Retrieval by Color Features in Image Databases

Prof. Ms. V.M.Deshmukh, Sohel A. Bhura
Pravin R. SatavSuresh, S. Asole, Sachin Agrawal
Department of Computer Science and Engineering PRMIT &R Badnera, Amravati (M.S)
Lecturer, Department of Information Technology BNCE Pusad (MS)
Lecturer, Department of Information Technology BNCE Pusad (MS)
Lecturer, Department of Computer Science and Engineering COET Akola (MS)

ABSTRACT

Content-based search and retrieval techniques for multimedia content utilize various descriptors for different image features. However, most content-based image retrieval systems use low-level visual features for representation and retrieval of images. The **color** content, being one of the easily computable features is often used in content-based search for retrieving images in large image databases. The paper presents a simple and fast technique for image indexing and retrieval based on a description of the spatial distribution of dominant colors. The idea in this study is to empirically evaluate the effect of color features to the query. Furthermore, some similarity measures are defined that allow the definition of effective visual queries for fast interactive image database browsing applications.

We use the image histograms to characterize the global and local color properties of an image and to preserve its intrinsic geometric information.

The color feature is probably the most visible visual feature for the most humans. This is the new approach in the image organization and retrieval of image, in image databases. Our approach combines the representation of the color as the global image feature and the color as the representation of the spatial information in the images.

Categories and Subject Descriptors

Image Processing

1.INTRODUCTION:

With the exponential growth of computer networks (the Internet) and of the relevant services, multimedia contents are becoming more and more important. New image and video data are continuously generated at an increasing rate from a wide range of different sources (tele-conferencing, biomedicine, distributed databases, etc). All of these data require to be frequently accessed for the purpose of storage, processing and transmission over heterogeneous platforms and networks. To avoid the expense and limitations of text, there is considerable interest to develop content-based image retrieval (CBIR) techniques based on automatically extractable visual attributes such as colors or textures. A CBIR system facilitates the formulation of visual queries by novice users in which a query can be initiated simply by supplying a sample image. *"The system extracts the visual features from the query image and matches it against those stored in the database."* The result is a set of images that are similar to the query image rather than an exact match. Color is one of the most used image features.

Defining color similarity criteria, which corresponds to human visual perception, is an unsolved problem. In addition, individual differences in perception of images make the problem even more complicated. This evidence has motivated us for investigating new content based retrieval system. Because of diversity and complexity of the photograph images, it is difficult to capture their visual properties such as shape, structure and complexity.

In this paper we describes our goal is creation of an advanced query capability that supports query by image example. The process of content-based image retrieval can be divided into the following sub processes

1.1   Feature extraction

Which, involves the creation of good mechanisms to code global and local lower-level features such as color etc., which are semantically related to the contents of images.

1.2   Defining Similarity

Color features are among the most important and extensively used low-level features in image database retrieval. They are usually robust in noise, resolution, and orientation and resizing. Due to their little semantic meaning and its compact representation, color features tend to be more domain independent compared to other features. Combined with image segmentation, color features can be used to describe the appearance of the image and even generate semantic annotations.

• *Image Data Extraction:*

The local features are calculated over parts of an image – single pixel or segment. Since an image may have several pixels or segments, these features usually result in feature value sets for the whole image. This complicates calculations but it has additional information and can produce better retrieval results. We chose to use very simple segmentation approach which divides the whole image into equal non-overlapping blocks and it represents the information about the spatial relations between the objects in the image. Every image I at pixel level (physical image) is represented in the image database with two type color descriptors and its name and location as:

$I = \{\{c_i, p_i\}, I = 1,..,M\}, \{c_i, j\}, i = 1.N, j = 1 \ldots N\}, Path\}$

This symbolic description of image I is stored in an Image Database. It has been shown that in the early perception stage human visual system performs identification of dominant colors by eliminating fine details and averaging colors with small areas. Consequently, on the global level, humans perceive images only as a combination of few most prominent colors, even though the color histogram of the observed

image may consist of significantly more colors. Because of this we reduce the number of colors to 10 basic colors from our color code book.

We chose to use the colors as a characteristic of the photograph images content. The system extracts and describes the color content of the images automatically by two structures. **First** is based on the dominant colors in every image. **Second** – represents local color distribution. These two structures are stored in the feature database with the path to the image file.

