Automated Quality Assurance Applied to Mammographic Imaging

Lilian Blot

School of Information Systems, University of East Anglia, Norwich NR4-7TJ, UK
Email: lb@sys.uea.ac.uk

Anne Davis

Portsmouth Hospitals NHS Trust PO3 6AD, Portsmouth, UK
Email: Anne.DavisMedPhysics@porthosp.nhs.uk

Mike Holubinka

Portsmouth Hospitals NHS Trust PO3 6AD, Portsmouth, UK
Email: mike.holubinka@porthosp.nhs.uk

Robert Martí

School of Information Systems, University of East Anglia, Norwich NR4-7TJ, UK
Email: marti@sys.uae.ac.uk

Reyer Zwiggelaar

School of Information Systems, University of East Anglia, Norwich NR4-7TJ, UK
Email: rz@sys.uae.ac.uk

Quality control in mammography is based upon subjective interpretation of the image quality of a test phantom. In order to suppress subjectivity due to the human observer, automated computer analysis of the Leeds TOR(MAM) test phantom is investigated. Texture analysis via grey-level co-occurrence matrices is used to detect structures in the test object. Scoring of the substructures in the phantom is based on grey-level differences between regions and information from grey-level co-occurrence matrices. The results from scoring groups of particles within the phantom are presented.

Keywords and phrases: automatic quality control, mammographic images, grey-level co-occurrence matrices, image segmentation.

1. INTRODUCTION

Meticulous quality control of mammography was advocated [1] as a mandatory requirement of the National Health Service Breast Screening Programme prior to the establishment of the NHSBSP. The quality control is with respect to evaluating system performance and for long term monitoring. Standards of acceptability have been established based upon subjective interpretation of image quality test phantoms, for example, Leeds test objects TOR(MAM), TOR(MAX), or TOR(MAS). Leeds TOR(MAM) is widely used for this purpose, and has been reported as being among the most consistent and sensitive [2] of those available commercially.

This study is aimed at applying image analysis methods that will be employed for the analysis of phantom films (which were digitised) for quantitative objective assessment of different imaging systems, monitoring temporal performance, and evaluating the impact of introducing new imaging systems, for example, direct digital mammography.

Analysis of phantom images relies upon scoring by experienced observers according to the visibility of details. Scoring schemes have been proposed whereby the observer allocates a value according to visibility of the test details (i.e., 0 = detail not seen, 1 = barely visible/threshold, 2 = less visible/faint, 3 = detail easily seen). Experienced observers have shown good self-consistency but have yielded substantially different scores when scoring images from the same phantom and when using the same protocol [3]. These interobserver variations clearly have the potential to call into question the
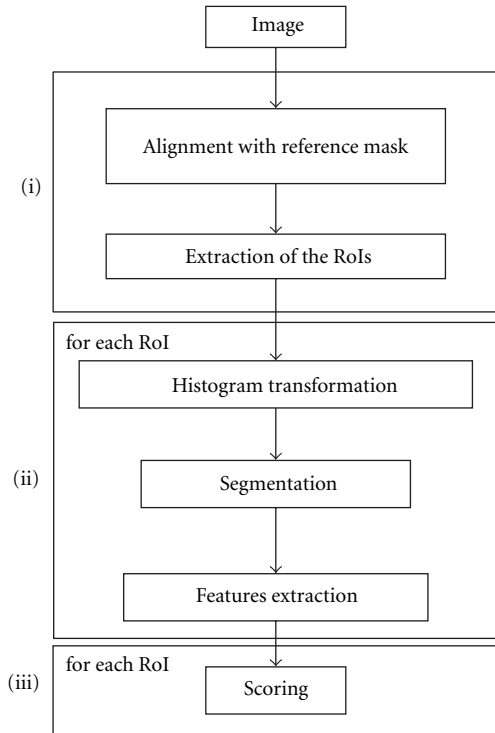
FIGURE 1: Overview of the method.

FIGURE 2: Leeds TOR(MAM) test object.

validity of conclusions drawn from such image quality measurements.

Automated computer analysis has been applied to the evaluation of image quality test phantom images and has yielded results comparable to, but more consistent than, the human observer [4, 5, 6]. Computer-assisted analysis also has some role in providing decision support to those analysing analogue phantom images, in the training of inexperienced observers and in the optimisation process.

This preliminary study is focused on the detection and scoring of groups of microparticles covering sizes from 63 to 354 microns (see also Table 1). The microparticle groups are one of the three principal features of the TOR(MAM) test object (see Figure 2). An expert viewed the films in a darkened room and made a visual assessment of the films using the 0-1-2-3 scoring scheme. The computerized automated method involves three principal stages, each of which is described in detail. Figure 1 depicts the various stages of the method. Briefly, they are the following:

(i) alignment of the phantom with a reference image and extraction of six regions of interest (RoI) around each microparticle group (see Section 2.1);

(ii) extraction of the features from each RoI (see Section 2.6). This is based on grey-level co-occurrence matrices (GLCMs) and a distance measure to compare GLCM's is introduced and described in Section 2.3;

(iii) scoring the detected features using a statistical classification method (k-nearest neighbour), which is described in Section 2.7.
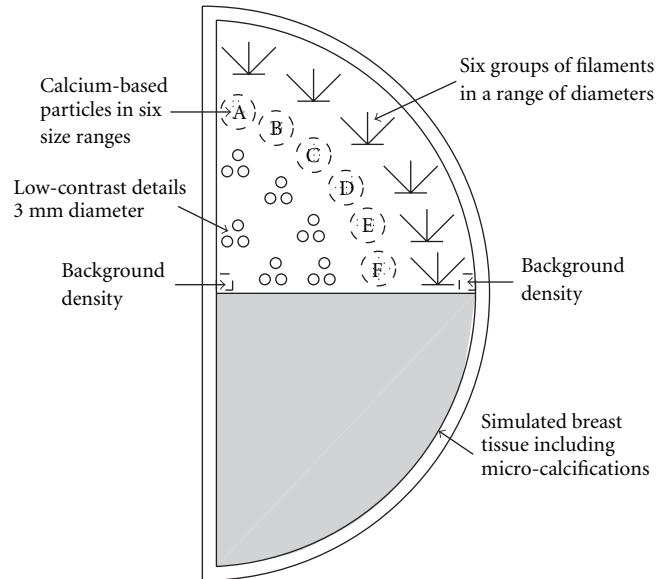
The results obtained by the expert are compared with those obtained by the application of the automated detection method, as is the self-consistency of the expert. This is discussed in Section 3.

