL}' denote the correct localization probability for this classifier. We will show that $P_{CL}' \leq P_{CL}^*$. Note that for any region R, $R = \cup_{j=0}^L (R \cap R_j^*) = \cup_{j=0}^L (R \cap R_j')$. Since $P_{FP}' = P_{FP}^*$,

$$\int_{\Re^N - R_0^*} p(\mathbf{x}|H_0)d\mathbf{x} = \int_{\Re^N - R_0'} p(\mathbf{x}|H_0)d\mathbf{x}$$
$$\Rightarrow \int_{(\Re^N - R_0^*) \cap (\Re^N - R_0')} p(\mathbf{x}|H_0)d\mathbf{x} + \int_{(\Re^N - R_0^*) \cap R_0'} p(\mathbf{x}|H_0)d\mathbf{x}$$
$$= \int_{(\Re^N - R_0^*) \cap (\Re^N - R_0')} p(\mathbf{x}|H_0)d\mathbf{x} + \int_{(\Re^N - R_0') \cap R_0^*} p(\mathbf{x}|H_0)d\mathbf{x}$$
$$\Rightarrow \int_{(\Re^N - R_0^*) \cap R_0'} p(\mathbf{x}|H_0)d\mathbf{x} = \int_{(\Re^N - R_0') \cap R_0^*} p(\mathbf{x}|H_0)d\mathbf{x} \quad (7)$$

Therefore,

$$\begin{aligned}
(1-p(H_0))P^*_{CL} &= \sum_{i=1}^{L}\int_{R_i^*} p(\mathbf{x}|H_i)p(H_i)d\mathbf{x} \\
&= \sum_{i=1}^{L}\Big[\sum_{j=1}^{L}\int_{R_i^*\cap R_j'} p(\mathbf{x}|H_i)p(H_i)d\mathbf{x} + \int_{R_i^*\cap R_0'} p(\mathbf{x}|H_i)p(H_i)d\mathbf{x}\Big] \\
&\geq \sum_{i=1}^{L}\sum_{j=1}^{L}\int_{R_i^*\cap R_j'} p(\mathbf{x}|H_j)p(H_j)d\mathbf{x} + \sum_{i=1}^{L}\int_{R_i^*\cap R_0'} p(\mathbf{x}|H_i)p(H_i)d\mathbf{x} \\
&= \sum_{j=1}^{L}\Big[\int_{R_j'} p(\mathbf{x}|H_j)p(H_j)d\mathbf{x} - \int_{R_j'\cap R_0^*} p(\mathbf{x}|H_j)p(H_j)d\mathbf{x}\Big] + \sum_{i=1}^{L}\int_{R_i^*\cap R_0'} p(\mathbf{x}|H_i)p(H_i)d\mathbf{x} \\
&\geq (1-p(H_0))P'_{CL} - \sum_{j=1}^{L}\int_{R_j'\cap R_0^*}\tau^* p(\mathbf{x}|H_0)d\mathbf{x} + \sum_{i=1}^{L}\int_{R_i^*\cap R_0'}\tau^* p(\mathbf{x}|H_0)d\mathbf{x} \\
&= (1-p(H_0))P'_{CL} - \tau^*\int_{(\Re^N - R_0')\cap R_0^*} p(\mathbf{x}|H_0)d\mathbf{x} + \tau^*\int_{(\Re^N - R_0^*)\cap R_0'} p(\mathbf{x}|H_0)d\mathbf{x} \\
&= (1-p(H_0))P'_{CL} \qquad (8)
\end{aligned}$$

This proves the lemma.

2.2. Costs for which the modified GLRT minimizes the Bayes Risk

The Bayes risk is given by:

$$\mathcal{R} = \sum_{i=0}^{L}\sum_{j=0}^{L} c_{ij}p(H_i|H_j)p(H_j) \qquad (9)$$

Here, c_{ij} is cost of deciding H_i when the true hypothesis is H_j and $p(H_i|H_j)$ is the probability of deciding H_i when the true hypothesis is H_j. The classifier minimizing the Bayes risk is given by[8,9]:

$$\text{Decide } H_k \Leftrightarrow k = \arg\min_{i\in\{0,\cdots,L\}} C_i(\mathbf{x}) \text{ where } C_i(\mathbf{x}) = \frac{1}{p(\mathbf{x})}\sum_{j=0}^{L} c_{ij}p(\mathbf{x}|H_j)p(H_j) \qquad (10)$$

