

Using the Low-Pass Monogenic Signal Framework for Cell/Background Classification on Multiple Cell Lines in Bright-Field Microscope Images

Firas Mualla · Simon Schöll · Björn Sommerfeldt · Andreas Maier ·
Stefan Steidl · Rainer Buchholz · Joachim Hornegger

Received: 31 July 2013 / Accepted: 29 November 2013

Abstract

Purpose Several cell detection approaches which deal with bright-field microscope images utilize defocusing to increase image contrast. The latter is related to the physical light phase through the transport of intensity equation (TIE). Recently, it was shown that it is possible to approximate the solution of the TIE using a low-pass monogenic signal framework. The purpose of

this paper is to show that using the local phase of the aforementioned monogenic signal instead of the defocused image improves the cell/background classification accuracy.

Materials and methods The paper statement was tested on an image database composed of three cell lines: adherent CHO, adherent L929, and Sf21 in suspension. Local phase and local energy images were generated using the low-pass monogenic signal framework with axial-derivative images as input. Machine learning was then employed to investigate the discriminative power of the local phase. Three classifier models were utilized: random forest (RF), support vector machine (SVM) with a linear kernel, and SVM with a radial basis function (RBF) kernel.

Results The improvement, averaged over cell lines, of classifying 5 x 5 sized patches extracted from the local phase image instead of the defocused image was 7.3% using the RF, 11.6% using the linear SVM, and 10.2% when a RBF kernel was employed instead of the linear one. Furthermore, the feature images can be sorted by increasing discriminative power as follows: at-focus signal, local energy, defocused signal, local phase. The only exception to this order was the superiority of local energy over defocused signal for suspended cells.

Conclusions Local phase computed using the low-pass monogenic signal framework considerably outperforms the defocused image for the purpose of pixel-patch cell/background classification in bright-field microscopy.

Keywords Monogenic signal · Cell detection · Bright-field microscopy · Transport of intensity equation · Local phase · Machine learning

* The final publication is available at springer.com.

Firas Mualla
Pattern Recognition Lab,
Friedrich-Alexander University Erlangen-Nuremberg
Tel.: +49 9131 85 27826
Fax: +49 9131 303811
E-mail: firmualla@cs.fau.de

Simon Schöll
Pattern Recognition Lab,
Friedrich-Alexander University Erlangen-Nuremberg
ASTRUM IT GmbH, Erlangen
SAOT Graduate School in Advanced Optical Technologies

Björn Sommerfeldt
Institute of Bioprocess Engineering,
Friedrich-Alexander University Erlangen-Nuremberg

Andreas Maier
Pattern Recognition Lab,
Friedrich-Alexander University Erlangen-Nuremberg
SAOT Graduate School in Advanced Optical Technologies

Stefan Steidl
Pattern Recognition Lab,
Friedrich-Alexander University Erlangen-Nuremberg

Rainer Buchholz
Institute of Bioprocess Engineering,
Friedrich-Alexander University Erlangen-Nuremberg

Joachim Hornegger
Pattern Recognition Lab,
Friedrich-Alexander University Erlangen-Nuremberg
SAOT Graduate School in Advanced Optical Technologies

1 Introduction

Detecting cells in microscope images is a crucial step in cell image analysis. Several approaches on different image modalities tackle the problem as a classification problem. A fixed-size square patch is sampled at each pixel and used to train a cell/background classifier. The features can be either the patches themselves as in [10] or the patches after applying traditional feature extraction schemes as in [12, 13, 17].

It is known that bright-field microscopy delivers insufficient contrast at focus especially for adherent cells [1, 2, 24]. More contrast can be obtained by defocusing the microscope [1]. This fact was utilized in the literature to improve cell-detection accuracy [2, 4, 16].

Moreover, in quantitative phase microscopy (QPM) approaches [18], the physical light phase can be reconstructed computationally from amplitude information in order to get both more contrast and more object details. A QPM approach in [3] suggests approximating the TIE (Section 2.1) solution in the monogenic signal (Section 2.2) domain. In fact, the obtained results approximate the local phase and the local energy of the physical light phase.

It is expected that a defocused image delivers higher discrimination between background and cells compared to an at-focus image. In this paper, we show that using the local phase, as obtained in the particular monogenic framework of [3], instead of a defocused image yields even higher discriminative power for the cell/background classification problem. Section 2.3 discusses the details of the used classifier models and features.

The experiments were performed on bright-field images of unstained cells from three cell lines: adherent L929, adherent CHO, and Sf21 in suspension. Section 2.4 clarifies the details of the image acquisition and labeling. Section 3 shows the results of these experiments which are further discussed and summarized in Section 4.

