CONTENT-BASED VISUAL RETRIEVAL ON MULTIPLE FEATURES IN THE IMAGE DATABASES OBTAINED FROM DICOM FILES

Liana Stanescu, Dan Dumitru Burdescu
Faculty of Automation, Computers and Electronics,University of Craiova, Bvd. Decebal, Craiova, Romania

Keywords: Content-based visual retrieval, color, color texture, region detection, color set back-projection algorithm.

Abstract: The paper presents the results of some experiments effectuated in the content-based visual query process applied on color medical images extracted from the DICOM files provided by the medical tools. The color feature was considered first, and the study implied more quantization methods (HSV color space at 166 colors, RGB color space at 64 colors and CIE-LUV color space at 512 colors) and several methods of computing the dissimilitude between the query and the target images (Euclidian distance, the histogram intersection and the quadratic distance between histograms). The content-based visual query on color texture feature was tested using two important methods of texture detection: the co-occurrence matrices and the Gabor filters. Also, the accurateness of the color set back-projection in detecting color regions representing sick tissue in medical images was studied. The effectuated statistics encourage the use of this algorithm in keeping track of the patient evolution under a certain treatment, with performances both in quality and speed.

1 INTRODUCTION

Retrieving the visual information is important in many applications starting with the artistic domain (art galleries, museums), to security and medical fields, which are in fact the most important (Del Bimbo, 2001). The purpose of this process is to retrieve from a database the relevant images for a query. The extraction of data that describe as accurately as possible the visual content is essential. Visual elements such as color, texture, shape that directly describe the visual content are used for retrieving images with a similar content from the database (Del Bimbo, 2001).

One of the domains in which the use of the content-based visual retrieval is needed is the medical one. This is mainly due to the fact that in the process of patient diagnosis, medical tools that offer images to the doctor are used on a large scale (Muller et al, 2004). A big part of these medical tools generates a standard DICOM file including images and associated data. This process led to very large medical image databases. Except for the traditional information retrieval in these databases (taking into account the patient name, the doctor name, the diagnosis), it is necessary to have a content-based visual query for the following reasons:

- The diagnosis process can be clarified in certain cases;
- The education and research activity can be improved by using the access visual methods;
- The visual characteristics allow not only the retrieving of the patients having the same disease, but also the cases where the visual similitude exists, but the diagnosis differs;

There are still few systems that are really integrated into the medical diagnosis process. The work for the application of the most suitable algorithms in image processing and features extraction continues (Muller et al, 2004). A large part of the images given by the medical apparatus are color, in which case the characteristics like color, color texture and shape must be considered.

In this article, the research has been effectuated on color images extracted from DICOM files provided by medical tools used in the diagnosis process. The images are stored in a database on which is applied the content-based image query on color and color texture features because there are

Stanescu L. and Dumitru Burdescu D. (2006).
CONTENT-BASED VISUAL RETRIEVAL ON MULTIPLE FEATURES IN THE IMAGE DATABASES OBTAINED FROM DICOM FILES.
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications*, pages 347-350
DOI: 10.5220/0001572203470350

some diseases that are characterized by the change of the color and the texture of the affected tissue, for example: colitis, esophagitis, polyps, ulcer and ulcerous tumor. Also, an important medical application of the color set back-projection algorithm, which is a method for detecting color regions, is presented.

2 CONTENT-BASED VISUAL RETRIEVAL ON COLOR FEATURE

The color is the visual feature immediately perceived on an image. In content-based visual query on color feature is important the used color space and the level of quantization, meaning the maximum number of colors (Del Bimbo, 2001). Because there is not a unanimously accepted solution on the appropriate color space to be used in the content–based image query on color feature, the study realized on color images extracted from the DICOM files takes into consideration three solutions, like:

1.The transformation of the RGB color space to HSV and the quantization at 166 colors (Smith, 1997)-**M1**

2.The use of the RGB color space quantized at 64 colors – **M2**

3.The transformation of the RGB color space at the CIE-LUV and the quantization at 512 colors (Smith, 1997)– **M3**

There are considered different color spaces and different levels of quantization to determine the way they affect the retrieval quality. It should be mentioned that in this study there are displayed three sets of result images that correspond to calculation mode of the distances between query image and target image. There have been taken into consideration the Euclidian distance (**D1**), the intersection of histograms (**D2**) and the quadratic distance between histograms (**D3**) (Smith, 1997).

