

# Computer-assisted detection of mammographic masses: A template matching scheme based on mutual information

Georgia D. Tourassi<sup>a)</sup> and Rene Vargas-Voracek  
*Department of Radiology, Duke University Medical Center, Durham, North Carolina 27710*

David M. Catarious, Jr.  
*Department of Biomedical Engineering, Duke University, Durham, North Carolina 27710*

Carey E. Floyd, Jr.  
*Department of Radiology, Duke University Medical Center, Durham, North Carolina 27710 and Department of Biomedical Engineering, Duke University, Durham, North Carolina 27710*

(Received 29 October 2002; accepted for publication 12 May 2003; published 24 July 2003)

The purpose of this study was to develop a knowledge-based scheme for the detection of masses on digitized screening mammograms. The computer-assisted detection (CAD) scheme utilizes a knowledge databank of mammographic regions of interest (ROIs) with known ground truth. Each ROI in the databank serves as a template. The CAD system follows a template matching approach with mutual information as the similarity metric to determine if a query mammographic ROI depicts a true mass. Based on their information content, all similar ROIs in the databank are retrieved and rank-ordered. Then, a decision index is calculated based on the query's best matches. The decision index effectively combines the similarity indices and ground truth of the best-matched templates into a prediction regarding the presence of a mass in the query mammographic ROI. The system was developed and evaluated using a database of 1465 ROIs extracted from the Digital Database for Screening Mammography. There were 809 ROIs with confirmed masses (455 malignant and 354 benign) and 656 normal ROIs. CAD performance was assessed using a leave-one-out sampling scheme and Receiver Operating Characteristics analysis. Depending on the formulation of the decision index, CAD performance as high as  $A_z = 0.87 \pm 0.01$  was achieved. The CAD detection rate was consistent for both malignant and benign masses. In addition, the impact of certain implementation parameters on the detection accuracy and speed of the proposed CAD scheme was studied in more detail. © 2003 American Association of Physicists in Medicine.  
[DOI: 10.1118/1.1589494]

Key words: mammography, computer-assisted detection (CAD), knowledge-based, mutual information, receiver-operating characteristic (ROC)

## I. INTRODUCTION

Breast cancer is one of the most devastating and deadly diseases for women.<sup>1</sup> While there are many exciting new techniques on the horizon, for the time being mammography remains the screening test in the battle against breast cancer. Patients with early-detected malignancies have a significantly lower mortality rate.<sup>2,3</sup> Unfortunately, it is reported that up to 30% of breast lesions go undetected in screening mammograms<sup>4-6</sup> and up to 2/3 of those lesions are visible in retrospect.<sup>7</sup> Breast masses comprise a significant portion of missed cancers.<sup>4,5</sup> The clinical significance of early diagnosis and the difficulty of the diagnostic task have generated a tremendous interest in developing computer-assisted detection (CAD) schemes for mammographic interpretation. Several studies have demonstrated that CAD technology has a positive impact on early breast cancer detection.<sup>5,7,8</sup> However, there are still unresolved issues related to the clinical role of CAD in mammography. For example, the CAD detection accuracy is reportedly lower for masses than for calcifications.<sup>9,10</sup> Since high sensitivity is essential in screening mammography, CAD is often compromised by a higher

false-positive rate in the detection of breast masses. The impact of false-positive CAD cues on the recall rate of mammograms is under investigation.<sup>5,8</sup> Generally, it is assumed that the radiologists will be able to discard easily most of the false-positive cues. However, a recent study has challenged this belief.<sup>11</sup> The study also showed that low-performing CAD tools degrade radiologists' performance in noncued areas. Therefore, it is recommended that a cueing CAD tool should be used by an experienced interpreter to effectively process all cues.<sup>10</sup> However, the medical and legal implications of dismissing CAD cues are currently unknown.<sup>10</sup> Consequently, CAD research efforts in mammography are ongoing.

Thus far, the overwhelming majority of CAD techniques follow a two-step approach (e.g., Refs. 9, 12-23). Initially, traditional image processing is performed to identify suspicious mammographic regions. Subsequently, morphological and/or textural features are automatically extracted from these regions. The features are merged with linear classifiers or artificial intelligence techniques to further refine the detection and often the diagnosis (benign versus malignant) of potential abnormalities. The suspicious mammographic re-

gions detected by the CAD system serve as cues to the radiologists. Commercially available products are designed to operate as black boxes that provide diagnostic cues but not comprehensible decision models. In addition, some researchers have developed mathematical models to describe the statistical nature of mammograms.<sup>24,25</sup> Such models could be potentially extended to perform as CAD tools.

The purpose of this study is to develop a knowledge-based (KB) CAD scheme for the detection of breast masses in digitized mammograms. Generally, knowledge-based CAD (KB-CAD) systems aim to provide evidence-based decision support using a knowledge databank. Much like a physician relates a present case to those seen in the past, a KB system relates a new case to similar cases stored in its knowledge databank. Based on the similar cases, a diagnosis is assigned to the new case by analogy or by copying the answer if the match is close enough. The main benefits of using KB-CAD systems are the following: (1) KB-CAD systems take full advantage of growing data libraries without further re-training of the CAD system and, (2) they can be interactive allowing physicians to formulate their own questions and get interpretable answers.