For our classification task, the following cost constraints seem reasonable:

$$\begin{aligned}
c_{ij} &= C_{FN}, \text{ if } i=0,\ j\in\{1,\cdots,L\} && \text{(11-a)}\\
c_{ij} &= C_{FP}, \text{ if } j=0,\ i\in\{1,\cdots,L\} && \text{(11-b)}\\
c_{ij} &= C_{FL}, \text{ if } i,j\in\{1,\cdots,L\},\ i\neq j && \text{(11-c)}\\
c_{ii} &= C_{CL}, \text{ if } i\in\{1,\cdots,L\} && \text{(11-d)}\\
c_{00} &= C_{TN} && \text{(11-e)}\\
C_{FL} > C_{CL} &\quad \text{and} \quad C_{FP} > C_{TN} && \text{(11-f)}
\end{aligned}$$

where the subscripts FN = false negative, FP = false positive, TN = true negative, CL = correct detection and correct localization, and FL = correct detection but false localization. Equation (11-a) above expresses the constraint that the false negative costs are the same regardless of the true location of the signal. Equation (11-b) indicates that the false positive costs are the same regardless of the reported location. Equation (11-c) indicates that the false localization costs are the same regardless of the distance of the reported location from the true location. Equation (11-d) says that the correct localization costs are independent of the reported (or true) location. Equation (11-e) above merely develops notation and is not a constraint. Equation (11-f) expresses some natural constraints on the costs.

With these costs, we get

$$
\begin{aligned}
C_0(\mathbf{x}) &= \frac{1}{p(\mathbf{x})}\left[C_{TN}p(\mathbf{x}|H_0)p(H_0) + C_{FN}\sum_{j=1}^{L} p(\mathbf{x}|H_j)p(H_j)\right] \\
C_i(\mathbf{x}) &= \frac{1}{p(\mathbf{x})}\left[C_{FP}p(\mathbf{x}|H_0)p(H_0) + C_{FL}\sum_{j=1}^{L} p(\mathbf{x}|H_j)p(H_j) - (C_{FL}-C_{CL})p(\mathbf{x}|H_i)p(H_i)\right] \;\; i \in \{1,\cdots,L\} \\
C_i(\mathbf{x}) &< C_{i'}(\mathbf{x}) \;\; i,i' \in \{1,\cdots,L\} \text{ and } i \neq i' \Leftrightarrow p(\mathbf{x}|H_i)p(H_i) > p(\mathbf{x}|H_{i'})p(H_{i'}) \qquad (12) \\
C_0(\mathbf{x}) &< C_i(\mathbf{x}) \;\; i \in \{1,\cdots,L\} \\
&\Leftrightarrow (C_{FP}-C_{TN})p(\mathbf{x}|H_0)p(H_0) > (C_{FL}-C_{CL})p(\mathbf{x}|H_i)p(H_i) + (C_{FN}-C_{FL})\sum_{j=1}^{L} p(\mathbf{x}|H_j)p(H_j) \\
C_0(\mathbf{x}) &< C_i(\mathbf{x}) \;\forall i \in \{1,\cdots,L\} \\
&\Leftrightarrow \frac{(C_{FP}-C_{TN})p(H_0)}{(C_{FL}-C_{CL})} > \max_{i\in\{1,\cdots,L\}} \frac{p(\mathbf{x}|H_i)p(H_i)}{p(\mathbf{x}|H_0)} + \frac{(C_{FN}-C_{FL})}{(C_{FL}-C_{CL})}\sum_{j=1}^{L}\frac{p(\mathbf{x}|H_j)p(H_j)}{p(\mathbf{x}|H_0)}
\end{aligned}
$$

With the additional assumption that $C_{FN} = C_{FL}$, we get

$$
C_0(\mathbf{x}) < C_i(\mathbf{x}) \;\forall i \in \{1,\cdots,L\} \Leftrightarrow \frac{(C_{FP}-C_{TN})p(H_0)}{(C_{FL}-C_{CL})} > \max_{i\in\{1,\cdots,L\}} \frac{p(\mathbf{x}|H_i)p(H_i)}{p(\mathbf{x}|H_0)} \qquad (13)
$$

Thus, from (12) and (13), we may observe that the optimal Bayes classifier reduces to the modified GLRT with $\tau = \frac{(C_{FP}-C_{TN})p(H_0)}{(C_{FL}-C_{CL})}$.

Thus, we have established that for our first simple model of a detection and localization task, the decision strategy in (4) maximizes the area under the LROC curve and that it minimizes the Bayes risk with the cost constraints in (11-a)-(11-f) and the additional constraint $C_{FN} = C_{FL}$. In Sec. 5, we further discuss the applicability of this additional constraint.

3. INCLUDING SEARCH TOLERANCE

A more realistic model of an LROC study (though not completely realistic because location will still not be considered a continuous quantity) is the following: The observer (classifier) searches for a signal in $\mathbf{x}$ and she reports a candidate location. We decide that she has correctly localized the signal if the reported location is reasonably close to the true location, i.e. within a tolerance region surrounding the true location.

We will establish some notation as we introduce our model that incorporates search tolerance. Here, we will denote the power set (i.e. the set of all subsets) of a set S by $\mathcal{P}(S)$. We model the observer as deciding upon one of $\tilde{L}+1$ actions. For a direct imaging system where the signal appears localized in $\mathbf{x}$, $\tilde{L}$ is naturally chosen equal to the number of pixels N. We will denote the actions taken by the observer as $A_i,\ i \in \{0,\cdots,\tilde{L}\}$. In the LROC task, action A_0 corresponds to deciding that no signal is present and each action $A_i,\ i \in \{1,\cdots,\tilde{L}\}$ corresponds to reporting signal present at one of $\tilde{L}$ locations. The data will still have $L+1$ classes (hypotheses). Class H_0 corresponds to the signal-absent class and the signal-present class $H_i,\ i \in \{1,\cdots,L\}$ is defined as before. We do not require the number of actions to equal the number of classes, though this might typically be the case.