2 Materials and methods

2.1 Transport of intensity equation

As mentioned in the introduction, defocusing a bright-field microscope yields more contrast in the acquired images. In fact, there is a relation between this contrast and the physical phase of light. The transport of intensity equation (TIE) [23] models this relation:

$$-\frac{2\pi}{\lambda} \frac{\partial I(x, y)}{\partial z} = \nabla \cdot (I(x, y) \nabla \varphi(x, y)) \quad (1)$$

where λ is the wavelength of light, z is the axial distance to the focus position, I is the intensity image at focus, φ

is the physical phase of light, ∇ is the gradient operator in the two lateral dimensions x and y , i.e. inside the image plane, and $\nabla \cdot$ is the corresponding divergence operator.

2.2 Monogenic signal

2.2.1 One-dimensional case:

The monogenic signal is a 2D generalization of a fundamental concept in signal processing called analytic signal [8]. The latter is defined for a real-valued one-dimensional signal $f(x)$ by the following equation [19]:

$$f_a(x) := f(x) + i f_h(x) \quad (2)$$

where $i^2 = -1$ and $f_h(x)$ is the Hilbert transform of $f(x)$:

$$f_h(x) := \mathcal{H}(f(x)) := f(x) * \frac{1}{\pi x} = \frac{1}{\pi} \rho \int_{-\infty}^{+\infty} \frac{f(\zeta)}{x - \zeta} d\zeta \quad (3)$$

where ρ stands for the Cauchy principal value of the improper integral. In Fourier domain, it can be shown that Eq. (3) is equivalent to:

$$F_h(\omega) = -i \operatorname{sign}(\omega) F(\omega) \quad (4)$$

where

$$\operatorname{sign}(\omega) := \begin{cases} +1 & : \omega > 0 \\ 0 & : \omega = 0 \\ -1 & : \omega < 0 \end{cases} \quad (5)$$

As can be seen in Eq. (4), the Hilbert transform of a signal is a phase shift of its frequency components by $\pm \frac{\pi}{2}$. Therefore, a signal and its Hilbert transform are commonly termed *quadrature pair*. Combining Eq. (2) and Eq. (4) yields:

$$\begin{aligned} F_a(\omega) &= F(\omega) + i F_h(\omega) \\ &= F(\omega) + \operatorname{sign}(\omega) F(\omega) \\ &= (1 + \operatorname{sign}(\omega)) F(\omega) \end{aligned} \quad (6)$$

In other words, the analytic representation of $f(x)$ can be obtained by discarding its negative frequency components. Moreover, due to the fact that Hilbert transform of a real signal is also real, Eq. (2) can be written in Euler form as:

$$f_a(x) = A(x) e^{i\phi(x)} \quad (7)$$

where $A(x) = \sqrt{f^2(x) + f_h^2(x)}$ is the local energy and $\phi(x) = \arctan \frac{f_h(x)}{f(x)}$ is the local phase. In practice, the analytic signal, and hence the local phase and energy, are computed for a band-passed version of $f(x)$ in order to

improve the frequency localization and make the result invariant to the signal energy (by removing the DC) [5]. In addition, the band-pass filter is usually designed as an even filter $e(x)$ because it has a constant phase, and thus, it does not change the phase information of the original signal $f(x)$ [5]. Based on these justifications, the analytic signal is computed in practical applications by the following equation:

$$\hat{f}_a(x) = f(x) * e(x) + i\mathcal{H}(f(x) * e(x)) \quad (8)$$

According to the convolution property [19] of Hilbert transform:

$$\begin{aligned} \hat{f}_a(x) &= f(x) * e(x) + i f(x) * \mathcal{H}(e(x)) \\ &= f(x) * (e(x) + ie_h(x)) \\ &= f(x) * e_a(x) \end{aligned} \quad (9)$$

In other words, finding the analytic representation of the signal filtered by $e(x)$ is equivalent to convolving this signal with a *quadrature filter* $e_a(x)$ which is the analytic representation of $e(x)$. Furthermore, the Hilbert transform of a real even function is a real odd function $o(x)$. Accordingly, one can write:

$$\hat{f}_a(x) = f(x) * (e(x) + io(x)) \quad (10)$$

Consequently, the local energy and phase are computed in practice as:

$$\hat{A}(x) = \sqrt{(f(x) * e(x))^2 + (f(x) * o(x))^2} \quad (11)$$

$$\hat{\phi}(x) = \arctan \frac{f(x) * o(x)}{f(x) * e(x)} \quad (12)$$

Several band-pass filters have been considered in the literature: Gabor, Gaussian derivatives, difference of Gaussians, and others. A thorough discussion about the choice of the quadrature filters can be found in [5].

