Journal of Imaging Science and Technology® 50(2): 187–192, 2006.

Estimation of Cell Count from Cell Culture Images

Hai-Shan Wu, Veneta Hanson and Joan Gil
Department of Pathology, Box 1194, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, New York 10029
E-mail: haishan.wu@mssm.edu

Abstract. *In this paper, we present an algorithm to estimate the cell counts from cultured rat B104 neuroblastoma cell images. Assuming that cells are alike, the algorithm identifies the representative cell based on processing the sorted size sequence of the threshold-segmented regions. The size of the representative cell is used to estimate the number of cells in each cluster where cells are attached and inseparable. Preprocessing procedures include the homomorphic filtering for improving the evenness of the image background, and gray-level morphologic dilation and erosion operations for filling the hollow cells. Results on the B104 neuroblastoma cell images are provided and compared with the numbers derived from manual cell count by pathologists. © 2006 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.(2006)50:2(187)]

INTRODUCTION

Quantitative analysis of the change of cell number over time is used to evaluate various physiological phenomena such as cell proliferation, differentiation, and cell death in the presence or depletion of biologically active compounds including growth or tropic factors.[1] B104 neuroblastoma cells display morphological and physiological features of neurons[2] and are widely used as a model for studying molecular *in vitro* analysis of phenomena in basic and applied neuroscience.[3] Cellular loss or gain as a consequence of cell death, proliferation or other mechanisms in an *in vitro* culture system can only be determined by an accurate count of cells remaining on a culture plate. Biochemical methods of cell death and proliferation such as flow cytometry, thymidine incorporation, and TUNEL assay often introduce methodological variables that may compromise the accuracy of actual cell count, therefore, it has been suggested that quantification should include microscopic evaluation of cell count.[4] The estimation of cell count in an image requires accurate segmentation of the cell regions.[5,6] Images of red blood cells and spores without significant crowding can be well segmented by thresholding and watershed algorithms.[7] However, it is difficult to segment images of overcrowding cells of large clusters such images of B104 neuroblastoma cells.

Quantification based on sufficiently large microscopic fields can be approached by digital image analysis. Such analysis can be hampered by cell clusters and by unevenness of examination fields resulting from illumination problems during photography with the inverted microscope. Although B104 neuroblastoma cells generally display homogeneously distributed bipolar cytoarchitecture with long neurite-like processes and a rounded cell body,[8] the presence of cell clumps is often observed. Such cell clusters, as well as unevenly illuminated microscopic fields can compromise digital cell counts. Here we present an algorithm for accurate digital estimation of cell count that compensated for cluster and illumination deficiency of microscopic cell culture images.

The paper is organized as follows. The section "Preprocessing" describes the preprocessing procedures of homomorphic filtering that improves the homogeneousness of the cell image background for the threshold segmentation of cells, and morphological dilation and erosion operations for the filling of the low intensity holes inside each cell. The "Identifying the Representative Cell" section presents an approach to identify the representative individual cell based on processing the size sequence of threshold-segmented regions. Results are provided in the section "Results" and discussions are presented in the "Conclusions and Discussions" section.

PREPROCESSING

A cell culture image usually contains scattered bright cells in a dark background that has uneven light. To differentiate the cell pixels from the background, a simple threshold may separate the higher intensity cell pixels from the lower intensity background. However, there are two problems that should be addressed before the threshold segmentation.

The first problem is the uneven background that may cause significant misclassification of pixels when a global threshold is applied. The homomorphic filtering with a high-pass filter can be used to reduce the slow changing unevenness in the background and retain the fast changing information.[9] Let the original image be represented by $x(i,j)$, for $i=0,1,\ldots,N_1-1$ and $j=0,1,\ldots,N_2-1$. The homomorphic filter is shown in Fig. 1 where a linear two-dimensional high-pass filter is applied on the logarithmic image

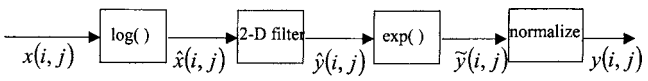

Figure 1. Homomorphic filtering reducing the slow changing background intensities.

Received Feb. 16, 2005; accepted for publication May 20, 2005.
1062-3701/2006/50(2)/187/6/$20.00.

$\hat{x}(i,j)=\log[x(i,j)]$ followed by an inverse exponential operation $\tilde{y}(i,j)=\exp[\hat{y}(i,j)]$. To ensure the output is still in the image intensity range of [0,255], a normalization mapping is performed such as $y(i,j)=255[\tilde{y}(i,j)-\tilde{y}_{\min}]/(\tilde{y}_{\max}-\tilde{y}_{\min})$, where $\tilde{y}_{\min}$ and $\tilde{y}_{\max}$ are the maximum and minimum values of $\tilde{y}(i,j)$, for $i=0,1,\ldots,N_1-1$ and $j=0,1,\ldots,N_2-1$, respectively.

