Morphological Multi-Cell Discrimination for Robust Cell Segmentation

HYUN-GYU LEE$^{1,2}$ AND SANG-CHUL LEE$^{1}$, (Member, IEEE)

$^1$Department of Computer Engineering, Inha University, Incheon 22212, South Korea
$^2$DoAI Inc., Yongin 17097, South Korea
Corresponding author: Sang-Chul Lee (sclee@inha.edu)

This work was supported by the Inha University Research Grant.

ABSTRACT Cell segmentation is challenging owing to the existence of various experimental configurations, cell shapes that cannot be mathematically defined, and ambiguous cell boundaries. We propose a cell segmentation method using cell region discriminator and multi-cell discriminator trained using heterogeneous machine-learning techniques such as logistic regression, expectation-maximization, and support vector machine (SVM). The cell-region discriminator identifies the regions where cells are found in images obtained from microscopes via a secondary logistic regressor, and its features use statistical information as well as the distribution of neighbor intensities. The SVM-based multi-cell discriminator determines whether multiple cells are present in the region detected by the cell-region discriminator and whether the region should be divided using the expectation-maximization algorithm. We suggest features for the boundary sectional area and the least square error for cell surface fitting to train the multi-cell discriminator. Using the features and the SVM, the multi-cell discriminator can be trained without overfitting, even for small training data. During this process, the proposed convex cell surface enhances the clustering performance. In experiments, our method based on two discriminators stably divided connected cells even when the contrast between a cell and the background area was small, and it outperformed state-of-the-art methods in terms of cell detection and segmentation accuracy.

INDEX TERMS Image segmentation, machine learning, regression analysis, support vector machine, expectation-maximization algorithm.

I. INTRODUCTION

Optical microscopy has been widely used for quantifying single-cell characteristics, such as cell size and intracellular density, by visualizing cellular compartments, such as the cytoplasm, nucleus, sub-cellular appearance of proteins, and DNA elements. The quantitative analysis of cellular compartments is used directly for disease diagnosis. For example, a pap smear test is a procedure typically used in cervical cancer screening for assistance in the grading of premalignant and malignant cancer. Cytologists can diagnose cervical cancer using information obtained from the shape and morphological structure of cancerous cells, e.g., the nuclei to cytoplasm ratio. These quantitative microscopic image analyses workflows typically comprise sample preparation, microscopic image acquisition, nuclear or cytoplasmic image segmentation, feature extraction, and cell population analysis.

Accurate cell quantification is highly dependent on the results of cell segmentation performed on images obtained from microscopes. Cell segmentation algorithms do not converge to a single solution with good performance; in addition, they have been developed in various ways depending on the characteristics of the target cell. Imaging hardware and analysis software platforms [1] have developed rapidly, but such cell segmentation studies have been relatively lagging.

Cell segmentation is challenging owing to the following three reasons. 1) Various experimental configurations, such as cell types and imaging protocols, produce images with different shapes and brightness characteristics. 2) Because cells generally have dynamically changing shapes over time, their shapes cannot be mathematically defined. 3) The boundaries of some cells that are in contact with each other during cleavage or migration may be unclear, and experts may have different opinions on whether one or more cells are connected.

To overcome the various cell brightness problems encountered owing to image conditions, representative image
binarization methods such as the Otsu method [2], [3] and watershed transformation [4] were improved for brightness-invariant localization or adaptive binarization. These methods are simple to use without any additional parameters, but it is difficult to obtain good performance with complicated backgrounds and split overlapping or adjacent objects. The active contour model (ACM) [5], [6] is a representative energy-minimization technique that generates adequate results on noisy images based on initial points defined by a user. Graph cut (GC) [7]–[11], another segmentation method based on energy minimization, finds a global optimal solution for a given initial value.

Machine-learning-based methods typically exhibit more than a certain level of segmentation performance in various datasets [12]–[16]. In particular, an unsupervised learning-based cell segmentation method using a blob detector produces boundaries around cells that are similar to those perceived by humans [17], [18].

Segmentation methods based on deep learning, which have recently achieved substantial success, have yielded remarkable results compared to conventional segmentation techniques with regard to both natural and medical images [19]–[22]. Deep learning exhibits significant abstraction ability for image features by learning vast amounts of data; however, it has a characteristic that the result is strongly dependent on whether the data used was observed previously. In the medical imaging field, the reliability of the annotated image is relatively lower than that of the natural image because the annotation information is generated subjectively using expert knowledge and experience. Using such annotated images, a deep-neural-network model with high performance cannot be learned.

Conventional cell segmentation studies usually consist of seed detection to find the approximate location of a cell and cell splitting to divide the region. In these studies, precise seed detection must be achieved for the accuracy of cell division. Furthermore, because these techniques require numerous parameters as well as appropriate parameter selection depending on the types of target cells and the imaging conditions, the segmentation results are sensitive to parameter configuration.