There exists a point-to-set mapping, which we call the correct localization mapping, given by $\mathcal{Y} : \{1,\cdots,L\} \rightarrow \mathcal{P}(\{1,\cdots,\tilde{L}\})$. This mapping indicates that class $H_j,\ j \in \{1,\cdots,L\}$ is correctly identified when any of the actions in the set $\mathcal{Y}(j)$ are taken. For our LROC task, the set $\mathcal{Y}(j)$ corresponds to all pixel locations that fall within a tolerance region around the true location j. In the absence of any search tolerance, with $L = \tilde{L}$, this point-to-set mapping will reduce to a point-to-point identity mapping. There also exists an inverse point-to-set mapping $\tilde{\mathcal{Y}} : \{1,\cdots,\tilde{L}\} \rightarrow \mathcal{P}(\{1,\cdots,L\})$ which indicates that action $A_i,\ i \in \{1,\cdots,\tilde{L}\}$ is equivalent to considering all the classes in the set $\tilde{\mathcal{Y}}(i)$ as possibly correct candidates. We may equivalently represent these two point-to-set mappings with a bipartite graph as shown in Fig. 3. Figure 3 corresponds to an $N = 4$ pixel object and

$L = \tilde{L} = 4$. The mappings are illustrated by the bipartite graph (middle) and the definitions for $\mathcal{Y}, \tilde{\mathcal{Y}}$ on either side of the graph. For example, if H_3 were correct, the classifier could choose actions A_2, A_3, A_4 to obtain a correct classification. In this case, there is a search tolerance of ± 1 pixel.

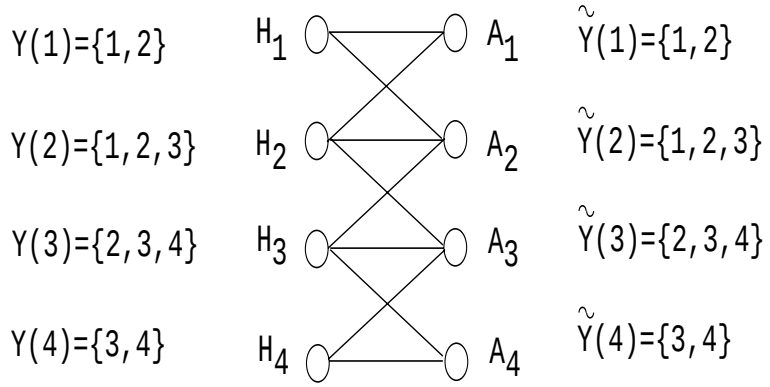

Figure 3. Bipartite graph illustrates search tolerance effects. Here $L = \tilde{L} = N = 4$.

Any classifier for this task can be specified with the tuple of disjoint regions $(R_0, R_1, \cdots, R_{\tilde{L}})$ where we take action A_i if the data vector $\mathbf{x}$ belongs to region R_i. Note that there are now $\tilde{L} + 1$ disjoint regions, not $L + 1$ disjoint regions as in Section 2. The false positive rate and the probability of correct localization are now given by:

$$\tilde{P}_{FP} = \sum_{i=1}^{\tilde{L}} \int_{R_i} p(\mathbf{x}|H_0) d\mathbf{x}. \tag{14}$$

$$\tilde{P}_{CL} = \frac{1}{1 - p(H_0)} \sum_{j=1}^{L} p(H_j) \sum_{i \in \mathcal{Y}(j)} \int_{R_i} p(\mathbf{x}|H_j) d\mathbf{x} = \frac{1}{1 - p(H_0)} \sum_{i=1}^{\tilde{L}} \int_{R_i} \sum_{j \in \tilde{\mathcal{Y}}(i)} p(\mathbf{x}|H_j) p(H_j) d\mathbf{x}. \tag{15}$$

Once again, we wish to find an optimal classifier that maximizes A_{LROC}. It turns out that this optimal classifier is given by:

$$\begin{aligned} t(\mathbf{x}) &= \max_{i \in \{1, \cdots, \tilde{L}\}} \sum_{j \in \tilde{\mathcal{Y}}(i)} \frac{p(\mathbf{x}|H_j) p(H_j)}{p(\mathbf{x}|H_0)} \\ i(\mathbf{x}) &= \arg \max_{i \in \{1, \cdots, \tilde{L}\}} \sum_{j \in \tilde{\mathcal{Y}}(i)} \frac{p(\mathbf{x}|H_j) p(H_j)}{p(\mathbf{x}|H_0)} \\ & \quad \text{Take action } A_{i(\mathbf{x})} \text{ if } t(\mathbf{x}) > \tau, \text{ otherwise take action } A_0 \end{aligned} \tag{16}$$

