

Multispectral Texture characterization: Application to Mitotic Detection in Breast Cancer Histopathology

Humayun Irshad, Ludovic Roux
University Joseph Fourier Grenoble
France

humayun.irshad, ludvoic.roux@ipal.cnrs.fr

Alexandre Gouallard
Institution2
First line of institution2 address

secondauthor@i2.org

Daniel Racocanu
University Pierre and Marie Curie
France

daniel.racocanu@upmc.fr

Abstract

Accurate detection of mitosis in breast cancer histopathology plays a critical role in the grading process. Manual detection and counting of mitosis is tedious and subject to considerable inter- and intra-reader variations. State-of-the-art methods have not significant accuracy on multispectral imaging in histopathology. This study aims at improving the accuracy of mitosis detection by selecting the multispectral spatial features including intensity and textures to enable clinical routine-compliant quality of mitosis discrimination from other objects. The proposed framework includes comprehensive analysis of spectral bands and z-stack focal planes and selection of multispectral spatial features, as a study of candidate detection in different spectral bands. This framework has been evaluated on MITOS multispectral dataset and achieved — % detection rate, — % precision and — % F-Measure.

1. Introduction

Breast Cancer (BC) is the most commonly diagnosed cancer after the skin cancers, and the second leading cause of cancer death, following lung cancer among U.S. women [26]. In 2012, an estimated 226,870 new cases of invasive BC and 39,510 BC deaths in women are reported in U.S. In addition, BC incidence and mortality rates have been increasing rapidly in economically less developed countries. According to World Health Organization, the reference process for breast cancer prognosis is histologic grading that combine tubule formation, nuclei atypia and mitotic counts [3, 10]. This assessment of tissue sample is synthesized

into a diagnosis that would help the clinician determine the best course of therapy. Several CAD solutions exist for the detection of tubule formation [31, 30] and nuclei atypia [7, 8, 5, 9], but only few are dedicated to mitotic counts (MC) [18, 19].

In histopathology, H&E is a well-established staining technique that exploits intensity of stains in the tissue images to quantify the nuclei and other structures related to cancer developments [2]. In this context, image-processing techniques are devoted to accurate and objective quantification and localization of cancer evolution in specific regions of the tissue such as cytoplasm, membranes and nuclei [29]. From the chromatic viewpoint, nuclear regions are characterized by non-uniform stain intensity and color, thus preventing a trivial classification based on color separation. In addition, the superposition of tissue layers, as well as the diffusion of the dyes on the tissue surface, may bring the stains to contaminate the background or other cellular regions, which are different from their specific target. Recent work [27, 18, 19] show great potential for CAD of histopathological datasets for breast cancer diagnosis.

One of the most difficult fields in histopathology imagery is spatial analysis, more specifically automated nuclei detection and classification [12]. The objective of nuclei classification is to assign different labels to different type of nuclei as normal, cancer, mitotic, apoptosis, lymphocytes etc. In addition, quantitative characterization is important not only for clinical applications (e.g., to reduce/eliminate inter- and intra-observer variation in diagnosis) but also for research applications (e.g., to understand the biological mechanisms of the disease process [16]).

Medical image analysis has been studied for several decades in cytology and numerous solutions [37, 35, 6, 32,

15] have thus been proposed in literature. The application of these solutions to histopathology is rather complicated due to radically different imaging modalities and highly complex image characteristics. Indeed, in the case of histology images, cellular structures and functions are studied embedded in the whole tissue structure, presenting various cells architecture (gland formation, DCIS) very difficult to handle with usual pattern recognition techniques.

