Published in final edited form as:

Acad Radiol. 2008 May ; 15(5): 626–634.

Evaluating the Effect of Image Preprocessing on an Information-Theoretic CAD System in Mammography

Georgia D. Tourassi, PhD¹, Robert Ike III, BS^{1,2}, Swatee Singh, BS^{1,3}, and Brian Harrawood¹

¹ Digital Advanced Imaging Laboratories, Department of Radiology, Duke University Medical Center, Durham, NC 27705

² Medical Physics Program, Duke University Medical Center, Durham, NC 27705

³ Department of Biomedical Engineering, Duke University, Durham, NC 27710

Abstract

Rationale and Objectives—In our earlier studies we reported an evidence-based Computer Assisted Decision (CAD) system for location-specific interrogation of mammograms. A content-based image retrieval framework with information theoretic (IT) similarity measures serves as the foundation for this system. Specifically, the normalized mutual information (NMI) was shown to be the most effective similarity measure for reduction of false positive marks generated by other, prescreening mass detection schemes. The objective of this work was to investigate the importance of image filtering as a possible preprocessing step in our IT-CAD system.

Materials and Methods—Different filters were applied, each one aiming to compensate for known limitations of the NMI similarity measure. The study was based on a region-of-interest database that included true masses and false positive regions from digitized mammograms.

Results—Receiver Operating Characteristics (ROC) analysis showed that IT-CAD is affected slightly by image filtering. Modest, yet statistically significant performance gain was observed with median filtering (overall ROC area index A_z improved from 0.78 to 0.82). However, Gabor filtering improved performance for the high sensitivity portion of the ROC curve where a typical false positive reduction scheme should operate (partial ROC area index $_{0.90}A_z$ improved from 0.33 to 0.37). Fusion of IT-CAD decisions from different filtering schemes markedly improved performance ($A_z=0.90$ and $_{0.90}A_z=0.55$). At 95% sensitivity, the system's specificity improved by 36.6%.

Conclusion—Additional improvement in false positive reduction can be achieved by incorporating image filtering as a preprocessing step in our information-theoretic CAD system.

Keywords

CAD; mammography; image processing; information theory

Corresponding Author: Georgia D. Tourassi, PhD, Duke Advanced Imaging Laboratories, Duke University Medical Center, 2424 Erwin Road, Hock Plaza, Suite 302, Durham, NC 27710, Telephone: (919) 684-1447, FAX: (919) 684-1491, E-mail: georgia.tourassi@duke.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

INTRODUCTION

Despite advances in treatment, breast cancer remains the second leading cause of cancer death in women [1]. The role of screening mammography in the battle against breast cancer is well established; women with malignancies detected at an early stage have a significantly better prognosis [2]. However, it is also recognized that the diagnostic interpretation of mammograms continues to be challenging for radiologists with a documented 20% false negative rate [3–6]. The clinical significance of early breast cancer diagnosis and the higher than desired false negative rate of screening mammography have motivated the development of Computer Aided Detection (CADe) systems for decision support. These systems typically involve a hierarchical approach, first applying elaborate image preprocessing steps to enhance suspicious structures in the image and then employing morphological and textural analysis to better classify these structures between true abnormalities and false positives. Detailed reviews of image processing techniques for mammographic image analysis and related CADe systems can be found elsewhere [7–10]. In addition, several CADe systems are already available commercially for both screen film mammography as well as full field digital mammography [7]. Although their true clinical impact is often debated [e.g., 11–19], the scientific community continues to work towards improving the diagnostic performance and clinical integration of CADe technology. Ongoing CADe research efforts focus mainly on the reduction of false positive computer marks as well as improving the detection rate of breast masses.

In our earlier studies, we presented a knowledge-based CADe system for breast mass detection in screening mammograms [20–22]. The system is interactive and it is designed to operate as a second opinion for mammographic locations that are deemed suspicious of containing breast masses. These suspicious locations are areas that attract the radiologist's attention or they are marked as suspicious by other automated mass detection schemes. Thus, our system is designed for location-specific interrogation of mammograms. The interrogation relies on a database of mass and normal examples with known ground truth. These examples serve as the knowledge database. Basically, the system compares the query location with the knowledge examples. The comparison is performed using featureless, information-theoretic (IT) similarity measures [21]. Such measures are based on the concept of image entropy [23] and they are calculated directly from the image pixel intensity values. Although we explored various IT measures, our IT-CADe system using either mutual information (MI) or its normalized version (NMI) was shown to be the most effective [21,22].

