2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control

Application of Support Vector Machine to Recognize Trans-differentiated Neural Progenitor Cells for Bright-field Microscopy

Bo Jiang*†, Xinyuan Wang*‡, Qunxia Gao*§, Ziqi Lin*¶, Rui Zhang*‖ and Xiao Zhang***
*Guangzhou Institute of Biomedicine and Health
Chinese Academy of Sciences, Guangzhou, Science Park
†Email: jiang.bo@gibh.ac.cn
‡Email: wang.xinyuan@gibh.ac.cn
§Email: gao.qunxia@gibh.ac.cn
¶Email: lin.ziqi@gibh.ac.cn
‖Email: zhang.rui@gibh.ac.cn
**Email: zhang.xiao@gibh.ac.cn

***Abstract*—One possible solution of the investigation of the cell fate decision and its function is the study of cell morphology. Bright-field imaging analysis allow us to use a labeling free and non-invasive approach to measure the morphological dynamics during cellular reprogramming, which includes induced pluripotent stem cells (iPSCs), and trans-differentiated neural progenitor cells (NPCs) from somatic cell source. In order to automatically analyze and cultivate cells, a system classifying NPCs under bright-field microscopic imaging is necessary. In this paper, we investigate the use of support vector machine (SVM) based on a set of features for this task. The results illustrate that such a data driven approach has remarkable recognition and generalization performance.**

keywords: machine learning, support vector machine, trans-differentiated neural progenitor cells, cell recognition, bright-field microscopy.

I. Introduction

Cellular reprogramming opens the door for personalized regenerative medicine especially in fight with chronic and degenerative disease. We have established a technology converting the cells from urine into neural progenitor cells (NPC), so called trans-differentiation. This technology can allow us easily obtain source cells in a noninvasive approach. However, further studies we found that, during the trans-differentiation process, along with positive NPCs, the negative colonies have different sub type of morphology and polarity. As we understand, the cellular polarity strongly linked with gene expression, cell cycle and other cellular regulation may explain the mechanism regarding the different route of reprogramming. For example cell polarity changes between MET and EMT which linked with induced pluripotent stem cell (iPSCs) colony formation or tumor genesis. Hence, the regulation of cell fate changes has a strong link with morphology as a read out.

Classification of the morphological changes in cellular reprogramming process in a quantitative way can guide us normalize the cell cycle, gene expression, and protein expression during this cell fate transition period to determine the cell function, which provides richer and more precise information rather than the incubation period. Studying with trans-differentiating NPC cell cycle progression can be different from batch to batch, hence cross comparison between different experiments in morphology need to be quantified. If people could classify NPCs and non-NPCs forming from digitized approach rather than traditional artificial incubation time period, this progress could be a cornerstone to identify different cell phases of cellular reprogramming from somatic cell source to NPCs.

Automated fluorescence microscopy provides an effective method to observe and study nuclei dynamically and is an important quantitative technique in the fields of cell biology and systems biology [18] [15]. However, the traditional method to study the NPC differentiation and its related function involves staining and cell lysis, which cannot materialized further for the clinical uses [12] [13]. In order to automatically, non-invasively, non-labelled analyze and cultivate cells, a recognition system classifying NPCs under bright-field microscopic imaging is necessary.

Bright-field microscopy is the most ubiquitous and widely available form of microscopy, and therefore bright-field-specific solutions will have a wide audience. Nevertheless, the vast amount and complexity of image data acquired from automated microscopy renders manual analysis unreasonably time-consuming. Accurate automatic cell classification is a difficult issue in cell biology studies using bright-field microscopy due to the immense variability of cell appearance. In images obtained with transmitted light illumination, there is a greater variation of whole-cell size and shape, which is different from images acquired by a fluorescent probe having a characteristic color and show considerable uniformity with respect to size and shape. In addition, cellular debris and other forms of "trash" can be similar in appearance to intact cells.

