

# Segmentation and Classification of Cell Cycle Phases in Fluorescence Imaging\*

Ilker Ersoy<sup>1</sup>, Filiz Bunyak<sup>1</sup>, Vadim Chagin<sup>2,3</sup>, M. Christina Cardoso<sup>2</sup>,  
and Kannappan Palaniappan<sup>1</sup>

<sup>1</sup> Department of Computer Science, University of Missouri Columbia, USA

<sup>2</sup> Department of Biology, Technische Universität Darmstadt, Germany

<sup>3</sup> Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia

**Abstract.** Current chemical biology methods for studying spatiotemporal correlation between biochemical networks and cell cycle phase progression in live-cells typically use fluorescence-based imaging of fusion proteins. Stable cell lines expressing fluorescently tagged protein GFP-PCNA produce rich, dynamically varying sub-cellular foci patterns characterizing the cell cycle phases, including the progress during the S-phase. Variable fluorescence patterns, drastic changes in SNR, shape and position changes and abundance of touching cells require sophisticated algorithms for reliable automatic segmentation and cell cycle classification. We extend the recently proposed graph partitioning active contours (GPAC) for fluorescence-based nucleus segmentation using regional density functions and dramatically improve its efficiency, making it scalable for high content microscopy imaging. We utilize surface shape properties of GFP-PCNA intensity field to obtain descriptors of foci patterns and perform automated cell cycle phase classification, and give quantitative performance by comparing our results to manually labeled data.

## 1 Introduction

The spatial distribution and temporal dynamics of proteins within living cells are being quantitatively studied using a variety of microscopy imaging modalities to understand the interaction between sub-cellular processes and cell behavior. The current method for live-cell visualization and tracking of proteins is to use translational fusion with fluorescent proteins. The analysis of cell cycle dependent changes is only now becoming feasible with the discovery of suitable markers that allow identification of cell cycle phases in proliferating cells [1]. Current experimental techniques use fusion proteins in combination with fluorescence time-lapse microscopy to mark sub-cellular structures in the nucleus to identify the cell cycle phase. Several cell cycle labeling approaches are being pursued including RFP-Ligase for localizing DNA methyltransferase, GFP-PCNA (proliferating cell nuclear antigen fused to green fluorescent protein) where PCNA is involved in DNA replication and repair, and YFP-RAD18 to

---

\* Research partially supported by the US NIH NIBIB award R33-EB00573 (KP).

label postreplication repair of damaged DNA or immunostaining these endogenous proteins to characterize all cell cycle phases [1, 2, 3]. The focus of this paper is on fluorescent labeling using GFP-PCNA which enables cell cycle phases to be distinguished by characteristic patterns of GFP-PCNA at different points of cell cycle: M-phase or mitosis, followed by G1-phase, early, mid and late S-phases, and G2-phase (Figure 1). GFP-PCNA produces a complex distribution of foci patterns in different stages of the S-phase, but is fairly homogenous during G1- and G2-phases, and very diluted during M-phase. In order to identify cell cycle phases, individual nuclei need to be detected, segmented and classified. Cell segmentation, classification and tracking require robust and sophisticated algorithms in order to deal with noise, shape changes, texture and touching cells [4, 5, 6, 7, 8]. In this paper we describe a novel technique for fluorescent nuclei detection and segmentation using a fast implementation of multi-phase graph partitioning active contours (fastGPAC). Fluorescent particles corresponding to the localization of GFP-PCNA have characteristic distributions that we utilize to train a support vector machine to classify the segmented nuclei into four cell phases and three sub-phases in S-phase.

## 2 Methods

### 2.1 Segmentation Using FastGPAC

To segment HeLa cell nuclei we use level set-based multi-phase fast graph partitioning active contours (FastGPAC) which is our novel efficient implementation of graph partitioning active contours (GPAC). FastGPAC reduces the  $O(N^4)$  computational complexity and memory requirements of the original GPAC algorithm [9, 10] to  $O(N^2)$  computational complexity and  $O(n \times L)$  constant memory where  $N \times N$  is the image size,  $L$  is number of histogram bins, and  $n$  is the number of phases. Graph partitioning active contours (GPAC) is introduced in [9, 10] as a new powerful curve evolution framework. GPAC can be implemented using explicit snake-based or implicit level set-based active contours. Level set-based implementation where a curve  $\mathcal{C}$  is represented implicitly via zero-level curve of a Lipschitz function  $\phi$   $\mathcal{C} = \{(x, y) | \phi(x, y) = 0\}$ , provide advantages such as eliminating the need to reparameterize the curve and automatic handling of topology changes [11]. The variational cost function that minimizes pairwise dissimilarity within regions is written as [10]:

