

Computer-Aided System for Microscopic Images: Application to Breast Cancer Nuclei Counting

Pornchai Phukpattaranont*, Member,
Somchai Limsiroratana, and Pleumjit Boonyaphiphat,

ABSTRACT

We developed a computer-aided system for analyzing microscopic images called Cell Image Analyzer (CIA) and demonstrate its application to nuclear stained breast cancer cell counting. The algorithm for image analysis is composed of three steps. First, the cancer cells in the microscopic image are segmented based on neural network and mathematical morphology. Next, the features of each cell consisting of average values of L^* ; a^* , b^* in CIELab color space, circularity ratio, and area are extracted by the system. Finally, in order to demonstrate the potential application of the system, the classification of breast cancer cell nuclei is performed using the Euclidean distance of selected features, i.e. the average values of L^* ; a^* , b^* . Results from our computer-aided analysis system on a microscopic image from breast cancer show that the cancer cells are appropriately segmented. The classifications of segmented cell type based on the Euclidean distance in CIELab color space agree with visual inspection very well. Quantitative evaluations of our computer-aided analysis with the expert also provide similar agreement to image visualization. In other words, sensitivity and positive predictive value of cell segmentation are 88% and 82%, respectively. Moreover, sensitivity, positive predictive value, specificity, and negative predictive value of color classification are 94%, 99%, 91%, and 78%, respectively.

Keywords: Quantitative immunohistopathology, Image segmentation, Cancer cell images

1. INTRODUCTION

Microscopic images from a variety of techniques such as immunohistochemistry are of a great value for medical diagnosis. Immunohistochemistry is a technique used for detecting *in situ* a tissue antigen by a

specific antibody. An antigen-antibody reaction is visualized by the color development of specific dye and can be seen by light microscope. The tissue antigen is presented at any part of the cell, i.e., cell membrane, cytoplasm or nucleus. Therefore, it is a useful technique to demonstrate the protein markers including cancer cell. Estrogen receptor (ER) and progesterone receptor (PR) are prognostic markers for breast cancer detected by this method. Evaluation of ER and PR positive cells are useful for hormonal therapy. Brown nuclear stains are recognized as positive cells and blue ones are negative cells. Traditionally, the percentage of positive cells of those markers is semi-quantitatively counted. It is time consuming, costly, subjective and tedious. In order to overcome these problems, a computer-aided system of image analysis that previously requires manual operations is performed on the basis of the developments in computer capabilities and image processing algorithms [1] [2] [3].

There are a number of benefits that result from a computer-aided analysis. These include an acceleration of the process, a reduction in cost for image analysis, as well as a decrement in a false inspection due to fatigue. Additionally, the automated analysis provides a quantitative description. Based on this quantitative measurement, the analysis result is objective. Furthermore, the correlation of the quantitative categorization with patient symptoms may allow for an automated diagnostic system [4]. However, it is not expected that computer-aided image analysis will replace pathologist's experience. It is only an aid to the pathologist for the repeated routine work and yields quantitative results that complement and enhance interpretations by pathologists. Visual examination by the pathologist is still required where unusual or abnormal cells that the method is not trained to deal with are encountered.

We have recently proposed a new strategy for segmenting cancer cells in a microscopic image of immunohistological nuclear staining of breast cancer tissue based on the color of pixel [5]. This is motivated from the way that a pathologist determines the positive and negative of tumor cells by using their color contents manually. In [6] we develop a computer-aided system, which is capable of analyzing microscopic images. Our goal is that pathologists can use

Manuscript received on March 26, 2009 ; revised on July 9, 2009.

* Corresponding author.

P. Phukpattaranont is with Department of Electrical Engineering, Prince of Songkla University, Thailand

E-mail address: pornchai.pepsu.ac.th

S. Limsiroratana is with Department of Computer Engineering, Prince of Songkla University, Thailand

P. Boonyaphiphat is with Department of Pathology, Prince of Songkla University, Thailand

this tool as an assistance to make their image analysis and interpretation easier and more effective.

Based on our knowledge from the previous work, a user friendly interface computer-aided system called Cell Image Analyzer (CIA) is developed. We hope that the CIA will be routinely used by a pathologist in the future. The capabilities of the computer-aided analysis tool consist of the separation of cells from backgrounds, the extraction of dominant features from each cell (color, circularity ratio, and area) and the classification of cells based on the selected features. In this paper we demonstrate the potential of the CIA by applying it on nuclear stained breast cancer cell counting. Moreover, results of cell counting from the CIA were quantitatively evaluated with those from an expert.

