Spectral image microscopy for label-free blood and cancer cell identification

Mark Gesley, Robert Goldsby, Stephen Lane, Romin Puri

Mark Gesley, Robert Goldsby, Stephen Lane, Romin Puri, "Spectral image microscopy for label-free blood and cancer cell identification," Proc. SPIE 10890, Label-free Biomedical Imaging and Sensing (LBIS) 2019, 108900H (4 March 2019); doi: 10.1117/12.2507474

SPIE.

Event: SPIE BiOS, 2019, San Francisco, California, United States

Spectral image microscopy for label-free blood and cancer cell identification

Mark Gesley[a], Robert Goldsby[b], Stephen Lane[d], and Romin Puri[a]

[a]Spynsite LLC, Oakland, CA
[b]UCSF Benioff Children's Hospital, San Francisco, CA
[d]UC Davis, Sacramento, CA

ABSTRACT

New forms of cancer cell identification coupled with faster detection and better accuracy may enhance diagnostic capabilities. The purpose of this study is to improve recognition of minimal residual disease from peripheral blood samples. Spectral images are generated by optical microscopy using filtered broadband visible light elastically scattered from human blood and cancer cells. Exogenous tags, like CD markers may introduce a label bias and are not required. A training cell may be validated without detailed knowledge of intra-cellular spectra used to classify random cells. Spectral object classification is scalable to any number of cell types. Small samples of erythrocytes, leukocytes, Jurkat cancer and non-small lung cell adenocarcinoma are accurately classified and associated with unique spatial-spectral characteristics.

Keywords: Label-free cytology, spectral imaging, learning, classification

1. INTRODUCTION

Advanced methods to identify abnormal cells in the blood, bone marrow, or other tissue have enhanced diagnostic and prognostic capabilities with improved risk assessment, most notably for breast cancer. Detection of minimal residual disease by flow cytometry has become a powerful predictor of outcome in patients with hematologic malignancies and is currently being used to tailor therapy for children and adults with acute lymphoblastic leukemia. Novel spectral-based methods may provide complementary capabilities by identifying rare cells in large populations. There is evidence that visible light spectra can be used to distinguish cancer from normal cells[1,2] and fetal- from adult-nucleated red blood cells.[3] In some cases if circulating tumor cells can be detected at low concentration peripheral blood tests might replace or ameliorate invasive medical procedures.

In this paper methods to classify cells and characterize intra-cellular spectral structure are described. A spectral image microscope illuminates a dispersed thin film sample with broadband light optics filtered by a monochromator over narrow bands in the visible range. A high throughput system architecture is described in Ref. 4. Small samples of erythrocytes (Rbc), leukocytes (Wbc), Jurkat T-lymphocyte cancer (Jur) and adinocarcinoma (H1975) are used as examples.

2. OVERVIEW

Cell classification is accomplished in a supervised machine learning environment. A field-of-view is rendered as pixelized intensity measurement spectra from a frame array of monochromatic grey images, Fig. 1. A training sample selected from a test population of known cell types may be identified from one or more monochromatic grey images, a color composite, or otherwise substantiated by a verification authority. Samples identified on a basis separate from constituent measurements allows knowledge about a cell, such as morphology or other spatial - spectral features, but possibly none about underlying structure interaction with elastically scattered light spectra.

Training cells are used as a basis for classification, e.g. spectral images of cell A and cell B are used to classify a random cell $\hat{O}$, which is assigned to class : A as $(\hat{O} : A)$, class : B as $(\hat{O} : B)$, or an unknown class as $(\hat{O} : X)$ evaluated as a not A and not B class. This tri-class method scales to $(K + 1)$ cell types associated with K training samples and a new, untrained, or unknown class.

Further author information: send correspondence to M. Gesley, e-mail: gesley@spynsite.com, phone: 1 510 918 5752.

Label-free Biomedical Imaging and Sensing (LBIS) 2019, edited by Natan T. Shaked, Oliver Hayden, Proc. of SPIE Vol. 10890, 108900H · © 2019 SPIE
CCC code: 1605-7422/19/$18 · doi: 10.1117/12.2507474

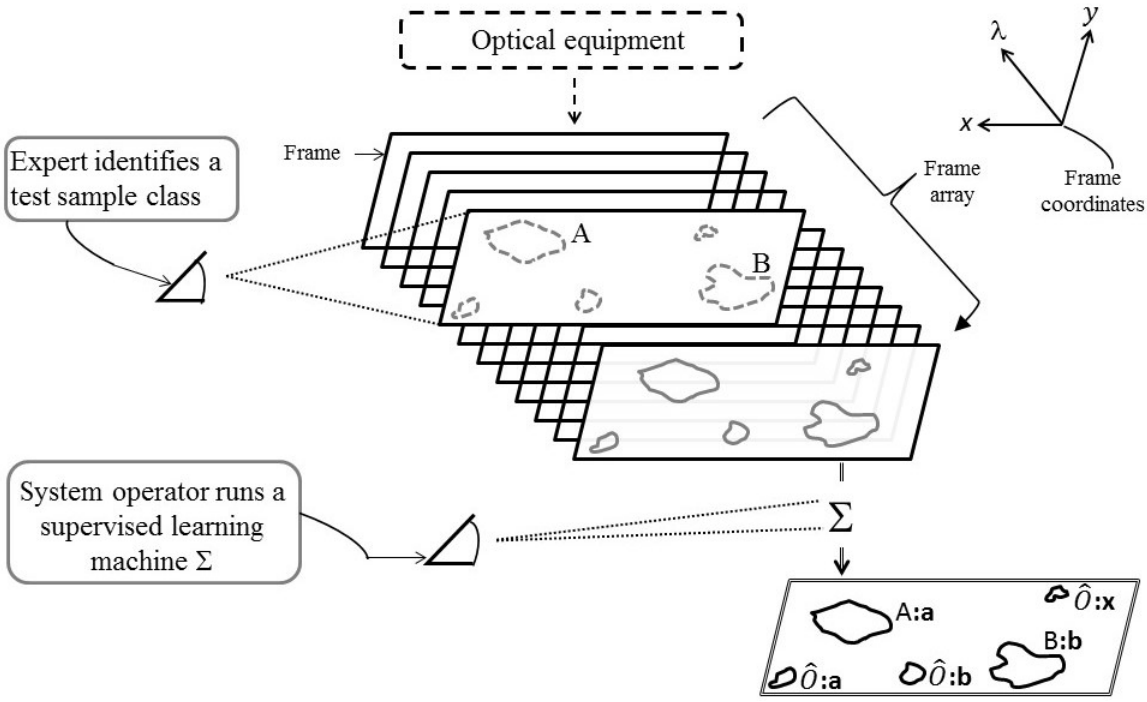

Figure 1. Supervised learning with cells rendered and classified from spectral images.

3. BACKGROUND AND PROBLEMS ADDRESSED

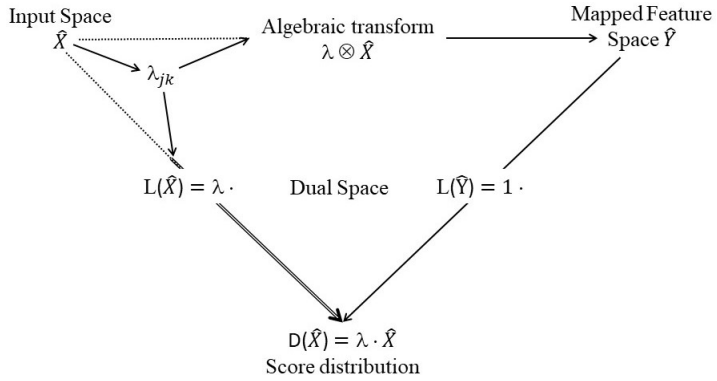

Figure 2. Vector spaces, linear functionals, and distributions.

A variety of methods address discrimination or classification of populations, Fig. 2. Algebraic or functional transformations to a mapped feature or dual space can increase training sample separation and improve binary decision accuracy. One path maps measurement spectra by an algebra with a product $\otimes$, such that a class set A is transformed to a linear scaled (mapped feature) space of p-dimensional scale vectors

$$\hat{Y}(A) \equiv \lambda \otimes \{A\} = \lambda \otimes \{\hat{x}_A\} = \{\lambda \otimes \hat{x}\}_A = \{\lambda_1 \hat{x}_1, \ldots, \lambda_p \hat{x}_p\}_A. \quad (1)$$

Statistical hypothesis testing interprets feature space components as class conditional probabilities for pattern recognition. A functional for each dimension is given by a probability calculated from conditional densities. A Bayes decision compares a likelihood to a threshold-bias. However, estimates of a multi-variate conditional density may not be available or difficult to compute, [5, pp.51-55].