2.2.2 Two-dimensional case:

The Riesz transform generalizes the Hilbert transform for n-dimensional signals [22]:

$$\mathcal{R}(f(\mathbf{x})) := (\mathcal{R}_1(f(\mathbf{x})), \dots, \mathcal{R}_n(f(\mathbf{x})))^T \quad (13)$$

$$\mathcal{R}_l(f(\mathbf{x})) := h_l(\mathbf{x}) * f(\mathbf{x}), l = 1, \dots, n \quad (14)$$

$$h_l(\mathbf{x}) := \frac{\Gamma((n+1)/2)}{\pi^{(n+1)/2}} \frac{x_l}{\|\mathbf{x}\|^{n+1}} \quad (15)$$

where Γ is the Gamma function [19], and $\mathbf{x} = (x_1, \dots, x_n)$. Eq. (15) can be written in Fourier domain as:

$$H_l(\mathbf{u}) = i \frac{u_l}{\|\mathbf{u}\|} \quad (16)$$

where $\mathbf{u} = (u_1, \dots, u_n)$ is the n-dimensional frequency vector. For $n = 1$, this transfer function expresses the Hilbert transform¹.

Without loss of generality, the monogenic signal is defined for 2D signals $f(x, y)$ as:

$$\begin{aligned} f_m(x, y) &= f(x, y) + ih_1(x, y) * f(x, y) \\ &\quad + jh_2(x, y) * f(x, y) \end{aligned} \quad (17)$$

The filters h_1 and h_2 are given by Eq. (15). $f_m(x, y)$ is defined in a quaternion space [9] whose imaginary units are i, j , and k ($i^2 = j^2 = k^2 = -1$) and the k -component is zero. One can see the monogenic signal of the 2D function $f(x, y)$ as a quaternion-valued function whose real part is the signal itself and whose vector part is the Riesz transform of the signal.

Similar to the one-dimensional case, in practice, a band-passed version of the signal is used:

$$\begin{aligned} \hat{f}_m(x, y) &= e(x, y) * f(x, y) + ih_1(x, y) * e(x, y) * f(x, y) \\ &\quad + jh_2(x, y) * e(x, y) * f(x, y) \end{aligned} \quad (18)$$

This can be reformulated as follows:

$$\begin{aligned} \hat{f}_m(x, y) &= f(x, y) * (e(x, y) + ih_1(x, y) * e(x, y) \\ &\quad + jh_2(x, y) * e(x, y)) \\ &= f(x, y) * e_m(x, y) \end{aligned} \quad (19)$$

This equation is similar to Eq. (9) in the one-dimensional case. It states that computing the monogenic representation of $f(x, y)$ filtered with $e(x, y)$ is equivalent to convolving the signal with a 2D quadrature filter given by the monogenic representation of $e(x, y)$.

Local energy is defined as the magnitude of the monogenic quaternion:

$$\hat{A}(x, y) = \sqrt{(f_e)^2 + (f_e * h_1)^2 + (f_e * h_2)^2} \quad (20)$$

where $f_e := f(x, y) * e(x, y)$. The specification of the domain (x, y) was omitted in order to simplify the notation. The local phase is defined as the angle between the vector part and the real part of the monogenic quaternion:

$$\hat{\phi}(x, y) = \arctan \frac{\sqrt{(f_e * h_1)^2 + (f_e * h_2)^2}}{f_e} \quad (21)$$

¹ There is a minus sign difference due to the incompatibility of the definitions between different authors [8, 19]. This incompatibility, however, is irrelevant for the discriminative power.

2.2.3 Solving the TIE using the monogenic signal:

In [3], a link between physical phase and local phase was established using the monogenic signal. According to [3], it is possible to use the monogenic signal framework to approximate the solution of Eq. (1) under two conditions: Firstly, the derivative image, i.e. the left side of Eq. (1), is used as an input instead of the image itself. Secondly, a low-pass filter is used in the monogenic signal framework instead of the typically-used band-pass filter. The employed low-pass filter in [3] was a Mellor-Brady filter [14] given by the following equation in the spatial domain:

$$\Omega(r, \alpha, \beta) := \frac{1}{r^{\alpha+\beta}} - \frac{1}{r^{\alpha-\beta}} \quad (22)$$

where $r = \sqrt{x^2 + y^2}$, while α and β are the filter parameters. Depending on these parameters, Ω behaves either as a low-pass or a band-pass filter. For solving the TIE, a low-pass filter was employed corresponding to $\alpha = \beta = 0.25$. Therefore, $e(x, y)$ in Section 2.2.2 was set to $\Omega(r, 0.25, 0.25)$. More details about the justification of the chosen parameter values can be found in [3].

2.3 Learning

We employ machine learning to investigate the discriminative power of the local phase as defined in Section 2.2 in the cell/background separation problem. Obviously, it is possible to measure the difference in the discriminative power of two features by learning a classifier for each of them and then comparing the test errors.

