DOI: 10.1007/s12541-014-0641-1

Determination of Histogram Type of Cell Images in Microsystem for Cell Tracking

Dong Pyo Hong[1], **Jaesool Shim**[2], and **Migyung Cho**[3,#]

1 Mechanical System Engineering, Chonbuk National University, 567, Baekje-daero, Deokjin-gu, Jeonju-si, Jeollabuk-do, South Korea, 561-756
2 School of Mechanical Engineering, Yeungnam University, 280, Daehak-ro, Gyeongsan-si, Gyeongsangbuk-do, South Korea, 712-749
3 Department of Media Engineering, Tongmyong University, 428, Sinseon-ro, Nam-gu, Busan, South Korea, 608-711
Corresponding Author / E-mail: mgcho@tu.ac.kr, TEL: +82-51-629-1214, FAX: +82-51-629-1129

KEYWORDS: Cell segmentation, Image histogram, Uni-modal distribution, Bi-modal distribution, Polygonal approximation

The process of separating cell objects from a background is very important in cell image analysis and cell tracking. Most cell image histograms that have uneven illumination during microscopic cell imaging have a uni-modal distribution. In general, it is more difficult to segment objects in images having a uni-modal distribution than in those having a bi-modal or multi-modal distribution; therefore, images having a uni-modal distribution require a special segmentation algorithm that is more complicated and requires computing costs. If the histogram of the given cell image is known, an appropriate segmentation algorithm is applied for an accurate segmentation. In this paper, we proposed an algorithm that automatically determines the histogram distribution of an inputted cell image. The proposed method uses the polygonal approximation method. We tested this algorithm on various cell images and the error was found to be less than 5% on average.

Manuscript received: July 10, 2014 / Revised: July 27, 2014 / Accepted: July 28, 2014

NOMENCLATURE

$f(k)$ = the frequency of an intensity value k ($0 \le k \le 255$)
P_k = a pair of coordinates (k, f(k))
H = the point set that constitutes the histogram of a cell image
L_{ij} = a straight li ne that passes the starting point P_i and the last position P_j

1. Introduction

The segmentation of a cell object from the cell image obtained from an optical microscope is one of the indispensable operations in the bio-imaging field. This field is an interdisciplinary combination of medicine, nano-biology, and biology, which involve the study of cell tracking, automatic cell counting, and three-dimensional cell restoration from a cell image. Extensive research has recently been carried out in this field.[1-3] Cell segmentation is a very challenging problem when the cell image is not in a good state because of background noises and uneven light source encountered when obtaining the cell image. Segmentation algorithms proposed in the image processing field such as Voronoi-based algorithms, histogram-based clustering or threshold algorithms, and active contours are used for isolating cell objects from a cell image.[2]

If a cell image is noise-free and has a clear background, we can segment cell objects easily by using a segmentation algorithm from the image processing field. The problem is that a general segmentation algorithm cannot appropriately segment cells if the cell image is not clear. If the intensity histogram is analyzed, a clear cell image, in most cases, has a bi-modal or multi-modal distribution. In such cases, accurate cell segmentation can be easily carried out by using a global threshold algorithm such as the Otsu algorithm. However, the application of a conventional algorithm to a cell image, which has different dimness in the background areas due to uneven reflection of light during imaging or severe noise, results in incorrect segmentation. Analysis of intensity histograms of these images typically reveals uni-modal distributions. A cell image with a uni-modal distribution requires a segmentation algorithm for uni-modal distribution or a local threshold algorithm.

