J Digit Imaging (2012) 25:662–669
DOI 10.1007/s10278-012-9478-2

Verification of DICOM GSDF in Complex Backgrounds

David L. Leong · Louise Rainford · Tamara Miner Haygood · Gary J. Whitman · Philip M. Tchou · William R. Geiser · Selin Carkaci · Patrick C. Brennan

Published online: 26 April 2012

Abstract While previous research has determined the contrast detection threshold in medical images, it has focused on uniform backgrounds, has not used calibrated monitors, or has involved a low number of readers. With complex clinical images, how the Grayscale Standard Display Function (GSDF) affects the detection threshold and whether the median background intensity shift has been minimized by GSDF remains unknown. We set out to determine if the median background affected the detection of a low-contrast object in a clustered lumpy background, which simulated a mammography image, and to define the contrast detection threshold for these complex images. Clustered lumpy background images were created of different median intensities and disks of varying contrasts were inserted. A reader study was performed with 17 readers of varying skill level who scored with a five-point confidence scale whether a disk was present. The results were analyzed using reader operating characteristic (ROC) methodology. Contingency tables were used to determine the contrast detection threshold. No statistically significant difference was seen in the area under the ROC curve across all of the backgrounds. Contrast detection fell below 50 % between +3 and +2 gray levels. Our work supports the conclusion that Digital Imaging and Communications in Medicine GSDF calibrated monitors do perceptually linearize detection performance across shifts in median background intensity. The contrast detection threshold was determined to be +3 gray levels above the background for an object of 1° visual angle.

Keywords Image perception · ROC-based analysis · Digital display · Contrast threshold · GSDF

D. L. Leong (✉)
Analogic Corporation,
8 Centennial Drive,
Peabody, MA 01960, USA
e-mail: dleong@analogic.com

D. L. Leong · L. Rainford
Diagnostic Imaging,
UCD School of Medicine and Medical Sciences,
University College Dublin,
Belfield,
Dublin 4, Ireland

T. M. Haygood · G. J. Whitman · S. Carkaci
Department of Diagnostic Radiology,
The University of Texas MD Anderson Cancer Center,
Houston, TX 77030, USA

P. C. Brennan
Brain and Mind Research Institute, Faculty of Health Science,
University of Sydney,
Sydney, NSW 2006, Australia

P. M. Tchou · W. R. Geiser
Department of Imaging Physics,
The University of Texas MD Anderson Cancer Center,
Houston, TX 77030, USA

Introduction

With the widespread use of digital imaging in radiography, it is easy to overlook the important role that the display of the image plays in interpretation. Object contrast is one of the key aspects that affect image interpretation. Since January 1998, with the release of Digital Imaging and Communications in Medicine (DICOM) standard's part 14 on Grayscale Standard Display Function [1], diagnostic display devices have been adjusted to make the display of grayscale medical images perceptually equivalent on different display devices. Perceptually linearizing the display reduces the affect of image brightness on object contrast. The methodology described by DICOM has become the monitor calibration

method recommended by organizations such as the American Association of Physicists in Medicine [2], the American College of Radiology [3], and Integrating the Healthcare Enterprise [4].

The DICOM standard is based on Barten's [5] model of the human visual system. By using Barten's model, a monitor can be calibrated so that the display of the gray levels in an image is closely tied to just noticeable differences (JND) in luminance changes. Each step along the calibrated grayscale changes the display's luminance by approximately the same number of JNDs, making the gray levels perceptually linear. It is also documented that the human visual system adapts to the average luminance of an image and that the JND thresholds vary with the average luminance level [5]. Barten's model assumes that the eye has adapted to the average luminance. There is also a general variation in the JND threshold between individuals. The DICOM standard uses a simplified Barten model that defines a standard observer as a person with good vision between the ages of 20 and 30 years, detecting a standard object. When visualizing objects in an image, the amount of contrast between the object and its background, the object shape and size, and the level of anatomical information all affect detection [6].

