Journal of Quantitative Spectroscopy & Radiative Transfer 224 (2019) 453–459

Contents lists available at ScienceDirect

Journal of Quantitative Spectroscopy & Radiative Transfer

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

Rapid classification of micron-sized particles of sphere, cylinders and ellipsoids by diffraction image parameters combined with scattered light intensity

Wenjin Wang[a,b], Yuhua Wen[a,b], Jun Q. Lu[a,c], Lin Zhao[a,d], Safaa A. Al-Qaysi[c,e], Xin-Hua Hu[a,c,*]

[a] Institute for Advanced Optics, Hunan Institute of Science and Technology, Yueyang, Hunan 414006, China
[b] School of Physics & Electronics, Hunan Institute of Science and Technology, Yueyang, Hunan 414006, China
[c] Department of Physics, East Carolina University, Greenville, NC 27858, USA
[d] School of Information Science, Hunan Institute of Science and Technology, Yueyang, Hunan 414006, China
[e] College of Pharmacy, Al-Mustansiriyah University, Baghdad, Iraq

ARTICLE INFO

Article history:
Received 10 October 2018
Revised 6 December 2018
Accepted 6 December 2018
Available online 7 December 2018

Keywords:
Single light scattering
Diffraction imaging
Image analysis
Nonspherical particle analysis

ABSTRACT

Spatial distributions of light scattered by single particles correlate closely with their morphologies in terms of refractive index (RI) distribution. Diffraction imaging of scattered light under coherent excitation presents a unique approach to acquire and extract feature parameters for particle classification. A validated method has been applied in this study to accurately simulate diffraction imaging of light scattered by homogeneous particles and obtain calculated diffraction image (DI) data. The feature parameters of DI data have been extracted by the gray-level co-occurrence matrix (GLCM) algorithm. We have developed an unsupervised machine learning algorithm based on Gaussian mixture model (GMM) to classify 1965 particles made of single and double spheres, cylinders and ellipsoids with varied RI values in parameter space. It has been shown that selected GLCM parameters combined with integrated forward scatter intensity can provide effective markers for accurate and morphology based classification. For 1791 particles of the same RI, the mean accuracy values of classifying particles into 3 particle types range from 82.6% to 97.2%. These results demonstrate the strong potential of diffraction imaging method for rapid analysis and classification of nonspherical and homogeneous particles by the GMM classifiers that is very challenging in comparison to distinguishing biological cell types.

1. Introduction

Rapid analysis and recognition of micron-sized particles have wide applications ranging from atmospheric and oceanic to biological sciences [1–3]. While fast methods such as flow cytometry and machine-learning based microscopy profiling are available, these approaches often require fluorescence staining and yield limited morphology information like 2D organelle sizes only [4]. In comparison, imaging of light scattered elastically by single particles under coherent excitation allows 3D characterization by taking advantage of strong correlations between image features and particle morphology. We have previously developed a method of diffraction imaging flow cytometry (DIFC) that can rapidly acquire diffraction images (DIs) from single particles including biological cells with no need for staining [5–7]. Classifying single spheres of different diameters and cells of different phenotypes have been empirically demonstrated using feature parameters of the measured DIs [8–12]. Still, the correlations between the image features of DI data and imaged particle's morphology remain to be quantified and explored for classification purpose.

Among various algorithms used for extracting features of DI data, we found that the gray-level co-occurrence matrix (GLCM) algorithm [13] is the most effective one in terms of robustness and computational cost. Combined with machine learning, GLCM presents a powerful tool for analysis of large populations of particles in real time by the DIFC method. GLCM applies second-order statistics to characterize textures of an input image by matrix elements expressed as the relative frequencies of paired pixel values along one of four array directions. One can extract up to 15 paraments for texture characterization from the matrix [13–15]. Unlike Fourier or Gabor transforms, GLCM parameters are very abstract for interpretation according to visual perception such as spatial

* Corresponding author at: Department of Physics, East Carolina University, Greenville, NC 27858, USA.
E-mail address: hux@ecu.edu (X.-H. Hu).

https://doi.org/10.1016/j.jqsrt.2018.12.010
0022-4073/

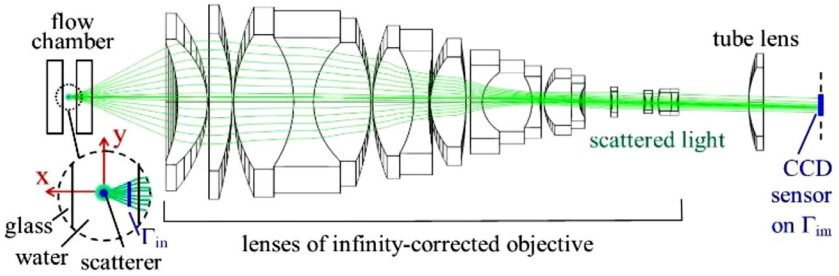

Fig. 1. Schematic of DIFC imaging configuration used for simulations with the green lines representing "rays" of light scattered by a particle at the origin by a laser beam incident along the z-axis pointing into paper. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

period and rotational symmetry among image pixels. It is therefore of significant interest to quantify the correlations between particle morphology and the DI textures characterized by different sets of GLCM parameters. Furthermore, development of rapid and label-free methods based on these correlations become possible to group particles into different morphology types by combining GLCM characterization of DI data with machine learning based classification. In this report, we present the results of a study on classification of 3 particle types by GLCM parameters extracted from DI data. A previously developed model of DI simulation has been validated and applied to calculate unpolarized DI data from given morphology of nonspherical particles made of two identical objects of spheres, cylinders and ellipsoids. For comparison, each particle type also includes single spheres, cylinders and ellipsoids. A total of 1965 particles have been generated and investigated with varied sizes, orientations and refractive index (RI) values. Different sets of GLCM and scattering intensity parameters were selected to evaluate their performance for type recognition by Gaussian mixture model (GMM) based classifiers. We found that selected GLCM parameters combined with integrated forward scattering intensity can serve as effective markers for characterization of the diffraction patterns and rapid classification of homogeneous particle.