In this paper, to overcome the above-mentioned limitations of deep learning, we propose a cell-segmentation method using a cell region discriminator R that detects a cell region using a logistic regression algorithm and a multi-cell discriminator M that determines whether a cell region is divided via an expectation-maximization (EM) algorithm (see Fig. 1). R identifies regions of interest for cells using linear regression analysis as well as features of statistical cell imaging and distribution characteristics of image brightness. The region of interest (ROI) is divided into two cells using expectation maximization for the features of the pixel coordinates of the detected ROI and the coordinates of the local maximum of each ROI location. In the process of dividing the region, M determines whether to re-segment the ROI. This is done by finding a hyperplane for the surface error in the cell area and the area of the segment boundary. This study makes the following three contributions:

- The proposed method does not require seed detection because a region made up of connected cells is recursively divided using the characteristics of the boundaries between the connected cells and surface fitting error.
- Using various learning models trained with the data of each image, the ROI detection and cell division performed in this study yield a high accuracy without parameter configuration.
- Without learning a large amount of data, the proposed model exhibits a stable segmentation performance using the morphological features of cells.

II. RELATED WORKS

The basic method of identifying cells automatically is to use intensity-based segmentation techniques, including Otsu-based methods [2], [3] and watershed transformations [4]. Such intensity-based techniques are easy to use; however, it is difficult to achieve a good performance for complex backgrounds or for splitting two or more adjacent, overlapping cells using these methods. Recently, a method of using the distance map based on the Otsu method as the initial value for watersheds has been proposed to improve segmentation accuracy [23].

Energy-minimization-based image segmentation techniques can produce better results than intensity-based approaches in the above-mentioned difficult scenarios. The ACM is a representative energy-minimization technique that employs area-based partitioning [5] and edge-based partitioning [6]; in addition, it exhibits relatively better performance with noisy images. However, ACM-based techniques require initial points; therefore, it is impossible to perform fully automatic segmentation, and the initial points affect the
accuracy of ACM-based techniques. Moreover, because of the high computational complexity arising from the iterative convergence process, ACM-based techniques are not suitable for finding multiple cells simultaneously.

Gradient flow tracking (GFT) [24], [25], GC [7]–[9], and level set [26]–[28] are energy minimization-based techniques that are more suitable than ACM-based techniques for multiple cell segmentation. GFT identifies cells by clustering the slope vector using the characteristic that the slope vector becomes very congested around a cell, but in a noisy environment, the slope at the cell boundary is quite small, and the direction is unreliable. GC-based algorithms [7]–[11] are widely used because they are guaranteed to find a global optimal solution for pixel boundaries between distinct regions. However, these methods may produce boundaries of an uneven step-like shape that are different from the boundaries perceived by a human; hence, a more advanced segmentation method is required.

Feature definition and classifier design have been suggested in various machine learning-based cell segmentation techniques [12]–[16]. These learning-based image segmentation techniques generally have good segmentation ability; however, they cannot perform segmentation at a very precise level for a specific problem.

Dividing the connected-cell ROI by assuming the shape of the ROI to be a 2D Gaussian mixture model (GMM) and using the EM method produces cell boundaries that are close to the ones perceived by humans [17], [18]. However, because there are numerous parameters, it is necessary to set the parameters depending on the image data set. Furthermore, the segmentation quality is poor for cell regions with small differences in contrast between the foreground and background.

Recently, most deep learning-based studies have demonstrated substantially high accuracies with regard to computer vision problems [19]–[22]. A generative adversarial network (GAN) is an unsupervised deep learning approach that has recently received considerable attention in the field of artificial intelligence. In the medical imaging field, GAN-based super-resolution techniques can produce MRIs and CT images at higher resolutions than the original resolution [29], [30]. Using convolutional networks, for example, the U-net for biomedical image segmentation, the accuracy was dramatically improved; subsequent applications of this approach have also been studied [19], [31]. The deep adversarial nuclei segmentation approach, which uses a conditional GAN (cGAN) [32], was proposed to classify overlapping and clumped nuclei with a high accuracy [33]. HoVer-Net was proposed to simultaneously conduct nuclei segmentation and classification under conditions such as large intra-class variability and the occurrence of clustered nuclei [34].

However, deep learning-based techniques always require a tremendous amount of data and accurate annotation information to train the network models. With regard to datasets that do not satisfy these two conditions, deep learning-based techniques cannot achieve high accuracy for each task because the models are not well-trained. Therefore, a novel approach that can be applied to small and inaccurate training datasets and is not sensitive to hyperparameters is necessary.

III. CELL REGION DISCRIMINATOR

Recent methods that determine the approximate coordinates of cell regions using a blob detector, such as Laplacian of Gaussian (LoG) or maximally stable extremal region (MSER), as well as techniques that employ local binarization and global binarization, exhibit good performances with regard to cell region detection [17], [18]. However, when the contrast between the background and a cell region is low, such methods cannot satisfactorily detect the cell region using global and local binarization. Therefore, a special image-normalization method capable of increasing the contrast difference with respect to the image characteristics is required. In this paper, we propose a pixel-level cell region discriminator R using statistical as well as histogram features and logistic regression to solve the degradation of cell region detector performance owing to low image contrast. The statistical feature to be used in the logistic regression analysis is defined as a five-dimensional feature vector composed of the median, mean, standard deviation, maximum value, and the difference between the maximum and the minimum values. The feature is extracted for the pixels of the local square image with one side comprising (s−∑min/φ − 1)/9 pixels, where ∑min is the minimum cell area. In addition, the distribution feature that directly expresses the distribution characteristics of pixel brightness is defined as a brightness histogram that divides the range between the minimum and the maximum values in the image into n classes. To concentrate on the low-intensity values where the brightness values of the cell area and the background area are densely distributed, the bin corresponding to the lowest 25% of all the bins are used. The two kinds of features are used to categorize parts of the images as the either the cell region or the background, and the feature that is used to identify the cell region is expressed via the following logistic regression:

\[
\log \frac{p(y = 1 | x; \theta)}{p(y = 0 | x; \theta)} = \beta^T x, \tag{1}
\]

where \( \sigma \) is the sigmoid function, \( \theta \) is the learned regression parameter vector, and \( x \) is the statistical feature or histogram feature vector, \( y \) is the label of \( x \) and is equal to 1 if it is a cell area or 0 if it is the background.

