Neurocomputing 149 (2015) 39–47

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Accurate segmentation of touching cells in multi-channel microscopy images with geodesic distance based clustering

Xu Chen [a,b], Yanqiao Zhu [a], Fuhai Li [c], Ze-Yi Zheng [d], Eric C. Chang [d], Jinwen Ma [a,*], Stephen T.C. Wong [c]

[a] Department of Information Science, School of Mathematical Sciences and LMAM, Peking University, Beijing 100871, China
[b] Unit of 91635, PLA, China
[c] NCI Center for Modeling Cancer Development, Department of Systems Medicine and Bioengineering, The Methodist Hospital Research Institute, Weill Cornell Medical College of Cornell University, Houston, TX 77030, USA
[d] Department of Molecular and Cellular Biology, The Breast Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

ARTICLE INFO

Article history:
Received 1 July 2013
Received in revised form
22 January 2014
Accepted 26 January 2014
Available online 1 August 2014

Keywords:
Cell segmentation
Color information
Riemannian metric
Clustering analysis
Quantitative evaluation

ABSTRACT

Multi-channel microscopy images have been widely used for drug and target discovery in biomedical studies by investigating morphological changes of individual cells. However, it is still challenging to segment densely touching individual cells in such images accurately and automatically. Herein, we propose a geodesic distance based clustering approach to efficiently segmenting densely touching cells in multi-channel microscopy images. Specifically, an adaptive learning scheme is introduced to iteratively adjust the clustering centers which can significantly improve the segmentation accuracy of cell boundaries. Moreover, a novel seed selection procedure based on nuclei segmentation is suggested to determine the true number of cells in an image. To validate this proposed method, we applied it to segment the touching Madin-Darby Canine Kidney (MDCK) epithelial cells in multi-channel images for measuring the distinct N-Ras protein expression patterns inside individual cells. The experimental results demonstrated its advantages on accurately segmenting massive touching cells, as well as the robustness to the low signal-to-noise ratio and varying intensity contrasts in multi-channel microscopy images. Moreover, the quantitative comparison showed its superiority over the typical existing cell segmentation methods.

1. Introduction

High content screening (HCS) has been widely used for discovering novel drugs and targets by investigating the morphological changes of interested proteins inside individual cells [1–3]. Along with the advance of automated image acquisition equipments, large scale of cell images with multiple fluorescent markers is being generated in experiments. Whereas, great challenges have been posed on the automated quantification of individual cell morphology due to their complex appearances, uneven intensity, low signal-to-noise ratio (SNR), and cell touching [4–7]. Automated detection and accurate segmentation of touching cells are crucial for biological studies such as cell morphological analysis, cell tracking and cell phase identification [8–10]. However, it remains being an open problem due to the aforementioned challenges.

A series of approaches of automated cell segmentation in different kinds of images have been reported [11,12]. In general, the widely used segmentation approaches, for example, watershed method [9,13] and deformable models like the snake model and level set methods [14,15], are sensitive to initializations (e.g., seeds selection), and intensity variation inside cells. Specifically, the watershed method is often involved in the over-segmentation problem, and also sensitive to the intensity noise. Active contour or level set based deformable models are very sensitive to boundary initializations. In addition, the computational cost is heavy for these methods. Therefore, they are unsuitable for processing a large scale image dataset in high content screening studies. Clustering methods and statistical approaches have also been employed for cell segmentation [16–18]. These methods are efficient but usually provide incomplete segmentations because they operate on the single pixel level and neglect the fact that cells are continuous regions. Especially, these methods have limited ability in delineating boundaries of touching cells in cell membrane segmentation. Moreover, large intensity variation on cell membrane and inside cells has made it a much more challenging

* Corresponding author.
E-mail address: jwma@math.pku.edu.cn (J. Ma).

http://dx.doi.org/10.1016/j.neucom.2014.01.061
0925-2312/

problem to find the dividing boundaries of cells, which usually lead to biased segmentation results.

In order to segment touching cells more accurately, several new techniques have been investigated in recent years, including certain adaptive learning approaches to separating overlapped nuclei or cells. In fact, Jung et al. [19] utilized an unsupervised Bayesian classification scheme based on parametric EM algorithm for separating overlapped nuclei. Moreover, the single-path voting followed by a mean-shift clustering process was employed to locate the cell centers by Qi et al. [20]. Furthermore, the graph-cuts-based algorithm was successfully applied by Yousef et al. [21] to develop a nuclei segmentation method as well as a powerful toolkit software named "FARSIGHT". On the other hand, some classical conventional methods have also been carried forward, such as the adaptive thresholding watershed algorithm [9] and the adaptive active physical deformable model [22]. To solve the cell segmentation problem more effectively via the advanced mathematical model, Jones et al. [23] proposed a new distance metric defined on image manifolds which combined image gradients and inter-pixel distance together. This metric demonstrated its effectiveness, but the segmentation results strongly depended on the initial selection of seeds. Although the centers of nuclei could be adopted as the seeds in CellProfiler [24], the segmentation results became worse as cell nuclei are touched together. Besides, CellProfiler accomplished the pixel clustering after only one time of assignment, the boundaries were obtained without optimization. Along the direction of new distance metric, we recently proposed a novel clustering method with the geodesic distance for gray scale images in [25] with little a priori knowledge. Although it is robust to the initialization of seeds, the determination of number of clusters is still a bottleneck in improving the segmentation performance further. Over-segmentation occurs when the number of initial seeds is more than the true number of cells in an image, which is a well known inherent disadvantage of *K*-means clustering methods.

Aiming for accurate segmentation of the touching Madin-Darby Canine Kidney (MDCK) epithelial cells in multi-channel (color) images, in this paper, we propose a nuclei segmentation based *K*-means clustering method using the geodesic distance. The main contribution of this paper is that an adaptive learning scheme is established to adjust the clustering centers with the geodesic distance for improving the segmentation accuracy. Technically, color information is taken into consideration to select the number of reasonable initial seeds for clusters through a well-designed nuclei segmentation procedure. As a result, the over-segmentation problem is effectively overcome, more accurate dividing boundaries are delineated, and the convergence of the improved *K*-means clustering method is also speeded up.

The rest of this paper is organized as follows. We briefly introduce the geodesic distance based clustering method in Section 2. Section 3 presents the details of our adaptive cell segmentation procedure. The experimental results on MDCK images are demonstrated in Section 4. Finally, a brief conclusion is provided in Section 5.

