SPECTRAL BAND SELECTION FOR MITOSIS DETECTION IN HISTOPATHOLOGY

Humayun Irshad^{1,3}, Alexandre Gouaillard^{4,5}, Ludovic Roux^{1,3}, Daniel Racocanu^{2,3}

¹University Joseph Fourier, Grenoble 1, ²Sorbonne Universités, UPMC Univ Paris 06,

³IPAL CNRS UMI 2955, ⁴CoSMo Software, Boston, MA, USA, ⁵Temasys Communications, Singapore

ABSTRACT

This study aims at evaluating the accuracy of mitosis detection on multispectral histopathological images by developing a solution specifically designed to take advantage of multispectral information. The proposed framework includes a selection of spectral bands and focal plane, detection of candidate mitotic regions, computation of morphological & multispectral statistical features (MMSF) and study of different state-of-the-art classification methods for mitosis classification. This framework achieved 74% TPR, 76% PPV and 74% F-Measure on MITOS dataset. Our results indicate that selected multispectral bands contain discriminant information for mitotic figures, being therefore a very promising exploration area to improve the quality of the diagnosis assistance in histopathology.

Index Terms— histopathology, multispectral images, spectral band selection, mitosis detection

1. INTRODUCTION

Breast Cancer is the most commonly diagnosed cancer after skin cancers, and the second leading cause of cancer death, following lung cancer, among U.S. women. According to the World Health Organization, the reference process for breast cancer prognosis is histologic grading that combine tubule formation, nuclei atypia and mitotic counts. This assessment of tissue sample is synthesized into a diagnosis that would help the clinician to determine the best course of therapy. In histopathology, hematoxylin & eosin (H&E) is a well-established staining technique, exploiting intensity of stains in the tissue images to quantify the nuclei and other structures related to cancer developments. Manual counting of mitosis is normal practice in histopathology, but automating the process could reduce its time, costs, and inter- and intra-reader variations.

Multispectral imaging (MSI) has the advantage to retrieve spectrally resolved information of a tissue image scene at specific frequencies across the electromagnetic spectrum. MSI captures images with accurate spectral content, correlated with spatial information, by revealing the chemical and

the anatomic features of histopathology [1]. This modality provides option to biologists and pathologists to see beyond the RGB image planes to which they are accustomed. Recent publications [1, 2] have begun to explore the use of extra information contained in such spectral data. The added benefit of MSI for analysis in routine H&E histopathology, however, is still largely unknown, although some promising results are presented in [2, 3]. As far as we know, there is no existing study for mitosis detection in multispectral histopathology.

In our study, we propose a framework for mitosis detection in breast cancer MSI. The proposed framework addresses two important questions: First, does the spatial-spectral analysis on selected spectral bands (SBs) (as opposed to spatial analysis on single SB or spatial-spectral analysis of all the SBs) suffice for efficient classification of mitotic and non-mitotic figures. An obvious advantage of using selected SBs is its reduced computational and storage complexity. Second, is the multiple features set more effective for discrimination as compared to one type of features set.

2. DATASET

We evaluated the proposed framework on multispectral MITOS dataset [4]. The data set is made up of 200 images coming from five different slides scanned at 40X magnification using a 10 SBs microscope. The training (TR) data set consists of 140 images containing 224 mitotic figures and the evaluation (EV) data set consists of 60 images containing 98 mitotic figures. All SBs are in the visible spectrum. There is some spectrum overlapping for the SBs. In addition, for each spectral band, the digitization has been performed at 17 different focal planes, each focal plane being separated from the other by 500 nm.

3. PROPOSED FRAMEWORK

The proposed framework has five main steps as shown in Fig. 1. Step one performs a selection of the most informative focal plane. Step two is responsible for the selection of relevant SBs for the objective of mitotic figures detection. Candidates for mitotic figures are detected in step three. Then, in step four, a MMSF signature vector of intensity and texture information across selected SBs is computed for each

This work is supported in part by the French National Research Agency ANR, Project MICO under reference ANR-10-TECS-015.

Fig. 1. Proposed Framework.

detected candidate. During step five, candidates are classified into mitosis and non-mitosis classes using decision tree (DT), multilayer perceptron (MLP) as well as linear and non-linear support vector machine (L- and NL-SVM) classifiers.

3.1. Focal plane Selection using Maximum Gradient

For selection of focal plane, average gradient of mitotic figures from background regions was computed on all the focal planes. The focal plane that has maximum average gradient (best focus) is selected for the next steps of the framework.

3.2. SBs Selection

3.2.1. Method 1: Tissue Spectral Absorption

The main tissue components visible in the data set images can globally be categorized into adipose (fat), cytoplasm, mitotic nuclei and non-mitotic nuclei. We selected 200 image patches for each tissue components and computed the spectral absorption responses of each tissue components for the available 10 SBs as shown in Fig. 2. Overall, mitotic and non-mitotic nuclei have higher absorption response than cytoplasm and adipose. This is in accordance with the fact that hematoxylin stain concentrates mostly on nuclei, while cytoplasm is stained to a lower degree by eosin. The adipose curve has very low absorption response as it is stained neither by hematoxylin nor by eosin. Mitotic and non-mitotic nuclei have higher absorption response in green spectrum (SBs 5,6,7) than in red spectrum (SBs 8,9,0) or blue spectrum (SBs 2,3,4). While cytoplasm absorption response is also higher in green spectrum. As a whole, mitotic regions, as well as cytoplasm regions, have highest absorption response in SB 5. These absorption responses of multispectral data are thus able to differentiate between nuclei and other tissue components.

