Multispectral Band Analysis: Application on the Classification of Hepatocellular Carcinoma Cells in High-Magnification Histopathological Images

Oranit B[1]*, Chamidu A[2], Hiroshi N[2], Kota A[2], Fumikazu K[3] and Masahiro Y[3]

[1]Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Japan
[2]Department of Imaging Science and Engineering Laboratory, Tokyo Institute of Technology, Japan
[3]Department of Global Scientific Information and Computing Center, Tokyo Institute of Technology, Japan

Abstract

The nuclei structure has a significant interpretation for cancer analysis in histopathological microscopic images. In this paper, we analyze hepatocellular carcinoma in 100x magnification from nuclear chromatin patterns. The multispectral imaging is a new potential technique for histopathology. It may provide an alternative to pathologists to see additional information. This paper utilizes multispectral images which have spatial and spectral information for nuclear analysis. The proposed framework is based on texture analysis of nuclei. The system aim to analyze the significant of multispectral bands for discriminating cancer and non-cancer nuclei. The textural features were extracted using Gabor descriptors. We present nuclei textural feature with 30 Gabor patterns at different scales and orientations. Bag-of-visual-word model with random forest classifier is employed to classify normal and cancer cells. Moreover, we remove irrelevant Gabor parameters using optimization algorithm, which achieve high recognition performance significantly. Experimental result shows that our approach achieves approximately 99% of classification accuracy.

Keywords: Gabor descriptor; Hepatocellular carcinoma; Histopathological images; Bag-of-visual-word model; Multispectral images

Introduction

Histopathology is a microscopic study of tissue especially in cancer diagnosis. Sub-cellular components of tissue play an important role for diagnosing cancers. Most cancer types are diagnosed by visualizing histopathology images [1]. To diagnose cancers, pathologists use tissue samples obtained from patients via biopsy. Pathologists mainly examine tissues such as nuclei, glands and lymphocyte and discriminate them on normal, benign and malignant [2]. The morphological appearance of these structures is an important indicator for analysis. Since most of normal cells appear as a round or ellipsoidal shape, the sizes and morphological features of nuclei tend to be uniform. In contrast, cancer nuclei have irregular shapes, variables in size and atypia chromatin patterns [3]. Consequently, the characteristics of nuclei indicate cancerous condition. In this paper, we focus on analysis of hepatocellular carcinoma (HCC) stained by hematoxylin and eosin (HE).

The visual interpretation of cancer diagnosis depends on pathologists' experiences for analyzing quantitative information. Except diagnosis perspective, the quantitative analysis based on chromatin texture in cancer cells [1] will assist pathologists to explore the reason behind a specific disease identification. Therefore, computer aided diagnosis (CAD) systems would support them to make decisions and to provide more useful information to clinical treatment. This kind of system is based on image analysis, pattern recognition and machine learning techniques.

The multispectral imaging technology has been extended from a remote sensing field to medical and biological fields as an effective technique for medical image analysis. Multispectral images provide spatial and spectral information. Therefore, this information has an advantage to perform histopathological analysis and broaden pathological knowledge as a new potential [4].

Histopathology analysis based on multispectral imaging technologies may support a novel system and provide an additional option to pathologists to see over traditional color images. Over the years, researchers have invented various multispectral imaging systems for histopathological analysis of biological tissue specimen [5-7]. These systems take advantage of using combined spatial-spectral information to detect and classify diseases. CAD systems in cancer diagnosis have been proposed in cervical cancer [8,9], breast cancer [10], and white blood cell [11] by utilizing multispectral images. Some publications analyzed cancers by using multispectral images. Their result showed that multispectral images can enhance the system performance compared with the conventional color images. Wu et al. [8] and Xin et al. [10] studies have shown the possibility of multispectral imaging technologies in pathological analysis. Moreover, multispectral images also contribute to researchers to get more knowledge on tissues and cells.

This study proposes a classification approach of HCC with high recognition performance. It was applied to 100x magnification histopathological multispectral images of HE stained tissue slides. The proposed system utilizes spatial information of individual multispectral band to discriminate between cancer and non-cancer nuclei. We apply an optimization algorithm to select relevant features, which

appropriate for all multispectral bands. Moreover, the proposed system is aimed to achieve high classification accuracy with free-conditions of band selections.

This paper is organized as follows: Section 2 reviews related studied. Section 3 describes the process of image acquisition, image pre-processing and basic techniques applied in proposed system. The overview of classification model, optimized Gabor parameters and nuclei segmentation procedure are provided in section 4. Our experiments and results are provided in section 5. Section 6 concludes the entire study.

