Feature extraction, selection and classifier design in automated time-lapse fluorescence microscope image analysis

M. Wang

School of Life Sciences, Nanjing University, Nanjing, People’s Republic of China, 210093.

Using automated time-lapse fluorescence microscope imaging to measure drug response, e.g. using taxanes and the vinca alkaloids to cancers, has become a common practice in cancer research and drug discovery. A mature system usually consists of the following subsystems: image acquisition subsystems, image pre-processing subsystems, image analysis subsystems, database subsystems and data analysis subsystems. Particularly, the image analysis subsystems play a key role in determining the overall performance of this system. The image analysis subsystems comprise of a dynamic cellular segmentation module, a feature extraction, a cell-cycle phase identification module, a tracking module and a descriptor quantification module. In this chapter, we will systematically introduce the feature extraction/selection and cell-cycle phase identification module. To illustrate the general techniques mentioned above, an applied laboratorial system which includes the many descriptors representing nuclei, feature selections techniques and online-SVM will be introduced. These techniques can easily be applied for users’ own research purposes.

Keywords:  microscopy, image processing, feature selection, feature extraction, classifier design

1. Background

In current drug-discovery practices, particularly in cancer researches, measuring cell cycle progression and initiation of apoptosis in individual cells are useful for better understanding of how apoptosis is induced by anti-mitotic drugs like Taxanes and the vinca alkaloids. However, most viability arrays and other techniques are endpoint assays and incapable of being performed on the same population over time. Furthermore, viability measurements are often based on bulk populations, concealing important local information in subpopulations.

Automated time-lapse fluorescence microscope imaging makes it possible to measure drug responses in a dynamic fashion without the need for cell synchronization as in conventional techniques. With the fluorescent protein markers, dynamic cell information becomes available by monitoring mitosis and apoptosis in living cells over the extended periods of imaging. Thus, the requirement of analyzing these images necessitates the inventions of a spectrum of new algorithms. Key algorithms include image acquisition, image pre-processing, image analysis, feature extraction, feature selection, cell tracking and classifier design. Once intergraded into an independent software package, the algorithms can be used to characterize the dynamics and evolutions of treated cells in time and space.

1.1 Antimototic Drugs- induced Apoptosis

Taxanes and vinca alkaloids are two commonly used drugs for curing cancers, but the mechanisms utilized by these drugs for inducing apoptosis remains to be mined. Though it is known that taxanes perturb mitosis by binding to tubulin, the subsequent events that lead to apoptosis have not been studied thoroughly. Other drugs inhibiting apoptosis include monastral and hesperadin. Monstral inhibits normal spindle formation during mitosis by the inhibition of the mitotic motor protein Eg5, which results in mitotic arrest via activation of the spindle checkpoint; by interfering with normal chromosome segregation, hesperadin inhibits cytokinesis. However, it remains unknown how perturbation of mitosis leads to the initiation of apoptosis. It is also unknown how taxol acts in combination with inhibitors of Eg5 or Aurora kinases and whether these combinations potentiate the apoptotic response in clinical settings.

![Fig.1](image)

Fig.1 Changes in the appearance of a nucleus during cell mitosis. From (a) to (h) consecutive image subframes form a sequence showing nuclei size and shape changes during cell mitosis [1].
To understand the mechanisms utilized by anti-mitotic drugs to induce apoptosis and how drug resistance might arise, we need to measure cell cycle progression, particularly the mitosis (M) phase including prophase, metaphase, anaphase and telophase, to differentiate normal cells from abnormal ones, and to detect initiation of apoptosis in individual cells as a function of time. However, most viability assays and Fluorescence-Activated Cell Sorting techniques cannot measure the same cell population over time. Furthermore, these techniques can only provide statistics of the bulk population, blurring important trends in subpopulations of cells. This problem is especially exacerbated when cells are not synchronized with respect to cell cycle stage. Though chemical synchronization methods can be used to overcome this difficulty, these methods are problematic as the treatments are cytotoxic.

1.2 Time-Lapse Microscopy

Fluorescence Microscopy is highly complex machine, traditionally requiring total reliance on well trained operators. To achieve automation, a microscopy should be capable of generating images from multi-well plates; auto-focusing when moved to a new position; and replacing filters for multiplexing colours. In the field of industry, automated microscopy is also being developed for pharmaceutical companies to endow them with capabilities such as imaging based screening. In addition, the time-lapse microscopy can acquire images over multiday periods, while measuring drug response in a dynamic fashion without requiring synchronization. This technique provides us unprecedentedly large amounts of new cellular image information of both spatial and temporal dimensions. The recent availability of fluorescent protein markers, such as GFP-tagged histones, allows the monitoring of mitosis and apoptosis in living cells over long periods of imaging.

