

A Novel Method of HCC Pathological Images Recognition based on Morphological Distribution of Liver Cell Cords

Ao-shuang DONG^{1,*}, Su JIANG², Hui-yan JIANG³ and Ke-ming MAO⁴

^{1, 2, 3, 4} Software College, Northeastern University, Shenyang 110819, China

*Correspondence should be addressed to Ao-shuang DONG;
dongaoshuang@mail.neu.edu.cn

Keywords: Liver Cell Cords, HCC, Pattern Recognition, Pathological Image.

Abstract. The morbidity of hepatocellular carcinoma (HCC), which has become a significant killer to human life, is getting higher over the years. Not only the shape and structure of liver cells are abnormal in pathological images, but also liver cell cords show some unusual phenomena, such as incassation and disorder. So our paper proposes a novel method of HCC pathological images recognition based on morphological distribution of liver cell cords, to assist pathologist to diagnose hepatic pathology. In the method, we firstly apply binarization and morphological operation to hepatic pathological images, and screen connected regions with their area and circularity. Adopted regions are considered as candidate centers of liver cell cords (vein center). Secondly we define different features (irregularity feature), according to the number of candidate centers in a pathological image, to recognize how structured liver cell cords are. Finally, we utilize the feature and classifier to identify HCC pathological images. The experiments indicate that the proposed method shows high accuracy of all HCC pathological images in well, moderately, poorly differentiated and normal forms.

Introduction

With the deterioration of the environmental pollution and variation of human's life style, the morbidity of hepatocellular carcinoma (HCC) is getting higher over the years[1-3]. Now liver cancer has gradually become a major disease endangering human health[4-5]. Therefore, the sooner the disease is diagnosed, the better effect the treatment will get.

There're not so many researches on computer-aided diagnosis for liver disease[6-7]. Kadah[8] utilized texture measure based on the first order and second order gray-level statistics and statistical neural classification algorithm to divide liver ultrasound diagnostic data into 3 categories, which is normal tissue, fatty liver and liver cirrhosis. Asvestas[9] utilized the fractal descriptor of power difference method and a fuzzy C-means cluster to divide B ultrasound liver tissue into 3 categories, which were normal tissue, canceration and hemangioma. Lee[10] made use of texture features generated from Gabor wavelet and the support vector machine (SVM) in CT images to recognize normal tissue, cyst, hemangioma and canceration. Liu Wei[11] divided CT images into 3 categories, which were normal tissue, HCC and hemangioma, with co-occurrence matrix texture analysis and probabilistic neural network classification in the research on texture analysis and artificial neural network. Ye Jun[12] extracted texture features, and then designed a new classifier with neural network, k-NN and SVM. Po-Whei Huang[13] extracted fourteen features based on six types of characteristics for HCC classification after segmentation with watershed algorithm.

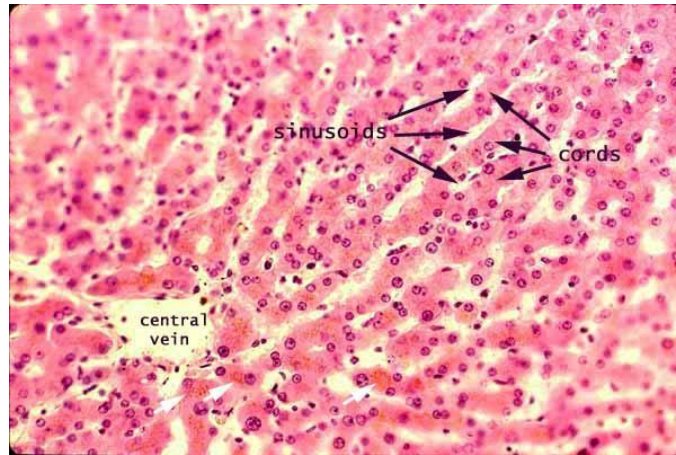


Fig.1 Schematic of liver cell cords.

Most of recognition algorithms utilize texture features [14-17]. While in this paper, we identify HCC pathological images by liver cell cords and irregularity of cell distribution. Liver cells are radially distributed and the central vein is the axis of them. The structure of cells on the slices is cord-like, which is the reason that they are called hepatic cell cords, or liver cell cords [18], as shown in Fig.1. If liver lesion exists, liver cell cords will break up[19].

