Songklanakarin J. Sci. Technol.
32 (3), 231-239, May - Jun. 2010

Songklanakarin Journal of Science and Technology
http://www.sjst.psu.ac.th

Original Article

Evaluation of color based breast cancer cell images analysis

Chalit Primkhajeepong[1]*, Pornchai Phukpattaranont[1], Somchai Limsiroratana[2], Pleumjit Boonyaphiphat[3] and Kanita Kayasut[3]

[1] *Department of Electrical Engineering,*

[2] *Department of Computer Engineering, Faculty of Engineering,*

[3] *Department of Pathology, Faculty of Medicine,*
Prince of Songkla University, Hat Yai, Songkhla, 90112 Thailand.

Received 8 December 2009; Accepted 4 May 2010

Abstract

This paper describes a simple yet effective algorithm for automatically counting stained breast cancer cell images based on color contents. The procedure for the approach consists of four steps. First, the cancer cell image in red-green-blue (RGB) color space is transformed with Haar wavelet. Second, the wavelet transformed image is changed to gray-scale image. Next, the gray-scale image is segmented with global thresholding using Otsu's method and morphological operations using opening, region filling, border clearing, and watershed segmentation. Third, the wavelet transformed image is changed to CIE$L^*a^*b^*$ color space. The feature, i.e. average value of b^* of each isolated cell, is extracted. Finally, the classification is applied by using the extracted feature. If the average value of b^* is positive that indicates yellow, the cancer cell is a positive cell. In addition, if the average value of b^* is negative that indicates blue, the cancer cell is a negative cell. Results show that the classified cancer cells by the proposed algorithm are in good agreement with the expert perception. In addition, the algorithm is practical to be used by a pathologist due to its simplicity and effectiveness.

Keywords: quantitative immunohistophathology, biomedical image processing, image segmentation, wavelet transformation, morphological operation

1. Introduction

Breast cancer is the most common cancer found in females worldwide and it is the second one for Thai women. In the pathological analysis of the breast cancer, the suspect tissue biopsy or fine needle aspiration is first analyzed by the pathologist visualization. If the cancer is beginning from the duct, estrogen receptor (ER) is stained by using immunohistochemistry (IHC). It is a prognostic marker for breast cancer detection. Evaluation of ER is useful for hormone therapy. Therefore, the quantitative report of ER is more significant than semi-quantitative or qualitative report. Traditionally, the evaluation of ER is assessed by the pathologist visualization. It is time consuming, costly, and tedious. A computer-aided system of image analysis is used in order to overcome these problems.

Previously, the analysis and evaluation of ER in breast cancer cell image was done by using Photoshop (Adobe Systems Inc., USA) (Lehr *et al*., 1997; Modifi *et al*., 2003) and WinROOF (Mitani Co., Ltd., Japan) (Hatanaka *et al*., 2003), which are commercial software. They are costly, providing not an automatic analysis, and they are not specific software for breast cancer cell image analysis. After that, an automatic algorithm is developed to analyze the cancer cell image by

* Corresponding author.
Email address: chalit.pr@gmail.com

using the image processing algorithms. Liu *et al*. (2007) presented the microscopic image analysis algorithm consisting of adaptive threshold segmentation based on the RGB and HIS color spaces, feature extraction using canny operator and eight-chain code tracking, and classification using morphological operations and colourometry. It can efficiently segment cell images and receive higher accuracy of cancer cell diagnosis. Phukpattaranont *et al*. (2008) presented an evaluation of computer-aided analysis for nuclear stained breast cancer cell images. The algorithm for image analysis consists of segmentation using neural network and mathematical morphology, with feature is consisting of an average value of L^*, a^*, b^* of each cell are extracted, and classification using Euclidian distance in CIE $L^*a^*b^*$ color space. However, the segmentation procedure is not automatic because the training data are required to be selected by a user.

In this paper we present an algorithm for automatically analyze the breast cancer cell image and its performance evaluation compared with an expert. The advantages of the proposed method against the conventional methods are its capability of automatic counting of the cancer cells and its confirmation of performance by the comparison with the result from a pathologist. The automatic cell counting technique is very important from a practical view point. In other words, each slide of breast cancer cell tissue is composed of 100-200 digital images. It is not convenient and it is time consuming if each digital image needs an input from a user when it is counted by the algorithm like cell counting technique based on neural network technique by Phukpattaranont *et al*. (2008).