Over the past decades, light microscopy have become a standard way of studying cellular dynamics, including changes in cell shape in response to the environment, biogenesis of organelles, intracellular traffic of vesicles, pathogens, nucleic acids, proteins and lipids, etc. These images link cell structure and cell physiology. Recent technical advances, such as faster and more reliable image acquisition techniques, increased computation power, new fluorophores, automated motorized microscopes and quantum dots, have led to the dramatically increased ability of capturing multi-spectra data during time-lapse live cell imaging in multiple dimensions. The volume of image output has consequently reached another scale of quantity and complexity.

1.3 High-Throughput Image Analysis

An automated cellular image analysis system usually consists of automated segmentation, feature extraction, pattern recognition, and classification algorithms. As the most critical step in analysis, both prior knowledge of the morphology of cells and the distribution of fluorescence signals in them are required for the design of a segmentation algorithm. The simplest and most intuitive methods are shareholding methods and its variations, which select a threshold to maximize mostly based on histograms [2-3]. If the entropy measure is adopted, the desirable threshold corresponds to the maximum sum of the objects and background's entropy values. To circumvent the failures of global threshold methods caused by uneven background or illustration, the adaptive thresholding method where a different threshold is used for different regions in the image was adopted. Other widely used segmentation algorithms include watershed, level set, k-means, and stochastic image processing.

To be processed by computer, every cell's features have to be extracted and kept in the form of numerical descriptors [4-6]. An array of cell features have been proposed, such as area, size, shape, intensity, texture, moment, Haralick texture and Gabor wavelet features. Other information like the change of size and shape of nuclei during and after mitosis can be deployed to either track or identify the progression of one cell during this process or of many cells simultaneously. The gold standard for validating these methods is to test if they can work for automatically acquired, unbiased, low magnification, auto segmented image movies of numerous cells present in real application scenarios.

Tacking techniques such as mean-shift and contour tracking are widely used for military purposes. Similar technologies can be employed to track the change in position, size and shape of an individual cell or its offspring during mitosis or to track a subpopulation of cells for studying their trends. This algorithm will turn out to be a powerful tool for studying cellular's apoptotic response to perturbations of drug treatments.

1.4 Statistical Analysis and Data Mining

After extracting the image features and storing them in a database in the form of vectors, a variety of techniques from pattern recognition, statistics and data mining, can be applied to interpreting results, building classification models and analyzing cellular response to perturbations of drugs. The first step is to analyze and quantize spatial and temporal descriptors and cellular phenotypes from microscopy images. With an appropriate set of statistical methods, an array of tasks can be accomplished: (1) evaluating image segmentation and other algorithms, (2) systematically understanding cell cycle changes under different perturbations, (3) predicting cancer cells' responses to various anti-mitotic drugs.
1.5 Image analysis for Drug discovery and Cell biology

Image-based assay and image analysis can be directly applied to drug discovery and cell biology. One typical application is to design an assay for measuring the shuttling of transcription factor between cytosol and nucleus. Different enrichment levels at nucleus and cytosol might indicate the activation of certain pathway involving this transcription factor. In another example, as mentioned before, the process of how apoptosis is induced by anti-mitotic drugs and how drug resistance might arise could be recorded by microscopy and analyzed by specially designed algorithms. In a more interesting example, the number of centrosomes of each drug-treated cell was counted by computer algorithms, which makes it much easier to identify the chemicals that affect centrosome duplication.

2. Image segmentation and Feature Extraction

2.1 Image Acquisition and Pre-processing

Images were acquired on an automated epifluorescence TE2000-E Eclipse microscope (Nikon Instruments Inc, USA) with a motorized XYZ-plane stage. Light was from a mercury arc lamp with 2 neutral density filters. SimplePCI was used to control image acquisition. A custom designed microscope incubator set at 37°C was used to keep a constant environment while acquiring images. Representative fields were chosen and the starting X, Y and Z coordinates were used to seed the auto-focusing. Auto-focusing occurred on the first pass and subsequent 10 passes to compensate for motor drift. Images were acquired using a 0.2s exposure time, every 15 minutes for 48 hours, giving a total of 200 images for each position, which were then exported from SimplePCI as 16 bit uncompressed TIFF files to a 7TB network attached storage (NAS) arrays for processing.

