Incorporation of a Laguerre–Gauss Channelized Hotelling Observer for False-Positive Reduction in a Mammographic Mass CAD System

Alan H. Baydush,^{1,2} David M. Catarious,³ Joseph Y. Lo,^{4,5} and Carey E. Floyd^{4,5}

Previously, we developed a simple Laguerre–Gauss (LG) channelized Hotelling observer (CHO) for incorporation into our mass computer-aided detection (CAD) system. This LG-CHO was trained using initial detection suspicious region data and was empirically optimized for free parameters. For the study presented in this paper, we wish to create a more optimal mass detection observer based on a novel combination of LG channels. A large set of LG channels with differing free parameters was created. Each of these channels was applied to the suspicious regions, and an output test statistic was determined. A stepwise feature selection algorithm was used to determine which LG channels would combine best to detect masses. These channels were combined using a HO to create a single template for the mass CAD system. Results from free-response receiver operating characteristic curves demonstrated that the incorporation of the novel LG-CHO into the CAD system slightly improved performance in high-sensitivity regions.

KEY WORDS: CAD, mass detection, classification, image processing, breast cancer

INTRODUCTION

For women, breast cancer is the second most common cause of cancer death. The American Cancer Society estimates that in 2002, breast cancer would be diagnosed in 203,500 women.¹ One of the most effective methods for detecting breast cancer is through screening mammography. It has been shown that early detection of suspicious regions helps improve patient outcomes. Several computer-aided detection (CAD) systems are available commercially, and some preliminary results on the use of these systems have shown an increase in cancer detection.^{2,3} Once again, this improved early detection is vital to positive patient outcomes. We firmly believe that improvements in

CAD techniques will have a great impact on early detection and hence on overall patient outcome.

Most CAD systems can be viewed as two distinct stages. The first stage generally uses a type of initial linear processing, with high sensitivity and low specificity, to detect an initial set of potential positive regions. The second stage typically consists of then classifying these suspicious regions using predictive modeling techniques, such as linear discriminants, neural networks, cluster analysis, etc., to reject a large number of false positives. Several commercial systems have been developed that detect mammographic masses using this type of two-step approach. From the clinician's point of view, one of the larger problems with these systems is the false positives per image (FPI) at a reasonable true-positive detection rate. Any efforts

¹From the Department of Radiation Oncology, Wake Forest University School of Medicine, Medical Center Blvd., Winston-Salem, NC 27157, USA.

²From the Department of Biomedical Engineering, Wake Forest University School of Medicine, Medical Center Blvd., Winston-Salem, NC 27157, USA.

³From the Washington, DC 20007, USA.

⁴From the Department of Biomedical Engineering, Duke University, Durham, NC USA.

⁵From the Department of Radiology, Duke Advanced Imaging Laboratories, Duke University Medical Center, Durham, NC USA.

Correspondence to: Alan H. Baydush, PhD., Department of Radiation Oncology, Wake Forest University School of Medicine, Medical Center Blvd., Winston-Salem, NC 27157, USA; tel: +1-336-7136508; fax: +1-336-7136565; e-mail: abaydush@wfubmc.edu

Copyright © 2007 by Society for Imaging Informatics in Medicine

doi: 10.1007/s10278-007-9009-8

toward the development of new techniques that can be used to reduce the number of false positives should be beneficial, and this focus is the goal of the current study.

A preliminary CAD system⁴ has been developed by our lab and has shown promising results. We are using this system as the backbone for the study presented in this paper. The first stage of this system involves several steps. The initial step is filtration with a difference of Gaussian (DOG) filter.⁵ This filter has been empirically determined and is applied using normalized cross-correlation. Next, a thresholding technique is used for initial suspicious region localization seeds. Each seed region is then passed on to a region segmentation routine⁴, which uses an iterative, linear classifier to determine inside and outside pixels for the segmentation. These steps are all used to create an initial set of suspicious regions, which are passed on to the second stage for false-positive reduction. The false-positive reduction stage of our system consists of feature extraction, followed by feature selection, classification, and false-positive reduction.

In this study, we discuss the continuing development and incorporation of a Laguerre–Gauss channelized Hotelling observer (LG-CHO) as an additional feature to be used in the false-positive reduction stage of our developing CAD system in hopes that it can be used to help further reduce false positives and improve the performance of the overall system.