Our algorithm consists of image preprocessing using wavelet transformation and color space transformation, segmentation using global thresholding and morphological operations, feature, i.e. average value of b^* of each cell is extracted, and classification using the extracted feature. This work is part of a system of computer-aided breast cancer cell image analysis that can automatically analyze and report the result to a pathologist. Our goal is that pathologists can use this tool as an assistance to make their image analysis and interpretation easier and more effective.

2. Breast Cancer Cell Images

Figure 1 shows an example of stained cancer cell image from microscope with a magnification of 40x. The stained cancer cells are classified into two categories according to their nuclei color contents, i.e. brown and blue. The brown and blue colors indicate the positive (P) and negative (N) results, respectively. In other words, the brown cells have an estrogen receptor, while the blue cells do not have the estrogen receptor in the cancer cell nuclear. The ratio of the total positive cells (P) and the total number of cancer cells (P+N) in the whole image is used by a doctor for medical planning and treatment.

3. Cancer Cell Image Analysis

Figure 2 shows the procedures of our algorithm for automatic cancer cell image analysis, i.e. the image preprocessing, segmentation, feature extraction, and classification. Details of each procedure are given below.

3.1 Image Preprocessing

In the image preprocessing procedure, the wavelet transformation (WT) and color space transformation are performed. The details of each stage are given below.

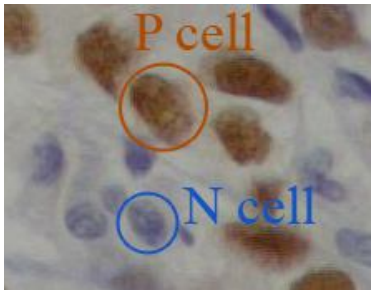

Figure 1. An example of stained cancer cell image. The brown and blue nuclei are representative samples of positive and negative staining of estrogen receptor of breast cancer cells, respectively.

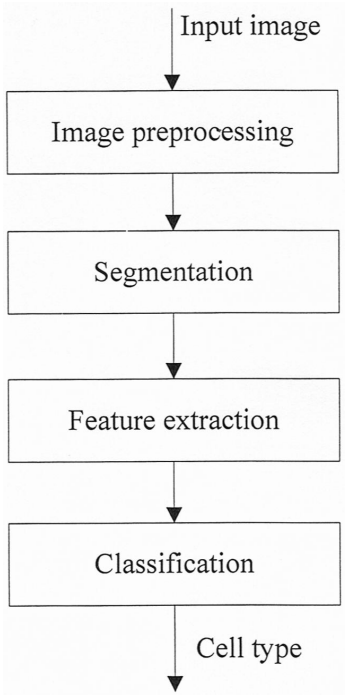

Figure 2. An automatic system for cancer cell image analysis.

3.1.1 Wavelet Transformation

The wavelet transformation is a useful image processing tool for 2-D images (Chai *et al.*, 2006). There are two objectives for using the WT in our proposed algorithm. First, the down-sampling process in the WT reduces the image size. Therefore, computational complexity of the algorithm such as morphology and feature calculations can be significantly decreased. These also lead to the reduction in processing time. Second, in order to remove high frequency noise in the image, we retain only low frequency components for further processing. In addition, Haar wavelet was selected due to its simplicity and effectiveness. The Haar wavelet function $\psi(x)$ and scaling function $\varphi(x)$ can be described as (Misiti *et al.*, 2007)

$$\psi(x)=\begin{cases}1 & if \quad 0\le x<0.5\\ -1 & if \quad 0.5\le x<1\\ 0 & otherwise\end{cases}, \tag{1}$$

$$\varphi(x)=\begin{cases}1 & if \quad 0\le x<1\\ 0 & otherwise\end{cases}. \tag{2}$$

By using WT, the original image is passed through low-pass and high-pass filters to generate low-low (LL), low-high (LH), high-low (HL), and high-high (HH) sub-bands as shown as Figure 3. The LL sub-band image is used for the next process.

3.1.2 Color Space Transformation

The RGB image from WT stage is converted to gray-scale image. The conversion of RGB image to gray-scale image can be expressed as (McAndrew, 2004)

$$Gray=(0.299\times R)+(0.587\times G)+(0.114\times B), \tag{3}$$

where *Gray* is gray-scale image and *R*, *G*, *B* are red, green, blue components, respectively.

3.2 Segmentation

In the segmentation procedure, the global thresholding and morphological operations are performed. The details of each stage are given below.

