On Generating Cell Exemplars for Detection of Mitotic Cells in Breast Cancer Histopathology Images

Nada A. Aloraidi[1], Korsuk Sirinukunwattana[2], Adnan M. Khan[3], Nasir M. Rajpoot[4,*]
*nasir.rajpoot@ieee.org

***Abstract*— Mitotic activity is one of the main criteria that pathologists use to decide the grade of the cancer. Computerised mitotic cell detection promises to bring efficiency and accuracy into the grading process. However, detection and classification of mitotic cells in breast cancer histopathology images is a challenging task because of the large intra-class variation in the visual appearance of mitotic cells in various stages of cell division life cycle. In this paper, we test the hypothesis that cells in histopathology images can be effectively represented using cell exemplars derived from sub-images of various kinds of cells in an image for the purposes of mitotic cell classification. We compare three methods for generating exemplar cells. The methods have been evaluated in terms of classification performance on the MITOS dataset. The experimental results demonstrate that eigencells combined with support vector machines produce reasonably high detection accuracy among all the methods.**

I. INTRODUCTION

According to the world health organisation, breast cancer is the most common cancer in women [1]. The gold standard for breast cancer grading largely remains to be the work of human experts so far, whereby histopathologists visually inspect the tissues slides under the microscope and assign a grade to the slide based on their experience. The grading of cancer is important, since it represents the aggressiveness of the tumour. Thus the information provided by the grading process helps as a guide for the treatment options.

There are different available grading systems, but the modified Scarf-Bloom-Richardson [2] is the most widely used tumour grading system for breast cancer. It consists of three components: nuclear pleomorphism, degree of tubule formation and mitotic activity. Nuclear pleomorphism is a measure of the difference in size and shape of nuclei in the tumour cells compared to the normal cells. Degree of tubule formation assesses the extent of normal duct structure in tumour tissue and mitotic activity indicates how fast the tumour cells are growing. In general, each component is given a score of 1 to 3 (1 being the best and 3 the worst) and the scores of all three components are added to determine the breast cancer tumour grade. The lowest possible score (1+1+1=3) is given to tumours that all form tubules, have well differentiated nuclei and have a low mitotic rate in 10 High Power Fields (HPFs), which are regions of interest of the cancerous tissue slide examined at high magnification. The highest possible score is 9 (3+3+3=9) which indicates high grade tumour. A group of researchers has also shown that *mitotic rate* alone can be as predictive as the three factors combined [3].

A vast majority of the computerised approaches for MC classification (such as those presented in [4]–[9]) work by first identifying candidate objects or locations that are then accepted (or rejected) as MCs based on some similarity criterion [10]. The candidate extraction phase often makes use of the colour distinctiveness of MCs by building statistical models [4] or by performing thresholding. This is because the intensity of MCs is noticeably darker than normal epithelial nuclei and only comparable to apoptotic, necrotic or compressed (artefact from the tissue preparation) and lymphocyte nuclei. These local intensity minima detection or pixel-wise classification methods are sometimes followed by refinement of the detected regions by morphological operations and/or active contours segmentation [6], [7]. Khan *et al*. [4] add an extra layer of preprocessing where they perform stain normalisation [11] and tumour segmentation [12] to facilitate the MC detection. In the second stage, more specialised features ranging from basic features (morphological, geometrical and textural) to more specialised features (such as those learned from deep convolutional neural networks) are used to train a classification framework [4]–[8].

In this paper, we test the following hypothesis: can a cell in a histopathology image be *effectively* represented using a linear combination of some *exemplar* cells? In other words, we address the following question: can we extract some exemplar cells which can be used as a basis for representing cells? Our test of *effective representation* is whether or not the coefficients of a cell represented as a linear combination of the exemplar cells be used for discriminating between mitotic and non-mitotic cells in breast cancer histopathology images? We use a broad definition of exemplar cells in that the *exemplar* cells are derived from sub-images consisting of cell nuclei and surrounding context. In that sense, the closest work to this one is another approach recently developed in our group by the name of *Cell Words* which employs a discriminative dictionary learning approach to compute dictionary atoms for minimising the reconstruction error and maximising the discrimination performance [13].

