Comparison study of distinguishing cancerous and normal prostate epithelial cells by confocal and polarization diffraction imaging

Wenhuan Jiang
Jun Qing Lu
Li V. Yang
Yu Sa
Yuanming Feng
Junhua Ding
Xin-Hua Hu

Comparison study of distinguishing cancerous and normal prostate epithelial cells by confocal and polarization diffraction imaging

Wenhuan Jiang,[a] Jun Qing Lu,[a] Li V. Yang,[b] Yu Sa,[c] Yuanming Feng,[c] Junhua Ding,[d] and Xin-Hua Hu[a,*]
[a]East Carolina University, Department of Physics, Greenville, North Carolina 27858, United States
[b]East Carolina University, Department of Internal Medicine, Brody School of Medicine, Greenville, North Carolina 27834, United States
[c]Tianjin University, Department of Biomedical Engineering, 92 Weijin Road, Tianjin 300072, China
[d]East Carolina University, Department of Computer Science, Greenville, North Carolina 27858, United States

Abstract. Accurate classification of malignant cells from benign ones can significantly enhance cancer diagnosis and prognosis by detection of circulating tumor cells (CTCs). We have investigated two approaches of quantitative morphology and polarization diffraction imaging on two prostate cell types to evaluate their feasibility as single-cell assay methods toward CTC detection after cell enrichment. The two cell types have been measured by a confocal imaging method to obtain their three-dimensional morphology parameters and by a polarization diffraction imaging flow cytometry (p-DIFC) method to obtain image texture parameters. The support vector machine algorithm was applied to examine the accuracy of cell classification with the morphology and diffraction image parameters. Despite larger mean values of cell and nuclear sizes of the cancerous prostate cells than the normal ones, it has been shown that the morphologic parameters cannot serve as effective classifiers. In contrast, accurate classification of the two prostate cell types can be achieved with high classification accuracies on measured data acquired separately in three measurements. These results provide strong evidence that the p-DIFC method has the potential to yield morphology-related "fingerprints" for accurate and label-free classification of the two prostate cell types. © *2015 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JBO.21.7.071102]

Keywords: single-cell analysis; cell classification; confocal imaging; light scattering; flow cytometry; prostate cancer.

Paper 150544SSR received Aug. 12, 2015; accepted for publication Oct. 26, 2015; published online Nov. 30, 2015.

1 Introduction

Prostate cancer is the most frequent nonskin cancer among men in the United States and European countries. Active research has been pursued to develop new surveillance methods with improved sensitivity and specificity over existing ones such as the prostate-specific antigen (PSA) test for accurate diagnosis and effective treatment in clinics.[1] Among various approaches, detection and enumeration of circulating tumor cells (CTCs) in peripheral blood samples after enrichment present promising potentials by providing CTC number as the surveillance biomarker and prognosis predictor in prostate cancer patients.[2,3] A widely used CellSearch technique approved by the U.S. Food and Drug Administration employs an approach of immunomagnetic enrichment before CTC enumeration through immunofluorescence microscopy.[2,3] Other methods have been reported for sample enrichment with microposts or nanosheets coated with EpCAM antibodies on the microfluid technology platform and CTC capture by imaging.[4] The immuno-based methods to select cells with EpCAM expression, however, may fail due to the variation of the targeted expression. It has been shown that downregulation of EpCAM can occur in CTCs captured from the blood of prostate cancer patients as a result of epithelial-mesenchymal transitions, which could account for the inconsistency between the small number of CTCs in patients with a confirmed diagnosis of prostate cancer.[5,6] This leads to the desire for

exploration of label-free approaches to analyze and classify different types of prostate cells. The first step in this direction is to examine the feasibility of any new method, which may prepare ground to develop practical approaches for detection of CTCs in enriched samples since the numbers of CTCs are extremely small in fresh blood samples. In this report, we present a feasibility study of prostate cell classification through diffraction imaging in comparison to the confocal imaging based three-dimensional (3-D) morphology characterization.

Light elastically scattered by single cells illuminated with a laser beam remains highly coherent as well and its spatial distribution patterns correlate with the 3-D distribution of intracellular refractive index relative to the host medium. Therefore, investigation of light scattering and its spatial distribution patterns provides a route to rapidly acquire information on the 3-D morphology of the scatterer and molecular polarization.[7–12] A method of polarization diffraction imaging flow cytometry (p-DIFC) has been developed to image coherent light scattered by single particles or cells using a microscope objective at off-focus positions to increase image contrast and adjust the angular cone of detection.[13–16] With this method, cells of high similarity in their morphology can be distinguished by the texture parameters of the cross-polarized diffraction images extracted with automated algorithms.[17–19] We have also employed a confocal microscopy based method to quantify 3-D morphology of cells.[14,18,20] An algorithm of a support vector machine (SVM)

---

has been applied to map the selected parameters into a high-dimensional feature space for classifying the two cell types with the parameters acquired by the p-DIFC method. The results of classification are presented here with two types of human prostate cells by acquisition of a 3-D parameter with the confocal imaging method and cross-polarized diffraction image parameters with the p-DIFC method. We conclude with a discussion on the dependence of the cross-polarized diffraction image parameters on 3-D morphology and molecular polarization to understand these initial results and their implications for potential application on CTC detection.