978-1-4673-7723-2/15 $31.00 © 2015 IEEE
DOI 10.1109/IMCCC.2015.52

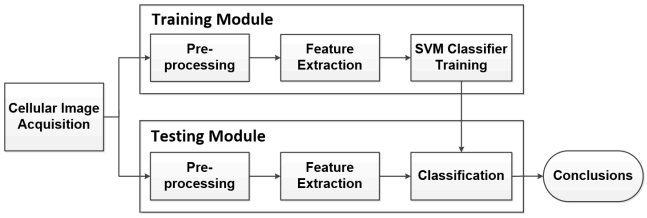

Fig. 1. The recognition system based on SVM.

Detection of unstained viable cells in bright-field images is an inherently difficult task due to the immense variability of cell appearance. Long et al. [11] use neural networks to automatic recognize of cultured cells in bright-field images.

Recently, with the rapid development of machine learning algorithms, some applications based on microscopic images have adopted further techniques in order to improve detection and classification performance. Among these algorithms, Support Vector Machines (SVMs) are broadly known and have shown a high classification performance on many applications, including cell nuclei detection and recognition [6], cell yeast cells on suspension in bioreactors [17], blood cell sorting and tissue cells [14], and cells in culture using fluorescent microscopy [16].

To train the SVM classifier, different categories of feature are explored to characterize the individual cells. When features characterizing different properties of cells are combined properly, the accuracy of classification cells generally improved. Though a variety of features can be estimated from cells, it is important to determine the most relevant features that give the highest classification accuracy because all features are not equally relevant and redundancy among the features affects the performance [**?**]. In this paper, a feature set extracting different kinds of cell features: Zernike moment for shape, morphological features [1], Daubechies (Db) wavelets for spatial features with various scales [8], and Gabor wavelets for spatial frequency features with various orientations and scales [4], of an image of a cell is used to characterize NPCs and non-NPCs.

To circumvent drawbacks of manual classification of NPCs, we investigate the possibility of the SVM for NPC recognition task, where a recognition system based on SVM is proposed to provide a tool to classify NPCs and non-NPCs. Experimental results based on our own cultured cell data support that the proposed NPCs cell recognition system provides an excellent performance in classification of NPCs and non-NPCs.

II. Recognition System with Support Vector Machine

A system based on bright-field microscopy in conjugation with supervised machine learning technique SVM and a image feature set with various cell features automates the recognition of NPCs and non-NPCs. The cell image recognition system with SVM is shown in Fig. 1. The main idea of the system is to train the system with cell samples in order that the system learns from the example images some criterion for distinguishing NPCs from non-NPCs just based on their visual appearance. In this training or learning processing, image features are extracted to support the classification.

The system is composed of two main modules: a training and a test module. In the training module, our own cultured cells are captured with a laboratory microscopy under bright-field settings. Then, an image patch of each NPC is collected within a rectangular window with $M \times N$ pixels around the detected cell center. After pre-processed, features are computed for the $M \times N$ sized image patches prior to performing training. A class label 0 is assigned to the feature vector of each NPC; while a class label 1 is assigned to that of each non-NPC. All of these labelled feature vectors are used to train the SVM classifier.

After training the classifier, it can be used to investigate new cell cultures, in which cell category is unknown. The test images are processed in the similar way as in the training module: capture of micrographs under bright-filed microscopy, detect cell by biological experts, and extract features, where the features are the same as in the training module. After that, the category of each tested cell is determined with the trained SVM classifier.

In our recognition system, the pre-processing process mainly use the Autolevels (AL) algorithm to enhance acquired images. In the feature extraction process, a features set is extracted by Zernike moments, Db wavelets, and Gabor wavelets. The SVM is used in the training and classification process. These processes will be introduced in the following sections with more details.

III. Autolevels

Almost all acquired images under bright-field microscopy have very low dynamic range. To further improve the differences existing in cells, an image enhancement algorithm is applied. This algorithm can be any one of automatic image enhancement algorithms, such as histogram equalization or homomorphic filtering [9]. However, we choose the automatic contrast stretch algorithm commonly referred to as Autolevels (AL) [9] which increases the brightness and contrast of an image by remapping the dynamic range of the data to the dynamic range of the display, eliminating the impact of a small number of outliers in the histogram of the image.

