Journal of Advanced Engineering Research
ISSN: 2393-8447
Volume 1, Issue 1, 2014, pp.97-101

Detection and Classification of Lung Tissue Using Block Based Intensity Features

S. Sumaiya Banu[1,*], S. Syed Farmhand[2], P. Prabaharan[2], L. Malathi[2]

[1]Department of Information Technology, Er.Perumal Manimekalai College of Engineering, Hosur, India
[2]Department of Computer Science Engineering, Vivekanandha College of Engg for Women, Thiruchangode, India

*Corresponding author email: sweetsumi.banu@gmail.com

ABSTRACT

Content based image classification address the problem of retrieving images relevant to the user needs from image databases on the basis of low-level visual features that can be derived from the images. Grouping images into meaningful categories to reveal useful information is a challenging and important problem. Clustering is a data mining technique to group a set of unsupervised data based on the conceptual clustering principal: maximizing the intra class similarity and minimizing the interclass similarity. Proposed framework focuses on color as feature. Color Moment and Block Truncation Coding (BTC) are used to extract features for image dataset. Experimental study using K-Means clustering algorithm is conducted to group the image dataset into various clusters.

Keywords – CBIR, intensify feature

1. INTRODUCTION

Most of the Content-based Image Classification (CBIR) systems such as QBIC [1], Virage [2], Photo book [3] and Netra [4] use a weighted linear method to combine similarity measurements of different feature classes. Stricker and Dimai's method [5] segments each image into five partially overlapping fuzzy regions and extracts first two colour moments of each region both weighted by membership functions of the region to form a feature vector for the image. A colour space for CBIR is presented which provides both the ability to measure similarity using fuzzy logic and psychologically based set theoretic similarity measurement. These properties are shown to be equal or superior to the conventional colour space. C. Vertan et al. propose fuzzy colour histogram that classifies fuzzy techniques as crude fuzzy, fuzzy paradigm based, fuzzy aggregational and fuzzy inferential [6].

The Fuzzy Hamming Distance (FHD) is an extension of Hamming Distance for real valued vectors. Because the feature space of each image is a real-valued, the fuzzy Hamming Distance can be successfully used as image similarity measure. FHD is applied for colour histograms of the two images. The authors claim that FHD not only considers the number of different colours but also the magnitude of this difference. Certain research works supports concept-based image classification as well as the inexact match with a fuzzy triple matching performed when evaluating queries. An image is represented by a set of segmented regions each of which is characterized by a fuzzy feature reflecting colour, texture and shape properties. The resemblance between two images is then defined as the overall similarity between two families of fuzzy features and quantified by unified feature matching. Non-Boolean fuzzy and similarity predicates are used to rank tuples according to fuzzy based algebra. Soft queries in image classification systems present the use of soft computing and user defined classifications in multimedia database systems for content based queries.

2. PROPOSED WORK

Fig. 1 shows the block diagram of the proposed texture classification.

Generally, there are two types of classification methods: text based and content-based. Text-based classification is a popular method that annotates images by text and uses text-based data base management systems to perform image classification. However, there are two major difficulties with this method, especially when the image data base is huge. One is the vast amount of labor required in manual image annotation. The other difficulty results from the rich content of images and the subjectivity of human perception. Subjective perception can lead to imprecise annotations that may produce incorrect search results in subsequent classification processes.

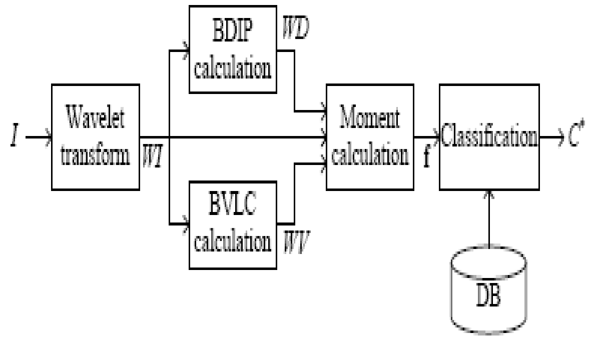

Fig. 1 Texture classification

The other method, content- based image classification (CBIR), does not suffer from these difficulties, for instead of relying on manual annotation, it indexes images by their visual features such as color, shape, and texture as given below

1. Color is one of the most widely indexed features in CBIR. It is invariant to image size and orientation. It can be indexed by feature descriptors such as color moment, color histogram, or color structure descriptor (CSD) which is one of MPEG-7 color descriptors.