The Otsu algorithm is efficient in terms of calculation and is cost-effective, but a segmentation algorithm for uni-modal distribution requires a considerable amount of time and a relatively high computing cost. Therefore, for efficient and accurate cell segmentation, it is

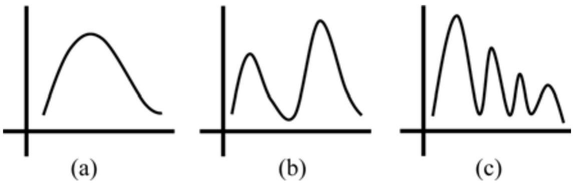

Fig. 1 Three types of histogram: (a) Uni-modal distribution (b) Bi-modal distribution (c) Multi-modal distribution

desirable to apply the Otsu algorithm which has low computing cost in the case of a bi-modal or multi-modal distribution, and to apply a special algorithm in the case of a uni-modal distribution after analyzing the histogram of the inputted cell image. In this study, we proposed and implemented an algorithm that determines the type of distribution that a given image has by analyzing the histogram of the inputted cell image. This algorithm analyzes the shape of a histogram after smoothing the given intensity histogram using the polygonal approximation.

2. Histogram Types of Cell Images

An intensity histogram for an image is a frequently used tool for image analysis. A grayscale image has intensity values ranging from 0 to 255 and the x-axis of the histogram has an intensity value ranging from 0 to 255 and the y-axis displays the frequency of its intensity exhibited on the image. In the case of a color image, a histogram is obtained by converting it into a grayscale image. The shape of the histogram varies depending on the characteristics of the images. In general, the histogram can have one of the three types of distributions: uni-modal, bi-modal, and multi-modal distributions.

In statistics, a large number of studies have been conducted to determine whether given values have a uni-modal distribution.[4,5] Representatively, it is known that the sufficient condition under which an independent variable X with a cumulative distribution function F is a uni-modal distribution satisfies Eq. (1).[4] In Eq. (1), μ denotes average; μ, median; and σ, standard deviation.

$$\frac{|\theta-\mu|}{\sigma}\le\sqrt{3}\ ,\ \frac{|\nu-\mu|}{\sigma}\le\sqrt{0.6}\ ,\ \frac{|\theta-\nu|}{\sigma}\le\sqrt{3} \qquad (1)$$

However, it is difficult to use the conditions of Equation 1 to decide whether the cell image histogram has a uni-modal distribution. In statistics, a uni-modal distribution is defined to be one with the maximum value of a single mode (i.e., a single peak), and thus, the peaks that have local minima are ignored. However, an image histogram has various forms of local minima, and the values of some local minima have important implications such that in many cases, they should be considered as a single mode. Consequently, Eq. (1) conditions are insufficient to determine whether the histogram of an image has a uni-modal distribution. Therefore, we attempt to develop a method that determines the type of histogram distribution for given cell images by using the shapes of the histograms.

First, the distribution characteristics of a histogram, which reflects the quality of the cell image, are examined. Fig. 2 shows a group of clear cell images and a group of unclear cell images that are used in the experiment. Fig. 2(a)~(c) are the cell images with relatively good quality and (d)~(f) are the cell images of low quality due to noise and an uneven light source. The cell image in Fig. 2(c) was extracted from Ref. 6, and the other cell images were extracted from Ref. 7.

