Segmentation and Grading of Hepatocellular Carcinoma in Biopsy Images

Yee-Jee Jan[a], Shou-Kuo Dai[b], Yan-Hou Lai[c], Hsi-Chi Chen[b], Shu-Jiuan Lin[a]
[a]Department of Pathology, Taichung Veterans General Hospital, VACRS
[b]Department of Information management, Chaoyang University of Technology
[c]Department of Computer Science, Chung-Hsing University

摘要

正確的肝癌病理分級診斷對於之後的預後和治療是很重要的資訊，然而常會因為經驗的不足和人類的主觀認定造成分級結果有著很大的不確定性。我們在這篇論文中將提出一個客觀的量化自動分級系統，我們首先利用一些特殊的影像處理技術將細胞切割出來，再根據醫師所認定的肝癌分級標準抽取適當的特徵，最後再以支持向量機將之分成正常和 4 個分級，我們自患者的肝臟病理切片中取出的 296 張影像作為實驗對象，實驗結果顯示我們的方法可以達到 95%的正確性。

關鍵詞：肝癌，病理切片影像，重建，支持向量機

Abstract

Accurate grading of Hepatocellular carcinoma in biopsy images greatly improves prognosis. However, human visual grading is very subjective due to interobserver and intra-observer variations. In this paper, we propose a repeatable and quantitative method for this task. We first segment the cells from the biopsy images with some special techniques. Then extract the features according to the criteria of HCC tumor cell. Support vector machine classify HCC into four grades according to there extracted features. We analyze 296 images captured from biopsy images. Experimental results show our method achieve 95% accuracy.

Keywords: Hepatocellular Carcinoma; Biopsy image; Reconstruction; Support Vector Machine

1. Introduction

Liver cancer is a major health problem worldwide. Especially in Taiwan, according to the report of the Department of Health, liver Cancer is still the most deathful disease in Taiwan. Hepatocellular Carcinoma (or HCC) is the most common histological type of primary liver cancer. The prognoses and treatments can be variance for different grading of HCC. Therefore, the pathological discrimination between grades of HCC is very important. However, pathological analysis of tissue is time consuming and requires the visual interpretation of microscopic images. This is usually based on subjective assessment and can lead to significant inter-observer variation in grading. Computer-based analysis of microscopic images can provide quantitative results and improve prognosis.

Considerable analysis methods of pathological images have been proposed during the last few years. Michael et al. [1] proposed an analysis of microscopic cell images by texture feature. A method for automatic recognition of cancerous tissues based on shape and the size analysis of the observed cells is presented in [2]. Frank et al. [3] provided a Biopsy Analysis Support System (BASS) which can detect the breast cancer nuclei in a section which is stained with monclonal antibodies. Abdeltahim et al. [4] developed an automated algorithm for the categorization of normal and cancerous colon mucosa with some of texture features. Barbara et al. [5] provide a method based syntactic structure analysis in automated classification. In [6], method employed the multiwavlets feature to analyze entire image not individual cells. However, degree of cell differentiation is still important for cancerous grading. In this work, we present an automatic

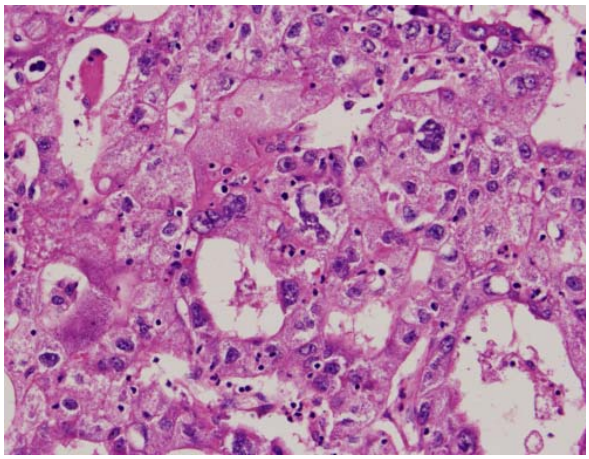

Fig.1 A HCC biopsy image.

method to analyze and grading the HCC biopsy images.

The methods of segmentation of HCC and the feature extraction are described in section two. Section three is the results of the experiments and discussion. Summary are in the session four.

2. Materials and Methods

2.1 Acquisition of the Images

In this study, every image is obtained with the same tissue processing and acquisition method, briefly described below. Each tissue is embedded in paraffin cubes after chemical processing, and cut into very thin sections. These sections are placed on glass slides and stained with colored dyes, Hematoxylin and Eosin. The images are acquired a set of technical equipment which includes a high quality optical microscope, a high resolution CCD camera and an acquisition system. Each image was acquired with a magnifying factor of 400 on the microscope. There are 296 images captured from biopsy images. These images are viewed by pathologist and classified into normal and grade one to grade four by pathologist in advance. Then we randomly select 20 of each class as training set, others are testing set.