A second path is to make a binary decision, which compares a transformed spectrum dual score $D(\hat{\mathbf{x}})$ to a threshold-bias (h), e.g. if $D(\hat{\mathbf{x}}) \geq h$, the spectrum is associated with decision a, $(\hat{\mathbf{x}} : a)$; else $(\hat{\mathbf{x}} : b)$. Linear transformations map training sample measurement spectra $\hat{\mathbf{x}}$ to a dual score distribution,

$$D(\hat{\mathbf{x}}) = \lambda \cdot \hat{x} = \sum_{j=1}^{p} \lambda_j \hat{x}_j \tag{2}$$

with λ a weight, linear functional, or scale vector depending on the context. A perceptron iteratively parameterizes weights and threshold-biases.[6] However parameters may not converge or sample order creates several solutions.[5,7]

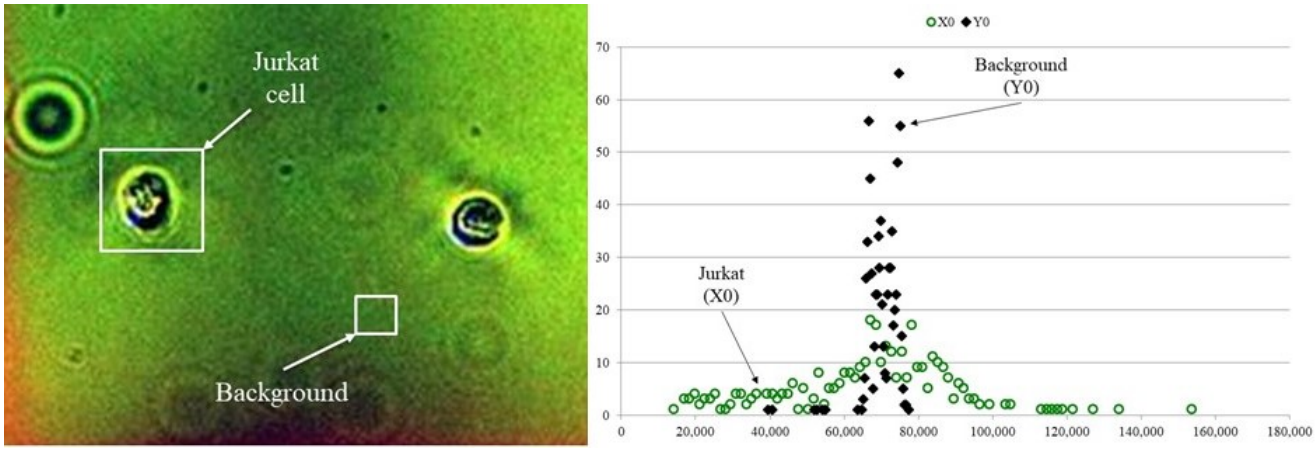

Figure 3. [left] RGB image bright field; [right] Jurkat cell ($X0$) and background ($Y0$) input score histograms.

Discriminating populations is fundamentally a nonlinear problem.[8] Neural networks model nonlinearities by a variety of constraints, interconnections, and states when calculating weights. However, a network architecture might be difficult to find or scale [9, pp.265-286] and the approach may be less effective if further insight to an underlying process is desired [10, pg.408]. A support vector machine (SVM) can evaluate a binary decision with a nonlinear kernel operator $K(w, \hat{\mathbf{x}})$ and weight vector parameter (w) to generate a score distribution,[11] e.g. if $K(w, \hat{\mathbf{x}}) \geq h$, associate $\hat{\mathbf{x}}$ with decision a $(\hat{\mathbf{x}} : a)$, else $(\hat{\mathbf{x}} : b)$. However, a heuristic kernel selection is required.

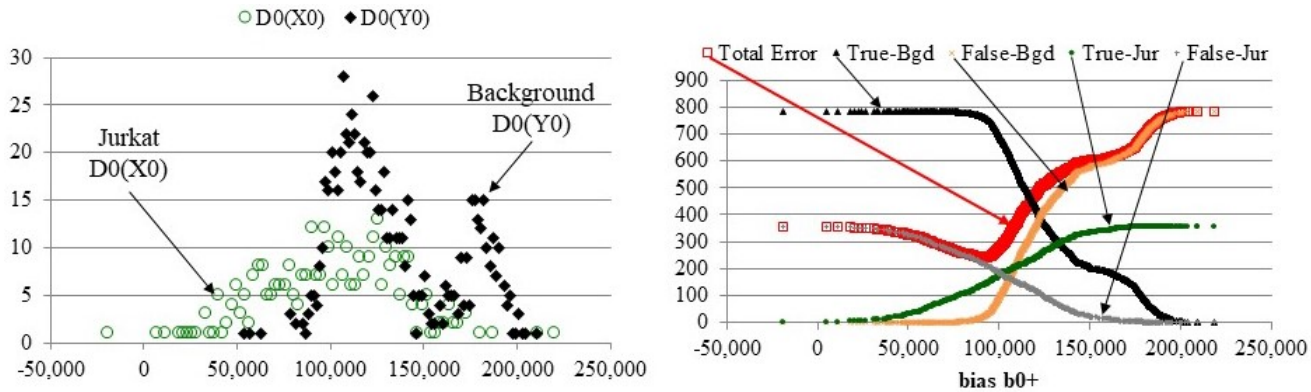

Figure 4. [left] dual score histograms for training sample pair: Jurkat cell $D0(X0)$ and background $D0(Y0)$; [right] corresponding binary decision error distributions.

It is simpler to make a binary decision using a nonparametric transformation of training samples to dual score distributions. The problem of poor discrimination associated with linear methods is addressed here by a

hierarchy of methods summarized as: a minimum error threshold-biases is derived, an iterative and convergent sieve process generates a binary decision ensemble, and a generalized accuracy measures cell spectra composition.

An example training pair comprises a Jurkat cancer cell ($X0$) and background sample ($Y0$), Fig. 3. An early method to distinguish two populations maximizes the ratio of mean separation to joint variance.[8] A scale vector $\lambda_{AB} = [\lambda_1, \cdots \lambda_p]_{AB}$ is sought to discriminate a training sample pair $\hat{\mathbf{x}}_\mathbf{A} = [\hat{x}_1, \cdots \hat{x}_p]_A$ and $\hat{\mathbf{x}}_\mathbf{B} = [\hat{x}_1, \cdots \hat{x}_p]_B$. A transformation is derived by a linearization condition to increase statistical separation, which produces

$$\lambda_{AB} = \mathbb{C}^{-1} < \hat{\mathbf{x}}_\mathbf{A} - \hat{\mathbf{x}}_\mathbf{B} > \tag{3}$$

with scale vector (λ_{AB}) components the linear discriminants (λ_j); the covariance matrix $\mathbb{C} = [C_{jk}] = (< \delta x_j \delta x_k >_A + < \delta x_j \delta x_k >_B)$ is averaged over a pooled set of training sample pairs (A) and (B) for within-class variations ($\hat{\delta x}_j = \hat{x}_j - < x_j >_T$) with class ($T = A$ or B). Covariance matrix stability provides a condition between measurement dimension (p) and sample size (n), Appendix A. If sampling is sparse, dimensional reduction may be required. A challenge is that insufficient dimensional reduction can generate an ill-conditioned transformation, while overcompensation diminishes information. For the cases discussed here, cell spectra dimensional reduction for measured intensity vectors with ($p = 12$) components and sample sizes (n) > 50 is not required.

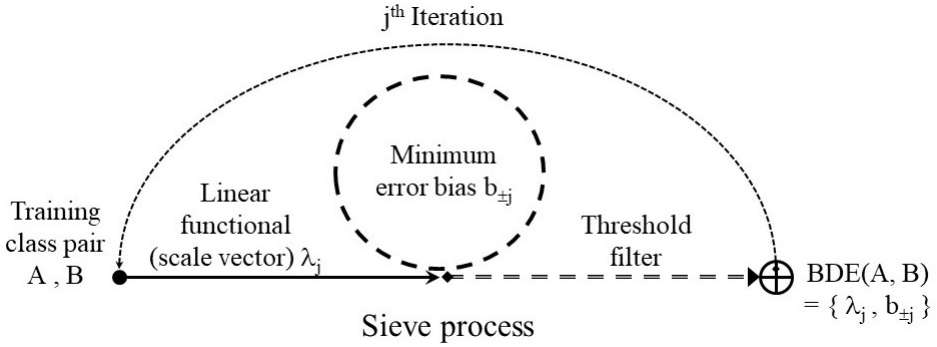

Figure 5. Sieve transforms a training sample pair to a binary decision ensemble (BDE).

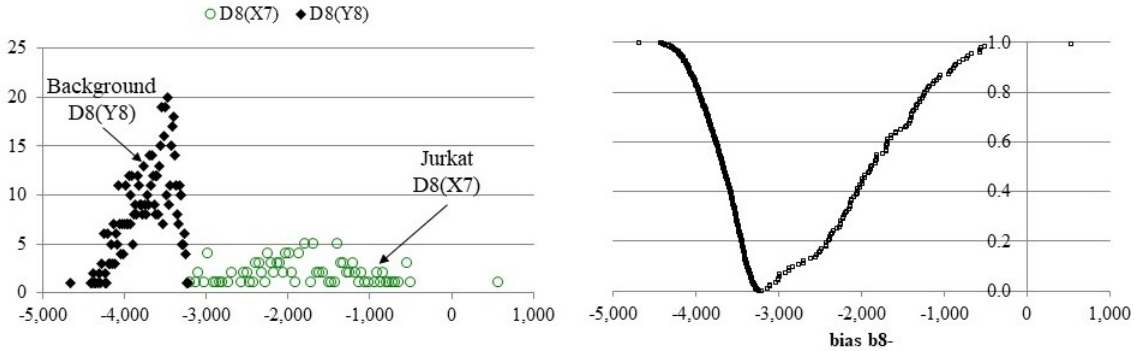

Figure 6. (left) a disjoint dual score histogram and (right) normalized binary decision error distribution.

