A New Color Information Entropy Retrieval Method for Pathological Cell Image

Xiangang Jiang, Qing Liang, Tao Shen

A New Color Information Entropy Retrieval Method for Pathological Cell Image

Xiangang Jiang ,Qing Liang , Tao Shen

Basic Science School of East China Jiaotong University,
Nanchang, Jiangxi, China (330013), jxg_2@tom.com

Abstract. For the distribution characteristics in a slice of pathological cell image, the system transforms them into the characteristic vectores by quantization and clustering in HSV color model, it promotes the concerned isolated pixel color description into the color feature ralative to its neighborhood's color histogram and color moments. By choosing appropriate neighborhood window size, it uses color information entropy method to gain efficient primary classification of gastric slice image and the similar retrieval results from the sample library images. Experiments show that this method has a better image classification and retrieval result.

Keywords: Mutual information,Color moments,Information entropy, Content based image retrieval

1 Introduction

With the unceasing development of the modern imaging technology and information processing technology, the number of images in medical scopes are more and more. By the submitting image, how to use the existing technical methods to retrieve and classify the comparative images from the qualitative sample library is crucial, which is used for clinical diagnosis and contrast, it is becoming one of the research task for efficient medical diagnosis.There are many kinds of image retrieval methods which focus on textual ,geometrical, and color features.In allusion to the pathological cell images, the color information are special and unique compered with other properties,and in the image registration and retrieval, Shannon mutual information is a exact representation of information of interactivity implicated in two color images. The paper makes use of neighborhood color moment histogram to extract color information and utilizes the theory of information entropy to realize shortcut image retrieval.

2 Shannon Mutual Information

Representation of mutual information derives from information theory, which is the measure of correlations of two variables. If use H to represent entropy, I represent an image, a represent the color values of an image, suppose the probability density function of I is $p_i = p(a=i), i=1,2,...,m$,then the Shannon Entropy of I is $H(I)$：

$$H(I) = \sum_{i \in I} p_i(i) \times \log \frac{1}{p_i(i)} \tag{1}$$

Where as $p_i(i) = 0$, supplementary definition：

$p_i(i) \times \log \frac{1}{p_i(i)} = 0$,as: $\lim_{p \to 0^+} p \log \frac{1}{p} = 0$.

To two images I_1 and I_2,their color values are described by 256 kinds of vectors, their joint entropy is defined as：

$$H(I_1, I_2) = \sum_{i=0}^{255} \sum_{j=0}^{255} P_{ij}(i,j) \log \frac{1}{P_{ij}(i,j)} \tag{2}$$

Where $P_{ij}(i,j)$ is the representation of color joint probability density distribution.

Their Shannon mutual information $I(I_1,I_2)$ is defined as：

$$\begin{aligned} I(I_1, I_2) &= \sum_{i=0}^{255} \sum_{j=0}^{255} P_{ij}(i,j) \log \frac{P_{ij}(i,j)}{p_i(i) p_j(j)} \\ &= H(I_1) + H(I_2) - H(I_1, I_2) \end{aligned} \tag{3}$$

Where $P_i(i)$,$p_j(_j)$ is the color probability density distribution of I_1 and I_2, $P_{ij}(i,j)$ is color joint probability density distribution,$H(I_1,I_2)$ is Shannon joint entropy. When $I(I_1,I_2)$=0, it shows that $I_1$and I_2 are independent mutually, if the value of $I(I_1,I_2)$ is bigger, which means the similarity is higher in the color distribution of two images.

3 Color Space Conversions

RGB color models can be better matched with the fact that people can fiercely percept the tricolor such as R, G, and B and express them in software developing language. However, RGB color model is not accorded to the human beings' visual perception dealing with distinguishing color similarity better than HIS, HSV, Luv, Lab and so on. Commonly used HSV's three component separately represent Hue, Saturation and Value[2]. This system firstly transformes r, g, and b in RGB color model to h,s and v in HSV color model to define color vector. For a cell image such as shown in Fig.1,the h component in the HSV color model is mainly concentrated in range $[0,100]$ and $[300,360]$ for a cell image, but the distribution of s and v component are relative uniformity. According to the characteristics, it will use the method in reference [1] for quantization and clustering of colors. Suppose the color vectors of pixel (i,j) are $p_{ij} \in [0,255]$.

4 Neighborhood Color Moment Mutual Information

4.1. Description of color moment

Any color distribution can be represented by its moments. In addition, as the color distribution information is mainly concentrated in lower-order-moments, so it is enough to express the color distribution of an image only by its first moment, second moment and third moment of colors. Comparing to color histogram, this method is good at the aspect that no needs to quantize the color characteristics. The lower-order-moments of color are expressed in mathematics as follows[3]:

$$\begin{cases} \mu = \frac{1}{N_A}\sum_{j=1}^{N} p_j \\ \sigma = (\frac{1}{N_A}\sum_{j=1}^{N}(p_j - \mu)^2)^{\frac{1}{2}} \\ s = (\frac{1}{N_A}\sum_{j=1}^{N}(p_j - \mu)^3)^{\frac{1}{3}} \end{cases} \quad (4)$$

Where p_j is the definition of color value of pixel j, N_A represents the whole number of pixels in an image.