2.2 Nuclei Segmentation

HCC biopsy images contain many undesirable elements such as cytoplasm. Since the nucleus of cell contains the major features for grading. It is essential to segment nuclei from the rest of the image also it is not an easy task. As the HCC biopsy image shown in Fig.1, cells can be located

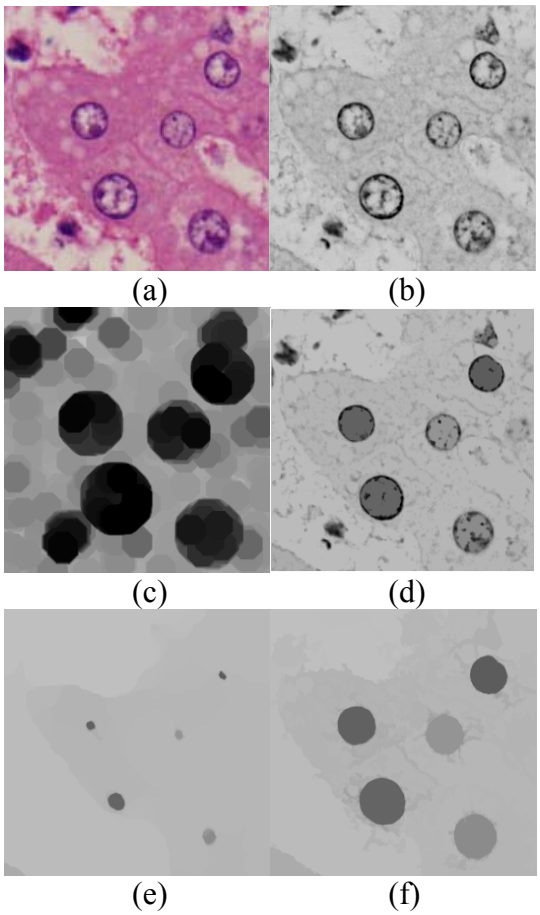

Fig 2 Example of Reconstruction algorithm.

in upper or lower position of tissue that the focus of microscope cannot fit for every cell. If a threshold method is applied, blurred nuclei could generate an unstable edge and confuse with other undesirable elements. For this reason, we develop a method combined the morphological reconstruction algorithm and Snake model to solve this problem.

Morphological reconstruction algorithm [7, 8, 9] based on geodesic dilation operation with two same size input images which are marker and mask image denoted by I and M respectively. M is act as a limit for dilation operation and I≤M. The geodesic dilation of I is given by

$$\delta_I(I) = \min((I \oplus B), M), \quad (1)$$

Where B is a structure element , the symbol $\oplus$ is dilation operation with limit by M and min is the pixel-wise minimum operation. Then we apply the geodesic dilation recursively until stability is reach. Fig. 2 shows an example to illustrate how we use reconstruction to segment the HCC nuclei.

Fig.2 (a) is the original image with RGB color model. Since our target area is nuclei in blue

