Computational Biology and Bioinformatics
2014; 2(4): 52-56
Published online September 20, 2014 (http://www.sciencepublishinggroup.com/j/cbb)
doi: 10.11648/j.cbb.20140204.11
ISSN: 2330-8265 (Print); ISSN: 2330-8281 (Online)

Human cell detection in microscopic images through discrete cosine transform and Gaussian mixture model

Faezeh Rohani[1], Hamid Hassannia[2], Mohammad Amin MoghaddasiFar[1], Elham Sagheb[3]

[1]Department of Computer Science, Sanabad Institute of Higher Education, Golbahar , IRAN
[2]IEEE Member, Tehran, IRAN
[3]Ferdowsi University of Mashhad, Mashhad, IRAN

Email address:

faezeh.rohani@ut.ac.ir (F. Rohani), Hamid.h.h@ieee.org (H. Hassannia), Amin.moghaddasi@ut.ac.ir (M. A. MoghaddasiFar), e_sagheb@yahoo.com (E. Sagheb)

To cite this article:

Faezeh Rohani, Hamid Hassannia, Mohammad Amin MoghaddasiFar, Elham Sagheb. Human Cell Detection in Microscopic Images through Discrete Cosine Transform and Gaussian Mixture Model. *Computational Biology and Bioinformatics.* Vol. 2, No. 4, 2014, pp. 52-56. doi: 10.11648/j.cbb.20140204.11

Abstract: Automatic detection of human cell is still one of the most common investigation methods that may be used as part of a computer aided medical decision making system[1]. In this paper a statistical method based on Gaussian Mixture Model is applied to human cell detection in microscopic images[2]. 120 normal microscopic images of human cell from our research laboratory were used for analysis. Texture and grayscale features extracted from blocks of these images are given to Gaussian Mixture Model as input. It is used to model this data into three classes which are cell, extra cellular space and cell membrane [3]. Our proposed algorithm is applied on a sample dataset and experimental results show that this model is both accurate and fast with overall detection rate of around 91.23%. Error rate for cell detection was 1.82%.

Keywords: Human Cell Detection, Cell Segmentation, DCT, Gaussian Mixture Model

1. Introduction

The Human body contains about 100 trillion cells, each working together in a complex symphony of interactions to perform everything from providing structure and stability and providing energy and means of reproduction for an organism[4]. All living organism are composed of the cells and depended on these cells to function normally. A typical cell size is 10 µm and a typical cell mass is 1 nanogram. Automatic cell detection method offers a new powerful tool for understanding the chemical biology of the complex cellular processes and provides opportunities for identifying new targets in the discover of disease. Many clinical applications require precise, reliable and fully automated measurements of cell tissue dimensions, even when the data is degraded by noise and imaging artifacts. Detection of the human cells is the fundamental process which involves in partitioning of a data space in electron microscope image into meaningful salient regions. While image segmentation and object detection in general has been studied for a log period in machine vision and digital image processing, the use of existing methods in human cell detection has been very limited so far.

Some common digital image processing techniques that are developed for human cell detection are based on the image threshold which could be implemented by pixel classification [5], [6], [7].

There are other approaches that use region based information but not applicable, because the darker background regions may be misclassified as cells and lighter cell regions may be classified as background [8]. One of the fundamental tasks of automated cell detection method is to obtain accurate cell segmentation that often precedes other analyses such as cell morphology, tracking and behavior.

The present paper is focused on a novel design and integration of Discrete Cosine Transform (DCT) and Gaussian Mixture Model (GMM) to perform human cell detection on microscopy images. The accurate and computationally efficient detection of closely grouped cells is the focus of this paper.

In this paper we proposed an algorithm based on DCT GMM to detect the human cells in an accurate and reliable manner. The goal of the work is to investigate whether DCT and GMM could be used for cell detection process.