Hence, in this paper, we adopt the degree of irregularity of cell distribution to recognize HCC. The samples in use include images in well, moderately, poorly differentiated and normal forms. The data are acquired from a large hospital in Shenyang, which are real and solid as well as representative. In the method, we firstly apply binarization and morphological operation to hepatic pathological images, and screen connected regions with their area and circularity. Adopted regions are considered as candidate centers of liver cell cords (vein center). Secondly we define different features (irregularity feature) according to the number of candidate centers in a pathological image, to recognize whether liver cell cords exist and, if they exist, what the structure is like. Finally, we utilize the feature and classifier to identify HCC pathological images.

Organization of the Text

The rest of the paper is organized as follows. We introduce the proposed method in Section 2, including the process of the method, feature extraction and classifier construction. Then we estimate the method with experiments in Section 3 and conclude the paper in Section 4 in the end.

A Classification Method Based on Morphological Distribution of Liver Cell Cords

Framework

The section headings are in boldface capital and lowercase letters. Second level headings are typed as part of the succeeding paragraph (like the subsection heading of this paragraph).

Normal hepatocytes are distributed in neat cord-like morphological structure, and liver cell cords always have a vein center. So we start from recognizing whether the center of liver cell cords exists and how structured the distribution of hepatocyte is, to identify HCC. Fig.2 shows the process of our method.

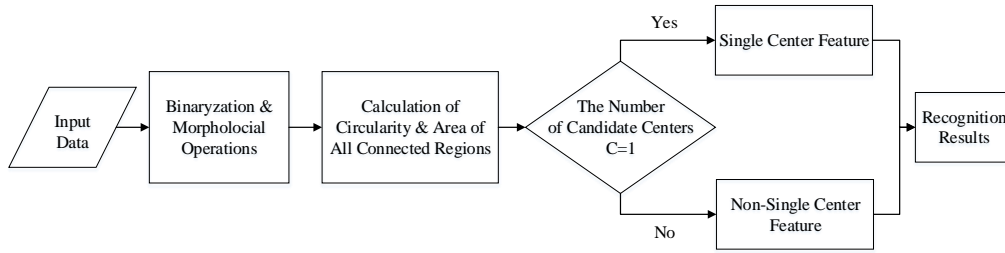


Fig.2 Flowchart of the proposed method.

As shown in Fig.2, after inputting a pathological image and thresholding the images, appropriate morphological operations are carried out. Then we calculate the area and circularity of all of the connected regions in the picture. If both of circularity and area of a region are above a threshold, the region will be considered as a candidate cord center. Vein center is the center of liver cell cords, and cords are radically arranged. However, sinusoid in hepatocyte pathology images may present similar shape after binarization and morphological operations. Therefore, sinusoid possibly satisfies the threshold and forms a region easily taken as a vein center. So a region satisfy the conditions may not be a real vein center. That's why we name those regions as candidate centers. After that, we extract different irregularity features based on whether the number of candidate centers in an image equals one. With single center features or non-single center features, classifier divides hepatocyte pathology images into two classes, normal liver and canceration.

Feature Extraction and Representation

Definition of Studied Variables. In the method, after screening connected regions through circularity and area, we obtain candidate liver cell cords centers. According to the differences between normal and HCC pathological images, various operations are adopted based on the number of candidate centers in a picture, aiming to recognize HCC by the morphological structure of hepatic cell cords. When acquiring candidate liver cord centers, calculating the circularity of connected regions after binarization and morphological operations is necessary.

What's more, the extraction of features is mainly based on the calculation of main direction of each block, which present the diversity of irregularity degree of hepatocyte arranged in the pictures. As a consequence, the operation of calculating the main direction of a block is considerable, which has huge influence on the recognition results.

Here are the definitions of circularity and main direction.

Def.1 Circularity

The method firstly calculates the number of candidate centers in a picture. In order to ascertain the number, circularity of each connected regions obtained after binarization is requisite.

The formula used here for circularity is defined as

$$C = 4\pi A / (P^2). \quad (1)$$

In this formula, C presents the circularity of regions, A presents the area and P presents the perimeter.

Hepatic cell cords are syllabically and radically arranged in normal hepatocyte pathological images, with vein center as its center. Most of the vein centers show high circularity after binarization. While there're many quasi-circle regions in HCC

pathological images as well. This is because there might be fat droplets or sinusoid that happens to have the shape of high circularity. For this reason, the regions that satisfy the threshold are named as candidate centers.