Previously, we have developed and tested a simple LG-CHO for incorporation into our mass CAD system.⁶ This LG-CHO was trained using a database of suspicious regions generated from the initial detection stage of our CAD system, thus using computer-selected positives and false positives as the training set. This observer was empirically optimized for a number of channels and additional free parameters. For testing purposes, the LG-CHO was used as a filter template and applied using normalized cross-correlation to the images in our database. This output was used to generate four additional features, which were extracted from each image for each suspicious region: mean, standard deviation, peak value, and the value at the centroid. These four features were then incorporated into our mass CAD system, and the incorporation of this LG-CHO improved the overall performance.

For the study presented in this work, we wish to create a more optimal mass detection observer

model based on a novel combination of LG channels and incorporate this new LG-CHO as an additional feature to be used in the false-positive reduction stage of the CAD system we are developing. The resultant performance of this reoptimized system will be compared with the previous performance.

MATERIALS AND METHODS

This section will give an overview of the creation of our image database, a background information on HOs, a description of the CHO, a description of LG-CHOs, and finally, the general procedure used in this study.

a) Image Database

The mammograms that were used for this study were extracted from the University of South Florida's Digital Database for Screening Mammography.⁷ No institutional review board protocols were necessary as the database is anonymized and publicly available for download. The image study database consists of 181 cranial–caudal view mammograms that were scanned with a Lumisys scanner at 50 $\mu\text{m}/\text{pixel}$ and a bit depth of 12. These images were then resampled to 200 $\mu\text{m}/\text{pixel}$ for this study to decrease computational complexity. Of the 181 images, 43 images represented 43 patients and contained a total of 48 malignant masses. Forty images represented 40 patients and contained a total of 50 benign masses. In other words, there were 41 malignant-only images, 38 benign-only images, and 2 images which contained both malignant and benign masses. This totals 81 mass images. One-hundred images represented 96 patients and contained no abnormalities. In the cases where more than one image was used from a patient, different breasts were used. Masses with diameters greater than 60 mm were excluded from the study database. The study cases were randomly chosen so that the distribution of reported mass descriptors closely resembled the distribution for the complete Lumisys image set.

To create the suspicious region database for this study, we used the image database described above as the initial input to the first stage of our CAD system. After the steps of DOG filtration, thresholding, region localization, and region segmentation, we are left with an initial set of suspicious regions. These suspicious regions were

extracted based on their centroid location, and they constitute the suspicious region database. The initial sensitivity of the front end of the system was 95.83% of the malignant masses (2 of the 48 were not detected) with approximately 8.42 FPI for a total of 1,570 (1,524 + 46) suspicious regions.

For the study presented in this paper, a positive detection is defined as a malignant cancer being correctly identified. Cases without any abnormalities and with benign masses were considered negative.

b) Hotelling Observers

The Hotelling observer (HO) is a mathematically defined observer that effectively discriminates between a two-class system: positive and negative cases. This observer has been developed from the human vision model and requires information about image signal, background, and noise to discriminate the two-class system. In white noise, the HO reduces to a matched filter. In a medical image, which has correlated noise, the HO attempts to decorrelate the noise to separate the signal from the background noise.⁸ Previous research has shown that HOs can be used to effectively track the performance of human observers for detection tasks.^{9–14}

Mathematically, the HO reduces to a set of weights (also called a template), which can be applied to an image (vector) to give an output test statistic. This output test statistic, L , can then be used to optimally separate the two classes if the signal, background, and noise are modeled correctly. The weights or template for the HO is defined as follows:

$$W = [\langle S + B \rangle - \langle B \rangle] * K^{-1} \quad (1)$$

where S is the signal, B is the background, $S + B$ is the signal in the background, $\langle \rangle$ represents the mean, and K is the covariance matrix. This covariance matrix is calculated as the weighted mean of the signal and background covariance matrices. To calculate the output test statistic, L , we take the dot product of these weights and the image vector pixels.

$$L = \sum_{p=1}^{NumPixels} W_p * I_p \quad (2)$$

If modeled properly, this output test statistic will be larger when a signal is present and smaller when absent. Optimally, this test statistic will divide the positive and negative cases perfectly,

but this rarely happens. The effectiveness of this observer can be examined by using receiver operating characteristic (ROC)^{15–17} analysis. The most common metric examined with ROC analysis is the area under the curve (AUC).

In the case of actual medical images, however, the exact signal and background are not known. Therefore, estimates of the signal, background, and the necessary covariance matrix have to be used to generate the HO, and this makes the derived observer suboptimal. In typical practice, the average positive signal, the average background, and the weighted covariance matrix are calculated directly from the data.