Image pre-processing was performed to enhance image quality for further image processing such as image segmentation, feature extraction and cell phase identification, etc. Noise removal, image restoration and contrast enhancement are three techniques widely used [7-8]. Nonlinear filters such as centre weighted median filters [9-10] and adaptive weighted median filters [11] can effectively preserve high-frequency information such as textures and edges in microscope images, which are important for better deconvolution results. Fast deconvolution algorithms like the preconditioned conjugate gradient (CG) method [12] and the generalized minimal residual (GMRES) method [13] were often applied to degrade large size microscope images based on the point spread function (PSF) of the microscope imaging system. This step is time-consuming. High-performance computation techniques can speed up this process. Image contrast enhancement facilitates its following processes. For example, homomorphic filtering is designed to correct for non-uniform intensity caused by light illumination.

2.2 Cell Segmentation

Image segmentation is critical to successful feature extraction, tracking and phase identification. Capitalizing on the property that cells have higher intensity than their background, straightforward image segmentation approaches, such as Otsu [14-15] and ISODATA algorithms [16], can segment the cell objects from the background. However they may fail in more complicated situations: (1) touching objects, (2) objects on the borders, and (3) incomplete cells. A large array of segmentation schemes have been proposed, among which Watershed algorithms [17] are specially suited for segmenting touching objects. Therefore, a watershed-based segmentation scheme was proposed for cell segmentation, which includes three major steps: binarization, cell detection and seeded-watershed based segmentation.

2.2.1 Binarization

Utilizing the property that cells have higher intensity than their background, a thresholding scheme can be used to generate the binary image for the distance transform [18]. The thresholding process is based on a data-driven background correction algorithm [19-20], wherein the background is estimated using cubic B-spline. Drawing on the huge intensity gap between the background and the object, each pixel is classified either as part of the object, if the gap between its intensity and that of the estimated background image is greater than a given threshold, or as the background, if otherwise.

2.2.2 Cell Detection and Seeded Watershed Algorithm

Local intensity maxima are often used in cell detection. Meanwhile, the distance image obtained by applying the distance transform on the binary image gives the cells’ shape information. To employ both the shape and the intensity information, the original image is superimposed on the distance image. The combined image is then filtered with a Gaussian filter with standard deviation $\sigma = 3$ [21]. In the filtered image, the noise has been suppressed and the local maxima are more likely to correspond to the cell centres. Thus, the cell detection problem task can be reduced to detecting local maxima in the filtered image.
Knowing that in the gradient vector field (GVF) the gradient vectors point to the local maxima, we put one particle on each object pixel and allowed the particles to move inside the GVF along the direction determined by the gradient vector at this pixel. This process was repeated for each particle. The local maxima could easily be detected by thresholding the number of particles accumulated at each pixel, since no or very few particles accumulated at non-maxima and noise pixels [1] [22]. The centres of cells are therefore being represented by these detected local maxima. With these cell centres as seeded watershed [23], seeded watershed separated cells from the background. A comparison between the proposed method and the strategy combining traditional watershed and fragments merging algorithms is given in Figure 2. The proposed method has fewer over-segmentation errors.

**Fig.2** Segmentation results of the proposed cell segmentation algorithm. (a) Segmentation result with the proposed method. (b) Segmentation result with traditional watershed based method.

### 3. Feature Selection

Feature vectors, each containing 211 features, are generated from the segmented nuclei. Each feature vector is composed of 10 general image features about shape, size, and intensity (max intensity, min intensity, deviation of gray level, average intensity, length of long axis, length of short axis, long axis/short axis, area, perimeter) [24]; 14 Haralick co-occurrence textural features [25]; 47 Zernike moment features [26] [27]; 85 features generated by Gabor transformation [28]; and 54 shape features [1]. Feature extraction is time-consuming and computationally demanding. Moreover, some of the features are redundant or irrelevant to the classification task. All these factors encourage us to adopt the feature selection strategy to find the optimal/sub-optimal feature subset to improve the prediction accuracy and computation speed. A prediction risk-based feature selection method is employed to choose the sub-optimal feature sets [29-30], aiming to remove the irrelevant features and improve the performance of the learning system. This method employs an embedded feature selection criterion of prediction risk, which evaluates features by calculating the change when the corresponding feature is replaced by its average value. Compared with wrapper and filter models, it has several advantages. (1) The embedded feature selection model couples with learning machines being used. Its prediction accuracy relies both on the feature subset and the learning machine. It can reach a higher accuracy than the filter model, while often having lower computation complexity than the wrapper model. (2) According to experiments with multi-class SVMs on 10 public data sets of University of California Irvine (UCI), the prediction risk criterion has outperformed Optimal Brain Damage [30]. (3) It is easy to implement this method. 58 features are kept for cell phase identification, comprised of 37 Gabor features, 1 geometric feature, 14 moment features, 2 texture features, etc. The geometric feature is "perimeter". Gabor features can describe the nuclear both in the time and frequency domain, which may explain why so many Gabor features are kept. Figure 3 illustrates the trend of the prediction accuracy while increasing the number of features used in the prediction risk-based feature selection method.

