International Journal of Information Technology and Knowledge Management
July-December 2011, Volume 4, No. 2, pp. 585-589

# DESIGN & DEVELOPMENT OF COLOR MATCHING ALGORITHM FOR IMAGE RETRIEVAL USING HISTOGRAM AND SEGMENTATION TECHNIQUES

Krishna Kumar Pandey[1] & Nishchol Mishra[2]

Color based image retrieval is an important research area in digital image processing and Color is the useful property of the image. The challenging task is to find an efficient and accurate color images similar value and some limitations between two different types of image are matching such as compressed color image and RGB color image, RGB image and HSV color image, color image and 90 degree rotate color image but this matching technique remove these problems. This paper Proposed two method first, a new matching technique to find the similar value between query color image and database color image using histogram, spatiogram and bins and second, developed a new segmentation based matching technique using threshold selection to matching between two color image based on histogram database. This method uses RGB and HSV color space. This is a matching technique based on histogram and spatiogram, spatiogram is a generalization of histogram.

Keywords: Color Based Image Retrieval, Color Space, Histogram, Spatiogram, Segmentation.

## 1. Introduction

The color image processing techniques applicable to color images, they are far from being exhaustive, color images are handled for a variety of image processing tasks. The color image processing subdivide into three principal areas: (1) color transformations also called color mapping (2) spatial processing of individual color planes and (3) color vectors processing. The first category deals with processing the pixels of each color plane based strictly on their values and not on their spatial coordinates. This category is analogous to the material in dealing with intensity transformations. The second category deals with spatial (neighborhood) filtering of individual color planes and is analogous to the spatial filtering. The third category deals with techniques based on processing all components of a color image simultaneously. Because full-color images have at least three components, color pixels can be treated as vectors. For example, in the RGB system, each color point can be interpreted as a vector extending from the origin to that point in the RGB coordinate system[1].

Color representation is based on the classical theory of Thomas Young (1802). The study of color is important in the design and development of color vision systems. Use of color in image displays is not only more pleasing, but it also enables us to receive more visual information. While we can perceive only a few dozen gray levels, we have the ability to distinguish between thousands of colors. The perceptual attributes of color are brightness, hue, and saturation. Brightness represents the perceived luminance. The hue of a color refers to its "redness", "greenness", and so on. For monochromatic light sources, differences in hues are manifested by the differences in wavelengths. Saturation is that aspect of perception that varies most strongly as more and more while light is added to a monochromatic light. These definitions are somewhat imprecise because hue, saturation, and brightness all change when the wavelength, the intensity, the hue, or the amount of white light in a color is changed [9].

Image retrieval is the process of browsing, searching and retrieving images from a large database of digital images. The collection of images in the web are growing larger and becoming more diverse. Retrieving images from such large collections is a challenging problem. One of the main problems they highlighted was the difficulty of locating a desired image in a large and varied collection. While it is perfectly possible to identify a desired image from a small collection simply by browsing, more effective techniques are needed with collections containing thousands of items. To search for images, a user may provide query terms such as keyword, image file/link, or click on some image, and the system will return images "similar" to the query. The similarity used for search criteria could be Meta tags, color distribution in images, region/shape attributes, etc. Unfortunately, image retrieval systems have not kept pace with the collections they are searching. The shortcomings of these systems are due both to the image representations they use and to their methods of accessing those representations to find images. The problems of image retrieval are becoming widely recognized, and the search for solutions an increasingly active area for research and

[1,2]School of IT, Rajiv Gandhi Prodyogiki Vishwavidyalaya, Bhopal (M.P.)
E-mail:[1]krishna24it@gmail.com, [2]nishchol@rgtu.net

development. In recent years, with large scale storing of images the need to have an efficient method of image searching and retrieval has increased. It can simplify many tasks in many application areas such as biomedicine, forensics, artificial intelligence, military, education, web image searching. Most of the image retrieval systems present today are text-based, in which images are manually annotated by text-based keywords and when we query by a keyword, instead of looking into the contents of the image, this system matches the query to the keywords present in the database. This technique has its some disadvantages: (a) Firstly, considering the huge collection of images present, it's not feasible to manually annotate them (b) Secondly; the rich features present in an image cannot be described by keywords completely [5].