The experiments were performed in the following conditions:

1. The images and alphanumeric data were extracted from DICOM files by applying the necessary algorithms (DICOM, 2006, LEAD, 2006).

2. It was created the test database with 920 color medical images extracted from the DICOM files, representing stomach and duodenum ulcers, ulcerate cancer, hernias and esophagus varicosis.

3. Each image from the database was processed before the execution of any query.

4. For each experimental query, an image was chosen like query image and there were established by a human factor, specialist medic, the images considered relevant for the respective query.

5. Each of the relevant images for the considered query was utilized, one by one, for querying the database containing images. The final values of the precision and the recall represent an average of the values resulted in the case of each image taken one by one as query image.

6. For each experimental query was constructed the graphic of the precision reported to the recall for each of the three distances in the case of each quantization method. Also there were presented under a tabular form the values that represent the number of relevant images, existing in the first 5, respectively 10 retrieved images, on the one hand, and on the other hand the number of images which can be retrieved for retrieving among them the first 5, respectively 10 relevant images.

The values from table 1 represent an average of the resulted values in the case of each image taken, one by one, as query image for the query that considers the stomach and duodenum ulcers images.

Table 1: Stomach and Duodenum Ulcers.

	D1	D2	D3
M1	5(9)	5(9)	5(9)
	8(13)	6(11)	7(12)
M2	5(9)	5(9)	5(9)
	7(18)	6(11)	5(11)
M3	4(7)	5(9)	4(7)
	8(23)	6(13)	7(17)

3 CONTENT-BASED VISUAL RETRIEVAL ON COLOR TEXTURE FEATURE

Together with color, texture is a powerful characteristic of an image, existing in nature and medical images, where a disease can be indicated by changes in the color and texture of a tissue. A series of methods have been studied to extract texture feature (Del Bimbo, 2001). Among the most representatives methods of texture detection are the co-occurrence matrices and Gabor representations, studied in this paper (Del Bimbo, 2001, Palm et al., 2000). There are many techniques used for texture extraction, but there isn't a certain method that can be considered the most appropriate, this depending on the application and the type of images taken into

account. Over the past few years, research has been done in color textures recognition, proving that the color information improves the color texture classification (Palm et al., 2000).

The experiments were performed in the following conditions:

1. the same database with 920 color images was used;
2. the co-occurrence matrices are computed for each component R,G,B and three vectors containing the 6 sizes (energy, entropy, maximum probability, contrast, inverse difference moment, correlation) are generated (Del Bimbo, 2001); the matrices are computed for d=1 and ϕ=0; in this case the characteristics vector has 18 values;
3. the Gabor characteristics vector containing 12 values (computed for 3 scales and 4 orientations) is generated (Palm et al., 2000);
4. the characteristics vectors generated at points 2 and 3 are stored in the database;

In order to make the query the procedure is:

1. a query image is chosen;
2. the dissimilitude between the query image and every target image from the database is computed for each specified criteria (the vector generated on the basis of the co-occurrence matrices and the vector for Gabor method);
3. 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 its 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 2. Met 1 represents the query on color texture feature using Gabor method and Met 2 represents the query on color texture feature using co-occurrence matrices.

The values in the table represent the number of relevant images in the first 5 images retrieved for each query and each method.

Table 2: The experimental results.

Query	Met 1	Met 2
Polyps	2.2	2.8
Colitis	1.7	1.7
Ulcer	1.8	2.2
Ulcerous Tumor	1.1	1.5
Esophagitis	1.7	2.5

4 DETECTING COLOR REGIONS

The application of an automated algorithm for detecting the color regions in medical images has two important utilizations:

1. in content-based region query on medical images collections, the specialist chooses one or more of the detected regions for querying the database, the purpose being the retrieval of images with similar color regions; this can be useful for clarifying some uncertain diagnosis. This problem was studied in Stanescu et al., 2004, interesting results being presented;
2. during the evolution in time of the disease of patients that follow a certain treatment;

For detecting color regions it was chosen the color set back-projection algorithm, introduced initially by Swain and Ballard and then developed in the research projects at Columbia University (Smith, 1997). For each detected regions the color set that generated it, the area and the localization are stored in the database. All the information is necessary further on in studying the evolution of the patients. The region localization is given by the minimal bounding rectangle (MBR). The region area is represented by the number of color pixels, and can be smaller than the minimum bounding rectangle.