2. CELL IMAGE ANALYSIS

Feature of cancer cells and details of computer-aided method for classifying cancer cells based on their color and shape analysis are described in this section. The procedure for the analysis approach is composed of three steps. First, the image is segmented in order to separate cancer cells from the background. Next, the feature extraction process is applied in order to extract the useful information from each segmented cell. Finally, the classification is operated using the selected characteristics extracted from the previous stage. Details of each step are given below.

2.1 Features of cancer cells

Figure 1 shows an example of stained cancer cell image from microscope with a magnification of 400. This staining procedure is utilized to demonstrate the existing of estrogen or progesterone receptors in the breast cancer cells. In other words, stained cancer cells are classified into two categories according to their nuclear color contents, i.e. brown (P) and blue (N). The brown color indicates a positive staining while the blue one demonstrates a negative result. The brown and blue cells shown in figure 1 are representative samples of positive and negative staining of estrogen receptor of cancer cells, respectively. The ratio of the total number of positive cancer cells to the total number of cancer cells in the whole image is used by a doctor for medical planning and treatment. It is also noticeable in figure 1 that there are color and intensity variations both in the cancer cells and in the background (non-cancer cells). The goal of our research project is to develop a computer-aided system for counting positive and negative cancer cells on immunohistological stained slides from breast cancer tissue.

2.2 Image segmentation

The originally acquired image is in the red-green-blue (RGB) color space. That is, the color image is formed by the combination of red, green, and blue monochrome images. In the first step, we classify color pixels in the image into one of three categories, i.e. background, P, or N, based on their RGB components. There are many classifiers that can be used for partitioning color of a pixel. However, a neural network is chosen in this paper due to its well known as a successful classifier for many applications [7] [8]. The R-, G- and B-components of pixel values from background, P, and N reference regions (Boxes B1, B2, and B3) shown in figure 1 are used as an example demonstrating the input vectors for training neural network. After cells are separated from background, morphology operations are used for addressing the spatial characteristics of cells. Finally, in order to obtain accurate cell counting results, the marker-controlled watershed is applied for separating attached multiple cells into distinct single cells. Details of an algorithm for segmenting cancer cells based on their colors and sizes are given in [5].

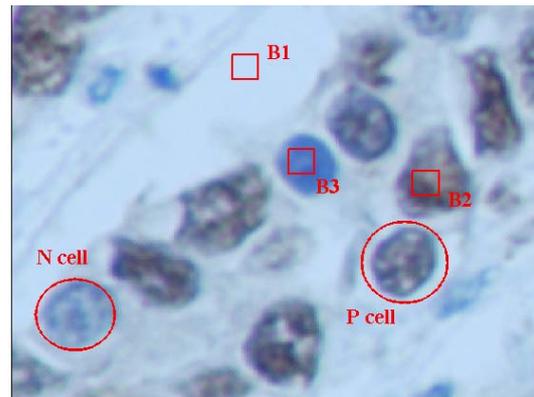


Fig.1: An example of stained cancer cell image. The brown and blue nuclei in the right and left circles are representative samples of positive and negative staining of estrogen receptor of breast cancer cells, respectively. Boxes **B1**, **B2**, and **B3** indicate reference regions from background, P, and N, respectively. Pixel values in the box regions are used for training neural network.

2.3 Feature extraction

After all cancer cells in the image are successfully segmented, the features of each isolated cancer cell consisting of average values of L^* ; a^* , b^* , circularity ratio, and area, are computed. Firstly, average values of L^* , a^* , b^* from the CIELab space are calculated. The CIELab space is designed to approximate human vision. It is chosen due to the close correlation between its Euclidean distances and human perception

of colors. The CIELab space can be defined by [9]

$$\begin{aligned} L^* &= 116 \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - 16 \\ a^* &= 500 \left[\left(\frac{X}{X_n} \right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} \right] \\ b^* &= 200 \left[\left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - \left(\frac{Z}{Z_n} \right)^{\frac{1}{3}} \right] \end{aligned} \quad (1)$$

for $X/X_n, Y/Y_n, Z/Z_n > 0.01$. The values X_n, Y_n, Z_n are the CIE (Commission Internationale de l'éclairage) 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 CIELab color space is shown in figure 2. Secondly, the circularity ratio is computed. The circularity ratio is the ratio of the area of the shape to the area of a circle having the same perimeter, which is expressed mathematically as

$$M = (4\pi \times \text{area}) / (\text{perimeter})^2 \quad (2)$$

M is equal to one for a circle and it is less than one for any other shape. For example, it is $\pi/4$ for a square and it is zero for an infinitely long and narrow shape. Thirdly, the area is obtained using the total number of pixels in each separated cell.