2 Materials and Methods

2.1 Microscopy Measurement and Three-Dimensional Reconstruction

We have investigated two prostate cell types that are denoted in this report as PC3 for the cancer cell line and PCS for the normal one. A PC3 human prostate cancer cell line of high metastatic potential (CRL-1435, ATCC) was maintained in RPMI-1640 (Gibco BRL, Life Technologies) supplemented with 10% fetal calf serum. The culture media were supplemented with penicillin 100 U/ml, streptomycin 100 μg/ml, and glutamine 0.1 mg/ml. The normal human prostate epithelial cells (PCS440010, ATCC) were maintained in the prostate epithelial cell basal medium (PCS440030, ATCC) supplemented with the prostate epithelial cell growth kit (PCS440040, ATCC). The adherent cells in their logarithmic phases of growth were detached from culture plates with trypsin–EDTA solution, resuspended in culture medium, and kept on ice before the confocal and p-DIFC measurements. Viability of the suspended cells was checked by a trypan blue exclusion test before measurement and percentages of viable cells were found to be ~95%. The concentrations of the cell suspension samples were adjusted to a value of about 1×10^6 cells/mL for p-DIFC measurement.
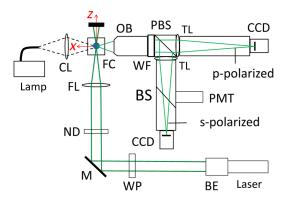
For confocal imaging, the cells were first double-stained for nucleus and mitochondria with fluorescent dyes (Syto-61 and Mitotracker-Orange, Life Technologies) with protocol detailed in Zhang et al.[20] A laser scanning confocal microscope (LSM 510, Zeiss) was used to acquire image stacks with a 63× water-immersion objective of 1.2 in NA and a 4× digital zoom option provided by the Zeiss image acquisition software to reduce pixel size to 0.07 μm. The number of two-dimensional (2-D) fluorescence image slices in each stack ranges from about 40 to about 70 with 0.5 μm in the translation step in air along the direction perpendicular to the slices. The acquired confocal image slices in a stack were segmented using an in-house developed software.[10,14,20] The segmented slices were then used to add image slices through interpolation to make final slice separation, after correction due to light refraction, approximately the same as the pixel size of the image slices to obtain cubic voxels for 3-D reconstruction. The details of the image segmentation and reconstruction have been provided elsewhere.[20] A total of 29 voxel-based parameters were calculated to quantify the morphology of the different organelles of cytoplasm, nucleus, and mitochondria in reconstructed cells. We used the SPSS software (Version 19, IBM) to perform the two-sample t-tests of the morphology parameters and assess the statistical significance of the differences in the 3-D parameters between the two cell types. The definitions and a table of 29 morphologic parameters are provided online.[21]

2.2 Diffraction Imaging Flow Cytometric Measurement

Design details of the p-DIFC system have been published elsewhere for cell positioning through hydrodynamic focusing in a square flow channel and the imaging of scattered light.[13–15,19] Briefly, a continuous-wave solid state laser (MGL-III-532-100, CNI) was used to produce an incident beam of 532 nm in wavelength and up to 180 mW in power. A spherical lens of 75 mm in focal length was used to focus the incident beam onto the core fluid carrying the cells with a spot diameter of about 30 μm. The profile of a linearly polarized incident beam propagating along the z-axis is close to a Gaussian distribution, and the power was measured as P_0 before the focusing lens and adjusted with neutral density filters. The loss of the optical power by the index-mismatch interfaces from the focusing lens to the imaged cell was estimated to be about 17%. Figure 1 presents the schematic of the p-DIFC system.

The polarization direction was set to one of the three directions of horizontal (hor), vertical (ver), or 45 deg from horizontal with a half-wave plate. The light scatter from flowing cells was collected by an infinity-corrected 50× objective of 0.55 in NA (378-805-3, Mitutoyo) within an angular cone, which was centered at the scattering polar angle $\theta_s = 90$ deg along the x-axis and of a cone angle θ_{sm} in water. A polarizing beam splitter was employed to divide the scattered light into the s- and p-polarized beams for acquisition of two cross-polarized diffraction images of 640×480 pixels and a 12-bit pixel depth by two CCD cameras (LM075, Lumenera). Camera-triggering signals were produced with a photomultiplier, and the exposure time was set to 0.3 ms to reduce image blurring for the imaged cells flowing at a speed of about 6 mm/s.

To vary the angular cone of the imaged light and increased image contrast, the imaging unit consisting of the objective, optics, and cameras was translated toward the flow chamber by a distance of $\Delta x = 150$ μm from the focusing position conjugate to the imaged cell or core fluid. It has been shown that at these nonconjugate positions, the acquired images present patterns of diffraction in high-fidelity because of the unique correspondence between the angular positions of the coherent light scatter and the imager pixel positions.[14,16] Furthermore, the maximum cone angle θ_{sm} of the scattered light passing through

Fig. 1 Schematic of an experimental p-DIFC system for acquisition of s- and p-polarized diffraction images. BE: beam expander; WP: half-wave plate; M: mirror; ND: neutral density filters; FL: focusing lens; FC: flow chamber; CL: condenser lens; OB: objective; WF: 532 nm wavelength filter; PBS: polarizing beam splitter; TL: tube lenses; BS: beam splitter; PMT: photomultiplier; CCD: camera. The x-axis and z-axis are labeled by red lines.

