CLASSIFICATION OF CERVICAL CELL NUCLEI USING MORPHOLOGICAL SEGMENTATION AND TEXTURAL FEATURE EXTRACTION[†]

Ross F. Walker*   Paul Jackway*   Brian Lovell*   I. D. Longstaff*

* Dept. of Electrical and Computer Engineering, University of Queensland, Brisbane 4072, AUSTRALIA

This paper presents preliminary results for the classification of Pap Smear cell nuclei, using Gray Level Co-occurrence Matrix (GLCM) textural features. We outline a method of nuclear segmentation using fast morphological gray-scale transforms. For each segmented nucleus, features derived from a modified form of the GLCM are extracted over several angle and distance measures. Linear Discriminant Analysis is performed on these features to reduce the dimensionality of the feature space, and a classifier with hyperquadric decision surface is implemented to classify a small set of normal and abnormal cell nuclei. Using 2 features, we achieve a misclassification rate of 3.3% on a data set of 61 cells.

1   Introduction

The area of machine-assisted Pap smear analysis is one of continuing and wide spread research, despite over 35 years of world-wide attention by researchers. Most published research to date utilises image processing techniques to extract morphometric/photometric features at the cell level[11], while other techniques include contextual (smear demographics, cell distribution) features [4] or Malignancy Associated Change (MAC) measurements[12].

It is thought that the initiation of cell dysplasia occurs in the cell nucleus, and is indicated by visible nuclear structural changes, including increased nuclear size and shape distortion, increased DNA Integrated Optical Density (IOD), and changes in textural appearance of the nucleus. It has been shown that textural features extracted from animal cell nuclei can be successfully used to discriminate between normal and abnormal cells using special staining techniques[1]. However, as the internationally used staining method for cervical cells is the Papanicolaou (Pap) method, we apply our method of cervical cell classification to regularly stained Pap smear images.

In sections 2, 3, and 4 we detail the imaging system used to capture the cell samples, the resulting data set, and nucleus segmentation method. Sections 5 and 6 discuss feature selection and classification, while section 7 details the results of our work and discusses points raised by the results. We finish with a discussion on the future direction of our work and conclusions in sections 8 and 9.

2   Imaging System

Image capture was performed by the use of an imaging system consisting of a Nikon Labophot 2A microscope and JVC TK1070E colour video camera. A ×100 oil-immersion lens with numerical aperture of 1.25 provided images with resolution near the optical limit ($0.25\mu m$). The digitised images were of 7-bit photometric resolution with a per-pixel spatial resolution of $0.12\mu m$. Each image was stored as gray-scale, using standard colour-to-grayscale RGB weighting.

3   The Data Set

The data consists of a set of 61 cells captured from 6 cervical slides processed using ThinPrep® slide preparation[1] and Papanicol-

---
[1] Cytyc Corporation, Massachusetts, U.S.A.

297

aou staining. A total of 31 abnormal cells, ranging from mild dysplasia (CIN1) to Carcinoma-in-situ (CIS) were captured from 5 abnormal slides, while 30 normal cells were captured from both the 5 abnormal and 1 normal slides. All cells were captured during the one capture session, and the imaging of normal and abnormal cells were randomly interspersed. All cells were classified through the microscope before capture, by a Cytotechnologist. Examples of typical normal and abnormal cells are shown in Figure 1. It can be appreciated from this figure that it is quite difficult for the untrained observer to distinguish visual differences between normal and abnormal cells in isolation.
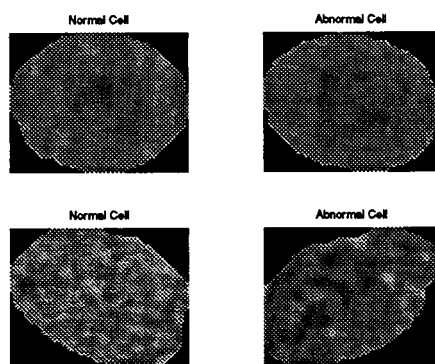

