Proceedings of International Conference on Image
Processing and Character Recognition

# An Automatic Image Analysis System for Quantitative Assessment of Liver Fibrosis

Yung-Nien Sun and Ming-Huwi Horng

Institute of Information Engineering, National Cheng Kung University Hospital, Tainan, Taiwan,

## Abstract

*Liver biopsy is the standard for the evaluation of liver diseases. The severity of fibrosis obtained by biopsy test is considered to reflect the stage of chronic liver disease. However, the current biopsy reading remains quite subjective, or at best semiquantitative. Therefore, developing an objective and reliable image analysis system for evaluating the liver fibrosis will be a major advance in the diagnosis and staging of chronic liver diseases. In this paper, we develop an automatic image analysis system, which consists of a microscope, a computer-driven slide-driver, and the software for image acquisition, processing and data analysis. Some image analysis procedures, including color model selection, histogram-based normalization, clustering, moment-persevering thresholding and ranking filter, have been developed and employed in this system for tissue characterization. In addition, the computerized motor driver and the x-y directional stage to move specimens, are designed and installed on an optical microscope to construct a fully automated system. The system is capable of computing the percentage of fibrous area to the whole liver tissue area as an index, called computer morphometry (CM) score to reflect the severity of liver fibrosis. Experimental results show that the proposed method correlates well these methods as expected, however, is more stable and quantitative than these conventional methods.*

*Keywords: Automatic assessment, liver fibrosis, color image analysis, computer morphometry*

## I. Introduction

Evaluating hepatic fibrosis on pathological sections greatly facilitates the staging and follow-up of chronic liver disease. The disease severity may range from a healthy carrier to a decompensated cirrhosis. However, erroneous assessment arises from the subjective experience and eye measurement. With the accelerated developments of microcomputer and moderately mechanical control technology in recent years, an automatic pathological section analysis system coupled with optical microscopy plays an increasingly important role in examining the pathological sections. The high-speed computation of a microcomputer makes analyzing the amount of images captured from the pathological section a reality. The sensor technology, modern charge-coupled devices (CCD) cameras can achieve high spatial resolution and high sensitivity measurement of signals captured from an optical microscope. Additionally, an accurate mechanical apparatus collaborating with electronic control provides automatic and precise section positioning capability. This progression motivates our ongoing developments of an automatic system for pathological section analysis.

Previous studies introduced image analysis in the morphological assessment of hepatic fibrosis[1]. This technique is based on employing an image analyzer. In this system a computer equipped with a unique software and imagery card acquires images by using a color video-camera placed on a light microscope. The system includes two high resolution video graphic array monitors: one used to display the results of regions segmentation and the other used to manually follow the image acquisition by the imagery card. The image segmentation is basically a process of partitioning the color images by interactive thresholding. Such a partitioning leads to a binary image whose regions of fibrosis appear in white and the regions of normal cells appear in black. Determining the threshold value depends on several factors, e.g. the lighting of the optical microscope, the thickness of the section and the stain discoloration of fibrosis. The section's thickness is initially checked at regular intervals. Each liver section assessed using this system is only randomly acquired four or five images at 10x magnification times for image analysis. The limitation causes lower reliability in quantifying the liver fibrosis.

This paper presents an automatic computer system for objectively and quantitatively assessing liver fibrosis in each liver section. The system is called ACM system, which is an acronym for Automatic

Computer Morphometry system. The ACM system consists of three subsystems: image acquisition and section positioning subsystem, sample-training subsystem and processing subsystem. For each liver section, a large number of liver images, each containing many normal cells and liver fibrosis, must be captured for analysis. To provide the continuous image acquisition, the positioning subsystem adopts an x-y directional stage driven by two stepping-motors to accurately locate a proper tissue area in the field-of-view of the microscope for capturing images without human intervention. Additionally, the sample-training subsystem offers a training mechanism to overcome discoloration of the histological section.

The processing subsystem consists of most of the software modules that make the system operate, including the control of other subsystems and the liver region segmentation. In addition, a new severity indicator of measuring the percentage of liver fibrosis occupied within the complete liver section called Computer Morphometry (CM) score is proposed in this system. Thirty-one liver sections stained by Masson stain are recruited for this study. These results are compared with other clinical assessments in our experiments.

