FCTH: FUZZY COLOR AND TEXTURE HISTOGRAM A LOW LEVEL FEATURE FOR ACCURATE IMAGE RETRIEVAL

Savvas A. Chatzichristofis and Yiannis S. Boutalis
Department of Electrical & Computer Engineering
Democritus University of Thrace
12, Vas. Sofias, 67100 - Xanthi, Greece
schatzic@ee.duth.gr, ybout@ee.duth.gr

Abstract

This paper deals with the extraction of a new low level feature that combines, in one histogram, color and texture information. This feature is named FCTH - Fuzzy Color and Texture Histogram - and results from the combination of 3 fuzzy systems. FCTH size is limited to 72 bytes per image, rendering this descriptor suitable for use in large image databases. The proposed feature is appropriate for accurately retrieving images even in distortion cases such as deformations, noise and smoothing. It is tested on a large number of images selected from proprietary image databases or randomly retrieved from popular search engines. To evaluate the performance of the proposed feature, the averaged normalized modified retrieval rank was used.

An online demo that implements the proposed feature in an image retrieval system is available at: http://orpheus.ee.duth.gr/image_retrieval.

1. Introduction

Content based image retrieval, known as CBIR, undertakes the extraction of several features from images, which, consequently, are used for the retrieval procedure. The visual content of the images is mapped into a new space, named the feature space. The features have to be discriminative and sufficient for the description of the objects. Basically, the key to attain a successful retrieval system is to choose the right descriptors that represent the images as “strong” and unique as possible.

Regarding their type, CBIR systems can be classified in systems that use color information, those that use texture information and finally in systems that use shape information. It is very difficult to achieve satisfactory retrieval results by using only one of these feature categories. Most of the so far proposed retrieval techniques adopt methods in which more than one feature types are involved. For example, color and texture features are used in the QBIC [1], SIMPLIcity [2] and MIRROR [3] image retrieval systems.

A question however that emerged the past years is how these features will become more compact. The characterization of an image with a high dimensional vector may have very good retrieval scores but it significantly delays the retrieval procedure and increases the storage requirements.

The descriptors that were proposed by MPEG-7 [4] [5], for indexing and retrieval, maintain a balance between the size of the feature and the quality of the results. These descriptors appear to be in place to describe well enough the visual content of the image.

This paper proposes a new low level descriptor that includes in one quantized histogram color and texture information. This feature (FCTH) results from the combination of 3 fuzzy units.

Initially the image is segmented in a preset number of blocks. Each block passes successively from all the fuzzy units.

In the first unit, a set of fuzzy rules undertake the extraction of a Fuzzy-Linking histogram [6]. This histogram stems from the HSV color space. Twenty rules are applied in a three-input fuzzy system in order to generate eventually a 10-bin histogram. Each bin corresponds to a preset color.

As second unit, this paper proposes a two-input fuzzy system, in order to expand the 10-bins histogram into 24-bins histogram, importing thus information related to the hue of each color that is presented. The process is described in section 2.

Next, in the third unit, each image block is transformed with Haar Wavelet transform and a set of texture elements are exported. These elements are used as inputs in a third fuzzy system which converts the 24-bins histogram in a 192-bins histogram, importing texture information in the proposed feature. In this unit, eight rules are applied in a three-input fuzzy system and the process is described in section 3 and 4.

With the use of the Gustafson Kessel [7] fuzzy classifier, 8 regions are shaped, which are then used in or-

978-0-7695-3130-4/08 $25.00 © 2008 IEEE
DOI 10.1109/WIAMIS.2008.24

der to quantize the values of the 192 FCTH factors in the interval {0-7}, limiting thus the length of the descriptor in 576 bits per image. The process is described in section 5.

Section 6 deals with a similarity metric that can be used in order to calculate the distance of images according to the proposed feature. Sections 7 comprise the experimental results of an image retrieval system that implements the proposed features and the MPEG-7 descriptors. Finally the conclusions are given in Section 8.

2. Fuzzy Color Segmentation

In [6], a fuzzy system was proposed in order to produce a fuzzy-linking histogram, which regards the three channels of HSV as inputs, and forms a 10 bins histogram as an output. Each bin represents a preset color as follows: (0) Black, (1) Gray, (2) White, (3) Red, (4) Orange, (5) Yellow, (6) Green, (7) Cyan, (8) Blue and (9) Magenta. These colors were selected based on works that had presented in the past [8].

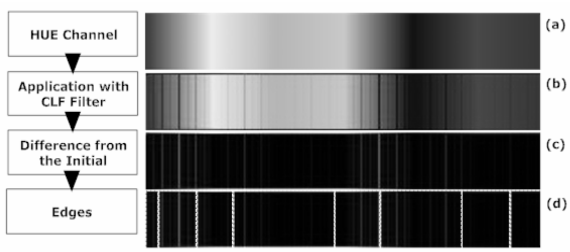

Figure 1: Edges extraction on H channel.

In this paper the work presented in [6] is further improved by recalculating the input membership value limits and resulting to a better mapping in the 10 custom colors.

These new limits are calculated based on the position of the vertical edges of images that represent the channels H (Hue), S (Saturation) and V (Value). Figure 1 shows the vertical edges of the channel H, which were used for determining the position of membership values of Figure 2. The selected hue regions are stressed by dotted lines in figure 1(d). The membership values limits of S and V are identified with the same process.

The use of coordinate logic filters (CLF) [9] was found to be the most appropriate among other edge detection techniques for determining the fine differences and finally extracting these vertical edges. In the procedure followed, each pixel is replaced by the result of the coordinate logic filter "AND" operation on its 3 × 3 neighborhood. The result of this action, stresses the edges of the image. Receiving the difference between the initial and the filtered image, the total of edges is exported.

Based on these edges, the inputs of the system are analyzed as follows:

Hue channel is divided into 8 fuzzy areas. Their borders are shown in figure 2 and are defined as: (0) Red to Orange, (1) Orange, (2) Yellow, (3) Green, (4) Cyan, (5) Blue, (6) Magenta and (7) Blue to Red.

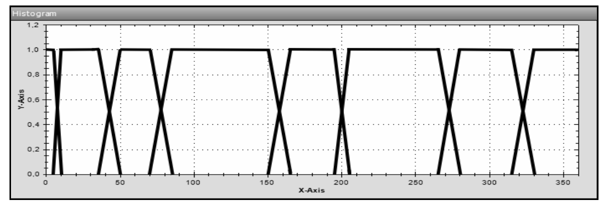

Figure 2: Hue membership functions.

Channel S is divided in 2 fuzzy areas. This channel defines the shade of a color based on white. The first area, in combination with the fuzzy area that is activated in channel V, is used to define if the color is clear enough to be ranked in one of the categories which are described in H histogram, or if it is a shade of white or gray color.

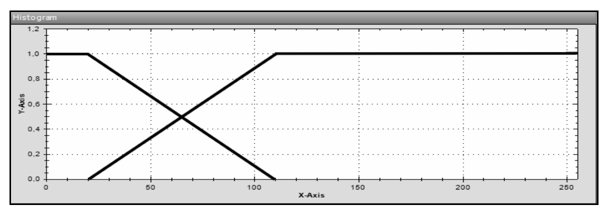

Figure 3: Saturation membership functions.

The third input, channel V, is divided in 3 areas. The first one is actually defining substantially when the input will be black, independently from the values that gives to the other inputs. The second fuzzy area, in combination with the value of channel S gives the gray color.

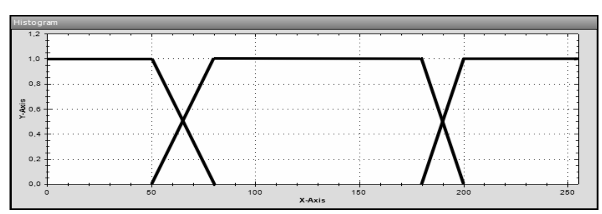

Figure 4: Channel Value membership functions.

A set of 20 TSK-like rules [10] with fuzzy antecedents and crisp consequents have been used. In the consequent part there are actually the variables that count the number of the original image blocks, which are mapped to each specific bin of the 10 bin histogram. Four of the rules depend on two only inputs (S and V). For these rules the decision is independent from the H value.