2.4 Classification

In this step, we use the features based on color components to classify cancer cells. The average values of L_r^* , a_r^* and b_r^* from P and N representative cells are selected to serve as the reference markers. The relative perceptual differences between reference colors and the color of the i -th cell under consideration in CIELab color space can be approximated by treating each color as a point in a three dimensional space and taking the Euclidean distance between them. The Euclidean distance in CIELab color space, ΔE_{ab}^* , can be expressed as

$$\Delta E_{ab}^* = \sqrt{(L_i^* - L_r^*)^2 + (a_i^* - a_r^*)^2 + (b_i^* - b_r^*)^2} \quad (3)$$

The Euclidean distances between the average values of L^* , a^* and b^* of the considering cell and those of all reference markers are compared. The smaller distance will tell us that the cell is more closely matches that reference marker. For example, if the distance between the considering cell and the

N reference marker is the smaller, then it would be labeled as the N cancer cell.

3. MATERIALS AND METHODS

3.1 Acquisition of the images

images used in this paper were breast cancer tissue stained with the antibody to ER in the nuclei. The positive cells were marked by brown color and the negative ones were blue color. In addition, the inflammatory cells such as lymphocytes, plasma cells and some connective tissue cells were labeled with blue color in the nuclei. However, these non-tumor cells were identified by their contours and pattern histologic arrangements. The tissue sections were observed under a microscope with a magnifying factor of 400. 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. The cell images were acquired in color using the Eclipse 80i advanced research microscope (Nikon Instech Co., Ltd., Japan). The digital image was saved as a color 2560×3200 JPEG files for processing. The total numbers of 323 nuclei from six different images were used in evaluation.

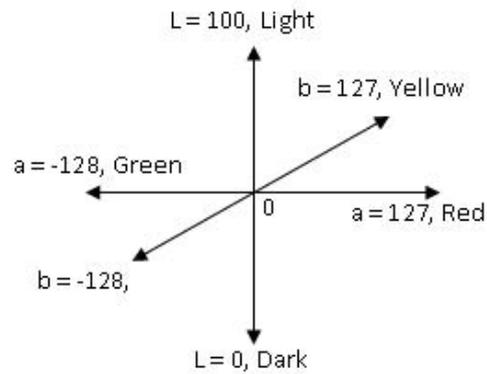


Fig.2: A three-dimensional representation of CIELab color space.

3.2 Performance evaluation

Two measures were defined in order to evaluate the performance of the algorithm in the 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 is given by

$$SS1 = \frac{TP1}{TP1 + FN1}, \quad (4)$$

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 have been marked 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. The $PPV1$ can be expressed as

$$PPV1 = \frac{TP1}{TP1 + FP1}, \quad (5)$$

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

In addition, there are four measures were defined in order to evaluate the performance of the algorithm for nucleus color classification after nuclei were correctly segmented. Sensitivity ($SS2$) is the probability that a P nucleus will be detected in case it has been marked positive by the experts. Positive predictive value ($PPV2$) is the probability that the detection of a P nucleus is actually associated with a P nucleus marked by the experts. Specificity ($SC2$) is the probability that an N nucleus will be detected in case it has been marked negative by the experts. Negative predictive value ($NPV2$) is the probability that the detection of an N nucleus is actually associated with a negative nucleus marked by the experts. The $SS2$, $PPV2$, $SC2$, $NPV2$ can be expressed as

$$SS2 = \frac{TP2}{TP2 + FN2}, \quad (6)$$

$$PPV2 = \frac{TP2}{TP2 + FP2}, \quad (7)$$

$$SC2 = \frac{TN2}{TN2 + FP2}, \quad (8)$$

$$NPV2 = \frac{TN2}{TN2 + FN2}, \quad (9)$$

4. RESULTS AND DISCUSSION

4.1 Cell image analysis

Figure 3 shows the segmented image superimposed on the original RGB image resulting from the algorithm described in Section 2. Details of this result can be found in [5]. Note that we discard cells at the boundary of image because of their partial appearance. The total number of segmented cancer cells is 24. It can be clearly seen that the perception of segmented cancer cells is in agreement with their original visualization.

