Learning to Detect Different Types of Cells under Phase Contrast Microscopy

Jiyan Pan[1], Takeo Kanade[1], and Mei Chen[2]

[1]Robotics Institute, Carnegie Mellon University, 5000 Forbes Avenue, PA 15213, U.S.A.
[2]Intel Labs Pittsburgh, 4720 Forbes Avenue, PA 15213, U.S.A.

***Abstract*— We propose a learning-based algorithm to detect various types of cells in phase contrast microscopy images. The algorithm automatically adapts to different cell types by learning variations in cell appearances and shapes, while staying discriminative against non-cell distractions in the background. Benefiting from a rich set of carefully designed feature descriptors, the proposed algorithm is able to detect considerably different types of cells with high accuracy. With experimental results achieving an average detection F-measure of 90%, the statistical learning framework proposed in this paper has proven to be promising in evolving into a highly adaptive and robust cell detection system.**

I. Introduction

Computer vision based automated analysis of stem cell behaviors plays a crucial role in facilitating rapid biological discoveries and effective clinical applications. Unlike humans, computers can monitor cell population continuously for an extended period of time, during which many otherwise unseen or unnoticed phenomena could be discovered [1].

In automated systems, single-frame cell detection is a key component, as cells must be reliably detected in each frame before further analysis such as tracking could be effectively performed [2]-[4]. Cell detection entails identifying a subset of pixels belonging to the cell region, and grouping them to each individual cell. Although temporal information could be employed to help disambiguate detection results [2], [8], the overall system performance would benefit from improved single-frame cell detection. However, robust single-frame cell detection is challenging, as shown in a phase contrast microscopy image of C2C12 muscle stem cells on the left side of Fig. 1, where hundreds of cells take a large variety of shapes, appearances and sizes in a background with many distractions (such as the black rings). The task becomes harder if the detection algorithm is expected to be applicable to various types of cells under different imaging conditions without tuning any parameters. For example, both the cells and the background in the right image of Fig. 1 exhibit different characteristics than those in the left image. This paper focuses on developing an effective single-frame cell detection algorithm that could automatically adapt to intra-type and inter-type variations of cell shapes and appearance in phase contrast microscopy images. To our knowledge, little effort has been devoted to solving this problem in the literature.

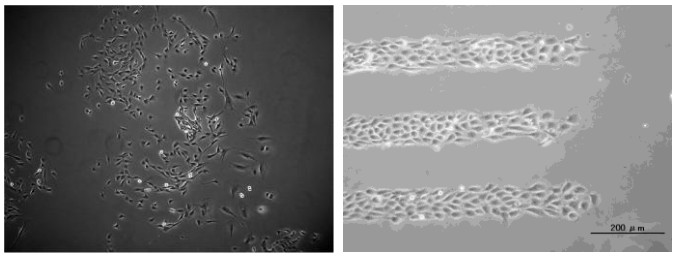


Fig. 1. Examples of two different types of cells under phase contrast microscopy. Left: C2C12 muscle stem cells. Right: bovine aortic endothelial cells

Many existing approaches detect cells by making use of the fact that cells in phase contrast microscopy images appear as darker regions surrounded by brighter halos. Usaj et al. convolve an inverted Laplacian of Gaussian (LoG) with the image and pick the local extrema as cell centers [5]. This approach is ineffective when the sizes of cells vary across the image. Another method is proposed by Ersoy et al. which selects local extrema of image intensity as candidate cell points, and filters them by comparing their principal intensity curvature to a threshold [6]. This method is vulnerable to background distractions as they also form local extrema with large intensity curvature. In a different approach, intensity histograms are used to distinguish cells from background with Otsu's thresholding [7] and/or maximum *a posteriori* criterion [8], [9], [11]. However, such a generative approach based on pixel intensity alone is not discriminative when cells are brighter or background is darker than usual. Furthermore, this method is not applicable to cases in which cells and background have similar intensity distributions. To improve the performance of cell detection, Li et al. propose a powerful preconditioning method that transforms the captured image so that cell regions obtain higher intensity in the transformed image [13]. Nevertheless, prior knowledge of the imaging system must be acquired, and the parameters of this method require careful manual tuning for each cell type.

Grouping detected cell pixels into individual cells is equally challenging. The most straightforward way takes each local extrema as a cell [5]. Another approach applies a threshold to the cell-likelihood of the detected pixels, and the connected regions in the thresholded image are associated with

individual cells [8], [9], [11], [13]. The method in [6] finds accurate cell boundaries by employing geodesic active contour evolution. However, this algorithm is not able to separate different cells when they touch on one another. An enhanced watershed algorithm is developed to separate the detected cell regions that are connected [7]. Nevertheless, this method works well only when the connected regions share a relatively short boundary.

A common problem facing all the aforementioned algorithms is that they make strong assumptions of cell structures and rely on rigid criteria. As a consequence, they are not able to adapt to a wide range of complex scenarios.

To enhance adaptability, we propose to detect cells using statistical learning techniques that enable the algorithm to learn the underlying structures of both cells and background. After candidate detection points are generated by preprocessing steps, a rich set of image features (called unary features) is extracted around each of those points, and a discriminative model is trained to determine if a candidate detection point indeed belongs to a cell. Another set of image features (referred to as pairwise features) is computed for each pair of cell points to train another discriminative model that predicts if the two cell points belong to the same cell or different cells, which is then used by a grouping algorithm to cluster points belonging to the same cells. The overall scheme of the proposed algorithm is illustrated in Fig. 2.