Related Works

Basically, an automated cancerous nuclei classification contains two main parts; i) segmentation of sub-cellular structures, ii) cancer classification. Researchers have proposed segmentation algorithms to identify subcellular structures from multispectral images. Most of segmentation algorithms utilized both of spatial and spectral information. The pixel-based methods are widely used in multispectral and hyper spectral images for segmentation such as Spectral Angle Mapper (SAM) [12] and Spectral Information Divergence (SID) [13]. The idea of SAM and SID techniques is to compute the spectral similarity between two spectral vectors. Guan et al. [14] applied SID algorithm in white blood cell segmentation on hyperspectral images and achieved promising result. Furthermore, conditional random field model (CRF) [15] and support vector machine (SVM) [11] were proposed to segment nuclei and other sub-cellular structures. For classification in multispectral images, Qi et al. [10] classified breast cancer tissue microarrays as well as low-magnification. They utilized a bag-of-visual-word technique with texture descriptors. The experiment was performed by applying SVM classifier with radial basis function for each individual multispectral band. The bands that achieved higher classification rate than their transformed gray scale RGB were chosen as candidates in final decision. The majority voting strategy was applied for final classification result. In addition, they compared classification performance with conventional RGB images. Their experimental result showed multispectral images achieved higher classification rate than RGB images. From above mentioned, CAD system for HCC diagnosis have not been proposed extensively in multispectral imaging technology.

Band selection [16-20] is one of the popular techniques and widely used in remote sensing. This technique refers to select the optimal subset of multispectral bands. The advantage of band selection is removing redundancy of data and reducing number of bands for analysis. In particular, the computational time of utilizing subsets of bands is faster than dealing with whole multispectral bands. However, the aim of this study focuses on different point of view. The present paper aims to analyze the significance of individual multispectral band in terms of characterizing cancer and non-cancer cell.

For feature extraction, the utility of Gabor filters have been beneficial. Particularly, Gabor filters have accomplished in many applications related to texture analysis. Gabor descriptors are the effective features to represent the characteristic of an image. However, the parameters of Gabor filter bank depend on a database. The design of filter bank is done by setting parameters manually. Therefore, the insignificant filters for representing the data are included. In color image, researchers have proposed feature selection techniques to remove irrelevant Gabor patterns such as Sequential Forward Selection (SFS), Sequential Backward Selection (SBS), Mutual information [21] and Minimum Redundancy and Maximum Relevance (mRMR)
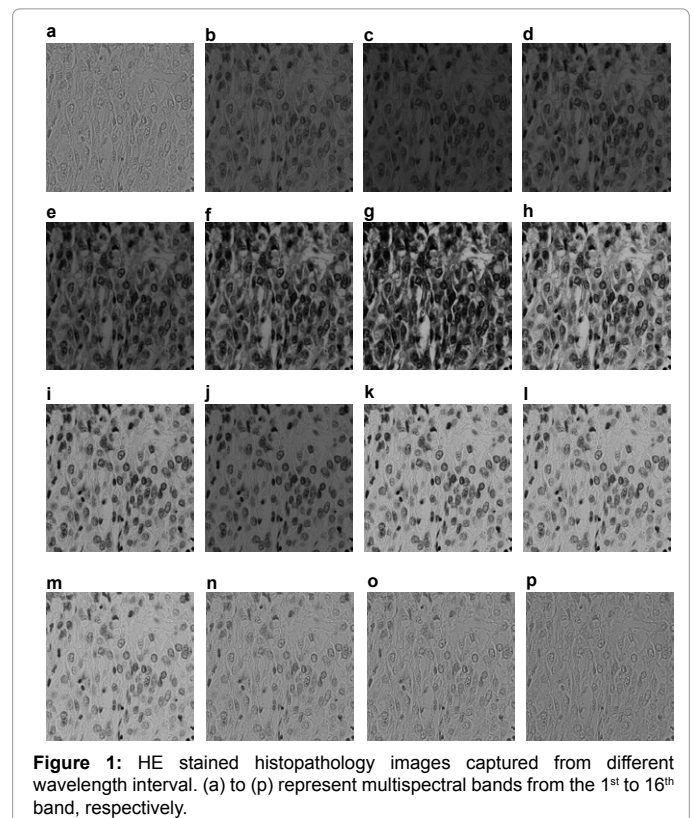
[22]. These techniques could not apply directly to our framework in multispectral images. Therefore, our objective of utilizing optimization algorithm is to find relevant Gabor patterns which suitable for overall multispectral bands in term of improving classification accuracy. The algorithm can mainly take the advantage to reduce dimensionality of feature and speed up computational time.
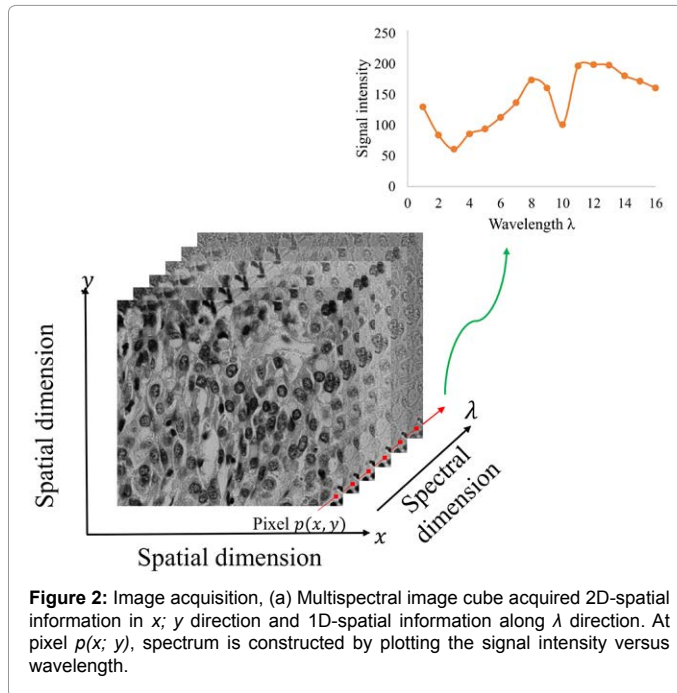
Materials