A training sample pair ($X0, Y0$), Fig. 3 [*right*] generates a scale vector (Eq. 3), which transforms cell spectra ($\hat{\mathbf{x}}$) to a dual score distribution (Eq. 2). Fig. 4 [*left*] shows transformed distributions $D_0(\hat{\mathbf{x}}_\mathbf{A}) = \lambda_{AB} \cdot \hat{X0}$ and $D_0(\hat{\mathbf{x}}_\mathbf{B}) = \lambda_{AB} \cdot \hat{Y0}$. A nonparametric binary decision uses a threshold-bias and scale vector derived solely from the training sample pair. A training class binary decision may have two possible outcomes, either a true or a false value. To accomodate this uncertainty, error distributions of both types are evaluated and the one with fewest false decisions generates a minimum total error threshold-bias. Total error and false-condition distributions are related by:

$$\epsilon(b_\pm) = n[FalseA(b_\pm)] + n[FalseB(b_\pm)] = n(A : b\ ;\ b_\pm) + n(B : a\ ;\ b_\pm),$$

where $n(A:b\ ;\ b_{\pm})$ and $n(B:a\ ;\ b_{\pm})$ are numbers of false decision at threshold-bias ($b_{\pm}$), Fig. 4 (*right*).

An iterative sieve process filters false spectra to improve discrimination, Fig. 5. The residual set of true training class members are input to a next sieve iteration: $\{\hat{x} \in A : a\} = A_1 \subseteq A_0$ and $\{\hat{x} \in B : b\} = B_1 \subseteq B_0$. Sieving is terminated with disjoint training class subset dual score distributions $\{D(\hat{x} \in A : a)\} \cap \{D(\hat{x} \in B : b)\} = \emptyset$, Fig. 6 (left). A sample pair (A, B) is transformed with j sieve iterations to a binary decision ensemble $BDE(A, B) = \{\lambda_j, b_{\pm j}\}$ of scale vectors and minimum error biases. Sample size can bias a decision. Here both large and small sets are assumed equally important so a size bias correction (normalization) is made. A training class sample-A spectrum is assigned a weight $1/n(A)$, resulting in a normalized total error distribution, Fig. 6 (*right*).

4. SIGNAL PROCESSING

Spectral image signal processing and cell classification share a common problem where distributions may not be sufficiently distinct. In signal processing, discrimination filters background regions of noninterest prior to rendering spectral objects, (cell pixel-spectra). Background from a brightfield (BF) or darkfield(DF) is discriminated by a training sample pair comprised of one or more cells and regions of non-interest. Subsequently, cell spectral images are rendered. For classification, the problem is to distinguish one cell type from another. While the two applications differ in the training samples input and use of the output, they both employ the sieve to transform a training sample pair to a binary decision ensemble (BDE), Fig. 5.

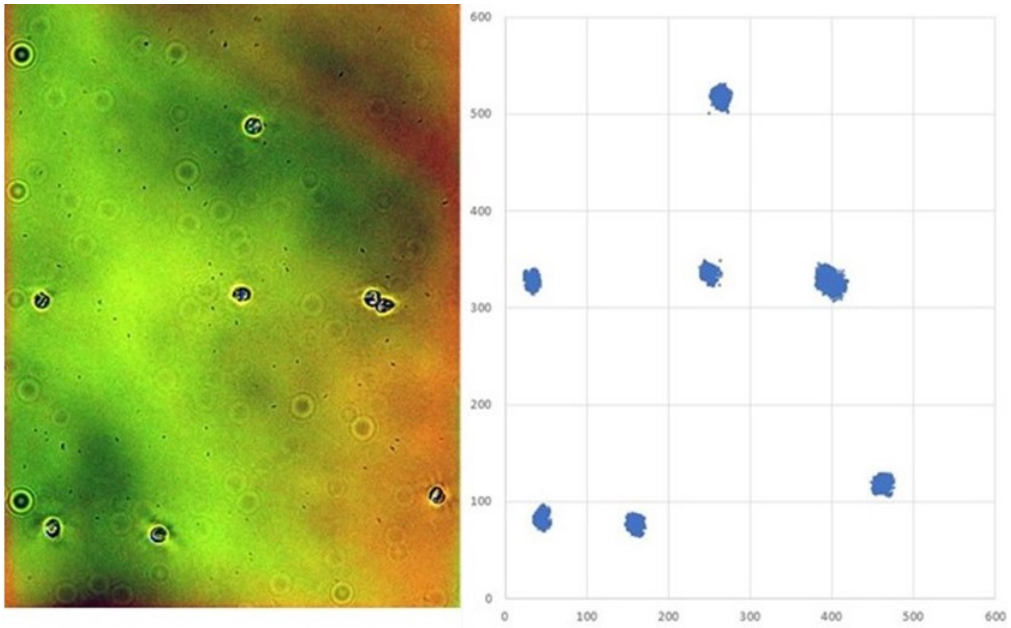

Figure 7. (left) Bright field-of-view RGB-image and (right) rendered Jurkat cells

Signal processing includes system calibrations, chromatic distortion corrections,[4] background discrimination, and spectral image cell rendering. An example of the latter two are shown in Fig. 7. A training sample pair taken from a bright field-of-view Fig. 3 (right) is transformed to a disjoint pair of distributions, Fig. 6 (left). The field-of-view (FOV) Fig. 7 (left) is discriminated by filtering spectra associated with background binary decisions. The rendered Jurkat cells are shown in Fig. 7 (right).

5. CELL CLASSIFICATION

For classification, training cells are sampled from a known, discriminated and rendered test population, e.g. a cell pair $Jurkat(J1c)$ and $Wbc(W1c)$ Fig. 8 (left). The sieve generates $BDE(J1c, W1c)$ and disjoint training class subsamples after eight iterations, Fig. 8 (right.)

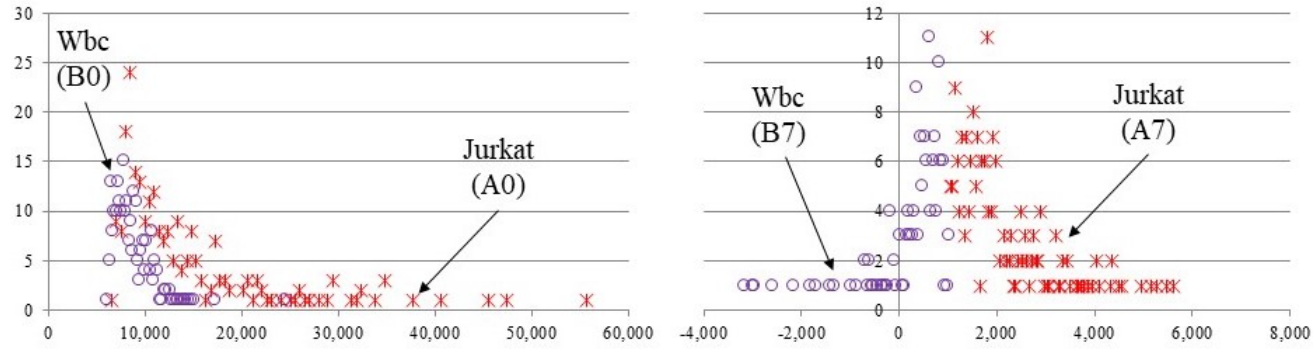

Figure 8. Dark field histograms: (left) input training cells Jur ($J1c$) and Wbc ($W1c$) and (right) sieved dual scores.

5.1 Cell partitions

Table 1. Jurkat training cell tri-bin partitions by BDE($J1e,\ W1c$)

Jurkat Training cell			Wbc Training cell		
J1c, jur	194	82%	W1c, jur	67	28%
J1c, wbc	12	5%	W1c, wbc	147	62%
J1c, xl	31	13%	W1c, xl	24	10%
S(J1c)	237	100%	S(W1c)	238	100%

A random cell's spectra is partitioned by the $BDE(A,\ B)$ of training class sample pair $(A,\ B)$. A sequence of binary decisions is generated per spectrum, e.g. $a(b)\cdots a(b)$ where $a(b)$ is a or b. A sequence is binned by: (i) if decisions are unanimous, spectrum $(\pi_i; a\cdots a)$ is assigned to bin $(xObj; a)$; or (ii) $(\pi_i; b\cdots b)$ is assigned to bin $(xObj; b)$; otherwise if decisions are mixed, (iii) pixel-spectrum $(\pi_i; mix)$ is assigned to bin $(xObj; x)$. A tri-bin partition of $xObj = \{\pi_i;$ t-sequence$\}$ is a union of three subsets: $xObj = (xObj; a \cup xObj; b \cup xObj; x)$. TABLE 1 shows training class cell partitions generated by their own BDE($J1e,\ W1c$). Relative bin size is the ratio of bin members to sample size and is another adjustment for size bias.