2. Image Description and Approximations

In our system we use a color code book, created on the basis of vector quantization in RGB color space with 9 colors (white-1, gray-2, black-3, red-4, brown-5, yellow-6, green-7, blue-8 and purple-9) and unrecognized colour-0. The goal of color feature extraction is to obtain compact, perceptually relevant representation of the color content of image retrieval.

2.1 Color Descriptor

Color has been widely used in content-based image retrieval applications. In such applications the color properties of an image are usually characterized by the probability distribution of the colors in the image. A distance measure is then used to measure the (dis-)similarity between images based on the descriptions of their color distributions in order to quickly find relevant images. The linear representation of color distributions is then used to derive new compact descriptors for color-based image retrieval.

Dominant colors are used to characterize the color content of an image. A color clustering is performed in order to obtain its representative dominant colors and its corresponding percentage. Each representative color and its corresponding percentage form a pair of attributes that describe the color characteristics in an image region. The dominant color histogram feature descriptor is defined to be a set of such attribute pairs

$$Sim\ (I, Q) = \left| \frac{I \cap Q}{I \cup Q} \right|$$

CD={{ci,pi}for i=1toM, *M dominant colors***)**

Where *M* is the total number of color clusters in the image, is an Arabic number corresponding to the color, *pi* is its percentage (Σpi =1).

First, the most dominant color in the image is added and then the less dominant color follows. This means that the color covering the largest area of an image is considered as the main color feature and its pair is the first one in the color descriptor. The second largest area of the same image defines the second one and the rest of pairs follow. (*Note that the colors that are not seen in the image will not be added in the descriptor.*) Some images may not contain all colors, so some color descriptors have less than 10 elements. Therefore, the size of the color descriptor can be at most ten; however, it must include at least one pair. For example, a color descriptor can be

"{5,41}{1,20}{2,12}{0,6.5}{8,5.9}{6,5}{7,4.5}{3,3.5}{4,1.1}{9,0.5}", OR

"{7,53.5}{8,21.7}{1,15.3}{2, 6.5}{5,3}".

There is only one descriptor per image

2.2 Color Descriptor Matrix

Color description matrix is used to represent the **spatial** color information in images. The image is divided into blocks MxM (where M = 4, 8, 16 and 32). In matrix are stored the dominant colors for the relevant image blocks This matrix stores the dominant colors for the image blocks. This matrix stores the coefficient of the dominant color from the selected color codebook in the corresponding part of the image.

CDM={ci,j},i=1.N,j=1…N}

This method for color feature representation can be considered as a quantized version of the color histogram and its scheme has several **advantages**. **First,** it best represents the original color content in terms of minimizing the mean square error using a small number of colors. **Second**, this color feature is very compact.

3. Defining Similarity:

Algorithms of similarity measures are used to perform the degree of similarity between the Query Image and the Images from database, which provide a ranking to arrange the order of images. Image retrieval systems usually support several types of queries – query by global features, query by image example or query by user sketch.

This formula returns a value of [0,1]. We expect the value of 1 when the query image and the database image are identical. The representation of **Jaccard's coefficient** as a decimal fraction, such as "0.75", may not be easy to understand for general users. Therefore we use a common way to illustrate the semantic of the image similarity – the representation of percentage score, such as "75%".

4. Conclusion

Image is analyzed pixel by pixel and feature extraction using color and inserting it into the database. the Color Descriptor may be a problem since those images "visually similar to each other" may have close but different Color Descriptors with different orders of colors, due to the fact that the histograms of those images are only slightly different.

If we select the same image is as a query image then it retrieve similar image from the database with 100% similarity.

5. REFERENCES:

[1] A. Goodrum, "Image Information Retrieval: An Overview of Current Research," *Special Issue on Information Science Research*, vol. 3, no. 2, 2000.

[2] Mojsilovic, J. Kovacevic, J. Hu, R. J. Safranek, K. Ganapathy, "Matching and retrieval based on the vocabulary and grammar of color patterns," *IEEE Trans. on Image Proc.*, vol. 9, no. 1, pp. 38-54, January 2000.

[3] B. Berlin and P. Kay, Basic Color Terms: Their Universality and Evolution, Berkeley:University of California, 1969.5. Boris Rachev, Irena Valova, Silyan Arsov "An Approach For Image Organization and Retrieval In Realistic Image Databases," *7th GIS Workshop Potsdam*, June 2001, Germany