We shall refer to this classifier as the max-sum classifier. Note that if the point-to-set mappings constitute a complete bipartite graph, i.e. we do not care about the location reported, then the max-sum classifier reduces to the two-class ideal observer that computes a test statistic $q(\mathbf{x}) = \frac{\sum_{i=1}^{L} p(\mathbf{x}|H_i) p(H_i)}{p(\mathbf{x}|H_0)}$ and decides signal-present if $q(\mathbf{x}) > \tau$. In the case that the point-to-set mappings constitute an identity mapping, i.e. $L = \tilde{L}$ and $\tilde{\mathcal{Y}}(i) = i$, the max-sum classifier reduces to the modified GLRT in (4). The optimality of the max-sum classifier can be proven with the following lemma:

Lemma For all classifiers satisfying $\tilde{P}_{FP} = \alpha$, the classifier in (16) maximizes $\tilde{P}_{CL}$.

We will provide a proof of this lemma that is very similar to the first proof in Sec. 2.1. We have also derived a different proof similar to the second proof in Sec. 2.1, but do not provide it here for lack of space. Using (15)

and (14), we now need to maximize:

$$E = \frac{1}{1-p(H_0)} \sum_{i=1}^{\tilde{L}} \int_{R_i} \Big(\sum_{j \in \tilde{\mathcal{Y}}(i)} p(\mathbf{x}|H_j)p(H_j) + (1-p(H_0))\lambda p(\mathbf{x}|H_0) \Big) d\mathbf{x} \quad - \lambda\alpha \tag{17}$$

The maximum is achieved when the region $R_i,\ i \in \{1, \cdots, \tilde{L}\}$ corresponds to
$i = \arg\max_{k \in \{1,\cdots,\tilde{L}\}} \sum_{j \in \tilde{\mathcal{Y}}(k)} p(\mathbf{x}|H_j)p(H_j) + \lambda(1-p(H_0))p(\mathbf{x}|H_0)$ and
$\sum_{j \in \tilde{\mathcal{Y}}(i)} p(\mathbf{x}|H_i)p(H_i) + \lambda(1-p(H_0))p(\mathbf{x}|H_0) > 0$.
This corresponds to the max-sum classifier with $\tau = -\lambda(1-p(H_0))$.

Now, we will show that for a choice of costs very similar to those in Sec. 2.2, the optimal classifier in (16) minimizes the Bayes risk. The Bayes risk is now given by:

$$\mathcal{R} = \sum_{i=0}^{\tilde{L}} \sum_{j=0}^{L} c_{ij} p(A_i|H_j)p(H_j) \tag{18}$$

The classifier minimizing the Bayes risk[10] is given by:

$$\text{Take action } A_k \ \Leftrightarrow k = \arg \min_{i \in \{0,1,\cdots,\tilde{L}\}} C_i(\mathbf{x}) \text{ where } C_i(\mathbf{x}) = \frac{1}{p(\mathbf{x})} \sum_{j=0}^{L} c_{ij} p(\mathbf{x}|H_j)p(H_j) \tag{19}$$

Suppose we choose the following reasonable cost constraints similar to those in Sec. 2.2:

$$\begin{aligned}
c_{ij} &= C_{FN}, \text{ if } i = 0,\ j \in \{1, \cdots, L\} \\
c_{ij} &= C_{FP}, \text{ if } j = 0,\ i \in \{1, \cdots, \tilde{L}\} \\
c_{ij} &= C_{FL}, \text{ if } j \in \{1, \cdots, L\},\ i \in \{1, \cdots, \tilde{L}\},\ i \notin \mathcal{Y}(j) \\
c_{ij} &= C_{CL}, \text{ if } j \in \{1, \cdots, L\},\ i \in \{1, \cdots, \tilde{L}\},\ i \in \mathcal{Y}(j) \\
c_{00} &= C_{TN} \\
C_{FL} > C_{CL} &\quad \text{and} \quad C_{FP} > C_{TN} \\
\text{with } C_{FN} &= C_{FL}
\end{aligned} \tag{20}$$

With these costs, we get

$$\begin{aligned}
C_0(\mathbf{x}) &= C_{TN} \frac{1}{p(\mathbf{x})} p(\mathbf{x}|H_0)p(H_0) + C_{FL} \frac{1}{p(\mathbf{x})} \sum_{j=1}^{L} p(\mathbf{x}|H_j)p(H_j) \\
C_i(\mathbf{x}) &= \frac{1}{p(\mathbf{x})} \big[C_{FP} p(\mathbf{x}|H_0)p(H_0) + C_{FL} \sum_{j=1}^{L} p(\mathbf{x}|H_j)p(H_j) - \\
&\qquad (C_{FL} - C_{CL}) \sum_{j \in \tilde{\mathcal{Y}}(i)} p(\mathbf{x}|H_j)p(H_j) \big] \quad i \in \{1, \cdots, \tilde{L}\} \\
C_i(\mathbf{x}) &< C_{i'}(\mathbf{x})\ i, i' \in \{1, \cdots, \tilde{L}\}\ i \neq i' \Leftrightarrow \sum_{j \in \tilde{\mathcal{Y}}(i)} p(\mathbf{x}|H_j)p(H_j) > \sum_{j \in \tilde{\mathcal{Y}}(i')} p(\mathbf{x}|H_j)p(H_j)
\end{aligned} \tag{21}$$