## II. Automatic Computer Morphometry System

As mentioned above, an automatic computer morphometry system for quantitatively assessing the severity of the liver fibrosis is proposed in this paper. The ACM system employs image processing techniques associated with mechanical control techniques to continuously acquire images and characterize them. The system diagram depicted in Fig. 1 consists of three subsystems: image acquisition and section positioning subsystem, sample-training subsystem and processing subsystem. Figure 2 presents the processing flow of the ACM system. Each subsystem and the appropriate design consideration are described as follows.

### II.1. Image acquisition and section positioning subsystem

The image acquisition subsystem consists of a microscope (Olympus CH-2, Tokyo, Japan) coupled with a color video camera and a color frame grabber (IMASCAN, Imagraph, TX, USA). Each stained section is placed upon a microscopic x-y directional stage with a pair of zoom lens and viewed in a transmitted light. To obtain the consistent image quality, the microscope light must remain constant from one examination to another and the magnification time remains constant at 108. Each image is captured as $512 \times 480$ pixels with RGB color format. In this system, a specimen with the size $1.2\text{cm} \times 0.8\text{cm}$ is captured into about ninety images by using the section positioning mechanism to automatically locate each image at proper position. Each pixel width in this magnification time is about 0.00208 mm in real world. Owing to the structure of a human eye, all colors are seen as variable combinations of three base color features in the so-called primary RGB (red, green, blue) model. The color images with RGB model are captured by the color video camera and digitized by the 24-bit frame grabber. This process generates a 3-D vector for each pixel whose color features range from 0 to 255.

The positioning subsystem attempts to accurately locate the area of interest of the liver section in the field-of-view of the microscope for acquiring images. An x-y directional stage is designed for this purpose, consisting of a stage driven by two pivots and a pathological section locating fixture used to fix the liver section. Two stepping-motors controlled by a dual 8255 controller of the subsystem are used to drive the two pivots along a given direction. One of the two motors moves the x-y directional stage forward or backward and the other can move right or left. To provide more accurate positioning capability, the adequate rotation steps of the motors are explored in advance. The positioning capability can then be automatically or manually maneuvered under the conduct of software in the processing subsystem. Figure 3 presents the graph of the positioning subsystem.

### II.2. Sample-Training subsystem

The sample-training subsystem provides manually training mechanism to extract the powerful features for analysis. The system includes the training methods and a database for storing these features. In applying the liver tissue image segmentation, the canonical analysis is performed to obtain the powerful color features for separating different liver regions. The training features are stored in a database for later segmentation.

The color model is a mathematical or geometrical representation of color. Most of the color models attempt to relate the representation of human color perception. However, no general model can be applied to all

applications. Thus, the initial step of color image segmentation entails defining a highly effective color model for processing the liver tissue images. The canonical transform is based on the statistical properties of the Fisher discriminate functions to linearly project the 3-D RGB color model to a 2-D color model whose chromatic features are rich in powerful information for liver tissue image segmentation. Experimentally, three region classes are presented in liver tissue image: regions of the liver fibrosis, regions of normal cells and background. Thus, segmenting the liver tissue image is a 3-class data classification problem. The liver cell region is labeled here as $C_1$, the liver fibrosis region as $C_2$ and the background as $C_3$ for training. Assume that a set of samples, $x_1, x_2, ...., x_n$ exist with 3-D RGB color model from some training images. Every region class $C_i$ consists of $n_i$ training samples. The generalization for the within-class scatter matrix $S_w$ and the total scatter matrix $S_T$ are defined as

$$S_W = \sum_{i=1}^{3} S_i , \tag{1}$$

$$S_B = \frac{1}{3}\sum_{i=1}^{3} (m_i - m)(m_i - m)^t \tag{2}$$

where

$$S_i = \frac{1}{n_i - 1}\sum_{x \in C_i} (x - m_i)(x - m_i)^t, \tag{3}$$

$$m_i = \frac{1}{n_i}\sum_{x \in C_i} x \text{ and } m = \frac{1}{n}\sum_{x} x \tag{4}$$

The Fisher linear discrimination is then defined as that linear function $y = W^t X$ for which the below criterion function is maximized, i.e.

$$J(W) = \frac{\left|W^t S_B W\right|}{\left|W^t S_W W\right|} \tag{5}$$

In this method, the columns of an optimal $W$ are the generalized eigenvectors that correspond to their own eigenvalues in the following equation.

$$S_B W_i = \lambda_i S_W W_i \tag{6}$$

The largest and next large eigenvalues can be found according to the roots of the characteristic polynomial $\left|S_B - \lambda_i S_W\right| = 0$. Generalized eigenvectors corresponding to the two eigenvalues can then be resolved. The new color model defined by Eq. (5) is with the maximum ratio of between-class scatter to within-class scatter. Here, the 2-D color model is defined as $Y_1 Y_2$ color model.

