An Improving Technique of Color Histogram in Segmentation-based Image Retrieval

Zhenhua Zhang, Wenhui Li
College of Computer Science and Technology
Jilin University
Changchun , China
zhangzh@jlu.edu.cn

Bo Li
Department of Radiology of the Second Hospital
Jilin University
Changchun , China

***Abstract*—The distribution of pixel colors in an image generally contains interesting information. Recently, many researchers have analyzed the color attributes of an image and used it as the features of the images for querying [1,2,3]. Color histogram [1, 2, 3] is one of the most frequently used image features in the field of color-based image retrieval. The color histogram is widely used as an important color feature indicating the contents of the images in content-based image retrieval (CBIR) [4][5] systems. Specifically histogram-based algorithms are considered to be effective for color image indexing. Color histogram describes the global distribution of pixels of an image which is insensitive to variations in scale and easy to calculate. However, the high-resolution color histograms are usually high dimension and contain much redundant information which does not relate to the image contents, while the low-resolution histograms can not provide adequate discriminative information for image classification. And an image often includes a part of colors but not all, So there will be many accounts of colors are zeros. In order to save space, we shouldn't need store them. In this paper, a color high-resolution, non-uniform quantized color histogram is proposed and the improving representation about histogram is proposed too. Major color, major segmentation block, and a new Gray scale co-existing matrix's method are proposed.**

Keywords- Image Retrieval; Color Histogr; Content-based Image Retrieval (CBIR); Image Segmentation

I. Introduction

In recent years, more and more attention is focused on Content-Based Image Retrieval (CBIR), which is a sub-problem of Content Based Retrieval (CBR). The tremendous growth of the numbers and sizes of digital image and video collections on Web is making it necessary to develop power tools for retrieving this unconstrained imagery. In addition, CBIR is also the key technology for improving the interface between user and computer.

Among various low-level features, the color information has been extensively studied because of its invariance with respect to image scaling and orientation. In 1990, a simple and effective color-indexing scheme based on color histograms was proposed by Swain and Ballard. Further, attempts have been made to develop general purpose image retrieval systems based on color feature, such as QBIC system in IBM, Visual Seek system in Columbia University [6]. These systems contribute lots for the research of CBIR.

Content-based image retrieval uses the visual contents of an image such as color, shape, texture, and spatial layout to represent and index the images. But, the color features are absolutely necessary features in image retrieval, especially color histogram which is easy to compute and tolerate against the small changes of the view points, so it is used most widely. With the view to color feature's importance, we just focus on color features to study.

In this paper, a color high-resolution, non-uniform quantized color histogram is proposed and the improving representation about histogram is proposed too. The new color quantization gets much higher precision and the improving representation about histogram saves space. In order to testify our theory, the query system is finished. In the system, color images and gray images are all queried. The rest of the paper is organized as follows: In Section 2, we introduce the color space and color quantization. The improving of color histogram storage is listed in section 3. Major color, major segmentation block, a new kind of color quantization and a new Gray scale co-existing matrix's method are proposed in Section 4. In Section 5, we present the experiment and the conclusions.

II. Color Space and Color Quantization

Obviously, of all image features, color and texture are fundamental characteristics of the contents of all images. Color is also an intuitive feature and color features play more important role in image matching. The models of human perception of color differences are described in the form of color spaces [7], so the research on color image must be studied in a given color space. RGB, YIQ, YUV, HSV etc. are the most frequently used color spaces.

A. HSV Color Space

The representation of the colors in the RGB space is quite adapted for monitors, but for a human being, this is not a useful definition. To provide a user representation in the user interfaces, programmers prefer the HSL color space or HSV color space [8][9]. The acronym stands for Hue, Saturation, and Luminosity.

HSV color space is the often used one because of its accordance with human visual feature. HSV color space has two distinct characteristics: one is that lightness component is independent of color information of images; the other is that hue and saturation component is correlative with manner

978-0-7695-3744-3/09 $25.00
DOI 10.1109/IAS.2009.156

of human visual perception. So programmers would like to transform RGB color space to HSV(HIS) color space in image indexing.

B. *Color Quantization*

Color quantization is a process that reduces the number of distinct colors used in an image, usually with the intention that the new image should be as visually similar as possible to the original image.

In order to reduce the number of colors in images before color feature extraction, we should convert all our colors into a subset which is called quantization. J.R.Smith[10] designed a quantization scheme to produce 166 colors. Li Guohui designed a non-uniform quantization method to produce 72 colors. Along with the advanced analysis of color features in HSV space, we propose a new dividing method to quantize the color space into 12*4*4 colors.