Table 1 shows the features extracted from each cell consisting of the average values of L^* , a^* , b^* , the area and circularity ratio. The maximum and minimum values of each feature are shown using bold typeface. The cell number 11 has the largest area while the cell number 6 has the smallest area. The cell number 14 has the most similar shape to a circle but the cell number 1 has the most different shape from a circle. The Euclidean distances in CIELab color space

Table 1: FEATURES EXTRACTED FROM SEGMENTED CANCER CELLS

No.	L	a^*	b^*	Area	M	D1 ¹	D2 ²	Type
1	59	-1	-20	4,144	0.53	18	28	P
2	53	3	-22	5,518	0.72	14	24	P
3	45	2	-16	4,758	0.67	5	30	P
4	52	2	-15	6,818	0.67	10	31	P
5	57	1	-18	8,372	0.74	15	29	P
6	60	-3	-37	2,542	0.83	31	15	N
7	61	-3	-27	5,054	0.75	24	23	N
8	42	5	-21	6,816	0.84	9	26	P
9	49	1	-20	7,461	0.73	10	26	P
10	45	4	-19	7,287	0.83	7	27	P
11	53	0	-17	8,418	0.77	11	29	P
12	48	2	-20	7,850	0.82	9	26	P
13	61	-2	-33	3,272	0.76	28	18	N
14	54	1	-42	4,567	0.89	32	6	N
15	50	1	-26	5,599	0.87	16	20	P
16	53	2	-18	2,597	0.71	12	28	P
17	53	1	-22	5,680	0.87	14	24	P
18	51	3	-16	6,812	0.78	9	30	P
19	61	-2	-26	4,408	0.73	23	24	P
20	49	3	-16	8,020	0.77	7	29	P
21	49	3	-19	6,878	0.57	9	27	P
22	56	0	-16	5,332	0.72	14	31	P
23	56	0	-18	5,900	0.6	15	28	P
24	51	1	-26	5,842	0.84	17	20	P

¹The average values of Lr^* , ar^* and br^* from the P representative region (box **B2**) are 43, 4, and -12, respectively.

²The average values of Lr^* , ar^* and br^* from the N representative region (box **B3**) are 49, 4, and -46, respectively.

of each cell from P and N representative regions are shown in column D1 and D2, respectively. Based on these two distances, the classification of cell type is shown in the last column. That is, the cell number 6, 7, 13, and 14 are classified as N cells while the others are P cells. Note that the Euclidean distances in CIELab color space of cell number 7 and 19 are very close. The threshold distance for classification can be adjusted by a specialist. In addition, area and circularity ratio are prepared for more quantitative assessment of cells in the future.

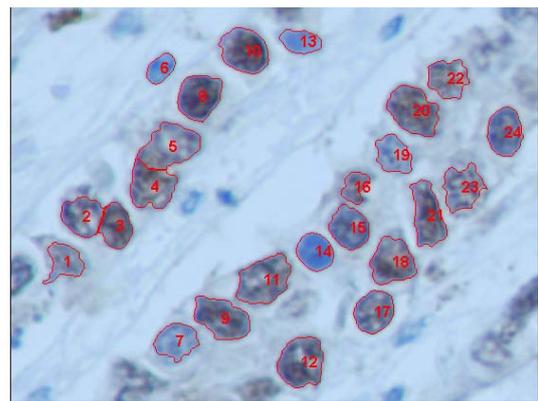


Fig. 3: Segmented image superimposed on the original RGB image.

4.2 Evaluation of cell counting

Figure 4 shows the CIA and its graphic user interface. The original image is shown in the left win-