As modern learning algorithms could account for much within-class variation, the proposed algorithm is highly flexible. Meanwhile, the discriminatively trained models make the algorithm effective in rejecting non-cell regions. While statistical learning techniques have found much success in natural image understanding [10], [12], [14], they are seldom applied to cell detection in microscopy images. Although some researchers use learning-based approaches to analyze microscopic images, most of them are for purposes other than cell detection [21], [22]. This paper makes an initial attempt to address the cell detection problem within the framework of statistical machine learning. A high level of detection accuracy (with F-measures around 90%) is achieved for both the disparate cell types in Fig. 1 without the need to manually tune any parameters.

The remainder of this paper is organized as follows. Section II describes the preprocessing steps that generate candidate detection points. The classification of candidate detection points into cell or background is detailed in Section III. Section IV discusses the grouping of multiple detection points belonging to the same cell. Experimental results are given in Section V, and Section VI concludes the paper.

II. Preprocessing

The purpose of preprocessing is to mask out evident non-cell pixels, and generate a relatively small number of candidate detection points for subsequent feature extraction, classification, and grouping. Here, a 100% recall is required while the number of points for further processing should be kept as low as possible.

The preprocessing involves three steps. Firstly, a large

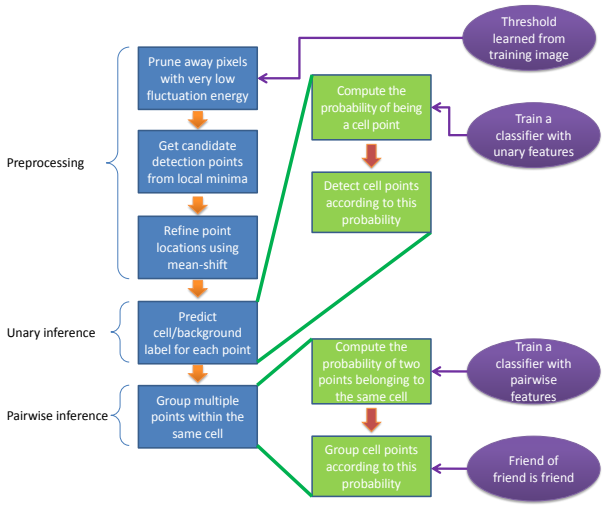


Fig. 2. Overall scheme of the proposed algorithm

portion of the texture-less image region can be directly determined as background and pruned away from further processing. To perform this task, a filter bank shown in Fig. 3 is applied to the image. The filter bank consists of a set of anisotropic Laplacian filters at 6 different orientations and an isotropic Laplacian filter [16], [20], all of which have 6 different scales. The squared responses of all the filters are summed to obtain the fluctuation energy at each pixel. The fluctuation energy of individual pixels is further normalized by the average fluctuation energy over the entire image to make this measure contrast invariant. Any pixel whose *normalized* fluctuation energy is lower than a pruning threshold is regarded as a background pixel and discarded. To maintain a high recall, the pruning threshold is set to be the *lowest* normalized fluctuation energy over the cell region in the *training* image for which the ground truth of cell segments is known.

Secondly, among the remaining pixels, the local minima of image intensity are selected as candidate detection points, based on the observation that all cells under phase contrast microscopy contain at least one local minimum of intensity, despite the great variation of their appearances. The size of the image patch over which a local minimum is defined is the *smallest* spatial span of a cell which is obtained from the training image. This way, no cell is missed out and a high recall is preserved.

Thirdly, the locations of candidate detection points are refined. Many candidate detection points that do belong to cells lie close to cell boundaries. For the convenience of further processing, it would be desirable to center those detection points. To achieve this goal, an algorithm similar to mean-shift mode detection [17] is employed. The location, $\boldsymbol{x}$, of a candidate detection point is updated iteratively. For each iteration,

Fig. 3. The filter bank consists of six groups of such filters, with standard deviations increasing from 1.5 pixels with a half-octave increment.

(a) (b)

Fig. 4. The effect of refining candidate detection points using mean-shift. In (a) and (b), candidate detection points are represented by red dots; the left and right image shows the situation before and after the mean-shift operation, respectively.

$$\boldsymbol{x}^{(t+1)} = \frac{\Sigma_{\boldsymbol{y}\in\mathcal{G}_{\boldsymbol{x}}^{(t)}}\left[\left(1-I(\boldsymbol{y})\right)-\min_{\boldsymbol{y}\in\mathcal{G}_{\boldsymbol{x}}^{(t)}}\left(1-I(\boldsymbol{y})\right)\right]\cdot\boldsymbol{y}}{\Sigma_{\boldsymbol{y}\in\mathcal{G}_{\boldsymbol{x}}^{(t)}}\left[\left(1-I(\boldsymbol{y})\right)-\min_{\boldsymbol{y}\in\mathcal{G}_{\boldsymbol{x}}^{(t)}}\left(1-I(\boldsymbol{y})\right)\right]}, \quad (1)$$

where $\mathcal{G}_{\boldsymbol{x}}^{(t)}$ is a square grid centered at $\boldsymbol{x}^{(t)}$, and $I(\boldsymbol{y})$ is the image intensity at position $\boldsymbol{y}$, ranging from 0 to 1. Note that both $\boldsymbol{x}$ and $\boldsymbol{y}$ can take non-integer values. Therefore, $I(\boldsymbol{y})$ needs to be interpolated. The size of $\mathcal{G}_{\boldsymbol{x}}^{(t)}$ is twice the standard deviation of the smallest filters in the filter bank. Unlike conventional methods, the weight of each point is the relative darkness at that point with respect to the minimal darkness over $\mathcal{G}_{\boldsymbol{x}}^{(t)}$. Such normalization results in much faster convergence by emphasizing relative difference in darkness.

