Microscopic Image Segmentation with Two-Level Enhancement of Feature Discriminability

Yang Song, Weidong Cai, David Dagan Feng
BMIT Research Group, School of IT, University of Sydney, Australia
{ysong,tomc,feng}@it.usyd.edu.au

***Abstract*—Microscopic cellular image segmentation has become one of the most important routine procedures in modern biological applications. The segmentation task is non-trivial, however, mainly due to imaging artifacts causing highly inhomogeneous appearances of cell nuclei and background with large intensity variations within and across images. Such inconsistent appearance profiles would cause feature overlapping between cell nuclei and background pixels and hence lead to misclassifiation. In this paper, we present a novel method for automatic cell nucleus segmentation, focusing on tackling the intensity inhomogeneity issue. A two-level approach is designed to enhance the discriminative power of intensity features, by first a reference-based intensity normalization for reducing the inter-image variations, and then a further localized object discrimination for overcoming the intra-image variations. The proposed method is evaluated on three different sets of 2D fluorescence microscopy images, and encouraging performance improvements over the state-of-the-art results are obtained.**

I. Introduction

Cell nuclear segmentation is one of the most basic processing steps in many biological image applications. With the vast amount of data, automatic segmentation is highly desirable comparing to time-consuming manual processes. However, while being visually distinguishable, it is very challenging to segment the cell nuclei automatically, due to problems such as image artifacts and intensity variations between the cell nuclei.

In particular, the intensity levels of the cell nuclei or background are often highly inhomogeneous *within* or *between* images, which are referred to as the *intra-* and *inter*-image variations. For example, intra-image variations can be seen from Fig. 2a, with some cell nuclei appearing much brighter than the others, and Fig. 2e showing the background with varying intensities; and comparison between Fig. 1a and 1e illustrates the inter-image variations, with cell nuclei in the latter image exhibiting intensities even lower than some background areas in the other image. While the traditional segmentation techniques, such as thresholding, watershed and merging [1], are easy to implement, they usually deliver relatively low segmentation accuracies. More advanced contour-based techniques are especially effective in separating touching cells [2], [3]; however, these images seem to be less affected by the inhomogeneity problem.

To explicitly address the intra-image variations, a simple way is to introduce a preprocessing step, such as the contrast filter [4]. More intelligently, an h-dome transformation based on grayscale morphology [5] has been proposed to enhance the objects for better discrimination from the background. It is also suggested that by accommodating the intensity variations into the level set model [6], inhomogeneous cell nuclei could be better delineated. Another different type of approach aims to reconstruct the ideal image, with the inhomogeneity modeled as a bias term for intensity correction, and has been shown effective for both microscopic images [7] and other imaging modalities [8].

These techniques, however, do not characterize the inter-image variations. This is rather acceptable for these methods since they perform the segmentation based on features extracted from the testing image itself. On the other hand, prior information gathered from other images can be valuable to effectively guide the segmentation, and hence various classification-based approaches have been proposed for cell image segmentation, with customized classifier [9] or feature set [10]. However, to differentiate between cell nuclei and background accurately, a clear separation between the feature ranges of the two classes is essential. Therefore, it is important to address the inter-image variation problem to avoid overlapping between the feature ranges.

Recently proposed approaches include global normalization of all image [11] and normalization between each pair of testing and reference images [12]. Such normalization techniques, however, are to simply align the intensity ranges across images. Without incorporating the discriminative information between object and background, the resultant redistribution of intensities would not necessarily lead to better feature separation. Especially for images with a large intensity span (e.g. Fig. 1e), the low-intensity cell nuclei might not be brightened much after the normalization, and would thus still appear quite similar to the background.

