Discrimination of Cancer Cell by Fuzzy Logic in Medical Images

Cheol-Hun Na, *Member, KIMICS*

Abstract—A new method of digital image analysis technique for medical images of cancer cell is presented. This paper deals with the cancer cell discrimination. The object images were the Thyroid Gland cell images that were diagnosed as normal and abnormal. This paper proposes a new discrimination method based on fuzzy logic algorithm. The focus of this paper is an automatic discrimination of cells into normal and abnormal of medical images by dominant feature parameters method with fuzzy algorithm. As a consequence of using fuzzy logic algorithm, the nucleus were successfully diagnosed as normal and abnormal. As for the experimental result, average recognition rate of 64.66% was obtained by applying single parameter of 16 feature parameters at a time. The discrimination rate of 93.08% was obtained by proposed method.

Keywords—Discrimination, Cancer Cell, Fuzzy Logic

I. INTRODUCTION

Pattern Recognition technique has a wide variety of applications in medical image, remote-sensing, geology, and robotics[1]. An example of application in medical image is the evaluation of roentgenograms in classifying normal and abnormal interstitial pulmonary patterns[2].

This paper deals with the cancer cell discrimination that calls attentions to the pathologists. The object cell images were those of Thyroid Gland cells image which were diagnosed respectively as normal cell, follicular neoplastic cell, and papillary neoplastic cell.

The Clinical Cytology which detects the cancer cells by analyzing the microscopic images was introduced by Papanicolaou[3]. The Clinical Cytology is the inspection method of detecting the cancer cells by analyzing the microphotographs of cells in medical image processing. Cells are taken from the internal organs of human body and checked for the existence of cancer cells. It is a necessary inspection method of detection of the various types of cancers for early diagnosis and treatments. However, discriminations were achieved by human visual system. Until the digital process of medical image began early 1960' dealing with the microscopic images, X-ray images, and Computer Tomographic(CT) images. Digital image processing methods has been applied to Clinical Cytology[4]. But, the Clinical Cytology has many problems to the engineers. Medical features are difficult for for engineers to understand. And there are various features in every type of cells. The discrimination experiment uses the multiple parameters instead of simple parameters to increase the discrimination rate. And the discrimination experiment is carried out by using the fuzzy logic which proposed in this paper.

II. FEATURE PARAMETERS

A feature of image is a distinguishing primitive characteristic or attribute of an image. Some features are natural in the sense that they are defined by the visual appearance of an image, while others are artificial result from specific manipulations of an image. This employs 16 feature parameters for experiment that were already proposed in the prior paper[5].

1. Nuclear Area(AR)

The area of nucleus is defined as the number of pixels that was included in nucleus. This feature parameter can be obtained by simply counting the number of pixels within the boundary.

2. Nuclear x-size, and Nuclear y-size(SX and SY)

The length(x-size and y-size) of a contour is one of its simplest descriptors. Simply counting the number of pixels along the contour gives a rough approximation of the length.

3. Perimeter(PM)

The perimeter of a region is the length of its contour. For a chain-coded curve with unit spacing in both directions, it can be obtained by counting the vertical and horizontal components are counting as 1 and the diagonal components as $\sqrt{2}$.

4. Gray Level Mean(GM)

Amplitude measurements may be made at specific image points, *e.g.* the amplitude $F(j,k)$ at coordinate (j,k), and amplitudes of neighborhood pixels centered at (j,k). For example, the average or mean gray level in a $W \times W$ pixel neighborhood is given by

$$GM = \frac{1}{W^2}\sum_m\sum_n F(j+m,k+n) \qquad (1)$$

5. Averaged Power(AP)

$$AP = \left[\Sigma\Sigma GV^2\right] / AR \qquad (2)$$

Manuscript Received December 8, 2005.

Cheol-Hun Na is professor in Mokpo Nat'l Univ., Do-Rim, Chungkye, Moo-An, Chonnam, 534-729, Republic of Korea, (e-mail : chna@mokpo.ac.kr)

where, GV is the Gray level Value of the nucleus.

6. Autocorrelation(AC)

The autocorrelation function has been suggested as the basis of a texture measure[6].

