Set-Permutation-Occurrence Matrix based Texture Segmentation

Reyer Zwiggelaar[1], Lilian Blot[1], David Raba[2], and Erika R.E. Denton[3]

[1] School of Information Systems, University of East Anglia, Norwich, UK
reyer.zwiggelaar@sys.uea.ac.uk
[2] Computer Vision and Robotics Group, University of Girona, Girona, Spain
[3] Department of Breast Imaging, Norfolk and Norwich University Hospital, Norwich, UK

Abstract. We have investigated a combination of statistical modelling and expectation maximisation for a texture based approach to the segmentation of mammographic images. Texture modelling is based on the implicit incorporation of spatial information through the introduction of a set-permutation-occurrence matrix. Statistical modelling is used for data generalisation and noise removal purposes. Expectation maximisation modelling of the spatial information in combination with the statistical modelling is evaluated. The developed segmentation results are used for automatic mammographic risk assessment.

1 Introduction

Texture is one of the least understood areas in computer vision and this lack of understanding is reflected in the ad-hoc approaches taken to date for texture based segmentation techniques. Although no generic texture model has emerged so far a number of problem specific approaches have been developed successfully [1–4]. Although the described approach is developed with one particular application in mind, we do believe that it is generic within the field of medical image understanding.

Since Wolfe's [5, 6] original investigation into the correlation between mammographic risk (i.e. the risk of developing breast cancer) and the perceived breast density a number of automatic approaches have been developed [7–11]. Some of these methods are based on grey-level distributions whilst others incorporate some aspect of spatial correlation or texture measure. While all these methods achieve some correlation with manual visual assessment in general they are not good enough to progress to clinical trials. We have investigated a process to separate the relevant background texture from other image structures [8, 12]. This showed that based on only background texture similar classification results could be obtained when compared to results based on the full image information. It is also important to note that the breast density can change over time for a number of reasons [13].

It is our thesis that the relative size of segmented image regions, representing distinct anatomical tissue classes, is correlated with mammographic risk assessment. Statistical modelling in combination with expectation maximisation (EM) [14] is used for the segmentation of mammographic images. To our knowledge, we introduce a new concept, the set-permutation-occurrence matrix, as a texture feature vector. Realistic

texture modelling is possible as spatial information is implicitly incorporated. Statistical modelling has been used as a pre-processing step to generalise the data whilst at the same time remove some noise aspects. Initial results from this automatic segmentation of mammographic images are promising with a good correlation with annotated regions. We show results for automatic mammographic risk assessment [5] and a comparison with expert manual classification is discussed.

To achieve segmentation a number of steps are required: a) information gathering, b) texture feature extraction, c) statistical modelling, d) EM clustering, and e) image segmentation.

2 Methods

A Gaussian mixture model G with k classes is defined as

$$G(x|\varphi) = \sum_{i=1}^{k} w_i g(x|m_i, v_i) \tag{1}$$

where x is an observation vector, φ is a vector with parameters w_i (weight), m_i (mean) and v_i (covariance) for each class, and $g(x|m_i, v_i)$ is defined as

$$g(x|m_i, v_i) = \frac{1}{(\sqrt{2\pi})^n (\sqrt{det(v_i)})} e^{-\frac{(x-m_i)^T v_i^{-1} (x-m_i)}{2}} \tag{2}$$

where $()^T$ indicates the transpose of a vector, v_i^{-1} indicates the inverse of v_i and $det(v_i)$ stands for the determinant of v_i. The *likelihood* function is a function that gives a measure of how well the probability density function defined by the parameters fits the data. If a set of parameter maximises the likelihood, then these parameters will be the optimum set for the given problem. The likelihood function is defined as

$$\mathcal{L}(\varphi) = \prod_{x \varepsilon \chi} p(x|\varphi) \tag{3}$$

where χ is the data set and $p(x|\varphi)$ is the probability density function. Here the assumption of independence for all data χ is made. Usually, the *log-likelihood* function is used, mainly to use a sum instead of a product and to reduce the magnitude of the result.

$$\mathcal{L}_{log}(\varphi) = \sum_{x \varepsilon \chi} log(p(x|\varphi)) \tag{4}$$

The EM algorithm [14] is a numerical method to estimate a set of parameters that describe a probability distribution, based on data that belongs to this distribution. On each iteration of the algorithm, two steps are performed: first, the E-step evaluates a probability distribution for the data using the parameters of the model estimated on the previous iteration, then the M-step finds the new parameters that maximises the likelihood function. It can be proven mathematically that on each iteration the likelihood increases [14]. One of the problems of the EM algorithm in application to Gaussian Mixture Models is the initialisation [15], with the end results depending on the initial

starting point. It is common to select a random starting point of the data set χ for the centre of each class. To make the overall classification more robust we initialise the centre of the class with the result of the *k-Means* algorithm [15].