2. MATERIALS AND METHODS

This preliminary study is based on a test set of images taken from 16 films of a Leeds TOR(MAM) test object (for more detail see [7]). These films have been scanned with a Lumisys Lumiscan 85, at a colour depth of 12 bits per pixel and a pixel size of $50 \mu m$.

A diagram of the phantom is shown in Figure 2. Three different types of structures are present in the phantom, these are groups of microparticles, radiating patterns of linear structures and groups of three disks of diminishing contrast. The groups of microparticles simulate mammographic microcalcification clusters. For each of the six groups, there is an upper and lower bound to the size of the microparticles and it should be noted that there is some overlap in the size range for subsequent groups (see Table 1); this is a feature of the phantom indicating a continuous transition between the groups. The linear structures simulate fibrous filament details. Again, there are six of these groups (of filaments) and the variation is in the diameter of the filaments. The groups containing three disks simulate lesion-like structures. The variation of contrast is used as an indication for their detectability. In this study, we concentrate on the automatic detection and classification of the six groups of particles (labeled A–F in Figure 2). From the phantom image, six RoIs centred around each group of particles are extracted.

Ninety-six regions are obtained and constitute our dataset. For the set of 16 phantom images, Table 2 shows the distribution of scores (i.e., 0 = detail not seen, 1 = barely visible/threshold, 2 = less visible/faint, 3 = detail easily seen) for

TABLE 1: Particles size range for the six groups.

| | | Group | | | | | |
		A	B	C	D	E	F
Particle size range	maximum size (μm)	354	283	226	177	141	106
	minimum size (μm)	224	180	150	106	90	63
	maximum size (*pixels*)	7	6	5	4	3	2
	minimum size (*pixels*)	5	4	3	2	2	1

TABLE 2: Description of the dataset for observation 1.

| | | Group of particles | | | | | |
		A	B	C	D	E	F
Scores	3	16	16	4	0	0	0
	2	0	0	12	6	0	0
	1	0	0	0	9	2	0
	0	0	0	0	1	14	16

the various group of particles. The scoring of each group of particles refers to a scoring process of one expert (referred to as observation 1), which will be considered as ground truth in this study. A previous score was completed by the same expert one year earlier and will be referred to as observation 2. This provided information about the intravariability of the observer (see Section 3). This indicates that, for our dataset, the details for the groups A and B are clearly visible, and that for the groups F no details can be seen.

2.1. Automated extraction of the region of interest

Automated evaluation of the phantom requires that the computer correctly identifies the individual details. The physical alignment of the test object within the imaged area and the location of the digitised data within digitised data files varies from image to image.

We restrict the transformation to be linear because the phantom used is a rigid plastic structure, which does not allow any local deformation (i.e., no bending nor compression). To extract the RoI automatically for the whole dataset, one phantom image was manually annotated with the positions of the RoI. Rigid registration [8] was used to globally align all phantom images in the dataset. Subsequently, all the RoI are extracted based on their position in the reference image. To remove local mis-alignment, the individual RoI are registered with the relevant RoI from the reference image.

2.2. Fundamentals

In our method, we use GLCMs to extract textural information from the image and segment particles. In order to reduce the size of the GLCMs, a linear transformation of the RoI's histogram is applied. The resulting images have a colour depth of sixty-four grey levels.

In an image, let N_g denote the set of grey levels, $|N_g|$ the number of grey-levels, and $\vec{t}$ a translation. The grey-level co-occurrence matrix [9] of the image based on the translation $\vec{t}$ is an $|N_g| \times |N_g|$ matrix where an entry $m(i, j)$ denotes the number of pixel pairs, separated by a translation $\vec{t}$, which have grey-level values i and j. The GLCM is normalised by dividing each entry by the number of pairs of pixels separated by the translation $\vec{t}$.

In principle we want to compare GLCMs and to be able to do so, we need to introduce a distance in the GLCM-space. Our approach is based on two assumptions:

(1) there is a statistical difference between the GLCMs for image regions which contain only background texture and regions which also include image structures (e.g., particles);

(2) regions with image structures are mostly surrounded by regions without any structures. This condition is essential for the approximation of the background. This assumption is correct for the phantom images.

An example of our approach can be found in Figure 3, where Figures 3a and 3b show the same image, except that in Figure 3b a cluster of particles has been added. For both images, a GLCM can be obtained and in this ideal case, the difference between the two GLCMs (see Figure 3c) is caused by the blob-like structures in Figure 3b. This difference GLCM can be seen as being caused by perturbations of the normal background texture and hence, can be used to reconstruct a perturbation image of the image structures, the results of which are shown in Figure 3d. As can be seen, most of the blob-like structures have been enhanced.

2.3. Distance metric in the GLCM-space

An $m \times n$ matrix can be considered as a vector of dimension mn. As GLCMs have entries values restricted to $[0, 1]$ an $n \times n$ GLCM can be seen as a vector of $[0, 1]^{n^2}$.

