

A Phantom Study for Assessing the Effect of Different Digital Detectors on Mammographic Texture Features

Yan Wang, Brad M. Keller, Yuanjie Zheng, Raymond J. Acciavatti,
James C. Gee, Andrew D.A. Maidment, and Despina Kontos

Department of Radiology, University of Pennsylvania, Philadelphia, PA USA
wangyan1@sas.upenn.edu

Abstract. Digital mammography (DM) is commonly used as the breast imaging screening modality. For research based on DM datasets with various sources of x-ray detectors, it is important to evaluate if different detectors could introduce inherent differences in the images analyzed. To determine the extent of such effects, we performed a study to compare the effects of two DM detectors, the GE 2000D and DS, on texture analysis using a validated breast texture phantom (Yaffe et. al, University of Toronto). DM images are acquired in Cranio-Caudal (CC) view, and texture features are generated for both raw and post-processed DM images. Image intensity profiles and texture features are compared between the two detector systems. Our results suggest that there are inherent differences in the images. For raw and processed images, the image intensity cumulative distribution function (CDF) curves reveal that there is a scaling and shifting factor respectively between the two detectors. Image normalization with z-score can reduce detector differences for grey-level intensity and the histogram-based texture features. The differences between co-occurrence and run-length texture features persist after intensity normalization, suggesting that simple z-scoring cannot alleviate all the detector effects, potentially also due to differences in the spatial distribution of the intensity values between the two detectors.

Keywords: Digital mammography, detectors, breast phantom, texture analysis.

1 Introduction

Breast cancer is considered a major health problem in western countries, as it comprises 10.4% of the cancer incidence among women, making it the second most common type of cancer. Early screening and proper treatment after diagnosis for individual women are both important aspects of current breast cancer research, and digital mammography (DM) is the main screening tool for cancer detection [1].

The Gail Model [2] has been shown be able to estimate breast cancer at the population level, however with limited capacity at the individual level. There's intensive research for individual breast cancer risk estimation, and mammographic density, estimated as the percent of dense tissue area within the breast, has been shown to be the strongest risk factor for breast cancer after age [3, 4]. Studies [5-8] also support a relationship between mammographic texture and breast cancer risk, as mammographic texture features may be able to quantify the local distribution of the

parenchymal pattern, potentially providing complementary information for breast cancer risk estimation.

In the process of developing proper imaging biomarkers for risk estimation, it is commonly the case that studies use DM images acquired with different imaging systems and different x-ray detectors. It may be important and necessary to treat the detector source as an additional parameter in the analysis, as different detectors may possibly introduce inherent differences in the DM images [9]. To determine the effects of x-ray detectors in mammographic texture analysis, we designed a physical breast phantom study. The image intensity and extracted texture features [10-12] from the breast phantom are compared between two x-ray detectors for both raw and the vendor post-processed (a.k.a., processed) images. The rationale is that since images are generated from the same phantom, quantitative imaging features should be consistent between the images (e.g., affected only minimally by noise). Instead of limiting to a square of region of interest (ROI) behind the nipple as has been done in prior work [6, 8], the textures of the whole breast region are used for the analysis. The results of this comparison study could guide the proper choice of more robust texture features that are less sensitive to detector differences. Our study can be potentially helpful for any studies utilizing texture analysis in digital mammography, including breast cancer risk assessment, breast tissue classification and computed aided lesion detection, as a method for assessing potential detector differences.

2 Methods

The Gammex 169 “Rachel” breast phantom was used in our experiments (Yaffe. et al, University of Toronto) [13]. Image acquisition was performed on GE Healthcare 2000D and DS FFDM system at 0.1mm/pixel resolution, 14 bit gray-levels. On each machine, the clinically optimized phototimed setting of (kVp, mAs) was chosen, which was 29 kVp, 71 mAs for the 2000D; and 29 kVp, 90 mAs for the DS system. The image acquisition process was repeated 5 times for both machines. The average of the 5 images was used to reduce the effects of noise in the imaging process.