The computational demands of maintaining, indexing, and querying a large knowledge databank have limited the application of these tools in mammography. Furthermore, defining similarity between two images is nontrivial. There is not a single similarity metric that is known to produce the best results in all applications. Common practice is to select diagnostically important features and feature-based distance metrics to determine similarity. Case-based reasoning (CBR) is a typical example of a KB system and it has been successfully applied for mammographic diagnosis based on radiologist-extracted BIRADS findings.<sup>26,27</sup> In addition, Chang *et al.* showed the feasibility of using a KB-CAD system for the detection of mammographic masses.<sup>28</sup> Their system employed a feature-based similarity metric that required segmentation of the suspected masses.

In contrast, our proposed KB-CAD scheme follows an image retrieval approach that is not feature-based but uses template matching with a global similarity metric. Template matching requires comparison of a given image with a template image. Each mammographic case stored in the knowledge databank serves as a template. Given a query mammographic region, the KB-CAD scheme retrieves similar cases from its knowledge databank. The focus of this study is to investigate mutual information (MI) as a potential similarity metric for knowledge-based detection of masses in screening mammograms.

MI is a fundamental concept in information theory.<sup>29</sup> It is defined in terms of two objects (i.e., images) and it measures how much one object can explain the other. Thus, MI captures the similarity or the amount of relevant information between two objects.<sup>29</sup> In medical imaging, MI has been a very effective similarity metric for image registration tasks.<sup>30</sup> The basic idea is that when two images are properly aligned, their MI is maximal. Our study aims to evaluate if MI can serve as a similarity metric in a template-matching scheme for the detection of mammographic masses. We hypothesize

that if two mammographic regions depict similar structures, they should contain relevant diagnostic information for each other. Therefore, by measuring their MI we can potentially quantify their diagnostic similarity. Furthermore, the MI is calculated directly from the images without any preprocessing. By using MI as a global similarity metric, we avoid issues related to image segmentation, feature extraction, and feature selection that are typically associated with feature-based similarity metrics or feature-based CAD schemes.

## II. MATERIALS AND METHODS

### A. The image database

The CAD system was developed and evaluated using the Digital Database for Screening Mammography (DDSM) that was collected at the University of South Florida under the DOD Breast Cancer Research Program Grant No. DAMD17-94-J-4015.<sup>31</sup> DDSM is intended as a benchmark database for CAD tools on screening mammograms. The database includes normal, cancer, and benign cases. A DDSM mammogram is considered normal if no further evaluation was required and the patient had a normal screening exam at least four years later. A cancer case is a screening mammogram with at least one biopsy-proven malignancy. A benign case is a screening mammogram with a suspicious finding that was determined to be benign by pathology or additional imaging.

DDSM includes three volumes, each containing mammograms digitized with a different digitizer (LUMISYS, HOWTEK, and DBA). Each DDSM screening exam consists of two images for each breast (standard craniocaudal and mediolateral oblique views). Our study focused on the DDSM mammograms digitized using the LUMISYS scanner. Initially, these mammograms were downloaded and archived. From those, all mammograms with annotated masses were selected. Specifically, all malignant masses present in the sets "cancer\_02," "cancer\_05," "cancer\_09," and "cancer\_15" were identified. Similarly, all benign masses present in the sets "benign\_01," "benign\_04," "benign\_06," "benign\_13," and "benign\_14" were identified. There were 260 studies with malignant masses and 146 studies with benign masses. Some masses were visible in one mammographic view only.

The DDSM includes information describing the location of the masses.  $512 \times 512$  pixel regions of interest (ROIs) centered on the known location of each annotated mass were extracted. In addition,  $512 \times 512$  pixel ROIs depicting normal tissue were also extracted. The normal ROIs were extracted from the sets "normal\_09" and "normal\_10." The two sets included 82 patients with normal screening mammograms. Two  $512 \times 512$  pixel ROIs were randomly chosen from each view per breast. Thus, eight ROIs were extracted from each DDSM patient with a normal screening exam. There were 1465 ROIs in total; 455 ROIs depicting a biopsy-proven malignant mass, 354 ROIs with a benign mass, and the remaining 656 ROIs were normal. To facilitate detailed analysis according to the difficulty level of the detection task, all extracted ROIs were further indexed according to the den-

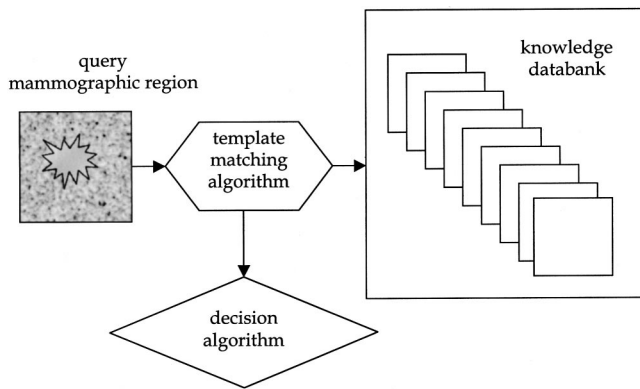


FIG. 1. Overview of the KB-CAD scheme.

sity rating of their corresponding mammogram. The ACR density rating is part of the associated patient information that is provided in the DDSM database.

