# Content Based Image Retrieval in Database of Segmented Images

**Szabolcs Sergyán**

John von Neumann Faculty of Informatics, Budapest Tech
Bécsi út 96/B, H-1034 Budapest, Hungary
sergyan.szabolcs@nik.bmf.hu

*Abstract: The great proportion of the content based image retrieval systems operates by color based searching, using global informations of the image. However between the color, the position of the homogeneous, nearly monochrome region, the segments has important information in an image. This paper introduces a color histogram based image searching system, which takes the positions of the segments into account. The algorithm which is presented in the paper uses an automated number of bins determination method for defining the segments.*

*Keywords: content based image retrieval, image segmentation*

## 1 Introduction

Image databases can be indexed by defining keywords or using the information of the pixels of image [7]. In the case of keyword indexing to introduce a new index needs enormous manual work, while in the case of content based indexing it can be solved effortlessly. It must have been one cause of the great evolution of the content based image retrieval systems.

A lot of realization of the content based retrieval systems applies the global color histogram of the images, nevertheless the color and the position of the nearly homogeneous color regions of the image have important information. I have realized an image searching system which takes the color and the position of the segments into account as well.

The Section 2 gives a review of the applied segmentation algorithm, and presents the automation of the number of bins which was used as a paramater of the segmentation. I have defined a distance measure between two images which contains segments. This measure takes the color histogram and position of the images into account.

Section 3 presents the way of the testing of the algorithm and the results, and in the Section 4 some inferences is drawn.

# 2 Project Description

In my project I achieved a content based image retrieval system, which at first divides images into regions with similar colors, and then takes the position and the color histogram of the segments into account to generate the indexes of the database. I made experiments to search similar images to a query image in a database of 1200 images.

I applied the algorithm introduced in Subsection 2.1 to define the homogeneous regions. To generate the parameters of this algorithm I used the method [6] which is presented in the Subsection 2.2. The distance of segmented images is defined in the Subsection 2.3.

## 2.1 Histogram Cluster Based Image Segmentation Algorithm

To search continuous, similar color regions of an image a modification of a clustering method [1] was applied which was published in [2]. The substance of this procedure is to make the two-variable histogram of the original RGB image above the chromaticity plane using some bins. To switch from the RGB space to the chromaticity plane is necessary to use the following equations:

$$ch_r = \frac{R}{R+G+B}, \tag{1}$$

$$ch_g = \frac{G}{R+G+B}, \tag{2}$$

where R, G and B means the red, green and blue intensity value of a given image, respectively. The $ch_r$ and $ch_g$ values fall into the $[0,1]$ interval, and in the case of $R = G = B = 0$, $ch_r$ and also $ch_g$ is considered as 0. The $[0,1]$ interval is divided into *k* pieces equal parts, where *k* is the so-called number of bins. In this way the unit square of the $(ch_r, ch_g)$ place can be divided into $k^2$ pieces coincident squares. Then the number of the orignal pixels which falls into a given bin of the chromaticity plane is determined and this number is stored. For example a division shown in Figure 1 can be obtained.

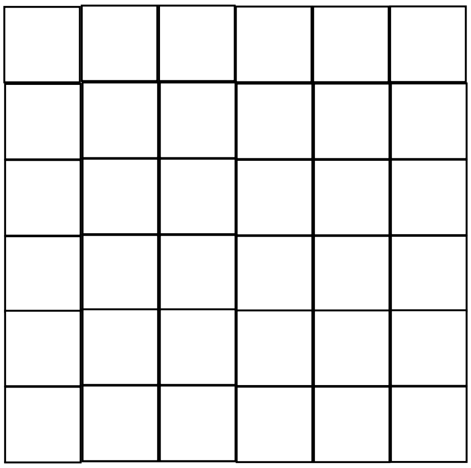

Figure 1

A possible two-variable histogram above the chromaticity plane with 6*6 bins

To clustering the generated histogram is required. In the applied clustering procedure the greatest values of the direct neighbours is assigned to each value, if this neighbour value is not greater than the actual value. If there is not any neighbour with greater value than the actual, then itself is assigned (see Figure 2). The assignments unambiguously determine the clusters. That values belong to the same cluster which has assignment of any direction. The obtained clusters can be seen in Figure 3.