$$E_{WR} = \iint_{\Omega} \iint_{\Omega} w(p_1, p_2) H(\phi(p_1)) H(\phi(p_2)) dp_1 dp_2 + \iint_{\Omega} \iint_{\Omega} w(p_1, p_2) (1 - H(\phi(p_1))) (1 - H(\phi(p_2))) dp_1 dp_2 \quad (1)$$

where  $\Omega$  is the whole image domain,  $w()$  is a pixel-to-pixel dissimilarity measure,  $H$  is the heaviside function, an indicator function for the points inside  $R_i(\mathcal{C})$ , and outside  $R_o(\mathcal{C})$  of the curve ( $H(\phi)$  and  $(1 - H(\phi))$ ) respectively. Curve evolution equation is obtained with steepest descent minimization. The complete curve

evolution equation (with regularization term, normalization factors  $\alpha$  and  $\beta$ , and weights  $\lambda_{1,2}$  and  $\mu$ ) is [10]:

$$\frac{\partial \phi(p_2)}{\partial t} = \delta(\phi(p_2)) \left[ \lambda_2 \beta \iint_{\Omega} w(p_1, p_2) (1 - H(\phi(p_1))) dp_1 - \lambda_1 \alpha \iint_{\Omega} w(p_1, p_2) H(\phi(p_1)) dp_1 + \mu \operatorname{div} \left( \frac{\nabla \phi(p_2)}{|\nabla \phi(p_2)|} \right) \right] \quad (2)$$

which is discretized as:

$$\frac{\Delta \phi(p_2)}{\Delta t} = \delta_{\epsilon}(\phi(p_2)) \left[ \lambda_2 \beta \sum_{p_1 \in R_o(\mathcal{C})} w(p_1, p_2) - \lambda_1 \alpha \sum_{p_1 \in R_i(\mathcal{C})} w(p_1, p_2) + \mu \mathcal{K} \right] \quad (3)$$

While powerful in terms of region description, heavy computational and memory requirements prevent GPAC's direct application to large images. Our FastGPAC approach reduces both computational and memory requirements of the *original* GPAC, without approximations such as dissimilarity computation at block or superpixel level [9, 10], and makes segmentation of large images with GPAC approach possible. The bottleneck in the original GPAC is the computation of the 2-D regional (inside and outside) sums in Eq. 3,  $\sum_{p_1 \in R_o(\mathcal{C})} w(p_1, p_2)$  and  $\sum_{p_1 \in R_i(\mathcal{C})} w(p_1, p_2)$ . To speed the process in [9], dissimilarities of every image point to every image point are pre-computed and stored in a  $N^2 \times N^2$  lookup table  $W$  for an  $N \times N$  image. But this  $O(N^4)$  table quickly becomes impractical for large images (i.e. over a terabyte of memory for a  $1024 \times 1024$  grayscale image). FastGPAC speeds up 2-D regional sum computations by maintaining two histograms  $\mathbf{h}_i$  and  $\mathbf{h}_o$  for regions  $R_i(\mathcal{C})$ ,  $R_o(\mathcal{C})$ . When  $w(p_1, p_2)$  does not incorporate spatial distance between points  $p_1$  and  $p_2$ ,  $w(p_1, p_2)$  can be rewritten as  $w(p_1, p_2) \equiv D(F(p_1), F(p_2))$  where  $F(p)$  is a feature extracted from the point  $p(x, y)$ , and  $D$  is a similarity/dissimilarity measure defined on  $F$  (i.e. for  $w(p_1, p_2) = |I(p_1) - I(p_2)|$ ,  $F(p)$  is grayscale intensity  $I(p)$  and  $D$  is  $L_1$  metric.)