## B. Overview of the CAD scheme

Figure 1 highlights the three critical components of our KB-CAD scheme: (1) the knowledge databank, (2) the template matching algorithm, and (3) the knowledge-based decision algorithm. The knowledge databank contains mammographic ROIs that depict masses of known truth or normal tissue. Each ROI stored in the databank serves as a template. A query suspicious mammographic region is compared to the stored templates using the template-matching algorithm. Based on their information content, all similar templates in the databank are retrieved. A decision algorithm effectively combines the similarity indices and known truth of the retrieved templates into a prediction regarding the presence of a mass in the suspicious query mammographic region.

## C. The template-matching algorithm

This section describes the algorithm employed in the study to measure the similarity between a query mammographic region and a template ROI stored in the knowledge databank. The algorithm utilizes mutual information (MI), a similarity index borrowed from information theory.<sup>29</sup>

Mutual information is a measure of general interdependence between two random variables  $x$  and  $y$ .<sup>29</sup> The MI concept can be easily extended to images. Given two images  $X$  and  $Y$ , their MI  $I(X;Y)$  is expressed as

$$I(X;Y) = \sum_x \sum_y P_{XY}(x,y) \log_2 \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}, \quad (1)$$

where  $P_{XY}(x,y)$  is the joint probability density function (pdf) of the two images based on their corresponding pixel values.<sup>29</sup> Equation (1) assumes that the image pixel values are samples of two random variables  $x$  and  $y$ , respectively.  $P_X(x)$  and  $P_Y(y)$  are the marginal pdfs. The basic idea is that when two images are similar, pixels with a certain intensity value in one image should correspond to a more clustered distribution of the intensity values in the other image.<sup>30</sup> The more the two images are alike, the more information  $X$

provides for  $Y$  and vice versa. Therefore, the MI can be thought as an intensity-based measure of how much two images are alike. In the template-matching context, the MI increases when the query image  $X$  and the template image  $Y$  depict similar structures. Then, the pixel value in image  $X$  is a good predictor of the pixel value at the corresponding location in image  $Y$ .

Theoretically, MI is a more effective and robust similarity metric than traditional correlation.<sup>32</sup> Correlation techniques assume a linear relationship between the intensity values in the two images. MI measures general dependence without making any *a priori* assumptions.

The MI estimation of two mammographic ROIs requires computation of the joint and marginal pdfs as shown in Eq. (1). There are two published methods for the task: (1) Parzen windows,<sup>33</sup> and (2) the histogram approach.<sup>34</sup> We followed the histogram approach since it is quick and easy to implement. Time efficiency is very important for a knowledge-based CAD system.

According to the histogram method, a pdf is approximated using a histogram. For each histogram bin, the probability is estimated by counting the number of pixels that fall into a particular bin and dividing that number by the total number of pixels. Then, the MI of two images  $X$  and  $Y$  can be computed according to Eq. (1).

The number of bins selected for histogram approximation is a critical issue.<sup>35</sup> More bins allow for more detailed representation of the pdfs. However, these details may be nothing more than noise caused by the small sample size in each bin. The potential estimation error can substantially alter the results of a study.<sup>36</sup> Since the DDSM images considered in our study are 12 bit images, the  $4096 \times 4096$  2D histogram required for the estimation of the joint pdf of two mammographic ROIs will be very sparse leading to serious MI estimation errors. Following typical practices of image registration applications, the pdfs were estimated using a reduced number of 256 equal-sized intensity bins for the histogram approximation technique. Furthermore, since the distribution for the pixel values can vary substantially among ROIs we applied the following rules. For each ROI, the mean  $\mu$  and standard deviation  $\sigma$  of the ROI pixel values were calculated. Then, the interval  $[\mu - 2\sigma, \mu + 2\sigma]$  was divided into the pre-selected number of equal segments (i.e., 256). Any rare pixel values falling outside the predetermined interval were assigned to the extreme left or right bins when calculating the histograms. The above rules were followed consistently for all ROIs.

## D. The knowledge-based decision index

The knowledge-based decision index was computed using the level of similarity and the ground truth of the best-matched templates. Two experiments were performed to determine the most effective way to use the CAD system as a computer aid for the detection of mammographic masses. In the first experiment, the knowledge databank included only

TABLE I. ROC performance of the CAD scheme for two decision indices ( $D_1, D_2$ ) and for varying number of the top matches considered ( $k=1, 10, 50, 100, 200, 400, \text{ALL}$ ). The MI calculations were based on 256 histogram bins and the full resolution  $512 \times 512$  ROIs.

	1	10	50	100	200	400	ALL
$D_1$	$0.71 \pm 0.01$	$0.71 \pm 0.01$	$0.71 \pm 0.01$	$0.72 \pm 0.01$	$0.73 \pm 0.01$	$0.74 \pm 0.01$	$0.75 \pm 0.01$
$D_2$	$0.71 \pm 0.01$	$0.79 \pm 0.01$	$0.84 \pm 0.01$	$0.85 \pm 0.01$	$0.85 \pm 0.01$	$0.86 \pm 0.01$	$0.87 \pm 0.01$

mammographic ROIs that contained a mass. In the second experiment, the knowledge databank included both normal and mass ROIs.