The second problem is that the cells themselves are not even in intensities. The cells show the bright membrane rings with different thickness and black holes of variant sizes. An approach to fill the hollowed cell bodies is the gray-level morphologic closing that is a cascade of dilation and erosion operations of circular structure element of radius R, such as $g(i,j)=ero[dil[y(i,j)]]$, where $dil[y(i,j)]=\max_{m^2+n^2\leq R^2}[y(i+m,j+n)]$ and $ero[y(i,j)]=\min_{m^2+n^2\leq R^2}[y(i+m,j+n)]$, for $i=0,1,\ldots,N_1-1$ and $j=0,1,\ldots,N_2-1$. The dilation operation completely fills the dark holes inside the cells if R is large enough. The distortion of the cell regions resulting from the dilation is restored by a subsequent reverse procedure of erosion operation with the same structure element.

IDENTIFYING THE REPRESENTATIVE CELL

With the images of filled cells in clean and homogeneous background after preprocessing, satisfactory segmentation of cell regions can be achieved by a global thresholding separating cells of high intensities from darker background. Let the threshold-segmented image be represented by $g_T(i,j)=1$ (true) if $g(i,j)\geq T$, and 0 (false) otherwise. A pixel at the coordinate (i,j) is considered belonging to a cell region if $g_T(i,j)=1$, to the background if $g_T(i,j)=0$. If the threshold, T, is properly selected, there should be a number of isolated cell regions of different shapes and sizes sparsely located in the black background. Each of the cell regions may contain one or more cells if its area is large enough. Separating the individual cells from a large region of mingled cells is difficult since they are often attached very tightly. In this section, we describe an approach to identify the representative cell that can be used to estimate the number of cells in each isolated region in the threshold-segmented image, $g_T(i,j)$.

Suppose there are totally K segmented regions that are addressed by the index k for $k=1,2,\ldots,K$. The background region is referred by index of 0 which is treated separately. Let the sequence $h_0(k)$, for $k=1,2,\ldots,K$, represent the number of pixels in the kth segmented region. When rearranging the sequence $h_0(k)$ in an ascending order, we obtain the sorted sequence $h(k)$, for $k=1,2,\ldots,K$, where $h(k)\leq h(m)$ if $k<m$. If the cells have a similar size, we can expect the regions of the isolated single cells are located together in the sorted sequence since they have the similar size and significantly different from those much larger regions with multiple cells. To have a more stable and smooth curve of sequence, we use a moving average filter with window of size $2M_s+1$, such as $h_s(k)=1/(2M_s+1)\Sigma_{m=-M_s}^{M_s}h(k+m)$, for $k=M_s+1,\ M_s+2,\ldots,K-M_s$. Based on the assumption of similar size of cells, $h_s(k)$ will have a relatively leveled period with small slopes corresponding to the single cell regions. Let the derivative sequence $h_d(k)=h_s(k)-h_s(k-1)$, for $k=M_s+2,\ M_s+3,\ldots,K-M_s$. Large derivatives that are basically located in the high end can be excluded since they correspond to two regions with different number of cells. We seek the cutting point at K_2 by searching $h_d(k)$ from M_s+2 upward until $h_d(K_2-1)<T_2$ and $h_d(K_2)\geq T_2$, where T_2 is a specified constant so that the (K_2-1)th and (K_2)th regions have different number of cells. Thus, T_2 should be larger than the average size of cells. While T_2 can be significantly larger than the average cell size, limiting its value can make the computations more simple. On the other hand, much smaller regions such as regions less than a few pixels should not be cell regions. While the derivatives are small for these regions, the relative derivatives to the size of the regions defined as $h_r(k)=h_d(k)/h(k)$ are high. Again, we use a moving average filter to smooth the relative derivative such as $h_{sr}(k)=[1/N_s(k)]\Sigma_{m=-M_s}^{M_s}h_r(k+m)$, $k=M_s+1,\ M_s+2,\ldots,K-M_s$, where $N_s(k)$ is the number of points in both the moving window of size $2M_s+1$ and the signal support $[M_s+1,\ K-M_s]$. $N_s(k)$ is equal to $2M_s+1$ if $2M_s\leq k<K-2M_s$, and smaller otherwise.