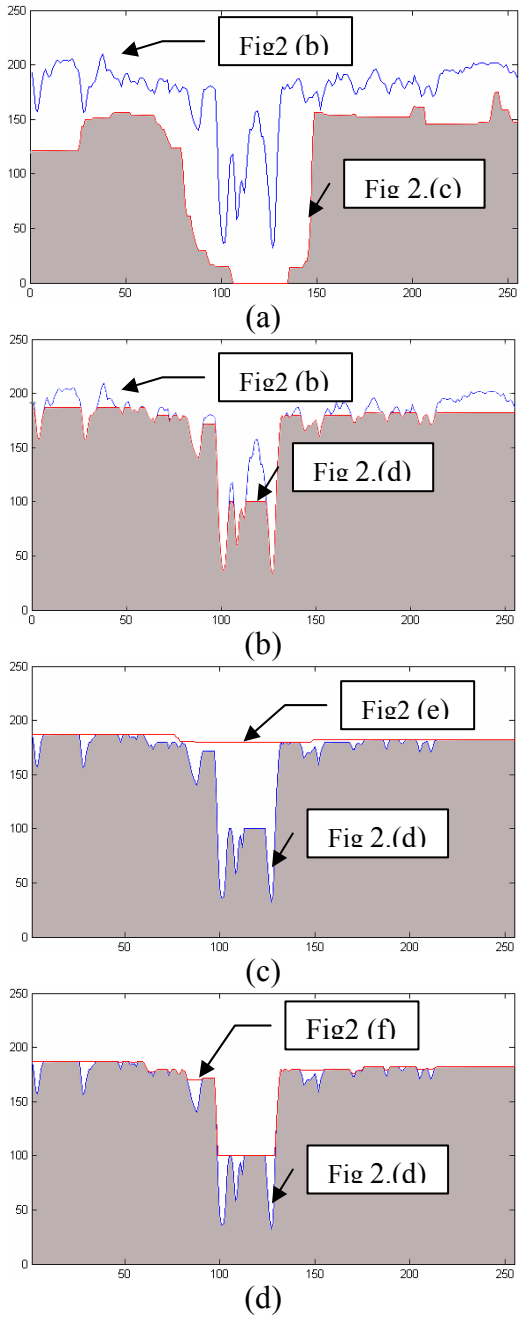

Fig 3.The gray level of row 178 of Fig 2(b)-(f).

color, while others are red and white. Therefore we extract red plane Fig.2(b) from RGB color image Fig.2(a). Fig2(c) is given by erode Fig.2(b) with a disk shape structure element. We adopt Fig2(c) and Fig.2(b) as maker and mask image respectively. By using reconstruction algorithm, Fig.2(d) is generated. Fig.2(e) is the result of dilation of Fig.2(d). The maker image Fig.2(d) and the mask image Fig.2(e) are taken as input for the reconstruction algorithm. The image Fig.2(f) is final result. It is correspondence between the nuclei in Fig.2(a) and Fig.2(f). Fig. 3 is the gray level extracted from row 178 of Fig 2(b)-(f). From Fig. 3(d), the nucleus and background can be discriminated easily. A thresholding operation can segment the nuclei from image. Where the shape of nucleus is an important information for grading. In the next step, a snake algorithm is applied to refine the contour of the nucleus.

A Snake is a curves $f(s) = [x(s), y(s)]$, where s is a normalized arc length. The actions of snake are controlled by minimizing energy function [10, 11, 12].

$$E = \int_0^1 E_{in}(f(s)) + E_{ext}(f(s))ds \qquad (2)$$

$$E_{in}(f(s)) = \alpha\left|f'(s)\right|^2 + \beta\left|f''(s)\right|^2 \qquad (3)$$

$$E_{ext}(f(s)) = -\gamma\left|\nabla[G_\sigma * I(f(s))]\right|^2 \qquad (4)$$

where E_{in} and E_{ext} are internal and external energy term respectively, while $f'(s)$ and $f''(s)$ denote the first and second derivatives of $f(s)$ with respect to s. G_σ is a Gaussian function with standard deviation σ and ∇ is the gradient operator. $\nabla[G_\sigma * I(f(s))]$ is the gradient of Gaussian filtered image *I*. α and β control the snake's tension and rigidity. α, β, γ together control the weights of internal and external energy. A snake that minimizes E must

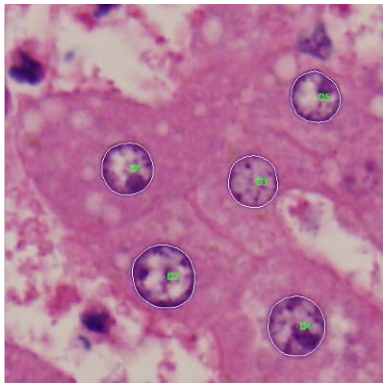

Fig.4 Result of Segmentation

satisfy the Euler equation

$$\alpha f''(s) - \beta f''''(s) - \gamma\nabla E_{ext}(f(s)) = 0 \qquad (5)$$

This can be viewed as a force balance equation

$$F_{in} + F_{ext} = 0, \qquad (6)$$

where

$F_{in} = \alpha f''(s) - \beta f''''(s)$ and $F_{ext} = -\gamma\nabla E_{ext}(f(s))$.

The internal force F_{in} discourages stretching and

bending while the external force F_{ext} pulls the snake toward the desired image edges.

To find a solution to (5), snake $f(s)$ is being a function of time t and s. Its partial derivative with respect to t is then set equal to the left hand side of (5) as follows:

$$f_t(s,t) = \alpha f''(s,t) - \beta f''''(s,t) - \gamma \nabla E_{ext}(f(s,t)) \quad (7)$$

When $f(s,t)$ stabilizes, a solution of (5) is achieved. Fig.4 is the result of segmentation.