2.1 Texture Feature

In general the usage of the EM approach for image segmentation is based on the grey-level information at a pixel level with no direct interaction between adjacent pixels. However, it is well known that texture based segmentation should incorporate spatial correlation information. The modelling should not be based on a single grey-level value but incorporates spatial information implicitly. This is why we extract information from a set of points. The information is extracted at several levels of a scale-space representation of the image.

Scale-Space Representation The first step in obtaining the texture features is the generation of an image-stack which is a scale-space representation. At the smallest scale the original grey-level values are used and to obtain the larger scale images we have used a recursive median filter [16], denoted $\otimes$, and a circular structuring element, R (the diameter of the structuring element increases with scale σ). The resulting image-stack is a set of images

$$\bigcup_{\sigma \in \Gamma} \{I_\sigma\} = \bigcup_{\sigma \in \Gamma} \{I \otimes R_\sigma\}, \tag{5}$$

where Γ is an ordered set of scales. This effectively represents a blurring of the original data and at a particular level in the image-stack only features larger than σ can be found. An alternative representation of the image-stack is given by

$$\bigcup_{\sigma \in \Gamma} \{\overline{I}_\sigma\} = \bigcup_{\sigma \in \Gamma} \{I \otimes R_{\sigma-1} - I \otimes R_\sigma\}, \tag{6}$$

where Γ is a set of scales. This represents the differences between two scales in I_σ and hence the data in the image-stack at a particular level will only contain features at a particular scale σ.

Sampling Points To capture the texture information over a set of scales a feature vector will need to be extracted from all levels of the image-stack. It can be seen that small size aspects (like noise and small objects) are represented at the top (least amount of smoothing) of the image-stack. On the other hand, large size aspects (large and background objects) are represented at the bottom (after smoothing at the appropriate scale) of the image-stack.

The developed method uses a model that can be seen as a generalisation of normal co-occurrence matrices [1]. Indeed, if we just look at the co-occurrence of grey-level values the information can be captured in matrix format, where the rows and columns represent the grey-level values at two sample points. This process can include a set

of points S_{xy}. An example of the points used is shown in Fig. 1. In the experiments described below we have used

$$S_{xy} = \bigcup_{\varepsilon \in D} \{(x, y+\varepsilon), (x+\varepsilon, y)\} \tag{7}$$

where $D = \{-32, -16, -8, -4, -2, 0, 2, 4, 8, 16, 32\}$. This particular set was chosen as it contains short and long range spatial, and directional information. Depending on the level in the scale-space representation this can be used to emphasize small and large scale structures in the image. In the case described here we generate the co-occurrence between all the points in the set of sample points. This is illustrated in Fig. 2 for one particular point, but it should be noted that the same approach is used in a round-robin way or in other words the points are fully connected. When using $\{\overline{I}_\sigma\}$ (a similar notation can be obtained when using $\{I_\sigma\}$), this representation of the texture information in the form of a matrix is given by

$$\overline{\Psi}^\sigma(x, y) = \left[\psi^\sigma_{i,j}\right]_{i,j \in N_g} \tag{8}$$

and

$$\psi^\sigma_{i,j} = \#\{(p, p') \in S_{xy} \times S_{xy} \mid \overline{I}_\sigma(p) = i,\ \overline{I}_\sigma(p') = j\} \tag{9}$$

where $\#$ denotes the number of elements in a set and N_g denotes the set of grey-level values. It should be noted that this approach provides a different description than that would be provided by using a set of co-occurrence matrices.