In this work, we present a new method for automatic segmentation of cell nuclei. Compared to previous works, we tackle both inter- and intra-image variations with two-

level enhancement of feature discriminability. Specifically, we design a two-level approach by first normalizing the intensity ranges between images for the cell nuclei (and background) based on reference images to reduce inter-image variations, and then a further localized classification for the initially detected cell nuclei (and clusters of cells) with local contrast information to handle the intra-image variations. These two steps are each optimized for better discrimination between the cell nuclei and background, with only simple region-based intensity features. We have successfully applied our method on three different sets of fluorescence microscopy images, and shown promising results compared to the state-of-the-art performances reported on the same data sets. Furthermore, our method is essentially designed for foreground-background segmentation based on distinctive intensity ranges, and relies on few other assumptions specific to the problem domain; therefore, we suggest that the proposed method can be generally applicable to other imaging modalities as well.

II. Proposed Method

A. Basic Segmentation Approach

Given an image $I = \{i\}$ of N pixels i, we formulate the segmentation objective as a pixel-wise classification problem: find the optimal labeling $L = \{l_i\}$ with pixel labeling $l_i = \{0, 1\}$, representing the background and foreground (i.e. cell nucleus) types respectively. Since individual pixels provide limited information, pixel-level processing would often produce noisy labelings, especially for the foreground. Such a problem can be reduced with region-based techniques, by clustering the nearby pixels into homogeneous superpixels, to incorporate the local spatial information.

We thus first cluster the image I into regions (i.e. superpixels) using quick-shift clustering [13]. Each region r is then characterized by a set of features f_r, and its labeling $l_r = \{0, 1\}$ is derived using a binary support vector machine (SVM) with linear kernel. Since the foreground pixels normally exhibit distinctive image intensities from the background, we choose to define a simple two-dimensional feature vector: (i) a_r the average intensity of r; and (ii) $(a_r - \max_{r'} a_{r'})$ the intensity difference with the neighboring regions r'.

B. The Inhomogeneity Problem

The feature set f_r is essential for obtaining accurate labeling; however, the discriminative power of f_r is restricted by intensity inhomogeneity across and within images, as shown in Fig. 1 and 2; and it would thus be difficult to obtain a clear separation between the foreground and background for all images using a single classifier. While it is intuitive to consider normalizing the images, it is not normally recognized that the variations are actually specific to each cell (true and false positive ones), not individual pixels or the whole images.

Therefore, we propose that a more effective way is to address the inhomogeneity issue at cell-level. However, to be able to process at cell-level, the images need to be segmented first to identify the individual cells. We thus design a two-level method to enhance the feature discriminability: first, prior to the feature computation, the images are normalized to the similar intensity ranges based on reference images; and second, after an initial region-based segmentation, the area surrounding each cell nucleus or cluster is further classified for refined discrimination. The details are described in the following two sections.

C. Level-1: Reference-based Intensity Normalization

At the first level, our objective is to normalize across the images, so that all images display similar *average* foreground intensities and similar contrasts between the foreground and background. As a result, the inter-image variations are reduced, and the feature ranges of foreground and background are better separated to suit a single classifier. Note that within a certain image, after the normalization, the foreground (and background) might still exhibit inhomogeneous intensity levels. Example normalization outputs are shown in Fig. 1 and 2.

To do this, for image I, a linear normalization is performed for each pixel I_i:

$$\hat{I}_i = I_i / p_I \cdot C \tag{1}$$

where C is a scaling constant (e.g. 64), p_I is an image-specific reference intensity, and $\hat{I}_i$ is capped at 255 (for grayscale image). As a result, pixels of value p_I are rescaled to C and other pixels linearly redistributed, so that all images can be aligned in a new intensity range with C as the origin. The problem then becomes how to determine the reference value p_I.

The ideal case is that after normalization, all images would exhibit standardized intensity distributions and such a distribution establishes a good contrast between the foreground and background. Therefore, we expect that p_I would serve as a separation point between the two types. A straightforward idea is thus to compute a threshold (e.g. Otsu's). However, such a computation would not be robust, mainly due to cell nuclei being highly inhomogeneous. Especially for images containing very bright areas (e.g. Fig. 1e), the resultant threshold would be usually quite high, causing the dark cell nuclei to fall into the background group. Consequently, computing p_I directly from I itself is quite difficult, and we choose to leverage on other images with segmentation ground truths to derive p_I.

The intuition is that, if images I and J are similar in their intensity distributions, then p_I is likely to be similar to p_J.