As classifier features, patches of size 5×5 pixels are used. Cell areas in our data are considerably larger than the area of the chosen patch. The advantage of using a small patch-size is reducing the sensitivity of the extracted feature vectors to the variability of cell orientation. In Section 3.5, we conduct a patch-size analysis.

As a classifier model, we use support vector machine (SVM) [21] and random forest (RF) [6]. Two kernels were utilized for the SVM: the radial basis function (RBF) and the linear kernel. The SVM cost parameter and the RBF parameter γ were set to the default values of LibSVM [7]. The number of trees in the RF and the number of the randomly selected variables at each node were set after [11] to 500 and $N_f/5$, respectively, where N_f is the feature number. The data was z-scored for the SVM, while it was used without normalization for the RF.

2.4 Materials

This study was performed on images acquired with an inverted Nikon Eclipse TE2000U microscope using Nikon USB camera. The used microscope's objective has a numerical aperture of 0.45, a working distance of 7.4 mm, and 20x magnification. Image resolution is 1280 x 960 pixels with 0.49 $\mu\text{m}/\text{pixel}$.

Table 1 Cell lines used in the evaluation

Cell line	Description	Image pairs	Cells
L929	L929 adherent cells	5	1078
Sf21	Sf21 cells in suspension	5	1001
CHO	CHO adherent cells	6	1431

Three cell lines were considered for the evaluation: adherent CHO, adherent L929, and Sf21 in suspension. The acquired data consists of 16 pairs of images. Each pair consists of an image at focus (Figure 1(a)) and another positively defocused image (Figure 1(b)) of the same scene. The number of image pairs and the total number of cells in each cell line can be seen in Table 1. The defocus distance was +30 μm for adherent cells and +15 μm for suspended cells. The total number of cells is more than 3500, all of which were labeled by two bioprocess engineering experts. This was done by manually delineating the cell borders in the defocused images using an annotation software [20].

The software SePhaCe [2] was used to generate the local phase (Figure 1(c)) and the local energy (Figure 1(d)) images for each image pair. As mentioned in Section 2.2.3, the axial derivative should be used as input for the monogenic signal framework. The difference between an at-focus image and its corresponding defocused image can be considered as a finite-difference approximation of this derivative. Therefore, it was used as input for the monogenic framework in SePhaCe.

3 Evaluation

3.1 Patch extraction

One can extract a patch (Section 2.3) at each pixel from all images. However, this is computationally expensive. Therefore, only P patches are randomly sampled from each training/testing image. Unless otherwise specified, P was set to 100. In order to achieve balanced learning, one-half of the P patches are sampled from background while the other half sampled from cells. The class of each patch is obtained from the labeled ground truth (Section 2.4). The ground truth for inhomogeneous patches,

