A STUDY OF TWO COLOR SYSTEMS USED IN CONTENT-BASED IMAGE QUERY ON MEDICAL IMAGERY

Liana Stanescu, Dumitru Burdescu, Cosmin Stoica and Marius Brezovan
University of Craiova, Faculty of Automation, Computers and Electronics

Keywords: Image feature extraction, image processing, content-based visual query, color, histogram, HSV color space, l color space, medical images.

Abstract: The article presents a comparative study over two methods used in content-based visual query. The two methods refer at two different color systems used for representing color information from images: HSV quantized to 166 colors and l1l2l3 quantized to 64 colors. The originality of the study comes from the fact that it was made on a database with medical images from digestive tract area, captured by an endoscope. The scope was to check the quality of the content-based visual query on images representing five different diagnoses (colitis, ulcer, esophagitis, polyps and ulcerous tumor) and taking into consideration that some parameters were modified during the capturing process: viewing direction, intensity and direction of the illumination, parameters that affect mostly the medical images captured during the diagnosis process.

1 INTRODUCTION

As the world is in the middle of the digital era, the quantity of visual information is increasing (Sebe and Lew, 2001). More than 2700 digital pictures are made in every second (in total 85 billion images yearly). For example, PhotoWorks includes tens of millions of images on its web site. The Internet allows us to have access to a big part of these images. The common images are completed by images with special purpose, like medical images with an estimation of 2 billion per year. Because of the tendency for digital (television, movies) and because everybody will have access to everything, the number of images will be increasing. The world production of digital information in 2007 is estimated to be more than 10^9 GB (250 MB for each man on the planet, ignoring his technological development). It is estimated that in the next 10 years, each of us will manipulate terabytes of information (video, static images, music, photos and documents) every day.

These image databases are associated with the problem of the content-based retrieval, solved in two steps (Sebe and Lew, 2001).

In the first step, when inserting a new image, it will be pre-processed and some features will be automatically extracted: color, texture and shape. The result will be a characteristics vector that will be stored in the database.

In the second step the content based retrieval is made, by choosing a query image, calculating the characteristics vector, comparing this vector with each vector of the images stored in the database and viewing the most similar images.

The color is one of the base image properties. In content based retrieval on color feature it is desired to find the images from the database having the color composition closest to the color composition of the query image (Del Bimbo, 2001, Gevers and Smeulders, 1999).

The color content of an image is best represented by color histograms.

Comparing color histograms of the query image and target image is done by histograms intersection or by the quadratic distance between histograms that takes into consideration the conceptual similitude between two colors (Sebe and Lew, 2001, Smith, 1997).

As was said before, one of the domains where a large number of images are accumulated is the medical domain. The advantages of using content-based visual query on medical images are on the following directions (Muller et al, 2004):

- Medical teaching
- Medical research
- Diagnostic aid
- Electronic patient records

Stanescu L., Burdescu D., Stoica C. and Brezovan M. (2007).
A STUDY OF TWO COLOR SYSTEMS USED IN CONTENT-BASED IMAGE QUERY ON MEDICAL IMAGERY.
In *Proceedings of the Fourth International Conference on Informatics in Control, Automation and Robotics*, pages 337-340
DOI: 10.5220/0001621303370340

The medical images are being produced directly by medical equipment used in patient diagnosis, or by digitalizing the images stored on other devices.

In each of these methods some changes can occur:

- Changes in viewing direction
- Changes in direction of the illumination
- Changes in the intensity of the illumination

As a result, the purpose of the paper is to make a comparative study of the content-based query results effectuated on medical images database where the color information is represented by HSV and l1l2l3 color systems.

The originality of the study is given by the fact that the experiments are made on medical images from digestive area produced by an endoscope. The ill area is seen from different directions and in different illumination intensity. This study, unlike the others made on CorelDraw images, uses images produced in real condition, in patient diagnosis.

The paper has the following structure: In section 2 the two color systems are presented. In section 3 the conditions and the results of the experiments are presented, and in section 4 the conclusions of the comparative study are discussed.

2 CONTENT-BASED IMAGE QUERY ON COLOR FEATURE

The color is the visual feature that is immediately perceived on an image. The color space used for representing color information in an image has a great importance in content-based image query, so this direction of research was intensely studied (Del Bimbo, 2001).

There is no color system that it is universal used, because the notion of color can be modeled and interpreted in different ways. Each system has its own color models that represent the system parameters (Gevers, 2004).

There were created several color spaces, for different purposes: RGB (for displaying process), XYZ (for color standardization), rgb, xyz (for color normalization and representation), CieLuv, CieLab (for perceptual uniformity), HSV (intuitive description) (Gevers, 2001, Gevers, 2004). The color systems were studied taking into consideration different criteria imposed by content-based visual query (Gevers and Smeulders, 1999):

- The independence of the imaging device
- Perceptual uniformity
- Linear transformation
- Intuitive for user
- Robust against imaging conditions: invariant to a change in viewing direction, invariant to a change in object geometry, invariant to a change in direction and intensity of the illumination and invariant to a change in the spectral power distribution of the illumination.

The studies have shown that two of these color systems can be used, with good results in a content-based visual query process, namely HSV and l1l2l3 (Gevers et al, 2006).

It was proved that the HSV color system has the following properties (Gevers, 2004):

- It is close to the human perception of colors
- It is intuitive
- It is invariant to illumination intensity and camera direction

The studies made on nature and medical images have shown that in the case of the HSV, RGB and CieLuv color systems, the HSV color space produces the best results in content based retrieval (Stanescu et al, 2006).

Still, the HSV color space has several problems (Gevers, 2004) :

- Nonlinear (but still simple) transformation from RGB to HSV
- Device dependent
- the H component becomes instable when S is close to 0
- the H component is dependent of the illumination color

Gevers and Smeulders have proposed a new color system, named l, whose components are defined using the equations (Gevers and Smeulders, 1999):

$$l1(R,G,B) = \frac{(R-G)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2}$$
$$l2(R,G,B) = \frac{(R-B)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2} \quad (1)$$
$$l3(R,G,B) = \frac{(G-B)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2}$$

Where R, G, B are the color values in the RGB color space. They also showed that the l color system is invariant to viewing direction, illumination direction and intensity. In this case it is also a nonlinear, but simple, transforming from RGB space to l space.

In case of medical images the main problems are regarding changing illumination intensity and viewing direction. That is why the two color spaces presented above are chosen.

The operation of color system quantization is needed in order to reduce the number of colors used in content-based visual query: from millions to tens. The quantization of the HSV color space to 166 colors, solution proposed by J.R. Smith, is the idea used in this study (Smith, 1997). For the color space l1l2l3 the solution of quantization to 64 colors is chosen, keeping 4 values for each component of the system. The fact that a color system is quantized to 166 colors and the other to 64 colors does not influence the quality of the content-based image query process, the research studies showing clearly this aspect (Stanescu et al, 2006). The color histograms represent the traditional method of describing the color properties of the images. They have the advantages of easy computation and up to certain point are insensitive to camera rotating, zooming, and changes in image resolution (Del Bimbo, 2001). In case of both color systems, to compute the distance between the color histograms of the query image and the target image, the intersection of the histograms is used (Smith, 1997). The studies have also shown that using this metric in content-based visual query gives very good results as quadratic distance between histograms that is more difficult to calculate (Smith, 1997, Stanescu et al, 2006).

3 EXPERIMENTS

The experiments were performed in the following conditions.

A database with 520 color images from the field of the digestive apparatus was created. The images are from patients with the following diagnosis: polyps, ulcer, esophagitis, ulcerous tumors and colitis. For each image there are several images with affected area captured from 3 or 4 viewing directions. For each image in the database there is another identical image, but having the illumination intensity changed.

A software tool that permits the processing of each image was created. The software tool executes the following steps:

1. the transformation of image from RGB color space to HSV color space and the quantization to 166 colors
2. the transformation of image from RGB color space to l1l2l3 color space and the quantization to 64 colors
3. calculation of the two color histograms with 166, respectively 64 values, that represent the characteristics vectors and storing them in the database

In order to make the query the procedure is:

- a query image is chosen
- the dissimilitude between the query image and every target image from the database is computed, for each two specified criteria (color histograms with 166 colors and the color histogram with 64 colors);
- the images are displayed on 2 columns corresponding to the 2 methods in ascending order of the computed distance.

For each query, the relevant images have been established. Each of the relevant images has become in turn a query image, and the final results for a query are an average of these individual results.

The experimental results are summarized in table 1. Method 1 represents the query using the HSV color space quantized at 166 colors and Method 2 represents the query on color using the l1l2l3 color space quantized at 64 colors. The values in the table represent the number of relevant images of the first 5 images retrieved for each query and each of the methods, as an average of the values obtained on each executed query.

Table 1: Experimental results.

Query	Method 1	Method 2
Polyps	3.6	3.2
Colitis	3.5	3.1
Ulcer	3.2	2.9
Ulcerous Tumor	3.5	3.1
Esophagitis	3.4	3.1

It must be mentioned that the queries were made for each of the 5 diagnostics in part. The notion of relevant image was strictly defined. The images from the same patient captured at different illumination intensity and from different points of view were considered relevant for a query, and not the ones with the same diagnosis. The quality of the content-based image query process was strictly analyzed. In figure 1 there is an example of content-based image query considering the two specified methods for images categorized as colitis. The first column contains 5 images retrieved by Method1 and the second contains the images retrieved using Method2. In the first case there are 5 relevant images and in the second case, 4 relevant images.