If the $h(k)$ is too small, then the kth region may not be a cell region. Suppose $h(m'-1)<T'$ and $h(m')\geq T'$, where

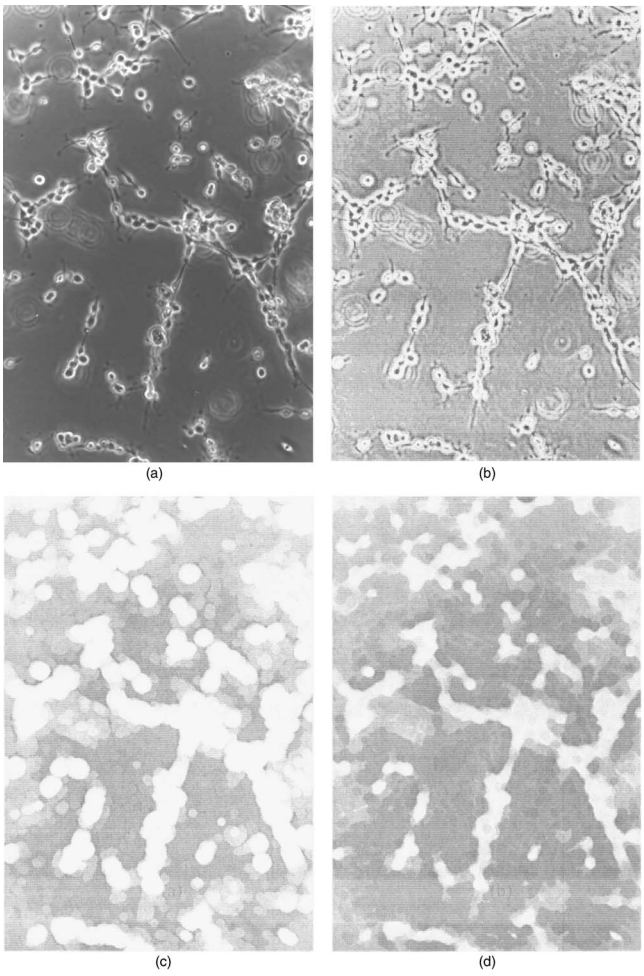

Figure 2. (a) Original cultured rat B104 neuroblastoma cell image; (b) normalized Homomorphically filtered image; (c) dilated image of (b); and (d) eroded image of (c).