**Fig.3** The trend of the prediction accuracy while increasing the feature number. C-SVM with C=100 and σ=1 was used in the prediction risk-based feature selection method.
4. Online SVM

SVM[31] has been proven as a classifier with good overall performance, mapping inputs into a higher dimensional space wherein an optimized linear division with least errors and maximal margin is sought. Its training process guarantees a globally optimized solution, avoidance of over-fitting, and insensitivity to feature dimensionality. A complete description of the theory of SVMs for pattern recognition is given in [31]. In contrast to conventional off-line learning algorithms for classification, the on-line adaptivity is incorporated into the algorithmic design to accommodate the ever-changing experimental conditions. An Online Support Vector Classifier (OSVC), which keeps removing support vectors from the old model and assigning new training examples weighted according to their importance, is thus proposed.

4.1 Problem formulation

Given a set of \( I \) samples, i.e., a series of input vectors \( \mathbf{x}_i \in \mathbb{R}^d \) (\( i = 1,...,I \)), where \( \mathbf{x}_i \) is the \( i^{th} \) vector, and \( \mathbb{R}^d \) is a Euclidean space with \( d \) dimensions. Since the multi-class identification problem can always be converted into a two-class identification problem, the formulation below is given for the two-class case only, without any loss of generality. Suppose the output is expressed by \( y_i \in \{+1,-1\} \) (\( i = 1,...,I \)),

where the indexes –1 and +1 represent two classes. The goal is to construct one binary classifier or derive one decision function from the available samples that has a small probability of misclassifying an unknown future sample. To achieve this goal in the batch setting, C-SVM[31] minimizes the following objective function:

\[
\min_{\omega,b} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^I \xi_i \\
\text{such that } \ y_i (\omega^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \\
\xi_i \geq 0, i = 1,...,I
\]

(1)

where \( C \) is a constant, \( \xi_i \) are slack variables, \( b \) is the bias term, \( \omega \) is the weight vector, and \( \phi \) is a function that maps an example into the feature space.

The Lagrange multiplier method is applied to solve the optimization problem:

\[
L(\omega, \xi, b, \alpha_i, \beta_i) = \frac{1}{2} \omega^T \omega + C \sum_{i=1}^I \xi_i - \sum_{i=1}^I \left( \alpha_i \left( y_i (\omega^T \phi(\mathbf{x}_i) + b) + \xi_i \right) + \beta_i \xi_i \right)
\]

(2)

where \( \alpha_i \geq 0, \beta_i \geq 0, \delta \geq 0 \) are all Lagrange multipliers. Its dual problem can be shown as:

\[
\max_{\alpha} \quad D(\alpha) = \sum_{i=1}^I \alpha_i - \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^I \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\
\text{such that } \alpha_i \leq C, i = 1,...,I \\
\sum_{i=1}^I \alpha_i y_i = 0
\]

(3)

The gradient of \( D(\alpha) \) is denoted as \( g_i = (g_{i1},...,g_{iL}) \cdot g_i \) is calculated by:

\[
g_i = \frac{\partial D(\alpha)}{\partial \alpha_i} = y_i - \sum_{j=1}^I \alpha_j y_j K_{ij} = y_i - \hat{y}(\mathbf{x}_i) + b
\]

(4)

where \( K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) \) is the kernel function, and \( \hat{y} \) is the predicted value.

The decision function is given in the form:

\[
\hat{f}(\mathbf{x}) = \text{sign} \left( \sum_{i=1}^I y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b \right).
\]

(5)

In this paper, the RBF kernel

\[
k(\mathbf{x}, \mathbf{x}') = \exp(-\frac{||\mathbf{x} - \mathbf{x}'||^2}{\gamma})
\]

is used.