The original IT-CADe prototype relied on raw image data without any preprocessing. Our present study reports on the effect of image preprocessing on the overall diagnostic performance of this system. Medical image registration studies using mutual information suggest that minimal preprocessing often improves image registration performance [24]. Consequently, in this study we explored the effect of various preprocessing image filters on our own IT-CADe system. The selection of each preprocessing filter targeted known limitations of the mutual information similarity measure. The resultant performance of the modified IT-CADe system was compared to that previously reported without the preprocessing filtering step. Such direct comparison is necessary to test the hypothesis that image preprocessing contributes to further improvement of the IT-CADe performance.

MATERIALS AND METHODS

A. Materials

Database—The image database used in this study has been described before in detail [20, 21]. Since the present study builds upon a system presented before, it is essential to demonstrate any incremental improvement using the same database. Here is a summary description of this database.

All mammographic cases were selected from the Digital Database of Screening Mammography (DDSM) [25]. These mammograms were scanned with a Lumisys scanner at 50 $\mu\text{m}/\text{pixel}$ and a bit depth of 12. There were 583 mammograms in total; 296 containing biopsy-proven malignant masses, 185 containing benign masses proven either by biopsy or additional imaging, and 82 normal mammograms. The available database was divided into 2 sets. One set contained 483 DDSM cases (256 cancer, 145 benign, 62 normal) and served as the knowledge database. The second set contained the remaining 100 cases (40 malignant, 40 benign, 20 normal) and served as the test database. Note that the test database was reserved from the beginning of our IT-CADe research efforts (prior to this study) to serve for final validation. The selection criteria were such that the test database represents a balanced mix of cases from all available DDSM/Lumisys volumes. The database did not contain any “benign-without-callback” cases, since these are considered easy cases to diagnose.

From each case, a 512×512 pixel ROI was extracted around the known location of any true mass present in the case. The mass locations are provided in the DDSM truth files. Dataset 1 (i.e., the knowledge database) contained 1,820 ROIs in total. Of those, 489 depicted a malignant mass, 412 depicted a benign mass, and 919 were normal. The mass ROIs serve as the system’s knowledge foundation of typical mass examples. The normal ROIs were initially selected from normal mammographic cases by randomly sampling the breast region. Such normal ROIs are essential to establish a knowledge foundation of normal breast parenchyma. Since the normal cases were few compared to the mass cases in the DDSM/Lumisys set, normal ROIs were also extracted from abnormal cases but only from imaged breasts that did not contain any physician annotations in either mammographic view.

In addition, 512×512 pixel ROIs were extracted around the known mass locations in the test database. There were 44 malignant mass ROIs and 40 benign mass ROIs in Dataset 2. In addition, 399 ROIs were extracted around mammographic locations marked as suspicious by a feature-based CADe system developed before in our laboratory [26,27]. Therefore, dataset 2 contained 483 ROIs in total. These ROIs served as queries to our IT-CADe system to determine whether the system can provide effective false positive reduction.

Overview of the IT-CADe System—The prototype IT-CADe system offers an evidence-based, second opinion regarding the presence of a possible mass in any mammographic location that is indicated by the CADe user. The basic IT-CADe system combines principles from content-based image retrieval and case-based reasoning. When an unknown query case is presented for evaluation to the system, the system compares the query to all known cases stored in the knowledge database. Similar cases are retrieved and they are used to make a prediction regarding the query case. The retrieval process relies on information theoretic measures. Such measures include mutual information, joint entropy, Kullback-Leibler divergence etc. [23]. Generally, these similarity measures are calculated using the image pixel intensity values directly, not any image features. The underlying assumption is that the co-occurrence of the intensity values in the two images is maximized when the images match well. The information-theoretic measures use the concept of entropy to measure the co-occurrence of pixel values [23].

