Differential Use of Image Enhancement Techniques by Experienced and Inexperienced Observers

Elizabeth A. Krupinski, Ph.D., Hans Roehrig, Ph.D., William Dallas, Ph.D., and Jiahua Fan, M.S.

Full-field digital mammography (FFDM) systems are currently being used to acquire mammograms in digital format, but digital displays are less than ideal compared to traditional film-screen display. Certain physical properties of softcopy displays [e.g., modulation transfer function (MTF)] are less than optimal compared to film. We developed methods to compensate for some of these softcopy display deficiencies, based on careful physical characterization of the displays and image-processing software. A series of 100 FFDM and 60 digitized images was shown to six observers—half experienced (mammographers) and half inexperienced (radiology residents). The observers had to decide if a mass or microcalcification cluster was present and classify it as benign or malignant. A window could be activated that brought the image detail within the window to full resolution and corrected for the nonisotropic MTF of the Cathode Ray Tube (CRT) display. Experienced readers had better diagnostic performance and took less time to view the images. Experienced readers used window/level more than inexperienced readers, but inexperienced readers used magnification and the MTF compensation tool more often. Use of the magnification and the MTF tool increased reader decision condence. Experienced and inexperienced readers use image-processing tools differently, with certain tools increasing reader confidence. Understanding how observers use image-processing tools may help in the development of better and more automated user interfaces.

KEY WORDS: Observer performance, mammography, softcopy display, optimization

INTRODUCTION

Full-field digital mammography (FFDM) has many advantages over traditional film display of mammograms, but its potential has yet to be realized fully. Many studies have shown diagnostic equivalence between digital mammography and traditional film, but a clear superiority with digital has yet to be shown.[1–4] Perhaps one reason is that the softcopy display has yet to be optimized for these large complex images. Part of the problem with softcopy displays is that they still have lower resolution than film, and certain characteristics such as the nonisotropic modulation transfer function (MTF) of the CRT display are less than optimal compared to film. Problems such as nonisotropic MTF may be overcome with the newer liquid-crystal display (LCD) technologies (these displays are isotropic in terms of horizontal and vertical MTF), but LCDs come with other limitations such as degradation of image quality with nonorthogonal viewing angles.[5,6]

Because the displays themselves are still imperfect, methods to compensate for their deficiencies are being investigated by a number of groups. An approach taken by Kundel et al.[7] was to create a "perceptually tempered" display based on estimating minimally detectable contrasts at different levels of display luminance, but it failed to show any significant differences in improvement when compared to a perceptually linear display (i.e., calibrated to the DICOM-14 Gray Scale Display Function Standard). A more common approach has been to investigate various image-processing

techniques.[8] Results of these types of studies have been mixed. Hemminger et al.[9] compared contrast-limited adaptive histogram equalization (CLAHE) and histogram-based intensity windowing (HIW) on simulated masses in mammograms and found that CLAHE was not effective at improving observer performance, but that HIW did hold some promise. Stefanoyiannis et al.[10] also found that CLAHE did not do well, but that a digital equalization technique that remaps gray level values by a correction factor accounting for thickness variations in breast periphery and breast density did improve visualization of anatomic features. In a study comparing performance of different FDDM systems, lesion types, and image-processing effects, Cole et al.[11] found that acquisition device and lesion type influenced performance but image processing did not.

Some image-processing techniques are designed to help compensate for deficiencies in either the acquisition or display devices themselves. Kallergi et al.[12] used a wavelet algorithm designed to attenuate image spectral characteristics for the long-range image correlation effects that can interfere with digital displays and found that it improved observer performance significantly compared to original images. Nunes et al.[13] developed a preprocessing technique that used information from the modulation transfer function (MTF) of the acquisition device to enhance the contrast of dense breast images. The study did not involve human observers, but a computer-based detection scheme had improved performance with the processed images compared to the originals. In another study, Krupinski et al.[14] found that a method to compensate for MTF deficiencies of the display improved observer performance significantly in the detection of microcalcifications. This study did not include masses and only used 512×512 regions of interest from mammograms instead of the complete image. The goal of this project was to further investigate the use of this MTF compensation technique using full mammograms and readers of different levels of experience.