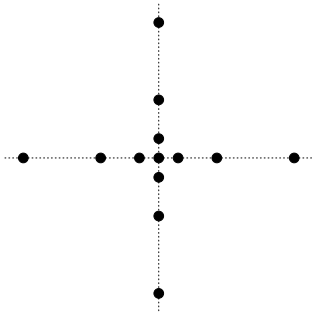

Fig. 1. Sample points S_{xy}.

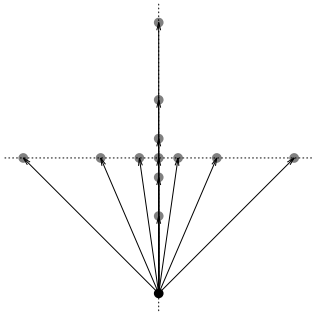

Fig. 2. Sample points connectivity.

Instead of using the co-occurrence of the grey-level values it is possible to use the occurrence of the grey-level difference. Again, this is using the same set of sample points S_{xy} (see Figs 1 and 2) at each scale (i.e. level in the image-stack). As we are using the occurrence of the grey-level difference values our co-occurrence grey-level value matrix reduces to a vector. When using the difference image-stack representation (see Eq. 6) the feature vector at a single scale is given by

$$\overline{\Phi}^{\sigma}(x,y) = \left[\phi_i^{\sigma}\right]_{i \in \delta_{N_g}} \tag{10}$$

where N_g is the set of grey-levels, σ a given scale, δ_{N_g} is the set of grey-level differences and

$$\phi_i^{\sigma} = \#\left\{(p,p') \in S_{xy} \times S_{xy} \mid \overline{I}_{\sigma}(p) - \overline{I}_{\sigma}(p') = i\right\} \tag{11}$$

where, again, $\#$ denotes the number of elements in a set.

One of the main attributes of the feature vector is that descending the original image-stack means that the occurrences of grey-level difference values becomes more localised. In the extreme all grey-level values are identical and the occurrence becomes a delta-function. The story with regard to the difference image-stack is less straight forward. In general the information is sparse and when a structure is present at a particular point and scale the representation changes which is represented as a set of side-bands in the histogram. It should be noted that for a side-band to be regarded as to be caused by an image structure its area should be related to the scale in the image-stack else it can be regarded as noise.

2.2 Statistical Modelling

The texture feature described above is extracted at a pixel level. Combining all the information results in a feature vector which can be used to generate a statistical model. In general such a model is used for noise removal and data generalisation. In this particular case the added bonus of data generalisation is a reduced dimensionality which speeds up the processing. Here we have used principal component analysis [17], but other statistical methods could have been used instead.

The principal components of a set of observation vectors $\{\mathbf{x}_j : j = 1..n\}$ (in our case the texture feature $\overline{\Phi}^{\sigma}$ or $\overline{\Psi}^{\sigma}$) are the characteristic vectors, $\mathbf{P}$, of the covariance matrix, $\mathbf{C}$, constructed from the data set. Projecting the data into its principal components generally results in a compact and meaningful representation in which the characteristic vectors associated with the largest characteristic values describe the major modes of data variation. The characteristic values give the variances associated with the principal components. An observation $\mathbf{x}_j$ can be approximated from the principal components using

$$\mathbf{x}_j \approx \mathbf{P}\ \mathbf{b}_j + \mathbf{m}, \tag{12}$$

where $\mathbf{m}$ is the average vector and $\mathbf{b}_j$ is a vector of weights. The dimensionality of the data set can be reduced by ignoring the principal components with low (or zero) characteristic values.

3 Results

For evaluation purposes we have used a subset of the Mammographic Images Analysis Society (MIAS) database was used [8, 18]. These are screening xray mammograms, and a detailed account of the database can be found in [18].

Although of interest, it is computationally impractical to base the EM modelling on the original texture feature vector as this has a large number of elements (a high dimensionality) and tends to be sparse. All the results presented in this section are based on a PCA reduced feature vector where we typically capture 95% of the data variation.

Segmentation results for example mammograms are shown in Fig. 3. The original mammograms are shown in Fig. 3a,b. The EM and statistical modelling process take only the breast area into account whilst excluding the pectoral muscle and the background. For the results shown in Fig. 3c-f six classes were used. The selection of six classes is based on an information theoretic approach [19]. In both cases the segmentation process produced plausible results which correlate with the original image.

