

Segmentation of Heavily Clustered Cell Nuclei in Histopathological Images

Rahul Singh¹(✉), Mukta Sharma¹, and Mahua Bhattacharya²

¹ Indian Institute of Information Technology and Management, Gwalior, India
rhlsingh720@gmail.com, mukta.24sharma@gmail.com

² Department of Computer Science and Engineering,
Indian Institute of Information Technology and Management, Gwalior, India
mahuabhatta@gmail.com

Abstract. Automated cell nuclei determination of stained images is of uttermost importance for diagnosis. In this work, we have proposed a novel efficient and accurate image segmentation technique for densely clustered overlapping cell nuclei. Firstly, we have extracted the cell body (foreground) from the background using global thresholding followed by local thresholding. Then, we have employed the fusion of seeded region growing technique and level-set algorithm. The initial seed points need to be selected accurately and precisely in order to generate appropriate outcomes from region growing framework. Initial contours for level-set evolution relies heavily on an output of this adaptive region growing approach and some morphological operations. Finally, Global Gaussian distribution with several means and variances is employed in an enhanced edge-based level-set approach for precise nuclei segmentation. We have performed our analysis on Nissl stained EMF exposed, and SHAM exposed cell images. The proposed framework is very much capable of extracting the cell nuclei from stained cell images. Experimental outcomes reveal that our approach has out-performed existing state of art techniques for cell nuclei extraction and segmentation.

Keywords: Nissl staining · EMF and SHAM exposed · Global and local thresholding · Seeded region growing and level-set approach

1 Introduction

Microscopic image analysis is an area for researchers when it comes to the domain of cell biology [1–5] computations. Several techniques have already been developed. Still, at the single cell level, the analysis of cell cultures is quite a task. Therefore, the techniques capable of accurately analyzing the cellular phenotypes images at the single cell level is the need of the hour as it plays a vital character in enhancing the treatment of human diseases. So, the basic necessity of such analysis is precise and accurate cell segmentation on the basis of intensity, morphology or texture. Therefore, several single cell detection techniques have been put-forward. Amongst them, the simplest of the methods rely on local or global thresholding which in most of the cases depends on image intensity histogram. Such methods usually have petite computational constraints

[6–9]. Some of the single cell segmentation techniques utilize inherent characteristics of image intensity for cell detection [10]. Machine learning techniques have also contributed a lot to such image analysis. But, when it comes to the complex spatial arrangements, all these above-mentioned techniques failed to detect multiple cells. So, a feasible way of overcoming the mentioned limitation could be to incorporate the segmentation algorithms with the prior shape details regarding the objects to be segmented. A common practice could be to fit defined rigid structures to datasets and recognize the equivalency [11–15]. These approaches can deal with the overlapping objects to quite an extent but fail completely when it comes to capturing the small shape variations. Several studies have already been put-forward regarding the automatic analysis of cell nuclei for the Immunohistochemistry (IHC) images. Active contours have gained popularity over the years and are quite handy for medical image analysis [16].

However, the conventional approach does not work satisfactorily to overlapping nuclei segmentation. But, its derivatives seem to work well for certain complex scenarios. An active contour employed with level set algorithm has gained popularity for nuclei segmentation of fluorescence images [17]. An unsupervised Bayesian classification approach on mammary invasive ductal carcinomas has been utilized for nuclei segmentation [18]. To segment the overlapping cells in breast cancer, an approach utilizing single-path voting following fusion of level set algorithm and mean-shift clustering has been employed [19]. Unsupervised clustering structure for nuclei segmentation has been used in case of prostate cancer [19].

