5recently, to solve the degradation of cell

5Recently, methods that find approximate coordinates of cell region by using Blob detector such as LoG or MSER, and find the cell region by using local binarization and global binarization show good performance in cell region detection 2021. However, in the case where the contrast between the background and a cell region is low, the cell region can not be detected by the global binarization and the local binarization complementing the problem can not yield sufficiently good results. Therefore, a special image normalization method capable of increasing the contrast difference according to the image is required. In this paper, we propose a pixel-level cell region discriminator R using statistical and histogram features and logistic regression to solve the degradation of cell region detector performance due 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 value and the minimum value for the pixel value of the local square image with one-side of pixels. In addition, the distribution feature which directly expresses the distribution characteristics of pixel brightness is defined as a brightness histogram that divides the range between the minimum value and the maximum value in the image into n classes.

To concentrate on the low intensity value where the brightness values of the cell area and the background area are densely distributed, the bin corresponding to the lower 25% of the whole are used. The two kinds of features are categorized into the cell region and the background, and the pixel of the feature is identified as the cell region through the following

logistic regression. (1) where is the sigmoid function, is the learned regression parameter vector, and is the statistical feature or histogram feature vector. is the label of and has 1 if it is a cell area or 0 if it is a cell area.

The pixel-level discriminator extracts features for every pixel of the image, so the training data is very large. Therefore, the parameter vector is optimized by using the stochastic gradient descent (SGD) defined as follows for fast convergence in learning.   (2) where is the learning rate and is the mini batch sample number. When identifying a cell region through logistic regression analysis, one of the statistical and distributional features has better discrimination power depending on the image. Therefore, we define probability values of pixels for two features estimated through logistic regression as a two-dimensional ensemble feature, and detect the cell region stably using the second-order regression. The classification threshold of each regressor is experimentally set to the value that best classifies the training data.

3.   Multi-cell discriminator 3. 1. Convex surface transform A cell region segment S detected in a pixel-level cell region discrimination R actually contains one to dozens of cells.

Therefore, S consisting of a plurality of cells must be divided into individual cells for precise cell segmentation. To this end, we adopt an existing study that assumes each cell as a 2D GMM and clusters the pixel coordinates to each cell using the Expectation-Maximization (EM) algorithm that find the parameters of the probability model. The features to be used in EM are the

pixel coordinates and the coordinates of the local maximum point for them, and the initial cluster is set through k-means clustering.

The local maximum point coordinate is an important factor that influences the performance of clustering, and the closer the maximum point coordinates in a cell are, the more accurate the division is. To reduce the variance of the local maximum point coordinates within a cell, we combine the original cell image with the following distance image to convert the cell surface to a more convex shape. where, is a local image containing only one region segment , and is a Gaussian blur kernel.

is the set of edge pixel coordinates of , and is all pixel coordinates of . If the number of components k of the GMM is equal to the number of actual cells in the region segment S, we can expect that a meaningful division result is generated through the EM algorithm 2021. Depending on various conditions such as brightness of , smoothness of its surface, and the boundary brightness between adjacent cells, the partitioning method 2021 of selecting which minimize the dissimilarity between the real cells and the virtual cells generated from the GMM parameter may be not working well. Therefore, not a method using dissimilarity, we propose a multi-cell discriminator M that divides a region into a binary tree structure by determining whether it is a multi-cell.

The feature of for multi-cell identification is defined as least square error for surface fitting using 3rd order polynomial for cell surface as follows where, is the number of pixels belonging to , and is a parameter for surface fitting. As shown in the first row of Fig. 3, the least square error of a single cell by the

third-order polynomial surface fitting of region segment is smaller than that of multiple cells. The difference between consisted of a single cell and consisted of multiple cells also appears in boundary sectional area between cells divided by EM. If EM is divided into multiple cells of S, the boundaries between the models will be formed at the boundary between the two cells, as shown in Fig.

3(e) and (f). If S of the single cell is divided into two cells, boundaries between the models will be formed at the vicinity of the center of the cell because there are no obvious Gaussian mixture distribution in S, as shown in Fig. 3(d). Therefore, the boundary sectional area between cells divided by EM in of multiple cells is smaller than the sectional area estimated in of single cell, so that it is suitable for discriminating between a single cell and multiple cells. Thus, we define the boundary sectional area feature of the cells divided by EM as follows. boundary area characteristics of divided cells for S of a single cell and two cells, and discriminates whether any S is multiple cells that can be divided. First, S is divided into two cells via EM in a learning stage of M.

is composed of a single cell in the ground truth, the class of the feature vector extracted by Eqs. (4) and (5) in the divided region S is assigned as single-cell class. If S is composed of multiple cells, the class of its feature vector is assigned as a multi-cell class.

With progressive partitioning on two divided regions and in S' given a multi-cell label, a binary tree structure is built, as shown in Algorithm 1. In the learning of M, SVM is trained from extracted features for all nodes of the

tree. When testing the cell segmentation, if S is divided into 2 cells by EM

and it is discriminated as a multi-cell by using the learned M, S is divided