In a representation framework using exemplar cells for the purpose of mitotic cell classification, we present a com-

[1]Nada A. Aloraidi is with College of Engineering, Qatar University, Qatar.
[2]Korsuk Sirinukunwattana is with College of Engineering, Qatar University, Qatar and Department of Computer Science, University of Warwick, UK.
[3]Adnan M. Khan is with College of Engineering, Qatar University, Qatar and Department of Computer Science, University of Warwick, UK.
[3]Nasir M. Rajpoot is with College of Engineering, Qatar University, Qatar and Department of Computer Science, University of Warwick, UK.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

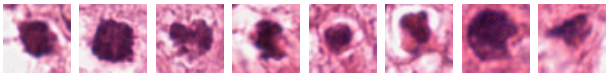

Fig. 1. Visual appearance of different cells in breast histopathological images. First 4 images (from left) are mitotic cells and last 4 images are non-mitotic cell images.

parison between three different approaches for generating exemplar cells: using the centroids of k-means clustering on sub-images containing different kinds of cells and their context, random projection sub-images, and eigencells [14]. We would like to emphasise here that the approach presented in [14] was tested on stained peripheral blood cells which are not very different in their appearance and context. In comparison, our case study of cells in breast cancer histopathology images is quite challenging (see, for example, various kinds of mitotic cells and non-mitotic cells in Fig. 1).

The remainder of this paper is organised as follows: Section II presents the methods which include method for generating candidate patches as well as the three different methods for generating cell exemplars using visual words, random projections and principal component analysis. Section III describes the experimental framework used to evaluate the performance of the the three methods. Section IV concludes the discussion with some possible directions for future research.

II. Materials & Methods

A. Materials

We evaluate our method on the publicly available MITOS dataset [15] that consists of 50 images corresponding to high power fields (HPFs) in 5 different breast tissue biopsy slides stained with hematoxylin and eosin (H&E). More than 325 MCs are visible in the MITOS dataset. Each HPF represents a $512\times512\mu m^2$ area, and is acquired using three different equipment: two digital slide scanners and a multispectral microscope. We consider images acquired by the Aperio slide scanner that has a resolution of 0.2456μm per pixel, Expert pathologists manually annotated all visible MCs. More details about the MITOS dataset can be found in [15].

We first perform candidate detection (as outlined in section II-B.1). Next, we divide all the candidate patches into two sets: S_{train} and S_{test}. Both S_{train} and S_{test} roughly contain 70% and 30% of total candidate patches respectively. RGB patches are converted to greyscale before using them as inputs to the proposed method.

B. Method

1) Candidate Detection: We extract candidate image patches using intensity thresholding. The positive candidates (patches that have MC) are segmented based on the ground truth, where the mitotic cells are marked in yellow. Negative candidates (patches that do not have MC) are segmented using intensity threshold for dark areas in the images. All the areas containing positive candidates are excluded from the search for negative patches. After the segmentation has been performed, all connected regions with an area between $10\mu m^2$ and $100\mu m^2$ are considered as candidates. Lastly, if centroids of the two candidate MCs are less than $4.5\mu m$ (20 pixels) apart, one of the two candidates is removed. A patch of 121×121 pixels around the centroid of the candidate MC is extracted.

2) Nuclear Alignment: In order to make the candidate patches translation and rotation invariant, candidate patch rotation is carried out to ensure that all the candidate MCs are aligned at the centre of the patches and have the same orientation. We use principal component analysis (PCA) to perform rotation. For this purpose, we first convert a patch into blue ratio image to get the spatial distribution of nuclear content of a patch [16]. PCA is then performed on the coordinates of nuclei pixels, to yield a rotation matrix in the form of principal component coefficients. In case of multiple nuclei, only the nucleus closest to the centroid of the image patch is considered. Finally, a candidate patch of size 51×51 is cropped at the centre of the patch for further processing. For details of the effect of candidate patch rotation, reader is referred to [13].