**GPAC Region Sum Theorem.** For cases where  $w(p_1, p_2)$  does not incorporate spatial distance, the 2-D regional sums,  $\sum_{p_1 \in R_r} w(p_1, p_2)$  (for  $R_r = R_i$  and  $R_r = R_o$ ) can be reduced to 1-D sums independent of the size or shape of the regions  $R_i(\mathcal{C})$  and  $R_o(\mathcal{C})$ .

$$\sum_{p_1 \in R_r} w(p_1, p_2) \equiv \sum_{j=0}^{L-1} h_r(j) D(F(p_2), j) \quad (4)$$

where  $\mathbf{h}_r$  is the histogram of the feature  $F$  in region  $R_r$ ,  $D()$  is a (dis)similarity measure,  $L$  is number of bins in  $\mathbf{h}_r$  and  $h_r(j) = \sum_{p \in R_r \wedge F(p)=j} 1$  is the  $j^{\text{th}}$  bin of  $h_r$  corresponding to the number of points  $p \in R_r$  whose features  $F(p)$  are in  $j^{\text{th}}$  bin ( $F(p) \in j$ ).

**Proof.** This equality is derived by grouping the points  $p$  into feature class bins  $F(p) \in j$ , and by separating the original sum into two sums as follows:

$$\begin{aligned} \sum_{p_1 \in R_r} w(p_1, p_2) &\equiv \sum_{p_1 \in R_r} D(F(p_1), F(p_2)) \\ &= \sum_{j=0}^{L-1} \sum_{p_1 \in R_r \wedge F(p_1) \in j} (D(F(p_2), j) \times 1) = \sum_{j=0}^{L-1} D(F(p_2), j) \times \underbrace{\sum_{p_1 \in R_r \wedge F(p_1) \in j} 1}_{h_r(j)} \end{aligned} \quad (5)$$

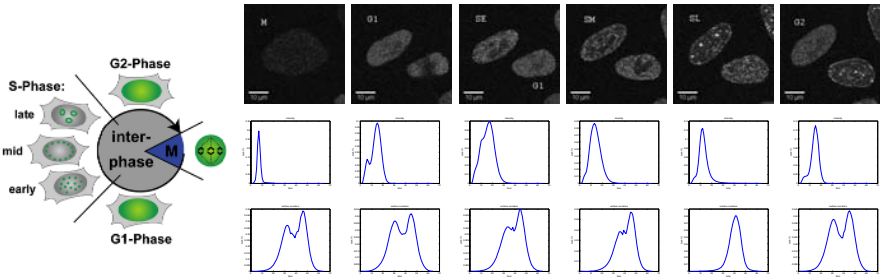
Using the GPAC region sum theorem, FastGPAC transforms GPAC curve evolution Eq. 3 into:

$$\begin{aligned} \frac{\Delta\phi(p_2)}{\Delta t} &= \delta_\epsilon(\phi(p_2)) \left[ \lambda_2 \beta \sum_{j=0}^{L-1} h_o(j) D(F(p_2), j) - \lambda_1 \alpha \sum_{j=0}^{L-1} h_i(j) D(F(p_2), j) + \mu \mathcal{K} \right] \\ &= \delta_\epsilon(\phi(p_2)) \left[ \sum_{j=0}^{L-1} [\lambda_2 \beta h_o(j) - \lambda_1 \alpha h_i(j)] D(F(p_2), j) + \mu \mathcal{K} \right] \end{aligned} \quad (6)$$

This transformation reduces  $N^4$  pairwise dissimilarity computations (from each pixel to each pixel) to  $N^2 2L$  dissimilarity computations (from each pixel to each of the  $L$  histogram bins) where  $L$  is constant and  $L \ll N^2$ . In our application, appearance of the nuclei changes during the different phases of the cell cycle. Use of two-phase schemes risk false misses, particularly during mitosis when signal-to-noise ratio drops near to background levels. Due to this, we use 4-phase segmentation. In [10], GPAC is extended to multi-phase in a way similar to Vese and Chan’s multi-phase extension [12]. As in the case of 2-phase GPAC, each sum in the multi-phase GPAC is transformed to its efficient form using the GPAC regional sum theorem.

