We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Computer Aided Diagnosis of Glandular Tumours for Internet-Based Telemedicine

Toshiyuki Tanaka
Keio University
Japan

1. Introduction

In developing countries and isolated islands, lack of medical staff is a major problem. When a person feels unwell, he or she usually attends hospital and is assessed by a clinician and given a provisional diagnosis. When the clinician suspects cancer, a pathological diagnosis is required. This requires a pathologist, but pathologists are badly off and scarce in many countries. Since pathological diagnosis represents a subjective assessment by each pathologist, several pathologists must examine the same specimen and collate the results for accurate diagnosis, and each pathological diagnosis accordingly requires a substantial amount of time. A diagnosis support system using objective and quantitative approaches would thus be extremely useful. Moreover, the use of the Internet for diagnosis is expected to alleviate the problem of access to medical professionals in developing countries and isolated islands. Glandular cancers, such as those of the stomach, colon, prostate, and breast, are common, and we have therefore focused on these cancers in developing a computer aided diagnosis system using the Internet.

In the field of medical image engineering, automatic and quantitative diagnosis systems have previously been studied by many researchers. Most such research has aimed at the development of diagnosis support systems addressing the lack of pathologists. A representative system is the pre-screening system for uterine cytodiagnosis that is currently in use. The pathologist must still perform many image processing steps to use these diagnosis systems. In research for tumour imaging, human input is still required for detection of the region of breast cancer, for identifying the glomeruli in the kidney; and so on. Although the types of tumour involved increase year by year, most research centres on tumours of the lung, colon, prostate, and ovaries, with few studies examining gastric tumours.

Well-differentiated glandular cancer and adenoma differ from each other, and the classification of these tumours has recently been discussed by pathologists. However, no adequate automatic diagnosis systems for glandular tumours are available for such studies as described above, and pathologists have been eagerly awaiting a system for glandular tumours. Morphological classification of colorectal microscopic images has been reported as a recent diagnosis support system. This method allows features of tumours to be obtained from the whole image, but has a low rate of correct classification for images with a relatively large background region. With the aim of formulating an Internet-based diagnosis system,

we first created an automatic diagnosis system for glandular cancer. In the second stage of our research, we will integrate the Internet-based diagnosis system.

In this chapter we explain our computer-aided diagnosis system for glandular tumours based on morphological features of the cytoplasm and nucleus. In this system, glandular tumours are classified into malignant tumours or adenoma, using numerical features obtained from the morphological assessments of pathologists. For numerical conversion of morphological features, manual extraction of regions of interest (ROIs) was performed, with pre-processing selection of colour components comprising image, contrast enhancement, binarisation using the Laplacian histogram and discriminant methods. In the pre-processing stage, each tumour was classified in terms of its glandular structure and nucleus. From the regions selected, many morphological and texture were identified, and using the obtained features, glandular tumours were classified by the discriminant method with the stepwise method. A small number of features were selected for diagnosis by the stepwise method. Finally, tumours were classified into various categories: non-neoplastic, gastric adenoma, and gastric cancer, using the obtained numerical features.

2. System summary

In this section we summarise the Internet-based diagnosis system, which allows doctors to obtain diagnostic results anytime, anywhere. Figure 1 shows the essential components of the system. First, we upload a glandular image to the system using the internet. After some processing, the system displays the uploaded image and the diagnosis. This upload window will be created using Perl. Recently, Java and PHP have been recognised as useful internet languages, and C language can also create diagnostic programs. Since the diagnosis system has to perform a large amount of processing, Perl cannot deliver sufficient computing speed. Figure 2 shows the algorithm used by the diagnosis system. One need only upload the lesion image, and the diagnosis system analyses the severity of the lesion. Since this algorithm is essential for this system, it may be very useful in future systems.
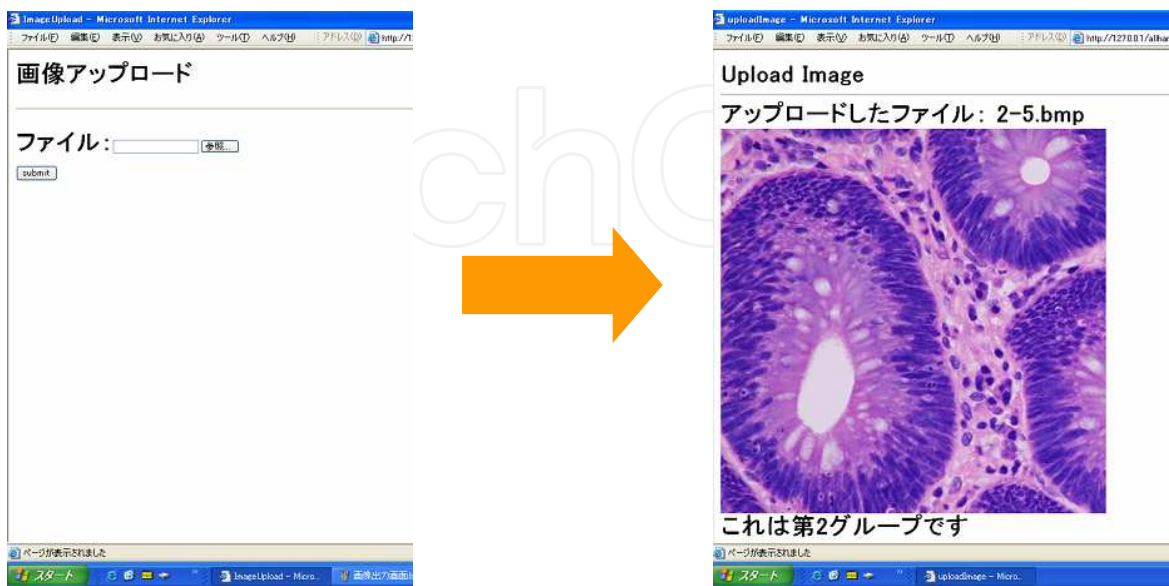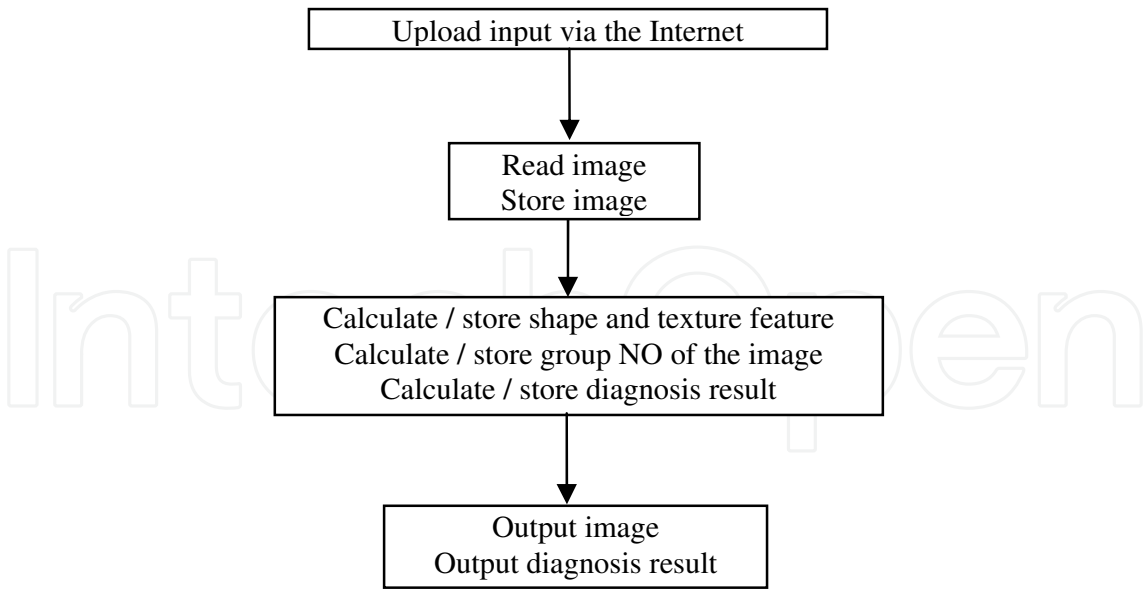

Fig. 1. Internet-based diagnosis system

```
┌─────────────────────────────────┐
│   Upload input via the Internet │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│          Read image             │
│          Store image            │
└─────────────────────────────────┘
                 │
                 ▼
┌───────────────────────────────────────────┐
│   Calculate / store shape and texture feature │
│   Calculate / store group NO of the image  │
│   Calculate / store diagnosis result       │
└───────────────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│          Output image           │
│      Output diagnosis result    │
└─────────────────────────────────┘
```

Fig. 2. Algorithm of the internet-based diagnosis system

3. Severity grading of glandular tumours

Pathologists classify glandular tumours into several categories as shown in Table 1, based on morphological features of the nucleus and cytoplasm as shown in Figure 3. Group 1 includes normal tissue and obviously benign lesions, and group 5 corresponds to frank malignancy. Group 2 to 4 are borderline cases between benign tissue and malignant tumor. Classification of group 2 to 4 is difficult, even when tissue samples are analysed by pathologists. We classify gastrointestinal glandular tumours into group 1, group 3, and group 5. As an example we show a typical case of classification of colon tumour according to morphological features of the nucleus and cytoplasm. The original images are classified into the correct category by the pathologist.