Fig. 2. Normalized absorption spectra of four tissue components in 10 SBs. SB 1 is separated from other SBs by dotted line as it covers the whole visible spectrum.

3.2.2. Method 2: H and E Spectral Absorption

To illustrate the possible correlation between SBs and the staining characteristics of the spectral samples, the plot of H and E dyes spectral absorptions are shown in Fig. 3 (this plot is derived from the work of Bautista and Yagi [5]). The H–E plot shows the difference of absorption between hematoxylin and eosin. The bands for which H–E is maximum are more suitable for discrimination between nuclei and cytoplasm. The absorption response of H is maximum in SBs 7 and 8 with almost zero E response. Hence, in this method, we reconstruct the spectrum of a pixel by using staining characteristics of tissue components for selection of the optimal number of SBs for mitosis discrimination in H&E stained MSI.

3.2.3. Method 3: mRMR Technique

In this method, mRMR technique [6] is used for SBs selection. Selection is based on two criteria; minimum redundancy $R(S, c)$ and maximum relevance $D(S, c)$. The relevancy of selected SBs to class labels has been measured by average of mutual information (MI) between each SB and class label.

Fig. 3. Normalized plot of the H and E dye absorption spectra in MSI and the difference of H and E.

Table 1. SBs Mutual Information (MI) Measure.

SBs	MI	Accumulated MI	Accumulated MI%
SB 8	3.60	3.60	33%
SB 9	3.59	0.95	42%
SB 7	3.38	0.94	51%
SB 6	3.18	0.93	60%
SB 2	3.16	0.92	69%
SB 1	3.11	0.91	78%
SB 3	3.05	0.89	86%
SB 0	2.99	0.88	91%
SB 4	2.94	0.85	95%
SB 5	2.85	0.82	100%

Their redundancy is measured by an average of MI between each pair of SBs. The average relevancy of selected SBs is:

$$D = \frac{1}{|S|} \sum_{s_i \in S} \text{MI}(s_i; c_j) \quad (1)$$

where S denotes the selected SBs set, $|S|$ denotes the number of selected SBs, c_j denotes j^{th} class label in class set C , s_i denotes i^{th} SBs in S and MI is mutual information between SB s_i and class label c_j . MI is computed using entropy as

$$\text{MI}(S; C) = H(S) - H(S|C) \quad (2)$$

where $H(S)$ and $H(S|C)$ are entropy functions that calculates the uncertainty of the SBs and the class labels. By maximizing D for full SBs set S_T , we can select a SBs set S with maximum relevance for discrimination of mitotic candidates with observing SBs set S_T .

It is likely that selected SBs have rich redundancy. Therefore, the following minimum redundancy $R(S, c)$ is added:

$$R = \frac{1}{|S|^2} \sum_{s_i, s_j \in S} \text{MI}(s_i; s_j) \quad (3)$$

$\text{MI}(s_i; s_j)$ is maximum when two SBs s_i and s_j have functional dependency and $\text{MI}(s_i; s_j) = 0$ if s_i and s_j are statistically independent. By minimizing R for selected SBs, we selected SBs set with minimum redundancy. The incremental search method was used to find the n SBs from the set $\{S_T - S_{n-1}\}$. The image samples, used in computation of spectral absorption of different tissue components, were divided into two classes. The non-mitosis class consisted of three tissue components including adipose, cytoplasm and nuclei, and the remaining samples belonged to mitosis class. We performed mRMR on these image samples and their MI, with ranking, are shown in Table 1.

3.3. Candidate Detection and Feature Computation

We performed candidate detection on the selected SB that has the highest MI. On the selected SB, we performed thresholding followed by morphological processing to eliminate small

Table 2. Different Rankings of SBs.

Method 1		Method 2		Method 3	
SB	Mitosis—Cytoplasm	SB	H—E	SB	MI
7	0.47	7	0.96	8	3.6
8	0.45	8	0.91	9	3.59
9	0.36	9	0.64	7	3.38
3	0.33	1	0.39	6	3.18
2	0.31	6	0.33	2	3.16
6	0.30	0	0.23	1	3.05
1	0.30	2	0.23	3	3.05
4	0.29	3	0.21	0	2.99
0	0.28	5	0.04	4	2.95
5	0.27	4	0	5	2.83

regions and fill holes. Finally, we selected candidates by filtering based on size of candidates and took a patch of size $16.65\mu\text{m} \times 16.65\mu\text{m}$. On patch, we computed co-occurrence and run length features in selected SBs like [7]. We also computed intensity and morphological features using segmented regions as explained in [7]. These morphological & multi-spectral statistical features (MMSF) are used to train different classifiers like DT, MLP, L-SVM and NL-SVM and eventually classify these candidates as mitosis and non-mitosis.