2.3 Feature Extraction

HCC grading is based on several criteria, including nucleocytoplasmic ratio, anisonucleosis, hyperchromatism, and nuclear deformity [13]. All images are classified into normal and grade one to grade four by pathologist in advance. Then we observe these graded images and selected the corresponding features for classification. Following are the description of selected features.

- **Circularity** [14]: Because of poorly differentiate, the shape of nuclei can be deformity and no longer keep the round shape. So we obtain a numerical value by following formula as one of criteria.

$$\text{Circularity} = 4 \times PI \times \frac{Area}{Circumference^2}. \quad (8)$$

- **Area**: Higher grade implies higher probability of larger nucleus.
- **Nuclear density**: Create a minimum spanning tree to connect every cell and search the shortest path.
- **Variance**: The variance of gray level of nuclear area.
- **B/D spot ratio**: Calculating the ratio of bright and dark spots in nucleus, which are derived from top-hat and bottom-hat algorithm [15, 16].
- **B/D area ratio**: Calculating the ratio of bright and dark area in nucleus.

Following features are obtained from the co-occurrence Matrix (GLAM) p [18, 19, 20, 21].

- **Contrast**= $\sum_{i,j} |i-j|^2 p(i,j)$. (9)
- **Energy**= $\sum_{i,j} p(i,j)^2$. (10)
- **Homogeneity**= $\sum_{i,j} \frac{p(i,j)}{1+|i,j|}$. (11)

2.4 Classification

For the features described above, we must classify the analyzed image, decide biopsy image is normal or grade 1~4. In this study, we select SVM [22] as the classifier. The software is downloaded from this website [23].

The goal of SVM is to produce a model which predicts class label of instances in the testing data set. The set of labeled training patterns (x_i, y_i), i=1,..,l where $x_i \in R^n$ and $y \in \{1,-1\}^l$, the solution of the following optimization problem:

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \quad (11)$$

$$\text{subject to } y_i\left(w^T \phi(x_i) + b\right) \geq 1 - \xi_i,$$

$$\xi_i \geq 0.$$

Function ϕ maps x_i into a higher dimensional space, C is the penalty parameter of the error.

Table 1. The Distribution of Images

	number of images
Normal	78
Grade 1	46
Grade 2	53
Grade3	67
Grade 4	52

Table 2. Feature Set Description.

	Features
FS.1	Area
FS.2	Circularity
FS.3	Contrast, Energy, Homogeneity
FS.4	B/D spot ratio and B/D area ratio
FS.5	Nuclear density, Variance
FS.6	FS.1~FS4
FS.7	All the selected features

We select radial basis function (RBF) as the kernel function,

$$K(x_i, x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right), \gamma > 0, \quad (12)$$

where γ, r, and d are kernel parameters.

The penalty parameter C will affect the accuracy of prediction. In order to get a higher accuracy, we will use cross validation process to decide penalty parameter C.

3. Experimental Results

Since SVM required a supervised training technique, the images were divided into 5 classes which are normal and grade one to grade four. Table 1 shows the distribution of these images. In this study we randomly select 20 images from each class as training set for SVM training process, while testing set contains 196 images. To evaluate the performance of selected features, we divide experiment into seven parts, each with different feature set (or FS). Table 2 is the description of these feature sets. Results of the Training model accuracy are in the Table 3, where we test the penalty parameter C with 1, 10, 100 and 1000. Table 4 gives the accuracy of the HCC grading of testing set. From these results, we can conclude that parameter C larger then 10 is a better choice. FS.3 and FS.5 are more discriminative than others. The best accuracy we achieved is about 95%.

Table 3. Accuracy of training model.

	C=1	C=10	C=100	C=1000
FS.1	56.504	63.415	64.634	65.447
FS.2	71.545	75.203	76.422	76.016
FS.3	58.537	70.325	68.699	69.919
FS.4	51.22	58.567	65.04	64.228
FS.5	70.732	76.423	76.829	78.455
FS.6	83.333	94.715	95.122	95.935
FS.7	91.057	96.748	97.561	97.561

Table 4
Accuracy of HCC Grading of testing set.

	C=1	C=10	C=100	C=1000
FS.1	61	63	63	59.6
FS.2	65.1	77.4	75.3	74.7
FS.3	50.7	60.3	70.5	72.6
FS.4	58.9	60.3	58.9	63
FS.5	62.3	75.3	78.1	77.4
FS.6	82.9	92.5	93.8	91.8
FS.7	85.6	95.8	95.2	95.2

4. Summary

This paper proposes a method for segmentation of HCC nuclei, which based on reconstruction algorithm and snake model. Then we extract several image features according to the criteria of HCC tumor cell. Support Vector Machine is employed for HCC grading. We analyze 296 images captured from HCC biopsy images. Experimental results demonstrate that can result in highly significant classification and discrimination of HCC grading.