Testing by Barten and others [7–10] used a sinusoidal grating pattern on a uniform background to cover the luminance range of the visual system. While a uniform background is necessary for model development, it does not represent clinical images. These papers also used a two-alternative forced-choice signal known exactly (SKE) experimental methodology, removing search from the task even though radiologists use search when interpreting clinical images. Important papers on contrast such as those by Burgess et al. [11, 12] and Peli [13, 14] either used an early form of Barten's work to perceptually linearize the monitors or used uncalibrated monitors. In these experiments, search was also eliminated. Wang et al. [15] investigated the effect of object shape, comparing a disk to a mass-like sphere, and used search, but only considered a uniform background. Now that DICOM-calibrated monitors are the norm for diagnostic imaging interpretation, we wanted to determine whether adaptation to a median background luminance affects the detection threshold when looking at complex images where search is involved. This research aimed to substantially increase the number of observers used in the study over the published literature average of three.

This study had one hypothesis and one goal. The hypothesis was that shifts in the median background intensity would affect object contrast detection threshold in complex images when the object luminance is relatively close to the median luminance. The goal was to determine the detection threshold level with a complex background typical of a screening mammogram.

Materials and Methods

Thirty synthetic clustered lumpy background images, 512× 512 pixel in size, were created using the method developed by Bochud et al. [16] using IDL 7.0 (IIT Visual Solution, Boulder, CO, USA) software. All of the clustered lumpy backgrounds were generated to have an approximate range of pixel values of ±60 gray levels from the median of the image. The range of the pixel values was designed to minimize the effect of local image adaptation yet give the images an appearance reasonably reminiscent of breast parenchyma on a mammogram. Figure 1 shows one of the images used in this study. Table 1 displays the metrics of the initial 30 images used in this study. An attempt was made to minimize the effects of the image brightness range while still maintaining a clinically realistic image. Using Barten's correction factor for surrounding illumination [17], we calculated a maximum correction factor of 1.066, which represents a 6.6 % change in contrast sensitivity for contrast disks. This small error was deemed acceptable for this study. In addition, synthetic images were selected over true clinical images for two reasons. The first reason was to control the gray level range of the image, and the second reason was to reduce any possible memory effect by eliminating any incidental anatomical markers that would assist in recognition [18].

The 30 initial background images were modified to create five images with different median levels for each of the initial backgrounds. This produced a total of 150 images. To vary the median background intensity, an offset was either subtracted or added to the base background to produce the different median levels (Table 2). Table 2 also includes the estimated median luminance of the background

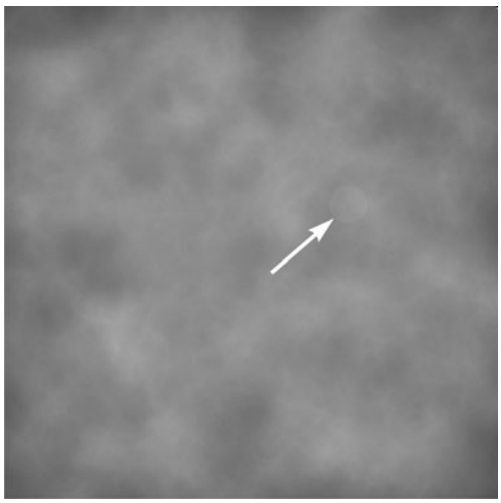

Fig. 1 A sample clustered lumpy background image showing the highest contrast disk that was added to the backgrounds. The left lower edge of the disk is marked with a *white arrow*. As intended, the background variation remains visible through the disk

Table 1 Image parameters

Image number	Mean	Std dev	Skewness	Kurtosis	Median	Min value	Max value
1	87.450	12.329	−0.761	0.408	90	36	120
2	89.424	14.334	−0.105	−0.218	90	34	133
3	87.784	14.275	−0.336	−0.091	90	26	130
4	88.216	13.110	−0.653	0.381	90	36	122
5	89.091	14.276	−0.331	−0.107	90	36	127
6	87.859	13.881	−0.402	−0.395	90	37	121
7	89.043	12.970	−0.608	0.491	90	33	128
8	88.038	14.276	−0.481	−0.410	90	36	124
9	89.213	13.646	−0.209	−0.170	90	41	130
10	87.443	14.104	−0.498	−0.201	90	36	124
11	88.783	13.356	−0.584	0.633	90	32	127
12	88.203	14.064	−0.543	−0.435	90	43	121
13	88.588	13.504	−0.355	−0.207	90	40	126
14	89.714	13.989	−0.190	−0.255	90	39	133
15	87.979	12.865	−0.630	0.468	90	33	124
16	87.815	13.170	−0.620	0.001	90	40	118
17	87.294	12.722	−0.815	0.376	90	37	117
18	87.774	13.674	−0.445	−0.192	90	37	126
19	88.517	13.935	−0.296	0.017	90	38	131
20	88.693	12.732	−0.764	0.931	90	29	121
21	87.728	13.900	−0.520	0.021	90	36	124
22	88.356	13.900	−0.389	0.103	90	41	133
23	88.761	13.584	−0.496	0.167	90	38	126
24	89.843	13.809	−0.346	0.462	90	35	133
25	88.870	13.393	−0.538	0.374	90	32	124
26	88.102	13.077	−0.646	0.552	90	36	124
27	87.887	13.942	−0.440	−0.071	90	36	133
28	90.067	13.629	−0.030	−0.527	90	41	135
29	89.165	13.722	−0.316	−0.106	90	41	131
30	89.497	13.333	−0.151	0.136	90	45	135
Averages	88.507	13.583	−0.450	0.071	90	36.7	126.7
Minimum	87.294	12.329	−0.815	−0.527	90	26	117
Maximum	90.067	14.334	−0.030	0.931	90	45	135