Figure 1: Examples of normal and abnormal cervical cell nuclei, segmented using the technique described in Section 4

4 Nuclear Segmentation

Following image capture, each nucleus was segmented from its surrounding cytoplasm using a series of automated fast morphological transforms with octagonal structuring elements[10]. Each gray-scale cell image I was first globally thresholded, resulting in an incomplete segmentation of the nucleus in binary form I_T. Cytoplasmic artifacts were removed by performing a closing of the image using a structuring element smaller than the smallest nucleus, and nuclear inhomogeneity was corrected by an opening of similar size. The resulting

binary image was then used as a mask for extracting the nucleus from the gray-scale image,

$$I_S = I * ((I_T \bullet B_1) \circ B_2), \qquad (1)$$

where I_S is the resulting segmented image, I_T is the thresholded image, and B_1 and B_2 are structuring elements. The symbols $\bullet$ and $\circ$ denote morphological closing and opening, respectively. The method proved quite robust, however on some darkly stained cells the initial threshold required adjustment. It is felt that local thresholding may be more appropriate and this will be investigated at a future date, allowing the segmentation to be fully automated.

5 Gray Level Co-Occurrence Matrix Texture Estimates

Gray level co-occurrence matrices (GLCM) provide a well known statistical method of extracting textural features from images. Connors[3], Haralick[8], and others used GLCM features as measures of texture in satellite images, while Albregtsen[1] used these measures for the classification of mouse liver cell nuclei. Connors and Harlow[2] have shown that GLCM is a more powerful technique than Gray Level Difference Matrix (GLDM), Gray Level Run Length Method (GLRLM), and the Power Spectral Method (PSM). See [2] and [7] for an overview of GLCM and other texture operators.

Prior to matrix calculation, we requantise each image to 16 levels while maintaining the histogram shape. We calculate the following 7 GLCM feature measures[2, 7]:

Energy:

$$Energy = \sum_{i,j} P(i,j)^2; \qquad (2)$$

Inverse Difference Moment (Local Homogeneity):

$$IDM = \sum_{i,j} \frac{1}{1 + (i-j)^2} P(i,j); \qquad (3)$$

Entropy:

$$Entropy = -\sum_{i,j} P(i,j) \log P(i,j); \qquad (4)$$

298

Correlation:

$$Corr = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)P(i,j))}{\sigma_i \sigma_j}; \quad (5)$$

Inertia:

$$IN = \sum_{i,j}(i - j)^2 P(i,j); \quad (6)$$

Cluster Shade:

$$CS = \sum_{i,j}((i - \mu_i) + (j - \mu_j))^3 P(i,j); \quad (7)$$

Cluster Prominence:

$$CP = \sum_{i,j}((i - \mu_i) + (j - \mu_j))^4 P(i,j). \quad (8)$$

Each feature measure is obtained for 8 distances ($d = 1, 3, 5, \ldots, 15$) and 4 angles ($\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$). For each distance, we average across the 4 angles, to provide "rotation invariance" to the technique. We then use a statistical method based on a χ^2 measure, to find the 3 distances d_i which best capture the nuclear texture information[14]. This results in a measurement space of $7 \times 3 = 21$ dimensions.

Our method of GLCM calculation differs slightly from the standard approach, in terms of the distance measure d and in the method of averaging the features over the 4 angles. The metric d is commonly defined as $\max\{|k - m|, |l - n|\}$, where (k, l) and (m, n) are the Cartesian coordinates of the image pixel pairs. While being somewhat inaccurate for angles of 45° and 135°, it allows quicker computation for digitised images. We propose a more precise method by using exact distance and angle measures, and use bilinear interpolation of gray levels for angles of 45° and 135°.

Averaging across the four angles is usually accomplished by averaging the four GLCM matrices and calculating features from this one matrix. We choose instead to calculate features for each of the four matrices, and then average the values of the corresponding features. This approach, while computationally more expensive, reduces the effect of averaging on features such as Energy, Cluster Shade, and Cluster Prominence, which are calculated through raising matrix elements to powers greater than unity.