Even though these mentioned techniques yield satisfactory results, but when it comes to the segmenting densely clustered cell nuclei, these approaches failed miserably. Addressing such an issue is quite a challenging task because of several complications. Firstly, tissue sections and color stains are often not produced evenly because of which separating the nuclei from the cluttered background is a cumbersome task. Secondly, huge intensity swings in the nuclei make the segmentation of such cell images difficult. Thirdly, there are always chances of under-segmentation because the 3-dimensional entity is being represented as 2-dimensional. The reason behind that is different planes of histological sections may seem to be oriented and overlapped when observed from a different perspective. Watershed algorithms are the most common technique to segment the overlapping cell nuclei [20]. Although the conventional watershed algorithm employs the regional minima as initial points which might result in over-segmentation. So, rather than using regional minima as initial points, researchers first detected the seeds and then employed those seeds as initial points [21, 22]. For satisfactory operation of such approach greatly depends on precise detection of these seeds. An iterative voting approach has been proposed to find the centers of cluttered cell nuclei which in turn is used for efficient detection of these seeds [23]. Laplacian of Gaussian (LoG) with distance constraints can also be used detect the seeds [24]. These techniques yield satisfactory results. However, these techniques fail when heavily clustered nuclei are there along with poorly defined boundaries.

In this paper, we have introduced a novel overlapping cell nuclei segmentation technique of Histopathology images. Our approach is capable of segmenting the heavily clustered and overlapped nuclei avoiding over-segmentation and

under-segmentation at the same time. Initially, we have separated the foreground (cell body) and background using global thresholding technique followed by local thresholding approach. Secondly, we have employed the fusion of seeded region growing method along with level set algorithm to segment the cell nuclei. The proposed method has edge over the conventional techniques of nuclei segmentation. Segmentation precision and computational time have been enhanced using this framework comprising of region growing and level-set approach. Some of them are (1) The results obtained by the seeded region growing technique is automatically fed as an initial contour to level-set approach; (2) Mean and variances Gaussian distribution are embedded to level-set approach. The remaining portion of the paper has been structured as following. In Sect. 2, we demonstrated our proposed cell nuclei segmentation technique and relevant theoretical knowledge. Results and discussion are explained in Sect. 3 of the paper. In Sect. 4, we have stated conclusion from our end.

2 Methodology

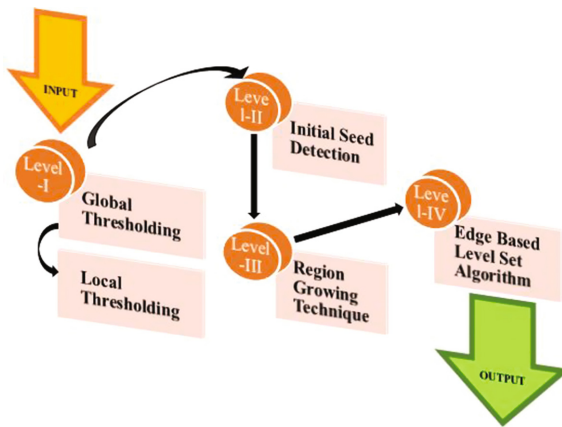


Fig. 1. Flow chart of proposed algorithm

Our proposed methodology comprises of three primary steps. The first step deals with the classification of foreground (cell body) and background. The second step deals with detection of initial seeds for region growing. Thirdly, region growing is used as prior information for the initialization of level-set. Apart from this, informatics regarding Gaussian distribution has also been availed to level-set. The proposed algorithm flowchart has been depicted in Fig. 1. Let us discuss each one of them in detailed manner.

2.1 Classification of Cell and Background

The intensity level of foreground that is the pixels possessing to the cell nuclei can be separated from the background by employing global thresholding approaches [25]. But, when it comes to the stained cell images, the mean intensity levels tend to under drastic changes throughout the entire image background because of which nuclei border is pretty much susceptible to be misclassified. Thereby, the global thresholding techniques do not yield acceptable results. We have employed both global as well as local thresholding approaches to counter this issue. Local thresholding method uses an adaptive valley point value in local regions. Therefore, it can encounter the intensity

variations through the background [26]. Local thresholding technique further processes those pixels which are distinguished as foreground (nuclei) by global thresholding approach. The thresholding accuracy is greatly influenced by window size of local regions and step-size of the sliding window. The size of the selected window must be greater than that of cell nuclei to efficiently extract the cells from the image (background). If the size of the window is more than that of nuclei, then it would lead to under-segmentation. The same holds good for the opposite case too. So, it becomes important to select proper window size. In Fig. 2, we have shown the results that acquired using various window sizes. For the sake of analysis, we have empirically set the window size to 60×60 . The window size defines the linear horizontal movement which is governed by the following relation,

$$1 \leq d \leq s;$$

Where s is the window size while d is the linear horizontal movement. In Fig. 3, we had demonstrated the results obtained when different horizontal distances were used. It is pretty clear that when small horizontal distances are used, most of the nuclei pixels are missed. Despite the fact that small distances work fine when it comes to the tracking the borders of heavily clustered nuclei, but it fails miserably for weakly stained nuclei. We have used the horizontal distance of 30 pixels. Once cell body is extracted from image background. Now, we will find the seeds for proper functioning of region growing technique.