Figure 4 displays a testing liver tissue image. Ninety blocks of interest of an image, each containing $5 \times 5$ pixels, are selected herein to demonstrate the capability for separating the three region classes by using $Y_1 Y_2$ color model. The ninety blocks of interest are equally divided into three classes: liver fibrosis, normal cell and background. The mean of the color features of the ninety blocks are determined. Figure 5 presents these scatter plots based on the $Y_1 Y_2$ color features and original RGB color features. As this figure demonstrates most of the data points of foreground (including the liver cell and liver fibrosis) and background can be divided by using the $Y_1$ color feature. Moreover, the $Y_2$ color features can effectively segment most data points of liver fibrosis regions and liver cell regions. Consequently, these parameters $W$ are stored in the database for later image segmentation.

## II.3. Processing subsystem

Processing tasks such as control of the subsystem, image acquisition and data analysis are all performed on a personal computer. The liver section positioning process is generally time consuming since the driving speed of mechanical stage is rather slow and can not be improved with the present setup. Thus, the system must be accelerated by facilitating the color image segmentation. The processing system is designed to provide an efficient processing capability by incorporating several image analysis procedures, including the color model transform, pattern classification and statistical analysis method. Figure 6 depicts the complete color image segmentation algorithm. The control of subsystem, region segmentation method and data analysis method are discussed in detail as follows:

### *II.3.1. Control of the subsystems*

The processing subsystem controls the entire processing flow of the ACM system. Initially, a trained image is selected to include the three above-mentioned regions that are manually acquired to train the parameters of the canonical transform via the sample-training subsystem. The parameters of canonical transform are stored in a database for later region segmentation. Then, the processing subsystem drives the positioning subsystem to acquire images within the liver section of test specimens for analysis. The areas of fibrosis and the areas of normal cells of each acquired

image are counted by partitioning the three regions by an automatic thresholding method in the processing subsystem. This processing continues until all images have been analyzed.

***II.3.2. Image Segmentation***

Each liver tissue image acquired from liver section is initially transformed to $Y_1Y_2$ color model using the training features stored in the above database. The histogram modification is then performed on the tissue image to normalize the range of the $Y_1$ and $Y_2$ distributions, thereby making the further processes consistent. The $Y_1$ color feature is used to separate the original image into two region classes that are foreground and background. Then, seperating foreground into fibrosis and normal regions is based on the thresholding of the $Y_2$ color feature. Finally, a rank filter is used to filter out the smaller fragments and smoothen the contours of segmented objects.

*Pre-processing*--- Each color image with RGB format is transformed into one image with $Y_1Y_2$ color model. The ramp histogram transform is then performed to normalize its color distibution [2]. The transform is represented as follows:

$$p' = \frac{p''}{N}[H_q(q'')]^{1/2} \tag{7}$$

where $p'$ and $q'$ represent the output and input pixel color values, respectively, $p''$ is the maximum allowable color values, $N$ is the square root of the total number of pixels in the image, and $H_q(q'')$ is the value of the cumulative distribution of $Y_1Y_2$ color features at $q'$.

*Region segmentation*-- The separation of three region classes is performed by an atuomatic theshold selection scheme called moment-preserving thesholding. The moment-preserving threshold involves selecting a threshold value such that all below-threshold gray values in original image $f$ are replaced by $Z_0$ and all above-threshold gray values replaced by $Z_1$. Afterwards the first three moments of image $f$ are still preserved in the resulting bilevel image $g$. The thresholding method is discussed as follows.