The Autolevels algorithm is an image enhancement via the 2-parameter $(low, high)$ gray level transformation defined by

$$\begin{aligned} \text{for } (l &= 0;\ l < L;\ l{+}{+}) \\ T[l] &= \text{Quantize}\left(L\frac{l - low}{high - low}\right); \end{aligned} \tag{1}$$

where $0 \le low < high \le L-1$, l is the gray level of the original image, and L is possible gray levels per pixel, e.g., an 8-bit image $L = 256$. And,

$$\text{Quantize}(x) = \begin{cases} 0 & x < 0 \\ \lfloor x \rfloor & 0 \le x < L \\ L-1 & x \ge L \end{cases} \tag{2}$$

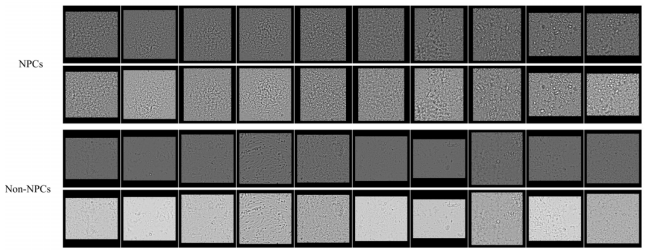

Fig. 2. Sample cell patches in the learning set. (Top-row) NPCs, (Second-row) enhanced NPCs, (Third-row) non-NPCs, and (Bottom-row) enhanced non-NPCs.

where $\lfloor \cdot \rfloor$ denotes the Floor function, and low and $high$ are two extreme values, which control the amount of gray level clipping. That is, the user supplies two small, positive real-valued parameters, $clow$ and $chigh$, $(clow + chigh < 1)$ that specify the amount of low-end and high-end clipping desired respectively. Typical values for $clow$ and $chigh$ are 0.005 in which case approximately 0.5% of the image pixels will have their value clipped to black and another 0.5% will be clipped to white. The two algorithm parameters, low and $high$, are then determined by solve the two equations

$$\begin{aligned} P[low] &= clow, \text{ and} \\ 1 - P[high - 1] &= chigh, \end{aligned} \tag{3}$$

for low and $high$ respectively where $P[\cdot]$ is the cumulative gray level distribution of a being processed image.

IV. Features

To appropriately characterize being classified objects, such as cells, a set of image features for automated recognition is significant. Different features, such as Zernike moments, Haralick texture features, wavelet features, run length features, etc., highlight different properties of cells [**?**]. Morphological features characterize the size of the objects in the cells, the intensity of edges, or the contour of the cells [1] [16]. Zernike moment features are computed from a set of Zernike polynomials and are good shape descriptors of cells [1]. The Db wavelet feature [8] and Gabor wavelet features [5] both give interpretation of cells in a multi-resolution way from spatial or frequency domain respectively [**?**].

A. Zernike Features

Morphological features describe various characteristics of objects and edges in the cellular image as well as the entire cells. These features usually characterize the size of the objects, the intensity of edges, or the contour of the cells. Zernike moments are calculated using an orthogonal basis of Zernike polynomials [1]. Zernkie moments are the projections of an image onto the orthogonal basis functions. They are the magnitude of a set of orthogonal complex moments that are spatially and rotationally invariant. Generally, people calculate Zernike moment features by using the Zernike polynomials with orders ranging from 2-20. In our paper, we calculated 72 Zernike features with order 15.

B. Wavelet Features

Wavelet packets are a generalization of orthonormal and compactly supported wavelets [8]. The coefficients of decomposition serve as distinct features of the original image. Initially, the image was decomposed up to 3rd level and 37 wavelet features are extracted to represent the frequency information. Here, we used Db3 wavelet to decompose an image to level 3, and the average energies, mean, standard deviations, entropy of the three high-frequency image at each level are used as features. Also, the entropy of the low-frequency image is calculated.