2. Geodesic distance based clustering method

2.1. Riemannian metric

We firstly introduce the assumed Riemannian metric defined at each pixel in an image $\mathcal{I}$ [23]. Specifically, $G(\cdot)$ is defined as

$$G(\cdot)=\frac{\nabla g(\cdot)\nabla g^T(\cdot)+\lambda \mathrm{E}}{1+\lambda}, \tag{1}$$

where E is the 2×2 identity matrix, $\lambda \geq 0$ is a regularization parameter and ∇ is a gradient operator. The function g is introduced to reduce the effect of noise by weighted averaging the pixels within a certain neighborhood. The infinitesimal distance at each pixel is then calculated by

$$\| dx \|^2_{G(\cdot)} \equiv dx^T G(\cdot)\, dx = \frac{(dx^T \nabla g(\cdot))^2 + \lambda\, dx^T\, dx}{1+\lambda}. \tag{2}$$

The item $(dx^T \nabla g(\cdot))^2$ increases greatly along the direction of the largest gradient in the image and captures the boundary information essentially. The operator $G(\cdot)$ becomes Euclidean as λ increases to infinity, while $\lambda = 0$, the operator considers only gradient information. More details can be found in [23].

2.2. Clustering scheme

Under the assumed Riemannian metric, the geodesic distance from any pixel to another can be calculated along the shortest path using Dijkstra's algorithm [23]. In general, the distance calculation propagates from a smaller neighborhood of the starting pixel to a larger one, and finally reaches the object pixel.

Based on the geodesic distance introduced above, we propose an improved version of *K*-means clustering algorithm to delineate cell boundaries. It can be summarized as the following four main steps:

1. Select initial seeds.
2. Compute pixels' geodesic distances to each seed according to the Riemannian metric and assign each pixel to the cell with the nearest geodesic distance.
3. Update each seed to the center of its current enclosing cell.
4. Repeat step 2 and step 3 until the algorithm has converged or the maximum number of iterations is reached.

The clustering centers will gradually converge to cells' spatial centers after a small number of iterations as long as each cell has one initial seed in it. Thus, the algorithm is robust to the initial seeds selection. The inherent characteristics of the Riemannian metric ensure that the algorithm can accurately delineate touching cell boundaries in images with high noise, poor contrast, and large intensity variations. After the algorithm has converged, cell boundaries are easy to draw according to the final clustering result.

3. Cell segmentation procedure

In the implementation of the improved clustering method for cell segmentation, it is important to determine the number of clusters and to set the initial seeds of the clusters. Clearly, the number of clusters should correspond to that of cells in the image and it is reasonable and efficient to set the initial seeds as the centers of the nuclei. In order to solve these problems, we can execute a nuclei segmentation procedure independently. In fact, accurate segmentation of the nuclei regions is usually a very important prerequisite of many cell segmentation methods to get some critical parameters, and even nuclei segmentation itself has become an independent research direction. Here, we need to determine the number of cells in an image as well as their locations. At this stage, it is not necessary to delineate the accurate boundaries of nuclei fields. However, a fixed global thresholding scheme seems to be unsatisfactory because of the inhomogeneity of intensity distribution. In fact, a lower thresholding scheme would fail in segmenting touching nuclei fields, while a higher one erases nuclei regions with low intensity.

To overcome this difficulty, we utilize a simple adaptive nuclei segmentation procedure based on the mathematical morphology and a priori knowledge. Actually, we can easily distinguish nuclei fields in an color cell image because nucleoli and nuclear heterochromatin are stained clearly with basic dyes. So the blue channels representing the nuclei fields in our MDCK epithelial cell images can be extracted and used as input images. We firstly select the pixels around the boundary lines by their statistical characteristic of local intensity variation, and break the touching nuclei fields by setting the gray value of these pixels to 0. Then, an adaptive local thresholding scheme can be implemented to get the binary image. With the help of a morphological method, we can get the final binary image where all the nuclei are separated from each other and thus take the centroids of the white connected regions as the initial clustering centers for our proposed clustering method. Specifically, we present the nuclei segmentation procedure in the following three subsections.

3.1. Boundary pixel selection

We begin to consider an input image as a topographic surface where the gray value of each pixel in the blue channel is regarded as its altitude. Then, the mountain peaks are composed of nuclei pixels, while the boundary pixels make up the valleys, as illustrated in Fig. 1. For any pixel p, let i be its row number and j be its column number in the image, we compare its gray value with those of its 5×5 neighborhood pixels, count the number of pixels which have greater gray values than that of p, and denote it by N_g. On the other hand, we also count the number of pixels with the gray values less than that of p and denote it by N_s. We determine the approximate location of p by calculating L_p as follows:

$$L_p = N_g - N_s = \sum_{s=i-2}^{i+2}\sum_{t=j-2}^{j+2} \mathrm{sgn}(g_{s,t} - g_{i,j}) \tag{3}$$

where $\mathrm{sgn}(\cdot)$ is the sign function, $g_{s,t}$ is the gray value of the neighbor pixel with line number s and row number t, $g_{i,j}$ is the gray value of p. If p is a pixel in a valley, the most neighborhood pixels around it may have a greater gray value than that of p so that N_g will be much greater than N_s, for instance, $L_p > 3$, we thus consider p as a boundary pixel. Otherwise, if p is on mountain slope, N_g and N_s would be almost the same, it can hardly be around the boundary lines. A calculation instance is illustrated in Table 1.

Fig. 2(a) shows a binary image derived from a blue channel image by using its mean gray value as the global threshold, while Fig. 2(b) demonstrates the boundary pixels we extract from the original blue channel image. For the convenience of further processing, we break the touching nuclei fields by setting the gray value of these selected pixels to 0. The benefits of doing this will be exhibited in the next subsection.