$$C_0(\mathbf{x}) < C_i(\mathbf{x})\ \forall i \in \{1, \cdots, \tilde{L}\} \Leftrightarrow \frac{(C_{FP} - C_{TN})p(H_0)}{C_{FL} - C_{CL}} > C_{FL} \max_{i \in \{1,\cdots,\tilde{L}\}} \sum_{j \in \tilde{\mathcal{Y}}(i)} \frac{p(\mathbf{x}|H_j)p(H_j)}{p(\mathbf{x}|H_0)} \tag{22}$$

We may observe that as in Sec. 2.2, the classifier in (19) reduces to the classifier in (16) with $\tau = \frac{(C_{FP} - C_{TN})p(H_0)}{C_{FL} - C_{CL}}$.

Thus, we have established that for the discrete model including search tolerance, the decision strategy in (16) maximizes the area under the LROC curve and that it minimizes the Bayes risk with the cost constraints in (20).

4. LOCATION AS A CONTINUOUS PARAMETER

So far, we have modeled location as a discrete parameter with finitely many values, but it is more naturally treated as a continuous parameter. Now, we shall model signal location $\mathbf{r}$ as continuous, though $\mathbf{x}$ remains an N-dimensional vector. Let $p_{\mathbf{R}}(\mathbf{r})$ denote the pdf of the true location of the signal under the signal-present hypothesis. Given a data vector $\mathbf{x}$, our decision strategy will decide whether it belongs to R_0 (signal absent) or R_1 (signal present), and also report a location $\mathbf{r}(\mathbf{x})$ where the signal may be present. If the true location of the signal is $\mathbf{r}'$, we deem correct localization to have occurred if $|\mathbf{r}(\mathbf{x}) - \mathbf{r}'| < T$, where T is a tolerance radius. (We need not choose a ball-type tolerance region to obtain our results below, but we make this assumption without loss of generality simply because it seems quite reasonable.) We get:

$$\breve{P}_{FP} = \int_{R_1} p(\mathbf{x}|H_0) d\mathbf{x} \tag{23}$$

$$\breve{P}_{CL} = \int p_{\mathbf{R}}(\mathbf{r}') d\mathbf{r}' \int_{\mathbf{x} \in R_1 : |\mathbf{r}(\mathbf{x}) - \mathbf{r}'| < T} p(\mathbf{x}|\mathbf{r}', H_1) d\mathbf{x} = \int_{R_1} \int p_{\mathbf{R}}(\mathbf{r}') p(\mathbf{x}|\mathbf{r}', H_1) circ(\frac{|\mathbf{r}(\mathbf{x}) - \mathbf{r}'|}{T}) d\mathbf{r}' d\mathbf{x} \tag{24}$$

It turns out that the optimal decision strategy maximizing A_{LROC} is given by:

$$\begin{aligned} t(\mathbf{x}) &= \max_{\mathbf{r}''} \int \frac{p(\mathbf{x}|\mathbf{r}', H_1)}{p(\mathbf{x}|H_0)} p_{\mathbf{R}}(\mathbf{r}') circ(\frac{|\mathbf{r}'' - \mathbf{r}'|}{T}) d\mathbf{r}' \\ \mathbf{r}(\mathbf{x}) &= \arg\max_{\mathbf{r}''} \int \frac{p(\mathbf{x}|\mathbf{r}', H_1)}{p(\mathbf{x}|H_0)} p_{\mathbf{R}}(\mathbf{r}') circ(\frac{|\mathbf{r}'' - \mathbf{r}'|}{T}) d\mathbf{r}' \\ & \quad \text{Decide } H_1 \text{ if } t(\mathbf{x}) > \tau \text{ else decide } H_0 \end{aligned} \tag{25}$$

We shall again refer to this strategy as the max-sum decision strategy since it is almost the same as the max-sum strategy in (16) with the summation over the discrete locations now replaced by a integral over the continuous location parameter $\mathbf{r}'$. Note that as $T \to 0$, the max-sum decision strategy reduces to a modified GLRT in continuous space and as $T \to \infty$, the max-sum decision strategy reduces to a 2-class ideal observer. The optimality of this decision strategy can be proven with the following lemma:

Lemma For all decision strategies satisfying $\breve{P}_{FP} = \alpha$, the decision strategy in (25) maximizes $\breve{P}_{CL}$.