5.2 Partition Proximity Measure

Table 2. Training class samples ($J1c$) and ($W1c$) partition proximity.

Partition χ^2 proximity	
(J1c - W1c) : jur	62
(J1c - W1c) : wbc	114
(J1c - W1c) : x	1
χ2 (J1c - W1c)	177

A χ^2-statistic (Eq. 4) measures a relative difference (proximity) between two cell partitions A and B having k bins and A_i and B_i sizes of respective ith bin[12]

$$\chi^2(A-B) = \sum_{i=1}^{k} (\sqrt{B/A}A_i - \sqrt{A/B}B_i)^2/(A_i + B_i). \tag{4}$$

Other goodness-of-fit statistics for discrete multivariate data may yield better results, e.g. if bin sizes are small.[13] The relative proximity of training class samples ($J1c$) and ($W1c$) is shown in Table 2.

5.3 Learning and cell category accuracy

Learning may improve training accuracy by resampling a known test population. Many algorithms focus on binary decision accuracy, but this is not necessarily the only, or preferred, way to accurate cell classification. Similar binary decisions may result from intra-cellular light scattering characteristics common to other classes.

Class knowledge is generalized from vectors (spectra) and binary decisions to objects (cells) and categories. A category is defined by test cells relative proximity to a training class sample pair. A test object ($tObj$) relatively closer to training class sample $S(A)$ than to $S(B)$ is categorized ($tObj : cA$). The cell category condition can be stated : if $\Delta\chi^2[tObj|S(A) - S(B)] > 0$, then $(tObj : cB)$, else $(tObj : cA)$, where

$$\Delta\chi^2[tObj|S(A) - S(B)] = \chi^2[tObj - S(A)] - \chi^2[tObj - S(B)]. \quad (5)$$

A special case is if a test cell is a training class sample: if $tObj = S(A)$, $\Delta\chi^2[S(A)|S(A)-S(B)] < 0 \Rightarrow S(A) : cA$. Thus, a training class sample, if used as a test cell, is consistently associated with its own category. An object represented by a set of vectors has been considered using kernel methods and distribution modeling, however only the special case of an object containing a single vector was developed for monochromatic images or handwriting recognition.[14] A broader interpretation of objects beyond sets of vectors has been used with kernel methods and applied to video sequences, but did not develop a dimensional reduction or error analysis.[15]

Before, an equivalence relation compared training class and binary decision per spectrum. A test population characterized as known meant the equivalence relation was evaluated as true or false per measurement vector (cell spectrum). Now, a generalized equivalence relation compares training class and category per spectral object (cell). The logical equivalence between test class and category is evaluated: if in closer proximity to their associated training class, known test cells are categorized true. For example, if a class A test object ($\mathbf{t\hat{O}} \in A$) is in closer proximity to training sample A than B, its assignment to category ($\mathbf{t\hat{O}} : cA$) is true. A generalized category accuracy is the ratio of cells categorized true to total test population. If category accuracy is acceptable, the BDE is used to classify random cells. Otherwise, learning is executed with another training sample selected from a test population.

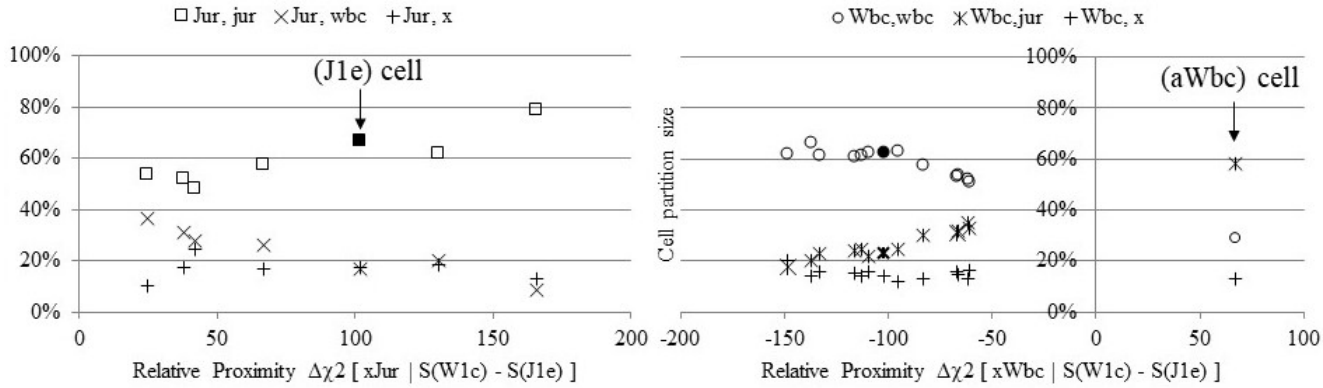

Figure 9. Test populations categorized by training class samples: S(J1e) and S(W1c): (left) Jurkat and (right) Wbc cells.

A learning example uses training class samples Jurkat $S(J1e)$ and Wbc $S(W1c)$. Relative proximity to the pair is plotted as the abscissa in Fig. 9 for test cells Jurkat $\{tJur\}$ (left) and Wbc $\{tWbc\}$ (right). The training class cell $S(J1e)$ [*solid box*, *Fig.* 9 (*left*)] is in relative proximity $\chi^2[S(J1e) - S(W1c)] = 102$ to training class cell $S(W1c)$ [*bold circle*, *Fig.* 9 (*right*)]. Accurate Jurkat test categorization ($\Delta\chi^2 > 0$) occurs with all cells closer to training class sample $S(J1e)$.

A need for learning is shown in Fig. 9 (right). The Wbc test population $\{xWbc\}$ is accurately categorized except for a cell located on an image boundary with a relative proximity $\Delta\chi^2(aWbc|S(W1c) - S(J1e)] = 67$. Boundary objects can be excluded, however the cell fragment is retained as a marker. Various learning methods can be used to sample a test population. If there are N A-class test objects and M B-class test objects, one approach is to evaluate all $N \cdot M$ pairs. The category error in Fig. 9 (*right*) is eliminated by selecting Jurkat test sample cell ($J1c$) located at $\Delta\chi^2[J1c|S(W1c) - S(J1e)] = 25$, Fig.9 (left).

5.4 Binary class regions

Binary class regions are defined by two test populations. A region is determined by test cells relative partition size and proximity to the training class pair. Unknown cells $\{xCell\}$ are classified by comparing random cell partitions size and relative location to the binary class regions. However as no truth condition can be evaluated to measure category accuracy, a cell may be classified, but with an accuracy evaluated only a posteriori.

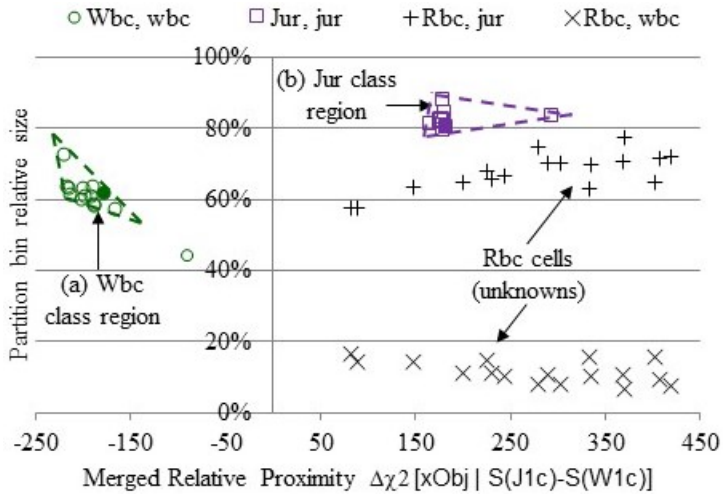

Figure 10. Class regions: (a) Wbc and (b) Jurkat. Rbc cells serve as unknowns.

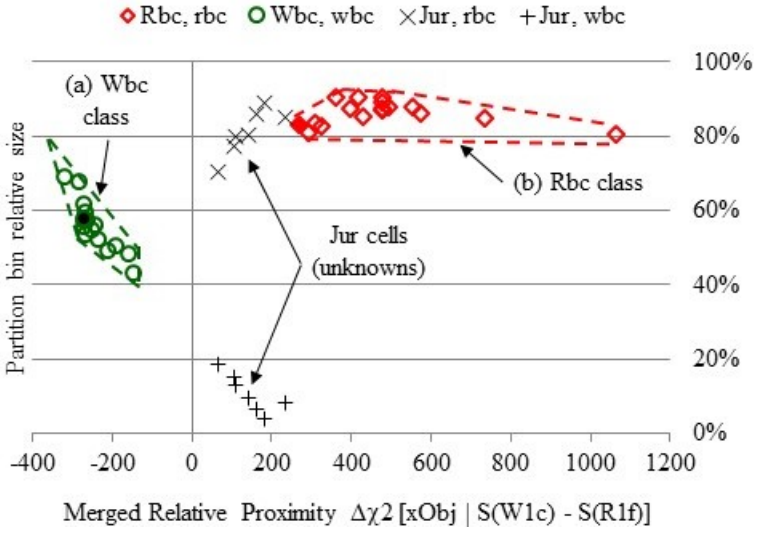

Figure 11. Class regions: (a) Wbc and (b) Rbc. Jur cells serve as unknowns.