images for each of the five median gray level values. All images were created at eight bits for 256 shades of gray. The dataset was split into two parts. Fifty images were left as created and were used as the normal cases for the reader operating characteristic (ROC) analysis (10 images at five different median intensities). One hundred images had a single contrast disk inserted into the image. The 100 images with the contrast disks were comprised of five disk contrast levels × five median intensities × four replicates. The inserted contrast disk had a fixed radius of 20 pixels and was randomly placed in the image, by adding the desired contrast value, ranging from +1 to +5 gray levels, to the background pixel value at the location of insertion. This was done to maintain the overall background variation and still maintain a specific contrast delta as seen in Figs. 1 and 2. The size of the disk was selected to be approximately 1° of visual angle when viewed at 65 cm on our test monitors.

Table 2 Image parameters

Median background[a]	Estimated median luminance (cd/m^2)	Contrast disk (gray levels)
65	5.2	+1
100	10.6	+2
135	19.2	+3
175	35.0	+4
205	53.2	+5

[a] The median background is the median gray level value for the entire image (without the disk) which has a total range of 0–255

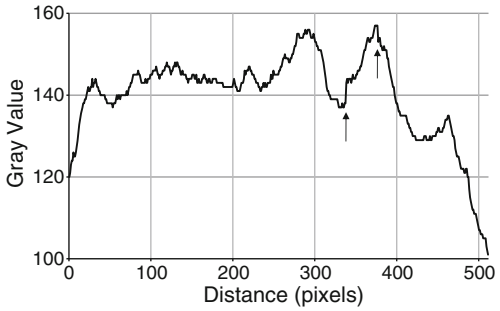

Fig. 2 The plot shows a horizontal line profile of the image shown in Fig. 1 through the center of the contrast disk. The left and the right edges of the disk are shown by the *black arrows*. The background variation remains independent of the disk and is seen as an offset

While it is known that contrast detection is affected by object size [5, 9], it was not practical to have observers run the full range of object sizes needed to get full contrast detail curves. By using moderate-size objects, extrapolation to objects with other sizes could be reasonably made. This experiment is an SKE test because the reader knew the size and the shape but not the location of the object and had to search for the object in the image.

To perform the reader study and to facilitate as many readers as possible, we used eight standard personal computers (PCs) with liquid crystal display (LCD) monitors at three different geographical locations in Sydney, Australia; Houston, Texas; and Boston, Massachusetts. To compensate for the variations existing between monitors in terms of luminance output and grayscale rendition, all of the monitors were calibrated to be compliant with DICOM part 14 using a single VeriLum calibration meter and software (Image Smiths, Bethesda, MD, USA). This resulted in the monitors typically having a calibration maximum luminance of around 100 cd/m^2 and a minimum luminance of around 0.7 cd/m^2. This allows for approximately 1.6 JND steps for each of the gray levels, making each of the possible shades of gray presented in this study perceivable on a uniform background, even with small variations in the calibration from monitor to monitor. The ambient lighting conditions were controlled to be below 20 lux using indirect lighting. An ROC reader study was run using the ViewDEX 1.0 software [19] (Sahlgrenska University Hospital, Göteborg, Sweden) both to display the images and to record user ratings. Readers were given unlimited time to review each image before rendering their decisions in the ViewDex software. Images were displayed in a different random order for each reader to reduce possible reading order bias caused by a regular ordering pattern of background brightness. Window/level and pan/zoom functions in the ViewDEX software were disabled and the images were displayed at a fixed window width of 256 and a level of 127. The resolution of the monitors was 1,280×1,024 and a black surround was used to fill the unused display area, allowing for 1:1 image display. To score the ROC study, a five-point confidence scale was used, which can be seen in Table 3.

