

An Improved Image Retrieval Based on Color Histogram

Denghui Li^{1,a}, Yanhong Wang^{2,b}

¹Institute of Information Technology Guilin University of Electronic Technology, GUILIN, China

²Guilin University of Electronic Technology, GUILIN, China

^aldhui@guet.edu.cn, ^bwangyh@guet.edu.cn

Keywords: image retrieval; color histogram; euclidean distance

Abstract. Image retrieval based on the color histogram is one of the technologies based on the content. The processing time of the method is shorter and easy to implement, but its precision need to be improved. In this paper, the algorithm is improved. The same image library is retrieved by the image retrieval method based on cumulative histogram and the normalized cumulative histogram. The recall and precision are obviously improved. The results show that the improved image retrieval algorithm based on the color histogram is more effective.

Introduction

In recent years the information technology is developed rapidly. How to get the useful information from the vast image library is the focus of research in recent years. The color, texture and shape of the image characteristics are more intuitive and more important. The color histogram meaning the pixel distribution is extracted simply and be not sensitive to rotation. So the image retrieval based on the content is used widely.

Color is the most significant underlying characteristic of image^[1]. The histogram based on global color is the most classic image retrieval. The method is quick and easy, but it can't reflect the color spatial information^[2]. Later, many improved histograms based on histogram are proposed. The algorithm based on space model is adopted by literature 3. The changing information of image edge is described, but the color information of the edge is not considered.^[3]The image retrieval algorithm based on image edge is adopted by literature 4. The color, direction and the space of image is not considered.^[4] The improved algorithm based on color component quantization image retrieval is put forward in the paper above analysis.^[5-7]

Feature Extraction and Quantitative

In the common color space model, the RGB color space model is difference from the human eye perception largely. The shade, brightness and saturation of the HSV color space model are more compared with human visual perception. So the RGB model is converted to the HSV model in order to represent the image features effectively.^[8]

In HSV space, in order to reduce dimension and calculation, the H channel, S and V channel converted need to be quantified.

Assuming that quantitative values are 8,3and 3 respectively, the 72-dimensional color histogram can be obtained finally. If the position any pixel in image is set to (x, y) , the corresponding value of each pixel is $HSV(x, y) = n, n \in \{1, 2, \dots, 72\}$.

First, the histogram of H component, S component and V component is got by calculating. The formula corrected by gamma is as follows:

$$\begin{bmatrix} V \\ S \\ H \end{bmatrix} = \begin{bmatrix} \max \\ (\max - \min) / \max \\ H_1 \end{bmatrix} \quad (1)$$

$$H_1 = \left\{ \begin{array}{l} (G - B) / \delta \quad R = \max \\ 2 + (B - R) / \delta \quad G = \max \\ 4 + (R - G) / \delta \quad B = \max \end{array} \right\} \quad (2)$$

Here, $\delta = \max - \min$, max is the maximum of RGB, min is the minimum of RGB.

Second, the Euclidean distance from the retrieval image to the image in the library is calculated. The distance is such as (3) below:

$$D(M_1, M_2) = \sqrt{\sum_i (H_1(r) - H_2(r))^2} \quad (3)$$

Here, $H_1(r)$ is the normalization histogram of image M_1 . $H_2(r)$ is the normalization histogram of image M_2 .

Finally, the images are displayed according to the distance in turn from the library, and the image retrieved is realized. The steps are as follows:

- (1) The reference image is read in, and the color histogram of the HSV space model is calculated.
- (2) The image in library is read in turn, and the three components of H, S and V are calculated.
- (3) The Euclidean distance from the retrieval image to the image in the library is calculated.
- (4) The images are displayed according the sorting distance.

Image Retrieval Based on Color Histogram Accumulated

The image retrieval accuracy based on the general color histogram is not high for the similarity of adjacent color. The misdetection cannot be avoided in general.

For example, Euclidean distance is measured between the red, pink and blue. The similarity in red and pink is equivalent to the similarity in blue and red. This doesn't accord with visual perception. The question can be solved by secondary Euclidean distance. But the method is not used due to the complicated calculation.

The problem can be solved by the accumulative color histogram. The accumulative histogram is color as the X-axis, accumulative frequency as Y-axis. In accumulative histogram, the adjacent color is relevant. The zero values in the histogram are eliminated and the bad retrieval effect can be overcome due to gross quantization. The image feature is extracted based on the normalized accumulative histogram in order to eliminate the influence of the translation, scaling and rotation. The steps are as follows:

- (1) The retrieval image is read in, and the normalized accumulative histogram of HSV is calculated.
- (2) The image in library is read in turn, and the three normalized accumulative components of H, S and V are calculated.
- (3) The Euclidean distance from the retrieval image to the image in the library is calculated.
- (4) The images are displayed according the sorting distance.

Improved Local Cumulative Histogram

Deferent shade signals are confused based on the color histogram for complex scene image. The confusion is reduced based on local cumulative histogram dividing the similar area.

1° is used on the quantization interval, and the local area of 60° is as a basic area. The three color components are converted to one-dimensional vector according to the quantization. The formula is as follow:

$$L = HQ_s Q_v + SQ_v + V \quad (5)$$

Here, Q_s and Q_v is quantitative series of S and V respectively.

The cumulative histogram is calculated for each local area. If the misdetection is more because

of the larger quantization interval, then quantization again and calculation again is done based on the first time. The two histograms are added and the average as the image feature of is taken. The steps are as follows:

- (1) The retrieval image is read in, and the normalized accumulative histogram of HSV is calculated. The two local quantitative values and the average are calculated respectively.
 - (2) The image in library is read in turn, cumulative histogram and the image features are calculated respectively.
 - (3) The Euclidean distance of local feature from the retrieval image to the image in the library is calculated.
 - (4) The images are displayed according the sorting distance.
- The image retrieval accuracy based on the general color.

