COLOR IMAGE RETRIEVAL USING FUZZY SIMILARITY MEASURES AND FUZZY PARTITIONS

M. Nachtegael[a], D. Van der Weken[b], V. De Witte[a], S. Schulte[a], T. Mélange[a], E.E. Kerre[a]

[a] Ghent University, Dept. of Applied Mathematics and Computer Science
Fuzziness & Uncertainty Modelling, Krijgslaan 281 (building S9), 9000 Ghent, Belgium
mike.nachtegael@ugent.be
[b] VRT - Research & Innovation, August Reyerslaan 42, 1043 Brussel

ABSTRACT

Color image retrieval is becoming more and more important, and so is the quest for automated and reliable retrieval systems. In this paper we present and illustrate a new color image retrieval system, of which the novelty lies in the use of a fuzzy partition of the HSI color space and fuzzy similarity measures. The system has the advantage that the images don't have to be characterized in advance using several features, and it is quite flexible since the database images are not required to have the same dimensions.

***Index Terms*—** color image, image retrieval, similarity measure

1. INTRODUCTION

The increasing availability of images and the corresponding growth of image databases, and users, make it a challenge to create automated and reliable image retrieval systems. A very extensive overview of content-based image retrieval systems up to 2000 can be found in [15]; some more recent papers include [2, 3, 9, 10, 11, 12, 17]. We consider the situation in which a reference image is available, and that similar images from a database have to be retrieved. A main drawback of a lot of existing systems is that the images are characterized by textual descriptors (describing features like color, texture, morphological properties, and so on), which usually have to be made by a person [4, 5, 7, 16]. In other words, the retrieval is not fully automated. In this paper we will present an image retrieval system that does succeed in extracting images from a database of images without relapsing into the characterization of images by assigning some fundamental properties. This is realized by partitioning the HSI color space into dominant color bins and by using a specific fuzzy similarity measure. The results show that this approach can contribute to the design of performant automated image retrieval systems.

This paper is organized as follows. After some basic notions in Section 2, we discuss the fuzzy partitioned HSI color space and explain how the similarity between two images in an image database is calculated. In Section 4, our approach is illustrated with two experiments. Finally, Section 5 summarizes our conclusions and some remarks regarding future research.

2. BASIC NOTIONS

2.1. Fuzzy sets, fuzzy numbers and similarity measures

A fuzzy set A [19] in a universe X is characterized by a X – $[0,1]$ mapping χ_A, which associates with every element x in X a degree of membership $\chi_A(x)$ of x in the fuzzy set A. In the following, we will denote the degree of membership by $A(x)$ and the class of all fuzzy sets in X as $\mathcal{F}(X)$. The set of all elements that have a non-zero membership degree in a fuzzy set A is called the support of A, i.e., $s(A) = \{x | x \in X \text{ and } A(x) > 0\}$.

A trapezoidal fuzzy number A [8] is a fuzzy set with a trapezoidal-shaped membership function. This trapezoidal membership function usually depends on four scalar parameters $a, b, c,$ and d, as given by:

$$A(x;a,b,c,d) = \begin{cases} 0 & \text{if } x \leq a, \\ \frac{x-a}{b-a} & \text{if } a < x < b, \\ 1 & \text{if } b \leq x \leq c \\ \frac{d-x}{d-c} & \text{if } c < x < d \\ 0 & \text{if } d \leq x \end{cases}$$

In the literature a lot of measures can be found to express the similarity or equality between two fuzzy sets [20]. There is no unique definition, but the most frequently used one is the following: a similarity measure is a fuzzy binary relation in $\mathcal{F}(X)$, with X the universe of grid points of the image, i.e., a mapping $S : \mathcal{F}(X) \times \mathcal{F}(X) \rightarrow [0,1]$ satisfying the following properties: reflexivity ($\forall A : S(A,A) = 1$), symmetry ($\forall A,B : S(A,B) = S(B,A)$), and min-transitivity ($\forall A,B,C : S(A,C) \geq \min(S(A,B), S(B,C))$).

2.2. HSI color space

The RGB color space (Red, Green, Blue) is widely used to represent colors, e.g. on computer screens. However, for color image retrieval purposes it is more convenient to use a color model that characterizes color with one dimension instead of three. Therefore, we prefer the HSI color space.