3.2.1 Global Thresholding

The gray-scale image is performed by using the global thresholding. It is defined by (Awcock and Thomas, 1995)

$$g(x,y)=\begin{cases}1 & , \quad f(x,y)>T\\ 0 & , \quad f(x,y)\le T\end{cases}, \tag{4}$$

where $g(x, y)$ is the threshold image, $f(x, y)$ is the gray-level value at (x, y) pixel, and T is the threshold value from Otsu's method (Otsu, 1979).

3.2.2 Morphological Operations

After finishing the global thresholding that generates the black and white (BW) image, the morphological operations, i.e. the opening, region filling, border clearing, and watershed segmentation are performed.

a) The opening is used to remove large and small objects that are not cancer cells. The morphological opening is performed with disk shape structure element (Adam, 1993; Jones and Soille, 1996) of radius 2 on the BW image. Moreover, the morphological area opening is performed to remove the objects that are higher than 450 pixels and lower than 40 pixels.

b) The region filling is performed to fill the cell regions and holes by using morphological reconstruction (Soille, 2002).

c) The border clearing is performed to remove the imperfect cell on the image border by using morphological reconstruction (Soille, 2002).

d) The watershed segmentation is a segmentation technique used to divide the overlapping cells (Essafi *et al.*, 2006). It uses the Fernand Meyer's algorithm (Mayer, 1994) with cityblock distance transformation (Rosenfeld and Pfaltz, 1966). The cityblock distance transformation can be defined by

$$d_{cityblock}=|x_1-x_2|+|y_1-y_2|, \tag{5}$$

where $d_{cityblock}$ is the distance between pixel (x_1, y_1) and (x_2, y_2).

3.3 Feature Extraction

After all cancer cells in the image are successfully segmented, we compute the feature of each isolated cancer cell, i.e., the average values of b^* of CIE $L^*a^*b^*$. The CIE $L^*a^*b^*$ color space is designed to approximate human vision. It can be defined by

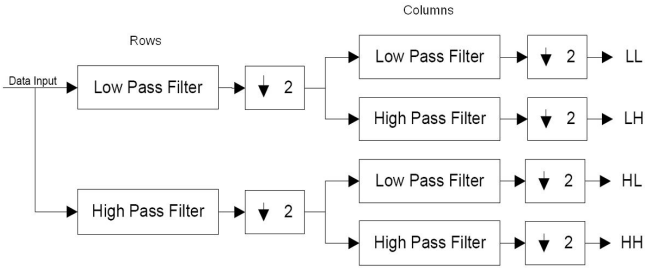

Figure 3. Wavelet transformation diagram.

$$L^* = 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16, \quad (6)$$

$$a^* = 500\left[\left(\frac{X}{X_n}\right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}}\right], \quad (7)$$

$$b^* = 200\left[\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - \left(\frac{Z}{Z_n}\right)^{\frac{1}{3}}\right], \quad (8)$$

for X/X_n, Y/Y_n, $Z/Z_n > 0.01$. The values X_n, Y_n, Z_n are the CIE (Commission International de l'Eclairage) tristimulus values of the reference white under the reference illumination, and X, Y, Z are the tristimulus values, which are mapped to the CIE color space. The L^* component represents intensity, which is the measure of brightness. $L^* = 0$ yields black and $L^* = 100$ indicates white. The values of a^* and b^* are in the range from -128 to 127. The a^* and b^* values are proportional to red-green and yellow-blue color contents, respectively. That is, for a^* negative values indicate green while positive values indicate red. For b^*, negative values indicate blue and positive values indicate yellow. A three-dimensional representation of CIE$L^*a^*b^*$ color space is shown in Figure 4 (Phukpattaranont *et al*., 2008).

The average value of b^* of each isolated cancer cell can be defined by

$$\overline{b_i^*} = \frac{\left(\sum x\right)_i}{n_i} - 127, \quad (9)$$

where $\overline{b_i^*}$ is the average value of b^* of the ith isolated cancer cell, $\left(\sum x\right)_i$ is the summary of pixel value in the ith isolated cancer cell, n_i is the area of the ith isolated cancer cell, and 127 is the normalization value.