*Experiment 1:* Given a query mammographic ROI  $Q_i$ , a decision index was calculated based on the MI between the query ROI and each known mass  $M_j$  in the knowledge databank. The decision index  $D_1(Q_i)$  was the average MI of the  $k$  best mass matches:

$$D_1(Q_i) = \frac{1}{k} \sum_{j=1}^k \text{MI}(Q_i, M_j). \quad (2)$$

Theoretically, a query ROI depicting a mass should match better with the databank of mass ROIs than a query ROI depicting normal breast tissue, thus resulting in a higher  $D_1(Q_i)$ .

*Experiment 2:* Given a query mammographic ROI  $Q_i$ , a decision index  $D_2(Q_i)$  was calculated as the difference of two terms. The first term measures the average MI between the query ROI and its  $k$  best mass matches  $M_j$ . Similarly, the second term measures the average MI between the query ROI and its  $k$  best normal  $N_j$  matches,

$$D_2(Q_i) = \frac{1}{k} \sum_{j=1}^k \text{MI}(Q_i, M_j) - \frac{1}{k} \sum_{j=1}^k \text{MI}(Q_i, N_j). \quad (3)$$

Theoretically, a query ROI depicting a mass should have a higher  $D_2(Q_i)$ .

## E. Performance evaluation

The diagnostic performance of the CAD system was evaluated using a leave-one-out sampling scheme.<sup>37</sup> Given the database of 1465 mammographic ROIs, each ROI was excluded once to serve as a query case. In experiment 1, the remaining mass cases were used to establish the knowledge databank. In experiment 2, the remaining 1464 cases were used to establish the databank. The experiments were repeated until every ROI served as a query ROI. The calculated decision indices  $D_1$  and  $D_2$  were analyzed based on Receiver Operating Characteristic (ROC) analysis methodology. The ROCKIT software package developed by Metz *et al.* (available at [www-radiology.uchicago.edu/krl/toppage11.htm](http://www-radiology.uchicago.edu/krl/toppage11.htm)) was used to fit ROC curves to the two decision indices implemented in this study. For both indices, ROC performance was estimated for varying values of the top matches (parameter  $k$ ) considered.

## F. Influence of implementation parameters

In a knowledge-based system, comparing a query case with every archived case can be computationally expensive. This is certainly a concern with image databases and global similarity metrics such as the mutual information. One way to reduce the computation time is by reducing the number of histogram bins employed for the MI estimation. We repeated the previous experiments estimating the MI using 64 and 128 histogram bins to evaluate the impact of this implementation parameter on the overall performance of the CAD scheme. In addition, we studied the effect of image sub-sampling. Since a knowledge-based CAD system requires individual comparisons of the query ROI with all stored ROIs, it can be computationally more effective if the comparisons are performed on reduced-resolution ROIs. We repeated the above-mentioned experiments with sub-sampled ROIs ( $256 \times 256$ ,  $128 \times 128$ , and  $64 \times 64$ ) to determine if the CAD detection rate degrades for sparsely sampled ROIs.

## III. RESULTS

The experimental results are presented in two sections. Each section addresses an important issue: (1) overall ROC performance, (2) influence of the implementation parameters on performance and time efficiency of the proposed CAD scheme.

### A. Overall ROC performance of the CAD scheme

No particular trend was observed in obtaining higher MI values with template ROIs extracted from the same mammogram as the query ROI. Therefore, the overall detection performance of the KB-CAD scheme was analyzed on a per ROI basis, not on a per-case basis. Table I shows the performance of the CAD system as measured by the ROC area index ( $A_z$ ) for each one of the decision indices  $D_1$ ,  $D_2$  and for varying number of the top matches considered (parameter  $k$ ).

Several observations can be made based on Table I. The performance of the KB-CAD scheme varied substantially depending on the decision algorithm. Overall, the CAD system had a significantly better ROC performance when the decision index was calculated using the knowledge databank that includes both mass and normal templates ( $D_2$ ). Furthermore, CAD performance improved as more matched cases were considered in the calculation of the decision index  $D_2$ . The CAD system achieved its best ROC performance ( $A_z = 0.87 \pm 0.01$ ) when all archived cases were included in the calculation of  $D_2$ . However, when the detection decision was

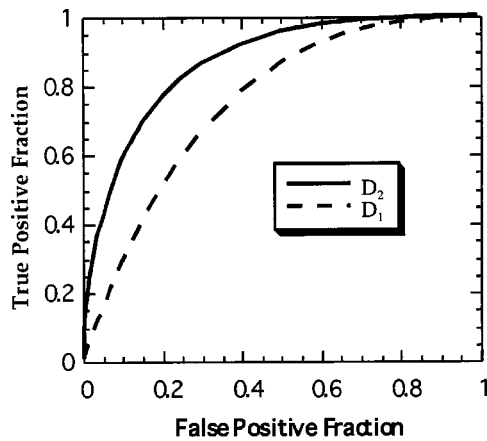


FIG. 2. ROC curves of the KB-CAD scheme based on the two decision algorithms ( $D_1$  and  $D_2$ ). The calculation of the two decision indices includes all archived templates.

based only on the best mass matches ( $D_1$ ), the ROC area index was statistically significantly lower ( $A_z=0.75\pm 0.01$ ) but substantially less dependent on the parameter  $k$ . Figure 2 shows the corresponding ROC curves of the CAD system based on the two decision algorithms ( $D_1$  and  $D_2$ ) for  $k = \text{ALL}$ . As the figure shows the best performing knowledge-based CAD scheme achieved 90% sensitivity while safely eliminating 65% of the normal regions.