From now on, a GLCM $M_{|N_g|, |N_g|} = [m_{i,j}]_{(i,j) \in N_g^2}$ will be referred to as a vector $X = [x_k]_{k \in \Lambda}$ where $\Lambda = \text{ind}(N_g^2)$, $x_{\text{ind}(i,j)} = m_{i,j}$, and ind is bijective and defined as

$$\text{ind} : \mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N},$$
$$(i, j) \longmapsto \text{ind} F(i, j) = k. \tag{1}$$

To compare the similarity of two GLCMs, we introduce the distance δ defined as

$$\delta : [0, 1]^N \times [0, 1]^N \longrightarrow \mathbb{R}_+,$$
$$(X, Y) \longrightarrow \delta(X, Y) = \sum_{i=1}^{N} \Delta_i(X, Y), \tag{2}$$

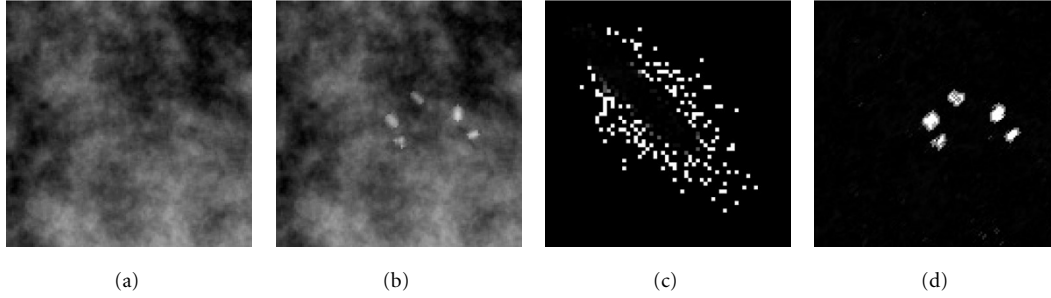(a)   (b)   (c)   (d)

FIGURE 3: Example images of the described approach.

where $X = [x_1, \ldots, x_N]$, $Y = [y_1, \ldots, y_N]$, and

$$\Delta_i(X, Y) = \begin{cases} 0 & \text{if } x_i = y_i = 0, \\ \dfrac{|x_i - y_i|}{|x_i + y_i|} & \text{otherwise.} \end{cases} \quad (3)$$

It should be noted that δ is not a distance on $\mathbb{R}^N$ (see Appendix A for a formal proof that δ is a distance for the normalised GLCM space).

This distance has been used instead of the usual Euclidean distance because it takes into account the relative value of an entry. If we consider a small perturbation ε and three vectors X, Y, and Z where

$$\begin{aligned} y_k &= \begin{cases} x_k & \text{if } k \neq i, \\ x_i + \varepsilon & \text{if } k = i; \end{cases} \\ z_k &= \begin{cases} x_k & \text{if } k \neq j, \\ x_j + \varepsilon & \text{if } k = j, \end{cases} \end{aligned} \quad (4)$$

with $i \neq j$, $x_i \gg x_j$. We obtain, on the one hand, $d(X, Y) = d(X, Z)$—where $d(X, Y)$ is the Euclidean distance—but on the other hand, $\delta(X, Y) \ll \delta(X, Z)$. Using the distance δ we emphasise that a small variation on a high value in a GLCM is negligible compared to the same variation on a smaller GLCM value.

2.4. Background texture approximation

To approximate the textural properties for a RoI we need to compare the GLCM of the RoI with the GLCM of the local neighbourhood. Nine subregions of the image are used (see Figure 4), 8 regions constitute the local neighbourhood (white squares) and are centred around the region of interest (grey square). There is no overlap between the subregions. The approximation $\overline{X}$ of the GLCM for the neighbourhood is given by

$$\overline{X} = \frac{1}{8} \sum_{k=1}^{8} X^k = [\overline{x_i}]_{i \in \Lambda}, \qquad \overline{x_i} = \frac{1}{8} \sum_{k=1}^{8} x_i^k, \quad (5)$$

where $X^k = [x_i^k]_{i \in \Lambda}$ is the GLCM from the kth square region of the neighbourhood.

Region of interest
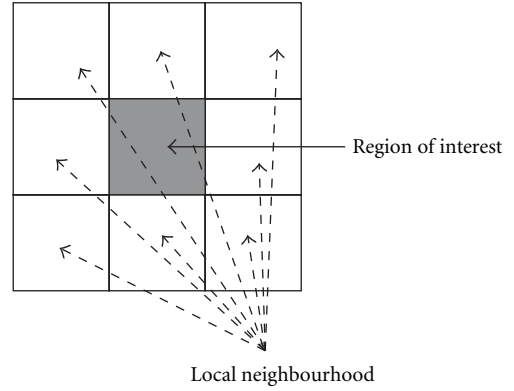
Local neighbourhood

FIGURE 4: Local decomposition of the image.

2.5. Pixel classification

A GLCM conforms to the definition of compositional data [10] when considering the following properties of a GLCM $X = [x_i]_{i \in \Lambda}$:

(1) for all $i \in \Lambda, 0 \leq x_i \leq 1$,
(2) $\sum_{i \in \Lambda} x_i = 1$.

This means that a variation of one entry in the co-occurrence matrix influences others entries to conserve the second property. When we increase (resp., decrease) an entry by ε, we must decrease (resp., increase), the other entries to keep the sum equal to one. All this the individual entries must obey the first property.

We want to investigate the perturbation caused by a positive variation ε on one component of a GLCM. To do so, consider two GLCMs, $X = [x_i]_{i \in \Lambda}$ and $X_\varepsilon^k = [x_i^k]_{i \in \Lambda}$, with X_ε^k given by

$$x_i^k = \begin{cases} 0 & \text{if } x_i = 0 \text{ and } i \neq k, \\ x_i - \dfrac{\varepsilon}{|\Theta - \{k\}|} & \text{if } x_i > 0 \text{ and } i \neq k, \\ x_k + \varepsilon & \text{if } i = k, \end{cases} \quad (6)$$

where $\Theta = \{j \in \Lambda \mid x_j > 0\}$, $|\Theta|$ is the number of elements in the set Θ and $0 < \varepsilon < |\Theta| \min(x_j)_{j \in \Theta}$, which ensures that negative values do not occur in the resulting vector X_ε^k. This

definition of X_ε^k preserves the compositional data properties. We notice that a positive variation in one entry automatically induces a negative variation in some of the others entries. The distance δ between these two GLCMs is

$$\delta(X, X_\varepsilon^k) = \frac{|\varepsilon|}{|2x_k + \varepsilon|} + \sum_{j \in \Theta - \{k\}} \frac{|\varepsilon/|\Theta - \{k\}||}{|2x_j - \varepsilon/|\Theta - \{k\}||}. \quad (7)$$