Our previous publications [20,21,22,28] addressed issues related to the composition of the knowledge database, the case retrieval process, the construction of the decision index, as well as the effect of the similarity metric. Based on our previous findings, the prototype system operates as follows. First, a 512×512 pixel ROI is extracted around the suspicious mammographic location indicated by the CADe user or marked by another detection algorithm. The ROI serves as the query case for the system. Then, the ROI is compared to all examples (or templates) stored in the system’s knowledge database. Examples with similar entropy as that of the query are quickly identified using an entropy-based indexing scheme we presented

before [22]. The entropy-based indexing scheme serves as a search mechanism to sort through the knowledge database fast and identify the stored examples that are more relevant to the specific query. Then, detailed pairwise comparisons are performed between the query Q and each relevant knowledge example (or template T). This detailed comparison is based on the normalized mutual information (NMI) similarity measure [22]. NMI captures the statistical dependence between two images and it is always bounded between 0 and 1. A value of 1 suggests perfect match between the query case Q and the stored template T . In contrast, a value of 0 indicates no statistical dependence or shared information between the two cases. Some studies in image registration have shown that NMI is often more successful and robust than MI [24,29,30]. Although our previous studies showed that both MI and NMI are equally effective in IT-CADe [21], the bounded nature of NMI makes it a more attractive option. Finally, a decision index is calculated measuring how well the query case matches on average the retrieved mass templates compared to the retrieved normal templates. In a clinical setting, an optimal threshold needs to be determined for the final decision. If the query's decision index exceeds the threshold value, then the query mammographic location is marked as a true mass. Otherwise, the query location is marked as normal.

B. Methods

Preprocessing Filters—To test the effect of image preprocessing on the system performance, we applied several different filters. These filters were selected to compensate for potential limitations of the NMI similarity measure such as lower robustness in the presence of noise, lack of spatial information, and questionable perceptual relevance. Specifically, five different filters were investigated. Two of them were popular denoising filters, namely the median and the adaptive Wiener filter. The third choice was a perceptually driven Gabor filter. Finally, two texture filters were also investigated, an entropy-based and a localized standard deviation filter. Image filtering was performed using the MATLAB programming environment (The MathWorks, Inc., Natick, MA).

(a) Median filter: Several image registrations studies suggested that noise reduction techniques are essential for more accurate MI-based image registration [24,31]. We have explored the same issue for our IT-CADe system. Specifically, we applied median filtering before the calculation of the similarity measures. Median filtering is a standard noise reduction technique. Furthermore, it is a reasonable preprocessing step for mass detection since it preserves the edge information of suspicious areas while reducing noise [32]. Median filters with different size kernels were explored (3×3 , 5×5 , 7×7 , 9×9 , 11×11 , 15×15 , 21×21 pixels) for the task.

(b) Adaptive Wiener filter: Similar to the median filter, the adaptive Wiener filter [33] was applied for denoising tailored on statistics estimated from the local neighborhood of each image pixel. The amount of smoothing performed by the filter depends on the local image mean and variance around the pixel of interest. The Wiener filter is a popular linear filter but its adaptive implementation preserves better the high frequency parts of the image. The same size kernel sizes were explored as with the median filter.

(c) Gabor filter: Another promising filter for image denoising and texture analysis is the Gabor filter [34]. This type of multi-channel filtering is considered an excellent preprocessing choice for image registration [35–37] due to its perceptual relevance [38]. Specifically, it has been shown that Gabor filters model the spatial frequency and orientation responses of simple cells in the primary visual cortex [39,40]. The Gabor representation has been shown to be optimal in the sense of minimizing the joint two-dimensional uncertainty in space and frequency [41]. Since the Gabor filter bank is derived from a wavelet basis with dilations and orientations, they are essentially bandpass filters.

A 2D symmetric Gabor filter was implemented as described in Eq. 1:

$$f(x, y) = e^{\left\{-\frac{1}{2}\left[\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right]\right\}} \cdot \cos(2\pi\mu_0(x\cos\theta + y\sin\theta)) \quad [1]$$

where μ_0 is the frequency of a sinusoidal plane, θ is the orientation, and σ_x and σ_y are standard deviations (or spatial spread) of the 2-D Gaussian envelope [42]. An octave bandwidth of 1 was used in our study as psychophysical studies in the past have confirmed that an octave bandwidth of 1 is a reasonably good estimate of the human eye when tuned to a frequency [43]. Central frequencies of 0.5, 1, 2, 4, 8, 16 and 32-cycles/degree with orientations at 0°, 45°, 90° and 135° were used in this study.

(d) Entropy-based filtering: Since NMI is calculated using only intensity information of corresponding pixels between two images, it has an inherent limitation. It ignores possible relationships between neighboring pixels. Since image texture is typically captured by such relationships, NMI ignores a potentially critical diagnostic component. To address this limitation, we investigated an entropy-based filter as a preprocessing step for all images. The filter replaces the intensity value of each image pixel with a new value that captures the local image entropy around the pixel [44]. Thus, each pixel value is replaced with a new value that contains localized textural information. This filter was implemented using the *entropyfilt* function in the MATLAB Image Processing Toolbox. Multiscale analysis was investigated repeating this filtering step at several scales by varying the neighborhood size of the entropy-based filter (3×3, 5×5, 7×7, 9×9, 11×11, 15×15, and 21×21 pixels).