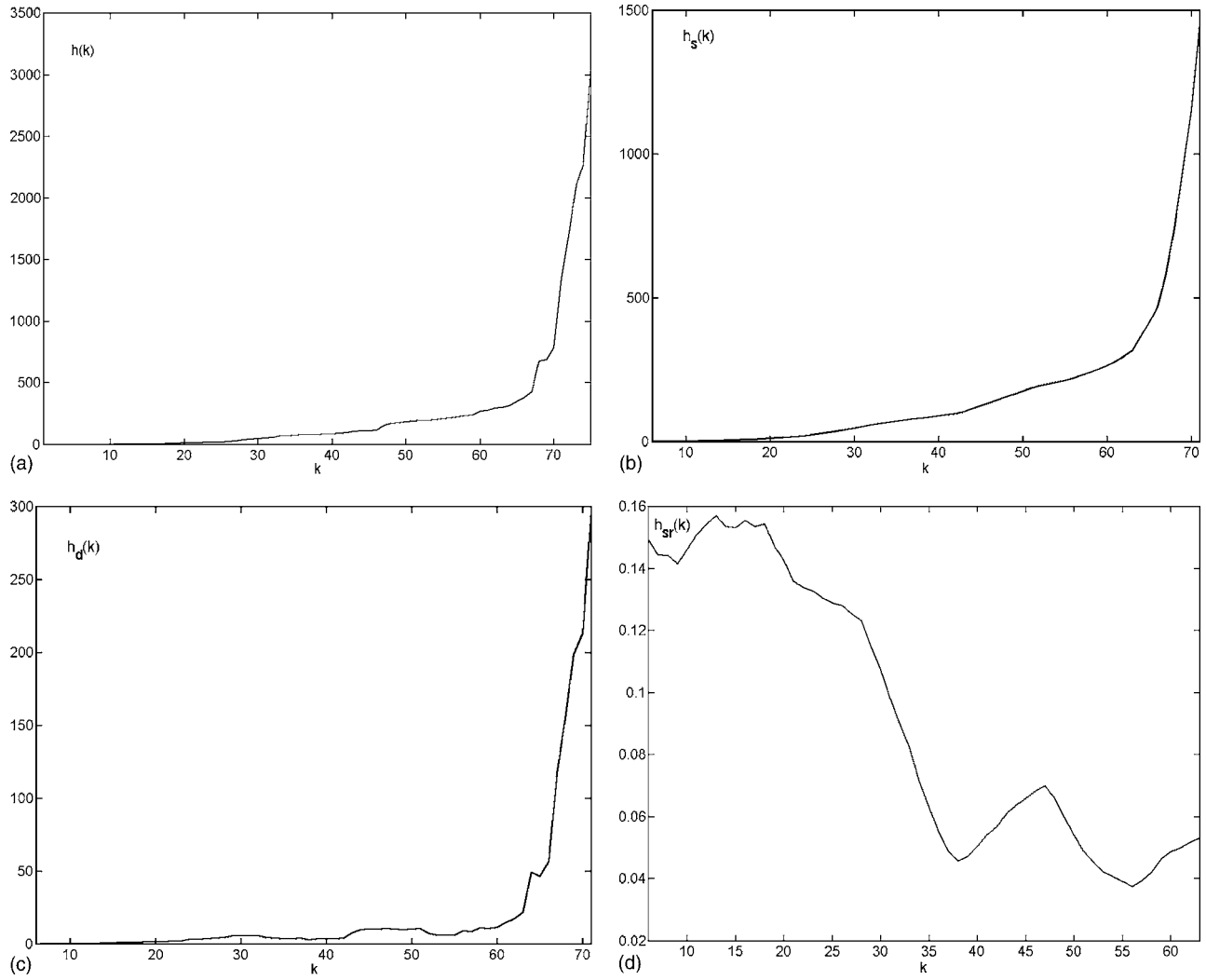

Figure 3. The sorted size sequence of the segmented regions and its derivatives.

T' is a small number. Assign the larger one of m' and M_s+1 to n', i.e., n' equals m' if $m' > M_s+1$, M_s+1 otherwise. The sequence of $h_{sr}(k)$ for $m' \leqslant k \leqslant K_2$ is an approximately decreasing sequence. The portion of regions in the range $m' \leqslant k \leqslant K_2$ with lower $h_{sr}(k)$ is the range of possible single cells. To separate the low $h_{sr}(k)$ from high $h_{sr}(k)$, we apply a threshold that is the middle of the total dynamic range, such as $\frac{1}{2}[\max(h_{sr})+\min(h_{sr})]$, where $\max(h_{sr})$ and $\min(h_{sr})$ are the maximum and minimum of $h_{sr}(k)$, for $m' \leqslant k \leqslant K_2$, respectively. Search $h_{sr}(k)$ upward from $m'+1$ for the index K_1 such that $h_{sr}(K_1-1) > \frac{1}{2}[\max(h_{sr})+\min(h_{sr})]$ and $h_{sr}(K_1) \leqslant \frac{1}{2}[\max(h_{sr})+\min(h_{sr})]$.

So far, we have found two indices K_1 and K_2 that set the lower and upper limits to the representative cell. While the isolated single cells are located in the index range of $[K_1, K_2]$, the isolated regions of double cells are likely included in this range as well since there may be a significant number of such regions in the image. However, regions of more than two cells should be very few if any included in this range since their small number of regions result in extraordinary high $h_d(k)$.

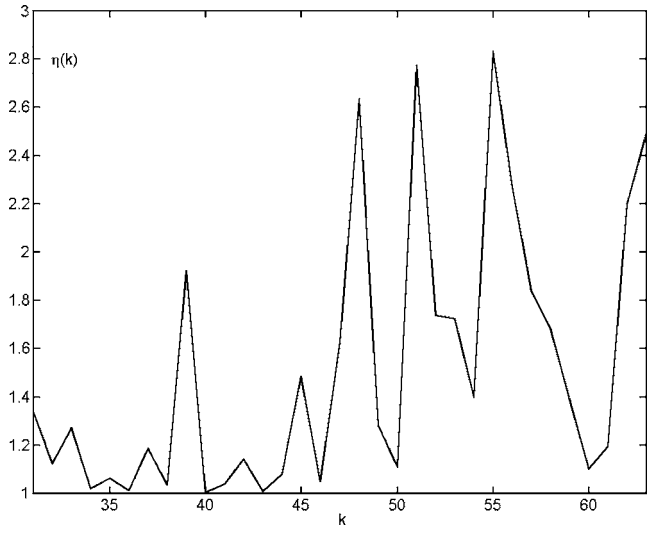

Figure 4. The circularity factor, $\eta(k)$, for $k = K_1, K_1+1, \ldots, K_2$. A lower factor corresponds to a region of shape more similar to a circle, while 1, the lowest possible value, corresponds to an ideal circular disk.