To train the pixel-level discriminator, features are extracted for every pixel in the image. Therefore, a considerable amount of time is required to train the logistic regression parameter \( \theta \) using common optimization techniques such as gradient descent. For fast learning convergence, we optimized the parameter vector \( \theta \) using stochastic gradient descent, defined as follows:

\[
\theta = \theta + \eta g, \\
g = \frac{1}{m'} \nabla_{\theta} \sum_{i=1}^{m'} L(x^{(i)}, y^{(i)}, \theta), \\
L(x^{(i)}, y^{(i)}, \theta) = -\log p(y^{(i)|x^{(i)}}; \theta), \tag{2}
\]
where $\eta$ is the learning rate; $m'$ is the mini-batch sample number, and $L$ is the per-example loss function often used as a cost function for machine learning algorithms.

When identifying a cell region through logistic regression analysis, either statistical or distributional features have better discrimination power, depending on the image. Therefore, we define the pixel probability values of pixels for two features estimated through logistic regression as a two-dimensional ensemble feature and detect the cell region stably using second-order regression. The classification threshold of each regressor is experimentally set to the value that best classifies the training data.

IV. MULTI-CELL DISCRIMINATOR

A. CONVEX SURFACE TRANSFORM

A cell region segment, $S$, detected in a pixel-level cell region discriminator, $R$, contains from one to dozens of cells. Therefore, an $S$ consisting of numerous cells must be divided into individual cells for precise cell segmentation. For this purpose, we adopted an existing study that assumes each cell as a 2D GMM and clusters the pixel coordinates to each cell using the EM algorithm, which determines the parameters of the probability model. The features to be used in EM are the pixel coordinates and the corresponding coordinates of the local maximum point, where the initial cluster is set through k-means clustering. The local maximum coordinate is an important factor influencing the clustering performance, and the closer the maximum point coordinates in a cell, the more accurate the division. To reduce the variance of the local maximum coordinates within a cell, the cell surface is converted to a more convex shape by combining the original cell image $I_S$ with the distance image $D$, as shown in Fig. 2(a) - (c). In the distance image, the high responses mean cell regions far from cell boundaries. Although the images are 2D information, the intensities are related to the height of the cell, assuming a cell shape is close to an ellipsoid. Therefore, we consider the intensity values as the 3D depth of the cell and conduct surface fitting to the cell height. The convex cell image $C$ is defined as follows:

$$C = \frac{1}{2} \left( \frac{I_S}{\max(I_S)} + \frac{D}{\max(D)} \right) \ast K,$$

$$D(x) = \|x - \arg \min_{x_B} (x - x_B)\|, \quad x_B \in B,$$

where $I_S$ is a local image containing only one region segment $S$, and $K$ is a Gaussian blur kernel. $B$ is the set of edge pixel coordinates $x_B$ of $S$, and $x$ is all pixel coordinates of $S$.

We estimate the local maximum coordinates of each pixel coordinate using gradient ascent for pixel brightness in the convex image $C$ of the cell surface. As shown in Fig. 2(d) and (e), because the maximum local point is detected closer to the cell center than the original cell surface in the convex image generated by Eq. (3), more accurate cell division is possible.

B. CLASSIFIER LEARNING TO DISTINGUISH BETWEEN SINGLE AND MULTIPLE CELLS

With the local maximum coordinates, an adequate number of components $k$ should be determined to generate the GMM. If the number of components $k$ in the GMM is equal to the number of actual cells in region segment $S$, we can expect that a meaningful division result will be generated through the EM algorithm [17], [18]. Depending on various conditions, such as the brightness of $S$, the smoothness of its surface, and the boundary brightness between the adjacent cells, the partitioning methods [17], [18] of selecting $k$ to minimize the dissimilarity between the real cells and the virtual cells generated from the GMM parameter may not work well.

To overcome these types of possible failure cases, we propose a multi-cell discriminator $M$ that divides a region $S$ into a binary tree structure by determining whether or not it is a multi-cell region. The feature of $S$ for multi-cell identification is defined as the least square error $\epsilon = \{\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4\}$ for surface fitting using the third-order polynomial function $f$ for cell surface, as follows: $\epsilon_1 = \frac{1}{m} \|I_S - f\|_1$, $\epsilon_2 = \|I_S - f\|_2$, $\epsilon_3 = \frac{1}{m} \|I_S - f\|_3$, and $\epsilon_4 = \frac{1}{m} \|I_S - f\|_4$.

Where $m$ is the number of pixels belonging to $S$. As shown in the first row of Fig. 3, the least square error of a single cell using a third-order polynomial surface fitting of region segment $S$ is smaller than that of multiple cells.

