Computational analysis of mitochondrial placement and aggregation effects on wide-angle cell scattering patterns

Patrick M. Pilarski[a], Xuan-Tao Su[b], D. Moira Glerum[c], and Christopher J. Backhouse[a]

[a]Dept. of ECE, University of Alberta, T6G 2V4, Edmonton, AB, Canada;
[b]Dept. of Physics, University of Alberta, T6G 2G7, Edmonton, AB, Canada;
[c]Dept. of Medical Genetics, University of Alberta, T6G 2H7, Edmonton, AB, Canada

ABSTRACT

Image analysis and pattern recognition are key elements of many biomedical analysis schemes. In this work, we show the use of pattern recognition and classification for the study of an interesting biomedical problem—the prediction of organelle arrangement within a cell based on wide-angle light scattering patterns. Organelle distribution is known to relate to disease and drug resistance. However, up until this point it has been unclear how changes to organelle distribution relate to the composition of wide-angle light scattering patterns. As such, we use a rapid new scattering simulation method and standard pattern analysis techniques to demonstrate clear correlations between scattering pattern composition and organelle distribution. The texture of scattering images—specifically the spot and edge content of samples—is found to directly relate to the type and size of organelle distributions within a cell. These relationships are used to quickly classify organelle distributions to a high degree of accuracy.

Keywords: biomedical image analysis, wide-angle light scattering, mitochondria, rapid simulation, medical diagnostics, organelle distributions, pattern classification, texture analysis, attribute selection, cytometry

1. INTRODUCTION

Image analysis and pattern recognition are key elements of many biomedical analysis schemes.[1] In a number of cases, the decomposition of an experimental image into a set of representative features and attributes allows biomedical data to be quickly analyzed, classified, and used to supplement clinical diagnostic practices.[1,2] One relevant new domain for this kind of analysis is the prediction of organelle arrangement within a cell from a sample of its wide-angle light scattering pattern.

Recent work has shown that the distribution and aggregation of organelles within a cell can have great predictive value.[3,4] Changes to organelles and their distribution have been related to the progression of a number of common disorders, ranging from cancer to neurodegenerative disease,[5,6] and can also predict drug resistance (*e.g.* a cell's response to treatments like chemotherapy[3,4]).

Unfortunately, there are few non-invasive methods to detect changes to the distribution of organelles within a cell.[3] Wide-angle cytometers provide one viable analysis method, as they do not rely on invasive labeling techniques and are able to capture the wide-angle light scattering from a single live cell.[7,8] However, up to this point it has been unclear whether changes to the distribution of organelles will lead to detectable and consistent changes to the composition of a cell's wide-angle scattering pattern. This is primarily a question of pattern analysis and correlation detection.

A number of groups have demonstrated pattern analysis methods able to identify a link between light scattering patterns and other aspects of cell structure. Methods are available to detect cell size,[9–12] organelle content (cell granularity),[12] hemoglobin concentration,[9] and optical properties[9–11] from a two-angle and/or a one-dimensional scattering sample. As yet there have been no demonstrated methods that can relate wide-angle scattering patterns to organelle distribution.

In the present work we demonstrate clear correlations between the arrangement (distribution) of organelles within a cell and the composition of wide-angle scattering patterns, thus identifying scattering trends that are

Send correspondence to: patrick.pilarski@ualberta.ca

Biomedical Applications of Light Scattering III, edited by Adam Wax, Vadim Backman
Proc. of SPIE Vol. 7187, 71870J · © 2009 SPIE · CCC code: 1605-7422/09/$18 · doi: 10.1117/12.809730

independent of cellular rotation and the inherent randomness of organelle position within a cell. Small changes to scatterer organization can dramatically alter the resulting pattern, making this a non-trivial problem.[13,14] In addition, we have obtained qualitative and quantitative descriptions of the identified scattering trends, and present an example of their potential for use in automated cell classification.

1.1 Rapid Scattering Simulation

To pursue pattern-based correlation analysis, a comprehensive database of labeled images is typically required.[1,2] For the present case, a database would ideally contain scattering patterns from a wide range of different cellular orientations and organelle distributions. This would make it possible to identify correlations that can determine *organelle distribution* independent of *specific organelle placement*.

An image database with these specifications was previously very difficult or impossible to generate for the case of cellular light scattering, due to the large computational cost of simulation (*e.g.* using Finite Difference Time Domain methods[7,13] (FDTD)), the lack of well-labeled testing data, and restrictive simulation models.[15] However, we recently demonstrated a computational method—the `mtPatterns` algorithm—that is capable of rapidly simulating wide-angle light scattering from 3D distributions of organelles within a single cell.[7,15] Experiments show that `mtPatterns` scattering simulations agree well with both experimental scattering patterns and FDTD simulations,[15] providing labeled scattering data without the high computational cost of FDTD—each pattern can be generated in minutes on a desktop computer.[15] This makes it possible to create a large and varied database of labeled wide-angle scattering images. In the present work, we expand the `mtPatterns` algorithm to allow the placement of organelles of varying size in constrained distributions, and apply it to the study of the interaction between the distribution of intracellular organelles and their observed light scattering patterns.

1.2 Background on Mitochondria and Disease

For this work we focus on one specific intracellular organelle: the mitochondrion. Not only are mitochondria responsible for the majority of wide-angle light scattering,[16] making them a primary determinant of scattering pattern composition,[15] but their form and function have been found to relate to the progression of a number of ailments.[3–6,16,17] Specifically, mitochondria are critical to cellular energy production. Mutations that affect mitochondria and related proteins have been identified in disorders such as cancer,[5,6,16] cardiomyopathy,[5] diabetes,[6] and common neurodegenerative disorders like Parkinson and Alzheimer Diseases.[6,7]

Despite a debate surrounding the exact shape, size, and distribution of mitochondria within human (and other mammalian) cells, it appears their morphology is highly cell-dependent,[16] and they are commonly held to be able to alter their shape, size, aggregation, and location within a cell based on external stimulus.[3,4,18] Table 1 presents a summary of experimentally and theoretically obtained mitochondrial geometries for selected cell types relevant to this work. Mitochondrial clustering can take many forms, including *perinuclear* (distributed around the nucleus), *peripheral* (distributed around the outside of the cell), *diffuse* (evenly distributed within a cell), and *aggregate* (in tight clumps within a cell). These distributions are shown in Fig. 1.