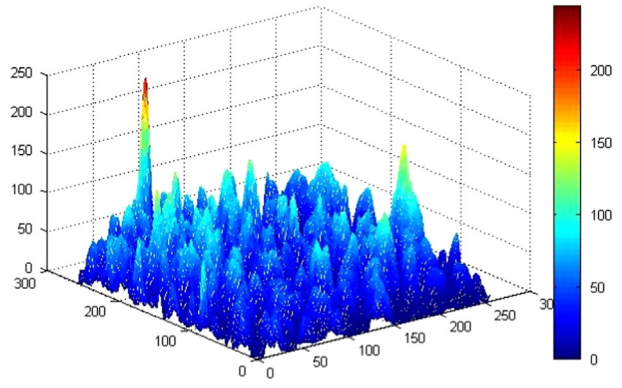


Fig. 1. The topographic surface of a cell image where the gray value of each pixel in the blue channel is regarded as its altitude. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

3.2. Adaptive local thresholding operation

We further adopt an adaptive local thresholding operation after breaking the touching nuclei fields to get a binary image in which white area represents nuclei fields. The threshold is ascertained by combining the global mean gray value M_g and the local mean value M_l, where M_g is defined as the average over all pixels in an image, while M_l is defined as the average over a predefined neighborhood of a pixel, p. If the gray value of p, denoted by g_p, satisfies $g_p > 1.2\cdot M_g$ or $g_p > \max\{0.6\cdot M_g, 2\cdot M_l\}$, we set the corresponding pixel of the binary image to 1; otherwise, we set it to 0. Mathematically, the thresholding function can be given as follows:

$$f(p) = \begin{cases} 1 & \text{if } g_p > 1.2\cdot M_g \text{ or } g_p > \max\{0.6\cdot M_g, 2\cdot M_l\}; \\ 0 & \text{else.} \end{cases} \tag{4}$$

The reason of the constraint $g_p > 0.6\cdot M_g$ is for suppressing the false positive points in poor intensity area caused by random fluctuation. For example, if $g_P = 5$, $M_l = 2$ and $M_g = 24$, $f(p)$ would be 1 without this constraint. All these parameters can be selected by experience, and will not be changed in the whole processing procedure for all the images. Fig. 2(c) illustrates a binary nuclei image after the adaptive local thresholding operation on the broken touching nuclei fields. In comparison with Fig. 2(a), we find the fact that the boundaries of adjacent nuclei fields are very fragile after our breaking operation, which thereby reminds us to utilize the mathematical morphologic opening. In mathematical morphology, opening is the dilation of the erosion of an image by a structuring element. So the broken boundaries can easily be split through an morphologic opening [26]. Fig. 2(d) shows the result of the morphologic opening.

3.3. Seed selection scheme

With the above preparations, we can present the seed selection scheme. It may be easy to set the initial seeds as the centers of these separated white areas corresponding to the nuclei fields, but there may emerge a new problem from our previous operations that an actual nuclei area might be broken up into some separate parts. So, we must merge them together, otherwise, the number of initial seeds will be greater than the number of cells in the image. In fact, a very simple merging method can solve this problem. We compute the geometrical center of each connected region and merge the centers that are close enough. The final centers of regions provide the selected seeds, which certainly lead to the number of cells in the image as well as their locations. Although there have been already many adaptive model selection clustering methods (e.g., [27–30]) that can determine the number of clusters automatically from the dataset, they are time-consuming and often lead to a wrong result when the actual clusters are overlapped or touched. Here, we implement the geodesic distance based K-means clustering method with this seed selection scheme on massive images of different types. The selection results on four specific images with very different characteristics are illustrated in Fig. 3.

In summary, we establish a geodesic distance based K-means clustering method for segmenting densely touching cells in a multi-channel microscopy image. Moreover, an efficient nuclei segmentation procedure using the color information of the image is suggested to select the number of reasonable initial seeds for clusters or cells.

Table 1
(a) A typical pixel around boundary line with gray value "4"; (b) The gray value comparison matrix of its 5×5 neighborhood pixels. Here, "-1" means that the gray value is less than that of the central pixel, while "1" means "greater" and "0" means "equal to". The deviation $N_g - N_s = 15$ can be calculated by summing over all the elements in the comparison matrix.

(a)				
24	23	21	21	21
11	10	8	10	3
4	1	*4*	2	13
3	5	9	13	15
10	11	13	14	15
(b)				
1	1	1	1	1
1	1	1	1	−1
0	−1	*0*	−1	1
−1	1	1	1	1
1	1	1	1	1

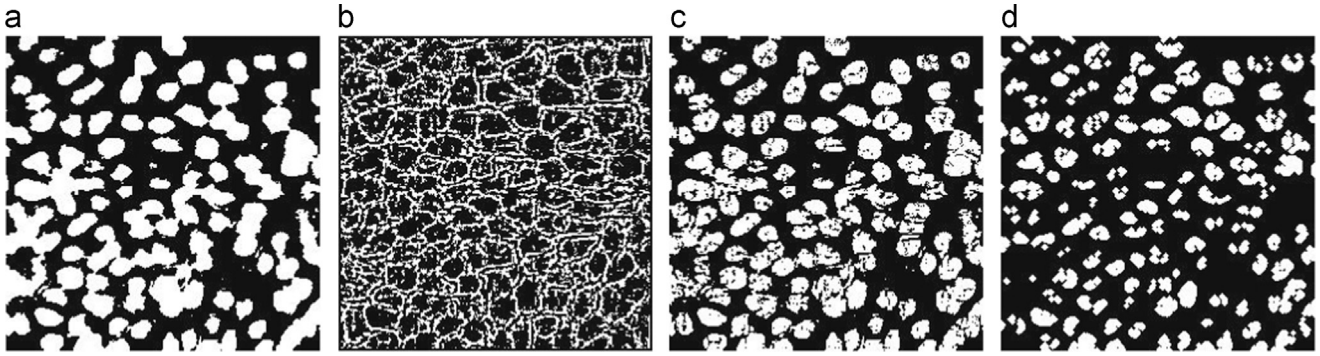


Fig. 2. (a) A binary image derived from a blue channel image by using its mean gray value as the global threshold; (b) The boundary pixels that satisfy the selection condition; (c) A binary nuclei image after having wiped the boundary pixels out and implemented the adaptive local thresholding operation; (d) The final segmented nuclei image after the morphological open operation.

4. Experimental results

We implemented the proposed geodesic distance based *K*-means clustering method on the MDCK epithelial cell images from dog kidney that were used to study the effects of drug compounds on regulating Ras protein levels, which were determined by measuring the fluorescence intensity of a GFP-tagged N-Ras reporter. Thus, the accurate segmentation of cells is crucial for measuring the GFP-tagged Ras signal changes between the control and drug-treated cells. But the obtained cell images have high noise, varying densities and complex morphologies, as illustrated in Fig. 3.