Therefore, we would like to find a set of images that are similar to I, and use their reference values to approximate p_I. The problem is then formulated as – for image I and given Q reference images $\{J\}$, the reference value p_I can be estimated as weighted combination of p_J:

$$p_I = \frac{1}{\sum_J \alpha(I,J)} \sum_J \alpha(I,J) \cdot \beta(I,J) \cdot p_J \qquad (2)$$

where α and β are two types of weights, and p_J is the average value between the minimum and mean intensity of the foreground in J. We choose such a p_J since we want it to create a good separation between the background and foreground, but not with the stringent requirement as a threshold; and higher values could cause under-segmentation of cell nuclei, while lower ones often result in too much overlap with the background.

The reference images $\{J\}$ are randomly selected from the data sets and need not all to be similar to I, and hence a similarity measure between I and J is important to determine the weight of contribution from J. We thus define the weights as (i) $\alpha(I,J)$ the resemblance factor between I and J, and (ii) $\beta(I,J)$ the equalization factor between I and J. Here α is computed using a sparse-regularized regression model:

$$\min_{\alpha} \parallel h_I - H\alpha \parallel_2^2, \; s.t. \parallel \alpha \parallel_0 \leq K \qquad (3)$$

where $\alpha \in \mathbb{R}^{Q\times 1}$ denotes the vector of all weights, $h_I \in \mathbb{R}^{n\times 1}$ is the grayscale histogram of I (n is the histogram dimension) and $H \in \mathbb{R}^{n\times Q}$ the histogram matrix of $\{J\}$, and K is a constant ($K < Q$). A higher $\alpha(I,J)$ means a higher resemblance between I and J and hence a higher contribution from J to p_I. We also restrict the number of non-zero elements in α to K, so that only K reference images would contribute to the estimation, to avoid overfitting. The minimization problem is solved efficiently using the OMP algorithm [14]. The weight $\beta(I,J)$ is computed as the ratio between the mean intensities of I and J, to adapt p_J to the suitable range for image I. The $\beta(I,J)$ is necessary since even the most similar images might exhibit different intensity ranges from I, and we find a mean-intensity based equalization quite effective for the adaptation.

Take Fig. 1e as an example. Originally the image contains one bright and multiple very dark cell nuclei. If a standard intensity normalization is performed by assuming desired intensity range, the dark cell nuclei would still remain similar to the background. Here by having a random half of the image data sets as the reference set J, the regression model Eq. 3 derives the weight of contribution $\alpha(I,J)$ for each reference image; and together with $\beta(I,J)$, the reference value p_I is then computed using Eq. 2. In this case, p_I is 8.12, which is roughly the desired

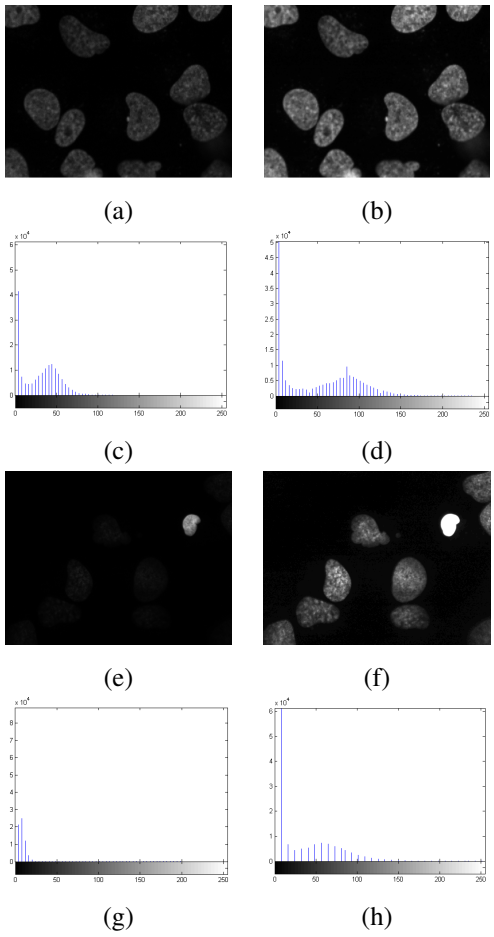

Fig. 1. Illustration of level-1 based on two examples from data set 1. (a) and (e): Parts of the original images. (b) and (f): Intensity normalized images. (c) and (g): Histograms of the original images. (d) and (h): Histograms of the normalized images.