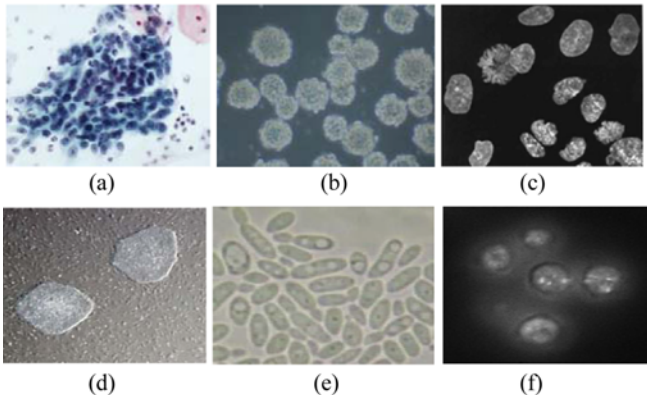

Fig. 2 Cell images: (a)~(c) Clear cell images, (d)~(f) Degraded cell images

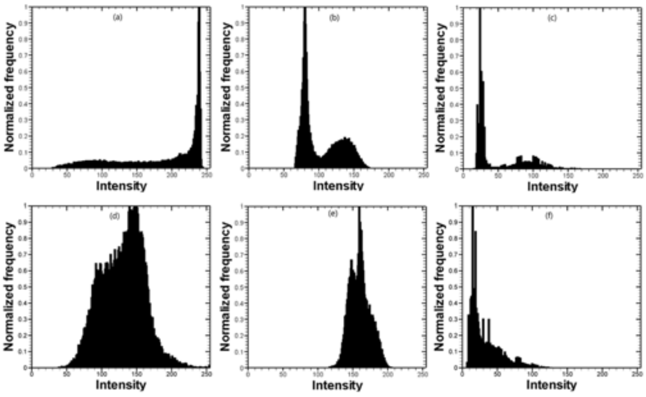

Fig. 3 Normalized histogram distribution corresponding to the cell images shown in Fig. 2

Fig. 3 shows the normalized histograms corresponding to the images shown in Fig. 2. Fig. 3 (a)~(c) show a similar distribution as bi-modal distribution; (d)~(f) images show rather complex shapes, but they are more similar to uni-modal distribution than to bi-modal distribution.

The cell segmentations of the images of the two groups yielded widely differing results. The segmentation results were accurate for the images with clear backgrounds when the Otsu algorithm, which is the most widely used among segmentation algorithms, was applied, but the cell images with a dim background could not perform accurate segmentation. Consequently, in order to obtain accurate segmentation results for a given cell image with a certain quality level, global thresholding algorithms such as the Otsu algorithm may be applied to a bi-modal distribution, whereas a special segmentation method should be applied to cell images with a uni-modal distribution.

3. Smoothing a Histogram using Polygonal Approximation

To determine the distribution on a histogram, the first process is the smoothing of the shape of the histogram by applying the polygonal approximation technique. If the intensity value of the image is drawn as a bar graph, the upper part shows a jagged shape as illustrated in Fig. 3. The graph's shape information has to be utilized in order to determine the histogram type, but the jagged shape hampers the accurate determination of the distribution type. The polygonal approximation technique is used for extracting the accurate shape of an object in the field of image processing or computer vision.[8] The polygonal approximation that is used for extracting the complex shape of an object in the field of image processing is diverse and complex. We use the polygonal approximation method simply to smooth the jagged shape of a histogram and not to extract the complex object, which is unnecessary the use of the approximation method that requires a complex calculation process.

The following polygonal approximation method was used in this study. Let the intensity value of the histogram be k ($0 \le k \le 255$) and the frequency of each intensity value be f(k). A set H is consist of points that constitutes histograms, then H is $\{P_k = (k, f(k): 0 \le k \le 255\}$. Let the starting point where polygonal approximation is applied be $P_i \in H$ and the point at the position displaced from P_i by a constant displacement (offset) be $P_j \in H$ ($j = i + \text{offset}$). The offset is set with an integer greater than 3.

Algorithm 1: Smoothing a histogram using polygonal approximation

[Step 1] Obtain the equation of a straight line it L_{ij} that passes the starting position P_i and the last position P_j.

[Step 2] Define the midpoints between P_i and P_j as set M, yielding $M = \{P_{i+1}, P_{i+2}, \ldots, P_{i+j-1}\}$. Find the perpendicular distance value d_{i+1}, d_{i+2}, ..., d_{i+j-1} for all points belonging to M to the straight line L_{ij}. Then, the distance values of all the points between P_i and P_j, and the straight line L_{ij} that passes through P_i and P_j can be expressed by Eq. (2).

$$dist(L_{ij}, M) = \sum_{k=i+1}^{i+j-1} d_k \tag{2}$$

[Step 3] Repeat Steps 1 and 2 by comparing dist(L_{ij}, M) with the fixed threshold value D, deleting all the points in set M if they are smaller than D, and setting P_j as the starting point. If the value of dist (L_{ij}, M) is greater than D, set P_{i+1} as the starting point and repeat Steps 1 and 2. This recursive process is repeated until i's value becomes 255 - offset.

[Step 4] Repeat Steps 1 to 3 until either one of the two following conditions becomes false. The first condition is to have less than 15 points that are not deleted. The second condition is to repeat Step 4 less than 15 times.