Data acquisition

In this experiment, multispectral image dataset is captured using multispectral microscopic-camera developed by Akasaka National Vision Research Center, Japan [23]. This camera system is composed of a 16-band rotating filter wheel and a 2048×2048 pixels CCD camera. The wavelength range spans the visible spectrum i.e., 400-750 nm. Each multispectral image is captured in different narrow band of wavelengths at the same position and represented in a gray scale image. Figure 1 shows HCC multispectral images of 16 spectral bands. Multispectral can be also visualized as a 3D-image cube in x, y and λ directions which represent 2D-spatial information along x and y directions and 1D-spectral wavelength information in λ dimension as shown in Figure 2. A spectrum can be formed by plotting the spatial coordinates against wavelength dimension at the same position. Each spectrum represents a specific characteristic of its spatial point in image cube. Normal liver and HCC tissue microarray specimen were purchased from US Biomax, Inc., (Rockville, MD, USA).

All tissue samples were collected with informed consent from patients. Multispectral dataset contains normal and HCC images of HE stained histopathology liver tissues from 19 patients, captured from the tissue microarray. One multispectral image is classified as normal or cancer by experienced pathologists. The multispectral dataset includes

Figure 1: HE stained histopathology images captured from different wavelength interval. (a) to (p) represent multispectral bands from the 1st to 16th band, respectively.

Figure 2: Image acquisition, (a) Multispectral image cube acquired 2D-spatial information in *x; y* direction and 1D-spatial information along *λ* direction. At pixel *p(x; y)*, spectrum is constructed by plotting the signal intensity versus wavelength.

162 images at 100x magnification with no overlapping location. One sample contains 16 wavelength bands.

Image preprocessing

The first process of multispectral image collection was to set a CCD exposure time and light intensity before cap-turing multispectral image. Then, multispectral images were captured under the same conditions to make sure that the data were comparable. However, image artifacts can be generated by factors such as non-uniform illumination during the multispectral image acquisition. It can affect image quality. To remove the artifacts, we acquired an additional reference image $B(x,y;\lambda_j)$ under exactly the same acquisition condition in a blank area of the tissue slide. *x* and y arespatial coordinates of the image and λ_j is wavelength interval of jth band. Therefore, normalized data $I(x,y;\lambda_j)$ can be expressed as

$$I\left(x,y;\lambda_j\right)=\frac{S\left(x,y;\lambda_j\right)}{B\left(x,y;\lambda_j\right)} \qquad (1)$$

where $S(x,y;\lambda_j)$ and represent sample image and a blank slide image in multispectral data respectively.

Bag of Words model

Over the years, bag-of-visual-words (BoW) approaches have been used in various kinds of computer vision applications as one of the powerful techniques. BoW model can transform a set of high-dimensional local features into a single feature vector for representing images. This representation has shown the effectiveness in image classification and categorization. For medical image analysis, BoW model has been successfully applied and achieved promising result in histopathology images in several types of cancers. We extracted textural features of each pixel in nuclear region. Based on BoW model, dictionary is generated by k-means algorithm. As a result, each cluster center represents a feature vector of its class, called visual word. Consequently, each pixel in region of interest finds the nearest visual word index and form histogram frequencies. This histogram is a novel representation of interested region.

Feature extraction by Gabor descriptors

In our experiment, we utilized Gabor descriptors. Gabor filtering is one of the textural feature extraction techniques, and widely used in various kinds of recognition applications including medical image diagnosis system [24].

