Nucleus Recognition of Uterine Cervical Pap-Smears using FCM Clustering Algorithm

Kwang-Baek Kim, *Member, KIMICS*

Abstract— Segmentation for the region of nucleus in the image of uterine cervical cytodiagnosis is known as the most difficult and important part in the automatic cervical cancer recognition system. In this paper, the region of nucleus is extracted from an image of uterine cervical cytodiagnosis using the HSI model. The characteristics of the nucleus are extracted from the analysis of morphemetric features, densitometric features, colormetric features, and textural features based on the detected region of nucleus area. The classification criterion of a nucleus is defined according to the standard categories of the Bethesda system. The fuzzy C-means clustering algorithm is employed to the extracted nucleus and the results show that the proposed method is efficient in nucleus recognition and uterine cervical Pap-Smears extraction.

Index Terms— Uterine Cervical Cytodiagnosis, Automatic Cervical Cancer Recognition System, HSI Model, Bethesda System, Fuzzy C-means Clustering Algorithm.

I. INTRODUCTION

Cervical cancer is one of the most frequently found diseases in Korean women but can be cured if detected early enough. Conquering the disease is a very important matter. Previous research shows that cervical cancer occupies 16.4 ~ 49.6% of malignant tumors in Korea and occupies 26.3 ~ 68.2% of malignant tumors in woman [1], [2]. The best method for a completely curing cervix cancer is to prevent the cell from developing into cervical cancer. For this purpose, there have been many efforts to completely or at least partially automate the process of cytodiagnosis during the last 40 years [3].

Diagnosis of the region of interest in a medical image is largely consisted of area segmentation, feature extraction and characteristic analysis. In area segmentation, a medical specialist detects abnormal regions of a medical image based on his expert knowledge. In feature extraction, features are extracted from the separated abnormal region. A medical doctor diagnoses a disease by using character analysis which is deciphering the extracted features to analyze and compare clinical information. Area segmentation methods used on a nucleus differs according to the target image. The approaches can be largely divided into the pixel-center method and the area-center method [4], [5]. Pixel-center method assigns an independent meaning to each pixel according to a predefined criterion. Pixel-center method can use the overall characteristic [4]. Although area-center method relatively needs more calculation time than the pixel-center method, it provides the usage of regional characteristics [5].

In this paper, the following simplification process allows the nucleus to be more easily detected in the image: (i) converting the extracted image of cervix uteri cytodiagnosis to a gray scaled image, (ii) removing noise using brightness information, and (iii) applying a 5×5 fuzzy gray morphology operation. To divide the nucleus area in the simplified image, the dark area of the nucleus is separated. Next, the following characteristic information is extracted: 13 morphometric features, and 8 densitometric features, 18 colorimetric features, and a textural feature. Extracted information is categorized into 4 degrees based on the extent of abnormality in each nucleus by using the fuzzy c-means clustering algorithm.

II. NUCLEUS AREA SEGMENTATION OF CERVIX UTERI CYTODIAGNOSIS

The proposed algorithm to extract the nucleus of cervix uteri cytodiagnosis is shown in Fig. 1.

In this paper, noise (leukocyte, etc.) is removed by using the information that the size of noise is smaller than that of the nucleus. Small-sized cervix uteri are normal and can therefore be removed because they do not have an influence on detecting an abnormal cell [6]. In order to analysis the changes and characteristics of a nucleus, 13 morphometric features, 8 densitometric features, 18 colorimetric features, and a textural feature are extracted after detecting the nucleus in cervix uteri. The classification criterion of extracted nucleus is defined according to the standard categories of the Bethesda system. The fuzzy c-means clustering algorithm is applied to classify the malignancy degree of the extracted nucleus according to the standard criterion.

Manuscript received January 29, 2007.

Kwang-Baek Kim is with the Division of Computer and Information Engineering, Silla University, Busan, 617-736, Korea (Tel: +82-51-999-5052, Fax: +82-51-999-5657, Email: gbkim@silla.ac.kr)