Def.2 Main direction

The direction of texture is a basic feature of the texture of images, and its description has important practical significance [20, 21]. In this paper, the computed main direction adopts the approach in literature [22]. The algorithm utilizes edge information of images to compute the direction of texture, which is consistent with the fact that the trend of edge image is accordant with that of the original image.

For the sake of improving efficiency in the implementation, angles are divided into 10 groups, i.e. $0^{\circ} \sim 18^{\circ}$ as group 1, $19^{\circ} \sim 36^{\circ}$ as group 2, $163^{\circ} \sim 180^{\circ}$ as group 10 and so on. Therefore, the result of calculation of main direction ranges from 1 to 10. In later calculation, 1 and 10 present horizontal direction, while 5 and 6 present vertical direction

The algorithm has two advantages. Firstly, by edge information extracted from edge detection operator, the trend of texture will be estimated. Then the influence of noise gets restrained, and helps to improve estimation accuracy. Further, estimating the direction of texture by edge information excludes the calculation of the interference factor, which greatly reduces the computational complexity, and increases speed. A large number of simulation experiments show the effectiveness of the algorithm.

This article reflects the arrangement of the cells indirectly by the direction of the gap between cells. When liver cells are neatly arranged in cords, the gaps between these liver cells are also distributed in cords. The gaps are long and slender with similar direction of cords. But once hepatic cords are destroyed or get incassation, the direction of gaps becomes disorder. If hepatocyte arrange in cords, the main directions of neighborhood will be close. While if they don't, then the main directions of neighborhood will be widely divergent.

Single Candidate Center. Fig.3 (a) shows the operation flowchart when there's only one candidate center. There are N hepatic pathological images, and C_i presents the number of candidate centers in the i th image ($i=1, 2, 3, \dots, N$). When there is only one candidate center, then segment of up, down, left and right blocks around the candidate center will be put into practice. The division is shown in Fig.4 (d). The size of four blocks are different in different pictures. After the candidate center is determined, a bounding rectangle is obtained. The up and down blocks are rectangles above and under the center respectively, both with triple width of the width of the bounding rectangle. While the left and right blocks are rectangles on the left and right of the center respectively, both with triple height of the height of the bounding rectangle. After all the four blocks are determined, the next step is to calculate main directions of all four blocks. If the candidate center is really the center of cords, i.e. the vein center, then the main directions of up and down blocks are close to vertical direction, and those of left and right blocks are close to horizontal directions. Since the results of main directions range from 1 to 10, where 5 and 6 both present the vertical direction, and 1 and 10 both present the horizontal direction. Then we get an 8-dimensional vector, named single center feature.

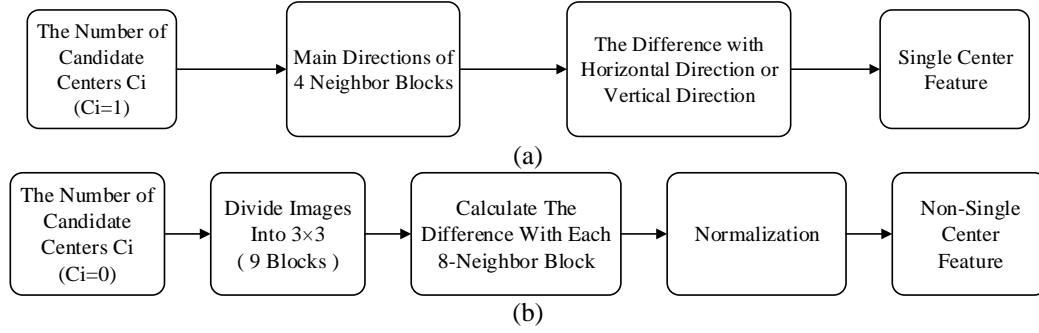
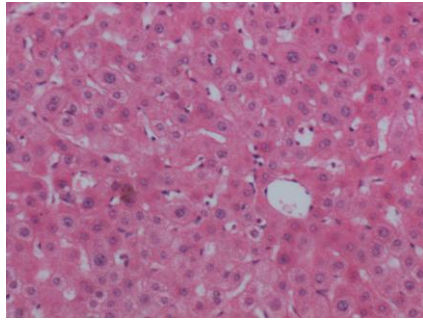
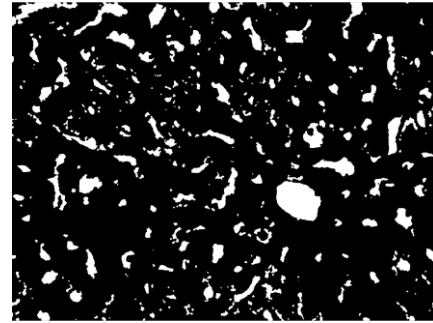