Acknowledgement

This research work was supported by National Science Council of ROC under contract NO. NSC -95-2221-E- 324-041.

References

[1] Michael Beil, Theano Irinopoulou, Jany Vassy, Gunter Wolf, "A Dual Approach to Structural Texture Analysis in Microscopic Cell Images," ***Computer Methods and Programs in Biomedicine***, Vol.48, pp. 211-219, 1995.

[2] Jean-Philippe Thiran, Benoit Macq, "Morpho- logical Feature Extraction for the Classifi- cation of Digital Images of Cancerous Tissues," ***IEEE Transaction on Biomedical Engineering***, Vol. 43, NO. 10, pp. 1011-1020, 1996.

[3] Frank et al., "Computer-Aided Detection of Breast Cancer Nuclei," ***IEEE Transaction on Information Technology in Biomedical***, Vol. 1, NO. 2, pp. 128-140, 1997.

[4] Abdeltahim Nasser Esgiar, Raouf N. G. Naguib, "Microscopic Image Analysis for Quantitative Measurement and Feature Identification of Normal and Cancerous Colonic Mucosa," ***IEEE Transaction on Information Technology in Biomedical***, Vol. 2, NO. 3, pp. 197-203, 1998.

[5] Barbara et al., "Computer-Assisted Differem-tial Diagnosis of Maligant Mesothelioma Based on Syntactic Structure Analusis," ***Cytometry***, Vol. 35, pp. 23-29, 1999.

[6] Kourosh Jafari-Khouzani, Hamid Soltanian-Zadeh, "Multiwavelet Grading of Patho-logical Images of Prostate," ***IEEE Transaction on Biomedical Engineering***, Vol. 50, NO. 6, pp. 697-704, 2003.

[7] R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image analysis using mathematical morphology," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, no. 4, pp. 532-549, July 1987.

[8] **S.** Sternberg, "Gray scale morphology," ***Computer Vision, Graphics, Image Processing,*** vol. 35, pp. 333-355, 1986.

[9] R. M. Haralick and L. G. Shapiro, ***Computer***

and Robot Vision, vol. 1. Reading, MA: Addison Wesley, 1992

[10] Kass M, Witkin A, Terzoulos D., "Snake: Actour contour models." ***International journal of Computer Vision***, Vol. 1, pp. 321–331, 1987.

[11] Xu C, Prince JL., "Snakes, shapes, and gradient vector flow." ***IEEE Transaction on Image Processing***, Vol.7, pp. 359–369, 1998.

[12] Chen et al., "Cell-Based Dual Snake Model: A New Approach to Extracting Highly Winding Boundaries in the Ultrasound Images," ***Ultrasound in Med. & Biol.***, Vol. 28, No. 8, pp. 1061–1073, 2002.

[13] Nzeako et al., "Tumor Pathology and Survival in HCC," ***Cancer***, Vol. 74, No. 4, pp. 579-588, 1995.

[14] Bottema, M.J., "Circularity of Objects in Images**," ICASSP '00. Proceedings. 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2000.**

[15] J. Serra, *Image Analysis and Mathematical Morphology*. London: Academic, 1982.

[16] A. Moragas, M. Garcia-Bonafe, I. de Torres, and M. Sana, "Textural analysis of lymphois cells in serous effusios, a mathematical morphologic approach," ***Anal. Quantum Cytol.***, Vol. 15, pp. 1656-1 170, 1993.

[17] Q. A. Holmes, D. R. Nuesch, and R. A. Shuchman, "Texture analysis and real-time classification of sea-ice types using digital SAR data," ***IEEE Trans. Geosci. Remote Sensing***, Vol. GE-22, pp. 113–120, 1984.

[18] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," ***IEEE Trans. Syst., Man, Cybern.***, Vol. SMC-3, pp. 610–621, 1973.

[19] R. M. Haralick, "Statistical and structural approaches to texture," ***Proc. IEEE***, Vol. 67, pp. 786–803, 1979.

[20] R. W. Conners and C. A. Harlow, "A theoretical comparison of texture algorithms," ***IEEE Trans. Pattern Anal.Machine Intell.***, Vol. PAMI-2, pp. 204–222, 1980.

[21] P. P. Ohanian and R. C. Dubes, "Performance evaluation for four classes of textural features," ***Pattern Recognition.***, Vol. 25, pp. 819–833, 1992.

[22] Cortes, C. and V. Vapnik Support-vector network. ***Machine Learning*** 20, pp. 273–297, 1995.

[23] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm