Design of a Fuzzy Distance for a CBIR System

Oscar D. Robles-Sánchez[1] and Enrique Frías-Martínez[2]
[1]Departamento de Tecnología Fotónica, Universidad Politécnica de Madrid
Campus de Montegancedo s/n, 28660 Boadilla del Monte, Madrid, Spain
Phone: +34-91-336-6944, Fax: +34-91-3367372
email: orobles@dtf.fi.upm.es
[2] Computer Science Department, New York University
715 Broadway, Room 719, New York, NY, USA
Phone: +1-212-998-3493
email: frias@cs.nyu.edu

ABSTRACT: This paper describes the design of a fuzzy distance for content-based image retrieval applied to 2D color images within an image database. The technique herein presented computes first a multiresolution representation using the Haar transform. Then, a color based primitive is obtained from the color level energy distribution at different resolution levels of the wavelet transform. Fuzzy membership functions are constructed from the values of the color based primitive of the test vector. The values of the elements of the database images are used to compute a fuzzy distance between the test and the database images. The experimental results given in this paper include the success rate achieved with the proposed distance using different membership functions and aggregation operators over a database composed of 298 color images and 68 tomographic volumes with test sets formed respectively by 1655 images and 805 volumes.

KEYWORDS: CBIR primitives, Data processing, Visual information systems, Fuzzy numbers, Fuzzy distance

INTRODUCTION

Current Content-Based Image Retrieval (CBIR) systems (e.g. [1], [2], [3]) use techniques inherited from computer vision and from database systems to represent and manage the available data. However, they differ from classical recognition systems in the working domain, where the set of objects that the system is able to recognize is very restricted. CBIR systems differ on the primitives used to retrieve the images: color, texture, scheme, shape, volume, spatial restrictions, objective and subjective attributes, movement, text labels, etc.; that is, any feature that can be used to represent image contents and has a discriminatory property. Some other classification criteria for this kind of systems can be the way the image's primitives are extracted –automatically, semi-automatically or manually-, the primitive abstraction level or the independence between the different conceptual domains.

Instead of working with the raw input data, a wavelet transform is applied to the images to obtain a multiresolution representation of the features searched. The development of the wavelet transform theory has spurred new interest in multiresolution methods and has provided a more rigorous mathematical framework. Wavelets give the possibility of computing compact representations of functions or data. Additionally, they allow variable degrees of detail or resolution to be achieved, and they are attractive from a computational point of view [4].

Fuzzy logic is a multi-valued logic designed to represent and manage knowledge with uncertainty. It is emerging as a powerful tool for any kind of application that tries to solve an ill-defined problem or a problem where the data or the decision mechanism has to deal with uncertainty, like image processing tasks with user interaction. Some of the sources of this uncertainty are imprecision in computations, ambiguity of representations and general problems in the interpretation of complex scenes.

One of the first connections of fuzzy set theory to image processing was made by Prewitt [5] who suggested that the results of image segmentation should be fuzzy subsets rather than crisp subsets of the image plane. Since then, fuzzy logic has been used constantly in image processing applications such as segmentation [6], boundary detection, object recognition and image enhancement [7], among others. In all those applications the representation capability of fuzzy logic is flexible and intuitive, and the results of the algorithms are good, producing not only crisp decisions when necessary, but also a corresponding degree of support.

This paper presents an application of fuzzy logic in a CBIR system, with a fuzzy distance into its minimum distance classifier. The selected primitive is based on the image's energy [8].

IMPLEMENTED PRIMITIVE

THE HAAR TRANSFORM

Wavelet transforms can be seen as a reformalization of the multiresolution methods of the 80's ([9], [10]). The information they provide is similar to that obtained from Fourier-based techniques, with the advantages of working with local information using base functions with compact support and keeping both frequency and spatial or temporal information of the original data. Wavelet transform coefficients show variations in object features at different resolution or scale levels ([9], [10], [11]). Roughly speaking, the detail coefficients of the wavelet transform can be considered as a high-frequency extraction process of the objects appearing on the images, while the analysis coefficients behave complementary: the lower the resolution level, the more homogeneous the regions they produce. This could be equivalent to a successive application of low-pass filters with a signal subsampling operation. The inverse of this process allows the reconstruction of the original signal by the so-called synthesis process.

The Haar transform has several properties that make it very appealing for our purposes:

- It keeps the energy of the wavelet transform coefficients invariant for all resolution level j [12].

$$\sum_{j,m} \gamma_{j,m}^2 = k \quad \forall j = 0,\ldots,n; \quad m = 0,\ldots,2^j - 1 \tag{1}$$

- The low order complexity of this transform, *O(n)*, allows an efficient implementation of the whole process.

The Haar transform will be used as a tool to extract some features of the transformed image that will allow to perform a discrimination process between the queries and the images stored in the information system [8], [13].

SIGNATURE BASED ON THE COEFFICIENT'S ENERGY

Our objective is to define a primitive that collects color based information of the original image at different resolution levels. It has been considered that the analysis coefficients pick up the most significant image information from the different resolution levels. Additionally, detail coefficients provide local information, which allows to discriminate between similar input images. Working with grey-level images, the wavelet transform of the original image provides all the information needed to compose the signature representing the contents of the image: concatenating the energy values for all the considered regions of coefficients at all the resolution levels, we obtain a vector of features collecting the intensity of the input data. When we deal with color images, for example in the RGB color space, we must apply the described process for the monochrome images over each one of the planes of the color images.

The energy of the analysis and the diagonal detail coefficients at each resolution level of the Haar transform for RGB images over each channel was chosen based on the results achieved in [8].

DESIGN OF THE FUZZY DISTANCE

The CBIR system compares the energy of each channel R, G and B at each one of the resolution levels of the query image with the corresponding primitive elements of each one of the images of the database. Once the system has compared the query image against all the images of the database, those images are ordered according to the value of their similarity distance.

To obtain the distance between the query image and any image of the database, the system constructs a fuzzy membership function $A_{i,j}$ for each level j of each one of the i channels of the query image. The definition is done by giving the sizes of both the support and the kernel. The support of a fuzzy membership function A is defined as:

$$\text{supp}(A) = \{x / \mu_A(x) > 0\} \tag{2}$$

The kernel of a fuzzy membership function A is the subset of elements of X that verifies:

$$\text{kernel}(A) = \{x / \mu_A(x) = 1\} \tag{3}$$

The kernel ($Kernel_{i,j}$) and support ($Support_{i,j}$) of each membership function $A_{i,j}$ are obtained as:

$$Kernel_{i,j} = \frac{e_{i,j} Ker}{100} \quad (4)$$

$$Support_{i,j} = \frac{e_{i,j} Sup}{100} \quad (5)$$

where $e_{i,j}$ is the energy of the query image on the channel i and level j, *Ker* is the percentage of $e_{i,j}$ that defines the kernel, and *Sup* the percentage that defines the support. The membership function $A_{i,j}$ is defined as a fuzzy number defined by $Kernel_{i,j}$ and $Support_{i,j}$, and centered in $e_{i,j}$. In this context, the kernel of the membership functions defines the degree up to which two energy levels are considered equal and the support the degree to which the system accepts the similarity of two energy levels.

Once the system has constructed the membership functions, it obtains the similarity between the energy of channel i and level j, $E_{i,j}$, of the image of the database and the membership function $A_{i,j}$ of the query image. This is done obtaining the degree of truth $\mu_{i,j}$ of $E_{i,j}$ in the membership function $A_{i,j}$:

$$\mu_{i,j}(E_{i,j}) = A_{i,j}(E_{i,j}) \quad (6)$$

The membership functions implemented have been the trapezoidal, quadratic and τ-Zadeh types (see Figure 1). The definition of the slope of these membership functions is given in (7),(8) and (9) respectively.

$$\mu_A(x) = \begin{cases} \frac{x-a}{b-a}, & a < x < b \\ 1, & x \geq b \end{cases} \quad (7)$$

$$\mu_A(x) = \begin{cases} \left(\frac{x-a}{b-a}\right)^2, & a < x < b \\ 1, & x \geq b \end{cases} \quad (8)$$

$$\mu_A(x) = \begin{cases} 2(\frac{x-a}{b-a})^2, a < x < c \\ 2(\frac{x-b}{b-a})^2, c \leq x < b \\ 1, x \geq b \end{cases} \quad (9)$$

Figure 1 presents an example of the left slope of these membership functions. The right slope of each function is defined in an analogous way.