Given an image $f$ with $n$ pixels whose color chromatic value at pixel $(x,y)$ is denoted by $f(x,y)$. Let the $i$th moment $m_i$ of $f$ be defined as

$$m_i = (1/n)\sum_x \sum_y f^i(x,y), \quad i=1,2,3. \tag{8}$$

Moments can also be computed from the histogram of $f$ in the manner,

$$m_i = (1/n)\sum_j n_j (z_j)^i = \sum_j p_j (z_j)^i, j = 0,1,...,255. \tag{9}$$

where $n_j$ is the total number of pixels in the image with color chromatic value $z_j$ and $p_j = n_j/n$. We also define $P_0$ and $P_1$ as the fractions of the blow-threshold pixels and the above-threshold pixels in $f$, respectively, then the first three moments of $g$ are just

$$m_i' = \sum_j p_j (z_j)^i, \ j = 0,1. \tag{10}$$

This selection of threshold is based on the following equation:

$$m_i' = m_i, = 1,2,3. \tag{11}$$

To obtain the desired threshold values, Eq. (11) can be solved to find $P_0$ and $P_1$, and then the threshold $t$ can be selected so that

$$P_0 = (1/n)\sum_{z_j \le t} n_j \tag{12}$$

Once the $P_0$, $P_1$ and $t$ are determined, the desired segmentation for the input $Y_1$ and $Y_2$ feature images can be readily accomplished.

*Post-processing* ---- Since using moment-preserving thresholding on the $Y_1$ and $Y_2$ color model does not resolve the spatial relationship among pixels, the subsequent image will be ragged in shape on the tissue contour or with scattered small fragments. To remedy this problem, a rank filter is designed to filter out the fragments and smoothen the contours of objects. The principle of designed rank filter is based on voting logic. The rank filter in our cases is represented as follows:

$$Y(x) = \begin{cases} W_h & \text{for } L_h > T \\ X(x) & \text{otherwise} \end{cases} \tag{13}$$

Where

$Y(x)$= the class of pixel $x$ in the output image,
$X(x)$= the class of pixel $x$ in the input image,
$W_h$= the $h$'th class considered here,

$L_i = \|A_i\|$,

$A_i = \{x \mid X(x-u) = W_i, \forall u \in B\}$, i=1,2,3.

$L_h = \max_i L_i$.

For each pixel $x$ in the image, the new class label of $x$ is determined based on a voting process with the $5 \times 5$ mask B depicted in Fig. 7. In the region of B, if the majority of the neighbors of x, exceeding 65%, belongs to class $W_i$, then the label of x is updated to $W_i$;

otherwise it is unchanged.

### II.4. Statistic Analysis

Total pixels of normal cells and fibrosis in the entire liver section are recorded to calculate the severity index referred to as CM score. The CM score is the percentage of liver fibrosis regions with respect to the whole liver section area. Equation (14) expresses the CM score as follow:

$$CM\,score = \frac{area\ of\ the\ regions\ of\ liver\ fibrosis}{total\ area\ of\ whole\ liver\ tissued} \quad (14)$$

## III. Experimental Result and Discussion

This study has successfully developed an ACM system to quantitate the fibrosis of liver section. The system integrates the techniques of mechanical control together with color microscopic image analysis. In this system, all programs are coded by Visual C++ version 1.5 with a Pentium 60 computer under MS-Windows 3.1 environment. The proposed system provides object-oriented programming and a user-friendly interface for training the sample data. The average processing time of each image, including the acquisition, tissue positioning and image analysis, is about 12.5 seconds.

Thirty-one liver sections with Masson stain are collected for examining the accuracy of assessment and the system's reliability. These sections are obtained from thirty-one patients where sixteen of them are with chronic hepatitis B, twelve have chronic hepatitis C, and three are alcoholics. To demonstrate the accuracy of region segmentation, two liver tissue images taken from different liver sections are illustrated. Figures 8 summarize these results. Figure 8(b) presents the segmented images. Each segmented region is represented by one of three specific colors. Blue denotes liver fibrosis, red presents normal cell and white denotes background. Finally, the edge maps are overlaid to illustrate the segmentation's accuracy, as indicated in Fig. 8(c). Experiments with the set of thirty-one sections reveal that the results are more than 95% consistent with the human estimation in the area ratios.