The histogram approximation result depends on the distance threshold D obtained in Step 3. If the distance threshold is very small, the histogram simplification effect cannot be obtained; if it becomes very large, points that should not be deleted might be deleted. Therefore, it is important to determine the optimum distance threshold. After experimenting with various cell images, the distance threshold set by using the standard deviation showed the best results. The formula used for calculating the distance threshold in this study is as follows:

$$D = K \times \sqrt{\frac{(p(x,y)-\mu)\times(p(x,y)-\mu)}{n}} \tag{3}$$

In Eq. (3), μ denotes average intensity of the image and the constant K has a value between 0 and 1. Fig. 4 illustrates the result when K is 0.7. As the value of the constant K decreases, the threshold value D of Equation 3 also decreases; therefore, there are few midpoints that are deleted, and hence, the original histogram shape can be maintained.

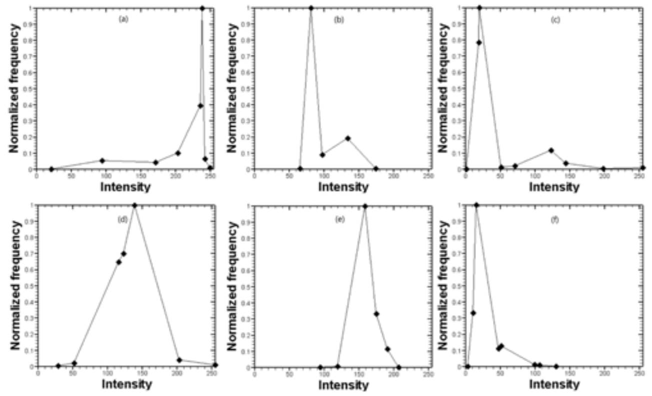

Fig. 4 Result of applying polygonal approximation to Fig. 3's histogram

A closer look at the smoothing result of the histograms shown in Fig. 4 reveals that, except for points in the places where rapid changes occurred, most points are removed from the histograms shown in Fig. 4 as a result of smoothing. The purpose of histogram smoothing is to simplify the shape of the histogram except for the positions where the important changes occurred. In order to achieve this purpose, in this study, we set the K value as 0.7.

4. Determination of the Type of a Histogram

Upon the completion of the histogram smoothing process using polygonal approximation, the histogram type determination process begins. We classified the histogram types into the three distribution types associated with the clear differences in results when a segmentation algorithm is applied: uni-modal, bi-modal, and multi-modal distributions. The method of determining a histogram type by analyzing the histogram smoothing results of the inputted cell image is as follows:

Algorithm 2: Determination of the type of a histogram

[Step 1] If we define a point set constructing a smoothed histogram as set SH, then SH can be expressed as $\{P_k = (x(k), f(x(k))\}$. In SH, k represents all the remaining points not removed during the smoothing process. Find a histogram sign sequence (HSS) value of two consecutive points P_k and P_{k+1} for all elements in the set SH by using Eq. (4).

$$HSS(k) = \begin{cases} + & if\ \ f(x(k)) \le f(x(k+1)) \\ - & if\ \ f(x(k)) > f(x(k+1)) \end{cases} \tag{4}$$

[Step 2] If the number of the continuing same signs is one in the results of Step 1, then calculate the intensity value difference, i.e., the difference between x(k) and x(k+1), at that position. For example, the HSS of Fig. 4(f) is (++-+---). In this HSS, the continuing number of the third and fourth signs is one for each. In this case, we calculate the intensity value difference between P_3 and P_4, $|x(4) - x(3)|$, and that between P_4 and P_5, $|x(5) - x(4)|$. Then, convert the sign at the position to the same sign of the previous sign if the intensity value difference is smaller than 5. $|x(4) - x(3)|$ is 12, and $|x(5) - x(4)|$ is 3. Therefore,

Table 1 Performance of the proposed method

Type of cell image	Clear cell images			Degraded cell images		
Null hypothesis (H_0)	uni-modal	bi-modal	multi-modal	uni-modal	bi-modal	multi-modal
Type 1 error	0.0	0.048	0.0	0.05	0.09	0.00
Type 2 error	0.0	0.024	0.0	0.00	0.09	0.00

maintain the third sign as it is and convert the fourth sign to – as the same sign of the previous one. Hence, the final HSS of Fig. 4(f) becomes (++-----).