The best performing CAD scheme was analyzed in more detail. First, detection accuracy was evaluated separately for malignant and benign masses. The CAD scheme showed robust performance among the two groups of masses:  $A_z(\text{malignant masses versus normal})=0.88\pm 0.01$  and  $A_z(\text{benign masses versus normal})=0.86\pm 0.01$ . A small subset of mammographic ROIs (57 out of 809 mass regions) contained both a mass and microcalcifications. Significant degradation in ROC performance was observed for this subset ( $A_z=0.80\pm 0.04$ ) compared to the remaining set of mass regions ( $A_z=0.89\pm 0.01$ ).

To assess the effect of case difficulty, the best performing CAD scheme was further analyzed for each subgroup of masses according to their DDSM subtlety rating. The mass subtlety rating is not a BI-RADS standard. It is simply a subjective impression of the DDSM radiologist on the subtlety of the lesion. A higher subtlety rating indicates a more obvious lesion. Table II shows that the overall ROC area index of the CAD tool is fairly robust regardless of the reported subtlety of the mass ROIs. The only exception is the subgroup of masses with Subtlety rating 2. For this subgroup, the KB-CAD had a statistically significantly lower ROC performance than the other subgroups.

TABLE II. Effect of mass subtlety rating on the overall ROC performance of the KB-CAD scheme.

	Subtlety=1	Subtlety=2	Subtlety=3	Subtlety=4	Subtlety=5
ROC $A_z$	$0.87\pm 0.04$	$0.79\pm 0.03$	$0.86\pm 0.02$	$0.85\pm 0.01$	$0.89\pm 0.01$

TABLE III. Effect of mammographic density on the ROC performance of the KB-CAD scheme.

Mammographic density	No. of mass ROIs	No. of normal ROIs	$A_z$
1: fatty breast	193	96	$0.98\pm 0.01$
2: fibroglandular breast	362	272	$0.91\pm 0.01$
3: heterogeneous breast	195	208	$0.87\pm 0.02$
4: dense breast	59	80	$0.64\pm 0.05$

Since mass detection is more challenging in dense breasts, we also analyzed the CAD performance for each subgroup of ROIs based on the DDSM density rating of the mammogram from which they were extracted. Table III summarizes those results. Table III shows that the ROC area varied significantly, starting from almost perfect performance in fatty breasts ( $A_z=0.98\pm 0.01$ ) and progressively degrading in fibroglandular ( $A_z=0.91\pm 0.01$ ) and heterogeneous breasts ( $A_z=0.87\pm 0.02$ ). The CAD performance was dramatically lower for dense mammograms ( $A_z=0.64\pm 0.05$ ) than for all remaining categories. Since the ROIs extracted from dense mammograms comprised only 10% (139/1465) of the whole data set, it is unclear if the inferior performance can be partially contributed to the lower representation of dense ROIs in the knowledge databank.

## B. Influence of implementation parameters on CAD performance

Tables IV and V demonstrate the impact of two implementation parameters on the overall ROC area index of the KB-CAD scheme. The first parameter is the number of histogram bins used in the calculation of the MI between two ROIs. The second parameter is the sub-sampling factor of the mammographic regions. Table IV shows the impact of both parameters on decision index  $D_1$ . Table V corresponds to decision index  $D_2$ . The calculation of both decision indices was based on all archived cases ( $k = \text{ALL}$ ).

Two important conclusions can be drawn from the above-mentioned tables. First, when estimating the MI between two ROIs, the number of histogram bins should be selected carefully. CAD performance can be significantly degraded as the number of histogram bins increases. The degradation is particularly strong with the coarser ROIs; using a large number of bins introduces serious estimation errors due to the smaller number of pixels available in each bin. However, there is no such concern with the full-resolution ROIs. Sec-

TABLE IV. Effect of image sub-sampling ( $256\times 256, 128\times 128, 64\times 64$ ) and the number of histogram bins (64, 128, 256) on the overall ROC area index of the KB-CAD scheme for decision index  $D_1$ . The full resolution ROIs are  $512\times 512$ . The reduced size ROIs were created by sub-sampling accordingly the full resolution ROIs.

	$512\times 512$	$256\times 256$	$128\times 128$	$64\times 64$
64 bins	$0.75\pm 0.01$	$0.75\pm 0.01$	$0.75\pm 0.01$	$0.72\pm 0.01$
128 bins	$0.75\pm 0.01$	$0.75\pm 0.01$	$0.73\pm 0.01$	$0.59\pm 0.01$
256 bins	$0.75\pm 0.01$	$0.73\pm 0.01$	$0.71\pm 0.01$	$0.51\pm 0.01$

TABLE V. Effect of image sub-sampling ( $256 \times 256$ ,  $128 \times 128$ ,  $64 \times 64$ ) and the number of histogram bins (64,128,256) on the overall ROC area index of the KB-CAD scheme for decision index  $D_2$ . The full resolution ROIs are  $512 \times 512$ . The reduced size ROIs were created by sub-sampling accordingly the full resolution ROIs.