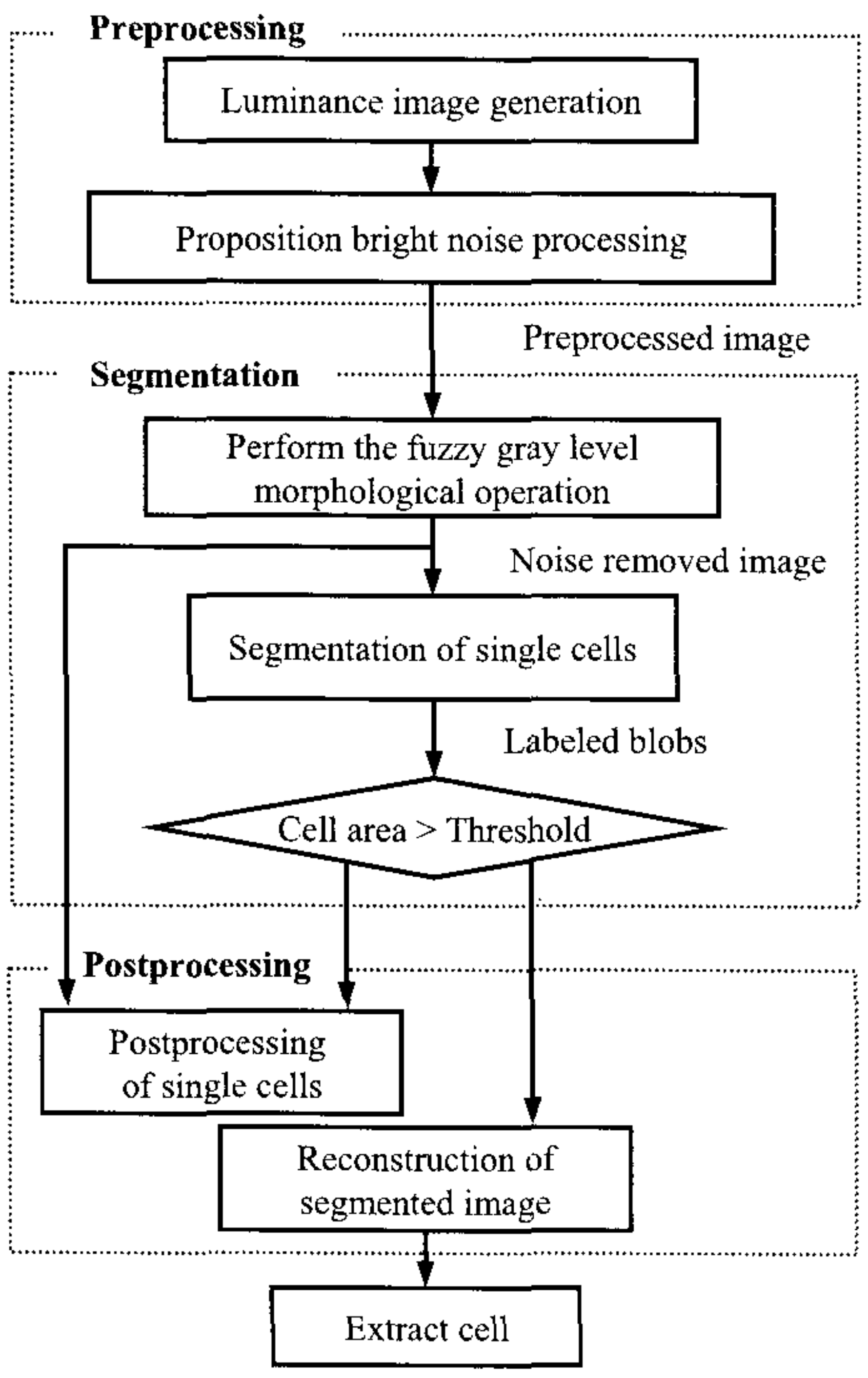


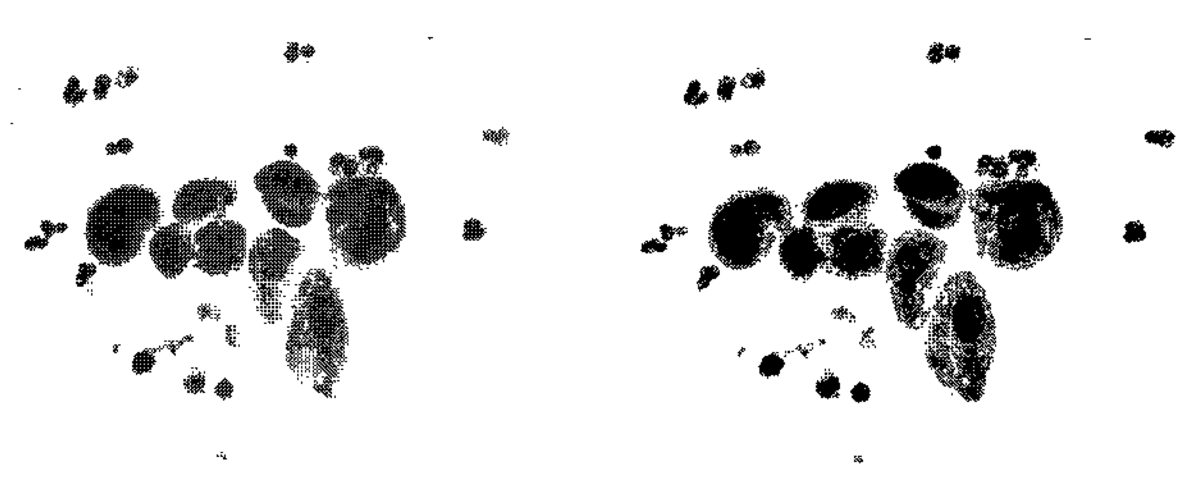
Fig. 1 Process to extract nucleus of cervix uteri cytodiagnosis

A. Noise exclusion using HSI information

In the pre-treatment process, color images are changed into grey images. To improve the quality of the image, noise is removed by using variation of brightness which can be acquired by p-tile stretching. After noise is removed from the image, it is divided into 8×8 blocks, in which Eq. (1) is applied to each block.

$$z' = \frac{b'-a'}{b-a}\times(z-a)+a' \qquad (1)$$

In Eq. (1), a and b are the values of brightness in the original image and z is $a \le z \le b$. Here, a is the lowest brightness value + (30% of the highest brightness value) and b is 80% of the highest brightness value. a' and b' are 0 and 255 respectively.