C. Gabor Features

Information captured by nonorthogonal Gabor wavelets is mostly the derivative information of an image such as edges [5]. Gabor wavelets are a set of basis functions generate through dilation and rotation of a mother Gabor wavelet. The input image was convolved with a Gabor filter at a specific scale and at a particular orientation. The mean and standard deviation of the responses are taken as texture features. We have used 7 different scales and 8 different orientations, yielding a total of 56 Gabor texture features.

Zernike moments, Daubechies wavelets, and Gabor wavelets represent an image with a set of orthogonal basis functions or polynomials but extracting different properties: shape, spatial scale, and spatial frequency, respectively. Two datasets including NPCs or non-NPCs consist different cells. Therefore, they will yield different sets of features. In all, 165 features are extracted to combine together to represent NPCs and non-NPCs, which are the input of the classifier SVM to train or classify NPCs and non-NPCs.

V. Support Vector Machine

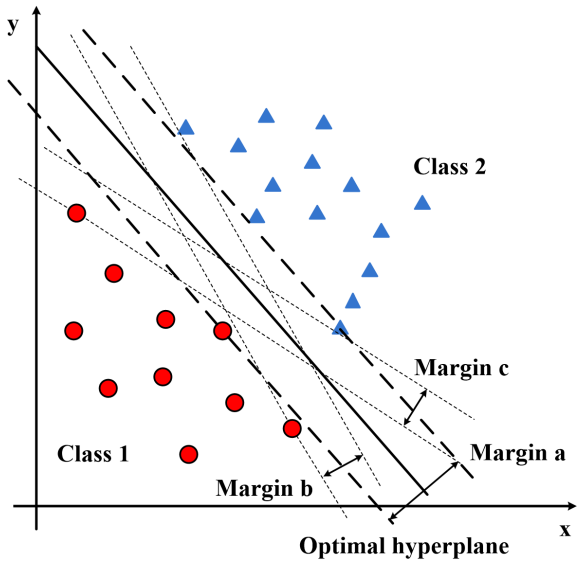

Fig. 3. SVM finds the hyperplane separating two classes with maximum distance.

SVMs, based on the principle of structural risk minimization form a well established approach in the application of machine learning algorithms and are proving to be particularly promising when used to construct accurate models based on large feature spaces [7] [3] [Mjolsness and DeCoste(2001)]. Particularly, SVMs deliver state-of-the-art performance in real-world applications [7]. They have some superiorities over other approaches, especially: (a) global minimum solution, and (b) learning and generalization in huge dimensional input spaces [7] [3]. Essentially, they use a hypothesis space of linear functions in a high dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory. The aim is to find a hyperplane which can classify two classes of data correctly, by maximizing the distance between the two classes of data and the hyperplane, in a space of higher dimension. From Fig. 3, we can see class 1 and class 2 can be separated by many hyperplanes but only the optimal hyperplane separates two classes with maximum margin. Margin b and c are shorter than margin a. Those points lying on the margin generated by the optimal hyperplane are support vectors (SV) [6].

A SVM classifier has to be initially trained before use for classification. During training, samples of two classes are presented using a training procedure. After analyzing these training data, the classifier finds the hyperplane which separates two classes with maximum margin in a given dimensional space. Then, the SVM [2] performs pattern classification based on the separating hyperplane at a maximum distance determined by the SV in the training set.