Direct application of the HO to a large region of interest (ROI) is prohibitive as too many image samples are needed to properly estimate the covariance matrix.¹⁸ For example, for an ROI of size 128×128 pixels, the covariance matrix would be of size 16 k (128×128) by 16 k elements. This covariance matrix would require approximately five- to tenfold that number of samples for accurate assessment. It is intractable to expect to be able to collect this amount of “real” data; therefore, alternative solutions are necessary.

c) Channelized Hotelling Observers

Historically, to reduce the number of data samples necessary, the dimensionality of the HO has been reduced by applying channelized models to create a channelized Hotelling observer (CHO).¹⁹ Mathematically, a channel model is used by applying channels to the input data to reduce the dimensionality of the data and therefore the dimensionality of the signal and background. In most works using CHOs, a system of radially symmetric channels has been chosen for simplification. Each of these radial channels (templates themselves) would be applied to the data by using a dot product to give a single output. Generally, frequency averaging of certain expected important frequency bands is attempted. Each individual channel output is then used as the input of a HO as described previously.

The use of this type of CHO reduces the dimensionality of the covariance matrix to N by N , where N is the number of channels. For instance, in the case presented above, for ROIs of size 128×128 pixels, the covariance matrix is size 16×16 k. If a 10-channel model is used, the covariance matrix is reduced to 10×10 . This massive reduction in dimensionality allows for the

estimation problem to now be tractable with a reasonably sized dataset of images.

d) Laguerre–Gauss Channelized Hotelling Observers Next, the question arises of what channel model should be used. Barrett et al.¹⁸ suggest that smooth functions should be more favored. Following the work of Barrett et al., we have chosen to use functions based on Laguerre–Gauss (LG) channels. LG channels are formed as the product of Laguerre polynomials and Gaussians. Laguerre polynomials are defined as:

$$L_n(x) = \sum_{m=0}^n (-1)^m \binom{n}{m} \frac{x^m}{m!}, \quad \text{where } \binom{n}{m} = \frac{n!}{(n-m)!m!}. \quad (3)$$

Multiplying these polynomials with a Gaussian gives LG channels, where each n value corresponds to a different channel. Each channel is then multiplied by an appropriate channel weighting (α_n) determined by applying an HO to the channels, and the sum of all the channels is taken to form the final LG-CHO template, w . In polar coordinate notation, the final template looks like:

$$w(r) = \sum_n \alpha_n \exp\left(\frac{-\pi r^2}{a^2}\right) L_n\left(\frac{2\pi r^2}{a^2}\right) \quad (4)$$

where n is the number of channels. Figure 1a shows a 3D representation of a sample 25-channel LG-CHO template, whereas Figure 1b shows a profile through the midline.

e) Procedure

Previously, we have developed and tested a simple LG-CHO for incorporation into our mass CAD system. This observer was empirically optimized for a number of channels and additional free parameters. As stated previously, this LG-CHO was used as a filter template and applied to all of the images in our database, and four additional features were extracted from this output: mean, standard deviation, peak value, and the value at the centroid for each suspicious region. These features were incorporated into our mass CAD system, and the incorporation of this LG-CHO slightly improved the performance.

For the study presented in this work, we wish to create a more optimal mass detection observer based on a novel combination of LG channels. To perform this study, a large matrix of LG channels

with differing free parameters was created. To create this large matrix, the channel numbers were varied from one to 40, and the a value (see Eq. 4) was varied from five to 100 in increments of five. These values gave a large matrix of 840 separate observer channels.

Each of these separate observer channels was applied to each region in the suspicious region database, which consisted of 1,524 false-positive regions and 46 positive regions. An intermediate output test statistic, $W \cdot I$, was determined for each region for each LG channel. These output test statistics were then used as inputs to a stepwise feature selection (SWFS) algorithm, which used ROC AUC values as the objective function. In general, SWFS reduces the number of features used in a particular model by selecting the features (or channels in this case) that potentially combine the best to discriminate a multiclass system. For this study, the SWFS was used to choose which channels from the above large matrix were helpful in discriminating masses from false positives and to optimize how the channel information was

Fig 1. The 3D (a) and 2D (b) representations of a sample 25-channel LG-CHO template.