Test sample populations bounded by convex hulls (dashed lines) define class regions (a) Wbc and (b) Jur, Fig. 10. The marker boundary cell in Fig. 9 (right) is located at $\Delta\chi^2 = -90$ outside the Wbc class region. The unknown objects are located in a merged relative proximity view where partition difference is referenced to the abscissa origin, e.g. training class samples are located at: x = +177 (J1c) and x = -177 (W1c). Rbc cells are measured outside the class regions and thus are classified ($Rbc : X$), an unknown class, (: X = [*Not* Wbc and *Not* Jurkat]).

Class regions (a) Wbc and (b) Rbc are defined by respective test populations and training samples ($R1f$, dark diamond) and ($W1c$, dark circle) with Jurkat cells $\{xJur\}$ the unknown population, Fig. 11 (left). Figure 12 plots

class regions derived from Jurkat $t(J1e)$ and Rbc $t(R1f)$ test cells. The relative proximity $\chi^2[S(J1e)-S(R1f)] = 259$ corresponds with training class cells located at: $(J1e) = -259$ (solid purple square) and $(R1f) = +259$ (solid red diamond). Boundary cells and Rbc cell clusters are excluded from the cell region in this case. In Figs 10 - 12 unknown random object populations are located exterior to class regions generated by three training class sample pairs and corresponding test populations. The cells serving as unknowns are distinct from the test populations, which may be inferred as an accurate classification a posteriori.

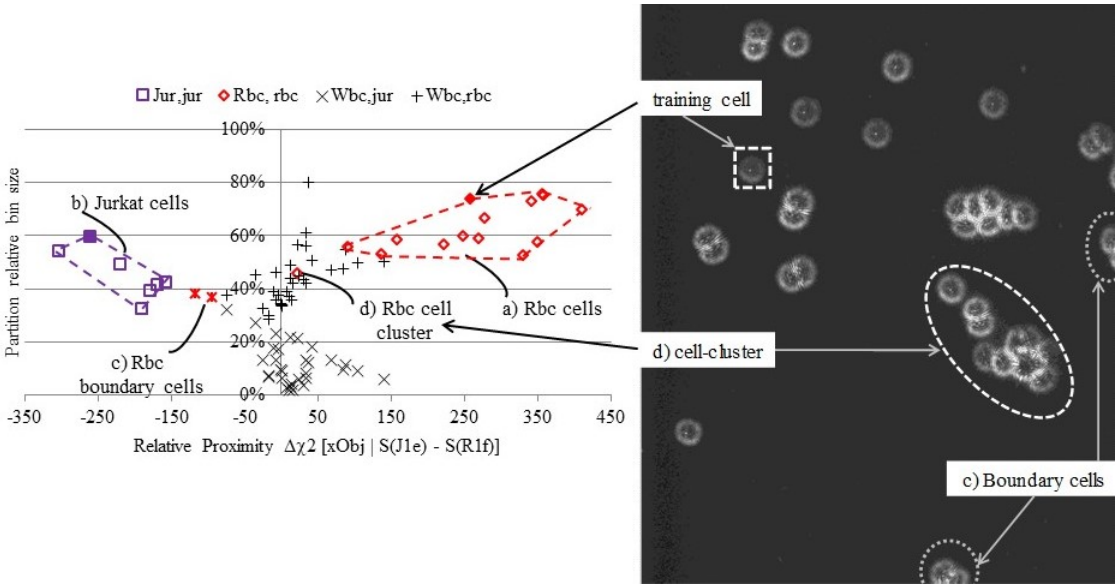

Figure 12. (left) Jurkat and Rbc class regions, (right) Rbc cells in a dark field-of-view.

5.5 Multi-classification

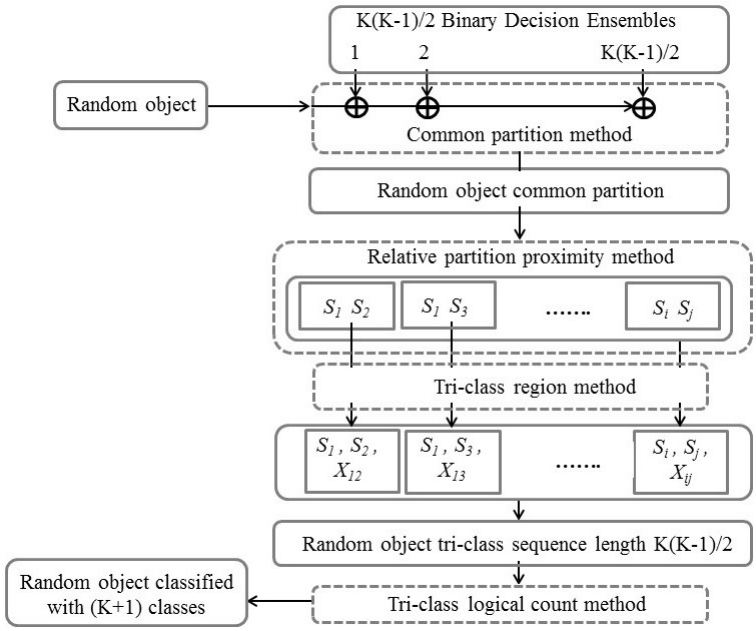

Figure 13. Spectral object multi-classification with a common partition.

It can be useful to have many training class samples. Hundreds of clusters of differention exist for leukocytes[16] and there are a number of medical imaging applications.[17] Availability of multiple cell types can generate new insights and greater accuracy.

Table 3. Pixel spectra common blocks for 3 Classes and 3 Binary Decisions

Cell pixel-spectrum	S(A)^S(B)	S(B)^S(C)	S(C)^S(A)	Common block count
	Binary decision tri-bin outcomes			
π_j	a	x2	a	2 votes for block a
π_k	b	b	c	2 votes for block b
π_n	x1	c	c	2 votes for block c
π_l	a	b	x3	1 vote for block "mix"
π_m	x1	x2	x3	1 vote for block "other"

Spectral object multi-classification is scalable to K cell samples, Fig. 13. A random cell common partition is generated by $\binom{\mathbf{k}}{\mathbf{2}}$ BDE's, which generate per spectrum a tri-bin sequence of length $K(K-1)/2$. Table 3 shows example block assignments where a tri-bin a or b is associated with unanimous binary decisions and tri-bin (x) = $(not\ a)$ and $(not\ b)$ outcome. A vote counts number of times a tri-bin occurs per spectrum, which is assigned a common block according to highest number of votes. There are at most $(K-1)$ tri-bin votes per spectrum. For vector (π_j) a unanimous BDE outcome of tri-bin partition (a) occurs twice and so provides two votes for common block (a). A single vote block $(x1+x2+x3)$ identifies member vectors not having any unanimous binary decisions. A $(1vote.mix)$ partition block contains other single decision combinations. A common partition block size is the number of referenced cell spectra.

Table 4. Cell partitions by binary decision ensemble (columns) and training sample (rows).

	S(J1c) ^ S(W1c)			**S(R1f) ^ S(W1c)**			**S(R1f) ^ S(J1e)**		
	(xRbc:X)			(xRbc:Rbc)			(xRbc:Rbc)		
	tRbc, jur	340	58%	tRbc, x2	52	9%	tRbc, jur	45	8%
	tRbc, wbc	84	14%	tRbc, wbc	49	8%	tRbc,rbc	435	74%
	tRbc, x1	165	28%	tRbc, rbc	488	83%	tRbc, x3	109	19%
tRbc = R1f	R1f	589	100%	R1f	589	100%	R1f	589	100%
	(xJur:Jur)			(xJur:X)			(xJur:Jur)		
	tJur, jur	351	81%	tJur, x2	45	10%	tJur, jur	208	48%
	tJur, wbc	16	4%	tJur, wbc	59	14%	tJur, rbc	96	22%
	tJur, x1	64	15%	tJur, rbc	327	76%	tJur, x3	127	29%
tJur = J1c + J1e	J1c + J1e	431	100%	J1c + J1e	431	100%	J1c + J1e	431	100%
	(xWbc:Wbc)			(xWbc:Wbc)			(xWbc:X)		
	W1c, jur	67	28%	W1c, x2	14	6%	xW1c, jur	42	18%
	W1c, wbc	147	62%	W1c, wbc	147	62%	xW1c, rbc	92	39%
	W1c, x1	24	10%	W1c, rbc	77	32%	xW1c, x3	104	44%
tWbc = W1c	W1c	238	100%	W1c	238	100%	W1c	238	100%

A random cell is located in a tri-class region. Two regions are defined by test cell common partitions relative proximity and size to a training class sample common partition pair, and a third unknown class region by a $[(not\ A)$ and $(not\ B)]$ region outside the first two. As a result three tri-class regions are generated per BDE. A cell located in a tri-class region $\binom{\mathbf{k}}{\mathbf{2}}$ times results in a tri-class sequence of length $K(K-1)/2$. A cell is assigned at least one of $(K+1)$ classes by counting a tri-class occurrence frequency.