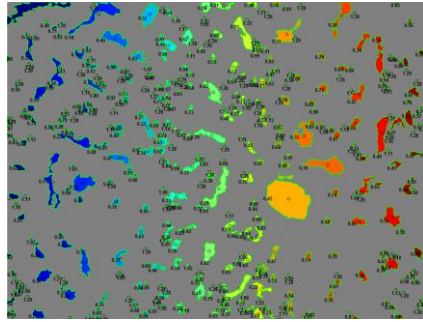
Fig.3 (a) shows the flowchart of feature extraction when there's only one candidate center. Fig.3 (b) shows the flowchart of feature extraction when there's not only one candidate center.



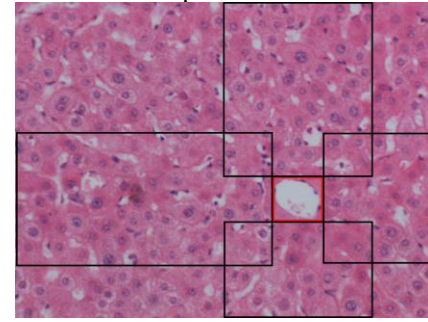
(a) The original image



(b) The result of binarization, morphological operations



(c) The circularity of every connected region



(d) The division based on the candidate center

Fig.4 (a) is an original image. Fig.4 (b) shows the result of binarization, morphological operations. And Fig.4(c) shows the circularity of every connected region. After above process, the region rounded by red line in (d) is considered as a candidate center.

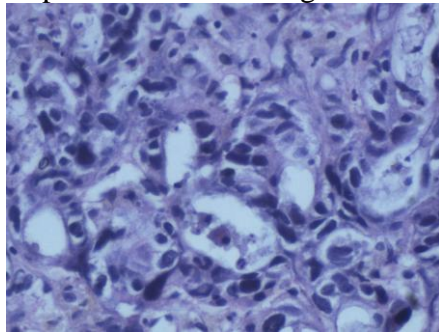
As shown in Fig.4, after binarization and appropriate morphological operations, Fig.4 (a) has been identified as one with single candidate center. Then we can obtain a single center feature presented by

$$D = \{ D_i \mid i = 1, 2, 3, \dots, 8 \}. \quad (2)$$

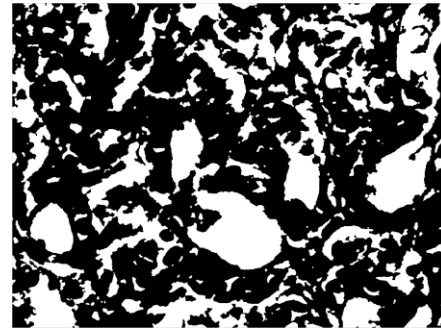
where $D_1 \sim D_2$ present the differences between the main direction of up block and vertical direction, $D_3 \sim D_4$ present the differences between the main direction of down block and vertical direction, $D_5 \sim D_6$ present the differences between the main direction of left block and horizontal direction and $D_7 \sim D_8$ present the differences between the main direction of right block and horizontal direction. Since vertical direction happens to be the adjacent boundary between 5 and 6 and horizontal direction happens to be the adjacent boundary between 1 and 10 when directions are presented by the main direction, two values are obtained when every block is compared with vertical direction or horizontal direction. Afterwards, the 8-dimensional vector will be used as

feature with SVM to recognize whether its original image belongs to HCC pathology images.

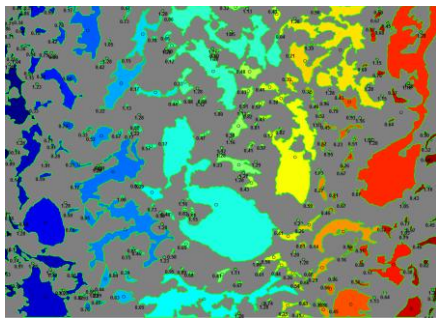
Non-Single Candidate Center. Fig.3 (b) shows an operation flowchart when there's not only one candidate center. When the number of candidate centers doesn't equal one, the images are divided into 3×3 parts. Due to the need of reserving the contact between directions of the blocks and trying not to lose direction information of each part of the image, we choose to divide the original images into 3×3 parts based on the resolution of images, i.e., a total of nine pieces. After that, main direction of each block is obtained to compare with that of neighbor blocks.