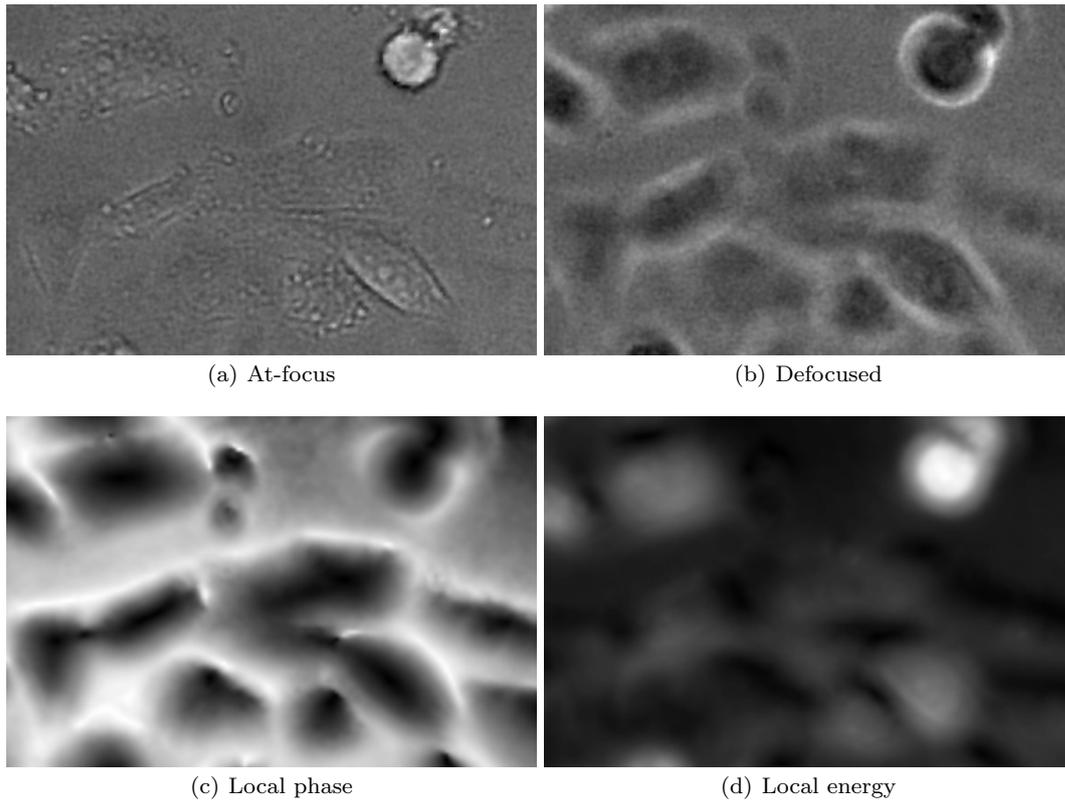


Fig. 1 Examples extracted from the L929 cell line. The histograms of the shown image parts were linearly stretched for clarity.

i.e. patches which contain both labels (usually near cell boundary), is not reliably known. Consequently, unless otherwise stated, these patches are discarded.

3.2 Comparison between local phase, local energy, at-focus signal, and defocused signal

One of the five defocused L929 images I_q was used to train three classifiers: linear SVM, RBF SVM, and RF. The learned models were then applied on the other defocused images in the same cell line and the average classification rate R_q over these test images was computed. This was repeated for each defocused L929 image, i.e. for each q value, and the mean of the R_q values was obtained.

The previous experiment was repeated 10 times with one mean classification rate obtained from each repetition. The mean and the standard deviation of all these mean classification rates can be seen in the first column of Table 2. The second, third, and fourth column of the same table show the results when the same process was applied on the at-focus, local phase, and local energy images, respectively. Table 3 and Table 4 show the results of the same procedure applied on Sf21 and CHO.

Tables 2, 3, and 4 reveal that the four features can be sorted by increasing discriminative power as follows: at-focus signal, local energy, defocused signal, local phase. The only exception for this order is that local energy is more discriminative than defocused signal for suspended cells (Sf21).

3.3 Comparison between the input space and the output space of the monogenic signal

In this section, we assess the use of the two monogenic outputs together for cell/background classification and compare it with the joint use of the two monogenic inputs. In this case, at a given pixel, a patch from a local phase image and another patch at the same pixel position from its corresponding local energy image are extracted. The values of the two patches are then concatenated. Therefore, the dimensionality of the resulting feature space is 50. The discrimination power of this feature space was compared with another 50-dimensional feature space: the monogenic input space. The latter is formed by using an at-focus image and its corresponding defocused image together for patch extraction instead of the local phase and energy images.

Table 2 L929: Comparison between local phase, local energy, at-focus signal, and defocused signal using the cell/background classification rate.

	Defocused	At-focus	Phase	Energy
Linear SVM	67.7%±0.5%	51.2%±1.0%	82.4%±1.1%	64.2%±1.3%
RBF SVM	68.2%±1.5%	56.2%±1.5%	81.7%±0.9%	64.5%±1.9%
RF	68.1%±1.1%	54.9%±0.9%	80.3%±1.1%	60.6%±1.3%

Table 3 Sf21: Comparison between local phase, local energy, at-focus signal, and defocused signal using the cell/background classification rate.

	Defocused	At-focus	Phase	Energy
Linear SVM	82.5%±1.2%	59.6%±1.2%	94.9%±0.7%	88.5%±0.5%
RBF SVM	84.1%±0.8%	73.8%±2.0%	94.9%±0.6%	88.8%±0.8%
RF	87.4%±1.3%	70.8%±2.9%	94.6%±0.7%	88.4%±1.1%

Table 4 CHO: Comparison between local phase, local energy, at-focus signal, and defocused signal using the cell/background classification rate.

	Defocused	At-focus	Phase	Energy
Linear SVM	60.9%±1.2%	49.8%±0.8%	68.5%±1.7%	57.4%±1.9%
RBF SVM	61.7%±0.9%	52.4%±0.6%	68.0%±2.4%	55.7%±1.3%
RF	61.1%±1.5%	52.5%±0.5%	63.7%±1.5%	54.8%±0.9%

The first column of Table 5 shows the cell/background classification rate on L929 when both an at-focus image and a defocused image are used together to train the classifiers. The second column shows the classification rate when both local phase and local energy are used to train the classifiers. The same can be seen in Table 6 for Sf21 and Table 7 for CHO. The classification rate was estimated in a similar way to Section 3.2. However, compared to Section 3.2, the dimensionality of the feature space is 50 instead of 25. Tables 5, 6, and 7 reveal that the compound signal of local phase and local energy is more discriminative than the compound signal of an at-focus image and a defocused image.