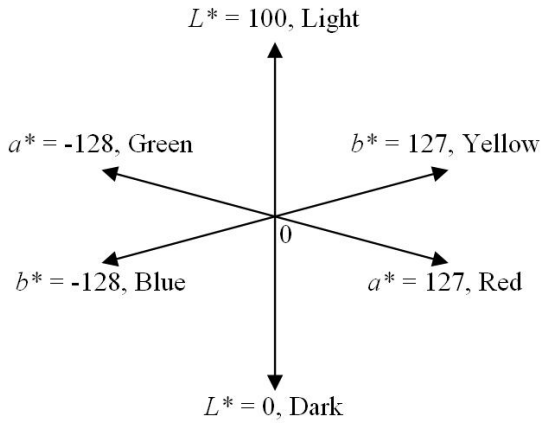

Figure 4. A three-dimensional representation of CIE $L^*a^*b^*$ color space.

3.4 Classification

In this procedure, we use the features based on color contents to classify cancer cells. The P cell is brown, which consists of yellow content and the N cell is blue, which consists of blue content. If the average of b^* is positive, the cancer cell is P cell. In addition, if the average of b^* is negative, the cancer cell is N cell. The cancer cell classification can be defined by

$$C_i = \begin{cases} P & , \quad if \quad \overline{b_i^*} \geq 0 \\ N & , \quad if \quad \overline{b_i^*} < 0 \end{cases}, \quad (10)$$

where C_i is the type of cancer cell, P is the positive cell, N is the negative, and is the average value of b^* of the ith isolated cancer cell.

4. Materials and Methods

4.1 Acquisition of the Images

The images used in this paper were cancer tissue stained with the antibody to ER in the nuclei. The tissue sections were observed under a microscope with a magnifying factor of 40x. The contrast and intensity of staining were manually corrected in such a way that the digitized image became visually acceptable for further manual classification by a specialist. Twenty-two cancer cell images were acquired in color using the Eclipse 80i advanced research microscope (Nikon Instech Co., Ltd., Japan). The digital images were saved as a 24-bits color 640x512 JPEG files.

4.2 Performance Evaluation

Two measures were defined in order to evaluate the performance of the algorithm in segmentation of nuclei from background. Sensitivity ($SS1$) is the probability that a nucleus will be detected in case it has been marked by the experts. It can be expressed as

$$SS1 = \frac{TP1}{TP1 + FN1}, \quad (11)$$

where $TP1$ (true positive) are the number of nuclei, which have been marked by the experts and also detected by the algorithm, and $FN1$ (false negative) are the number of nuclei, which are marked by the experts but are not detected by the algorithm. Positive predictive value ($PPV1$) is the probability that the detection of a nucleus is actually associated with a nucleus marked by the experts. It can be expressed as

$$PPV1 = \frac{TP1}{TP1 + FP1}, \quad (12)$$

where $FP1$ (false positive) are the number of nuclei, which are not marked by the exports but are detected by the algorithm.

In addition, four measures were defined in order to evaluate the performance of the algorithm for nucleus color classification after the nuclei were segmented. Sensitivity (*SS*2) is the probability that a P nucleus will be detected in case it has been marked positive by the experts. Positive predictive value (*PPV*2) is the probability that the detection of a P nucleus is actually associated with a P nucleus marked by the experts. Specificity (*SC*2) is the probability that a N nucleus will be detected in case it has been marked negative by the experts. Negative predictive value (*NPV*2) is the probability that the detection of an N nucleus is actually associated with a N nucleus marked by the experts. The *SS*2, *PPV*2, *SC*2, and *NPV*2 can be expressed as

$$SS2 = \frac{TP2}{TP2 + FN2}, \tag{13}$$

$$PPV2 = \frac{TP2}{TP2 + FP2}, \tag{14}$$

$$SC2 = \frac{TN2}{TN2 + FP2}, \tag{15}$$

$$NPV2 = \frac{TN2}{TN2 + FN2}, \tag{16}$$

where *TP*2 (true positive) are the number of nuclei, which the experts and the algorithm mark to P cells, *TN*2 (true negative) are the number of nuclei, which the experts and the algorithm mark to N cells, *FP*2 (false positive) are the number of nuclei, which the experts mark to N cells but the algorithm mark to P cells, and *FN*2 (false negative) are the number of nuclei, which the experts mark to P cells but the algorithm mark to N cells (Phukpattaranont *et al*., 2008).