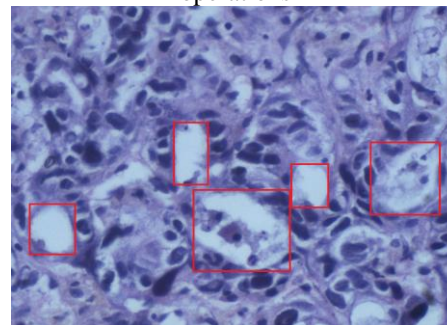
(a) The original image



(b) The result of binarization, morphological operations



(c) The circularity of every connected region

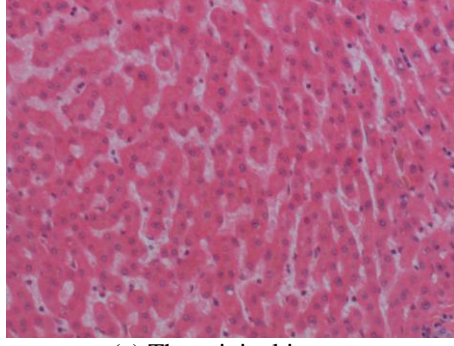


(d) The candidate centers in red rectangles

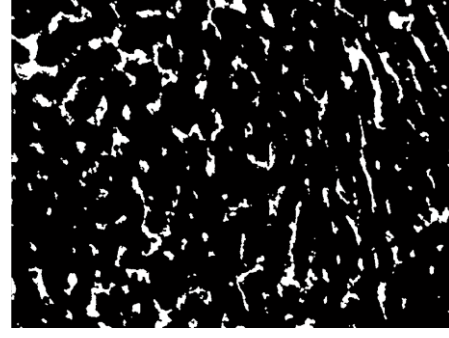
Fig.5 (a) is an original image. Fig.5 (b) shows the result of binaryzation, morphological operations. And Fig.5 (c) shows the circularity of every connected region. After above process, the region rounded by red line in (d) is considered as candidate centers.

As shown in Fig.5, after binarization and appropriate morphological operations, Fig.5 (a) has been identified as with more than one candidate center. Similarly, after almost the same operations, Fig.5 (e) has been identified as with no candidate center. Both of them are taken as images with non-single candidate center. There are N hepatic pathological images, and C_i presents the number of candidate centers in the i th image ($i = 1, 2, 3, \dots, N$). When there are more than one candidate centers, we can summarize that the distribution of hepatocyte is out of cords, or the cords do exist but they are disorder. When there is no candidate center, whether there exist liver cell cords is also uncertain. Therefore, the pictures are likely to be HCC pathological images. But the final recognition results need help from SVM.

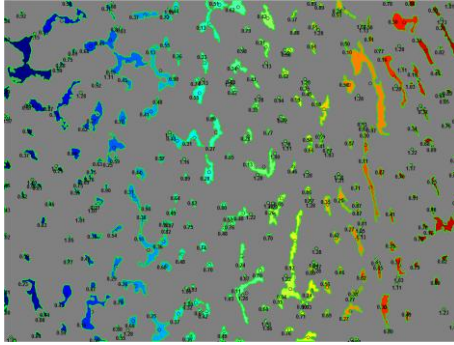
Since normal liver cells obviously have cord-like structure, so they are arranged in order. There's a similar case that the gaps between hepatocyte also have a certain regularity. These gaps are mostly slender, and also have a certain direction, consistent with the direction of the cell typically arranged. Therefore, calculation of the main directions in this article takes advantages of this fact.



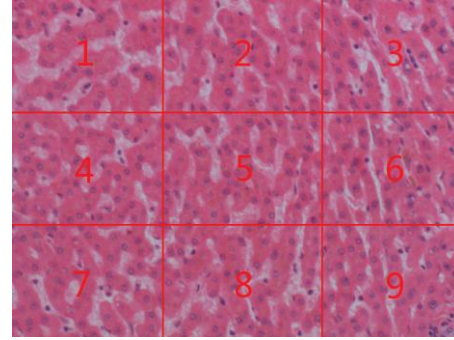
(a) The original image



(b) The result of binarization, morphological operations



(c) The circularity of every connected region



(d) The division when there's no candidate centers

Fig.6 (a) is another original image. Fig.6 (b) shows the result of binarization, morphological operations. And Fig.6 (c) shows the circularity of every connected region. Then it is identified as with non-single center. So the image is divided into 9 blocks as shown in (d) to identify whether the distribution of hepatocyte is in order.