The difference between an $S$ with a single cell and one with multiple cells also appears in the boundary sectional area between the cells divided by EM. If EM divides $S$ into multiple cells, then the boundaries between the models will be formed at the boundary between two cells, as shown in Fig. 3(e) and (f). If $S$ of the single cell is divided into two cells, the boundaries between the models will be formed in...
the vicinity of the cell center because there are no obvious Gaussian mixture distributions in S, as shown in Fig. 3(d).

Therefore, the boundary sectional area between the cells divided by the EM for a multi-cell S is smaller than that estimated for a single-cell S; hence, the boundary sectional area is suitable for discriminating between a single cell and multiple cells. Thus, we define the boundary sectional area feature \( A = \{ A_1, A_2, A_3, A_4, A_5 \} \) of the cells divided by EM as follows:

\[
\begin{align*}
A_1 &= \frac{1}{2m} \sum_{i=1}^{n} I_S(x_B^{(i)}), \\
A_2 &= \frac{1}{2nm} \sum_{i=1}^{n} I_S(x_B^{(i)}), \\
A_3 &= \frac{1}{nm_{\text{min}}} \sum_{i=1}^{n} I_S(x_B^{(i)}), \\
A_4 &= \frac{1}{n} \sum_{i=1}^{n} I_S(x_B^{(i)}), \\
A_5 &= \frac{1}{n} \sum_{i=1}^{n} I_S(x_B^{(i)}),
\end{align*}
\]

where \( n \) is the number of boundary pixels of the divided cells; \( I_S(x_B^{(i)}) \) is the brightness of the boundary pixel coordinate \( x_B^{(i)} \), and \( m_{\text{min}} \) is the number of pixels of the smaller of the two cells split by EM. \( x_B^{(i)}_{\text{min}} \) and \( x_B^{(i)}_{\text{max}} \) are the corresponding pixel coordinates belonging to cells of smaller and larger areas among the divided cells. \( A_1 \) is an average boundary sectional brightness, and other features are its normalized values. \( A_2 \) is normalized by the area of the cell to be divided, and \( A_3 \) is normalized by the area of the smaller of the two cells. \( A_4 \) and \( A_5 \) are normalized by the summation of the brightness of the smaller cell and the larger cell, respectively.

The multi-cell discriminator M is based on a support vector machine (SVM) that learns the surface fitting error and the boundary area characteristics of divided cells for a single-cell and double-cell S and determines whether any S contains multiple cells that can be divided. First, S is divided into two cells via EM in the learning stage of M. If S is composed of a single cell in the ground truth, the feature vector extracted by the least square error for cell surface fitting and the boundary sectional area in the divided region S is assigned the single-cell class. If S is composed of multiple cells, its feature vector is assigned the multi-cell class. Given a multi-cell label, with progressive partitioning on two divided regions \( S'_1 \) and \( S'_2 \) in S and their paired children nodes \( N_{\text{left}} \) and \( N_{\text{right}} \), a binary tree structure is built, as shown in Algorithm 1.

In the learning phase of M, an SVM is trained from extracted features \( \epsilon, A \) for all nodes of the tree. Therefore, using the surface shape obtained by the least square fitting \( \epsilon \) and boundary height \( A \) obtained by the EM algorithm, SVM classifies whether an area is composed of single cells or connected or overlapping cells. When testing the cell segmentation, if S is divided into two regions by EM and the S is determined as a multi-cell region using the learned M, the S is divided into two sub-cells. As in the binary tree structure shown in Fig. 1, the sub-cells will be reclassified and divided by the SVM until all leaf nodes are classified as a single cell.

Algorithm 1 Build Binary Single-Cell Tree

**Variable description:**

\( N \): Node of binary single cell tree

\( S \): cell region

**procedure** TRAIN_MTree(\( N, S \))

\( N \leftarrow \text{CreateChild}(N) \)

\( \epsilon \leftarrow \text{LeastSquareError}(S) \)

\( [S'_1, S'_2] \leftarrow \text{EM}(S) \)

\( A \leftarrow \text{BoundaryBrightness}(S'_1, S'_2) \)

\( y \leftarrow M(\epsilon, A) \)

if \( y = 1 \), then

\( [\epsilon, A] \in F_s \)

else

\( [\epsilon, A] \in F_m \)

\( N_{\text{left}} \leftarrow \text{TRAIN_MTree}(N_{\text{left}}, S'_1) \)

\( N_{\text{right}} \leftarrow \text{TRAIN_MTree}(N_{\text{right}}, S'_2) \)

end if

return \( N \)

end procedure

V. EXPERIMENTAL RESULTS

A. EXPERIMENTAL SETUP

To assess segmentation performance against a variety of imaging conditions, we tested our algorithm on the BBBC006 and BBBC020 datasets from the Broad Bioimage Benchmark Collection [35] and ISBI 2009 U2OS cell [36] dataset. BBBC006 is relatively larger than BBBC020 and ISBI, but some of the ground truth in BBBC006 is not correct. Although the size of the ISBI dataset is small, its ground truth is more reliable than those of others. BBBC020 is the smallest dataset, and some of the ground truth is lacking in this dataset. Because the BBBC020 dataset showed the lowest signal-to-noise ratio, this dataset is noisier than the others. Our experimental evaluation in Table 4 shows that the proposed method robustly segments the cells in noisy images.
TABLE 1. Data characteristics analysis by mean μ, standard deviation σ, and signal-to-noise ratio (SNR).