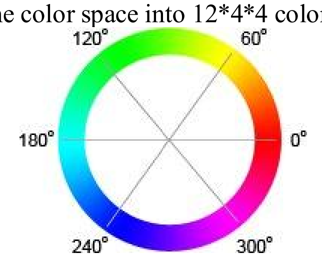

Figure 1. color ring

From above color ring, we find that each color is not uniform according to the range of hue. For instance, three primaries of red, green and blue occupy more space than other colors, three primaries' ranges are the most widest, secondary colors take second place, other colors' ranges are most narrowest. So we adopt non-uniform quantized method. So it's reasonable to quantize primaries into more colors and quantize secondary colors less colors. According to the reason, we quantize Hue just like this: (25,45],(45,75](75,95],(95,145],(145,165],(165,195],(195,215],(215,265],(265,285],(285,315],(315,335], (335,25].

Except for color component, saturation and value also adopt uniform quantization. But saturation is quantized into four parts, and value is also quantized into four parts, because they have less effect on color.

$$s=\begin{cases}0 & s\in[0,0.25)\\ 1 & s\in[0.25,0.5)\\ 2 & s\in[0.5,0.75)\\ 3 & s\in[0.75,1]\end{cases}$$

$$v=\begin{cases}0 & v\in[0,0.25)\\ 1 & v\in[0.25,0.5)\\ 2 & v\in[0.5,0.75)\\ 3 & v\in[0.75,1]\end{cases}$$

III. COLOR HISTOGRAM

Color is usually represented by color histogram, color correlogram, color coherence vector, and color moment under a certain color space. The color histogram serves as an effective representation of the color content of an image if the color pattern is unique compared with the rest of the data set. The color histogram is easy to compute and effective in characterizing both the global and local distribution of colors in an image. In addition, it is robust to translation and rotation about the view axis and changes only slowly with the scale, occlusion and viewing angle. So Color histograms are widely used for the Content-Based Image Retrieval.

The high-resolution color histograms cost much storage. In fact, there are some colors which they don't appear in some images. We don't need to store them. In this paper, we propose a method which doesn't store the information about the numbers of the colors are zeros, and give the distance arithmetic between two color histograms.

A. *Definition*

At first, we quantize the colors in images using above method. And then, we make color histograms;

$A(j)=i+H(i)\quad i\in[0,\ Nc),\ j\in[0,\ Nb),H(i)\neq 0$

Nc is the number of quantized color. Nb is the number of nonzero of H(i). i is the sequence of the ith color. We use A(j) to store those colors histograms which are nonzero. A(j) is a dynamic array.

B. *Similarity Calculation*

INT(A(j)) is a function which returns integer part of A(j) throwing away the decimal fraction. Ap(i) expresses A(i) in picture P and Aq(i) expresses A(i) in picture Q. Sim(P,Q) expresses the similarity between image P and image Q. Ncp is the number of nonzero in the color histogram of image P. Ncq is the number of nonzero in the color histogram of image Q.

```
i=0;
j=0;
Sim(P,Q)=0;
While (i<Ncp-1&&j<Ncq)
 Do{
  If(INT(Ap(i))== INT (Aq(j))) {
     Sim(P,Q)= Sim(P,Q)+min(Hp(i),Hq(i));
     i++;
     j++;
  }
  else if(INT(Ap(i))< INT (Aq(j)))
     i++;
  else
     j++;
}
```

IV. SEGMENTATION

In most cases, features are extracted in whole images or in fixed blocks such as [11][12][13]. There are disadvantages in the two cases. Space information is lost at the first case, and the invariance is lost to displacement and rotation

[11][12][13]. So we adopt segmentation algorithm, the extract feature in segmentation block. Higher-level problems such as recognition and image indexing can also be solved by making use of segmentation results in matching. So In this system, we use segmentation technique.

We regard an image as a graph, every pixel as a vertex. We use edges to describe neighboring vertex's relation. Weights on each edge measure the distance between vertexes. There will be an edge if two neighborings8-) vertex's color distance is very little, less than parameter k. The distance between two neighboring vertexes is described by L1-distance. However, unlike the classical methods, our technique adaptively adjusts the segmentation criterion based on the degree of variability in neighboring regions of the images.

$$w=(w_y\times|y_1-y_2|+w_u\times|u_1-u_2|+w_v\times|v_1-v_2|)/(|Y|+|U|+|V|) \quad (1)$$

w is a distance between two vertexes. y_i means luminancy, u_i and v_i means chromatism(i∈(1,2)). |Y| means the number of luminancy, |U| and |V| means the number of chromatism. w_y , w_u , w_v are three color channels' weight.

When $w \le \frac{1}{M\times N\times O}\times k$, we consider the two vertexes have an edge.