The 2D-Gabor filters explained in [25] can be defined as

$$g_{uv\theta}\left(x,y\right)=\exp\left(-\frac{x'^2+\gamma^2 y'^2}{2u^2}\right)\cos\left(2\pi\frac{x'}{v}+\psi\right)$$
$$x'=x\cos(\theta)+y\sin(\theta),$$
$$y'=y\cos(\theta)-x\sin(\theta) \qquad (2)$$

where $g_{uv\theta}$ denotes a Gabor filter with spatial aspect ratio , standard deviation of Gaussian *u*, phase offset *ψ* and orientation *θ*. *x, y* are spatial coordinates. *v* is the wavelength and $1/v$ is spatial frequency of the cosine factor. The ratio u/v determines the spatial frequency bandwidth *bw*. The relationship between the half-respond of spatial frequency bandwidth *bw* and the ratio u/v can be expressed as

$$bw=\log_2\frac{\frac{u}{v}\pi+\sqrt{\frac{\ln 2}{2}}}{\frac{u}{v}\pi-\sqrt{\frac{\ln 2}{2}}}, \frac{u}{v}=\frac{1}{\pi}\sqrt{\frac{\ln 2}{2}}\frac{2^{bw}+1}{2^{bw}-1} \qquad (3)$$

The image features can be extracted by convoluting original image $I(x, y)$ with a set of Gabor patterns as defined

$$G_{uv\theta}\left(x,y\right)=I\left(x,y\right)*g_{uv\theta}\left(x,y\right) \qquad (4)$$

where $G_{uv\theta}(x, y)$ represents the convolution result corresponding to the original image and Gabor filter $g_{uv\theta}$.

Gabor feature may provide useful information from region of interest. The main key of generating Gabor filter bank is scales and orientation parameters. In particular, the combination of their parameters will produce different results. Therefore, this study focuses on orientation *θ* and spatial frequency bandwidth *bw* parameters for creating a set of Gabor filters.

Methods

Texture classification model

In multispectral images, the classification of HCC is based on BoW model of textural features. Figure 3 shows the overview of classification procedures for single multispetral band, while the experiments were performed for every multispectral band. Firstly, we extracted features of nuclear texture utilizing Gabor filter bank. Then, BoW model is applied for representing nuclei as a histogram frequency. This histogram of visual words provides a discriminative characteristic for each nucleus in HCC images. A random forest classifier [26] is adopted in classification step. The histogram is used as an input for generating model of random forest in training step and also used in testing step. As a result of the training, posterior probability of each class is represented. The class that has maximum posterior probability is selected as the final class of that nucleus. Finally, majority voting strategy is utilized for making a final decision of each patient. Let *qc* be the total number of nuclei which belong to class *c* and *r* be total number of nuclei. Also let the probability of a patient classified into class *c* be *pc=qc/r*. Given a threshold *T*, if maximum of *pc* is greater than *T*, then the patient is classified into class *c*.
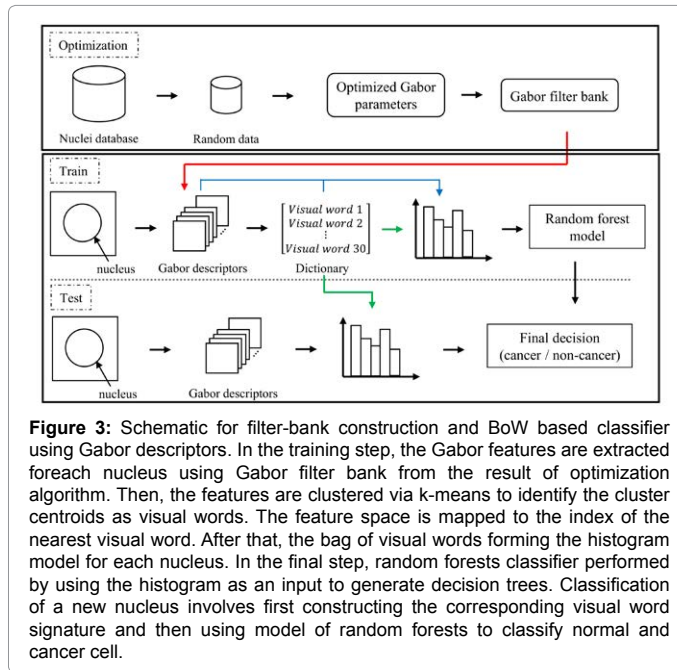
Figure 3: Schematic for filter-bank construction and BoW based classifier using Gabor descriptors. In the training step, the Gabor features are extracted foreach nucleus using Gabor filter bank from the result of optimization algorithm. Then, the features are clustered via k-means to identify the cluster centroids as visual words. The feature space is mapped to the index of the nearest visual word. After that, the bag of visual words forming the histogram model for each nucleus. In the final step, random forests classifier performed by using the histogram as an input to generate decision trees. Classification of a new nucleus involves first constructing the corresponding visual word signature and then using model of random forests to classify normal and cancer cell.

