International Journal of Engineering & Technology, 7 (2.33) (2018) 1160-1163

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Develop Automation Program of Quality Control Standard Phantom of Mammography using Digital Image Processing Technology

Mi-Hwa Lee[1], Hong-Ryang Jung[2*]

[1] Department of Health Care, Hanseo University, Seosan-si, CO 31962, Korea
[2] Department of Health Care, Hanseo University, Seosan-si, CO 31962, Korea
**Corresponding author Email: hrjung@hanseo.ac.kr*

Abstract:

The quality control using phantom in mammography is visual inspection so that the deviation can be occurred depending on the evaluator's subjective decisions, automated detection program was developed to minimize the errors. The study subjects were 165 phantom images in 74 hospitals nationwide that passed the test of the special medical equipment by Korean Institute for Accreditation of Medical Image from May 2015 to April 2016. We analyzed the intra-rater reliability and inter-rater reliability. In the analysis outcomes to analyze reproducibility of the visual inspection, inter-rater reliability ICC was shown to be low (Fiber=0.349, Specks=0.265, Masses=0.212, Total lesions=0.378) while intra-rater reliability mean ICC was to be approximately 60%. Upon the comparison results for test objects applying the existing visual inspection method and the automated detection program based on the developed image processing technique, the latter method was evaluated to be higher utility as the new analytical method for the standardization of standard phantom quality control, detecting 0.38 unit of fiber and one unit of mass. The automated detection program can overcome the biases by manual processes, and improve the reproducibility so as to automate the quality control system consistently, simply, and correctly within the short time.

Keyword: *Mammography, Phantom radiography, Quality control of image, Visual inspection, Quantitative analysis, Automation*

1. Introduction

Mammography images are stored in Picture Archiving and Communications System (PACS). After passing the era of analog images of film systems, it is changing into digital images(Kang Yeo Dong, 2013). Mammography images with patient information and various image informations are stored in international standard Digital Imaging and Communications in Medicine (DICOM) files for transfer and storage. Header information section of DICOM files have information about the size of each pixel that compose the image as well as grayscale that provides useful information for digital image processing(X. Q. Zhou, et al., 2001).

While all other items (personnel inspection, quality control inspection, phantom imaging and clinical imaging inspection) of mammography equipment among quality control inspection are digitalized and objective evaluation is possible, items that must depend on subjective evaluation by humans are phantom image inspection and clinical image inspection. In phantom image inspection, inspectors (radiology specialists of Korean Institute for Accreditation of Medical Image) need to visually observe a total of 10 or more test objects, including 4 or more fibers, 3 or more specks, and 3 or more masses among 16 test objects (6 fibers, 5 specks, and 5 masses) within phantom image before the equipment can receive usage approval. However, there could be deviations due to subjective elements of inspectors. Among inspectors, the rate of replication for measurement has been low. Therefore, the objective of this study was to develop an automated detection program to minimize such error. The automated quantitative analysis method was compared to the manual inspection method. In addition, a new quantitative analysis index was created.

2. Research Methods

2.1. Research Phantom

For mammography, American College of Radiology (ACR) certified phantom Nuclear Associates Model 18-220, RMI Model 156, CIRS Model 015 or phantoms with identical certifications are used. Five are within the phantom was composed of nine and fibers. Microcalcification was simulated using Al_2O_3. Masses were simulated using a lens-shaped mass, shown in (Figure 1).

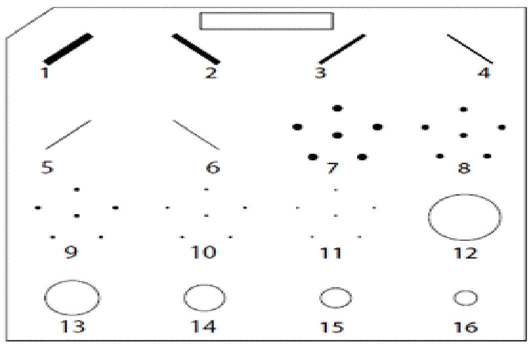

Fig. 1:. Location of the test objects in the wax insert.

Test objects were composed of fibers in six locations (from 1 to 6) with diameters of 1.56 mm, 1.12 mm, 0.89 mm, 0.75 mm, 0.54 mm, and 0.40 mm, respectively. Specks were in five locations (from 7 to 11) with diameters of 1.56 mm, 1.12 mm, 0.89 mm, 0.75 mm, 0.54 mm, and 0.40 mm, respectively. Masses were in five locations (from 12 to 16) with diameters of 2.00 mm, 1.00 mm, 0.75 mm, 0.50 mm, and 0.25 mm, respectively(Handrick R. E. et al., 1999), (Andrew P. and Smith , 2003) shown in (Table I).