the exit pupil of the objective decreases linearly as Δx increases, which allows the variation of the angular cone viewed on the acquired images. At $\Delta x = 150\ \mu m$, $\theta_{sm} = 23.3$ deg for the objective used in the imaging unit.[16] The throughput of the p-DIFC measurement was maintained at about 1 cell/s and mainly limited by the frame rate of cameras triggered externally. Before extraction of p-DIFC image parameters, the acquired image pairs were first filtered with an in-house developed pre-processing software. The overexposed and underexposed image pairs were removed, which were defined, respectively, as those with one image of saturated pixels more than 1% of the total pixels or both images of average pixel intensities <2% of the saturated pixel values ($= 4095$ for 12-bit images). Additionally, image pairs with strip patterns of high symmetry or large speckles were also removed, since these have been shown to associate with spherical particles or aggregated small particles or cellular debris instead of intact cells.[18]

The remaining diffraction image pairs were converted linearly from the 12-bit images of the acquired data into normalized 8-bit images in which the minimum and maximum pixel intensities in the 12-bit image were set to 0 and 255. The bit reduction was designed to speed up the subsequent parameter extraction by the gray-level-co-occurrence-matrix (GLCM) algorithm without significant loss of dynamic range.[17,22] Using the GLCM algorithm, a total of 38 image parameters have been extracted as g_{im} with $m = 1, 2, \ldots, 38$ to characterize the texture and pixel intensity of the normalized image pair for the i'th imaged cell.[19] The list of the 38 diffraction image parameters and their definitions are provided online.[21]

2.3 Diffraction Image Analysis and Cell Classification

With either 3-D parameters from the confocal image data or parameters of the cross-polarized diffraction image data, the i'th imaged cell can be represented by a parameter vector given by $\mathbf{c}_i = \sum_{m=1}^{N_m} g_{im}\mathbf{u}_m$, where $\mathbf{u}_m$ are unit vectors defining a parameter space Σ_p of N_m-dimension. The set of N_m parameters, $\{g_{im}\}$, consists of either all or a portion of the 3-D parameters with $1 \leq N_m \leq 29$ or of the GLCM parameters with $1 \leq N_m \leq 38$. To classify the imaged cells as represented by their parameter vectors, we chose a statistical learning algorithm of SVM for its well-recognized balance between training complexity and test performance in comparison to other machine-learning algorithms such as the neural network method.[23,24] Instead of direct classification by $\mathbf{c}_i$ of the training data in Σ_p, the SVM approach maps the input vectors into a high-dimensional feature space E by a kernel function $K(\mathbf{c}_i, \mathbf{c}_j)$ with a training data set consisting of N_{tra} cells.

Specifically, SVM constructs a matrix Q of rank N_{tra} and defines its elements $Q_{ij} = t_i t_j K(\mathbf{c}_i, \mathbf{c}_j)$ with the type identifier of t_i and t_j ($= 1$ or -1 for two types of cells) and the index i or j ranging from 1 to N_{tra}. The mapping from Σ_p to E allows classification in E and solves for $\boldsymbol{\alpha}$ as a quadratic optimization problem by minimizing $(1/2)\boldsymbol{\alpha}^T \cdot Q\boldsymbol{\alpha} - \mathbf{e}^T \cdot \boldsymbol{\alpha}$ under the constraints of positive definite $\boldsymbol{\alpha}$ and $\mathbf{t}^T \cdot \boldsymbol{\alpha} = \Delta$.[25] By defining $\boldsymbol{\alpha}^T = (\alpha_1, \ldots, \alpha_{N_{tra}})$ and $\mathbf{e}^T = (1, \ldots, 1)$ in E, one can obtain $\boldsymbol{\alpha}$ together with a bias parameter b from a training data set in terms of the vectors $\mathbf{c}_i$ with N_m parameters as components and a kernel function. These define an SVM model with a decision function F as the classifier

$$F(\mathbf{c}) = \sum_{i=1}^{N_{tra}} t_i \alpha_i K(\mathbf{c}_i, \mathbf{c}) + b, \qquad (1)$$

where $\mathbf{c}$ is the parameter vector of an "unknown" cell drawn from test data. The sign of F determines the type of cell for $\mathbf{c}$. We have employed an open-source code package of SVM (LIBSVM 2.86)[25] to investigate the classification of the two prostate cell types with four types of kernel functions: linear, polynomial, Gaussian radial basis function (RBF), and sigmoid.

By assessing the performance, we define the following numbers to measure the outcomes of the classification according to the values of F: TP as the number of correctly identified image pairs acquired from the given PC3 cells with $F > 0$, TN as the number of correctly identified image pairs from the PCS cells with $F < 0$, FP as the number of image pairs of PCS cells that are incorrectly identified as PC3 cells with $F > 0$, and FN as the number of image pairs of PC3 cells that are incorrectly identified as PCS cells with $F < 0$. SVM models were evaluated by their classification accuracy A on a given data set from the above values as

$$A = \frac{TP + TN}{TP + TN + FP + FN}. \qquad (2)$$

3 Results

3.1 Confocal Measurement and Quantitative Characterization of Three-Dimensional Morphology

We have performed confocal imaging and 3-D reconstruction of the detached PC3 and PCS cells after double-staining of the nucleus and mitochondria. Following reconstruction, a total of 29 parameters were obtained.[20] Selected confocal image slices and perspective 3-D views of two PC3 and two PCS cells are presented in Fig. 2. Table 1 lists the mean values and standard deviations of 17 key parameters together with the p-values to test the statistical significance of the difference between the two prostate cell types. From these data, one can clearly see that the PC3 cells' cell and nuclear volumes are larger on average than those of the PCS cells. Similar morphologic differences of statistical significance can also be observed in the cell shapes as indicated by the distribution of the membrane voxels' distances R_c to the centroid.