The autocorrelation function is defined as

$$AC = \sum_j \sum_k F(j,k)F(j-m,k-n) \quad (3)$$

for computation over a W×W window with $-T \leq m$, $n \leq T$ pixel lags.

7. Standard Deviation of Gray Level Value(SD)

The standard deviation of neighbor is one of the useful image amplitude features of the image, which can be computed as

$$SD = \frac{1}{W^2}\left[\sum_m \sum_n \left[f(j+m,k+n) - GM\right]^2\right]^{\frac{1}{2}} \quad (4)$$

8. Average of X-Size and Y-Size (MS)

$$MS = \frac{(SX+SY)}{2} \quad (5)$$

9. Denseness(DS)

A parameter in measuring of denseness of a nucleus can be defined as

$$DS = \frac{PM^2}{AR} \quad (6)$$

where, AR is the area of nucleus.

10. Valid Rate(VR)

The Valid Rate is a ratio of nuclear area to area of circumscribed square to the nucleus and this feature parameter can be expressed as

$$VR = \frac{AR}{D} \times 100\% \quad (7)$$

where, D is the area of the circumscribed square to the nucleus.

11. Entropy(ET)

Entropy is a statistical measure of uncertainty. For a given ensemble of pattern vectors, a good measure of intraset dispersion is the entropy, given by

$$ET = -H\,(\log 2 / \log P) \quad (8)$$

Where, $H = \sum_m \sum_n F(j+m,k+n)$ and $\boldsymbol{P}$ is the probability density of the pattern population, and H is the expectation operator with respect to $\boldsymbol{P}$. The entropy concept can be used as a suitable criterion in the design of feature parameters.

12. The ratio of perimeter versus average of X-size and Y-size (RP)

$$RP = \frac{PM}{MS} = \frac{PM}{\left[\frac{(SX+SY)}{2}\right]} \quad (9)$$

13. Inactivity(IN), Inverse Difference(ID) and Absolute Value(AV)

Consider two pixels *A(j,k)* and *A(l,m)* that are located at coordinates *(j,k)* and *(l,m)*, respectively.

The first-order probability distribution of the amplitude of a quantized image may be expressed as[6]

$$\begin{aligned} P(a) &= P\left[A(j,k) = d_a\right] \\ P(b) &= P\left[A(l,m) = d_b\right] \end{aligned} \quad (10)$$

where, d_a and d_b represent quantized pixel amplitude value for $0 \leq a \leq X-1$ and $0 \leq b \leq X-1$.

If an image region contains fine texture, the two-dimensional histogram of pixel pairs will tend to be uniform, and for coarse texture the histogram values will be skewed toward the diagonal of the histogram. Consider the pair of pixels and that are separated by radial units at an angle with respect to the horizontal axis.

* Inactivity(IN)

$$IN = \sum_a \sum_b (a-b)^2 P\left[A(j,k) = a, A(l,m) = b\right] \quad (11)$$

* Inverse Difference(ID)

$$ID = \frac{\sum_a \sum_b P\left[A(j,k) = a, A(l.m) = b\right]}{\left[1+(a-b)\right]} \quad (12)$$

*Absolute Value(AV)

Consider two pixels ***A(j,k)*** and ***A(m,n)*** that are located at coordinates and ***(j,k)***, respectively. They are separated by distance d at an angle ϕ with respect to the horizontal axis. The histogram estimation of the second order distribution is

$$J(a,b) \approx B(a,b) / AR \quad (13)$$

where, AR is the total number of pixels in the measurement window and ***B(a,b)*** denotes the number of occurrences for which $A(j,k)=d_a$ and $A(m,n)=d_b$. Then the absolute value can be expressed as

$$AV = \sum_a \sum_b |a-b| J(a-b) \quad (14)$$

III. FUZZY LOGIC ALGORITHM FOR RECOGNITION OF CANCER CELL

Generally, the discrimination of nucleus was carried out using single parameters in Clinical Cytology. Because the discrimination rate of some feature parameters are enough to be used for automatic discrimination. Thus the parameter of having high discrimination rate was required in discrimination problems.