The mean-shift operation refines the locations of candidate detection points in two aspects. Firstly, they are more centered within cells, as shown in Fig. 4(a). The reason behind this effect is that mean-shift converges when the "pulling forces" coming from around a point are balanced. In addition to centering the detection points, some of the points that are close together within the same cell end up converging to the same point, as shown in Fig. 4(b). Note that if two detection points belong to two different regions separated by a boundary, they will not get merged even if they are close to each other.

III. Predicting Cell/Background Labels

In this section, we discuss classifying each candidate detection point generated by the preprocessing steps into either belonging to cells or to the background. Due to the high variability of cell shapes and appearances, a discriminative model is expected to outperform a generative model. Therefore, we train a discriminative model to directly map image features to class labels (i.e. cell or background).

For this discriminative model, the image region to extract features of a point is a square block centered at that point. The orientation and scale of the block is the orientation and scale of the filter that yields the largest response over the filter bank (referred to as the "best filter" in what follows). When the best filter is an isotropic Laplacian filter, the orientation of the block is set as zero. The block is then divided into 9 sub-blocks, as is illustrated in Fig. 5. Five types of features are extracted for each sub-block, which are detailed below.

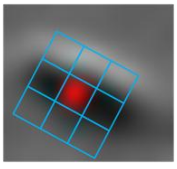


Fig. 5. Illustration of the image block and sub-blocks from which features are extracted.

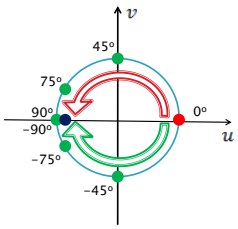


Fig. 6. Illustration of the mapping of relative gradient orientations to points on the unit circle. The red dot represents the orientation of the best filter, the green dots indicate several relative gradient orientations, and the blue dot shows the average of 75° and -75°.

As cells under phase contrast microscopy usually have bright halos, the first feature we use is normalized mean intensity, defined as the mean intensity of the sub-block normalized by that of the center sub-block. This feature reflects the center-surround intensity ratio.

Since the spatial distribution of edgeness around a cell point is quite different from that of a background point, the second feature is selected as the mean gradient magnitude of the sub-block.

To further characterize the orientation distribution of the edges, the Histogram of Oriented Gradient (HOG) [15], [12] is used as the third feature. HOG computes the histogram of gradient orientations weighted by gradient magnitudes. Here, the size of orientation bins is set as 45°.

Since a cell could be rather small, the sub-block sometimes does not contain sufficient pixels to generate a reliable histogram. To remedy this problem, we extract the fourth feature: mean relative gradient orientation weighted by gradient magnitude. Relative gradient orientation $\theta(\boldsymbol{x})$ of a pixel $\boldsymbol{x}$ is defined as

$$\theta(\boldsymbol{x}) = \begin{cases} \theta'(\boldsymbol{x}) - \theta_0 + \pi, & \theta'(\boldsymbol{x}) < \theta_0 - \pi/2 \\ \theta'(\boldsymbol{x}) - \theta_0, & |\theta'(\boldsymbol{x}) - \theta_0| \le \pi/2, \\ \theta'(\boldsymbol{x}) - \theta_0 - \pi, & \theta'(\boldsymbol{x}) > \theta_0 + \pi/2 \end{cases} \quad (2)$$

where $\theta'(\boldsymbol{x})$ is the gradient orientation at $\boldsymbol{x}$; θ_0 is the orientation of the best filter at the candidate detection point located in the center block. Both $\theta'(\boldsymbol{x})$ and θ_0 are within $[0,\pi)$. Equation (2) enforces the relative gradient orientation $\theta(\boldsymbol{x})$ to be within $[-\pi/2,\pi/2)$, meaning the polarity of gradient orientation does not matter, since cell boundaries normally contain edges of opposite polarities.

To obtain the mean relative gradient orientation, one cannot simply compute the weighted average of $\theta(\boldsymbol{x})$ because of its cyclic nature. For example, the weighted average of 75° and -75° should be 90° (or equivalently, -90°), rather than 0° (suppose the weights are 1). This problem is solved by mapping $\theta(\boldsymbol{x})$ to a point on a unit circle through $u(\boldsymbol{x}) = \cos[2\theta(\boldsymbol{x})]$, $v(\boldsymbol{x}) = \sin[2\theta(\boldsymbol{x})]$, where $u(\boldsymbol{x})$ and $v(\boldsymbol{x})$ are the coordinates of the point corresponding to $\theta(\boldsymbol{x})$, as is illustrated in Fig. 6. The mean relative gradient orientation is then represented by the weighted average of $u(\boldsymbol{x})$ and $v(\boldsymbol{x})$ within the sub-block.

Note that the mean coordinate $(\bar{u}, \bar{v})$ usually lies *within* the unit circle. The interpretation of such a coordinate is that its angle reflects the "hue" of the mean orientation, while its radius reflects the "saturation" of the mean orientation. The closer the coordinate is to the unit circle, the more uniform the orientations are within the sub-block. Using this definition, the paradox caused by the cyclic nature of the relative gradient orientation no longer exists. As is shown by the blue dot in Fig. 6, the mean of 75° and -75° is now indeed 90°, with a decreased saturation (i.e. uniformity).

To eliminate false detections that are too big or too small to be a cell, the feature set also includes the scale of the best filter at the candidate detection point.