In the recent decade, Knodell *et al.* [3] propose a popular score widely used to measure liver fibrosis. According to the physician's observation, the score system records five numerical scores which are 0: no fibrosis; 1: fibrous portal expansion, 3: bridging fibrosis; 2: between score 1 and score 3, 4: cirrhosis. Obviously, it is not precise, moreover, it suffers from the clinical expert's subjective observation. Jimenez *et al* proposed an alternative approach named colorimetric method [4]. In this method, the Sirius red and the Fast green are used to bind collagen and non-collagenous proteins. Then, the spectrophotometer accumulates the absorption spectrums of collagen and non-collagenous protein to define the percentage of fibrosis as the following equations.

$$\frac{mg\ collagen}{mg\ total\ proteins} = \frac{mg\ collagen}{mg\ collagen + mg\ noncollagenous\ proteins} \quad (15)$$

$$mg\ collagen = \frac{absorbance\ 540nm - 7.78\% absorbance\ 630nm}{37} \quad (16)$$

$$noncollagenous\ proteins = \frac{absorbance\ 630\ nm}{3} \quad (17)$$

The Knodell's score, colorimetric method and CM score are used to assess thirty-one liver tissue sections. The validity of CM score in measuring liver fibrosis is verified by comparing with the Knodell's score and colorimetric method. The Pearson correlation coefficient, $r_{xy}$, between the CM scores and colorimetric scores is also computed by

$$r_{xy} = \frac{\frac{\sum (X-\bar{X})(Y-\bar{Y})}{N-1}}{\sqrt{\frac{\sum (X-\bar{X})^2}{N-1}}\sqrt{\frac{\sum (Y-\bar{Y})^2}{N-1}}} \quad (18)$$

where the two severity score values are recorded by X and Y.

Figure 9 depicts the correlation relationship between the results of CM scores and colorimetric method, which is the linear regression line with 95% confidence level. As this figure indicates, CM score is highly correlated with the one by colorimetric method. The correlation coefficient is 0.9014 ($p<0.0001$).

The validity of the CM score is also evaluated by comparing with the ranking of the Knodell's score. Two physicians determine the Knodell's ranking of each patient according to their observations and patient's medical records. The ANOVA and point-bisexual correlation analysis is used to compare means of the CM scores according to the number of ranks in the Knodell's score. The test results demonstrate rank 0: $1.800 \pm 0.860$, rank 1: $4.5167 \pm 1.246$, rank 2: $6.6571 \pm 1.829$, rank 3: $12.3975 \pm 3.358$ and rank 4: $19.0444 \pm 10.495$. A convenient nonparametric measurement called Spearman rank coefficient is used to test the correlation between the ranking of CM scores and Knodell's score. The coefficient is given on the basis of Eq. (19).

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \qquad (19)$$

where the two sets of values are ranked from 1 to $n$ and $d_i$ is the difference in ranking for each pair of observations. The correlation is 0.9136 (95% confidence level and $p<0.001$). A similar test is also used to transfer the colorimetric scores into five scores. Consequently, the Spearman rank coefficient between the colorimetric method and Knodell's score is only 0.8191 (95% confidence level and $p<0.01$). These results demonstrate that CM score is much more correlated with the Knodell's score conventionally used in clinical diagnosis.

## V. Conclusion

The paper presents a reliable and powerful ACM system to effectively deal with the fibrosis quantification of liver section. The proposed system provides functions of auto-driving, auto-acquisition, image processing and data analysis to quantitatively assess the severity of liver fibrosis. Some problems due to tissue discoloration are naturally overcome by the designed mechanism of sample-training. The positioning subsystem equipped with an x-y directional stage driven by two stepping-motors can locate the interested tissue area properly in the field-of-view of the microscope for image acquisition. The processing subsystem coordinates all subsystems to provide the automatic assessment capability. Additionally, the proposed image analysis procedures can effectively separate liver tissues into three different regions and then compute the statistics of liver fibrosis.

Experimental results demonstrate that the proposed CM score is a highly reliable indicator for the severity of liver fibrosis. Among the three adopted methods in this paper, Knodell's score is the easiest and most subjective, thereby making it prone to generate intra and inter-observer errors due to variation in human observations. The colorimetric method is an alternative to assess liver fibrosis. However, the cumbersome steps of deparaffinning and staining cause the loss of tissue during wash procedure and generally suffer from elution error of the stained tissue. Since the tissue has been destroyed, it is also impossible to reproduce the assessment parameters or even to observe the original specimens again. The CM score generated by ACM system is the most objective and fastest among the three methods. Only 14.5 minutes are required to process a liver section with size of $1.2 \times 0.8\ cm^2$ which is digitized into ninety $512 \times 480$ color images. The inspection covers the complete area of the prescribed section that significantly improves the current scoring based on partial samples. The processing speed also satisfies the requirement for clinical practice. The system is implemented in a Pentium 60 personal computer with a user-friendly interface by object-oriented programming techniques. The proposed system can be readily extended to the other pathological analyses by using the visual programming hierarchy.