(e) Standard deviation filter: Similar to the entropy-based filter, the standard deviation (STD) filter replaces each pixel value of the grayscale image with the local standard deviation of a neighborhood around the pixel of interest. This preprocessing filter was implemented using the *stdfilt* function of the MATLAB Image Processing Toolbox and it was also evaluated for variable size neighborhoods as the entropy-based filter.

Figure 1 shows a representative, unprocessed ROI (a) depicting a malignant mass along with its filtered versions using the following filters: (b) median, (c) locally adaptive Wiener, (d) Gabor, (e) entropy-based, and (f) standard deviation-based.

Evaluation Methods—Both Datasets 1 and 2 were preprocessed using the previously described filters. For each filter separately, the IT-CADe system was tested using Dataset 1 as the knowledge database and Dataset 2 as the test bed for the discrimination of true masses from false positive findings. Detection performance was evaluated with Receiver Operating Characteristic (ROC) Analysis [45]. ROC curves were fitted with the ROCKIT software, available by Charles Metz at the University of Chicago. The overall ROC area A_z and the partial ROC area $_{0.90}A_z$ were used as the reported performance indices. Although A_z is the most common performance index for binary diagnostic tasks taking into account all possible decision thresholds, the partial ROC area index summarizes the detection performance for decision thresholds corresponding only to the high sensitivity portion (>90%) [46]. For our study, $_{0.90}A_z$ is certainly a more appropriate performance index since any false-positive reduction scheme is expected to perform at a high cancer detection rate for a clinically effective cancer-screening CADe system.

RESULTS

A. Effect of Image Filter

First, the effect of the kernel size on IT-CADe performance was investigated carefully for the median, Wiener, entropy-based, and standard deviation-based filters. Figure 2 shows the

corresponding ROC area index A_z (Fig. 2a) and partial ROC area index $_{0.90}A_z$ (Fig. 2b) for all kernel sizes considered. For the median filter, the 3×3 kernel resulted in the highest ROC performance with the 5×5 kernel producing a slightly (yet not statistically significantly) lower performance. As the kernel size of the median filter increased, the performance of the system steadily decreased. This result is expected due to the resulting oversmoothing of the images. The kernel size of the Wiener filter had minimal impact on the system's performance, at least for the size range evaluated in this study. Similar to the median filter, the neighborhood size of the texture filters also affected system performance. Performance peaked for the 9×9 neighborhood size with the entropy-based filter. The improvement was statistically significant compared to all other neighborhood sizes with the exception of the 11×11 neighborhood where the difference was borderline significant (2-tailed p-value=0.05 for the partial ROC area index). For the standard-deviation filter, ROC performance peaked for the 21×21 neighborhood size but it was not significantly better compared to the other neighborhood sizes. Further increase of the neighborhood size resulted in severe drop of the system's performance.

Table 1 summarizes the results of this study for all preprocessing scenarios considered. The table shows the performance indices for the IT-CADe system depending on the image preprocessing scheme. For simplicity, Table 1 shows only the system performance for each preprocessing filter operating with its best performing kernel size. As a point of reference, the Table also includes the performance of the original IT-CADe system without any image preprocessing ("none"). In addition, the table reports the specificity achieved by the system at 95% detection rate for masses.

The table highlights several interesting trends. Overall, both texture filters resulted in dramatic decline of the IT-CADe diagnostic performance with respect to all performance indices. The decline of the ROC area index was significant for both the entropy-based and standard deviation-based filters (2-tailed p-value<0.001). With respect to the partial area index, the decline was borderline significant for the standard deviation filter (2-tailed p-value=0.05) but not significant for the entropy filter (2-tailed p-value=0.12). These results suggest that the texture filters investigated in this study are not appropriate choices, if they are to be used as an independent preprocessing step. However, since they capture textural information, such filters have potentially incremental diagnostic value.