The images were firstly down-sampled to 256×256 from 512×512. Function g in the distance definition was taken as the Gaussian low pass filter in 3×3 neighborhood and λ was set to 2.4 in our experiments. In our seed selection procedure, we calculated M_l over a 15×15 neighborhood of each pixel. A flat diamond-shaped structuring element with a radius of 3 pixels was used in the morphologic opening operation. Some of the typical segmentation results of our proposed method on four representative MDCK images with very different characteristics are illustrated in Figs. 4 and 5, being compared with the segmentation results of the manual drawing by expert, our previous method [25], CellProfiler [24] and the adaptive thresholding watershed algorithm [9].

It can be found from Figs. 4 and 5 that the segmentation results of our proposed method is very similar to those of the manual drawing by expert and much better than those of the other methods. Clearly, CellProfiler led to certain "narrow band" segmentation error, while the previous method fell into over-segmented problem. Moreover, the watershed method could not detect the correct boundaries of the cells.

We further try to evaluate and compare those cell segmentation results quantificationally. Since there is neither universal evaluation method nor standard database in this respect so far, we use three reasonable evaluation indexes together to evaluate the cell segmentation results more precisely. A widely used quantitative evaluation method is to compare the results with manual segmentations quantificationally on different aspects. Actually, we can consider the accuracy of cell number, the average ratio of overlapped areas of cells as well as the average excursion of boundary pixels of cells, which can be defined and calculated as follows.

The accuracy of cell number, denoted by R_n, can be defined and calculated by

$$R_n = 1 - \frac{|N_a - N_m|}{N_m} \tag{5}$$

where N_a is the cell number obtained by our proposed clustering segmentation method, while N_m is that of the manual segmentation. As for the average ratio of overlapped areas of cells in the manual and clustering segmentation images, we can compute it as follows. For each segmented cell P_{mi} in the manual segmented image, we calculate its geometric center (C_{ix}, C_{iy}) by taking an average on all the row coordinates and column coordinates of pixels in P_{mi}, respectively. Then, we can find the corresponding segmented cell P_{ai} which encloses (C_{ix}, C_{iy}) in the clustering segmented image, calculate the overlapped area between P_{mi} and P_{ai}, and get the overlapping ratio by dividing it by the area of P_{mi}. Averaging on all possible P_{mi} in the manual segmented image,

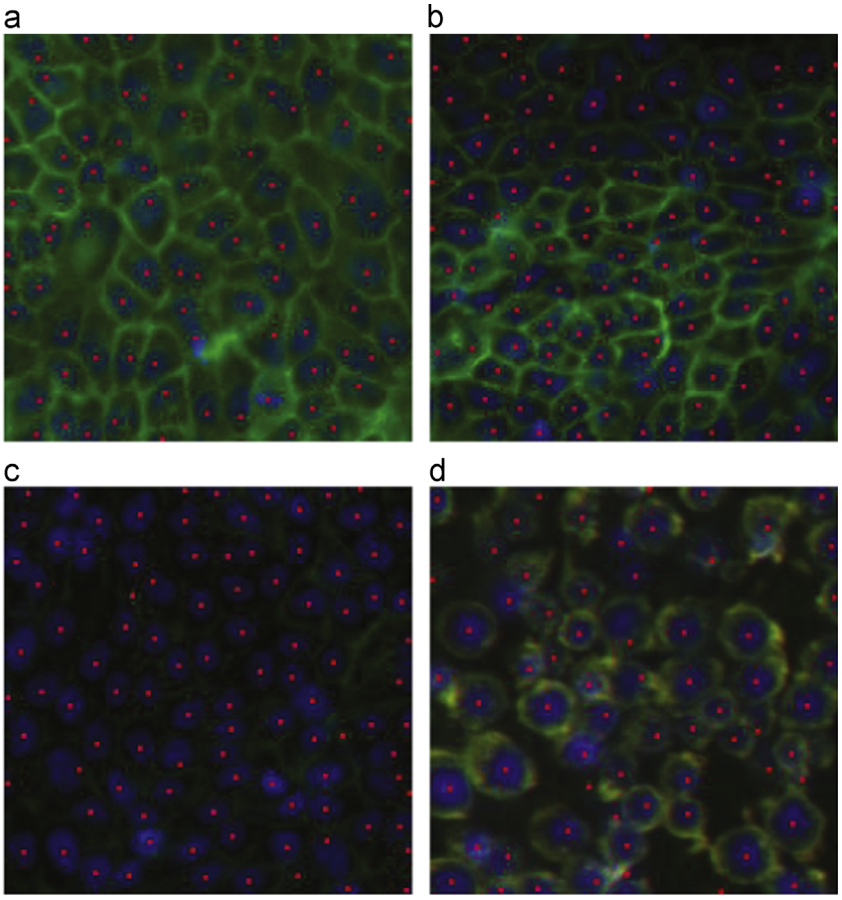


Fig. 3. The seeds (white points in black and white or red points in color) selection results in four typical MDCK images with distinct characteristics: (a) strong RAS signal (white rings in black and white or green rings in color); (b) uneven RAS signal; (c) low RAS signal; and (d) noise and varying RAS signal. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

we finally get the average ratio as follows:

$$R_s = \frac{1}{N_m}\sum_{i=1}^{N_m}\frac{S(P_{mi}\cap P_{ai})}{S(P_{mi})}, \tag{6}$$

where $S(A)$ denotes the area of a field A.

We now turn to the average excursion of boundary pixels of cells. The offset of boundary pixels is the distance between the automatic segmentation results and the manual segmentation results in the orthogonal or normal direction of local boundary [23]. We consider a statistic of the offset as follows:

$$P_{b,k} = \frac{\sum_{i=0}^{k} N_{bi}}{N_b}, \tag{7}$$

where $P_{b,k}$ represents the proportion of boundaries pixels with an offset no more than a predefined positive integer k (along the normal direction), while N_{bi} is the number of boundary pixels with an offset of i, N_b is the total number of boundary pixels delineated by the manual segmentation. Since $P_{b,k}$ can benefit from over-segmentation and result in a ridiculous result, we construct another statistic, $R_{b,k}$, by amending $P_{b,k}$ with a regularization term, which can be formulated by

$$R_{b,k} = \delta \times P_{b,k} + (1-\delta)\times\frac{1}{1+M_k}, \tag{8}$$

where M_k is the average number of boundary pixels in an $l\times l$ neighborhood of each cell center with $l=[k/2]$, where $[x]$ denotes the largest integer less than x, δ is a regularization parameter satisfies $0\le\delta\le 1$. Since M_k is a meaningful measure of dispersion of the boundary pixels away from their centers, the second term of Eq. (8) is a penalty term for the over-segmentation phenomenon because over-segmentation introduces redundant boundaries such that there are more boundary pixels being close to a cell center. So, by combining the two terms together, $R_{b,k}$ gives a more reasonable index to evaluate the segmentation result of the cell image. A feasible selection of parameter δ is $\delta_k = 1/(1+k)$. In this way, the accuracy of boundaries and credibility of $P_{b,k}$ decreases as k increases, while the excursion degree of boundaries away from cell centers becomes more important. Clearly, this index is reasonable only when k is less than the average cell diameter.