## Acknowledgments

The authors would like to thank the National Science Council, R.O.C and the Department of Public Health for partially supporting this work under Grant Nos. NSC 85-2213-E-006-075 and DOH85-HR-416, respectively. Additionally, the authors also wish to thank Dr. Lin Xi-Zhang for his help of preparing and diagnosing these pathological liver tissues.

## References

[1] **Erler BS, Chein K, Marchevsky AM:** An image analysis workstation for the pathology laboratory, *Modern Pathology*, 6: 612-618, 1993.

[2] **Sun YN, Wu CS, Lin XZ and Chou NH:** Color image analysis for liver tissue classification. *Optical Engineering*, 31:1609-1615. 1993.

[3] **Knodell G, Ishak KG , Black WC, *et al*:** Formulation and application of numerical scoring system for assessing histological activity in asymptomatic chronic active hepatitis. *Hepatology*, 1: 431-435, 1981.

[4] **Jimenez W, Pares A, Cabalieria J. *et al*:** Measurement of fibrosis in needle liver biopsies. *Hepatology*, 5: 815-8, 1985.

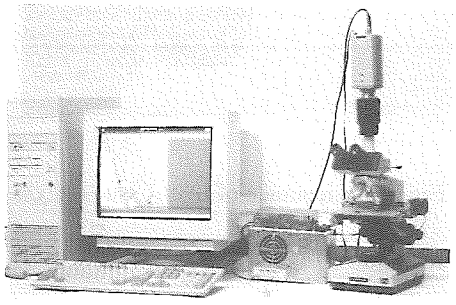

Fig 1. ACM system for analysing the liver

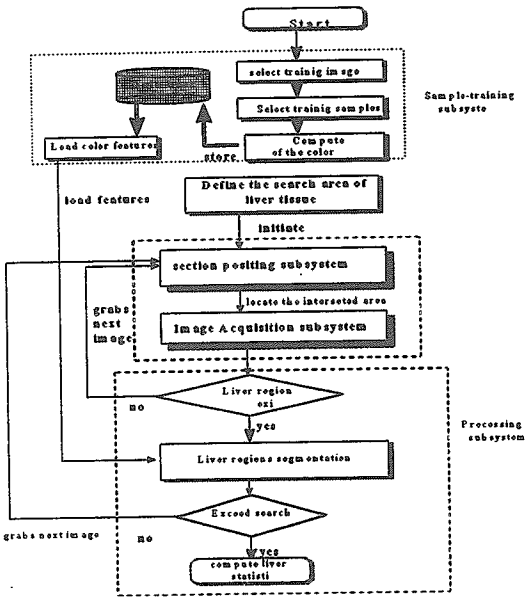


Fig 2.Processing flowchart of the ACM system.

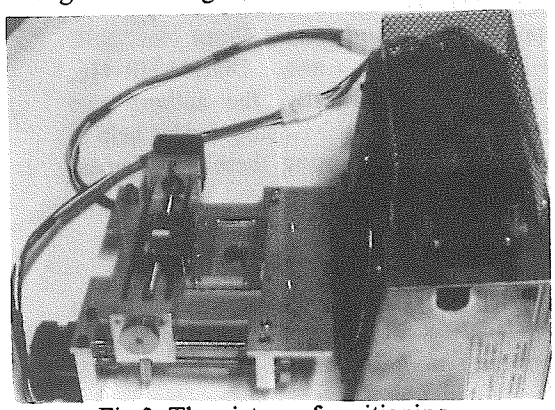

Fig 3. The picture of positioning subsystem.