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 and reveals the chemical and anatomic features of histopathology [21, 23]. This modality provides option to biologists and pathologists to see beyond the RGB image planes that they are accustomed to. Recent publications [11, 24, 39, 20] have begun to explore the use of extra information contained in such spectral data. Specifically, there have been comparisons of spectral methodologies which demonstrate the advantage of multispectral data [22, 14]. The added benefit of MSI for analysis in routine H&E histopathology imagery, however, is still largely unknown, although some promising results are presented in [33, 11, 20, 39]. As far as we know, there is no existing study of the advantage of MSI for automation of MCs in breast cancer histopathology. We proposed here to extend the already successful work of [19] to support MSI and illustrate the automation of MCs in BC histopathology.

The reminder of the paper is organized as follows. Section 2 reviews the state-of-the-art multispectral methods, particularly in object or region detection in histopathology, related to this research work. Section 3 describes the proposed framework for mitotic figure detection. Experiment and results are presented in section 4. Finally, the concluding remarks with future work are given in section 5.

2. Literature Review

The main idea for extracting texture features from MSI is the use of combined spectral and spatial information for discrimination of region or objects. We found few methods in the MSI literature for texture characterization of histopathological images. Some of them employed single band of MSI and other used multiple selected bands of MSI. Fernandez et al. [11] coupled high-throughput Fourier transform infra-red (FTIR) spectroscopic imaging of tissue microarrays with statistical pattern recognition of spectra indicative of endogenous molecular composition and demonstrate histopathologic characterization of prostatic tissue. They explicitly defined metrics consisting of spectral features that have a physical significance related to tissue biochemistry and facilitating the measurement of cell types.

The approach in [38] used hyperspectral images of colon biopsy slides whereby a classification algorithm was based

on spectral analysis to discriminate between normal and cancerous biopsies of the colon tissue. In this study of hyper-spectral cancer analysis, they used Laplacian eigenmaps to take into account the non-linear geometry in the design of learning algorithms and evaluated two approaches for spectral feature selection: Haar wavelet packet best bases (active sensing) and random projections. Boucheron et al. [4] presented an analysis of the utility of multispectral versus standard RGB imagery for pixel level classification of nuclei in H&E stained histopathological images. Recently, Masood et al. [28] proposed a colon biopsy classification method based on spatial analysis of hyperspectral image data from colon biopsy samples. Initially, using circular local binary pattern algorithm, spatial analysis of patterns is represented by a feature vector in selected spectral band. Later, classification is achieved using subspace projection methods like principal component analysis, linear component analysis and support vector machine.

Khelifi et al. [20] proposed a spatial and spectral gray level dependence method in order to extend the concept of gray level co-occurrence matrix by assuming the presence of texture joint information between spectral bands. Malon et al. [27] demonstrated a segmentation based features with convolutional neural networks using additional focal planes and spectral bands for identification of mitotic nuclei in breast cancer histopathology and achieved best classification accuracy (F-measure = 59%) on multispectral dataset during ICPR contest 2012 [34].

Recently, Wu et al. [39] proposed a multilayer conditional random field model using a combination of low-level cues and high-level contextual information for nuclei separation in high dimensional data set obtained through spectral microscopy. In this approach, the multilayer contextual information was extracted by an unsupervised topic discovery process from spectral images of microscopic specimen, which efficiently helps to suppress segmentation errors caused by intensity inhomogeneity and variable chromatin texture.

In the proposed methodology, we address the shortcomings of previous works, including (1) comprehensive analysis of multispectral spatial feature vector in all bands rather than single band [28, 40, 39] and (2) extracting multispectral spatial feature vector in order to discriminate mitotic figures from other nuclei and microscopic objects. The main novel contributions of proposed work are: (1) a multispectral spatial and morphology feature vector computation which inherit discriminant information from other nuclei and (2) an projection features from multispectral features vector for classification of mitotic figures in breast cancer histopathological images.