The HSI color space (Hue, Saturation and Intensity) attempts to produce an intuitive representation of color [14, 13]. Hue is the color as described by wavelength, for instance the distinction between red and yellow. Saturation is the amount of the color that is present. The saturation describes how much a certain color differs from white light, for instance the distinction between red (high saturation) and pink (low saturation). The intensity is the amount of light, for instance the distinction between dark red and light red or between dark grey and light grey. To produce a color we can simply adjust the hue; to change the amount of white light we adjust the saturation, and to make the color darker or lighter we alter the intensity.

The HSI color space can be modelled with cylindrical coordinates. The hue is represented as the angle, varying from $0°$ to $360°$. Saturation corresponds to the radius, varying from 0 to 1 (or from 0 to 255). The intensity varies along the z-axis with 0 being black and 1 (or 255) being white.

3. A NEW IMAGE RETRIEVAL SYSTEM

Suppose that our image database contains n digital color images, and that we have a source image A_0. In order to obtain the most similar image w.r.t. A_0 from the database, we calculate the similarity between the source image A_0 and every database image A_j. The images are then ranked with respect to decreasing similarity, and only the most similar images (e.g. the first 10) are retrieved from the database. So in this case, the calculation of the similarity between two color images is the key to a successful image retrieval system. Therefore, we introduce the fuzzy HSI color space.

3.1. Fuzzy partition of the HSI color space

As discussed before, since the image retrieval efficiency is highly dependent on the colors present in the images, we prefer a space that allows us to characterize a color with only one dimension. Therefore, the HSI color space is very suitable, since the hue component is enough to recognize the color, except when the color is very pale or very somber. In order to perform an extraction based on dominant colors, we limit ourselves to 8 fundamental colors, that are modelled with trapezoidal fuzzy numbers [3]. In that way we obtain a fuzzy partition in the sense of Ruspini [6] of the hue component (see Figure 3.1).

In those cases where there is nearly no color present in the image we will use the intensity component to identify the dominant "color". Also for this component we use a fuzzy partition to model the intensity component (see Figure 3.2).

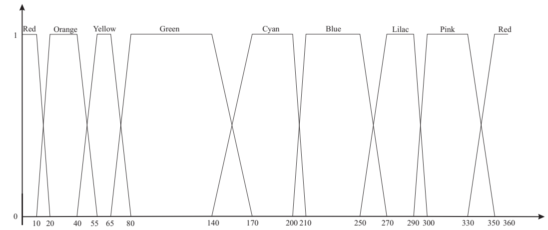

Fig. 3.1: Fuzzy partition of the hue component.

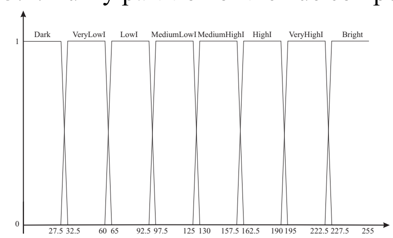

Fig. 3.2: Fuzzy partition of the intensity component.

3.2. Image retrieval using fuzzy similarity measures

3.2.1. Similarity w.r.t. hue

First, we calculate the membership degrees of all the pixels in every image with respect to the fundamental colors modelled by the trapezoidal fuzzy numbers (see Figure 3.1). In that way we obtain 8 new "images".

Secondly, we consider the histogram of each of these 8 images, and normalize these histograms by dividing all the values by the maximum value of each of the histograms. In that way we obtain for each image 8 fuzzy sets, representing the frequency distribution of the membership degrees with respect to the 8 fundamental colors. For an image A_j and a color c, these histograms will be denoted as $h^c_{A_j}$.

To calculate the similarity $S(h^c_{A_0}, h^c_{A_j})$ between two histograms $h^c_{A_0}$ and $h^c_{A_j}$, we use the common fuzzy similarity measure that turned out to be useful for histogram comparison (see [18] for an extensive discussion):

$$S(h^c_{A_0}, h^c_{A_j}) = \frac{|h^c_{A_0} \cap h^c_{A_j}|}{|h^c_{A_0} \cup h^c_{A_j}|} = \frac{\sum\limits_{x \in s_c} \min(h^c_{A_0}(x), h^c_{A_j}(x))}{\sum\limits_{x \in s_c} \max(h^c_{A_0}(x), h^c_{A_j}(x))},$$

where s_c is the support of the fuzzy number representing the color c. This value can be considered as the degree of similarity between A_0 and A_j w.r.t. the color c.