Optimization algorithm for Gabor parameters

The process of extracting features is also important for representing nuclei before performing BoW framework. A result of feature extraction directly influences classification performance. This study utilizes Gabor filters by setting different frequencies and orientations. Therefore, Gabor parameters construct a filter bank. However, not all of frequencies and orientations play an important role in liver cancer cell classification. We propose a concept to select subset of relevant Gabor parameters. The idea takes an advantage of improving the overall computational complexity. The objective of our proposed algorithm is to search the most effective parameter which is appropriate to overall multispectral bands. Moreover, this optimization technique also makes the data comparable under the same condition. In this study, the proposed optimization algorithm focused on selecting a subset of relevant parameters in Gabor filter, that is, orientation and spatial frequency bandwidth. The overview of proposed algorithm is shown in the optimization part of Figure 3.

The algorithm is described as follows.

Given a data set of nuclei $D \subset \mathbb{Z}^{N \times 16}$, where N denotes the total number of nuclei in the multispectral database, we randomly select K, $(K<N)$ data from D for optimizing Gabor parameters. A and B denote a set of orientations and spatial frequency bandwidths of Gabor filter bank respectively where $A=\{a_1,a_2,...,a_m\}$ and $B=\{b_1,b_2,...,b_s\}$. The optimal values in A and B are selected from the minimum ranking of mean square error (MSE) of texture classification.

We describe our proposed parameter optimization procedure in Algorithms 1 and 2. The Algorithm 1 is aimed to find the optimal subset of parameters A and B. Firstly, MSE of each parameter in A (MSE_A) and B (MSE_B) is calculated by Algorithm 2. After that, first n-rank and p-rank of descending order of MSE_A and MSE_B respectively are selected as optimal values in A ($optA$) and B ($optB$). n and p are the number of selected elements in A and B respectively. The MSE_A is calculated using Algorithm 2 with input d, A, B. Secondly, the MSE_B is also computed from Algorithm 2 with input d, B, A. The optimization

of Gabor parameters is described in Algorithm 2. The input consists of data d and parameter sets A and B. Firstly, a set of Gabor pattern f is generated by utilizing a_i with a set of B. Then, Gabor descriptors in each multispectral bands is calculated by taking convolution between original image I_j at band j and f. We calculate error classification rate e_j at band j from texture classification model as described in texture classification model.

We iteratively compute e_j for every multispectral band. MSE of a_i is computed, where

$$\mu = \frac{1}{16} \sum_{j=1}^{16} e_j.$$

Consequently, we iteratively change a_i, $i=1,2,...,m$. Finally, the result of Algorithm 2 return MSE_A. Moreover, the MSE_B is also computed in Algorithm 2 as described above.

Algorithm 1

INPUT: data $d \subset \mathbb{Z}^{K \times 16}$, set of parameters A and B, and n and p are numbers of optimized A and B respectively.
OUTPUT: $optA$ and $optB$
$MSE_A = OptimizedGaborParameter(d, A, B)$
$MSE_B = OptimizedGaborParameter(d, B, A)$
$optA = $ first n — value of $SortDescent(MSE_A)$
$optB = $ first p — value of $SortDescent(MSE_B)$

Algorithm 2 Optimized Gabor Parameter (d, A, B)

INPUT: data $D \subset \mathbb{Z}^{K \times 16}$, set of parameters A and B
OUTPUT: MSE_A
for $i = 1 : m$ **do**
$a(i) = \{a(i) | i = 1,....,m\}$
for $j = 1 : 16$ **do**
$F_j \leftarrow I_j * f(a(i), B)$
$e_j \leftarrow Classify(F_j)$
end for
$MSE_{a(i)} = \frac{1}{16} \sum_{j=1}^{16} (e_j - \mu)^2$
end for
$MSE_A = \{MSE_{a(i)} | i = 1,...,m\}$
return MSE_A

Nuclei segmentation

In our data set, a normal liver and HCC images mainly consist of four regions: nucleus, cytoplasm, lymphocyte and background regions. Our segmentation purpose is to find regions of nuclei. However, we perform multiclass segmentation to get more precise and accurate segmentation. The segmentation method is based on pixel-based classification and enhancement. We obtained experts' annotated 15 sample images. Each image is annotated into four regions: nucleus, cytoplasm, and lymphocyte and background regions. Generally, features of nucleus and lymphocyte regions are sometimes overlapped. Therefore, we group nucleus and lymphocyte regions as one class. We utilized 147 multispectral images from 19 patients as experimental dataset.

From eq. (1), we have a multispectral image $I(x, y; \lambda_j)$, $j=1,....,16$. The spectrum at a point (x,y) can be defined as a vector $V(x,y)$

$$V(x,y)=\{I(x,y;\lambda_1),I(x,y;\lambda_2),...,I(x,y;\lambda_{16})\} \tag{5}$$

The spectrum given by equation (5) represents a pixel feature vector. In pixel-based classification, random forest classifier [26] is

utilized for classification. The feature vectors of annotated regions were used as an input for generating model of random forest in training step. Then, we obtained prediction probability of each class as a result of random forest training. Since our goal focuses on nuclei segmentation, we consider only prediction probability of nucleus-lymphocyte class. Subsequently, we normalized these probabilities into gray scale image as shown in (Figure 4b).