By comprehensive analysis, we can find that these three indices are reasonable and valuable for evaluating the cell segmentation results on an MCDK or general image assuming that the manual segmentation results by expert are correct and reliable. In general, an expert can draw the correct cell segmentation results as we expect. However, in certain cases, it is very difficult even for an expert to distinguish whether there is a cell or not in some blurred area. So, the reliability of the manual cell segmentation is still questionable, but we neglect this problem here. On the other hand, these three indices evaluate the cell segmentation results on different aspects and none of them is so perfect. R_n just focuses on cell number, while R_s cares about only the overlapped areas between the clustered and manual segmented cells. As for $R_{b,k}$, the reasonable selection of δ is controvertible and needs to be further investigated. For more reasonable and effective evaluation on the cell segmentation results, we use all the three indices together in the following analysis and comparison.

Most of our MDCK images are quite similar to Fig. 3(a) and (b). In fact, our proposed method has a very good and consistent performance on these two kinds of MDCK images. To make the validation more convincing, we implemented the proposed

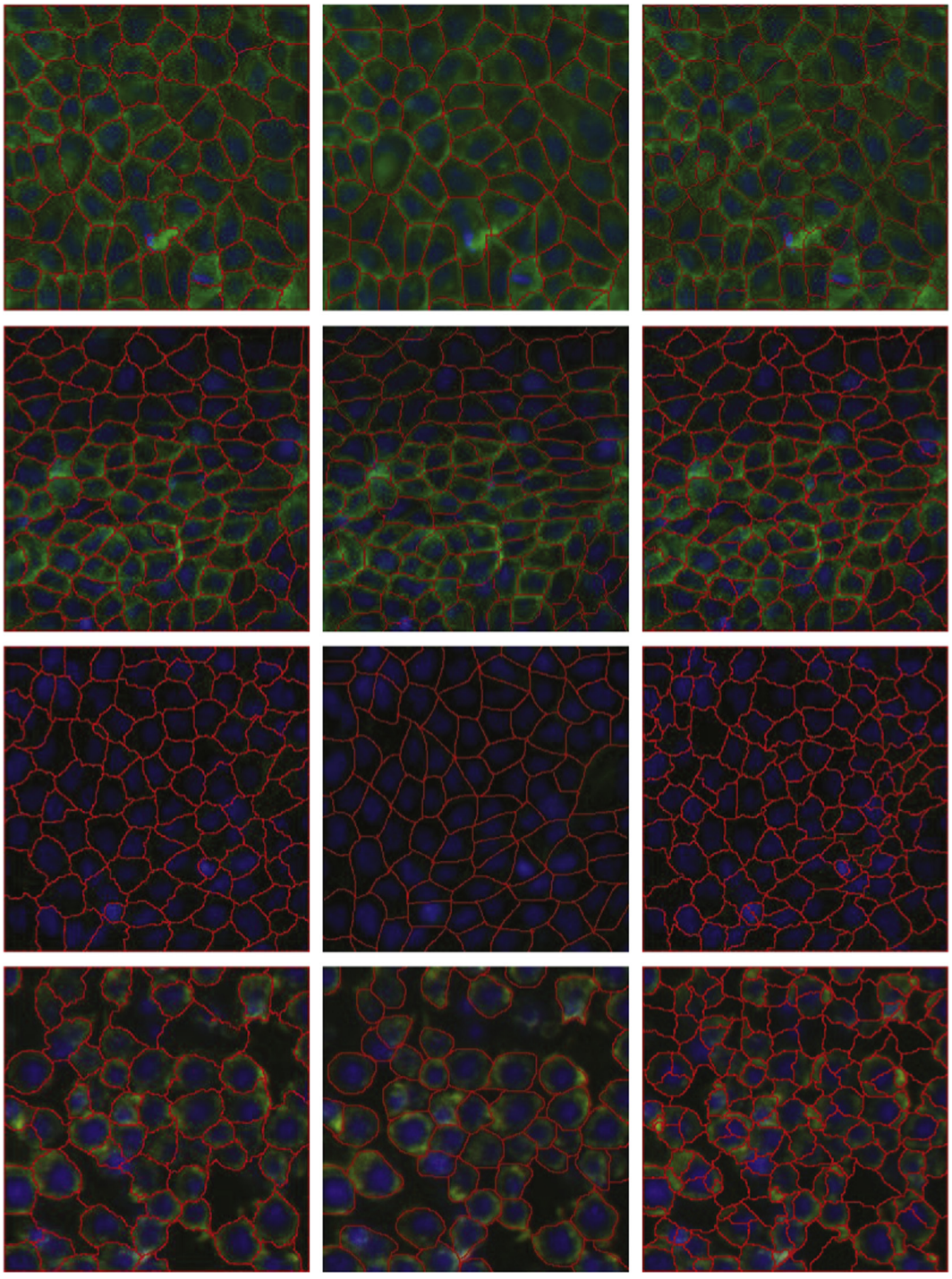


Fig. 4. The segmentation results of the proposed method (the left column), the manual drawing by expert (the middle column) and our previous method (the right column) on four representative images.

method to the four kinds of MDCK images shown in Fig. 3, denoted by (a)–(d). The accuracy of cell number for each of the four images is illustrated in Table 2, and Table 3 shows the average ratios of overlapped areas of cells for the four images in comparison with our previous method. We do not provide the corresponding indexes of R_n and R_s for K-means and CellProfiler here because finely divided segmentations and fragmentary boundaries make them meaningless.