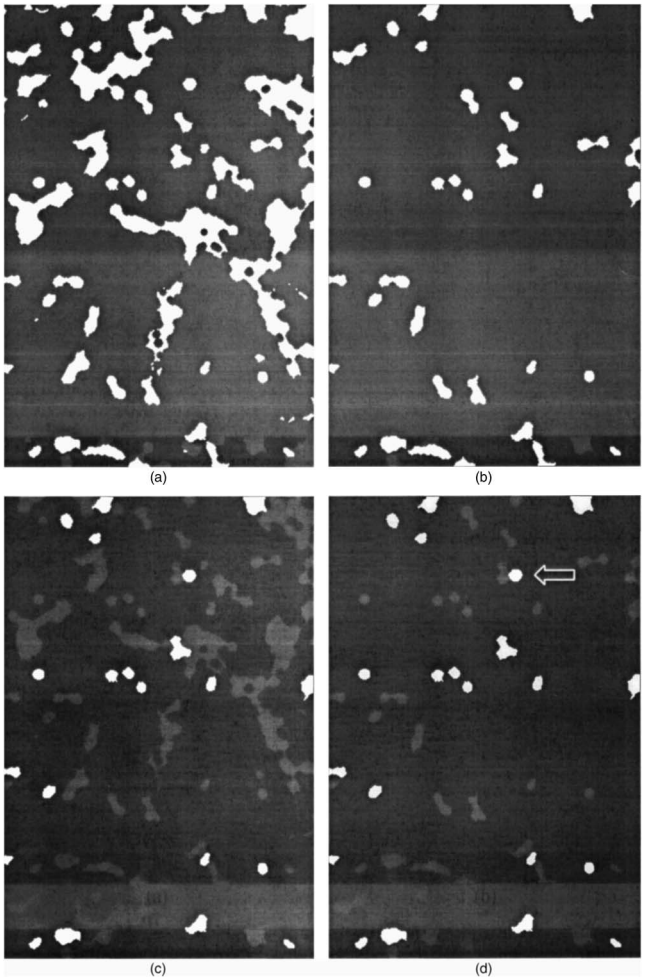

Figure 5. (a) The initial segmentation; (b) limited regions containing single, double or triple cells; (c) selected from (b) the regions that are closer to round disks with smaller circularity factors; and (d) the single region identified as the representative cell region shown in white pointed by an arrow while the rest cells set to lower gray level from (c). Estimated 186 cells based on the selected representative cell.

It is observed that the cells from cultured rat B104 neuroblastoma are generally nearly round in shape. Regions of double cells are much longer with two touched circles and significantly different from a round shape. Based on the shape information, it is efficient to differentiate between regions containing single cells and regions with double cells. We define a circularity factor, η, that measures the difference of a region of area A to a circular disk of the radius $\sqrt{A/\pi}$, such as $\eta(k)=(2\pi/A_k^2)\Sigma\Sigma_{g(i,j)=k}[(i-i_k)^2+(j-j_k)^2]$, $\forall(i,j)$, where A_k is the number of pixels in the kth region and (i_k,j_k) is the centroid of the kth region. The minimum value of $\eta(k)$ is 1 that corresponds to an ideal circular disk. A larger value of $\eta(k)$ corresponds to a region more different from a round disk such as a longer and narrower shape. It is noticed that regions of single cells show more roundness and thus have smaller η values. If a region whose η value is higher than a given value T_η, the region is then not considered a region of isolated single cell and should be excluded. In the remaining regions, the one with the median size is considered as the best representative region for the cells. With the representative cell region, the number of cell count of any other region can then be estimated by comparing their sizes. If the size of the representative region is measured as α, the total number of cells in a given region is estimated by $\langle h(k)/\alpha\rangle$, where the operator $\langle\cdot\rangle$ is to round the enclosed variable to its closest integer. The total number of cells in the whole image is then estimated as the sum of cell counts in all regions such as $\Sigma_{k=K_1}^{K}\langle h(k)/\alpha\rangle$.

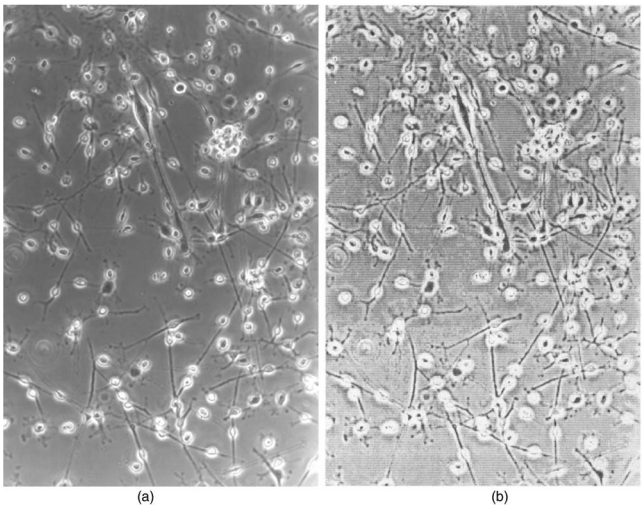

Figure 6. (a) Another original cultured rat B104 neuroblastoma cell image and (b) normalized homomorphically filtered image.