6 Feature Selection and Classification

Many of the 21 measures are correlated, as all were derived from the seven texture measures detailed above. To find an optimal feature set of reduced dimension, we implemented a feature selection method of the "generalised simultaneous forward selection of n/backward elimination of m" class[6]. In particular, our algorithm forward-selects 2 new features and backward-eliminates one feature, thus capturing feature pairs with higher-order discriminatory powers. We begin with an exhaustive search for the best feature pair, and then augment this subset using the add-2/subtract 1 selection method. Kittler[9] reported that this method almost always gave optimal results and computationally was comparable to less optimal approaches.

We chose the parametric Bhattacharyya distance measure J_B[6] for determining feature separation, as it allows closed form evaluation and provides upper and lower bounds on the Bayes classification error ε.

Classification was performed by a general Bayes decision function for Gaussian feature distributions with unequal variance-covariance matrices[5]. The resulting decision boundary ($d_i = d_j$) is of hyperquadric form:

$$d_i(x) = \log P_{\omega_i} - \frac{1}{2}\log|C_i| \\ -\frac{1}{2}\left[(x - \bar{x}_i)' C_i^{-1}(x - \bar{x}_i)\right]. (9)$$

To obtain a more accurate estimation of the true classification performance, and given the small data set, we applied 3-fold cross-validation[13].

For each of the 3 trials, a training set of 41 cells and a test set of 20 cells was selected at random from the data. For each trial the training set was used to select features according to the above procedure and the performance of the resulting classifier was evaluated on the test set.

7 Results and Discussion

For all seven feature types, the χ^2 measure indicated maximum GLCM structure content for the distances $d = 1, 3$, and 5, resulting in a feature space of 21 dimensions. Subsets of the 21 measures of varying dimension were determined using the feature selection algorithm previously described. For each trial and each feature set, the corresponding estimated average Bayes misclassification error and the actual error were determined. Table 1 details the best feature and best feature pair, together with their Bhattacharyya distance measure J_B and actual misclassification rate for each trial. It must be stated that the classification error rates quoted in Table 1 are approximate, due to the small data set size. However, as will be shown later, there is strong separation between the two classes.

	Trial 1	Trial 2	Trial 3
Feat-ures	$IN_{d=3}$	$IN_{d=3}$	$IN_{d=3}$
J_B	1.2	1.1	1.4
Misclass-ification	2/20	1/20	1/20

(a)

	Trial 1	Trial 2	Trial 3
Feat-ures	$IN_{d=3}$ $IDM_{d=3}$	$IN_{d=3}$ $IDM_{d=3}$	$IN_{d=3}$ $IDM_{d=3}$
J_B	1.8	1.6	1.8
Misclass-ification	1/20	0/20	1/20

(b)

Table 1: Feature sets, their corresponding Bhattacharyya distance measure J_B, and actual classification error: (a) for best single feature; (b) best pair of features.

From Table 1 we can see that the same features have been selected in each trial indicating stability in the feature selection process. Additionally, the distance measures and actual classification performances are similar across trials providing some measure of confidence in the performance of the classification scheme.

Figure 2 shows the scatter plot of the two best features, $IN_{d=3}$, and $IDM_{d=3}$. It can be seen that both classes of cells are, to a large degree, well clustered and separated, indicating that two features alone provide good discrimination between the classes. Indeed, the misclassification rate does not decrease further until a total of five features are used. There also appears to be negative correlation between these features, and this is not surprising. The feature IN weights diagonal elements less than off-diagonal elements, and the converse is true for IDM. On the surface it may seem surprising that IDM be chosen as the second feature, considering its correlation with IN. However, as shown in Figure 2 it is the significant difference in class *means* that is providing the increased discriminatory power.