value measuring from the histogram Fig. 1g. Then by the linear normalization Eq. 1, all pixel intensities are scaled up by a factor of $C/p_I \approx 8$ (maximum intensity is 255); and as a result, the cell nuclei become much brighter as shown in Fig. 1f. The histograms Fig. 1d and 1h of the two different images are also better aligned, hence a global classifier would be more effective. Similar effects can also be seen in Fig. 2.

D. Level-2: Localized Object Discrimination

While the normalization step helps to reduce the inter-image intensity variations, within a certain image, the foreground and background can still be largely inhomogeneous, i.e. intra-image variations. Such a problem often causes difficulties in delineating the border areas of cell nuclei, especially for those with lower intensities or if enclosed by relatively bright background. Nevertheless, for localized area surrounding a cell nucleus, the foreground normally

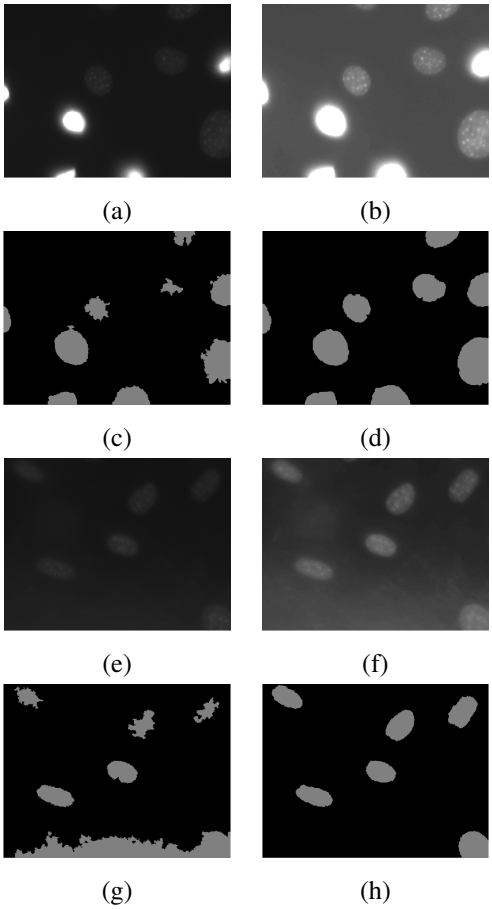

Fig. 2. Illustration of level-2 based on two examples from data set 2. (a) and (e): Parts of the original images. (b) and (f): Intensity normalized images. (c) and (g): The initial segmentation outputs. (d) and (h): The final segmentation results after localized object discrimination.

exhibits quite distinctive intensities from the background. This thus motivates us to perform a second level enhancement based on localized contrast information surrounding individual cell nucleus (or cluster) rather than global image-level processing.

To do this, we first compute the feature f_r (Section II-A) based on the normalized intensities, and derive an initial segmentation using a binary SVM (Fig. 2c and 2g). A connected-component analysis is then performed on the segmentation mask to obtain isolated foreground areas, with each denoted as R. Subsequently for each localized R, we refine its labeling with the following (Fig. 2d and 2h).

First, we want to identify if R is likely a high-intensity background, cell, or cluster. We differentiate the latter two types since: (i) if R is a cell (or a few tightly connected ones), the initial segmentation could be over- or under-segmented, and inclusion of its surrounding area is necessary to incorporate the local contrast information; and (ii) if R is a cluster, it normally contains cells that are connected with brighter-than-normal background (as shown in second example of Fig. 2), and the local contrast information depicted in R itself is more suitable for better discrimination. We thus use a three-class SVM for the differentiation, based on the mean intensity, standard variation, size, grayscale threshold (Otsu's), and solidity of R. We then discard the pure background region from further processing; and if R is likely a cell, it is expanded via contour dilation with a constant width (set to 20 pixels in this study) to include its surrounding area.