For our particular problem we substitute the GLCM $\overline{X}$, the approximation of the neighbourhood, for X in (7) and we can measure the perturbation due to a variation ε on one component $\overline{x_i}$. A negative variation is caused by a structure replacing the background whilst a positive variation indicates an addition of a structure. For one single structure there is a high correlation between the negative and the positive variation (but this correlation is reduced when multiple structures are added). As we are interested in the added structures, only positive variations are considered. We define a perturbation function pert() as

$$\mathrm{pert}\,(x_i, \varepsilon)$$

$$= \begin{cases} 0 & \text{if } \varepsilon \le 0, \\ \delta\left(\overline{X}, \overline{X_\varepsilon^i}\right) = \dfrac{|\varepsilon|}{|2\overline{x_i} + \varepsilon|} \\ \qquad + \sum_{j \in \Theta_i} \dfrac{|\varepsilon/|\Theta_i||}{|2\overline{x_j} - \varepsilon/|\Theta_i||} & \text{otherwise, using (7),} \end{cases}$$

$$(8)$$

where $|\varepsilon| \ll 1$, $\Theta = \{k \in \Lambda | \overline{x_k} > 0\}$ and $\Theta_i = \Theta - \{i\}$.

This perturbation function is computationally expensive (as we calculate the perturbation for each entry of a GLCM). Using problem specific facts (see Appendix B), we can approximate the function pert() as

$$\mathrm{pert}\,(x_i, \varepsilon_i) = \begin{cases} 0 & \text{if } \varepsilon_i \le 0, \\ \dfrac{|\varepsilon_i|}{|2\,\overline{x_i} + \varepsilon_i|} + K\varepsilon_i & \text{otherwise,} \end{cases} \quad (9)$$

where $K = \sum_{j \in \Theta} 1/|\Theta| |2\overline{x_j}|$ and $X = [x_i]_{i \in \Lambda}$ is the GLCM from the RoI.

From this perturbation function we are able to construct a probability image representing the probability that a pixel is a part of a particle. If we consider two pixels p_1 and p_2 with intensity I_{p_1} and I_{p_2}, respectively, and separated by a translation $\vec{t}$, pixel p_1 adds one occurrence to the element $x_{\mathrm{ind}(I_{p_1}, I_{p_2})}^{\vec{t}}$ of the GLCM using translation $\vec{t}$ denoted by $X^{\vec{t}}$. The pixel p_1 can be seen as a part of the component $x_{\mathrm{ind}(I_{p_1}, I_{p_2})}^{\vec{t}}$ of $X^{\vec{t}}$. From that we can define the probability $P^{\vec{t}}(p_1)$ of pixel p_1 to be part of a structure by

$$P^{\vec{t}}(p_1) = \mathrm{pert}\left(x_{\mathrm{ind}(I_{p_1}, I_{p_2})}^{\vec{t}}, x_{\mathrm{ind}(I_{p_1}, I_{p_2})}^{\vec{t}} - \overline{x}_{\mathrm{ind}(I_{p_1}, I_{p_2})}^{\vec{t}}\right). \quad (10)$$

To improve the detection of particles and to remove noise, we use a set of translations Γ. Due to the small size of

the particles, a scale of 1 pixel has been used with four orientations ($0°$, $45°$, $90°$, and $135°$). The four translations cover the 8-connected neighbour pixels and the usage of additional orientations did not add any information. The inclusion of additional scales did not increase the sensitivity. The probability for a pixel p_i is given by

$$P(p_i) = \min_{\vec{t} \in \Gamma}\left(P^{\vec{t}}(p_i)\right). \quad (11)$$

When we have the probability for a pixel to be part of a particle, the segmented image is obtained by applying a threshold to the probability image. All pixels having a probability higher than the threshold are classified as particles.

2.6. Features

The features used for the automated scoring in this study are based on GLCM information and particle detection and described by

(1) the distance between the GLCM from the neighbourhood and the GLCM from the RoI is determined by $\delta_+(X, \overline{X})$. The definition of the distance is derived from (3), where only positive differences in the RoI are taken into account (see (12)). This feature gives a measure of the difference in the texture of the two regions. $\delta_+(X, \overline{X})$ is defined as

$$\delta_+(X, \overline{X}) = \sum_{i,(x_i - \overline{x_i}) > 0} \frac{|x_i - \overline{x_i}|}{|x_i + \overline{x_i}|}; \quad (12)$$

(2) the mean grey-level difference ($\mathrm{MGLD}p$) between the detected particles and the rest of the RoI. An increase in the difference means an increased visibility of the particles and hence a higher score

$$\mathrm{MGLD}p = \log\left(|P|\right)\left(\sum_{p \in P} \frac{I(p)}{|P|} - \sum_{p \in B} \frac{I(p)}{|B|}\right), \quad (13)$$

where P is the set of pixels segmented as particles, B is the set of pixels classified as background, $I(p)$ is the intensity of pixel p, and $|P|$ is the number of elements in P; $\mathrm{MGLD}p$ is weighted by $\log(|P|)$ to discriminate between a small number of pixels with high contrast and a larger number of pixels with the same high contrast. A larger number of high contrast pixels means that the group is more significant.

2.7. Classification

A k-nearest-neighbour classifier [11], based on the two features described previously, is used to classify groups of particles into four classes. The choice of four classes is made to emulate the human observer process where a 0-1-2-3 scoring system is used. The features have been normalised to unit variance and zero mean. Furthermore, an equal a priori likelihood for each score is imposed by applying a correction to the number of neighbours in each class to take into account the different sample sizes of the four classes.

A leave-one-out approach is used for all scoring results. This means that each sample is scored by a classifier trained on all the other samples. Results are evaluated using confusion matrices, which allow a more detailed inspection of the scoring errors.

3.   RESULTS

3.1.   Comparing features and expert classification

Each RoI was characterised by the feature δ_+ (see (12)). Figure 5a illustrates the relationship between the scores assessed by an expert and δ_+. Figure 5a shows that there is an overlap in the range of δ_+ between adjacent scores. However, we note that a good separation exists between a score of 0 or 1 and a score of 3. In addition, there is a general trend of an increased δ_+ measure when the expert's score increases.