Table 3 ROC confidence scale used

Confidence scale	Scores
Very confident object present	5
Confident object present	4
Somewhat confident object present	3
Confident object not present	2
Very confident object not present	1

A total of 17 readers from three institutions were recruited to participate with varying levels of clinical experience from engineers with no experience reading clinical images to attending radiologists. Seven of the readers were practicing radiologists with an average of 13.2 years of experience postresidency (range of 3–35 years). The remaining readers were medical physicists (n=4), radiology technologist lecturers (n=2), engineers (n=2), and senior radiology technologist students (n=2). Readers ranged in age from their early 20s to their early 60s, with nine women and eight men participating. Readers were given the opportunity to take breaks to prevent fatigue, but none chose to take a break. Readers completed the study in an average of 45 min with a range of 30–60 min.

All readers ran a training set of four images before starting the reading to become familiar with the ViewDex software and the general appearance of the clustered lumpy backgrounds and the contrast disks. The training and reading were done in one session. Readers were made aware that contrast disks were randomly located in the image on approximately 66 % of the images. To identify a difference in detection performance based on different median background intensities, the reader data were analyzed using the Trapezoidal/Wilcoxon method in DBM-MRMC [20–26] 2.2 software (University of Chicago, Kurt Rossmann Laboratories for Radiologic Image Research, Chicago, IL, USA), by comparing the area under the ROC curve (AUC). Using JMP®, Version 9 statistical software (SAS Institute Inc., Cary, NC, USA), contingency tables were generated and used to determine the contrast threshold over all median backgrounds and to perform comparisons of reader scores at each background level. Additionally, JMP was used to perform the test for correlation between the median background and the AUC scores with Hoeffding's D method. Hoeffding's D method was chosen for its more generalized test for independence over the more common Spearman's Rho. Power estimates were calculated using SAS 9.1.3 SP4 (SAS Institute Inc.) with scripts developed at, and downloaded from, the Medical Image Perception Laboratory at the University of Iowa [23, 27], using a power of 0.8 and an alpha of 0.05.

Results

The results of the ROC analysis are shown in Fig. 3. Figure 3 is a composite of all of the areas under the ROC curve for each reader and median background level with *box–whisker* plots used to highlight the distribution of the scores. The ANOVA analysis results from the DBM-MRMC software for random cases and random readers, testing for a difference in disk detection between median background levels demonstrated no significance (p value 0.2729). In the test for correlation between median background level and AUC scores, Hoeffding's D showed no statistical significance with a p value of 0.62. Based on the retrospective power calculations (alpha of 0.8) in this study, a difference in AUC value of 0.095 can be detected across the median background levels.

Figure 4 shows a mosaic plot which is a graphical representation of a contingency table of the scores for all readers and all median background levels. In Fig. 5, the reader scores are grouped into "Yes" when the reader scored a disk as being present in the image with a score of 3–5, or "No" for a score of 1 or 2. This allowed better visualization of the contrast detection threshold. In Fig. 6, the mosaic plot is broken down to show how the readers' scores varied around different median background levels for each of the disk contrast levels.

All of the mosaic plots are in the same format. There are two parts of the mosaic plot, the main graph on the left and a summary bar on the right. The summary bar on the right represents the distribution of scores across all columns in the main section. The numbers next to this summary show the reader score and its relation to the color. White represents a score of 5, and black represents a score of 1. As the distribution changes, the different colors are used to determine the scores in each column. The contrast label along the bottom, in the main section of the graph, is the contrast of the disk as compared with the median background of the image. Images with no disk are in the right-most column. The y-axis label on the left is the cumulative percent scale;

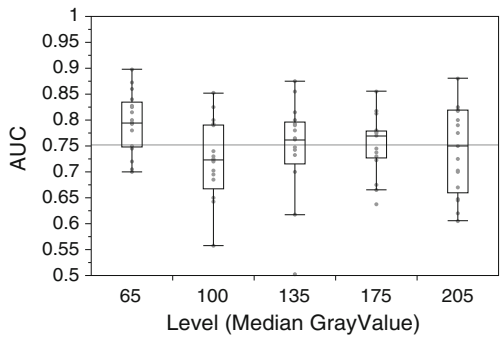

Fig. 3 The graph shows the area under the curve (AUC) for each reader, as a *gray dot*, with a *box–whiskers* plot to show the distribution of the sores at each median background level. The *horizontal line* is the mean of all AUC scores