In the original phantom image, the outer bounding case appeared in the image (Figure. 1). Therefore, as an additional preprocessing step, in order to avoid the operational artifacts (e.g. the phantom may not be centered perfectly on the two detectors and the outer case artifacts), we cropped a region of interest (ROI) corresponding to the breast area, as shown in the middle of Figure 1, and used a synchronic threshold scheme to generate the breast masks for the two detectors. Based on the assumption that the breast area in the image is the same using the two detectors, we optimize the thresholds of post-processed image intensity by solving the optimization problem:

$$\arg \min_{(t_1, t_2)} |BA(2000D, t_1) - BA(DS, t_2)| \quad (1)$$

here breast area is denoted as BA , t_1 , t_2 as the intensity threshold for the post-processed images from the two detectors. $BA(\text{detector}, t) =$ the cardinal of the set $\{p \mid \text{the pixel } p \text{ is in the ROI and the post-processed image intensity at } p \geq t\}$.

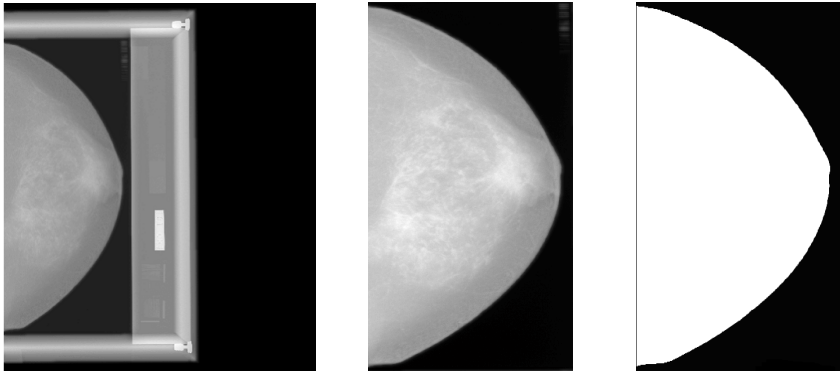


Fig. 1. Processed DM image acquired on the 2000D system; Left: original phantom image, Middle: manually cropped ROI, Right: the final breast region mask

Multiple texture features are extracted using an automated breast image analysis software pipeline [14], including 1) grey-level histogram features, 2) co-occurrence texture features, and 3) run-length features. These features have been shown in previous studies to have value in breast cancer risk estimation [5-7]. The texture images are generated by calculating texture features within a series of adjacent square regions covering the original breast region, with the side length of the square equals to 16 pixels. A total of 26 texture features were computed. A summary of these texture features is shown in Table 1.

Table 1. Texture descriptors included in parenchymal texture analysis

Grey-Level Histogram	5th /5thmean/95th/95thmean/ max/mean/min/sum entropy/kurtosis/sigma/skewness
Co-occurrence	cluster shade/ inverse difference moment correlation/energy/entropy/inertia Haralick correlation
Run-Length	grey level nonuniformity/ run length nonuniformity high grey level run emphasis/ long run emphasis low grey level run emphasis/short run emphasis run percentage

Our comparison study is based on comparing the image intensity and texture feature profiles of both the raw and processed DM phantom images, where the cumulative distribution function (CDF) curves of each image feature are computed and compared between the two x-ray detectors. All features in Table 1 are generated for both original images and the z-scored images, and the effects of z-scoring on detector differences are also evaluated. The Kolmogorov-Smirnov distance, which defined as the maximum of the absolute vertical difference between two CDF curves [15, 16], is used as the distance between two CDF curves in the result.

3 Results

The size of the ROI is 1842×775 pixels for the images of both detectors, with optimized threshold of $(t_1^*, t_2^*) = (1068, 456)$, with $BA(2000D, t_1^*) = 939916$ (pixels), and $BA(DS, t_2^*) = 940598$ (pixels). The breast area difference/ $BA(2000D, t_1^*) = 0.07\%$.

On the 2000D system, for raw and processed images, the standard deviation of the 5 times of phototimed imaging in terms of the maximum/minimum/standard deviation of the image intensity is $2.25/7.29/0.52$ and $9.4/4.2/0.0821$ respectively. On the DS system, the corresponding values are $24.4/5.39/1.1$ and $5.1/4.3/0.12$. In the following analysis, the average of the 5 phototimed images was used for each detector to reduce the effects of noise in the imaging process. The cumulative distribution function (CDF) of the image intensity within the breast region is shown for the two detectors (Fig. 2).

