Observer Evaluations of Wavelet Methods for the Enhancement and Compression of Digitized Mammograms

Maria Kallergi[1], John J. Heine[1], and Bradley J. Lucier[2]

[1] H. Lee Moffitt Cancer Center & Research Institute, Cancer Control and Prevention, 12902 Magnolia Drive, Tampa, Florida, USA, 33620
{kallergi, heinejj}@moffitt.usf.edu
[2] Department of Mathematics, Purdue University, West Lafayette, IN
Lucier@purdue.edu

Abstract. Two observer experiments were performed to evaluate the performance of wavelet enhancement and compression methodologies for digitized mammography. One experiment was based on the localization response operating characteristic (LROC) model. The other estimated detection and localization accuracy rates. The results of both studies showed that the two algorithms consistently improved radiologists' performance although not always in a statistically significant way. An important outcome of this work was that lossy wavelet compression was as successful in improving the quality of digitized mammograms as the wavelet enhancement technique. The compression algorithm not only did not degrade the readers' performance but it improved it consistently while achieving compression rates in the range of 14 to 2051:1. The proposed wavelet algorithms yielded superior results for digitized mammography relative to conventional processing methodologies. Wavelets are valuable and diverse tools that could make digitized screen/film mammography equivalent to its direct digital counterpart leading to a filmless mammography clinic with full inter- and intra-system integration and real-time telemammography.

1 Introduction

Wavelets have found several applications in medical imaging including mammography. Applications range from image compression to image enhancement, feature extraction and segmentation to image reconstruction.[1] Depending on the selected type of wavelet, the outcome even within the same application may be dramatically different. In addition, a single wavelet processing may yield multiple effects, e.g., enhancement and compression, enhancement and segmentation.

We have experimented with several wavelet methods for a variety of processes of digitized and digital mammograms.[2],[3],[4],[5],[6],[7]. In this paper, we will report the results from the wavelet-based enhancement [8] and compression [9] of the same set of digitized mammograms that were evaluated by the same radiologists in similar experiments. The results, significant on their own, are analyzed here simultaneously to obtain a better understanding of the effect of the wavelet analysis on the images as well as the observer.

Susan M. Astley et al. (Eds.): IWDM 2006, LNCS 4046, pp. 482 – 489, 2006.

The work presented here is based on high-resolution digitized mammograms as opposed to direct digital mammograms. The reason for this lies in our past efforts and deep interest to find ways to integrate screen/film (SFM) with full field digital mammography (FFDM), a process that is currently facing serious impediments due to the advent of FFDM and the shift of interest, not unjustifiably, to the latter. However, film mammography is the current standard of practice worldwide with a major share in the international system market. Furthermore, mammography can no longer stay outside the filmless radiology department. Hence, methodologies that provide solutions to a filmless SFM are urgently needed.

2 Materials and Methods

2.1 Wavelet Enhancement Method

The purpose of enhancing digitized mammograms was to obtain high quality images that could be used for primary diagnosis from computer monitors (softcopy display and interpretation). For this application, we used multiresolution statistical analysis [4],[10] based on the orthogonal wavelet expansion of the original images and Fourier spectral characterization.[5] The 12-coefficient wavelet basis was used that is nearly symmetric with the mother wavelet having a large, almost symmetric, center lobe that resembles to some degree to the profile of the average calcification. More details of the method are given in Ref. [8].

2.2 Wavelet Compression Method

The images in this application were decomposed using a biorthogonal wavelet decomposition. Specifically, we used the biorthogonal, fifth-order accurate wavelets with piecewise constant duals of Cohen, Daubechies, and Feauveau, found on page 272 of Ref. [11]. The fifth-order wavelet was used for compression because it was found to give measurably smaller RMS errors at the same compression rates that the lower order wavelets. More details of the method are given in Ref. [9].

2.3 Evaluation Experiments

Two evaluation studies were performed for the two methodologies. First, a localization response operating characteristic (LROC) experiment was conducted. The LROC evaluation involved both signal likelihood and signal location tasks that, theoretically, offer a more complete analysis of observer performance. The LROC test was followed by a localization experiment that resembled the multiple alternative forced choice (MAFC) setup.[12] The results of both LROC evaluations are reported in detail elsewhere [8],[9] and will be briefly summarized here. The second evaluation test is the focus of this work.

The same database and readers were used for all tests. The set consisted of 500 single view mammograms, 250 of which were negative, 131 benign, and 119 cancer cases. A total of 375 findings were present in the benign and cancer cases, 182 of which were masses (98 benign and 84 cancer) and 193 calcification clusters (100 benign and 93 cancer). Negative cases were selected from negative mammograms

with at least two years of negative follow-up. Negative views matched the abnormal ones (benign or malignant) in terms of breast parenchymal density and size. Films were digitized at 30 μm and 16 bits per pixel with an ImageClear R3000 scanner (DBA Inc., Melbourne, FL).