3) Cell Exemplars Generation: Let $\mathbf{s_i} \in \mathbb{R}^n$ denote a vector containing grey intensities of a patch i, with n be the total number of pixels in the patch. Here, we consider three approach for cell exemplar generation.

Visual Words: In the methodology of bag-of-word representation [17], image descriptors are first extracted from each training image. These descriptors often have high-dimensionality. To handle these descriptors in an economical way, a small set of exemplars (or visual words) which represent some common characteristics among the descriptors are then constructed. Various approaches have been proposed to construct visual words [18], [19]. One of the most simplest and commonly used method is k-means clustering algorithm, in which we use a centroid of each cluster of image descriptors as a visual word representing all the descriptors in the cluster. An image can then be represented through a feature calculated based on a set of visual words, which in general has significantly lower dimensionality as compared to the original image descriptors.

A set of k visual words $\{\mathbf{v}_1,...,\mathbf{v}_k\}$, $\mathbf{v}_j \in \mathbb{R}^n$ is constructed from $\mathbf{s}_i \in S_{train}$ using k-means. For each patch i, a feature vector $\mathbf{x}_i$ is defined by

$$\mathbf{x}_i = [x_{i,1},...,x_{i,k}] \tag{1}$$

where $x_{i,j}$ is the dot product between $\mathbf{s}_i$ patch and the visual word $\mathbf{v}_j$. The value of the dot product can be interpreted as the similarity measure between two vectors. The greater the value, the more similar the two vectors.

Random Projection: A random projection is a dimensionality reduction method that project the high-dimensional data into a low-dimensional subspace using a random matrix. It has an attractive theoretical property in which the distance in low-dimension representation of data is well-preserved and the computational cost involved in the projection is significantly lower than that of commonly used methods such as PCA [20]. Random projection method has been successfully applied to various computer vision tasks such

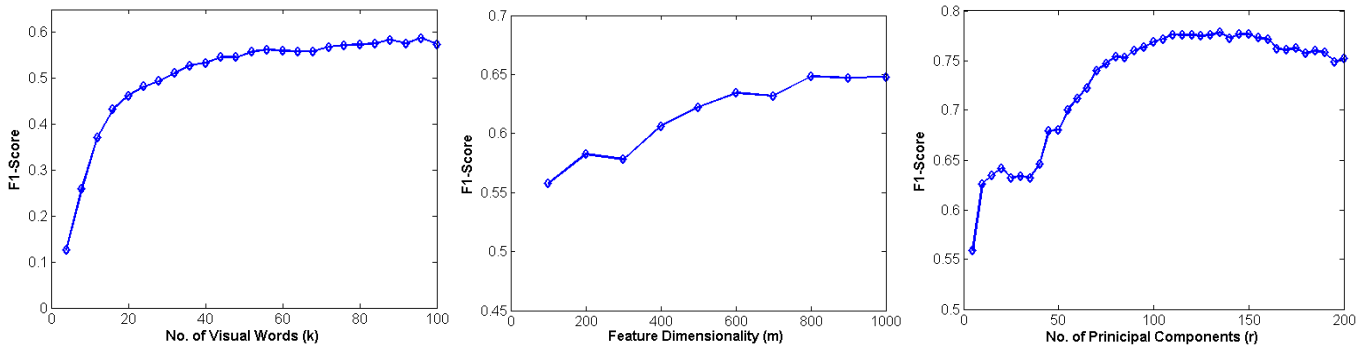

Fig. 2. (left) Effect of the number of visual words on the classification performance of the proposed system; (middle) Effect of the number of dimensions of image features (m) on the classification performance of the proposed system; (right) Effect of the number of principal components on the classification performance of the proposed system.