Table I : Test objects sizescontained within the mammography accreditation phantom

	Fibers (diameter, mm)	Specks (diameter, mm)	Masses (thickness, mm)
1	1.56	0.54	2.00
2	1.12	0.40	1.00
3	0.89	0.32	0.75
4	0.75	0.24	0.50
5	0.54	0.16	0.25
6	0.40	-	-

2.2. Phantom Image Acquisition and Data Collection Methods

By obtaining image with ACR phantom, images were transferred to PACS and stored as DICOM files, shown in (Figure 2).

Fig. 2. Image acquisition.

According to the decision of the Minister of Health and Welfare, hospitals should conduct phantom image evaluation every six months. Because they must conduct document inspection for Korean Institute for Accreditation of Medical Image as a registered quality management inspection institution, visual (qualitative) evaluation figures of the phantom and images are usually obtained by a radiology specialist. The study subjects were 165 phantom images in 74 hospitals nationwide that passed the test of the special medical equipment required by Korean Institute for Accreditation of Medical Image from May 2015 to April 2016.

2.3. Automated Quantitative Analysis Program Development

Using Matlab R2010a (The Mathworks Inc., Natick, MA, USA), a new automated quantitative analysis program was developed. After obtaining the original image, original ROI image was extracted. The shape of the image to be analyzed was placed in 4 phantom location settings rotated by 90° each. It was always fixed as an image with arrangement angle of 0°. By obtaining negative image using brightness invert function, the entire matrix was divided into 16 pixel coordinates to separate areas of test objects arranged 4 × 4. It was divided into 3 object areas (fiber, specks, masses). After applying normalization technique after designating test object area, binary image was created through applying median filtering technique and adaptive threshold technique. By applying the labeling technique, automated quantitative index was analyzed.

2.4 Assessment Methods

2.4.1. Inter-Rater Reliability

Blind experiment was conducted by 10 radiology specialists to cross evaluate these images. To familiarize the phantom image analysis manual, training was conducted for 30 minutes to evaluators before the evaluation. Training was conducted by evaluating ideal valuation standards. A total of 11 evaluations per image were conducted, including 1 evaluation based on Korean Institute for Accreditation of Medical Image submission and evaluations by 10 evaluators. Fiber, Specks, Masses, and Total lesions were evaluated to calculate ICC (intra-class correlation coefficient). For ICC value, 0.80 to 1 indicates excellent reliability, 0.60 to 0.79 represents acceptable reliability, and under 0.59 represents poor reliability(Streiner D. L, and Norman G. R, 2003).

2.4.2. Intra-Rater Reliability

After conducting primary evaluation by 10 radiology specialists, secondary evaluation was conducted with a four-week interval. To measure reliability within evaluators, the first evaluation must be erased from memory. To remove this memory effect, at least four weeks' interval is needed before evaluation6. Mean ICC was calculated to analyze the reliability between evaluation and reevaluation.

2.4.3. Comparison of Qualitative Analysis and Quantitative Analysis

To provide descriptive statistics on the difference between test objects detected by the automation program and by visual inspections for fiber, specks, and masses, mean values and standard deviations were calculated to determine if there were differences between the two methods. Paired t-test was conducted for verification. Analysis of data was done using SPSS 21.0 (Statistical Package for Social Science for Window TM release 21.0 SPSS Inc, Chicago, U.S.A.). Statistical significance was considered when P value was less than 0.05.

3. Results

3.1. Inter-Rater Reliability

The ICC values for fiber, specks, masses, and total lesions were 0.349 (95% CI: 0.287-0.419), 0.265 (95% CI: 0.209-0.331), 0.212 (95% CI: 0.162-0.274), and 0.378 (95% CI: 0.316 -0.449), respectively. Thus, visual evaluation of test objects among the 11 evaluators was not reliable, shown in (Table II).

Table Ii: Inter-rater reliability (n=165)

lesions	ICC	95% Confidence Interval	P
Fiber	0.349	0.287-0.419	<0.001
Specks	0.265	0.209-0.331	<0.001
Masses	0.212	0.162-0.274	<0.001
Total lesions	0.378	0.316-0.449	<0.001

*ICC: Intra-class Correlation Coefficient

For visual evaluation of each test object by the 11 evaluators, although the same images were tested, there were various variations (fiber: 3.83~4.67, specks: 2.90~3.78, masses: 3.25~4.09, and total lesions: 10.45~12.11), shown in (Table III).

