

Methods of establishing the similarity in searching by content in multimedia database with medical images

D.D. GARAIMAN ⁽¹⁾ GABI-DANIELA GARAIMAN ⁽²⁾

(1) IT Department, University of Medicine and Pharmacy of Craiova (2) University of Craiova

ABSTRACT The content based image search and retrieve systems are grounded on their images characteristics: the colour histogram, the colour binary set, the texture or granulation which are materialized into images representation models. The search and retrieval of a content based image implies comparisons between the images. This thing involves the use of a huge amount of resources and for a very long period of time. That is why the comparison takes place between the image representation models. The measuring of the similarity between the two images can be determined in the following ways: by calculating the Minkowski distance, the Hamming distance, the square distance, generalised measure Jaccard and correlation measure Pearson. The study was realized for a set of three images compose by a one interrogation endoscopic image and two target endoscopic images (one relevant image and one irrelevant image). They are compared results for all methods of establishing image similitude using the images representation models in the RGB colour space and the HSV color space.

KEYWORDS *medical image, multimedia database, similitude, histogram, colour binary set, texture*

Introduction

The increasing of information volum in all medical disciplines makes necessary the introduction of some quickly and efficient methods of storage and retrieval of all available dates.

The systems of searching by content in multimedia database with medical images allows the storage of a large amount of images modeled being able to return the similar images with the interrogation image. The models are built beginning from the images features about color content (normalized color histogram, binary color set) or texture(grain).

Representation models of images are secret distributions through vectors in a metric space N -dimensional (S^N), where N can be:

- the number of used colors:
- the number of grained size.

Searching and retrieval of an image based on content means comparisons between images. In practice, taking to consider the reduction of hardware necessary resources, the comparison runs between the representative models of images.

The comparison between two medical images, one of interrogation (I_I) and other target (I_T) the comparison is made by calculating the distance $d(v_I, v_T)$ or by calculating the approach (v_I, v_T) between vector's points of representation v_I and v_T , in the S^N space. The distance or the approach between v_I and v_T which belong to S^N is a number which satisfy the following conditions [17a]:

1. Identity: $d(v_I, v_T)=0$, $m(v_I, v_T)=\max$;
2. Non-negativity: $d(v_I, v_T)\geq 0$, $m(v_I, v_T)\geq 0$;

3. Commutativity or symmetry: $d(v_I, v_T)=d(v_T, v_I)\geq 0$, $m(v_I, v_T)=m(v_T, v_I)\geq 0$;

4. Triangle inequality: $d(v_A, v_C)\leq d(v_A, v_B)+d(v_B, v_C)$, $m(v_A, v_C)\geq m(v_A, v_B)+m(v_B, v_C)$.

Two images I_1 and I_2 are similar for one feature if the distance $d(v_1, v_2)$ and the approach $m(v_1, v_2)$ between associated points of characteristic vectors is smaller or equal or equal and higher or equal with a threshold of similarity τ .

Assessment of the similarity

If x and y are two points in a metric space S^N of dimension N having associated the $x = (x_1, x_2, \dots, x_N)$ and $y = (y_1, y_2, \dots, y_N)$, measuring the similarity of two points can be made using the following methods:

1. Minkowski distance [1], [2]:

$$d_{M,p}(x, y) = \left(\sum_{i=1}^N |x_i - y_i|^p \right)^{1/p},$$

where $p \in \mathbb{N}^*$;

2. Hamming distance [1], [3]:

$$d_H(x, y) = \frac{\sum_{i=1}^N |x_i - y_i|}{\sum_{i=1}^N x_i \sum_{i=1}^N y_i};$$

3. quadratic distance [1], [3]:

$$d_2(x, y) = \sum_{i=1}^N \sum_{j=1}^N |x_i - y_i| a_{i,j} |x_j - y_j|,$$

where $a_{i,j}$ is the similarity between vectors elements with indexes i and j , also $a_{i,j} = a_{j,i}$;

4. generalized Jaccard measure [2]:

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

5. correlated Pearson measure [2]:

$$d_p(x, y) = \frac{\sum_{i=1}^N |x_i - \bar{x}| |y_i - \bar{y}|}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}.$$

Two images I_1 and I_2 are similar in terms of one feature if the distance $d(v_1, v_2)$ and the approach $m(v_1, v_2)$ between associated points of characteristic vectors is smaller or equal or higher or equal with a threshold of similarity τ .

The values for each method are contained in an interval (upward for distance and downward for measure) to the left end corresponds to the maximum similarity (identical images in terms of representation model), and the to the right end for minimum similarity (opposite images).

The values of similarity threshold are considered to be correct for similarity after comparing two images with same values between in the first quarter of the interval for each method.

The Representation models used in two color spaces RGB and HSV are quantified at a number of 27, 125 and 36, 162 colors.