(a) (b)

Fig. 2 Image of cervix uteri cytodiagnosis with noise reduction: (a) Image of cervix uteri cytodiagnosis, (b) Image of cervix uteri cytodiagnosis which noise is removed

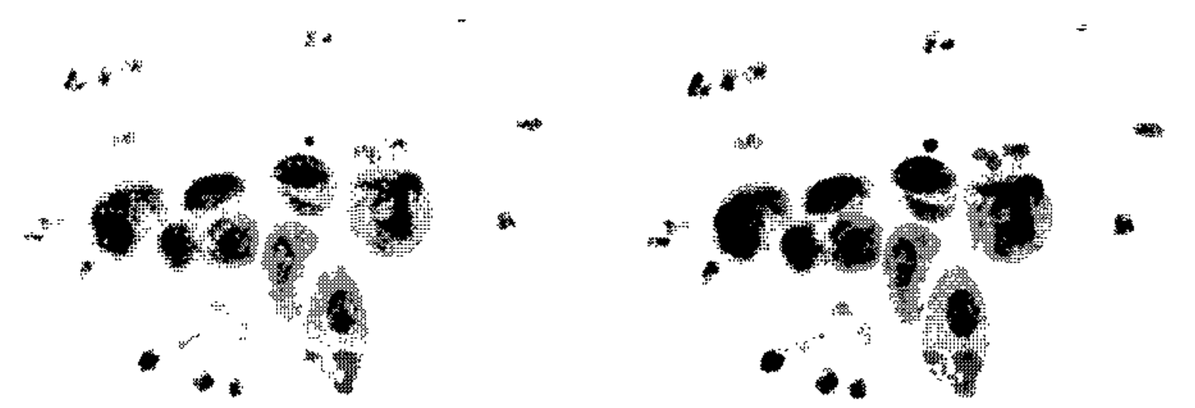
The image of cervix uteri cytodiagnosis that is used in this paper is in Fig. 2(a). The grey image in which noise is removed by the proposed pre-treatment process is in Fig. 2(b).

B. Morphology operation using fuzzy gray morphology

If noise is removed by the proposed pre-treatment process, partial information of the normal cell nucleus and cancer cell nucleus is lost, making it difficult to precisely extract the nucleus. Therefore, by applying a 5×5 fuzzy grey morphology operation to improve this problem, the extracted nucleus of the normal cell and abnormal cell can be precisely extracted. Fuzzy morphology operation is in Eq. (2) and (3). The resulting image that is applied to a 5×5 fuzzy grey Erosion · Dilation morphology operation is shown in Fig. 3. Here, a is the original image and b is the 5×5 mask.

$$A \ominus B = \{(x, \mu_{A\ominus B}(x)) \mid x \in E^N\},$$
$$\mu_{A\ominus B}(x) = \inf_{z\in E^N} \min\,[1, MRF(\mu_A(z), \mu_{(B:x)}(z))], \qquad (2)$$
$$MRF(a,b) = \begin{cases} 0 & \text{if } b=0 \\ a/b & \text{otherwise,} \end{cases}$$

$$A\oplus B = \{(x, \mu_{A\oplus B}(x)) \mid x \in E^N\},$$
$$\mu_{A\oplus B}(x) = \sup_{z\in E^N} [\mu_A(z)\times \mu_{(B:x)}(z)]. \qquad (3)$$



(a) (b)

Fig. 3 Images of the closing results: (a) Erosion, (b) Dilation

C. Area segmentation of nucleus using repeat threshold choice method

The threshold is chosen by using an iterative threshold selection method in 45% to 100% section of the histogram that is based on the simplified image using the proposed pre-treatment process as shown in Fig. 4. Fig. 5 is the resulting image that comes from dividing the nucleus using the repeat threshold choice method.