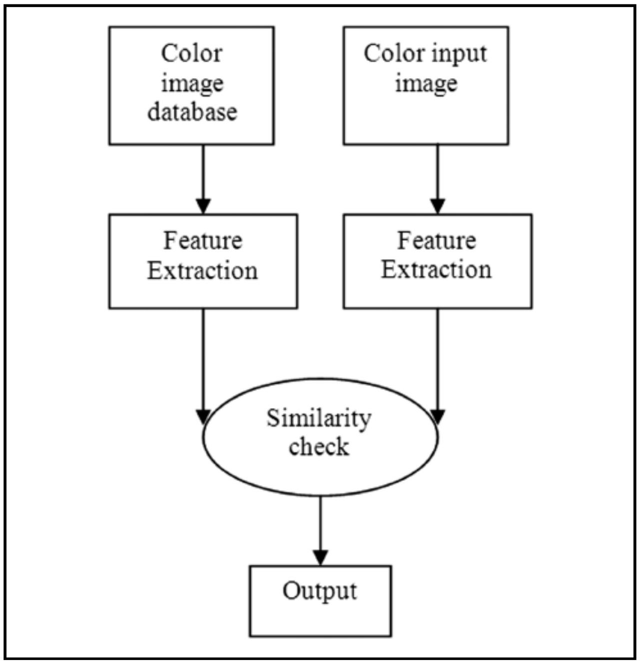


Fig.1: Block Diagram of Color Based Image Matching

### 1.1 RGB Color Model

The RGB color model has three basic primary colors: red, green and blue. All other colors are obtained by combining them. This model can be thought as a cube, where three non – adjacent and perpendicular corners are R, G and B. As can be seen, RGB is an additive color model, since the combination of green, red and blue gives white. This is the color model that is most commonly used in computer graphics, since it matches the way the color is stored in video memory. Used for color model and broad class of color of color models corners, RGB color are at three points at corners and at basic scale. Magenta, yellow and cyan are at three remaining points at corner. Black is situated at origin and white is situated at the farest point. The line joining black and white shows gray scale. All other colors are inside the cube in the form of points and are defined by vectors extending from origin. The number of pixels used to represent a pixel is called pixel depth. In general application, we require 8 bits for presenting a color, for RGB, we needs = 8*3 = 24 bits, so total number of colors = 2^24 = 16,777, 216. But this large number of colors are practically not usable that why we use only 256 colors. The set of colors that can be used instead of whole colors is called subset of colors. This is also called safe RGB colors.

### 1.2 HSV Color Model

The HSV stands for the Hue, Saturation, and Value. We treat the hue-saturation-value (HSV) space as a cone: for a given point (HSV), h and sv are the angular and radial coordinates of the point on a disk of radius vat height v; all coordinates range from 0 to 1 [2]. HIS model decouples the color and gray levels. It is the method that will provide information about image than is easily described and interpreted. The HIS color model was designed having in mind the way graphic designers and artists think of colors. Artists use terms like saturation (the pureness of a color), hue (the color in itself) and intensity (the brightness of the color). This is exactly what the HSI color model represents. The color space is strange, since it is not orthogonal, in this color space, like in the others, a color is a vector. H(hue) is the angle of the vector over the basic triangle, starting from red (0 degree). S(saturation) is the proportional size of the module of the projection of the vector over the basic triangle, and I(intensity), is the distance from the end of the vector to the basic triangle. There is a conversion from RGB to HIS is quite complicated.

