2009 International Forum on Computer Science-Technology and Applications

Research of Image Retrieval Based on Color

Bai Xue, Liu Wanjun
School of Software, Liaoning Technical University, Huludao, 125105,China
xiaopanghuamao@163.com

ABSTRACT: ***Research on image retrieval technology based on color feature, for the color histogram with a rotation, translation invariance of the advantages and disadvantages of lack of space, a color histogram and color moment combination Image Retrieval. The theory is a separate color images and color histogram moment of extraction, and then two methods of extracting color feature vector weighted to achieve similar distance, similar to the last distance based on the size of the return search results, based on the realization of the characteristics of the color image Retrieval system. The results show that the method is rotation, translation invariance, a single method of extracting color features, enhanced image search and improve the accuracy of the sort.***

KEYWORDS: ***image retrieval; color feature; histogram; similarity distance***

I. INTRODUCTION

With the rapid development of computer technology and network communication technology, especially the emergence and popularization of Internet, rapid expansion of the size of the multimedia database such as digital image, every day hundreds of millions of military or civilian image data is stored in the database. Information Retrieval[1] (Information Retrieval) was proposed by Calvin Moores in1951. It is a process that organized and stored the information according to a certain way, and in accordance with the needs of users to find the interrelated information, it is also called "Information Storage and Retrieval". However at that time it only means by the text of the document retrieval. By age 70, database experts began to study how to manage image data effectively, The main method of image files is to establish Keywords or text description of the title as well as some additional information, and then establish a link between storage path and the keywords of the image, which is text-based image retrieval. However, with the storage capacity of images to start using GB or TB, the own shortcomings of text-based image retrieval technology led to two difficulties in the retrieval: First, it has been impossible to note each image; Second, the subjectivity and non-precision of image annotation may lead to the adaptation in the retrieval process. In order to overcome these problems, Content-Based Image Retrieval (CBIR) was proposed in the 90's.

II. THE TECHNOLOGY OF IMAGE RETRIEVAL BASED ON CONTENT

Content-Based Image Retrieval (CBIR) [2] According to the user-supplied bottom characteristics, directly find out images containing specific content from the image library.

The basic process: First of all, do appropriate pre-processing of images, and then extracte image characteristics needed from the image according to the contents of images to keep in the database. When we retrieve to identify the image , extract the corresponding features from a known image and then retrieve the image database to identify the images which are similar with it, also we can give some of the characteristics based on a queried requirement , then retrieve out the required images based on the given eigenvalues.

In the whole retrieval process, feature extraction is crucial [3] , it is closely related to all aspects of the future. Thus the research of content-based image feature extraction has been the focus of people's attention, but also the focus of this paper, the characteristics of images mainly refers to the bottom of the image features, such as color, shape, texture and space, in recent years, many scholars began to study high-level semantic features of images. The color, shape and texture characteristics thereof as the most basic visual features of images, is the most commonly used features in image retrieval. In practice, a feature or a combination of several characteristics is often used to search for images, then let's study on the various issues of image color feature extraction.

III. COLOR MODEL AND COLOR QUANTIZATION

A. Color model

Color feature is one of the most significant features of image retrieval. There are many color models to express color such as the RGB color model, YUV color model and the HSV color model

HSV model thereof is most consistent with the induction of the human visual model[4]. *H* represents color hue, and it is the wavelength of the light reflected from an object or throughout the object; *S* represents color saturation, means how much white is added to the color; *V* represents brightness (vale), is the degree of color shading. But, the computer can only identify the RGB color component of an image, in which *R* represents the red component, *G* represents the green component, *B* represents the blue component. Therefore, we need the following formula for the image conversion from RGB color space to HSV color space:

978-0-7695-3930-0/09 $26.00 © 2009 IEEE
DOI 10.1109/IFCSTA.2009.74

$$h' = \begin{cases} \dfrac{(g-b)}{\delta} & if \quad r = \max \\ \dfrac{2+(b-r)}{\delta} & if \quad g = \max \\ \dfrac{4+(r-g)}{\delta} & if \quad b = \max \end{cases} \qquad \begin{aligned} h &= h' * 60 \\ s &= \frac{\max - \min}{\max} \\ \mathrm{v} &= \frac{\mathrm{r}+g+b}{3} \end{aligned} \tag{1}$$

Thereinto, max=MAX(r,g,b), min=MIN(r,g,b), δ=max-min, $h \in [0,360]$, $(s,v) \in [0,1.0]$

B. Color quantization

For a true color image, the number of the kind of the colors up to 2^{24} = 16777216, so directly extract color feature from true color will lead to large computation. In order to reduce the computation, without a significant reduction in image quality, some representative color is extracted to represent image, thereby to reduce storage space and enhance the purpose of processing speed[5]. There are 36, 72 and 256 quantiative color quantization and we can depend on different need to select different levels of quantitative methods. Now we introduce the 256 quantitative.