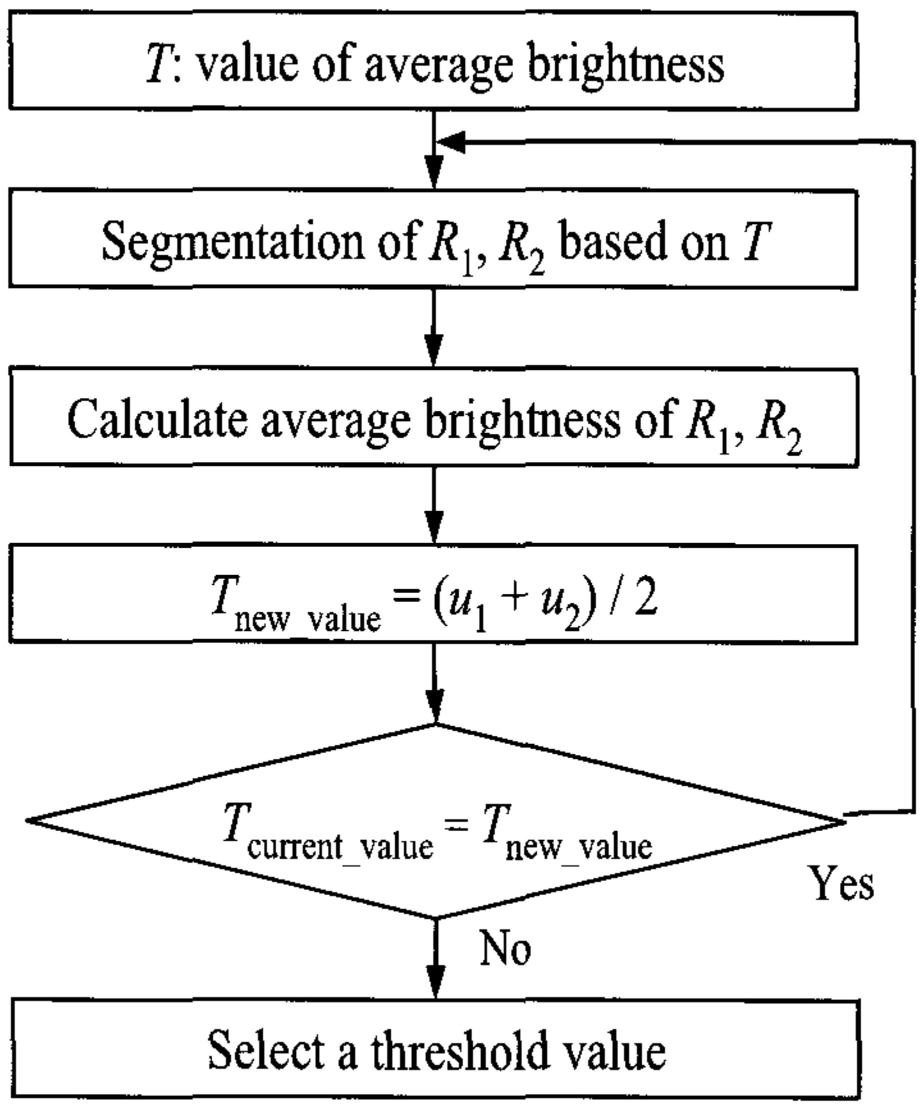


Fig. 4 Algorithm for the selection of the critical value

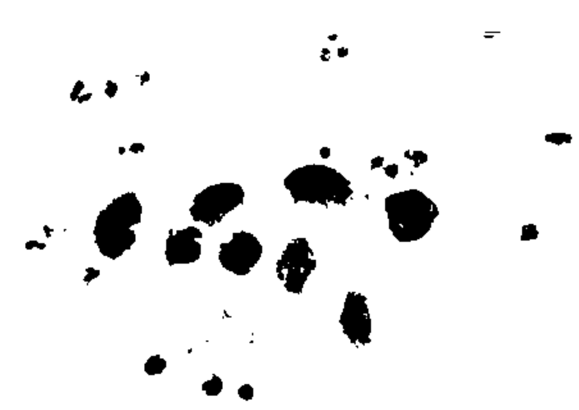


Fig. 5 Image of segmentation for the region of nucleus

D. Nucleus characteristic extraction for cancer cell recognition

A normal nucleus in cervix uteri cytodiagnosis appears small and pale and the nucleus-cytoplasm ratio is also small. On the other hand, an abnormal cell nucleus has a large size and longish or irregular shape compared to a normal cell. Also, because the dyeing rate is different in an abnormal cell, an abnormal nucleus appears dark and the chromatin in the nucleus does not appear even but rough [7], [8].

In this paper, to classify these characteristics, the characteristics of the nucleus and cell image were extracted. First, the following features are extracted for nucleus characteristic: area of nucleus, circumference of nucleus, ratio between circumference of nucleus and circumference of quadrilateral, degree of roundness in nucleus' shape, reciprocal of degree of roundness in nucleus' shape, $\log_{10}$(height/width) in the smallest area of quadrilateral, the longest interior line in horizontal and vertical directions, ratio between area of nucleus and area of quadrilateral. Second, we calculate the area that includes the exterior area of the nucleus and the area of the convex hull wrapped around the nucleus in the convex range.

