Computerized Medical Imaging and Graphics 38 (2014) 390–402

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Multispectral band selection and spatial characterization: Application to mitosis detection in breast cancer histopathology

H. Irshad [a,c,*], A. Gouaillard [d,e], L. Roux [a,c], D. Racoceanu [b,c]

[a] University Joseph Fourier, Grenoble 1, France
[b] Sorbonne Universités, UPMC Univ Paris 6, France
[c] French National Center for Scientific Research, IPAL UMI 2955, Singapore
[d] CoSMo Software, Boston, MA, USA
[e] Temasys Communications, Singapore

ARTICLE INFO

Article history:
Received 26 September 2013
Received in revised form 25 March 2014
Accepted 11 April 2014

Keywords:
Histopathology
Breast cancer
Multispectral images
Spectral bands selection
Texture characterization
Classification
Object detection
Features extraction

ABSTRACT

Breast cancer is the second most frequent cancer. The reference process for breast cancer prognosis is Nottingham grading system. According to this system, mitosis detection is one of the three important criteria required for grading process and quantifying the locality and prognosis of a tumor. Multispectral imaging, as relatively new to the field of histopathology, has the advantage, over traditional RGB imaging, to capture spectrally resolved information at specific frequencies, across the electromagnetic spectrum. This study aims at evaluating the accuracy of mitosis detection on histopathological multispectral images. The proposed framework includes: selection of spectral bands and focal planes, detection of candidate mitotic regions and computation of morphological and multispectral statistical features. A state-of-the-art of the methods for mitosis classification is also provided. This framework has been evaluated on MITOS multispectral dataset and achieved higher detection rate (67.35%) and *F*-Measure (63.74%) than the best MITOS contest results (Roux et al., 2013). Our results indicate that the selected multispectral bands have more discriminant information than a single spectral band or all spectral bands for mitotic figures, validating the interest of using multispectral images to improve the quality of the diagnostic in histopathology.

1. Introduction

According to the International Agency for Research on Cancer (IARC) in its GLOBOCAN 2012 estimated cancer incidence, mortality and prevalence worldwide [1], breast cancer is the second most commonly diagnosed cancer worldwide after lung cancer. In 2012, there was an estimated 1.677 million new breast cancer cases and 522,000 deaths from breast cancer.

The reference process for breast cancer prognosis, recommended by the World Health Organization, is histologic grading that combines tubule formation, nuclei atypia and mitotic count [2,3]. This assessment of tissue sample is synthesized into a diagnosis that would help the clinician to determine the best course of therapy. Several computer aided diagnostic solutions exist for the detection of tubule formation [4,5] and nuclei atypia [6–9] but only a few are dedicated to mitosis detection [10,11].

In histopathology, hematoxylin and 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 [13]. 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 [14]. 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.

One of the most difficult challenges in histopathological dataset analysis is spatial analysis, more specifically automated nuclei detection and classification [15,12]. The objective of nuclei classification is to assign different labels to different types of nuclei as normal, cancer, mitotic, apoptosis, lymphocytes, etc. In addition,

* Corresponding author at: French National Center for Scientific Research, IPAL UMI 2955, Singapore. Tel.: +65 64082601.
E-mail address: humayun.irshad@gmail.com (H. Irshad).

http://dx.doi.org/10.1016/j.compmedimag.2014.04.003
0895-6111/

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].

Image analysis in cytology has been studied for years and numerous solutions [17,30–33] have been proposed in the literature. The application of these solutions to histopathology is rather complicated due to the radical differences between the two imaging modalities and to the 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. Nevertheless, recent works [34,10–12] show great potential for computer assisted diagnostic of histopathological datasets for breast cancer grading.

Multispectral imaging has the advantage over traditional RGB imaging to retrieve spectrally resolved information of a tissue image scene at specific frequencies across the electromagnetic spectrum. Multispectral imaging system captures images with accurate spectral content, correlated with spatial information, by revealing the chemical [29] and anatomic features of histopathology [35,36]. This modality provides option to biologists and pathologists to see beyond the RGB image planes to which they are accustomed. Recent publications [37–40] have begun to explore the use of extra information contained in such spectral data. Specifically, a comparison of spectral methodologies demonstrates the advantage of multispectral data [41,42]. The added benefit of multispectral imaging for analysis in routine H&E stained histopathological images, however, is still largely unknown, although some promising results are presented in [43,37,40,39]. As far as we know, there is no existing study of the use of multispectral imaging for automation of mitosis detection in breast cancer histopathology.