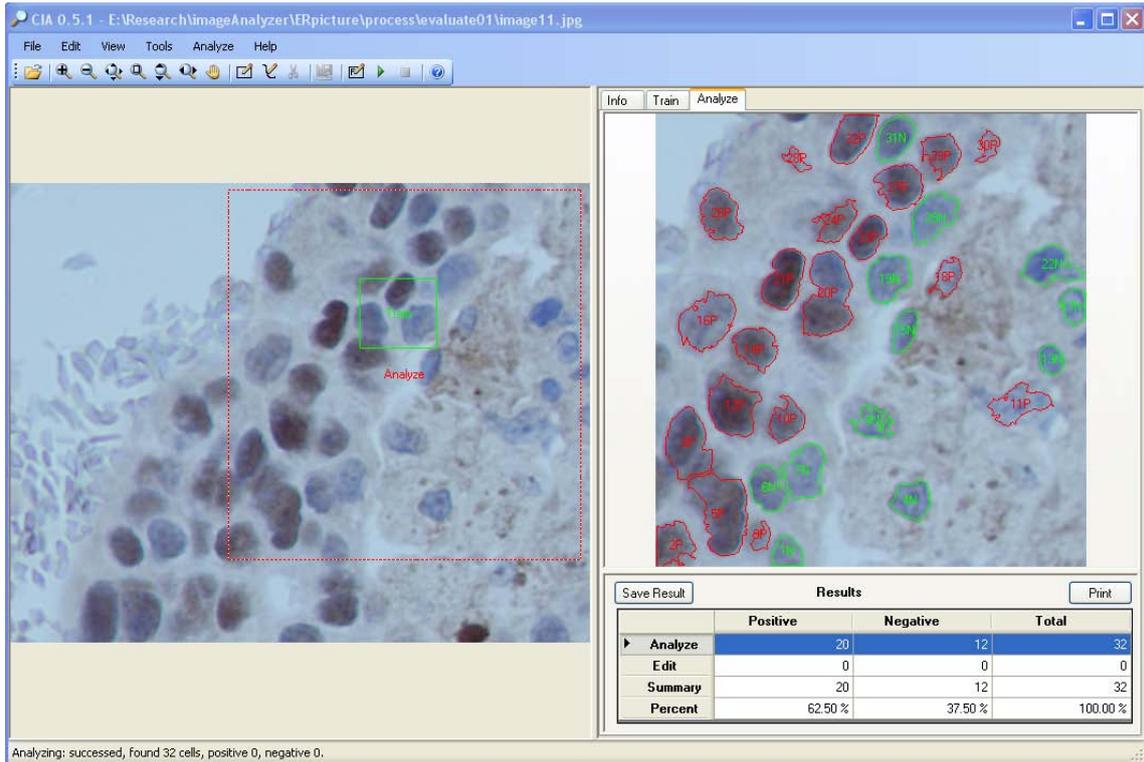


Fig.4: The CIA and its graphic user interface.

dow. The selected training region is inside the green rectangular area. After neural network training, the testing region inside the red rectangular area is chosen for analysis and cell counting results are shown in the right window. The numbers of total positive and negative cells are reported in the table under the analyzing result.

Figure 5 shows an example of the segmented image superimposed on the original RGB image of cancer cells. While nuclei with contour lines are marked by the algorithm, nuclei with rectangular windows are marked by the expert. Forty eight nuclei are marked correctly and shown in pink. Nine fault positive and three fault negative nuclei are shown in red contours and green rectangles, respectively. Moreover, SS1 and PPV1 are 94% and 85%, respectively.

We apply the CIA to the other five images. Results are reported in Table 2 and 3. Table 2 shows performance of the nucleus segmentation algorithm for all six images. The average values of SS1 and PPV1 are 88% and 82%, respectively. Table 3 shows performance of the nucleus color classification for all six images. The average values of SS2, PPV2, SC2, NPV2 are 94%, 99%, 91% and 78%, respectively.

5. CONCLUSIONS

We present our progress in the development of computer-aided system for microscopic images called Cell Image Analyzer (CIA) and the evaluation of the system for analyzing nuclear stained breast cancer

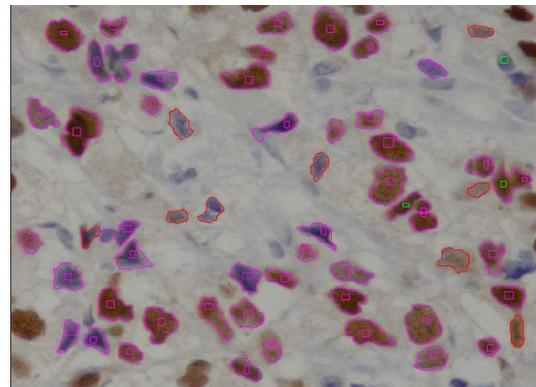


Fig.5: Segmented image superimposed on the original RGB image. Nuclei marked by the expert are shown with rectangular windows. Nuclei marked by the algorithm shown with contour lines.