As the ability to distinguish a image based on lower-order-moments of single pixel color is suboptimal, it doesn't make comprehensive consideration to its local color distribution of pixels. It is better to consider the neighborhood of a pixel as an analysis unit and it concerns the local color spatial distribution information in a pixel's around window. Suppose p_{ij} is the color value of concerned pixel (i, j), which is quantified and clustered in HSV color model, then its first, second and third order central moment are as follows:

$$\begin{cases} \mu_{ij}^{M} = \frac{1}{M \times N}\sum_{m=i-\frac{N-1}{2}}^{i+\frac{N-1}{2}}\sum_{n=j-\frac{M-1}{2}}^{j+\frac{M-1}{2}} p_{mn} \\ \sigma_{ij}^{M} = [\frac{1}{M \times N}\sum_{m=i-\frac{N-1}{2}}^{i+\frac{N-1}{2}}\sum_{n=j-\frac{M-1}{2}}^{j+\frac{M-1}{2}} (p_{mn} - u_{ij}^{M})^2]^{\frac{1}{2}} \\ s_{ij}^{M} = [\frac{1}{M \times N}\sum_{m=i-\frac{N-1}{2}}^{i+\frac{N-1}{2}}\sum_{n=j-\frac{M-1}{2}}^{j+\frac{M-1}{2}} (p_{mn} - u_{ij}^{M})^3]^{\frac{1}{3}} \end{cases} \quad (5)$$

The moving window is respectively selected as 3×3, 5×5, 7×7,the range of numerical values in extraction of three lower-order-moments of all pixels (i, j) is limited in [0，255].In allusion to each pixel in an image, it makes use of quantization and clustering color histogram of image referred above to obtain nine neighborhood color moments, (for a window 3×3), which are described in the region [0，255] just by rounding all float numbers. It makes the global statistics and normalization to the entire image and gets normalization histograms of nine neighborhood lower-order central moments.Fig.2 shows as：the normalization histograms of the first, second and third order central moment in 3×3 neighborhood window, where the red curve represents $H(\mu^3)$, the green one represents $H(\sigma^3)$,the blue one represents $H(s^3)$ and the black one represents the color histogram by quantified and clustered color vectors.

To 256 colors in a picture, it can make a statistics separately for the frequency of the first, second and third order central moment in the image, suppose the probability distribution of the k order central moment is q_{ik} ,then：

$$q_{ik} = N_{ik} / N_A \qquad (6)$$

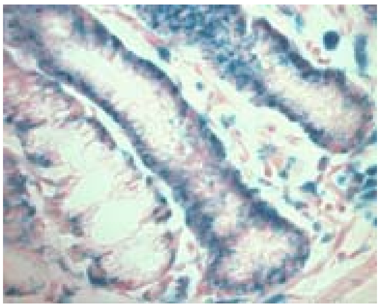

Fig.1. Pathological cell image

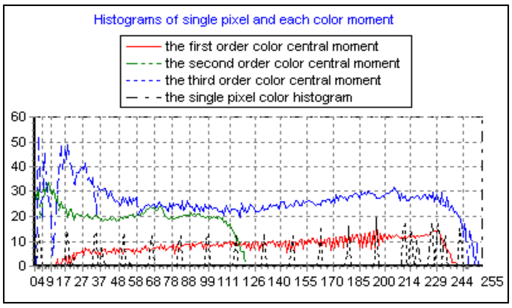

Fig.2. The compared histograms of the first, second,third order moment and the entire image's quantified and clustered color vectors in a 3×3 neighborhood window

Where N_{ik} represents the occurrence times of the k order color central moment, N_A is the representation of all pixel numbers in an image. Define the color joint probability density distribution of the k order central moment of two images by the same methods, which is supposed as q_{ijk} , where $i, j = 0,1,...255$, $k = 1,2,3$.

4.2 The calculation of mutual information of each order multi-neighborhood color moments

In order to define the mutual information of multi-neighborhood color moments, instead q_{ik} , q_{ijk} of $p_i(i)$, $P_{ij}(i,j)$ to obtain the definition of entropy and mutual information of multi-neighborhood color moments[4].Then the definition of entropy of each order multi-neighborhood color moments is：

$$H(I) = \sum_{i \in I} q_{ik}(i) \times \log \frac{1}{q_{ik}(i)} \qquad (k = 1,2,3) \qquad (7)$$

The mutual information of each order multi-neighborhood color moments between two images I_1, I_2 is defined as:

$$I(I_1, I_2) = \sum_{i=0}^{255}\sum_{j=0}^{255} q_{ijk}(i,j)\log\frac{q_{ijk}(i,j)}{q_{ik}(i)q_{jk}(j)} \qquad (k=1,2,3)$$
$$= H(I_1) + H(I_2) - H(I_1, I_2) \tag{8}$$

In all, the classification and retrieval for the pathological cell images can be gained not only by the mutual information of each order multi-neighborhood color moments, but also by the mutual information of synthesis vector normalized by each order neighborhood color moment.