Next, we cluster the pixels in R into background and foreground using k-means clustering ($k = 2$), without prior trainings. Note that since R is a localized area, its intensity inhomogeneity is much lower than that of the entire image I; and with the rather clear local contrast between foreground and background, a simple k-means clustering is thus quite effective in discriminating the two types. And for each pixel i in R, we compute its probability of representing the foreground using:

$$Pr(l_i = 1|R) = \min\{\max\{\frac{I_i - u_{R,0}}{u_{R,1} - u_{R,0}}, 0\}, 1\} \quad (4)$$

where I_i is the normalized intensity of i; and $u_{R,0}$ and $u_{R,1}$ are the average intensity of the clustered background and foreground in R. And the probability of i representing the background is thus:

$$Pr(l_i = 0|R) = 1 - Pr(l_i = 1|R) \quad (5)$$

While the clustering step only considers the global intensities in R, the local contrast information between neighboring areas is also a useful hint. In particular, if the neighboring areas display similar intensities, they are likely to take the same label. Especially for pixels with intensities around the middle of $u_{R,0}$ and $u_{R,1}$, analyzing their similarities with the surrounding pixels would help to improve the spatial consistency of labeling outputs. To model such structural information for R, rather than labeling each pixel individually, we construct a conditional random field (CRF) [15] to further improve the labeling:

$$E(L_R|R) = \sum_r \phi(l_r) + \sum_{r,r'} \psi(l_r, l_{r'}) \quad (6)$$

Here $\phi(l_r)$ represents the unary cost of region r taking the label l_r. The regions $R = \{r\}$ are created using quick-shift clustering, and

$$\phi(l_r) = 1 - Pr(l_r|R) \quad (7)$$

$$Pr(l_r|R) = \frac{1}{N_r} \sum_i Pr(l_i|R) \quad (8)$$

where N_r is the number of pixels in region r. The pairwise term $\psi(l_r, l_{r'})$ penalizes the labeling difference between the neighboring regions r and r', following the usual Potts

model based on the average intensity differences between r and r':

$$\psi(l_r, l_{r'}) = \exp(-\frac{|I_r - I_{r'}|}{2\gamma_R})\mathbf{1}(l_r \neq l_{r'}) \quad (9)$$

where γ_R is the normalization factor as the average of all intensity distances between neighboring regions in R. Empirically, the unary and pairwise terms are assigned equal weights. The labeling set L_R is then obtained by minimizing $E(L_R|R)$ using graph cut [16].

Take Fig. 2e as an example. The initial labeling produced (Fig. 2g) assigns a large portion of the background as foreground while some parts of the actual cell nuclei are misclassified as the background, due to the intra-image variations. Then with the three-class SVM, five cell nuclei and one cluster are identified and further labeling is performed individually. For the cluster, although it exhibits low contrast internally, the k-means clustering step is still able to produce good separation between the foreground and background. For the cell nucleus, a contour dilation is necessary to include its surrounding areas, so that more pixels would be possibly labeled as foreground. And based on the CRF construct, the final labeling output shows much better delineation of the cell nuclei (Fig. 2h).

III. Experimental Results

A. Data Sets

The experiment is performed on three different sets of 2D fluorescence microscopic images of cell nuclei, with two in grayscale [1], and one of color images [4]. The three data sets consist of 48, 49 and 10 images respectively, with a total of 1831, 2178 and 254 cell nuclei, all including segmentation ground truth. For each set, we use the leave-one-out scheme for the intensity normalization, but limit the number of real contributing reference images (K) to half of the data size. Training of classifiers is conducted for each data set separately, with 5 images selected from each data set as the training set. Testing is performed for all images for the first two data sets [6], and the remaining 5 images are used for testing on the third data set [10]. Our results are also compared with the state-of-the-art performances reported in [6] and [10] for the same data sets.

