Original Paper

A Comparative Study for Methods of Content Search in Multimedia Databases with Endoscopic Images

D.D. GARAIMAN[(1)], A. SAFTOIU[(2)]

[(1)]IT Department, University of Medicine and Pharmacy of Craiova; [(2)] Department of Gastroenterology, University of Medicine and Pharmacy of Craiova

ABSTRACT This is a comparative study for nerecursive algorithms content search in multimedia databases with endoscopic images are based three methods of deciding the similarities between the models of images: Minkowsky distance, generalized measure Jaccard and correlated measure Pearson. The performance of the search has been measured according to three parameters: reapel, precision and the quality of retrieval. The model of representing the images are the normalized color histogram. The color space of the images are RGB reduced to 125 and 256 colors and HSV reduced to 162 and 256 colors. The study was realized in a database containing 360 images grouped in 23 categories. The results are presented both in tables and graphs.

KEY WORDS *multimedia database, endoscopic images, methods of deciding the similarities, color space, nerecursive algorithms content search*

Introduction

Currently, most areas of activity of medical life facing an explosion of information on the nature of multimedia in general, and especially image. Have been developed and continue to develop effective systems management information both multimedia retrieval by content and its. Most systems are implemented on the investigation and treatment equipment, but in recent years the foundations of the development of databases conventional medical data and medical images at improving medical and decision act, but also medical education and research activities.

The systems of searching according to contents in a multimedia databases with medical images allow for the storage of large quantities of modeled images as they are able to come up with images similar to the target image. They utilize in subsidiary algorithms of content search, algorithms that are based methods of deciding the similarities between two image models.

The study aims at determining and comparing the extremely algorithms content search using three different methods of deciding the similarities for one multimedia dabase with endoscopic images with models to represent in color spaces RGB and HSV.

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

1. Reappeal (R) measured the ability of the system to retrieve relevant information from the data base. Reappeal 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 defined as the proportion between the number of retrieved relevant items and the total number of retrieved items.

3. The quality of the retrieval (CR) is established by the order of the retrieved items, and it is defined according to the following formula:

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

where n is the total number of relevant items.

Additionally, as will take account of harmonic measure of reappeal and precision parameters, given by:

harmonic measure = 2 * reappeal * precision / (reappeal + precision) (2)

Multimedia database

The study was executed on a data base which contains the models of 360 endoscopic images organized in 23 categories containing 7 to 36 images. Images were processed previously under study in order to standardize the size and eliminate the inconclusive.

The color spaces utilized in modeling the images are the RGB space reduced to a number of 125 and 256 color respectively, and HSV color space, reduced to a number of 162 and 256 colors respectively.

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

$$H_I[m = 1..M] = \left(\sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} \begin{cases} 1 \text{ if } I[x,y]) = m \\ 0 \text{ in rest} \end{cases} \right) / M \quad (3)$$

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

Table I. The organization on categories of multimedia database

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.	induced gastritis	10
12.	papular gastritis	10
13.	syphilitic gastritis	24
14.	gastropatie	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.	Ulcer gastric	18
22.	Varice esofagiene	15
23.	Varice gastrice	16

The Method of determining similarities

In this study three methods of determining similarities were used:

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

$$d_M(X,Y) = \left(\sum_{i=1}^{N} |x_i - y_i|^2 \right)^{1/2} \quad (4)$$

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).

- A method which used the generalized measure Jaccard between the vectors of the models of image representation is defined by the formula:

$$m_J(x,y) = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2 + \sum_{i=1}^{N} y_i^2 - \sum_{i=1}^{N} x_i y_i} \quad (5)$$

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).

- A method which used the correlated measure Pearson between the vectors of the models of image representation is defined by the formula:

$$m_P(X,Y) = \frac{\sum_{i=1}^{N} |x_i - \bar{x}||y_i - \bar{y}|}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \sum_{i=1}^{N} (y_i - \bar{y})^2}} \quad (6)$$

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).

Results

Establish performance was achieved by calculating the average parameters obtained by comparing each image with the rest of the base images from the database.

Tables II, III, IV presents the results of the reappeal and precision between the harmonic measure is optimal for the algorithms using Minkowski distance, generalized measure Jaccard and correlated measure Pearson.

Similarity thresholds used are:
- 0.175 for Minkowski distance;
- 0.850 for generalized measure Jaccard;
- 0.850 for correlated measure Pearson.