As an example, the three training samples of Table 4 are employed with a fourth adenocarcinoma population $\{xH1975\}$ serving as unknowns. A common partition relative size and proximity to training class sample common partition pairs is measured with Eq.(4). Table 5 shows the three training class cell common partitions and their

relative proximity. Fig. 14 is a metric view of training class pair separations (in-plane) and relative bin size (out-of-plane axis-a). Common partitions tend to increase training sample pair separation, e.g. relative proximity differences 177, 269, and 259 increase to 266, 318, and 302 respectively.

Table 5. Training class common partitions and their relative proximity.

Training class common partitions				Partition differences $\chi 2$		
		%	Group	S(Rbc)	S(Jur)	S(Wbc)
Rbc, 2vote.jur	34	6%		0	175	14
Rbc, 2vote.wbc	31	5%		0	6	210
Rbc, 2vote.rbc	371	63%	U(Rbc)	0	111	92
Rbc, 1vote.mix	149	25%		0	6	3
Rbc x1+x2+x3	4	1%		0	3	0
n(R1f)	589	100%		0	302	318
Jur, 2vote.jur	197	46%	U(Jur)	175	0	45
Jur, 2vote.wbc	9	2%		6	0	202
Jur, 2vote.rbc	80	19%		111	0	4
Jur, 1 vote.mix	145	34%		6	0	11
Jur x1+x2+x3		0%		3	0	4
n(J1c+J1e)	431	100%		302	0	266
Wbc, 2vote.jur	33	14%		14	45	0
Wbc, 2vote.wbc	129	54%	U(Wbc)	210	202	0
Wbc, 2vote.rbc	28	12%		92	4	0
Wbc, 1vote.mix	46	19%		3	11	0
Wbc, x1+x2+x3	2	1%		0	4	0
n(W1c)	238	100%		318	266	0

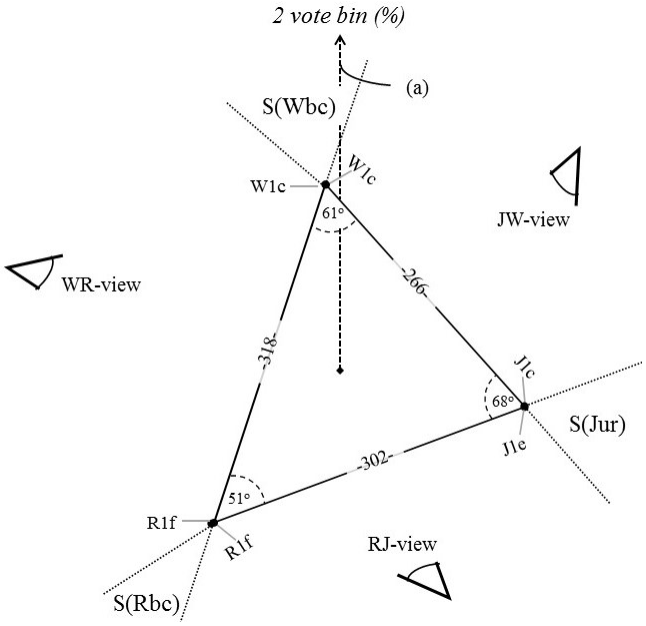

Figure 14. Training sample common partition proximity for (K=3) classes.

5.6 A new view of cell structure

Cell structure is characterized by a combination of spatial and spectral features produced from elastically scattered visible light. Spectral image microscopy of an unstained red blood cell (Rbc) superposes phase object characteristics on the Gaussian image, Fig. 15. Rayleigh resolution is $0.44 \sim 0.54$ $[\mu m]$ at 440 $[nm]$ wavelength

for dark field apodization and bright field imaging respectively.[18] The 6.8 [μm] Rbc cell diameter bounded by the box is estimated from the dark fringe boundary in the bright field image. The reddish interior corresponds to an order of magnitude greater molar extinction coefficient at 440 [nm] compared to 620 nm.[19] The biconcave cell results in a bright 2x2 pixel central region. Partially coherent illumination by unscattered light emitted in a relatively narrow 6^o emission semi-angle from the thin central region may explain the nature of the interference rings. The diffraction rings extend to ~ 11.2 [μm] diameter at $\lambda = 440$ [nm].

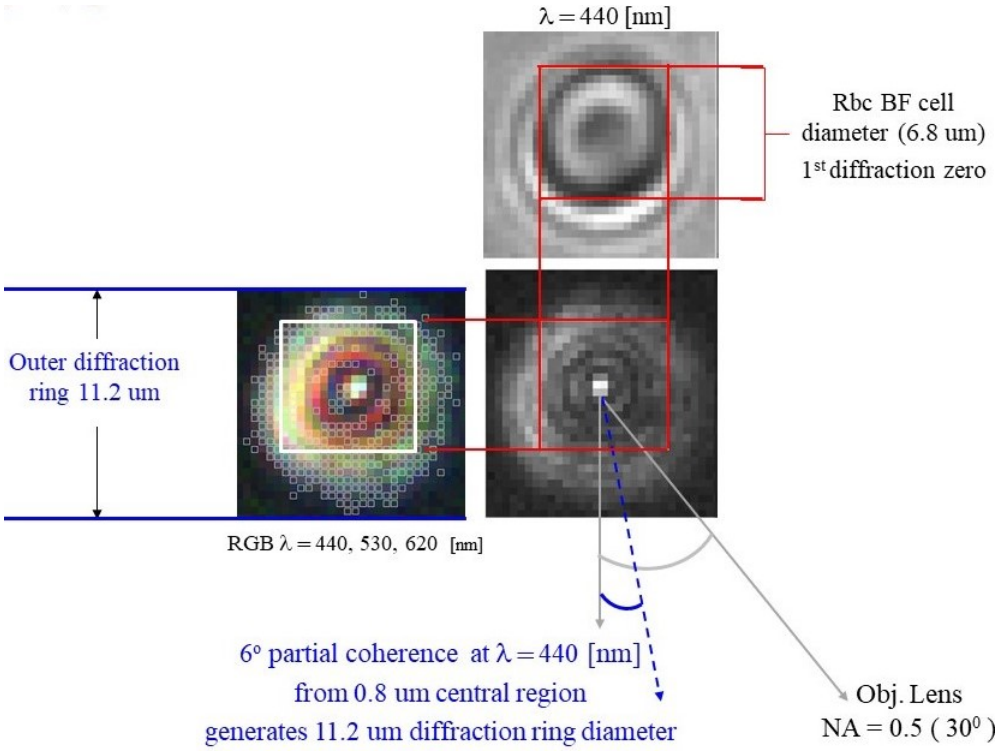

Figure 15. Rbc cell spectral image features, 0.4 [$\mu m/pixel$])

Unique spectra determined from the cell partitions reveal a characteristic spatial distribution for each class. Figure 16 shows unique intra-cellular spectra (white boxes) labeled U(Rbc), U(Jur), and U(Wbc) in TABLE 5 associated with unanimous decisions for a respective class.

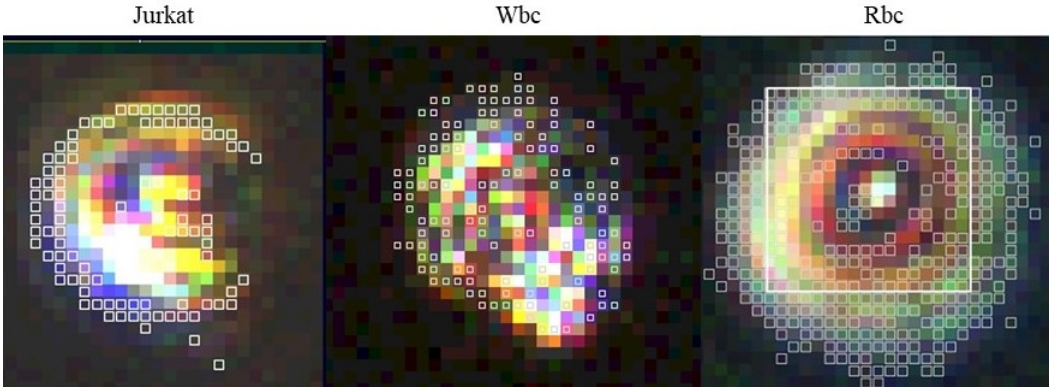

Figure 16. Unique pixel spectra: (left) Jurkat (J1C); (middle) Wbc (W1C); (right) Rbc (R1f) with large 6.8μm white square enclosing the first diffraction zero. 1 pixel length= 0.4 [μm]. RGB images at 620,545,440 [nm].

