Proceedings of the Federated Conference on Computer Science and Information Systems pp. 707–710
ISBN 978-83-60810-51-4

A comparative study between recursive and non-recursive algoritms content search in a dual multimedia databases with images

Gabi Daniela Garaiman
University of Craiova, A.I. Cuza 13, 200585, Craiova, Romania Email: grm.dana@yahoo.com

Dan Dumitru Garaiman
University of Medicine and Pharmacy of Craiova, Petru Rares 2, 200349, Craiova, Romania
Email: dangaraiman@yahoo.com

***Abstract*—This is a comparative study for recursive and non-recursive content search algorithms in a dual multimedia databases with medical (endoscopic) and different (natural) images. The recursive algorithms in two stages use the same method of determining similarity at each stage and models to represent color content of images in quantified uniform spaces. The non-recursive algorithms use a single iteration and models to represent color content of images in unevenly quantified color spaces. The performance of the search has been measured according to four parameters: reappeal, precision, the quality of retrieval and the cost of retrieval. These are based on two methods of deciding the similarities between the models of images: Minkowski distance and Jaccard generalized measure. The model of representing the images in multimedia database used here are the normalized color histogram. The color space of the images are RGB reduced to 64 and 125 colors and HSV reduced to 60 and 162 colors. The study was realized in a dual database containing 360 endoscopic images grouped in 23 categories and 280 different images grouped in 10 categories. The results are presented both in tables and graphs.**

I. Introduction

CONTENT search algorithms in multimedia databases with images and multidisciplinary and interdisciplinary projects, attempt to answer to the multimedia information explosion that most of fields are facing today. Content search algorithms in multimedia databases have lately seen an unprecedented development, both at a conceptual level, as well at the implementing level. Content search systems allow the storing a high quantity of modeled images, and can return images similar to the search image in a short time.

The study will determine and compare the performances of recursive and non-recursive content search algorithms in multimedia databases in general, as well as in the particular case of medical endoscopic images. The purpose of this study is improving the methods of search in databases containing endoscopic images, as well as trying to apply these methods to other categories of medical images or other image databases.

II. Elements of the comparative study

A. Database

The study was executed on a dual database which contains:

- the models of 360 endoscopic images (END) organized in 23 categories containing 7 to 36 images;
- the models of 280 different images (DIF) organized in 10 categories containing 7 to 47 images;

Images were processed previously under study in order to standardize the size and eliminate the inconclusive.

B. Color spaces

The color spaces utilized in modeling the images are:

- the RGB (Red, Green, Blue) color space reduced to a number of 64 and 256 colors;
- the HSV (Hue, Saturation, Value) color space reduced to a number of 60 and 162 colors.

For recursive algorithms, the RGB color space was quantified by reducing the points that represent the middle

Table I.
The organization of categories for endoscopic images

No. categ.	The name of categorie	No. imag.
1.	adenopathy	12
2.	gastric atrophy	10
3.	esophageal cancer	27
4.	gastric cancer	36
5.	antral gastric cancer	16
6.	cyst	10
7.	duodenal diverticulum	17
8.	esophageal diverticulum	15
9.	reflux oesophagitis	16
10.	fitobezoar	7
11.	NSAID induced gastritis	10
12.	papular gastritis	10
13.	syphilitic gastritis	24
14.	gastropathy	34
15.	hiatus hernia	9
16.	normal major papilla	10
17.	polyp	11
18.	resected stomach	14
19.	duodenal ulcer	10
20.	esophageal ulcer	13
21.	gastric ulcer	18
22.	Esophageal varices	15
23.	Gastric varices	16

TABLE II.
THE ORGANIZATION OF CATEGORIES FOR DIFFERENT IMAGES

No. categ.	The name of categorie	No. imag.
1.	animals	24
2.	sunsets and sunrises	16
3.	flowers	40
4.	documents	7
5.	parks	20
6.	winter landscapes	13
7.	mountain landscapes	38
8.	urban landscapes	47
9.	seascapes	30
10.	portraits	45

of the intervals obtained from dividing one the axes of the coordinate system into 4, and the other into 5 equal intervals. The HSV color space was quantified in a similar way, by dividing the axes into (15,2,2) and (18,3,3).

For non-recursive content search algorithm in unevenly quantified spaces, each axis was divided in intervals of different sizes, which are obtained after analyzing the representation models in the uniform quantified spaces. The centers of the intervals are the new points of the quantified space. The sizes of the new intervals are obtained starting from the standard size by increasing or decreasing the according to the pertaining number of images.