Each RoI was also characterised through measurement of the MGLDp feature (see (13)). Figure 5b compares the results of the MGLDp measure with the expert scores in observation 1. The same trend as for the feature δ_+ exists.

The combination of the two features enhances the separation of the four scores as illustrated in the scatter plot shown in Figure 6.

3.2.   Classification results

Classification results obtained using two features ($\delta_+(X, \overline{X})$, MGLD$p$) and a 4-nearest-neighbour classifier. The first experiment uses observation 1 as ground truth. The confusion matrix in Table 3a shows the details of the scoring. The second experiment uses observation 2 as ground truth and the details of the results are presented in Table 3b. Table 4 is the confusion matrix for the observer when the film scoring was repeated (i.e., observation 1 versus observation 2) to assess intra-observer consistency.

An intrascore agreement between the expert and the automatic approach of 83% has been achieved. As the scores are not well defined and are subjective, exact agreement is hard to obtain. Table 4 shows that the same observer cannot obtain an exact agreement (a consistency of 76% has been achieved).

Using the conventional scoring system, the phantom score for the groups of particles is obtained by adding the score of the six groups A-B-C-D-E-F. The intra-observer standard deviation for these scores is 1.36 compared to 0.88 for the computer (between experiment (a) and (b) described in Table 3). The mean variation is 0.56 for the observer against 0.12 for the computer. In addition, the intrascore agreement in both experiments is similar (82% and 83%). These observations show a good consistency of the results. However, for the human observer no deviation larger than one in the score has been recorded, but in 4% of cases the computer delivered a score difference greater than one (i.e., 2).

4.   DISCUSSION

Seventy percent of the computer's score deviation came from the last three groups of particles (D, E, and F). This is mainly
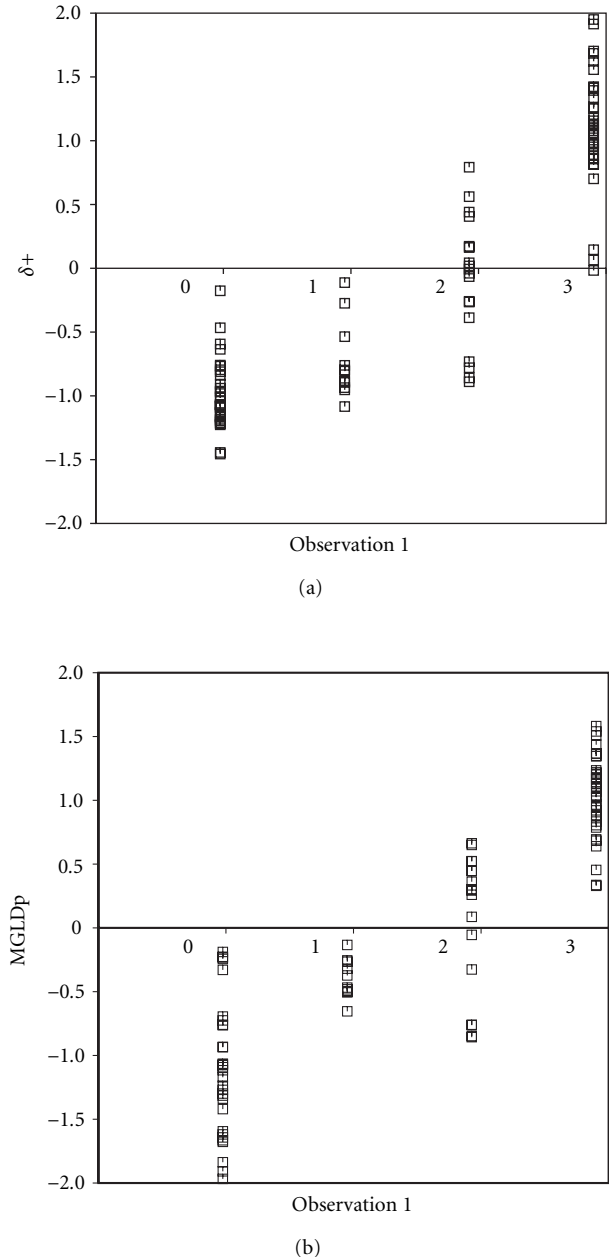

(a)

(b)

FIGURE 5: Scatter plot of the expert scores versus (a) $\delta_+(X, \overline{X})$ and (b) MGLDp.

due to the digitisation of films. Particles, especially small particles, are blurred when data are binned into individual pixels during digitisation. Figure 7 illustrates the problem for particles with a size of around $100\,\mu$m (e.g., 2 pixels at $50\,\mu$m/pixel). The expert observer has noticed a loss of readability of the digitised image compared with the initial film. Whilst visible in the film image, particles from groups E and F become obscured as a result of the combined effects of the pixel size limitation and the noise added into the digitised image by the digitiser.
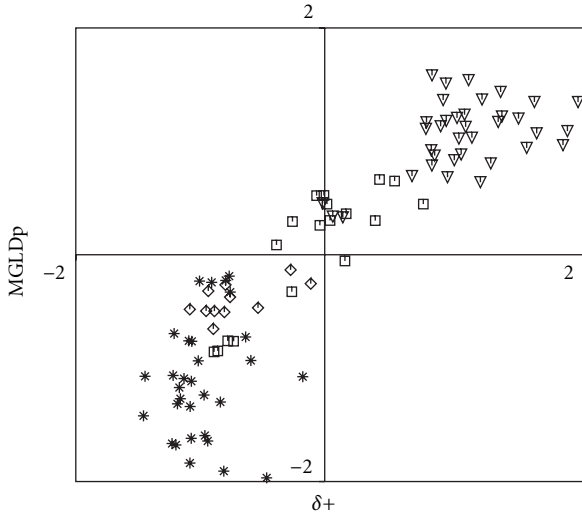
FIGURE 6: Scatter plot of $\delta_+(X, \overline{X})$ versus MGLDp where $*$ is score 0, $\Diamond$ is score 1, $\square$ is score 2, and $\triangledown$ is score 3.

TABLE 3: Classification results, where observation 1 (a) and observation 2 (b) are used as ground truth.