as face recognition [21], noise reduction [20], tumour segmentation [22] and object tracking [23]. Let $R \in \mathbb{R}^{m \times n}$ be a matrix whose elements are sample from a standard normal distribution, and its rows are linearly dependent and has unitary Euclidean norms. Here, $m \ll n$ is a dimension of a transformed data. A random projection feature $\mathbf{x}_i$ of a given patch $\mathbf{s}_i$ writes,

$$\mathbf{x}_i = R \times \mathbf{s}_i \in \mathbb{R}^m \tag{2}$$

EigenCells: Eigencells are generated using PCA that projects the data onto a low-dimension orthogonal subspace such that the correlation between observed variables is minimised. This results in a new coordinate system where the first coordinate (the first principal component) lies in the direction that captures the highest variation in the data. The variations of the data in the second coordinate and so on are descending in order. We perform PCA on the grey intensity descriptors of the whole data set, and use the first $r \ll n$ principal components of the projected data $\mathbf{x}_i \in \mathbb{R}^r$ as a feature vector.

III. Experimental Results & Discussion

In order to reduce the effect from random partition of S_{train} and S_{test}, we perform 50 repetitions of each experiment. For each repetition i, we generate the set of candidates for S_{test} and count the number of True Positives $N_{tp,i}$ (i.e. detections whose centroids are closer than $8\mu m$ from the ground truth centroid), False Positives $N_{fp,i}$ and False Negatives $N_{fn,i}$. The total numbers of True Positives, False Negatives, and False Positives are given by $(N_{tp} = \sum_{i=1}^{50} N_{tp,i})$, $(N_{fn} = \sum_{i=1}^{50} N_{fn,i})$, and $(N_{fp} = \sum_{i=1}^{50} N_{fp,i})$, respectively. Then following performance measures are calculated: precision (P = $N_{tp}/(N_{tp} + N_{fp})$), recall (R = $N_{tp}/(N_{tp} + N_{fn})$) and F1-score (F1 = 2PR/(P+R)).

In the experiments, we construct classifiers from linear support vector machine. A value of a block constraint parameter which penalises misclassification is set to 1. In Fig. 2, we evaluate the effect of the number of visual words (k), the dimensionality of random projected feature (m), and the number of principal components (r) on the performance of classification models based on cell exemplars generated via k-means, random projection, and eigencells, respectively.

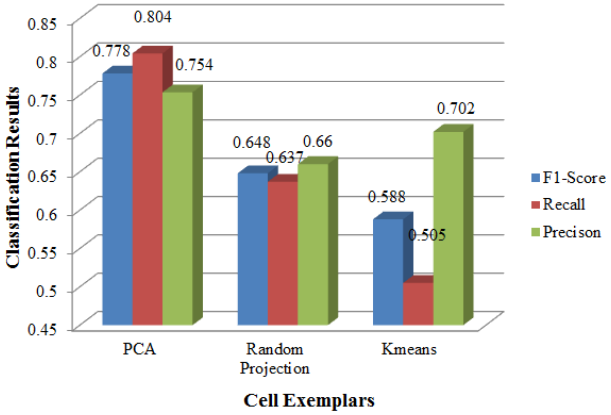

Fig. 3. Comparative performance of the three cell exemplars.

Generally, the classification performance tends to increase and converges to some values as the dimensionality of exemplars increase. This implies that variation of cells is better captured in high dimensional space. However, the trend of classification performance of eigencells is noticeably different from the others as the performance drops down after some point. This means that after $r = 135$, the principal components start to provide more non-discriminative features among mitotic and non-mitotic classes. This, therefore, reduces the discrimination power of the classification model.