An array of algorithms has been developed to solve the SVM QP problem [31-32] (such as eq.1). They work by continually searching from the current support vector \( \mathbf{u} \) and extending along the specified feasible direction \( \mathbf{w} \). The Sequential Minimal Optimization (SMO) algorithm chooses the direction with only two non-zero elements, which are determined by the so-called \( \tau \)-violating pair. However data is supplied to all these algorithms in batches and thus a large amount of computation is involved. Recently, various online SVM algorithms [1, 33-35] have been proposed to extend the SVM to the online setting. The standard online SVM algorithms are discussed in the binary classification setting without addressing the issues that ensue when different classes are of different levels of importance. Due to the ever-changing experimental conditions, the model has to be updated periodically. Biologists hope that after labelling only one or two “movies” of microscopy images, the updated classifiers will automatically classify new examples with better accuracy. However, before applying online SVM to the task of cell phase identification; three difficulties have to
be circumvented. First, the data sets are critically imbalanced. The classification accuracy will be undesirably biased toward the classes with more samples. Second, sometimes, the classes with fewer samples may be more important than other classes. For example, prophase plays an important role for the identification of the starting point of the mitosis process, but there are only about 140 examples of prophase in a movie of 200 frames of microscope images. Last, the classification problem should be addressed in the multi-class setting.

4.2 Online Support Vector Classifier

Given a previously trained model and a set of new examples at hand, i.e., a series of input vectors \( x_i \in \mathbb{R}^d \), the online learning will be discussed under the framework proposed by Shalev-Shwartz and Singer [36]. In the online setting, on trial \( t \), where \( t \in [1, T] \), the online learning task can be regarded as solving the following optimization problem:

\[
\min_{\omega \in \mathbb{R}^d} P(\omega) = \frac{1}{2} \omega^T \omega + C \sum_{i=1}^{m} \xi_i
\]

with the first \( t-1 \) examples being \( \{(x_1, y_1), \ldots, (x_t, y_t)\} \).

Similarly, the dual objectives function of \( \min_{\omega \in \mathbb{R}^d} P(\omega) \) is given by:

\[
D(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j)
\]

(7)

From the definitions of \( D(\alpha) \) and \( D(\omega) \), we can deduce that \( D(\alpha_1, \ldots, \alpha_m) = D(\alpha_2, \ldots, \alpha_m, 0, \ldots, 0) \). Thus, the online-SVM can be treated as an incremental solver of the problem \( \max_{\omega \in \mathbb{R}^d} P(\omega) \) s.t. \( \forall i x_i \in \mathbb{R}^d \). This problem can be solved in a sequential manner. For example, on trial \( t \), \( D(\alpha) \) only depends on the first \( t \) observed variables. Intuitively, the larger the increase in the dual objective on each trial, the better the online SVC adjusts itself [36].

An online support vector is proposed for the problem of the cell phase identification in the aforementioned framework. Three improvements were made to meet the real applications' demands. First, only the misclassified examples are fed into the program to update the model. Because new examples are critically imbalanced, the dominant class, i.e. interphase, will bring overwhelming information compared to other classes. Second, when being added into the current support vector set, the coefficient \( \alpha_i \) of the new sample is initialized with different weights according to the importance of each class [36,37]. Finally, the support vectors of the old model will be discarded once they become obvious non-support vectors during the optimization process. “One-Verse-One” strategy is adopted to convert binary SVM into the multi-class version to handle multiple classes [31].

This program maintains four pieces of information: the set \( S \) of potential support vector indices, the set \( S_{old} \) of support vector indices belonging to the old model, the coefficients \( \alpha_i \) of the old kernel expansion and the partial derivative \( g_i \). The procedures INSERTION and UPDATING are two basic blocks constituting OSVC (Online Support Vector Classifier). The basic idea of OSVC can be summarized as below. (1) Load the previously trained model, (2) the procedure INSERTION attempts to insert the misclassified example into the current kernel expansion, and (3) the procedure UPDATING updates the model.

Algorithm (Online Support Vector Classifier)

1) \textbf{Initialization:}
   
   Load the old support vector model.
   