In addition, we enhanced the result of pixel-based classification by applying image processing techniques to segment the nuclei effectively. The first step is to apply the morphological opening operator. After that, binarize the image by Otsu's binarization algorithm [27]. Then, the morphological closing operator is applied to remove small regions. Since the size of lymphocyte is always smaller than nucleus region, we set the threshold value for removing lymphocyte region. The result is shown in (Figure 4c). Finally, boundaries of nuclei in images were detected. (Figure 4d) shows the result of nuclei segmentation. The out-of-focus nuclei were not detected in this method, since they have unclear boundary and chromatin pattern. As a result of pixel-based classification, the probability of nucleus class of that area is not high. From this reason, the spectrum may similar to other sub-cellular structures. Furthermore, by applying morphological operators, out-of-focus nuclei are disappeared. Our system discriminates cancer regions from chromatin patterns. Therefore, only focused nuclei were used for analysis.

Experiments and Results

Implementation

After the segmentation of nuclei, we used these nuclei to discriminate cancer and non-cancer cells in classification step. In the following experiments, Gabor descriptor is utilized to extract textural patterns from each nucleus. A set of Gabor filters has different frequencies and orientations. We set $\theta=\{0°, 30°, 45°, 60°,90°,135°\}$ in orientation, the spatial frequency bandwidth $bw=\{0.6, 0.7, 0.8, 0.9, 1\}$ spatial aspect ratio $\gamma=0.6$ and wavelength $v=2$. Thus, we obtain 9-Gabor

Figure 4: Step-by-step processes of nuclei segmentation (a) An example of original image in band 8th. (b) Prediction probability of nuclei class was normalized into gray scale image. (c) Final result by applying morphological operators. (d) The boundary of nuclei were detected.

Band	All filters	Optimization
1	92.32 ± 5.09	99.11 ± 1.03
2	88.58 ± 11.39	99.64 ± 0.31
3	88.45 ± 9.07	98.61 ± 1.98
4	90.00 ± 9.78	99.84 ± 0.25
5	87.81 ± 9.99	99.13 ± 1.06
6	87.53 ± 7.73	98.85 ± 1.16
7	94.22 ± 4.49	98.55 ± 0.98
8	97.88 ± 2.94	99.94 ± 0.09
9	93.67 ± 5.11	99.78 ± 0.31
10	88.06 ± 9.88	99.84 ± 0.31
11	95.10 ± 2.87	99.21 ± 0.79
12	94.00 ± 3.23	99.75 ± 0.61
13	95.82 ± 3.21	98.58 ± 1.27
14	93.74 ± 4.65	98.09 ± 1.43
15	92.50 ± 4.19	97.04 ± 2.25
16	90.70 ± 6.58	97.35 ± 2.74

Table 1: Comparison of classification accuracy (%) and its standard deviation between use of all of Gabor filters and optimized Gabor parameter.

patterns. The set of Gabor patterns may contain irrelevant patterns and it may affect the classification accuracy. Therefore, we perform optimization of Gabor parameter on our proposed algorithm. Our study randomized 25% of nuclear database to find relevant parameters which are suitable for all multispectral bands. From this experiment, we selected first 3 ranks of values in parameter θ and bw by the proposed optimization algorithm. Thus, we obtain 9-Gabor patterns. After extracting feature descriptors, classification procedure is performed in patient level for each individual multispectral band as shown in (Figure 3). Our experiment created 30 visual vocabularies of dictionary. This number of visual words provides the optimal classification rate. The final decision is specified by majority voting strategy for each patient. Furthermore, k-fold cross validation (CV) is applied for evaluating the performance. The average accuracy over k-fold is obtained as a result of CV accuracy. Specifically, we compared the classification performance by utilizing all Gabor parameters and optimized Gabor parameters in this study. In addition, we performed experiment by combining the multispectral bands into several color spectrums according to the reference of wavelength interval of each color spectrum as shown in (Table 2). The classification was performed by concatenating the histogram of visual words in each band. Consequently, random forest classifier and majority voting strategy were performed for final decision.

Parameter selection

The optimization of Gabor parameters is resulted by selecting the first 3 ranks in a subset of parameter θ and bw. In this experiment, the relevant parameter in $\theta=\{0°, 30°, 60°\}$ and $bw=\{0.6, 0.7, 0.8\}$ are suitable for discriminating cancer and non-cancer nuclei in HCC images. Therefore, the feature vector for representing nucleus pixel reduces 30% of utilizing all parameters.

Classification on single spectral band

The classification results of individual multispectral bands are presented in Table 1. We evaluated the performance of the system with 10-fold CV and its standard deviation (SD) in patient level. We compared the classification accuracy with optimized and non-optimized Gabor parameters for discriminating normal and cancer nuclei. We can conclude the result as follows.