Fig. 3 compares the highest classification performance (f1-score, precision, and recall) of different classification models. The classification model based on k-means attains the highest f1-score of 0.588 at $k = 96$. The model based on random projection attains the highest f1-score of 0.648 when $m = 800$, and the model obtained from eigencells attains the highest f1-score of 0.778 when $r = 135$. We can see that within the methods for exemplar cell extraction, eigencells yields the best performance with the reasonable number of principal components. Fig. 4 illustrates cell exemplars from k-means and eigencells.

IV. Conclusions

We present a comparison of three distinct methods to extract exemplar cells which can be used as a basis for

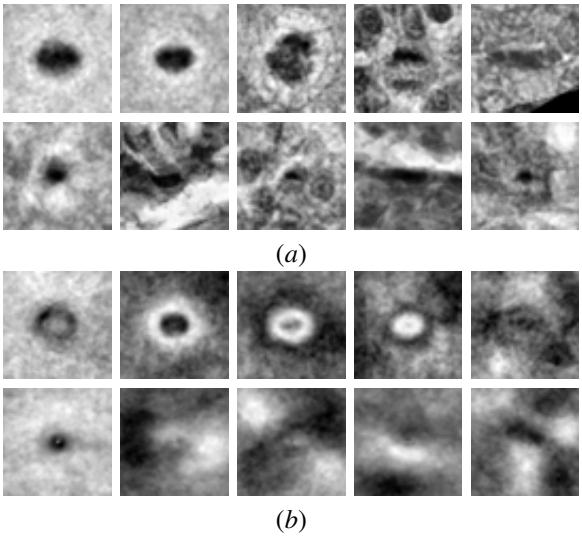

Fig. 4. (a) Top 5 exemplars obtained using kmeans for mitotic (first row) and non-mitotic class (second row); (b) Top 5 exemplars obtained using eigencells for mitotic (first row) and non-mitotic class (second row);

representing mitotic cells in breast histopathology images. Experimental results demonstrate high detection accuracy when exemplars are generated using principal component analysis. Potential future directions include extension of the same paradigm for detecting other types of cells [24] in differently stained histological images [25].

ACKNOWLEDGEMENTS

This work is partly supported by QNRF grant no NPRP-5-1345-1-228.

REFERENCES

[1] W. health Organisation (WHO), "Breast cancer: prevention and control." http://www.who.int/cancer/detection/breastcancer/en/, 2014. [Online; accessed 7-April-2014].

[2] C. Elston and I. Ellis, "Pathological prognostic factors in breast cancer. I. the value of histological grade in breast cancer: experience from a large study with long-term follow-up," *Histopathology*, vol. 19, no. 5, pp. 403–410, 1991.

[3] F. Clayton, "Pathologic correlates of survival in 378 lymph node-negative infiltrating ductal breast carcinomas. Mitotic count is the best single predictor," *Cancer*, vol. 68, no. 6, pp. 1309–1317, 1991.

[4] A. M. Khan, H. ElDaly, and R. N. M, "A Gamma-Gaussian mixture model for detection of mitotic cells in breast cancer histopathology images," *Journal of Pathology Informatics*, vol. 4, no. 11, 2013.

[5] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 411–418, Springer, 2013.

[6] H. Irshad, "Automated mitosis detection in histopathology using morphological and multi-channel statistics features," *Journal of Pathology Informatics*, vol. 4, 2013.

[7] M. Veta, P. van Diest, and J. Pluim, "Detecting mitotic figures in breast cancer histopathology images," in *SPIE Medical Imaging*, vol. 867607, International Society for Optics and Photonics, 2013.

[8] A. Tashk, M. S. Helfroush, H. Danyali, and M. Akbarzadeh, "An automatic mitosis detection method for breast cancer histopathology slide images based on objective and pixel-wise textural features classification," in *5th International Conference on Information and Knowledge Technology (IKT)*, pp. 406–410, IEEE, 2013.

[9] C. D. Malon and E. Cosatto, "Classification of mitotic figures with convolutional neural networks and seeded blob features," *Journal of Pathology Informatics*, vol. 4, 2013.