C. The models of representation of images

One model of images representation was used in this study, the normalized color histogram which represents the color distribution of the images, being represented according to this formula:

$$H_I[m=1..M]=\left(\sum_{x=0}^{X-1}\sum_{y=0}^{Y-1}\begin{cases}1 \text{ if } I[x,y])=m\\ 0 \text{ otherwise}\end{cases}\right)/M \qquad (1)$$

where M is the degree of quantification of the color space and X and Y are the dimentions of the I image.

D. The methods of determining similarities

In this study two methods of determining similarities were used:

1) A method which uses Minkowski distance (MD) between the vectors of the models of image representation. It is defined by the formula:

$$d_M(X,Y)=\left(\sum_{i=1}^{N}|x_i-y_i|^2\right)^{1/2} \qquad (2)$$

in which X,Y are the vectors of the models of representation of images, with the dimension N. If the distance has the value 0, the representation models are considered identical (maximum similarity of the images) and for the value 1.41 of the distance, the models of representation are considered to be opposite (minimal similarity of the images).

2) A method which used the generalized measure Jaccard (GMJ) between the vectors of the models of image representation is defined by the formula:

$$m_J(x,y)=\frac{\sum_{i=1}^{N}x_i y_i}{\sum_{i=1}^{N}x_i^2+\sum_{i=1}^{N}y_i^2-\sum_{i=1}^{N}x_i y_i} \qquad (3)$$

in which X, Y are the vectors of the models of image representation, with the dimension N. For the value 1 of the measure, the models of representation are considered identical (maximum similarity of the models) and for the value 0 of the measure the models of representation are considered to be opposite (minimal similarity of the images).

E. The algorithms content search

In this study, two algorithms have been subjected to tests:

- recursive algorithms in two stages using the same method of determining similarity at each stage;
- non-recursive algorithms in one stage using the models of representations of images in unevenly quantified color spaces.

F. The performance of the retrieval

The performance of the retrieval process is measured according to four parameters:

1. reappeal measured the ability of the system to retrieve relevant information from the data base. R is defined as the proportion between the number of retrieved relevant items and the total number of relevant items;
2. precision (P) measured the accuracy of the retrieval process. It is defines as the proportion between the number of retrieved relevant items and the total number of retrieved items;
3. the quality of the retrieval (QR) is established by the order of the retrieved items, and it is defined according to the following formula:

TABLE III.
RESULTS OF RECURSIVE ALGORITHMS IN TWO STAGES USING NORMALIZED COLOR HISTOGRAM FOR DIFFERENT IMAGES

ID	Reapeal (R)		Precision (P)		Quality of retieval (QR)		Cost (C)	
RGB	**64**	**125**	**64**	**125**	**64**	**125**	**64**	**125**
MD	*0.49*	*0.51*	*0.50*	*0.47*	*0.51*	*0.51*	*19.5*	*38.2*
JGM	*0.61*	*0.65*	*0.54*	*0.61*	*0.51*	*0.52*	*36.4*	*71.0*
HSV	**60**	**162**	**60**	**162**	**60**	**162**	**60**	**162**
MD	*0.50*	*0.55*	*0.53*	*0.48*	*0.56*	*0.54*	*16.1*	*43.4*
JGM	*0.57*	*0.60*	*0.69*	*0.72*	*0.52*	*0.50*	*31.8*	*85.6*

$$QR = \frac{\sum_{i=1}^{n} \begin{cases} i \text{ if is relevant} \\ 0 \end{cases}}{\sum_{i=1}^{n} i} \quad (4)$$

where n is the total number of relevant items;

4. cost (C) measured the reaction speed of the algorithm. It is defined as the number of standard operations executed until the result is given. It is measured in million operations/ second.

III. RESULTS

The performances were determined at the level of the entire database as the average of the performances obtained for each category. At the category level, the performances were determined as the average of the performance for each image.

Table III presents the results for endoscopic images and table IV for different images in which the harmonic measure between reappeal and precision is high. Figure 1 shows the performance diagram (reappeal-precision) for the recursive algorithms in two stages that uses the methods of determining the similarity Minkowski distance, respectively generalized measure Jaccard for the normalized color histogram.

Similarity thresholds used are:

- for Minkowski distance and endoscopic images: 0.275 for first step and 0.155 for second step;
- for Minkowski distance and different images: 0.475 for first step and 0.280 for second step;
- for Jaccard generalized measure and endoscopic images: 0.675 for first step and 0.900 for second step
- for Jaccard generalized measure and different images: 0.475 for first step and 0.700 for second step.