VI. Experimental results

A. Expermental Data

To culture NPCs, urine samples are collected from different donors. For reprogramming, an oriP/EBNA1-based pCEP4 episomal vector containing the OCT4, SOX2, KLF4, and SV40LT genes20 and a pCEP4 vector carrying the miR302-367 precursor21 were co-transfected into urine cells via nucleofection (Amaxa Basic Nucleofector Kit for primary mammalian epithelial cells with the T-013 program, Lonza). Transfected urine cells were directly plated to Matrigel-coated six-well plates ($1 - 3 \times 105$ cells per well) in urine cell culture medium. On day 2 post-transfection, the media were changed into reprogramming media mTeSR or 5i (mTeSR supplemented with 5i 0.5 μM A83-01, 1 μM PD0325901, 3 μM CHIR99021, 0.5 μM thiazovivin and 0.2 μM DMH1). Medium was changed every 2 day during the reprogramming. Fifteen days after transfection, colonies were picked up and transferred onto a new Matrigel plate for further passaging.

After sixteen days culturing, the plate is taken a picture from day 6 post-transfection. The images were taken with a instrument Solentim Cell Metric under bright-field microscopy. Each acquired image with 17702×17684 pixels has various NPCs and non-NPCs. Then, biologists manually detect and find out NPCs and non-NPCs, and clipped these areas from the acquired images and save each one as a PNG image. The resolutions of the clipped images are range from about 200×200 to 500×500. Therefore, these images are normalized to the same size in the pre-processing process. So far, our image data consists of 90 images. Among these images, 40 images contained NPCs. Some samples are shown in Fig. 2. The Fig. 2 demonstrate that the NPC images are more like texture images, but with different variations. However, the non-NPC images are varied from including only plate without or less cells, cell debris, dead cells, to unsuccessfully differentiated NPCs. Therefore, the recognition task for NPCs under bright-field microscopy is an interesting challenge.

B. Results

To validate our SVM-based recognition system, the images of each class: NPC and non-NPC are split into training and test images. The percent of training image is varied from 90% to 10%. Therefore, the effectiveness of the system is determined by the percentage of test images that are classified correctly using the training images.

Table I shows the performance of the SVM-based recognition system for NPC and non-NPC images. With the percentage of training samples decreasing, the accuracy of test is reduced from 100% to 56.8% as expected. However, even the percentage of the training samples is less than the test samples, the accuracy is sometimes still good enough, e. g. 90.7% accuracy for 40% training samples. The accuracy, precision, and recall are satisfied while the percentage of training samples is larger than 50%.

VII. Conclusions

When combining with effective image feature set and appropriate feature selection algorithm, the SVM can help to generate a simple, suitable, and high performance classifier for various applications. According to this fundamental idea, we proposed a machine learning system based on SVM to classify NPC and non-NPC images acquired from bright-field microscopy in this paper. Our approach shows a 90% success rate based on our cultured cells with enough training samples. This supports the high performance of machine learning approach in this application, which can pave a pathway for biologists to analyze cell cycle process in trans-differentiating processes of NPCs based generation medicine.

In the future work, we are extending the current work to more cell images. And feature selection process with certain algorithm such as linear discriminant analysis or genetic algorithm will be included in the system to determine the best subset of these features according to certain criteria so that the best performance can be achieved. The advantages of feature selection can be versatile: for instance, reducing dimensionality, enhancing system robustness, increasing recognition rate, etc. More importantly, the system has to be automated by localizing cells without human intervention. After that, they system could be easily applied to the microscopic image for different applications.

Acknowledgements

The author wishes to thank the Ministry of Finance Life Science Instrumentation Development Program managed by

TABLE I
COMPARISONS OF TEST ACCURACY WITH VARIOUS SPLITS OF NPC AND NON-NPC IMAGES.

% of training samples	90%	80%	70%	60%	50%	40%	30%	20%	10%
Accuracy of test	100%	100%	100%	94.4%	93.3%	90.7%	81%	58.3%	56.8%
Precision of test	100%	100%	100%	100%	100%	100%	100%	100%	100%
Recall of test	100%	100%	100%	87.5%	85%	79.2%	57.1%	6.25%	2.78%
Errors/Test samples	0/9	0/18	0/27	2/36	3/45	5/54	12/63	30/72	35/81

Chinese Academy of Sciences with grant No. ZDYZ2012-3 for the funding, which made this work possible.