Group 1	Normal tissue and benign lesion
Group 2	Benign lesion with aberrant tissue
Group 3	Boundary case between benign and malignant tumor
Group 4	Tissue at increased risk for cancer
Group 5	Complete cancer

Table 1. Severity classification of gastric biopsy specimens

The severity of the tumour is decided by the shape of glands and the distribution of typically shaped glands. Although histological images can provide only 2D information about gland structure, pathologists need information on 3D structure. Pathologists mentally construct the 3D structure from the 2D shape and thereby determine the severity of the tumour. Figure 3 shows one examples of the relationship between 2D shape and 3D structure of glands. Many structures with a similar 2D appearance have different structures in 3D. Thus substantial experience is required for pathological diagnosis. In this research, statistical processing is used to recognise various 3D patterns.
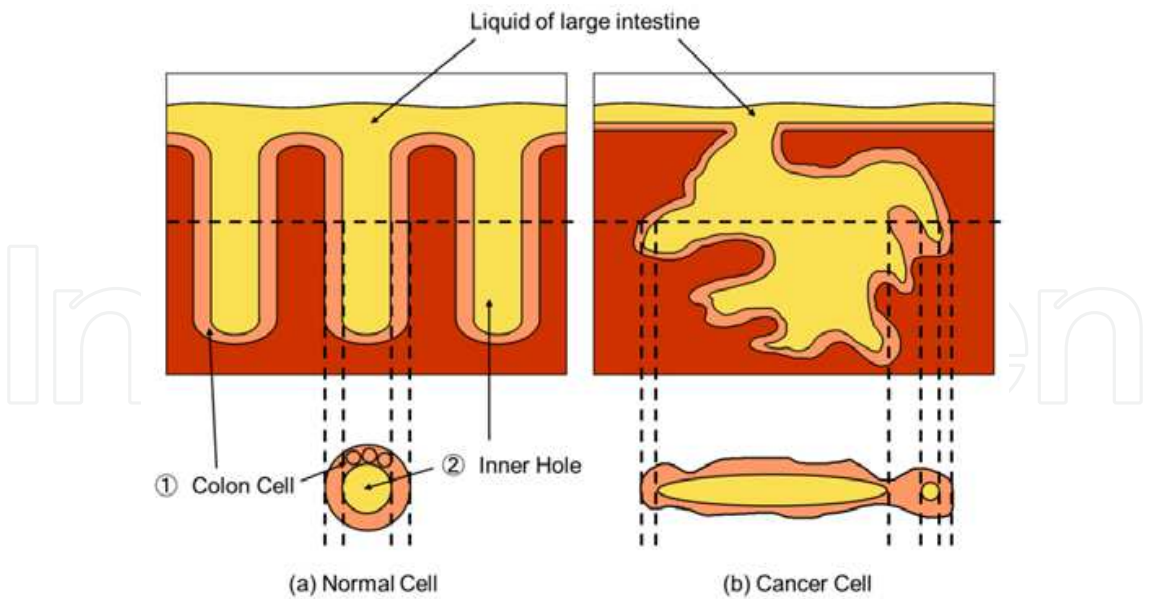
Fig. 3. Relation between 2D shape and 3D structure of glands

4. Methodology

4.1 Region extraction

In this study we used original 2240x1680 pixel images with a 24-bit bitmap file format. As shown in Figure 4, rectangular regions containing glands are manually removed from the original image, and are used as the input images. The entire gland must be visible in these images.
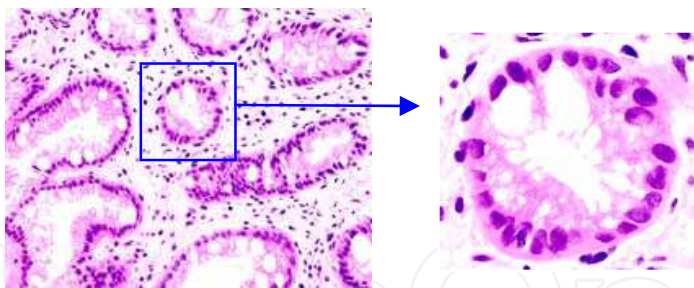

Fig. 4. Original image and extracted gland image

First of procedure in this research we precisely extracted the nuclear regions and the glandular regions for computation of morphological features. Figure 5 shows the flowchart of the region extraction method for nuclear and glandular regions. A different binarisation method was used for these two types of region extraction: strong contrast between the nucleus and cytoplasm is better for binarisation for nuclear region extraction while strong contrast between the gland and background is preferred for binarisation for gland region extraction. In this study we used the red component of the RGB image for nuclear region extraction, and the green component image for glandular region extraction. The dotted line in Figure 5 shows that the nuclear regions are extracted from the extracted gland regions. In the automatic diagnosis system the discrimination of tumour severity was performed by the features computed from the extracted nuclear and gland shapes.
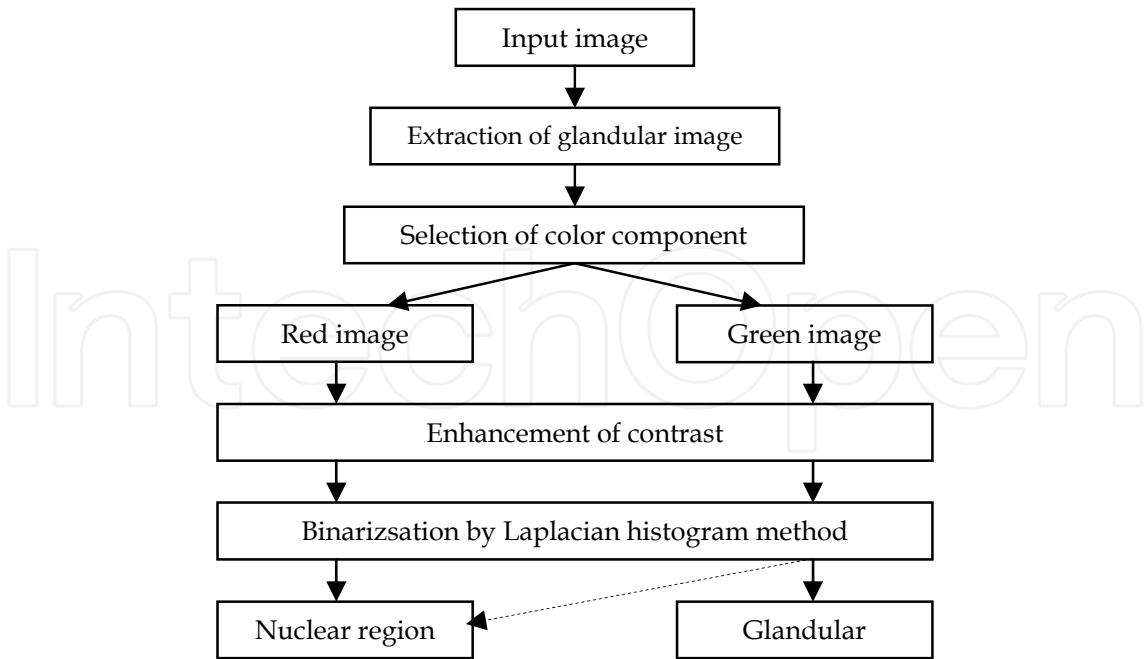
Fig. 5. Flowchart of region extraction method for nuclei and glands

4.2 Pre-processing
4.2.1 Selection of colour components for pre-processing

Since the nuclei and glands are different colours in the images and these colours also vary slightly according to the staining protocol, it is difficult to extract the shape of nuclei and glands from colour images. We therefore converted the colour image into a greyscale image. There are various methods for doing this; one using one component in the RGB image and one converting the RGB basis into the YIQ basis. We used the second method, which is well known, using the Y component of the image in the YIQ basis. We show the conversion equation from the RGB basis to the YIQ basis below.