Assessment of similarity for color histograms

Taking to consider histograms as being normalized, so values of histogram vector x_i are included in the interval $[0,1]$. Both for the distance and for the measures, are compared identical elements for target histogram H_x with the elements of interrogation histogram (figure 1).

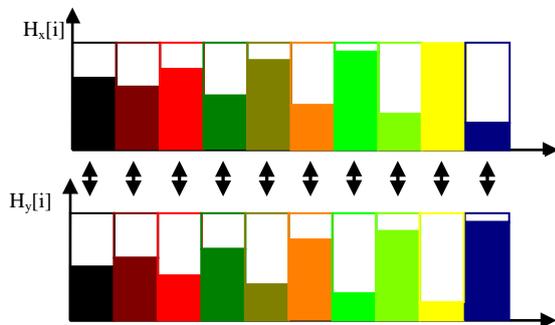


Figure 1. The comparison of color histogram elements for distance and measure

For quadratic distance are compared and weighted all target and interrogation histogram elements (figure 2).

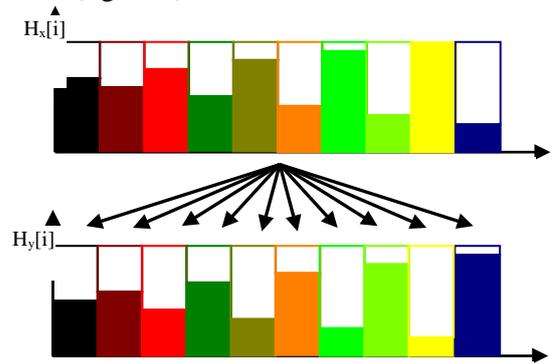


Figure 2 – The color histogram elements comparison for quadratic distance

In the table 1 are presented in a comparative way 3 endoscopic medical images (two images with esophageal ulcer and another one with syphilitic gastritis), the results of the assessment of similarity using methods based on distances and measures.

Table 1 – Comparisons between methods for color histograms

Color Space RGB			
Image	RGB Unreduced	RGB27	RGB125
target			
interrogation (1) relevant			
interrogation (2) irrelevant			
Method	Similarity values range		
Minkowski distance	[0;1.4142]	0.1010 (1) 0.6895 (2)	0.0798 (1) 0.5259 (2)
Hamming distance	[0;2]	0.2389 (1) 1.2791 (2)	0.2647 (1) 1.5684 (2)
quadratic distance	[0;2]	0.0072 (1) 0.4533 (2)	0.0056 (1) 0.3398 (2)
generalized Jaccard measure	[1;0]	0.9584 (1) 0.1485 (2)	0.9514 (1) 0.1267 (2)
correlated Pearson measure	[1;0]	0.9780 (1) 0.3108 (2)	0.9740 (1) 0.2596 (2)
Color Space HSV			
Image	HSV Unreduced	HSV36	HSV162

target			
interrogation (1) relevant			
interrogation (2) irrelevant			
Method	Similarity values range		
Minkowski distance	[0;1.414]	0.1140 (1) 0.8756 (2)	0.1507 (1) 0.5901 (2)
Hamming distance	[0;2]	0.2166 (1) 1.4962 (2)	0.4051 (1) 1.5122 (2)
quadratic distance	[0;2]	0.0090 (1) 0.9930 (2)	0.0144 (1) 0.4126 (2)
generalized Jaccard measure	[1;0]	0.9589 (1) 0.1958 (2)	0.8508 (1) 0.1538 (2)
correlated Pearson measure	[1;0]	0.9777 (1) 0.3579 (2)	0.9185 (1) 0.3104 (2)

Assessment of similarity for color binary sets

Values x_i of binary color set vector are included in the set $\{0,1\}$. As with histogram for distance are compared identical images of binary color target set S_x with binary color set elements of interrogation S_y (figure 3).

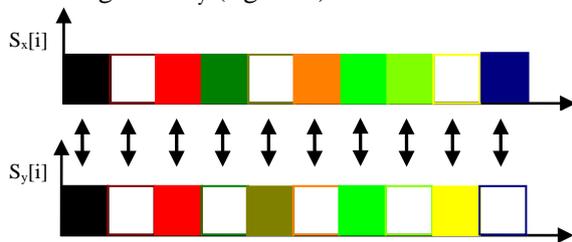


Figure 3 . Comparing the elements for binary color set

In table 2 are presented in a comparativ way for the three medical images of endoscopical nature the results of similarity assessment using the method based on Minkowski distance.