The present contribution is expected to draw more attention from digital image processing community to human cell detection and its application. 120 normal microscopic images of human cell from our research laboratory were used for this analysis. Texture and color features extracted from blocks of these images are given to GMM as input. GMM is used to model this data into three classes which are cell, extra cellular space and cell membrane. Our proposed algorithm is applied on a sample dataset and experimental results show that this model is both accurate and fast with overall detection rate of around 91.23%. Error rate for cell detection was 1.82%.

The rest of this paper is arranged as follows. In section 2 and 3 we describe the fundamental concepts of DCT and GMM. Section 4 illustrates our proposed algorithm by using DCT and GMM. Section 5 and 6 are the experimental results and conclusion respectively.

2. DCT

Discrete Cosine Transform (DCT) is a longstanding technique in the digital image processing literature. DCT expresses a sequence of finitely many data points in terms of a sum of cosine functions oscillating at different frequencies. DCTs are important to numerous applications in science and engineering, from Lossy compression of audio (e.g. MP3) and images (e.g. JPEG), where small high-frequency components can be discarded, to spectral methods for the numerical solution of partial differential equations [9]. The DCT is similar to the discrete Fourier transform and transforms a signal or image from the spatial domain to the frequency domain [9].

The general equation for two dimensional (Image Size: N×M) DCT is defined as follow:

$$F(u,v) = \left(\frac{2}{N}\right)^{\frac{1}{2}}\left(\frac{2}{\mathrm{M}}\right)^{\frac{1}{2}} \sum_{\mathrm{i}=0}^{N-1}\sum_{\mathrm{j}=0}^{\mathrm{M}-1} \mathrm{A}(i).\mathrm{A(j)}.\mathrm{B} \qquad (1)$$

$$\mathrm{B}= \cos\left[\frac{\pi.\mathrm{u}}{2.\mathrm{N}}(2i+1)\right]\cos\left[\frac{\pi.v}{2.M}(2j+1)\right].\mathrm{f(i,j)}$$

Where f (i, j) is the image intensity function and $F(u,v)$ is a 2D matrix of DCT coefficients. Generally, the DCT coefficients are divided into three bands or sets; low frequencies, middle frequencies and high frequencies. Low frequencies are correlated with the illumination conditions and high frequencies represent noise and small variations. Middle frequencies coefficients contain useful information and construct the basic texture of the image [9].

3. GMM

In this section we shall begin with a short introduction to Gaussian Mixture Models (GMM), followed by a explanation of its equations. GMM is among the most statistically mature methods for clustering (though they are also used intensively for density estimation)[10]. It is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as vocal-tract related spectral features in a speaker recognition system.

A Gaussian mixture model is a weighted sum of M component Gaussian densities as given by the equation:

$$p\,(X|\delta) = \sum_{i=1}^{M} w_i\, g(X|\mu_i, \sum_i\,) \qquad (2)$$

Where X is a D-dimensional continuous-valued data vector (i.e. measurement or features), W_i, i = 1, . . . ,M, are the mixture weights, and $g(X|\mu_i, \sum_i\,)$, i = 1, . . . ,M, are the component Gaussian densities. Each component density is a D-variant Gaussian function of the form:

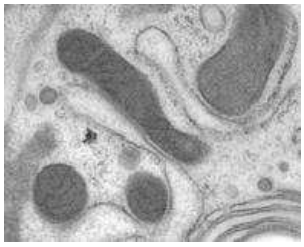
$$g(X|\mu_i, \sum_i\,) = \frac{1}{(2\pi)^{D/2}\,|\sum_i|^{\frac{1}{2}}} \exp\left\{\frac{-1}{2}\,(X-\mu_i)\sum_i^{-1}(X-\mu_i)\right\} \qquad (3)$$

With mean vector μ_i and covariance matrix $\sum_i$. The mixture weights satisfy the constraint that $\sum_{i=1}^{M} w_i = 1$.

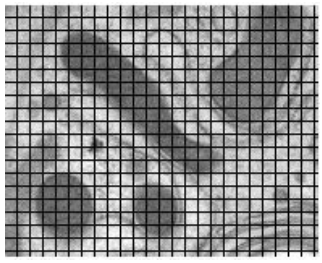
The complete Gaussian mixture model is parameterized by the mean vectors, covariance matrices and mixture weights from all component densities.

