HISTOGRAMS IN BREAST DENSITY CLASSIFICATION: ARE 4 BETTER THAN 1?

Izzati Muhimmah[a], Erika RE Denton[b], dan Reyer Zwiggelaar[c]
[a]*Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia*
Kampus Terpadu UII, Jl. Kaliurang Km 14.5 Yogyakarta, E-mail: emma@fti.uii.ac.id
[b]*Department of Breast Imaging, Norwich and Norfolk University Hospital, NR4 7UY, UK*
[c]*Department of Computer Science, University of Wales Aberystwyth, SY23 3DB, UK*

Abstract

Breast density is known to be an important indicator of breast cancer risk. Quantitative estimation approaches have been developed and some of these are based on grey level or texture metrics. Developing an example of the former, we combined multi-resolution histograms and pattern matching approaches. The results show good agreement compared to expert radiologist classification. The presented approach provides similar results to existing methods and limitations are discussed.

***Keywords**: multi-resolution histogram technique, mammography, breast density classification*

1. Introduction

Breast density is known to be an important indicator of breast cancer risk [1,2]. It is common practice to divide breast density into a number of percentage density classes [2], whilst sometimes the range of classes is collapsed to two or three [3]. Many studies have been reported on (semi-)automatic breast density segmentation or classification, some of these approaches were based on histogram information [4-8]. In general, the published histogram based approaches for automatic density estimation produced robust and reliable results.

In contrast with fatty breast tissue which appears dark in mammographic images, dense breast tissue appears brighter [9]. This phenomenon plausibly leads to model breast dense tissue using histograms, which are a simplified representation of images. The study by Zhou et al. [7] showed that there were some typical histogram patterns for each density class. However, they also pointed out that there are similar histogram patterns that represent different risks. On the other hand, a recently published paper by Hadjidemetriou et al. showed that different generic texture of binary images with similar histograms can be discriminated by a multi-resolution approach [10]. Based on these findings, our aim in this study is to investigate whether it is possible to automatically estimate mammographic density using a combination of multi-resolution histogram information and traditional classification approaches.

The remainder of this paper is outlined as follows: the proposed multiresolution histogram, features and classifiers are described in Section 2. Sections 3 and 4 give respectively, results of the proposed method and discussion on our findings. Finally, conclusions appear in Section 5.

2. Histogram Based Classification

A brief description of Hadjidemetriou's multi-histogram texture classification approach [10] can be found below.

The main aim is to obtain feature vectors which can be used to discriminate between the various mammographic density classes. A feature vector representing a mammogram is derived from a set of histograms {*h_0* , *h_1* , *h_2*, *h_3*}. *h_0* is obtained from the original mammogram, and histograms *h_1* , *h_2* and *h_3* are obtained after Gaussian filtering the mammogram by 5x5, 9x9 and 13x13 kernels, respectively. For all four histograms only grey level information from the breast area (ignoring the pectoral muscle and background areas) is used and the histograms are normalised with respect to this area. For increasing scales this shows the general shift to lower grey-level values and the narrowing of the peaks in the histogram data. It should be noted that these histograms deviate significantly from those described by Hadjidemetriou et al. [10] which start with delta function peaks which broaden on smoothing.

Subsequently, the set of histograms are transformed into a set of cumulative histograms {c_0 , c_1 , c_2, c_3\}. The feature vector for each mammogram is constructed from the difference between subsequent cumulative histograms: {c_0 - c_1 , c_1 - c_2 , c_2 - c_3}. Between scales this shows a shift to lower grey-level values, but the overall shape of the data remains more or less constant. The dimensionality of the resulting feature space is equal to 768.

The final stage in the classification process is to use the feature vectors in combination with a k-nearest-neighbour approach. Here we have used three neighbours, an Euclidean distance (in [10] a *L_1* norm was used), and Bayesian probability or major voting (e.g. in the under-represented classes).

Data used for validation were mammograms in MIAS database, which consists of pairs of mammograms (all medio-lateral views, L/R). It is known that mammographic intensities vary with exposure levels and film characteristics [11, 12]. In an imaging session, a woman likely had the mammogram captured using similar films and/or exposure levels. To minimise bias, we used a *leave-one-woman-out* strategy in training.

3. Experimental Results

We did two independent experiments to validate this methodology. Both classification results were presented in a form of confusion matrix.

Firstly, we evaluated this methodology on 321 images of the Mammographic Image Analysis Society (MIAS) database, which provide three density categories, namely: Fatty, Fatty-glandular, and Dense-glandular, as a gold standard [3]. This is done to provide direct comparison with some of the published works.

Secondly, we asked an expert radiologist to classify 319 of these images (two images were not included due to technical problems), based on Boyd's SCC [2], twice and the consensus rating was found in 259 cases that were used for training, see Table 1. It should be noted that the low and high density class examples were under-represented in the database.

Table 1. Expert radiologist: intra-observer variation. Within the tables the proportion of dense tissue is represented as 1: 0%, 2: 0-10%, 3: 11-25%, 4: 26-50%, 5: 51-75% and 6: 76-100%.

		Expert Classification					
		1	2	3	4	5	6
Expert Classification	1	6	0	0	0	0	0
	2	5	59	5	0	0	0
	3	0	7	54	12	0	0
	4	0	1	8	66	7	0
	5	0	0	0	6	57	0
	6	0	0	0	0	9	17

The result of the first experiment is presented as a confusion matrix in Table 2, in an agreement of *61.56%*.

Table 2. Comparison between automatic, histogram based, and expert classification. Within the tables the proportion of dense tissue is represented as 1: Fatty, 2: Fatty-glandular, and 3: Dense-glandular.

		Expert Classification		
		1	2	3
Auto	1	78	32	7
	2	23	53	38
	3	5	19	66

The validation of the multi-resolution histogram based classification using SCC is presented as a confusion matrix in Table 3. The results in comparison with Expert Assessment 1, see Table 3 (a), showed an agreement of *52.65%* and *94.36%* when minor classifications deviation is allowed. Whilst results in comparison with Expert Assessment 2, see Table 3 (b), gave an agreement of *55.17%*; allowing minor classifications deviations, this brings the overall agreement to almost *95.29%*.

Table 3. Comparison between automatic, histogram based, and expert classification. Within the tables the proportion of dense tissue is represented as 1: 0%, 2: 0-10%, 3: 11-25%, 4: 26-50%, 5: 51-75% and 6: 76-100%.

		Expert Classification 1					
		1	2	3	4	5	6
Automatic Classification	1	0	5	0	0	0	0
	2	11	43	15	3	0	0
	3	0	14	26	13	2	0
	4	0	5	24	54	26	4
	5	0	0	2	12	41	9
	6	0	0	0	2	4	4

(a)

		Expert Classification 2					
		1	2	3	4	5	6
Automatic Classification	1	0	5	0	0	0	0
	2	11	43	15	3	0	0
	3	0	14	26	13	2	0
	4	0	5	24	54	26	4
	5	0	0	2	12	41	9
	6	0	0	0	2	4	4

(b)

4. Discussion

The results in Table~\ref{tab:FGD} are similar to those reported by Masek et al. [8], i.e *62.42%* when using an Euclidean distance. Their method is based on direct distance measures of average histogram of original images for each density class. It should be noted that we used less data for training due to *leave-one-pair-out* stategy. Moreover, this is inline with our own single histogram (h_0) results, which were *61.99%* for triple MIAS calssification and *57.14%* for SCC based classification. These results might indicate there is little benefit in using the multi-resolution histogram approach.

We found the density estimation not to be robust. As can be seen in Table 1, the intra-obeserver agreement of an expert radiologist was *81%*. Furthermore, the results in Tables 3 (a) and (b) presented different agreements. Thus, it might be essential to include mammographic MRI data to provide a more robust gold standard.

