Turk J Elec Engin, VOL.13, NO.3 2005, © TÜBİTAK

Color Image Profiling Using Fuzzy Sets

Amine AÏT YOUNES
CReSTIC-MODECO, Université de Reims Champagne Ardenne,
rue des Crayères BP 1035, 51687 Reims cedex 2, France
e-mail: amine@leri.univ-reims.fr
Isis TRUCK
MSH Paris-Nord, Université Paris 8
4 rue de la Croix Faron, Plaine Saint-Denis, 93210 Saint-Denis, France
e-mail: truck@ai.univ-paris8.fr
Herman AKDAG
LIP6, Université Paris 6
8 rue du Capitaine Scott, 75015 Paris, France
e-mail: herman.akdag@lip6.fr

Abstract

In this article, a software for image indexing and retrieval is presented. The classification proposed here is based on the dominant color(s) of the images. The process consists in assigning a colorimetric profile to the image in HLS space (Hue, Lightness, Saturation). First, the definition of hue is done thanks to a fuzzy representation to take into account the non-uniformity of colors distribution. And then, lightness and saturation are represented through linguistic qualifiers also defined in a fuzzy way. Finally, the profile is built through fuzzy functions representing the membership degree of the image to different classes. In order to improve the performances i.e. to define more accurate profiles, we propose to consider zones of pixels, instead of pixels individually. Those zones can be constructed thanks to an edge detection algorithm. A sample of pixels is chosen inside a zone to determine the color of the zone. According to the detected dominant colors, such a software may be used to classify indoor/outdoor images for example, or harmonious/disharmonious images ...

Key Words: *Color image classification, image profiling, dominant colors, fuzzy membership functions.*

1. Introduction

Image retrieval is an important problem that is useful in many fields [1], [2], [3],[4]. In medical applications, it is important to retrieve images in order to help medical expert forecasts, for example. Another example lies in web content detection: Hammami *et al.* classify images in determining whether they contain a lot of skin texture or not in order to detect adult and sexual contents [2].

There are several works on image classification based on the determination of a similarity degree between images. This kind of classification can be done through several technics, for example: statistical approach like Support Vector Machines [3], [5], [6] and fuzzy logic [7], [8], [9]. Barla & al address the problem of classifying images by exploiting color and illumination features, using histograms intersection.

The histogram intersection is used as a kernel function for SVMs and allows one to classify images by similarity of histograms [5]. Another approach is presented by Wang & Du: they propose an algorithm for indexing and retrieving images based on region segmentation, and they also compute similarities between images in order to classify them [10].

The aim of our work is not to make a classical classification but to retrieve images according to their dominant(s) color(s) expressed through linguistic expressions, and in the future, to adapt the classification to the user sensibility, since color perception is often very subjective.

In the field of medical applications, the work we detail in this article can be used to propose a general methodology to classify medical images sets or sequences in order to help medical expert forecasts and analysis, like tumors detection, for example. In industrial applications such as cosmetics it can be interesting to work on skin color to help the make-up manufacture. Another example lies in advertising where our process can help the business man to find more easily and quickly the image that corresponds to his selection criteria.

In the process we propose, profiles are assigned to images and depend on the quantity of pixels or region of pixels that belong to color classes.

The paper is organized as follows: section 2 explains about our choices for color spaces while section 3 is devoted to the problem of color representation where fuzzy membership functions are used. In section 4 we focus on the profile determination for each new entry (image) in the database. Some improvements are proposed in section 5: an edge detection and a division in *zones* can be performed on each image to refine the search for the dominant color. Finally the software we have developed is presented in section 6 with screen captures and section 7 concludes this article.

2. Color spaces

To represent colors, the most common color space is probably the RGB space, where the three primary colors Red, Green and Blue take their values between 0 and 1 or between 0 to 255. The lack of color ("black" color) is symbolized by the triplet (0,0,0). On the other hand the point (255,255,255) corresponds to the maximum of color, i.e. the "white" color. The representation of the colors in this space gives us a cube (cf. figure 1).