3. Proposed Framework

3.1. Dataset

We evaluated the proposed framework on multispectral MITOS dataset [1], a freely available mitosis dataset. The data set is made up of 50 high power fields (HPF) coming from five different slides scanned at 40X magnification using a 10 bands multispectral microscope. There are 10 HPFs per slide and each HPF has a size of $512 \times 512 \mu\text{m}^2$ (that is an area of 0.262mm^2). The spectral bands are all in the visible spectrum. In addition, for each spectral band, the digitization has been performed at 17 different focus planes (17 layers Z-stack), each plane being separated from the other by 500 nm. For one HPF, there are 170 gray scale images (10 spectral bands and 17 layers Z-stack for each spectral band). These 50 HPFs contain a total 322 mitotic cells. The training data set consists of 35 HPFs containing 224 mitotic cells and evaluation data set consists of 15 HPFs containing 98 mitotic cells [34]. Figure 1 shows the spectral coverage of each of the 10 spectral bands of the multispectral microscope.

3.2. Proposed Framework

In this paper, we propose a framework for MC based on spatial analysis of MSI in breast cancer histopathology as shown in Figure 2. Initially, a z-stack plane is selected based on gradient and band is selected based on histogram analysis. Then, candidates for mitotic figures are computed on selected band and z-stack plane. A multispectral feature vector is computed for all detected candidates including intensity and texture features in all bands of multispectral images. In addition, using segmented regions of detected candidates, morphological features are also computed. A feature selection algorithm is employed on this multispectral features vector to save the computation cost and to discard any redundancy in the data in order to improve classification accuracy. Classification is achieved using support vector machine (SVM), Bayesian network (BN) as well as decision tree (DT). A side advantage of performing the spatial analysis on a multiple band is to investigate whether improvement in accuracy can be achieved with carefully selected multispectral features over those methods [28, 40, 39] which use single band data.

3.2.1 Candidate Detection

Initially, we perform gradient analysis on all the focus planes and select focus plane 10 that is having averagely maximum gradient value. Using selected focus plane, we compute the spectral responses of mitotic nuclei and background regions for all the available 10 spectral bands as represented in Figure 3 while responses of mitotic and non mitotic nuclei are presented in Figure 4. The peak of the

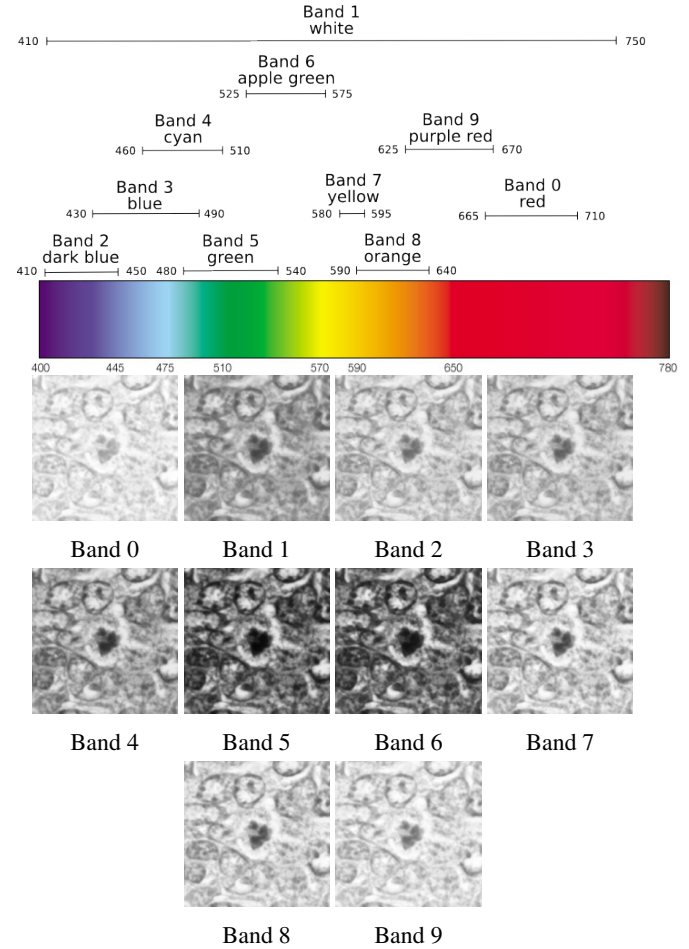


Figure 1: Spectral bands of the multispectral microscope and examples for each band.

mitotic and non-mitotic nuclei is almost similar but mitotic nuclei have different peaks than background regions. It can be said that multispectral data is able to discriminate between different tissue parts but spectral response of mitotic and non-mitotic nuclei is not distinguishable. This motivates us to perform spatial analysis on multispectral data for achieving reasonable classification nuclei into mitotic and non-mitotic nuclei. Using selected focus plane and spectral band, we perform binary thresholding followed by morphological processing to eliminate small regions and fill holes and later select the candidates by filtering based on minimum size ($37 \mu\text{m}^2$ i.e., 200 pixels) and maximum size ($1000 \mu\text{m}^2$ i.e., 5405 pixels) of mitotic nuclei while keeping the number of candidates to classify as low as possible.

3.2.2 Multispectral Features Computation

Instead of single band as in [4, 28, 40, 39], we compute multispectral spatial feature vectors having intensity and tex-

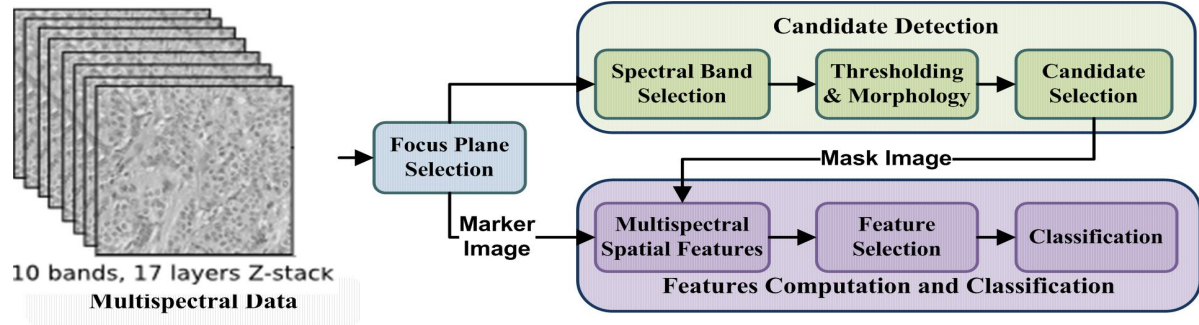


Figure 2: Proposed Framework.