		Computed score			
		0	1	2	3
Observation 1	0	25%	3%	4%	0%
	1	3%	8%	1%	0%
	2	0%	1%	17%	1%
	3	0%	0%	4%	33%

(a)

		Computed score			
		0	1	2	3
Observation 2	0	14%	3%	0%	0%
	1	9%	17%	2%	0%
	2	2%	2%	18%	0%
	3	0%	0%	0%	33%

(b)

Figure 8 compares a digitised film image with an image of the same microparticle group acquired directly, using a Small Field Digital (SFD) mammography system (Siemens Opdima, Siemens-Elema, Sweden). Whilst in this case the pixel sizes are comparable (50 μm for film and 44 μm for the SFD), the effect of the higher noise content of the digitised film all but obscures the microparticles. The addition of degrading factors into the computer image (pixel size and digitisation noise) clearly imposes a limitation when using this methodology as a tool in monitoring the quality of the mammography imaging system. This would be less important if the magnitude of these effects is well below the limits
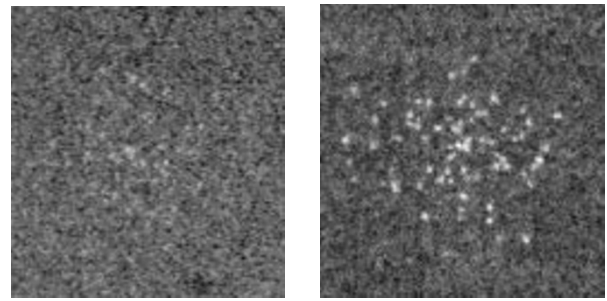
TABLE 4: Intra-observer agreement.

		Observation 1			
		0	1	2	3
Observation 2	0	18%	0%	0%	0%
	1	15%	9%	3%	0%
	2	0%	2%	16%	4%
	3	0%	0%	0%	33%
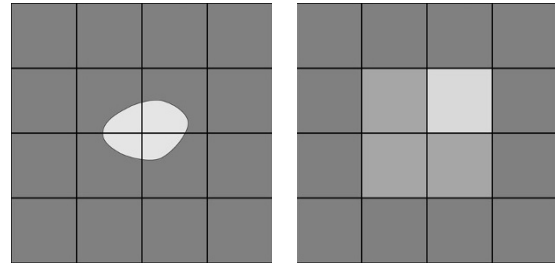
(a)   (b)

FIGURE 7: (a) Original particle at intercept of four pixels, (b) digitised grey-level representation of each pixel.

(a)   (b)

FIGURE 8: (a) Digitised film and (b) direct digital acquisition.

for the key parameters of spatial resolution and noise of the primary imaging transducer, in this case the film. This could be achieved with higher quality film digitisation.

Small Field Digital mammography systems are already entering into routine clinical use for the imaging of symptomatic women or those initially identified with high suspicion from the screening film. Full Field Digital mammography systems, having a pixel size of typically 100 μm, are being clinically evaluated against traditional film-screen systems and are likely to become increasingly used for breast imaging. Systems using computed radiography (CR) are also being evaluated for their effectiveness compared with conventional film-screen. The benefits of computerised detection with all these modalities is self-apparent as data is acquired directly and can be analysed without the need to undergo a secondary

conversion process. Differences between the detection results and that of the observer can be used to enhance the performance of the observer and identify weaknesses in the image display to the observer. The optimisation process must hence take into account both ways in which the image is acquired, but also displayed. Work is in progress using image data from the Siemens Opdima SFD mammography system.

For a fully automated evaluation of the test object the next step is to analyse and score the low contrast linear (fibre) structures, and the low contrast circular (lesion) structures. Comparison with the ground truth based upon the known content of the test object must also replace the subjective opinion of the expert and be exploited in a "scoring methodology" specific to computerised detection. This has the potential for greater objectivity and sensitivity concerning overall image quality.

5. CONCLUSION

We have investigated the automated scoring of the microparticle features of the Leeds TOR(MAM) test object used in quality control for mammographic imaging. The results are encouraging as they show good agreement with the expert observer (better than 82%), and a higher consistency. Variability due to subjectivity in the assessment has been removed. This is important if true variations in mammography system performance are to be identified and monitored. There is also the potential to use this methodology in optimising system performance. The consistent evaluation of the image is essential for the many images that would need to be acquired under different acquisition conditions as part of the optimisation process. It would be difficult for the observer to maintain consistency and concentration during the time-consuming evaluation of many images. In order for this to be viable, higher quality film digitisation or direct digital acquisition will be required so as to avoid degrading the image information.

APPENDICES

A. THE DISTANCE MEASURE δ

We prove that δ is a distance on the space $[0, 1]^N$. To be a distance, δ must have the following properties:

(P1) for all $X, Y \in [0, 1]^N$, $\delta(X, Y) \geq 0$;
(P2) for all $X \in [0, 1]^N$, $\delta(X, X) = 0$;
(P3) for all $X, Y \in [0, 1]^N$, $\delta(X, Y) = \delta(Y, X)$;
(P4) for all $X, Y, Z \in [0, 1]^N$, $\delta(X, Y) + \delta(Y, Z) \geq \delta(X, Z)$.

Clearly (P1), (P2), and (P3) are correct. The property (P4) is proved below.