The union of the five types of features of the nine sub-blocks forms the unary features for each candidate detection point. Using the unary features, we employ a non-linear Support Vector Machine (SVM) [18] with a Gaussian kernel to discriminate cell points from background ones. Here, instead of directly using the *hard* predictions of the SVM, the training labels are fitted against the *soft* outputs of the trained SVM using logistic regression. The *soft* SVM outputs of the testing data are converted into posterior class probabilities $P\big(cell|F(\boldsymbol{x})\big)$ using the learned logistic regression model [19]:

$$P\big(cell|F(\boldsymbol{x})\big) = \frac{1}{1+exp(-wS[F(\boldsymbol{x})]-b)}, \quad (3)$$

where $F(\boldsymbol{x})$ is the unary feature vector associated with point $\boldsymbol{x}$; $S[\cdot]$ is the soft SVM output, w and b are the parameters of the fitted logistic regression model. All the points with $P\big(cell|F(\boldsymbol{x})\big)$ above a threshold are classified as belonging to cells. This threshold is automatically determined using cross-validation on the training image.

After predicting the labels of every candidate detection point, all those points labeled as background are removed.

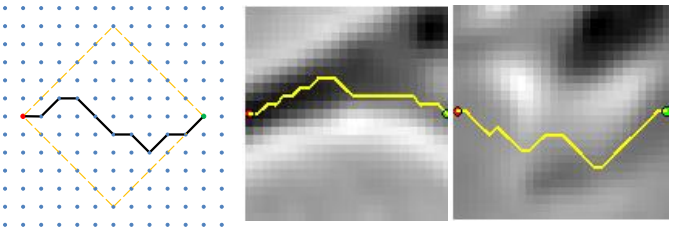
IV. Grouping Detection Points Within The Same Cell

Having obtained the final detection points (i.e. cell points) using the cell/background classifier, we need to explicitly define a model to predict if two cell points are within the same cell, and if yes, group them. The following sub-sections describe the discriminative model that distinguishes within-cell/between-cell relationships and the algorithm to group cell points after estimating the relationships.

A. Predicting within-cell/between-cell relationships

For a given pair of cell points, a set of eight pairwise features are extracted to predict if they are within the same cell or across different cells.

If two cell points are within the same cell, there is typically a path from one point to the other that does not need to "climb" the bright ridge formed by the halos near cell boundaries. Therefore, we need to find a path that connects a pair of cell points with the least connecting effort. Here, "the least connecting effort" means that the maximum intensity along the optimal path is the lowest over all possible paths connecting the two points. That is,



(a) (b)

Fig. 7. (a): Illustration of the search of the optimal connecting path. The dashed orange diamond delineates the search region, and a possible connecting path is illustrated with the black curve. (b): Two examples of the optimal connecting path which is indicated by the yellow curve.

$$\hat{\boldsymbol{c}} = \arg\min_{\boldsymbol{c}}\{\max_{\boldsymbol{x}\in\boldsymbol{c}} I(\boldsymbol{x})\}, \quad (4)$$

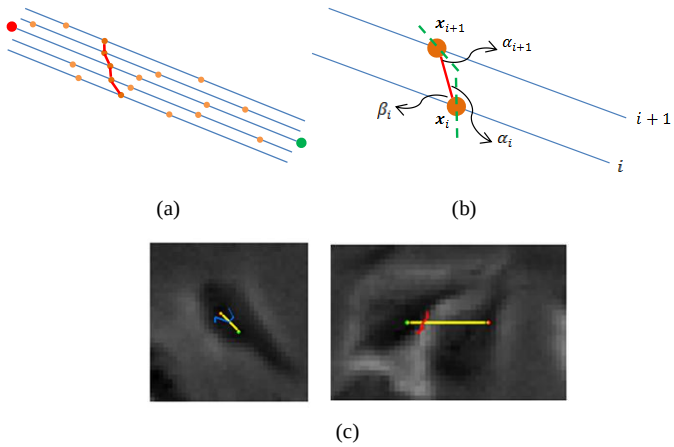
where $\boldsymbol{c}$ denotes a path connecting the two points and $\hat{\boldsymbol{c}}$ is the optimal path with the least connecting effort. This path is able to circumvent isolated bright points within a cell and is therefore robust in determining if the two points belong to the same cell.

To find such an optimal path efficiently, the image patch containing the two points is rotated such that the two points are on the same horizontal line. Then, dynamic programming is employed, in which a pixel is allowed to be linked with its six immediate neighbors to the left and right, and the search region is confined within the orange diamond-shaped region illustrated in Fig. 7(a). When several paths have identical connecting effort, the path with the smallest "bumpiness" is chosen. Here, the *bumpiness* of a path is defined as

$$B = \frac{1}{N_{\boldsymbol{c}}-1}\sum_{i=2}^{N_{\boldsymbol{c}}}|I(\boldsymbol{x}_i) - I(\boldsymbol{x}_{i-1})|, \quad (5)$$

where $N_{\boldsymbol{c}}$ is the number of pixels on the connecting path, and $\boldsymbol{x}_i$ is the i^{th} pixel along the path. Two examples of the optimal connecting path are shown in Fig. 7(b).

After the optimal connecting path is obtained, several



(a) (b)

(c)

Fig. 8. (a): Illustration of generating the best possible boundary. The red and green dots represent a pair of detection points. The red curve is the best hypothesis boundary. (b): Illustration of angular relationships among lines. The green dotted lines represent the direction *perpendicular* to the gradient orientation at the orange dots. (c): Two examples of the best possible boundary which is indicated by the blue/red curve. The color of the curve reflects its hypothetic boundary strength. More reddish means larger strength.

features can be derived from it. The first feature is its intensity profile, obtained by equally dividing the path into five segments, and averaging the intensity of the pixels within each segment. This is essentially a down-sampled version of the pixel intensities along the optimal connecting path. In addition, the bumpiness of the optimal path (see Equation 5), the least connecting effort, and the ratio of the least connecting effort with respect to the intensities of the two cell points are also added to the pairwise feature set.