5. Results and Discussion

Figure 5 is an example of an original image in RGB color space. In the image preprocessing procedure, the WT and color space transformation are performed. The original image is transformed with Haar wavelet family. The LL subband image from WT is shown in Figure 6(a). The WT image in RGB is converted to gray-scale color space as shown in Figure 6(b). In the segmentation procedure, the global thresholding and morphological operations are performed. The global thresholding is performed on gray-scale image by using Otsu's method. The histogram and threshold value from Otsu's method of gray-scale image is shown in Figure 6(c). The result from global thresholding is BW image as shown in Figure 6(d). Next, the morphological operations, i.e. opening, region filling, border clearing, and watershed segmentation are performed to remove the large and small objects that are not cell, fill the cell regions and holes, remove the imperfect cell in the image border, and segment the overlapping cells, respectively. The result from this procedure is shown in Figure 6(e). The comparison of the segmented result by our algorithm marked with contour lines and the expert marked with the rectangular windows is shown in Figure 6(f). The result from our algorithm is almost accurately comparing with the expert.

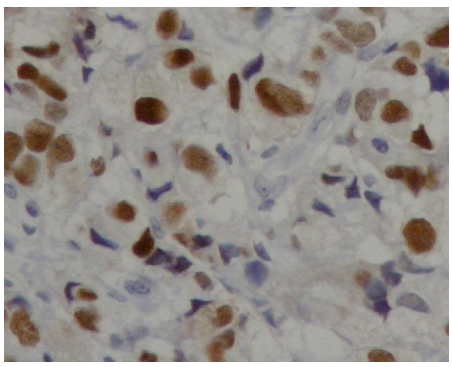

Figure 5. Original breast cancer cell image No. 10.

After all cancer cells are successfully segmented, the feature extraction will be performed. The WT image in RGB color space is converted to CIE *L*a*b** color space consisting of *L**, *a**, and *b**. Figure 7(a) shows a scatter plot of the average value of *a** and *b** of each isolated cell. The cancer cells are divided into 2 groups using the average value of *b**. In the classification procedure, the cancer cells are classified to P and N cells by using the extracted feature. If the average value of *b** is positive, the cancer cell is P cell. In addition, if the average value of *b** is negative, the cancer cell is N cell. Figure 7(b) shows the comparison of the classified result by the algorithm (P and N cells marked with green and red contour lines, respectively) and the expert (P and N cells marked with green and red rectangular windows, respectively).

The performance of the cancer cells segmentation of all 22 images is shown in Table 1. The average value of *SS*1 and *PPV*1 are 86±5 and 81±5, respectively. In addition, the performance of the cancer cells classification is shown in Table 2. The average value of *SS*2, *PPV*2, *SC*2, and *NPV*2 are 98±3, 99±1, 93±16, and 86±21, respectively.

From the extracted feature, it can accurately classify the cancer cells to P and N cells very well. There are some images where the performance is low, e.g. image No. 08 and image No. 15. The image No. 08 has *SC*2 = 40%. This means that the algorithm cannot detect N cell well in this image. Many N cells need to be manually added by an expert. Also, the image No. 15 has *NPV*2 = 38%. This indicates that the algorithm makes an over prediction of the N cells. Many N cells need to be manually erased by the expert. This is due to the fact that the average value of *b** of segmented cells is close to zero as shown in the scatter plot in Figure 8. At the same time, it is difficult to classify by the expert's visualization because its color is not clearly brown or blue. The threshold value or criteria should be revised under the