Proof.

$$(P4) \iff \sum_{i=1}^{N} \left(\Delta_i(X, Y) + \Delta_i(Y, Z) - \Delta_i(X, Z) \right) \geq 0$$
$$(P4) \impliedby \forall i = 1, \ldots, N, \quad \Delta_i(X, Y) + \Delta_i(Y, Z) - \Delta_i(X, Z) \geq 0. \tag{A.1}$$

We study the sign of $\Delta_i(X, Y) + \Delta_i(Y, Z) - \Delta_i(X, Z)$. Consider $i \in \mathbb{N}, 1 \leq i \leq N$

(1) if $x_i = z_i$ or $x_i = y_i$ or $y_i = z_i$, the result is trivial;
(2) now consider the case where x_i, y_i, and z_i are distinct, then

$$\Delta_i(X, Y) + \Delta_i(Y, Z) - \Delta_i(X, Z) = \frac{|x_i - y_i|}{|x_i + y_i|} + \frac{|y_i - z_i|}{|y_i + z_i|} - \frac{|x_i - z_i|}{|x_i + z_i|}. \tag{A.2}$$

As $x_i, y_i, z_i \in [0, 1]$ the sign of $\Delta_i(X, Y) + \Delta_i(Y, Z) - \Delta_i(X, Z)$ is the same as A_i defined by

$$\begin{aligned} A_i = &\ |x_i - y_i| (y_i + z_i)(x_i + z_i) \\ &+ |y_i - z_i| (x_i + z_i)(x_i + y_i) \\ &- |x_i - z_i| (y_i + z_i)(x_i + y_i). \end{aligned} \tag{A.3}$$

Now, we study the sign of A_i. To remove the absolute values, we need to split the problem in six cases:

(i) $0 \leq x_i < y_i < z_i \leq 1$,

$$\begin{aligned} A_i = &\ (y_i - x_i)(y_i + z_i)(x_i + z_i) \\ &+ (z_i - y_i)(x_i + z_i)(x_i + y_i) \\ &- (z_i - x_i)(y_i + z_i)(x_i + y_i) \end{aligned} \tag{A.4}$$

after simplification,

$$\begin{aligned} A_i = &\ (z_i - y_i)(x_i - z_i)(x_i - y_i), \\ &\ z_i - y_i > 0, \qquad x_i - z_i < 0, \\ &\ x_i - y_i < 0 \quad \text{so } A_i \geq 0. \end{aligned} \tag{A.5}$$

We proceed in the same way for the other cases;
(ii) $0 \leq x_i < z_i < y_i \leq 1$,

$$\begin{aligned} A_i = &\ (3x_i y_i^2 - 3x_i z_i^2) + (y_i^2 z_i - y_i z_i^2) + (x_i^2 y_i - x_i^2 z_i), \\ &\ 3x_i y_i^2 - 3x_i z_i^2 \geq 0, \qquad y_i^2 z_i - y_i z_i^2 \geq 0, \\ &\ x_i^2 y_i - x_i^2 z_i \geq 0 \quad \text{so } A_i \geq 0; \end{aligned} \tag{A.6}$$

(iii) $0 \leq y_i < x_i < z_i \leq 1$,

$$\begin{aligned} A_i = &\ (3z_i x_i^2 - 3z_i y_i^2) + (z_i^2 x_i - z_i^2 y_i) + (x_i^2 y_i - y_i^2 x_i), \\ &\ 3z_i x_i^2 - 3z_i y_i^2 \geq 0, \qquad z_i^2 x_i - z_i^2 y_i \geq 0, \\ &\ x_i^2 y_i - y_i^2 x_i \geq 0 \quad \text{so } A_i \geq 0; \end{aligned} \tag{A.7}$$

(iv) $0 \leq y_i < z_i < x_i \leq 1$,

$$\begin{aligned} A_i = &\ (3x_i z_i^2 - 3x_i y_i^2) + (x_i^2 z_i - x_i^2 y_i) + (z_i^2 y_i - y_i^2 z_i), \\ &\ 3x_i z_i^2 - 3x_i y_i^2 \geq 0, \qquad x_i^2 z_i - x_i^2 y_i \geq 0, \\ &\ z_i^2 y_i - y_i^2 z_i \geq 0 \quad \text{so } A_i \geq 0; \end{aligned} \tag{A.8}$$

(v) $0 \leq z_i < x_i < y_i \leq 1$,

$$\begin{aligned} A_i = &\ (3z_i y_i^2 - 3z_i x_i^2) + (z_i^2 y_i - z_i^2 x_i) + (y_i^2 x_i - x_i^2 y_i), \\ &\ 3z_i y_i^2 - 3z_i x_i^2 \geq 0, \qquad z_i^2 y_i - z_i^2 x_i \geq 0, \\ &\ y_i^2 x_i - x_i^2 y_i \geq 0 \quad \text{so } A_i \geq 0; \end{aligned} \tag{A.9}$$

(vi) $0 \leq z_i < y_i < x_i \leq 1$,

$$A_i = (x_i - y_i)(z_i - x_i)(z_i - y_i),$$

$$x_i - y_i > 0, \qquad z_i - x_i < 0, \qquad \text{(A.10)}$$

$$z_i - y_i < 0 \quad \text{so } A_i \geq 0.$$

In all cases $A_i \geq 0$ so $\Delta_i(X, Y) + \Delta_i(Y, Z) - \Delta_i(X, Z) \geq 0$ and finally the property (P4) is verified. Thus δ is a distance on the space $[0, 1]^N$. $\qquad\square$

B. APPROXIMATION OF THE pert() FUNCTION

The definition of the function pert() given by (8) (see below and also Section 2.5) is computationally too expensive to be used. An approximation of this function is given

$$\text{pert}\,(x_i, \varepsilon_i)$$

$$= \begin{cases} 0 & \text{if } \varepsilon_i \leq 0, \\ \dfrac{|\varepsilon_i|}{|2\overline{x_i} + \varepsilon_i|} + \displaystyle\sum_{j \in \Theta_i} \dfrac{|\varepsilon_i/|\Theta_i||}{|2\overline{x_j} - \varepsilon_i/|\Theta_i||} & \text{otherwise, using (7),} \end{cases}$$

$$\text{(B.1)}$$

where $|\varepsilon| \ll 1$, $\Theta = \{k \in \Lambda | \overline{x_k} > 0\}$, $\Theta_i = \Theta \setminus \{i\}$ and $|\Theta_i|$ is the cardinality of the set Θ_i.