Table Iii : Visual assessment of imitation lesions of the rater (n=165)

	Fiber		Specks		Masses		Total lesions	
	Mean	±SD	Mean	±SD	Mean	±SD	Mean	±SD
1	4.67	± 0.49	3.72	± 0.44	3.74	± 0.44	12.04	± 1.09
2	4.59	± 0.56	3.62	± 0.46	3.59	± 0.48	11.79	± 1.27
3	4.61	± 0.54	3.42	± 0.47	3.25	± 0.45	11.28	± 1.13
4	4.27	± 0.36	3.56	± 0.39	3.70	± 0.34	11.52	± 0.80
5	4.46	± 0.54	3.16	± 0.42	4.09	± 0.41	11.70	± 1.09
6	4.16	± 0.27	3.39	± 0.41	3.89	± 0.42	11.43	± 0.68
7	4.57	± 0.49	3.01	± 0.53	3.78	± 0.42	12.11	± 0.95
8	3.85	± 0.56	2.90	± 0.53	3.78	± 0.53	10.45	± 1.64
9	3.83	± 0.62	3.78	± 0.36	3.89	± 0.47	10.65	± 1.52
10	3.98	± 0.72	3.00	± 0.46	3.99	± 0.59	10.85	± 1.61
11	4.30	± 0.51	3.36	± 0.45	3.77	± 0.46	11.38	± 1.18

3.2. Intra-Rater Reliability

Between evaluation and re-evaluation, the mean ICC for fiber was 0.693 (min.: 0.568, max.: 0.919) and mean ICC for specks was 0.655 (min.: 0.339, max.: 0.916). Mean ICC for masses was 0.667 (min.: 0.415, max.: 0.840) and mean ICC for total lesions was 0.684 (min.: 0.482, max.: 0.881). The average degree of reliability for all test objects was statistically significant. Thus, visual evaluations of test objects between evaluation and reevaluation by the 10 evaluators had medium degree of reliability, shown in (Table IV).

Table iv : intra-rater reliability (n=165)

lesions	rater	ICC	mean ICC
Fiber	1	0.568[a]	0.693
	2	0.729	
	3	0.636	
	4	0.570	
	5	0.785	
	6	0.635	
	7	0.678	
	8	0.919[b]	
	9	0.659	
	10	0.746	
Specks	1	0.814	0.655
	2	0.740	
	3	0.489	
	4	0.446	
	5	0.339[a]	
	6	0.768	
	7	0.846	
	8	0.916[b]	
	9	0.553	
	10	0.642	
Masses	1	0.611	0.667
	2	0.690	
	3	0.829	
	4	0.829	
	5	0.796	
	6	0.522	
	7	0.415[a]	
	8	0.840[b]	
	9	0.555	
	10	0.584	
Total lesions	1	0.840	0.684
	2	0.810	
	3	0.618	
	4	0.680	
	5	0.482[a]	
	6	0.541	
	7	0.511	
	8	0.881[b]	
	9	0.692	
	10	0.787	

* ICC : Intra-class Correlation Coefficient
a. lowest
b. best

3.3. Comparison of Qualitative Analysis and Quantitative Analysis

For fiber, the automated detection constitutive analysis method showed 0.83 more superior result compared to the subjective evaluation method. For specks, there was no statistically significant difference between the two methods. For masses, the quantitative analysis method showed around one more superior result compared to the subjective evaluation method.

Compared to the subjective visual evaluation method, the quantitative automatic method was superior. There was statistically significant difference between the two methods, shown in (Table V).

Table V : Comparison of visual evaluation and proposed algorithm

Variable	Visual evaluation	Proposed algorithm	t	p*
Fiber	4.59 ± 0.55	4.97 ± 0.71	-2.61	0.013
Specks	3.56 ± 0.50	3.56 ± 0.56	0.00	1.000
Masses	3.79 ± 0.41	4.79 ± 0.41	-9.67	<0.001

* by paired t-test to p<0.05.

4. Discussion and Conclusion

In case of specks, it has a dot in the middle. It is surrounded by six dots in a star-shape which is difficult to distinguish from artifact. The brains of human can recognize such test objects even though they cannot classify between noise and test object by predicting the star shape. This is the same for fiber and masses. These are the disadvantages of visual evaluation.