2. Theoretical methods

In an experimental DIFC system, single particles are rapidly moved through an incident laser beam with scattered light collected by an imaging unit that includes a microscope objective for acquisition of DI data [5–7,10,16]. With a previously developed method, we can perform simulations of the diffraction imaging process by a configuration based on the experimental system with details given elsewhere [17]. Briefly, a core fluid of particle suspension is hydrodynamically focused by a sheath fluid in a square channel of the glass flow chamber to force the moving particles in single file. An incident laser beam of $\lambda_0 = 532$ nm in wavelength in vacuum propagates along the *z*-axis and is focused on the core fluid in a spot of about 30 μm in diameter. As a particle passes through the beam focus, it emits coherent light as scatter passing through the host medium of water and chamber glass before collection by an imaging unit aligned along the *x*-axis. Fig. 1 shows a schematic of the imaging configuration. The imaging unit in our experimental system consists of an infinity-corrected 50x objective of 0.55 in NA (378-805-3, Mitutoyo), a polarizing beam splitter for separating scattering light into s- and p-polarized components, two tube lenses and two cameras. The unit, with camera sensors fixed to the focal planes of tube lenses, is translated away from the focused position conjugate to the core fluid for DI acquisition. We have shown that the off-focus translation of $\Delta x > 0$ to non-conjugate positions toward the flow chamber make it possible to adjusts angular cone of light collection with increased contrast for acquired DI data [6,17]. For DI calculations performed in this study, we set $\Delta x = 150$ μm in simulation which is the same as our experimental setting which affects image contrast and GLCM parameters of calculated DIs [9,17].

Simulations with the DIFC imaging configuration illustrated in Fig. 1 were performed in two steps of light scattering modeling and simulation of diffraction imaging process. First, we consider a particle with its RI given by $n_p(\mathbf{r}, \omega)$ in a rectangular volume V_p containing the particle only and being immersed in a host medium of volume V_h of constant RI n_h, where ω is the angular frequency of the incident light. The electric field $\mathbf{E}_s$ of scattered light can be obtained from the following volume integral equation [18]

$$\mathbf{E}_s(\mathbf{r}, \omega; \mathbf{k}_s) = ({k_h}^2 \overleftrightarrow{I} + \nabla\nabla) \int_{V_p} \frac{e^{i\mathbf{k}_s\cdot(\mathbf{r}-\mathbf{r}')}}{4\pi|\mathbf{r}-\mathbf{r}'|} (m^2(\mathbf{r}', \omega) - 1)\mathbf{E}(\mathbf{r}', \omega; \mathbf{k}_s) d\mathbf{r}' \quad \forall \ \mathbf{r} \in V_h \cup V_p. \tag{1}$$

where $\mathbf{E} = \mathbf{E}_i + \mathbf{E}_s$ in the volume integral term is the total field with $\mathbf{E}_i$ as the field of incident light with $\mathbf{k}_h = (2\pi/\lambda_h)\mathbf{z}$ as the wavevector in the host medium and $\mathbf{z}$ as unit vector of z-axis, $\overleftrightarrow{I}$ the unit dyad, $\mathbf{k_s}$ the wave vector of scattered wavefields with $|\mathbf{k}_s| = |\mathbf{k}_h|$ for elastic scattering and (θ_s, φ_s) for its polar and azimuthal angles, $m(\mathbf{r}, \omega) = n_p(\mathbf{r}, \omega)/n_h$ the relative RI of the particle. It is clear from Eq. (1) that the scattered light intensity in a far-field region, $I_s(\mathbf{k}_s) \propto |\mathbf{E}_s(\mathbf{k}_s)|^2$, can be regarded as a mapping function of $m(\mathbf{r}')$ for an incident light of given $\omega = 2\pi c/(\lambda_h n_h) = 2\pi c/\lambda_0$ with c as the light speed in vacuum.

To solve for the scattered field $\mathbf{E}_s$, we employed an ADDA code that has been developed by Yurkin and Hoekstra for implementation of the discrete-dipole-approximation (DDA) algorithm and is available in public domain [19,20]. The code discretizes V_p of a particle scatterer into cubic voxels called as dipoles of size d and defines the ratio of λ_h/d as dpl (dipole-per-wavelength) to quantify the discretization accuracy, which should be set larger than 10 for values of m close to 1. By importing the relative RI values of $m(\mathbf{r})$ at λ_h for each voxel, Eq. (1) can be solved in voxels of V_p for **E** field as a linear equation group given $\mathbf{E}_0$ of an incident light beam. The Eq. (1) is solved again for $\mathbf{E}_s$ in a far-field region of V_h defined by $|\mathbf{r}| >> |\mathbf{r}'|$ and $|\mathbf{r}| >> \lambda_h$. For this study, we consider particles made of single or double homogeneous spheres, cylinders and ellipsoids with the n_p values set either to $n_{ph} = 1.558$ for polystyrene particles [21] or $n_{pl} = 1.400$ for cell-like ones. The host medium is treated as water with $n_h = 1.334$ at $\lambda_0 = 532$ nm. The imaginary RI values for both particle and host medium were set to 1.5×10^{-5}. The ADDA simulations were performed in parallel mode on a single workstation equipped with two Xeon E5-2650 CPU of 12 cores each in the high-performance computing cluster at the Institute for Advanced Optics. The incident light beam was set to a plane-wave function and dpl values to 15 or larger. For particles of single-object, size and orientation were varied to obtain different DIs for cylinders and ellipsoids. For particles made of two identical objects, the center connection vector **C** was varied by |**C**| and its polar and azimuthal angles, (θ_c, φ_c). After $\mathbf{E}_s$ of scattered light