To examine the difference in morphology closely, we provide in Fig. 3 the scatter plots of the 3-D parameters selected from Table 1 with p-values <0.05 for the imaged cells. While most of the cells in each type overlap each other in these scatter plots, the PC3 cells as a group appear to have significantly smaller spreads in their values of cellular and nuclear parameters than those of the PCS cells, which can also be noted from the standard deviations of most of the other parameters in Table 1. Taken together, the quantitative characterization provides insight into the morphologic differences between the two cell types and demonstrates that the 3-D parameters alone are not sufficient for accurate classification, which is confirmed by the SVM classification results to be presented later.
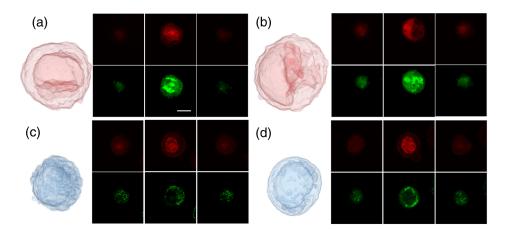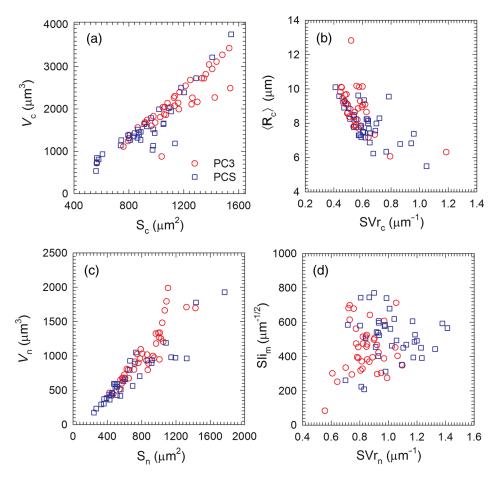
Fig. 2 Perspective views of 3-D structures (not to scale) and selected slices from confocal image stacks acquired from two PC3 and two PCS cells with nucleus and cytoplasm stained and imaged in red (top rows) and mitochondria and cytoplasm stained in green (bottom rows) channels. The cell type and values of cell volume in μm^3, nucleus-to-cell volume ratio, and mitochondria-to-cell volume ratio are: (a) PC3, 2140, 28.1%, 4.06%; (b) PC3, 2073, 32.4%, 11.3%; (c) PCS, 1185, 46.2%, 22.8%, and (d) PCS, 1451, 40.9%, 6.89%. Bar = 10 μm.

Table 1 Morphologic parameters of the two prostate cell types.

Parameter	Symbol	Unit	Mean ± standard deviation		p^a
			PC3 ($n = 40$)[a]	PCS ($n = 38$)[a]	
Cell surface area[b]	S_c	μm^2	1135 ± 226	918.6 ± 229	7.0×10^{-4}
Cell volume[c]	V_c	μm^3	2116 ± 623	1543 ± 665	1.9×10^{-4}
Cell surface to volume ratio	SVr$_c$	μm^{-1}	0.5615 ± 0.123	0.6386 ± 0.134	0.011
Cell surface irregularity index[d]	SIi$_c$	$\mu m^{-1/2}$	242.8 ± 25.7	231.6 ± 26.6	0.66
Average distance of cell membrane voxels to centroid	$\langle R_c \rangle$	μm	8.788 ± 1.19	7.881 ± 1.015	1.0×10^{-3}
Standard deviation of R_c	ΔR_c	μm	2.297 ± 0.781	2.002 ± 0.657	0.076
Nuclear surface area	S_n	μm^2	830.5 ± 231	665.9 ± 344	0.015
Nuclear volume	V_n	μm^3	1022 ± 383	679.5 ± 379	1.7×10^{-4}
Nuclear surface to volume ratio	SVr$_n$	μm^{-1}	0.8451 ± 0.121	1.018 ± 0.175	2.5×10^{-6}
Nuclear surface irregularity index	SIi$_n$	$\mu m^{-1/2}$	254.2 ± 33.4	245.3 ± 62.4	0.43
Mitochondrial surface area	S_m	μm^2	546.9 ± 309	629.4 ± 347	0.27
Mitochondrial volume	V_m	μm^3	160.7 ± 108	148.8 ± 95.7	0.61
Mitochondrial surface to volume ratio	SVr$_m$	μm^{-1}	4.272 ± 1.93	4.993 ± 1.59	0.077
Mitochondrial surface irregularity index	SIi$_m$	$\mu m^{-1/2}$	441.2 ± 138	513.3 ± 136	0.023
Nucleus-to-cell centroid distance	D_{nc}	μm	0.1400 ± 0.0477	0.1552 ± 0.0475	0.12
Nucleus-to-cell volume ratio	Vr$_{nc}$	—	0.4933 ± 0.135	0.4351 ± 0.128	0.054
Mitochondrion-to-cell volume ratio	Vr$_{mc}$	—	0.0795 ± 0.053	0.1056 ± 0.0739	0.076

[a]n = number of imaged cells, p-values were obtained by a two-sample t-test method.
[b]$S = N_s \cdot s_0$ with N_s as the number of voxels on the membrane of the organelle and s_0 as the diagonal plane area of voxel.
[c]$V = N_v \cdot v_0$ with N_v as the number of voxels inside the organelle of interest and v_0 as voxel volume.
[d]$SIi = N_s \cdot a_0/(V)^{1/2}$ with a_0 as the side length (= 0.07 μm) of voxel.

