Classification of Human Carcinoma Cells Using Multispectral Imagery

Umut Çinar*[a], Yasemin Y. Çetin[a], Rengul Cetin-Atalay[a], Enis Çetin[b]
[a]Graduate School of Informatics, Middle East Technical University, Ankara, Turkey; [b] Electrical and Electronic Engineering, Bilkent University, Ankara, Turkey

ABSTRACT

In this paper, we present a technique for automatically classifying human carcinoma cell images using textural features. An image dataset containing microscopy biopsy images from different patients for 14 distinct cancer cell line type is studied. The images are captured using a RGB camera attached to an inverted microscopy device. Texture based Gabor features are extracted from multispectral input images. SVM classifier is used to generate a descriptive model for the purpose of cell line classification. The experimental results depict satisfactory performance, and the proposed method is versatile for various microscopy magnification options.

Keywords: Multispectral imaging, automatic classification, cancer cells, Gabor features, microscopy

1. INTRODUCTION

Automatic recognition of cancerous cell groups is an important task for accurate decision making process in medical diagnosis of cancer types [1]. With the advances of recent imaging and microscopy technologies, human cell morphology can be manually identified by expert pathologists. However, the manual diagnosis is time consuming and subjective. Therefore, automatic classification methods are desired.

In the literature, microscopy combined with multispectral imaging has been studied by various researchers. In the first group of studies, cell and tissue segmentation from microscopy images was considered. For example, in study [2], microscopic images were captured by a multispectral camera, and different categories of features such as textural and spectral were investigated in terms of their effect on overall classification results. The best cell segmentation accuracy was observed when morphologic features were employed together with spectral features. Similarly, in another study [3], a random field model is proposed to improve cell segmentation using both spectral and spatial features. In study [4], a contour based cell segmentation model is elaborated. For this purpose, multiple Bayesian classifiers are trained with three spectral bands, and several biological particles were segmented effectively. The study in [5] focuses on localization and classification of cell membrane activities over time. TIRF microscopy images were segmented using adaptive thresholding in Laplacian of Gaussian domain. Then, a probability map indicating change amount is constructed using statistical properties of segmented time sequences. In study [6], a Hidden Markov Model (HMM) based segmentation framework is developed to be used with sophisticated microscopy data, in which many different categories of biological material exist. Super pixel level segments obtained from HMM operation are combined into object level segments by the help of transition probabilities. The empirical results pose a better segmentation accuracy when compared with general purpose segmentation algorithms.

Apart from segmentation focused papers, automatic cell line classification problem is also widely researched. In [7], the authors present an automatic method for detecting different stages of human bladder cancer cells. Principle component analysis with a scoring scheme and linear regression were employed in order to define spectral characteristics of the disease stages. The study [8] employs wavelet features to classify cell boundary symmetries in microscopy images. A multiscale method with rotation invariance is proposed to determine the location of junctions in the cell line images.

*ucinar@metu.edu.tr; phone +(90 312) 210-3741; fax +(90 312) 210-3745; www.ii.metu.edu.tr

Medical Imaging 2016: Digital Pathology, edited by Metin N. Gurcan, Anant Madabhushi, Proc. of SPIE Vol. 9791, 97911C · © 2016 SPIE
CCC code: 1605-7422/16/$18 · doi: 10.1117/12.2217022

A framework for early detection of lung cancer cells is illustrated in [9]. An existing diagnosis method, called Bronchoscope-guided bronchoalveolar lavage (BAL), is further improved with the introduction of multiresolution analysis. In study [10], nuclei of the individual cells are classified from digital microscopy images. For this purpose, sample images captured from glioma biopsies are annotated by expert neuropathologists, and both textural and spectral features are employed for training an accurate classifier. For the classification model, Sequential Floating Forward Selection (SFFS) is used to select descriptive features before training the classifier model with Quadratic Discriminant Analysis.

In this study, the cell lines established from different cancer patients are investigated to devise an efficient method for categorizing cancer types. We mainly focus on Gabor filter-bank features obtained from filtering by different orientations and scales as a base feature descriptors [11]. Then, we train a classifier model by Support Vector Machines (SVM) method in order to generate an automatic classification method [12].

2. METHODOLOGY

In the study, we follow a linear classification schema containing image retrieval, feature extraction and classification.