For many disorders, there is a noticeable genetic or structural difference between the mitochondrial populations of healthy and damaged cells.[3,7,16] In particular, the distribution and aggregation of mitochondria within a cell has been shown to relate to different disease-related metabolic states, and can help predict a damaged cell's response to treatment.[3] The presence of *perinuclear* mitochondrial clustering in cultured cancer cells has been shown to indicate a sensitivity to chemotherapy, while a *peripheral distribution* in the same culture indicates potential chemotherapy resistance.[3,4] In a similar fashion, a differentiation between *diffuse* and *perinuclear* clusters can help distinguish between normal and transformed cells,[3] and the detection of *aggregates* can help in cell classification.[3,16]

If it were possible to rapidly assess mitochondrial distribution, it could be used as a powerful indicator of cellular heath. Unfortunately, there are few—if any—current techniques that can non-invasively identify mitochondrial clustering trends in mixed cell populations.[3] However, as shown in this work, wide-angle light scattering analysis could provide one method to perform non-invasive, label-free distribution analysis when supplemented with image analysis and pattern recognition tools.

Table 1. Comparison of selected mitochondria dimensions presented in the literature

Length	Width	Cell Type	Number	Shape	Source
0.25–1.0μm	0.25–1.0μm	Human (diseased, Raji[a])	83–677	sphere	([7])
0.68–0.73μm	0.68–0.73μm	Human (normal)	$10^{12} cm^{-3}$	sphere, rod	([16,19])
0.75–0.83μm	0.75–0.83μm	Human (diseased[b])	$10^{12} cm^{-3}$	sphere, rod	([16,19])
1.0–2.0μm	0.5–1.0μm	Human (diseased)	varies	oval	([20])
2.1/2.3μm	2.1/2.3μm	Human (diseased, Siha/MCF7[c])	varies	sphere	([21])

[a] Raji (Burkitt's lymphoma) cells have an outer diameter of 16μm and a nuclear diameter of 8μm.[7]
[b] Deficiencies relating to cytochrome oxidase enzyme.[19]
[c] Siha/MFC7 (cervical carcinoma) cells have an outer diameter of 15μm and a nuclear diameter of 9μm.[21]

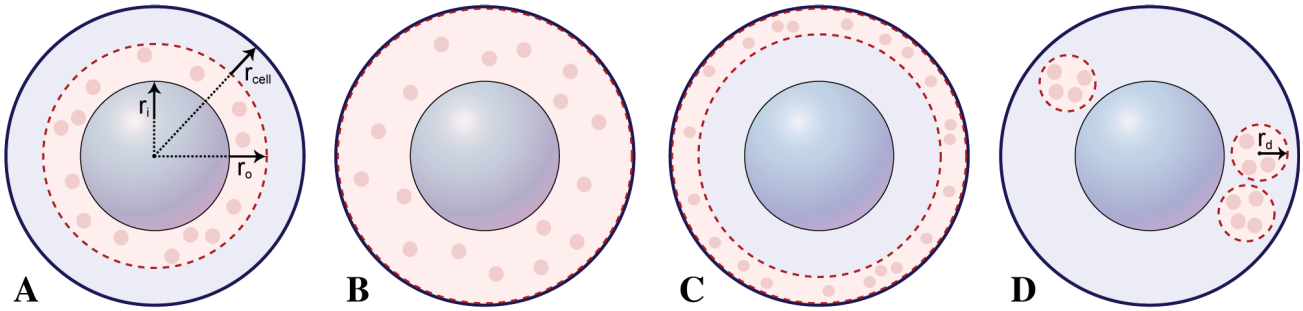

Figure 1. Four different organelle distributions: (A) a *perinuclear distribution*,[3] where organelles are distributed close to the nuclear wall ($r_i = r_{nucleus}$) but still inside an outer sphere of radius r_o; (B) a *diffuse distribution*[3] where organelles are randomly placed throughout the entire volume between the cell wall and the nuclear wall; (C) a *peripheral distribution*,[3] where organelles are placed in a spherical shell of inside radius r_i and outer radius $r_o = r_{cellwall}$; and (D) an *aggregate distribution*,[3,16] where organelles are distributed in one or more tightly packed spherical regions of radius r_d (*e.g.* a highly organized collection of mitochondria[16]).

2. METHODS

To evaluate the correlation between mitochondrial placement and scattering pattern composition, we utilized a rapid light-scattering simulation method—the `mtPatterns` algorithm[15]—to generate wide-angle scattering patterns for the different diseases-related mitochondrial distributions described in the previous section. The resulting database of scattering images was evaluated using a set of standard biomedical image analysis routines.[1,22] From this evaluation, image characteristics with the greatest predictive value were identified using a bank of established attribute selection methods.[2] The selected attributes were then analyzed for their relationship to organelle distribution, and used in four different classifiers to demonstrate their predictive ability. Image generation and simulation setup, image analysis methods, and attribute selection methods are described in detail below. Results and a discussion of the data will be presented in the following section.

2.1 Simulation Setup

For the tests described in this study, we based the simulation model parameters on an immune cell from the Raji celline—an example of Burkitt's lymphoma. This cell model has been previously described in Su *et al.* 2007[7] and Pilarski *et al.* 2008.[15] As indicated in Tab. 1, it assumes an outer cell radius of 8μm, a nuclear radius of 4μm, and approximately 83–677 mitochondria distributed within the available cell volume (*i.e.* between the outer wall and the nucleus, a volume of approximately 151μm^3.) Refractive index values for this cell model are described by Su *et al.*[7] Using the Raji cell parameters as a base, we examined the four mitochondrial distributions shown in Fig. 1—*perinuclear, diffuse, peripheral,* and *aggregate* clustering.