   \( S \leftarrow S_{old} \); \( b \leftarrow b_{old} \)

2) \textbf{Online Iteration:}
   
   Set \( \{x_k, y_k\}, \) for \( k = 1, 2, \ldots, l \)
   
   for \( k = 1, \ldots, l \) do
   
   Obtain a new example \( S_i = \{x_i, y_i\} \)
   
   Compute \( y = \sum_{j=1}^{m} \alpha_j y_j K(x_i, x_j) + b \)
   
   if example \( x_i \) is misclassified then
   
   INSERTION ( \( k \) )
   
   UPDATING ( \( k \) )
   
   end if
   
   end for
In **INSERTION**, the misclassified example \(x_k, k \notin S\) is inserted into the current support vector set. The coefficient \(\alpha_k\) is assigned with the preset values \(C_p\) for positive example and \(C_n\) for negative example [36-37]. The aim of assigning different weights is twofold. (1) The algorithm’s performance is also impaired when the unbalanced training sets are used. When the class of interest only has limited training samples, its prediction accuracy will undesirably decrease. To solve this problem, we can assign greater weight to the class of interest and thus improve its accuracy. (2) When the experimental condition changes dramatically, the new examples can be assigned with larger weights to reflect the changes in the experiment. Consequently, the online adaptivity of our algorithm can be improved. Finally, a direction search will be performed to update the coefficients \(\alpha_i\) and \(g_i\).

**PSEUDO CODE of INSERTION** (\(k\)):

Obtain a new example \(S_i = \{x_i, y_i\}\)

Compute \(\hat{y} = \sum_{i \in S} \alpha_i K(x_i, x_k) + b\)

if example \(x_k\) is misclassified, then

if \(y_k = +1\) then

\(\alpha_k \leftarrow C_p, g_k \leftarrow y_k - \sum_{i \in S} \alpha_i y_i K(x_i, x_k), S \leftarrow S \cup \{k\}\)

\(i \leftarrow k, j \leftarrow \arg\min_{i \in S} g_i\ \text{with} \ \alpha_i < C_n\)

else

\(\alpha_k \leftarrow C_n, g_k \leftarrow y_k - \sum_{i \in S} \alpha_i y_i K(x_i, x_k), S \leftarrow S \cup \{k\}\)

\(i \leftarrow k, j \leftarrow \arg\max_{i \in S} g_i\ \text{with} \ \alpha_i < C_p\)

end if

if \((i, j)\) is a \(\tau\)-violating pair

\(\lambda \leftarrow \min\left\{\frac{g_i - g_j}{K_{ij} + K_{jj} - 2K}, C_p - \alpha_i, C_n - \alpha_j\right\}\)

\(\alpha_i \leftarrow \alpha_i + \lambda, \alpha_j \leftarrow \alpha_j - \lambda\)

\(g_i \leftarrow g_i - \lambda(K_{ii} - K_{ij})\ \forall s \in S\)

end if

end if

The procedure **UPDATING** keeps searching for the \(\tau\)-violating pair from the current support vector set and updating coefficients \(\alpha\) to increase the dual objective until there are no such pairs left. Analogous to **REPROCESS** in LASVM, the obvious non support vectors belonging to the old model will be removed[35].

**PSEUDO CODE of UPDATING** (\(k\)):

While there exists the \(\tau\)-violating pair \((i, j)\) do

\(\hat{\lambda} \leftarrow \min\left\{\frac{g_i - g_j}{K_{ij} + K_{jj} - 2K}, C_p - \alpha_i, C_n - \alpha_j\right\}\)

\(\alpha_i \leftarrow \alpha_i + \hat{\lambda}, \alpha_j \leftarrow \alpha_j - \hat{\lambda}\)

\(g_i \leftarrow g_i - \hat{\lambda}(K_{ii} - K_{ij})\ \forall s \in S\)

\(i \leftarrow \arg\max_{i \in S} g_i\ \text{with} \ \alpha_i < C_p\)

\(j \leftarrow \arg\min_{j \in S} g_j\ \text{with} \ \alpha_j < C_n\)

for each \(\alpha_s = 0\) and \(s \in S_{old}\)

if \(y_s = -1\) and \(g_s \geq g_i\), then \(S = S \setminus \{s\}\); end if

if \(y_s = +1\) and \(g_s \leq g_i\), then \(S = S \setminus \{s\}\); end if

\(b \leftarrow (g_i + g_j)/2\)

end for each

end while
5. Experiments and Results

Sensitivity and specificity are two common measures for the evaluation of experimental results. Suppose TP, TN, FP and FN stand for the number of true positive, true negative, false positive and false negative samples respectively after the completion of cell phase identification. Sensitivity is defined as \( \text{sensitivity} = \frac{TP}{(TP+FN)} \), and specificity is defined as \( \text{specificity} = \frac{TN}{(TN+FP)} \). In other words, sensitivity is a statistical measure of how well the positive cells are classified, while specificity reflects the ability to identify negative cells correctly. We can calculate the sensitivity and specificity for each class if we take one class as positive and other classes as negative.