5 The Experimental Analysis and Conclusions

In allusion to three hundred images downloaded from www.phoenixpeptide.com &www. microscopyu.com, sixty cancer images are included and the other ones are normal. It clearly shows that the main distinctions between cancer images and normal ones are the differences of their color distribution by comparing and analyzing their information entropy of color moment. In the experiment, it employs a new color information entropy retrieval method which promotes the concerned isolated pixel color description into the color feature of its neighborhood region color histogram and makes use of the theory of information entropy to gain the retrieval and primary classification to a further compared classification between the key image and the qualitative sample library. There are three test patterns designed in this experiment：(1)based on color mutual information of single pixel; (2)by making use of the isolated mutual information of the pre-three order color moment in window size of 3×3，5×5，7×7 separately; (3)based on the mutual information of the normalized characteristics vector of the integerated three order color moments. This system uses a as a key aimage and finds out the most similar eight images from the sample image library.

This system implements the retrieval and classification for the stomach cell images,it uses the mutual information between a key image's and sample library image's based on the single pixel to obtain the most similar eight images being showed in Fig.3. Fig.4 shows the retrieval results by employing the method based on the mutual information of the first order color moment in the window size of 5×5 .Each method's retrieval efficiency are showed as the table 1.

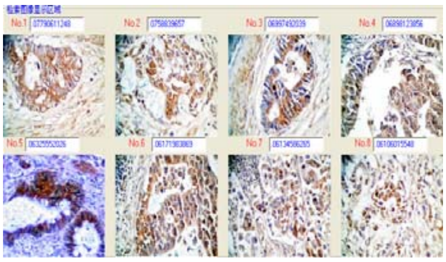

Fig.3. The retrieval result by the method

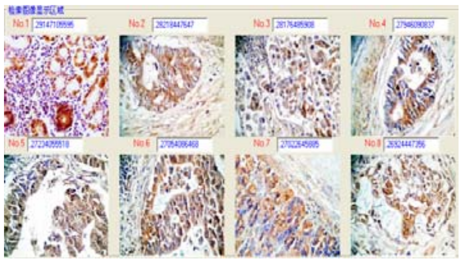

Fig.4. The retrieval result by the method

based on the single pixel based the first order color moment with a 5×5 window

Table 1. The comparison of methods of color mutual information retrieval

Retrieval method		Precision (%)	Precision of using moments synthesis (%)	Time (ms)	Total time of using moments synthesis(ms)
Window size	Single pixel color moment	80.65		6141	
3 × 3 window	The first order color moment	88.25	91.23	8828	18012
	The second order color moment	90.19		8875	
	The third order color moment	90.19		9234	
5 × 5 window	The first order color moment	87.88	92.56	8625	26412
	The second order color moment	91.30		12375	
	The third order color moment	92.65		13025	

From above experiments and above comparson table, we know that the retrieval efficiencies of the methods of mutual information of each order color moment regarding neighborhood region are higher than the one based on the single pixel, meanwhile, the retrieval efficiencies relative to the color moment order trend to be positive growth, the way based on synthesis of mutual information of each order neighborhood color moment has the best precision. At the point of consuming-time, the method of each color moment is longer relatively to the one based on the single pixel and it is in direct proportion to the order. Considering the neighborhood window sizes, while the geometric size of each tissue in an image is small, it will be better to choose a smaller neighborhood window, which has a high retrieval efficiency and low consuming-time, otherwise selecting a bigger neighborhood window to a bigger average tissue size in an image, it would avoid the characteristic fuzzy of each tissue in an image by adopting a flexible window size.

References

1. Chen Qin,Tai Xiaoying, Ba Teer: Gastroscope Image Retrieval Based on Histogram of Neighborhood Color Moments. J.Computer Engineering and Applications. 44(11),205-208(2008,).
2. Wang Yeqin: Expression on the Basis of the Timber Surface Color Characteristic of the Histogram and Color Moment Method. J.Forestry Science & Technology. 31(5), pp.56-58 (2006)
3. Dong Weijun, Zhou Mingquan, Geng Guohua: Image Retrieval Technique Based on Combined Features.J.Computer Applications and Software. 22(10), 34-36(2005)
4. Huang Xiaoling, Chen Guangxi: Color Image Retrieval Based on Color and Texture. J.Popular Science & Technology. 126(2), 13-15 (2010)