	512×512	256×256	128×128	64×64
64 bins	0.87±0.01	0.87±0.01	0.87±0.01	0.87±0.01
128 bins	0.87±0.01	0.87±0.01	0.86±0.01	0.84±0.01
256 bins	0.87±0.01	0.86±0.01	0.84±0.01	0.81±0.01

ond, decision index  $D_2$  appears to be more robust to the above effects.  $D_2$  is basically the difference of two terms. If both terms are over- or underestimated, their difference can still reasonably maintain its relative discriminant power. Our experimental results support this hypothesis.

The above-mentioned experiments were performed on a Sun Sparc Ultra-80 workstation with 4 450 MHz processors (Sun Microsystems, Mountain View, CA). Using a single processor, the time requirements to calculate the mutual information between two mammographic ROIs ranged from 0.01 to 0.21 s depending on the ROI size and number of histogram bins selected for the MI estimation. Therefore, the proposed knowledge-based CAD scheme can be easily translated into a real-time CAD system. It takes 2.5 min to compare a mammographic region with 1000 archived cases. The above-mentioned calculation assumes  $512 \times 512$  ROIs and 64 histogram bins. If the comparison is made using sub-sampled ROIs ( $64 \times 64$ ), the CAD response time can be reduced to 10 s per query mammographic ROI.

#### IV. DISCUSSION

In this study, we presented a knowledge-based mass detection scheme for screening mammograms. The proposed CAD scheme is designed to provide a prediction regarding the presence or absence of a mass in a query mammographic region based on similar cases stored in the system's knowledge databank. In its present state, the CAD scheme can function as an interactive tool to help radiologists analyze mammographic regions that attract visual attention. However, the proposed algorithm could be combined with other mass detection schemes for evidence-based reduction of false positive CAD cues. Based on our study, the system was able to maintain 90% sensitivity while effectively eliminating 65% of the normal regions. The performance was consistent for both malignant and benign masses. Since breast masses span a wide range of shapes, sizes, and contrast, the performance of a knowledge-based CAD scheme can be easily compromised if its knowledge databank is limited. Our CAD scheme was developed and evaluated based on a large number of examples from a publicly available database. It has been reported that the database contains really challenging cases.<sup>38</sup> Overall, the estimated performance of our CAD scheme compares favorably with published results from other CAD systems.<sup>14,28</sup> However, direct comparison is not feasible since the results were obtained from different databases. Further studies are needed to evaluate our approach in

other data sets and larger populations of screened women. Furthermore, since the proposed CAD capitalizes on continuously depositing cases in the databank, it is important to assess the impact of the digitization process. The present study was based on DDSM cases digitized with the same digitizer. Studies are under way evaluating how well the CAD system can generalize to other DDSM cases digitized using a different digitizer.

The reported CAD performance was fairly robust regardless of the mass subtlety rating. However, analysis according to breast density showed that CAD performance degrades substantially in dense breasts as it is clinically known. This issue needs investigation due to the lower representation of dense mammograms in the dataset. It is possible that augmenting the knowledge databank with more examples from dense mammograms will improve the CAD performance. Another potentially promising strategy is to design the KB-CAD scheme so that each query ROI is only compared to archived ROIs that were extracted from mammograms with similar density rating as the query ROI. We acknowledge that although indexing the ROIs according to their mammographic density may improve the overall performance of the knowledge-based scheme during the development stage, it may also introduce a serious bias. Observer variability in the reporting of BI-RADS findings is a well-documented issue. Specifically, a study indicated that the overall agreement across observers for the BI-RADS reporting of the mammographic density is only moderate.<sup>39</sup> The same study also showed very poor agreement among observers in use of the category "heterogeneous" breast. Since the DDSM density rating was reported by several different radiologists at various clinical sites, it is expected that any CAD tool developed on the data set will be more fault-tolerant than a CAD tool developed based on cases collected from a single site and read by a single radiologist. However, this issue needs careful investigation.

The main innovation of this study is the application of the mutual information as the similarity metric in a knowledge-based system. MI is a statistical tool that measures to what degree one image can be predicted from another. In image databases, similarity is typically feature-based and often demands substantial image preprocessing. In contrast, the MI between two images is calculated directly without the burden and potential variability of segmentation, object recognition, and feature selection. Therefore, critical CAD issues such as optimized feature selection and merging are bypassed in the proposed KB-CAD system. Considering the difficulty of the mass detection task, the presented concept could be generalizable to other imaging modalities and diagnostic tasks. However, special attention is required when selecting certain implementation parameters. Our study showed that parameters such as the image sub-sampling factor and the number of histogram bins used to estimate the MI affect the overall performance of the detection scheme. For the detection of mammographic masses, if the number of histogram bins is kept reasonably low, then the overall ROC performance of the system remains very robust to image sub-sampling. Continuing research on the formulation of information-theoretic

similarity metrics can be a promising alternative to feature-based CAD techniques.