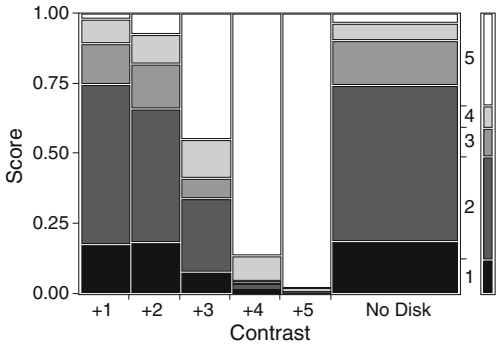

Fig. 4 Mosaic plot showing the proportion of all reader confidence score as the contrast in the disk changed for all background levels. The scores relate to the confidence scale listed in Table 3. *White* indicates a score of 5 (very confident that an object is present) and *black* a score of 1 (very confident that an object is not present). The *y-axis* on the left is the percent distribution of the score. The *right axis* label is the score number

from this scale, the percentage of each score at each contrast level can be determined. The height of each column represents the proportion of the dataset for each of the contrast settings. For example, in Fig. 4 in the +3 column, the dark gray, second from the bottom, represents the percent of score of 2 "Confident Object Not Present." This represents approximately 26 % of the total scores received for this contrast level. Scores of 5 or "very confident" that a disk is present are represented by the white section at the top of the column. This is approximately 45 % of the total scores. If we look at the "No Disk" column and analyze the white, light gray, and medium gray areas, for scores 3–5, we see that these total approximately 25 % of the scores. This means that 25 % of the time a reader scored an image as having a disk when none was present in the image. These are false-positive marks. There were 150 images in the dataset and 50 images were "No Disk," so the width of the right-most column represents one third of the total.

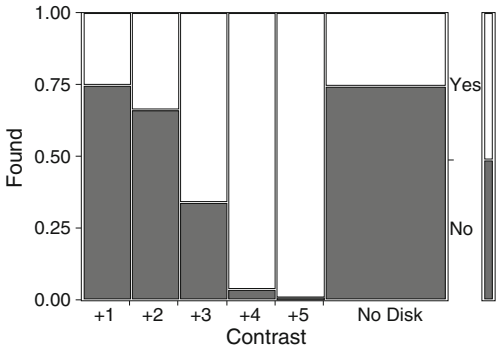

Fig. 5 Mosaic plot showing the proportion of readers who scored a disk as either absent (score of 3–5, *dark gray*) or present (score of 1–2, *white*) as the contrast in the disk changed. The *y-axis* on the left is the percent distribution of the score. The *right axis* is the detection label. As contrast increased, readers were more likely to score correctly those images actually containing disks