## 2.2 Cell Cycle Phase Classification

In order to classify detected nuclei into one of the six classes (M, G1, S (early, mid, late) and G2), we utilize the characteristic appearance of GFP-PCNA in fluorescent nuclei images. As shown in Figure 1, different phases are manifested by a rich textural information that can be captured by using the histograms of intensity as well as intensity surface curvature. We do not use 2-D geometric features of the cells in order to make the feature vector robust to changes in the shape of nuclei. Our choice of feature vector (64 bins of intensity histogram and 64 bins of intensity surface curvature histogram) captures the characteristic texture information of each class without more elaborated feature vector computations such as [13, 14, 7]. Shape-based properties of the intensity surface can be utilized for blob and ridge analysis [6]. The GFP-PCNA in nucleus produces spikes and blob-like patterns hence the utilization of blob detection methods to obtain a texture signature is theoretically sound. Ridges and blobs can be defined as local extrema of principal curvatures of the intensity surface. Principal curvatures and directions of a hypersurface  $L$  correspond to the eigenvalues



**Fig. 1.** GFP-PCNA fluorescence-based patterns during six periods of cell cycle and corresponding signatures. Left to right: M phase, G1 phase, early, mid, late S phases and G2 phase. Top to bottom: sample cell image, intensity histogram signature, surface curvature histogram signature.

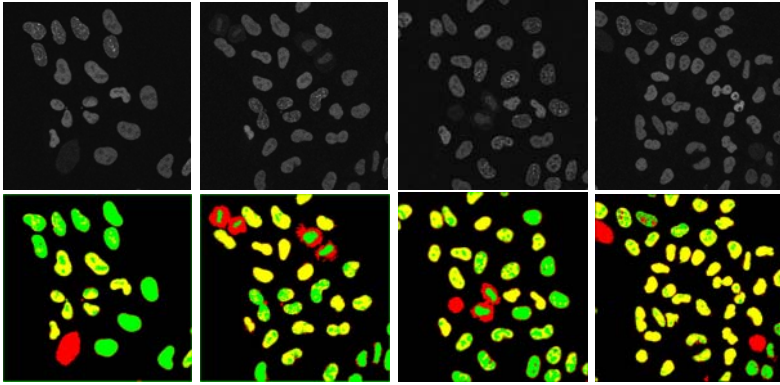
$\kappa_1 \geq \dots \geq \kappa_{n-1}$  and eigenvectors  $\xi_1 \geq \dots \geq \xi_{n-1}$  of the shape operator matrix on the tangent space  $W$ . In 2-D case,  $W$  is given as a function of the first and second fundamental forms. Since computation of principal curvatures is expensive, mean curvature  $H = \frac{1}{2}(\kappa_1 + \kappa_2) = \frac{1}{2}\text{trace}(W)$  is often used to classify surface patches ( $H < 0$  : peak, ridge, or saddle ridge;  $H = 0$  : flat or minimal surface;  $H > 0$  : pit, valley, or saddle valley). In generalization of local extrema for real-valued functions of a vector variable, a point  $x_0$  is classified as maximum if  $\nabla L(x_0) = 0$  (critical point) and  $\mathcal{H}(L(x_0))$  is negative definite (all eigenvalues  $\lambda_i < 0$ ) where  $\mathcal{H}$  is the Hessian matrix:

$$\mathcal{H} = \begin{bmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{bmatrix} \quad (7)$$

For critical points ( $\nabla L(x_0) = 0$ ), eigenvalues  $\lambda_i$  and eigenvectors  $v_i$  of the Hessian matrix correspond to principal curvatures  $\kappa_i$  and principal directions  $\xi_i$  respectively. We utilize the histograms of  $\lambda_1(\mathcal{H})$  ( $|\lambda_1| > |\lambda_2|$ ) and intensity to obtain a characteristic signature for each cell cycle phase. By binning the intensity and surface curvature histograms into 64 bins we obtain a 128-D feature vector. Figure 1 shows the average signatures of each class. We train a support vector machine [15] with a test set of nucleus images for each of the six classes using these signatures as described in the following section.