In the information about brightness, pixels with the following characteristics in which the values are over 60 are extracted: average value, standard deviation, dispersion of brightness, histogram's maximum light and darkness value, central value, smallest light and darkness value, brightness value. Information about color is calculated using the average and standard deviation from the components of red, green, blue, hue, saturation, and intensity. HVS divides texture information into channels in which the energy vector and energy deviation are calculated. Texture characteristic vector that uses energy is calculated by the following.

$$q_{mn} = C_{mn} \sum_{\omega} \sum_{\theta} [p_\theta(\omega)]^2, \tag{4}$$

$$e_{mn} = \log(1 + p_{mn}). \tag{5}$$

Here $p_\theta(\omega)$ represents the value in frequency space of each channel. C_{mn} is a constant for the normalization value. The calculation of the texture feature using energy deflection is as follows [9]:

$$q_{mn} = \sqrt{D_{mn} \sum_{w} \sum_{\theta} [(p_\theta(w))^2 - p_{mn}]^2}\ , \tag{6}$$

$$d_{mn} = \log(1 + q_{mn}). \tag{7}$$

Where, D_{mn} is a constant for the normalization value. The calculated values from equation (4) to (7) and the texture representation that displays the texture feature of a nucleus using the average value and standard deviation of an image is expressed in Eq. (8).

$$Descriptor_{texture} = \begin{bmatrix} dc & std & e_{00} & e_{01} & \dots \\ e_{45} & d_{00} & d_{01} & \dots & d_{45} \end{bmatrix} \tag{8}$$

III. NUCLEUS CLASSIFICATION AND RECOGNITION USING FCM CLUSTERING ALGORITHM

Basically, the fuzzy c-means clustering algorithm resembles the K-means clustering algorithm in that it divides each cluster by using the Euclidian distance of a data set [10]. The only difference is that it does not present whether data belongs to each cluster (0 or 1) but presents a degree (a real number between 0 and 1) of belonging with classification matrix. Also the beginning point is defined in the middle of the whole date by classification matrix that is made temporarily when the beginning point is set. Fuzzy c-means clustering algorithm that is used in this paper is shown in Fig. 6.

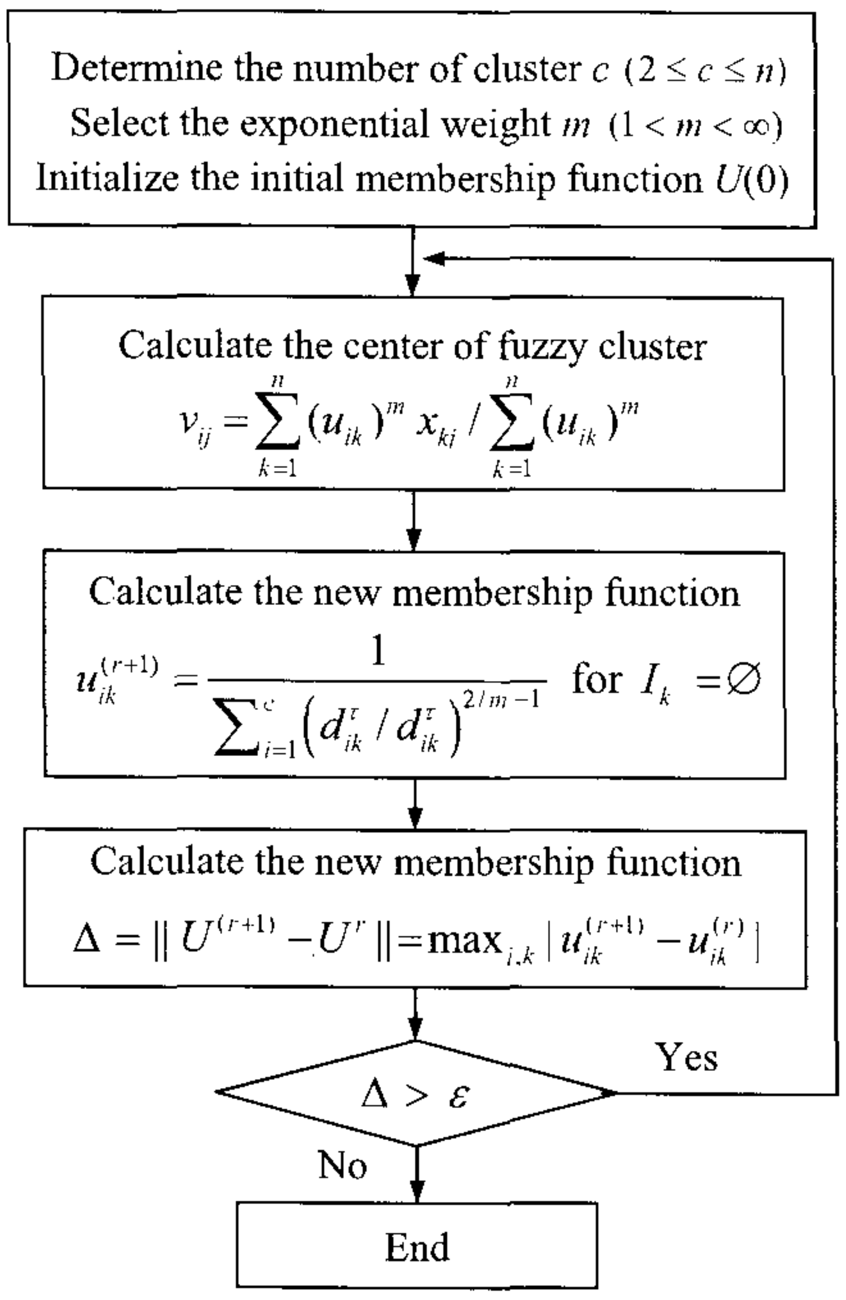
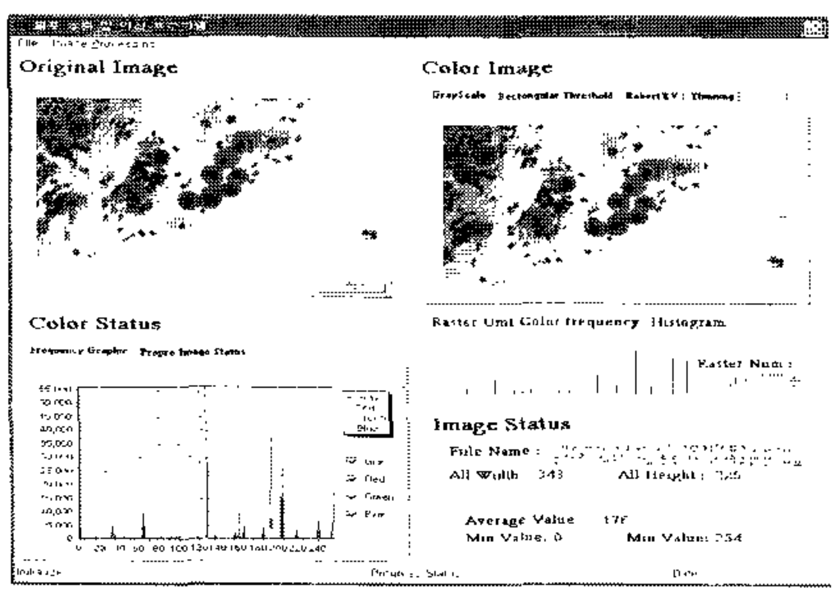