Pathologist do not diagnosed norm or abnormal cells by using only one factor. The discrimination of the nucleus in medical cells image was handled in spatial domain. Form the cytological point of view, considered as a whole from the cytological point of view, the difference between normal and abnormal Thyroid Gland cells was basically caused by nuclear membrane, shape of nucleus, and the size of nucleus. It is expected that the difference between normal and abnormal Thyroid Gland cells will be distinguished by the parameters of nuclear area, nuclear x-size and y-size, perimeter, denseness, and entropy of nucleus.

The fuzzy logic algorithm provides effective method for regulation from uncertain information. The Feature Selection used to reduce the order of initial data that contains the feature in pattern recognition. A design method for discrimination system that has language variable input was proposed by Nath[9]. It consists of discrimination part and fine control part. And the fine control part used the sample data to control the input data. This design method adapted to medical examination system. This paper uses the dominant feature parameters from the 16 feature parameter to create the fuzzy algorithm. According to this algorithm, it adapted to discrimination of cell of thyroid gland.

This paper defines the fuzzification function as following. The fuzzification function of i^{th} nucleus, j^{th} feature parameter, k^{th} nuclear type (k=1, 2, 3; normal, follicular, papillary neoplastic) is

$$F_{ijk} = \begin{cases} \left| \dfrac{Parm_{ijk} - cell_{ij}}{Parm_{I+1jk} - Param_{ijk}} \right| & \text{for } Param_{ijk} \le cell_{ij} \le Param_{I+1\,jk} \\ \left| \dfrac{Parm_{ijk} - cell_{ij}}{Param_{ijk} - Param_{I-1\,jk}} \right| & \text{for } Param_{I-1\,jk} \le cell_{ij} \le Param_{ijk} \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

$Param_{Ijk}$ is the feature variable of j^{th} feature parameter, $cell_{ij}$ it was 1 ; normal, k^{th} nuclear type, is the j^{th} feature variable of i^{th} nuclear.

Value of F_{Zzijk} $= 1 - F_{ijk}$
Value of F_{ZZijk} $= (1 - F_{ijk})^{1/2}$
where,

$$0 \le F_{ijk} \le 1. \quad (16)$$

The output of i^{th} nuclear about k^{th} nuclear type is

$$Output_{ik} = \sum_{j=1}^{16} Value \text{ of } F_{zz_{ijk}} \quad (17)$$

For example, the Fuzzy Logic Algorithm will increase the variable assigned for normal cell if the discrimination by one feature parameter is normal. If the result is abnormal, algorithm will increase proper variable and proceeds to next parameter. Finally, the Fuzzy Logic Algorithm will discriminate the cell by selecting of variable with biggest value. By using the Fuzzy Logic Algorithm, the discrimination rate will be highly improved.

IV. EXPERIMENT

A. Materials

The samples used in the experiment of this paper are Thyroid Gland cells images. They are microscopic images of Thyroid Gland cells taken by surgical operation and are respectively diagnosed as normal, follicular neoplastic and papillary neoplastic. After dying by Papanicolaou technique, sample images of each class were obtained by microscope and captured by camera. The size of object image was 512×512 with 8 bit resolution of gray leveled value. Actual size of sample image was 102.4 μm.

B. Thresholding

Many images can be characterized as images containing some object of interest of reasonably uniform brightness placed against a background with different brightness.

Several analytic approaches to the setting of a luminance threshold have been proposed[7]. A bi-modal histogram often suggests that the image which contains an object with a narrow amplitude ranges against a background with different amplitude. Where an image is clearly distinguishable from background object, the shape of gray level histogram is bi-modal[5].

C. Segmentation

* Contour Following Method with directional angle

The cells image contains nucleus, cytoplasm, red blood cell, and extra-cellular materials : *e.g.*, colloid, blood plasma. Thus segmentation for classification of the nucleus from cells image was required[3-5]. The Contour Following method searches with two directions only. Searching by two directions is quite within the realms of possibility of missing the diagonal pixels. Thus, when the bug searches for the neighbor pixel, it should search for not only vertical or horizontal direction but also every other direction. A reasonable choice for the searching directions is directions with angle of multiples of π/4. If the increment of searching direction for neighbor pixel is multiples of π/4, there is no missing pixel[8].