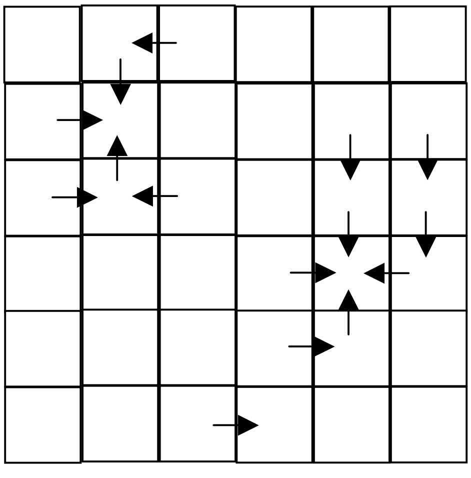

Figure 2

Assigning the histogram values

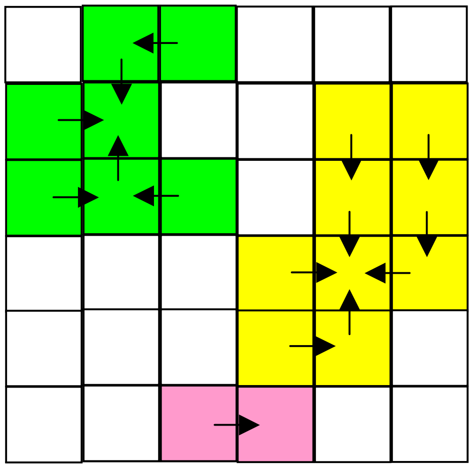

Figure 3
Histogram clusters

With this we have to assay where can be found pixels in the original image, which corresponding chromaticity values fall into the same histogram cluster.

## 2.2 Automated Determination of Number of Bins

We apply the previously introduced region detection algorithm to an *n*-element image database with , and bins. In case of all number of bins region sequences can be obtained, which are denoted by , and respectively. Only that regions are taken into account, which number of pixels is more than 5% of number of pixels of the whole image.

In similarly way we run the color region detection method with the normalized images of the database with , and bins. (We use the color cluster rotation for normalizing [4], which was proved useful earlier [5] too.) The obtained region sequences are denoted by , and respectively.

After that we make the grayscale segmentation of the grayscale image according to the original color image. The region sequence of it is denoted by .

Let

(3)

be the number of elements of the *j*-th color region sequence of the *i*-th image. Similarly let

(4)

be the number of elements of the grayscale region sequence of the *i*-th image. For the *k*-th element of the grayscale region sequence, where , we search the closest region of the region sequence, which is denoted by , where . In this case closest refers to the Euclidean distance of the centre of mass of regions. In this way for all region is assigned unambiguosly one region.

Let we define the distance between and the corresponding in the following way:

(5)

where means the simmetrical difference of two regions and refers to the number of pixels of a region.

Let denote the distance of the and the region sequences, which is defined in the following way:

, (6)

where *l* is the index of that color region which was assigned to the *k*-th grayscale unambiguosly.

Six distance can be generated in the case of the examined *i*-th image, inasmuch as color region sequences were made in six way, that is to say . Let we examine which distance is the smallest of the six distances, and let denote this distance:

. (7)

Let we search that *j* index, where

(8)

Based on our algorithm we can say the obtained procedure generates the best regions, so that number of bins is useful, which belongs to the *j*-th procedure. We can say that as well, applying normalization is necessary or not.

### 2.3 Similiraty Measure of Segmented Images

I defined the numerical similarity of image and on the following way.