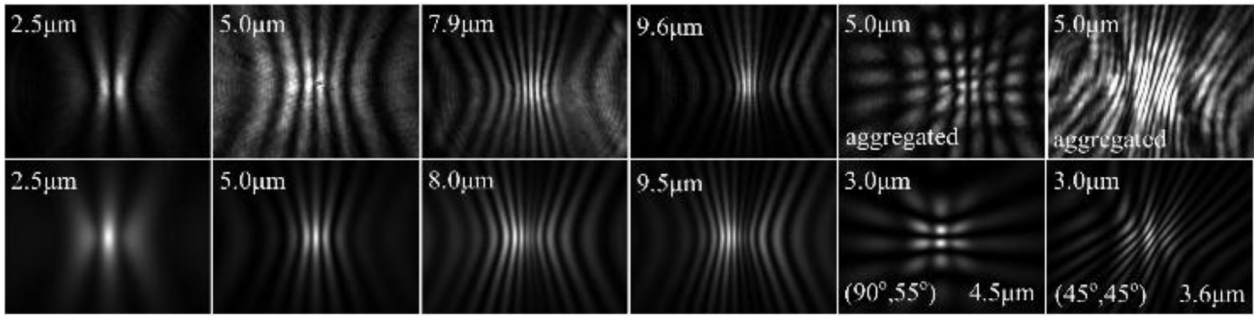

Fig. 2. Measured (top row) and calculated (bottom row) DIs with $\Delta x = 150\,\mu m$ and $\lambda = 532\,nm$. Values of sphere diameter D_s are labeled at the top of each image and **C** vectors at the bottom of calculated DIs for double spheres.

is obtained in a far-field region by executing the ADDA code, the output data is saved in the form of angle-resolved Mueller matrices of $S_{ij}(\theta_s, \varphi_s)$ with i, j = 1, 2, 3 4 and (θ_s, φ_s) representing the direction of $\mathbf{k}_s$ [18]. The simulation times of Mueller matrix data by different particles ranged from 5 minutes to 20 minutes.

In the second step of DI simulation, an in-house developed code built on MATLAB (MathWorks, 2013a) reads selected matrix elements of S_{ij} and project them on a virtual input plane Γ_{in} shown in Fig. 1. This produces an input image $I_{in}(x_0, y, z)$ with $\theta_s \in [60°, 120°]$ and $\varphi_s \in [150°, 210°]$ and $-x_0$ is the intercept of Γ_{in} to the x-axis. Different combinations of S_{ij} for producing I_{in} can be derived according to the polarizations of incident and scattered light. Without loss of generality, the calculated DIs is obtained from $S_{11}(\theta_s, \varphi_s)$ as follows

$$I_{in}(-x_0, y, z) = \frac{|\cos\phi_s \sin\theta_s|}{x_0^2(1+\tan^2\phi_s + \frac{1+\tan^2\phi_s}{\tan^2\theta_s})} S_{11}(\theta_s, \phi_s)\cdot \quad (2)$$

Extension to cross-polarized DI pairs is straightforward [22] and conclusions remain the same on GLCM characterization of texture features and related classification. With the input image I_{in}, the MATLAB code then calls a commercial software for optical design (Zemax, 2009) to trace the scattered light "rays" from Γ_{in} located at $x = -150\,\mu m$ inside the sheath fluid in flow chamber to the image plane Γ_{im} as shown in Fig. 1. The ray tracing is carried out for each pixel of I_{in} with initial directions given by (θ_s, φ_s) through the sheath fluid as host medium, flow chamber glass, air and imaging unit to obtain an image I_{out} of 640×480 pixels at Γ_{im} as the calculated DI. Comparison of the calculated DI data against the measured ones presented in our previous study [17] and here in Fig. 2 shows clearly that the coherent nature of the scattered light in DI is accurately modeled by the simulation method described above despite the fact that ray-tracing is based on geometric model. We further note that the imaging unit shown in Fig. 1 is a simplified configuration for DI simulation without the polarizing beam splitter and the second arm of tube lens and camera.

3. Results

3.1. Comparison to the measured data in the sphere cases

To demonstrate the accuracy of DI simulations, we first compare in Fig. 2 the calculated DIs with $n_{ph} = 1.558$ and $n_h = 1.334$ to the measured data acquired with different polystyrene sphere suspensions. The calculated DI data of sphere particles were selected with diameter D_s close to the nominal D_s values of suspensions supplied by vendors with coefficient of variation ≤ 20% or 7.5% [9]. The two measured DIs to the right of top row in Fig. 2 are obviously due to aggregated spheres. By varying (θ_c, φ_c) of the center connection vector **C** of two touching spheres of $D_s = 3.0\,\mu m$, we could obtain calculated DIs with similar patterns. The high degree of similarity in patterns between the two DI sets thus provide additional validation data in support of the simulation method [17] given the errors in Δx and particle positioning relative to the incident beam focus for the measured data.

Table 1
Particle morphology and RI for calculated DIs.