Fig 4. An instance of liver tissue images

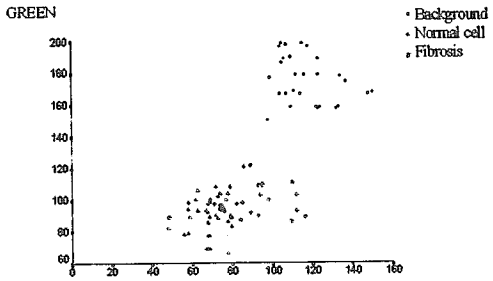


Fig 5(a)

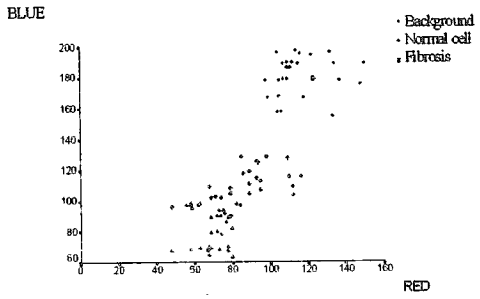


Fig 5(b)

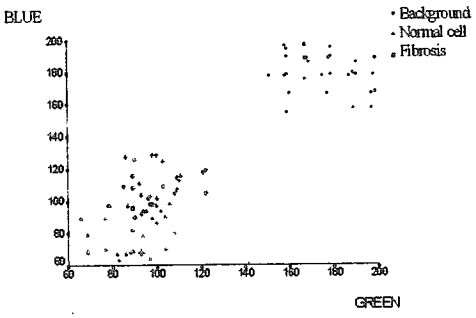


Fig 5(c)

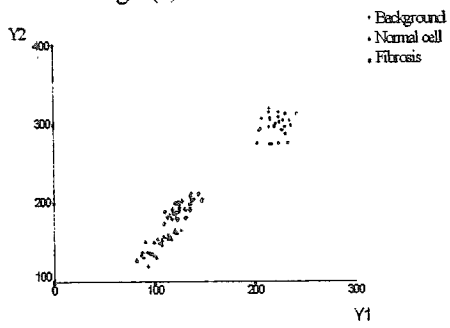


Fig 5(d).

Fig. 5(a). The red-green scatter plot of the samples of Fig. 4. Fig 5(b). The red-blue scatter plot of the samples of Fig. 4. Fig 5(c). The green-blue scatter plot of the samples of Fig. 4. Fig. 5(d). The $Y_1 - Y_2$ scatter plot of the samples of Fig. 4.

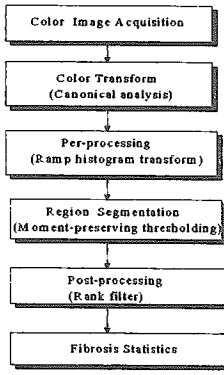


Fig 6. Flow chart of image processing.

| | | | | |
|---|---|---|---|---|
| 0 | 1 | 1 | 1 | 0 |
| 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | X | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 |
| 0 | 1 | 1 | 1 | 0 |

X: center of mask
1: masked region

Fig 7. The mask B used for rank filter.

Fig 8a

Fig 8(b)

Fig 8(c)

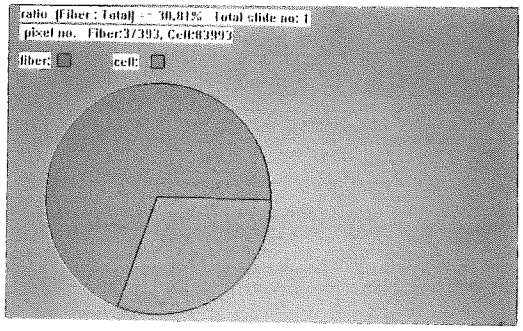


Fig 8(d).

Fig 8. Segmentation examples: (a) original image (b) segmentation results (c)edge maps (d)statistical result.

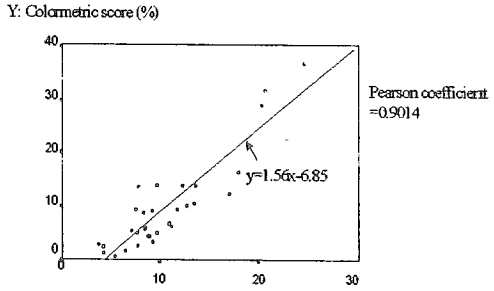


Fig 9. The relationsfip between colorimetric score and computer morphometry score.