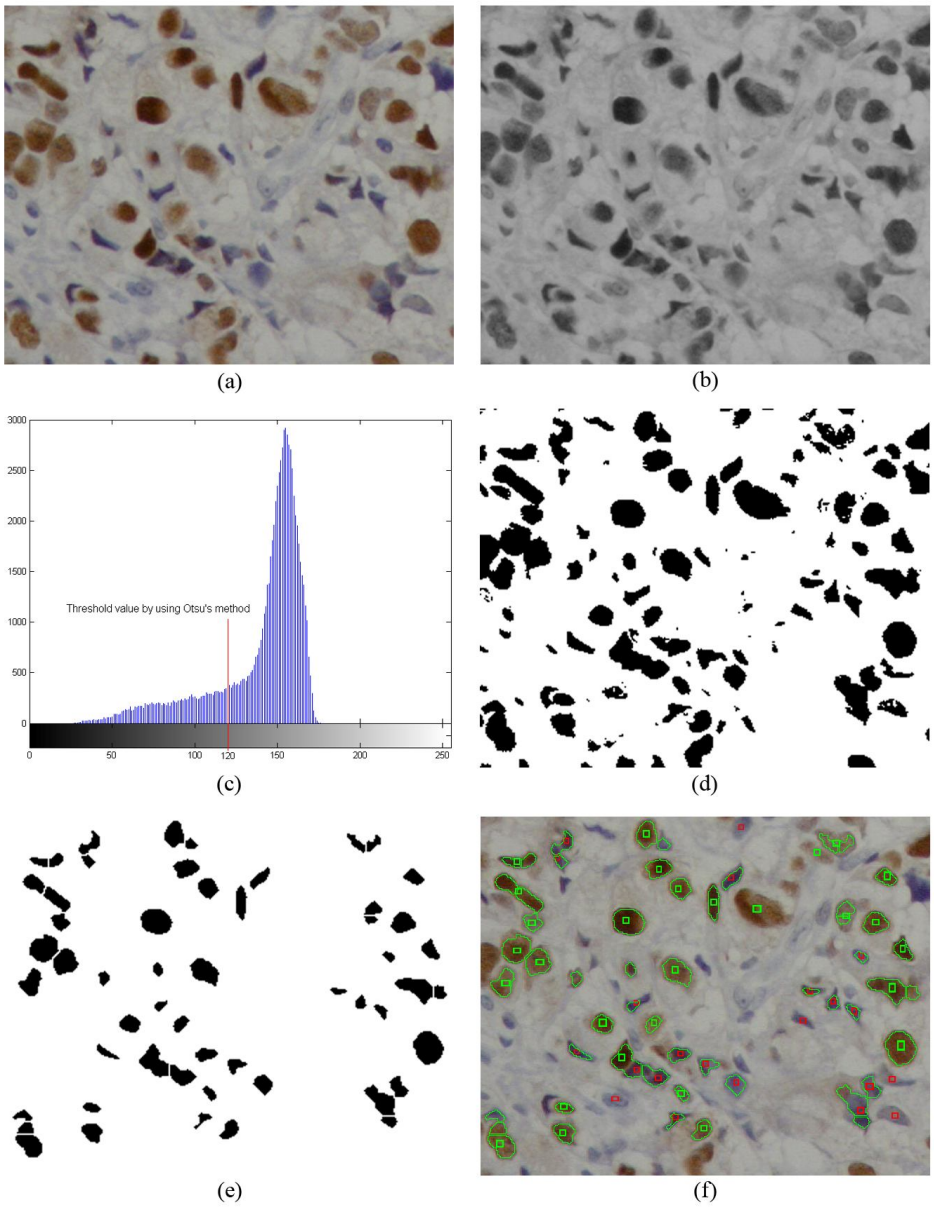

Figure 6. The results of the image processing and segmentation procedure, (a) wavelet transformed image, (b) gray-scale image, (c) histogram and threshold value by using Otsu's method, (d) BW image from using global thresholding, (e) BW image from using morphological operations, and (f) comparison of the segmented result by the algorithm (contour lines) and the expert (rectangular windows).

suggestion from the expert. For example, the majority vote of pixels in a single cell may be used to determine the cell type.

6. Conclusions

This paper presents an algorithm for automatic classification of cancer cells in a microscopic image. The procedure for the algorithm consists of image preprocessing, segmentation, feature extraction, and classification. Wavelet transformation and color space transformation are performed in image preprocessing procedure. Global thresholding and morphological operation are performed in segmentation procedure. The average value of b^* of CIE $L^*a^*b^*$ color space is extracted and used as a feature. Finally, comparison of average value of b^* is performed in classification procedure. If the average value of b^* is positive, the cancer cell is P cell. In addition, if the average value of b^* is negative, the cancer cell is N cell. Results show that the segmented and classified cancer cells by our algorithm are in good agreement with the expert perception.

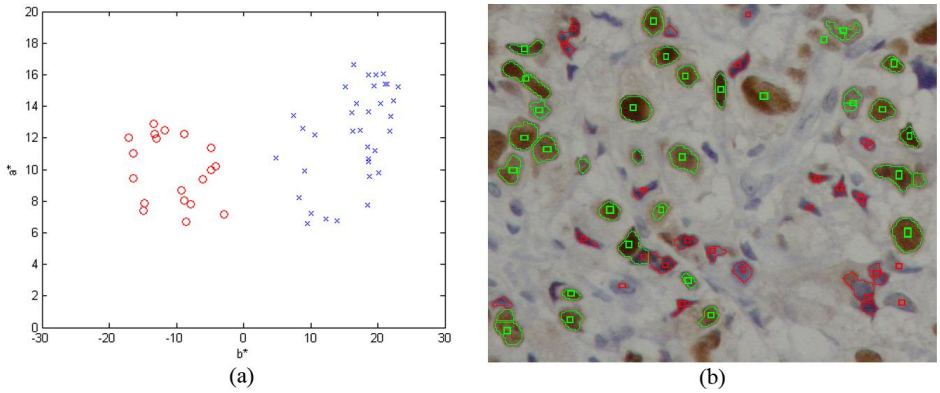

Figure 7. The results of classification procedure. (a) scatter plot between average value of a^* and b^* of each isolated cell. (b) comparison of the classified results by the algorithm (P and N cells marked with green and red contour lines, respectively) and the expert (P and N cells marked with green and red rectangular windows, respectively).