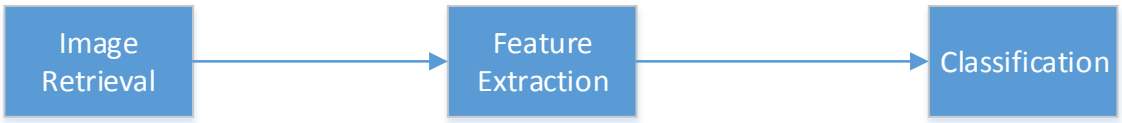
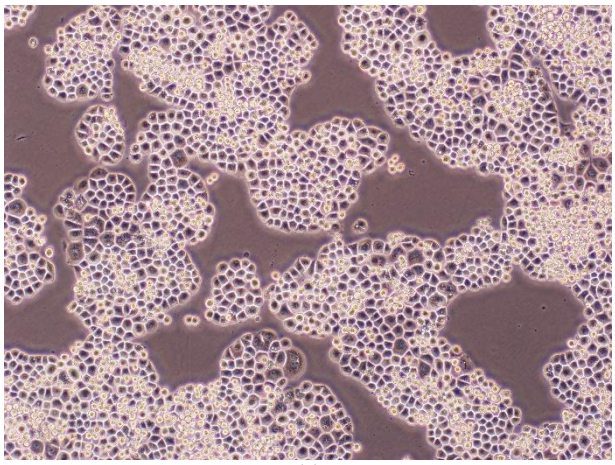


Figure-1. Flow chart of the procedure

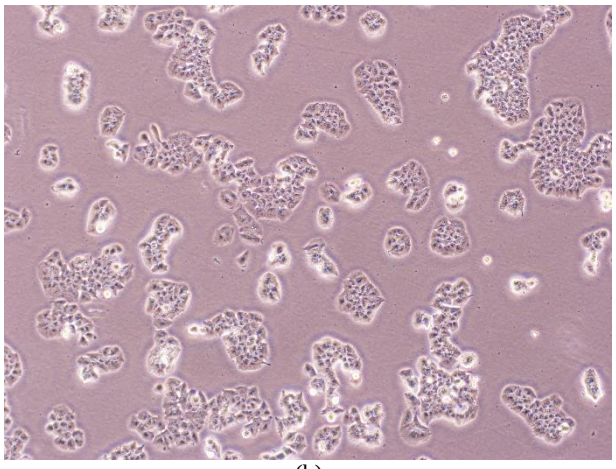
2.1 Image Retrieval

Cell line images are captured by Olympus CLX41 inverted microscope using Olympus DP72 camera with various magnification options including 10x, 20x and 40x [13].

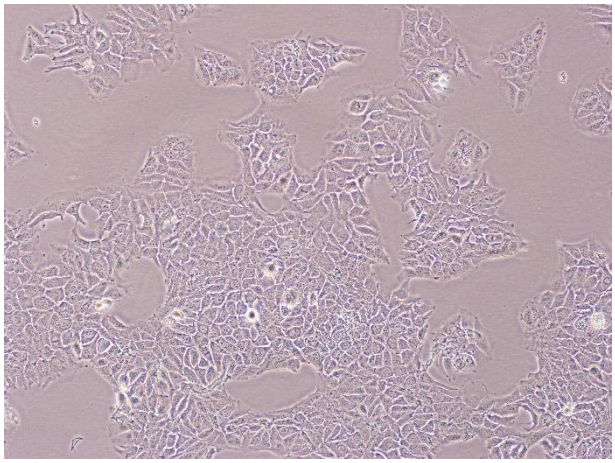
In this study, we study the application of multispectral classification methods on human carcinoma cell image dataset [13]. The dataset contains 14 distinct cell lines with 7 classes of breast and 7 classes of liver cancer cells. There are 840 images captured by RGB camera in the dataset. The size of images is 3096x4140 pixels. Sample images from the dataset can be seen in Figure-2.