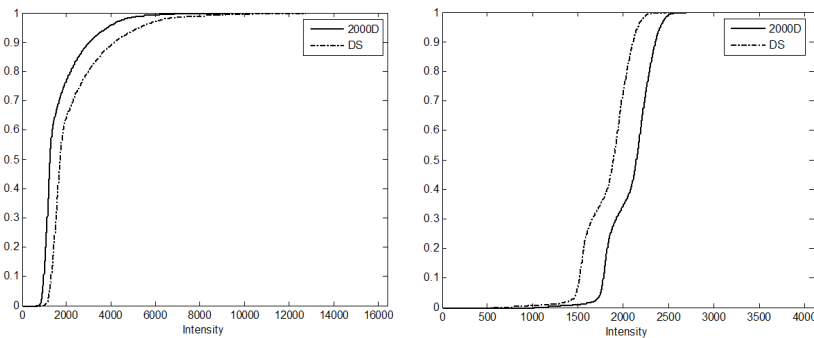


Fig. 2. The image intensity CDF of the original image. Left: raw, right: processed images.

The CDF comparison in the raw/processed images indicates that the intensity differences between the images from two detectors may be affected by a scaling/shifting factor respectively. After z-scoring the image intensity within the breast area, the differences in the CDF of the image intensity are alleviated, as shown in Figure 3. The distance between CDF curves of the two detectors are reduced from $0.4730/0.4731$ to $0.0249/0.0263$ for raw/processed images respectively.

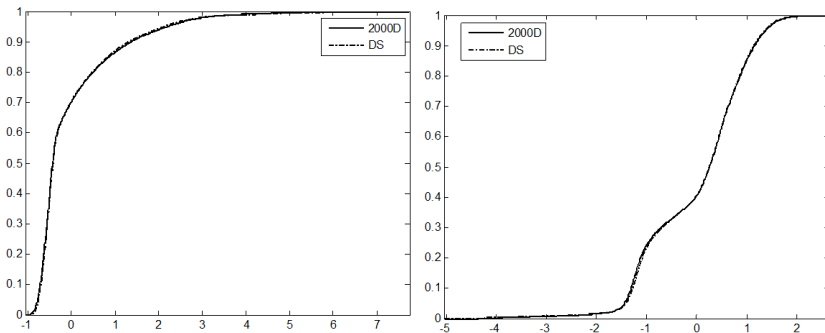


Fig. 3. The image intensity CDF of z-scored image. Left: raw, Right: processed image.

Comparisons are also done for all the texture features listed in Table 1. Our results indicate that texture features can be broadly categorized into three groups, according to how they are affected by the detector differences and z-scoring. Certain features are not affected by detectors, others are affected but compensated by z-score, and some are affected regardless.

Specifically, the first group of features that is minimally affected by detectors includes the grey-level histogram features (e.g., entropy, kurtosis, sigma, skewness). The CDF curves of the texture feature kurtosis is shown as an example in Figure 4, for which the distance between the CDF curves is 0.0839/0.0732 for the original raw and processed images, after z-scoring, the distance remains the same 0.0839/0.0732.

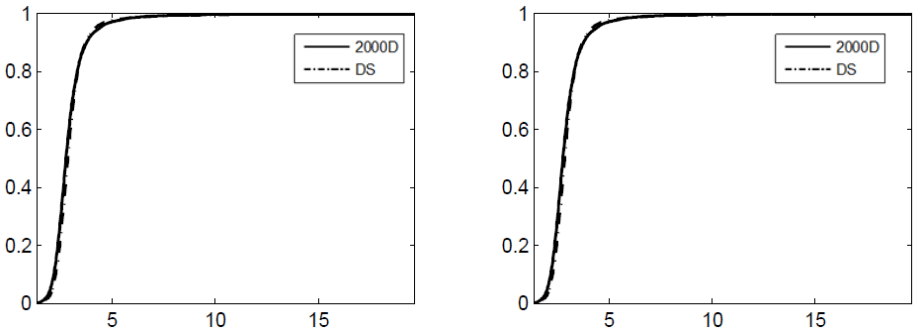


Fig. 4. Left: CDF of feature image from original image, right: CDF of feature image from z-scored image. (Grey-level histogram feature: kurtosis)

