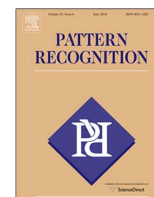
Pattern Recognition 61 (2017) 234–244

Contents lists available at ScienceDirect

Pattern Recognition

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



Pattern recognition and classification of two cancer cell lines by diffraction imaging at multiple pixel distances

He Wang [a], Yuanming Feng [a], Yu Sa [a], Jun Q. Lu [b], Junhua Ding [c], Jun Zhang [a], Xin-Hua Hu [b,*]

[a] Department of Biomedical Engineering, Tianjin University, Tianjin 300072, China
[b] Department of Physics, East Carolina University, Greenville, NC 27858, USA
[c] Department of Computer Science, East Carolina University, Greenville, NC 27858, USA

ARTICLE INFO

Article history:
Received 9 February 2016
Received in revised form
23 May 2016
Accepted 22 July 2016
Available online 25 July 2016

Keywords:
Single-cell assay
Image pattern analysis
Diffraction imaging
Cell classification
Light scattering
Flow cytometry
Cancer cells

ABSTRACT

Rapid and label-free imaging methods for accurate cell classification are highly desired for biology and clinical research. To improve consistency of classification performance, we have developed an approach of pattern analysis by gray level co-occurrence matrix (GLCM) algorithm to extract textural features at multiple pixel distances from cross-polarized diffraction image (p-DI) pairs, which were acquired with a method of polarization diffraction imaging flow cytometry using one time-delay-integration camera for significantly reduced blurring. Support vector machine (SVM) based classification was performed to discriminate HL-60 from MCF-7 cells using the GLCM features and consistency of optimized SVM classifiers was evaluated on three test data sets. It has been shown that the classification accuracy of the best performing SVM classifiers at or above 98.0% can be achieved among all four data sets for each of the three incident beam polarizations. These results suggest that the p-DI pair data provide a new platform for rapid and label-free classification of single cells with high and consistent accuracy.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Cell classification by recognition of image patterns is of fundamental interest and can have wide applications in life science and clinics [1–8]. Compared to microscopy, flow cytometry (FCM) with imaging capability allows rapid data acquisition and extraction of pattern parameters from large numbers of cells on a single-cell basis [9]. The high rate of image acquisition through FCM demands development of automated image analysis algorithms to process and analyze the big data of acquired images. Based on previous studies of coherent light scattering [10–15], we have developed a polarization diffraction imaging flow cytometry (p-DIFC) method to measure the spatial distribution of light scattered by single cells illuminated by a linearly polarized laser beam [16–22]. Different from fluorescence imaging, the cross-polarized diffraction image (p-DI) pair data acquired by the p-DIFC method record intensity distribution of coherent light scattered by a cell due to the heterogeneous 3D distribution of intracellular refractive index and needs no cell staining. The acquired p-DI pair data are results of coherent superposition of wavefields emitted by the induced dipoles of molecules in the imaged cell and thus present speckle patterns carrying molecular and morphological information of the cell [12,13,15]. We have shown that the p-DI pair data can be used to accurately distinguish two cell lines of high morphological similarity derived from human T and B cancer cells and two types of prostate cells derived from cancer and normal cells [20,22].

Conventional cell images are typically acquired with non-coherent fluorescent light through a microscope or imaging FCM which present 2D projections of a 3D object. Image segmentation is generally needed for further analysis [9]. In contrast, the speckle patterns in a p-DI pair result from superposition of coherent wavefields from all excited intracellular molecules as "digital holograms" and needs no segmentation. A pixel-based global image processing algorithm often suffices to quantify the patterns or textures of such images that can be automated for rapid processing. We have developed a gray-level-co-occurrence-matrix (GLCM) based software to quantitatively characterize textures of p-DI pair data [22–25]. The GLCM parameters were determined from a matrix of elements given by the co-occurring probabilities of paired pixels separated by **d** as the displacement vector with $|\mathbf{d}| = 1$. A total of 38 parameters were obtained from each p-DI pair to form feature vectors, which were used to train support vector machine (SVM) classifiers with different kernels. With the above algorithms, we have shown that the p-DIFC method performs well for accurate and label-free cell classification through automated image pattern recognition.

Despite the attractive qualities of the p-DIFC method, however,

* Corresponding author.
E-mail address: hux@ecu.edu (X.-H. Hu).

http://dx.doi.org/10.1016/j.patcog.2016.07.035
0031-3203/© 2016 Elsevier Ltd. All rights reserved.

we have found that the performances of the trained SVM classifiers are not consistent when they were applied to test sets of p-DI pairs data acquired in different runs of measurements. The causes of inconsistency may relate to the variation of diffraction patterns in different data sets, which are due to cell speed fluctuations leading to different degree of blurring for images acquired with conventional CCD cameras, positioning errors of cells relative to the focus of incident beam and the high sensitivity of extracted GLCM parameters on fine pattern changes among pairs of nearest-neighbor pixels that are unrelated to cell morphology or molecular composition. In our previous study of T versus B cell lines and PC3 versus PCS prostate cells [20,22], SVM classifiers had to be trained and tested with the p-DI pairs acquired in the same measurement run for achieving high values of classification accuracy A. The value of A decreases markedly if the SVM classifier trained with data from one run was applied to the test data acquired in another run on a different day [22]. For example, the value of A reduces from 99.5% to 62.8% in the case of PC3 versus PCS cells. Such reduction in performance prevents application of the p-DIFC method with pre-trained SVM classifiers on p-DI data acquired later.