Simulation Analysis

In order to verify the effectiveness of the improved algorithm in this paper, the library is selected. The whole process is realized with MATLAB simulation environment. The image library we used contains 100 images, including flowers, animals, buildings and the sea.

The retrieval image is selected from the library and five images are extracting from each category. The top 20 images are as retrieval result. When the flower is as the example, the retrieval result is as shown in figure 1, figure 2 and figure 3. The retrieval effectiveness used cumulative histogram is better than the retrieval used global histogram comparing the fig 1 and fig2. The disadvantage is overcome by the second quantization and extracting local cumulative histogram as the feature from fig3. The images retrieved are more and the retrieval effect is better when the same flower is retrieved.

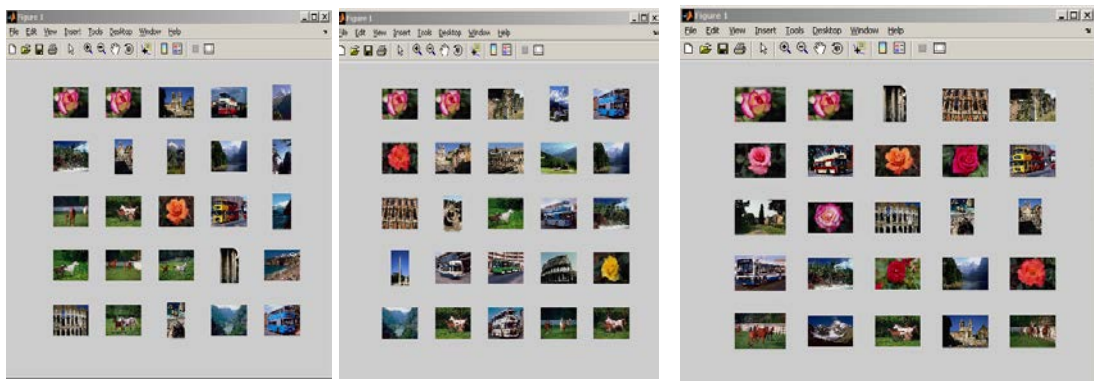


Fig1 CH algorithm

Fig2 CHA algorithm

Fig3 LCH algorithm

The objective standard, precision, is adopted in order to compare the performance of the algorithms. The five types of flowers, cars, buildings and so on are selected from library. Assuming that the retrieval images are n_1 , the images returned are n_2 , the precision is n_1/n_2 . The precision of algorithms are get, such as in table1.

TABLE I. IMAGE PRECISION OF ALGORITHMS

| algorithm | flower | bus | building | sea | horse |
|-----------|--------|------|----------|------|-------|
| CH | 0.32 | 0.42 | 0.35 | 0.52 | 0.41 |
| CHA | 0.35 | 0.51 | 0.45 | 0.55 | 0.52 |
| LCH | 0.82 | 0.75 | 0.65 | 0.73 | 0.65 |

The precision of improved algorithm is better than other algorithms from table1.

Conclusion

The color is the important feature in retrieval image based on content. The improved algorithm based on local cumulative histogram in the paper is adopted to overcome the shortcoming of the similar color, scaling and rotating.

The result shows that the retrieval accuracy is improved and the dimension of space is reduced. The part of image attribute is expressed by the single color feature, and the visual demand can't be satisfied commonly. The texture feature is also an important feature of the image. Therefore, the texture feature is integrated into image retrieval in the next step. The retrieval precision is improved based on the texture feature and the color feature in the later research.

Acknowledgment

This work was supported by Guangxi Key Lab of Wireless Wideband Communication & Signal Processing (GXKL0614105)

References

- [1] Samuel Rota Bulò, Massimo Rabbi, Marcello Pelillo. Content-based image retrieval with relevance feedback using random walks[J]. *Pattern Recognition* . 2011 (9): 152-156.
- [2] Junding Sun, Fang Yuan. Content-based image retrieval technology[J]. *Computer Systems & Applications*, 2011, 20(8) : 240 -244.
- [3] Gauri Deshpande, Megha Borse. Image Retrieval with the use of different color spaces and the texture feature [C]. *International Conference on Software and Computer Applications(ICSCA 2011)*, 2011-07-01 :69-77.
- [4] Rongguo Zhang; Zongjian Zhang; Li Li and Liyang Peng. Edge-Enhanced ShapeFeature Extraction for Image Retrieval [C]. *2012 International Conference on Network and Computational Intelligence (ICNCI 2012)*: 332-335.
- [5] Guang-Hai Liu, Jing-Yu Yang. Content-based image retrieval using color difference histogram[J]. *Pattern Recognition* . 2013 (1):188-198.
- [6] Chandan Singh, Pooja. An effective image retrieval using the fusion of global and local transforms based features[J]. *Optics and Laser Technology* . 2012 (7):2249-2259.
- [7] Guang-Hai Liu, Zuo-Yong Li, Lei Zhang, Yong Xu. Image retrieval based on micro-structure descriptor[J]. *Pattern Recognition* . 2011 (9):2123-2133.
- [8] B.J. Chen, H.Z. Shu, H. Zhang, G. Chen, C. Toumoulin, J.L. Dillenseger, L.M. Luo. Quaternion Zernike moments and their invariants for color image analysis and object recognition[J]. *Signal Processing* . 2012;92 (2):308-318.