B. Results on Data Set 1 & 2

We employ the same performance measures as used in [6], including the Dice coefficient for pixel-level accuracy, the normalized sum of distances (NSD) and Hausdorff distance for contour delineation, and the number of false positives (FP) and false negatives (FN) for object-level detection rate. While higher Dice corresponds to better performance, the other four measurements favor lower values. As shown in Table I and II, our proposed method achieves overall higher performance than the level set method [6]. It is especially encouraging that while our method is based on region-based classification, rather than contour-based modeling, we actually obtain smaller NSD and Hausdorff distance. On the other hand, our method seems to trade-in FP for FN, especially for the first data set, due to our classification routine in level-2 keeping most cell nuclei or clusters rather than discarding as background. Furthermore, we also assess the following variations of our proposed method: (i) basic segmentation (Section II-A); (ii) segmentation after intensity normalization (level-1 only); (iii) basic segmentation then localized object discrimination, without the intensity normalization (level-2 only). The results illustrate the benefits of level-1 and level-2 comparing to the basic region-based segmentation, and the advantage of the two-level approach. Note that data set 2 is more challenging than data set 1, with much higher intensity inhomogeneities and more cell nuclei clustered with bright background; and hence larger improvements can be observed when comparing the proposed method with the basic segmentation.

TABLE I
The segmentation performance on data set 1.

	Dice	NSD	Hausdorff	FP	FN
Level set [6]	0.94	0.06	13.3	0.5	3.9
Basic	0.91	0.06	13.9	5.3	0.7
Level-1 only	0.93	0.03	10.0	3.3	0.7
Level-2 only	0.92	0.04	10.1	2.2	2.7
Proposed	0.94	0.02	9.5	2.0	2.5

TABLE II
The segmentation performance on data set 2.

	Dice	NSD	Hausdorff	FP	FN
Level set [6]	0.83	0.14	16.5	1.7	11.3
Basic	0.65	0.39	16.4	5.1	4.5
Level-1 only	0.78	0.25	10.7	4.0	1.7
Level-2 only	0.80	0.18	11.5	2.1	5.7
Proposed	0.85	0.14	10.5	1.6	3.4

C. Results on Data Set 3

To show our method extensibility to color images, and images having cytoplasm in addition to cell nuclei, we evaluate further on this data set. The pixel- and object-level accuracies [10] are measured, and including NSD and Hausdorff distance as contour-based metrics. The intensity normalization is performed for each RGB channel, and we use RGB, LUV and grayscale intensities as the feature vector for classification; and due to the presence of cytoplasm, the k-means clustering in level-2 is performed to cluster

TABLE III
THE SEGMENTATION PERFORMANCE ON DATA SET 3.

	Pixel Acc	Obj Acc	NSD	Hausdorff
Feature [10]	0.851	0.840	–	–
Basic	0.837	0.845	0.06	3.9
Level-1 only	0.849	0.903	0.05	3.8
Level-2 only	0.846	0.877	0.04	4.0
Proposed	0.852	0.907	0.05	3.8

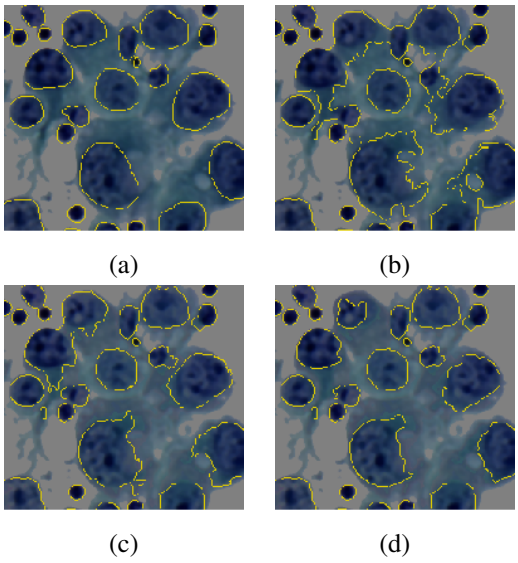

Fig. 3. Example results on data set 3. (a) Part of original image (yellow contours indicating the segmentation ground truth). (b) Segmentation outputs using basic approach. (c) Level-1 enhancement only. (d) Our proposed method.