A weakness of the optimal connecting path is that it could sometimes "sneak" through a boundary gap when two points belong to adjacent cells. As a remedy, we compute *hypothetic boundary strength* as the fifth feature. This feature measures the strength of the most likely boundary between two points even if there is actually no boundary, in which case the strength is low.

The most likely boundary between two points is obtained as follows. Firstly, local maxima of gradient magnitude are identified on each of the five lines parallel to the line joining the two points. These lines have a total width of 5 pixels, as is illustrated in Fig. 8(a), where the local maxima are represented by orange dots. Secondly, dynamic programming is used to select one local maximum $\boldsymbol{x}_i$ on each line i, such that the following objective function is maximized:

$$J_{BS}(\{\boldsymbol{x}_i\}) = \sum_{i=1}^{4} \left(\frac{2}{\pi}|\beta_i|\right)^a \left\{\left(1-\left(\frac{2}{\pi}|\alpha_i|\right)^b\right) G(\boldsymbol{x}_i) + \left(1-\left(\frac{2}{\pi}|\alpha_{i+1}|\right)^b\right) G(\boldsymbol{x}_{i+1})\right\}, \quad (6)$$

where α_i reflects the consistency between the local edge direction and boundary direction, and β_i measures the angle between the boundary direction and the joining direction, as is illustrated in Fig. 8(b). The range of α_i and β_i is always between $-\pi/2$ and $\pi/2$. Equation 6 encourages a boundary with large gradient magnitude, being smooth (i.e. small α_i), and perpendicular to the joining direction of the two cell points (i.e. large β_i). a and b are positive parameters which control the degree of regularization of angles. The specific values of a and b are not critical; setting $a = 0.3$ and $b = 3$ suffices for most scenarios. Two examples of the most likely boundary are shown in Fig. 8(c).

The hypothetic boundary strength is the value of J_{BS} corresponding to the most likely boundary:

$$f_{BS} = \max_{\{\boldsymbol{x}_i\}} J_{BS}(\{\boldsymbol{x}_i\}), \quad (7)$$

In addition to the features extracted above, we also notice that when the two points are within the same cell, their best filter orientations are typically aligned with their joining direction. Therefore, the sixth feature in the pairwise feature set is the relative orientations of the two cell points with respect to the direction of the line joining them.

As additional clues, the pairwise feature set also includes the gradient magnitude and scale of the two cell points.

The union of the eight types of pairwise features is used to train a non-linear SVM with a Gaussian kernel as well as a logistic regression model for probabilistic interpretation, as discussed in the previous section. Here, the probabilistic output is the probability of a pair of cell points belonging to the same cell, $P(within\_cell|F(\boldsymbol{x}_1, \boldsymbol{x}_2))$, given the pairwise feature vector $F(\boldsymbol{x}_1, \boldsymbol{x}_2)$.

B. *The grouping algorithm*

The intuition of the grouping algorithm is: two cell points should be grouped together if they have a high probability of belonging to the same cell according to the discriminative model described above.

Ideally, we could compute $P(within\_cell|F(\boldsymbol{x}_1, \boldsymbol{x}_2))$ for all possible pairs of cell points. However, this turns out to be infeasible and unnecessary. Instead, each cell point $\boldsymbol{x}_0$ *only* has pairwise relationship with its five nearest cell points $\boldsymbol{x}_{N1}, \ldots, \boldsymbol{x}_{N5}$. For each of those five nearest neighbors $\boldsymbol{x}_{Ni}$, it is assigned to the same group as $\boldsymbol{x}_0$ if $P(within\_cell|F(\boldsymbol{x}_0, \boldsymbol{x}_{Ni}))$ is above a threshold which is automatically learned by cross-validation. In many cases, two cell points do not have direct relationship, yet they may still be grouped together through a chain of cell points that are considered as belonging to the same cell in a pairwise sense. In other words, the grouping process has a transitive property, making it unnecessary to deal with all possible pairs of cell points.

V. Experimental Results

We first evaluate the performance of the proposed algorithm on bovine aortic endothelial cells imaged by Olympus IX71 microscope at X10 magnification. The first frame in a time-lapse sequence is used as the training image, and 10 other frames with a very different look are randomly chosen as testing images. The training image and one of the testing images are shown in Fig. 9. We continue to use this testing image as an example in what follows. Results on the other testing images are comparable.

After preprocessing, most of the background region is excluded from further analysis according to the pruning threshold learned from the training image. To evaluate the effectiveness of pruning, the initial candidate detection points computed over the pruned image region are displayed on the right side of Fig. 10, whereas the points computed over the entire testing image are shown on the left side of Fig. 10. (Note that in practice there is no need to compute the initial candidate detection points over the entire image.) After the pruning, 100% of the candidate detection points that belong to cells survive the process, while the total image area to focus on shrinks by 68.28%, and the number of candidate detection points reduces from 5,229 to 701. This suggests that the learned pruning threshold is highly effective.

The pruned points then undergo mean-shift refinement. The refined points are displayed on the left side of Fig. 11. After the refinement, the number of candidate detection points further reduces to 642, and many points become more

centered within cells.