Three “movies” are manually labelled for validating the OSVC algorithm. These movies are taken from three successive experiments with different experimental conditions. The first two movies are untreated and the third is treated with taxol. Each movie contains 240 images. SVM [31] and LASVM [35] are used as baseline algorithms for comparisons. The second and third movies are each separated into two halves. The first halves are fed into algorithm as training sets and the second as testing sets. In the testing set of the second movie, there contain 9,249 interphase cells, 71 prophase cells, 183 metaphase cells and 180 anaphase cells. In the second half of the third movie, there are 13,930 interphase cells, 48 prophase cells, 539 metaphase cells and 225 anaphase cells. For the untreated case, SVM [31] is trained with the cells in the first movie and the first half of the second movie. For the LASVM and OSVC, the “old” model is generated with the first movie while the first half of second movie is used to update the “old” model with the proposed online learning algorithm. For the treated case, the combination of the first and the third movie is used, in contrast to the untreated case wherein the first and the second movie are used. The parameters obtained with cross validation test are used for all algorithms: \( C = 0.707 \) and the \( \sigma \) value for RBF kernel is 0.25. In OSVC, both weighted and non-weighted cases are tested.

In weighted OSVC, the weights for interphase, prophase, metaphase, and anaphase are 1, 30, 10, and 10 respectively. In tables 1, 2, 3 and 4, the third rows of "OSVC weighted" correspond to the weighted cases. In the non-weighted case, the \( C_1 \) and \( C_k \) in \text{UPDATING} (k) are initialized with zeros. Tables 1 to 4 demonstrate the sensitivity and specificity of both treated and untreated cases. Each line is an approach and each column is a phase. From table 2 and 3, the sensitivity of prophase with LASVM decreases while OSVC increases the sensitivity greatly for both the untreated and treated cases. The weights of OSVC are allowed to be tuned adaptively, a useful feature when the focus is on the classes with limited training samples.

<p>| Table 1. Sensitivity of cell phase identification for the untreated case |</p>
<table>
<thead>
<tr>
<th>PHASE</th>
<th>INTER</th>
<th>PRO</th>
<th>META</th>
<th>ANA</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>99.6%</td>
<td>56.3%</td>
<td>67.2%</td>
<td>66.6%</td>
</tr>
<tr>
<td>LASVM</td>
<td>99.8%</td>
<td>45.1%</td>
<td>77.2%</td>
<td>82.0%</td>
</tr>
<tr>
<td>WOSVC</td>
<td>94.9%</td>
<td>87.3%</td>
<td>66.1%</td>
<td>84.0%</td>
</tr>
<tr>
<td>OSVC</td>
<td>99.2%</td>
<td>77.1%</td>
<td>87.9%</td>
<td>89.4%</td>
</tr>
</tbody>
</table>

<p>| Table 2. Sensitivity of cell phase identification for the treated case |</p>
<table>
<thead>
<tr>
<th>PHASE</th>
<th>INTER</th>
<th>PRO</th>
<th>META</th>
<th>ANA</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>99.4%</td>
<td>18.8%</td>
<td>56.4%</td>
<td>77.8%</td>
</tr>
<tr>
<td>LASVM</td>
<td>98.5%</td>
<td>22.9%</td>
<td>75.1%</td>
<td>86.7%</td>
</tr>
<tr>
<td>WOSVC</td>
<td>90.9%</td>
<td>77.0%</td>
<td>79.2%</td>
<td>77.3%</td>
</tr>
<tr>
<td>OSVC</td>
<td>98.5%</td>
<td>25.0%</td>
<td>70.9%</td>
<td>83.6%</td>
</tr>
</tbody>
</table>
Table 3. Specificity of cell phase identification for the untreated case

<table>
<thead>
<tr>
<th>PHASE</th>
<th>INTER</th>
<th>PRO</th>
<th>META</th>
<th>ANA</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>67.7%</td>
<td>99.7%</td>
<td>99.8%</td>
<td>99.9%</td>
</tr>
<tr>
<td>LASVM</td>
<td>80.7%</td>
<td>99.9%</td>
<td>99.8%</td>
<td>99.9%</td>
</tr>
<tr>
<td>WOSVC</td>
<td>95.9%</td>
<td>99.6%</td>
<td>99.8%</td>
<td>99.8%</td>
</tr>
<tr>
<td>OSVC</td>
<td>80.1%</td>
<td>99.6%</td>
<td>99.0%</td>
<td>98.6%</td>
</tr>
</tbody>
</table>