REFERENCES

[1] Michael V Boland and Robert F Murphy. A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of hela cells. *Bioinformatics*, 17(12): 1213–1223, 2001.

[2] Christopher JC Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.

[3] Nello Cristianini and John Shawe-Taylor. An introduction to support vector machines, 2000.

[4] Scott Doyle, Shannon Agner, Anant Madabhushi, Michael Feldman, and John Tomaszewski. Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 496–499. IEEE, 2008.

[5] Simona E Grigorescu, Nicolai Petkov, and Peter Kruizinga. Comparison of texture features based on gabor filters. *Image Processing, IEEE Transactions on*, 11(10):1160–1167, 2002.

[6] Ji Wan Han, Toby P Breckon, David A Randell, and Gabriel Landini. The application of support vector machine classification to detect cell nuclei for automated microscopy. *Machine Vision and Applications*, 23 (1):15–24, 2012.

[7] Marti A. Hearst, Susan T Dumais, Edgar Osman, John Platt, and Bernhard Scholkopf. Support vector machines. *Intelligent Systems and their Applications, IEEE*, 13(4):18–28, 1998.

[8] Kai Huang and Robert F Murphy. From quantitative microscopy to automated image understanding. *Journal of biomedical optics*, 9(5):893–912, 2004.

[9] Bo Jiang, Glenn A. Woodell, and Daniel J. Jobson. Novel multi-scale retinex with color restoration on graphics processing unit. *Journal of Real-Time Image Processing*, 2014.

[10] liu2011features Song Liu, Piyushkumar A Mundra, and Jagath C Rajapakse. Features for cells and nuclei classification. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 6601–6604. IEEE, 2011.

[11] Xi Long, W Louis Cleveland, and Y Lawrence Yao. Effective automatic recognition of cultured cells in bright field images using fisher's linear discriminant preprocessing. *Image and Vision Computing*, 23(13):1203–1213, 2005.

[Mjolsness and DeCoste(2001)] Eric Mjolsness and Dennis DeCoste. Machine learning for science: state of the art and future prospects. *Science*, 293(5537):2051–2055, 2001.

[12] Tim W Nattkemper, Helge J Ritter, and Walter Schubert. A neural classifier enabling high-throughput topological analysis of lymphocytes in tissue sections. *Information Technology in Biomedicine, IEEE Transactions on*, 5(2):138–149, 2001.

[13] Tim W Nattkemper, Thorsten Twellmann, Helge Ritter, and Walter Schubert. Human vs. machine: evaluation of fluorescence micrographs. *Computers in biology and medicine*, 33(1):31–43, 2003.

[14] H Wang, C Zheng, Y Li, H Zhu, and X Yan. [application of support vector machines to classification of blood cells]. *Sheng wu yi xue gong cheng xue za zhi= Journal of biomedical engineering= Shengwu yixue gongchengxue zazhi*, 20(3):484–487, 2003.

[15] Meng Wang, Xiaobo Zhou, Randy W King, and Stephen TC Wong. Context based mixture model for cell phase identification in automated fluorescence microscopy. *BMC bioinformatics*, 8(1):32, 2007.

[16] Meng Wang, Xiaobo Zhou, Fuhai Li, Jeremy Huckins, Randall W King, and Stephen TC Wong. Novel cell segmentation and online svm for cell cycle phase identification in automated microscopy. *Bioinformatics*, 24 (1):94–101, 2008.

[17] Ning Wei, Jia You, Karl Friehs, Erwin Flaschel, and Tim Wilhelm Nattkemper. An in situ probe for on-line monitoring of cell density and viability on the basis of dark field microscopy in conjunction with image processing and supervised machine learning. *Biotechnology and bioengineering*, 97(6):1489–1500, 2007.

[18] Xiaobo Zhou, Xinhua Cao, Zach Perlman, and Stephen T.C. Wong. A computerized cellular imaging system for high content analysis in monastrol suppressor screens. *Journal of Biomedical Informatics*, 39(2): 115 – 125, 2006.