3.4 Comparison between local phase and the input space of the monogenic signal

Figure 2 shows a comparison between the learning curve of a RBF SVM trained using both a defocused image and an at-focus image compared to the learning curve of the same classifier model trained using a local phase image. For each point in the curve, i.e. for each number of patches $P_L = 100L^2, L = 1..10$, the classification rate was estimated in a cross-validation loop similar to the loop described in Section 3.2. The learning curve shows that a local phase image is more discriminative than the two images which were used to generate it even when more training data is incorporated in order to compensate for the increased dimensionality of the feature space.

Table 5 L929: Comparison between the input space and the output space of the monogenic signal using the cell/background classification rate.

	At-focus and defocused	Phase and energy
Linear SVM	72.8%±1.2%	81.6%±1.1%
RBF SVM	71.4%±1.7%	81.5%±0.8%
RF	73.2%±1.5%	79.1%±1.9%

Table 6 Sf21: Comparison between the input space and the output space of the monogenic signal using the cell/background classification rate.

	At-focus and defocused	Phase and energy
Linear SVM	84.9%±1.0%	96.2%±0.6%
RBF SVM	86.1%±1.5%	97.1%±0.6%
RF	87.2%±2.4%	95.7%±0.6%

Table 7 CHO: Comparison between the input space and the output space of the monogenic signal using the cell/background classification rate.

	At-focus and defocused	Phase and energy
Linear SVM	58.0%±1.2%	67.6%±1.4%
RBF SVM	55.7%±0.6%	67.0%±1.7%
RF	57.1%±1.6%	65.0%±1.6%

3.5 Patch-size analysis

All experiments in our evaluation were performed so far with 5 x 5 sized patches. In this section, we investigate other patch sizes. Table 8 shows the classification rate

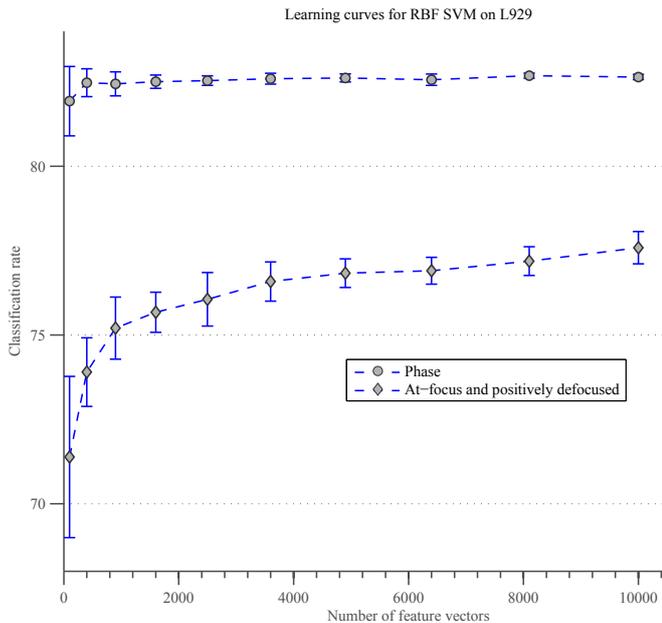


Fig. 2 Comparison between learning curves of a RBF SVM on L929 using two feature spaces: 1) local phase 2) the monogenic input space.

of a RBF SVM classifier on L929 employing the same evaluation scheme described in Section 3.2 but using 23×23 , 33×33 , and 43×43 sized patches. In order to give the reader a feeling about the ratio between these patch dimensions and cell dimensions, we point out that the length of the minor axis of the L929 cells in our data is 32.95 ± 10.72 pixels.

Increasing the patch-size increases the dimensionality of the corresponding feature space, and consequently, the number of samples needed for training. Table 9 shows the same experiment reported in Table 8, but with 900 random patches per training/testing image instead of 100.

As stated in Section 3.1, only homogeneous patches are used in training and testing. This is a plausible choice when the patch’s area is small compared to cell’s area (e.g. 5×5 sized patches). However, with larger patches, this will exclude more cell pixels from the evaluation scheme and hence degrade the generalizability of the derived conclusions. Table 10 shows the results when no patches are excluded from the evaluation. In this case, the label of the patch’s center is considered.

Tables 8, 9, and 10 reveal that the superiority of local phase over defocused signal holds for larger patches even when more samples are employed in training. On the other hand, unlike the discriminative power of local phase, the discriminative power of defocused signal benefits from employing near-boundary patches in training.

Table 8 The effect of the patch-size on the cell/background classification rate. The cell line is L929, the classifier model is RBF SVM, and the number of patches per training/testing image is 100.