We created examples of each distribution type at three mitochondria concentration levels: when they contained 83 mitochondria, 250 mitochondria, and 677 mitochondria—the full range expected for the Raji immune cell.[7] The effective size of mitochondria—in terms of their achievable spacing within a cell—was also varied for

Table 2. Breakdown of texture analysis classes and attributes used

Parameter	Labels and Attributes
Class Labels (7):	Perinuclear (r=4.0–4.8μm), Perinuclear (r=4.0–5.6μm), Perinuclear (r=4.0–6.4μm), Aggregate (r_d = 1.0μm), Aggregate (r_d = 1.5μm), Peripheral (r=7.8–8.0μm), Diffuse (r=4.0–8.0μm)
Texture Attributes (21):	variance, skewness, kurtosis, haralick-f1-homogeneity, haralick-f2-contrast, haralick-f3-correlation, haralick-f4-sumofsquares, haralick-f5-local-homogeneity, haralick-f6-sumavg, haralick-f7-sumvar, haralick-f8-sumentropy, haralick-f9-entropy, haralick-f10-diffvar, haralick-f11-diffentropy, law-l5, law-e5, law-s5, law-s5x2, law-s5x4, law-r5, law-w5
Data Points (170)	85 ($d_{mt} = 0.25\mu m$ data), 85 ($d_{mt} = 0.75\mu m$ data)

each test model, and results are presented for organelles placed using an effective diameter of $d_{mt} = 0.25\mu m$ and $d_{mt} = 0.75\mu m$ (spherical mitochondria, as per the examples shown in Tab. 1; scattering as described by Pilarski *et al.*[15]). For each distribution type and concentration level, we performed the simulation five times, each with a different randomly generated organelle arrangement (i.e. five different random seeds) within the specified distribution to simulate cell rotation and other variability inherent in real experimental systems. This resulted in a total of 170 different cell models for simulation and analysis.

Using these cell models, we applied the `mtPatterns` algorithm[15] to generate the final database of 170 scattering images. The `mtPatterns` algorithm was implemented as described by Pilarski *et al.*,[15] and initialized with an incident light wavelength of $\lambda = 632nm$, a receptive field area of $3mm \times 3mm$, and with the center of the cell model placed $d = 5mm$ below the center of the receptive plane. As indicated by Pilarski *et al.*, this setup gives a viewable side-scatter region between 77.3° and 106.7° in both the θ and ϕ axes (*e.g.* a cone of $\simeq 30°$).[15]

2.2 Image Analysis

Image analysis is an important part of most biomedical analysis systems,[1] where it is commonly used to extract a set of shape and texture attributes from experimental images. These attributes are then evaluated for medical significance, usually through the use of attribute selection and classification routines.[1] Within the set of possible attributes, texture attributes have the advantage that they are largely objective with respect to the determination of region boundaries and ground truth values;[1] as such, they provide a robust numerical platform for examining the image composition of complex light scattering images. For this work, we utilized three common approaches to image texture analysis, as described by Rangayyan: *statistical texture measures, Haralick texture measures,* and *Law texture energy.*[1] These measures each focus on different aspects of an image's composition in terms of grey-level intensity, and together form a compact but descriptive image signature.

Statistical texture measures evaluate the distribution of an image's grey-level histogram. Histograms are analyzed for *variance*, *skewness* (how off-center a distribution is) and *kurtosis* (the peakedness of a distribution); for this work, computation was performed as described by Rangayyan.[1] While statistical metrics are based on an intensity histogram, irrespective of spatial arrangement, the fourteen Haralick texture metrics (F1–F14) are based on local relationships in the image derived from a grey-level correlation matrix (GLCM).[1] This matrix is a measure of how likely a pixel of a given intensity level will be next to—at a given distance and angle—a pixel of another intensity level (computed for all different intensity levels). As per Rangayyan, we used a pixel distance of one, and averaged over relationships at $0°, 45°, 90°$, and $135°$.[1] After forming a GLCM for each image, a number of Haralick measures were extracted that indicate the level of homogeneity, correlation, variance, consistency, and entropy in a scattering image. In the interest of space, the exact form and meaning of each Haralick equation used in this work can be found in Rangayyan 2004.[1]

Table 3. Attribute selection results

Method	**Selected Attributes or Subsets** (top five, in order of decreasing rank)
$\mathbf{d_{mt} = 0.25\mu m}$	
CFS Subset Eval.[23]	**law-e5**, law-s5, law-s5x2
ReliefF[23]	**law-s5x2**, law-e5, law-s5, law-r5, law-w5, ...
InfoGain[23]	**law-s5x2**, law-e5, law-s5, law-w5, law-r5, ...
Consistency[23]	**haralick-f7-sumvar**, law-e5, law-s5
Wrapper (BayesNet)[23]	**haralick-f3-correlation**, law-e5, law-s5x2
ChiSquared[2]	**law-e5**, law-s5, law-s5x2, law-w5, law-r5, ...
$\mathbf{d_{mt} = 0.75\mu m}$	
CFS Subset Eval.[23]	**law-e5**, law-s5, law-s5x2, law-w5
ReliefF[23]	**law-s5x2**, law-e5, law-s5, law-w5, law-s5x4, ...
InfoGain[23]	**law-s5x2**, law-s5, law-e5, law-w5, law-r5, ...
Consistency[23]	**haralick-f7-sumvar**, haralick-f8-sumentropy, haralick-f11-diffentropy, law-e5, law-s5x2, ...
Wrapper (BayesNet)[23]	**haralick-f9-entropy**, haralick-f10-diffvar, law-s5x2
ChiSquared[2]	**law-s5x2**, law-e5, law-s5, law-w5, law-r5, ...

Bold indicates the most relevant attribute, as ranked by each routine.

Law texture measures give a more qualitative view of image texture, and are based on the convolution of a 5×5 filter kernel with the target image. The summation of intensity (energy) in the filtered image can be used to evaluate a texture's edge content (*law-e5*), spot content (*law-s5*), wave content (*law-w5*), and ripple content (*law-r5*). For this work we also implemented two additional spot measures, *law-s5x2* and *law-s5x4*, that measured spot content at two and four times the scale of the standard *law-s5* filter. This was done by reducing the size of the filter's input image by a factor of two and four, respectively. A more detailed discussion of Law texture measures and their meaning can be found in Rangayyan 2004.[1]

All tests were done for an image reduction level to 25 pixels by 25 pixels. After analysis, each vector of features (statistical, Haralick, and Law; twenty-one attributes in total) was labeled with the class of the organelle distribution used to generate the image. A full list of attributes and class labels used in the work can be found in Tab. 2. Once labeled, the data could be used for attribute selection and classification.

2.3 Attribute Selection

Attribute selection is the process of determining which of a number of attributes (in this case texture measures) contains the most information about the class label of a data point. There are a number of different attribute selection techniques, each with their own areas of specialty and modes of operation.[23] As such, it is important to use a suite of different attribute selection techniques, to allow for better identification of relevant features. Attributes indicated (voted for) by the majority of selection techniques are most likely to hold predictive information about the class of a sample.