Table II. Optimal results for Minkowski distance and the normalized color histogram

Color space	1. Reappeal (R)		1. Precision 2. (P)		1. Quality retrival 2. (CR)	
RGB	2. 125	3. 256	3. 125	1. 256	1. 125	1. 256
	3. 0.47	4. 0.48	4. 0.63	2. 0.58	2. 0.48	2. 0.49
HSV	4. 162	5. 256	5. 162	3. 256	3. 162	3. 256
	5. 0.43	6. 0.45	6. 0.63	4. 0.63	4. 0.44	4. 0.45

Table III. Optimal results for generalized measure Jaccard and the normalized color histogram

Color space	Reappeal (R)		Precision (P)		Quality retrival (CR)	
RGB	125	256	125	256	125	256
	0.32	0.28	0.77	0.86	0.37	0.29
HSV	162	256	162	256	162	256
	0.41	0.42	0.55	0.61	0.40	0.41

Table IV. Optimal results for correlated measure Pearson and the normalized color histogram

Color space	Reappeal (R)		Precision (P)		Quality retrival (CR)	
RGB	125	256	125	256	125	256
	0.52	0.43	0.49	0.68	0.51	0.45
HSV	162	256	162	256	162	256
	0.46	0.50	0.47	0.63	0.45	0.45

Figure 1 shows the performance diagram (reappeal-precision) for the three nerecursives algorithms which uses methods of deciding the similarities between the models of images Minkowsky distance, generalized measure Jaccard

and correlated measure Pearson and a standard diagram for the database lower than 500 images.
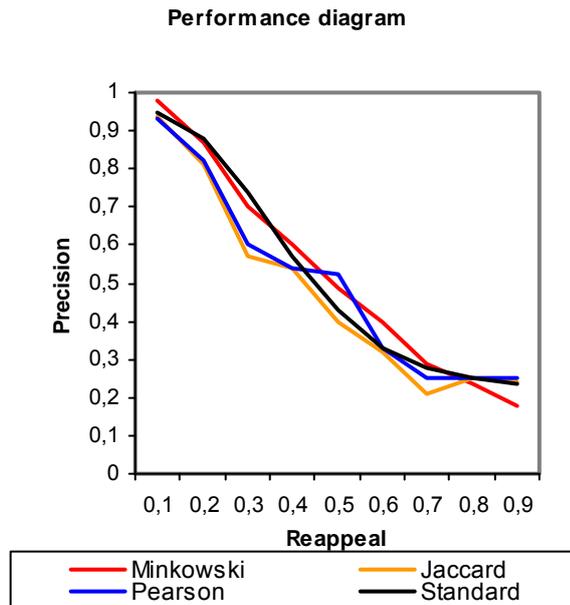
Performance diagram

Figure 1. Standard performance diagram and for three used algorithms

Conclusions

It will make two types of comparisons:

1. comparations of the performance parameters obtained for the three methods in points for the harmonic measure is optimal, with usual parameters of the database;

2. comparations between reappeal-precision diagrams for the three methods

Normal values for the parameters considered normal performance database of 500 images are: 0.450 for reappeal (**RN**), 0.500 for precision (**PN**) and 0.475 for quality retrival (**CRN**).

Media differences in the three methods in two spaces for reappeal ($dR=R_{RGB}-R_{HSV}$), precision ($dP=P_{RGB}-P_{HSV}$) and quality retrival ($dCR=CR_{RGB}-CR_{HSV}$) are:

- dR is negative, but low, having the value -0.019;

- dP is pozitive, but medium, having the value 0.127;

- dCR is negative, but low, having the value -0.042.

It is found that the three methods of determining similarity is within normal parameters for a small database.

From the performance diagram is inferred that the algorithms that have at base the Minkowski distance respond the best, been followed by those using correlated measure Pearson and at the final generalized measure Jaccard.

References

1. Abraham A., Grosan C., Ramos V., *Swarm Intteligence in Data Mining*, 2006.
2. Baeza-Yates R., Ribeiro-Neto B., *Modern Information Retrieval*, ACM Press/Addison-Wesley, 1999.
3. 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.
4. 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.
5. Smith J.R., Chang S.H., *Tools and Techniques for Color Image Retrieval*, Symposium on Electronic Imaging. Science and Technology. Storage and retrieval for Image and Video Database, San Jose, 1996.
6. Stănescu L., Burdescu D.D., Mihai G., Brezovan M., Stoica-Spahiu C., *Multimedia Elements for Medical e-Learning Improvement,* Proceedings of Intelligent Interactive Multimedia Systems and Services, Mogliano Veneto, Italy, 2009.
7. Stoica-Spahiu C., Stanescu L., Burdescu D.D., Brezovan M. *Multimedia Database Server Implementing Content Based Retrieval,* Proceedings of Third International Workshop on Enterprise Systems and Technology, Sofia, Bulgaria, 2009.

Correspondence Adress: D.D. Garaiman, IT Department, University of Medicine and Pharmacy of Craiova, Str Petru Rares nr. 4, 200456, Craiova, Dolj, România; e-mail: dangaraiman@yahoo.com