It can be found from Table 2 that the segmentation result of our proposed method is quite good in images (a)–(c), but not very good in image (d). Actually, in image (d), apart from its high noise and blur, there are also many blue nuclei fields which are not surrounded by cell plasma, and manual segmentation excludes them from independent cells. According to the average ratios of overlapped areas of cells given in Table 3, it can be further found that even for images (c) and (d), our proposed algorithm can get a near or over 80% average ratio on overlapped areas of cells. Moreover, it performs much better than our previous method. Obviously, it is just the over-segmentation behaviour that makes our previous method to not perform so good.

Fig. 6 illustrates the sketches of the average excursions of boundary pixels of cells within 7 pixels of offset by taking an

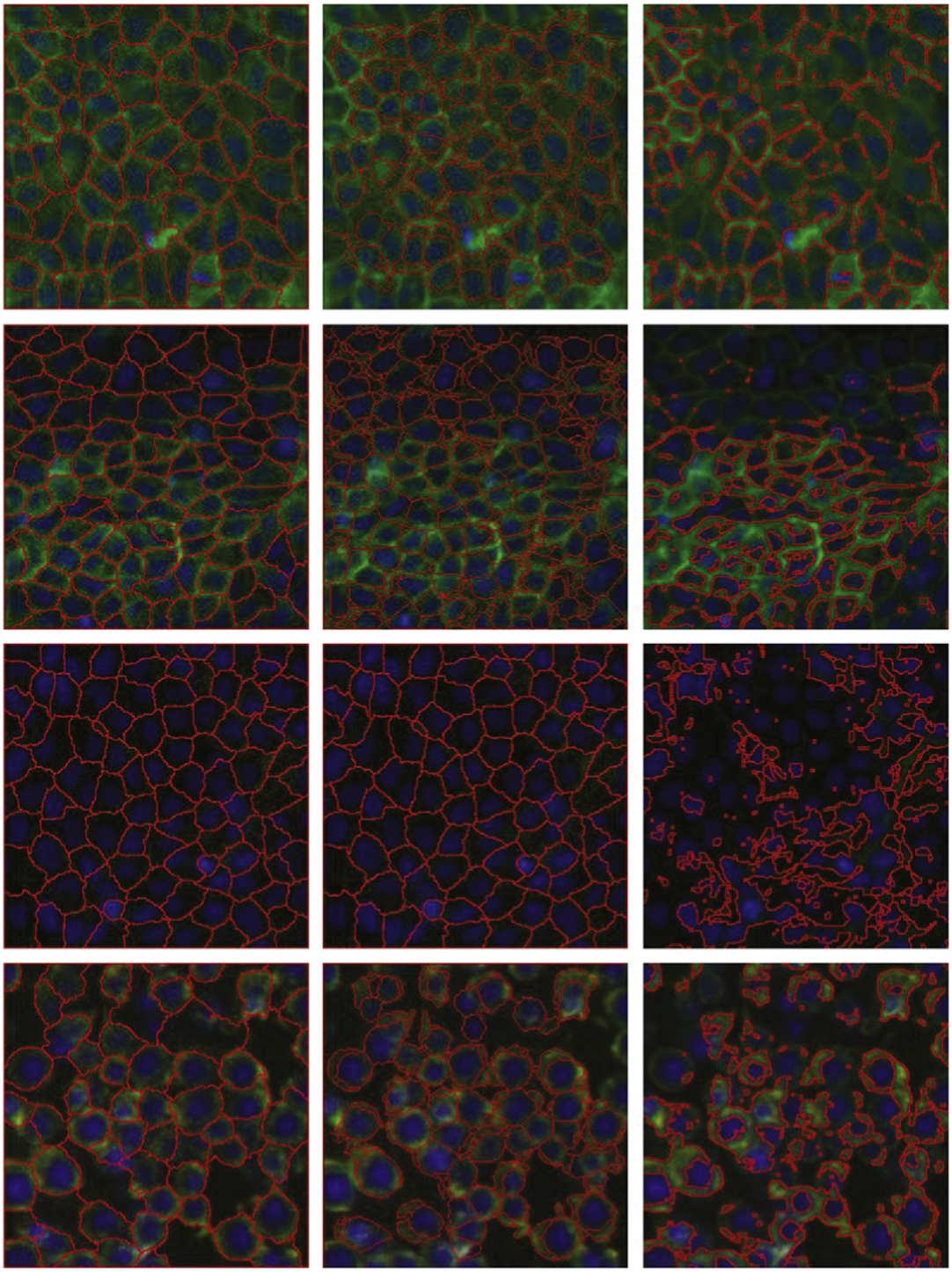


Fig. 5. The segmentation results of the proposed method (the left column), CellProfiler (the middle column) and the watershed algorithm (the right column) on four representative images.

Table 2
The accuracies of cell number for the four representative images.

Image	N_m	N_a	R_n
(a)	86	80	0.93
(b)	120	112	0.93
(c)	93	100	0.92
(d)	55	66	0.80

Table 3
The average ratios of overlapped areas of cells for the four representative images in comparison with our previous clustering method in [25].

Image	N_m	Previous method	New method
(a)	86	0.7832	0.8842
(b)	120	0.8560	0.8797
(c)	93	0.7120	0.7813
(d)	55	0.6272	0.8467

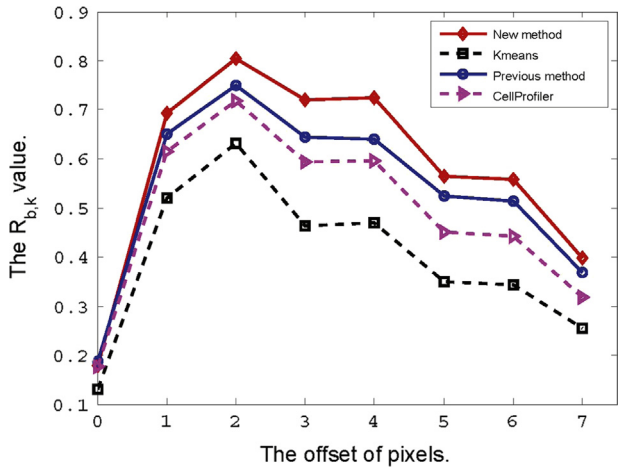


Fig. 6. The sketches of the average dispersions of boundary pixels of cells within 7 pixels of offset of the four cell segmentation methods.

average over 20 typical images in MDCK database for our proposed new method, our previous method, conventional *K*-means algorithm, and CellProfiler. Since the watershed algorithm did not perform well on cell segmentation of MDCK images, we do not consider it in this case. It can be clearly observed that our proposed method performs better than the other three methods. In fact, it can be observed from Fig. 4 that CellProfiler generated a lot of false boundaries and our previous method delineated some accurate boundaries but over-segmented. Actually, the (conventional) *K*-means algorithm led to many debris and fluctuated boundaries. Thus, this index confirms the advantages of our proposed method on accurate segmentation of cells.