Fig 2. Slice through the optimized LG-CHO, which used three selected channels.

merged. We chose to use a Hotelling observer to combine the individual channels to determine a final output value and the AUC as our test metric to determine which combinations worked best.

The SWFS was used to determine which channels combined best to identify masses from suspicious regions. Instead of using all 840 different channels, the SWFS was used to reduce this set to something more manageable and realistic. The final reduced set of selected features was then combined using an HO to form a new overall template for false-positive reduction. This overall template was then applied to the entire image database using normalized cross-correlation, and the four previously mentioned values (mean, standard deviation, peak, and value at the centroid) were then incorporated into our mass CAD system as additional features to the ones that were previously measured and used by our system. All of these features were then used in the CAD systems feature selection and reduction stage, and overall system performance was measured using free-response ROC (FROC) analysis.

RESULTS

The goal of this study is to present a comparison of our original CAD system and the CAD system which uses the empirically determined

LG-CHO to the CAD system using our novel method of optimizing channel selection of LG-CHOs. The empirically determined LG-CHO used 40 channels and a value of 65. When the full channel matrix was used in a feature selection step to optimize the chosen channels and the a parameter, and the objective function of AUC increment of 0.01 was used, three channels were selected in the stepwise selection process. The a values were 70, 55, and 15, whereas the channel numbers were 7, 1, and 6, respectively, for the new novel LG-CHO template. Figure 2 shows a profile plot through the center of the three-channel optimized LG-CHO after the channels had been weighted and summed using the HO. This new, optimized LG-CHO was incorporated into the CAD system.

The original CAD system had selected and used five total features for its base level performance. When we incorporated the empirical LG-CHO, the system was reoptimized and selected four features, one of which was the value of the empirical LG-CHO at the centroid of the suspicious region. When we further incorporated the additional four features from the new, optimized LG-CHO, the system again selected four features, and one of the features selected was the mean value of the optimized LG-CHO in the suspicious region. The stepwise selection of these LG-CHO

Fig 3. FROC output of the original CAD system (solid gray), the CAD system incorporating the empirical LG-CHO (dashed), and the CAD system incorporating the optimized LG-CHO (solid with circles).

features shows that they are indeed adding information, which aids in the decision process for false-positive reduction.

Figure 3 shows an FROC comparison of the three CAD systems: the original CAD system (solid gray line), the CAD system incorporating the empirical LG-CHO (dashed line), and the CAD system incorporating the optimized LG-CHO (solid line with circles).

DISCUSSION

Previously, we had shown that incorporation of an empirically determined LG-CHO into this system helped improve FROC performance in the sensitivity range below 91%. While this improvement was important as most CAD systems operate below that level, we wished to develop a better model to help maintain high sensitivity over a wider range of specificities.

The purpose of this study was to investigate the use of a novel methodology for optimized channel selection to incorporate LG-CHOs into a mass detection CAD system. A large matrix of LG channels was formed, and these channels were applied to each of the ROIs in a suspicious region database. These outputs were then used as inputs to a stepwise feature selection stage. In this manner, each individual LG channel is considered independently as a separate “feature”, and stepwise selection is used to determine which channels combine best to improve discrimination performance. These selected channels were then combined to form an optimized LG-CHO model. In our study, three channels were selected to form this new observer. This new observer model was then used to generate four new features for our CAD system. Our CAD system selected one of these new features during its training procedure. This selection demonstrates that the system improved with the incorporation of this LG-CHO model.

When this newly developed CAD system is compared to the original system and the system incorporating the empirical LG-CHO, the FROC curves show promising results for this methodology. As already stated, the empirical LG-CHO system showed improvement at levels below 91% sensitivity. The new system tracks exactly with the empirical LG-CHO system up to 86% sensitivity but definitely outperforms the LG-CHO

system and the original system at high sensitivity. The new system maintains maximum sensitivity down to having 1.8 vs 2.35 FPI for the original system. At approximately 94% sensitivity, the original, empirical LG-CHO, and new LG-CHO system have 2.3, 2.2, and 1.6 FPI, respectively. Additionally, at 75% sensitivity, the original, empirical LG-CHO, and new LG-CHO system have 0.7, 0.55, and 0.5 FPI, respectively. At both of these levels of sensitivity, the new system shows an approximate 30% reduction in the number of FPI vs the original system.

For the results presented in this paper, the same dataset was used for both training and testing purposes. We feel, however, that this is not an issue because only three channels were used in creating the optimized HO. With the number of positive and negative regions used to train the system and with the results being evaluated over the entire image and not ROIs, we feel that there is very little chance of overtraining. Additional tests with independent datasets will be performed in future work to demonstrate this assertion.