In the investigated problem, $\varepsilon_i = x_i - \overline{x_i}$, where $X = [x_i]_{i \in \Lambda}$ is the GLCM of the RoI and $\overline{X} = [\overline{x_i}]_{i \in \Lambda}$ is the GLCM which approximates the background texture. In addition, $|\Theta| - 1 \leq |\Theta_i| \leq |\Theta|$ and $|\Theta| \gg 1$ so we obtain a first approximation given by

$$\sum_{j \in \Theta_i} \frac{|\varepsilon_i/|\Theta_i||}{|2\overline{x_j} - \varepsilon_i/|\Theta_i||} \approx \sum_{j \in \Theta_i} \frac{|\varepsilon_i|}{|\Theta||2\,\overline{x_j} - \varepsilon_i/|\Theta||}. \quad \text{(B.2)}$$

Moreover, for all $k \in \Theta$, for all $i \in \Lambda, \overline{x_k} \gg \varepsilon_i/|\Theta|$ so $|2\,\overline{x_j} - \varepsilon_i/|\Theta|| \approx |2\overline{x_j}|$ and

$$\sum_{j \in \Theta_i} \frac{|\varepsilon_i/|\Theta||}{|2\overline{x_j} - \varepsilon_i/|\Theta||} \approx \sum_{j \in \Theta_i} \frac{|\varepsilon_i|}{|\Theta||2\overline{x_j}|}. \quad \text{(B.3)}$$

Finally, for practical reasons the following approximation is done

$$\forall i \in \Lambda, \quad \sum_{j \in \Theta_i} \frac{|\varepsilon_i|}{|\Theta_i||2\overline{x_j}|} \approx \sum_{j \in \Theta} \frac{|\varepsilon_i|}{|\Theta||2\overline{x_j}|} \quad \text{(B.4)}$$

to obtain

$$\text{pert}\,(x_i, \varepsilon_i) = \begin{cases} 0 & \text{if } \varepsilon_i \leq 0, \\ \dfrac{|\varepsilon_i|}{|2\overline{x_i} + \varepsilon_i|} + K\varepsilon_i & \text{otherwise,} \end{cases} \quad \text{(B.5)}$$

where $K = \sum_{j \in \Theta} 1/|\Theta||2\overline{x_j}|$ and $X = [x_i]_{i \in \Lambda}$ is the GLCM from the RoI.

REFERENCES

[1] P. Forrest, "Breast cancer screening report to the Health Ministers of England, Wales, Scotland and Northern Ireland by a working group chaired by Professor Sir Patrick Forrest," DHSS HMSO, 1986.

[2] K. Faulkner and J. A. Law, "A comparison of mammographic phantoms," *British Journal of Radiology*, vol. 67, pp. 174–180, 1994.

[3] A. C. Underwood, J. Law, and C. Clayton, "Reproducibility of leeds TOR(MAM) mammographic test object plates," *British Journal of Radiology*, vol. 70, pp. 186–191, 1997.

[4] K. W. Brooks, J. H. Trueblood, and K. J. Kearfott, "Automated analysis of mammography quality control images," *Med. Phys.*, vol. 20, no. 3, pp. 881, 1993.

[5] N. Karssemeijer, D. P. C. Treinekens, and M. A. O. Thijssen, "Automated computation of contrast detail curves of mammographic imaging equipment," *Eur. Radiology*, vol. 5(suppl.), no. s6, 1995.

[6] A. D. Castellano-Smith, I. A. Castellano-Smith, and D. R. Dance, "Objective assessment of phantom image quality in mammography: a feasibility study," *British Journal of Radiology*, vol. 71, no. 841, pp. 48–58, 1998.

[7] A. R. Cowen, D. S. Brettle, N. J. Coleman, and G. J. S. Parkin, "A preliminary investigation of the imaging performance of photostimulable phosphor computed radiography using a new design of mammographic quality control test object," *British Journal of Radiology*, vol. 65, no. 774, pp. 528–535, 1992.

[8] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, 1992.

[9] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

[10] J. Aitchison, *The Statistical Analysis of Compositional Data*, Chapman and Hall, New York, NY, USA, 1986.

[11] J. G. McLachan, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York, NY, USA, 1992.

Lilian Blot obtained a Bachelor of Science in computer science at the University of Caen (France) and graduated in 1994. He also obtained a Bachelor of Science in mathematics in 1997 and an M.S. degree in computer science in 1998 at the University of Caen. He subsequently passed a Research Masters in artificial intelligence and computer vision at the same institution before commencing a Doctorate in computer vision at the University of Portsmouth (United Kingdom). He is currently continuing his Ph.D. at the University of East Anglia (United Kingdom). Lilian has a particular interest in medical image analysis with respect to mammography and volumetric data analysis.

Anne Davis is a Principal Physicist working for Portsmouth Hospitals NHS Trust. She works within the Radiological Sciences Group of the Medical Physics Department providing scientific support and advice to users of a wide range of diagnostic x-ray and CT equipment. She has 10 years experience of providing physics support to breast screening units working within the NHS Breast Screening programme, and is a member of the UK Mammography Physics Group. She gained her physics degree at the University of Exeter in 1989 and is a corporate member of the Institute of Physics and Engineering in Medicine and a State Registered Clinical Scientist.

Mike Holubinka is a Consultant Medical Physicist heading up the Radiological Sciences Group of the Portsmouth Medical Physics Service. He obtained his first degree from the University of Kent at Canterbury in 1979 and a Master degree at the University of Surrey in 1981. He started his career in Medical Physics at Southampton University Hospital specialising in radiological physics and radiation protection in 1981 and has continued this work in Portsmouth since 1986. He has coauthored more than 20 conference and journal papers. His main research and development interests are in the application of image analysis techniques to the quantification of image quality in diagnostic radiology.

Robert Martí received his B.S. degree in 1997 and his M.S. degree in 1999 both in computer science at the University of Girona, Spain. He is currently in the third year of his Ph.D. at the University of East Anglia, after spending the first two years at the University of Portsmouth. His research interests include computer vision and medical image analysis, specially focusing on image registration, feature extraction and similarity measures.

Reyer Zwiggelaar is a Senior Lecturer at the University of East Anglia. He obtained his first degree from the State University Groningen in the Netherlands in 1989. Subsequently, he studied at University College London (UK) to obtain a Ph.D. in 1993. Dr. Zwiggelaar has coauthored more than fifty conference and journal papers. His main research interest is in medical image understanding, especially concentrating on mammographic data, pattern recognition, statistical methods and feature detection techniques.