For the task of identifying texture attributes that relate to organelle distribution, we applied five of the attribute selection mechanisms implemented by Hall *et al.* in their 2003 benchmarking study.[23] This includes `CFS Subset Evaluation, ReliefF, Information Gain, Consistency`, and a `Wrapper Bayes Network`.[23] For completeness, we also used a `Chi-Squared` selection routine.[2] As done by Hall *et al.*, all selection tasks were performed using the Weka environment (v3.4.11), a widely-used Java-based data mining toolkit.[2]

All twenty-one texture attributes (described in Tab. 2) were presented to each attribute selection routine. The top five results of each selection process are shown in Tab. 3, in decreasing order of relevance (left to right); some techniques selected a subset of less than five attributes. As shown in Tab. 4, attributes were then ranked by the number of votes they received—*i.e.* the number of times they were selected in the top five or top three by a given selection routine. As determined by these tests, the top three attributes were Law features measuring

Table 4. Attribute selection results: votes per attribute, out of twelve selection tests

Attribute	**Rank**	**Votes** (in top 5)	**Votes** (in top 3)
law-s5x2	1*	11	10
law-e5	2**	11	10
law-s5	3	9	8
law-w5	4	7	0
law-r5	5	5	0
haralick-f7-sumvar	6**	2	2
haralick-f3-correlation	7**	1	1
haralick-f9-entropy	7**	1	1
haralick-f8-sumentropy	8	1	1
haralick-f10-diffvar	8	1	1
haralick-f11-diffentropy	8	1	1
law-s5x4	9	1	0

* selected as most relevant attribute 5 out of 12 times, more than any other attribute.
** selected as most relevant attribute at least once.

the edge content of the image, *law-e5*, and the spot content of the images at two different scales, *law-s5* and *law-s5x2* (Tab. 4, above the dotted line). These attributes were consistently selected in the top three by almost all of the attribute selection routines.

Given that wide-angle mitochondrial scattering images are semi-ordered or disordered arrangements of bright spots, known to vary in size and spacing with the distribution and number of organelles,[15,24] it logically follows that these selected Law measures would contain predictive information about distribution class. For the remainder of this work, data trends will be presented in terms of these three attributes; the trends shown are also mirrored by a number of the other attributes, but with a lesser degree of class separation.

2.4 Extensions to the `mtPatterns` Algorithm

For use in this work, the `mtPatterns` algorithm of Pilarski *et al.* was extended by adding a collision checking routine to ensure the realistic placement of each scattering point given a set of physical constraints—when placing organelles within the available distribution volume, a spherical boundary was used to check for collisions with the cell wall, nucleus, and other organelles. In this way, the scattering points placed by the `mtPatterns` algorithm could maintain realistic spatial relationships, and avoid positions that would be not physically achievable for organelles with volumes similar to those of the mitochondria in Tab. 1. While this resulted in a slightly more demanding modeling algorithm, it was found to reduce the variability in recorded distribution data due to unobtainable organelle positions, and more closely matches the optical conditions present in a real cell.

3. RESULTS & DISCUSSION

We observed distinct qualitative and quantitative differences between the scattering patterns of each distribution type. In the aggregate case, the scattering pattern consisted of a small number of broad and smoothly changing bright intensity patches (Fig. 2, A). For the peripheral distribution, the number of intensity regions was much larger, while the size of each region was only a fraction of the image's width (Fig. 2, E). The other distributions were found to fall between these two extremes, in relation to distribution size and type (Fig. 2, B,C,D).

Tables 5 and 6 show that this change in image composition between distributions is reflected in the *law-e5*, *law-s5*, and *law-s5x2* texture energy measures. As noted above, these measures represent an image's edge content (*law-e5*) and spot content at two different scale sizes (*law-s5* and *law-s5x2*). Table 5 shows that the aggregate distribution did in fact have a much lower edge and spot content than the other distributions; this corresponds to the smoothness observed in Fig. 2, A. For the case of $d_{mt} = 0.25\mu m$ mitochondria, as the size of the distribution increased, there was a corresponding increase in the values of *law-e5*, *law-s5*, and *law-s5x2* (Tab. 5, top), significant in terms of the standard deviation of the samples (Tab. 5, bottom). Though not shown,

A **B** **C** **D** **E**

Figure 2. Qualitative comparison of five different mitochondrial distributions: (A) an *aggregate distribution* of 125 small organelles clustered within a $2\mu m$-diameter spherical region next to the nucleus; (B) a *perinuclear distribution* of 250 organelles, tightly clustered within $0.8\mu m$ of the nucleus; (C) a *perinuclear distribution* of 250 organelles, clustered within $1.6\mu m$ of the nucleus; (D) a *diffuse distribution* of 250 organelles; and (E) a *peripheral distribution* of 250 organelles, tightly clustered within $0.8\mu m$ of the outer cell wall.

Haralick attributes also indicated a greater level of local homogeneity in aggregate samples, when compared to the other distributions.

These trends were also found to be present for the case of $d_{mt} = 0.75\mu m$ mitochondria (Tab. 6). However, through a comparison of the data in Tabs. 5 and 6, we found that spot and edge attributes were on average slightly higher for each distribution in the $d_{mt} = 0.75\mu m$ mitochondria case. This is likely due to the greater separation between scattering points due to physical constraints; as demonstrated in our recent work, spacing between scattering points is inversely proportional to the size and spacing of intensity regions in the scattering image.[15] Spacing in fact appears to be the dominant cause of changes to scattering image composition, as changes to the size and arrangement of a distribution were much more closely linked to consistent image attribute variation than were changes to the mitochondria concentration. Correlations between image texture and mitochondria concentration will be explored at length in future work.

The importance of the *law-e5*, *law-s5*, and *law-s5x2* texture attributes for classification and prediction can be observed by plotting them in relation to each other, where data points are labeled with the class of each distribution. This is shown in Figs. 3 and 4 for the plot of *law-e5* energy v.s. *law-s5x2* energy (the two most informative attributes, as identified in Sec. 2). As seen in Fig. 3, there is a clear differentiation between the four major distribution types, with almost no overlap between the class clusters for aggregate distributions (purple stars and diamonds; bottom left, $0.25\mu m$ plot only), perinuclear distributions (blue circles, crosses, and downward triangles; middle), diffuse distributions (green upward triangles; mid to top right), and peripheral distributions (red stars; top right). Aggregates are not shown for the $d_{mt} = 0.75\mu m$ case (Fig. 4), as it was impossible to pack 83 or more organelles within either the $r_d = 1.0\mu m$ or $r_d = 1.5\mu m$ aggregate containing volume.