MATERIALS AND METHODS

A series of 160 mammographic cases [cranio-caudal (CC) and medio-lateral oblique (MLO) views of the right and left breast] was used in a Receiver Operating Characteristic (ROC) study. One hundred were FFDM images, 46 acquired using the Trex FFDM system (Hologic LORAD Division, Danbury, CT) and 54 acquired using the GE Senographe FFDM system (GE Medical Systems, Waukesha, WI). Sixty were digitized screen-film images (Lumiscan 85, Kodak Corp., Rochester, NY). Half of the images contained masses and half contained microcalcification clusters. Half of the mass and half of the microcalcification cluster images were benign and the other half was malignant. All cases were biopsy-proven.

Six observers participated. Three were mammographers certified by the Mammography Quality Standards Act who read mammograms on a daily basis. Three were radiology residents (third and fourth year) who had been through at least one mammography rotation. The study was IRB-approved, and all observers gave informed consent to participate. The images were viewed in three sessions lasting about 1 h each. The observers were shown all four images (CC and MLO, right and left) from a case on two high-resolution (5 megapixels) CRT monitors (Siemens SMM21201P, Siemens Medical Systems, Erlangen, Germany) that were calibrated to the DICOM-14 Gray Scale Display Function Standard. The images were downsampled so all four could be shown at once. Ambient room lights were turned off. Viewing time was unlimited.

The observers were instructed to examine the images for masses and microcalcifications. They were told to report whether a mass or calcification cluster was detected and then report their confidence in that decision on a six-point scale where 1 = absent, definite and 6 = present, definite. They were also instructed to report whether they thought the detected lesion was benign or malignant. During their search of the images, they could use window/level operations and/or activate a specialized image-processing window. The window could be used to bring a region of interest to full resolution while autoranging (a.k.a. gray-level stretching, an image-processing technique used to maximize the brightness and contrast of the image data) the area inside the window. With an additional click of a mouse button, they could activate an MTF compensation algorithm[14,15] to improve the detectability of image details. Whether or not they used window/level, the magnifier and the MTF compensation function was recorded. Viewing time was also recorded.

The MTF compensation technique is essentially the same one used in.[14] The monitor MTF was derived from the line spread function (LSF). The small signal approximation was used because of the nonlinearity of the display.[16,17] Stimuli of square CRT fields of uniform background luminance, with the exception of a horizontal or vertical line in the middle, were imaged by a charge-coupled device camera. Because the display functions of soft copy systems are usually expressed as luminance vs. digital input, the digital data were converted to luminance values and then processed by a Wiener–Helstrom filtering algorithm to approximate a compensation filter, which has the primary goal of compensating for the mid- to high-frequency contrast losses of the particular CRT monitor. The MTF compensation processing is implemented as two one-dimensional filters in the Fourier domain. The measured vertical and horizontal MTF functions form the bases for constructing the filters. It is emphasized that only the spatial frequency amplitude attenuation expressed by the MTF is compensated for. There is no attempt to process the phase of the optical transfer function (OTF).

RESULTS

The Multi-Reader Multi-Case (MRMC) ROC method[18] was used to analyze observer performance. An initial analysis of the confidence data and the image-processing use data revealed that there were no statistically significant differences as a function of image type (Trex, GE or digitized), so the results presented here are based on the combined sets of observer data. The overall ROC area under the curve (Az) results for the calcification and mass cases for each of the six observers are shown in Figure 1. Readers 1–3 are the residents and 4–6 are the experienced mammographers.