Table V presents the results for endoscopic images and table VI for different images in which the harmonic measure between reappeal and precision is high. Figure 2 shows the performance diagram (reappeal-precision) for the non-recursive algorithms in unevenly quantified spaces that uses the methods of determining the similarity Minkowski distance, respectively generalized measure Jaccard for the normalized color histogram.

Similarity thresholds used are:

TABLE IV.
RESULTS OF RECURSIVE ALGORITHMS IN TWO STAGES USING NORMALIZED COLOR HISTOGRAM FOR ENDOSCOPIC IMAGES

ID	Reapeal (R)		Precision (P)		Quality of retieval (QR)		Cost (C)	
RGB	**64**	**125**	**64**	**125**	**64**	**125**	**64**	**125**
MD	*0.59*	*0.57*	*0.65*	*0.60*	*0.57*	*0.55*	*25.1*	*49.1*
JGM	*0.59*	*0.60*	*0.65*	*0.74*	*0.52*	*0.56*	*46.7*	*91.2*
HSV	**60**	**162**	**60**	**162**	**60**	**162**	**60**	**162**
MD	*0.62*	*0.61*	*0.64*	*0.64*	*0.59*	*0.60*	*20.6*	*55.7*
JGM	*0.59*	*0.64*	*0.52*	*0.60*	*0.49*	*0.50*	*40.8*	*110*

- for Minkowski distance and endoscopic images: 0.165;
- for Minkowski distance and different images: 0.300;
- for Jaccard generalized measure and endoscopic images: 0.860;
- for Jaccard generalized measure and different images: 0.650.

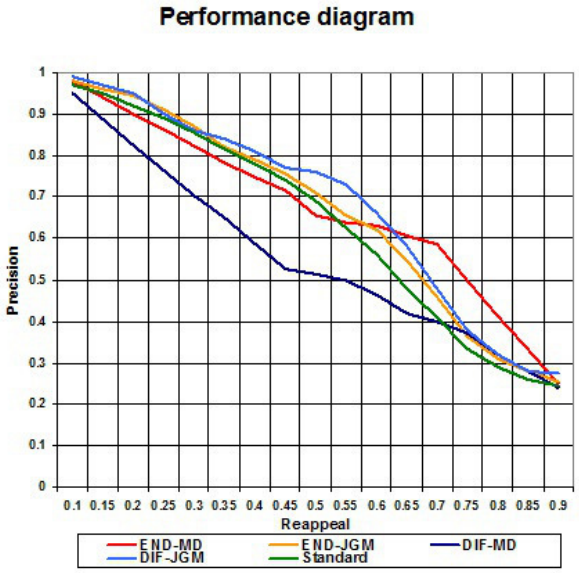

Fig. 1 Performance diagrams of recursive algorithms in two stages using normalized color histogram

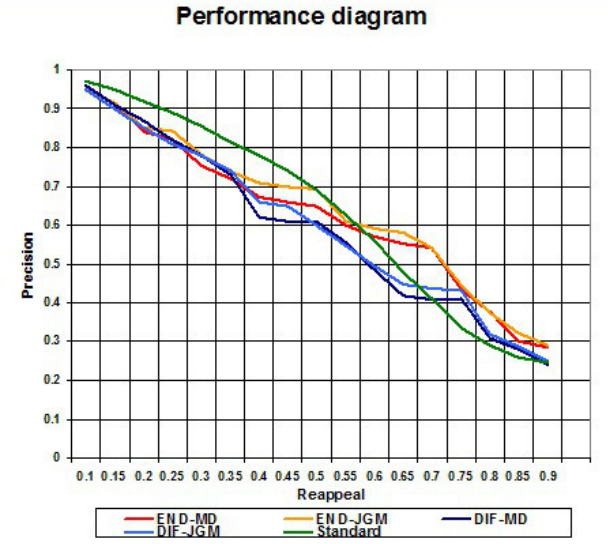

Fig. 2 Performance diagrams of non-recursive algorithms in unevenly quantified spaces using normalized color histogram

IV. CONCLUSIONS

The conclusions are based on the comparison between the performance of algorithms content search, starting of the criteria: type algorithm, method of determining similarity, color space used, comparasion with standard values for multimedia database with small-medium size (250-1000 images). Comparations values are summarized in table VII.