The choice of model configuration (number of components, full or diagonal covariance matrices, and parameter tying) is often determined by the amount of data available for estimating the GMM parameters and how the GMM is used in a particular biometric application [11], [12].

4. Proposed Algorithm



(a)



(b)

Figure 1. *(a) Original Image, (b) using block of size 8 × 8.*

The main purpose of the proposed algorithm is to detect the human cell in the microscopic images. This algorithm is based on DCT and GMM. We categorized the features into two categories, texture features and grayscale features. Texture and grayscale features extracted from blocks of the sample images are given to GMM as input. We used

energies of DCT coefficient in regions of frequency space corresponding to wavelet decomposition for the texture feature [9]. The energies of DCT regions are efficient in texture analysis and classification.[13][14] This feature can be considered as equation (1) for an N×M image size. In this paper we used a block of size 8×8, and then both of N and M are equal to 8. As you see in figure 1, we selected a block size of 8×8 because the experimental results showed that this block is big enough to capture a sample of largest texture of our microscopic images. As we mentioned earlier, the middle frequencies coefficients of DCT contain useful information about the basic texture of the image. Hence, these middle frequencies were selected as the particular candidates for cell detection in this paper.

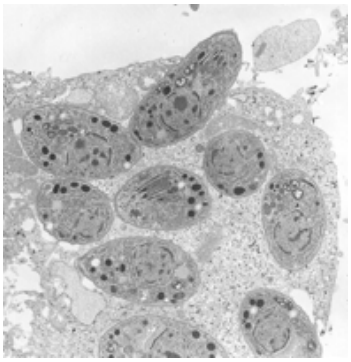
For grayscale features, we used grayscale mean, median, histogram and covariance values information over a block.

Mean is a method to derive the central tendency of a sample space. If we have sample space {a[1], a[2],… a[n]}, then arithmetic mean is defined via the equation (4) [15] :

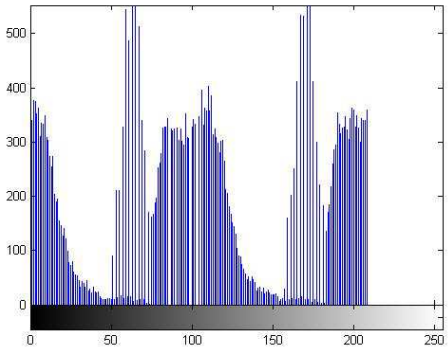
$$mean = \frac{1}{n} \sum_{i=1}^{n} ai \quad (4)$$

Median is described as the numeric value separating the higher half of a sample, a population, or a probability distribution, from the lower half [15].

Covariance matrix is symmetric and only half of it including the diagonal is included in the feature vector. Figure 2 shows a sample microscopic image of a human cell and its histogram.



(a)



(b)

Figure 2. *A sample of cell microscopic image and its histogram. (a) Input image. (b) Histogram.*

Therefore, the feature vector is a 17 dimensional vector, two coming from the grayscale mean and median, three coming from the histogram (hue, intensity and saturation), five coming from the covariance and seven coming from grayscale DCT. Table 1 shows the different extracted features and their length.

Table 1. *Texture and grayscale extracted features.*

Feature Row	Length	Feature Name
1	1	Grayscale mean
2	1	Grayscale median
3	1	Hist. Hue mean
4	1	Hist. Intensity mean
5	1	Hist. Saturation mean
6-10	5	Covariance
11-17	7	Grayscale DCT

We can combine grayscale and texture features into one feature vector, but it is necessary to define a confidence value for this type of classification. To achieve the optimal combination method we use the statistical model titled "feature selection" by weka software [16]. It takes into account the dependencies between the grayscale feature and the texture feature.

Figure 3 shows the block diagram of our proposed method to human cell detection in microscopic image.