|                  | BBBC006 | ISBI   | BBBC020 |
|------------------|---------|--------|---------|
| μ(T)             | 0.10    | 0.05   | 0.06    |
| μ(s(I))          | 0.09    | 0.08   | 0.14    |
| μ(SNR(I))        | 21.18   | 21.32  | 17.13   |

The statistical and distribution features used in cell region discriminator R were extracted from the original image filtered by a Gaussian kernel with a standard deviation of 0.8. Histogram class \( n \) was experimentally set to 4, and the distribution feature only used values of fewer than 10 bins, corresponding to 25% of all bins. In the learning phase of the logistic regressor in R, mini-batch size \( m' \) and learning rate \( \eta \) were set to 10,000 and 3, respectively. Learning was terminated after the training tolerance of the learned parameter \( w \) was 0.001 or the number of learning iterations was more than 10,000.

In the multi-cell discriminator M, the surface-fitting error and boundary area features were extracted from the original image filtered by a Gaussian kernel with a standard deviation of 2.4. The Gaussian kernel size was determined by the equation \( 2[2\sigma] + 1 \). Training and testing of R and M were performed with five-fold cross-validation. For the final result, we used 11 ensemble classifiers using M taught with different feature combinations. For the SVM classification, we used the \texttt{fitcsvm} function in Matlab. In this experiment, the radial basis function was used, and the kernel scale was set to autoscale.

As shown in Fig. 4, the segmentation accuracy of the proposed discriminator trained using only one image was not significantly different from those of the discriminator trained using 600 images. Because the proposed features are well designed, the proposed method using the logistic regression and SVM with good generalization capability does not require extensive data.

B. EVALUATION METRICS AND TESTED ALGORITHM

Seven evaluation metrics for measuring cell region detection and cell segmentation performance were used to compare the results of our proposed method with those of other methods. In terms of accuracy at the pixel level, we used the Dice similarity coefficient (DSC) [38], sensitivity (SEN) [39], Jaccard index (JAC) [40], and accuracy (ACC), which was evaluated by the formula \( \text{ACC} = (TP + TN)/(TP + TN + FP + FN) \).

To accurately estimate the cell-region partitioning performance, only the cell region and not the background was evaluated as an object. The error aggregation method was evaluated by counting the number of over-added, missed, over-merged, and over-split cells. This method was introduced by Coelho et al. [36] and is extensively used for the evaluation of nucleus segmentation in fluorescence images [41], [42].

The performance of the proposed algorithm was compared with those of seven segmentation algorithms: GC [7], MINS [14], Otsu-based segmentation [3], ilastik [15], NOER [13], LoG-EM [17], MSER-EM [18], U-Net [19], and Deepcell [37]. The proposed method and supervised-learning-based methods, such as ilastik and NOER, were trained and tested for each dataset. The parameters of the comparison algorithms were set to values given in previous studies [7], [13], [17], [18]. The histograms, mean, and variance of neighboring pixels provided in the feature list were used as feature vectors in ilastik. To generate as well-trained a model as possible, NOER and U-Net were tested using ten-fold cross-validation that included more training data than that used in the proposed method. During the training of the U-net, we augmented images with vertical and horizontal flips, width and height shifts, rotation, and shear. The input image size was 256 x 256, the batch size, epoch, and steps per epoch were set to 16, 10, and 43, respectively. To compare our method with a more accurate deep learning method, Deepcell [37], [43] was tested with five-fold cross-validation. During the training of Deepcell, images were augmented with operations such as vertical and horizontal flips, rotation, and scaling of 0.8 to 1.2. The input image size was 348 x 260, the batch size, epoch size, and steps per epoch were set to 1, 3, and 603, respectively. Because Deepcell was developed to solve segmentation problems similar to our study, we used parameters that were the same as the ones in the author’s codes.