In our study, we propose a framework [11] for mitosis detection in breast cancer multispectral histopathology. This framework addresses two important questions: first, does the spatial–spectral analysis on selected spectral bands (as opposed to spatial analysis on single spectral or spatial–spectral analysis of all the spectral bands) suffice for efficient classification of mitotic and non-mitotic figures. An obvious advantage of using selected spectral bands is its reduced computational and storage complexity. Second, how effective are the multiple features for discrimination of mitotic and non-mitotic figures, as compared to one type of features? Some examples of mitotic and non-mitotic nuclei are shown in Fig. 1.

The remainder of the paper is organized as follows. Section 2 reviews the state-of-the-art of multispectral methods, particularly in object or region detection in histopathology, related to this research work. Section 3 describes the dataset used for this study. Section 4 describes the proposed framework for mitosis detection in breast cancer multispectral histopathology. Experiments and results are presented in Section 5. Section 6 contains the discussion part. Finally, the concluding remarks with future work are presented in Section 7.

2. Literature review

Multispectral imaging uses more than three spectral filters (not like RGB color images) to capture a series of images having spectral and spatial information. This spectral and spatial information has been used in remote sensing [18–21], biometric systems [22,23], plant and animal disease identification [24,25,36,26,27]. The main idea for extracting discriminative information from multispectral imaging is the use of combined spectral and spatial information for discrimination of regions or objects. In earlier works, Holmquist et al. [28] utilized spectrophotometric information for cervical cell segmentation in Papanicolaou-stained images. They found that two spectral bands were more appropriate for cervical cells segmentation as compared to single spectral band.

Fernandez et al. [37] coupled high-throughput Fourier transform infra-red spectroscopic imaging of tissue microarrays with statistical pattern recognition of spectra indicative of endogenous molecular composition for histopathological characterization of prostate 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.

We found few methods in the multispectral literature for spectral and spatial characterization of histopathological images. Some of them employed single spectral band of multispectral imaging [44,34] and others used multiple spectral bands of multispectral imaging [40,39,45]. Some methods computed one type of features on single spectral band for quantitative analysis. Masood and Rajpoot [44] proposed a colon biopsy classification method based on spatial analysis of hyperspectral images. First, spectral band 588 nm was selected, as it is the one that seemed to contain more textural information. Then, using circular local binary pattern algorithm, spatial analysis of patterns was represented by a feature vector in the selected spectral band. Later, classification was achieved using subspace projection methods like principal component analysis (PCA), linear component analysis (LCA) and support vector machine (SVM).

Some methods computed different types of features on single spectral band for quantitative analysis. Malon and Cosatto [34] combined manually designed nuclear features with the learned features extracted by convolutional neural network for mitosis detection and achieved the best *F*-Measure (59%) on multispectral dataset during ICPR contest 2012 [46]. First, focal plane number five was selected as it was clearly focused. Second, two spectral bands were selected using PCA to extract the top two eigenvectors from a set of 10 spectral bands of H&E stained multispectral images. Third, two step thresholding was applied on first eigenvectors (hematoxylin image) to obtain candidate blobs. Fourth, a set of shape, contour, pixel and texture features was computed on the selected spectral band only. Fifth, log likelihoods of class membership were computed using convolutional neural network classifier for each patch of candidate blob. In the last, the SVM classifier was used to classify each blob as either mitotic or non-mitotic blob using output of convolutional neural network along with feature vector. This approach discards additional potentially relevant information from other spectral bands.

Instead of limiting themselves to a single spectral band, some authors use multiple and sometime even all spectral bands, from a given dataset. Boucheron et al. [45] presented a study in which the additional spectral bands have additional useful information for nuclear classification in histopathology as compared to the three standard bands of RGB imagery. Using all spectral bands, they reported a 0.79% improvement in performance compared to the next best performing image type. Similarly, Wu el 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 is extracted to interpret spectral data with dynamically imposed pairwise constraints along the neighboring spectral bands. It is an unsupervised process, which efficiently helps to suppress segmentation errors caused by intensity inhomogeneity and variable chromatin texture. Khelifi et al. [40] 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. Some spectral bands have more relevant information for specific object or