5.7 H1975 as an unknown cell type

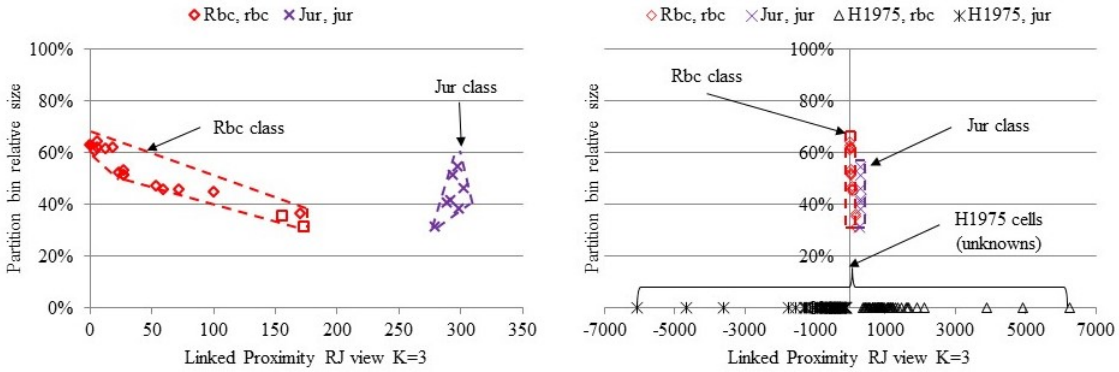

Figure 17. Rbc and Jur test cells and H1975 cancer as unknown cells.

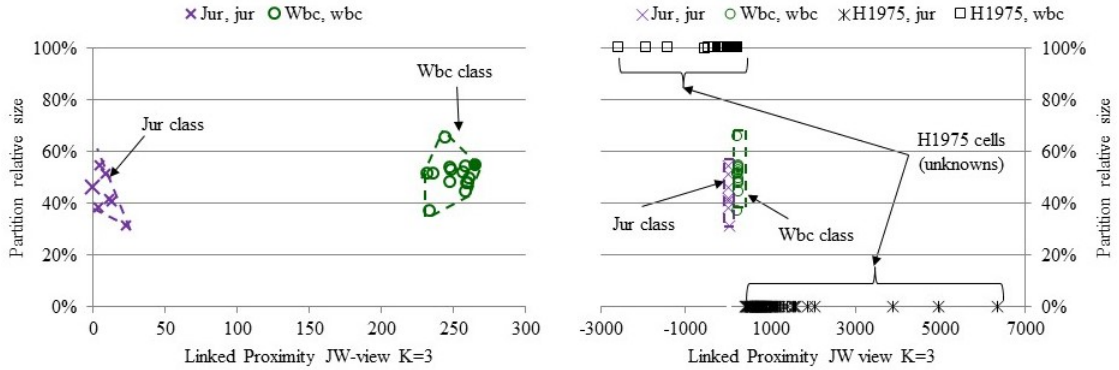

Figure 18. Wbc and Jur test cells and H1975 cancer as unknown cells.

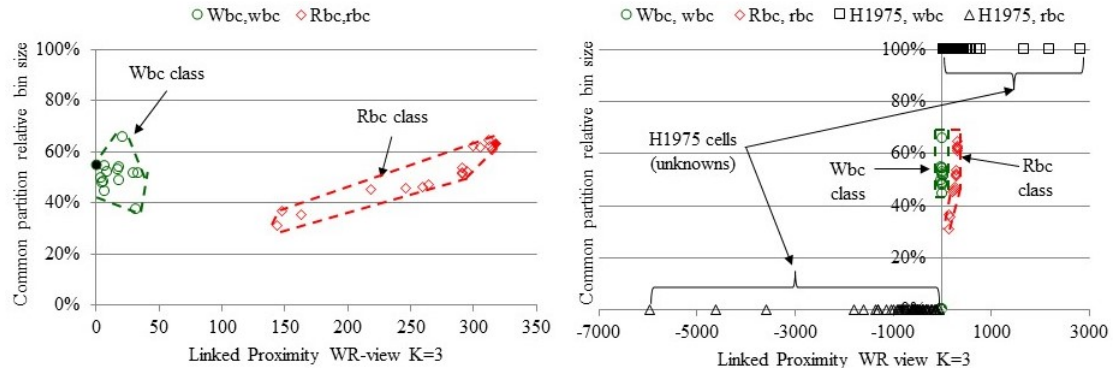

Figure 19. Wbc and Rbc test cells and H1975 cancer as unknown cells.

Unknown cells can be classified, albeit with an accuracy inferred a posteriori. As an example, non-small cell lung cancer adenocarcinoma cells (H1975)[20] are introduced as unknown spectral objects and compared to the Table 5 training class samples. A linked proximity view provides a convenient alternative to display multiple binary test classification regions. Training class samples $S(A)$ and $S(B)$ and a random object $xObj$ with bins a and b are related by: $\chi^2[xObj - S(A)]$ for $xObj, bin = a$ and $\chi^2[S(A) - S(B)] - \chi^2[xObj - S(B)]$ for $xObj, bin = b$. A linked *AB-view* displays a training class sample A located at the x-axis (abscissa) origin and training class

sample B is plotted as a relative proximity to A. H1975 lung cancer cells, serving as unknowns, are widely separated from the Rbc, Wbc, and Jur test populations, (right): Figs. 17 - 19. Note the expanded abscissa scales required with H1975 cells as a result of relatively distant proximity compared to training class sample pairs.

6. CONCLUSION

Spectral object classification uniquely characterizes heterogeneous cell structure and is scalable to large numbers of types. Small samples of erythrocytes, leukocytes, and Jurkat and H1975-adenocarcinoma have been accurately classified at 0.4 $[\mu m]$ resolution with unique pixel-spectra identified. Spectral image microscopy with visible light elastic scattering makes feasible label-free blood and cancer cell identification.

APPENDIX A. COVARIANCE MATRIX STABILITY CONDITION

A matrix stability condition is developed to determine a class sample size dependent optimal dimensionality for binary decisions, which use the Eq. 3 linear functional and covariance matrix. LU decomposition is a standard method to determine the existence of a matrix inverse, [12, pp.48-54]. If an inverse exists, it is nonsingular and has a nonzero determinant. If a matrix is singular and LU decomposition fails other approaches must be used. Singularity resolution involves handling near rank deficiency and ill-conditioned behavior as part of effective dimensional reduction with adequate tolerancing of numerical errors.

Singular value decomposition (SVD) is a technique to diagnose a matrix singularity [12, pp.65-69]. Scope is reduced when evaluating a symmetric p-by-p covariance matrix $\mathbb{C} = \mathbb{C}^T$ with transpose $\mathbb{C}^T$. The singular value decomposition (SVD) of a covariance matrix is $\mathbb{C} = \mathbb{U}\Sigma\mathbb{V}^T$ with orthogonal matrices $\mathbb{U} = [u_1, \cdots, u_p]$ and $\mathbb{V} = [v_1, \cdots, v_p] \in \mathbb{R}^{pxp}$; u_i and v_i ith left and right singular vectors, and the diagonal p-by-p matrix $\Sigma = diag(\sigma_1, \cdots \sigma_p)$ contains singular values [21, pg.70]. The inverse is $\mathbb{C}^{-1} = \mathbb{V}[diag(1/\sigma_1, \cdots 1/\sigma_p]\mathbb{U}^T$ [12, pp.69].

A condition number specified as

$$\kappa(\mathbb{C}) = \|\mathbb{C}\|\|\mathbb{C}^{-1}\| \geq \|\mathbb{C}\mathbb{C}^{-1}\| = 1 \tag{6}$$

uses a matrix norm, which satisfies the submultiplicative property $\|\mathbb{AB}\| \leq \|\mathbb{A}\|\|\mathbb{B}\|$. A convention is that $\kappa(\mathbb{C}) = \infty$ for singular $\mathbb{C}$. If $\kappa(\mathbb{C})$ is large, then $\mathbb{C}$ is ill-conditioned. Frobenius and p-norms satisfy this inequality and other bounds, which is convenient as the norms are then equivalent, [21, pp. 56-57]. As a result a condition number is the ratio of largest to smallest singular value, [12, pp. 69, SVD of a Square Matrix].

SVD space decomposition is $Rank(\mathbb{C}) + Nullity(\mathbb{C}) = p$ where $Nullity(\mathbb{C}) = nullspace$ dimension. If $r =$ rank of $\mathbb{C} = dim(range(\mathbb{C})) < p$ then $\mathbb{C}$ is rank deficient. A rank deficient singular matrix has null space components equal to zero or numerically so small, values may be dominated by roundoff error. To avoid ill-conditioned behavior one approach solves a pseudoinverse

$$\mathbb{C}^{-1} = \mathbb{V}[diag(1/\sigma_1, \cdots 1/\sigma_r, 0, \cdots, 0)]\mathbb{U}^T \tag{7}$$

for $\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots \sigma_p = 0$ [21, pg.257]. A corresponding vector with reduced dimensionality may then be calculated, [12, pp.69-70]. However there may be numerical difficulties where rounding errors or fuzzy data can make rank determination a nontrivial exercise, [21, pg.72].