three classes, with the lowest intensity class as the cell nuclei. As shown in Table III, while our method is based on simple intensity features only, it achieves comparable pixel-level and much higher object-level accuracies, compared to [10]. The intensity variations between cell nuclei are also quite high in this data set, and hence the improvement of level-1 over the basic segmentation is clear. Furthermore, since the cell nuclei are surrounded by cytoplasm with intensities similar to the cell nuclei, classification at the border is more difficult, and the inclusion of level-2 is thus helpful in better delineation of the cell nuclei. The effects can also be visually perceived from Fig. 3.

IV. CONCLUSIONS

We proposed a new method for automatic segmentation of cell nuclei from 2D fluorescence microscopy images. While being visually distinguishable, the intensity inhomogeneity within and between images tends to be a major issue in intensity-based classification approach. We thus design a two-level approach to enhance the feature discriminability by reference-based intensity normalization and localized object discrimination. We have demonstrated the effectiveness of our method on three different data sets. We also suggest that by using more comprehensive features for region classification, in addition to the intensity features, we expect higher segmentation accuracies; and that our proposed method is also easily adaptable to other imaging domains.

REFERENCES

[1] 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. ISBI*, pp. 518–521, 2009.
[2] C. Zhang, C. Sun, and T. D. Pham, "Clustered nuclei splitting using curvature information," *in Proc. DICTA*, pp. 352–357, 2011.
[3] M. Horn and M. R. Berthold, "Towards active segmentation of cell images," *in Proc. ISBI*, pp. 177–181, 2011.
[4] O. Lezoray and H. Cardot, "Cooperation of color pixel classification schemes and color watershed: a study for microscopical images," *IEEE Trans. Image Proc.*, vol. 11, no. 7, pp. 783–789, 2002.
[5] S. H. Rezatofighi, R. Hartley, and W. E. Hughes, "A new approach for spot detection in total internal reflection fluorescence microscopy," *in Proc. ISBI*, pp. 860–863, 2012.
[6] J. P. Bergeest and K. Rohr, "Fast globally optimal segmentation of cells in fluorescence microscopy images," *in MICCAI LNCS*, pp. 645–652, 2011.
[7] K. Li and T. Kanade, "Nonnegative mixed-norm preconditioning for microscopy image segmentation," *in IPMI LNCS*, pp. 362–373, 2009.
[8] C. Li, R. Huang, Z. Ding, J. C. Gatenby, D. N. Metaxas, and J. C. Gore, "A level set method for image segmentation in the presence of intensity inhomogeneities with application to mri," *IEEE Trans. Image Proc.*, vol. 20, no. 7, pp. 2007–2016, 2011.
[9] Z. Yin, R. Bise, M. Chen, and T. Kanade, "Cell segmentation in microscopy imagery using a bag of local bayesian classifiers," *in Proc. ISBI*, pp. 125–128, 2010.
[10] L. Cheng, N. Ye, W. Yu, and A. Cheah, "Discriminative segmentation of microscopic cellular images," *in MICCAI LNCS*, pp. 637–644, 2011.
[11] C. Chen, J. A. Ozolek, W. Wang, and G. K. Rohde, "A pixel classification system for segmenting biomedical images using intensity neighborhoods and dimension reduction," *in Proc. ISBI*, pp. 1649–1652, 2011.
[12] H. Chang, L. A. Loss, P. T. Spellman, A. Borowsky, and B. Parvin, "Batch-invariant nuclear segmentation in whole mount histology sections," *in Proc. ISBI*, pp. 856–859, 2012.
[13] A. Vedaldi and S. Soatto, "Quick shift and kernel methods for mode seeking," *in ECCV LNCS*, pp. 705–718, 2008.
[14] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 20, pp. 2231–2242, 2004.
[15] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," *in Proc. ICML*, pp. 282–289, 2001.
[16] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 2, pp. 147–159, 2004.