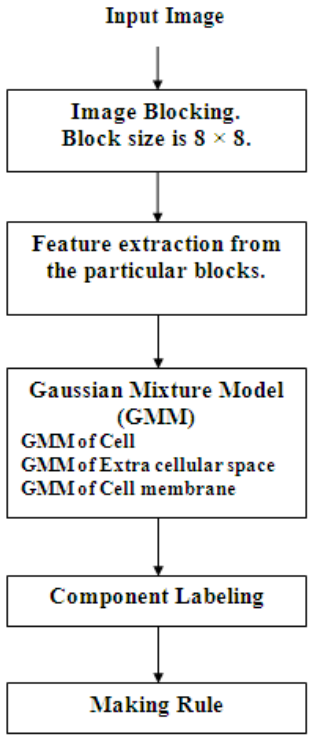


Figure 3. *Block diagram of our proposed method.*

Our proposed system has been implemented by C++ (C++ Builder XE2) and simulated by MATLAB R2009a, which have the enough strength to work with the digital images feature extraction and decision making systems.

5. Experimental Results

Here, we analyze and evaluate some of our results by applying our proposed method on real microscopic images. To prove the performance of our proposed GMM algorithm for human cell detection in microscopic image, we used 120 normal microscopic images of human cell from our research laboratory. The output of GMM is labeled with using bays rule and maximum likelihood. Each block is mapped to one of the three different groups which are labeled as 0, .5 and 1 indicating cell, extra cellular space and cell membrane respectively. The following materials were used in our experiments.

Human cell images were obtained with using Hitachi S-4500 SEM electron microscope on August to November 2011 and the size of images was 400 × 400 pixels. Our proposed system has been implemented by C++ (C++ Builder XE2) and simulated by MATLAB R2009a. We ran our system on a computer with Core2 Due 3.2 GHZ CPU and 8GB DDR2 RAM.

Two experiments will be presented in this section to show the implementation and the results of our proposed system. These are the True and False detection rates and the visual quality. Table 2 shows the corresponding values which we used for calculating the detection rate of our algorithm.

Table 2. *Values for calculating the detection rate.*

Values	Description
TC	True Cell Detection
TCM	True Cell Membrane Detection
FC	False Cell Detection
FCM	False Cell Membrane Detection

The general accuracy or true classification percentage (TCP) can be obtained as follow [….]:

$$\frac{TC+TCM}{TC+FC+TCM+FCM} \times 100 \qquad (5)$$

Hence, the false detection rate related to cell and cell membrane could be obtained by the following equations:

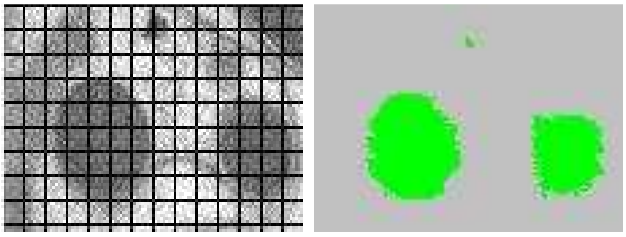
$$\text{False Cell Detection}=\frac{FC}{TC+FC+TCM+FCM} \times 100 \qquad (6)$$

Table 3 presents the results of different block size with their corresponding true and false detection rate.

Table 3. *True and False detections according to different block sizes.*

	Block Size 4×4	Block Size 8×8	Block Size 16×16
TCP	67.49	91.23	83.54
FC	4.41	1.82	3.02

Figure 4 shows the result of cell detection in 8×8 block sizes with its input image. This is clearly shown the visual quality of our proposed system. Green color indicates cells.



(a) (b)

Figure 4. *(a) Input image. (b) Result of cell detection.*

6. Conclusion and Future Work

In this paper we have used DCT and GMM for texture analysis of microscopic images to human cell detection. 120 normal microscopic images of human cell from our research laboratory were used for features analysis and our proposed algorithm was tested over 70 images of human cell. Figure 4 shows one those images and the algorithm's result. A general view of the system is demonstrated in figure 3.

Experimental results obtained from our algorithm shown that this model is both accurate and fast with around 91.23 percent of detection.