### 3 Experimental Results

Genetically modified human HeLa Kyoto cell lines were generated and validated to stably express the fused protein green fluorescent protein-tagged proliferating cell nuclear antigen (GFP-PCNA). The first step involved creating HeLa Kyoto lines containing a stably integrated Flp-recombination site (FRT). This was followed by site-specific integration of a construct containing the human EF1 $\alpha$  promoter to drive expression of the fusion gene, in this case GFP-PCNA, and a blasticidin resistance marker gene used for selection of the transgenic cells flanked by FRT sites. This strategy allows the Flp recombinase mediated integration of DNA into a specific site in the genome and a reliable and homogeneous



**Fig. 2.** Sample FastGPAC segmentation results for four sample frames. Top: Original frames, bottom: Recolored level set masks.

level of the fluorescent protein in every cell. Live cell analysis was performed by plating the cells on chambered glass coverslips before microscopy. The chambered glass coverslip was mounted onto the microscope stage and maintained in a humidified atmosphere of 5% CO<sub>2</sub> at 37°C on a microscope incubation system. For time lapse analysis, images were acquired with a Zeiss LSM 510 Meta laser scanning confocal microscope using the 488 nm laser line of an Argon ion laser at low power every 15 minutes. The image sequence has 174 1024 × 1024 frames. Figure 2 shows sample segmentation results for four frames using 4-phase level set FastGPAC method. When the 4-phase segmentation masks are recolored as in 2nd row (black-red-green-yellow in the order of increasing average phase intensity), some observations on the color scheme can be made i.e. just before mitosis, nuclei appear as solid red blobs (fluorescent intensity fades), just after mitosis daughter nuclei appear as red blobs with green centers. Since original GPAC requires terabyte size memory for 1024 × 1024 images, 220 × 160 regions are cropped and segmented for comparison. FastGPAC segments cropped images in 2.2 CPU seconds, original GPAC spends 415 CPU seconds (189× speedup). The actual completion time of original GPAC is longer due to extensive memory swaps. For the classification task, colored masks are binarized into foreground and background using a rule-based scheme to avoid merging of neighboring nuclei. 100 frames are labeled by an expert in the art to provide ground truth for quantification of the classification performance. A total of 1543 cells are chosen for training and testing. For each cell, a 128-D feature vector is derived from the intensity and surface curvature histograms. The feature vectors are used in 5-fold cross validation to obtain five runs for classifier performance training and testing. Each run of training is also performed in five folds to obtain the best parameters for SVM. Table 1 shows the average percentage confusion matrices of the proposed classification approach and of the Wndchrm method in [14] when applied to our data set for classification of four phases and three sub-phases in S-phase (columns do not add up to 100% due to rounding). Top 30% of the

**Table 1.** Average confusion matrix for 6 classes. Left: Wndchrm, right: proposed approach.

	M	G1	SE	SM	SL	G2	M	G1	SE	SM	SL	G2
M	90	1	1	0	0	0	95	1	0	0	0	1
G1	6	90	6	3	1	18	5	97	7	7	0	17
SE	1	2	80	12	1	0	0	0	90	16	0	0
SM	0	0	13	65	8	0	0	0	2	69	1	0
SL	0	0	1	20	90	1	0	0	0	9	98	0
G2	3	7	0	0	0	81	0	2	0	0	1	83

**Table 2.** Average confusion matrix for 4 classes. Left: Wndchrm, right: proposed approach.

	M	G1	S	G2	M	G1	S	G2
M	90	1	0	0	95	1	0	1
G1	6	90	3	18	5	97	4	17
S	1	2	97	1	0	0	96	0
G2	3	7	0	81	0	2	0	83

ranked 1025-D feature vector in [14] is used for classification as the best result. Proposed approach exceeds the accuracy of Wndchrm in all classes. As expected, separating G1 from G2 without resorting to temporal constraints is a challenge. Similarly the mid S-phase is highly confused since there are no clear cut boundaries between SE-SM and SM-SL. The proposed feature vector provides good performance in capturing the textural characteristics of the GFP-PCNA in nuclei. The computation of our feature vector for 1543 cells takes about 3 minutes, whereas the running time of Wndchrm on our data set is about 5 hours. The overall average accuracy of our classification approach is 92.3%, the same of Wndchrm is 86.4% where (worst case, best case) accuracies are (90.3%, 94.5%) for our approach and (84.9%, 87.5%) for Wndchrm. Table 2 shows the average percentage confusion matrices for the 4-phase classification.