For the evaluation of the consequent variables two methods have been used. Initially LOM (Largest of Maximum) algorithm was used. This method assigns the input to the output bin which is defined from the rule that gives the greater value of activation. Next a Multi Participate algorithm was tried. This method assigns the input to the output bins which are defined from all the rules that are being activated. Experimental results show that the second algorithm performs better.

Next, a second system undertakes to separate each color in 3 hues. This system forms a 24 bins histogram as an output. Each bin represents a preset color as follows: (0) Black, (1) Grey, (2) White, (3) Dark Red, (4) Red, (5) Light Red, (6) Dark Orange, (7) Orange, (8) Light Orange, (9) Dark Yellow, (10) Yellow, (11) Light Yellow, (12) Dark Green, (13) Green, (14) Light Green, (15) Dark Cyan, (16) Cyan, (17) Light Cyan, (18) Dark Blue, (19) Blue, (20) Light Blue, (21) Dark Magenta, (22) Magenta, (23) Light Magenta.

The design of a system that approaches these shades is based on the determinations of the subtle vertical edges appearing in images with smooth transition from the absolute white to the absolute black through a color. The use of coordinate logic filter (CLF) "AND" [9] was found to be appropriate for determining these vertical edges too.

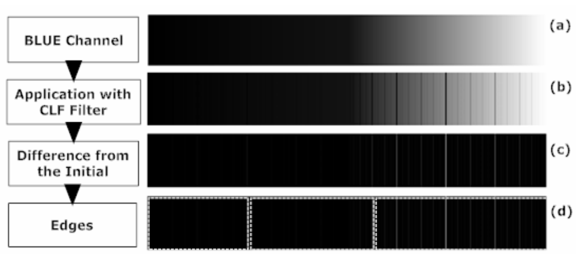

Figure 5: Vertical Edge extraction by using CL Filters.

The values of S and V from each block as well as the value of the bin (or the bins) resulting from the fuzzy 10-bins unit constitute entries in the 24-bins Fuzzy Linking system.

The second system inputs are analyzed as follows. Channel S as well as channel V are divided in 2 fuzzy regions as they appear in figure 6.

This system actually undertakes to classify the input block in one (or more) from the 3 hue areas derived after the vertical edge extraction procedure described above. These hues are labeled as follows: Dark Color (as Color is used the color that attributed by the first 10-Bins system) - Color and Light Color.

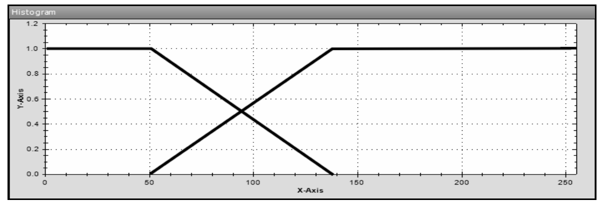

Figure 6: Membership functions of channel S (Saturation) and V (Value).

A set of 4 TSK-like rules [10] with fuzzy antecedents and crisp consequents have been used. For the evaluation of the consequent variables, the Multi Participate method was also used.

3. Fuzzy Texture Segmentation

For exporting texture information from the images, three features that represent energy in high frequency bands of wavelet transforms were used. These elements are the square root of the second order moment of wavelet coefficients in high frequency bands. To obtain these features, the Haar transform applied to the Y (Luminosity - that emanates from the YIQ color space) component of an image block. The derision of the block size depends on the image dimensions and is described in the following section. Suppose for example that the block size is 4 × 4. After a one-level wavelet transform, each block is decomposed into four frequency bands. Each band contains 2 × 2 coefficients. The coefficients in the HL band are $\{C_{kl}, C_{k,l+1}, C_{k+1,l}, C_{k+1,l+1}\}$. One feature is then computed as [2]:

$$f = \left(\frac{1}{4} \sum_{i=0}^{1} \sum_{j=0}^{1} C^{2}{}_{k+i,l+j} \right)^{\frac{1}{2}}$$

The other two features are computed similarly from the LH and HH bands. The motivation for using these features is their reflection of texture properties. Moments of wavelet coefficients in various frequency bands have proven effective for discerning texture. The intuition behind this is that coefficients in different frequency bands signal variations in different directions. For example, the HL band shows activities in the horizontal direction. An image with vertical strips thus has high energy in the HL band and low energy in the LH band. This texture feature is a good compromise between computational complexity and effectiveness [2].

Elements f_{LH}, f_{HL} and f_{HH} are normalized and used as inputs in a fuzzy system, which shape a histogram of 8 bins (areas) as output. These areas are analyzed as follows: (0) Low Energy Linear area, (1) Low Energy Horizontal activation, (2) Low Energy Vertical activation, (3) Low Energy Horizontal and Vertical activation, (4) High Energy Linear area, (5) High Energy Horizontal activation, (6) High Energy Vertical activation, (7) High Energy Horizontal and Vertical activation.

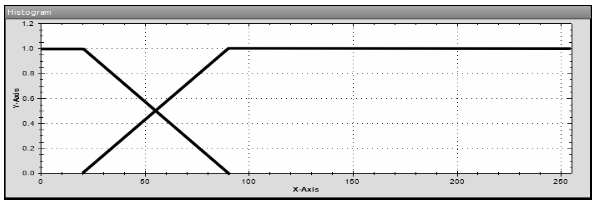

Figure 7: The membership functions of f_{LH} and f_{HL}.

The inputs of the system are analyzed as follows: f_{HL} and f_{LH} are divided into 2 fuzzy areas as shown in figure 7 and f_{HH} is divided into 2 fuzzy areas as shown in figure 8.

A set of 8 TSK-like rules [10] with fuzzy antecedents and crisp consequents have also been used. For the evaluation of the consequent variables, the Multi Participate method was also used.

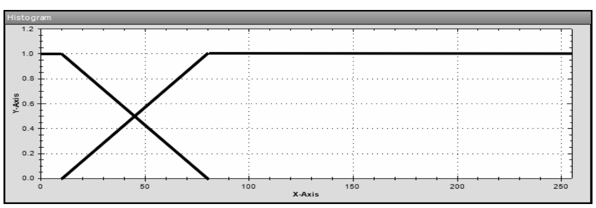

Figure 8: The membership functions of f_{HH}.

4. FCTH – Implementation

The configuration of FCTH is resolved as follows: The histogram is constituted by 8 regions, as these are determined by the fuzzy system that takes decision with regards to the texture of the image. Each region is constituted by 24 individual regions, as these results from the second fuzzy system. Overall, the output that results includes 8 × 24 = 192 bins. Based on the content of the bins the respecting final histogram is produced.

In order to shape the histogram, we firstly dispatch the image in 1600 blocks. This number of blocks was chosen as a compromise between the image detail and the computational demand. Each block passes successively from all the fuzzy systems. If we define the bin that results from the fuzzy system of texture detection as N and as M the bin that results from fuzzy system that shapes fuzzy 24-bins color linking histogram, then each block is placed in the bin position: N × 24 + M. Figure 9 illustrates the whole process, which is described next.

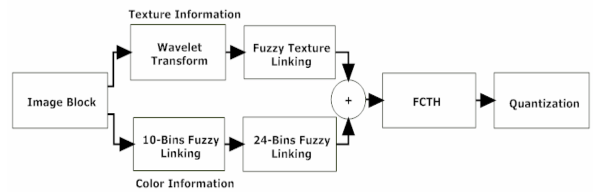

Figure 9: Combination of the three fuzzy systems.

Each block is transported in the YIQ color space and transformed with the Haar Wavelet transform. The f_{LH}, f_{HL} and f_{HH} values are calculated and through the third fuzzy system this block is classified in one of its 8 output bins. Assume, for example, that the classification assigned this block to the second bin that defines low energy horizontal action. Next, the same block is transported in the HSV color space and the mean H, S and V block values are calculated. These values constitute inputs in the fuzzy system that shapes the fuzzy 10-bins histogram described in section 2. Assume again, that the classification assigned this block to the fourth bin that defines that the color is red. Then, the second fuzzy system, using the mean values of S and V as well as the value of the bin (or bins) resulting from the 10-bins fuzzy linking system, calculates the hue of the color and shapes the fuzzy 24-bins histogram. Assume again that the system classifies this block in the fourth bin which defines that color is the dark red. The combination of the 3 fuzzy systems finally will classify the block in the 27^{th} bin (1 × 24 + 3). The process is repeated for all the blocks of the image. At the completion of the process, the histogram is normalized in the interval {0-1} and quantized in 3 bits/bin. The quantization process is described in the following section 5.