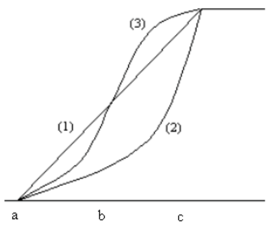

Figure 1.- Example of trapezoidal (1), quadratic (2) and τ-Zadeh (3) membership function.

Depending on the type of membership function used, the semantic associated with the slope is different. In a trapezoidal membership function the variation from 0 to 1 is lineal, which means that everything is penalized in the same way. In case a quadratic membership function is used, the shape of the function penalizes any value of energy that is not in the kernel, having a smaller degree of truth than in the case of the trapezoidal membership function. In case a τ-Zadeh membership function is used, the system does not penalize much the values of the energy that are near the kernel, but as the difference increases, the membership function penalizes more those values.

Once all the degrees of truth have been obtained, an aggregation operator is applied to obtain the final distance D between the two images compared:

$$D = \mathop{\Phi}_{i=1}^{N} (\mathop{\Phi}_{j=1}^{K} \mu_{i,j}(E_{i,j})) \quad (10)$$

where Φ is the aggregation operator implemented, $N=3$ is the number of channels, K is the number of energy levels, $E_{i,j}$ the energy of the database image for channel i and level j, and $\mu_{i,j}$ the membership function defined by the energy level j of channel i of the image to find in the database.

Once the mechanism to obtain the distance between two images has been defined, the most appropriate values to define the kernel (*Ker*), the support (*Sup*), the type of membership function (trapezoidal, quadratic or S-Zadeh) and the aggregation operator (Φ), have to be determined in order to obtain the optimum identification rate.

RESULTS

EXPERIMENT DESCRIPTION

We have used the same databases described in [8] and [13]. The fist one was formed by 298 RGB two-dimension 128x128 pixel size images collected from different sources, like [14] or Internet. The test set was generated introducing images which share the same concept but look quite different from those of the considered database,applying affine transformations or selecting regions of interest from the original database images. Following these guidelines, a test set formed by 1655 two-dimension color images was obtained.

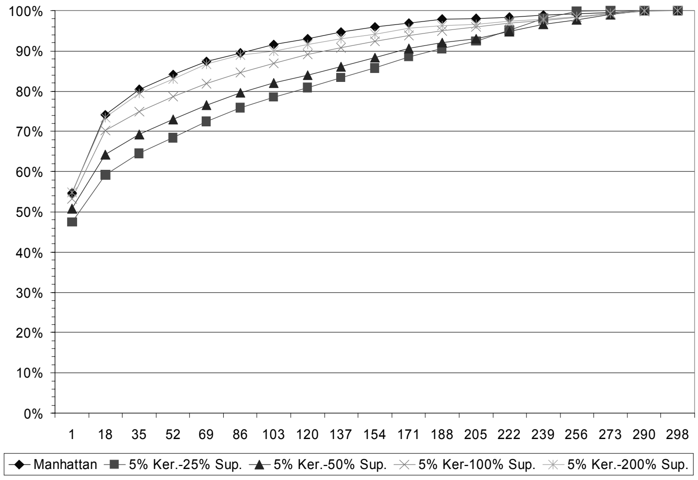

Figure 2.- 2D database: different retrieval rate results of the trapezoidal function.

The second test is the volumetric image database, formed by 68 tomographic images from different parts of a human body, each one of 64^3 voxels. The density of tissues, bones and organs is represented on a gray-level range. The associated test set was formed by 805 images, all of them generated by shifting 5 pixels the original volumetric images by means of all possible combinations over each one of the three Cartesian axes.

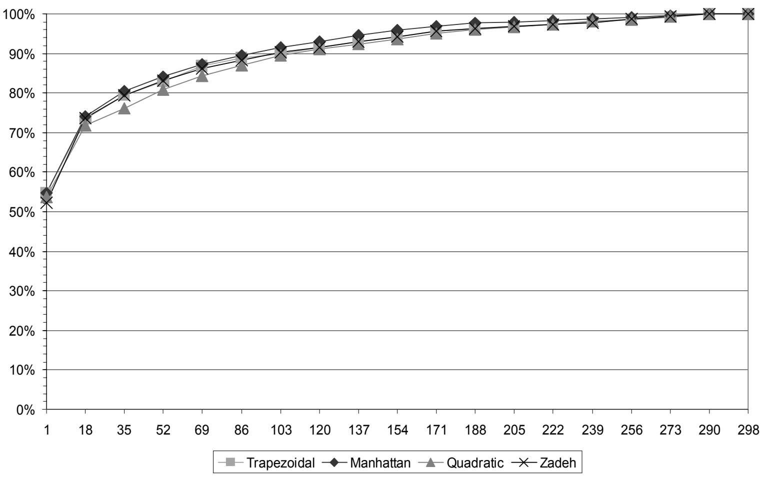

Figure 3.- 2D database: retrieval rate results obtained with different fuzzy functions.

In both cases, the experiments have consisted in querying all the images from each one of the test sets in order to retrieve the associated image stored in the corresponding database, selected as the representative image of its class in the query. For each input image, the result from the search is a list ordered according to the level of similarity of its signature with the signatures stored in the database. The classifier used in the matching process has been a minimum distance classifier. This list contains the best *n* matches of the query.

RESULTS ANALYSIS

We present two graphics, Figures 2 and 3 for the two-dimensional database, and another one, Figure 4, for the volumetric database. Each graphic compares the best results achieved by the signature described and the different membership functions, as well as the kernel and support values proposed. We have added to the Figures the results presented in [8], achieved using the minimum distance classifier (labeled as "Manhattan").

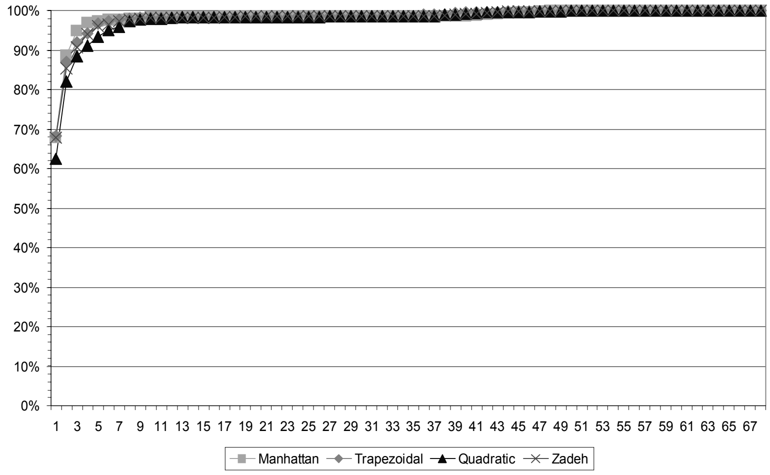

Figure 4.- 3D database: retrieval rate results obtained with different fuzzy functions.

The information represented by each curve of the Figures is the percentage of test set images which are placed up to the given position when a query is made. The abscissa axis shows the position obtained by the arrangement which produces the result of the query. The ordinate axis shows the accumulated frequencies histogram of the percentage of objects placed in a position up to that one pointed by the abscissa value.

The first set of tests was designed in order to obtain the values of kernel and support which give the best results for the trapezoidal membership function and using the addition as the aggregation operator (Fig. 2). The bests results were achieved setting the kernel to 5% and the support to 200% (54.80% of objects classified on the first position). This result is similar to the obtained by the Manhattan distance, although the computational cost of the fuzzy distance is lower.

Although the results achieved by the three membership functions are quite similar, the best results are obtained using the trapezoidal one (Figure 3) as could be expected, since the trapezoidal membership function can be used as a universal approximation of any other membership function. It also has to be noted that the cost of computing this function is the lowest, since the computation of the slope of the other two membership functions is more demanding. In our opinion, the worse behavior of the quadratic membership function for the volumetric database (Fig. 4) is due to the extremely similar images managed. If we reduce the resolution of the image, the disparity between the images diminishes too and this produces a reduction in the success rate since the slope of the quadratic membership function is the most restrictive of the three (see Fig. 1).