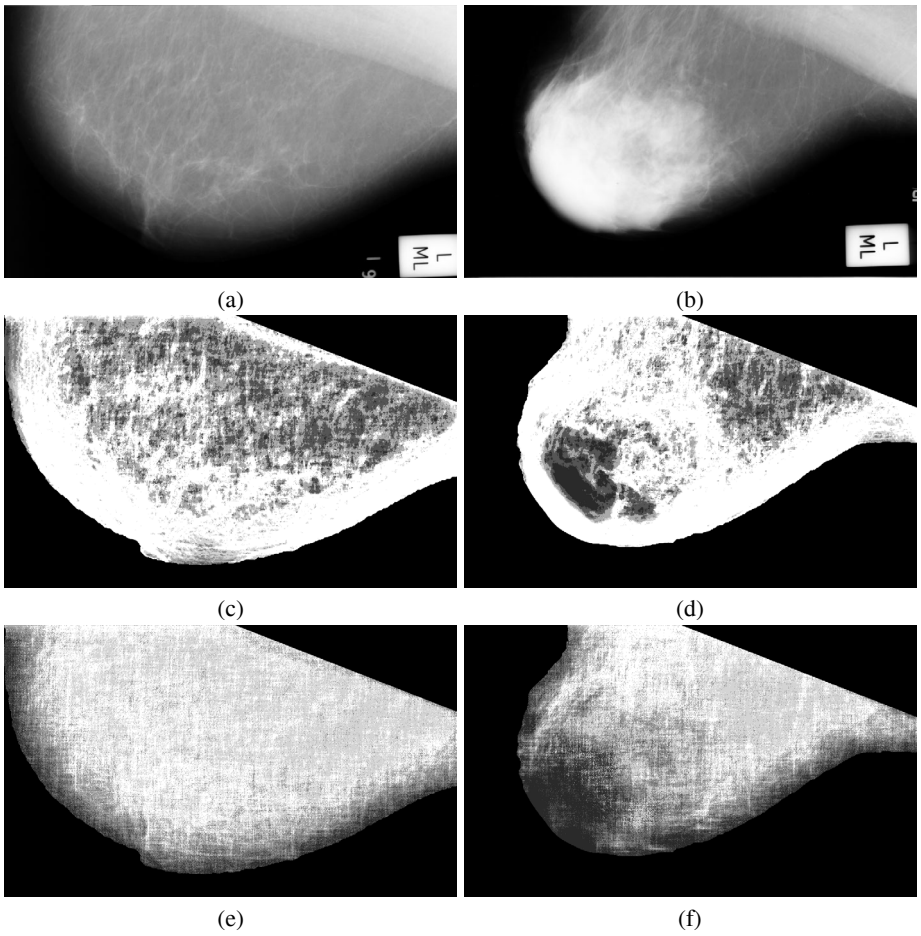

Fig. 3. Original fatty (a) and dense (b) mammographic images. Segmentation results where the EM modelling is based on (c,d) $\{I_\sigma\}$, and (e,f) $\{\overline{I}_\sigma\}$.

3.1 Risk Assessment

To evaluate the segmentation results for mammographic risk assessment all the images were assessed by mammographic experts who provided an estimate of the proportion of dense tissue (i.e. high intensity/non-fatty tissue, see also Fig. 3a,b) in each mammogram. The segmentation results, based on EM and statistical modelling using $\{I_\sigma\}$ or $\{\overline{I}_\sigma\}$, can also be used to obtain the relative size of the segmented regions for each class. This feature is used as our classification space. The correlation between the relative region size distribution and the estimated proportion of dense tissue, when using a nearest neighbour classifier on a leave-one-out basis, can be found in Table 1. This shows an agreement for 66% of the mammograms when using $\{I_\sigma\}$. This increases to 86% when using $\{\overline{I}_\sigma\}$. This compares well with an inter-observer agreement of 45%. The intra-observer agreement on the used dataset is 89%.