As shown in Fig.6 (d), what we need to do is to calculate absolute values of the difference between each block and its 8-neighbor blocks. For example, block 2, 4 and 5 are neighbor blocks of block 1. Once compared with main direction of neighbor blocks, whether the distribution of hepatocyte in a block is in order is certain. D_{ij} presents the absolute value of the difference between block i and block j , i.e. the absolute value of the difference between block 1 and block 2 is denoted by D_{12} . After that, a 20-dimensional vector is obtained, which is presented by

$$A = \{ D_{ij} \mid i = 1, 2, \dots, n, j \in N8(i), i < j \}. \quad (3)$$

In the formula, n presents the number of blocks, and $N8(i)$ presents a set of 8 neighbor blocks of i .

Normalization is necessary to eliminate the differences between the images. Afterwards, the vector will be used as feature with SVM to recognize whether its original image belongs to HCC pathology images.

Classifier Construction.

The classification technique used in this paper is the SVM method. In the SVM method, the sample space is mapped into a high-dimensional or even an infinite dimensional feature space (Hilbert Space) by a nonlinear mapping, to transform a nonlinear separable problem in the original sample space into a linear one in the feature space, i.e. rising dimension and linearization. Generally, rising dimension increases the computation complexity, while the SVM method solves the problem skillfully.

We choose variable kernel functions to generate corresponding SVM method. Commonly used kernel functions are linear kernel function, polynomial kernel function radial basis function (RBF) and layer neural network kernel function. In this paper, RBF is chosen, and SVM were constructed using LIBSVM [23].

Experiments

Data

In this paper, the proposed method is tested by experiments with high differentiation, moderately differentiation, low differentiation and normal hepar pathological images as data. The data are acquired from a large hospital in Shenyang in clinic, real and solid as well as representative.

There were 544 images with size of 320 by 240 pixels. According to the characteristics for identifying HCC, the images are divided into 4 groups by experienced pathologists, which are in well, moderately, poorly differentiated and normal forms.

Evaluation Criteria

In this paper, sensitivity rate and accuracy rate are considered as evaluation criteria. Sensitivity is presented by TPR (True Positive Rate), and accuracy rate is presented by Accy. Expression of the formula is presented as follows.

$$TPR = TP / (TP + FN). \quad (4)$$

$$Accy = (TP + TN) / (TP + TN + FP + FN). \quad (5)$$

In the formulas above, TP, TN, FP and FN present true positive, true negative, false positive and false negative respectively.

Comparison with Other Features

Most of HCC recognition methods utilize texture features. Under the circumstances, we choose LBP and Gray-level Co-occurrence Matrix to compare with irregularity feature. After extracting LBP, GLCM and irregularity features from all of the pictures, SVM is used to identify pathological images.

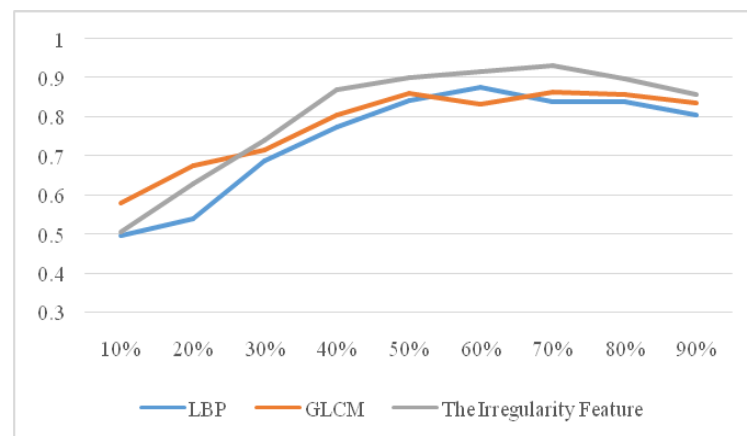


Fig.7 The recognition results of LBP, GLCM and the irregularity feature with different percentages of data as training dataset.

The proportion of training dataset may have influence on classification rate of the algorithm. In this subsection, training dataset is selected randomly with the proportion from 10% to 90%, and the remainder is used for testing. The testing is performed many times and the average classification rate is computed.

As shown in Fig.7, LBP shows the best performance when training dataset is selected with proportion of 60%. And GLCM gets the best result when training dataset accounts for 50%, and the performance of 70% is good as well. Irregularity features reach the highest accuracy when training dataset is selected with proportion of 70%. When the percentage gets higher, the accuracy doesn't become better as it supposed to be. On the contrary, the accuracy goes a little down because of overfitting.