With respect to the overall ROC area index, the median filter resulted in a statistically significant improvement of the diagnostic performance. The area index increased from 0.78 to 0.82 (2-tailed p-value=0.01). However, such improvement was not observed with respect to the partial ROC area index. Actually, the partial ROC area index declined slightly from 0.33 to 0.32 after median filtering (2-tailed p-value=0.20). The Wiener filter resulted in similar performance of the IT-CADe system as without any preprocessing (2-tailed p-values of 0.53 and 0.29 for the ROC and partial ROC area indices respectively). Finally, Gabor filtering produced a notable improvement of the partial ROC area index (from 0.33 to 0.37). However this improvement did not reach statistical significance (2-tailed p-value=0.12).

With respect to specificity at 95% mass detection rate, Gabor filtering was the most effective. The system achieved 34.1% specificity when including Gabor filtering as a preprocessing step compared to 31.3% specificity without filtering. This result represents a 9% improvement in system specificity.

B. IT-CADe Fusion

Although no filter emerged as a clearly superior choice, it is possible that a multi-filter fusion approach may be more effective. To test this possibility, we constructed a linear classifier that combined the predictions of the IT-CADe systems (each operating with a different preprocessing step) into one comprehensive decision. The underlying hypothesis is that fusing

the IT-CADe outputs based on multiple, complementary preprocessing filters may be superior to any one of the filters alone.

Specifically, linear classifiers were built combining the filter-specific IT-CADe outputs. For a given query ROI, the continuous decision indices of the filter-specific IT-CADe systems served as inputs to the fusion classifier. Thus, the fusion CAD system relied on stacked generalization where the level 0 classifiers were the knowledge-based, filter-specific IT-CADe systems and the level 1 combiner was a trainable linear classifier. We performed an exhaustive search, building a linear classifier for every possible combination of “filters” (i.e., filtered IT-CADe outputs). With 6 different filters considered, there were 57 possible combinations (i.e., 15 combinations merging the IT-CADe outputs of only 2 different filters at a time, 20 combinations merging 3 different filters, 15 combinations merging 4 different filters, 6 combinations merging 5 filters, and 1 combination including all 6 filtered IT-CADe outputs). Thus, 57 different linear classifiers were built. These classifiers were evaluated using leave-one out sampling on Dataset 2 since the clinical focus is on differentiating masses from false positives. Furthermore, in our previous experiments we observed that leave-one-out is an appropriate data handling scheme when stacking knowledge-based IT-CADe systems with a simple combiner such as a linear classifier [47]. The above experiments were performed using the R software package [48,49].

Table 2 highlights some of the most interesting trends of the IT-CADe fusion experiment. Specifically, the table shows which combination produced the best performing fusion classifier when the number of input filters is restricted (only 2 filters, only 3 filters, etc.). Overall, the fusion experiment showed that the synergistic approach of the linear classifiers using information from IT-CADe with different preprocessing schemes resulted in statistically significantly better performance compared to the original IT-CADe system with respect to all performance metrics. Fusing all 6 filtering schemes produced the best results ($A_z=0.90$, $_{0.90}A_z=0.55$). At a fixed 95% mass sensitivity rate, seventy-six additional false positive queries were correctly identified by the fusion system. This represents a 61% specificity improvement over the original IT-CADe system. However, combining the decisions of the IT-CADe system operating with only two preprocessing steps, namely the median and the adaptive Wiener filters, produced similar results ($A_z=0.88$, $_{0.90}A_z=0.52$, and 57.8% specificity improvement at a fixed 95% mass detection rate). This performance was not significantly lower than the best reported one using all 6 filters. Compared to the best performing median filter, the 2-filter fusion LDA increased specificity from 31.3% to 49.4%. Similarly, a 44.9% specificity improvement was observed over the system operating with the best performing Gabor filter (specificity increased from 34.1% to 49.4%). Although several combinations of 3, 4, and 5 filters resulted in incremental performance improvements (shown in Table 2), none of these were statistically significant compared to combining just the median and Wiener filters. This result suggests that the added complexity of additional filters is not justified.

DISCUSSION

The general concept of building and mining knowledge databases of imaging data in radiology is becoming increasingly relevant. In the digital era, it is important to capitalize on the growing number and variety of mammograms that are continuously acquired and stored. Our interactive, knowledge-based CADe system for location-specific interrogation of mammograms has such capacity without needing additional training of its decision-making module every time a new case is added in its knowledge database. Furthermore, our featureless approach for case similarity assessment eliminates any concerns regarding careful selection, extraction, and merging of image features for decision making. This is a particularly attractive property that facilitates easier knowledge transfer across databases (e.g., mammograms acquired with

different systems or digitized with different digitizers) as we have shown in previous studies [50,51].