M,N,O describe the color channel's number of quantization rank. k is an adjustive parameter, which is a decimal fraction and whose value varies with texture. k will be bigger if images have more obvious texture, and will be smaller on the other situation.

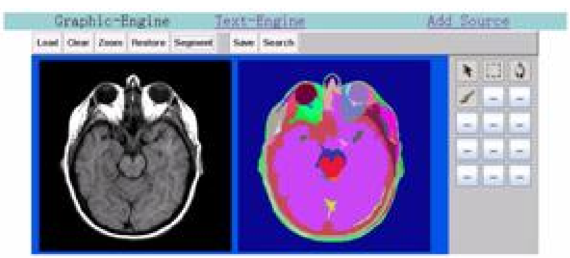

Figure 2. the result of segmentation

V. Experiments and Results

In this paper, our purpose is to describe a novel approach for replacing color histogram, so we just focus on color histogram.

We have developed the system using MATLAB 7.3. There are about 1000 images whose sizes are from 140×105,640×480 to 1024×768 in the experiment. There are two group experiments in this system. Figure 4 and figure6 adopts the conditional color histogram, and figure 5 and figures 7 adopts the new method. From the results, it proves that the new method is robust to translation and rotation, and is more excellent in based on originality, significance, quality and clarity. Certainly, to meet user's different need, we should consider feedback and which will be our further research issues.

Acknowledgment

This work is supported by the grants from 'China Postdoctoral Science Foundation' [Grant No. 20060183042], 'Jilin Planned Projects for Science Technology Development' [Grant No. 20060527, Grant No. 20070703], 'The National High Technology Research and Development Program of China' [Grant No. 2008AA10Z224], 'the National Natural Science Foundation of China'[Grant Nos. 60873147 and 60573182].The Project-sponsored by SRF for ROCS, SEM

References

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [1]. Where appropriate, include the name(s) of editors of referenced books. The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first . . ."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

[1] R. Brunelli, and O. Mich,"Histograms Analysis for Image Retrieval", Pattern Recognition, Vol. 34, 2001, pp.1625-1637.

[2] Y. K. Chan, and C. C. Chang, "A Color Image Retrieval System Based on the Run-Length Representation", Pattern Recognition Letter, Vol. 22, 2001, pp. 447-455.

[3] W. J. Kuo, "Study on Image Retrieval and Ultrasonic Diagnosis of Breast Tumors," Dissertation, Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi, Taiwan, R.O.C, January 2001.

[4] Datta Ritendrata, LiJ ia, Wang James Z. Content-based image retrieval-Approaches and trends of the new age[A]. In Proceedings of the 7th International Workshop on Multimedia Information Retrieval, in Conjunction with ACM International Conference on Multimedia[C], Singapore, 2005:253-262.

[5] Lu, W., Tan, Y. P., A Color Histogram-Based People Tracking System, The IEEE International Symposium on Circuits and Systems, Vo1.2(2001), pp. 137-140

[6] Smith J R, Integrated spatial and feature image system: Retrieval, analysis and compression ,Ph D dissertation, Columbia University, New York, 1997

[7] Datta Ritendrata, LiJ ia, Wang James Z. Content-based image retrieval-Approaches and trends of the new age[A]. In Proceedings of the 7th International Workshop on Multimedia Information Retrieval, in Conjunction with ACM International Conference on Multimedia[C], Singapore, 2005:253-262.

[8] WANG Tao, HU Shi-min, SUN Jia-guang, Image Retrieval Based on Color-Spatial Feature , Journal of Software, Vol.13, No.10 ,2002

[9] Brunelli,R.,and Mich,O., Histograms Analysis for Image Retrieval, Pattern Recognition, Vol.34, (2001) 1625-1637

[10] Smith J R, Integrated spatial and feature image system: Retrieval, analysis and compression ,Ph D dissertation, Columbia University, New York, 1997

[11] He Shan, Guo Baolong, Hong Junbiao, Region Entropy-based Image Retrieval, Computer Engineering, Vol.32 No.18 2006.9,pp214-216

[12] Tian Yumin,Mei Lixia, A Method of Image Segmentation Based on Color and Texture Features, computer engineering and application, 2004:p92-94

[13] Pan Hong, Wu Fei, A New Image Retrieval Method Based on Independent Keyblock, Jounal of Computer Research and Development , Vo.l40,No.7,2003.4 2007, pp. 57-64, doi:10.1109/SCIS.2007.357670.

Figure 3. the images retrieving result of traditional color histogram

Figure 4. he images retrieving result of the new method

Figure 5. the color images retrieving result of traditional color histogram

Figure 6. the color images retrieving result of he new method