The second group of features is affected by detectors, but the differences in the CDF can be alleviated by z-scoring the original image. This includes the remaining grey-level histogram features and the cluster shade co-occurrence feature. As an example in Figure 5, for the grey-level histogram feature mean, the distance between the CDF curves of the two detectors is 0.4732/0.4728 for raw and processed images, after z-scoring, the distance is decreased to 0.0255/0.0328.

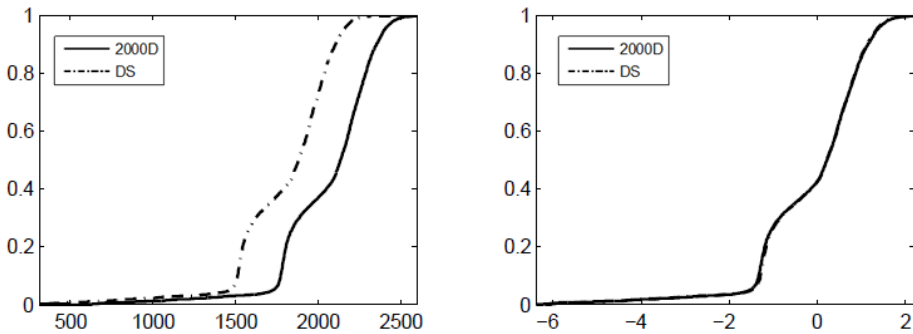


Fig. 5. Left: CDF of feature image from original image, right: CDF of the feature image from z-scored image. (Grey-level histogram feature: mean)

The third group of features is affected by detectors; however simple z-scoring cannot reduce the observed differences. This group mainly includes the co-occurrence texture features (except cluster shade) and run-length features. As an example in Figure 6, for the co-occurrence feature inverse difference moment, the distance between the CDF curves between two detectors is 0.2331/0.2301 for raw and processed images, however after z-scoring, the distance is increased to 0.3229/0.4145.

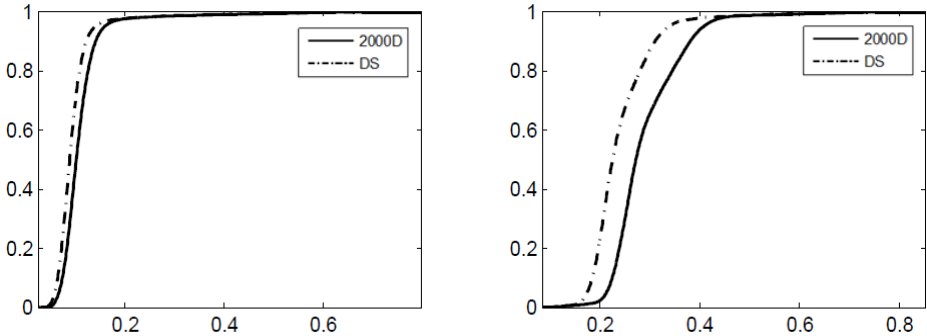


Fig. 6. Left: CDF of feature image from original image, right: CDF of feature image from z-scored image (Co-occurrence texture feature: inverse difference moment)

4 Discussion and Conclusion

In this study, we compared the image intensity and texture profiles of two GE DM x-ray detectors using a physical breast phantom. The rationale is that since images are generated from the same phantom, the resulting image features should remain very similar (e.g., effected only minimally by noise). Our results show that the CDF curves for processed and raw image intensity values between two detectors reveal a shifting and scaling pattern respectively. Comparing the different texture features suggests that the texture features can be broadly categorized into three groups. In summary, after z-scoring the image intensity of the original phantom images, the differences in the intensity values and the grey-level histogram features are alleviated, however the differences in texture features may depend not only on absolute gray-level intensity values in the image, but also on the spatial distribution of the image intensity values, such as most of the co-occurrence and run-length texture features differences between the two detectors, which are not reduced by simple intensity normalization z-scoring.