1. After performing optimization algorithm, the result shows that the 1st, 2nd, 4th, 5th and 8th-14th band successfully achieve high
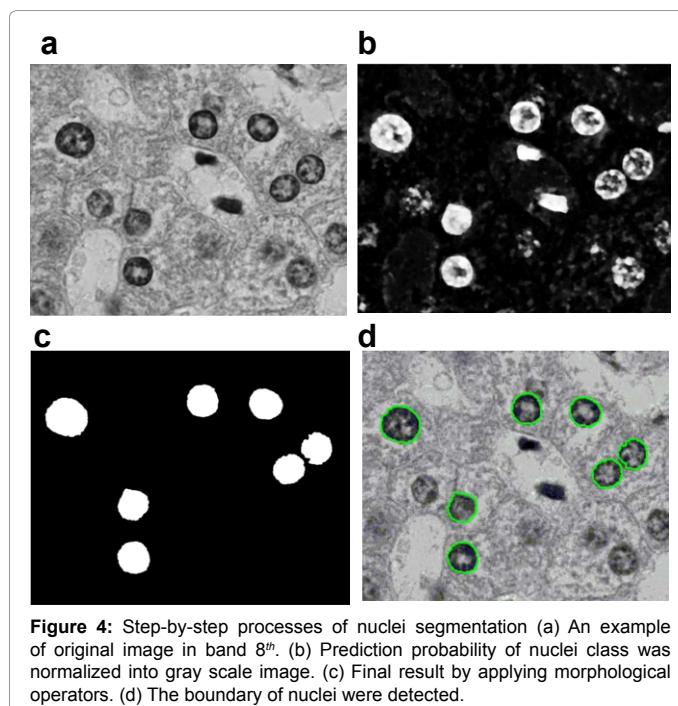
Color	Wavelength interval	Band	Classification rate (%) ± SD
Red	~700-635 nm.	12th-16th	98.34 ± 2.50
Orange	~635-590 nm.	10th-12th	99.51 ± 0.40
Yellow	~590-560 nm.	9th-10th	99.73 ± 0.29
Green	~560-490 nm.	5th-8th	99.82 ± 0.28
Blue	~490-450 nm.	1st-4th	99.69 ± 0.32

Table 2: The wavelength interval in each color [28]. Band column is grouped multispectral bands by mapping approximately the wavelength interval for each color.

classification rate approximately over 99%. Furthermore, classification performance in 6th, 7th, 13th, 14th band are around 98%.

2. The 15th and 16th band which is in the wavelength range 686-753 nm. have not much significant for classifying cancer and non-cancer nuclei in HCC images compare with other spectral bands.

3. Classification rates utilizing optimized Gabor parameters achieve higher performance than non-optimized for all multispectral bands.

The classification result showed that single multispectral band contains enough information for classifying cancer and normal nuclei in high-magnification of liver tissue. In other words, the combination of multispectral bands was not necessary. By utilizing feature optimization, a small number of features are used to select the most significant features. The experimental result also showed that there are no additional advantage of using more features. It was not get more beneficial information of chromatin patterns. In addition, parameter optimization assists to improve recognition performance and reduce computational time for generating feature and speed up the system efficiently.

Classification on color spectrum

In this experiment, we analyzed the impact of different color spectrum in visible wavelength interval on HCC classification. The classification results are shown in Table 2. Green spectrum reported the highest classification performance (99.82%). However, blue, yellow and orange spectrum obtained classification rate 99.5% approximately and red spectrum achieved recognition rate 98.34%. These five color spectrums have similar classification performance.

Conclusions

This study proposed a HCC cell classification system using multispectral images in 100x magnification. We applied Gabor descriptors with BoW model to extract chromatin pattern of nucleus. Consequently, this system analyzed the use of spatial information on one spectral band for characterizing cancer and non-cancer cells. The system enhances the classification performance of all multispectral bands by extracting more efficient features from Gabor patterns. Moreover, the experimental results show that most of multi spectral bands are significant for distinguishing cancer and normal nuclei. In particular, the 1st, 2nd, 4th, 5th and 8th-12th bands achieve highest classification accuracies approximately over 99%. However, the classification rates drop down around 0.5% and 2% in 3rd, 6th, 7th and 13th bands and 15th-16th bands respectively. In summary, the textures of nuclei obtained from wavelength 418-467 nm, 481-513 nm and 548-641 nm are adequate to classify normal and HCC in high-magnification. Our approach shows that multispectral images provide meaningful feature in terms of classifying normal and HCC nuclei. It was also proved that nuclei texture is sufficient to classify normal and HCC.

References

1. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, et al. (2009) Histopathological Image Analysis: A Review. IEEE Rev Biomed Eng 2: 147-171.