The feature $IN_{d=3}$ alone provides powerful discrimination, with the use of a simple threshold providing only 3 misclassifications. Finally, it is interesting to note that abnormal cells appear to form tighter clusters in feature space. This point was also noticed by Albregtsen[1].
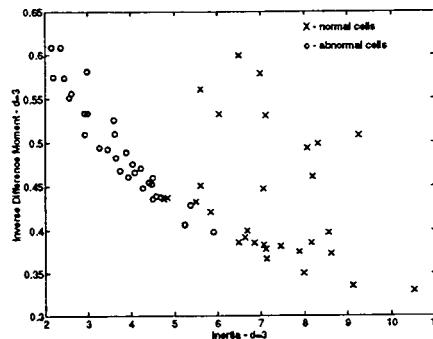

Figure 2: Scatter plot of best feature pair

8 Future Work

The limited data set size of 61 cells needs to be increased to better define the true discriminatory power of the selected features. To facilitate this the current data set will be augmented with other cytologist-classified cells.

The variability of cell staining between slides is well known for causing variability in feature values and also segmentation problems. While our feature extraction technique is invariant to inter-cell and inter-slide staining variations, our segmentation method proved to be not so robust. Further work in this area is envisaged with the possible use of local thresholding.

9 Conclusions

It has been shown that, based on a small sample set, GLCM texture features provide a means of discriminating between normal and abnormal Pap smear cervical cells. While many of the features provide little discriminatory power when used individually, groups of features with higher-order discriminatory powers have shown good classification performance. We have observed that a pair of features seem to adequately detect cell differences. Caution should be used when drawing conclusions from the results, due to the small data set used. However, the results achieved in this preliminary work are encouraging and provide a good foundation for further research in this area.

10 Acknowledgements

References

[1] Fritz Albregtsen, Yogesan Kanagasingam, George Farrants, and Havard E. Danielsen. Texture discrimination of normal and malignant mouse liver cell nuclei. In P. Johansen and S. Olsen, editors, *Theory and Applications of Image Analysis: Selected Papers from The 7th Scandinavian Conference on Image Analysis*, pages 324–335. World Scientific Publishing Company, 1992.

[2] Richard W. Connors and Charles A. Harlow. A theoretical comparison of texture algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2(3):204–222, 1980.

[3] Richard W. Connors, Mohan M. Trivedi, and Charles A. Harlow. Segmentation of a high-resolution urban scene using texture operators. *Computer Vision, Graphics, and Image Processing*, 25:273–310, 1984.

[4] G.L. Garcia. *Feasibility Of Contextual Analsyis in an Automated Cervical Prescreening System*. PhD thesis, Worcester Polytechnic Institute, 1986.

[5] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Addison-Wesley Publishing, U.S.A., 1992.

[6] D. J. Hand. *Discrimination and Classification*. John Wiley and Sons, USA, 1981.

[7] R.M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, 1979.

[8] R.M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3:610–621, 1973.

[9] J. Kittler. Feature set search algorithms. In C. H. Chen, editor, *Pattern Recognition and Signal Processing*. Sijthoff and Noordhoff, The Netherlands, 1978.

[10] Yong H. Lee. Algorithms for mathematical morphological operations with flat top structuring elements. *SPIE Applications of Digital Image Processing*, pages 33–45, 1982.

[11] Raymond J. Madachy and Yao S. Fu. Image analysis for automatic classification of mitotic cervical cells. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, pages 372–374, New Orleans, LA, USA, 4-7 November 1988. IEEE, New York, NY, USA.

[12] Branko Palcic and Calum MacAulay. Malignancy associated changes - can they be employed clinically? In G. Wied, P. Bartels, D. Rosenthal, and U. Schenck, editors, *Compendium on the Computerized Cytology and Histology Laboratory*, pages 157–165. Tutorials of Cytology, Chicago, U.S.A., 1994.

[13] Sholom M. Weiss and Casimir A. Kulikowski. *Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Networks, Machine Learning, and Expert Systems*. Morgan Kaufmann, San Mateo, 1990.

[14] Steven W. Zucker and Terzopoulos Demetri. Finding structure in co-occurance matrices for texture analysis. *Computer Graphics and Image Processing*, 12:286–308, 1980.