Finally, the similarities between A_0 and A_j with respect to the 8 fundamental colors are merged into one single overall similarity value $S^h(A_0, A_j)$ for the hue component, using the standard average as aggregation operator:

$$S^h(A_0, A_j) = \frac{\sum\limits_{c} S(h^c_{A_0}, h^c_{A_j})}{8}.$$

In those cases where both histograms $h^c_{A_0}$ and $h^c_{A_j}$ only contain values equal to zero, the value $S(h^c_{A_0}, h^c_{A_j})$ will not be taken into account to calculate the average (this means that this color is not present in both images, thus a comparison with respect to this color is not relevant).

3.2.2. Similarity w.r.t. intensity

It is necessary to consider the intensity component because in extreme cases, where there is hardly no color present in the images, black and white will be considered as highly similar. This is of course not satisfactory, and the intensity component will make a distinction between black (intensity equals zero) and white (intensity equals one). As mentioned before, we will use a similar fuzzy partition to model the intensity component (see Figure 3.2).

The procedure to calculate the overall similarity is the same as for the hue component. For an image A_j and an intensity degree d, the histograms will be denoted as $h^d_{A_j}$. The overall similarity value $S^i(A_0, A_j)$ for the intensity component is then given by:

$$S^i(A_0, A_j) = \frac{\sum\limits_c S(h^d_{A_0}, h^d_{A_j})}{8}.$$

3.2.3. Overall similarity and reducing calculations

The overall similarity between the images A_0 and A_j is defined as:

$$S(A_0, A_j) = \frac{S^h(A_0, A_j) + S^i(A_0, A_j)}{2}.$$

Calculating membership degrees of all the pixels with respect to the 8 fundamental colors and 8 degrees of intensity is a rather time-consuming process. However, since we consider the histogram after the calculation of the membership degrees, we can first consider the standard histogram, followed by calculating the membership degree of every bin in the histogram with respect to respectively the 8 fundamental colors and 8 degrees of intensity. The histograms $h^c_{A_0}$ and $h^c_{A_j}$ (or $h^d_{A_0}$ and $h^d_{A_j}$) are then calculated by multiplying the membership degrees of every bin by the value of that specific bin in the standard histogram.

4. EXPERIMENTAL RESULTS

We illustrate the resulting color image retrieval system on a set of synthetic images and on a set of natural images.

4.1. Flag images

In the first experiment we illustrate our proposed method with a database that contains 130 different 24-bit color images of country flags. Since we consider histograms to calculate the similarity between two color images it is not necessary that all the images have the same dimension, which is a big advantage (this is typical for systems that are based on color histograms). The results using the flag of Taiwan as query image are shown in Figure 4.1, where the 10 most similar images are displayed together with the calculated similarity value.

The results show that the nine most similar retrieved images all contain exact the same colors as the query image, which is of course desired. This illustrates the color-sensitivity of the proposed retrieval system.

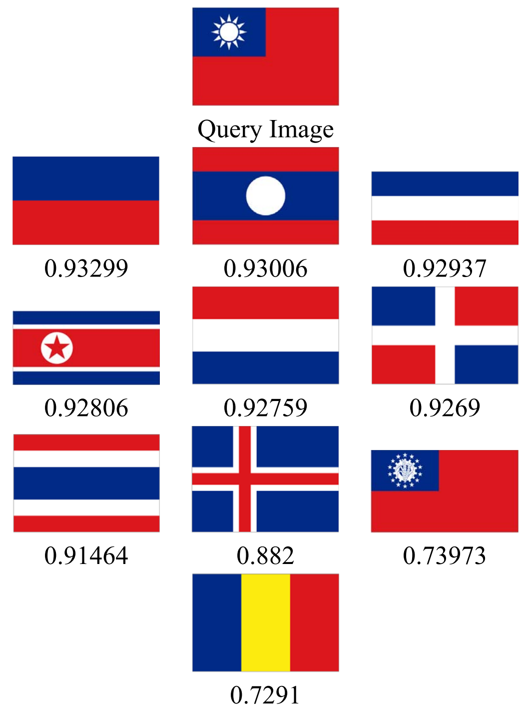

Fig. 4.1: Retrieval result for the flag experiment.

4.2. Natural images

Flag images typically contain not so much dominant colors, which makes their retrieval less difficult. In the second experiment we illustrate our proposed method with a database that contains over 500 natural images (animals, flowers, buildings, cars, texture images, ...). The results using a flower image as query image are shown in Figure 4.2, where the 10 most similar images are displayed together with the numerical result.

