Available online at www.sciencedirect.com
ScienceDirect
Information Processing and Management 43 (2007) 461–472
INFORMATION PROCESSING & MANAGEMENT
www.elsevier.com/locate/infoproman

Color image retrieval technique based on color features and image bitmap

Tzu-Chuen Lu [a,*], Chin-Chen Chang [b]

[a] Department of Information Management, Chaoyang University of Technology, Taichung 41349, Taiwan, ROC
[b] Department of Information Engineering and Computer Science, Feng Chia University, Taichung 40724, Taiwan, ROC

Received 19 May 2006; accepted 25 July 2006
Available online 11 October 2006

Abstract

The field of color image retrieval has been an important research area for several decades. For the purpose of effectively retrieving more similar images from the digital image databases, this paper uses the color distributions, the mean value and the standard deviation, to represent the global characteristics of the image. Moreover, the image bitmap is used to represent the local characteristics of the image for increasing the accuracy of the retrieval system. As the experimental results indicated, the proposed technique indeed outperforms other schemes in terms of retrieval accuracy and category retrieval ability. Furthermore, the total memory space for saving the image features of the proposed method is less than Chan and Liu's method.

Keywords: Image bitmap; Color distribution; Content-based image retrieval; Standard deviation

1. Introduction

Image retrieval has been a very active research topic since the 1970s. Most convenient image retrieval schemes are annotated-based that annotate each image in an image database by using keywords for similar image retrieval. However, the annotated-based retrieval method has two major problems that make it impractical. The first problem is that the size of the image database has become larger and larger such that to create keywords for each image is time-consuming. The second problem is that different people may give the same image different keywords. Afterwards, the content-based image retrieval (CBIR) techniques were proposed to solve the problems of the annotated-based image retrieval methods. In a CBIR system, images are automatically indexed by summarizing their visual contents through automatically extracted primitive features, such as shape, texture, color, size, and so on.

Many researchers used color feature to depict image contents for region matching, semantic categorization, and similarity searches (Flickner et al., 1995; Schettini, Ciocca, & Zuffi, 2001). However, the

* Corresponding author.
E-mail addresses: tclu@cyut.edu.tw (T.-C. Lu), ccc@cs.ccu.edu.tw (C.-C. Chang).

0306-4573/$ - see front matter © 2006 Published by Elsevier Ltd.
doi:10.1016/j.ipm.2006.07.014

performance is poor for a CBIR system, which only uses a color feature to search for similar images from a huge database. Thus, some researchers analyzed the color distribution of the image to increase the retrieval accuracy (Brunelli & Mich, 2001; Du & Wang, 2001; Fuh, Cho, & Essig, 2000; Hsieh, Grimson, Chiang, & Huang, 2000; Kankanhalli, Mehtre, & Huang, 2001; Schettini et al., 2001; Stehling, Nascimento, & Falcao, 2001; Wang & Du, 2001). For example, in 1996, Gong et al. proposed an image indexing and retrieval scheme. In their scheme, an image is split into nine equal sub-areas. They presented each sub-area by using a color histogram to model the color spatial information (Gong, Chuan, & Xiaoyi, 1996). In 1997, Stricker and Dimai split an image into an oval central region and four corner sub-regions for image indexing (Stricker & Dimai, 1997). Gagliardi and Schettini described the image in the CIELAB color space with two palettes and integrated different color information descriptions and similarity measurements to enhance the effectiveness of a CBIR system in Gagliardi and Schettini (1997). Kou used the mean value, the standard deviation, and the skewness of pixels from each bin in a color histogram as the image features to search similar images (Kou, 2001).

In 2003, Chan and Liu proposed a CBIR system based on color differences on edges in spiral scan order (Chan & Liu, 2003). With a view to increasing the retrieval accuracy, Chan and Liu combined the color feature with color differences among adjacent pixels for image retrieval. The color differences can be viewed as the local features of an image.

In this paper, we adopt another local feature, image bitmap, to investigate similar image from the huge database. The image bitmap is obtained from an image compressed with the block truncation coding method. The block truncation coding method uses the properties, the mean value, the standard deviation and the image bitmap, of an image for image compression. In this paper, the retrieval technique is based on these properties instead of numerous image features.

2. Related works

Swain and Ballard (1991) proposed an image indexing scheme based on color histogram in 1991. In their scheme, a color image is transformed into the gray-scale image. Then, they generated a histogram with 256 bins to record the total number of pixel values in the gray-scale image as image feature to search the similar images from the database. In order to reduce the memory space for storing the histogram, they reduce the number of bins in the histogram from 256 to 64. In addition, the color distributions of each bin were used to enhance the retrieve accuracy of the histogram-based retrieval scheme. After that, Stricker and Orengo (1995) used three statistic moments, the average, the variance, and the skewness to represent the image. In their scheme, a color image is transformed to three spectrums, H, S and V. The scheme, then, calculated the three statistic moments for each spectrum. The nine moments obtained from the three spectrums are the features of the image.

In Chan and Liu (2003) and Chang and Chan (2000) used color differences on edges in spiral scan order feature to depict the color, color complexity, and color differences among neighboring pixels in a color image. In their schemes, all pixels of the images in the database were divided into 64 clusters by using K-mean algorithm. Then, they classified each pixel of the image into the most similar cluster. After classification, the image was scanned in the spiral order to compute the difference between any two neighboring pixels. The scan order is shown in Fig. 1.

In their scheme, each pixel in the classified image was represented by a vector in three primary color spaces, red (R), green (G), and blue (B). Let $P_i = \begin{bmatrix} R_i \\ G_i \\ B_i \end{bmatrix}$ be the ith pixel of the classified image. The scanning process begins from the central pixel of the image in a spiral direction pixel by pixel. If the pixel P_i is different from the next pixel P_{i+1}, then the scheme computes the color difference between two pixels by using $d(P_i, P_{i+1}) = [(R_i - R_{i+1})^2 + (G_i - G_{i+1})^2 + (B_i - B_{i+1})^2]^{1/2}$.

Each cluster has its own bin to record the difference. The difference $d(P_i, P_{i+1})$ is added to the bin of the cluster of P_i. The final values of the bins are the features of the image. For example, assuming that the first