## 4 Conclusions

The proteins of interest in fluorescence-based imaging are involved in basic cellular processes such as DNA repair and replication. PCNA is a key component of the DNA replication machinery. GFP-PCNA-based cell cycle analysis is more precise since PCNA is directly linked to the DNA replication, and provides higher resolution as well as information about the progress of S-phase through patterns of different sized foci. The imaging noise, lower SNR in some phases, complex textural patterns, significant shape changes during cell division and large data volumes require the development of a multiclass region-based segmentation algorithm with topological flexibility. We extended the recently proposed multiphase GPAC algorithm for fluorescence-based cell nucleus segmentation by incorporating density functions to capture the variability of regions for reliable and accurate segmentation. GPAC has not been previously applied to large biomedical segmentation applications due to extensive memory (on the order of terabytes) and computational requirements for large images. We derive a FastGPAC algorithm that requires constant memory and is highly scalable for high content screening time-lapse microscopy images. Preliminary results indicate that the multi-phase implementation is able to accurately segment nuclei of proliferating cells imaged for more than 40 hours. We also use a support vector machine to classify segmented nuclei into one of the four phases and three sub-phases. Quantitative

results show highly accurate cell phase classification using intensity and surface curvature histograms without the need for elaborated feature extraction schemes that are computationally expensive. Future work includes improving the confusion of phases and incorporating these results into our multi-object multi-hypothesis tracker [4] to enforce temporal constraints and provide accurate lineage construction.

## References

1. Easwaran, H., Leonhardt, H., Cardoso, M.: Cell cycle markers for live cell analyses. *Cell Cycle* 4(3), 453–455 (2005)
2. Sporbert, A., Gahl, A., Ankerhold, R., Leonhardt, H., Cardoso, M.: DNA polymerase clamp shows little turnover at established replication sites but sequential de novo assembly at adjacent origin clusters. *Molecular Cell* 10(6), 1355–1365 (2002)
3. Leonhardt, H., Rahn, H.-P., Weinzierl, P., Sporbert, A., Cremer, T., Zink, D., Cardoso, M.: Dynamics of DNA replication factories in living cells. *J. Cell Biology* 149(2), 271–280 (2000)
4. Bunyak, F., Palaniappan, K., Nath, S., Baskin, T., Dong, G.: Quantitative cell motility for *in vitro* wound healing using level set-based active contour tracking. In: Proc. IEEE Int. Symp. Biomedical Imaging, April 2006, pp. 1040–1043 (2006)
5. Nath, S.K., Palaniappan, K., Bunyak, F.: Cell segmentation using coupled level sets and graph-vertex coloring. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 101–108. Springer, Heidelberg (2006)
6. Ersoy, I., Bunyak, F., Palaniappan, K., Sun, M., Forgacs, G.: Cell spreading analysis with directed edge profile-guided level set active contours. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 376–383. Springer, Heidelberg (2008)
7. Wang, M., Zhou, X., Li, F., Huckins, J., King, R.W., Wong, S.T.: Novel cell segmentation and online SVM for cell cycle phase identification in automated microscopy. *Bioinformatics* 24(1), 94–101 (2008)
8. Padfield, D., Rittscher, J., Thomas, N., Roysam, B.: Spatio-temporal cell cycle phase analysis using level sets and fast marching methods. *Medical Image Analysis* 13(1), 143–155 (2009)
9. Sumengen, B., Manjunath, B.: Graph partitioning active contours (GPAC) for image segmentation. *IEEE Trans. Patt. Anal. Mach. Intell.*, 509–521 (April 2006)
10. Bertelli, L., Sumengen, B., Manjunath, B., Gibou, F.: A variational framework for multi-region pairwise similarity-based image segmentation. *IEEE Trans. Patt. Anal. Mach. Intell.*, 1400–1414 (August 2008)
11. Sethian, J.: *Level set methods and fast marching methods*, 2nd edn. Cambridge Univ. Press, Cambridge (1999)
12. Vese, L., Chan, T.: A multiphase level set framework for image segmentation using the Mumford and Shah model. *Int. J. Computer Vision* 50(3), 271–293 (2002)
13. Boland, M.V., Murphy, R.F.: A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. *Bioinformatics* 17(12), 1213–1223 (2001)
14. Shamir, L., Orlov, N., Eckley, D.M., Macura, T., Johnston, J., Goldberg, I.: Wndchrm - An open source utility for biological image analysis. *Source Code for Biology and Medicine* 3(1) (2008)
15. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines (2001), <http://www.csie.ntu.edu.tw/~cjlin/libsvm>