Based on the comparison results on all the three indexes, we are confident that our proposed method can segment cells accurately and outperforms the other existing methods.

To sum up, our proposed method is a promising cell segmentation method. It inherits the advantages of our previous method such that it can delineate continuous and unique boundaries of cells even on images with low contrast, high noise, high intensity variation and varying cell density. Besides, it overcomes the over-segmentation problem and greatly improves the accuracy of cell number, and speeds up the convergence since the initial seeds have been already located properly and reasonable via a nuclei segmentation procedure. That is, our proposed method has been considerably improved to get such better segmentation results by utilizing more prior information including the color information.

5. Conclusion

We have proposed a geodesic distance based *K*-means clustering method that can well segment touching Madin-Darby Canine Kidney (MDCK) epithelial cells for discovering novel drugs regulating the spatial patterns of the N-Ras protein inside individual cells in multi-channel microscopy images. Actually, an adaptive learning scheme is introduced to adjust the clustering centers iteratively and efficiently. Moreover, a novel nuclei segmentation procedure is established to determine the number of reasonable initial seeds for the clusters in the *K*-means paradigm. Experimental results demonstrate that our proposed method is effective and efficient in delineating continuous touching boundaries between massive cells and robust to the low signal-to-noise ratio, varying intensity contrasts and high cell densities in microscopy images. In comparison with the common cell segmentation tool CellProfiler and the conventional *K*-means clustering method, our proposed method can generate more accurate and smooth cell boundaries between touching cells, even for very low contrast and high noise images. As compared with our previous clustering method, this new clustering method ascertains the number of cells in an image by utilizing the color information and thus overcome the over-segmentation problem.

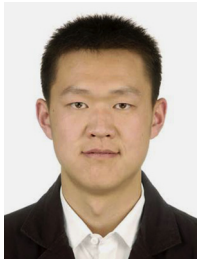
Acknowledgments

This work was supported by the Natural Science Foundation of China for Grants 61171138, NIH U54 CA149196-01, R01LM009161 and John S Dunn Research Foundation.

References

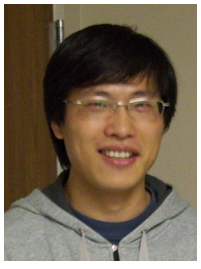
[1] F. Li, X. Zhou, J. Ma, S. Wong, Multiple nuclei tracking using integer programming for quantitative cancer cell cycle analysis, IEEE Trans. Med. Imaging 29 (1) (2010) 96–105.
[2] F. Li, X. Zhou, J. Ma, S. Wong, An automated feedback system with the hybrid model of scoring and classification for solving over-segmentation problems in RNAi high content screening, J. Microsc. 226 (2) (2007) 121–132.
[3] M. Wang, X. Zhou, F. Li, J. Huckins, R. King, S. Wong, Novel cell segmentation and online SVM for cell cycle phase identification in automated microscopy, Bioinformatics 24 (1) (2008) 94–101.
[4] P. Quan, Y. Cuirong, F. Yingle, C. Yu, Overlapped cell image segmentation based on distance transform, in: Proceedings of the 6th World Congress on Intelligent Control and Automation, 2006, pp. 9858–9861.
[5] E. Cheng, S. Challa, R. Chakravorty, Microscopic cell detection based on multiple cell image segmentations and fusion algorithms, in: 2nd International Conference on Biomedical Engineering and Informatics, Tianjin, China, 2009. BMEI'09. IEEE, 2009, pp. 1–7.
[6] O. Dzyubachyk, W. Niessen, E. Meijering, Advanced level-set based multiple-cell segmentation and tracking in time-lapse fluorescence microscopy images, in: Proceedings of The 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Paris, France, 2008 (ISBI'2008), IEEE, 2008, pp. 185–188.
[7] F. Li, X. Zhou, J. Zhu, J. Ma, X. Huang, S. Wong, High content image analysis for human h4 neuroglioma cells exposed to CuO nanoparticles, BMC Biotechnol. 7 (2007) 66.
[8] E. Ficarra, S. Cataldo, A. Acquaviva, E. Macii, Automated segmentation of cells with IHC membrane staining, IEEE Trans. Biomed. Eng. 58 (5) (2011) 1421–1429.
[9] X. Zhou, F. Li, J. Yan, S. Wong, A novel cell segmentation method and cell phase identification using Markov model, IEEE Trans. Inf. Technol. Biomed. 13 (2) (2009) 152–157.
[10] F. Li, X. Zhou, S. Wong, Optimal live cell tracking for cell cycle study using time-lapse fluorescent microscopy images, in: Machine Learning in Medical Imaging, Lecture Notes in Computer Science, vol. 6357, 2010, pp. 124–131.
[11] E. Meijering, Cell segmentation: 50 years down the road, IEEE Signal Process. Mag. 29 (5) (2012) 140–145.
[12] F. Li, Z. Yin, G. Jin, H. Zhao, S. Wong, Chapter 17: bioimage informatics for systems pharmacology, PLoS Comput. Biol. 9 (4) (2013) e1003043.
[13] E. Cheng, S. Challa, R. Chakravorty, Microscopic cell segmentation and dead cell detection based on CFSE and PI images by using distance and watershed transforms, in: Proceedings of International Conference on Digital Image Computing: Techniques and Applications (DICTA'09), IEEE, Melbourne, Australia, 2009, pp. 32–39.
[14] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, Int. J. Comput. Vis. 1 (4) (1988) 321–331.
[15] F. Li, X. Zhou, H. Zhao, S. Wong, Cell segmentation using front vector flow guided active contours, Med. Image Comput. Comput. Assist. Interv. 12 (Pt 2) (2009) 609–616.
[16] E. Bak, K. Najarian, J. Brockway, Efficient segmentation framework of cell images in noise environments, in: The 26th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society, San Francisco, CA, US, 2004 (IEMBS'04), vol. 1, IEEE, 2004, pp. 1802–1805.
[17] T. Muda, R. Salam, Blood cell image segmentation using hybrid k-means and median-cut algorithms, in: Proceedings of The 2011 IEEE International Conference on Control System, Computing and Engineering (ICCSCE), IEEE, 2011, pp. 237–243.
[18] G. Palacios, J. Beltran, Cell nuclei segmentation combining multiresolution analysis, clustering methods and colour spaces, in: Proceedings of the International Machine Vision and Image Processing Conference, 2007 (IMVIP'2007), IEEE, 2007, pp. 91–97.
[19] C. Jung, C. Kim, S.W. Chae, S. Oh, Unsupervised segmentation of overlapped nuclei using Bayesian classification, IEEE Trans. Biomed. Eng. 57 (12) (2010) 2825–2832.
[20] X. Qi, F. Xing, D. Foran, L. Yang, Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set, IEEE Trans. Biomed. Eng. 59 (3) (2012) 754–765.