Fig. 3 Scatter plots of 40 PC3 cells and 38 PCS cells with different combinations of 3-D parameters. See Table 1 for definitions of the symbols used as axis labels.

3.2 Measurement and Analysis of Diffraction Images

Cross-polarized diffraction image pairs have been acquired from cell suspension samples of about 2 mL in volume with the p-DIFC system shown in Fig. 1 in three measurements carried out in different weeks to confirm the repeatability of acquired data and subsequent classification. In each measurement, a small portion of the PC3 or PCS cell suspension sample was loaded into the core fluid syringe followed by alignment of the imaging unit to the same off-focus position of Δx and adjustment of the incident laser beam power P_0. About 1000 to 2000 cells were imaged from each cell sample for one of the three incident beam polarizations at the directions of ver, hor, and 45 deg.

Figure 4 shows examples of the normalized 8-bit image pairs acquired from single PC3 and PCS cells for three incident beam polarizations. It is clear from these normalized 8-bit images and the range of pixel values in the raw 12-bit images that both PC3 and PCS cells present stronger s-polarized light scatter for an incident beam that is also s-polarized (ver). Similarly, stronger p-polarized scatter can be observed for images acquired with an incident beam of p-polarization (hor). For 45 deg polarization of the incident beam, however, both PC3 and PCS cells yield much stronger scattered light of s-polarization, which can be understood by the fact that molecular dipoles induced by the incident laser beam within the illuminated cell are of higher efficiency to emit s-polarized than p-polarized light as scatter along the side directions. The dependence of scattered light intensity on polarization is shown in Fig. 5.

3.3 Classification of Two Prostate Cell Types and Comparison

With the 3-D morphology or p-DIFC image parameters, we performed cell classification study by the SVM algorithm to obtain the best SVM model with the highest value of accuracy A. For SVM classification with the 3-D parameters, the data were divided into a training data set of 30 cells/type and a test data set with the rest of cells. The data for GLCM parameters extracted from diffraction images were similarly divided, and Table 2 provides the number of cells in the training and test data sets acquired in three p-DIFC measurements.
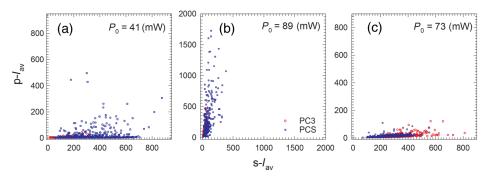
The search for the best SVM model to classify cells started by evaluation of the individual performance of 29 3-D parameters or 38 p-DIFC image parameters with different kernel functions based on the averaged values of A as A_{av} using a scheme of five-fold cross-validation with the training data set. The scheme divides the data into five equal parts with one part being used as a test data assembly and the remaining four parts as a training data assembly. The procedure was iterated five times with A calculated each time to obtain A_{av} followed by ranking of the single parameters in the order of decreasing A_{av}, which depends on the kernel functions used in SVM calculations. Different SVM models were then formed by a parameter vector $\mathbf{c}_i$ for cell i in the training data, with N_m selected parameters in the same sequence of ranking as components, and the corresponding kernel function. Each SVM model was trained in the feature space E with the training data and then applied to the test data to obtain A_{av} for evaluation.

Fig. 4 Examples of normalized 8-bit cross-polarized image pairs of two PC3 and two PCS cells acquired in measurement #1 for each incident beam polarization with white for pixel intensity 255 and black for 0. Each image is labeled with the cell type, polarization of the incident beam, polarization of the scattered light, and maximum, average, and minimum pixel intensities of the acquired 12-bit images.

For SVM classification with the 3-D parameters, the single parameters with highest $A_{av} = 71.7\%$ for the training data set are the cell's equivalent spherical radius (ER_c) using the kernel functions of the polynomial or RBF. The corresponding parameter for a sigmoid kernel function is the nucleus to cell volume ratio (Vr_{nc}) with $A_{av} = 68.3\%$ and cell volume (V_c) for linear with $A_{av} = 71.7\%$. SVM models obtained by including additional 3-D parameters ($N_m > 1$) according to

Fig. 5 Scatter plots of N_{tot} imaged cells with the average pixel intensity of the acquired s-polarized (s-I_{av}) versus that of the p-polarized diffraction image (p-I_{av}) acquired in measurement #1 with different incident beam polarizations: (a) vertical or s-polarized with $N_{tot} = 716$ for PC3 cells and $N_{tot} = 668$ for PCS cells; (b) horizontal or p-polarized with $N_{tot} = 681$ for PC3 cells and $N_{tot} = 623$ for PCS cells; (c) 45 deg with $N_{tot} = 770$ for PC3 cells and $N_{tot} = 378$ for PCS cells. The values of incident beam power P_0 are labeled.

Table 2 Experimental parameters and classification results with diffraction images.