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.144 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.522 & 0.331 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

Figure 6 shows the obtained colour components of a glandular image represented in greyscale. In this instance, the R component image in the RGB basis is suitable for extraction of the nuclear region and the G component image is suitable for extraction of the glandular region. Moreover, the Y component image in the YIQ basis is suitable for computation of textural features within the glandular region.
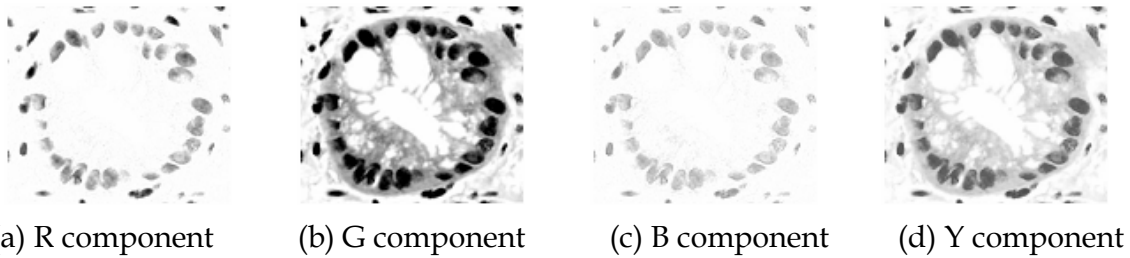

(a) R component   (b) G component   (c) B component   (d) Y component

Fig. 6. Colour components of glandular image

4.2.2 γ compensation

We performed γ compensation in order to optimise the results of binarisation. When the input pixel density is X and output pixel density is Y, the equation for γ compensation is shown below.

$$Y = \left(\frac{X}{255}\right)^{\frac{1}{\gamma}} \times 255 \tag{2}$$

In the above equation, γ indicates the parameter of the equation, and we use γ=5 for the nuclear region extraction and γ=3 for the glandular region extraction.

4.3 Binarisation
4.3.1 Binarisation by fixed threshold

A binarisation method is useful when we try to extract an object shape. Here we explain the fundamental approach. Binarisation converts an intensity image into a two-level black and white image, using one threshold for the whole image. In the intensity image each pixel has a value of 8 bits, which is from 0 to 255. Below is a fundamental equation of binarisation in which $f(x,y)$ indicates an intensity value of (x,y) coordinates in an image, $g(x,y)$ refers to the output value after binarisation, and a T means the threshold value.

$$g(x,y) = \begin{cases} 0 & f(x,y) \leq T \\ 255 & f(x,y) > T \end{cases} \tag{3}$$

Since an object region such as a nuclear or gland region has a lower intensity than a background region, object regions have an intensity of 0 and background regions have intensity of 255 after binarisation. Therefore we can select the object region after binarisation, and easily analyse morphological features from the selected object. The selection of threshold value is important in the binarisation because the shape of the object can change according to the selected threshold. In this research we used the Laplacian histogram method and discriminant method.

4.3.1 Binarisation by the discriminant method

Here we explain binarisation with the discriminant method. In this method the optimal threshold is determined by searching for the maximal interclass variance after the intensity histogram $h(k)$ is constructed from an image and is converted into two classes.

First, an intensity value is set as k. We define the following equation as an index for goodness of binarisation, in which $\eta(k)$ is the degree of separation used in the discriminant method, $\sigma_B^2(k)$ is the interclass variance, and σ_T^2 is the whole variance.

$$\eta(k) = \frac{\sigma_B^2(k)}{\sigma_T^2} \tag{4}$$

$\sigma_B^2(k)$ and σ_T^2 are computed in the next equation.

$$\sigma_B^2(k) = \frac{\{\mu_T \cdot \omega(k) - \mu(k)\}^2}{\omega(k)\{1 - \omega(k)\}} \tag{5}$$

$$\sigma_T{}^2 = \sum_{l=0}^{L-1} (l - \mu_T)^2 p(l) \tag{6}$$

When an occurrence probability is $p(l)$ using intensity l, the whole level is L (L=255 in this research), intensity histogram is $h(l)$, and the total pixel number is N, the whole average value μ_T is shown in the next equation.

$$p(l) = \frac{h(l)}{N} = \frac{h(l)}{\sum\limits_{l=0}^{L-1} h(l)} \quad \left(p(l) \geq 0, \sum_{l=0}^{L-1} p(l) = 1 \right) \tag{7}$$

$$\mu_T = \sum_{l=0}^{L-1} lp(l) \tag{8}$$

Moreover $\omega(k)$ and $\mu(k)$ are defined by the occurrence probability $p(l)$.

$$\omega(k) = \sum_{l=0}^{L-1} p(l) = \omega(k-1) + p(k) \tag{9}$$

$$\mu(k) = \sum_{l=0}^{L-1} lp(l) = \mu(k-1) + kp(k) \tag{10}$$

Since $\sigma_T{}^2$ is constant regardless of k value, the degree of selection $\eta(k)$ between classes becomes maximal when the value of $\sigma_B{}^2(k)$ becomes maximal. $\sigma_B{}^2(k)$ is computed by changing the value of k from 0 to L-1 in series, and an optimal k value is obtained. We perform the binarisation of a greyscale image using the obtained k value. We can objectively and automatically obtain an optimal threshold in this manner.

4.3.1 Laplacian histogram method

A problem arises when determining the threshold by the discriminant method: the obtained threshold shifts a little toward the intensity of a larger area when the ratio of the object area to background area is large. However, it is hoped that the intensity histogram for the discriminant method has a ratio of object area to background area of almost 1.

The Laplacian histogram method is one solution for this problem. In this method the range of the histogram is limited to the regions adjacent to an object and background using the Laplacian value of an input image. In the images we used, the regions adjacent to an object and background have a large absolute Laplacian value of each pixel. We use the following equation for computing the Laplacian value for all pixels within an image and construct the histogram using only pixels with a Laplacian value greater than a certain threshold.

$$\begin{aligned}
\nabla^2 f(x,y) &= \frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2} \\
&= f(x,y-1) + f(x,y+1) + f(x-1,y) + f(x+1,y) - 4f(x,y)
\end{aligned} \tag{11}$$

In this study we selected only the pixels within the top 20% of all computed Laplacian values.

4.4 Post-processing

The two-level image obtained by the Laplacian histogram and discriminant method has the following noise considerations.

1.   Salt-and-pepper noise in the background
2.   White space in the object region

For the first type of noise a median filter is effective, and a filter operator of 3x3 is used in this study. For the second type an area combination by closing processing is effective.

In the G component image after labelling processing for the two-level image, we select the largest area as the glandular region and eliminate other regions as noise in the image. For the white space in the gland we obtain the area and leave the largest area as the glandular cavity and eliminate any other white space.

In the two-level image of the R component we select the regions within the gland as nuclear and eliminate the regions external to the gland as noise. In other words, the nuclear region is determined after extraction of the glandular region.

Figure 7 shows the result of post-processing for the G component image. Figure 7(a) shows an image after binarisation of the G component image and Figure 7(b) shows the image obtained after post-processing. Figure 8 shows the result of post-processing for the R component image. Figure 8(a) shows the image after binarisation of R component image and Figure 8(b) shows an obtained image after post-processing.

(a) After binarisation   (b) After post-processing

Fig. 7. Glandular image

(a) After binarisation   (b) After post-processing

Fig. 8. Nuclear image

The median filter outputs the median value of intensity magnitude as intensity value of reference pixel after the pixel values within the masked region are permutated according to the intensity magnitude. This filter is suitable for elimination of salt-and-pepper noise in an image. For the filtering in this study we use a filter operator (mask operator) of 3x3 pixels. As the closing processing, several shrinkage treatments are performed after several expanding processing procedures.