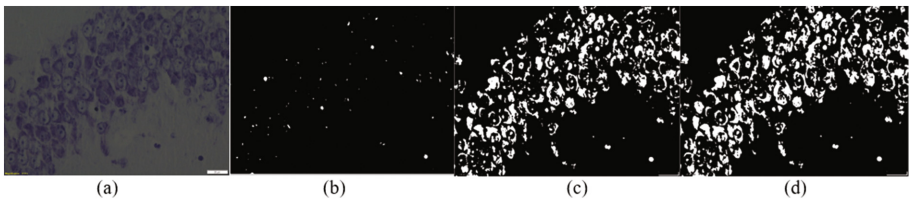


Fig. 2. Results obtained against various window sizes. ((a). Original image (b). Window size of 10×10 (c). Window size of 25×25 (d). Window size of 60×60)

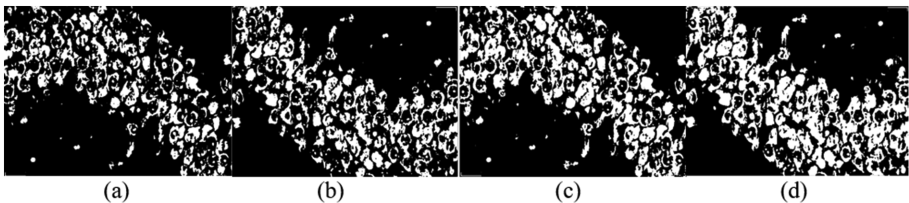


Fig. 3. Results obtained against various horizontal distances for (a). 1 pixel (b). 10 pixels (c). 15 pixels (d). 30 pixels

2.2 Selection of Seed Points

The initial selection of seed points: The gray values of cell nuclei structures and its adjacent anatomical structure's gray values are indistinguishable. These regions have small coverage area. To make the complex computation easy and the segmentation more authentic, we only process a local area along with entire cell nuclei region and some other regions. Primarily, we have selected the initial seed points manually from the original image of cell nuclei. Then we get slice image having initial seed at the center position. We have done shearing to reduce the disturbance of cell nuclei segmentation because it will filter out the bulk of cerebrospinal fluid, white matter, and gray matter. Slice images represents the features of cell nuclei better and hence make the segmentation more accurate.

2.3 Region Growing Framework

From given initial seeds, the neighboring pixels having equivalent properties are appended to grow the region. We have chosen standard deviation as the growth rule corresponding to the special position of cell nuclei. It can be formulated as:

$$|I_s(x, y) - \mu_{seed}| \leq T_{thres} = \varepsilon \sigma_{global} \quad (1)$$

Where, $I_s(x, y)$ is the slice image, μ_{seed} is the mean intensity of current seed points, σ_{global} is the global standard deviation of the slice image, and ε is the weight of σ_{global} . All the connected blocks derived by the quad-tree decomposition, are allowed to include due to this criterion, whose intensity does not differ from the seed point intensity more than manually define threshold value. According to this examination, σ_{global} can be formulated as:

$$\sigma_{global} = \sqrt{\frac{1}{p \times q} \sum_{x=1}^p \sum_{y=1}^q (I_s(x, y) - \mu_{seed})^2}, \quad (2)$$

$$\mu_{seed} = \frac{1}{p \times q} \sum_{x=1}^p \sum_{y=1}^q I_s(x, y)$$

where, $p \times q$ is the size of slice image and μ_{seed} is the mean intensity of slice image. Pixels belonging to the 8-connected neighborhood of seed points satisfy the growth criterion and add to the seed point set.