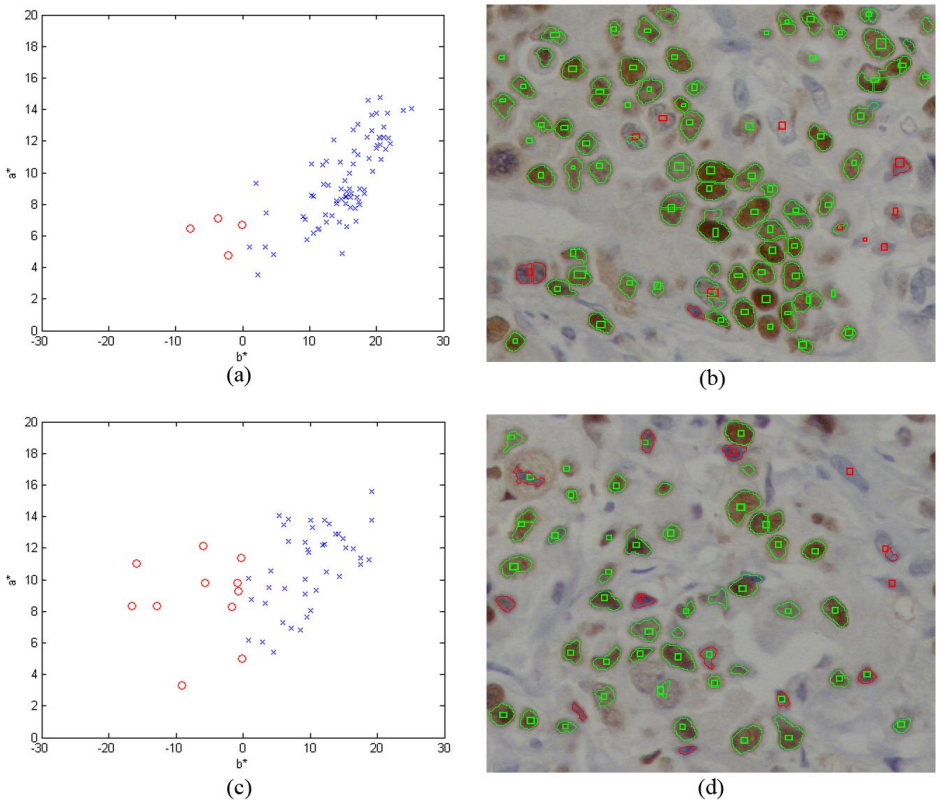

Figure 8. (a) scatter plot between average value of a^* and b^* of each isolated cell of image No. 08. (b) comparison of the classified results by the algorithm (P and N cells marked with green and red contour lines, respectively) and the expert (P and N cells marked with green and red rectangular windows, respectively) of image No. 08. (c) scatter plot between average value of a^* and b^* of each isolated cell of image No. 15. (d) comparison of the classified results by the algorithm (P and N cells marked with green and red contour lines, respectively) and the expert (P and N cells marked with green and red rectangular windows, respectively) of image No. 15.

Acknowledgements

This research is supported by the Department of Electrical Engineering and Department of Computer Engineering, Faculty of Engineering through contract no. ENG-51-2-7-02-0018-S, Department of Pathology, Faculty of Medicine, and Graduate School, Prince of Songkla University.

Table 1. Performance of cancer cell segmentation (Separation of cancer cells from background)

Image No.	*SS*1 (%)	*PPV*1 (%)
01	82	82
02	86	85
03	75	78
04	86	88
05	94	81
06	87	90
07	91	91
08	82	80
09	89	81
10	84	80
11	79	75
12	88	80
13	92	77
14	85	71
15	91	81
16	88	85
17	84	81
18	81	77
19	87	79
20	76	85
21	86	82
22	86	70
Average	82±5	81±5

Table 2. Performance of cancer cell classification (Separation of positive cancer cells from negative cancer cells).