Type[1]		Size (μm)[1]	n_p[2]	N_{DI}[3]
Spheres	single	$D_s \in [1, 9]$	1.558 (33) 1.400 (17)	50
	double	$D_s = 3.0$ $\|\mathbf{C}\| \in [0.1, 6]$	1.558 (532) 1.400 (32)	564 / 614
Cylinders	single	$D_c \in [1, 9]$ $h \in [1, 9]$	1.558 (130) 1.400 (27)	157
	double	$D_c = h = 3$ $\|\mathbf{C}\| \in [0.1, 6]$	1.558 (510) 1.400 (32)	542 / 699
Ellipsoids	single	$D_1 \in [1,9]$ $D_2/D_1 \in [0.2, 1]$ $D_3/D_1 \in [0.2, 1]$	1.558 (70) 1.400 (34)	104
	double	$D_1 = 3$ $D_2/D_1 = 0.5$ $D_3/D_1 = 1$ $\|\mathbf{C}\| \in [0.1, 6]$	1.55: (516) 1.400 (32)	548 / 652

[1] Particle types, D_s = sphere diameter, D_c = cylinder diameter, h = height, D_1, D_2, D_3 = principal diameters of ellipsoids.
[2] Numbers of particle or unpolarized DIs are in parenthesis.
[3] N_{DI}: left = number of unpolarized DIs or particles; right = total number for each type.

3.2. Diffraction imaging simulation of single- and double-object particles

For this study, we generated 1965 homogeneous single-object and double-object particles immersed in water to obtain calculated DIs. Among them, 1791 particles have high RI values of $n_p(\mathbf{r}) = n_{ph}$ representing different particle morphologies or orientations with **r** as intraparticle position vector. The rest consists of selected "copy" particles with low RI values of $n_p(\mathbf{r}) = n_{pl}$ to examine the effect of RI values. Table 1 summarizes 3 particle morphology types and related RI values with the ranges of major parameters that were used to obtain calculated DI data. The magnitude and orientation of **C** for double-object particles were varied with the center of **C** fixed to the coordinate origin defined in Fig. 1. The diffraction patterns of these mostly nonspherical particles can be visually divided into two categories of "line style" and "grid style", which have high degrees of symmetry due to the homogeneous RI inside each particle. The former includes spherical, ellipsoidal and double-object particles of **C** along certain special directions such as $\theta_c = 0$ for being parallel to the incident light beam direction along the z-axis or $\theta_c = 45°$ and $\varphi_c = 0°$ for equally between z-axis and x-axis as the light collection direction by the imaging unit. The particle morphologies responsible for "grid-style" patterns are those made of double-object particles with large |**C**| and unsymmetrical **C** directions.