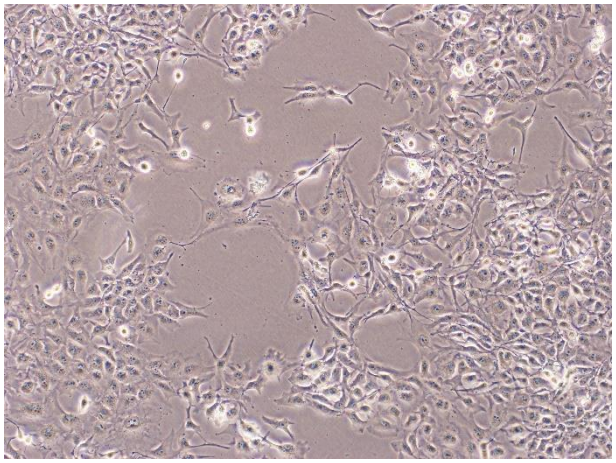
(a)



(b)



c)



(d)

Figure-2. Sample images from the dataset. (a) Cama1 (b) MDA-MB-361 (c) Huh7 (d) FOCUS

Detailed information about the dataset classes is depicted at Table-1.

Table-1. Summary of dataset cell line morphology

Cell Line	**Shape**	**Source**	**Disease**
BT-20	stellate	mammary gland breast	Adenocarcinoma
CAMA-1	grape-like	mammary gland breast	Adenocarcinoma
MDA-MB-157	stellate	mammary gland breast	Medullary carcinoma
MDA-MB-361	grape-like	mammary gland breast	Metastatic adenocarcinoma
MDA-MB-453	grape-like	mammary gland breast	Metastatic carcinoma
MDA-MB-468	grape-like	mammary gland breast	Metastatic adenocarcinoma
T47D	mass	mammary gland breast	Invasive ductal carcinoma
FOCUS	polygonal to spindle-shaped	liver	Hepatocellular carcinoma
Hep40	polygonal	liver	Hepatocellular carcinoma
HepG2	polygonalgrow as clusters	liver	Hepatocellular carcinoma
Huh7	polygonal	liver	Hepatocellular carcinoma
Mahlavu	polygonal to spindle-shaped	liver	Hepatocellular carcinoma
PLC	polygonal	liver	Hepatocellular carcinoma
SKHep1	polygonal to spindle-shaped	liver	Hepatocellular carcinoma

2.3 Feature Extraction

In this study, we are utilizing Gabor based textural features. Gabor features have been successfully used for variety of texture recognition problems, and they have ability to represent both spatial and directional relationships in 2D signal. For feature extraction an overlapping windowing operation is implemented. The purpose of the overlapping windowing operation is to optimally identify interest regions for the analysis. Since we are only interested in the regions containing cell structures, the image windows having the highest color entropy are selected for training the classifier. Total number of image windows for each image is limited by a pre-defined parameter S. Gabor based texture analysis provides a description technique close to human visual system [14] ; we model each image patch as a set of Gabor features. In order to create a feature encoding scheme, Gabor filters are applied for 8 scales and 4 directions for each image patch. The features extracted from Gabor response images are summarized in Table-2.

Table-2. Gabor features extracted

Name	**Formulation**
Local Energy	$LE = \sum_{p \in KxK} G(p)^2$
Mean of Real Part	$MM = \frac{\sum_{p \in KxK} REAL(G(p))}{K^2}$
Variance Magnitude	$VM = \sum_{p \in KxK} Var(G(p))$
* where G is response image and K is the window size	

2.4 Classification

SVM with RBF kernel is employed for classifying image features [15]. For classification, each image patch is considered as a sample, as a result, there are Sx(number_of_images) rows in the training dataset.

3. EXPERIMENTS

We adopt a 20-fold cross-validation strategy for the experiments. The dataset is divided into 20 disjoint subsets and each subset consisting of 14 images is used exactly once as the test set. The window size parameter K is left to vary and overlapping rate among the consecutive windows is empirically set to %50. For each input image, first 10 windows having highest entropy are selected for classification and the remaining windows are discarded. The detailed results regarding to the experiments are depicted in Table-3. The best classification accuracy is achieved when window size, K is set 400 pixels.

Table-3. Classification accuracy

Cross Validation Accuracy (**10x magnification**)	Cross Validation Accuracy (**20x magnification**)	Cross Validation Accuracy (**40x magnification**)	Patch Size **(KxK)**
79,1	77,1	75,5	50
84,2	75,0	80,2	100
88,9	81,9	82,7	150
89,6	82,3	85,1	200

95,1	87,1	88,5	250
95,4	92,7	88,3	300
98,1	95,1	91,4	350
99,2	96,4	**93,4**	400
98,5	**96,8**	92,6	450
98,0	94,1	91,1	500
94,2	95,1	88,6	550

Furthermore, we observed 0.1% of accuracy increase when 8 Gabor directions are employed instead of 4 directions, on the other hand it increased computation time significantly.