All digital images were reviewed on one or two high-resolution DR 110 monitors (Data-Ray Corp., Westminster, CO) with Md5/SBX boards (Dome, Waltham, MA) in an Ultra Sparc 2 workstation (Sun Microsystems, Santa Clara, CA). Each DR110 monitor provided a 2048×2560 pixel display with an 8-bit digital to analog (DAC) converter.

In the LROC studies, the 500 single-view mammograms were reviewed one at a time in three different formats (original, enhanced, compressed) randomly mixed by three expert mammographers. The observers reported the x,y coordinates of a detected lesion and rated the suspiciousness for each detected lesion and the overall view using a custom-made user interface.

In the localization experiment, the 250 abnormal images were matched with the 250 negative images in terms of size and breast density and presented in left/right pairs in three formats (original, enhanced, compressed) randomly mixed to the observers, who compared the two views, selected the suspicious one, and localized and rated abnormal finding(s) similar to the LROC test. As mentioned earlier, this setup is similar to the MAFC but it is not a true MAFC experiment because it involves many targets in different backgrounds. Nevertheless, our goal for this test was to determine the ability of the readers to identify the abnormal view from a pair, compare the result to LROC, and perform another relative comparison of the wavelet methodologies.

Our studies were approved by the institutional review board as a research study using existing medical records and exempted from individual patient consent requirements. The patient identifiers were obliterated from all images.

2.4 Data Analysis

First, the x,y coordinates selected by the readers from both tests were compared to a ground truth file to determine the number of correct and incorrect localizations. A finding was considered as a hit or correct localization, if its x,y coordinates were within ±200 pixels of those listed in the truth file. If the difference was greater than 200 pixels then the finding was considered as an incorrect localization or a miss.

The LROC program, version of 1998, was applied to the LROC data.[13] ROC and LROC fitted curves were generated in this case including estimates of the areas under these curves and their standard errors. Two performance indices were primarily considered and compared: the detection accuracy, which corresponds to the area under the ROC curve (A_{ROC}), and the localization accuracy (P_{CL}), which corresponds to the ordinate of the LROC curves.[8], [9]

For the localization experiment, performance was determined by analyzing the selections of the observers in terms of both lesions and views. Rates for “lesion hits”, “lesion misses”, “view hits”, and “view misses” were estimated based on the correct and incorrect view selections and lesion localizations as follows: (a) the “lesion hit”

rate, defined as the fraction of correctly identified abnormal views with at least one lesion correctly localized, (b) the "lesion miss" rate, defined as the fraction of correctly identified abnormal views but with none of the lesions correctly localized, (c) the "view miss" rate, defined as the fraction of negative images that were incorrectly selected as the abnormal ones. Note that the "view hit" rate, i.e., the fraction of abnormal images (benign or malignant) that were correctly selected as abnormal independent of whether the true lesion(s) was correctly localized can be determined as 1-"view miss" rate. In addition to the overall accuracy in lesion localization, the hits and misses of the observers were analyzed in terms of pathology (benign/malignant) and type of lesion (calcification cluster/mass).

3 Results

3.1 LROC Performance Indices

The results of the two that for these plots and calculations, we combined the benign and cancer cases LROC studies have been already analyzed and reported independently elsewhere.[8], [9] Figure 1 shows the ROC and LROC curves for all three readings modes, i.e, original, enhanced, and compressed mammograms for one of the three readers. Similar results were obtained from the other readers.

Tables 1-3 list the performance indices for all observers and for the three reading modes. Performance indices include the area under the ROC curve (A_z) and its standard error (SE), the area under the LROC curve, the localization accuracy (P(CL)) and its standard error. Note in one group, labeled "abnormal", and compared them to the negatives cases, "normal" group. This is different from what was previously published and focuses more on the detection than the diagnostic aspect of the studies.