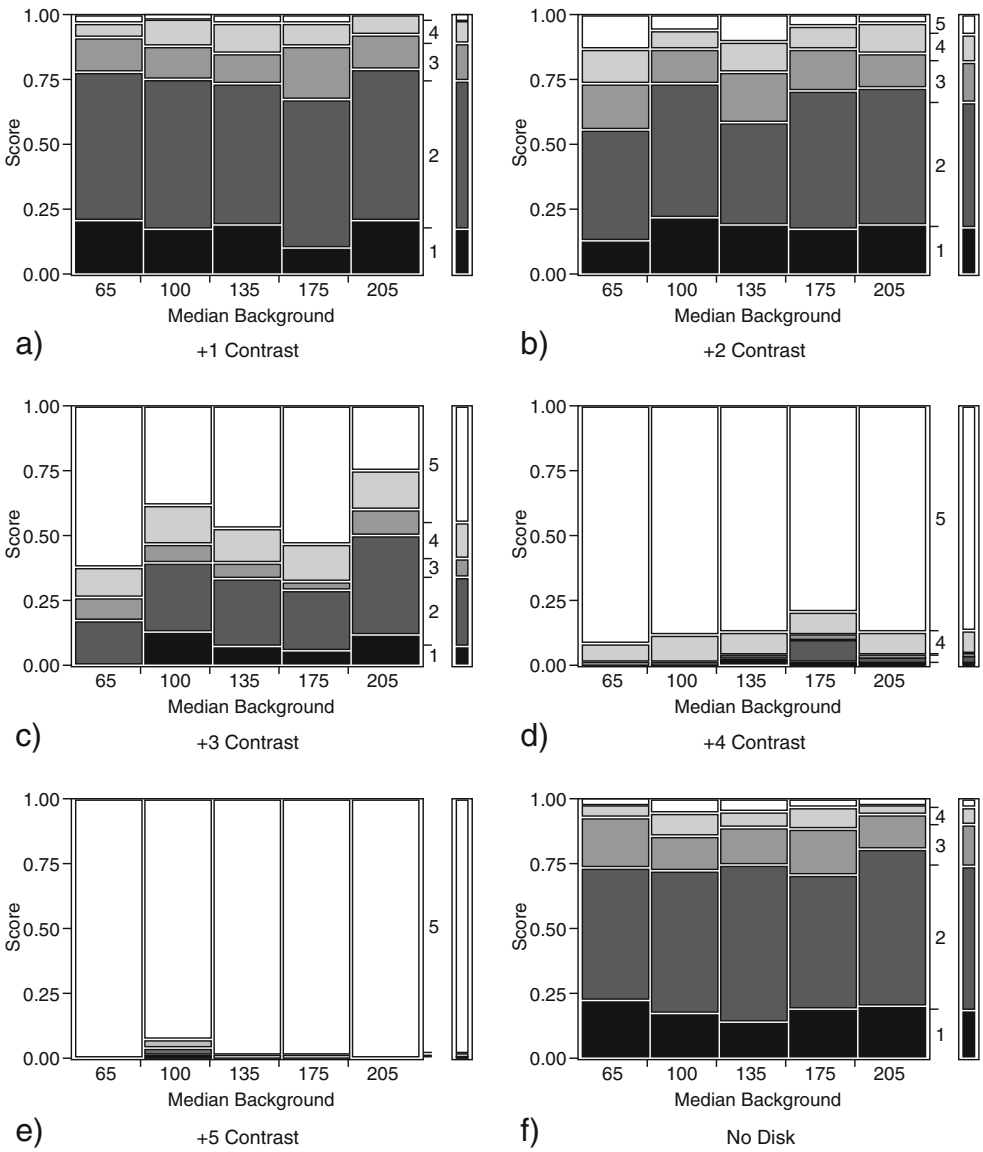

Fig. 6 Mosaic plot showing the reader confidence score as the median background changed for each of the five disk contrasts and for images with no disk. The scores relate to the confidence scale listed in Table 3. *White* indicates a score of 5 (very confident that an object is present) and *black* a score of 1 (very confident that an object is not present). The *y-axis* on the left is the percent distribution of the score. The *right axis* label is the score number

Even though no correlation was found between AUC and median background level, there was a possibility that a pattern would be present in the actual scores across median background level for each disk contrast level. As these were in the contingency tables, Pearson tests were used to check that the distribution of scores was statistically the same for the background levels at each contrast level. The tests for each contrast level produced p values of 0.90, 0.65, 0.01, 0.26, 0.20, 0.26, starting from +1 to +5 and finally "No Disk", respectively. These correspond to the mosaic plots in Fig. 6.

Discussion

With a p value of 0.27, no statistically significant difference in contrast detection across the different median background levels was detected for the population of readers and cases. This study was able to detect changes in AUC down to 0.095. Within the 0.095 difference, the DICOM standard does make monitors perceptually linear and means that, whatever the average luminance of the background, our ability to detect differences in contrast remains the same. This assumes that the illumination in the reading room is optimized.

The contrast detection threshold was found to occur at +3 gray levels (Fig. 5) using a threshold of 50 % to define the cutoff. This contrast level is specific to the objects which are 1° of visual angle in size on complex backgrounds similar to that of a mammogram. Since the JND steps for the 8-bit images was 1.6, on a uniform background, the contrast threshold would be expected to be one as shown by Wang et al. [15] The difference of 3 gray levels or 4.9 JND steps can be attributed to the anatomical type noise in the

background image. It is interesting to note that Wang et al. [15] obtained the same contrast threshold when using a disk that was nodule-like, having a taper to the edges on a uniform background as our simple disk on a complex background.