4.6 Morphological feature extraction

4.6.1 Features of area

The area of an object region is expressed by the number of pixels within the connected components. In this study we define the area as shown in Figure 9 and Table 2. Four types of area are computed as shown below.

1. The nuclear area: the sum of the areas of all nuclei within a gland
2. The cytoplasmic area: the area of A in Figure 9
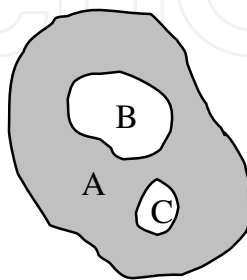3. The glandular cavity area: the area of B+C in Figure 9
4. The glandular area: the area of A+B+C in Figure 9

Fig. 9. Definitions of various areas

Cytoplasm	Hole	Gland
A	B+C	A+B+C

Table 2. The definitions of various areas

4.6.2 Features of length

The perimeter of an object region is expressed using the number of contour pixels obtained by contour tracing. The distance between obliquely lining pixels is multiplied by a weight of $\sqrt{2}$. Three features as shown below are computed as the features of length.

1. Perimeters of the gland and glandular cavity
2. Feret's diameter: the length of the diagonal line of the minimal rectangle surrounding a gland

4.6.3 Features of chord and axis

The five features computed as features of chord and axes are as follows.

1. The longest segment: the length of the maximal segment in the gland and the length of A in the horizontal direction as shown in Figure 10.
2. The maximal section: the length of the longest segment among segments in all directions and the length of B as shown in Figure 10.
3. The average number of chords comprising the gland in the horizontal axis and in the vertical axis
4. The average length of the vertical segment; the average length of the segment vertical to the maximal segment and the length of C as shown in Fig.10.

4.6.4 Features of equivalent shape

One feature of equivalent shape is defined as the length of the short axis in the ellipse which has the same area of the gland and half of the maximal segment. In this definition we obtain the following equation using the letters *a* and *b* in Fig.11.
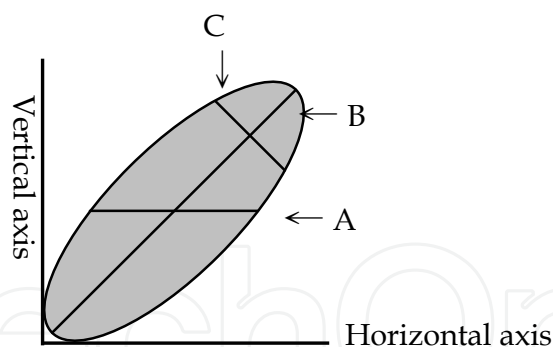
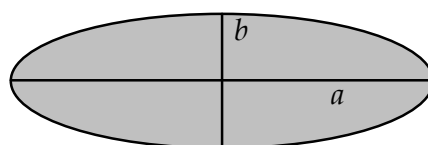Fig. 10. Definition of chord and axis features of the object

Fig. 11. The definition of equivalent oval

$$\begin{aligned} \text{particle area} &= \pi ab \\ \text{maximal segment} &= 2a \end{aligned} \tag{12}$$

The short axis of the equivalent ellipse is computed by the next equation.

$$2b = \frac{4 \times \pi ab}{\pi \times 2a} \tag{13}$$

4.6.5 Other features of equivalent shape
1. The short axis and long axis of the ellipse: the length of the short axis and long axis of the ellipse which has the same area and the same perimeter.
2. The ellipticity: the ratio of the long axis length to the short axis length. The ratio increases as the long axis length of the equivalent ellipse increases, and the ratio becomes 1 as the shape of equivalent ellipse approximates a circle.
3. The side ratio of rectangle: the ratio of the long axis length to short axis length in the equivalent rectangle. The side ratio of the rectangle increases as the rectangle becomes longer and thinner, and the ratio becomes 1 as the rectangle approximates a quadrate.

4.6.6 Features of morphology
1. The inertia moment: this feature shows the distribution of pixels to the centre of inertia.
2. The stretching factor: the ratio of maximal segment of the gland to the average vertical segment. The factor increases as the gland becomes larger and thinner.
3. The degree of dispersion: the ratio of the gland area to the minimal rectangle including a gland. The value ranges from 0 to 1, and becomes 1 as the gland shape approximates a rectangle.
4. Heywood diameter: the ratio of gland perimeter to circle perimeter with the same area as the gland. Heywood diameter becomes 1 as the shape of the object approximates a circle.
5. The hydraulic radius: the ratio of area to perimeter in a gland.

6. Waddel disk radius: the radius of a circle with the same area as the gland.
7. Nuclear to cytoplasmic ratio: this index is used by pathologists in actual diagnosis.

4.7 Textural feature extraction
4.7.1 Textural feature extraction

Essentially, the aim of our system is to diagnose gastric tumours using morphological features of the glands, cytoplasm, and nucleus. However, with increasing tumor grade it is difficult to separate the individual nuclei. Therefore we use the following textural features within the cytoplasm instead of computing the morphological features of individual nuclei. We determine textural features by intensity histogram, difference statistics, and co-occurrence matrix.

1. Intensity histogram

In an image the normalized histogram $p(l)$ is characterized by the following equation, using the intensity level L ($L=256$) and the intensity histogram $h(l)$ ($0 \leq l \leq L-1$).

$$p(l) = \frac{h(l)}{\sum_{l=0}^{L-1} h(l)} \tag{14}$$

The above normalized histogram indicates a probability distribution of intensity. By treating the intensity histogram as the probability distribution, the computed result does not depend on the area of region. We use seven scalar values computed by the probability distribution $p(l)$ as features in the intensity histogram. The equations for these features are shown below.

$$MEN = \sum_{l=0}^{L-1} lp(l) \tag{14}$$

$$CNT = \sum_{l=0}^{L-1} l^2 p(l) \tag{15}$$

$$VAR = \sum_{l=0}^{L-1} (l - MEN)^2 p(l) \tag{16}$$

$$SKW = \frac{1}{VAR^3} \left\{ \sum_{l=0}^{L-1} (l - MEN)^3 p(l) \right\}^2 \tag{17}$$

$$KRT = \frac{1}{VAR^2} \sum_{l=0}^{L-1} (l - MEN)^4 p(l) \tag{18}$$

$$EGY = \sum_{l=0}^{L-1} p^2(l) \tag{19}$$

$$EPY = -\sum_{l=0}^{L-1} p(l) \log p(l) \tag{20}$$

In the above equation, MEN means the average of the intensity values of all the pixels in an image. CNT indicates contrast and is the second moment around the origin. CNT becomes large when the histogram distribution slants to the strong intensity side. VAR means variance and is the second moment around the variance or the average. VAR becomes large when the number of pixels with values far from the average is large. SKW means skewness and is large when the histogram shape is distorted from the normal distribution. KRT means kurtorsis and shows a degree of concentration ("peakedness") compared with the normal distribution. EGY indicates energy and the angular second moment. EGY is large when the histogram distribution focuses on the pixels with a particular intensity. EPY means entropy and has a large value when the pixels within the reference region have different intensities.