5. FCTH Quantization

For the restriction of the FCTH length, a 3bits/bin quantization was used, limiting thus its total length in 192 × 3=576 bits. To calculate the quantization table, a sample of 10000 images was used. Initially, FCTH vectors were calculated for the total of images. The crowd of 10000 ×192 elements constitutes the entry of the fuzzy Gustafson Kessel classifier [7], which separates the crowd of samples in 8 regions. Basically this classification maps the bin values from the decimal area {0-1} to the integer area {0-7}.

Gustafson Kessel parameters were selected to be: Clusters: 8, Repetitions: 2000, e=0.002, m=2. The resulting quantization is given in table 1. The entries of the table have the following meaning. The values of the histogram appearing in bins 0-47 are assigned to one of the values {0-7} according to the minimum distance of each bin value from one of the eight entries in the first row of the table. The same procedure is followed for the entries of bins 48-143 and 144-191 where in this case the eight entries of the second and the third row respectively are used.

Bin 0-47							
13	931	2243	4312	8316	10143	17484	22448
Bin 48-143							
23	1732	3911	6933	7912	9098	16179	18472
Bin 144-191							
18	2373	4145	5391	6912	8198	9179	12472

Table 1: Normalization table (Multiplication by 10^7).

6. Similarity Measure

For the measurement of the distance of FCTH feature between the images we choose to use Tanimoto coefficient [11].

$$T_{ij} = t(x_i, x_j) = \frac{x_i^T x_j}{x_i^T x_i + x_j^T x_j - x_i^T x_j}$$

Where x^T is the transpose vector of x. In the absolute congruence of the vectors the Tanimoto coefficient takes the value 1, while in the maximum deviation the coefficient tends to zero.

7. Experiments

The proposed low level feature has been integrated in the retrieval software system *img(Rummager)*, which has been developed in the "Automatic Control Systems & Robotics" laboratory of "Democritus University of Thrace - Greece".

Initially, experiments were performed in the database of 1000 images that was used by James Wang [2] [12]. FCTH was used in the retrieval procedure and the results are compared with the corresponding results of the following MPEG-7 [4] [5] descriptors:

Color Descriptors: Dominant Color Descriptor (DCD QHDM), Scalable Color Descriptor (SCD), Color Layout Descriptor (CLD), Color Structure Descriptor (CSD).

Texture Descriptors: Edge Histogram Descriptor (EHD), Homogeneous Texture Descriptor (HTD).

To evaluate the performance of the image retrieval system, the objective measure called ANMRR (Averaged Normalized Modified Retrieval Rank) [5] is used. The ANMRR is always in range of 0 to 1 and the smaller the value of this measure is the better the matching quality of the query. ANMRR is the evaluation criterion used in all of the MPEG-7 color core experiments.

A set of ground truth images that are most relevant to query were identified. The ground truth data is a set of visually similar images. In particular experiments we used as ground truth, the groups of images as proposed in the MIRROR image retrieval system [3]. This system separates the Wang's database in 20 queries. In table 2 that follows certain indicative results appear while in table 3 the values of ANMRR for the total of the 20 queries have been calculated.

DESCRIPTOR	327	113	703	600	522
DCD	0.336	0.685	0.828	0.527	0.601
SCD	0.406	0.614	0.211	0.082	0.507
CLD	0.541	0.870	0.252	0.454	0.454
CSD	0.347	0.670	0.145	0.06	0.526
EHD	0.317	0.464	0.306	0.277	0.690
HTD	0.594	0.876	0.615	0.445	0.734
FCTH	**0.253**	**0.451**	**0.119**	**0.022**	**0.353**

Table 2: Query (image number) / NMRR values.