## 2. COLOR IMAGE SEGMENTATION

The human eyes have adjustability for the brightness, which we can only identified dozens of gray-scale at any point of complex image, but can identify thousands of colors. In many cases, only utilize gray-Level information can not extract the target from background; we must by means of color information. Accordingly, with the rapidly improvement of computer processing capabilities, the color image processing is being more and more concerned by people. The color image segmentation is also widely used in many multimedia applications, for example; in order to effectively scan large numbers of images and video data in digital libraries, they all need to be compiled directory, sorting and storage, the color and texture are two most important features of information retrieval based on its content in the images and video. Therefore, the color and texture segmentation often used for indexing and management of data; another example of multimedia applications is the dissemination of information in the network. Today, a large number of multimedia data streams sent on the Internet, However, due to the bandwidth limitations; we need to compress the data, and therefore it calls for image and video segmentation [3]. Commonly used for color image segmentation methods are histogram threshold, feature space clustering, region-based approach, based on edge detection methods, fuzzy methods, artificial neural network approach, based on physical model methods, etc.

## 3. SIMILARITY MEASURES

Finding good similarity measures between images based on some feature set is a challenging task. On the one hand, the ultimate goal is to define similarity functions that match with human perception, but how humans judge the similarity between images is a topic of ongoing research. The Direct Euclidian Distance between an database image D and query image Q can be given as the equation below [7]. Where, a and b be the feature vectors of database image D and Query image Q respectively with size 'n'. Histogram similarity measures namely, Histogram Intersection (HI), Histogram Euclidean Distance (HED) and Histogram Quadratic Distance Measures (HQDM) [8].

## 4. PROPOSED WORK

Proposed a new matching technique for color based image retrieval using histogram, spatiogram and bins firstly, select a RGB color space then converted its color space into HSV color space or gray level, in the case of HSV color space the results are more effective and accurate in comparison to the RGB color space, but the gray level similarity matching between two image the results are improve compare to the color space, developed a new formula and matching two color images. Proposed a another matching technique based on segmentation to find similar value between two color image, developed a new segmentation technique using threshold selection, in this method using mathematical formula and apply to two color images. The following technique used in this proposed works.

### 4.1 Color Histogram

The color histogram for an image is constructed by counting the number of pixels of each color [6]. In other words color histogram defines as a distribution of colors in an image. For digital images, a color histogram represents the number of pixels that have colors in each of a fixed list of color ranges. Those span the image color space the set of all possible colors. A visual representation of the histogram of an image is a simple but useful tool because it describes the image in terms of brightness and contrast. To find the color image histogram

$$S = \text{sum}\,[\text{sum}\,\{\text{sum}\,(\sqrt{h1} * \sqrt{h2})\}]$$

Where s is similarity value between two color image histogram, h1 is the query image histogram and h2 is the database image histogram.

### 4.2 Color Spatiogram

We introduce the concept of a spatiogram, which is a generalization of a histogram that includes potentially higher order moments. A histogram is a zeroth-order spatiogram, while second-order spatiograms contain spatial means and covariance's for each histogram bin. This spatial information still allows quite general transformations, as in a histogram, but captures a richer description of the target to increase robustness in tracking. We show how to use spatiograms in kernel-based trackers, deriving a mean shift procedure in which individual pixels vote not only for the amount of shift but also for its direction. Experiments show improved tracking results compared with histograms, using both mean shift and exhaustive local search [4].

$$C = 2 * \sqrt{(2 * pi)}$$

$$C2 = 1 / (2 * pi)$$

$$q = sigma1 + sigma2$$

$$Dist = (q * Q * z)$$

$$S = \text{sum}\{\text{sum}\,(\sqrt{h1} * \sqrt{h2} * dist.)\}$$

#### Algorithm for New Color Matching Algorithm Using Histogram

(1) Read the database image d(I) and query image q(I) and both image are RGB color images.

Where d(I) and q(I) are variables

(2) Convert d(I) and q(I) RGB image into HSV image.

RGB query image [q(I)] = HSV query image [q'(I)]

RGB database image [d(I)] = HSV database image [d'(I)]

(3) Extract a color histogram from each image h1 and h2.