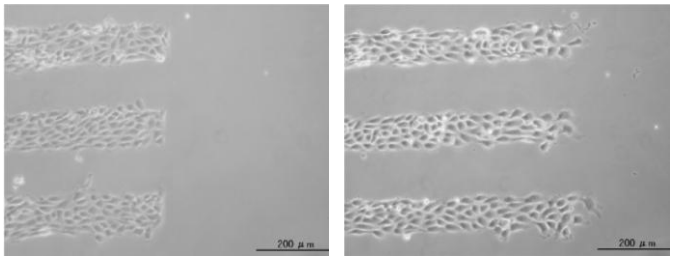


Fig. 9. Training and testing images for bovine aortic endothelial cells. Left: training image. Right: one of the testing images

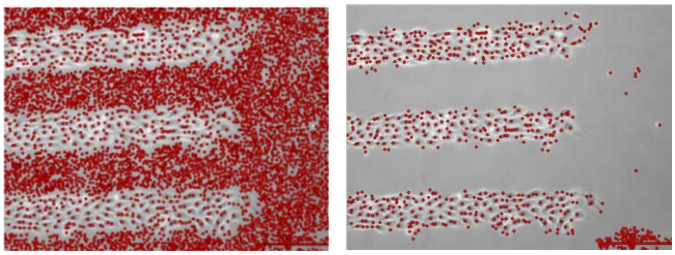


Fig. 10. Pruning of candidate detection points which are indicated in red. Left: before pruning. Right: after pruning.

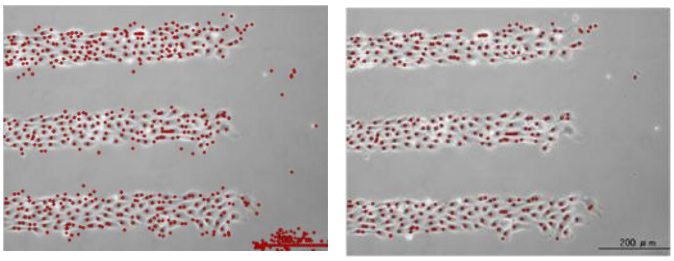


Fig. 11. Refining and classifying of detection points. Left: after refining the candidate detection points. Right: after removing the candidate detection points classified as background.

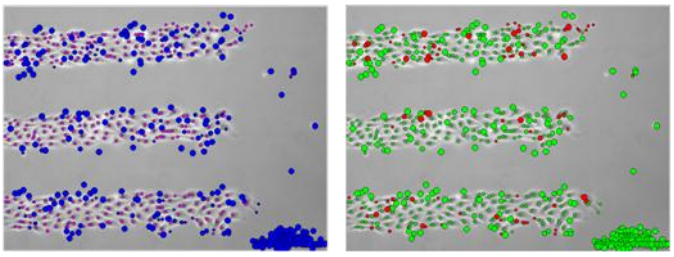


Fig. 12. Labeling of the detection points. Left: Magenta dots indicate cell points and blue dots represent background points. Right: green points indicate correct labels and red points indicate wrong labels.

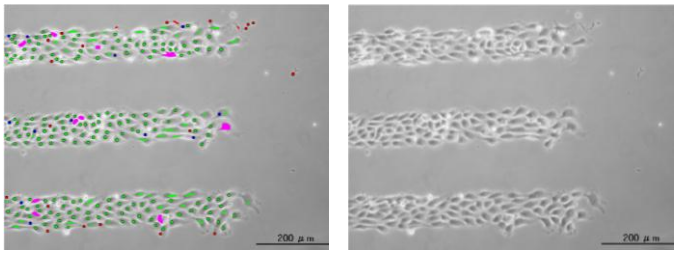


Fig. 13. Left: the final detection result. Right: the testing image.

The probability of each refined point belonging to a cell is computed using the learned cell/background discriminative model. Applying the learned threshold, we get the cell/background label for each refined point, as is shown in the left image of Fig. 12, where magenta dots indicate cell points, and blue dots represent background points. Larger circle size indicates higher classification confidence. The correctness of this labeling is displayed in the right image of Fig. 12, where green dots indicate correct labels, and red dots represent wrong labels. Larger circle size indicates higher degree of correctness (or incorrectness). As we can see, most of the wrong labels occur close to cell/background borders and/or in cluttered ambiguous regions. After removing the candidate detection points classified as background ones, the final detection points are shown in the right image of Fig. 11.

TABLE I
ACCURACY OF THE TWO CLASSIFIERS

	Classifier	Proposed	No Mean-shift
Bovine	Cell/background	88.47%	87.59%
	Within/between cell	90.91%	87.66%
C2C12	Cell/background	91.41%	90.96%
	Within/between cell	90.93%	89.68%

TABLE II
DETECTION PERFORMANCE UNDER VARIOUS SCENARIOS

		Proposed	No Group	No MRGO	No HBS	No MS
Bovine	P	86.59%	70.63%	84.47%	84.64%	75.68%
	R	96.17%	96.17%	94.89%	96.17%	97.82%
	F	91.13%	81.44%	89.38%	90.04%	85.33%
C2C12	P	90.66%	48.92%	90.46%	90.25%	93.61%
	R	89.18%	92.45%	87.55%	88.78%	83.67%
	F	89.92%	63.98%	88.98%	89.51%	88.36%

All the final detection points establish pairwise relationships with their five nearest neighbors, and the probability of a pair of points belonging to the same cell is computed for each pair using the learned pairwise classifier. The grouping algorithm is then applied to give the final result, as is shown in the left image of Fig. 13, where points within the same group are connected by lines. The green lines/dots indicate correct detections; the red lines/dots show incorrect detection points falling in the background; the blue lines/dots represent the detection points that are indeed cell points yet the cell has already been claimed by another group of points (i.e. over-segment of a cell); the magenta regions are the ground truth cell masks for the cells that are not detected. For easy observation, the original testing image is displayed on the right side of Fig. 13.

As we can see, many of the mistakes occur where the image region is highly ambiguous without resort to temporal information. Those mistakes could be corrected when

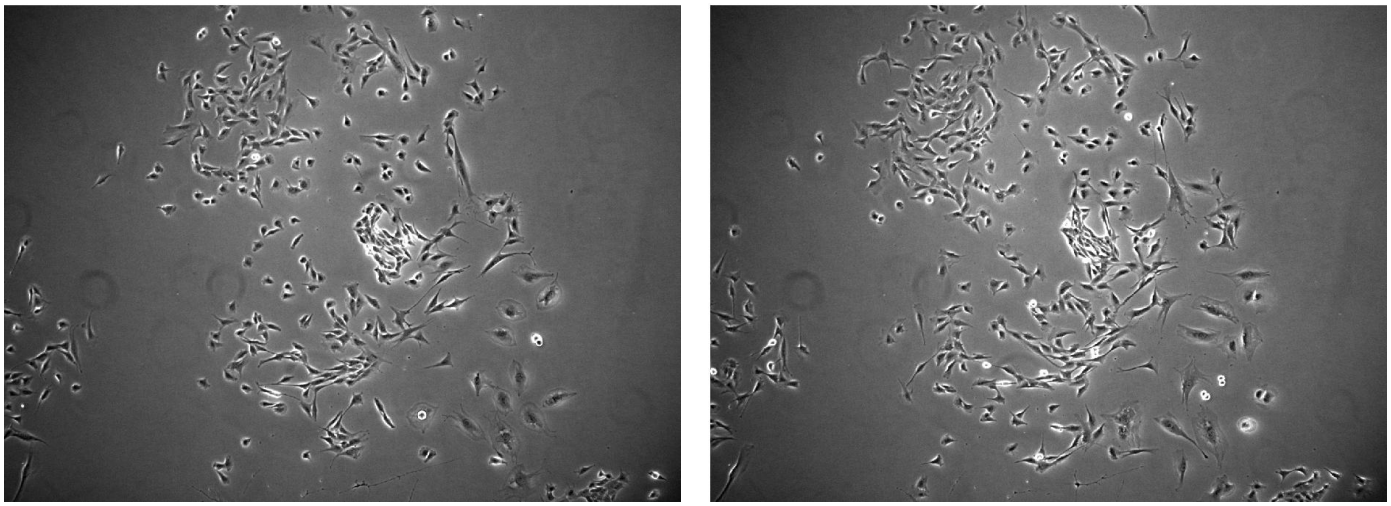


Fig. 14. Training and testing images for C2C12 muscle stem cells. Left: Training image. Right: one of the testing images.

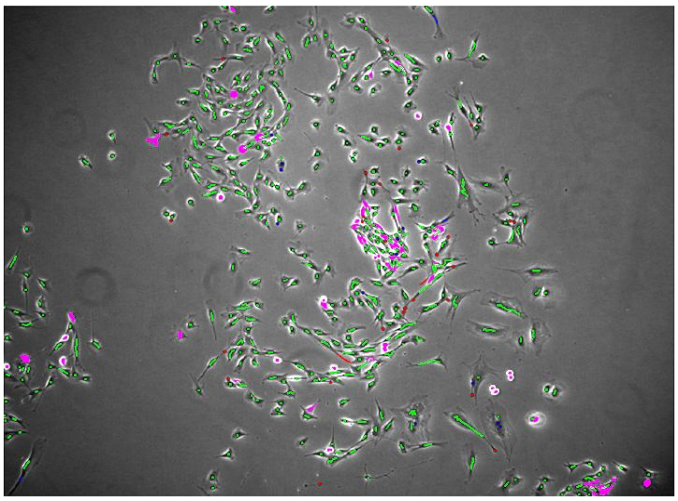


Fig. 15. The final detection result of C2C12 muscle stem cells for the example testing image.

temporal cues are incorporated in the future.

The proposed algorithm is further evaluated quantitatively. Table I lists the average accuracy of the *cell/background classifier* with and without the mean-shift refinement over all the testing images, as well as that of the *within-cell/between-cell classifier*. Both classifiers achieve an accuracy around 90%, and the mean-shift refinement does help improve classification accuracy in both cases.

The final detection result is shown in Table II, where the average precision (P), recall (R), and F-measure (F) over all the testing images are listed under various scenarios. We can see that the average F-measure of the proposed algorithm is above 91%, a promising result considering that only single-frame information is used. The precision is around 86%, due to both false detections in the background and multiple detections within the same cell. Here, precision is defined in a strict sense: even if the detection points are indeed within the cell region, they are still counted as false positives if the cell they are associated with has already been claimed by other detection points. If this condition is relaxed, the precision of the proposed algorithm reaches 91.57%.

The "No Group" column in Table II lists the result before detection points are grouped together. In this case, individual detection points are regarded as different cells. It is obvious that the precision would be low while the recall would be high. The goal of the grouping process is to raise the precision as much as possible while preserving the recall. Comparing the results in the "Proposed" and the "No Group" columns, we can see that the grouping algorithm successfully achieves this goal.

The "No MRGO" and the "No HBS" columns in Table II are the detection performance when the mean relative gradient orientation or the hypothetic boundary strength is removed from the feature set, respectively. The drop of performance after the removal of these features confirms their respective roles of complementing the HOG feature and the optimal connecting path feature, as is stated in Sections III and IV. Again, applying the algorithm without the mean-shift refinement results in much decline in performance.

The framework of statistical learning is intended to enable the proposed algorithm to adapt to different cell types without the need of any modification or tuning. To verify this claim,

we apply the proposed algorithm *as is* to the C2C12 muscle stem cells imaged by ZEISS Axiovert 135TV microscope at 05X magnification. Same as before, the first frame is for training and 10 other frames in the sequence are randomly chosen for testing. The training image and one of the testing images are displayed in Fig. 14.