A two-factor analysis of variance (ANOVA) was conducted with ROC Az as the dependent variable and experience level (experienced vs. inexperienced) and tool use (used or not for window/level, magnification, and MTF compensation) as independent variables. Overall, for both masses and microcalcifications, the experienced observers performed higher as would be expected (see Figure 1). For both microcalcifications and masses, there was a significant interaction effect between experience and use of window/level. ROC Az was higher when window/level was used by the experienced readers but not by the inexperienced readers ($F = 4.435$, $p = 0.0357$ for microcalcifications; $F = 15.452$, $p < 0.0001$ for masses). For magnification use, only the main effects were significant for both masses and microcalcifications. Experienced readers performed better than inexperienced ($F = 3266.582$, $p < 0.0001$ for microcalcifications; $F = 1945.583$, $p < 0.0001$
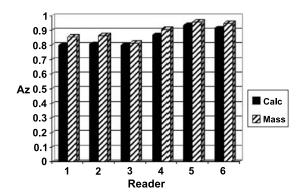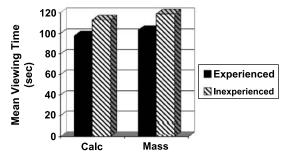

Fig 2. Mean viewing times for experienced and inexperienced readers on the mass and microcalcification images.

for masses), and both groups performed better with magnification use ($F = 7.937$, $p = 0.005$ microcalcifications; $F = 9.985$, $p = 0.0017$ masses) with no significant interaction effect. For MTF compensation tool use with microcalcifications, there was a significant interaction effect ($F = 8.540$, $p = 0.0036$), with experienced readers performing significantly higher with tool use, but inexperienced readers showing no difference in ROC Az with MTF tool use. For masses and MTF compensation tool use, the experienced readers performed better ($F = 1653.568$, $p < 0.0001$), and for both groups, performance was higher with tool use ($F = 6.004$, $p = 0.0146$) than without. There was no significant interaction between experience and tool use on performance.

Overall viewing time was significantly shorter for experienced readers (mean = 103.15 s, SD = 46.37 s) compared to inexperienced readers (mean = 119.03 s, SD = 53.81 s) for masses ($t = 3.462$, $df = 478$, $p = 0.0006$). Overall viewing time was significantly shorter for experienced readers (mean = 97.48 s, SD = 42.58 s) compared to inexperienced readers (mean = 113.39 s, SD = 52.66 s) for microcalcifications ($t = 3.640$, $df = 478$, $p = 0.0003$). The results are shown graphically in Figure 2.

Image-processing tool use also differed as a function of reader experience on both types of lesions. For the masses, the experienced readers (90% of the images) used window/level significantly more ($\chi^2 = 19.06$, $df = 1$, $p < 0.0001$) than the inexperienced readers (75% of the images). The inexperienced readers used the magnification (60%) and MTF compensation processing (40%) significantly more often ($\chi^2 = 25.22$, $df = 1$, $p < 0.0001$ for magnification; $\chi^2 = 20.83$, $df = 1$,

Fig 1. Az values for the six readers for the mass (stripes) and microcalcification (solid) images.

$p < 0.0001$ for MTF) than the experienced readers (38 and 21% for magnification and MTF, respectively). For the calcifications, the experienced readers (88%) again used window/level significantly more ($\chi^2 = 44.35$, $df = 1$, $p < 0.0001$) than the inexperienced readers (62%). The inexperienced readers used magnification (55%) and MTF compensation processing (47%) significantly more often ($\chi^2 = 4.03$, $df = 1$, $p = 0.0446$ for magnification; $\chi^2 = 14.85$, $df = 1$, $p = 0.0001$ for MTF) than the experienced readers (46 and 30% for magnification and MTF, respectively). The results are shown graphically in Figure 3.