2.   Difference statistics

We construct a histogram with frequencies with the intensity difference l between pixels at the positional relation of $\delta = (r, \theta)$ within a reference region, as shown in Figure 12. Like the intensity histogram, the intensity probability $p_\delta(l)$ is constructed from the frequency of the histogram. The seven above-mentioned scalar features are computed by replacing $p(l)$ by $p_\delta(l)$. In this study MEN, EGY, EPY, and VAR are selected by the after-mentioned test.
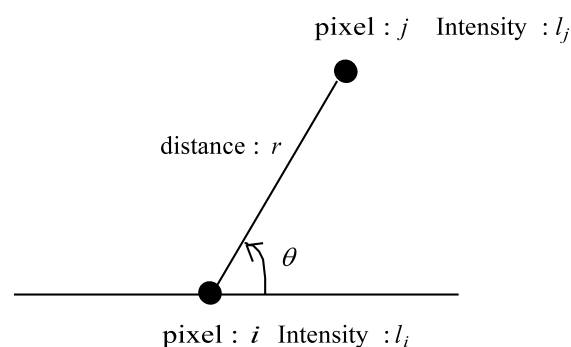

Fig. 12. The positional relationship of a pixel pair

3.   Co-occurrence matrix

Like the difference statistics, the co-occurrence matrix is the method that quantifies the relationship between pixels at the positional relationship of $\delta = (r, \theta)$. When the reference pixel is i and has an intensity of l_i, the pixel j is separately placed at distance of δ and has an intensity of l_j, and the probability is defined by $P_\delta(l_i, l_j)$. The matrix obtained by $P_\delta(l_i, l_j)$ is called a co-occurence matrix. When the intensity level is L, the co-occurence matrix has a size of $L \times L$. Figure 13 shows an example of a co-occurrence matrix. Figure 13(1) shows part of an image and each small square represents a pixel within the image. The numbers in the squares in Figure 13(a) represents the intensity values in each pixel. Figure 13(b) shows a co-occurence matrix when the distance $r=1$ and the angle $\theta=0°$, and Figure 13(c) shows a co-occurence matrix when the distance $r=1$ and the angle $\theta=90°$.

We use 13 textural features obtained from the co-occurrence matrix. Before computation of these textural features, we define $P_x(l_i)$, $P_y(l_j)$, $P_{x-y}(k)$, $P_{x+y}(k)$, μ_x, μ_y, σ_x^2, σ_y^2, HX, HY, HXY1, and HXY2 as follows.

$$P_x(l_i) = \sum_{l_i=0}^{L-1} P_\delta(l_i, l_j) \tag{21}$$

0	0	1	1	1	2
3	0	0	1	1	1
3	0	2	2	1	1
3	3	3	3	2	3
0	0	2	2	2	0
1	0	0	2	2	1

Intensity

	0	1	2	3
0	4	2	3	0
1	1	5	1	0
2	1	2	4	1
3	2	0	1	3

Intensity

	0	1	2	3
0	3	2	1	2
1	1	4	2	1
2	1	1	2	2
3	3	0	2	2

(a). Object image   (b). $p_\delta(l_i, l_j)$ ($\delta = (1, 0°)$)   (c). $p_\delta(l_i, l_j)$ ($\delta = (1, 90°)$)
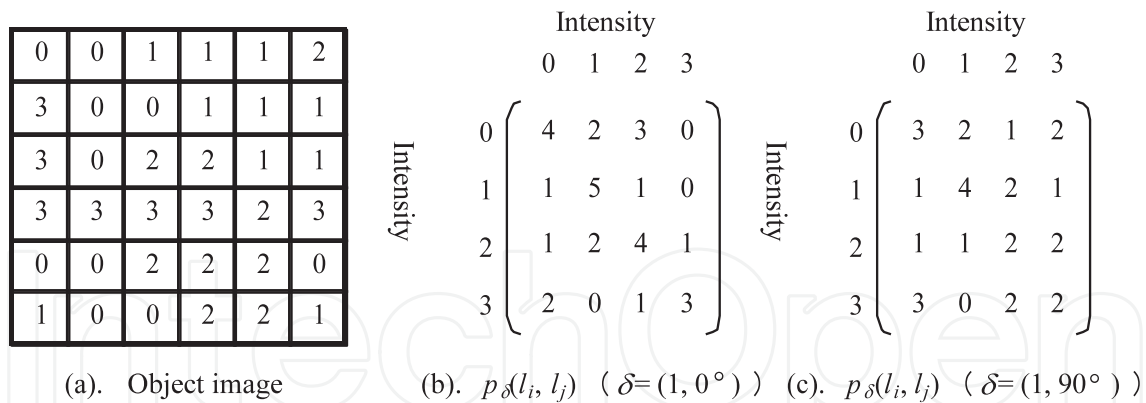
Fig. 13. Example of a co-occurrence matrix

$$P_y(l_j) = \sum_{l_i=0}^{L-1} P_\delta(l_i, l_j) \tag{22}$$

$$P_{x-y}(k) = \sum_{l_i=0}^{L-1}\sum_{l_j=0}^{L-1} P_\delta(l_i, l_j) \quad (0 \le k = |l_i - l_j| \le L-1) \tag{23}$$

$$P_{x+y}(k) = \sum_{l_i=0}^{L-1}\sum_{l_j=0}^{L-1} P_\delta(l_i, l_j) \quad (0 \le k = l_i + l_j \le 2L-2) \tag{24}$$

$$\mu_x = \sum_{l_i=0}^{L-1} l_i P_x(l_i) \tag{25}$$

$$\mu_y = \sum_{l_j=0}^{L-1} l_j P_y(l_j) \tag{26}$$

$$\sigma_x^2 = \sum_{l_i=0}^{L-1} (l_i - \mu_x)^2 P_x(l_i) \tag{27}$$

$$\sigma_y^2 = \sum_{l_j=0}^{L-1} (l_j - \mu_y)^2 P_y(l_j) \tag{28}$$

By using the above definitions, the textural features are computed as shown in the next equations.

$$\text{ASM} = \sum_{l_i=0}^{L-1}\sum_{l_j=0}^{L-1} \left\{ P_\delta(l_i, l_j) \right\}^2 \tag{28}$$

$$\text{CNT} = \sum_{k=0}^{L-1} k^2 P_{x-y}(k) \tag{29}$$

$$CRR = \frac{\sum_{l_i=0}^{L-1}\sum_{l_j=0}^{L-1} l_i l_j P_\delta(l_i,l_j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{30}$$

$$VAR = \sum_{l_i=0}^{L-1}\sum_{l_j=0}^{L-1} (l_i - \mu_x)^2 P_\delta(l_i,l_j) \tag{31}$$

$$IDM = \sum_{l_i=0}^{L-1}\sum_{l_j=0}^{L-1} \frac{1}{1+(l_i-l_j)^2} P_\delta(l_i,l_j) \tag{32}$$

$$SAV = \sum_{k=0}^{2L-2} k P_{x+y}(k) \tag{33}$$

$$SVR = \sum_{k=0}^{2L-2} (k-SAV)^2 P_{x+y}(k) \tag{34}$$

$$SEP = -\sum_{k=0}^{2L-2} P_{x+y}(k)\log\left\{P_{x+y}(k)\right\} \tag{35}$$

$$EPY = -\sum_{l_i=0}^{L-1}\sum_{l_j=0}^{L-1} P_\delta(l_i,l_j)\log\left\{P_\delta(l_i,l_j)\right\} \tag{36}$$

$$DVR = \sum_{k=0}^{L-1}\left\{k-\sum_{k=0}^{L-1} k P_{x-y}(k)\right\}^2 P_{x-y}(k) \tag{37}$$

$$DEP = -\sum_{k=0}^{L-1} P_{x-y}(k)\log\left\{P_{x-y}(k)\right\} \tag{38}$$

$$IM1 = \frac{EPY - HXY1}{\max\{HX,HY\}} \tag{39}$$

$$IM2 = \left[1-\exp\left\{-2\left(HXY2-EPY\right)\right\}\right]^{1/2} \tag{40}$$

$$HXY = -\sum_{l_i=0}^{L-1}\sum_{l_j=0}^{L-1} P_\delta(l_i,l_j)\log\{P_\delta(l_i,l_j)\} \tag{41}$$

$$HX = -\sum_{l_i=0}^{L-1} P_x(l_i)\log\{P_x(l_i)\} \tag{42}$$

$$HY = -\sum_{l_j=0}^{L-1} P_y(l_j) \log\{P_y(l_j)\} \tag{43}$$

$$HXY1 = -\sum_{l_i=0}^{L-1} \sum_{l_j=0}^{L-1} P_\delta(l_i, l_j) \log\{P_x(l_i) P_y(l_j)\} \tag{44}$$

$$HXY2 = -\sum_{l_i=0}^{L-1} \sum_{l_j=0}^{L-1} P_x(l_i) P_y(l_j) \log\{P_x(l_i) P_y(l_j)\} \tag{45}$$

Since the above features are mathematically defined by the co-occurrence matrix, it is difficult to explain their physical meanings.

In our study a small square of 16x16 pixels is clipped from the tumour region for textural analysis. We compute the textural features with respect to this small region. For the analysis of nuclear distribution, an area near the gland base is desirable. We use the next procedure for clipping the small regions.