For the example testing image, the learned pruning threshold preserves 100% of the candidate detection points while reducing the area of image under analysis by 71.38%, and the number of candidate detection points drops from 21,038 to 3,661. After the mean-shift refinement, that number further decreases to 3,447. After the cell/background classification, between-cell/within-cell classification, and point grouping, the final detection result is displayed in Fig. 15. Note that most of the missed cells are in the highly cluttered region where humans also have difficulty in labeling them. Many false detection points are located in the ends or corners of cells, which are also ambiguous to humans.

The quantitative results on C2C12 cells are also listed in Tables I and II. As we can see, although the cells and background in this experiment are very different from the previous one, the proposed algorithm again approaches an average F-measure of 90%, and the accuracy of both classifiers remain high. The other conclusions drawn in the experiment of bovine cells could still be arrived in this experiment, again confirming the effectiveness of the proposed algorithm.

VI. Conclusion

This paper proposes a single-frame cell detection algorithm in phase contrast microscopy images. It employs statistical learning techniques to capture cell appearances and inter-cell relationships, so that cells with a wide range of variations in shapes and appearances can be effectively discriminated from the background and correctly separated. The proposed algorithm achieves high detection accuracy on two different types of cells imaged using different microscopes with different settings, using single-frame clues alone and without manual tuning of parameters. This demonstrates the promise of this approach as a general framework for detecting various types of cells captured under different imaging conditions.

Future work includes extending the learning framework to temporal domain, and making the algorithm adaptive to different imaging modalities.

References

[1] M. M. Morelock et al., "Statistics of assay validation in high throughput cell imaging of nuclear factor B nuclear translocation," in *ASSAY Drug Dev Tech*., 2005, vol. 3, pp. 483–499.

[2] O. Debeir, P.V. Ham, R. Kiss, and C. Decaestecker, "Tracking of migrating cells under phase-contrast video microscopy with combined mean-shift processes," in *IEEE Transactions on Medical Imaging*, 2005, vol. 24, no. 6, pp. 697–711.

[3] S. Hadjidemetriou et al., "Detection and tracking of cell divisions in phase contrast video microscopy," in *Proc. of the Third MICCAI Worshop on Microscopic Image Analysis with Applications in Biology*, 2008.

[4] K. Li et al., "Cell population tracking and lineage construction with spatiotemporal context," in *Medical Image Analysis*, 2008.

[5] M. Usaj, D. Torkar, and D. Miklavcic, "Automatic cell detection in phase-contrast images for evaluation of electroporation efficiency in vitro," in Proc. *of the 11th Mediterranean Conference on Medical and Biomedical Engineering and Computing*, 2007, pp. 851–855.

[6] I. Ersoy, F. Bunyak, M.A. Mackey, and K. Palaniappan, "Cell segmentation using Hessian-based detection and contour evolution with directional derivatives," in Proc. of *International Conference on Image Processing*, 2008.

[7] P. Yan, X. Zhou, M. Shah, and S.T.C. Wong, "Automatic segmentation of high-throughput RNAi fluorescent celluar images," in *IEEE Transactions on Information Technology in Biomedicine*, 2008, vol. 12, no. 1, pp. 109–117.

[8] K. Li et al., "Online tracking of migrating and proliferating cells imaged with phase-contrast microscopy," in *Proc. of the 2006 Conference on Computer Vision and Pattern RecognitionWorkshop*, 2006.

[9] K. Li, and T. Kanade, "Cell population tracking and lineage construction using multiple-model dynamics filters and spatiotemporal optimization," in *Proc. of the 2nd International Workshop on Microscopic Image Analysis with Applications in Biology*, 2007.

[10] J. Shotton, M. Johnson, and R. Cipolla, "Semantic Texton Forests for Image Categorization and Segmentation," in *Proc. of the 2008 Conference on Computer Vision and Pattern Recognition*, 2008.

[11] K. Li, M. Chen, and T. Kanade, "Cell population tracking and lineage construction with spatiotemporal context," in *Proc. of the 10th International Conference on Medical Image Computing and Computer Assisted Intervention*, 2007.

[12] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proc. of the 2008 Conference on Computer Vision and Pattern Recognition*, 2008.

[13] K. Li, and T. Kanade, "Nonnegative mixed-norm preconditioning for microscopy image segmentation," in *Proc. of the 21st Biennial International Conference on Information Processing in Medical Imaging*, 2009.

[14] X. Ren, and J. Malik, "Learning a classification model for segmentation," in *Proc. of International Conference on Computer Vision*, 2003.

[15] N. Dalal, and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. of the 2005 Conference on Computer Vision and Pattern Recognition*, 2005, pp. I: 886–893.

[16] D.A. Forsyth, and J. Ponce, *Computer Vision: A Modern Approach*. Prentice Hall, 2004.

[17] D. Comaniciu, and P. Meer, "Mean shift: a robust approach toward feature space analysis," in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2002, vol. 24, no. 5, pp. 1–17.

[18] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[19] J. Platt, "Probabilities for SV machines," *Advances in Large Margin Classiers*, 2000, pp. 61–74.

[20] R. C. Gonzalez, and R. E. Woods. *Digital Image Processing, Second Edition*. Prentice Hall, 2002.

[21] M. Wang et al., "Novel cell segmentation and online SVM for cell cycle phase identification in automated microscopy," in *Bioinformatics*, 2008, vol. 24, no. 1, pp. 94–101.

[22] S. C. Chen, G. Gordon, and R. F. Murphy, "A novel approximate inference approach to automated classification of protein subcellular location patterns in multi-cell images," in *Proc. of the 2006 IEEE International Symposium on Biomedical Imaging*, 2006, pp. 558–561.