In this way, we can get 256 handles of one-dimensional histogram by the quantified image.

IV. EXTRACT METHOD OF COLOR FEATURE

A. Color histogram

Color have a certain stability, and it is not sensitive to size and direction, so the image retrieval technology using color characteristics has been pay great attention to. The retrieval method of using color characteristic was originally proposed by Swain and Ballard, they put forward t the color histogram [6] method of which the core idea is to use a certain color space quantization method for color quantization, and then do statistics for the proportion of each quantitative-channel in the whole image color. Abscissa represents the normalized color value, ordinate represents the sum of image pixels which corresponding to each color range. Image statistical histogram is a one-dimensional discrete function:

$$h_k = \frac{n_k}{n}, \quad k=0,1,\cdots,L-1 \tag{2}$$

The letter k presents eigenvalues of color,letter l presents the number of features of value . So we get the color histogram of the image P as follows:

$$H_{(p)}=[h_1, \ h_2, \ \ldots h_{L-1}] \tag{3}$$

There are many color histogram methods such as the global color histogram, cumulative histogram and sub-block histogram.

However, color histogram has its own drawbacks, such as the color histograms of different images may be the same: there are two unrelated images in Figure 1 , but they are the same as the color histogram in Figure 2.

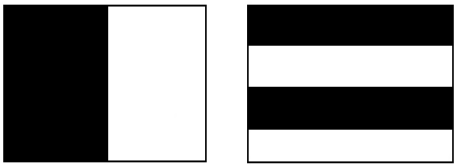

Figure 1. Two unrelated images

Figure 2. Color Histogram

B. Color Moment

Stricker and Orengo who propose the method of color moment [7] consider that the color information focus on the low-level color moment of the image, and they mainly do statistics for the thfirst order, second-order and third-order moment of each color component. For image retrieval, the color moment is a simple and effective representative method of color features. Such color moment as first-order (mean) and second (variance) and third-order (gradient), is proved to be very effective in presenting color distribution of images. The three colors moments are defined with figures as follows:

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} f_{ij} \tag{4}$$

$$\sigma_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(f_{ij}-\mu_i\right)^2} \tag{5}$$

$$\mathrm{s}_i = \sqrt[3]{\frac{1}{N}\sum_{j=1}^{N}\left(f_{ij}-\mu_i\right)^3} \tag{6}$$

In which, f_{ij} presents the probability of the color component of pixel j is i, N presents the number of pixels in the image. Each pixel has three color components, and each component has three low-order moment, so there are nine color characteristics components $\mu_1, \sigma_1, s_1, \mu_2, \sigma_2, s_2, \mu_3, \sigma_3, s_3$.

However, color moment is only the initial color characteristics extraction of image, and the effect of extraction is very rough[8], so is regularly integrated used of other extraction methods.

In practice, there are other color feature extraction methods such as color sets, color correlogram, color polymerization vector and consistent color vector.

V. THE IMPROVED EXTRACTION METHOD OF COLOR FEATURE

In view of the defects of the two extraction method of color feature color histogram and color moment method, we select an integrated approach of the two methods to extract the color feature in this paper in order to improve the retrieval accuracy and reform the ranking. The concrete steps are as follows:

First: Realizing the conversion of the color space:

We can get the HSV color value of each color in the image by using the formula for the image conversion from RGB color space to HSV color space.

Second: Color quantization:

All colors of the images in 256 color quantization as shown in formula.

Third: Using color histogram to extract color feature :

Extract the color histogram of original image by using formula, and then we will get the 256-dimensional color histogram:

$$H_{(P)}=(h_0,\ h_1,\ h_2,\ \ldots h_{255}) \tag{7}$$

Histogram intersection method is used to achieve the similar distance between image Q of the image library and the sample image P, the formula is as follows:

$$S_1 = 1-\sum_{l=0}^{255} \min(h^{P}_{l}, h^{Q}_{l}) \tag{8}$$

In which l = 1,2, ... 9, is quantitative series. From the similar distance formula above can be seen, if the color distributions of the image corresponding blocks are the same, the calculated result will be 0; else the result will be 1, that is, the value of the similar distance between image corresponding blocks is either 0 or 1.