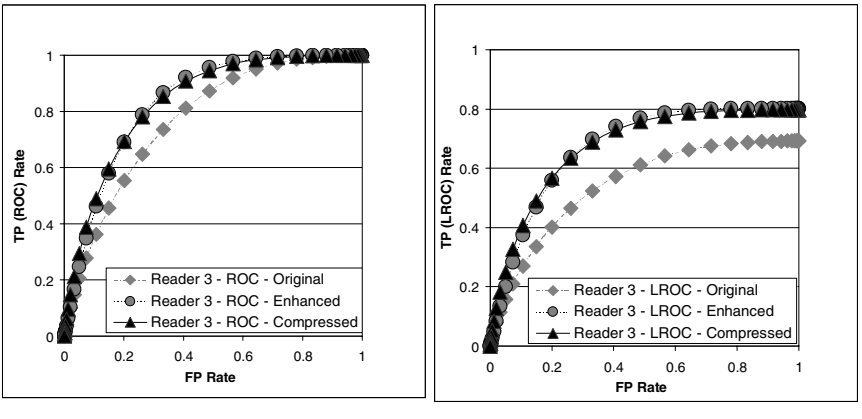

Fig. 1. Graphs for Reader 3 show fitted (a) ROC and (b) LROC curves obtained from the interpretation of original, enhanced, and compressed mammograms from patients with no findings (negative) versus mammograms from patients with benign or malignant findings. The performance indices of this reader are listed in Table 3.

Table 1. Performance indices obtained from the LROC analysis of the original data

Reader	ROC			LROC	
	A_z	SE	Area	P(CL)	SE (P(CL))
1	0.8013	0.0137	0.6027	0.7357	0.0224
2	0.7749	0.0137	0.5497	0.6964	0.0232
3	0.7718	0.0142	0.5435	0.6914	0.0249

Table 2. Performance indices obtained from the LROC analysis of the enhanced data. An average of 11% improvement was observed in localization accuracy with the enhanced images.

Reader	ROC			LROC	
	A_z	SE	Area	P(CL)	SE (P(CL))
1	0.8490	0.0128	0.6980	0.8064	0.0196
2	0.8081	0.0133	0.6163	0.7589	0.0201
3	0.8366	0.0131	0.6732	0.8016	0.0190

Table 3. Performance indices obtained from the LROC analysis of the compressed data. An average of 12% improvement was observed in localization accuracy with the compressed reconstructed images.

Reader	ROC			LROC	
	A_z	SE	Area	P(CL)	SE (P(CL))
1	0.8510	0.0128	0.7019	0.8092	0.0195
2	0.8164	0.0132	0.6328	0.7673	0.0200
3	0.8370	0.0132	0.6739	0.7971	0.0198

3.2 Detection and Localization Performances

Table 4 lists the number of correctly and incorrectly localized lesions and abnormal mammograms for all three readers; the corresponding rates are included in parentheses. We observe that for all readers the number of missed lesions was decreased with the enhanced and compressed images compared to the original data. A similar performance was observed for the number of correctly and incorrectly identified abnormal views.

Table 4. Correctly and incorrectly localized benign or malignant lesions (Lesion Hit and Lesion Miss) and mammographic views incorrectly identified as abnormal (View Miss) in the pair selection experiment. Corresponding rates are included in parentheses.

	Lesion Hit			**Lesion Miss**			**View Miss**		
Reader	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
Org	169 (68%)	155 (62%)	158 (63%)	32 (13%)	35 (14%)	34 (14%)	49 (20%)	60 (24%)	58 (23%)
Enh	197 (79%)	174 (70%)	186 (74%)	29 (12%)	39 (16%)	32 (13%)	24 (10%)	37 (15%)	32 (13%)
Comp	191 (76%)	182 (73%)	199 (80%)	21 (8%)	41 (16%)	28 (11%)	38 (15%)	27 (11%)	23 (9%)

Tables 5 and 6 break down the performance of each reader for the various types of abnormalities that were present in the mammograms, i.e., calcification clusters and masses, and pathology, i.e., benign and cancer. Both results indicate that all readers improved their localization performance with the enhanced and compressed reconstructed images. However, few differences were statistically significant.

Table 5. Number of correctly localized benign and malignant calcification clusters by each reader. Corresponding rates are included in parenheses. Note that the 250 abnormal mammographic views included a total of 193 calcification clusters (100 benign and 93 cancer).

	Calcification Clusters					
	Benign			Cancer		
Reader	1	2	3	1	2	3
Org	44 (44%)	43 (43%)	39 (39%)	44 (47%)	44 (47%)	50 (54%)
Enh	57 (57%)	48 (48%)	50 (50%)	52 (56%)	45 (48%)	52 (56%)
Comp	55 (55%)	50 (50%)	55 (55%)	48 (52%)	50 (54%)	50 (54%)

Table 6. Number of correctly localized benign and malignant masses by each reader. Corresponding rates are included in parenheses. Note that the 250 abnormal mammographic views included a total of 182 masses (98 benign and 84 cancer).