A difference between SVD matrix factorization and eigendecomposition disappears if a symmetric covariance matrix is used, [12, pp. 569-570]. An ordered set of singular values σ_r may then be handled similarly to the eigenvalues. Computing eigenvalues of a general m^{th} order matrix typically is an iterative process, since Galois theory and the Abel-Ruffini theorem state the general polynomial for degree $m \geq 5$ is not solvable by radicals, as there is no algebraic solution to the characteristic polynomial, [22, theorem 16, pg. 299] and [23, pp. 37-38]. A Jacobi method diagonalizes a covariance matrix by a sequence of orthogonal similarity transforms. A feature is superior accuracy, [23, pp. 43] compared to faster strategies, e.g. Givens and Householder methods [12, pp.570-589]. However, the termination of an iterative method depends on numerical tolerances and roundoff errors. One approach is to use perturbation theory to study eigenvalue sensitivity, [21, pg.320].

Alternatively, a relatively simple method is developed here based on covariance stability. A relationship between sample size and dimension is sought that ensures a well conditioned covariance matrix. An error tolerance helps resolve numerical difficulties inherent with iterative SVD and eigendecomposition techniques. Computational precision sets a lower bound on a tolerance δ where rank is determined from a spectrum of principal values contained in Σ , or eigenvalues if that approach is taken, such that

$$\sigma_1 \geq \cdots \geq \sigma_{r-1} > \delta_r > \sigma_{r+1} = \cdots = \sigma_m = 0. \tag{8}$$

The problem may then appear in two forms. If a tolerance is set too large so that rank is reduced, then useful results may be lost. Alternatively, if rank is too large, a singularity may be removed but the covariance inverse remains ill-conditioned. From inequalities (8) a covariance condition number is

$$\kappa(\mathbb{C}) \cong \sigma_1 \ /\delta_r. \tag{9}$$

Using Eq. 7 the covariance pseudo-inverse condition number is

$$\kappa(\mathbb{C}^{-1}) \cong (1 \ /\delta_r^{'}) \ /(1 \ /\sigma_1) = \sigma_1 \ /\delta_r^{'}. \tag{10}$$

The two tolerances δ_r and $\delta_r^{'}$ may differ as a result of numerical noise and iteration method. Dimensionality or sample size may be adjusted using the condition number as a guide. Unfortunately there is little correlation between the determinant of a matrix ($det\mathbb{A}$) and its condition number, [21, pg.82, Determinants and Nearness to Singularity]. As a result, at least in part, condition estimation evolved as a collection of refined algorithms based on norm bounds, [21, pp. 128-132]. However these may be relatively complex to implement and a simpler method is given to determine a well conditioned covariance inverse with maximum rank, which include the tolerances in Eqs. 9, 10. Using a tolerance δ_r as an estimate of the smallest non-zero value with $r =$ rank of $\mathbb{C}$, the covariance determinant is a product of diagonal components,

$$det\ \mathbb{C} \cong \sigma_1 \cdots \sigma_{r-1}\delta_r. \tag{11}$$

The determinant of the covariance inverse is

$$det\ \mathbb{C}^{-1} \cong 1 \ /\delta_r^{'} \cdot 1/\sigma_{r-1} \cdots 1 \ /\sigma_1. \tag{12}$$

A relation between the covariance and its inverse in terms of determinant and condition is given by combining Eqs. 9 - 12

$$det\ \mathbb{C} \cdot det\ \mathbb{C}^{-1} \cong \kappa(\mathbb{C}^{-1}) \ /\kappa(\mathbb{C}) \cong \delta_r \ /\delta_r^{'}. \tag{13}$$

A norm-independent and computationally convenient covariance stability condition can be specified to first order in the identities $\mathbb{C} \cdot \mathbb{C}^{-1} = 1$ and $det\mathbb{A} \cdot det\mathbb{B} = det(\mathbb{A} \cdot \mathbb{B})$, by an error tolerance ϵ_c,

$$det\ \mathbb{C} \cdot det\ \mathbb{C}^{-1} \cong 1 \pm \epsilon_c. \tag{14}$$

The maximum rank may then be determined by evaluating Eq.14 for a covariance of a given dimension. A dimensional bound is set by a tolerance ϵ_c with determinants evaluated as a product of singular values or eigenvalues. A tolerance lower bound ϵ_{IEEE} has 15 decimal digit precision set by the IEEE 754-2008 floating point specification for 64 bit binary double precision. The tolerance ϵ_c is set above the IEEE floating point precision to avoid numerical noise, $\epsilon_{IEEE} < \epsilon_c < 1$. Comparing Eqs. 13, 14 results in $\pm\epsilon_c \cong \Delta \ /\delta$ where $\Delta = \delta_r - \delta_r^{'}$. As an example, a minimum detectable difference $\Delta \sim 10^{-14}$ and a five decimal digit tolerance $\epsilon_c \sim 10^{-5}$ implies a $\delta \sim 10^{-9}$ lower bound on the singular values.

ACKNOWLEDGMENTS

The authors wish to thank our colleagues for their optics expertise: Zach Smith, Tim Thomas, and Doug Holmgren; providing cells: Suzanne Miyamoto; and for securing cell spectral images: Tim Zhang and Latevi Lawson. This work was partially supported by the Center for Biophotonics Science and Technology Center managed by the University of California Davis under Cooperative Agreement No. PHY0120999.

REFERENCES

[1] Malik, Z., Rothmann, C., Cycowitz, T., Cycowitz, Z., and Cohen, A., "Spectral morphometric characterization of b-cll cells versus normal small lymphocytes," *J. Histochem. Cytochem* **46** Oct.(10), 1113 – 1118 (1998).

[2] Katzilakis, N., Stiakaki, E., Papadakis, A., Dimitriou, H., Stathopoulos, E., Markaki, E., Balas, C., and Kalmanti, M., "Spectral characteristics of acute lymphoblastic leukemia in childhood," *Leuk. Res* **28** Nov.(11), 1159–1164 (2004).

[3] Lim, K.-H., Salahuddin, S., Qiu, L., Fang, H., Vitkin, E., I. Ghiran, I., Modell, M., Takoudes, T., Itzkan, L., Hanlon, E., Sachs, B., and Perelman, L., "Light-scattering spectroscopy differentiates fetal from adult nucleated red blood cells: may lead to noninvasive prenatal diagnosis," *Optics Letters* **34** No. 9, May 1, 1159–1164 (2009).

[4] Gesley, M. and Puri, R., "A high throughput spectral image microscopy system," *Rev. Sci. Instruments* **89**, 01375 1–5 (2018). https://doi.org/10.1063/1.4998725.

[5] Fukunaga, K., [*Introduction to Statistical Pattern Recognition*], Academic Press, San Diego, 2nd ed. (1990).

[6] Rosenblatt, F., [*Principles of Neurodynamics*], Spartan Books, New York (1962).

[7] Minsky, M. and Papert, S., [*Perceptrons: An Introduction to Computational Geometry*], MIT Press (1988).

[8] Fisher, R., "The use of multiple measurements in taxonomic problems," *Annals of Eugenics* **7**, 179–188 (1936).

[9] Rojas, R., [*Neural Networks A Systematic Introduction*], Springer-Verlag, Berlin (1996).

[10] Hastie, T., Tibshirani, R., and Friedman, J., [*The Elements of Statistical Learning*], Springer, New York, 2nd ed. (2009).

[11] Cristianini, N. and Shawe-Taylor, J., [*Support Vector Machines*], 'Cambridge Univ. Press (2000).

[12] Press, W., Teukolsky, S., Vetterling, W., and Flannery, B., [*Numerical Recipes*], Cambridge Univ. Press, Cambridge, 3rd ed. (2007).

[13] Reed, T. and Cressie, N., [*Goodness-of-Fit Statistics for Discrete Multivariate Data*], Springer, New York (1998).

[14] Kondor, R. and Jebara, T., "A kernel between sets of vectors," *Proc. 20th Int. Conf. on Machine Learning (ICML-2003)* (2003).

[15] Vishwanathan, S. and Smola, A., "Binet-Cauchy kernels," *Intl. J. Computer Vision* **73**, 95–119 (2007).

[16] Zola, H. and Swart, B., "The human leucocyte differentiation antigens HLDA workshops: the evolving role of antibodies in research, diagnosis and therapy," *Cell Research* **15**, 691–694 (2005).

[17] Lu, G. and Fei, B., "Medical hyperspectral imaging: a review," *Journal of Biomedical Optics* **19**(1), 1–23 (2014).

[18] Born, M. and Wolf, E., [*Principles of Optics*], vol. 472, Cambridge Univ. Press, 7th ed. (1999).

[19] Prahl, S., "Optical Absorption of Hemoglobin." https://omlc.org/spectra/hemoglobin/. (Accessed: 15 November 2018).

[20] "ATCC Number: CRL-5908." http://www.atcc.org/. (Accessed: 24 November 2018).

[21] Golub, G. and Loan, C. V., [*Matrix Computations*], Johns Hopkins University Press, Baltimore, 3rd ed. (1996).

[22] Dean, R., [*Elements of Abstract Algebra*], J.Wiley and Sons, Inc., New York (1966).

[23] Golub, G. and van der Vorst, H., "Eigenvalue computation in the 20th century," *J. Computational and Applied Mathematics* **123**, 35–65 (2000).