Proof We will provide a proof that is similar to the first proof in Sec. 2.1 and the proof in Sec. 3. We have also derived a different proof similar to the second proof in Sec. 2.1, but do not provide it here for lack of space. Using (23) and (24), we now need to find $\mathbf{r}(\mathbf{x})$ and R_1 in order to maximize:

$$E = \int_{R_1} \Big(\int p_{\mathbf{R}}(\mathbf{r}') d\mathbf{r}' p(\mathbf{x}|\mathbf{r}', H_1) circ(\frac{|\mathbf{r}(\mathbf{x}) - \mathbf{r}'|}{T}) + \lambda p(\mathbf{x}|H_0) \Big) d\mathbf{x} \ - \lambda\alpha \tag{26}$$

The maximum is achieved when $\mathbf{r}(\mathbf{x}) = \arg\max_{\mathbf{r}''} \big(\int p_{\mathbf{R}}(\mathbf{r}') d\mathbf{r}' p(\mathbf{x}|\mathbf{r}', H_1) circ(\frac{|\mathbf{r}'' - \mathbf{r}'|}{T}) + \lambda p(\mathbf{x}|H_0) \big)$ and the region R_1 corresponds to $\big(\int p_{\mathbf{R}}(\mathbf{r}') d\mathbf{r}' p(\mathbf{x}|\mathbf{r}', H_1) circ(\frac{|\mathbf{r}(\mathbf{x}) - \mathbf{r}'|}{T}) + \lambda p(\mathbf{x}|H_0) \big) > 0$. This corresponds to the max-sum classifier with $\tau = -\lambda$.

The Bayes risk in the continuous case is given by:

$$\begin{aligned} \mathcal{R} &= c_{00} p(H_0|H_0) p(H_0) + \int d\mathbf{r}\ c_{10}(\mathbf{r}) p(H_1, \mathbf{r}|H_0) p(H_0) + \int d\mathbf{r}'\ c_{01}(\mathbf{r}') p(H_0|H_1, \mathbf{r}') p(H_1, \mathbf{r}') \\ &\quad + \int d\mathbf{r}' \int d\mathbf{r}\ c_{11}(\mathbf{r}, \mathbf{r}') p(H_1, \mathbf{r}|H_1, \mathbf{r}') p(H_1, \mathbf{r}') \end{aligned} \tag{27}$$

where
c_{00} = cost of reporting that the signal is absent when the signal is absent,
$c_{10}(\mathbf{r})$ = cost of reporting that the signal is present at location $\mathbf{r}$ when the signal is absent,
$c_{01}(\mathbf{r}')$ = cost of reporting that the signal is absent when the signal is present at location $\mathbf{r}'$ and
$c_{11}(\mathbf{r}, \mathbf{r}')$ = cost of reporting that the signal is present at location $\mathbf{r}$ when the signal is located at $\mathbf{r}'$.

Lemma A set of costs for which the max-sum decision strategy minimizes the Bayes risk in (27) are given by:

$$\begin{aligned} c_{01}(\mathbf{r}') &= C_{FN} \\ c_{10}(\mathbf{r}) &= C_{FP} \\ c_{11}(\mathbf{r},\mathbf{r}') &= C_{FL}\big(1 - circ(\frac{|\mathbf{r}-\mathbf{r}'|}{T})\big) + C_{CL}circ(\frac{|\mathbf{r}-\mathbf{r}'|}{T}) \\ c_{00} &= C_{TN} \\ C_{FL} > C_{CL} \quad &\text{and} \quad C_{FP} > C_{TN} \\ \text{with } C_{FN} &= C_{FL} \end{aligned} \tag{28}$$

(Note that the expression for $c_{11}(\mathbf{r},\mathbf{r}')$ is slightly subtle. It says that if the reported location is within the search tolerance, then $c_{11}(\mathbf{r},\mathbf{r}') = C_{CL}$ as in (20). Otherwise, $c_{11}(\mathbf{r},\mathbf{r}') = C_{FL}$.)

Proof With the costs in (28), the Bayes risk becomes:

$$\begin{aligned} \mathcal{R} &= \int_{R_0} d\mathbf{x} C_{TN} p(\mathbf{x}|H_0)p(H_0) + \int_{R_1} d\mathbf{x} C_{FP} p(\mathbf{x}|H_0)p(H_0) + \int_{R_0} d\mathbf{x} \int d\mathbf{r}' C_{FL} p_{\mathbf{R}}(\mathbf{r}')p(H_1)p(\mathbf{x}|\mathbf{r}',H_1) + \\ &\quad \int_{R_1} d\mathbf{x} \int d\mathbf{r}' \Big[C_{FL}\big(1 - circ(\frac{|\mathbf{r}-\mathbf{r}'|}{T})\big) + C_{CL}circ(\frac{|\mathbf{r}-\mathbf{r}'|}{T})\Big] p(\mathbf{x}|\mathbf{r}',H_1)p_{\mathbf{R}}(\mathbf{r}')p(H_1) \\ &= \int d\mathbf{x} C_{TN} p(\mathbf{x}|H_0)p(H_0) + C_{FL}\int d\mathbf{r}' p_{\mathbf{R}}(\mathbf{r}') \int d\mathbf{x} p(\mathbf{x}|\mathbf{r}',H_1)p(H_1) \\ &\quad + \int_{R_1} d\mathbf{x}\Big((C_{FP}-C_{TN})p(\mathbf{x}|H_0)p(H_0) - \int d\mathbf{r}'(C_{FL}-C_{CL})circ(\frac{|\mathbf{r}-\mathbf{r}'|}{T})p(\mathbf{x}|\mathbf{r}',H_1)p_{\mathbf{R}}(\mathbf{r}')p(H_1)\Big) \\ &= C_{TN}p(H_0) + C_{FL}p(H_1) + \\ &\quad (C_{FL}-C_{CL})p(H_1)\int_{R_1} d\mathbf{x}\Big(\frac{(C_{FP}-C_{TN})p(H_0)}{(C_{FL}-C_{CL})p(H_1)}p(\mathbf{x}|H_0) - \int d\mathbf{r}' circ(\frac{|\mathbf{r}(\mathbf{x})-\mathbf{r}'|}{T})p(\mathbf{x}|\mathbf{r}',H_1)p_{\mathbf{R}}(\mathbf{r}')\Big) \end{aligned} \tag{29}$$