2.4 Morphological Operations

Group of nonlinear operations related to shape or morphology features of an image are called morphological operations. Primarily, we have calculated the perimeter of the homogenous region obtained by region growing framework. As a result, we get is not exactly the same as the cell nuclei region. So, we have applied the level set method developed from region growing method is applied to upgrade the output. The perimeter

fitting has been done to gain the convex polygon which prevents falling into local minima and further used to initialize the level set algorithm.

2.5 Level-Set Algorithm

We obtain some prior information from region growing and fuse global Gaussian distribution information with the standard level set model to increase the result accuracy in detection of cell nuclei. Global information plays the beneficiary role for regions having weak boundaries. Especially it is critical to minimize the sensitivity to the initialization of the contour. From statistical view interpretation, the proposed model fitting energy is the fusion of Li’s model and global Gaussian distributions having distinct means and variances, respectively. The global Gaussian distribution fitting energy can be formulated as:

$$\psi_{GDF} = \sum_{k=1}^2 \int_{\Omega_k} \log p_k(I(r))dx, k = 1, 2 \tag{3}$$

where, $\Omega_1 = \text{in}(C)$, $\Omega_2 = \text{out}(C)$ and

$$p_k(I(r)) = \frac{1}{\sqrt{2\Pi}\sigma_k} \exp\left(-\frac{(I(r)-\mu_k)^2}{2\sigma_k^2}\right) \tag{4}$$

where, μ_k and σ_k are global intensity means and standard deviation, respectively.

Suppose, $\Omega \in R^2$ be a two-dimensional image space and $I : \Omega \rightarrow R$ be a given gray image. Assume that image domain segregated into two regions. These two regions represent the two regions i.e. the inside and outside region of zero level ϕ ; i.e. $\Omega_1 = \{\phi > 0\}$ and $\Omega_2 = \{\phi < 0\}$. We can rewrite the global Gaussian distribution fitting energy function by using Heavside function H_ϵ as:

$$\psi_{GDF}(\phi(x)) = \int_{\Omega_1} -\log p_1(I(r)) H_\epsilon(\phi(x))dx + \int_{\Omega_2} -\log p_2(I(r))(1 - H_\epsilon(\phi(x)))dx \tag{5}$$

In this analysis, we have elucidated the energy function as follows:

$$\begin{aligned} E(\phi(x)) &= \mu T(\phi(x)) + \kappa K_g(\phi(x)) + \lambda S_g(\phi(x)) + \omega \psi_{GDF}(\phi(x)) \\ &= \frac{1}{2} \mu \int_{\Omega} (|\Delta\phi(x)| - 1)^2 dx \\ &\quad + \kappa \int_{\Omega} \xi_\epsilon^\pm(\phi(x)) |\Delta\phi(x)| dx + \lambda \int_{\Omega} \xi H_\epsilon(-\phi(x)) dx \\ &\quad + \omega \left[\int_{\Omega} -\log p_1(I(r)) H_\epsilon(\phi(x)) dx + \int_{\Omega} -\log p_2(I(r))(1 - H_\epsilon(\phi(x))) dx \right] \end{aligned} \tag{6}$$

By calculus of variations, it is interpreted that the parameters μ_k and σ_k^2 that reduce the energy functional in (above Eq. 6) satisfy the Euler’s Lagrange equation written below:

$$\int (I(r) - \mu_k)N_{k,\varepsilon}(\phi(x))dx = 0$$

$$\int (\sigma_k^2 - (I(r) - \mu_k)^2)N_{k,\varepsilon}(\phi(x))dx = 0$$
(7)

where, $N_{1,\varepsilon}(\phi(x)) = H_\varepsilon(\phi(x))$ and $N_{2,\varepsilon}(\phi(x)) = 1 - H_\varepsilon(\phi(x))$.

From Eq. 7, we obtain;

$$\mu_k = \frac{\int I(r)N_{k,\varepsilon}(\phi(x))dx}{\int N_{k,\varepsilon}(\phi(x))dx}$$
(8)

$$\sigma_k^2 = \frac{\int (I(r) - \mu_k)^2 N_{k,\varepsilon}(\phi(x))dx}{\int N_{k,\varepsilon}(\phi(x))dx}$$
(9)

which reduce the energy functional $E(\phi)$ for fixed ϕ .