2. 2 George NP (1928) New Cancer Diagnosis. Proceedings of the Third Race Betterment Conference.

3. Zink D, Fischer AH, Nickerson JA (2004) Nuclear structure in cancer cells. Nature Reviews Cancer 4: 677-687.

4. Li Q, He X, Wang Y, Liu H, Xu D, et al. (2013) Review of spectral imaging technology in biomedical engineering: achievements and challenges. Biomedical Optics 18: 100901-100901.

5. Papadakis A, Stathopoulos E, Delides G, Berberides K, Nikiforidis G, et al. (2003) A novel spectral microscope system: application in quantitative pathology. IEEE Transactions on Biomedical Engineering 50: 207-217.

6. Richard L, Joseph B, George M (2012) Spectral imaging in preclinical research and clinical pathology. Analytical Cellular Pathology 35: 339-361.

7. Nallala J, Gobinet C, Diebold MD, Untereiner V, Bouch O, et al. (2012) Infrared spectral imaging asa novel approach for histopathological recognition in colon cancer diagnosis. Biomedical Optics 17: 116013-1-116013-12.

8. Wu Q, Zeng L, Ke H, Zheng H, Gao X, et al. (2005) A multispectral imaging analysis system for early detection of cervical cancer. Proceedings of SPIE, Physics of Medical Imaging.

9. Gertych A, Galliano G, Bose S, Farkas DL (2010) Automatic recognition of abnormal cells in cytological tests using multispectral imaging. Proceedings of SPIE, Computer-Aided Diagnosis.

10. Xin Q, Fuyong X, David FJ, Lin Y (2011) Comparative performance analysis of stained histopathology specimens using RGB and multispectral imaging. Proceedings of SPIE.

11. Guo N, Zeng L, Wu Q (2007) A method based on multispectral imaging technique for White Blood Cell segmentation. Computers in Biology and Medicine 37: 70-76.

12. Kruse FA, Lefkoff AB, Boardman JW, Heidebrecht KB, Shapiro AT, et al. (1993) The spectral image processing system (SIPS) interactive visualization and analysis of imaging spectrometer data. Remote Sensing of Environment 44: 145-163.

13. Chang C (1999) Spectral information divergence for hyperspectral image analysis. Procceedings of Geoscience and Remote Sensing Symposium 1: 509-511.

14. Qingli L, Xiaofu H, Yiting W, Hongying L, Dongrong X, et al. (2012) Pathological leucocyte seg-mentation algorithm based on hyperspectral imaging technique. Optical Engineering 51: 053202-1-053202-7.

15. Wu X, Amrikachai M, Shah SK (2012) Embedding Topic Discovery in Conditional Random Fields Model for Segmenting Nuclei Using Multispectral Data. IEEE Transactions on Biomedical Engineering 59: 1539-1549.

16. Sotoca JM, Pla F, Sanchez JS (2007) Band Selection in Multispectral Images by Minimization of Dependent Information. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 37: 258-267.

17. Sotoca JM, Pla F, Klaren AC (2004) Unsupervised band selection for multispectral images using information theory. Proceedings of the 17th International Conference on Pattern Recognition.

18. Martinez-Uso A, Pla F, Gracia-Sevilla P, Sotoca JM (2006) Automatic band selection in multispectral images using mutual information-based clustering. Progress in Pattern Recognition, Image Analysis and Applications 4225: 644-654.

19. De Backer S, Kempeneers P, Debruyn W, Scheunders P (2005) A band selection technique for spectral classification. IEEE Geoscience and Remote Sensing Letters 2: 319-323.

20. Baofeng G, Gunn SR, Damper RI, Nelson JDB (2006) Band selection for hyperspectral image classification using mutual information. IEEE Geoscience and Remote Sensing Letters 3: 522-526.

21. Peng H, Fulmi L, Ding C (2005) Feature selection based on mutual information

criteria of maxdependency, maxrelevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 27: 1226-1238.

22. Monalisa M, Anirban M (2013) An improved minimum redundancy maximum relevance approach for feature selection in gene expression data. Procedia Technology 10: 20-27.

23. Bautista PA, Abe T, Yamaguchi M, Ohyama N, Yagi Y (2008) Multispectral image enhancement for H&E stained pathological tissue specimens. Proceedongs of SPIE, Visualization, Image-guided Procedures, and Modeling.

24. Castellano G, Bonilha L, Li LM, Cendes F (2004)Texture analysis of medical images. Clinical Radiology 59: 1061-1069.

25. Petkov N, Kruizinga P (1997) Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: bar and grating cells. Biological Cybernetics 76: 83-96.

26. Breiman L (2001) Random forests. Kluwer Academic Publishers, Hingham, MA, USA.

27. Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man and Cybernetics 9: 62-66.

28. Bohren CF, Clothiaux E (2006) Fundamentals of atmospheric radiation: An introduction with 400 Problems. Wiley-VCH 9.