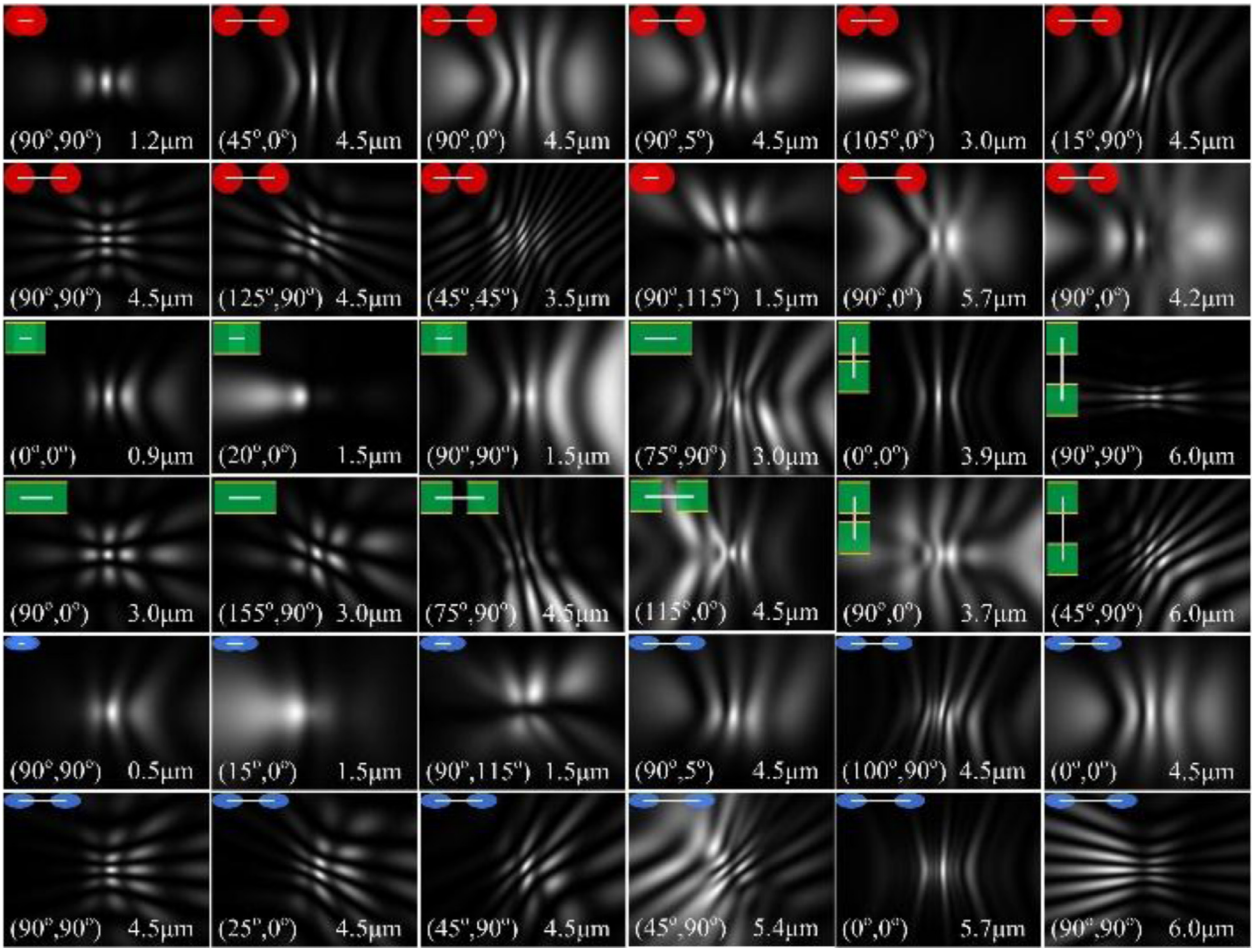

Fig. 3. Examples of normalized DIs calculated with double-object particles of identical spheres (top two rows), cylinders (middle two rows) and ellipsoids (bottom two rows) of $n_{ph} = 1.558$, $\Delta x = 150\,\mu m$ and $\lambda_0 = 532\,nm$. Each image is marked on top by a configuration drawing with white line representing C and on bottom by (θ_c, φ_c) and |C|. For DIs of double-cylinders, the yellow bars indicate cylinders' end surfaces. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Selected examples of calculated DI data by double-object particles of identical spheres, cylinders and ellipsoids are presented in Fig. 3 to illustrate their diffraction patterns with the high RI values of $n_{ph} = 1.558$ and $n_h = 1.334$. By examining these DI data, one can observe that the diffraction patterns of double-sphere particles exhibit relative high degrees of similarity to those of double-ellipsoid, while the patterns of both particle types show significant differences from those of double-cylinder particles. It is obvious that efficient methods of image analysis and machine learning algorithms for particle classification need to be developed to quantitatively investigate the variations in diffraction patterns and the correlations between these patterns and morphological parameters of the particles. For this purpose, we chose clustering algorithms of machine learning to classify the diffraction patterns in the GLCM parameter space because these classifiers require no training or prior knowledge on DI data type and provide insight on the ability of GLCM parameters to characterize the image data.

The 16-bit DI data calculated with particle morphologies shown in Table 1 were first normalized individually into 8-bit images for efficient calculations of GLCMs in 4 directions of neighboring pixels (0°, 45°, 90° and 135°). Each set of 4 GLCMs were averaged followed by extraction of 15 parameters to characterize image textures for particle classification with definitions and symbols given in Supplementary Material (see S1). In addition to the GLCM parameters of normalized DIs, we have also obtained two scattering intensity parameters of integrated forward scatter (IFS) and integrated side scatter (ISS) from the angle-resolved Mueller matrices output by ADDA. The parameter IFS was calculated by integrating $S_{11}(\theta_s, \varphi_s)$ over the forward scattering directions of $\theta_s \in [5°,15°]$ and $\varphi_s \in [175°,185°]$ while ISS over the side scattering directions of $\theta_s \in [85°,95°]$ and $\varphi_s \in [175°,185°]$. Different combinations of the 15 GLCM parameters with IFS and ISS, were selected to form different classifiers for evaluation of their performance in recognizing particle types in the multi-dimensional parameter space.

3.3. Classification of particles by GMM based clustering algorithm

Several clustering algorithms have been investigated for classification of particle types that include k-means and related variants, hierarchical, GMM and manifold learning. The best classifiers with high stability and computational efficiency for this study were obtained by combining the hierarchical with GMM clustering algorithms. In this approach, the hierarchical clustering was applied first to a chosen set of GLCM and intensity parameters extracted from the calculated DIs, which divided the corresponding particles into k clusters with $k = 3$. The results were then imported into GMM as the initial fitting parameters. The combination of hierarchical and GMM clustering algorithms is very effective to remove the sensitivity of GMM output to the initial parameters that are usually set randomly. As the first step of clustering, the hierarchical algorithm starts by assuming each DI, or associated particle, being a cluster of its own. It then iterates by linking two clusters in

Table 2
Confusion matrices of GLCM classifiers for all particles.

Cluster[1]	C_1	C_2	C_3	normalization scheme; (P_i); $\mu_A \pm \sigma_A$[2]
Sphere	99	**352**	163	uniform norm. (all 15 parameters) 48.0% ± 6.8%
Cylinder	**290**	126	283	
Ellipsoid	154	204	**294**	
Sphere	30	**491**	93	type norm. (all 15 parameters) 55.7% ± 17.2%
Cylinder	191	200	**308**	
Ellipsoid	**280**	232	140	
Sphere	135	**292**	187	uniform norm. (COR,SAV,SEN,DEN,CLS) 45.1% ± 1.9%
Cylinder	**300**	132	267	
Ellipsoid	125	235	**292**	
Sphere	44	**491**	79	type norm. (VAR,SVA,ENT,DVA,MAP) 65.8% ± 11.4%
Cylinder	**458**	65	176	
Ellipsoid	1	312	**339**	

[1] Rows represent ground truth. Bold numbers represent assigned particle types for three clusters derived by the hierarchical + GMM classifier.
[2] (P_i) represents the set of normalized GLCM parameters used by the classifier, μ_A and σ_A are the mean value and standard deviation of classification accuracies for 3 particle types.

each step according to the Ward's method [23]. In this method the distance between any two clusters, A and B, is measured by the increase of sum of squares for the linked cluster of A∪B from those of A and B given by

$$d_{AB}{}^2 = \sum_{i\in A\cup B} |\mathbf{p}_i - \mathbf{c}_{A\cup B}|^2 - \sum_{i\in A} |\mathbf{p}_i - \mathbf{c}_A|^2 - \sum_{i\in B} |\mathbf{p}_i - \mathbf{c}_B|^2 \quad (3)$$

where $\mathbf{p}_i$ and $\mathbf{c}_A$ (or $\mathbf{c}_B$ or $\mathbf{c}_{A\cup B}$) is respectively the position vector of *i*th particle and cluster A's (or B's or A∪B's) centroid in the GLCM parameter space. The cluster linking in each step proceeds by calculating $d_{AB}{}^2$ for all paired clusters and executing on the two of smallest $d_{AB}{}^2$. The process continues until the number of clusters is reduced to k, which is 3 for our study.