It was also possible to identify the different size sub-classes for each distribution. As shown in Fig. 3, top, there was a clear delineation between the $r_d = 1.0\mu m$ and $r_d = 1.5\mu m$ aggregate volumes. There was also an observed stratification within the three types of perinuclear distribution, with the average Law values being arranged in increasing order of distribution size (*i.e.* data points for larger values of r_o were closer to the top right). In the interest of space, only the relationships between *law-e5* and *law-s5x2* energy were included in this document; the observed relationships were also present in all other combinations of the three most relevant attributes, and to a lesser degree for other attributes selected by the routines in Sec. 2. These trends are very important for classification, as they allow the distinction of not only distribution class, but also sub-class size directly from image texture attributes. This facilitates the rapid classification of experimental images.

3.1 Classification of Distributions

To identify their use in the classification and prediction of organelle distributions in single cells, the *law-e5*, *law-s5*, and *law-s5x2* texture attributes were given as input to four different machine learning systems under 10-fold cross-validation (CV-10)—the data was divided into ten folds, trained on nine of the folds and then tested on the remaining, previously unseen, fold[2,25] (*i.e.* systems were tested for classification accuracy only on data they had never seen in training). For these tests, we used the standard Weka[2] implementation of

Table 5. Comparison of Law texture energy for scattering from mitochondria with $d = 0.25\mu m$ (averaged over 5 trials)

MEAN

Test Type	**E5**			**S5**			**S5x2**		
Perinuclear ($r_o = 4.8\mu m$)	132.5	144.5	139.5	20.0	22.1	22.8	58.7	68.0	68.5
Perinuclear ($r_o = 5.6\mu m$)	142.2	141.1	149.8	21.3	22.4	23.0	67.4	68.6	75.1
Perinuclear ($r_o = 6.4\mu m$)	150.1	154.0	154.7	23.9	25.3	24.3	78.2	84.6	82.5
Diffuse	172.6	171.6	168.6	29.9	29.4	28.4	109.9	105.7	102.1
Peripheral ($r_i = 7.2\mu m$)	190.4	192.6	193.2	40.0	42.5	42.4	142.4	147.1	148.0
Aggregate ($r_d = 1.0\mu m$)	19.9	–	–	11.3	–	–	20.3	–	–
Aggregate ($r_d = 1.5\mu m$)	27.1	–	–	12.6	–	–	22.7	–	–
	83mt	**250mt**	**677mt**	**83mt**	**250mt**	**677mt**	**83mt**	**250mt**	**677mt**

STANDARD DEVIATION

Test Type	**E5$_\sigma$**			**S5$_\sigma$**			**S5x2$_\sigma$**		
Perinuclear ($r_o = 4.8\mu m$)	8.96	8.45	3.19	1.17	1.60	0.65	6.91	4.66	3.27
Perinuclear ($r_o = 5.6\mu m$)	8.67	4.27	3.41	1.25	1.11	0.90	5.23	5.57	3.13
Perinuclear ($r_o = 6.4\mu m$)	3.42	13.62	4.59	1.31	1.99	0.95	5.60	9.76	5.08
Diffuse	5.37	3.83	5.59	1.69	1.57	1.55	7.97	3.01	6.04
Peripheral ($r_i = 7.2\mu m$)	3.16	6.73	2.49	3.07	4.56	2.21	5.91	10.92	3.78
Aggregate ($r_d = 1.0\mu m$)	1.52	–	–	0.75	–	–	1.76	–	–
Aggregate ($r_d = 1.5\mu m$)	2.45	–	–	0.87	–	–	1.65	–	–
	83mt	**250mt**	**677mt**	**83mt**	**250mt**	**677mt**	**83mt**	**250mt**	**677mt**

(–) dashes indicate tests where it was impossible to fit all organelles within the specified containing volume.

a random decision tree (`trees.RandomTree`), support vector machine (`functions.SMO`), multilayer perceptron (`functions.MultilayerPerceptron`, 2000 epochs), and Bayes network (`bayes.BayesNet`). This was done to cover a wide range of algorithmic assumptions and methods of handling the data. Classifiers were provided with the entire data set of 150 valid data points for both the $d_{mt} = 0.25\mu m$ and $d_{mt} = 0.75\mu m$ cases (models where all organelles could not be placed within a containing volume were not included). Each data point was labeled with one of four classifications: *aggregate, perinuclear, diffuse,* or *peripheral*.

The results of these tests are shown in Tab. 7. All classifiers showed a consistently high classification accuracy, corresponding to a misclassification of between 7 (Bayes network) and 10 (multilayer perceptron) instances. It was possible to accurately identify all four distribution types. From an examination of the confusion matrix for each classifier, in all cases except the multilayer perceptron, all misclassifications were between the diffuse and perinuclear classes. This is to be expected, as in Figs. 3 and 4 we can see some class overlap between clusters for the larger radius perinuclear distribution ($r_o = 6.4\mu m$) and the diffuse distribution, especially for the $d_{mt} = 0.75\mu m$ case. The higher error for the multilayer perceptron is due to its misclassification of one aggregate distribution as a perinuclear distribution, most likely due to confusion with the $r_o = 4.8\mu m$ perinuclear sub-class.