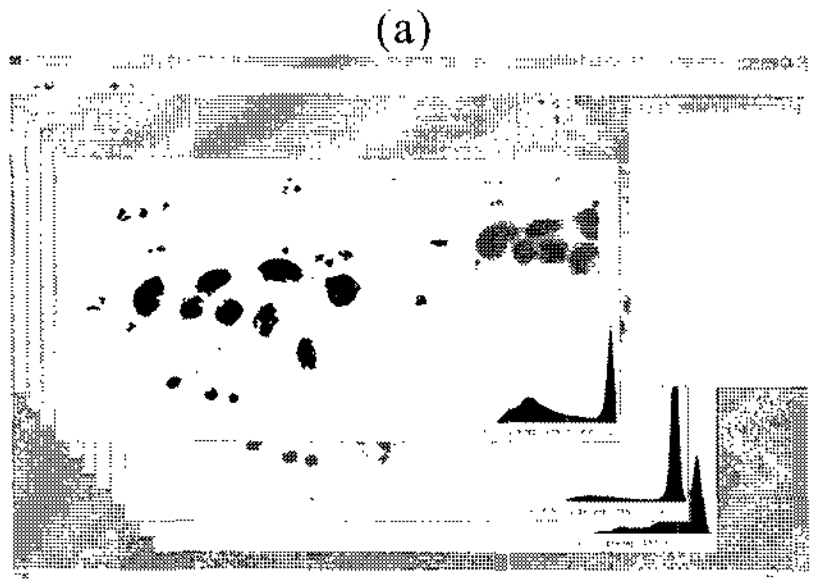
Fig. 6 Fuzzy C-means clustering algorithm

IV. EXPERIMENTAL RESULTS

The environment of the experiment is embodied by Visual C++ 6.0 and C++ Builder 6.0 in Pentium-IV PC of the IBM compatible. The specimen was 20 samples of 640×480 cervix uteri cytodiagnosis image size and it was acquired from Pusan university hospital. The analysis result of a cervix uteri cytodiagnosis image with the proposed method is shown in Fig. 7 (a), and the nucleus extraction result of the cervix uteri cytodiagnosis is shown in Fig. 7 (b).



(a)



(b)

Fig. 7 Nucleus extraction result of the cervix uteri cytodiagnosis: (a) Image analysis of proposed cervix uteri cytodiagnosis, (b) result of nucleus extraction of cervix uteri cytodiagnosis

Table 1 Nucleus extraction result

	Medical Specialist	Proposed method	Extraction rate of the proposed method
Nucleus Extraction	316	278	87.9%

To evaluate the performance of the proposed nucleus segmentation method on 20 actual images, the method was compared with the diagnostic results of a medical specialist. The nucleus number of cervix uteri cytodiagnosis extracted from a medical specialist in 20 samples is 316, and the number extracted from this research is 278. Table 1 displays the number of nucleus that was extracted by the proposed method.