		Expert Classification					
		1	2	3	4	5	6
Automatic Classification	1	0	0	0	0	0	0
	2	0	0.22	0	0	0	0
	3	0	0	0.08	0.08	0.06	0
	4	0	0	0.03	0.19	0.11	0
	5	0	0	0.06	0	0.17	0
	6	0	0	0	0	0	0

(a)

		Expert Classification					
		1	2	3	4	5	6
Automatic Classification	1	0	0	0	0	0	0
	2	0	0.17	0	0	0	0
	3	0	0	0.14	0.06	0	0
	4	0	0.06	0.03	0.22	0	0
	5	0	0	0	0	0.33	0
	6	0	0	0	0	0	0

(b)

Table 1. Comparison of the density estimate as given by an expert radiologist and automatic segmentation. (a) $\{I_\sigma\}$ and (b) $\{\overline{I}_\sigma\}$. 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%.

4 Conclusions

We have shown that a combination of EM and statistical modelling results in a robust approach to the segmentation of mammographic images. We have introduced a texture feature vector based on a set-permutation occurrence matrix which captures both spatial and local grey-level information. The use of this type of matrix, especially the size and shape of S_{xy}, will need further development to explore its limitations and full potential. We have shown that the segmentation results can be used to provide valuable information in mammographic assessment of density applications and therefor possibly such as risk assessment.

Acknowledgement

We would like to thank Glynis Wivell for marking the data.

References

1. M.W. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, 1979.
2. R.W. Conners and C.A. Harlow. A theoretical comparison of texture algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(3):204–222, 1980.
3. A.P. Pentland. Fractal-based description of natural scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):661–674, 1984.
4. T.R. Reed and J.M.H. Dubuf. A review of recent texture segmentation and feature-extraction techniques. *Computer Vision, Graphics and Image Processing*, 57(3):359–372, 1993.
5. J.N Wolfe. Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer*, 37(5):2486–2492, 1976.
6. A.M. Oza and N.F. Boyd. Mammographic parenchymal patterns: a marker of breast cancer risk. *Epidemiology Review*, 15:196–208, 1993.
7. N. Karssemeijer. Automated classification of parenchymal patterns in mammograms. *Phys. Med. Biol.*, 43:365–378, 1998.
8. L. Blot, E.R.E. Denton, and R. Zwiggelaar. Risk assessment: the use of background texture in mammographic imaging. 6^{th} *International Workshop on Digital Mammography*, Bremen, Germany:541–543, 2002.
9. J.W. Byng, M.J. Yaffe, G.A. Lockwood, L.E. Little, D.L. Tritchler, and N.F. Boyd. Automated analysis of mammographic densities and breast carcinoma risk. *Cancer*, 80(1):66–74, 1997.
10. J.J. Heine and R.P. Velthuizen. A statistical methodology for mammographic density detection. *Medical Physics*, 27:2644–2651, 2000.
11. R.Sivaramakrishna, N.A. Obuchowsky, W.A. Chilcote, and K.A. Powell. Automatic segmentation of mammographic density. *Academic Radiology*, 8(3):250–256, 2001.
12. R. Zwiggelaar. Separating background texture and image structure in mammograms. In *Proceedings of the* 10^{th} *British Machine Vision Conference*, pages 362–371, Nottingham, UK, 1999.
13. K. Polyak. On the birth of breast cancer. *Biochimica et Biophysica Acta*, 1552:1–13, 2001.
14. P. Demster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society B*, 39:1–38, 1977.
15. P. McKenzie and M.D. Alder. Initialising the em algorithm for use in gaussian mixture modelling. *Proceedings of the International Workshop on Pattern Recognition in Practice IV*, pages 91–105, 1994.
16. R. Zwiggelaar, T.C. Parr, J.E. Schumm, I.W. Hutt, S.M. Astley, C.J. Taylor, and C.R.M. Boggis. Model-based detection of spiculated lesions in mammograms. *Medical Image Analysis*, 3(1):39–62, 1999.
17. I.T. Jolliffe. *Principal Component Analysis*. Springer Verlag, 1986.
18. J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, and J. Savage. The mammographic images analysis society digital mammogram database. In Dance Gale, Astley and Cairns, editors, *Digital Mammography*, pages 375–378. Elsevier, 1994.
19. R. Zwiggelaar, P. Planiol, J. Marti, R. Marti, L. Blot, E.R.E. Denton, and C.M.E. Rubin. Em texture segmentation of mammographic images. 6^{th} *International Workshop on Digital Mammography*, Bremen, Germany:223–227, 2002.