3.2 Relationship Between Distribution Average Radius and Image Complexity

Figure 5 demonstrates the correlation between image complexity and distribution size—as the outer radius of the distribution increased, there was a corresponding increase in the complexity of the image (Fig. 5, A). This increase in complexity was much more pronounced as the *average* radius of the distribution increased (*i.e.* an increase in both the inner and outer radius values, as opposed to just an increase in outer radius; *image complexity* $\propto r_{avg} = (r_i + r_o)/2$). Scatter from a shell distribution with increasing r_{avg} is shown in Fig. 5, B; the increase in complexity is much more noticeable than in Fig. 5, A. This was further confirmed using Weka regression analysis with Pace Regression[2] and Linear Regression,[2] which showed that linear combinations of the *law-e5*, *law-s5*, and *law-s5x2* attributes could approximately model the relationship between image composition complexity—in terms

Table 6. Comparison of Law texture energy for scattering from mitochondria with $d = 0.75\mu m$ (averaged over 5 trials)

MEAN

Test Type	**E5**			**S5**			**S5x2**		
Perinuclear ($r_o = 4.8\mu m$)	140.8	151.0	–	21.1	23.7	–	64.9	75.4	–
Perinuclear ($r_o = 5.6\mu m$)	138.1	148.5	–	21.9	23.9	–	68.8	77.2	–
Perinuclear ($r_o = 6.4\mu m$)	153.2	162.9	168.2	24.8	26.9	28.0	82.3	93.7	98.6
Diffuse	171.0	170.3	178.4	29.2	30.2	33.2	106.8	106.4	120.9
Peripheral ($r_i = 7.2\mu m$)	190.9	194.7	199.5	43.3	46.3	50.8	148.4	154.3	163.8
Aggregate ($r_d = 1.0\mu m$)	–	–	–	–	–	–	–	–	–
Aggregate ($r_d = 1.5\mu m$)	–	–	–	–	–	–	–	–	–
	83mt	**250mt**	**677mt**	**83mt**	**250mt**	**677mt**	**83mt**	**250mt**	**677mt**

STANDARD DEVIATION

Test Type	**E5$_\sigma$**			**S5$_\sigma$**			**S5x2$_\sigma$**		
Perinuclear ($r_o = 4.8\mu m$)	7.59	5.79	–	1.37	0.33	–	6.04	3.90	–
Perinuclear ($r_o = 5.6\mu m$)	7.59	12.39	–	0.78	1.18	–	5.60	7.08	–
Perinuclear ($r_o = 6.4\mu m$)	7.23	4.58	7.25	1.41	1.10	1.16	5.67	2.89	4.30
Diffuse	5.92	6.99	3.85	1.37	1.90	2.20	6.72	6.97	5.84
Peripheral ($r_i = 7.2\mu m$)	3.84	5.35	3.39	1.76	3.71	3.09	5.41	7.10	7.21
Aggregate ($r_d = 1.0\mu m$)	–	–	–	–	–	–	–	–	–
Aggregate ($r_d = 1.5\mu m$)	–	–	–	–	–	–	–	–	–
	83mt	**250mt**	**677mt**	**83mt**	**250mt**	**677mt**	**83mt**	**250mt**	**677mt**

Underline indicates an increase from the $d_{mt} = 0.25\mu m$ case.

Table 7. Classification accuracy for different learning methods using CV-10 (four classes, 150 data points)

Learning Method	**% Correct Classifications**
Bayes Network	95.3
Random Tree	94.7
SVM	94.0
Multi-layer Perceptron	93.3

Four classes used: *perinuclear, aggregate, peripheral, diffuse.*
Attributes used in classification: *law-e5, law-s5, law-s5x2.*

of spot and edge content—and average distribution radius with the resulting correlation coefficients exceeding 0.97 in both cases (*e.g.* with models $r_{avg} = [-0.3581 + 0.0247(\textit{law-e5}) + 0.0191(\textit{law-s5}) + 0.0145(\textit{law-s5x2})]$ for Pace Regression, and $r_{avg} = [0.0238(\textit{law-e5}) + 0.0202(\textit{law-s5x2}) - 0.221]$ for Linear Regression).

4. CONCLUSIONS

In this work we used image analysis and pattern recognition methods to identify relationships between wide-angle scattering pattern composition and the distribution of organelles (in this case mitochondria) within a cell. This was done using a library of scattering patterns generated by the `mtPatterns` algorithm[15] and a series of standard image analysis, attribute selection, and machine learning tools.

Scattering pattern composition—in terms of twenty-one image texture measures[1]—was found to vary in a predictable and consistent way between different organelle distributions and aggregations. Notably, the size and spacing of intensity regions were inversely proportional to the radius of the scattering distribution; this was reflected in the amount of edge and spot content detected in the images using Law texture measures. Very little

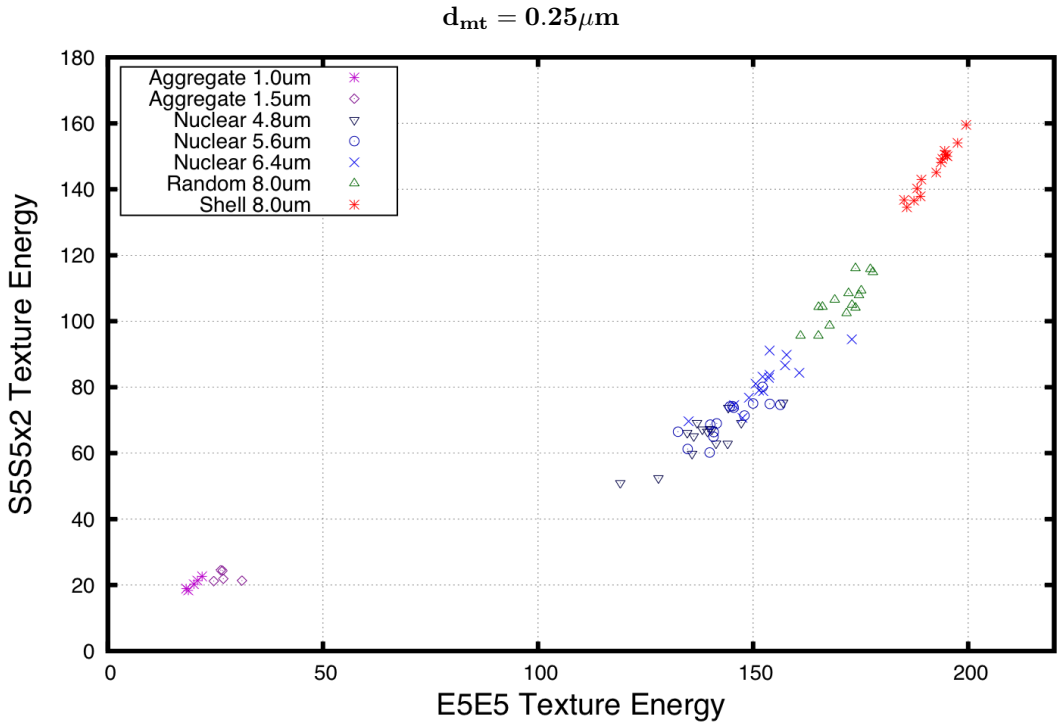

Figure 3. Plot of the *law-e5* and *law-s5x2* Law texture energy measures, labeled by mitochondria distribution class (colour and shape of each data point), for $0.25\mu m$ diameter organelles. Clear separation is visible between distributions that are clustered in aggregates (purple diamonds and stars; bottom left), around the nucleus (blue crosses, circles, and downward triangles; middle), diffusely (green upward triangles; mid to top right), and around the cell wall (red stars; top right). Similar trends were observed for all other combinations of *law-e5*, *law-s5*, and *law-s5x2* measures.

overlap between classes was observed in plots of edge v.s. spot content, and it was also possible to distinguish (to a first approximation) the average radius of a distribution from the data in these plots.