2. Shape is a feature that represents the contour of an object in an image. It is invariant to the size and location of the object. However, it is difficult to extract the contour of an object correctly. Shape can be represented by feature descriptors such as Fourier descriptor, chain code, moment invariant, and Zernike moments.

3. Texture refers to innate surface properties of an object and their relationship to the surrounding environment. In the early 1970s, Haralick *et al*. [7] proposed the co-occurrence matrix representation of texture features. Tamura texture was subsequently proposed as an enhanced version. In the early 1990s, after wavelet transform was introduced, researchers began to study the use of wavelet transform in texture representation. Among classification methods using wavelet transform, the method using means and ariances extracted from wavelet sub bands is known to produce excellent texture images. Wavelet transform-based methods have also been combined with other techniques to achieve better image classification performance. Additional texture indexing techniques that have been proposed include: Gabor transform, which reflects characteristics of the human visual system; PIM (picture information measure), which has a property of entropy operator; and edge histogram descriptor (EHD), which is an MPEG-7 texture descriptor that captures the spatial distribution of edges.

The difference of inverse probabilities (DIP) is an operator for extracting sketch features that contain valleys and edges subject to local intensities. In the DIP, the ratio of pixel intensity in an image window to the sum of all pixel intensities in a window is considered as a probability. So, the name DIP means the difference between the inverse of the probability for the center pixel in a window and that for the pixel of maximum intensity in the window. BDIP is defined as the difference between the number of pixels in a block and the ratio of the sum of pixel intensities in the block to the maximum in the block.

$$\mathrm{BDIP} = M^2 - \frac{\sum_{(i,j)\in B} I(i,j)}{\max_{(i,j)\in B} I(i,j)}$$

Where I(i,j) denotes the intensity of a pixel (I,j) and a block of size MxM . The larger the variation of intensities there is in a block, the higher the value of BDIP.

Variation of local correlation coefficients (VLCC) is known to measure texture smoothness well. It is defined as the variation, or the difference between the maximum and minimum, of local correlation coefficients according to four orientations. The second texture feature we propose, BVLC, is a block-based version of the VLCC. Each local correlation coefficient is defined as local covariance normalized by local variance. That is

$$\rho(k,l) = \frac{\frac{1}{M^2}\sum_{(i,j)\in B} I(i,j)I(i+k,j+l) - \mu_{0,0}\mu_{k,l}}{\sigma_{0,0}\sigma_{k,l}}$$

Where B denotes a block of size MxM and μ0, 0 and σ0, 0 denote the local mean and standard deviation of the block B, respectively. The notation (k, l) denotes a pair of horizontal shift and vertical shift associated with the four orientations like (-90’, 0’45’, -45’). As a result, and represent the mean and standard deviation of the block shifted by (k,l), respectively.

Then BVLC is calculated as follows

$$\mathrm{BVLC} = \max_{(k,l)\in O_4}[\rho(k,l)] - \min_{(k,l)\in O_4}[\rho(k,l)]$$
$$O_4 = \{(0,1),(1,0),(1,1),(1,-1)\}.$$

The block diagram for the proposed work (retrieval block diagram) is given in Fig. 2.

1. When a query color image enters the system, each color component image is divided into non overlapping blocks of size MxM