1. A raster scan is performed using a small square with 16x16 pixels within a two-level image of the gland region.
2. The raster scan is finished when all the intensity values are 0 in the small region.
3. The textural analysis is performed by Y component image with the coordinates that are obtained at the end of the scan.

Figure 14 shows the clipped 16x16-pixel region. Figure 14(a) shows a small square obtained by raster scanning and Figure 14(b) shows the Y component image superimposed on the specified small region.

(a) Binarized image   (b) Y image
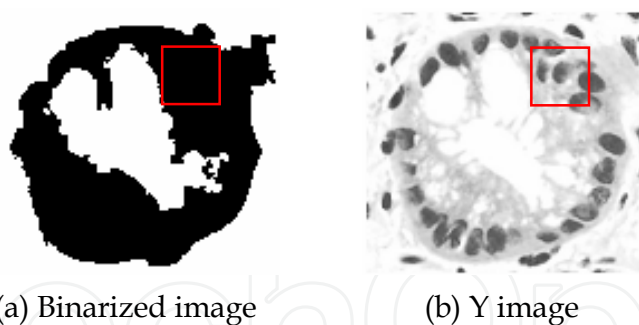
Fig. 14. The small region obtained for the computation of textural features

4.8 Principal component analysis

The principal component analysis is performed to reduce the number of dimensions after the computation of shape and textural features. When variables are x_i (i=1 to n) and coefficients are a_i (i=1 to n), the principal component Z is defined by the next equation.

$$Z = a_1 \cdot x_1 + a_2 \cdot x_2 + \cdots + a_n \cdot x_n \tag{46}$$

For the principal component, the coefficients a_i are obtained under the next condition.

$$a_1^2 + a_2^2 + \cdots + a_n^2 = 1 \tag{47}$$

The coefficients are obtained by the eigenvalues λ_i of the coefficient correlation matrix. The contribution rate of the obtained principal component is computed by the following equation. In this study the principal components are selected so that the cumulative contribution rate becomes about 90%.

$$\text{contribution rate} = \frac{\lambda_i}{p} \times 100 \, [\%] \qquad (48)$$

4.9 Discriminant method

The discriminant method is performed for classifying the cases into three classes (non-neoplastic, adenoma, and adenocarcinoma) after the shape and textural features and the principal component score are computed. The procedure of discrimination is shown below.

Step 1.   F test for homoscedasticity
Step 2.   t-test for average difference
Step 3.   Stepwise method
Step 4.   Discriminant method

In step 1 all 54 features are tested to determine whether each feature has similar variance in each tumour. In step 2 all the features are tested to determine whether they have similar averages in each tumour. Homoscedasticity must be tested because the method by which the t test is conducted varies depending on whether homoscedasticity is present or not. When a feature des not have average difference, the feature is removed from among the candidates for classification. In step 3 the features for step 4 are selected. In step 4 pattern classification is performed by the discriminant method.

5. Results and discussion

The tumour image information used in our system is shown in Table 3. For computation of shape features we used images with 896x672 pixels, which was reduced at a ratio of 40% of the original image. For computation of textural features we used images with 444x336 pixels, reduced at a ratio of 40% of the original image.

Tumor	Number of images	Number of glands
Not malignant	8	28
Gastric adenoma	19	84
Gastric cancer	6	46

Table 3. Number of images corresponding to each category

5.1 Pre-processing and Post-processing

The region extraction results for nuclei and glands are shown in Figures 15 and 16. Figures are photographed at the same magnitude. In these figures, (a) is the original image including a gland, (b) is the extracted nuclear region, and (c) is the extracted gland (cytoplasm) region.

Using the threshold determined by the Laplacian histogram and discriminant method, the nuclear and glandular regions were precisely extracted for most of the images. Although the nuclear region was precisely extracted for all the images, extraction was less precise for

glandular regions. Figure 17 shows one of the unsuccessful attempts. The reason for the lack of success is that the boundary between the background and gland is not clear.
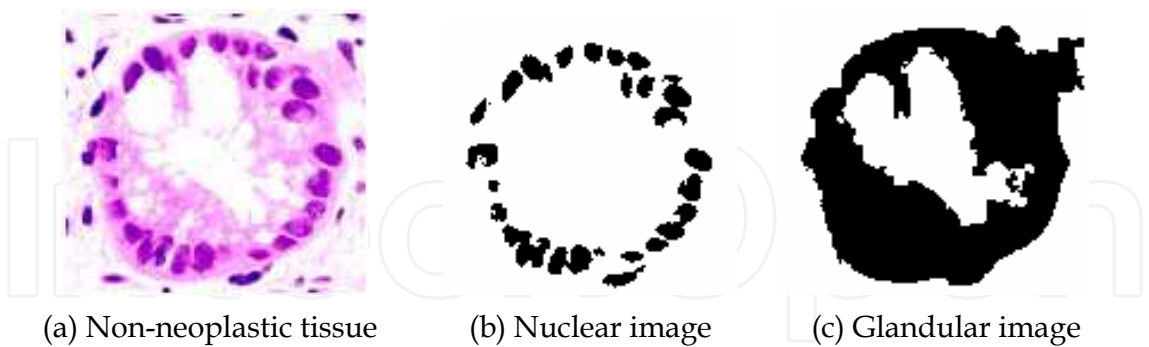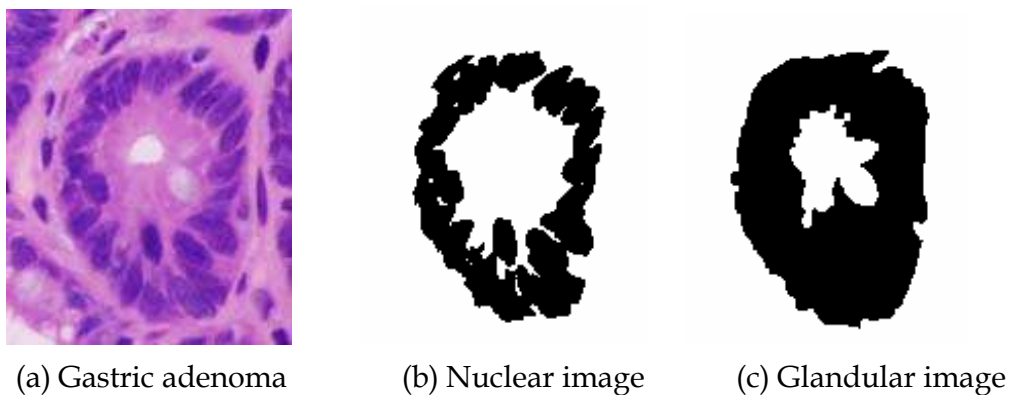

(a) Non-neoplastic tissue   (b) Nuclear image   (c) Glandular image

Fig. 15. Extraction results for non-neoplastic tissue

(a) Gastric adenoma   (b) Nuclear image   (c) Glandular image

Fig. 16. Extraction results for gastric adenoma
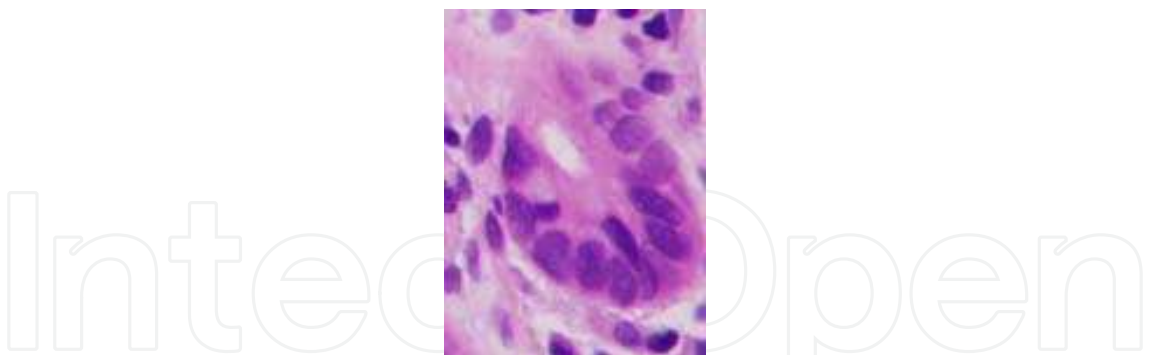
Fig. 17. Image for which extraction of the gland region was unsuccessful

5.2 Textural features

We selected 14 textural features among all the candidate features by the principal component analysis and discriminant methods. The following parameters were derived from difference statistics and co-occurrence matrix at the computation of these textural features.