C. CELL REGION SEGMENTATION AND CELL SEGMENTATION IN THE BBBC006 DATASET

The BBBC006 images were acquired from one 384-well microplate containing U2OS cells stained with Hoechst 33342 markers (to label nuclei). The ground truths of the images were manually examined by an expert to classify which focal planes corresponded to in-and out-focus images. The dataset has 768 microscope images with ground-truth images passively produced by experts. Each image has a pixel size of 696 x 520 and is in a 16-bit grayscale format.
| Method      | DSC   | SEN   | JAC   | ACC   | Added | Missed | Merged | Split | mP   | mC   |
|-------------|--------|-------|-------|-------|-------|--------|--------|-------|------|------|
| ilastik [15]| 0.64(0.74)| 0.60(0.82)| 0.49(0.62)| 0.68(0.91)| 5.38(5.07) | 11.35(8.33) | 19.12 | 0.04 | 0.60(0.73) | 0.97(0.95) |
| Otus [5]    | 0.70(0.89) | 0.59(0.82) | 0.54(0.81) | 0.71(0.96) | 9.13(9.09) | 0.32(0.24) | 12.81 | 0.18 | 0.63(0.86) | 5.61(4.67) |
| NOER [13]   | 0.79(0.82) | 0.67(0.71) | 0.66(0.70) | 0.77(0.93) | 3.3(3.00) | 2.23(0.85) | 1.94 | 1.11 | 0.72(0.78) | 2.08(1.92) |
| U-Net [19]  | 0.80(0.88) | 0.68(0.80) | 0.67(0.79) | 0.77(0.95) | 2.67(2.43) | 0.30(0.21) | 3.01 | 1.46 | 0.73(0.85) | 1.86(1.32) |
| GC [7]      | 0.80(0.92) | 0.69(0.87) | 0.68(0.86) | 0.78(0.96) | 5.15(4.21) | 0.18(0.14) | 2.09 | 12.25 | 0.74(0.90) | 4.92(2.17) |
| Deepcell [37]| 0.81(0.84) | 0.69(0.73) | 0.68(0.73) | 0.78(0.94) | 1.16(1.16) | 0.91(0.52) | 1.55 | 2.82 | 0.74(0.80) | 1.61(0.64) |
| MINS [14]   | 0.84(0.90) | 0.75(0.84) | 0.73(0.82) | 0.82(0.96) | 5.63(5.32) | 0.89(0.17) | 1.90 | 4.23 | 0.78(0.87) | 3.16(2.74) |
| MSER-EM [18]| 0.89(0.96) | 0.83(0.90) | 0.81(0.93) | 0.87(0.98) | 2.71(2.33) | 0.81(0.19) | 1.65 | 8.84 | 0.86(0.96) | 3.50(1.26) |
| LoG-EM [17] | 0.95(0.96) | 0.92(0.96) | 0.87(0.93) | 0.91(0.98) | 1.18(0.83) | 0.61(0.19) | 2.63 | 2.92 | 0.91(0.96) | 1.83(0.51) |
| Proposed    | 0.93(0.97) | 0.92(0.98) | 0.87(0.94) | 0.91(0.99) | 0.83(0.67) | 0.01(0.01) | 2.51 | 3.19 | 0.91(0.97) | 1.64(0.34) |

**TABLE 3.** Accuracy of cell segmentation for the ISBI dataset.

![FIGURE 5. Segmentation results for the BBBC006 dataset (a) original image, (b) ground truth, (c) ilastik, (d) OTSU, (e) NOER, (f) U-Net, (g) GC, (h) Deepcell, (i) MINS, (j) MSER-EM, (k) LoG-EM, and (l) the proposed method.](image-url)

10 images without ground-truth information were excluded from the experiments. In the BBBC006, the cell surfaces are relatively flat but have different brightness values, and the boundaries of the cells are quite uncertain.

As shown in Table 2, the proposed method showed the highest performance with regard to all measures at the pixel-level, including sensitivity, Jaccard index, Dice similarity coefficient, and accuracy for cell segmentation and cell region segmentation. The over-missed and the over-added regions in detecting the background region had the lowest values compared to other algorithms. It means that cell regions were effectively detected even in the case of a cell image with large brightness deviations. The sum of the over-merged and the over-split cells was lower for some of the models used for comparison with the proposed method; however, the two metrics were evaluated in the detected cell region only. Therefore, the four error-aggregation metrics (Added, Missed, Merged, and Split) at the cell-level should be analyzed comprehensively. In terms of the mean of cell-level performance mC, the proposed method showed the best performance.

**D. CELL REGION DETECTION AND SEGMENTATION IN THE ISBI DATASET**

The ISBI dataset was used to validate the proposed learning-based method in other challenging conditions. The dataset contained original images from two microscopes for two cell types stained with Hoechst 33342. It also contained images showing the hand-segmentation of the Hoechst images into regions containing single nuclei. Because images in the ISBI dataset have an uneven cell surface and much lower contrast between cell and background regions compared to the BBBC dataset, segmentation is more challenging than the BBBC dataset. The ISBI dataset consists...
H.-G. Lee, S.-C. Lee: Morphological Multi-Cell Discrimination for Robust Cell Segmentation

FIGURE 6. Segmentation results for the ISBI dataset (a) original image, (b) ground truth, (c) OTSU, (d) Deepcell, (e) MSER-EM, (f) ilastik, (g) GC, (h) NOER, (i) U-Net, (j) MINS, (k) LoG-EM, and (l) the proposed method.

TABLE 4. Accuracy of cell segmentation for the BBBC020 dataset.

| Method          | DSC | SEN | JAC | ACC |
|-----------------|-----|-----|-----|-----|
| Proposed        | 0.74(0.80) | 0.79(0.91) | 0.59(0.67) | 0.79(0.97) |
| Deepcell [37]   | 0.75(0.75) | 0.72(0.77) | 0.70(0.69) | 0.74(0.79) |
| U-Net [19]      | 0.73(0.74) | 0.70(0.72) | 0.71(0.69) | 0.74(0.78) |
| Otlu [3]        | 0.74(0.82) | 0.76(0.92) | 0.75(0.69) | 0.75(0.79) |
| GC [7]          | 0.68(0.64) | 0.75(1.00) | 0.61(0.48) | 0.68(0.92) |
| ilastik [15]    | 0.68(0.79) | 0.77(0.96) | 0.52(0.65) | 0.78(0.96) |
| MSER-EM [18]    | 0.68(0.74) | 0.77(0.92) | 0.53(0.60) | 0.77(0.96) |
| MINS [14]       | 0.71(0.79) | 0.78(0.97) | 0.56(0.65) | 0.79(0.96) |
| LoG-EM [17]     | 0.74(0.78) | 0.84(0.96) | 0.59(0.63) | 0.78(0.96) |
| NOER [13]       | 0.78(0.83) | 0.78(0.87) | 0.65(0.71) | 0.81(0.97) |

In terms of cell segmentation at the pixel-level, the proposed method showed the highest accuracy even without the brightness normalization technique used in LoG-EM, which considers characteristics of the image to enhance image contrast (see Table 3). In addition, mC was the second-lowest, and the cell regions were recognized stably even with very low contrast.