Finally, an important component of a KB-CAD system is the decision algorithm that combines the similarity level and the truth files of the retrieved cases into a prediction about the query case. Our study showed that using a databank of both mass and normal cases results in a CAD system with a statistically significantly better performance. However, the study also showed that the overall CAD performance varies depending on the number of top matches considered in the calculation of the decision index. Based on the results, the CAD performance was optimal when all stored cases contributed equally in the derivation of the decision index. These findings are attributed to the fact that mutual information is primarily a shape and size-driven similarity measure. Although a mass ROI is expected to match better with other mass ROIs than normal ROIs, the opposite is not true. We will elaborate on that. Theoretically, MI should be able to capture the dissimilarity of two ROIs if one depicts a mass and one depicts just normal tissue since MI is affected by morphology. If, however, both ROIs depict normal parenchyma, then their probability of matching is smaller since the structures and patterns present in normal parenchyma are much more variable. Thus, the MI of two normal ROIs is generally expected to be low. Decision index  $D_2$  capitalizes on this by taking the difference of two terms. If the query ROI contains a mass, then the difference reflects the substantial separation between the morphological properties of masses and normal regions. If the query ROI is normal, then the difference is small since the normal ROI should have low MI with either mass or normal cases. It is documented however, that in image registration MI can produce misleading matches in the presence of noise.<sup>30</sup> Nonetheless, the impact of the noise effect on  $D_2$  should be minimized as more archived cases are considered in the calculation of the decision index.

To summarize, the recent emergence of multimedia digital libraries has increased the interest on comprehensive similarity metrics that can capture effectively the content of images without requiring elaborate image preprocessing. Such metrics can play an important role in knowledge-based CAD systems in an effort to facilitate evidence-based diagnostic interpretation of medical images. Our study showed that mutual information is a promising similarity metric in a knowledge-based CAD scheme for the detection of masses in screening mammograms.

## ACKNOWLEDGMENTS

The authors would like to thank Brian Harrawood, BS for scientific programming. This work was supported in part by U.S. Army Medical Research and Materiel Command Grant No. DAMD 17-02-1-0367.

<sup>30</sup>Electronic mail: gt@deckard.mc.duke.edu

<sup>1</sup>R. T. Greenlee, M. B. Hill-Harmon, T. Murray, and M. Thun, "Cancer statistics, 2001," *Ca-Cancer J. Clin.* **51**, 15–36 (2001).

<sup>2</sup>U.S. D.H.H.S., "Healthy People 2010 (Conference Edition, in Two Volumes)," Washington, DC, 2000.