Using these identified trends, it was possible to classify different mitochondrial distributions directly from their wide-angle scattering patterns using Law texture measures for edge and spot content and standard machine learning techniques. This is important, as mitochondrial distribution has been shown to directly relate to cellular health and susceptibility to treatment.[3,4] Based on the agreement between `mtPatterns` simulations and published FDTD and experimental cytometry results,[7,15] these predictive trends are expected to extend well to the image composition of scattering patterns from actual patient samples. This will be explored in future work.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of the Natural Sciences and Engineering Research Council (NSERC) and the Canadian Institute for Photonic Innovations (CIPI). Patrick M. Pilarski is supported by funding from NSERC, the Informatics Circle of Research Excellence (iCORE), SPIE, and an Alberta Ingenuity Graduate Student Scholarship in Nanotechnology.

REFERENCES

[1] Rangayyan, R. M., [*Biomedical Image Analysis*], CRC Press (2004).

[2] Witten, I. H. and Frank, E., [*Data Mining: Practical Machine Learning Tools and Techniques*], Morgan Kaufmann, 2 ed. (2005).

[3] Sikder, S., Reyes, J. M. G., Moon, C. S., Suwan-apichon, O., Elisseeff, J. H., and Chuck, R. S., "Noninvasive mitochondrial imaging in live cell culture," *Photochemistry and Photobiology* **81**(6), 1569–1571 (2005).

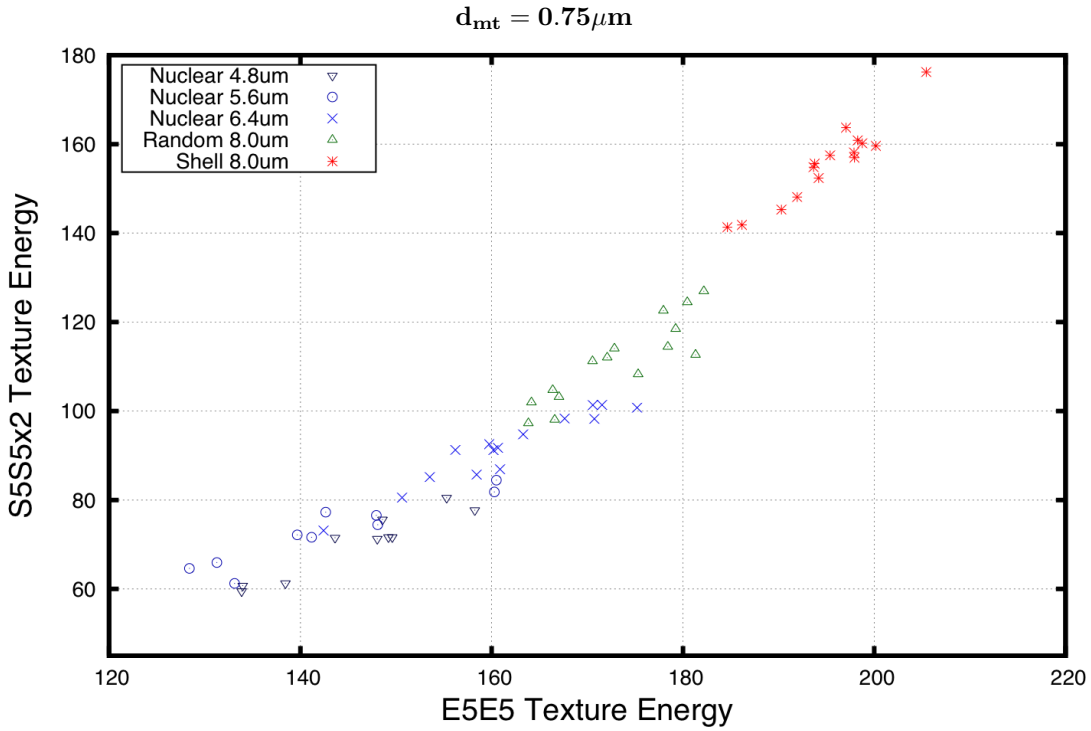

Figure 4. Plot of the *law-e5* and *law-s5x2* Law texture energy measures, labeled by mitochondria distribution class (colour and shape of each data point), for $0.75\mu m$ diameter organelles. Clear separation is visible between distributions that are clustered around the nucleus (blue crosses, circles, and downward triangles; middle), diffusely (green upward triangles; mid to top right), and around the cell wall (red stars; top right). Similar trends were observed for all other combinations of *law-e5*, *law-s5*, and *law-s5x2* measures.

[4] Villa, A. M. and Doglia, S. M., "Mitochondria in tumor cells studied by laser scanning confocal microscopy," *Journal of Biomedical Optics* **9**(2), 385–394 (2004).

[5] Wallace, D. C., "Mitochondrial diseases in man and mouse," *Science* **283**(5407), 1482–1488 (1999).

[6] Enns, G. M., "The contribution of mitochondria to common disorders," *Molecular Genetics and Metabolism* **80**(1-2), 11–26 (2003).

[7] Su, X.-T., Capjack, C., Rozmus, W., and Backhouse, C., "2d light scattering patterns of mitochondria in single cells," *Opt. Express* **15**(17), 10562–10575 (2007).

[8] Singh, K., Su, X., Liu, C., Capjack, C., Rozmus, W., and Backhouse, C. J., "A miniaturised wide-angle 2-d cytometer," *Cytometry A* **69A**, 307–315 (2006).

[9] Sem'yanov, K. A., Tarasov, P. A., Soini, J. T., Petrov, A. K., and Maltsev, V. P., "Calibration-free method to determine the size and hemoglobin concentration of individual red blood cells from light scattering," *Applied Optics* **39**(31), 5884–5889 (2000).

