Content-Based Image Retrieval Using Color Existence Features

Choong-woong Yoo¹, Myeong-chul Park², Yong-ho Moon³, Jing Zhang³, and Seok-wun Ha³*

¹Dept. of Naval Architecture, Sorabol College, Rep. of Korea
²Dept. of Computer Information, Songho College, Rep. of Korea
³Dept. of Informatics, Gyeongsang National University, Rep. of Korea
swha@gnu.ac.kr³*

Abstract

The most popular technique for image retrieval in a database of color images is the comparison of images based on their color histogram. The color histogram describes the distribution of colors in the color space of a color image. In the most image retrieval systems, the color histogram is used to compute similarities between the query image and all the images in the database. But, small changes in the resolution, scaling, and illumination may cause important modifications of the color histogram, and so two color images may be considered to be very different from each other even though they have completely related semantics. It is a result that color features are only limited to frequencies of colors. When the color histogram is computed, color existence information is included but not used in fact. So we adapted this color existence property and could earn good results in image retrieval. In this paper, we propose two color existence representation methods and compare them with the traditional color histogram methods. The experimental results reveal that the proposed is less sensitive to small changes in the image and that achieve higher retrieval performances than the traditional color histograms.

1. Introduction

Color is a very important feature in extracting information from a color image, and color histogram comparison has recently become a popular technique for image or video retrieval [1-9]. The color histogram comparison is used as a preliminary step for database indexing in order to reduce the number of candidate images for next steps which could use other features (e.g. texture, shape) to compare database images with a given query image. In general, the RGB color space is inappropriate for image retrieval due to the fact that small changes in the resolution, scaling, and illumination may cause important modifications of the color histogram, and two color images may be considered to be very different from each other even though they have completely related semantics [2], and so other color spaces like HSI, YIQ, or LUV are widely used for image retrieval [3,4,5].

Actually, the RGB color space is related to the color signal from the image capture devices like camera, video camera, or scanner, and so it is need to use directly the RGB color space without a color space translation to other color spaces [1]. In this paper, two color existence representation methods – color position feature and 2-dimensional projection maps method [9]– are proposed. Color position feature is constructed by extracting position information that colors are located from color histogram. And 2-dimensional projection maps are consisted of three 2-dimensional color spaces separated from 3-dimensional RGB color space [9].

2. Traditional Color Histograms

In the traditional color histogram, there are two methods, RGB color histogram and Hue histogram.

Usually, RGB color histogram calls histogram from RGB color space. A color histogram denotes the joint probabilities of the intensities of the three color channels. The color histogram is defined as follows:

\[
h_{\text{RGB}}[r,g,b] = N \cdot \text{Prob}[R = r,G = g,B = b] \quad (1)
\]

Where R, G and B are the three color channels and N is the number of pixels in the image [6].

The color histogram is computed by counting the number of pixels of each color within the image. Since the number of colors is finite, it is usually more convenient to transform the three channel histogram into a single variable histogram. Given an RGB image,
one transform is given by $m = r + N_r g + N_g b$, where $N_r$, $N_g$, and $N_b$ are the number of bins for colors red, blue and green, respectively. For example, figure 1 shows RGB color histogram of a aircraft image with quantization rate of 5 bits.

This color histogram has a defect that small changes of color in the image may cause a great modification of the color histogram, and also color histogram may vary greatly by changes in lighting and large view angle.

An effective way to retain light-independent color properties is to use only hue and saturation in HSI color space, and some kinds of improved methods to extract color features from the histogram of Hue component have studied[7,8]. But because an additional mathematical conversion processing from RGB space to HSI space is requested to use the Hue histogram, it is need to use directly the RGB color space without a color space translation to other color spaces for fast retrieval from the image database.

Histogram of the hue component is defined as follows:

$$h_{H}[h] = \sum \Pr(o_{h, b}) \text{ for } h $$

where H is the hue component and N is the number of pixels in the image and Hue component has one value between 0 and 1. Figure 2 shows the histogram of the hue component of the aircraft image in figure 1.

Because these traditional color histograms relate in frequency of pixels that have equal color, color histogram changes greatly to slight change of image contents. But, to slight change of image contents, if there is identical color continuously, feature of color may not change. In our study, we thought that color existence of color might be a very important factor in field of content-based image retrieval and so we propose new color feature representation methods emphasizing in existence of such color.

3. Color Existence Representation Methods

3.1 Color Position Feature (CPF)

Because Color histogram represents size of distribution area of any color, although two images include same colors they are classified into different images each other. So we need an improved property to overcome this defect. In general image retrieval system have two continuous steps, first is selection of candidate images from color feature and second is choice of similar images from the candidates using other feature, texture or shape, except color feature. Therefore during the first step of computation frequencies of any color in image if we store the existence property of the corresponding color and use this information we could earn an improved performance in image retrieval. This color existence property is called color position feature (CPF).

Equation 3 shows that if the color exists in image the corresponding position feature value is assigned to 1, otherwise the position feature value is assigned to 0.

$$h_{CPF}(p) = \begin{cases} 
1, & \text{at } h_{RGB}[r, g, b] \\
0, & \text{at other bins}
\end{cases}$$

Figure 3 shows properties of color position feature detected from the RGB color histogram of figure 1.
3.2 2-dimensional Projection Maps (2DPM)

A new method that is proposed to represent color features of a color image efficiently is changing 3-dimensional RGB color space map to three 2-dimensional projection maps. Figure 4 shows 3-dimensional RGB color space map representing the three brightness values in R, G, and B channels of the aircraft image.