CONCLUSIONS AND FUTURE WORK

The use of fuzzy set theory is growing in image processing as it is in all intelligent processing. In this paper we have introduced a fuzzy distance for a CBIR system. A fuzzy membership function captures by definition the fuzziness associated with an image, and this is used to obtain the degree of similarity between the query image an each one of the images of the database. This similarity is calculated using the energy of the image for each resolution level and channel. We have also presented the results of the CBIR system for different values of kernel, support, different type of membership functions and different aggregation operators. Our results conclude that the best solution is obtained with a trapezoidal membership function, the addition as the aggregation operation, and 5% and 200% as kernel and support values respectively. Also, this solution obtains the same success retrieval rate as the Manhattan distance but with a lower computational cost. Future versions of the fuzzy distance will introduce elements for a faster evaluation of the degrees of truth. Also, a more refined fuzzy inference will be introduced by defining membership functions not only for the energy levels of the query image, but also for the database image.

ACKNOWLEDGEMENTS

This work has been partially funded by the Spanish Commission for Science and Technology (grant CICYT TIC99-0947-C02-01) and the Madrid's Autonomous Community (grant CAM 07FT/0014/2001).

REFERENCES

[1] Minerva M. Yeung, Boon-Lock Yeo and Charles A. Bouman, editors. "Proceedings of the SPIE/EI'2000 Symposium on Storage and Retrieval for Media Databases, volume 3972. SPIE, January 2000.

[2] A. Del Bimbo. "Visual Information Retrieval". California: Morgan Kaufmann Publishers, San Francisco, 1999.

[3] Colin C. Venters and Matthew Cooper. "A review of content-based image retrieval systems". Technical report, Manchester Visualization Center. Manchester Computing. University of Manchester, July 2000.

[4] J.-L. Starck, F. Murtagh, "Image Processing and Data Analysis. The Multiscale Approach". Cambridge Press, 1998.

[5] J.M. Prewitt, "Object enhancement and extraction" in Picture Processing and Psychopictropics, B.S. Lipkin and A. Rosenfield (Eds.), Academic Press, New York, 1970, pp. 75-149

[6] X. Chenyang, D.L. Pahm, J.L. Price, "Finding the Brain Cortex Using Fuzzy Segmentation, Isosurfaces, and Deformable Surface Models" in XVth Internationl Conference on Informacion Processing in Medical Images, June 1997, pp 399-404.

[7] T. K. Michaelis, "On Fuzzy Enhancement of Megavoltage Images in Radiation Therapy", on Proc. of the IEEE Conference on Fuzzy Systems, Fuzzy IEEE 97, Barcelona, Spain, pp. 1399-404.

[8] Angel Rodríguez, Oscar D. Robles and Luis Pastor. "New features for Content-Based Image Retrieval using wavelets". In Fernando Muge and Rogério Caldas Pinto Moisés Piedade, editors, V Ibero-american Simposium on Pattern Recognition, SIARP 2000, Lisbon, Portugal, September 2000, pp. 517 – 518.

[9] Ingrid Daubechies. "Ten Lectures on Wavelets", volume 61 of CBMS-NSF Regional Conf. Series in Appl. Math. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992.

[10] S. G. Mallat. "A theory for multiresolution signal decomposition: the wavelet representation". IEEE Trans. on Pattern Analysis and Machine Intelligence, volume 11, number 7, pp. 674 – 693, 1989.

[11] Luis Pastor, Angel Rodríguez and David Ríos. "Wavelets for Object Representation and Recognition in Computer Vision". In Brani Vidaković and Peter Müller (Ed.), Bayesian Inference in Wavelet Based Models, volume 141 of Lectures Notes in Statistics, chapter 17. New York: Springer Verlag, 1999), pp. 267 – 290.

[12] E. J. Stollnitz, T. D. DeRose and D. H. Salesin. "Wavelets for Computer Graphics". San Francisco: Morgan Kauffman Publishers, 1996.

[13] O. D. Robles, A. Rodríguez and M. Luisa Córdoba: "A Study about Multiresolution Primitives for Content-Based Image Retrieval Using Wavelets". In Proceedings on International Conference on Visualization, Imaging and Image Processing, IASTED VIIP 2001, pp. 506 – 511. Marbella, Spain, September 2001.

[14] MIT Media lab. VisTex. Web, 1998. Colour image database.
http://www-white.media.mit.edu/vismod/imaginery/VisionTexture.

[15] A. Rosenfeld, editor. "Multiresolution Image Processing and Analysis", volume 12 of "Springer Series in Information Sciences", Springer Verlag, 1984.

[16] D. Marr and E. Hildreth. "Theory of Edge Detection". In Proceedings of the Royal Society, London, England, volume 207, 1980, pp. 187 – 217.