[10] Maltsev, V. P., "Scanning flow cytometry for individual particle analysis," *Review of Scientific Instruments* **71**(1), 243–255 (2000).

[11] Ulanowski, Z., Wang, Z., Kaye, P. H., and Ludlow, I. K., "Application of neural networks to the inverse scattering problem for spheres," *Applied Optics* **37**(18), 4027–4033 (1998).

[12] Pilarski, L., Baigorri, E., Mant, M., Pilarski, P., Adamson, P., Zola, H., and Belch, A., "Multiple myeloma includes cd20+ b and plasma cells that persist in patients treated with rituximab," *Clinical Medicine: Oncology* **2**, 275–287 (2008).

[13] Liu, C., Capjack, C. E., and Rozmus, W., "3-d simulation of light scattering from biological cells and cell differentiation," *Journal of Biomedical Optics* **10**(1), 014007 (12 pages) (2005).

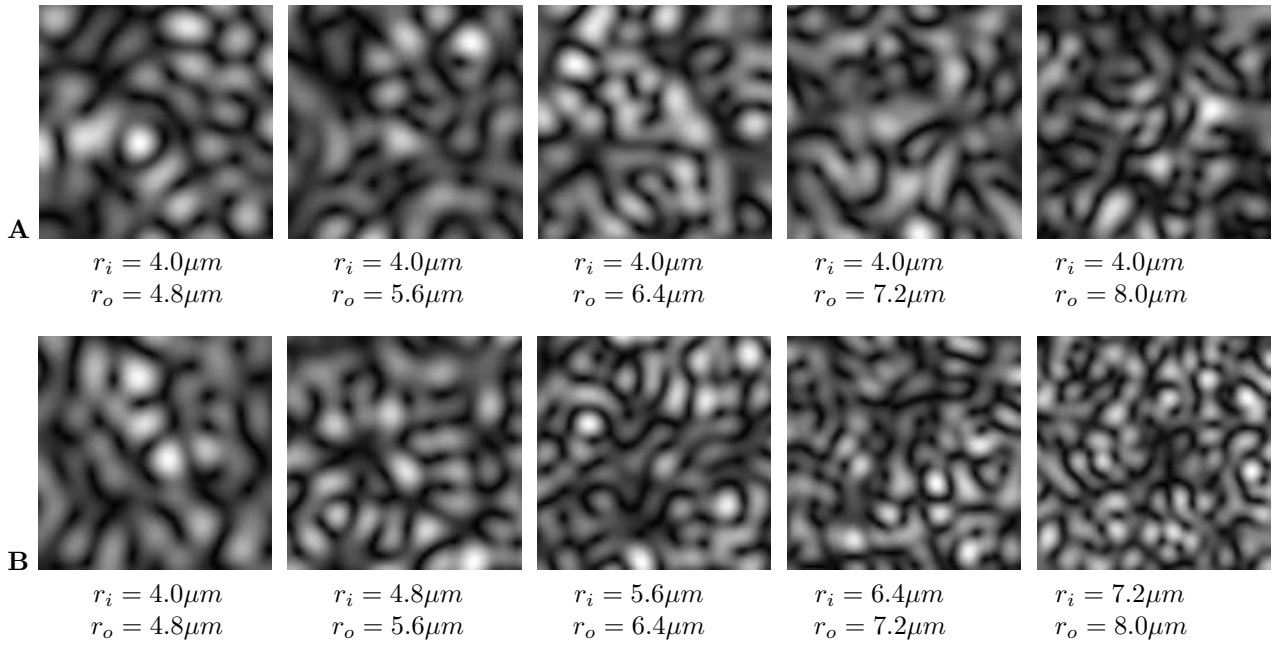

Figure 5. The relationship between the radius of an organelle distribution and scattering pattern intensity, for a fixed number of scatterers and two different distribution types: (A, top row) a perinuclear distribution with increasing radius—*i.e.* Fig. 1, C—and (B, bottom row) a shell of increasing radius—*i.e.* Fig. 1, B. Distributions contained 250 spherical organelles, each with an effective diameter of $0.75\mu m$.

[14] Bohren, C. F. and Huffman, D. R., [*Absorption and scattering of light by small particles*], Wiley, New York (1998).

[15] Pilarski, P. M., Su, X.-T., Glerum, D. M., and Backhouse, C. J., "Rapid simulation of wide-angle scattering from mitochondria in single cells," *Optics Express* **16**(17), 12819–12834 (2008).

[16] Gourley, P. L., Hendricks, J. K., McDonald, A. E., Copeland, R. G., Barrett, K. E., Gourley, C. R., and Naviaux, R. K., "Ultrafast nanolaser flow device for detecting cancer in single cells," *Biomedical Microdevices* **7**(4), 331–339 (2005).

[17] Zeviani, M. and Di Donato, S., "Mitochondrial disorders," *Brain* **127**(10), 2153–2172 (2004).

[18] Bereiter-Hahn, J. and Voth, M., "Dynamics of mitochondria in living cells: shape changes, dislocations, fusion, and fission of mitochondria," *Microscopy Research and Technique* **27**, 198–219 (1994).

[19] Gourley, P. L. and Naviaux, R. K., "Optical phenotyping of human mitochondria in a biocavity laser," *IEEE Journal of Selected Topics in Quantum Electronics* **11**(4), 818–826 (2005).

[20] Modica-Napolitano, J. S. and Singh, K., "Mitochondria as targets for detection and treatment of cancer," *Expert Reviews in Molecular Medicine* **4**(9), 1–19 (2004).

[21] Bartlett, M., Huang, G., Larcom, L., and Jiang, H. B., "Measurement of particle size distribution in mammalian cells in vitro by use of polarized light spectroscopy," *Applied Optics* **43**(6), 1296–1307 (2004).

[22] Shapiro, L. G. and Stockman, G. C., [*Computer Vision*], Prentice Hall (2001).

[23] Hall, M. A. and Holmes, G., "Benchmarking attribute selection techniques for discrete class data mining," *IEEE Trans. on Knowledge and Data Engineering* **15**(6), 1437–1447 (2003).

[24] Pilarski, P. M. and Backhouse, C. J., "A method for cytometric image parameterization," *Optics Express* **14**(26), 12720–12743 (2006).

[25] Duda, R. O., Hart, P. E., and Stork, D. G., [*Pattern Classification: second edition*], Wiley Interscience, New York (2001).