[Step 3] Insert the MARK sign in the position where the sign changes. For example, the HSS of Fig. 4(f) becomes (++MARK-----).

[Step 4] Determine the histogram as uni-modal distribution if the number of MARKs in HSS is 1, as bi-modal distribution if it is 2, and as multi-modal distribution if it is 3 or greater.

We divided the cell images to two groups: the clear image group and the unclear image group. Then, we carried out experiments to analyze the performance of the algorithm proposed in this paper. Each group was composed of 20 cell images. Around 70% of the clear image group showed a bi-modal distribution, and approximately 86% of the dim image group, a uni-modal distribution.

We counted the type 1 and type 2 error occurrences to analyze the accuracy of the algorithm. A type 1 error occurs when the null hypothesis is true but the system rejects it. For example, it occurs when the null hypothesis says “It is a uni-modal image,” but the algorithm determines it as a bi-modal or multi-modal image by rejecting its being uni-modal. A type 2 error occurs when the null hypothesis is false but the system accepts it. In other words, the null hypothesis says “It is a uni-modal image,” but the algorithm determines the given image as a uni-modal image even when it is not a uni-modal image.

Table 1 shows the experimental results of cell images by differently setting the null hypothesis for the two groups. The null hypothesis was set as “It is a uni-modal distribution,” “It is a bi-modal distribution,” and “It is a multi-modal distribution” according to the histogram distribution type. It could be observed that the experimental result of the clear cell image was relatively more accurate.

As for cell images, most of them have either a uni-modal or bi-modal distribution. The type 1 and type 2 errors of the proposed algorithm were confirmed to be less than 5% on average; therefore, it is considered very effective in automatically determining the type of histogram distribution.

5. Conclusions

In this paper, we proposed a method that automatically determines the type of histogram based on histogram shape of the cell image, which is simplified by applying the polygonal approximation technique on the cell image’s histogram. We experimentally tested the performance of the proposed algorithm on various cell images. The accuracy of the proposed method yielded very good results, with both type 1 and type 2 error occurrences being less than 5% on average. Therefore, the proposed algorithm is expected to serve as a useful tool of a cell tracking system for determining the distribution that an image has in the cell segmentation process.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0028562) and this work was partially supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (NRF-2012R1A1A2009392).

REFERENCES

1. Yin, Z., Bise, R., Chen, M., and Kanade, T., “Cell segmentation in Microscopy Imagery using a Bag of Local Bayesian Classifiers,” Proc. of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 125-128, 2010.
2. Du, X. and Dua, S., “Segmentation of Fluorescence Microscopy Cell Images using Unsupervised Mining,” The Open Medical Informatics Journal, Vol. 4, pp. 41-49, 2010.
3. Bise, R., Yin, Z., and Kanade, T., “Reliable Cell Tracking by Global Data Association,” Proc. of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1004-1010, 2011.
4. Johnson, N. L. and Rogers, C. A., “The Moment Problem for Unimodal Distributions,” The Annals of Mathematical Statistics, Vol. 22, No. 3, pp. 433-439, 1951.
5. Medgyessy, P., “On the Unimodality of Discrete Distributions,” Periodica Mathematica Hungarica 2, pp. 245–257, 1972.
6. Smith, P. J. and Errington, R. J., “Live Cell DNA Labeling and Multiphoton/Confocal Microscopy,” in: Cell Biology-A Laboratory Handbook, 3rd Ed., Celis, J. E., Hunter, T., Carter, N., Shotton, D., Small, J. V., Simons, K., (Eds.), Academic Press, Vol. 1, Chap. 37, 2006.
7. http://inperc.com/files/CellImages.zip
8. Rosin, P. L., “Techniques for Assessing Polygonal Approximations of Curves,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 6, pp. 659-666, 1997.