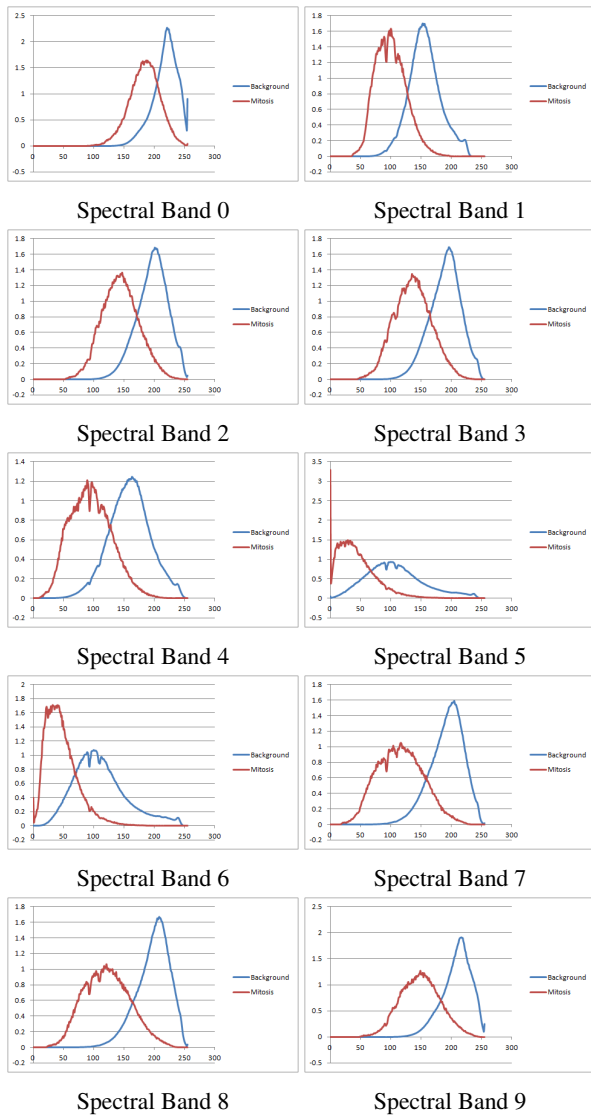


Figure 3: Histogram analysis of mitotic and background regions in 10 spectral bands.

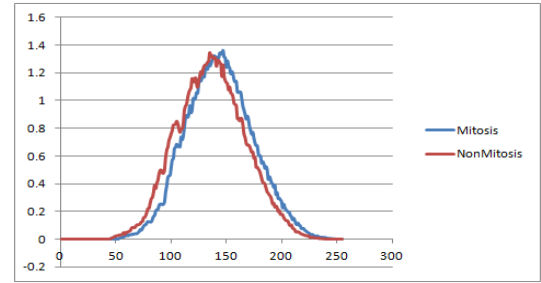


Figure 4: Mitosis and non-mitosis region histogram

tural features. In addition, we also include morphological features (such as area, roundness, elongation, perimeter and equivalent spherical perimeter) that are computed from regions segmented during candidate detection. The morphological features reflect the phenotype information of mitotic nuclei. Utilizing spatial information in multispectral bands, we compute five intensity-based features including mean, median, variance, kurtosis and skewness for candidate. This results in 50 multispectral intensity based features. Haralick co-occurrence (HC) [17] and run-length (RL) [13] features are computed with 1 displacement vector in four direction (0° , 45° , 90° , 135°) for all the spectral bands as in [19]. These multispectral textural features are rotationally invariant. So by making average in all four directions, eight HC and ten RL features are computed for each candidate in single spectral band. The resulted multispectral textural features vector consists of 80 HC features and 100 RL features for each candidate. The final multispectral features vector contains 235 features for each candidate.

3.2.3 Feature Selection and Classification

Conceptually, a large number of descriptive features are highly desirable for classification of object as mitosis or non-mitosis. However, when using all computed features (i.e. 235 features) for classification of candidate as mitosis or non-mitosis, the classification performance is poor.

Some features are irrelevant for classification and some features are redundant that represents duplication of features and does not provide additional class discriminatory information, degrading the classification performance. Later, we use consistency subset evaluation method [25] to select a subset of features that maximize the consistency in the class values. We evaluate the worth of subsets of features by the level of consistency in the class values using the projection of subset of features from training dataset. The consistencies of these subsets are not less than that of the full set of features. At last, we used these subsets in conjunction with a hill climbing search method, augmented with backtracking value 5, which looks for the smallest subset with consistency equal to that of the full set of features. This procedure achieved 86% reduction in the dimensionality of features set by selecting 35 features. The selected features contained two morphological, intensity and textural features in different spectral bands. The selected features set is used to train different classifiers like decision tree (Random Forest-RF), neural network (Multilayer Perceptron-MP) , Bayesian, linear SVM (L-SVM) and non-linear SVM (NL-SVM) classifier [36]. Throughout the experiments, the parameters used in RF classifier are MaxDepth = 10 and NumOfTree = 10, in MP are learning rate 0.3 and momentum = 0.2, in L-SVM classifier are bias = 1, cost = 1, eps = 0.01 and kernel = L2-loss SVM and in NL-SVM classifier are kernel = rbf, degree of kernel = 3, eps = 0.001 and loss = 0.1.