Noting the similarity between the integral in (26) and the integral in the last equality in (29), we can conclude that the Bayes risk is minimized by the max-sum strategy in (25) with $\tau = \frac{(C_{FP}-C_{TN})p(H_0)}{(C_{FL}-C_{CL})p(H_1)}$.

Our model in this section is a special case of the hybrid detection-estimation model treated in Sec. 13.3.9 of Barrett and Myers.[1] However, in Barrett and Myers,[1] the following optimal decision strategy is adopted: First apply a 2-class ideal observer to decide if the signal is present, and if present, estimate its location using a MAP estimator. It turns out that this strategy minimizes the Bayes risk for a particular choice of costs. In the notation comparable to (28), the costs take the form:

$$\begin{aligned} c_{01}(\mathbf{r}') &= C_{FN} \\ c_{10}(\mathbf{r}) &= C_{FP} \\ c_{11}(\mathbf{r},\mathbf{r}') &= C_{FL}\big(1 - rect(\frac{|\mathbf{r}-\mathbf{r}'|}{\epsilon})\big) \; \epsilon \to 0 \\ c_{00} &= C_{TN} \end{aligned} \tag{30}$$

This form of observer does not maximize A_{LROC} and, correspondingly, does not include a notion of search tolerance T. It may seem that our cost constraints (28) may be made a special case of (30) by taking the limit as $T \to 0$ and $C_{CL} = 0$. However, the derivation in Barrett and Myers[1] has an implicit assumption $C_{FN} \neq C_{FL}$. Our derivation requires the opposite constraint, $C_{FN} = C_{FL}$, and hence even in the case $T \to 0$, the two optimal strategies differ.

Thus, we have established that for the case of continuous signal location, the decision strategy in (25) maximizes the area under the LROC curve and that it minimizes the Bayes risk with the cost constraints in (28).

5. DISCUSSION

For the joint detection and localization task, we have proposed decision strategies optimal in the sense that they maximize A_{LROC}. While A_{LROC} maximization is useful in itself, the same decision strategy can *also* be shown to minimize Bayes risk under certain cost constraints previously described. All of these cost constraints are fairly non-restrictive with the exception of the one constraint that the cost for a false negative equals that for a false localization ($C_{FN} = C_{FL}$). Such an assumption seems appropriate in certain medical tasks, for example, in diagnostic mammography scenarios where a patient is subjected to biopsy only at the location identified as suspicious from the mammogram. This assumes that the primary cost incurred is in not detecting the lesion and the cost of conducting the biopsy is ignored*. Even if this cost constraint does not hold for an application, our decision strategies are still useful in maximizing A_{LROC}.

It is interesting to consider the task of detection only, but with signal location uncertainty. In this case, one would seek a decision strategy to maximize AUC, which is the scalar task performance metric for this case. The implementation of such a 2-class decision strategy is discussed in Barrett *et al.*[12] and Park *et al.* 2003,[13] with location a real-valued parameter. In terms of our own notation, for the problem setup in Sec. 2, the optimal decision strategy maximizing AUC then requires the computation of a likelihood ratio test statistic $q(\mathbf{x}) = \frac{\sum_{i=1}^{L} p(\mathbf{x}|H_i)p(H_i)}{p(\mathbf{x}|H_0)}$ and we decide signal-present if $q(\mathbf{x}) > \tau$.

We also note that our $(L+1)$ class classification problem in Sec. 2 has a superficial similarity to the 3-class classification problem discussed in Edwards, Metz and Nishikawa.[14] There, the classifier needs to first decide if a tumor is present or not and then, if present, decide if it is malignant or benign. However, a quantity analogous to $P_{CL}(\tau)$ is inappropriate for summarizing performance on this task since it aggregrates all correct tumor-present and correctly identified decisions (i.e. when we correctly identify a tumor as malignant or benign) and we lose the ability to separately observe how often we correctly identified a tumor as benign and how often we correctly identified a tumor as malignant. From a Bayes risk perspective, it seems unreasonable to assign equal costs (analogous to C_{FL}) for misclassifying benign and malignant tumors (assuming that it has been correctly established that a tumor is present). The assumption analogous to $C_{FN} = C_{FL}$, namely that the costs for missing tumors and misclassifying them are equal, also seems unreasonable for this problem. Hence a metric analogous to A_{LROC} seems to be inappropriate for this task.