Minimization of the energy functional $E(\phi)$ can be achieved by solving the gradient decent flow equation given below with respect to ϕ .

$$\frac{\partial \phi}{\partial t} = \mu \left[\Delta \phi - \operatorname{div} \left(\frac{\Delta \phi}{|\Delta \phi|} \right) \right] + \kappa \zeta_\varepsilon(\phi) \operatorname{div} \left(\xi \frac{\Delta \phi}{|\Delta \phi|} \right) + \lambda \xi \zeta_\varepsilon(\phi) - \omega \zeta_\varepsilon(\phi) (g_1 - g_2)$$
(10)

where,

$$g_1 = \log \left(\sqrt{2\Pi} \sigma_1 \right) + \frac{(I(r) - \mu_1)^2}{2\sigma_1^2}$$
(11)

$$g_2 = \log \left(\sqrt{2\Pi} \sigma_2 \right) + \frac{(I(r) - \mu_2)^2}{2\sigma_2^2}$$
(12)

$$H_\varepsilon(x) = \frac{1}{2} \left(1 + \frac{2}{\Pi} \arctan \left(\frac{x}{\varepsilon} \right) \right)$$
(13)

$$\zeta_\varepsilon(x) = \frac{1}{\Pi} \frac{\varepsilon}{\varepsilon^2 + x^2}$$
(14)

3 Results and Discussion

We have performed our analysis on the Nissl stained dataset which comprises of 64 EMF exposed images at 900 MHz and 56 sham exposed images. These Hippocampal CA3 region images were captured under $60 \times$ objective, i.e., 600 magnification. Female Swiss albino mice of 6–8 weeks of age were selected and exposed to 900 MHz (Power level +10 dBm, SAR 0.00831 W/kg) EMF exposure in GTEM cell for 120 days (6 days/week, and 2 h/days, with 5 min break after every 30 min). Sham-exposed mice were kept under the same condition without radiation exposure. These are courtesy of Department of Life Science Engineering, Manipal University, Manipal, Karnataka, India. We have used the window size of 60×60 pixels and horizontal moving distance of 30 pixels. Both these parameters were found empirically. A steadfast and systematic approach is required to find these parameters. It could also be an active area of research. The experiments were performed upon Intel i7 ACER with 4 GB RAM and 3.4 GHz processor operating under Windows 8 Pro. We have performed our algorithm on MATLAB[®] R2015a. The program can be executed on MATLAB platform.

In our proposed work, we have employed seeded region growing technique followed by level-set algorithm to extract and segmentation the cell nuclei from histopathological images. We have evaluated the flexibility of our algorithm on 64 EMF exposed images and 56 sham exposed images. Moreover, we have compared our results with some of existing conventional techniques and the results clearly reveals that our approach has out-performed those techniques by quite a margin. The well-known techniques for cell nuclei segmentation that are involved in the comparison are: Al-Kofahi et al. [24] which was evaluated on FARSIGHT [28], publicly available iterative voting procedure [23], conventional watershed approach on globally thresholded datasets [25] and extended minima dataset (E-min) [27]. We have demonstrated the results in Table 1. We have demonstrated the results of our proposed algorithm in Fig. 4.

Table 1. AS = Auto Segmentation, the number of segmented nuclei result. CD = Correct Detection. OS = Over Segmentation. US = Under Segmentation. Miss = Miss Segmented nuclei. FP = False Positive. AR = Average Accuracy Rate, the average correct detection rate. OR = Overall Accuracy Rate, CD/GD (Ground Truth).

	AS	CD	OS	US	Miss	FP	AR	OR
Global [25]	902	661	14	203	379	20	51.7%	53.4%
E-min [27]	923	784	31	76	368	31	60.8%	60.5%
Voting [23]	808	440	73	211	539	82	32.7%	36.4%
LoG [24]	1638	828	339	43	45	419	62.3%	67.1%
Proposed technique	1450	1058	51	72	121	69	82.8%	81.3%

As far as performance evaluation is concerned, we have considered various measurement criterions like false positive, missed nuclei, under and over segmentation and correct detection. The number of false positive nuclei in the over-segmented regions is defined as false positive. The number of cell nuclei that resides among the foreground and under-segmented regions is defined as missed nuclei number. The number of cell