In this study, we presented a range of image filtering techniques as potential preprocessing steps in an attempt to improve the performance of our IT-CADe system. The filters were selected so that they complemented the similarity metric in our IT-CADe system. Since normalized mutual information is sensitive to image noise, a smoothing median filter and an adaptive Wiener filter were considered as promising preprocessing steps. In addition, two texture-based filters were considered. Because NMI does not capture localized textural information, an entropy-based filter and a standard deviation-based filter were applied to quantify the local texture in the image. The final choice was a Gabor filter optimized according to the human perception system. This comparative study focused on the false positive reduction task since such task still represents one of the major challenges of existing CADe systems in mammography.

Our study was restricted to a small but diverse group of preprocessing filters. Overall, no particular filter emerged as the superior choice. Although median filtering resulted in significantly better performance with respect to the overall ROC area, Gabor filtering demonstrated superior performance for the clinically critical, high sensitivity portion of the ROC curve. However, the improvement did not reach statistical significance. The entropy-based and standard deviation-based texture filters were the only ones that deteriorated the diagnostic performance of the IT-CADe system. It should be noted that other texture-based filters that are better tailored to the clinical task could be potentially more successful. Finally, integrating all filters with a linear classifier achieved dramatic improvement with respect to all performance indices. These results highlight the significance of image preprocessing for our information-theoretic CADe system, especially when a fusion approach is considered where the filters are complementary in nature.

In conclusion, image preprocessing through carefully tailored filters should be investigated as a promising strategy to improve substantially upon the performance of our CADe system. Moreover, advanced fusion strategies that incorporate CADe decisions based on complimentary preprocessing steps hold most promise for providing even further improvements.

Acknowledgements

This work was supported by grant R01 CA101911 from the National Cancer Institute.

References

1. ACS, American Cancer Society. Cancer Facts and Figures 2002. Atlanta, GA: American Cancer Society; 2002.
2. Tabar L, Vitak B, Chen HH, Yen MF, Duffy SW, Smith RA. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. *Cancer* 2002;91:1724–1731. [PubMed: 11335897]
3. Beam CA, Conant EF, Sickles EA. Factors affecting radiologist inconsistency in screening mammography. *Acad Radiol* 2002;9:531–540. [PubMed: 12458879]
4. Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. *Radiol* 1992;184:613–617.
5. Birdwell RL, Ikeda DM, O'Shaughnessy KF, Sickles EA. Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. *Radiol* 2001;219:192–202.
6. Yankaskas BC, Schell MJ, Bird RE, Desrochers DA. Reassessment of breast cancers missed during routine screening mammography: A community-based study. *Am J Roentg (AJR)* 2001;177:535–541.