We note that ideal observers for the N-class classification task have been investigated in Edwards, Metz and Kupinski.[15] These ideal observers minimize the Bayes risk and an N-class version of the Neyman-Pearson lemma has been derived to show that they also minimize the hypervolume under a certain ROC hypersurface. However, these decision strategies are not designed to optimize any joint detection-localization performance measure.

We note that in Swensson,[2] under certain assumptions and for certain decision strategies, it was shown that $A_{LROC} = 2\ AUC - 1$. We make no assumptions in computing A_{LROC}. In fact, the assumptions are tantamount to committing to a particular search tolerance, and in our models, search tolerance is user-specified. Therefore, the relation $A_{LROC} = 2AUC - 1$ does not generally hold for our decision strategies.

We note that it may be possible to implement the optimal decision strategies (modified GLRT and max-sum) of Sec. 2.1, 3 and 4 for emission imaging systems with realistic background models by modifying the MCMC procedure developed for 2-class ideal observers in Park *et al.* 2003[13] and Park *et al.* 2005.[16] The performance of these optimal decision strategies might be used for imaging hardware design. It may also be used for assessing the efficiency[16] of human observers (or model observers emulating humans), which may be useful in optimizing the image processing that is performed on the data delivered by the imaging hardware. It remains to be seen whether the optimal system or data processing parameters change significantly when the task changes from a detection to a joint detection-localization task.

6. ACKNOWLEDGEMENTS

This work was supported by grant NIH-NIBIB EB02629.

*This mammography example was suggested by C. K. Abbey.

REFERENCES

1. H. H. Barrett and K. J. Myers, *Foundations of Image Science*, Wiley Interscience, 2003.
2. R. Swensson, "Unified measurement of observer performance in detecting and localizing target objects on images," *Medical Physics* **23**, pp. 1709–1725, Oct. 1996.
3. H. C. Gifford, M. A. King, R. Wells, W. Hawkins, M. Narayanan, and H. Pretorius, "LROC analysis of detector-response compensation in SPECT," *IEEE Trans. Med. Imaging* **19**, pp. 463–, May 2000.
4. T. Farncombe, H. C. Gifford, M. Narayanan, H. Pretorius, E. Frey, and M. A. King, "Assessment of scatter compensation strategies for Ga-67 SPECT using numerical observers and human LROC studies," *Journal of Nuclear Medicine* **45**, pp. 802–812, May 2004.
5. T. Farquhar, J. Llacer, J. Sayre, Y. Tai, and E. Hoffman, "ROC and LROC analyses of the effects of lesion contrast, size and signal-to-noise ratio on detectability in PET images," *Journal of Nuclear Medicine* **41**, pp. 745–754, Apr. 2000.
6. M. Kallergi, J. Heine, C. Berman, M. Hersh, A. Romilly, and R. Clark, "Improved interpretation of digitized mammography with wavelet processing: A localization response operating characteristic study," *American Journal of Roentgenology* **182**, pp. 697–703, Mar. 2004.
7. P. Khurd and G. Gindi, "Rapid computation of LROC figures of merit using numerical observers (for SPECT/PET reconstruction)," *IEEE Trans. Nuclear Science* , To appear 2005.
8. S. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*, Prentice Hall, 1998.
9. K. Fukunaga, *Statistical Pattern Recognition*, Academic Press, 1990.
10. R. Duda, P. Hart, and D. Stork, *Pattern Classification*, Wiley Interscience, 2000.
11. G. Casella and R. Berger, *Statistical Inference*, Brooks Cole, 2nd edition, 2001.
12. H. H. Barrett and C. K. Abbey, "Bayesian detection of random signals on random backgrounds," in *Info. Processing In Med. Imag.*, **1230**, pp. 155–166, Springer-Verlag, 1997.
13. S. Park, M. Kupinski, E. Clarkson, and H. H. Barrett, "Ideal observer performance under signal and background uncertainty," in *Info. Processing In Med. Imag.*, pp. 342–353, 2003.
14. D. Edwards, C. Metz, and R. Nishikawa, "Hypervolume Under the ROC Hypersurface of a "near-guessing" ideal observer in a three-class classification task," in *Conf. Rec. SPIE Med. Imag.*, **5372**, pp. 128–137, SPIE, Feb. 2004.
15. D. Edwards, C. Metz, and M. Kupinski, "Ideal Observers and Optimal ROC Hypersurfaces in N-class Classification," *IEEE Trans. Med. Imaging* **23**, pp. 891–895, July 2004.
16. S. Park, E. Clarkson, M. A. Kupinski, and H. H. Barrett, "Efficiency of the human observer detecting random signals in random backgrounds," *J. Optical Soc. America A* **22**, pp. 3–16, Jan. 2005.