Image No.	*SS*2 (%)	*PPV*2 (%)	*SC*2 (%)	*NPV*2 (%)
01	100	100	100	100
02	98	100	100	50
03	100	100	100	100
04	100	100	100	100
05	100	100	100	100
06	96	98	75	60
07	100	100	100	100
08	100	95	40	100
09	98	100	100	50
10	100	100	100	100
11	97	100	100	91
12	100	100	100	100
13	93	100	100	67
14	100	100	100	100
15	88	100	100	38
16	98	98	75	75
17	100	100	100	100
18	95	100	100	67
19	100	97	60	100
20	100	100	100	100
21	96	100	100	83
22	100	100	100	100
Average	98±3	99±1	93±16	86±21

References

Adams, R. 1993. Radial decomposition of discs and spheres. Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing. 55(5), 325-332.

Awcock G.J. and Thomas, R. 1995. Applied Image Processing, McGraw-Hill, Inc., New York, USA, pp. 144-146.

Chai, Y.H., Gao, L.Q., Lu, S. and Tian L. 2006. Wavelet-based watershed for image segmentation algorithm. Proceeding of the 6th World Congress on Intelligent Control and Automation, Dalian, China, June 21-23, 2006, 9595-9599.

Essafi, S., Doughri, R., M'Hiri, S., Romdhane, K.B. and Ghorbel, F. 2006. Segmentation and classification of breast cancer cells in histological images. Proceeding of the 2nd Information and Communication Technologies, Damascus, Syria, April 24-28, 2006, 1097-1102.

Hatanaka, Y., Hashizume, K., Nitta, K., Kato, T., Itoh, I. and Tani, Y. 2003. Cytometrical image analysis for immunohistochemical hormone receptor status in breast carcinomas. Pathology International. 53(10), 693-699.

Jones, R. and Soille, P. 1996. Periodic lines: definition, cascades, and application to granulometrie. Pattern Recognition Letters. 17(10), 1057-1063.

Lehr, H.A., Mankoff, D.A., Corwin, D., Santeusanio, G. and Gown, A.M. 1997. Application of Photoshop-based image analysis to quantification of hormone receptor expression in breast cancer. Journal of Histochemistry and Cytochemistry. 45(11), 1559-1565.

Liu, B., Yin, C., Liu, Z., Zhang, Z., Gao, J., Zhu, M., Gu, J. and Xu, K. 2007. Microscopic image analysis and recognition on pathological cells. Proceeding of Electrical and Computer Engineering, Canada, April 22-26, 2007, 1022-1025.

McAndrew, A. 2004. Introduction to Digital Image Processing with MATLAB. Thomson Leaning, Inc., Massachusetts, USA, pp. 378-379.

Meyer, F. 1994. Topographic distance and watershed lines. Signal Processing. 38(1), 113-125.

Misiti, M., Misiti, Y., Oppenheim, G. and Poggi, J.M. 2007. Wavelet Toolbox 4: User's Guide, The MathWorks, Inc., Massachusetts, USA.

Mofidi, R., Walsh, R., Ridgway, P.F., Crotty, T., McDermott, E.W., Keavny, T.V., Duffy, M.J., Hill, A.D.K. and O'Higgins, N. 2003. Objective measurement of breast cancer oestrogen receptor status through digital image analysis. Europeon Journal of Surgical Oncology. 29(1), 20-24.

O'Gorman, L., Sanderson, A.C. and Preston, K. 1985. A system for automated liver tissue image analysis: methods and results. The Institute of Electrical and Electronics Engineers Transactions on Biomedical Engineering. 32(9), 696-706.

Otsu, N. 1979. A threshold selection method from gray-level histograms. The Institute of Electrical and Electronics Engineers Transactions on Systems, Man, and Cybernetics Society. 9(1), 62-66.

Phukpattaranont, P., Limsiroratana, S. and Boonyaphiphat, P. 2008. Evaluation of computer-aided analysis for nuclear stained breast cancer cell images. Proceeding of the 3rd International Symposium on Biomedical Engineering, Bangkok, Thailand, November 10-11, 2008, 213-216.

Rosenfeld, A. and Pfaltz, J.L. 1966. Sequential operations in digital picture processing. The Association for Computing Machinery. 13(4), 471-494.

Soille, P. 2002. Morphological Image Analysis: Principles and Applications, Springer-Verlag, Germany, pp. 183-218.