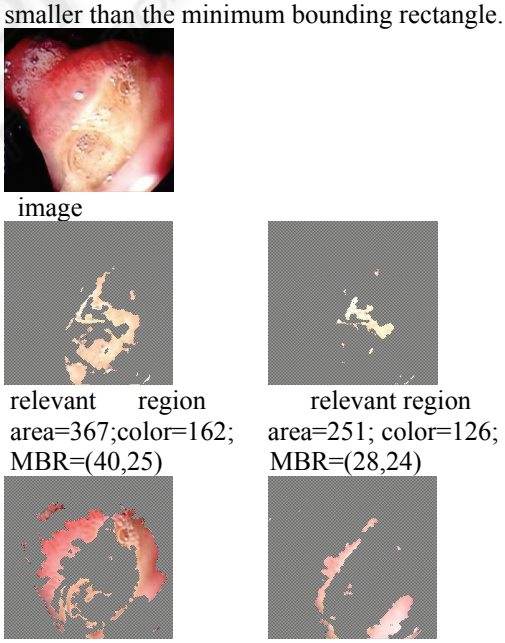

Figure 1: The detected color regions for the image.

The experiments were effectuated on 202 color images extracted from DICOM files and representing patients with peptic ulcer disease. The color set back-projection algorithm is applied on

each image for detecting the single color regions, and the region (s) that represents the sick zone were marked as relevant. By applying a certain drugs treatment on some patients, at some time intervals strictly established by the physician, the images were again collected from the same patients and the same algorithm is applied for detecting the color regions. The relevant region(s) that represents the sick tissue is marked again. The comparison between the new and old regions detected as relevant for the same patient, taking into consideration the number of pixels, can help the physician to establish in what percentage the sick region is reduced because of the administrated drugs. This approach may lead to a more rapid estimation and correct enough of the percentage in which the medication has a good effect in the ulcer diagnosis. This may come in the help of the patients, specialists, and the drugs producers that intend to test a medical product. The experiments showed that there were slight differences between human observer and computer system in order to appreciate healing staging. The speed of the retrieval process was also tested, comparing the time spent by the observer and the computer to find each patient's record in the database. This process was electronically measured and stored in the computer for statistics. The result is that the software has a significantly higher speed than human observer with no significant decrease of the retrieval quality.

5 CONCLUSIONS

As a result of effectuating an important number of experiments (synthetically presented here) in the de content-based visual retrieval process on databases with images extracted from DICOM files generated by medical tools, some conclusions are clear. In the case of the content-based image query on color feature, the series of effectuated tests indicated that the best results were obtained for the color space HSV quantized at 166 colors and using the histogram intersection for computing the image similitude. In the case of the content-based image query on color texture feature, better results are obtained using the co-occurrence matrices. The two algorithms (Gabor filters and co-occurrence matrices) have the same complexity $O(n^2)$ where n represents the maximum dimension of the image (Burdescu, 1998). As a result, the co-occurrence matrices method is recommended in this type of query. The retrieval system can combine two methods: one based on color feature and the other based on color texture detected with co-occurrence matrices, which complete one another. The statistic studies for the color set back-projection algorithm in keeping track of the patient evolution, during the treatment of a certain disease, emphasize a superior speed in sick region detection and a similar quality between the computerized mode and the process effectuated by the specialist.

In the future, the studies will be extended on more types of color medical images and new methods will be implemented and compared, taking into consideration the same factors: the complexity of the algorithm and the retrieval quality.

REFERENCES

Burdescu, D.D., 1998. *Analiza complexitatii algoritmilor*, Ed. Albastra. Cluj-Napoca.

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

DICOM Homepage, 2006. *http://medical.nema.org/*

LEAD Technologies, 2006. *http://www.leadtools.com/SDK/Medical/DICOM/ltdc1.htm*

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)

Palm, C., Keysers, D., Lehmann, T., Spitzer, K., 2000. Gabor Filtering of Complex Hue/Saturation Images For Color Texture Classification. In: *JCIS2000, 5th Joint Conference on Information Science*. Atlantic City, USA.

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., Mocanu, M., 2004. Detecting Color Regions and Content-based Region Query in databases with Medical Images. *Periodica Politechnica, Transactions on Automatic Control and Computer Science.* 49(63)