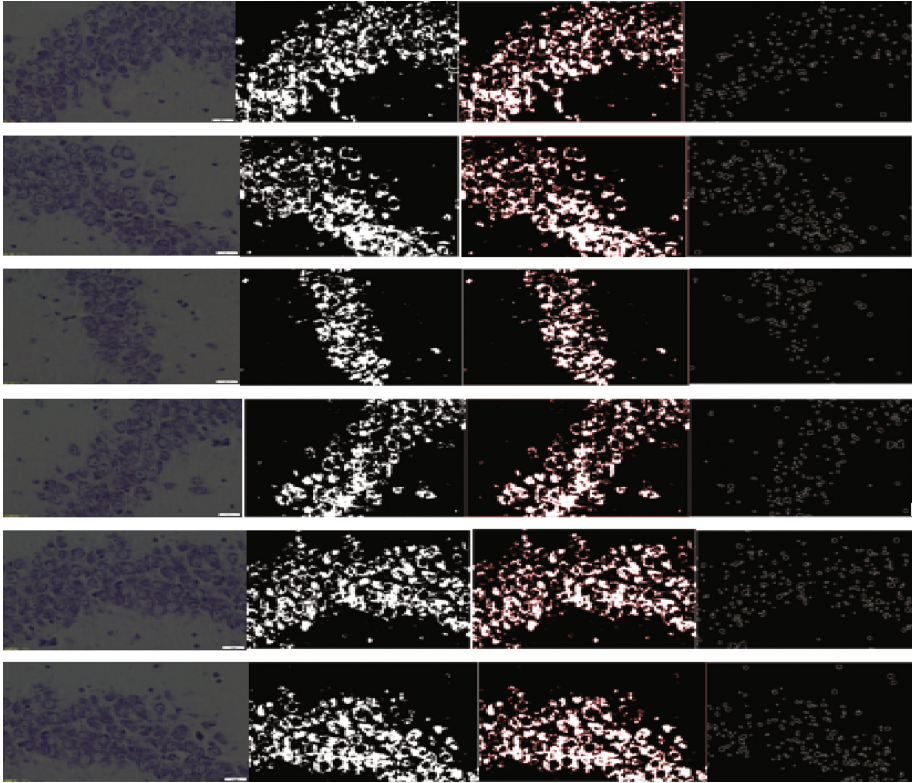


Fig. 4. First column depicts original cell images, Second column shows cell body separated from the background, Third column represents level-set framework output and Fourth column depicts the final nuclei segmentation.

nuclei detected truly is defined as correct detection. We have evaluated this value and represented in tabular for our proposed method as well as for existing in-class techniques in Table 1. It is evident from the table that our proposed approach maintains the balance amongst under and over segmentation besides providing significant correct detections and minimal missed ones as compared to conventional techniques.

4 Conclusion

Cell nuclei quantification plays an important role for strategic therapy. The objective approach is utter significance when it comes to the reduction of the subjective orientation of quantification. In our work, we have proposed novel cell nuclei segmentation technique for histopathological images. We have employed the Li's framework by involving various mean and variance Gaussian distributions into the level-set algorithm. This prevents the boundary leakage issue in case if there are any low contrast borders in our cell images. To address the issue of sensitivity, we have employed seeded region

growing which is used to initialize the level-set procedure. Experimental findings on large datasets also support our claim of being the best technique so as far as segmentation of heavily clustered cell nuclei is concerned. The similar techniques in this class need a lot of parameters to be calculated empirically. This work can be extended by considering and addressing the changes in the shape due to staining to achieve better segmentation outcomes.

Competing Interests

The authors announce that there is no conflict of interests when it comes to publishing the paper.

Acknowledgments. This work is supported by Department of Science and Technology (DST) and Department of Technology (DoT), India. Besides, these Hippocampal CA3 region EMF exposed and sham exposed images are courtesy of Department of Life Science Engineering, Manipal University, Manipal, Karnataka, India.