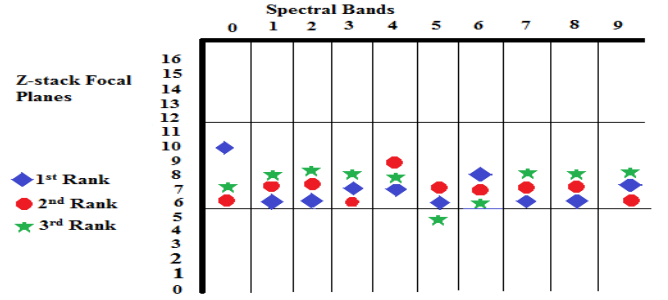
4. Experiments and Results

4.1. Experiments

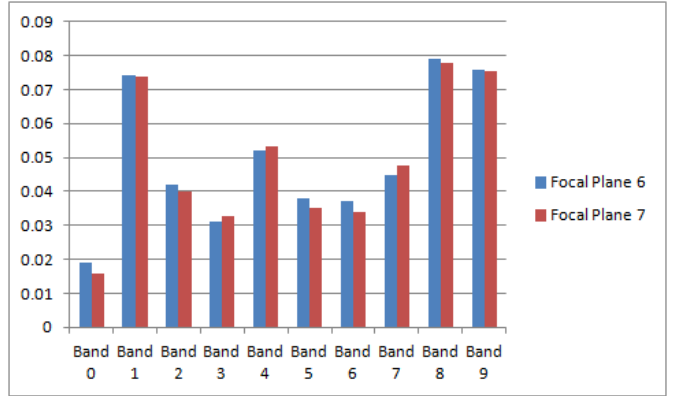
The proposed framework is evaluated on MITOS multispectral dataset [1], a freely available mitosis dataset. On training and evaluation sets, the candidate detection stage detects 5355 and 2235 candidates, containing 203 and 91 ground-truth (GT) mitosis from a total 224 and 98 GT mitosis, respectively. Therefore, among the entire candidate mitosis there are 5152 and 2144 non-mitosis in the training and evaluation sets. The candidate detection stage generates a large number of non-mitosis and misses 21 and 7 GT mitosis, from training and respectively, from evaluation sets.

4.2. Results

The results of candidate detection and classification methods is compared with ground-truth information provided along with the dataset. The metrics used to evaluate the mitosis detection of each method include: number of true positive (TP), number of false positives (FP), number of false negative (FN), sensitivity or true positive rate (TPR), precision or positive predictive value (PPV) and F-Measure.



(a) Top three ranks (F-Measure) focal planes.



(b) F-Measure in Focal planes 6 and 7.

Figure 5: Candidate Detection Results

4.2.1 Candidate Detection in Different Spectral Bands

To gain a better understanding of the relative contributions of specific spectral bands, we perform candidate detection in all available spectral bands and z-stack focal planes. The results of candidate detection is ranked according to F-Measure as reported top three rank results in Figure ?? . Z-stack focal planes 6 and 7 have more information for candidate detection as compared to other focal plane. Spectral band 8 in focal layer 6 (590 – 640nm) have less FP and FN as compared with other spectral bands and TP is also comparable with other spectral bands as well.

4.2.2 Candidate Classification

Using full multispectral features with all classifiers, we get good TP and F-Measure with all classifiers except Bayesian. By selecting features from multispectral features vector, F-Measure is increase a bit with all classifiers with more TP.

5. Discussion and Conclusion

An automated mitosis detection framework for breast cancer MSI based on multispectral spatial features vectors has been proposed. Initially, candidate detection is performed in selected spectral and z-stack focal plane. Then,

we compute multispectral spatial feature vectors for each candidate, a highly efficient model for capturing texture features for nuclei discrimination. In future work, we plan to use other multispectral feature models to improve the framework accuracy.