V. CANCER CELL DISCRIMINATION OF THYROID GLAND CELLS

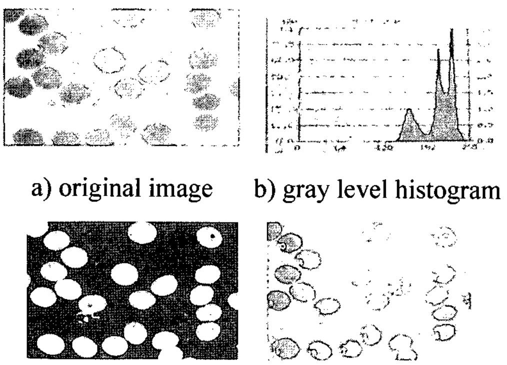
A. Selection of Dominant Feature Parameters

The feature parameters used in this paper are 16 parameters. At the beginning of the experiment, the distance between the data of Look-up table of each feature parameter and the data of each feature parameter of nucleus was calculated, and the discrimination was carried out based upon this distance. Then, calculate the discrimination rate for all the possible combinations of

16 feature parameters in order to find the combination for the best discrimination rate[8]. By the experimental result, the number of dominant feature parameters is 5 and the selected dominant feature parameters were ①SX, ②RP, ③DS, ④SD, ⑤AV.

The combinations of dominant feature parameters may be changed depending upon the object cell used because the features of other cells are not same as the features of Thyroid Gland cells used in this paper.

B. Sample result



a) original image b) gray level histogram

c) thresholded image d) segmentation result

Fig. 1 A Sample Result of Normal.

Figure 1 shows the case of normal nucleus. In this Figure, a) express the original cells image, b) shows gray level histogram of original image, c) gives the thresholded image of a), and d) represent the segmented result. The numbers shown in d) signifies the sequence number assigned to the isolated nucleus.

C. Experimental Results

Table 1 shows the summary of discrimination rates in each class by applying the feature parameters. The parameters of deneness(DS), entropy(ET), and absolute value(AV) gave high discrimination rate(around 78%) for all classes. Discrimination of nucleus by averaged power(AP), inverse difference(ID), and inactivity(IN) were effective on follicular cells but not for normal or papillary cells. The standard deviation of gray level of nucleus(SD) gave high discrimination rate(around 80%) for all class. Other parameters did not give high enough discrimination rate for all classes(around 40%)[8]. The experimental result was as follows. The best discrimination rate obtained by single parameter method was 79.97% for generally used feature parameters. The second is 79.75% which was by nuclear X-size. The worst was 35.46% which was obtained by averaged power of nucleus(AP). The best discrimination rate obtained by single parameter method was 79.69% for proposed feature parameters. The second is 78.10% which was by ratio of perimeter versus average of SX and SY(RP). The worst was 44.40% which was obtained by inverse difference(ID). Average discrimination rate obtained by applying 16 feature parameters, single parameter at a time, was 64.66%. 80.35% of discrimination rate obtained by Power Spectral Density function of FFT[10, 11]. As examined above, the discrimination rate of each parameter was high enough for automatic discrimination. However most feature parameters gave high discrimination rate for some class. According to the experimental results, discrimination of the cell by using single parameter may be risky. The discrimination of single parameter method can be overcome by dominant feature parameters method.

Discrimination rate obtained by Fuzzy Logic Algorithm was 93.08%. It was improved at least 12.73%.

Table 1 Summary of discrimination rates.

parameters \ class		Normal		Follicular		Papillary		Total	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%
No.	No. of Nuclei	191	100	879	100	752	100	1822	100
1	AR	64	33.51	814	92.61	399	53.06	1280	70.25
2	SX	84	43.98	828	94.20	541	71.94	1453	79.75
3	SY	96	50.26	804	91.47	473	62.90	1373	75.36
4	PM	76	39.79	772	87.83	570	75.80	1418	77.83
5	GM	61	31.94	372	42.32	321	42.69	754	41.38
6	AP	117	61.26	201	22.87	328	43.62	646	35.46
7	AC	110	57.59	282	32.08	316	42.02	708	38.86
8	SD	101	52.88	771	87.71	585	77.79	1457	79.97
9	AV	96	50.26	747	84.98	608	80.85	1451	79.63
10	MS	14	7.33	818	92.94	49	6.52	880	48.30
11	DS	101	52.88	739	84.07	612	81.38	1452	79.69
12	VR	23	12.04	797	90.67	318	42.29	1138	62.46
13	ET	119	62.30	733	83.39	449	59.71	1301	71.41
14	RP	102	53.40	768	87.37	553	73.54	1423	78.10
15	IN	113	59.16	638	72.58	555	73.80	1306	71.68
16	ID	88	46.07	407	46.30	314	41.76	809	44.40
17	Avg. of 1-16	85.31	44.67	655.69	74.59	436.94	58.10	1178.06	64.66
18	PSD of FFT	158	82.72	749	85.21	557	74.07	1464	80.35
19	Proposed	178	93.19	841	95.68	677	90.03	1696	93.08