[21] Y. Al-Kofahi, W. Lassoued, W. Lee, B. Roysam, Improved automatic detection and segmentation of cell nuclei in histopathology images, IEEE Trans. Biomed. Eng. 57 (4) (2010) 841–852.
[22] C. Plissiti, M.E. Nikou, Overlapping cell nuclei segmentation using a spatially adaptive active physical model, IEEE Trans. Image Process. 21 (11) (2012) 4568–4580.
[23] T. Jones, A. Carpenter, P. Golland, Voronoi-based segmentation of cells on image manifolds, in: Lecture Notes in Computer Sciences, vol. 3765, 2005, pp. 535–543.
[24] M. Lamprecht, D. Sabatini, A. Carpenter, et al., Cellprofiler: free, versatile software for automated biological image analysis, Biotechniques 42 (1) (2007) 71.
[25] X. Chen, Y. Zhu, F. Li, Z. Zheng, E. Chang, J. Ma, T. Wong, Stephen, A novel geodesic distance based clustering approach to delineating boundaries of touching cells, in: The Tenth International Symposium on Neural Networks, Dalian, China, (ISNN 2013), Part II, Lecture Notes in Computer Science, vol. 7952, IEEE, 2013, pp. 315–322.
[26] R.C. Gonzalez, R.E. Woods, Digital Image Processing, 3rd ed., Publishing House of Electronics Industry, Beijing, China, 2011.
[27] J. Ma, T. Wang, L. Xu, A gradient by harmony learning rule on gaussian mixture with automated model selection, Neurocomputing 56 (2004) 481–487.
[28] J. Ma, T. Wang, A cost-function approach to rival penalized competitive learning (RPCL), IEEE Trans. Syst. Man Cybern. Part B 36 (4) (2006) 722–737.
[29] J. Ma, J. Liu, The byy annealing learning algorithm for gaussian mixture with automated model selection, Pattern Recognit. 40 (4) (2007) 2029–2037.
[30] C. Fang, W. Jin, J. Ma, k'-means algorithms for clustering analysis with frequency sensitive discrepancy metrics, Pattern Recognit. Lett. 34 (2013) 580–586.



Xu Chen received his B.S. and M.S. degrees in Applied Mathematics from Peking University, in 2009 and 2013, respectively. He is now a Research Assistant of a government research institute. His main research interests include neural computation, statistical learning, and intelligent information processing.

Yanqiao Zhu received his B.S. degree in Mathematics and Applied Mathematics from Dalian University of Technology, in 2008 and Ph.D. degree in Applied Mathematics from Peking University, in 2013. His research interests include statistical learning and intelligent information processing.



Fuhai Li received his Ph.D. degree in Applied Mathematics from Peking University, in 2007 and is now an Instructor in Department of Systems Medicine and Bioengineering, Houston Methodist Hospital Research Institute. His research interests are biomedical image analysis, bioinformatics, and computational systems biology.

Ze-Yi Zeng received his B.Sc. degree in Biological Sciences and Technology from Zhejiang University (China), in 2001, and Ph.D. degree in Biomedical Sciences from the National University of Singapore (Singapore), in 2006. Since 2007, he is a Susan G. Komen Foundation Postdoctoral Fellow at the Lester & Sue Smith Breast Center of the Baylor College of Medicine in Houston, TX. His research interests include Ras oncogene and breast cancer biology.

Eric C. Chang is an Associate Professor in Baylor College of Medicine (Houston TX, USA). His lab is interested in a class of proteins that belong to the Ras GTPase superfamily which functions as binary switches to govern cell growth and differentiation. Abnormalities in these proteins are frequently found in human cancers. Ras proteins must localize to the plasma membrane to mediate the signal of growth from growth factors. Thus one of the major areas of research in the Chang lab centers on defining how plasma membrane localization of Ras proteins is controlled, such that this process may be targeted for drug development to treat cancers that are driven by oncogenic Ras proteins.

Jinwen Ma received his M.S. degree in Applied Mathematics from Xi'an Jiaotong University, in 1988 and the Ph.D. degree in Probability Theory and Statistics from Nankai University, in 1992. From July 1992 to November 1999, he was a Lecturer or Associate Professor at Department of Mathematics, Shantou University. From December 1999, he became a Full Professor at Institute of Mathematic, Shantou University. From September 2001, he has joined the Department of Information Science at the School of Mathematical Sciences, Peking University, where he is currently a Full Professor and Ph.D. tutor. During 1995 and 2003, he visited several times at the Department of Computer Science and Engineering, the Chinese University of Hong Kong as a Research Associate or Fellow. From September 2005 to August 2006, he worked as Research Scientist at Amari Research Unit, RIKEN Brain Science Institute, Japan. From September 2011 to February 2012, he visited as a Scientist at the Department of Systems Medicine and Bioengineering, Houston Methodist Hospital Research Institute. He has published over 100 academic papers on neural networks, pattern recognition, bioinformatics, and information theory.

Stephen T.C. Wong, Ph.D., PE, is John S. Dunn Distinguished Endowed Chair in Biomedical Engineering and Chief of Medical Physics, Houston Methodist Hospital, Founding Department Chair of Department of Systems Medicine and Bioengineering, Houston Methodist Research Institute at Texas Medical Center and Professor of Radiology, Neurosciences, Pathology and Laboratory Medicine of Weill Cornell Medical College of Cornell.