Despite that multi-resolution histogram technique is claimed to be robust to match either synthetic, Brodatz, or CUReT textures [10], our results could not confirm its application in mammographic density classification. We would like to investigate whether this is caused by nature of the mammographic texture patterns and/or imaging system effects. Thus, additional pre-processing to enhance the contrast between fatty and dense tissue, or to incorporate the X-ray imaging protocol information, are areas of future research.

It should be noted that our methodology slightly deviated from Hadjidemetriou et al. [10]. Their implementation of the multi-scale approach includes a subsampling step which makes a second normalisation essential. In our case, we only using the smoothing stage of the multi-scale approach without the subsampling. As such the second normalisation step is not used.

The underlying assumption on typical histograms for each density class as stated by Zhou et al. [7] may be false. Our visual observations on histogram patterns for six density categories could not infer a single typical histogram patterns that represent each density class. This is due to large variations on means and standard deviations within classes. Statistical analysis to test this hypothesis will be invesigated in the future.

The bottom line is that none of the described (or published) approaches shows a high precise correlation with expert data. It should be mentioned that the inter-observer correlation can also be low. Another strand of our research will investigate the combination of grey-level and texture information in the mammographic density classification process.

5. Conclusions

We have presented a novel approach to mammographic density classification, which uses multi-resolution histogram information. It was shown that the approach was insufficient when compared to the gold standard provided by an expert radiologist, but when minor classifications errors are allowed it provide a performance better than *95%*. Future work will include texture information and the use of MRI data as a gold standard.

Acknowledgements

A studentship for Izzati from Universitas Islam Indonesia and University of Wales Aberystwyth is gratefully acknowledged.

References

[1] J.N. Wolfe, *Risk for Beast Cancer Development Determined by Mammographic Parenchymal Pattern*, Cancer, Vol. 37, pp. 2486-2492, 1976.

[2] N. Boyd, J. Byng, R. Jong et al. *Quantitative Classification of Mammographic Densities and Breast Cancer Risk: Results from the Canadian National Breast Screening Study*, Journal of the National Cancer Institute, Vol. 87, No. 9, pp. 670-675, May 1995.

[3] J. Suckling, J. Parker, D. Dance et al. *The Mammographic Image Analysis Society Digital Mammogram database*. Exerpta Medica. International Congress Series, Vol. 1069, pp. 375-378, 1994.

[4] J. Byng, N. Boyd, E. Fishell et al. *The Quantitative analysis of mammographic densities*, Physics in Medicine and Biology, Vol. 39, pp. 1629-1638, 1994.

[5] N. Karssemeijer, *Automated Classification of Parechymal Patterns in Mammograms*, Physics in Medicine and Biology, Vol. 43, pp. 365-378, 1998.

[6] R. Sivaramakhrisna, N.A. Obuchowski, W.A. Chilcote et al. *Automatic Segmentation of mammographic density*, Academic Radiology, Vol. 8, No. 3, pp. 250-256, March 2001.

[7] C. Zhou, H.P Chan, N. Petrick et al. *Computerized Image Analysis: Estimation of Breast Density on Mammograms*, Medical Physics, Vol. 26, No. 6, pp. 1056-1069, June 2001.

[8] M. Masek, S. Kwok, C. deSilva et al. *Classification of Mammographic Density using Histogram Distance Measures*, CD-ROM Proceedings of the World Congress on Medical Physics and Biomedical Engineering, pp. 1, August 2003.

[9] R.P. Highnam & M. Brady, *Mammographic Image Analysis*, Kluwer Academic Publishers, 1999.

[10] E. Hadjidemetriou, M.D. Grossberg & S.K. Nayar, *Multiresolution Histograms and their Use for Recognition*, IEEE Trans. on PAMI, Vol. 26, No. 7, pp. 831-847, July 2004.

[11] FUJIFILM, *The Fundamentals of Medical Radiography*, in FUJIFILM Technical Handbook: Medical X-ray System, Fuji Photo Film Co. Ltd, 2003.

[12] FUJIFILM, *Fundamentals of Sensitized Materials for Radiography,* in FUJIFILM Technical Handbook: Medical X-ray System, Fuji Photo Film Co. Ltd, 2003.