The confidence data for the experienced and inexperienced observers were analyzed to determine if there was any relationship between confidence level and use of any of the image-processing tools. For the experienced observers on the mass images, there was no relationship between confidence and window/level use ($\chi^2 = 3.28$, $df = 5$, $p = 0.6571$), but there was between confidence and magnification use ($\chi^2 = 18.60$, $df = 5$, $p = 0.0023$) and MTF compensation tool use ($\chi^2 = 26.07$, $df = 5$, $p < 0.0001$). For both magnification and MTF compensation use, confidence ratings tended to be higher when the tools were used than when they were not. The same pattern of results for calcifications was observed, with window/level showing no significant relationship ($\chi^2 = 6.67$, $df = 5$, $p = 0.2465$) and magnification ($\chi^2 = 16.24$, $df = 5$, $p = 0.0062$) and MTF compensation ($\chi^2 = 44.22$, $df = 5$, $p < 0.0001$) showing higher confidence with tool use.

For the inexperienced observers on the mass images, there was no relationship between confidence and window/level use ($\chi^2 = 9.06$, $df = 5$, $p = 0.1068$), but there was between confidence and magnification use ($\chi^2 = 17.86$, $df = 5$, $p = 0.0031$) and MTF compensation tool use ($\chi^2 = 66.85$, $df = 5$, $p < 0.0001$). For both magnification and MTF compensation use, confidence ratings tended to be higher when the tools were used than when they were not. For calcifications, window/level showed no significant relationship ($\chi^2 = 6.40$, $df = 5$, $p = 0.2690$) and neither did magnification ($\chi^2 = 7.77$, $df = 5$, $p = 0.1697$), but MTF compensation did ($\chi^2 = 37.46$, $df = 5$, $p < 0.0001$) show higher confidence with tool use.

DISCUSSION

Overall, the experienced readers performed better (higher Az) than the inexperienced observers for both mass and microcalcification detection as was expected. Viewing times were also shorter for the experienced observers, and that was expected as well because both of these effects have been observed previously.[19,20] In terms of how the image-processing tools were used, there were some interesting differences seen between the experienced and inexperienced observers. The experienced observers used the window/level function significantly more than the inexperienced observers. However, for neither group was there a significant relationship between window/level use and decision confidence. The higher use of window/level by the experienced observers may be because of the fact that they normally use this function during their clinical interpretation of mammograms and thus have a better understanding of what they are looking for and how changes in window/level can affect the appearance of the structures they are looking at. This may be true for the group of radiologists used in this study because they read FFDM images on a daily basis, but may not hold true if we did the study with readers still using film. That it is not related to confidence is likely because of the fact that they simply use it on practically every image.

The inexperienced observers used the magnification and MTF compensation tools significantly more than the experienced observers, but for both groups, there was a significant increase in confi-
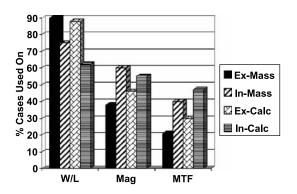

Fig 3. Percent of cases window/level (W/L), magnification (mag), and the MTF compensation processing (MTF) tools were used by the experienced and inexperienced readers on the mass and microcalcification images.

dence when those tools were used. One possible explanation for the difference in usage may be that the experienced observers are better able to judge the relevant characteristics of lesions (to determine benign vs. malignant) on a more global level than the inexperienced observers. The inexperienced observers needed to examine more closely (magnify) and enhance the features (MTF compensation), whereas the experienced observers could recognize these features without the enhancement aids.

Careful physical characterization of softcopy display devices and development of image-processing techniques to compensate for deficiencies in these displays can improve observer performance in the interpretation of mammographic images. Users of softcopy displays for mammographic interpretation need to be aware of some of the deficiencies in these displays and the fact that techniques are available to help in the interpretation process. The use of these techniques may take more time, but they can improve decision confidence.