RESULTS

Figure 2(a) shows an original cultured rat B104 neuroblastoma cell digital image of size 450×304 and 256 gray levels acquired by phase microscopy. It is seen that many cells are mingled and attached together in clusters where individual cells are difficult to be separated. Although most cells are attached to form large clusters, there are many loose individual cells that are isolated from the others. Our algorithm is to first identify those isolated cells and then find the typical cell that is representative to others on average in size. To improve the thresholding segmentation of image, we apply a homomorphic filtering to produce an image that has more uniform background followed by a dynamic linear mapping to expand the image range to the maximum of $[0,255]$. Figure 2(b) is the result after the homomorphic filtering and mapping. Note that cells are hollowed with bright membranes and dark cores. To have a segmentation of the whole cell regions, it requires a mechanism to fill those holes without significantly changing the cell sizes and shapes. We apply a gray-level dilation to fill the holes followed by an erosion to restore the cell regions. Figure 2(c) is the dilated image of Fig. 2(b) by a structure element of a solid circle of radius 5, while Fig. 2(d) is the eroded image of Fig. 2(c) by the same structure element. Figure 2(d) is the preprocessed image in which the hollow cells have been turned to solid ones for the subsequent thresholding segmentation. The image has an intensity range of $[0,255]$. We select the threshold $T=90$,

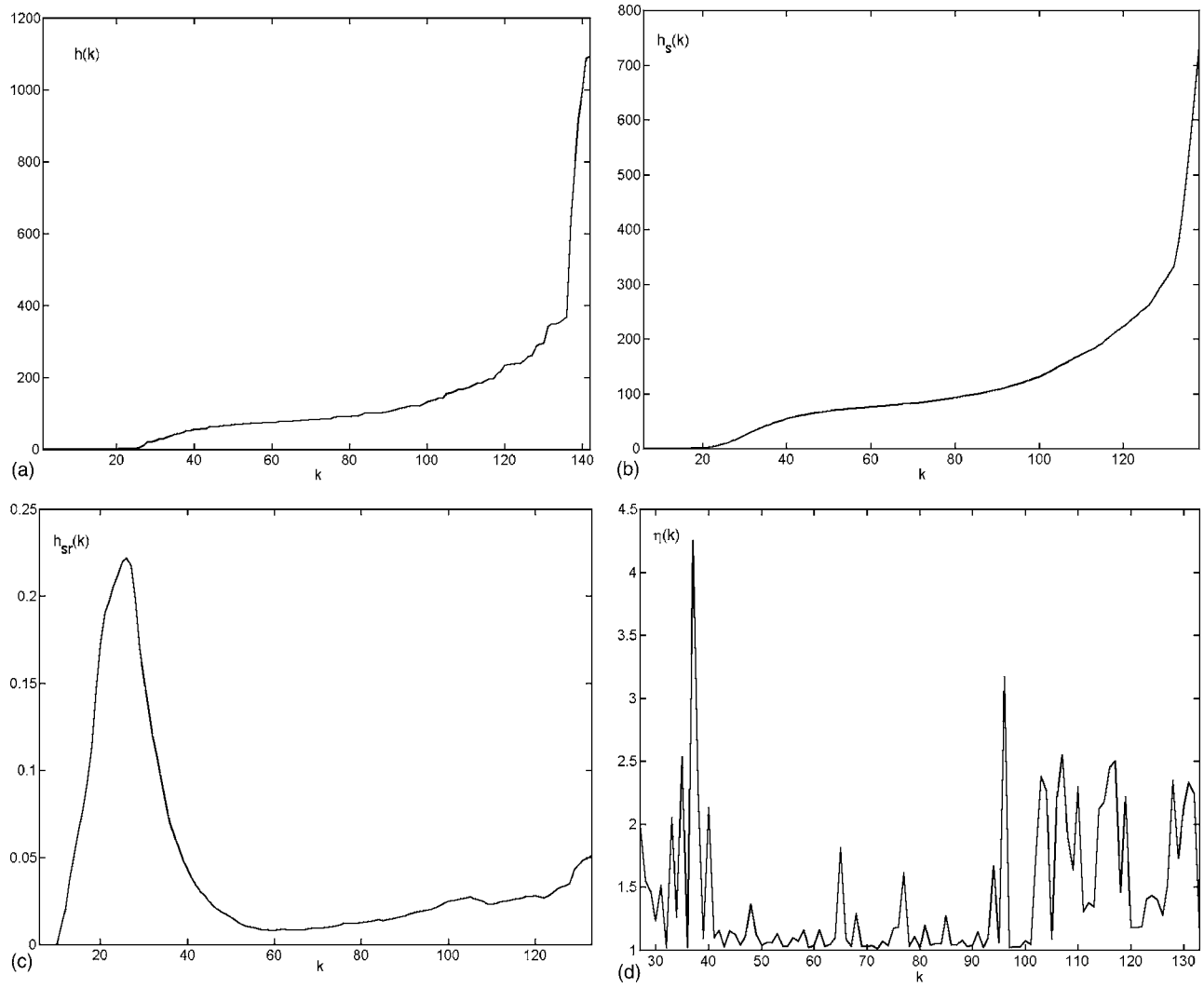

Figure 7. The sorted size sequence of the segmented regions, its derivatives, and the circularity factor.