For subsequent clustering, the GMM algorithm imports the hierarchical output of grouping DI data or particles into 3 clusters obtain a Gaussian probability density function (pdf) for each cluster in the given parameter space. Classification is optimized by maximizing iteratively a likelihood function L defined as the logarithmic sum of pdf's over all particles assigned to different clusters. Each iteration varies the cluster assignment and pdf parameters to increase L until its value stabilizes [24]. Afterwards, each cluster of C_1, C_2 and C_3 was assigned to one particle types defined in Table 1, sphere, cylinder and ellipsoid, based on the dominant type in a confusion matrix. An accuracy A is defined to measure the performance of a classifier by the number ratio of correctly identified particle type to the total number of particles. If two clusters are dominated by the same particle type, the assignment was made so that the mean value of A, defined as μ_A, for 3 types was maximized.

Particle classification with GLCM parameters has been investigated by different classifiers with parameter number increased from 3 to 15 to explore the possibility of unsupervised classification according to the diffraction patterns only. The values of each parameter were normalized to the range of [0, 1] with two schemes before clustering analysis. The uniform scheme (uniform norm.) refers to parameter normalization by the same minimum and maximum values of all DI data while the type scheme (type norm.) does that with the minimum and maximum values for each particle type. In Table 2 we present the confusion matrices of classifiers with all particles in 4 cases of maximum μ_A values using 15 or 5 GLCM parameters by the two normalization schemes. These results show clearly that the mean accuracy values of μ_A are very low for either normalization scheme. Because RI values are related to molecular polarity, the calculated DIs correspond to particles of different morphological types and "molecular" types as well.

To examine the effect of morphology on classifiers with the GLCM parameters only, we removed those DIs calculated with low RI values of $n_p = n_{pl} = 1.400$ and performed clustering analysis on the 1791 particles with high RI value of $n_p = n_{ph} = 1.558$. Table 3 presents selected results of 5-parameter classifiers. In this case, μ_A improves significantly to a very high value of 95.7% with the type normalization scheme but it reduces to 60.9% if the uniform scheme was applied. Furthermore, even for the best case of $\mu_A = 95.7\%$, reducing number of particles can lead to significant fluctuations in accuracy that indicates the instability of classifier performance.

In above clustering analysis with GLCM parameters only, the input DIs were individually normalized and thus carry no information on relative differences in integrated scattered light intensity among the particles. One thus expects that clustering accuracy and stability may improve by adding two parameters of IFS and ISS. It is interesting to first note that the mean accuracy μ_A remained very low if IFS and ISS were used without GLCM parameters for clustering, which is expected based on numerous studies by the conventional flow cytometry using only forward scatter and side scatter signals. In contrast, with addition of IFS to GLCM parameters, the stability of classification accuracy was found to increase significantly while adding ISS to GLCM parameters show little improvement. Table 4 lists the 4 best classifiers of selected GLCM parameters and IFS which show that high classification accuracy can be achieved in multiple combinations of GLCM parameters. In Fig. 4, we present the normalized histograms of the GLCM parameter COR or correlation and IFS for particles having $n_p = n_{ph}$ that is used to obtain highest accuracy $\mu_A = 97.2\%$ for the case of the second hierarchical+GMM classifier in Table 4. One can see from the two plots that values of these parameters overlap significantly among the three particle types. The rescaled curves of Gaussian pdf obtained by the hierarchical+GMM classifier and Gaussian curve fitted to the labeled data are also shown in Fig. 4 respectively as solid and dash lines. For the histogram of COR, the two Gaussian curves exhibit larger difference in the case of spheres than the curves for cylinders and ellipsoids, which is consistent with confusion matrix in Table 4 showing larger errors of classification in the sphere case.

4. Discussion

The DIFC method based on diffraction imaging combined with flow cytometry presents a different approach for rapid analysis and classification of micro-sized particles in comparison to con-

Table 3
Confusion matrices of GLCM classifiers for particles of $n_p = n_{ph} = 1.558$.

Cluster	C_1	C_2	C_3	normalization scheme; (P_i); $\mu_A \pm \sigma_A$
Sphere	84	93	**388**	uniform norm. (ASM/VAR/IDM/ENT/DIS) 60.9% ± 5.7%
Cylinder	**375**	221	44	
Ellipsoid	138	**324**	124	
Sphere	57	**503**	5	type norm. (COR/VAR/SAV/DEN/MAP) 95.7% ± 4.8%
Cylinder	0	0	**640**	
Ellipsoid	**574**	8	4	

Table 4
Confusion matrices of GLCM + IFS classifiers for particles of $n_p = n_{ph} = 1.558$.[1].