Table 2 – Comparisons for binary color set

Color Space RGB			
Image	Interval	RGB27	RGB125
Minkowski distance	[0;1.4142]	0.0740(1) 0.1851 (2)	0.0080 (1) 0.1120 (2)
Color Space HSV			
Image	Interval	HSV36	HSV162
Minkowski distance	[0;1.414]	0.0(1) 0.0833 (2)	0.0061(1) 0.0802 (2)

Assessment of similarity for texture

We consider textures as being normalized, so values x_i of texture vector are included in the interval $[0,1]$. In case of distance and measures are compared identical images of target textures T_x with elements of interrogation texture T_y (figure 4).

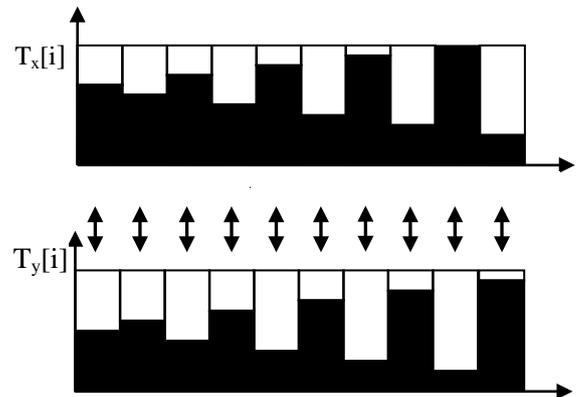


Figure 4 – Comparing texture elements for distances and measures

In table 3 are presented in a comparativ way for the three medical images of endoscopical nature the results of similarity assessment using methods based on distances and measure.

Table 3 – Comparisons for the texture

Color Space RGB			
Image	Interval	RGB27	RGB125
Minkowski distance	[0;1.4142]	0.1210 (1) 0.6004 (2)	0.2339 (1) 0.1838 (2)
Hamming distance	[0;2]	0.3325 (1) 1.4505 (2)	0.6368 (1) 0.6027 (2)
quadratic distance	[0;2]	0.0053 (1) 0.3967 (2)	0.0338 (1) 0.0305 (2)
generalized Jaccard measure	[1;0]	0.9142 (1) 0.0666 (2)	0.6545 (1) 0.7088 (2)
correlated Pearson measure	[1;0]	0.9785 (1) 0.2532 (2)	0.8629 (1) 0.8720 (2)
Color Space HSV			
Image	Interval	HSV36	HSV162
Minkowski distance	[0;1.414]	0.6853 (1) 0.9824 (2)	0.1713 (1) 0.3749 (2)
Hamming distance	[0;2]	1.4470 (1) 1.6540 (2)	0.4689 (1) 0.9609 (2)
quadratic distance	[0;2]	0.4880 (1) 1.1187 (2)	0.0162 (1) 0.1111 (2)
generalized Jaccard measure	[1;0]	0.0786 (1) 0.0183 (2)	0.7751 (1) 0.3004 (2)
correlated Pearson measure	[1;0]	0.2401 (1) 0.1148 (2)	0.8628 (1) 0.5221 (2)

Conclusions

From the analysis of the 3 table we have the following conclusions:

1. The values of distances and measures for relevant images must place in the first quarter of the values interval, and for irrelevant images in the second and in the third quarter.

2. For most of the used methods and models of representation similarity improves with the growth of the quantification grade of the color used space.

3. Normalized histograms answer the best to the requirements of appreciation similarity of endoscopic images, binary color sets can be used in the first stage in order to exclude dissimilar images, and the texture can give efficiency in combination with other representation models or for medical images of morphopathological nature.

4. As performance looking at similarity methods of calculation appreciation of it can be ranked like: Minkowski distance, quadratic distance, correlated Pearson measure, generalized Jaccard measure, Hamming distance.

5. We observe a relatively equality in terms of similarity performance appreciation for the space color used spaces: RGB and HSV.

Creating a system of searching by content in multimedia database with medical images allow you in a basis area of social life(medicine) to realise the implementation of modern techniques for archiving, training and diagnostic.

Establishing methods of similarity image appreciation represent an important step in implementation an efficient system of searching by content in multimedia database with medical images.

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. Scarlat R., Stanescu L., Popescu E., Burdescu D.D., Case-Based Medical E-assessment System, ICALT, p. 158-162, 2010.
3. Garaiman D.D., Garaiman G.D., Recursive algorithms content search in multimedia databases with endoscopic images, IMCSIT, 2009, pp. 471-475.
4. 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.
5. Smith, J.R., Chang, S.H., Automated Image Retrieval Using Color and Texture, IEEE Transactions on Pattern Analysis and machine Intelligence, 1999.
6. Smith, J.R., Integrated Spatial and Feature Image System. Retrieval, Compression and Analysis, Ph.D. thesis, Graduate Scholl of Arts and Sciences, Columbia University, 1997.

Correspondence Address: Dumitru-Dan Garaiman, IT Departament University of Medicine and Pharmacy of Craiova, Str. Petru Rares nr. 4, 200456, Craiova, Dolj, Romania, email: dangaraiman@yahoo.com