While these results are not statistically significant, the improvement in the overall performance of the CAD system, especially in the high-sensitivity region of the FROC curve where we wish to operate, demonstrates the usefulness of this type of novel channel selection process.

CONCLUSIONS

Future work in using this type of selection procedure with more advanced channel models should be investigated and should provide continued improvements in the development of false-positive reduction stages in our CAD system.

ACKNOWLEDGMENTS

We would like to gratefully acknowledge the support for this research from the DOD Breast Cancer Research Program, DAMD17-02-1-0367.

REFERENCES

1. ACS, American Cancer Society: Cancer Facts and Figures 2002. Atlanta, GA: American Cancer Society 2002, 2002
2. Freer TW, Ullissey MJ: Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center. *Radiology* 220:781–786, 2001

3. Castellino RA, Roehrig J, Zhang W: Improved Computer-aided Detection (CAD) Algorithms for Screening Mammography in Radiology. 2000, p 400
4. Catarious DM, Baydush AH, Floyd Jr, CE: Incorporation of an iterative, linear segmentation routine into a mammographic mass CAD system. *Med Phys* 31:1512–1520, 2004
5. Catarious DM, Baydush AH, Lo JY, Floyd Jr, CE: Characterization of difference of Gaussian filters in the detection of mammographic regions. *Med Phys* 33:4104–4114, 2006
6. Baydush AH, Catarious DM, Floyd Jr, CE: Incorporation of Laguerre–Gauss channelized Hotelling observer into a mammographic mass CAD system. In: *International Workshop on Digital Mammography*, 2004
7. Heath M, Bowyer KW, Kopans D: Current status of the digital database for screening mammography. In: Karssemeijer N, Thijssen M, Hendriks J Eds. *Digital Mammography*. The Netherlands: Kluwer Academic Publishers, 1998, pp 457–460
8. Eckstein M, Abbey C, Whiting J: Human vs model observers in anatomic backgrounds. In: *Proceedings of SPIE: Image Perception*, 1998, vol 3340, pp 16–26
9. Fiete RD, Barrett HH, Smith WE, Myers KJ: Psychophysical study to test the ability of the Hotelling trace criterion to predict human-performance. *J Opt Soc Am A–Opt Image Sci Vis* 3:P126, 1986
10. Fiete RD, Barrett HH, Smith WE, Myers KJ: Hotelling trace criterion and its correlation with human-observer performance. *J Opt Soc Am A–Opt Image Sci Vis* 4:945–953, 1987
11. Gifford HC, King MA, de Vries DJ, Soares EJ: Channelized Hotelling and human observer correlation for lesion detection in hepatic SPECT imaging. *J Nucl Med* 39:771, 1998
12. Gifford HC, Wells RG, King MA: A comparison of human observer LROC and numerical observer ROC for tumor detection in SPECT images. *IEEE Trans Nucl Sci* 46:1032–1037, 1999
13. Gifford HC, King MA, de Vries DJ, Soares EJ: Channelized Hotelling and human observer correlation for lesion detection in hepatic SPECT imaging. *J Nucl Med* 41:514–521, 2000
14. Wollenweber SD, Tsui BMW, Lalush DS, Frey EC, LaCroix KJ, Gullberg GT: Comparison of Hotelling observer models and human observers in defect detection from myocardial SPECT imaging. *IEEE Trans Nucl Sci* 46:2098–2103, 1999
15. Metz C: Evaluation of CAD methods. In: Doi K, MacMahon H, Giger ML, Hoffmann KR Eds. *Computer-Aided Diagnosis in Medical Imaging*. Amsterdam: Elsevier Science, 1998, pp 543–554
16. Metz CE: Basic principles of ROC analysis. *Semin Nuc Med* 8:283–298, 1978
17. Metz CE: ROC methodology in radiologic imaging. *Invest Radiol* 21:720–733, 1986
18. Barrett HH, Abbey CK, Gallas BG: Stabilized estimates of Hotelling-observer detection performance in patient-structured noise. In: *Proceedings of SPIE: Image Perception*, 1998, vol 3340, pp 27–43
19. Myers KJ, Barrett HH: Addition of a channel mechanism to the ideal-observer model. *J Opt Soc Am A–Opt Image Sci Vis* 4:2447–2457, 1987