4. CONCLUSION AND OUTLOOK

In this study, the main objective is to explore applicability of multispectral imaging for cancer cell line classification. The proposed method is successful in accurately classifying different types of cancer cell lines. The best classification performance is obtained for 10x magnification option. It can be inferred that lower magnification rates result in better descriptive power in terms of textural features. When compared to the study [13] with which we utilized the same dataset, our performance is very similar. However, the present paper's method is more generic due to its independence of magnification factor. That is, our algorithm can accurately work on different magnification rates including 10x, 20x and 40x objective lenses without changing any parameter. Since different scaling options are taken into account when collecting Gabor features at different scale levels, the window size parameter K is nearly the same for all three magnification options.

For the future, we are planning to expand our framework with hyperspectral imaging, and we are working on integrating our hyperspectral camera with the existing hardware setup. In this way, we expect to improve accuracy of classification by introducing detailed spectral features collected by HSI sensor attached to the microscopy device. In fact, detailed spectral information for the cells are expected to ease discrimination of similar cell line morphologies.

REFERENCES

[1] B. F. Guolan Lu, "Medical hyperspectral imaging: a review," *Journal of Biomedical Optics,* vol. 19, pp. 1-23, 2014.

[2] J. Thigpen and S. Shah, "Multispectral Microscopy for Cell Differentiation in Thyroid Cytology," in *Multisensor Fusion and Integration for Intelligent Systems*, 2008.

[3] X. Wu and S. Shah, "Random field model for cell segmentation in transmission mode multispectral microscopy images," in *Signals, Systems and Computers*, 2009.

[4] S. Shah, "Segmenting Biological Particles in Multispectral Microscopy Images," in *Applications of Computer Vision*, 2007.

[5] A. Basset, P. Bouthemy, J. Boulanger, J. Salamero and C. Kervrann, "Localization and classification of membrane dynamics in TIRF microscopy image sequences," in *Biomedical Imaging (ISBI)*, 2014.

[6] R. Shenoy, M.-C. Shih and K. Rose, "A probabilistic framework for simultaneous segmentation and classification of multiple cells in multi-marker microscopy images," in *Biomedical Imaging* , 2015.

[7] C.-P. Jen, C.-T. Huang, Y.-S. Chen, C.-T. Kuo and Hsiang-Chen Wang, "Diagnosis of Human Bladder Cancer Cells at Different Stages Using Multispectral Imaging Microscopy," in *elected Topics in Quantum Electronics*, 2014.

[8] Z. Puspoki and M. Unser, "Wavelet-based identification and classification of local symmetries in microscopy images," in *Biomedical Imaging*, 2014.

[9] T. Pengo, A. Muñoz-Barrutía and C. Ortiz-de-Solórzano, "A Novel Automated Microscopy Platform for Multiresolution Multispectral Early Detection of Lung Cancer Cells in Bronchoalveolar Lavage Samples," *Systems Journal,* vol. 8, no. 3, pp. 985-994, 2014.

[10] J. Kong, L. Cooper, F. Wang, C. Chisolm, C. Moreno, T. Kurc, P. Widener, D. Brat and J. Saltz, "A comprehensive framework for classification of nuclei in digital microscopy imaging: An application to diffuse gliomas," in *Biomedical Imaging: From Nano to Macro*, 2011.

[11] I. Fogel and D. Sagi, "Gabor filters as texture discriminator," *Biological Cybernetics,* vol. 61, no. 2, pp. 103-113, 1989.

[12] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning ,* vol. 20, no. 3, pp. 273-297, 1995.

[13] F. Keskin, S. Alexander, K. Kose, T. Ersahin, A. E. Cetin and R. Cetin-Atalay, "Image Classification of Human Carcinoma Cells Using Complex Wavelet-Based Covariance Descriptors," *PLOS ONE,* vol. 8, no. 1, pp. 1-10, 2013.

[14] S. Marčelja, "Mathematical description of the responses of simple cortical cells," *Journal of the Optical Society of America,* p. 1297–1300, 1980.

[15] C.-C. a. L. C.-J. Chang, "A library for support vector machines," *Transactions on Intelligent Systems and Technology,* vol. 2, no. 3, pp. 27:1--27:27, 2011.