ACKNOWLEDGMENTS

REFERENCES

1. Fischer U, Baum F, Obenauer S, et. al: Comparative study in patients with microcalcifications: full-field digital mammography *vs* screen-film mammography. Eur Radiol 12:2679–2683, 2002

2. Obenauer S, Hermann KP, Marten K, et. al: Soft copy *versus* hard copy reading in digital mammography. J Digit Imaging 16:341–344, 2003

3. Cole E, Pisano ED, Brown M, et. al: Diagnostic accuracy of Fischer Senoscan Digital Mammography *versus* screen-film mammography in a diagnostic mammography population. Acad Radiol 11:879–886, 2004

4. Pisano ED, Cole EB, Kistner EO, et. al: Interpretation of digital mammograms: comparison of speed and accuracy of softcopy *versus* printed-film display. Radiol 223:483–488, 2002

5. Krupinski EA, Johnson J, Roehrig H, Nafziger J, Fan J, Lubin J: Use of a human visual system model to predict observer performance with CRT *vs* LCD display of images. J Digit Imaging 17:258–263, 2004

6. Badano A, Flynn MJ, Martin S, Kanicki J: Angular dependence of the luminance and contrast in medical monochrome liquid crystal displays. Med Phys 30:2602–2613, 2003

7. Kundel HL, Weinstein SP, Conant EF, Toto LC, Nodine CF: A perceptually tempered display for digital mammograms. Radiographics 19:1313–1318, 1999

8. Pisano ED, Cole EB, Hemminger BM, Yaffe MJ, Aylward SR, et. al: Image processing algorithms for digital mammography: a pictorial essay. Radiographics 20:1479–1491, 2000

9. Hemminger BM, Zong S, Muller KE, Coffey CS, et. al: Improving the detection of simulated masses in mammograms through two different image-processing techniques. Acad Radiol 8:845–855, 2001

10. Stefanoyiannis AP, Costaridou L, Skiadopoulos S, Panayiotakis G: A digital equalization technique improving visualization of dense mammary gland and breast periphery in mammography. Eur J Radiol 45:139–149, 2003

11. Cole EB, Pisano ED, Kistner EO, Muller KE, Brown ME, Feig ME, et. al: Diagnostic accuracy of digital mammography in patients with dense breasts who underwent problem-solving mammography: effects of image processing and lesion type. Radiol. 216:153–160, 2003

12. Kallergi M, Heine JJ, Berman CG, Hersh MR, Romilly AP, Clark AP: Improved interpretation of digitized mammography with wavelet processing: a localization response operating characteristic study. Am J Roentgenol 182:697–703, 2004

13. Nunes FL, Schiabel H, Benatti RH: Contrast enhancement in dense breast images using the modulation transfer function. Med Phys 29:2925–2936, 2002

14. Krupinski EA, Johnson J, Roehrig H, Engstrom M et. al: Using a human visual system model to optimize soft-copy mammography display: influence of MTF compensation. Acad Radiol 10:1030–1035, 2003

15. Reiker GG, Blume H, Slone RM, Woodward PM, et. al: Filmless digital chest radiography within the radiology department. Proc SPIE Med Imag 3035:355–368, 1997

16. Roehrig H: The monochrome cathode ray tube display and its performance. In: Kim Y, Horii SC (Eds). Handbook of Medical Imaging Volume 3. Display and PACS. Bellingham, WA: SPIE Press, 2000, pp 155–220

17. Blume H, Steven P, Ho A, Stevens F, Abileah A, Robinson S, et. al: Characterization of liquid-crystal displays for medical images—part 2. Proc SPIE Med Imag 5029:449–473, 2003

18. Dorfman DD, Berbaum KS, Metz CE: Receiver operating characteristic analysis: generalization to the population of readers and patients with the jackknife method. Invest Radiol 27:723–731, 1992

19. Krupinski EA: Visual scanning patterns of radiologists searching mammograms. Acad Radiol 3:137–144, 1996

20. Nodine CF, Mello-Thoms C, Kundel HL, Weinstein SP: Time course of perception and decision making during mammographic interpretation. Am J Roentgenol 179:917–923, 2002