Forth: Using color moment to extract the features of color image.

Extract the color feature by using the color moment method on original image with formula and then we will get the nine color feature component of each color $\mu_1, \sigma_1, s_1, \mu_2, \sigma_2, s_2, \mu_3, \sigma_3, s_3$

Using Euclidean distance formula [5] to calculate the similar distance between image Q of image library and the sample image P:

$$S_2 = \sqrt{\sum_{i=1}^{9}(P_i - Q_i)^2} \tag{9}$$

In which *Pi* and *Qi* respectively indicates the color moment feature vector of image P and image Q. From the formula above can be seen, if the color distributions of the image corresponding blocks are the same, the calculated result will be smaller; else the result will be greater.

Fifth: Similar match:

After the step three and four, we get the similarity and similar distance by using two types of color feature extraction method. In this paper we use the method weighted combination of the two to get the total similar distance of color features between the image Q of image library and the sample image P:

$$S = \omega_1 S_1 + \omega_2 S_2 \tag{10}$$

In which, $\omega_1 + \omega_2 = 1$, here we take $\omega_1 = \omega_2 = 0.5$

Sixth: In accordance with the similar distance, according to the order from small to large to return to retrieval result.

VI. EXPERIMENTS AND CONCLUSIONS

In this paper, the verification system is developed by vc + +6.0. Experimental Gallery is composed of 1000 images chosen from a standard test images Corel including snow-capped mountains, flowers, dinosaurs and the interference images.

Judge the system performance by precision, recall and sorting.

1) *Precision:*

$$P = \frac{number\ \ of\ \ relationally\ \ retrieval\ \ images}{number\ \ of\ \ all\ \ retrieval\ \ images}$$

2) *Recall:*

$$P = \frac{number\ \ of\ \ relationally\ \ retrieval\ \ images}{number\ \ of\ \ all\ \ relational\ \ images}$$

3) *Sorting:*

Figure 3. the retrieval result of global color histogram method

Figure 4. the retrieval result of color moment method

Figure 5. retrieval result by the method proposed in this paper

The experimental results and analysis:

Fig.3 use global color histogram method, due to the lack of color space information, retrive images not in accord with requirements such as pieces of 5,6,9,10,11,13, and the obviously not relevant piece 5 in front of 6,7,8,12, so the sorting is less effective. But due to invariance of translation rotation, the flipped image of the example image were retrievaled.

Fig.4 use color moment method. feature extraction of this method is more rough ,and the piece 8 and 11 ranked behind piece 2 and 7, so the sorting is less effective, while the flipped image of example image were not retrievaled.

Fig.5 use the integrated method proposed in this paper , not only retrievaled the flipped image of the example image and arranged it in the second place but also the recall and precision had been improved and the image sorting was better.

TABLE I. COMPARISON OF THE THREE METHODS

Retrieval Algorithm	Precision	Recall	Sorting
Global histogram method	47%	52%	bad
Color moment	39%	63%	bad
Method proposed in this paper	76%	74%	good

REFERENCES

[1] Zhang Xinjun."Research of image retrieval based on color features" [D].Liaoning technical University. 2006, 9(2), pp.42-50.

[2] Rao Juan. "Intelligent image retrieval technique based on content"[D]. Lanzhou University of Technology. 2006, pp. 32-64.

[3] Ye Yu-guang, "Research of image retrieval based on fusing with multi-character", Hua Qiao university, 2007, pp.14-36

[4] Zeng Tai-sheng, "Study on the image retrieval of based on color and texture and feedback", Hua Qiao University,2006, pp.34-25.

[5] Hafner J, Sawhney H S , "Efficient color histogram indexing for quadratic form distance functions", IEEE Transactions on Pattern Analysis and Machine Intelligence, Intelligence,1995,17(7): pp.729-736.

[6] Swain MJ and Balland DH. "Color indexing of Computer Vision" [J], 1991-7(1):11-32.

[7] M. Stricker and A. Dimai. "Color indexing with weak spatial constraints" [C]. IS&T/SPIE Conf. on Storage and Retrieval for Image and Video Databases IV, Vol. 2670, San Jose, CA: 1996,pp.29-40.

[8] Wang Wen-hui , Zhou Liang-zhu. "Image retrieval technology research and development based on content" Computer Engineering and Applications, 2001-5, pp.54-56.