somewhat below 127, the middle level of the intensity range, since the background near zero in intensities has a significantly smaller dynamic range than the foreground of cells. The resulting segmentation is shown in Fig. 5(a) where the black represents the background while the white represents the segmented cell pixels.

The next step is to evaluate the size of each threshold-segmented region and rearrange the size sequence in an ascending order to obtain the sorted size sequence $h(k)$, for $k=1,2,\ldots,K$, as shown in Fig. 3(a), where the number of isolated cell regions counted as $K=75$ meaning there are 75 isolated white regions in Fig. 5(a). The smoothed sequence $h_s(k)$ by a moving average filter of window of size of 9 is shown in Fig. 3(b). The derivative of the smoothed sequence, $h_d(k)$, is displayed in Fig. 3(c) where the value keeps low until after $k=50$ when it increases abruptly. Set the threshold $T_2=40$ for $h_d(k)$ to find the cutting index K_2. A T_2 value of 40 is a large value to ensure that the K_2th and the (K_2-1)th segmented regions contain different number of cells. The upper index boundary is found to be $K_2=63$ in this case. The smoothed relative derivative, $h_{sr}(k)$, is shown in Fig. 3(d). The size of the moving average window is also selected as 9. We select the constant $m'=10$ implying that regions smaller in size than ten pixels are not considered as a whole cell. The assumption is reasonable since a whole cell region is generally much larger than ten pixels. The maximum and minimum of $h_{sr}(k)$, for $m' \leqslant k \leqslant K_2$, are $\max(h_{sr})=0.157$ and $\min(h_{sr})=0.0375$, respectively. The midlevel is therefore $\frac{1}{2}[\max(h_{sr})+\min(h_{sr})]=0.0972$. The lower index boundary is found to be $K_1=31$ by searching through the smoothed relative derivative $h_{sr}(k)$ for $m' \leqslant k < K_2$. Thus, we expect that all the isolated regions of single cells are located in the index range $[K_1,K_2]=[31,63]$. The corresponding segmented cell regions in the index range of [31,63] are displayed in Fig. 5(b) where the regions considered too large or too small are eliminated. The next step is to further exclude those regions of multiple cells in the index range of $[K_1,K_2]$. The circularity factor, $\eta(k)$, of each region in the interested index range $K_1 \leqslant k \leqslant K_2$ is calculated and shown in Fig. 4. As known, regions of single cells have significantly lower value of circularity factor η than regions