The current study focuses on pattern analysis of p-DI data and effect of blurring as a part of our research efforts to solve the accuracy-dropping problem. A new configuration of illumination and imaging has been developed to eliminate or significantly reduce motion blur with one time-delay-integration (TDI) CCD camera and decrease the sensitivity of image patterns of acquired data with enlarged focal spot for the incident beam [21]. With this imaging configuration we have investigated the GLCM approach of pattern analysis with $d \geq 1$ to improve the robustness of the SVM classifiers. In this report, we present the results of pattern recognition and classification on two cancer cell lines, HL-60 versus MCF-7, by the p-DI data acquired with the new imaging configuration. The dependence of p-DI parameters and classification accuracy on d has been analyzed to examine the benefits of using different values of d for improved performance. We have also blurred the measured p-DI data with no or little blurring by window smoothing to investigate the effect of blurring on classification. These results demonstrate that the new imaging configuration and GLCM analysis with different d can improve significantly the robustness of SVM based classification.

2. Related work

Automated image analysis for cell classification often proceed by extracting feature parameters related to cell morphology and/or image texture. Morphological parameters can be regarded as quantitative extension of human perception and obtained by characterizing geometric structures of organelles in images acquired by bright-field or fluorescent microscopy, which are typically performed with cells stained with contrast or fluorescent reagents. In deriving these parameters, image segmentation is necessary on the basis of unique correspondence between regions of interest (ROI) and individual organelles. By comparison, characterization of texture can be performed either globally in a given image or locally over clustered pixels to quantify patterns of intensity variation. In the latter case, proper image segmentation may also be needed. With wide availability of algorithms and ever increasing computing power per unit cost, the two types of image parameters are frequently combined to achieve best outcomes of classification. Below, we briefly review different methods of automated cell classification using morphology and texture features as discriminators in terms of their performance.

In an early study of multiclass classification, 9 morphology and 4 texture features have been extracted from 2D images to discriminate 7 *Eimeria* parasite species of different shapes. The parameters were combined as descriptors for training Bayesian classifiers and an overall correct classification rate or classification accuracy of 86.8% was obtained over a data set of 3891 oocyst images [2]. To expand from individual efforts, a contest of HEp-2 Cells Classification has been held in 2012 to evaluate different methods on a common test data set. The contest was designed to discriminate 6 antinuclear autoantibody fluorescence patterns on a given data set of 28 indirect immunofluorescence (IIF) images containing 1457 HEp-2 cells, with half of the IIF images used as the training data [6]. A total of 28 automated methods have been submitted and tested to extract morphological and/or texture features as the input to different classifiers including the SVM and k-nearest-neighbor methods, among others. The test data set consists of manually segmented images of single cells from 14 reserved IIF images. The mean accuracy of classifying the IIF patterns in the segmented cell images into 6 classes ranged from 21.4% to 70.4% for the submitted methods, which were less than the baseline value of 77.5% achieved by a specialist against the ground truth data of the 28 IIF images established by experts [6,26]. Notably, the best performing method utilized the image texture information by quantifying the co-occurrence of local binary differences of neighboring pixel intensities among image pixels [4]. Following the 2012 competition, additional studies have been performed to further improve the performance of automated classification of the HEp-2 cells [7]. It has been reported, for example, that combination of multiple SVM classifiers trained with different sets of sparsely encoded texture features could produce a mean classification accuracy of 87.1% on the IIF patterns of HEp-2 cells on heldout test data sets [8].

Several methods among those submitted to the HEp-2 cell contest discussed above utilized texture features extracted by the GLCM algorithm. Among these, the highest value of mean classification accuracy was achieved at 63.0% on the common test data set using 4 GLCM parameters coupled with additional features of morphology and gradient and multi-class SVM classifiers [5,6]. Textural and morphological features have also been extracted to classify histopathological images of colon tissues into three classes of normal, low- and high-grade adenocarcinoma and a mean value of 57.1% was achieved for accuracy using only the GLCM features [3]. Other applications of image texture analysis include study of apoptotic cells which identified correlations between the GLCM features in ROI of cell images to the changes of nuclear morphology characterizing apoptosis [1]. It is worth noting that in all studies discussed in this section, the image feature extraction was preceded by manual segmentation to remove regions in an image that do not relate to the organelles of interest. For the images acquired with the p-DIFC method presented here, segmentation is unnecessary since all light signals presented by the pixels in a diffraction image are from the imaged cell. Furthermore, each non-dark pixel represents the intensity of total electromagnetic fields emitted by induced molecular dipoles inside the imaged cell. Hence, the p-DI pairs present a new data platform to investigate cell classification with the attractive qualities of fast acquisition, low background noise and capacity for fully automated processing.

3. Methods

3.1. The p-DIFC system and cell measurement

Details of the p-DIFC system have been published elsewhere for cell positioning by hydrodynamic focusing in a square flow channel and imaging of coherent light scatter [16–18,20,21]. Briefly, a continuous-wave solid state laser (MGL-III-532-100, CNI) was used to produce an incident beam of 532 nm in wavelength. Two cylindrical lenses of 500 mm and 60 mm in focal lengths

Download English Version:

https://daneshyari.com/en/article/533064

Download Persian Version:

https://daneshyari.com/article/533064

Daneshyari.com