Bins = {4, 8, 12, 32 …}

(4) Compare their histogram, similarity (h1, h2).

$$S = \text{sum}\,\{\text{sum}\,(\sqrt{h1} * \sqrt{h2} * dist.)\}$$

If result = 0, very low similarity

Result = 0.9, good similarity

Result = 1, perfect similarity.

(5) Start the process of matching.

(6) Extract a color Spatiogram from each image s1 and s2.

Bins = {4, 8, 12, 32 …}

(7) Compare their spatiogram, similarity (s1, s2).

$$S = \text{sum}\,\{\text{sum}\,(\sqrt{h1} * \sqrt{h2} * dist.)\}$$

If result = 0, very low similarity

Result = 0.9, good similarity

Result = 1, perfect similarity.

(a) (b) (c) (d)

Fig. 2: Images

Proposed Method I

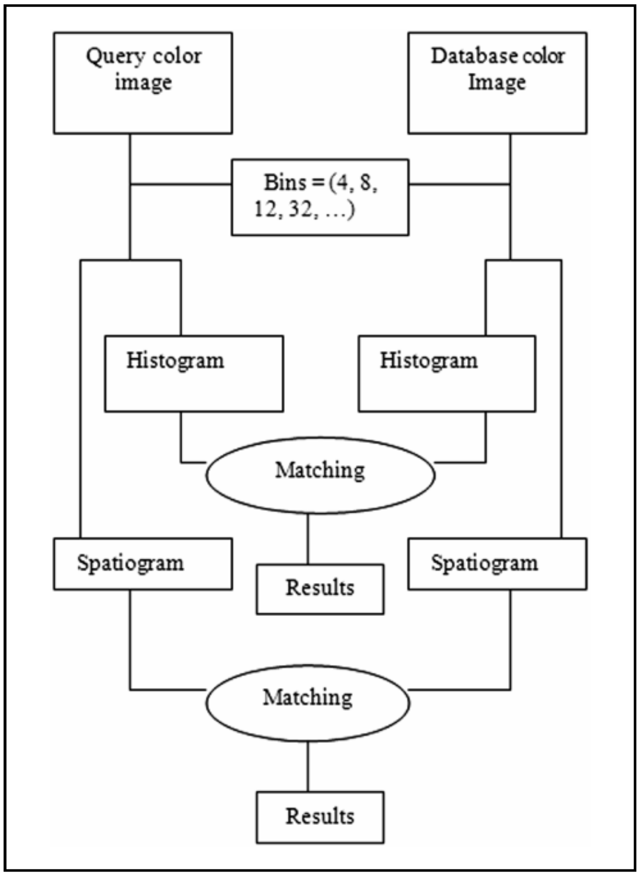


Table 1
Matching Results Using Histogram and Spatiogram.

| Image | Hbin | Sbin | Histogram | Spatiogram |
|---|---|---|---|---|
| (d, a) | 4 | 4 | 0.869045 | 0.791551 |
| (d, a) | 8 | 8 | 0.796508 | 0.675316 |
| (d, a) | 16 | 16 | 0.701256 | 0.532764 |
| (d, a) | 32 | 32 | 0.623712 | 0.421750 |
| (d, b) | 4 | 4 | 0.593771 | 0.514547 |
| (d, b) | 8 | 8 | 0.418037 | 0.290486 |
| (d, b) | 16 | 16 | 0.266407 | 0.145882 |
| (d, b) | 32 | 32 | 0.180057 | 0.067183 |
| (d, c) | 4 | 4 | 0.662087 | 0.602848 |
| (d, c) | 8 | 8 | 0.599680 | 0.525683 |
| (d, c) | 16 | 16 | 0.559177 | 0.465471 |
| (d, c) | 32 | 32 | 0.518520 | 0.410768 |

## Segmentation Based Matching Algorithm

(1) Read the input color image i.
(2) Convert RGB image into gray scale.
(3) Threshold selection.
(4) Repeat the step (1) to (3) for database image j.
(5) Similarity = histogram (I, j)/numel (I, j).
(6) Results.