TABLE V.
RESULTS OF NON-RECURSIVE ALGORITHMS IN UNEVENLY QUANTIFIED SPACES USING NORMALIZED COLOR HISTOGRAM FOR ENDOSCOPIC IMAGES

ID	Reapeal (R)		Precision (P)		Quality of retieval (QR)		Cost (C)	
RGB	**64**	**125**	**64**	**125**	**64**	**125**	**64**	**125**
MD	*0.53*	*0.54*	*0.62*	*0.58*	*0.46*	*0.47*	*12.0*	*23.5*
JGM	*0.59*	*0.53*	*0.70*	*0.68*	*0.42*	*0.33*	*30.9*	*60.5*
HSV	**60**	**162**	**60**	**162**	**60**	**162**	**60**	**162**
MD	*0.49*	*0.51*	*0.63*	*0.63*	*0.42*	*0.44*	*11.2*	*30.4*
JGM	*0.44*	*0.46*	*0.67*	*0.74*	*0.45*	*0.46*	*29.0*	*78.3*

TABLE VI.
RESULTS OF NON-RECURSIVE ALGORITHMS IN UNEVENLY QUANTIFIED SPACES USING NORMALIZED COLOR HISTOGRAM FOR DIFFERENT IMAGES

ID	Reapeal (R)		Precision (P)		Quality of retieval (QR)		Cost (C)	
RGB	**64**	**125**	**64**	**125**	**64**	**125**	**64**	**125**
MD	*0.52*	*0.57*	*0.53*	*0.51*	*0.49*	*0.50*	*9.4*	*18.3*
JGM	*0.58*	*0.44*	*0.43*	*0.57*	*0.39*	*0.47*	*24.1*	*47.1*
HSV	**60**	**162**	**60**	**162**	**60**	**162**	**60**	**162**
MD	*0.53*	*0.61*	*0.58*	*0.54*	*0.52*	*0.52*	*8.8*	*23.7*
JGM	*0.44*	*0.41*	*0.75*	*0.80*	*0.40*	*0.35*	*22.6*	*60.9*

V. REFERENCES

[1] Garaiman D.D., Saftoiu A., *A comparative study for methods of content search in multimedia databases with endoscopic images*, Current health sciences journal, vol. 37, no. 4, 2011.

[2] Stanescu L., Burdescu D.D, Brezovan M., *A comparative study of some methods for color medical images segmentation*, EURASIP J. Adv. Sig. Proc, 2011.

[3] Scarlat R., Stanescu L., Popescu E., Burdescu D.D., *Case-Based Medical E-assessment System,* ICALT, p. 158-162, 2010.

[4] Burdescu D.D., Brezovan M., Ganea E., Stanescu L., *A New Method for Segmentation of Images Represented in a HSV Color Space,* ACIVS, p. 606-617, 2009.

[5] Garaiman D.D., Garaiman G.D., *Recursive algorithms content search in multimedia databases with endoscopic images*, IMCSIT, 2009, pp. 471-475.

[6] Stanescu L., Burdescu D.D., Mihai G., Brezovan M., Stoica Spahiu C., *A Graphical Interface For Medical E-Learning*. CSREA EE, p. 37-42, 2009.

[7] Chen Y., Wang J.Z., Krovetz R., *Cluster-Based Retrieval of Images by Unsupervised Learning,* IEEE Trans. Image Processing, vol. 14(8), 2005.

[8] Jeong, S., Won C.S., Gray R.M., *Image Retrieval Using Color Histograms Generated by Gauss Mixture Vector Quantization,* Computer Vision and Image Understanding, vol. 9 (1–3), 2004.

[9] Müller H., Michoux N., Bandon D., Geissbuhler A., *A Review of Content-Based Image Retrieval Systems in Medicine. Clinical Benefits And Future Directions*, International Journal of Medical Informatics, volume 73, 2004.

[10] Aisen A.M., Broderick L.S., Winer-Muram H., Brodley C.E., Kak A.C., Pavlopoulou C., Dy J., Shyu C.R., Marchiori A., *Automated storage and retrieval of medical images to assist diagnosis. Implementation and preliminary assessment*, Radiology, 228(1), 2003.

TABLE .VII
COMPARASIONS REGARDING THE RESULTS OF THE ALGORITHMS STUDIED

Comparation type	Difference level
1. Recursive algorithms / Non-recursive algorithms in unevenly quantified spaces	++ / - (END) + / - (DIF)
2. Algorithms in RGB color space/ Algorithms in HSV color space	- / + (END) - / + (DIF)
3. Recursive algorithms studied / Standard values	- / - (END) - / + (DIF)
4. Non-recursive algorithms in unevenly quantified spaces / Standard values	- / - (END) - / + (DIF)
5. Algorithms using Minkowski distance / Algorithms using Jaccard generalized measure	- / + (END) - / + (DIF)