Measurement group	Incident polarization	Cell type	N_{tot}[a]	N_{tra}[a]	N_{tes}[a]	A_{av} (%)		M and kernel function of best SVM model[b]
						Training	Test	
#1	Vertical	PC3	716	500	216	99.1	97.1	10
		PCS	668	500	168			and linear
	Horizontal	PC3	681	500	181	93.7	84.5	10
		PCS	623	500	123			and polynomial
	45 deg	PC3	770	300	470	80.7	64.8	10
		PCS	378	300	78			and polynomial
#2	Vertical	PC3	998	800	198	76.9	74.8	13
		PCS	1393	800	593			and polynomial
	Horizontal	PC3	890	400	490	100	100	6
		PCS	578	400	178			and linear
	45 deg	PC3	897	600	297	76.3	78.2	5
		PCS	758	600	158			and RBF
#3	Vertical	PC3	1130	800	330	93.5	93.0	9
		PCS	1006	800	206			and linear
	Horizontal	PC3	1104	800	304	99.5	99.5	14
		PCS	1337	800	537			and polynomial
	45 deg	PC3	1137	800	337	86.0	89.0	1
		PCS	1092	800	292			and linear
All data groups combined	Vertical	PC3	2844	2100	744	88.3	87.8	14
		PCS	3067	2100	967			and polynomial
	Horizontal	PC3	2675	1700	975	80.1	75.4	8
		PCS	2538	1700	838			and polynomial
	45 deg	PC3	2804	1700	1104	73.4	79.2	13
		PCS	2228	1700	528			and polynomial

[a]N_{tot} = number of diffraction image pairs of viable cells for extraction of 38 image parameters; N_{tra} = number of diffraction image pairs in the training data set; $N_{tes} = N_{tot} - N_{tra}$ = number of diffraction image pairs in the test data set.
[b]M = number of image parameters used in the best SVM model for classification.

their ranks have been found to produce slightly larger or smaller values of A_{av} in comparison to the single parameter model with the top-ranked one. The performance results of SVM models with N_m up to 10 are presented in Fig. 6 with

different kernel functions for both training and test data sets. It is obvious that 3-D morphology parameters extracted from confocal image stacks of cells with stained nucleus and mitochondria do not yield accurate markers for
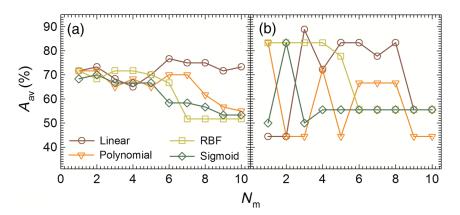

Fig. 6 Averaged accuracy A_{av} versus the maximum number of 3-D parameters N_m used for SVM classification with four different kernel functions for: (a) training data set and (b) test data set. The lines are for visual guide.
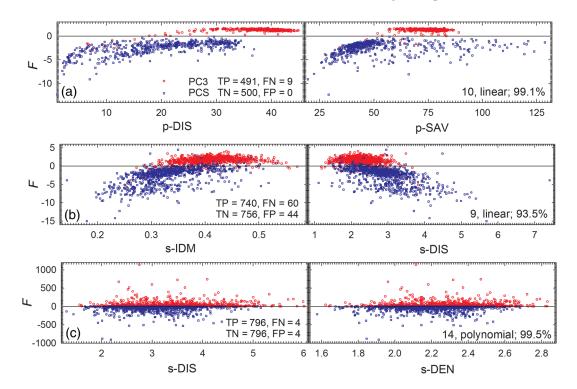
classification of the two prostate cell types, which is consistent with the data shown in Fig. 3.

To investigate classification with the p-DIFC image parameters, an SVM model was first optimized with the training data set. Table 2 includes the values of A_{av}, N_m, and kernel functions of the best SVM models established for the diffraction image pair data acquired with three different incident beam polarizations in three measurements. One can clearly see that the p-DIFC parameters provide a much improved performance in comparison to the 3-D parameter for classifying the two prostate cell types. However, A_{av} decreases significantly if we combine all data from the three measurements together as shown by the bottom section of Table 2. Similar decreases were observed by applying the best SVM model trained by the data of one measurement to the data of different measurements (not shown).

To demonstrate the effectiveness of the SVM algorithm with the p-DIFC image parameters, scatter plots of the training results are presented in Fig. 7 for three cases of cell classification with the best SVM model in each case on data acquired in the same measurement. The data show clearly that the SVM algorithm provides a powerful tool to improve cell classification with extracted image parameters by mapping them from the parameter space Σ_p into the feature space E using a kernel function. In the case of Fig. 5(a), the top two ranked single GLCM parameters of dissimilarity and sum average[22] extracted from p-polarized images yield, respectively, classification accuracies A_{av} of 91.0% and 87.7% for the training data. These values of A_{av} with single parameter values are significantly smaller than the

accuracy of 99.1% that can be achieved with the best SVM model of $N_m = 10$ parameters and the linear kernel function. A similar improvement in classification can be observed in the other two cases: the values of A_{av} were found to increase, respectively, from 77.9% for s-IDM and 74.4% for s-DIS alone to 93.5% with $N_m = 9$ and a linear kernel function in the case of Fig. 5(b) and from 61.3% for s-DIS and 61.2% for s-DEN alone to 91.3% with $M = 10$ and a polynomial kernel function in the case of Fig. 5(c). The GLCM parameters extracted from the normalized diffraction image pairs are available online for readers to investigate other classification methods.[26]

4 Discussion

Accurate classification of biological cells of the same tissue of origin is fundamentally challenging and also of practical interest in clinical applications, such as detection of CTCs.[27] In this report, we focus on the feasibility of diffraction imaging for accurate classification of the prostate epithelial cells of PC3 and PCS by comparison to the conventional morphology measurement through confocal imaging. Despite the statistically significant differences in the cell and nuclear volumes and other parameters as indicated by the p-values smaller than 0.05 in Table 1, the scatter plots of the imaged cells by these parameters in Fig. 3 and the SVM classification results in Fig. 6 show clearly that the 3-D parameters alone cannot yield accurate markers for classification, which stands in stark contrast to the use of arrangement patterns of the carcinoma cells in a