Table 2: PERFORMANCE OF THE NUCLEUS SEGMENTATION ALGORITHM

Image No.	SS1(%)	PPV1(%)
1	94	85
2	85	95
3	81	93
4	96	68
5	83	96
6	86	55
Average	88	82

Table 3: PERFORMANCE OF THE NUCCLOUS COLOR CLASSIFICATION

Image No.	SS2(%)	PPV2(%)	SC2(%)	NPV2(%)
1	100	100	100	100
2	100	100	100	100
3	100	97	60	100
4	80	100	100	33
5	89	97	87	57
6	97	100	100	80
Average	94	99	91	78

cell images. Results show that all segmented nuclei are in agreement with the human perception very well. In addition, the classification of cell type based on the extracted features is feasible. These results can be used as guidance for improving the algorithm. In other words, the shape and area of cancer cells should be incorporated into more sophisticated classifiers in order to improve accuracy. More evaluation of the system by a pathologist after algorithm improvement will be reported in the near future.

6. ACKNOWLEDGMENT

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

References

- [1] J. P. Thiran and B. Macq, "Morphological feature extraction for the classification of digital images of cancerous tissues," *IEEE Trans. Biomed. Eng.*, vol. 43, no. 10, pp. 1011-1020, Oct. 1996.
- [2] B. Fang, W. Hsu, and M. L. Lee, "On the accurate counting of tumor cells," *IEEE Trans. Nanobiosci.*, vol. 2, no. 2, pp. 94 -103, Jun. 2003.
- [3] S. Petushi, C. Katsinis, C. Coward, F. Garcia, and A. Tozeren, "Automated identification of microstructures on histology slides," in *IEEE International Symposium on Biomedical Imaging: Macro to Nano*, 2004, vol. 1, pp. 424 -427.
- [4] L. O'Gorman, A. C. Sanderson, and K. Preston Jr., "A system for automated liver tissue image analysis: Methods and results," *IEEE Trans. Biomed. Eng.*, vol. 32, no. 9, pp. 696-706, Sep. 1985.
- [5] P. Phukpattaranont and P. Boonyaphiphat, "Color based segmentation of nuclear stained breast cancer cell images," *ECTI Transaction on Electrical Eng., Electronics, and Communication*, vol. 5, no. 2, pp. 158-164, Aug. 2007.
- [6] P. Phukpattaranont and P. Boonyaphiphat, "Computer-aided analysis of nuclear stained breast cancer cell images," in *Proceedings of 2008 ECTI International Conference (ECTI-*

CON 2008), Krabi, Thailand, May 15-16, 2008, vol. 1, pp. 485-488.

- [7] E. Gelenbe, Y. Feng, and K. R. R. Krishnan, "Neural network methods for volumetric magnetic resonance imaging of the human brain," *Proc. IEEE*, vol. 84, no. 10, pp. 1488-1496, Oct. 1996.
- [8] W. E. Reddick, G. O. Glass, E. N. Cook, T. D. Elkin, and R. J. Deaton, "Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural network," *IEEE Trans. Med. Imag.*, vol. 16, pp. 911-918, 1997.
- [9] H. J. Trussell, E. Saber, and M. Vrhel, "Color image processing (basics and special issue overview)," *IEEE signal processing magazine*, vol. 22, no. 1, pp. 14-22, Jan. 2005.



P. Phukpattaranont was born in Songkla, Thailand. He received the B. Eng. and M. Eng. degrees in electrical engineering from Prince of Songkla University in 1993 and 1997, respectively, the Ph.D. degree in electrical engineering from the University of Minnesota, in 2004. He is currently an assistant professor of electrical engineering at Prince of Songkla University. His research interests are ultrasound contrast imaging,

ultrasound signal processing, medical image processing, and biomedical signal processing. Dr. Phukpattaranont is a member of the IEEE.



S. Limsiroratana was born in Thailand. He received the B.Eng. degree in electrical engineering from Prince of Songkla University in 1991, the M.Arg and Dr.Arg degrees from division of environmental science and technology, Kyoto University in 2000 and 2005 respectively for the detection of fruits on natural background research. He has been working at Department of Computer Engineering, Prince of Songkla University since 1991. His research interests are agricultural image processing, medical image processing, digital watermarking and AI.

His research interests are agricultural image processing, medical image processing, digital watermarking and AI.



P. Boonyaphiphat was born in Thailand. She received the B.Sc. degree from Faculty of Nursing, Prince of Songkla University, in 1980, the M.Sc. degree from Faculty of Science, Mahidol University, in 1986. She has been working at Department of Pathology, Faculty of Medicine, Prince of Songkla University since 1986. Her position is an assistant professor in anatomical pathology. Her research field is head and neck cancer,

breast cancer and image analyzer for histopathology.