![3-dimensional RGB color space map of the aircraft image in Fig. 1](image)

Fig. 4 3-dimensional RGB color space map of the aircraft image in Fig. 1

Projecting the 3-dimensional color space map to 2-dimensional RG, GB, and BR surfaces make three 2-dimensional projection maps. First, to generate the projection map to RG surface, correspond the integer value of the R channel brightness to the x coordinate bin and the integer value of the G channel brightness to the y coordinate bin by one to one for one pixel in the image, then the position (x, y) in the 2-dimensional map has value of 1, and then the same procedure for the all pixels can make a 2-dimensional projection map. Through equal process, other projection maps to GB surface and BR surface, $P_{RG}$, $P_{GB}$, and $P_{BR}$ are formed as follows:

$$
P_{RG}(x,y) = \begin{cases} 
1, & \text{at } x = v(r), y = v(g) \\
0, & \text{at other bins}
\end{cases} 
$$

$$
P_{GB}(x,y) = \begin{cases} 
1, & \text{at } x = v(g), y = v(b) \\
0, & \text{at other bins}
\end{cases} 
$$

$$
P_{BR}(x,y) = \begin{cases} 
1, & \text{at } x = v(b), y = v(r) \\
0, & \text{at other bins}
\end{cases} 
$$

Where, $v(r)$, $v(g)$, and $v(b)$ are brightness values of one pixel in the image. Three 2-dimensional projection maps of the sunset image are presented in Figure 5.

![Three 2-dimensional projection maps](image)

Fig. 5 Three 2-dimensional projection maps.

In the content-based image retrieval, in order to measure similarities between the query image and the object images in the database, these three 2-dimensional projection maps are connected and transformed to a 1-dimensional projection vector as follows:

$$
\{P_{RG}(1,1), P_{RG}(2^q,1^2), P_{RG}(1^2,1), P_{RG}(1,2^q), P_{RG}(2^q,2^q), P_{RG}(2^q,1^2), P_{RG}(1^2,2^q)\}
$$

(5)

Where, I is 1-dimensional projection vector, $q$ is the number of quantization bits per pixel and $q$ is 5 here.

![A 1-dimensional projection vector of three 2-dimensional projection maps](image)

Fig. 6 A 1-dimensional projection vector of three 2-dimensional projection maps

Therefore, length of the 1-dimensional projection vector is $2^q \times 2^q \times 3$. Because length of the color histogram is $2^4 \times 2^4 \times 2^9$, length rate between color histogram and 2-dimension projection map is $\frac{2^4}{3}$, and in case of $q=5$, the length rate is about 10.7. Therefore, we can find that the processing speed reduces greatly and will appear for the performance through following experiments.

4. Experimental Results

The proposed method of color feature representation was compared with the traditional color histograms. Their performance to retrieve similar images from a heterogeneous collection of images has been used as comparison criterions. For the color
histograms, because values that compose the color histogram are analogous, the Euclidean distance method is used to measure similarities between a query image and candidate images.

For the RGB color histogram and the histogram of Hue component, when the feature vector of the query image is $C_p$ and one of the candidate image is $C_c$, the distance between these two images, $E_{hist}$, is computed as follows.

$$E_{hist} = \sqrt{\sum_{i=1}^{2^3+2^2+2^2}(C_p(i)-C_c(i))^2}$$

For the color position feature and the 2-dimensional projection maps, because its 1-dimensional projection vector is a bit pattern of binary value, the Hamming distance was used to measure similarities between a query image and candidate images, when the feature vector of the query image is $I_p$ and one of the candidate image is $I_c$, the Hamming distance between these two images, $E_{2d}$, is as follows.

$$E_{2d} = \sum_{i=1}^{2^3+2^2+2^2}|I_p(i)-I_c(i)|$$

To estimate their performance, 3 aircraft query images are selected from database that stores 100 aircraft images and the selected images are represented in figure 7.

![Fig. 7 5 query images for retrieval experiments](image)

When these images were used as query images in turn, the experimental results were compared each other. First, when the first image of the selected sunset images was presented as a query image, figure 8 shows the results of, (a) the RGB color histogram, (b) the Hue histogram, (c) the color position feature, and (d) the proposed 2-dimensional projection maps.

![Fig. 8 Results of retrieval for second aircraft image as a query](image)

Therefore, from the results of above experiments, when the 2d projection map that was proposed for representing color features of an color image was used in image retrieval, we found that the proposed has better performances than other two methods. We also compared their retrieval performances for flower images and animal images and same results were obtained. As figure 9 and figure 10, we could know that the proposed 2d projection map could express better color characteristic of image.

![Fig. 9 Results of retrieval by (a) the color histogram, (b) the Hue histogram, (c) the color position feature, and (d) 2-dimensional projection maps for flower images](image)
5. Conclusions and Future Works

To improve color image retrieval performance using color existence properties, the color position feature and the 2-dimensional projection maps method have been proposed. For the given heterogeneous image collections, some experimental results show that the proposed methods have better performance than the color histogram and the Hue histogram in image retrieval.

Hereafter, a study for comparison between the proposed method and the traditional color feature representation methods would be done by more precise methods in image retrieval fields.

6. References