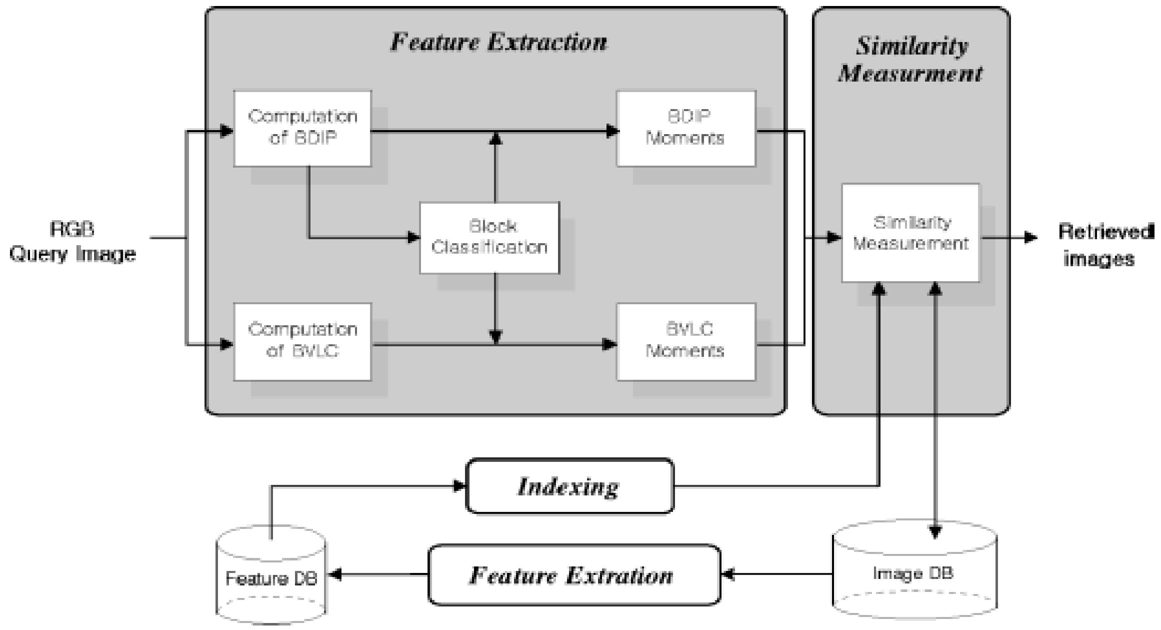

Fig. 2 Retrieval block diagram

2. The system then computes BDIP and BVLC in each block and classifies all the blocks into eight classes. The purpose of the classification is to reflect the characteristics of several homogeneous regions or objects, which an image generally contains, in the proposed features.

The block classification proceeds as follows. In the first step, all the blocks are classified into two groups, and the average of BDIPs over all blocks in the image is used as a threshold. In the second step,

All blocks in each of the two groups are classified again into two groups, but this time the average of BDIPs over all blocks in each group is used as a threshold. In the last step which repeats the same procedure as in the second step, all the blocks are classified into eight classes. After the block classification, the system computes the first and second moments of BDIP and BVLC for each class and combines these moments as a feature vector. The system finally calculates the distance between the feature vector of the query image and that of each target image in an image DB and retrieves a given number of the most similar target images.

3. IMPLEMENTATION

3.1 Algorithm

1. The entire given query image is taken one by one.

2. A function blocks bdip is called from the main file, so that bdip values are calculated for all the query images.

3. A function blocks bvlc is called from the main file, so that BVLC values are calculated for all the query images.

4. The query image is obtained as the input from the user.

5. The .mat file with an extension of feature.mat is compared and then the similar image files from the data base are displayed.

3.2 Blocks BDIP

1. Get the number of images in the current directory where all the query images are stored using imread command.

2. Get the number of horizontal and vertical blocks. Here it is taken as 4 and 4 respectively.

3. Then get one image and calculate the horizontal and vertical block length.

4. For the calculations to be done, first the image is converted in to float using double command.

5. Then in each block the maximum intensity pixel in found out using max command.

6. As per the formula of BDIP, we need to calculate the number of pixels in each block.

7. Then the total intensity of the image is calculated.

8. After that the following formula is worked out.

$$\mathrm{BDIP} = M^2 - \frac{\sum_{(i,j)\in B} I(i,j)}{\max_{(i,j)\in B} I(i,j)}$$

9. Then the partitioned images are shown in a figure window using subplot command.

3.3 Blocks BVLC

1. Get the number of images in the current directory where all the query images are stored using imread command.

2. Get the number of horizontal and vertical blocks. Here it is taken as 2 and 2 respectively.

3. Then get one image and calculate the horizontal and vertical block length.

4. For the calculations to be done, first the image is converted in to float using double command.

5. Based on the Chun's algorithm, we do compare the texture values of the blocks that we had divided into four blocks. From this we can estimate the smoothness of the image.