4. RESULTS AND DISCUSSION

4.1. SBs Selection

How many SBs are necessary for a good detection of mitotic figures? Which SBs are relevant for mitotic figure detection?

To discuss these two questions, we tried first to evaluate the contributions of each SBs using three different proposed methods as discussed in 3.2. The results are shown in Table 2. The SBs ranking in method one is based on difference between spectral absorption of mitotic nuclei and cytoplasm, while the SBs ranking in method two is based on difference between hematoxylin and eosin spectral absorption. These three rankings put the same SBs 7, 8 and 9 in top three positions. At the bottom of the table, there are SBs 4 and 5 for all three rankings. According to method two ranking, the difference between absorption response of H and E in SBs 4 and 5 are almost zero which represent that these two SBs are irrelevant for mitosis discrimination. Based on these analyses, we ignore SBs 4 and 5 for mitosis discrimination. Considering the available SBs and their rankings, our selection of SBs contains the following eight bands: 8, 9, 7, 6, 2, 1, 3, and 0. These methods took less than 1 second to compute SBs' ranking.

4.1.1. Classification Results using SBs Ranking

Using mRMR ranking, the results using different classifiers are shown in Figure 4. F-Measure (FM) increases while we

Fig. 4. Plot of FM using SBs selection. Result from using all SBs from left to the current, e.g. SB 2 result uses SB 8, 9, 7, 6, 2. This order is taken from the mRMR ranking.

Fig. 5. Comparison of proposed framework results with MITOS contest result.

add more SBs to the set of selected SBs. FM reaches a peak with a set of eight selected SBs, then it starts decreasing when adding more SBs. In case of few SBs MMSF, L-SVM, NL-SVM and DT report poor classification accuracy while MLP reports higher classification accuracy. As more SBs are selected L-SVM classifiers start performing better than other classifiers and reached maximum performance with first eight selected SBs.

The comparison of proposed framework results with MITOS contest results [4] are shown in Figure 5. Malon and Cosatto [3] method ranked first in the contest with 59% FM. In comparison with MITOS contestants, the proposed framework using selected SBs MMSF and L-SVM classifier, managed to achieve highest FM (74%). Our proposed framework has outperformed in comparison with selected MMSF on all SBs using feature selection technique [8]. In comparison with color framework [7], multispectral framework managed to get better results for mitosis detection on MITOS dataset. This clearly demonstrates that our proposed framework results in an improved ability to distinguish mitosis from other objects.

5. CONCLUSION

An automated mitosis detection framework for breast cancer MSI based on multispectral spatial features has been proposed. Based on MI of SBs and spectral absorption of

different tissue components and stains, SBs were selected for candidate detection and feature computation. The proposed framework outperformed the MITOS contest results with 25% improvement of F-Measure. In future work, we plan to investigate unmixing of bands as most SBs have overlapping area, which increase redundancy. The pre-selection of the focal plane (or volume) is also of great importance to reduce the complexity of the dataset and improve the actual performances to reach clinical operational acceptance expected by our professional consortia.

6. ACKNOWLEDGEMENT

This work is supported by the French National Research Agency ANR, project MICO, reference ANR-10-TECS-015.

7. REFERENCES

- [1] R M Levenson, A Fornari, and M Loda, "Multispectral imaging and pathology: seeing and doing more," *Expert Opinion on Medical Diagnostics*, vol. 2, no. 9, pp. 1067–1081, 2008.
- [2] X Wu, M Amrikachi, and S K Shah, "Embedding topic discovery in conditional random fields model for segmenting nuclei using multispectral data," *IEEE Trans. on Biomed. Eng.*, vol. 59, no. 6, pp. 1539–1549, 2012.
- [3] C D Malon and E Cosatto, "Classification of mitotic figures with convolutional neural networks and seeded blob features," *J of Path Inform*, vol. 4, pp. 9, 2013.
- [4] L Roux, D Racocanu, N Lomenie, M Kulikova, H Irshad, J Klossa, F Capron, C Genestie, G Le Naour, and M N Gurcan, "Mitosis detection in breast cancer histological images an icpr 2012 contest," *J of Path Inform*, vol. 4, pp. 8, 2013.
- [5] P A Bautista and Y Yagi, "Digital simulation of staining in histopathology multispectral images: enhancement and linear transformation of spectral transmittance," *J of Biomed Opt*, vol. 17, no. 5, 2012.
- [6] Long F Ding C Peng, H, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27 (8), pp. 1226–1238, 2005.
- [7] H Irshad, "Automated mitosis detection in histopathology using morphological and multi-channel statistics features," *J of Path Inform*, vol. 4, pp. 10, 2013.
- [8] H Irshad, *Automated Mitosis Detection in Color and Multi-spectral High-Content Images in Histopathology: Application to Breast Cancer Grading in Digital Pathology*, Ph.D. thesis, Univ Joseph Fourier, Grenoble, 2014.