Table 1. The results of recognition of various differentiated images

Categories	Sensitivity(Accuracy)
In well differentiated form	0.9574/-
In moderately differentiated form	0.9248/-
In poorly differentiated form	0.8973/-
Normal hepatocyte pathological images	-/0.9847

The results of experiments with the proposed method are shown in Table 1. The percentage of training dataset is 70%. It shows that the method has high recognition rates of all low, moderately, high differentiated HCC pathological images.

The structure of hepatic cords in normal hepatocyte pathological images is obvious. Incrassation of cords exists in high differentiation hepatocyte pathological images, and the structure of cords is getting vaguer as the level of differentiation of hepatocyte pathological images is getting lower. The proposed method can achieve high accuracy even for well differentiated hepatocyte pathological images, which are close to normal ones.

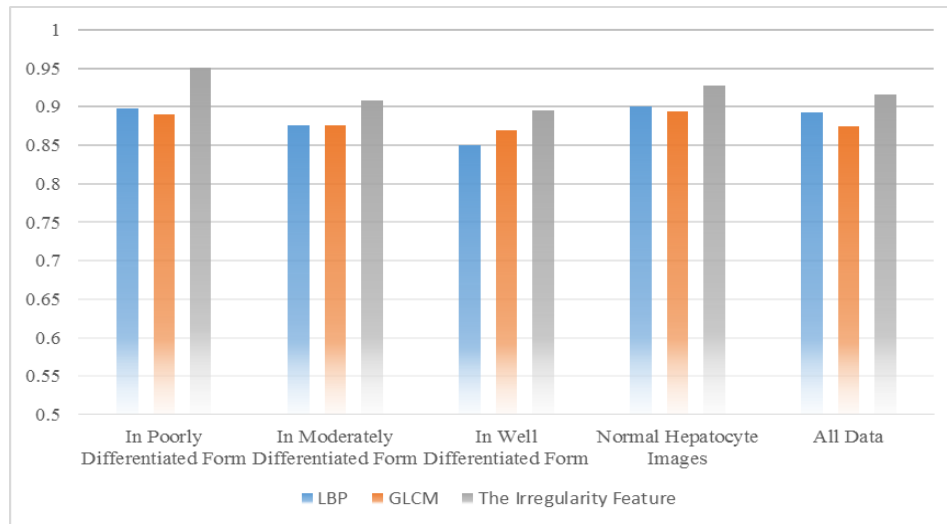


Fig.8 The results of recognition using LBP, GLCM and the irregularity feature respectively.

Fig.8 shows the results of HCC recognition with other features with histogram, in order to compare with the irregularity feature proposed in this paper. As shown in Fig.8, irregularity feature shows advantages of accuracy in all four kinds of images (low, moderately, high differentiated HCC pathological images and normal ones). In addition, on account of that, LBP and GLCM are both 256-dimensional. While the proposed

feature is 8-dimensinal (with single candidate center) or 20-dimentional (with non-single candidate center), the cost of extracting the proposed feature gets lower as a result.

Summary

The morbidity of liver disease rises fast, becoming an important factor that threatens human health and life. In order to recognize HCC as soon as possible and save human life, this paper presents a new method to improve the efficiency of HCC recognition. Liver pathological images are widely used in clinical medicine and meaningful in scientific research and clinical diagnosis. Hepatic cells with vein center as axis are syllabically and radically arranged, whose structure is called cell cords. As data displays, when liver cancer develops to an advanced stage, liver cell cords in pathological images become potentially thicken and may be destroyed in the end. So in the process of recognition of liver cancer cells, this paper starts from the distribution of hepatocyte to determine negative and positive hepatic tissue in images. First of all, binarization and morphological operations are carried out to input images. The next step is to calculate circularity and area of each connected regions to obtain candidate centers. Depending on the number of centers, using different ways to extract feature representing the irregularity of distribution. Experiments show that the proposed method can achieve good effect for the identification of high, medium and low differentiated hepatocellular pathological images data with low cost of computation.

Acknowledgement

The research is supported by the National Natural Science Foundation of China (No. 61472073). And thanks to the hospital and pathologists for offering experimental data and differentiate the images.