Fig. 7 Scatter plots of training data with values of decision function F versus the top two ranked p-DIFC image parameters used by the best SVM model established for: (a) data acquired in measurement #1 with vertical incident beam polarization and $N_{tra} = 500$ for each of the two cell types, p-DIS: dissimilarity of p-polarized images, p-SAV: sum average of p-polarized images; (b) data acquired in measurement #3 with vertical incident beam polarization and with $N_{tra} = 800$, s-IDM: inverse difference moment of s-polarized images; (c) same as (b) except with horizontal polarization, s-DEN: difference entropy of s-polarized images. The cells with $F > 0$ (above the line of $F = 0$) are classified by the SVM model as PC3 cells and those with $F < 0$ (below the line) as PCS cells. The values of TP, FN, TN, FP, N_m, kernel function, and A_{av} of the best SVM model are labeled.

stained tissue section as a part of the evidence for prostate cancer staging.[28]

With the p-DIFC method, we have shown that detection of the diffraction patterns of the coherent light scattered by single cells through polarization diffraction imaging can provide an accurate and effective approach to classify the two cell types for data acquired in the same measurement. By imaging the coherent side scatter, the diffraction image parameters $\mathbf{c}_i$ obtained with an optimized SVM model can serve as the morphology-related "fingerprints" of the cell i impressed by the coherent electromagnetic wavefields of the incident laser beam. Even though the fingerprints as a result of diffraction have been known to correlate strongly with cell morphology, they are formed through the complex interaction of the incident wavefields with the molecules inside the illuminated cell. Because of the unknown intracellular distribution refractive index, the detailed relations remain to be investigated between $\mathbf{c}_i$ extracted from a pair of 2-D cross-polarized diffraction images and the cell's 3-D morphology. Still the results presented here provide strong evidence that the p-DIFC method has the capacity to establish an empirical approach for accurate classification of normal and cancerous human prostate epithelial cells. With the powerful data mining tools like the SVM algorithm, the diffraction image data can be used to construct a high-dimension feature space E defined by the training data and a kernel function for significantly improved classification as shown by the results in Fig. 7. From the last part of Table 2, it is also clear that the diffraction images or their texture parameters are sensitive to the positioning of the flowing cell relative to the focused incident beam and the imaging unit on the scales of 10 μm. Since these positionings could not be accurately controlled with the current experimental system, the SVM model has to be retrained between measurements to achieve accurate classification. System improvement is underway to use two laser beams and forward scatter signals and improve the positioning of the cells carried by the core fluid.

Careful examination of the average pixel intensity data in Fig. 5 demonstrates that the incident beam polarization markedly affects the detection efficiency of side scatter. The same sensitivity to the incident polarization can also be observed in the values of A_{av} presented in Table 2. These data indicate clearly that the p-DIFC image parameters could provide "fingerprint" makers carrying rich information on intracellular biomolecules in terms of their ability to polarize in the wavefields of the incident beam. It is interesting to note further that among the three polarization directions, cell classification with data acquired at 45 deg tends to produce smaller values of A_{av} for each of the three measurements. Similar results have been observed in our previous classification study of the Jurkat T-cell line and Ramos B-cell line derived from cancerous white blood cells.[19] The less ability of the p-DIFC method with a 45-deg polarized incident beam to separate different cell types of highly similar morphology could be understood by the following considerations. For the incident beam propagating along the z-axis with polarization at 45 deg, the intracellular molecules can have induced dipoles to oscillate along both the x-axis and the y-axis. The equal probability of induced molecular dipoles reduces the selectivity of the p-DIFC method to contrast the differences among cells with different molecular responses to the incident wavefields. These considerations are corroborated by a visual inspection of the cross-polarized diffraction images, with limited but randomly selected examples presented in Fig. 4, in which the two images in each pair acquired with 45 deg polarization exhibit diffraction patterns of higher similarity than those acquired with vertical or horizontal polarizations.

5 Conclusion

A classification study of two types of prostate epithelial cells has been performed, and it has been shown that the cancerous cells can be accurately distinguished from the normal cells with the measured cross-polarized diffraction image pair data using the data acquired in the same measurement. The classification ability of the label-free p-DIFC method suggests strongly that diffraction imaging senses the molecular differences among the two different cell types in addition to the morphologic differences, which have been quantified by confocal imaging and 3-D reconstruction. The employment of the SVM classification algorithm allows significantly improved classification in comparison with the direct approach in the parameter space defined by the GLCM parameters. It should be pointed out that the p-DIFC method remains to be further enhanced in terms of the acquisition of high-contrast diffraction images at a faster rate, accurate positioning of the flowing cells, and development of superior algorithms for characterization of image textures with less sensitivity to image noises.