The results of the computed textural features are shown in Figure 18. Figure 18 (a) shows the values of SKW obtained from the intensity histogram and figure (b) shows those of VAR. In both figures blue plots show cancer cases, pink plots indicate adenoma cases, and yellow plots show the non-neoplastic cases. Values of SKW for cancer and non-neoplastic lesions

are similar, and those of cancer and adenoma are similar. For example, when we classify all cases into 3 classes, we first classify them into 2 classes according to Figure 18(a); one being cancer and non-neoplastic lesions and the other adenoma. Next we classify the class of cancer and non-neoplastic lesions into the class of cancer and the class of non-neoplastic lesions. In this way we can classify all cases into 3 classes.

Features	invariables	value
Gray-Level Difference Matrix	distance : r	4
	angle : θ	0°
co-occurrence matrix	distance : r	4
	angle : θ	0°

Table 4. Parameters for computation of difference statistics and co-occurrence matrix
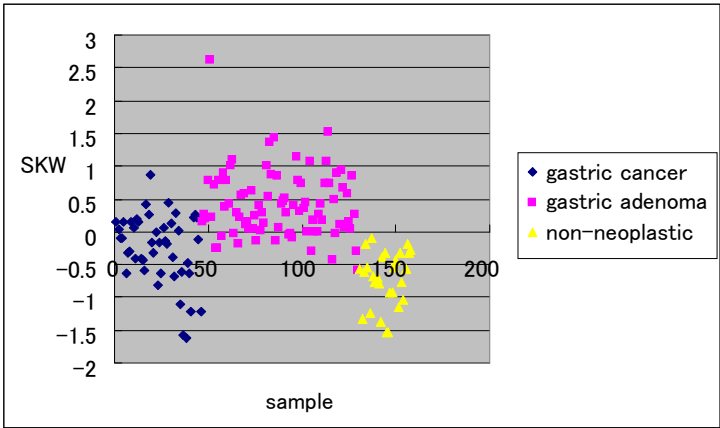
(a) SKW (density histogram)
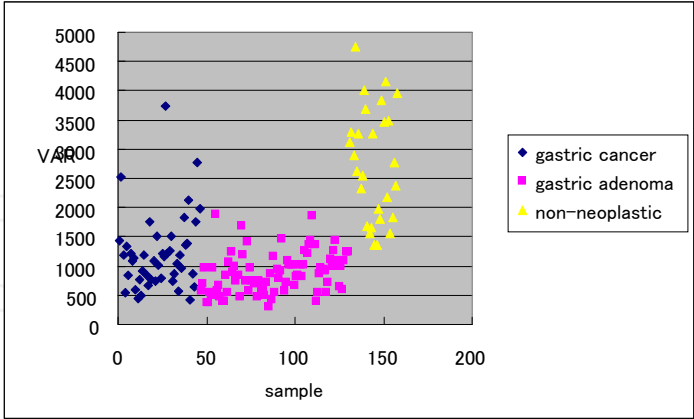
(b) VAR (density histogram)

Fig. 18. Results of computed textural features

5.3 Principal component analysis
In this study we computed 40 shape features and 14 textural features for discrimination of tumor severity. The relationship between factor loading and each principal component is shown in Figure 19. The vertical axis indicates factor loading and the horizontal axis shows

the number of principal components. In this study we selected the principal components with an eigenvalue of more than 1. Consequently, the 1st to the 9th principal components were selected, and the cumulative contribution rate was 91%.
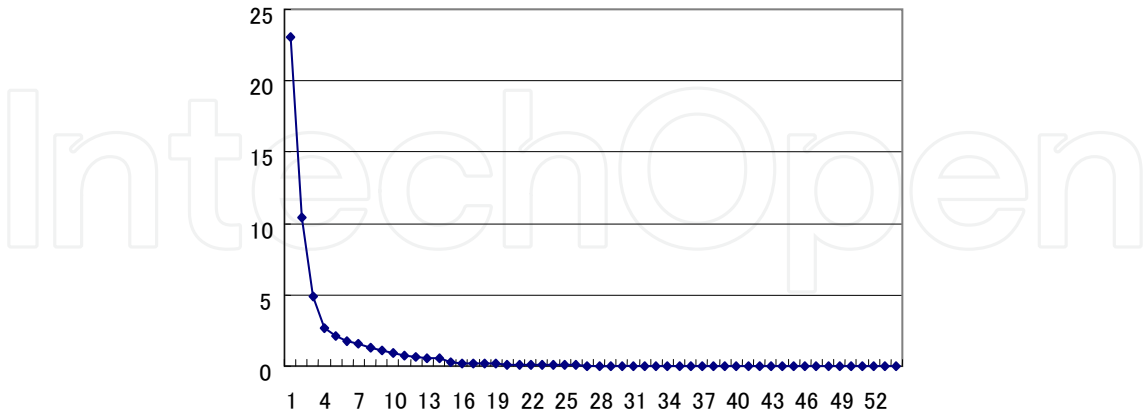

Fig. 19. Factor loading of each principal component

5.4 Stepwise method

In the previous section, 40 shape features and 14 textural features were computed and then selected by principal component analysis. However, some features revealed a strong correlation with each other or did not contribute towards discrimination. In this situation the features used for the discriminant analysis are selected by the stepwise method. The following seven features among the 54 features were selected by the stepwise method.

1. AVE in the intensity histogram
2. VAR in the intensity histogram
3. EPY in the co-occurrence matrix
4. Nuclear to cytoplasmic ratio
5. Stretching factor
6. The average number of chords in glands in the horizontal axis
7. SKW in the intensity histogram

5.5 Discriminant analysis
5.5.1 Discriminant analysis by principal component score

The Table 5 shows the result of pattern classification using eight principal components without the third principal component. "Actual tumour category" in this Table means the classification according to the pathologist and "assigned tumour category" means the classification according to the computer analysis.

According to the computer analysis, the correct classification rate was 88% and the incorrect classification rate was 12%. Further, both the false negative rate and false positive rate were about 6%.

5.5.2 Discriminant analysis by shape features and textural features

Table 6 shows the results by shape features and textural features using seven features that obtained by stepwise method. "Actual tumour category" in this Table means the classification according to the pathologist and "assigned tumour category" means the classification according to the computer analysis.

		Assigned tumour		
		Gastric cancer	Gastric adenoma	Not malignant
Actual tumour	Gastric cancer	37 (80%)	6 (13%)	3 (7%)
	Gastric adenoma	9 (11%)	75 (89%)	0 (0%)
	Not malignant	1 (4%)	0 (0%)	27 (96%)

Table 5. Result of discriminant analysis using principal components score

		Assigned tumour		
		Gastric cancer	Gastric adenoma	Not malignant
Actual tumour	Gastric cancer	38 (82%)	4 (9%)	4 (9%)
	Gastric adenoma	6 (7%)	78 (93%)	0 (0%)
	Not malignant	1 (4%)	0 (0%)	27 (96%)

Table 6. Result of discriminant analysis using features

According to this method, the correct classification rate was 91% and incorrect classification rate was about 9%. Further, both the false negative rate and false positive rate were about 5%.

5.5.3 Glandular images that were unsuccessfully classified
Here we consider the glandular images that were unsuccessfully classified. In tumour diagnosis in particular, it is important to reduce the false negative rate. Representative images that were unsuccessfully classified are shown below.
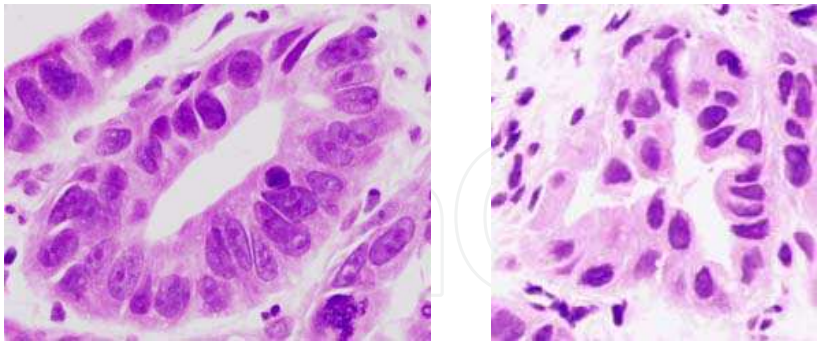

Fig. 20. Glandular images that were misclassified

The above images have glands that appear to suggest cancer, but they were classified as non-neoplastic. The reason for the misclassification is that the gland in the left image has low apytia and is similar to glands seen in non-neoplastic tissue. The gland in the right image also has low apytia and is not enlarged.