7. Sampat, MP.; Markey, MK.; Bovik, AC. Handbook of Image and Video Processing. 2. 2005. Computer-Aided Detection and Diagnosis in Mammography; p. 1195-1217.
8. Thangavel K, Karnan M, Sivakumar R, Kaja Mohideen A. Automatic Detection of Microcalcification in Mammograms– A Review. ICGST-GVIP Journal 2005;5(5):31–61.
9. Highnam, RP.; Brady, JM. Mammographic Image Analysis. Kluwer Academic Publishers; 1999.
10. Fitzpatrick, JM.; Sonka, M. Medical Image Processing and Analysis. 2. SPIE Press; 2000. Handbook of Medical Imaging.
11. Burhenne LJW, Wood SA, D’Orsi CJ, Feig SA, Kopans DB, et al. Potential contribution of computer-aided detection to the sensitivity of screening mammography. Radiol 2000;215:554–562.
12. Brem RF, Baum J, Lechner M, et al. Improvement in sensitivity of screening mammography with computer-aided detection: A multiinstitutional trial. Am J Roentg (AJR) 2003;181:687–693.
13. Gur D, Sumkin H, Rockette HE, Ganott M, Hakim C, Hardesty LA, et al. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. J National Cancer Institute 2004;96:185–190.
14. Taylor P, Given-Wilson RM. Evaluation of computer-aided detection (CAD) devices. Brit J Radiol 2005;78:26–30.
15. Birdwell RL, Bandodkar P, Ikeda DM. Computer-aided Detection with Screening Mammography in a University Hospital Setting. Radiol 2005;236:451–457.
16. Morton MJ, Whaley DH, Brandt KR, Amrami KK. Screening mammograms: Interpretation with computer-aided detection – Prospective evaluation. Radiol 2006;239:375–383.
17. Krupinski EA. Computer-aided detection in clinical environment: benefits and challenges for radiologists. Radiol 2004;231:7–9.
18. Fenton JJ, Taplin SH, Carney PA, Abraham L. Influence of Computer-Aided Detection on Performance of Screening Mammography. N Engl J Med (NEJM) 2007;356:1399–1409.
19. Hall FM. Breast Imaging and Computer-Aided Detection. N Engl J Med (NEJM) 2007;356:1464–1466.
20. Tourassi GD, Vargis-Voracek R, Floyd CE Jr. Computer-Assisted Detection of Mammographic Masses: A Template Matching Scheme based on Mutual Information. Med Phys 2003;30:2123–2139. [PubMed: 12945977]
21. Tourassi GD, Harrawood B, Singh S, Lo JY, Floyd CE Jr. Evaluation of information-theoretic similarity measures for content-based retrieval and detection of masses in mammograms. Med Phys 2007;34:140–150. [PubMed: 17278499]
22. Tourassi GD, Harrawood B, Singh S, Lo JY. Information-theoretic CAD system in mammography: Entropy-based indexing for computational efficiency and robust performance. Med Phys 2007;34:3193–3204. [PubMed: 17879782]
23. Cover, TM.; Thomas, JA. Elements of Information Theory. John Wiley & Sons; New York, NY: 1991.
24. Pluim JPW, Maintz JBA, Viergever MA. Mutual-information-based registration of medical images: a survey. IEEE Trans Med Imag 2003;22:986–1004.
25. Heath, M.; Bowyer, K.; Kopans, D., et al. Digital Mammography. Kluwer Academic Publishers; 1998. Current Status of the Digital Database for Screening Mammography. Available: <http://marathon.csee.usf.edu/Mammography/Database.html>
26. Catarious DM, Baydush AH, Floyd CE Jr. A mammographic mass CAD system incorporating features from shape, fractal, and channelized Hotelling observer measurements: preliminary results. SPIE 2003;5032:111–119.
27. Catarious DM, Baydush AH, Floyd CE Jr. Incorporation of an iterative, linear segmentation routine into a mammographic mass CAD system. Med Phys 2004;31:1512–1520. [PubMed: 15259655]
28. Tourassi GD, Harrawood B, Floyd CE Jr. Cross-Digitizer Robustness of a Knowledge-Based CAD System for Mass Detection in Screening Mammograms. SPIE 2007;6514:65141Y1–8.
29. Hajnal, JV.; Hill, DLG.; Hawkes, DJ. Medical Image Registration. CRC Press; Boca Raton, FL: 2000.
30. Studholme C, Hill DLG, Hawkes DJ. An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognition 1996;32:71–86.