The CDF curve information studied in this study is used as the first step comparison study on how different x-ray detectors may affect the image intensity and texture. Further work is underway to fully-investigate such differences, and to develop a comprehensive feature standardization scheme that can be potentially used to reduce effects introduced by the imaging system on the subsequent image analysis process.

Acknowledgement. This work was supported in part by American Cancer Society Grant RSGHP-CPHPS-119586, United States Department of Defense Concept Award BC086591, and National Institutes of Health PROSPR Grant 1U54CA163313-01.

References

1. Smith, K.L., Isaacs, C.: Management of women at increased risk for hereditary breast cancer. *Breast Disease* 27, 51–67 (2006)
2. Gail, M.H., Brinton, L.A., Byar, D.P., Corle, D.K., Green, S.B., Schairer, C., Mulvihill, J.J.: Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J. Natl. Cancer* 81(24), 1879–1886 (1989)
3. Boyd, N.F., Guo, H., Martin, L.J., Sun, L., Stone, J., Fishell, E., Jong, R.A., Hislop, G., Chiarelli, A., Minkin, S., Yaffe, M.J.: Mammographic density and the risk and detection of breast cancer. *New England Journal of Medicine* 356(3), 227–236 (2007)
4. Harvey, J.A., Bovbjerg, V.E.: Quantitative Assessment of Mammographic Breast Density: Relationship with Breast Cancer Risk. *Radiology* 230(1), 29–41 (2004)
5. Huo, Z., Giger, M.L., Wolverton, D.E., Zhong, W., Cumming, S., Olopade, O.I.: Computerized analysis of mammographic parenchymal patterns for breast cancer risk assessment: feature selection. *Medical Physics* 27(1), 4–12 (2000)
6. Li, H., Giger, M.L., Olopade, O.I., Margolis, A., Lan, L., Chinander, M.R.: Computerized Texture Analysis of Mammographic Parenchymal Patterns of Digitized Mammograms. *Academic Radiology* 12(7), 863–873 (2005)
7. Manduca, A., Carston, M.J., Heine, J.J., Scott, C.G., Pankratz, V.S., Brandt, K.R., et al.: Texture Features from Mammographic Images and Risk of Breast Cancer. *Cancer Epidemiol. Biomarkers Prev.* 18(3), 837–845 (2009)
8. Kontos, D., Bakic, P.R., Carton, A.K., Troxel, A.B., Conant, E.F., Maidment, A.D.A.: Parenchymal texture analysis in digital breast tomosynthesis for breast cancer risk estimation: A preliminary study. *Academic Radiology* 16(3), 283–298 (2009)
9. Williams, M.B., Yaffe, M.J., Maidment, A.D., Martin, M.C., Seibert, J.A., Pisano, E.D.: Image quality in digital mammography: image acquisition. *Journal of the American College of Radiology* 3(8), 589–608 (2006)
10. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics* 3, 610–621 (1973)
11. Galloway, M.D.: Texture classification using gray level run length. *Computer Graphics and Image Processing* 4, 172–179 (1975)
12. Amadasum, M., King, R.: Textural features corresponding to textural properties. *IEEE Transactions on Systems Man and Cybernetics* 19, 1264–1274 (1989)
13. Caldwell, C.B., Yaffe, M.J.: Development of an anthropomorphic breast phantom. *Medical Physics* 17(2), 273–280 (1990)
14. Zheng, Y., Keller, B., Wang, Y., Tustison, N., Song, G., Bakic, P.R., Maidment, A.D., Conant, E.F., Gee, J.C., Kontos, D.: A Fully-Automated Software Pipeline for Parenchymal Pattern Analysis in Digital Breast Images: Toward the Translation of Imaging Biomarkers in Routine Breast Cancer Risk Assessment. In: Quantitative Imaging Reading Room, the 97th Scientific Assembly and Annual Meeting of the Radiological Society of North America (RSNA) 2011, Chicago, IL (software exhibit) (2011)
15. Smirnov, N.V.: Tables for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics* 19, 279–281 (1948)
16. Rachev, S.T.: Probability Metrics and Stability of Stochastic Models. JohnWiley & Sons (1991)