It is important to note that, for the +1 and No Disk columns, the "Yes" percents are statistically equal. Based on the work of Starr et al. [28], the majority of detections of the +1 contrast disk can be attributed to incorrect localization. This occurred when a disk was truly present and undetected, yet the reader scored a disk as present based on false identification of some part of the background as a disk. Some similar false-positives or incorrect localizations, could occur at all of the other contrast levels as well, but it is reasonable to suppose that these incorrect localizations would become less prevalent as contrast increases because the true disks would become more and more obvious. Since it is not practical to determine the false-positive rates for the other contrasts, as a worst case scenario, we applied the same level of assumed false-positives to the +2 and +3 contrast levels as to the +1 contrast level. True positives were estimated to be actually close to 8 and 40 % for the +2 and +3 contrast levels, respectively, by subtracting 25 % from the yes scores. Even with this correction, the correct scores at +3 are still overall more numerous than the "No" scores (34.12 %), further supporting our determination of +3 being the contrast detection threshold.

As can be seen in Fig. 3, the variation for readers by median background was high. One of the largest differences was for reader 9, whose AUC scores ranged from 0.5 to 0.9 across the different median backgrounds. Since all readers' AUC scores varied randomly across the different median backgrounds, no experimental bias was found. Effort was made to analyze the cause for the high variance and, while a conclusive cause was not identified, it does appear that the false-positive detection rate in the images with no disk (see Fig. 4) may have an effect. The variance also highlights the need to employ a large number of readers in studies of this nature, if power is to be at a reasonably acceptable level. In addition, the high variance calls into question the power achieved in the previous detection threshold work where Burgess [11, 12] and Wang et al. [15] used only three and six readers, respectively. These results may also raise the question as to whether the experience of the reader needs to be standardized, even with tasks that do not require medical, or specifically radiological, expertise. This requires further study.

The data shows that for images without contrast disks, 13 readers scored images as "Confident" or "Very Confident" that a disk was present, with reader 1 scoring 22 out of 50 images as a disk present. A review of reader 1's AUC scores did not show any systematic difference to the other readers. Adding the rating "Somewhat Confident Object Present", all of the readers thought that a disk was present when none was present at least occasionally. The range was from two to 37 images out of 50 without disks. This was a surprising result. This study was designed to create a mammography background pattern that would be different from the disk so that the readers would not detect objects that were not present. Since the high false-positive level may be the cause of the large variance, the scores were broken down by median background level to see if the distributions of scores were statistically equal for all readers (Fig. 6). No statistical differences were seen. During the testing of the score distributions, only the +3 contrast was statistically significantly different across all median background levels with a p value of 0.01 using a Pearson test. This result is understandable in that this was the contrast level where detection changed. As can been seen in the +3 graph of Fig. 6, there is no pattern to the change between median background levels.

An additional method to test the performance of the Grayscale Standard Display Function would have been a test of equivalence. Since this study was primarily focused on the impact of differing median backgrounds on contrast detection with calibrated monitors, there was no control (noncalibrated monitors) to benchmark performance change and this coupled with the high variance, did not facilitate an effective test of equivalence. This issue is currently being addressed.

Conclusions

Comparing the detection threshold between the five median background levels when the disk of +1 to +5 gray levels was within ±60 gray levels of the median, did not result in any statistically significant difference. Within the detection capability of a difference of 0.095 in AUC value, the DICOM Gray Scale Display Function does minimize the effects of median background level on contrast detection for LCD monitors. The detection threshold was identified as being close to +3 gray levels for an object size that is 1° in visual angle and for the 8-bit images included in this study. The high rate of false-positives leads the researchers to the conclusion that a location-based methodology (where the location of the lesion is identified) might be best used for contrast detection threshold experiments. Future work will look at the effect of known location on the contrast detection threshold.

Acknowledgments The authors would like to acknowledge all of the readers who participated in this study for their support of this research.