31. Dekker N, Ploeger LS, van Herk M. Evaluation of cost functions for gray value matching of two-dimensional images in radiotherapy. *Med Phys* 2003;30:778–784. [PubMed: 12772984]
32. Bovik AC, Huang S, Manson DC. The effect of median filtering on edge estimation and detection. *IEEE Trans Pattern Anal Machine Intell (PAMI)* 1987;9:181–194.
33. Mayo P, Rodenas F, Verdu G. Comparing Methods to Denoise Mammographic Images. *IEEE EMBS* 2004;1:247–250.
34. Gabor D. Theory of communication. *J Inst Elec Eng* 1946;93:429–459.
35. Zheng Q, Chellappa R. A computational vision approach to image registration. *IEEE Trans Imag Processing* 1993;2:311–326.
36. Manjunath BS, Shekhar C, Chellappa R. A New Approach to Image Feature Detection with Applications. *Pattern Recognition* 1996;29:627–640.
37. Liu J, Vemuri BC, Marroquin JL. Local frequency representations for robust multimodal imageregistration. *IEEE Trans Med Imag* 2002;21:462–469.
38. Rubner, Y.; Tomasi, C. *Perceptual Metrics for Image Database Navigation*. Springer; 2001.
39. Hubel DH, Weisel TN. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J Neurophysiol* 1965;28:229–289. [PubMed: 14283058]
40. Campbell FW, Robson JG. Application of Fourier analysis to the visibility of gratings. *J Physiol* 1968;197:551–566. [PubMed: 5666169]
41. Daugman JG. Complete Discrete 2D Gabor Transforms by Neural Networks for Image Analysis and Compression. *IEEE Trans ASSP* 1998;36:1169–1179.
42. Chen CC, Chen CC. Gabor transform in texture analysis. *SPIE Proc* 2003;2353:237–245.
43. Watson, AB.; Braddick, OJ.; Sleigh, AC. *Physical and Biological Processing of Images*. Springer-Verlag; Berlin: 1983. Detection and recognition of simple spatial forms.
44. Gonzalez, RC.; Woods, RE.; Eddins, SL. *Digital Image Processing Using MATLAB*. Chapter 11. New Jersey: Prentice Hall; 2003.
45. Obuchowski NA. Receiver Operating Characteristic curves and their use in radiology. *Radiol* 2003;229:3–8.
46. Jiang Y, Metz CE, Nishikawa RM. A receiver operating characteristic partial area index for highly sensitive diagnostic tests. *Radiol* 1996;201:745–750.
47. Tourassi GD, Jesneck JA, Mazurowski ML, Habas PA. Stacked Generalization in Computer Assisted Decision Systems: Empirical Comparison of Data Handling Schemes. *Proceedings of International Joint Conference on Neural Networks (IJCNN)* 2007:1343–1347.
48. R Development Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2003.
49. The R project for statistical computing (<http://www.R-project.org>)
50. Tourassi GD, Harrawood B, Floyd CE Jr. Cross-Digitizer Robustness of a Knowledge-Based CAD System for Mass Detection in Screening Mammograms. *SPIE* 2007;6514:65141Y1–8.
51. Tourassi, GD.; Saunders, R.; Samei, E. Mass Detection in Full Field Digital Mammograms: Validation of an Information-Theoretic Knowledge-Based System. *RSNA 92nd Scientific Assembly*; Chicago, IL: 2006.

Figure 1. Example ROI depicting a malignant mass (a) unprocessed, and processed with the following filters: (b) 3×3 median, (c) 3×3 adaptive Wiener, (d) Gabor, (e) 9×9 entropy-based, and (f) 21×21 standard deviation-based.

(a)

(b)

Figure 2.

Effect of the filter kernel size on the ROC performance of the IT-CADe system with respect to the (a) overall ROC area index and the (b) partial ROC area index for the high sensitivity (>90%) portion of the ROC curve.

Table 1

Effect of image filtering as a preprocessing step on the performance of the IT-CADe system for the detection of masses in screening mammograms

Preprocessing Filter	$A_z (\pm\sigma)$	$0.90 A_z (\pm\sigma)$	Specificity at 95% sensitivity
None	0.778 $\pm$ 0.025	0.326 $\pm$ 0.055	31.3% (125/399)
Median (3 $\times$ 3)	0.816 $\pm$ 0.025	0.320 $\pm$ 0.065	29.6% (118/399)
Wiener (3 $\times$ 3)	0.785 $\pm$ 0.026	0.323 $\pm$ 0.057	31.1% (124/399)
Gabor	0.783 $\pm$ 0.024	0.368 $\pm$ 0.053	34.1% (136/399)
Entropy (9 $\times$ 9)	0.706 $\pm$ 0.028	0.268 $\pm$ 0.046	27.6% (110/399)
Standard Deviation (21 $\times$ 21)	0.667 $\pm$ 0.028	0.236 $\pm$ 0.042	23.8% (95/399)

Table 2

Performance of linear discriminant analysis (LDA) decision models that combine the filter-specific IT-CADe outputs. Different LDA models were built for each possible combination of filtering options (UN: unprocessed, M: median, W: adaptive Wiener, G: Gabor, H: entropy-based, STD: standard deviation-based). The table shows which combinations emerged as the superior ones depending on the number of inputs allowed in the LDA model.

LDA	$A_z (\pm\sigma)$	$0.99 A_z (\pm\sigma)$	Specificity at 95% sensitivity
2 filters: (M, W)	0.884±0.019	0.517±0.067	49.4% (197/399)
3 filters: (M, W, STD)	0.893±0.018	0.523±0.070	47.4% (189/399)
4 filters: (M, W, H, STD)	0.896±0.017	0.535±0.067	48.3% (193/399)
5 filters: (M, W, G, STD, UN)	0.897±0.018	0.549±0.067	49.9% (199/399)
ALL: (M, W, G, H, STD, UN)	0.898±0.018	0.548±0.068	50.4% (201/399)