DCD	SCD	CLD	CSD	EHD	HTD	FCTH
0.45	0.36	0.40	0.32	0.51	0.71	**0.27**

Table 3: ANMRR results in Wang's database.

An online demo of the application is available at: *http://orpheus.ee.duth.gr/image_retrieval* and the NMRR values for the MPEG-7 descriptors in Wang's database are available at [13].

Experiments have also been performed on the database that is incorporated in the application *img(Rummager)*. This database has 15000+ images, coming from several retrieval systems and images from private collection. Part of the images (100) comes from a random image selection using random keywords in Google (http//images.google.com).

DCD	SCD	CLD	EHD	FCTH
0.51	0.32	0.38	0.55	**0.25**

Table 4: ANMR results in img(Rummager) database.

Note that DCD, SCD, EHD and CLD implementation in *img(Rummager)* application match the XM implementation [14].

8. Conclusions

This paper presents the extraction of a new low level feature that contains, in one histogram, color and texture information and an extension of this feature so as to incorporate spatial information. This element is intended for use in image retrieval and image indexing systems. Experimental results show that the proposed feature can contribute in accurate image retrieval. Its main functionality is image-to-image matching and its intended use is for still-image retrieval, where an image may consist of either a single rectangular frame or arbitrarily shaped, possibly disconnected, regions. The increase of texture regions would definitely help in the improvement of the results but also in the use of FCTH for semantics image retrieval.

9. References

[1] Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., Flickner, "Query by image and video content: the QBIC system". *Computer 28 (9) P.Y.*, 1995, 23–32.

[2] James Z. Wang, Jia Li and Gio Wiederhold, "SIMPLIcity: Semantics-Sensitive Integrate, Matching for Picture Libraries", *IEEE transactions on pattern analysis and machine intelligence*, Volume 23, no. 9, September 2001.

[3] Ka-Man Wong, Kwok-Wai Cheung and Lai-Man Po, "*MIRROR:* an interactive content based image retrieval system", *Proceedings of IEEE International Symposium on Circuit and Systems 2005*, vol. 2, pp. 1541-1544. 23-26 May 2005, Japan.

[4] Jose M. Martinez, *Mpeg-7 overview* (http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm).

[5] B. S. Manjunath, Jens-Rainer Ohm, Vinod V. Vasudevan, and Akio Yamada, "Color and Texture Descriptors", *IEEE Transactions on Circuits and Systems for Video Technology*, Volume. 11, no. 6, 2001, pp. 703-715.

[6] S. Chatzichristofis and Y. Boutalis, "A Hybrid Scheme for fast and accurate image retrieval based on color descriptors", *IASTED International Conference on Artificial Intelligence and Soft Computing (ASC 2007)*, August, 2007, Palma De Mallorca, Spain.

[7] E.E. Gustafson, W.C. Kessel: *Fuzzy Clustering with a Fuzzy Covariance Matrix*, IEEE CDC, San Diego, California, 761- 766 (1979)

[8] K. Konstantinidis, A. Gasteratos, I. Andreadis, "Image Retrieval Based on Fuzzy Color Histogram Processing", *Optics Communications*, Volume 248, Issues 4-6, 15, pp. 375-386, 2005.

[9] B. Mertzios and K. Tsirikolias, "Coordinate Logic Filters: Theory and Applications" *Nonlinear Image Processing*, Chapter 11, S. Mitra and G. Sicuranza editors, Academic Press, ISBN: 0125004516, 2004.

[10] H.J. Zimmerman, *Fuzzy Sets, Decision Making and Expert Systems* (Kluwer Academic Publ., Boston MA, 1987).

[11] Chi, Z., Yan, H. and Pham, T., Fuzzy Algorithms: With Applications to image processing and pattern recognition, *Advance in fuzzy systems – Applications and theory*, Volume 10, World Scientific, 1996.

[12] Jia Li, James Z. Wang, ``Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1075-1088, 2003.

[13] "*MIRROR:* an interactive content based image retrieval system", http://abacus.ee.cityu.edu.hk

[14] ISO/IEC/JTC1/SC29/WG11: "Description of Core Experiments for MPEG-7 Color/Texture Descriptors, MPEG document N2929", Melbourne (1999)