References

- [1] Crawford JM. The liver and biliary tract. In: Cotran R, Kumar V, Robbin S, Schoen F, editors. Pathologic basis of disease. Philadelphia:Saunders, 1994. pp. 831–95.
- [2] Fernandez MDP, Redvanly RD. Primary hepatic malignant neoplasms. RCNA 1998; 36: 333–48.
- [3] Oka H, Kurioka N, Kim K, Kanno T, Kuroki T, Mizoguchi Y, Kobayashi K. Prospective study of early detection of hepatocellular carcinoma in patients with cirrhosis. Hepatology 1990; 12: 680–7.
- [4] Shashi Bala Paul, Manpreet Singh Gulati. Spectrum of hepatocellular carcinoma on triple phase helical CT: A pictorial essay [J]. Clinical Imaging, 2002, 26(4): 270-279.
- [5] S.C.H Yu, D. T. K Yeung, N.M.C So. Imaging features of hepatocellular carcinoma [J]. Clinical Radiology, 2004, 59(2): 145-156.
- [6] DOI K. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential [J]. Computer Medical Imaging Graph, 2007, 31(4-5): 198-211.
- [7] LIU Jianhua, WANG Jianwei. Liver cancer diagnosis based on CT image processing [J]. J Tsing Univ(Sci & Technol), (2014, 54(7):917-923.

- [8] Kadah YM, Frag AA, Zurada JM, et al. Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images [J]. *IEEE Trans Med Image*, 1996, 15 (4):466-478.
- [9] Asvestas P, Matsopoulos GK, Nikita KS. A power differentiation method of fractal dimension estimation for 2-D signals [J]. *Journal of Visual Communication and Image Representation*, 1998, 9(4):392-400.
- [10] Lee CC, Chen SH, Tsai HM, et al. Discrimination for liver diseases from CT images based on Gabor filters [C]. *Proceedings of IEEE Symposium on CBMS Conference*. 2006.
- [11] LIU Wei. Liver CT Image Recognition Based on Texture Analysis and Artificial Neural Network [D]. Tai'an: Taishan Medical University, 2006.
- [12] YE Jun. Research on Key Techniques in Multi-Phase CT Image Based Computer-Aided Hepatic Lesion Diagnosis System [D]. Shanghai: Shanghai Jiao Tong University, 2010.
- [13] Po-Whei Huang, Yan-Hao Lai. Effective segmentation and classification for HCC biopsy images [J]. *Pattern Recognition*, 2010, 43(4): 1550-1563
- [14] Jitendra, Virmani, Vinod, Kumar, etc. SVM-Based Characterization of Liver Ultrasound Images Using Wavelet Packet Texture Descriptor [J]. *Journal of Digital Imaging*, 2013, 26(3): 530-543.
- [15] Deepti Mittala, Vinod Kumara, etc. Neural network based focal liver lesion diagnosis using ultrasound images [J]. *Computerized Medical Imaging and Graphics*, 2011, 25(4): 315-323.
- [16] Saima Rathore¹, Muhammad Aksam, etc. Texture analysis for liver segmentation and classification: a survey [J]. *2011 Frontiers of Information Technology*, 2011, (10): 121-126.
- [17] Chamidu Atupelage, Hiroshi, Nagahashi, etc. Noninvasive classification of hepatic fibrosis based on texture parameters from double contrast-enhanced magnetic resonance images [J]. *Journal of Magnetic Resonance Imaging*, 2013, 36(5): 1154-1161.
- [18] Information on <http://baike.baidu.com/link?url=IoFDFoTyhUj-UQUPbTS2t2zbK-flkWYxGTxwtl5uxgFQW95iqcDGDal8PcMnTuWTmI0JtTyr4jbNLrFgYXAts2a>
- [19] <http://www.eurocytology.eu/en/course/412>.
- [20] Jae Hyun Jeon, Jae Young Choi, Sihyoung Lee, Yong Man Ro. Multiple ROI selection based focal liver lesion classification in ultrasound images. *Expert Systems with Applications* 40 (2013) 450–457.
- [21] WANG Zhen, WANG Zhi-quan, MAO Yao-bin. A Description Based on Texture Direction and the Clustering and Segmentation to Directional Texture Images. *Journal of Image and Graphics*, 2002, 7(12): 1279~1284.
- [22] ZHANG Hong-ying, WU Ya-dong, WU Bin. Estimation on the Main Direction of Texture Using Image Edge Information. *Journal of Southwest University of Science and Technology*, 2007, 22(2): 44~47.

[23] ACM Trans. Intell. Syst. Technol., 2 (27) (2011), pp. 1–27 Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>