As shown in Fig. 6(e) and the values in parentheses in Table 3, MSER-EM could not recognize cell regions owing to the low contrast. Deepcell performed better than U-Net under the condition of training with numerous data, but under the opposite condition, it yielded poor results owing to its complexity.

E. CELL REGION DETECTION AND SEGMENTATION IN THE BBBC020 DATASET
The BBBC020 images depict bone-marrow-derived macrophages from C57BL/6 mice. The image set consists of 25 images, each consisting of three channels: DAPI,
CD11b/APC, and their merger. Because we can see the nuclei in samples stained with DAPI, only the DAPI channel was used in this experiment. The brightness of the BBBC020 images is the most saturated of the three datasets, so their connected boundaries are ambiguous as well.

As shown in Table 4, the performance of our method was not the highest, but it achieved the second highest and third highest ranks in terms of performance. With regard to the saturated and noisy images, the proposed method worked robustly. NOER scored the highest in terms of ROI detection, but its ability to divide multiple nuclei was poor, resulting in unstable overall segmentation performance. As shown in Fig. 7(a), although some annotations were missed in the ground truth, the proposed method clearly divides the boundaries between the connected cells when compared with U-Net. Because this dataset size is the smallest among the three, Deepcell with complex structure showed the poorest results.

F. DISCUSSION
The proposed method achieved the best performance in terms of average mP and average mC of the three datasets, as shown in Fig. 8. Even for a dataset consisting of noisy or high-contrast images, our method ranked within the top 3 in terms of accuracy. In addition to the average scores, to verify the statistical significance between the results of the proposed and the nine comparison methods, the paired samples t-test was employed, as shown in Table 5. The results of the proposed method were significantly improved compared to those of the other methods, with p-values being less than 0.05. LoG-EM achieved high performance, but the proposed method performed better than LoG-EM in terms of mC in BBBC006 and mP in ISBI. In all statistical analyses, the accuracy of our method was similar to or higher than those of the other methods.

We summarized our experiments in Table 6. The proposed method showed the highest ranking in the three datasets. The morphology of detected cells was smooth, and boundaries between connected cells were well split. The number of hyperparameters required by the proposed method was small, and most of the hyperparameters were not individually tuned for each dataset. Even though the training data size was small and the data was not accurate, our method worked well owing to our novel handcrafted features, such as the boundary sectional area and the least square error for cell surface fitting. Deep learning-based methods could not show outstanding performance owing to the insufficient amount of training data and inaccuracy of labels paired with images. For such data, well-designed machine-learning approaches are more appropriate than the deep learning method.

VI. CONCLUSION
We proposed a supervised learning-based cell segmentation method that does not require parameter tuning or a tremendous amount of data. We used logistic regression to identify cell regions and SVM to determine whether a region split is...
necessary. The proposed method showed better performance than the current state-of-the-art methods. The cell region discriminator detected cell regions correctly even in low-contrast images, and the multi-cell region discriminator determined whether a cell region included a large number of cells and divided the cell region into a binary tree structure without seed information. In addition, in our proposed approach, a separate seed-detection process is not necessary. For a dataset that is small and has inaccurate annotation information, our well-designed features and machine-learning methods are more useful than the deep learning-based methods. Although the proposed method is suitable for simple backgrounds with low contrast and dark backgrounds, our future work will seek to improve it so that it can be applied to more complex histopathological images.

REFERENCES

[1] K. W. Eliceiri, M. R. Berthold, I. G. Goldberg, L. Ibáñez, B. S. Manjunath, M. E. Martone, R. F. Murphy, H. Peng, A. L. Plant, B. Roysam, N. Stuurman, J. R. Swedlow, P. Tomancak, and A. E. Carpenter, “Biological imaging software tools,” Nature Methods, vol. 9, pp. 697–710, Jul. 2012.

[2] M. Braik, A. Benzinou, K. Nasreddine, S. Labidi, and N. Hymery, “Segmentation of dendritic cells from microscopic images using mathematical morphology,” in Proc. 2nd Int. Conf. Adv. Technol. Signal Image Process. (ATSIP), Mar. 2016, pp. 282–287.

[3] N. Otsu, “A threshold selection method from gray-level histograms,” IEEE Trans. Syst., Man., Cybern., vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.

[4] K. Jiang, Q.-M. Liao, and Y. Xiong, “A novel white blood cell segmentation scheme based on feature space clustering,” Soft Comput., vol. 10, no. 1, pp. 12–19, Jan. 2006.

[5] T. F. Chan and L. A. Vese, “Active contours without edges,” IEEE Trans. Image Process., vol. 10, no. 2, pp. 266–277, Feb. 2001.

[6] V. Caselles, R. Kimmel, and G. Sapiro, “Geodesic active contours,” Int. J. Comput. Vis., vol. 22, no. 1, pp. 61–79, Feb. 1997.

[7] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, “Improved automatic detection and segmentation of cell nuclei in histopathology images,” IEEE Trans. Biomed. Eng., vol. 57, no. 4, pp. 841–852, Apr. 2010.