According to a research by Park(D. S. Park et al., 2011) the average numbers of recognizable fibers, specks, and masses in phantom images are 4.0 ± 0.5 (range, 2.7 ~ 5.0), 3.0 ± 0.3 (range, 2.0 ~ 3.7), and 3.5 ± 0.4 (range, 2.3 ~ 4.0), respectively. However, in this study, the range of difference in recognition was narrower. In this study, the numbers of fibers, specks, and masses were 3.83 ~ 4.67, 2.90 ~ 3.78, and 3.25 ~ 4.09, respectively. Although the range of difference in recognition was narrower in this study, satisfactory result between evaluators or within the evaluators was not obtained. This could be due to subjective application of each different standard and differences in innate visual ability. Therefore, Park(K. J. Park, 2004) has suggested a measurement standard for software that can evaluate resolution by visualizing optical density of mammography and the necessity for the development of a quantitative program to replace the visual evaluation.

Based on the comparison results for test objects using the existing visual inspection method and the automated detection program developed in this study, the latter method was found to have higher reliability as a new analytical method for the standardization of standard phantom quality control. It can detecting 0.38 unit more of fiber (4.97 - 4.59) and one unit more of mass (4.79 - 3.79).

According to studies by Larissa Cristina dos Santos Romualdo(Larissa Cristina dos Santos Romualdo and Marcelo Andrade da Costa Vieira 2009) Douglas(Douglas E and Kurvilla Verhse 2000), and Hunt(R. A. Hunt et al., 2005) microcalcification with diameters between 0.2 mm to 0.5 mm can be used to diagnose breast cancer early. For this reason, the detection ability of the visual inspection method for specks was found to be better than that for fiber or masses. However, there is no need to decrease the value of microcalcification diameter due to the lower evaluation reliability of the visual evaluation method. Furthermore, the automatic program has similar or better evaluation results with superior detection capacity.

In summary, the automated detection program developed in this study can overcome the biases caused by manual processes among inspectors and within inspection committee members. It can improve the reproducibility and automate the quality control system consistently, simply, and correctly within a short time period compared to the existing visual inspection method. Therefore, the automated detection method is expected to improve the efficiency of quality control for phantom in mammography by multiple inspectors. In addition, the convenience of its training might enhance human resource management for special medical equipment managers.

Acknowledgment

This study was funded by a 2016 Academic Research Support Project of Hanseo University, Republic of Korea.

References

[1] Andrew P, and Smith Ph. D, “Fundamentals of Digital Mammography, Physics, Technology and Practical Considerations”, Radiology management, published in United States. Vol.25, pp.18-24, 26-31, 2003.
[2] Douglas E, and Kurvilla Verhse, "Digital mammography images simulation using Monte Carlo", Med. Phys, Vol.27, pp.568-579, 2000.
[3] D. S. Park, H. H. Won, W. S. Lee, “Educational Evaluation”, Muneumsa, 2011.
[4] Handrick R. E, Bassett L, Botsco M. A, et al. “Mammography Quality control Manual”, Reston (VA) American College of Radiology, 1999.
[5] Kang Yeo Dong, “Image Comparison Analysis of Film-Screen Methods Which Use Mammo Phantom and DR Methods, Journal of advanced engineering and technology”, 6 (1), pp. 43-47, 2013.
[6] K. J. Park, "Study of Image Quality Control and Evaluation of the Glandular Exposure Dose for the Low Energy X-ray using the Phantom of Breast Tissue Equivalent", Graduate School, Kyonggi University, 2004.
[7] Larissa Cristina dos Santos Romualdo, Marcelo Andrade da Costa Vieira, Homero Schiabel, "Mammography Images Restoration by Quantum Noise Reduction and Inverse MTF Filtering", Sibgrapi, pp.180-185, 2009.
[8] R. A. Hunt, D. R. Dance et al, “Monte Carlo simulation of a mammographic test phantom”, Radiation Protection Dosimetry, Vol.114, pp.432-435, 2005.
[9] Streiner D. L, Norman G. R, “Health Measurement Scales : A Practical Guide to Their Development and Use", New York : Oxford University Press, 2003..
[10] X. Q. Zhou et al, “Authenticity and Integrity of Digital Mammography Images”, *IEEE Transactions on medical Imaging*, Johns Hopkins University, 20(8), August, 2001.