References

- [1] Mitos contest 2012. <http://ipal.cnrs.fr/ICPR2012>.
- [2] G. Avwioro. Histochemical uses of haematoxylin review. *J Pharm Clin Sci (JPCS)*, 1:24–34, 2011.
- [3] H. Bloom and W. Richardson. Histological grading and prognosis in breast cancer: a study of 1409 cases of which 359 have been followed for 15 years. *British Journal of Cancer*, 11(3):359, 1957.
- [4] L. E. Boucheron, Z. Bi, N. R. Harvey, B. Manjunath, and D. L. Rimm. Utility of multispectral imaging for nuclear classification of routine clinical histopathology imagery. *BMC Cell Biology*, 8(Suppl 1):S8, 2007.
- [5] A. R. Chaudhury, K. Iychettira, R. Iyer, and A. Sreedevi. Diagnosis of invasive ductal carcinoma using image processing techniques. In *International Conference on International Conference on Image Information Processing (ICIIP)*, pages 1–6. IEEE, 2011.
- [6] E. S. Cibas and B. S. Ducatman. *Cytology: Diagnostic principles and clinical correlates*. Saunders, 2009.
- [7] E. Cosatto, M. Miller, H. P. Graf, and J. S. Meyer. Grading nuclear pleomorphism on histological micrographs. In *19th International Conference on Pattern Recognition (ICPR)*, pages 1–4, Dec. 2008.
- [8] J.-R. Dalle, H. Li, C.-H. Huang, W. K. Leow, D. Racoceanu, and T. C. Putti. Nuclear pleomorphism scoring by selective cell nuclei detection. In *IEEE Workshop on Applications of Computer Vision*, 2009.
- [9] M. Dundar, S. Badve, G. Bilgin, V. C. Raykar, R. K. Jain, O. Sertel, and M. N. Gurcan. Computerized classification of intraductal breast lesions using histopathological images. *IEEE Transactions on Biomedical Engineering*, 58(7):1977–1984, July 2011.
- [10] E. Elston and I. Ellis. Method for grading breast cancer. *Journal of clinical pathology*, 46(2):189, 1993.
- [11] D. C. Fernandez, R. Bhargava, S. M. Hewitt, and I. W. Levin. Infrared spectroscopic imaging for histopathologic recognition. *Nature biotechnology*, 23(4):469–474, 2005.
- [12] T. J. Fuchs and J. M. Buhmann. Computational pathology: Challenges and promises for tissue analysis. *Computerized Medical Imaging and Graphics*, 35(7):515–530, 2011.
- [13] M. Galloway. Texture analysis using gray level run lengths. *CGIP*, 4:172–9, 1975.
- [14] S. M. Gentry and R. M. Levenson. Biomedical applications of the information-efficient spectral imaging sensor (ISIS). In *International Biomedical Optics Symposium (BiOS)*, pages 129–142. International Society for Optics and Photonics, 1999.
- [15] Y. Gong. Breast cancer: Pathology, cytology, and core needle biopsy methods for diagnosis. In M. K. Shetty, editor, *Breast and Gynecological Cancers*, pages 19–37. Springer New York, 2013.
- [16] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener. Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering*, 2:147–171, 2009.
- [17] R. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Trans. on Systems, Man and Cybernetics*, 3:610–21, 1973.
- [18] H. Irshad, S. Jalali, L. Roux, D. Racoceanu, L. Joo-Hwee, G. L. Naour, and F. Capron. Automated mitosis detection using texture, SIFT features and HMAX biologically inspired approach. *Journal of Pathology Informatics*, in press.
- [19] H. Irshad, L. Roux, and D. Racoceanu. Automated mitosis detection in histopathology using morphological and multi-channel statistics features. *Journal of Pathology Informatics*, in press.
- [20] R. Khelifi, M. Adel, and S. Bourennane. Multispectral texture characterization: application to computer aided diagnosis on prostatic tissue images. *EURASIP Journal on Advances in Signal Processing*, 2012(1):1–13, 2012.
- [21] R. M. Levenson. Spectral imaging perspective on cytomics. *Cytometry Part A*, 69(7):592–600, 2006.
- [22] R. M. Levenson, P. J. Cronin, and K. K. Pankratov. Spectral imaging for brightfield microscopy. In *Biomedical Optics 2003*, pages 27–33. International Society for Optics and Photonics, 2003.
- [23] R. M. Levenson, A. Fornari, and M. Loda. Multispectral imaging and pathology: seeing and doing more. *Expert Opinion on Medical Diagnostics*, 2(9):1067–1081, 2008. PMID: 23495926.
- [24] R. M. Levenson and J. R. Mansfield. Multispectral imaging in biology and medicine: slices of life. *Cytometry Part A*, 69(8):748–758, 2006.
- [25] H. Liu and R. Setiono. A probabilistic approach to feature selection - a filter solution. In *ICML*, pages 319–27, 1996.
- [26] J. Ma and A. Jemal. Breast cancer statistics. In A. Ahmad, editor, *Breast Cancer Metastasis and Drug Resistance*, pages 1–18. Springer New York, 2013.
- [27] M. Malon and E. Cosatto. Classification of mitotic figures with convolutional neural networks and seeded blob features. *Journal of Pathology Informatics*, in press.
- [28] K. Masood and N. Rajpoot. Texture based classification of hyperspectral colon biopsy samples using clbp. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro ISBI*, pages 1011–1014. IEEE, 2009.
- [29] G. Meijer, J. Belien, P. Van Diest, and J. Baak. Origins of image analysis in clinical pathology. *Journal of clinical pathology*, 50(5):365–370, 1997.
- [30] S. Naik, S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski. Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, pages 284–287. IEEE, 2008.
- [31] S. Petushi, F. U. Garcia, M. M. Haber, C. Katsinis, and A. Tozeren. Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer. *BMC Medical Imaging*, 6(1):14, 2006.