[8] K. Oyebode and J. Tapamo, “Adaptive parameter selection for graph cut-based segmentation on cell images,” Image Anal. Stereolog., vol. 35, no. 1, pp. 29–37, 2016.

[9] K. Oyebode and J.-R. Tapamo, “Selective cell segmentation using semi-automatic graph cut with adaptive distance penalties,” Biomed. Res., vol. 27, no. 4, pp. 1094–1098, Jan. 2016.

[10] V. Kolmogorov and R. Zabih, “What energy functions can be minimized via graph cuts?” IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 2, pp. 147–159, Feb. 2004.

[11] N. O. Loewe, S. Pai, C. Cordeiro, D. Black, B. L. King, C. H. Contag, B. Chen, T. M. Baer, and O. Solgaard, “Automated cell segmentation for quantitative phase microscopy,” IEEE Trans. Med. Imag., vol. 37, no. 4, pp. 929–940, Apr. 2018.

[12] A. Arbelle, N. Drayman, M. Bray, U. Alon, A. Carpenter, and T. R. Raviv, “Analysis of high-throughput microscopic videos: Catching up with cell dynamics,” in Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent., Cham, Switzerland: Springer, 2015, pp. 218–225.

[13] A. Artetta, V. Lemptsys, J. A. Noble, and A. Zisserman, “Learning to detect cells using non-overlapping extremal regions,” in Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. Berlin, Germany: Springer, 2012, pp. 348–356.

[14] X. Lou, M. Kang, P. Xenopoulos, S. Muñoz-Descalzo, and A.-K. Hadjantonakis, “A rapid and efficient 2D/3D nuclear segmentation method for analysis of early mouse embryo and stem cell image data,” Stem Cell Rep., vol. 3, no. 3, pp. 392–397, Mar. 2014.

[15] C. Sommer, C. N. Strählle, U. Köthe, and F. A. Hamprecht, “Ilastik: Interactive learning and segmentation toolkit,” in Proc. ISBI, Mar./Apr. 2011, vol. 2, no. 5, p. 8.
[35] V. Ljosa, K. L. Sokolnicki, and A. E. Carpenter, “Annotated high-throughput microscopy image sets for validation,” *Nature Methods*, vol. 9, no. 7, p. 637, Jul. 2012.

[36] L. P. Coelho, A. Shariff, and R. F. Murphy, “Nuclear segmentation in microscope cell images: A hand-segmented dataset and comparison of algorithms,” in *Proc. IEEE Int. Symp. Biomed. Imag., From Nano Macro*, Jun. 2009, pp. 518–521.

[37] D. A. Van Valen, T. Kudo, K. M. Lane, D. N. Macklin, N. T. Quach, M. M. DeFelice, I. Maayan, Y. Tanouchi, E. A. Ashley, and M. W. Covert, “Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments,” *PLOS Comput. Biol.*, vol. 12, no. 11, 2016, Art. no. e1005177.

[38] L. R. Dice, “Measures of the amount of ecologic association between species,” *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945, doi: 10.2307/1932409.

[39] E. Fernandez-Moral, R. Martins, D. Wolf, and P. Rives, “A new metric for evaluating semantic segmentation: Leveraging global and contour accuracy,” in *Proc. IEEE 4th Intell. Vehicles Symp.*, Jun. 2018, pp. 1051–1056.

[40] A. D. Jaccard, “Nouvelles recherches sur la distribution florale,” *Bull. Soc. Vaud. Sci. Nat.*, vol. 44, no. 163, pp. 223–270, 1908.

[41] J. C. Caicedo, J. Roth, A. Goodman, T. Becker, K. W. Karhohs, M. Broisin, M. Csaba, C. McQuin, S. Singh, F. J. Theis, and A. E. Carpenter, “Evaluation of deep learning strategies for nucleus segmentation in fluorescence images,” *Cytometry*, vol. 95, no. 9, pp. 952–965, Sep. 2019.

[42] A. El-Baz and J. Suri, *Level Set Method in Medical Imaging Segmentation*. Boca Raton, FL, USA: CRC Press, 2019.

[43] D. Bannon, E. Moen, E. Borba, A. Ho, I. Camplisson, B. Chang, E. Osterman, W. Graf, and D. Van Valen, “Deepcell 2.0: Automated cloud deployment of deep learning models for large-scale cellular image analysis,” *BioRxiv*, 2018, doi: 10.1101/505032.

HYUN-GYU LEE received the B.E. and Ph.D. degrees in computer engineering from Inha University, South Korea, in 2017. He worked as a Postdoctoral Researcher at the Human-Centered Computing Research Center, in 2017, and at LG Electronics as a Research Engineer, in 2018. He is currently a Principal Research Engineer with DoAI Inc. His research interests are in artificial intelligence, computer vision, and medical imaging.

SANG-CHUL LEE (Member, IEEE) received the B.E. degree in computer science and engineering from Inha University, South Korea, in 1998, and the Ph.D. degree in computer science from the University of Illinois at Urbana–Champaign, in 2006. He was a Research Programmer at the National Center for Supercomputing Applications, from 2006 to 2008. He was a Guest Researcher at the Information Technology Laboratory, National Institute of Standard and Technology (NIST), in 2012. He is currently a Professor with the Department of Computer Engineering, Inha University, South Korea. His research interests are in computer vision, machine learning, robotics, and medical imaging.