VI. CONCLUSION.

In this paper, a new method of discrimination for medical image analysis was studied by using the pattern recognition techniques. The focus of this paper is the automatic discrimination of cells into normal and abnormal cells. The object cells image used in this paper was microscopic image of Thyroid Gland cells. And new technique for discrimination of cells image by using the Fuzzy Logic Algorithm was also proposed.

The result of segmentation was very successfully executed. 16 parameters of Morphological features were extracted in spatial domain. Power spectral density function of FFT in frequency domain was also calculated for comparison.

As for the experimental result, average discrimination rate of 64.66% was obtained by applying single parameters at a time. Discrimination rate of 93.08% was obtained by dominant feature parameters method in spatial domain. Discrimination rate of 80.35% was obtained by PSD of FFT method in transform domain. As a consequency of using the fuzzy logic proposed in this paper(K=5, feature parameters are ①SX, ②RP, ③DS, ④SD, ⑤AV), the discrimination rate was improved by at least 12.73%.

REFERENCES

[1] Sing-Tze Bow, Pattern Recognition and Image Preprocessing, Marcel Dekker, New York, 1992.

[2] A. Khotanzad and R. Kashyap, "Feature Selection for Texture Recognition Based on Image Systhesis,"

IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC-17, pp. 1087-1095, Nov., 1987.
[3] W. Galvraith, et, al., "Studies on Papanicolaou Staining," Analytical and Quantitative Cytology, Vol. 1, No. 3, pp. 160-169, 1979.
[4] G. W. Gill, K. A. Miller, "In Compendium on Cytopreparatory Techniques," Edited by C.M. Keebler, Tutorials of Cytology, Vol. 9, No. 25, 1974.
[5] Cheol-Hun Na and Hyeon-Jae Kim, "Effective Discrimination of Cancer cells in Medical Images," Proc. of IEEE 1993 NSS and MIC, Vol. 2, pp. 1371-1375, San Francisco, CA, U. S. A., Nov. 1993.
[6] W. K. Pratt, Digital Image Processing, 2nd ed., John Wiley & Sons, New York, 1991.
[7] J. S. Weska, "A Survey of Threshold Selection Techniques," Computer Graphics and Image Processing, Vol. 7, No. 2, pp. 259-265, April 1978.
[8] Cheol-Hun Na and Sang-Jin Ryoo, "Cancer Cell Recognition by Dominant Feature Parameters Method," GSPx 2003, Dallas, TX, U.S.A.,2003.
[9] A. K. Nath and T. T. Lee, "On the Design of a Classifier with Linguistic Variables as Inputs," Fuzzy Sets Syst., vol. 11, pp. 265~285, 1983.
[10] Cheol-Hun Na et, al., "Discrimination of Cancer Cells by Distribution Function of Power Spectral Density of Fourier Transform," GSPx 2004, Santa Clara, CA, U.S.A., 2004.
[11] Cheol-Hun Na et, al., "Cancer Cell Recognition by Fuzzy Logic in Medical Images," GSPx 2005, Santa Clara, CA, U.S.A., 2005.

Cheol-Hun Na
Received B.S., M.S. and Ph.D. degree from Elecrtonic Engineering, Chonnam Nat'l Univ. in 1985, 1987 and 1994, respectively. From 1995, he worked as a professor of Mokpo Nat'l University. His research interest is in the area of Digital Images Processing, Communication Engineering, and Medical Imaging.