Table 4. Specificity of cell phase identification for the treated case

<table>
<thead>
<tr>
<th>PHASE</th>
<th>INTER</th>
<th>PRO</th>
<th>META</th>
<th>ANA</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>67.0%</td>
<td>99.6%</td>
<td>99.8%</td>
<td>99.6%</td>
</tr>
<tr>
<td>LASVM</td>
<td>80.6%</td>
<td>99.9%</td>
<td>99.8%</td>
<td>99.8%</td>
</tr>
<tr>
<td>WOSVC</td>
<td>95.9%</td>
<td>99.6%</td>
<td>99.8%</td>
<td>99.8%</td>
</tr>
<tr>
<td>OSVC</td>
<td>80.9%</td>
<td>99.6%</td>
<td>99.0%</td>
<td>98.6%</td>
</tr>
</tbody>
</table>

6. Discussion and Conclusion

In this chapter, the general framework of microscope image analysis as well as a laboratory system is introduced to demonstrate the basic ideas underlying image based drug discoveries. In particular, the feature extraction, selection and online SVM algorithms in a laboratory system are demonstrated. Future work is to construct a database from vast amounts of images of cancer cell lines under different drug perturbation conditions. Following this, the influence of various drug components in the mitotic process will be analyzed to find the key anti-mitotic cancer drug components.

Image segmentation, a hot research area, is one of the most critical step toward a successful automatic microscopy system. A watershed based segmentation scheme is applied because it is capable of handling the segmentation problem of touching and overlapping cells by computing the watershed line on the complement of distance transformation. However, we have to emphasize that there are many other choices: level set, clustering methods, histogram-based methods, region growing methods, model based segmentation, etc. Each method shows unique advantages under certain circumstances.

For the study of subcellular location, we can extract the subcellular features, developed by Dr. Murphy's Lab [38]. They are the number of fluorescent objects in the image, the Euler number of the image, the average of the number of above-threshold pixels per object, etc. A more commonly used and robust shape descriptor is the Fourier Descriptors, which achieves both better representation and better normalization [39].

In the OSVC, the initial weights of new samples being inserted into the current kernel set are assigned by the relative importance of the samples' class. These values are adaptive and still lack a theoretical guidance. In the updating procedure, the obliteration of non-support vectors in old models is determined by the optimization procedure. More constraints can be imposed on this process: such as ensuring that the newly added samples have less chances of being discarded than those older samples, etc.

7. Further reading and Advanced topics

As powerful as it is in classification tasks, SVM is nonetheless incapable of providing feature importance. SVM can combine with various feature selection strategies besides the method mentioned above. Some of them are “filters”: general feature selection methods independent of SVM. Another category is wrapper methods: modified SVMs both choose important features as well as conduct training/testing. Feature selection can also be conducted during the process of training SVMs, which induces some embedded feature selection methods [30]. Some popular methods include: F-score for feature selection; F-score and Random Forest for Feature Selection; and Random Forest and RM-bound SVM for feature selection. Interested readers can refer to the book [40] for more information.

In the military setting, a variety of pattern recognition techniques have been widely applied to identify and track potentially hazardous objects such as tanks or airplanes on radar images. As a crucial aspect of dynamic cell analysis, a good tracking algorithm allows us to track the behaviour of individual cells across image frames from different time points in the temporal sequences, such as during mitosis, and compare their morphological variations. Major hurdles for
a effective tracking system are: (1) existence of touching cells owing to under-segmentation; (2) existence of fragmental cells owing to over-segmentation. (3) hundreds of cells requiring to be tracked within each image across image frames at three different sites from the well; (4) there being about three to five thousand cells in each well; (5) the morphologic features of cells undergoing mitosis may vary greatly; (5) the complication of cell-cycle phases; (6) the nonlinearity of cell movement; (7) the non-Gaussian environments. A simple method may utilize the location information for the matching between cells on two consecutive frames. The correspondence ambiguity could be resolved by size and location information. However, other salient features and cell phase information can also be incorporated into the process of modelling and algorithm design [41]. Other choices include kalman filter [42], active contours [43], level sets analysis [44] and Monte Carlo method [45].

Acknowledgements This work is partially founded by initial funding of Nanjing University and a NIH grant R01 LM008696. JinJing Xu at Nanjing University has proofread this chapter.

References