- [32] M. E. Plissiti and C. Nikou. A review of automated techniques for cervical cell image analysis and classification. In *Biomedical Imaging and Computational Modeling in Biomechanics*, pages 1–18. Springer, 2013.
- [33] M. Roula, A. Bouridane, F. Kurugollu, and A. Amira. A quadratic classifier based on multispectral texture features for prostate cancer diagnosis. In *Seventh International Symposium on Signal Processing and Its Applications*, volume 2, pages 37–40. IEEE, 2003.
- [34] L. Roux, D. Racocanu, N. Lomenie, M. Kulikova, H. Irshad, J. Klossa, F. Capron, C. Genestie, G. L. Naour, and M. N. Gurcan. Mitosis detection in breast cancer histological images an icpr 2012 contest. *Journal of Pathology Informatics*, in press.
- [35] C. Stewart, J. Duncan, M. Farquharson, and J. Richmond. Fine needle aspiration cytology diagnosis of malignant lymphoma and reactive lymphoid hyperplasia. *Journal of clinical pathology*, 51(3):197–203, 1998.
- [36] Weka. Url <http://www.cs.waikato.ac.nz/ml/weka>.
- [37] W. H. Wolberg, W. N. Street, and O. L. Mangasarian. Breast cytology diagnosis via digital image analysis. *Analytical and Quantitative Cytology and Histology*, 15(6):396–404, 1993.
- [38] F. Woolfe, M. Maggioni, G. Davis, F. Warner, R. Coifman, and S. Zucker. Hyper-spectral microscopic discrimination between normal and cancerous colon biopsies. *IEEE Transactions on Medical Imaging*, 99(99), 1999.
- [39] X. Wu, M. Amrikachi, and S. K. Shah. Embedding topic discovery in conditional random fields model for segmenting nuclei using multispectral data. *IEEE Transactions on Biomedical Engineering*, 59(6):1539–1549, 2012.
- [40] X. Wu and S. K. Shah. A fast band selection method to increase image contrast for multispectral image segmentation. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, pages 1123–1126. IEEE, 2009.