5.5.4 Diagnosis by weighting combination
In the previous section each glandular image was diagnosed by the discriminant method. But in the actual hospital setting, pathologists formulate a diagnosis only by looking at the

shape and distribution of all glands. Here we propose a more comprehensive diagnosis method by averaging the classification score of each gland. Gastric cancer would score 3; gastric adenoma, 2; and non-neoplastic lesions, 1, as shown in Table 7.

tumor	weight
Gastric cancer	3
Gastric adenoma	2
Not malignant	1

Table 7. Scores for cancer, adenoma, and non-neoplastic lesions

Table 8 shows the diagnosis results of all the cases, obtained by computing an average of all glandular scores in the original image. This table shows that the correct diagnosis rate was 91% and the wrong diagnosis rate about 9% by this method. Moreover, the false negative rate was 0% and the false positive rate was about 9%.

		assigned tumor		
		Gastric cancer	Gastric adenoma	Not malignant
actual tumor	Gastric cancer	7 (88%)	1 (13%)	0 (0%)
	Gastric adenoma	2 (11%)	17 (89%)	0 (0%)
	Not malignant	0 (0%)	0 (0%)	7 (100%)

Table 8. Classification of gastric tumours using severity score

6. Conclusion

This chapter deals with computer-aided diagnosis of glandular tumours for internet-based telemedicine. In many countries including Japan, the scarcity of pathologists is one of the problems in medical diagnosis, and this Internet-based diagnosis system goes some way toward addressing this issue. In this chapter we have reported a diagnosis system for glandular tumours such as those of the stomach, colon, prostate, and breast. Here we explain the diagnosis system with a central focus on gastric tumour.

First, we discussed the difference between shape features and textural features of non-neoplastic lesions, adenoma, and cancer, and perform numerical conversion of atypism and anisocytosis. We created a system for extracting the nuclei and glands and tested this method in gastric tumours. As the result the system using binarisation of the red and green components of a tumour image can extract a region with an arbitrary density when the region is surrounded by a relatively clear boundary. Next, we computed 40 shape features and 14 textural features and determined the score of principal component analysis. The discriminant result reached 88% with eight principal components, and 91% with seven selected features. Moreover, when we performed weighting based on the discriminated gland and classified the original image into three categories of severity, the discriminant ratio reached 91% by the principal component method and 94% by the selected features method. Our proposed system therefore appears valid for the diagnosis of gastric tumours.

With this method, we must remove each gland region from the original image. In future work, we will focus on a method of automatically extracting each gland. In higher grade cases most of the glands are unformed, and creation of an automatic gland shape extraction algorithm is a difficult problem facing Internet-based diagnosis systems.

7. References

Bamford, P. & Lovell, B. (1998). Unsupervised cell nucleus segmentation with active contours, Signal Processing, Vol.71, pp.203-213

O'Gorman, L., Snderson, A.C. & Preston Jr, K. (1983). Image segmentation and nucleus classification for automated tissue section analysis, Proceedings of the IEEE Computer Society Conference on computer Vision and Pattern Recognition, pp.89-94

Okii, H. & Uozumi, T. (2003). Automatic Feature Extraction from Breast Tumor Images Using Artificial Organisms, IEICE TRANS. INF. & SYST., Vol.E86-D, NO.5 MAY pp.964-972

Marghani, K.A., Dlay, S.S., Sharif, B.S. & Sims, A.J. (2003). Automated morphological analysis approach for classifying colorectal microscopic images, Proc. Of SPIE, Vol.5267, pp.240-249

Esgair, A.N. & Naguib, R.N.G. (1998). Microscopic Image Analysis for Quantitive Measurement and Feature Idenitification of Normal and Cancerous Colonic Mucosa, IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDECINES, Vol.2, No.3, SEPTEMBER pp.197-203

Wu, H.-S. & Gil, J. (2003). Linear Clustering for Segmentation of Color Microscopic Lung Cell Images, JOURNAL OF IMAGING SCIENCE AND TECHNOLOGY, Volume 47, Number 2, March/April, pp.161-170

Cuisenaire, O., Romeo, E., Veraart, C. & Macq, B. (1997). Automatic segmentation and measurement of axons in microscopic images, SPIE proceedings SOCIETY of PHOTO-OPTICAL Instrumentally Engineers, vol.3661-94, No.9, pp.4211-4216

Baak, J.P.A., Delemarre, J.F.M., Langley & F.A. et al (1986). Grading ovarian tumors, Evaluation of decision making by different pathologists, Anal Quant. Cytol. Histol. 8,, pp.349-353

Chan, F.H.Y., Lam, F.K. & Zhu, H. (1998). Adaptive Thresholding by Variational Method, IEEE TRANSACTION ON IMAGE PROCESSING, Vol. 7, NO.3, MARCH, pp.468-473

Jafari-Khouzani, K. & Soltanian-Zadeh, H. (2003). Multiwavelet Grading of Pathological Images of Prostate, IEEE TRANSACTION ON BIOMEDEICAL ENGINEERING, Vol. 50, NO.6, JUNE, pp.697-704

Thiran, J.-P. & Macq, B. (1996). Morphological Feature Extraction for the Classification of Digital Images of Cancerous Tissues, IEEE TRANSACTION ON BIOMEDEICAL ENGINEERING, Vol. 43, NO.10, OCTOBER, pp.1011-1020

Tanaka, T., Uchino, Y. & Oka, T. (2006). Classification of Gastric Tumors using Shape Features of Gland, IEEJ Transaction on Electronics, Information and Systems, Vol.126, No.10, pp.1242-1248

Suzuki, A. & Tanaka, T. (2007). Telemedicine system of pathology image using the Internet, SICE Annual Conference 2007, pp.694-697
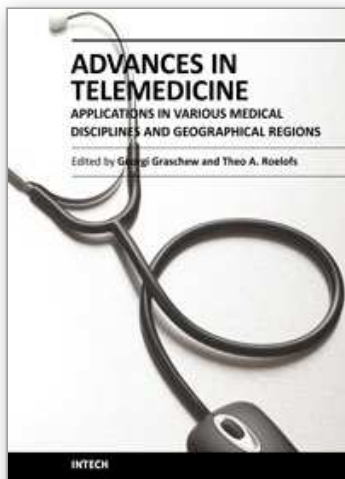
Advances in Telemedicine: Applications in Various Medical Disciplines and Geographical Regions

Edited by Prof. Georgi Graschew

ISBN 978-953-307-161-9

Hard cover, 296 pages

Publisher InTech

Published online 22, March, 2011

Published in print edition March, 2011

Innovative developments in information and communication technologies (ICT) irrevocably change our lives and enable new possibilities for society. Telemedicine, which can be defined as novel ICT-enabled medical services that help to overcome classical barriers in space and time, definitely profits from this trend. Through Telemedicine patients can access medical expertise that may not be available at the patient's site. Telemedicine services can range from simply sending a fax message to a colleague to the use of broadband networks with multimodal video- and data streaming for second opinioning as well as medical telepresence. Telemedicine is more and more evolving into a multidisciplinary approach. This book project "Advances in Telemedicine" has been conceived to reflect this broad view and therefore has been split into two volumes, each covering specific themes: Volume 1: Technologies, Enabling Factors and Scenarios; Volume 2: Applications in Various Medical Disciplines and Geographical Regions. The current Volume 2 is structured into the following thematic sections: Cardiovascular Applications; Applications for Diabetes, Pregnancy and Prenatal Medicine; Further Selected Medical Applications; Regional Applications.

INTECH
open science | open minds