Cluster	C1	C2	C3	(P_i); $\mu_A \pm \sigma_A$
Sphere	**515**	49	1	(COR/SEN/ENT/ DIS/IFS) 93.8% ± 2.6%
Cylinder	33	**598**	9	
Ellipsoid	7	8	**571**	
Sphere	**537**	27	1	(COR/IDM/SEN/ENT/ DIS/IFS) 97.2% ± 1.9%
Cylinder	1	**638**	1	
Ellipsoid	14	4	**568**	
Sphere	**467**	29	69	(CON/COR/IDM/SEN/ ENT/DIS/CLS/IFS) 83.4% ± 7.9%
Cylinder	0	166	**474**	
Ellipsoid	0	**547**	39	
Sphere	**463**	71	31	(COR/IDM/SEN/ENT/ DVA/DIS/CLS/CLP/IFS) 82.6% ± 8.0%
Cylinder	0	**468**	172	
Ellipsoid	0	43	**543**	

[1] All GLCM and IFS parameters were normalized by type scheme.

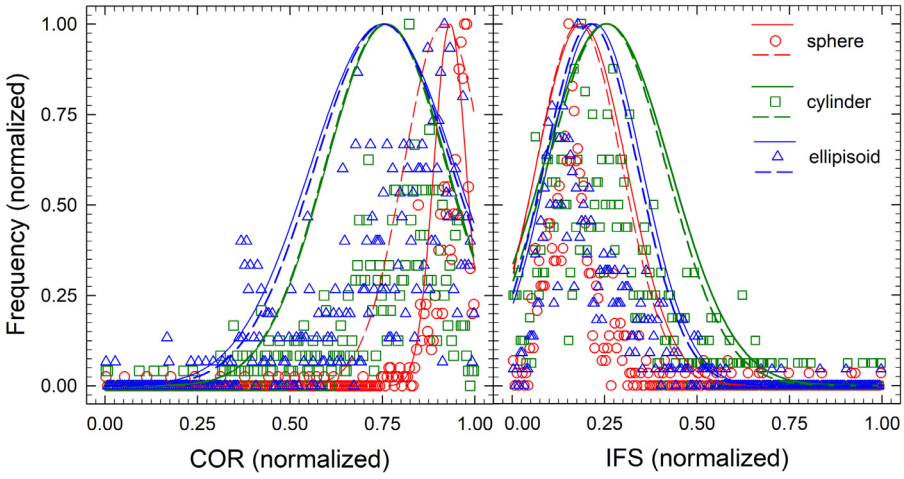

Fig. 4. Normalized histograms of COR and IFS parameters for the three particle types with $n_{ph} = 1.558$. The solid and dash lines represent respectively the rescaled curves of Gaussian pdf obtained by the hierarchical+GMM classifier and the Gaussian curves fitted to the histograms of the labelled data.

ventional microscopy and flow cytometry that requires fluorescent staining. Unlike digital holography and quantitative phase microscopy [25–27], the DIFC method is not to be used to retrieve m(**r**) of the scatterer particle in Eq. (1) from the measured distribution of I_s and thus needs no a priori knowledge on cell morphology and complex inverse calculations. Rather, the goal of DIFC approach is to extract effective pattern parameters from the DI data for morphology based analysis and classification of particles. In this report we have applied a validated method of simulation to calculate DI data from a large group of micro-sized homogeneous single particles made of single and double objects. Using the GLCM algorithm, the DI data have been characterized by up to 15 parameters per image that is formed by imaging the coherent light scattered along side directions. We have shown that accurate classification of 3 particle types can be achieved with statistical significance using a GMM based clustering algorithm in the parameter space of GLCM and angularly integrated forward intensity of scattered light even when they have homogeneous and identical intraparticle RI value.

We have previously shown that classification of biological cells according to their phenotypes can be achieved with the DIFC data in the cases of Jurkat versus Ramos derived from human lymphocytes [10], PC3 versus PCS from human prostate cells [12] and other cell lines [8,28]. In comparison, classification of regularly-shaped and homogeneous particles is expected to be difficult because only shape difference affects the diffraction patterns. This can be corroborated by comparing the ranges of GLCM parameters between the measured DIs of prostate cells [12,29] and calculated DIs of the three particle types in this report. For example, the parameter COR or correlation is one of the key GLCM parameters used for achieving the cases of high classification accuracy in Tables 3 and 4. By definition, COR measures the intensity correlation of the paired pixels for GLCM calculation and ranges between −1 and 1. The COR values of the calculated DIs for the three parti-

cle types are very high ranging from 0.992 to 0.999 while those for the measured DIs of prostate cells are much smaller ranging from -4.92×10^{-3} to 0.996 with mean values around 0.84. It is clear that the much stronger intensity correlations of the two neighboring pixels in our case here is a consequence of RI homogeneity and high degree in shape symmetry in comparison to those of biological cells. Thus, the failure of classification algorithms such as k-means clustering (results not shown) that require overlapping as small as possible in parameter space may be attributed to the mixed nature of COR and other parameter distributions as displayed in Fig. 4, where our results demonstrate the significant improvement in accuracy by GMM. Taken together, these results allow us to quantitatively assess the ability of GLCM parameters and scattering signals for label-free analysis of homogeneous particles, which is very useful for further increasing classifiers' performance on DI analysis with supervised machine learning tools such as deep neural networks. The outcomes of this study also provide fresh insight on the correlation between 3D shapes of homogeneous particles and pattern features of DI data. Finally, we note that in additional to COR other GLCM parameters including sum entropy (SEN), entropy (ENT), dissimilarity (DIS) and inverse difference moment (IDM) [13] are also effective in recognizing the particle types.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jqsrt.2018.12.010.