As can be concluded from Table 1, the accuracy of extraction rate is 87.9% in this research. When two or more cluster cells piled up, extraction did not occur correctly. A nucleus can be classified into normal cell nucleus, abnormal cell nucleus and cancer cell nucleus. Therefore, in this paper, the nucleus is divided into the following 5 classes based on the Bethesda System: WNL, ACUS, LSIL, HSIL, SCC. Here, WNL is a normal cell but as it progresses, it has high malignancy in an abnormal cell. Cells that are classified as SCC are cancer cells. Fig. 8 displays nucleus characteristic information of cervix uteri cytodiagnosis that was extracted using the proposed method in this paper.

The classification of nucleus characteristic information based on the Bethesda System by using nucleus characteristic information such as in Fig. 8 is in Fig. 9. By using standards of nucleus classification information to identify normal cells and cancer cells such as in Fig. 9, the accuracy of cancer diagnosis can be much more improved compared to classifying the whole nucleus that appears in the image.

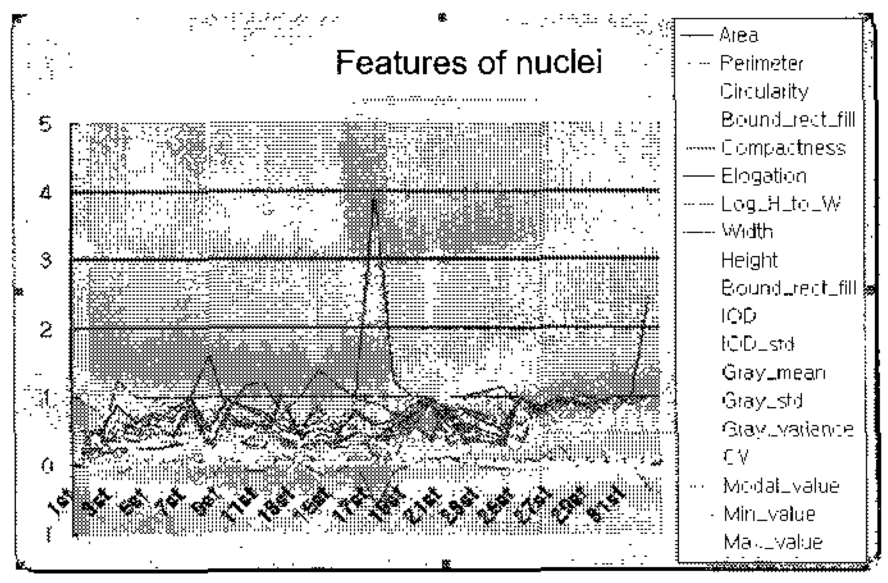


Fig. 8 Characteristic information of nuclei

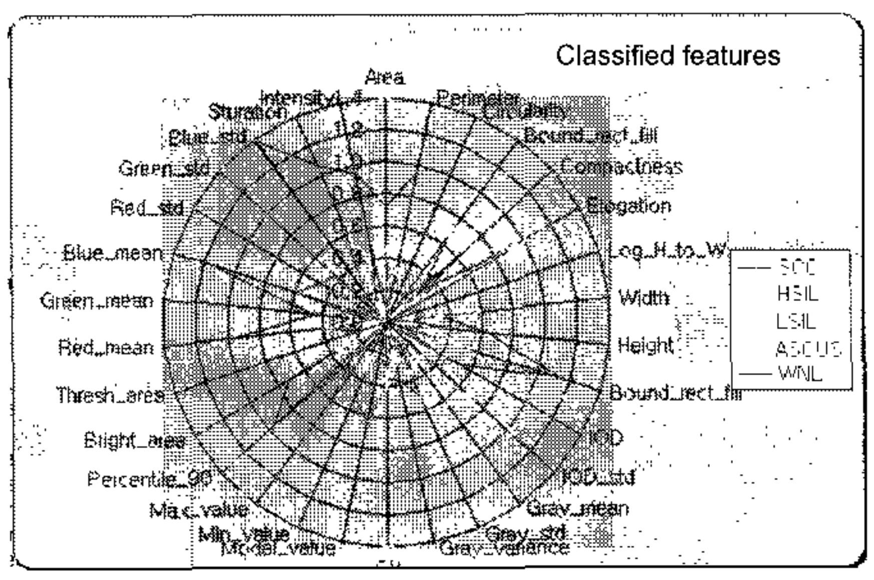


Fig. 9 Information of the standard category in the Bethesda system

In Fig. 8, fuzzy c-means clustering algorithm is applied to classify and distinguish the characteristic information of the extracted nucleus, normal cell, abnormal cell and cancerous cell. The results from classifying and distinguishing the state of the cell using final FCM clustering algorithm is in Table 2.