4 CONCLUSION

The paper presents the condition in which the quality of the content-based visual query process was studied, using a collection of medical images from digestive tract. The quality was measured

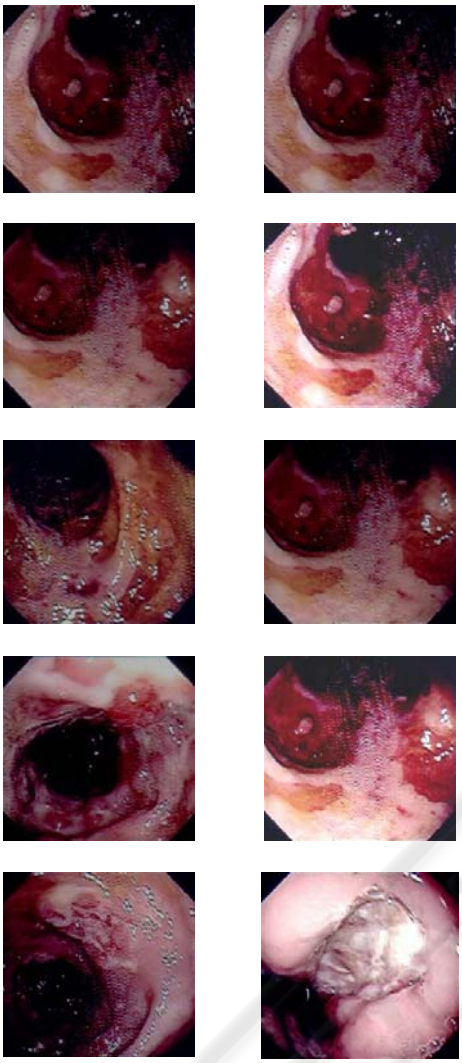

Figure 1: The retrieved images using the two specified methods.

calculating the precision and recall parameters. HSV system, quantized to 166 colors and l1l2l3 color system quantized to 64 colors were considered highlighting the way they influence the process of content-based visual query if some important parameters that often affects medical images are modified: viewing direction, direction and intensity of the illumination.

Several conclusions can be formulated after the experimental results were analyzed:

1. to find images representing the same ill area, that were captured by an endoscope from several viewing directions, the solution that uses HSV color system quantized to 166 colors gives the best results
2. for images representing the same ill area, captured to different illumination intensities, the solution that uses l1l2l3l color system quantized to 64 colors, gives the best results in querying process
3. globally, the solution that uses HSV color space gives most satisfying results, because the database includes both types of images

In general, for medical images, the first case, with images representing ill area captured from different angles is the most frequent case. So, that is why the use of HSV color space, quantized to 166 colors, is recommended. The situation in the database that was studied was the same, namely, the number of images captured from different angles was higher than the number of images where only the illumination intensity was different.

In the future the study will be extended by using a bigger database with much more images in order to see if this conclusion will be also confirmed. New experiments with images from other parts of the human body or images produced by other medical devices will be effectuated.

REFERENCES

Del Bimbo, A., 2001. *Visual Information Retrieval*, Morgan Kaufmann Publishers. San Francisco USA.

Gevers, T., Smeulders, W.M., 1999. Color-based object recognition. *Pattern Recognition.* 32, 453-464

Gevers, T., 2001. Color in Image Search Engines. In *Principles of Visual Information Retrieval*. Springer-Verlag, London.

Gevers, T., 2004. Image Search Engines: An Overview. In *Emerging Topics in Computer Vision*. Prentice Hall.

Gevers, T., Van de Weijer. J., Stokman, H., 2006. Color Feature Detection. In *Color Image Processing: Methods and Applications*. CRC Press.

Muller, H., Michoux, N., Bandon, D., Geissbuhler, A., 2004. A Review of Content_based Image Retrieval Systems in Medical Application – Clinical Benefits and Future Directions. *Int J Med Inform*. 73(1)

Sebe, N., Lew, M., 2001. Color-based retrieval. *Pattern Recognition Letters*. 22, 223-230

Smith, J.R., 1997. *Integrated Spatial and Feature Image Systems: Retrieval, Compression and Analysis, Ph.D. thesis*, Graduate School of Arts and Sciences. Columbia University.

Stanescu, L., Burdescu, D.D., Ion, A., Brezovan, M., 2006. Content-Based Image Query on Color Feature in the Image Databases Obtained from DICOM Files. In : *International Multi-Conference on Computing in the Global Information Technology*. Bucharest. Romania