	Defocused	Phase
5×5	68.2 % \pm 1.5 %	81.7 % \pm 0.9 %
23×23	63.4 % \pm 1.3 %	89.3 % \pm 0.8 %
33×33	56.5 % \pm 1.8 %	88.8 % \pm 0.6 %
43×43	52.2 % \pm 1.9 %	86.7 % \pm 1.7 %

Table 9 The effect of the patch-size on the cell/background classification rate. The cell line is L929, the classifier model is RBF SVM, and the number of patches per training/testing image is 900.

	Defocused	Phase
5×5	68.6 % \pm 0.4 %	82.4 % \pm 0.3 %
23×23	65.6 % \pm 0.4 %	89.8 % \pm 0.4 %
33×33	58.9 % \pm 0.7 %	87.9 % \pm 0.4 %
43×43	53.8 % \pm 0.6 %	82.4 % \pm 1.0 %

Table 10 The effect of the patch-size on the cell/background classification rate. The cell line is L929, the classifier model is RBF SVM, and the number of patches per training/testing image is 100. Inhomogeneous patches are included in training and testing.

	Defocused	Phase
5×5	65.2 % \pm 1.5 %	76.9 % \pm 0.7 %
23×23	66.2 % \pm 1.3 %	77.3 % \pm 1.0 %
33×33	65.5 % \pm 1.9 %	76.9 % \pm 0.8 %
43×43	67.7 % \pm 1.3 %	76.1 % \pm 1.3 %

4 Conclusion and discussion

We showed empirically that the pixelwise cell/background classification yields considerably better results when the local phase as obtained in [3] is used instead of the defocused image. More generally, the feature images can be sorted by increasing discriminative power as follows: at-focus signal, local energy, defocused signal, local phase. The only exception to this order was the superiority of local energy over defocused signal for suspended cells.

In addition, we showed that the monogenic output space is more discriminative than the monogenic input space. This is probably due to the following reason: The monogenic output delivers information about the physical light phase represented in a way which describes signal features. In fact, there is a relation between the signal features, e.g. edges and blobs, and the local phase and energy of this signal. Local energy is high at distinctive signal features while local phase determines the feature type [15].

In our experiments, a local phase image is an output of the monogenic signal framework with an at-focus

image and a defocused image used as inputs. The natural question which arises here is whether using both input images together could deliver the same discriminative power obtained by the local phase image. Due to the difference in the dimensionality between the two feature spaces, more samples need to be provided for the higher dimensional feature space in order to achieve a fair comparison. For this reason, the learning curve was utilized to compare the local phase with the input of the monogenic signal framework. The results show that by increasing the training data size, the local phase is still more discriminative than the monogenic input.

The ground truth was defined by delineating cell borders in the defocused images. The latter are blurred compared to the at-focus images and defocused cells tend to occupy larger area (cf. Figure 1(a) and Figure 1(b)). Therefore, comparing classification accuracy between a defocused image and an at-focus image is slightly biased. This bias is small because the random sampling and the exclusion of inhomogeneous patches make the probability of selecting a pixel which belongs to a defocused cell but not to the corresponding at-focus cell very low. In addition, as mentioned in the introduction, the superiority of the defocused image over the at-focus image in the cell/background separation is already known in literature, and hence it is not a main concern of our paper.

More than 3500 manually labeled cells were used in the evaluation. This relatively large number of cells supports the soundness of the paper statement. On the other hand, one might criticize the evaluation as being done using a fixed defocus distance, i.e. the distance of 30 μm or 15 μm described in Section 2.4. In fact, the selection of the defocus distance is not arbitrary. The very short distances do not deliver sufficient contrast. On the other hand, very long distances smear the image information due to the excessive blurring by the point spread function of the optical system. Therefore, there is an optimal distance which maximizes the contrast. During the image acquisition, we tried to select this optimal distance experimentally. However, this was judged subjectively. We are currently developing automatic methods to choose this distance objectively.

Another critique might be that the improvement in classification accuracy on the CHO cell line was considerably less than the improvement on L929 although both of them are adherent and have similar visual appearance. We noticed that the defocused CHO images were over-blurred compared to the images of L929. We thus anticipate that choosing a shorter defocus distance for CHO will improve the classification rate. As mentioned above, we are developing methods for automatic selection of the optimal defocus distance.

Acknowledgements The authors would like to thank the Bavarian Research Foundation BFS for funding the project COSIR under contract number AZ-917-10. In addition, we gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German Research Foundation (DFG) in the framework of the German excellence initiative.