[10] M. Veta, P. J. v. Diest, S. M. Willems, H. Wang, A. Madabhushi, A. Cruz-Roa, F. Gonzalez, A. B. L. Larsen, J. S. Vestergaard, A. B. Dahl, D. C. Cirean, J. Schmidhuber, A. Giusti, L. M. Gambardella, F. B. Tek, T. Walter, C.-W. Wang, S. Kondo, B. J. Matuszewski, F. Precioso, V. Snell, J. Kittler, T. E. de Campos, A. M. Khan, N. M. Rajpoot, E. Arkoumani, M. M. Lacle, M. A. Viergever, and J. P. Pluim, "Assessment of algorithms for mitosis detection in breast cancer histopathology images," Submitted to *Medical Image Analysis*, 2014.

[11] A. M. Khan, N. Rajpoot, D. Treanor, and D. Magee, "A non-linear mapping approach to stain normalisation in digital histopathology images using image-specific colour deconvolution," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 6, pp. 1729–1738, 2014.

[12] A. M. Khan, H. El-Daly, E. Simmons, and N. M. Rajpoot, "HyMaP: A hybrid magnitude-phase approach to unsupervised segmentation of tumor areas in breast cancer histology images," *Journal of Pathology Informatics*, vol. 4 (Suppl), 2013.

[13] A. M. Khan, K. Sirinukunwattana, and N. Rajpoot, "Cell words: Modelling the visual appearance of cells in histopathology images," *Computerized Medical Imaging and Graphics*, 2014.

[14] S. Sanei and T. Lee, "Bayesian classification of eigencells," in *International Conference on Image Processing (ICIP)*, vol. 2, pp. 929–932, IEEE, 2002.

[15] L. Roux, D. Racoceanu, N. Loménie, M. Kulikova, H. Irshad, J. Klossa, F. Capron, C. Genestie, G. Le Naour, and M. Gurcan, "Mitosis detection in breast cancer histological images," *Journal of Pathology Informatics*, vol. 4, no. 8, 2013.

[16] H. Chang, L. A. Loss, P. T. Spellman, A. Borowsky, and B. Parvin, "Batch-invariant nuclear segmentation in whole mount histology sections," in *9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 856–859, IEEE, 2012.

[17] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *ECCV Workshop on Statistical Learning in Computer Vision*, vol. 1, pp. 1–2, 2004.

[18] F. Jurie and B. Triggs, "Creating efficient codebooks for visual recognition," in *10th IEEE International Conference on Computer Vision (ICCV)*, vol. 1, pp. 604–610, IEEE, 2005.

[19] S. Lazebnik and M. Raginsky, "Supervised learning of quantizer codebooks by information loss minimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 7, pp. 1294–1309, 2009.

[20] E. Bingham and H. Mannila, "Random projection in dimensionality reduction: applications to image and text data," in *7th International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, pp. 245–250, ACM, 2001.

[21] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[22] A. M. Khan, H. El-Daly, and N. Rajpoot, "RanPEC: Random projections with ensemble clustering for segmentation of tumor areas in breast histology images," in *Medical Image Understanding and Analysis (MIUA)*, pp. 17–23, 2012.

[23] G. Tsagkatakis and A. Savakis, "A random projections model for object tracking under variable pose and multi-camera views," in *3rd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*, pp. 1–7, IEEE, 2009.

[24] A. M. Khan, K. Sirinukunwattana, and N. Rajpoot, "Geodesic geometric mean of regional covariance descriptors as an image-level descriptor for nuclear atypia grading in breast histology images," 2014. Submitted.

[25] A. M. Khan, A. F. Mohammed, S. A. Al-Hajri, H. M. A. Shamari, U. Qidwai, I. Mujeeb, and N. M. Rajpoot, "A novel system for scoring of hormone receptors in breast cancer histopathology slides," in *2nd Middle East Conference on Biomedical Engineering (MECBME)*, pp. 155–158, IEEE, 2014.