References

1. A. Heidenreich et al., "EAU guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent-update 2013," *Euro. Urol.* **65**(1), 124–137 (2014).
2. J. S. de Bono et al., "Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer," *Clin. Cancer Res.* **14**(19), 6302–6309 (2008).
3. A. Goldkorn et al., "Circulating tumor cell counts are prognostic of overall survival in SWOG S0421: a phase III trial of docetaxel with or without atrasentan for metastatic castration-resistant prostate cancer," *J. Clin. Oncol.* **32**(11), 1136–1142 (2014).
4. H. J. Yoon et al., "Sensitive capture of circulating tumour cells by functionalized graphene oxide nanosheets," *Nat. Nanotechnol.* **8**(10), 735–741 (2013).
5. A. J. Armstrong et al., "Circulating tumor cells from patients with advanced prostate and breast cancer display both epithelial and mesenchymal markers," *Mol. Cancer Res.* **9**(8), 997–1007 (2011).
6. T. M. Gorges et al., "Circulating tumour cells escape from EpCAM-based detection due to epithelial-to-mesenchymal transition," *BMC Cancer* **12**, 178 (2012).
7. A. N. Shvalov et al., "Particle classification from light scattering with the scanning flow cytometer," *Cytometry* **37**(3), 215–220 (1999).
8. J. Neukammer et al., "Angular distribution of light scattered by single biological cells and oriented particle agglomerates," *Appl. Opt.* **42**(31), 6388–6397 (2003).
9. J. Q. Lu, P. Yang, and X. H. Hu, "Simulations of light scattering from a biconcave red blood cell using the FDTD method," *J. Biomed. Opt.* **10**(2), 024022 (2005).
10. R. S. Brock et al., "Effect of detailed cell structure on light scattering distribution: FDTD study of a B-cell with 3D structure constructed from confocal images," *J. Quant. Spectrosc. Radiat. Transfer* **102**(1), 25–36 (2006).

11. K. J. Chalut et al., "Label-free, high-throughput measurements of dynamic changes in cell nuclei using angle-resolved low coherence interferometry," *Biophys. J.* **94**(12), 4948–4956 (2008).
12. X. Su et al., "Microscope-based label-free microfluidic cytometry," *Opt. Express* **19**(1), 387–398 (2011).
13. K. M. Jacobs, J. Q. Lu, and X. H. Hu, "Development of a diffraction imaging flow cytometer," *Opt. Lett.* **34**(19), 2985–2987 (2009).
14. K. M. Jacobs et al., "Diffraction imaging of spheres and melanoma cells with a microscope objective," *J. Biophotonics* **2**(8–9), 521–527 (2009).
15. Y. Sa et al., "Study of low speed flow cytometry for diffraction imaging with different chamber and nozzle designs," *Cytometry A* **83**(11), 1027–1033 (2013).
16. R. Pan et al., "Analysis of diffraction imaging in non-conjugate configurations," *Opt. Express* **22**(25), 31568–31574 (2014).
17. K. Dong et al., "Label-free classification of cultured cells through diffraction imaging," *Biomed. Opt. Express* **2**(6), 1717–1726 (2011).
18. J. Zhang et al., "Analysis of cellular objects through diffraction images acquired by flow cytometry," *Opt. Express* **21**(21), 24819–24828 (2013).
19. Y. Feng et al., "Polarization imaging and classification of Jurkat T and Ramos B cells using a flow cytometer," *Cytometry A* **85**(9), 817–826 (2014).
20. Y. Zhang et al., "Comparative study of 3D morphology and functions on genetically engineered mouse melanoma cells," *Integr. Biol.* **4**(11), 1428–1436 (2012).
21. "3D morphology and GLCM parameter definition tables," http://bmlaser.physics.ecu.edu/literature/3D_GLCM_Para_Def_Tables.pdf (23 September 2015).
22. R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE* **67**(5), 786–804 (1979).
23. V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York (1995).
24. C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Disc.* **2**, 121–167 (1998).
25. C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Syst. Technol.* **2**(3), 1–27 (2011).
26. "PC3_PCS_GLCM and intensity parameter files," http://bmlaser.physics.ecu.edu/literature/PC3_PCS_GLCM_Parameters.rar (23 September 2015).
27. P. Foggia et al., "Benchmarking HEp-2 cells classification methods," *IEEE Trans. Med. Imaging* **32**(10), 1878–1889 (2013).
28. P. A. Humphrey, "Gleason grading and prognostic factors in carcinoma of the prostate," *Mod. Pathol.* **17**(3), 292–306 (2004).

Wenhuan Jiang received his MS degree in medical physics and PhD degree in biomedical physics in 2012 and 2015 from East Carolina University. His current research interests include morphology and light scattering properties of biological cells and radiation treatment of cancers.

Jun Qing Lu received her BS/MS degrees in physics from Nankai University and her PhD degree from the University of California, Irvine. She is currently an associate professor at East Carolina University and mainly interested in numerical studies of light scattering in turbid media and with biological cells.

Li V. Yang is an associate professor at East Carolina University. He received his PhD degree in molecular biology and genetics from Wayne State University in 2002. His research interests include tumor biology, inflammation, and vascular biology.

Yu Sa received his BS and PhD degrees from Tianjin University and is a lecturer in the Department of Biomedical Engineering at Tianjin University. His research interests include instrument development, diffraction imaging and computational fluid dynamics modeling studies.

Yuanming Feng is a professor of biomedical engineering at Tianjin University. His research interests include diffraction imaging flow cytometry and measurement technique of radiation-induced apoptosis.

Junhua Ding is an associate professor of computer science with East Carolina University. He received his BS, MS and PhD, all in computer science, in 1994, 1997, and 2004, respectively. His research interests are in software engineering and data engineering, and published over 60 peer-reviewed papers in these fields.

Xin-Hua Hu received his BS and MS degrees from Nankai University, Tianjin, China, in 1982 and 1985, MS degree in physics from Indiana University in 1986 and PhD degree in physics in 1991 from University of California at Irvine. He joined the physics faculty in 1995 and is a professor at East Carolina University. His main research interests relate to the investigations of light scattering and their applications in probing tissues and cells.