Let denote the number of obtained segments in the image $f$, and similarly let be the number of obtained segments in the image $g$. That segment of the image $g$ is assigned to the $f_i$ segment of the image $f$, which centre of mass is the closest to the centre of mass of $f_i$. In this way one segment of can be assigned to each segment of , unconventionally not in injective way.

The distance of segments and which are assigned to each other can be defined in the following way:

, (9)

where means the euclidean distance of two vectors, and means the color histogram of a segment above the RGB place with bins [8].

The distance of the image and $g$:

, (10)

where is the index of the segment , which is assigned unambiguosly to the -th segment of .

## 3 Experiments

The testing of the algorithm which was presented in the previous section was fulfilled in the images of the frequently used coil-100 database [3]. In the original database can be found the photos of 100 several simple objects with solid background. 72 photos of each object are in the database which was made from differenct directions with deflection. For instance some images of the database can be seen in the Figure 4.


Figure 4
Some elements of the COIL-100 database

## 3.1 Conditions of the Experiments

12-12 representatives of the 100 several objects was chosen for the experiments from the original database. All images of one object was made from different directions with deflection.

In this new database 100 searches were terminated. In all search the 12 closest images were found to a randomly chosen representative of an object using the simililarty measure which was defined earlier.

The testing was fulfilled by MATLAB 7.0.

## 3.2 Results of the Experiments

Considering the 12 found images closest to the query image, two aspect was analyzed: (1) how many images contains the same object as the query image has, and (2) how many images are totally different from the query image.

Table 1 shows the number of occurences in the case of both examined questions.

| Number of the same objects | Number of occurences | Number of totally different images | Number of occurences |
|---|---|---|---|
| 0 | 12 | 0 | 60 |
| 1 | 16 | 1 | 8 |
| 2 | 21 | 2 | 3 |
| 3 | 7 | 3 | 10 |
| 4 | 7 | 4 | 10 |
| 5 | 8 | 5 | 3 |
| 6 | 7 | 6 | 2 |
| 7 | 7 | 7 | 3 |
| 8 | 4 | 8 | 3 |
| 9 | 6 | 9 | 2 |
| 10 | 3 | 10 | 1 |
| 11 | 2 | 11 | 1 |
| | | 12 | 0 |

Table 1

Results of the experiments

## Conclusions

The results of the experiments presents that a new procedure development was succeded, which method can find at least one representative of the object of the query image in 88% of the occurences. In 91% of the occurences the number of the totally different images from the query image is not more than 1/3.

The results of the experiments prove that the developed procedure can be used for content based image retrieval, if the images of the database have homogeneous regions taking notice of the color and the position of the regions.

## References


[1] K. Fukunaga: Introduction to Statistical Pattern Recognition, Academic Press, 1990

[2] J. Matas: Colour-based Object Recognition, PhD dissertation, University of Surrey, 1996

[3] S. A. Nene, S. K. Nayar, H. Murase: Columbia Image Object Library (COIL-100), Technical Report CUCS-006-96, February 1996

[4] D. Paulus, L. Csink, H. Niemann: Color Cluster Rotation, In: Proceedings of International Conference on Image Processing, 1998, pp. 161-165

[5] Sz. Sergyán, L. Csink: Consistency Check of Image Databases, In: Proceedings of 2nd Romanian-Hungarian Joint Symposium on Applied Computational Intelligence, Timisoara, Romania, May 12-14, 2005, pp. 201-206

[6] Sz. Sergyán, L. Csink: Experiments in Color Based Region Detection (in Hungarian), In: Proceedings of Informatika a Felsőoktatásban 2005 Konferencia, Debrecen, 2005, ISBN 963 472 909 6

[7] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain: Content-Based Image Retrieval at the End of the Early Years, In: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22(12), 2000, pp. 1349-1380

[8] J. R. Smith, S.-F. Chang: Tools and Techniques for Color Image Retrieval, In: IS & T/SPIE Proceedings, Vol. 2670, Storage and Retrieval for Image and Video Database IV, 1995