Table 2 Classification and recognition results of cell by FCM clustering algorithm

Medical specialist			Proposed method		
Normal cell(WNL)	88		Normal cell(WNL)	68	
Abnormal cell	ASCUS	78	Abnormal cell	ASCUS	90
	LSIL	40		LSIL	45
	HSIL	51		HSIL	54
Cancer cell(SCC)	21		Cancer cell(SCC)	21	

Listing in order, the 5 classes WNL, ACUS, LSIL, HSIL, SCC is divided by the Bethesda System, 5 clusters were set for classification as can be seen in Table 2. As can be concluded from Table 2, much more frequently in the proposed method were normal cells classified as abnormal cells than in the diagnosis of a medical specialist. This is because the extracted nucleus is dyed, making it open to being classified as an abnormal cell. But, it can be confirmed that the accuracy between abnormal cells and cancer cells, which a medical specialist diagnosis has little performance however the classification grade of abnormal cells differs. The reason is that FCM clustering algorithm does not sort correctly the dyeing density of nucleus information, which is partial information of the abnormal nucleus cell. But it can be concluded through Table. 2 that the proposed method is comparatively efficient in classifying abnormal cells and cancer cells, and can help a medical specialist in diagnosis.

V. CONCLUSIONS

Because cervix uteri cytodiagnosis is various and complicated, it is difficult to extract and identify cell nucleus efficiently with existing image processing methods.

In this paper, the extraction and identification method of cervix uteri cytodiagnosis using 5 classes WNL, ACUS, LSIL, HSIL, SCC based on the Bethesda System was proposed. In this paper, the nucleus is easily detected by using the following simplification process of the image: (i) converting the extracted image of cervix uteri cytodiagnosis to a grey-scaled image, (ii) removing noise using brightness information, and (iii) applying a 5×5 fuzzy grey morphology operation. To segment the nucleus area in the simplified image, the dark area of the nucleus is separated, and then the following characteristic information is extracted: 13 morphometric features, and 8 densitometric features, 18 colorimetric features, and a textural feature. Extracted information in each nucleus is categorized and recognized into normal cells, abnormal cells with 3 degrees, and cancer cells by the fuzzy c-means clustering algorithm.

As a result of experimenting with 20 samples of cervix uteri cytodiagnosis, it can be confirmed that the proposed method is efficient in recognizing abnormal cells and cancer cells and has little difference with the diagnosis of a medical specialist.

In the future, studies must be conducted to correctly extract characteristic information of nucleus by analyzing morphology and color characteristics of the nucleus and by making a standard of classification to reduce presumption error in nucleus classification by extracting much information.

REFERENCES

[1] Rutenberg, Mark R., "Neural Network Based Automated Cytological Specimen Classification Systems and Method," *United States Patent, Patent* No.4965725, 1990.

[2] Grohs, Heinz K., Husain, O. A. Nassem, "Automated Cervical Cancer Screening," *Igaku-shoin*, 1994.

[3] Seo , C. W., Choi, S. J., Hong, M. K., Lee, H. Y., Jeong, W. G.," Epidemiologic Observationof Diagnosis of Cervical Cancer and Comparative Study of Abnormal Cytologic Smear and Biopsy Result," *Journal of The Korean Academy of Family Medicine*, Vol. 17, No. 1, pp.76-82, 1996.

[4] 4. Banda-Gamboa, H, Ricketts, I, Cairns, A, Hussein, K, Tucker, JH, Husain, N.," Automation in cervical cytology: an overview," *Analytical Cellular Pathology*, Vol. 4, pp.25-48, 1992.

[5] Lee, J. D., Color Atlas Diagnostic Cytology: *Press of Korea Medical Publishing Company*, 1989.

[6] Kim, H. Y., Kim, S. A., Choi, Y. C., Kim, B. S., Kim, H. S., Nam, S. E.," A Study on Nucleus Segmentation of Uterine Cervical Pap-Smears using Multi Stage Segmentation Technique," *Journal of Korea Society of Medical Informatics*. Vol.5, No.1, pp.89-95, 1999.

[7] Kim, G, B., "Detection and Recognition of Uterine Cervical Carcinoma Cells in Pap Smear Using Kapur Method and Morphological Feature," *The Journal of The Korea Institute of Maritime Information & Communication Sciences*, Vol.11, Np.10, pp.1992-1998, 2007.

[8] Bishop, C. M., Neural Networks for Pattern Recognition, Oxford University Press., 1995.

[9] Manjunath, B. S., Ma, W. Y.," Texture Features for Browsing and Retrieval of Image Data," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol. 18, No. 8., 1996.

[10] Klir, George J., Yuan, Bo, Fuzzy Sets and Fuzzy Logic Theory and Applications, *Prentice Hall PTR*, 1995.

Kwang-Baek Kim
Received his M. S. and the Ph.D. degrees in Department of Computer Science from Pusan National University, Busan, Korea, in 1993 and 1999, respectively. From 1997 to present, he is an associate professor, Division of Computer and Information Engineering, and Silla University in Korea. He is currently an associate editor for Journal of Korea Multimedia Society and Journal of The Korean Institute of Maritime Information & Communication Sciences. His research interests include Fuzzy Neural Network and Application, Bioinformatics, Image Processing.