There are some aspects that can be improved in this method. For instance, we can take into neural networks with self learning ability to improve the accuracy and time consumption of these cases. We also intend to design and deploy N-Tier reusable software architecture for cell segmentation and detection purpose [17], [18]. Another future direction would be to increase the robustness of the approach to better handling the different kind of noise which exists in any type of digital images.

References

[1] Musen, Mark A., Blackford Middleton, and Robert A. Greenes. "Clinical decision-support systems." *Biomedical informatics*. Springer London, 2014. 643-674.

[2] Nogueira, Pedro A., and Luís Filipe Teófilo. "Automatic analysis of Leishmania infected microscopy images via Gaussian mixture models." *Advances in Artificial Intelligence-SBIA 2012*. Springer Berlin Heidelberg, 2012. 82-91.

[3] Mathivanan, Suresh, Hong Ji, and Richard J. Simpson. "Exosomes: extracellular organelles important in intercellular communication." *Journal of proteomics* 73.10 (2010): 1907-1920.

[4] Camazine, Scott. *Self-organization in biological systems*. Princeton University Press, 2003.

[5] K Wu, D Gauthier, MD Levine, "Live cell image segmentation," IEEE Transaction on Biomedical Engineering, vol 42, pp. 1-12, 1995.

[6] HS Wu, J Berba, J Gil, "Iterative thresholding for segmentation of cells from noisy images," Journal of Microscopy, vol 197, pp. 296-304, 2000.

[7] Tafti, A. P., Naji, H. R. Malakooti, M. V., “An Efficient Algorithm for Human Cell Detection in Electron Microscope Images based on Cluster Analysis and Vector Quantization,” Second International Conference on Digital Information and Communication Technology and it's Applications (DICTAP), pp. 125 – 129. IEEE, Bangkok, 2012.

[8] AA Aly, SB Deris, N Zaki, “Research review for digital image segmentation techniques,”International journal of compute science & Information technology, vol3, No 5, 2011.

[9] Bovik, A.C., The essential guide to image processing, Academic Press, 2009.

[10] McNicholas, Paul D., and Thomas Brendan Murphy. "Model-based clustering of microarray expression data via latent Gaussian mixture models." *Bioinformatics* 26.21 (2010): 2705-2712.

[11] SH Han, E Ackerstaff, R Stoyanova, S Carlin, W. Huang, J. A. Koutcher, J. K. Kim, G. Cho, G. Jang, and H. Cho, “Gaussian mixture model-based classification of dynamic contrast enhanced MRI data for identifying diverse tumor microenvironments: preliminary results, ” NMR in Biomedicine, vol 26, pp. 519-532, 2013.

[12] P Mayorga, C Druzgalski, RL Morelos, O. H. Gonzalez, and J. Vidales, “Acoustics based assessment of respiratory diseases using GMM classification,”Annual international conference of the IEEE Engineering in Medicine and Biology Society , Buenos Aires, 2010.

[13] Pun, Chi-Man, and Hong-Min Zhu. "Image Segmentation Using Discrete Cosine Texture Feature." *Rn* 1 (2010): 1.

[14] Rubel, Aleksey, Vladimir Lukin, and Oleksiy Pogrebnyak. "Efficiency of DCT-Based Denoising Techniques Applied to Texture Images." *Pattern Recognition*. Springer International Publishing, 2014. 261-270.

[15] Glob GH, Van Loan, Matrix computation. 2nd ed. Baltimore, Johns Hopkins University Press, 1989.

[16] http://www.cs.waikato.ac.nz/ml/weka/

[17] Tafti, A. P., Rohani, F., Moghadasifar, M., “Towards a scalable G2G framework for customs information system through N-Tier architecture,” International Conference on e-Learning and e-Technologies in Education , pp. 175 – 179. IEEE, Lodz, 2012.

[18] Tafti, A. P., Janosepah, S., Modiri, N., “Development of a Framework for Applying ASYCUDA System with N-Tier Application Architecture,” Second International Conference, ICSECS, Part III. Communications in Computer and Information Science, vol 181, pp. 533-541. Springer, 2011.