- <sup>3</sup>L. Tabar *et al.*, "Reduction in mortality from breast cancer after mass screening with mammography," *Lancet* **1**, 829–832 (1985).
- <sup>4</sup>R. E. Bird, T. W. Wallace, and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," *Radiology* **184**, 613–617 (1992).
- <sup>5</sup>R. L. Birdwell, D. M. Ikeda, K. F. O'Shaughnessy, and E. A. Sickles, "Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection," *Radiology* **219**, 192–202 (2001).
- <sup>6</sup>B. C. Yankaskas, M. J. Schell, R. E. Bird, and D. A. Desrochers, "Reassessment of breast cancers missed during routine screening mammography: A community-based study," *AJR, Am. J. Roentgenol.* **177**, 535–541 (2001).
- <sup>7</sup>L. J. W. Burhenne *et al.*, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," *Radiology* **215**, 554–562 (2000).
- <sup>8</sup>T. W. Freer and M. J. Ullissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," *Radiology* **220**, 781–786 (2001).
- <sup>9</sup>N. Petrick, B. Sahiner, H.-P. Chan, M. A. Helvie, S. Paquerault, and L. M. Hadjiiski, "Breast cancer detection: Evaluation of a mass-detection algorithm for computer-aided diagnosis—Experience in 263 patients," *Radiology* **224**, 217–224 (2002).
- <sup>10</sup>C. J. D'Orsi, "Computer-aided detection: There is no free lunch," *Radiology* **221**, 585–586 (2001).
- <sup>11</sup>B. Zheng, M. A. Ganott, C. A. Britton, C. M. Hakim, L. A. Hardesty, T. S. Chang, H. E. Rockette, and D. Gur, "Soft-copy mammographic readings with different computer-assisted detection cuing environments: Preliminary findings," *Radiology* **221**, 633–640 (2001).
- <sup>12</sup>W. Qian, L. H. Li, and L. P. Clarke, "Image feature extraction for mass detection in digital mammography: Influence of wavelet analysis," *Med. Phys.* **26**, 402–408 (1999).
- <sup>13</sup>B. Zheng, Y. H. Chang, X. H. Wang, W. F. Good, and D. Gur, "Feature selection for computerized mass detection in digitized mammograms by using a genetic algorithm," *Acad. Radiol.* **6**, 327–332 (1999).
- <sup>14</sup>W. Qian, L. H. Li, L. P. Clarke, R. A. Clark, and J. Thomas, "Digital mammography: Comparison of adaptive and nonadaptive CAD methods for mass detection," *Acad. Radiol.* **6**, 471–480 (1999).
- <sup>15</sup>H. P. Chan *et al.*, "Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces," *Med. Phys.* **25**, 2007–2019 (1998).
- <sup>16</sup>H. P. Chan *et al.*, "Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: An ROC study," *Radiology* **212**, 817–827 (1999).
- <sup>17</sup>D. L. Thiele, C. Kimme-Smith, T. D. Johnson, M. McCombs, and L. W. Bassett, "Using tissue texture surrounding calcification clusters to predict benign vs malignant outcomes," *Med. Phys.* **23**, 549–545 (1996).
- <sup>18</sup>Z. Huo, M. L. Giger, C. J. Vybotny, D. E. Wolverton, R. A. Schmidt, and K. Doi, "Automated computerized classification of malignant and benign masses on digitized mammograms," *Acad. Radiol.* **5**, 155–168 (1998).
- <sup>19</sup>F. Schmidt *et al.*, "An automatic method for the identification and interpretation of clustered microcalcifications in mammograms," *Phys. Med. Biol.* **44**, 1231–1243 (1999).
- <sup>20</sup>M. A. Gavrielides, J. Y. Lo, R. Vargas-Voracek, and C. E. Floyd, "Segmentation of suspicious clustered microcalcifications in mammograms," *Med. Phys.* **27**, 13–22 (2000).
- <sup>21</sup>S. Y. Yu and L. Guan, "A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films," *IEEE Trans. Med. Imaging* **19**, 115–126 (2000).
- <sup>22</sup>W. Qian, X. J. Sun, D. S. Song, and R. A. Clark, "Digital mammography: Wavelet transform and Kalman-filtering neural network in mass segmentation and detection," *Acad. Radiol.* **8**, 1074–1082 (2001).
- <sup>23</sup>S. Paquerault, N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Improvement of computerized mass detection on mammograms: Fusion of two-view information," *Med. Phys.* **29**, 238–247 (2002).
- <sup>24</sup>J. J. Heine, S. R. Deans, R. P. Velthuisen, and L. P. Clarke, "On the statistical nature of mammograms," *Med. Phys.* **26**, 2254–2265 (1999).
- <sup>25</sup>J. J. Heine and R. P. Velthuisen, "A statistical methodology for mammographic density detection," *Med. Phys.* **27**, 2644–2651 (2000).
- <sup>26</sup>C. E. Floyd, Jr., J. Y. Lo, and G. D. Tourassi, "Breast biopsy: Case-based reasoning computer-aid using mammography findings for the breast biopsy decisions," *AJR, Am. J. Roentgenol.* **175**, 1347–1352 (2000).
- <sup>27</sup>A. O. Bilaska-Wolak and C. E. Floyd, Jr., "Development and evaluation of

- a case-based reasoning classifier for prediction of breast biopsy outcome with BI-RADSTM lexicon," *Med. Phys.* **29**, 2090–2100 (2002).
- <sup>28</sup>Y.-H. Chang, L. A. Hardesty, C. M. Hakim, T. S. Chang, B. Zheng, W. F. Good, and D. Gur, "Knowledge-based computer-aided mass detection on digitized mammograms: A preliminary assessment," *Med. Phys.* **28**, 455–461 (2001).
- <sup>29</sup>T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
- <sup>30</sup>J. V. Hajnal, D. L. G. Hill, and D. J. Hawkes, *Medical Image Registration* (CRC, Boca Raton, FL, 2000).
- <sup>31</sup>M. Heath *et al.*, "Current status of the digital database for screening mammography," in *Digital Mammography* (Kluwer Academic, Dordrecht, 1998).
- <sup>32</sup>W. Li, "Mutual information functions versus correlation functions," *J. Stat. Phys.* **60**, 823–837 (1990).
- <sup>33</sup>W. M. Wells, P. V. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Med. Image Anal* **1**, 35–51 (1996).
- <sup>34</sup>F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodal image registration by maximization of mutual information," *IEEE Trans. Med. Imaging* **16**, 187–198 (1997).
- <sup>35</sup>A. Treves and A. Panzeri, "The upward bias in measures of information derived from limited data samples," *Neural Comput.* **7**, 399–407 (1995).
- <sup>36</sup>G. D. Tourassi, E. D. Frederick, M. K. Markey, and C. E. Floyd, Jr., "Application of the mutual information criterion for feature selection in computer-aided diagnosis," *Med. Phys.* **28**, 2394–2402 (2001).
- <sup>37</sup>B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, *Monographs on Statistics and Applied Probability*, edited by D. R. Cox *et al.* (Chapman & Hall, New York, 1993).
- <sup>38</sup>G. M. teBrake, N. Karssemeijer, and J. H. Hendricks, "An automatic method to discriminate malignant masses from normal tissue in digital mammograms," *Phys. Med. Biol.* **45**, 2843–2857 (2000).
- <sup>39</sup>W. A. Berg, C. Campassi, P. Langenberg, and M. J. Sexton, "Breast imaging reporting and data system: Inter- and intraobserver variability in feature analysis and final assessment," *AJR, Am. J. Roentgenol.* **174**, 1769–1777 (2000).