References

1. Moffat, J.G., et al.: Phenotypic screening in cancer drug discovery - past, present and future. *Nat. Rev. Drug Discov.* **13**(8), 588–602 (2014)
2. Celli, J.P., et al.: An imaging-based platform for high-content, quantitative evaluation of therapeutic response in 3D tumour models. *Sci. Rep.* **4**, 3751 (2014)
3. Singh, S., et al.: Increasing the content of high-content screening: an overview. *J. Biomol. Screen.* **19**(5), 640–650 (2014)
4. Uhlmann, V., et al.: CP-CHARM: segmentation-free image classification made accessible. *BMC Bioinform.* **17**(1), 51 (2016)
5. Martin, H.L., et al.: High-content, high-throughput screening for the identification of cytotoxic compounds based on cell morphology and cell proliferation markers. *PLoS ONE* **9**(2), e88338 (2014)
6. Hoggart, S.G.: *Mathematics of Digital Images: Creation, Compression, Restoration, Recognition.* Cambridge University Press, New York (2006)
7. Carpenter, A.E., et al.: Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**(10), R100 (2006)
8. Collins, T.J.: ImageJ for microscopy. *Biotechniques* **43**, 25–30 (2007)
9. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*, 3rd edn. Prentice-Hall, Upper Saddle River (2006)
10. Wang, X., et al.: Cell segmentation and tracking using texture-adaptive snakes. In: *Proceedings of 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2007, Metro Washington DC, USA*, pp. 101–104. IEEE (2007)
11. Perrin, G., et al.: A marked point process model for tree crown extraction in plantations. In: *Proceedings of 12th IEEE International Conference on Image Processing, ICIP 2005, Genova, Italy*, pp. 709–712. IEEE (2005)
12. Thevenaz, P., Unser, M.: The snakuscul. In: *Proceedings of 13th IEEE International Conference on Image Processing, ICIP 2006, Atlanta, GA, USA*, pp. 1633–1636. IEEE (2006)
13. Thevenaz, P., et al.: The ovuscul. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(2), 382–393 (2011)

14. Olivo-Marin, J.-C.: Extraction of spots in biological images using multiscale products. *Pattern Recognit.* **35**(9), 1989–1996 (2002)
15. Kimme, C., et al.: Finding circles by an array of accumulators. *Commun. ACM* **18**(2), 120–122 (1975)
16. Kass, M., et al.: Snakes - active contour models. *Int. J. Comput. Vis.* **1**(4), 321–331 (1987)
17. Bergeest, J.-P., Rohr, K.: Efficient globally optimal segmentation of cells in fluorescence microscopy images using level sets and convex energy functionals. *Med. Image Anal.* **16**(7), 1436–1444 (2012)
18. Jung, C., et al.: Unsupervised segmentation of overlapped nuclei using Bayesian classification. *IEEE Trans. Biomed. Eng.* **57**(12), 2825–2832 (2010)
19. Qi, X., et al.: Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set. *IEEE Trans. Biomed. Eng.* **59**(3), 754–765 (2012)
20. Vincent, L., Soille, P.: Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.* **13**(6), 583–598 (1991)
21. Yang, X., Li, H., Zhou, X.: Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy. *IEEE Trans. Circ. Syst. I Regul. Pap.* **53**(11), 2405–2414 (2006)
22. Jung, C., Kim, C.: Segmenting clustered nuclei using H-minima transform-based marker extraction and contour parameterization. *IEEE Trans. Biomed. Eng.* **57**(10), 2600–2604 (2010)
23. Parvin, B., et al.: Iterative voting for inference of structural saliency and characterization of subcellular events. *IEEE Trans. Image Process.* **16**(3), 615–623 (2007)
24. Al-Kofahi, Y., et al.: Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Trans. Biomed. Eng.* **57**(4), 841–852 (2010)
25. Ridler, T.W., Calvard, S.: Picture thresholding using an iterative selection method. *IEEE Trans. Syst. Man Cybern.* **8**(8), 630–632 (1978)
26. Wu, K., Gauthier, D., Levine, M.D.: Live cell image segmentation. *IEEE Trans. Biomed. Eng.* **42**(1), 1–12 (1995)
27. Soille, P.: *Morphological Image Analysis: Principles and Applications*, 1st edn., pp. 170–171. Springer, Heidelberg (1999)
28. FARSIGHT. http://www.farsight-toolkit.org/wiki/Main_Page