References

1. National Electrical Manufacturers Association: Digital Imaging and Communications in Medicine (DICOM) Part 14: Grayscale Standard Display Function. National Electrical Manufacturers Association, Rosslyn, 2008

2. Samei E, Badano A, Chakraborty D, Compton K, Cornelius C, Corrigan K, Flynn MJ, Hemminger B, Hangiandreou N, Johnson J, Moxley-Stevens DM, Pavlicek W, Roehrig H, Rutz L, Shepard J, Uzenoff RA, Wang J, Willis CE: Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report. Med Phys 32:1205–1225, 2005
3. American College of Radiology: Practice guideline for digital radiography. Reston: 2007
4. IHE Technical Framework volume I Integration profiles. Chicago: 2007
5. Barten PGJ: Contrast sensitivity of the human eye and its effects on image quality. SPIE Optical Engineering Press, Bellingham, 1999
6. Wang J, Langer S: A brief review of human perception factors in digital displays for picture archiving and communications systems. J Digit Imaging 10:158–168, 1997
7. Legge GE, Foley JM: Contrast masking in human vision. J Opt Soc Am 70:1458–1471, 1980
8. Pelli DG: Uncertainty explains many aspects of visual contrast detection and discrimination. J Opt Soc Am A 2:1508–1532, 1985
9. Rovamo J, Luntinen O, Nasanen R: Modelling the dependence of contrast sensitivity on grating area and spatial frequency. Vision Res 33:2773–2788, 1993
10. Nachmias J, Sansbury RV: Letter: Grating contrast: discrimination may be better than detection. Vision Res 14:1039–1042, 1974
11. Burgess AE, Jacobson FL, Judy PF: Human observer detection experiments with mammograms and power-law noise. Med Phys 28:419–437, 2001
12. Burgess AE, Jacobson F, Judy P: Mass discrimination in mammography: experiments using hybrid images. Acad Radiol 10:1247–1256, 2003
13. Peli E: Suprathreshold contrast perception across differences in mean luminance: effects of stimulus size, dichoptic presentation, and length of adaptation. J Opt Soc Am A Opt Image Sci Vis 12:817–823, 1995
14. Peli E, Arend L, Labianca AT: Contrast perception across changes in luminance and spatial frequency. J Opt Soc Am A Opt Image Sci Vis 13:1953–1959, 1996
15. Wang J, Xu J, Baladandayuthapani V: Contrast sensitivity of digital imaging display systems: contrast threshold dependency on object type and implications for monitor quality assurance and quality control in PACS. Med Phys 36:3682–3692, 2009
16. Bochud F, Abbey C, Eckstein M: Statistical texture synthesis of mammographic images with super-blob lumpy backgrounds. Opt Express 4:33–42, 1999
17. Barten PGJ: Formula for the contrast sensitivity of the human eye. SPIE, San Jose, 2004
18. Ryan JT, Haygood TM, Yamal J-M, Evanoff M, O'Sullivan P, McEntee M, Brennan PC: The "memory effect" for repeated radiological observations. AJR Am J Roentgenol 197:W985–W991, 2011
19. Borjesson S, Hakansson M, Bath M, Kheddache S, Svensson S, Tingberg A, Grahn A, Ruschin M, Hemdal B, Mattsson S, Mansson LG: A software tool for increased efficiency in observer performance studies in radiology. Radiat Prot Dosimetry 114:45–52, 2005
20. Dorfman DD, Berbaum KS, Lenth RV, Chen YF, Donaghy BA: Monte Carlo validation of a multireader method for receiver operating characteristic discrete rating data: factorial experimental design. Acad Radiol 5:591–602, 1998
21. Dorfman DD, Berbaum KS, Metz CE: Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method. Invest Radiol 27:723–731, 1992
22. Hillis SL: A comparison of denominator degrees of freedom methods for multiple observer ROC analysis. Stat Med 26:596–619, 2007
23. Hillis SL, Berbaum KS: Power estimation for the Dorfman–Berbaum–Metz method. Acad Radiol 11:1260–1273, 2004
24. Hillis SL, Berbaum KS: Monte Carlo validation of the Dorfman–Berbaum–Metz method using normalized pseudovalues and less data-based model simplification. Acad Radiol 12:1534–1541, 2005
25. Hillis SL, Berbaum KS, Metz CE: Recent developments in the Dorfman–Berbaum–Metz procedure for multireader ROC study analysis. Acad Radiol 15:647–661, 2008
26. Hillis SL, Obuchowski NA, Schartz KM, Berbaum KS: A comparison of the Dorfman–Berbaum–Metz and Obuchowski–Rockette methods for receiver operating characteristic (ROC) data. Stat Med 24:1579–1607, 2005
27. Medical Image Perception Laboratory. Available at http://perception.radiology.uiowa.edu/. Accessed July 27, 2010.
28. Starr SJ, Metz CE, Lusted LB, Goodenough DJ: Visual detection and localization of radiographic images. Radiology 116:533–538, 1975