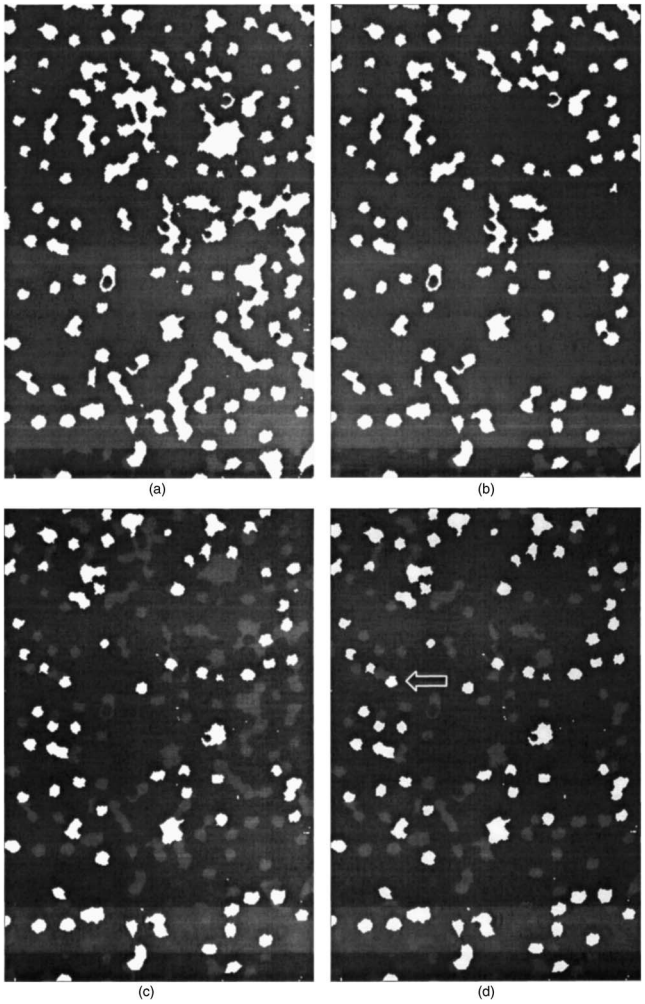

Figure 8. (a) The initial segmentation; (b) limited regions containing single, double or triple cells; (c) selected from (b) those regions closer to round disks with smaller circularity factors; and (d) the individual region identified as the representative cell in white intensity pointed by an arrow while the rest cells set to gray level from (c). Estimated 229 cells based on the selected representative cell compared to about 202 cells of manual count by pathologists.

with two cells. By eliminating all the regions whose η value is above $T_\eta = 1.5$, we obtain the segmentation of isolated regions of single cells as shown in Fig. 5(c) in which regions appear longer and less round are eliminated. The best representative cell region is selected as the one whose size is the median of the remaining regions. In this case the representative cell is indexed as $k = 43$ in the size sequence and marked by a pointing arrow in Fig. 5(d). The size of the representative cell region in Fig. 5(d) is measured as $\alpha = 105$ pixels. The estimated total number of cells is then estimated to be 186 compared to 184, the numbers of cells derived from manual cell count by pathologists.

Figures 6–8 show the results of the estimation of cell count from another B104 neuroblastoma cell image. All the choosing parameters are the same as those used in the previous experiment. The derived parameters are $K_1 = 27$ and $K_2 = 133$. The index of the representative cell region is found $k = 74$. The size of the representative cell is $\alpha = 85$. The total number of cells is estimated to be 229 compared to 202, the manual count by pathologists.

CONCLUSIONS AND DISCUSSIONS

We have presented an algorithm for automatic counting of cells from cultured rat B104 neuroblastoma cell images. A preprocessing procedure converts cell regions from shapes of hollow rings to shapes of solid circles by filling the holes in center of cells with morphological operations of dilation and erosion. Assuming that the cells are alike, the number of cells in each cell cluster is estimated based on the size of a representative cell. If the number of cells in an image is large enough, there are likely many cells that are loose and isolated from the others. The loose cells are located adjacent in the ascending size sequence of the segmented regions forming a wide plateau. The algorithm seeks all the loose single cells based on the derivatives of the sorted size sequence and the roundness of the segmented regions before selecting the region with median size as the representative cell region. The total number of cells is finally estimated based on the size of the representative cell.

The assumptions in the algorithm imply the limitation of its applications. If cells in an image are not alike, the estimation will be less accurate. If there are no loose cells that are individually isolated in the image, the algorithm will give inaccurate results. Fortunately, the images of B104 neuroblastoma cells satisfy these assumptions. The possible applications of the algorithm to other cell images remain to be investigated.

REFERENCES

[1] M. W. Miller, "Balance of cell proliferation and death among dynamic populations: A mathematical model", J. Neurobiol. **57**, 172–182 (2003).
[2] J. E. Bottenstein and J. H. Sato, "Growth of rat neuroblastoma cell line in serum free supplemented medium", Proc. Natl. Acad. Sci. U.S.A. **76**, 514–517 (1979).
[3] J. Luo and M. W. Miller, "Basic fibroblast growth factor and platelet-derived growth factor mediated cell proliferation in B104 neuroblastoma cells: Effect of ethanol on cell cycle kinetics", Brain Res. **770**, 139–150 (1997).
[4] A. Saraste, "Morphologic criteria and detection of apoptosis", Herz **24**, 189–195 (1999).
[5] H.-S. Wu and J. Gil, "Linear clustering for segmentation of color microscopic lung cell images", J. Imaging Sci. Technol. **47**, 161–170 (2003).
[6] H.-S. Wu, L. Deligdisch, and J. Gil, "Segmentation of microscopic nuclear images—a review", in *Recent Res. Devel. Electronics*, edited by S. G. Pandalai (Transworld Research Network, Kerala, India, 2004) Vol. **2**, pp. 1–17.
[7] S. A. Park, H. J. Park, B. I. Lee, Y. H. Ahn, S. U. Kim, and K. S. Choi, "Bcl-2 blocks cisplatin-induced apoptosis by suppression of ERK-mediated p53 accumulation in B104 cells", Molecular Brain Res. **93**, 18–26 (2001).
[8] J. C. Russ, *The Image Processing Handbook*, 4th ed. (CRC Press, Boca Raton, FL, 2002).
[9] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing* (Prentice-Hall, London, 1989).