The results are quite good: the three most similar retrieved images are flowers in the same color as the one in the query image. The other retrieved images do not contain flowers but have a very similar layout, i.e., they all show an object with a natural background. This illustrates that the proposed approach has potential w.r.t. color image retrieval.

Fig. 4.2: Retrieval result for the natural image experiment.

5. CONCLUSION AND FUTURE WORK

We presented a new color image retrieval system based on a fuzzy partition of the HSI color space. It makes use of a specific fuzzy similarity measure to calculate the similarity between histograms, which are determined by the membership degrees with respect to the fuzzy partitions of the hue and intensity component. To optimize the retrieval process, future research should focus on the use of more advanced aggregation operators [1] (OWA-operators, uninorms, the Sugeno integral, the Choquet integral). This can lead to better results, since the aggregation operators allow to emphasize the importance of a specific element, or to model correlations between different elements that have to be aggregated. Also an extensive comparative study, including other color image retrieval systems, should be made. This will not only lead to a better view on the existing systems and their performance, but may also lead to a new system that combines the best features of the different studied systems.

6. REFERENCES

[1] G. Beliakov, "Definition of general aggregation operators through similarity relations", Fuzzy Sets and Systems, vol. 114, pp. 437-453, 2000.

[2] R. Brunelli, and O. Mich, "Histograms analysis for image retrieval", Pattern Recognition, vol. 34, pp. 1625–1637, 2001.

[3] J. Chamorro-Martnez, J.M. Medina, C. Barranco, E. Galn-Perales, and J.M. Soto-Hidalgo, "An approach to image retrieval on fuzzy object-relational database using dominant color descriptors", in *Proceedings of the 4th Conference of the European Society for Fuzzy Logic and Technology*, EUSFLAT, 2005, pp. 676–684.

[4] S. Chang, "Content-based indexing and retrieval of visual information", IEEE Signal Processing Magazine, vol. 14(4), pp. 45–48, 1997.

[5] A. Del Bimbo, "Visual information retrieval", Morgan Kaufmann Publishers, 2001.

[6] D. Dubois, and H. Prade (editors), "Fundamentals of fuzzy sets", Kluwer Academic Publishers, 2000.

[7] A. Ekin, and A.M. Tekalp, "Robust dominant color region detection and color-based applications for sports video", in *Proceedings of the 2003 International Conference on Image Processing*, 2003, pp. 21–24.

[8] R. Fuller, "On product-sum of triangular fuzzy numbers", Fuzzy Sets and Systems, vol. 41(1), pp. 83–87, 1991.

[9] J. Han, and K.-K. Ma, "Fuzzy color histogram and its use in color image retrieval", IEEE Transactions on Image Processing, vol. 11(8), pp. 944-952, 2002.

[10] Q. Iqbal, and J.K. Aggarwal, "Image retieval via isotropic and anisotropic mappings", Pattern Recognition, vol. 35, pp. 2673–2686, 2002.

[11] G. Lu, and J. Phillips, "Using perceptually weighted histograms for colour-based image retrieval", in *Proceedings of the 4th International Conference on Signal Processing*, 1998, pp. 1150–1153.

[12] J.F. Omhover, M. Detyniecki, M. Rifqi, and B. Bouchon-Meunier, "Ranking invariance between fuzzy similarity measures applied to image retrieval", in *Proceedings of the 2004 IEEE International Conference on Fuzzy Systems*, IEEE, 2004, pp. 1367–1372.

[13] J.C. Russ, "The image processing handbook", CRC Press, 1999.

[14] G. Sharma, "Digital color imaging handbook", CRC Press, 2003.

[15] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, "Content-based image retrieval at the end of the early years", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22(12), pp. 1349-1379, 2000.

[16] P. Stanchev, "Using image mining for image retrieval", in *Proceedings of the IASTED International Conference on Computer Science and Technology*, 2003, pp. 214–218.

[17] P. Stanchev, D. Green, and B. Dimitrov, "High level color similarity retrieval", International Journal of Information Theories & Applications, vol. 10(3), pp. 283–287, 2003.

[18] D. Van der Weken, M. Nachtegael, and E.E Kerre, "Using similarity measures for histogram comparison", Lecture Notes in Artificial Intelligence, vol. 2715, pp. 396–403, 2003.

[19] L.A. Zadeh, "Fuzzy sets", Information Control, vol. 8, pp. 338–353, 1965.

[20] L.A. Zadeh, "Similarity relations and fuzzy orderings", Information Sciences, vol. 3, pp. 177–200, 1971.