References

[1] Sorensen CM. Light scattering by fractal aggregates: a review. Aerosol Sci Technol 2001;35:648–87.
[2] Snyder WA, Arnone RA, Davis CO, Goode W, Gould RW, Ladner S, et al. Optical scattering and backscattering by organic and inorganic particulates in U.S. coastal waters. Appl Opt 2008;47:666–77.
[3] Wang JJ, Han YP, Chang JY, Chen ZY. Light scattering of a Bessel beam by a nucleated biological cell: an eccentric sphere model. J Quant Spectrosc Radiat Transf 2018;206:22–30.
[4] Jones TR, Carpenter AE, Lamprecht MR, Moffat J, Silver SJ, Grenier JK, et al. Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. Proc Natl Acad Sci USA 2009;106:1826–31.
[5] Jacobs KM, Lu JQ, Hu XH. Development of a diffraction imaging flow cytometer. Opt Lett 2009;34:2985–7.
[6] Jacobs KM, Yang LV, Ding J, Ekpenyong AE, Castellone R, Lu JQ, et al. Diffraction imaging of spheres and melanoma cells with a microscope objective. J Biophotonics 2009;2:521–7.
[7] Wang H, Feng Y, Sa Y, Ma Y, Pan R, Lu JQ, et al. Acquisition of cross-polarized diffraction images and study of blurring effect by one time-delay-integration camera. Appl Opt 2015;54:5223–8.
[8] Dong K, Feng Y, Jacobs KM, Lu JQ, Brock RS, Yang LV, et al. Label-free classification of cultured cells through diffraction imaging. Biomed Opt Express 2011;2:1717–26.
[9] Sa Y, Zhang J, Moran MS, Lu JQ, Feng Y, Hu XH. A novel method of diffraction imaging flow cytometry for sizing microspheres. Opt Express 2012;20:22245–51.
[10] Feng Y, Zhang N, Jacobs KM, Jiang W, Yang LV, Li Z, et al. Polarization imaging and classification of Jurkat T and Ramos B cells using a flow cytometer. Cytometry A 2014;85:817–26.
[11] Yang X, Feng Y, Liu Y, Zhang N, Lin W, Sa Y, et al. A quantitative method for measurement of HL-60 cell apoptosis based on diffraction imaging flow cytometry technique. Biomed Opt Express 2014;5:2172–83.
[12] Jiang W, Lu JQ, Yang LV, Sa Y, Feng Y, Ding J, et al. Comparison study of distinguishing cancerous and normal prostate epithelial cells by confocal and polarization diffraction imaging. J Biomed Opt 2016;21:071102.
[13] Haralick RM. Statistical and structural approaches to texture. Proc IEEE 1979;67:786–804.
[14] Ohanian PP, Dubes RC. Performance evaluation for four classes of textural features. Pattern Recognit 1992;25:819–33.
[15] Shahbahrami A, Pham TA, Bertels K. Parallel implementation of gray level co-occurrence matrices and Haralick texture features on cell architecture. J Supercomput 2012;59:1455–77.
[16] Sa Y, Feng Y, Jacobs KM, Yang J, Pan R, Gkigkitzis I, et al. Study of low speed flow cytometry for diffraction imaging with different chamber and nozzle designs. Cytometry A 2013;83:1027–33.
[17] Pan R, Feng Y, Sa Y, Lu JQ, Jacobs KM, Hu XH. Analysis of diffraction imaging in non-conjugate configurations. Opt Express 2014;22:31568–74.
[18] Lu JQ, Yang P, Hu XH. Simulations of light Scattering from a biconcave red blood cell using the FDTD method. J Biomed Opt 2005;10:024022.
[19] Yurkin MA, Hoekstra AG. The discrete dipole approximation: an overview and recent developments. J Quant Spectrosc Radiat Transfer 2007;106:558–89.
[20] Yurkin MA, Hoekstra AG. The discrete-dipole-approximation code ADDA: capabilities and known limitations. J Quant Spectrosc Radiat Transfer 2011;112:2234–47.
[21] Ma X, Lu JQ, Brock RS, Jacobs KM, Yang P, Hu XH. Determination of complex refractive index of polystyrene microspheres from 370 to 1610nm. Phys Med Biol 2003;48:4165–72.
[22] Zhang J, Feng Y, Jiang W, Lu JQ, Sa Y, Ding J, et al. Realistic optical cell modeling and diffraction imaging simulation for study of optical and morphological parameters of nucleus. Opt Express 2016;24:366–77.
[23] Willett P. Recent trends in hierarchic document clustering: a critical review. Inf Process Manag 1988;24:577–97.
[24] Shental N, Bar-Hillel A, Hertz T, Weinshall D. Computing Gaussian mixture models with EM using equivalence constraints. In: Proceedings of the neural information processing systems; 2003.
[25] Choi W, Fang-Yen C, Badizadegan K, Dasari RR, Feld MS. Extended depth of focus in tomographic phase microscopy using a propagation algorithm. Opt Lett 2008;33:171–3.
[26] Merola F, Memmolo P, Miccio L, Savoia R, Mugnano M, Fontana A, et al. Tomographic flow cytometry by digital holography. Light-Sci Appl 2017;6:e16241.
[27] Steelman ZA, Eldridge WJ, Weintraub JB, Wax A. Is the nuclear refractive index lower than cytoplasm? Validation of phase measurements and implications for light scattering technologies. J Biophotonics 2017;10:1714–22.
[28] Wang H, Feng Y, Sa Y, Lu JQ, Ding J, Zhang J, et al. Pattern recognition and classification of two cancer cell lines by diffraction imaging at multiple pixel distances. Pattern Recognit 2016;61:234–44.
[29] Wang S, Liu J, Lu JQ, Wang W, Al-Qaysi SA, Xu Y, et al. Development and evaluation of realistic optical cell models for rapid and label-free cell assay by diffraction imaging. J Biophotonics 2018. doi:10.1002/jbio.201800287.