6. The blocks are taken and it is correlated with themselves as per the angle given below namely, -90', 0', +45' and -45'.

7. The BVLC values are computed as per the formula given below.

$$\rho(k,l) = \frac{\frac{1}{M^2}\sum_{(i,j)\in B} I(i,j)I(i+k,j+l) - \mu_{0,0}\mu_{k,l}}{\sigma_{0,0}\sigma_{k,l}}$$

and

$$\mathrm{BVLC} = \max_{(k,l)\in O_4}[\rho(k,l)] - \min_{(k,l)\in O_4}[\rho(k,l)]$$
$$O_4 = \{(0,1),(1,0),(1,1),(1,-1)\}.$$

8. The feature vectors are then extracted compared with those of the values stored in the mat file.

9. After comparison the values nearest to the target image is displayed using sub plot command.

4. RESULTS AND DISCUSSIONS

4.1 Screen shots of matlab results

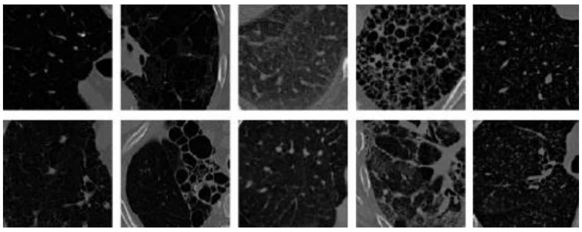

Fig. 3 Classified outputs

After the search process is over related images to lung tissue are as shown in output.

The advantages of the above work are

1. Annotations about the image are not required.

2. So data based required to store the annotations is reduced.

3. Picture information measure is better.

5. CONCLUSION

In image classification system, the content of an image can be expressed in terms of different features such as color, texture and shape. These low-level features are extracted directly from digital representations of the image and do not necessarily match the human perception of visual semantics. We proposed a framework of unsupervised clustering of images based on the color feature of image. Results are quite acceptable and showing that performance of BDIP and BVLC algorithm is better than color moments.

REFERENCE

[1] W.R. Webb, N.L. Muller and D.P. Naidich, High-Resolution CT of the Lung, *Philadelphia, PA: Lippincott Williams Wilkins,* 2008.

[2] A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P.A. Poletti and H. Muller, Building a reference multimedia database for interstitial lung diseases, *Comput. Med. Imag. Graph.,* 36 (3), 2012, 227–238.

[3] I. Sluimer, A. Schilham, M. Prokop and B. Van Ginneken, Computer analysis of computed tomography scans of the lung: A survey, *IEEE Trans. Med. Imag.,* 25 (4), 2006, 385–405.

[4] Y. Uchiyama, S. Katsuragawa, H. Abe, J. Shiraishi, F. Li, Q. Li, C. T. Zhang, K. Suzuki and K. Doi, Quantitative computerized analysis of diffuse lung

disease in high-resolution computed tomography, *Med. Phys.*, 30 (9), 2003, 2440–2454.
[5] M. Stricker, A. Dimai, E. Dimai, Color Indexing with Weak Spatial Constraints, *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1996.
[6] C. Vertan and N. Boujemaa, Embedded fuzzy logic in content based revival, *Procd. of 19th Int. Meeting North Americal Fuzzy International Society*, July 2010, 85-89.
[7] R.M. Haralick, K. Shanmugam and I. Dienstien, Textual Features for Image Classification, *IEEE Transactions on systems, man and cybernetics*, 3(6), 1973, 610-621.
[8] Y. Xu, M. Sonka, G. Mc Lennan, J. Guo and E.A. Hoffman, MDCT based 3-D texture classification of emphysema and early smoking related lung pathologies, IEEE Trans. Med. Image., 25 (4), 2006, 464–475.
[9] I.C. Sluimer, M. Prokop, I. Hartmann and B. Van Ginneken, Automated classification of hyperlucency, fibrosis, ground glass, solid, and focal lesions in high-resolution CT of the lung, Med. Phys., 33 (7), 2006, 2610–2620.