Conflicts of interest Firas Mualla, Andreas Maier, Stefan Steidl, Rainer Buchholz, and Joachim Hornegger declare that they have no conflict of interest. Simon Schöll is employee of ASTRUM IT GmbH, Erlangen, Germany. Björn Sommerfeldt is financed by the Bavarian Research Foundation until the end of December 2013.

References

1. Agero, U., Monken, C.H., Ropert, C., Gazzinelli, R.T., Mesquita, O.N.: Cell surface fluctuations studied with defocusing microscopy. *Physical Review E* **67**(5), 051,904 (2003)
2. Ali, R., Gooding, M., Szilágyi, T., Vojnovic, B., Christlieb, M., Brady, M.: Automatic segmentation of adherent biological cell boundaries and nuclei from brightfield microscopy images. *Machine Vision and Applications* **23**(4), 607–621 (2012)
3. Ali, R., Szilágyi, T., Gooding, M., Christlieb, M., Brady, M.: On the use of low-pass filters for image processing with inverse laplacian models. *Journal of Mathematical Imaging and Vision* pp. 1–10 (2010)
4. Becattini, G., Mattos, L., Caldwell, D.: A novel framework for automated targeting of unstained living cells in bright field microscopy. In: *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 195–198 (2011)
5. Boukerroui, D., Noble, J.A., Brady, M.: On the choice of band-pass quadrature filters. *Journal of Mathematical Imaging and Vision* **21**(1-2), 53–80 (2004)
6. Breiman, L.: Random forests. *Machine Learning* **45**(1), 5–32 (2001)
7. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* **2**, 27:1–27:27 (2011). Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
8. Felsberg, M., Sommer, G.: The monogenic signal. *IEEE Transactions on Signal Processing* **49**(12), 3136–3144 (2001)
9. Hamilton, W.R.: II. On Quaternions; or on a New System of Imaginaries in Algebra. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **25**(163), 10–13 (1844)
10. Jesper Sjöström, P., Frydel, B., Wahlberg, L.: Artificial neural network-aided image analysis system for cell counting. *Cytometry* **36**(1), 18–26 (1999)
11. Khoshgoftaar, T., Golawala, M., Van Hulse, J.: An empirical study of learning from imbalanced data using random forest. In: *Proceedings of the IEEE International Conference on Tools with Artificial Intelligence*, 2, pp. 310–317 (2007)
12. Long, X., Cleveland, W., Yao, Y.: A new preprocessing approach for cell recognition. *IEEE Transactions on Information Technology in Biomedicine* **9**(3), 407–412 (2005)
13. Long, X., Cleveland, W., Yao, Y.: Automatic detection of unstained viable cells in bright field images using a support vector machine with an improved training procedure. *Computers in Biology and Medicine* **36**(4), 339–362 (2006)

14. Mellor, M., Brady, M.: Phase mutual information as a similarity measure for registration. *Medical image analysis* **9**(4), 330–343 (2005)
15. Morrone, M.C., Ross, J., Burr, D.C., Owens, R.: Mach bands are phase dependent. *Nature* **324**(6094), 250–253 (1986)
16. Mualla, F., Schöll, S., Sommerfeldt, B., Maier, A., Hornegger, J.: Automatic cell detection in bright-field microscope images using SIFT, random forests, and hierarchical clustering. *Medical Imaging, IEEE Transactions on* **32**(12), 2274–2286 (2013)
17. Nattkemper, T., Ritter, H., Schubert, W.: Extracting patterns of lymphocyte fluorescence from digital microscope images. *Intelligent Data Analysis in Medicine and Pharmacology* **99**, 79–88 (1999)
18. Popescu, G.: *Quantitative Phase Imaging of Cells and Tissues*. McGraw-Hill (2011)
19. Poularikas, A.D.: *Handbook of Formulas and Tables for Signal Processing*, vol. 13. CRC Press (2010)
20. Russell, B., Torralba, A., Murphy, K., Freeman, W.: Labelme: A database and web-based tool for image annotation. *International Journal of Computer Vision* **77**(1), 157–173 (2008)
21. Scholkopf, B., Smola, A.J. (eds.): *Learning With Kernels*. MIT Press (2001)
22. Stein, E.M.: *Singular integrals and differentiability properties of functions* Elias M. Stein., vol. 2. Princeton university press (1970)
23. Teague, M.R.: Deterministic phase retrieval: a green's function solution. *Journal of the Optical Society of America* **73**(11), 1434–1441 (1983)
24. Tscherepanow, M., Zöllner, F., Hillebrand, M., Kummert, F.: Automatic segmentation of unstained living cells in bright-field microscope images. In: *Advances in Mass Data Analysis of Images and Signals in Medicine, Biotechnology, Chemistry and Food Industry, Lecture Notes in Computer Science*, vol. 5108, pp. 158–172. Springer (2008)