Fig. 3: Images

Proposed Method 2

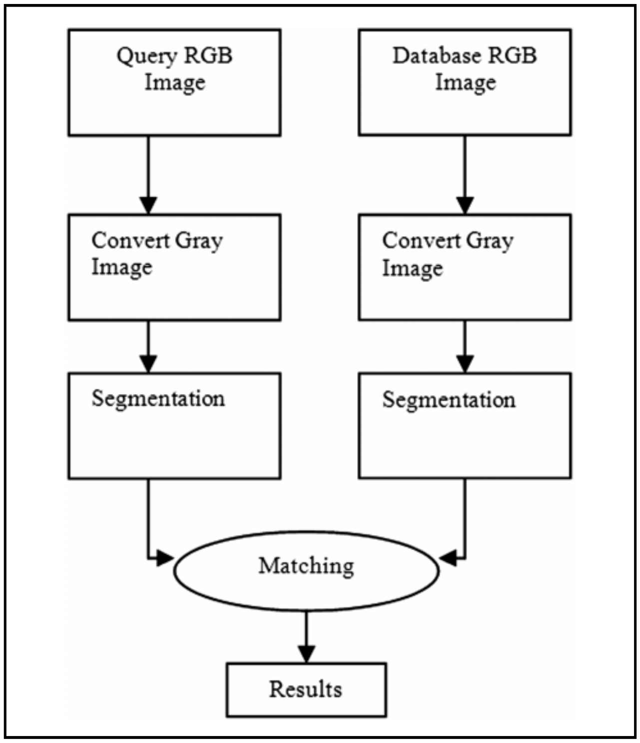


Table 2
Matching Results Using Segmentation.

| Image | Formula 1 | Formula 2 |
|---|---|---|
| (6, 1) | 0.0044 | 0.9988 |
| (6, 2) | 0.0179 | 0.9955 |
| (6, 3) | 0.0058 | 0.9985 |
| (6, 4) | 0.0292 | 0.9926 |
| (6, 5) | 0.0075 | 0.9980 |
| (6, 6) | 0.0000 | 1.0000 |

## 5. CONCLUSIONS

This paper have presented color image matching methods, which can be used to find the similar value between query image and database image based on histogram and spatiogram and the another proposed method are segmentation based using threshold selection, various color based image segmentation such as feature based color segmentation, image based color segmentation and physical based image segmentation, but this paper develop a new segmentation technique and find the similar value between query image and database image. The accuracy of any particular method in any given situation will depend on the histogram bin and segmentation methods. The main aim of this research is to find a similarity between two color images using histogram database. In this paper three condition are apply to find color similarity between two color images, first if results are 1, images color are perfect match, second, result are less than 1 and greater than 0.5 then color similarity are good and the third condition, results are less than 0.5 then color similarity are poor. This technique also finds the similarity between compressed color image, 90 degree rotate color image and RGB and HSV color image. The future work of this method is finding the similarity value between different format color images such as .jpeg, .png, .tif. This proposed method has simulated in MATLAB 7.5.

## REFERENCE

[1] Rafael C. Gonzalez and Richard E. Woods "Digital Image Processing Using MATLAB", 2010.

[2] Weisheng Li "New Color Cluster Algorithm For Image Retrieval", 2009.

[3] Jun Tang, "A Color Image Segmentation Algorithm Based on Region Growing", 2010.

[4] Yoichi MIYAKE and Kimiyoshi MIYATA "Color Image Processing Based on Spectral Information and Its Application", 1999.

[5] V.S.V.S. Murthy, E.Vamsidhar, J.N.V.R. Swarup Kumar "Content Based Image Retrieval Using Hierarchical and K-Means Clustering Techniques", 2010.

[6] Sangoh Jeong, "Histogram-Based Color Image Retrieval", 2001.

[7] N. K. Kamila "Image Retrieval Using Equalized Histogram Image Bins Moments", 2010.

[8] V. Vijaya Kumar "Integrated Histogram Bin Matching For Similarity Measures of Color Image Retrieval" 2009.

[9] A. K. Jain "Fundamentals of Digital Image Processing", 2006.