	Masses					
	Benign			Cancer		
Reader	1	2	3	1	2	3
Org	45 (46%)	36 (37%)	32 (33%)	36 (43%)	32 (38%)	37 (44%)
Enh	51 (52%)	44 (45%)	42 (43%)	37 (44%)	37 (44%)	40 (48%)
Comp	49 (50%)	48 (49%)	49 (50%)	39 (46%)	34 (35%)	45 (54%)

4 Discussion and Conclusions

Our current work focuses on issues related to the seamless integration of SFM and FFDM. This integration is seriously hindered by the lack of advanced tools and systems for the former and the significant delay in the development of such tools relative to FFDM that receives most of the attention. However, SFM is the current standard of clinical practice with millions of examinations performed worldwide. It is expected that digital will replace film in the future. Until then, however, film-based mammography clinics cannot afford to stay outside a filmless radiology department. Finding a solution to their integration should be an immediate priority.

The results of the two observer studies led to several interesting conclusions: (i) Our wavelet enhancement approach could significantly improve the detection of abnormalities in digitized softcopy mammography. The technique offers a robust and generally applicable approach independent of film digitization conditions or digitizer. Results could be further improved by modifying the algorithm to address challenging cases such as the mammograms of low breast density where the digital image quality is usually low or to better match the display medium characteristics. (ii) Our lossy wavelet compression method yielded high compression rates without compromising diagnostic performance. The mean compression rate was 59:1 for the negative

mammograms, 56:1 for the benign images, and 53:1 for the cancers.[9] Such high compression rates without visual losses, and hence, without losses in diagnostic power, could offer effective solutions to the problems of display, transfer, and storage of digitized, and possibly digital mammograms. (iii) The localization experiments are valuable in understanding the observer performance. The results of both tests indicate that the true lesions are not always accurately localized by the readers and critical signals are often missed or mispositioned. Most of the benign findings are easily and automatically discarded in the review process while detection of either benign or malignant lesions is seriously limited when a single view or limited information is presented. This has a major impact on the design of validation experiments and the selection of validation methodologies.

In conclusion, the experiments presented here supported our hypothesis that wavelets hold significant advantages for digitized mammography and could bridge the gap between digitized and direct digital mammography, thus facilitating the integration of film and filmless departments. Wavelet enhancement could support softcopy reading of digitized mammograms while wavelet compression could yield visually lossless, high-rate compression of the digitized films to facilitate storage and transmission. Interestingly, the two effects may be achieved through the same algorithm as suggested by our wavelet compression technique that showed improved tumor localization similar to the enhancement process.

References

1. IEEE Trans. Med. Imaging, Special Issue on Wavelets in Medical Imaging. Unser M, Alroubi A, and Laine A. (Eds.), 2003.
2. Kallergi M, Clarke LP, Qian W, et al. Interpretation of calcifications in screen/film, digitized, and wavelet-enhanced, monitor displayed mammograms: an ROC study. *Acad Radiol* 1996;3:285-293
3. Qian W, Clarke LP, Kallergi M, Clark RA. Tree structured nonlinear filters in digital mammography. *IEEE Trans Med Imaging* 1994;13:25-36
4. Heine JJ, Kallergi M, Chetelat SM, Clarke LP. Multiresolution wavelet approach for separating the breast region from the background in high resolution digital mammography. In: Karssemeijer N, Thijssen M, Hendricks J, van Erning L, eds. *Digital Mammography: proceedings of the 4th international workshop on digital mammography*. The Netherlands, Kluwer, 1998:295-298
5. Heine JJ, Velthuizen RP. A statistical methodology for mammographic density detection. *Med Phys* 2000;27:2644-2651
6. Lucier BJ, Kallergi M, Qian W, et al. Wavelet compression and segmentation of mammographic images. J Digit Imaging 1994; 7(1):27-38.
7. Yang Z, Kallergi M, DeVore R, et al. The effect of wavelet bases on the compression of digital mammograms. IEEE Eng Med Biol 1995; 14(5):570-577.
8. Kallergi M, Heine JJ, Berman CG, Hersh MR, Romilly AP, and Clark RA. Improved interpretation of digitized mammography with wavelet processing: A localization response operating characteristic study. AJR 2004; 182:697-703.
9. Kallergi M, Lucier BJ, Berman CG, Hersh MR, Kim JJ, Szabunio MS, and Clark RA. High-Performance Wavelet Compression for Mammography: Localization Response Operating Characteristic Evaluation. Radiology 2006; 238:62-73.

10. Heine JJ, Deans SR, Cullers DK, Stauduhar R, Clarke LP. Multiresolution statistical analysis of high resolution digital mammograms. *IEEE Trans Med Imaging* 1997;16:503-515
11. Daubechies I. Ten Lectures on Wavelets. SIAM, Philadelphia, 1992.
12. Burgess AE. Comparison of receiver operating characteristic and forced choice observer performance measurement methods. Med. Phys. 1995; 22(5):643-655.
13. Swensson RG. Unified measurement of observer performance in detecting and localizing target objects on images. *Med Phys* 1996;23:1709-1725.