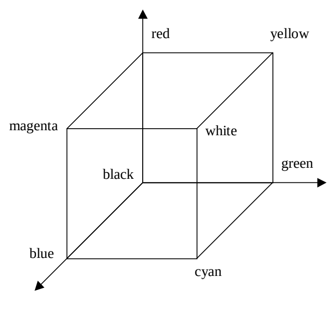

Figure 1. The RGB space.

This space is widely used in color histograms where the pixels are distributed on the three axes R, G and B. Many classification methods compute the similarities between histograms to determine a similarity between images [11].

However with this kind of histogram it is difficult to define a membership degree to a given color, for example how to define an "orange"? Or a "navyblue"?

Such questions lead us to favor another kind of spaces that permit to identify directly the color i.e. with only one dimension, instead of three as with RGB space. HLS space (Hue, Lightness, Saturation) is a space that characterizes the color directly thanks to its hue. Indeed hue is enough to recognize the color, except when the color is very pale or very somber. In this space saturation corresponds to the quantity of "white" in the color and lightness corresponds to the light intensity of the color. Thus, the identification of color is made in two steps : first H, then L, S.

Besides, it is to notice that various models of color representation use also a "two-step" identification of color. For example, Aron Sigfrid Forsius, Pantone Matching System, RAL (*ReichsAusschuß für Lieferbedingungen und Gütesicherung*), Munsel, ISCC-NBS (Inter-Society Color Council - National Bureau of Standards), etc. [12] use at first a color description by means of the hue then a refinement through the saturation and the lightness.

The HLS space can be represented through a cylinder or a bi-cone (cf. figure 2). The hue H is an angle, it means that its definition interval loops (0 and 256 are the same points). The "pure" red ((255,0,0) in RGB space) corresponds to an angle equal to 0 for h, a saturation s equal to 255 and a lightness l equal to 128.

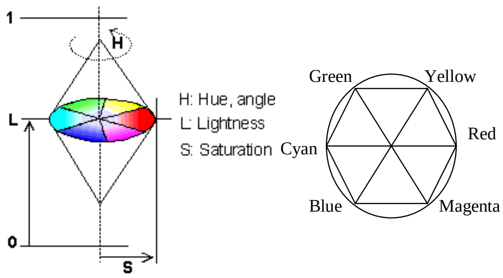

Figure 2. The HLS space.

In our work, we limit ourselves to the nine fundamental colors defined by the set $\mathcal{T}$ representing a good sample of colors (dimension H) :

$\mathcal{T} = \{$*red, orange, yellow, green, cyan, blue, purple, magenta, pink*$\}$

$\mathcal{T}$ corresponds to the seven colors of Newton [13] to which we have added color pink and color cyan, that are included in the rainbow color set.

3. Color Representation

Another important point about spaces is the problem of uniformity of the scale. HLS space is quite convenient for our problem but it is a non UCS (uniform color scale) space [15]. Indeed our eyes don't perceive small variations of hue when color is green ($h = \pm 85$) or blue ($h = \pm 170$) while they perceive it very well with orange ($h = \pm 21$) for example.

Thus to model the fact that the distribution of colors is not uniform on the circle of hues, Truck *et al.* propose to represent them with trapezoidal or triangular fuzzy subsets [16]. Several other works have been done in the field of non uniformly distributed scales: for example, Herrera and Martínez use fuzzy linguistic hierarchies with more or less labels, depending on the desired granularity [14].

Similarly, [16] associate colors with fuzzy sets. Indeed, for each color of $\mathcal{T}$ they built a membership function varying from 0 to 1 (f_t with $t \in \mathcal{T}$). If this function is equal to 1, the corresponding color is a "true color" (cf. figure 3). The linguistic colors definitions used in this process comes from `www.pourpre.com`.

For each fundamental color, the associated interval is defined according to linguistic names of colors. For example to construct f_{yellow}, one can use color "mustard" whose hue is equal to 55 and whose membership to f_{yellow} is equal to ± 0.5.

For some colors, we obtain a wide interval. It is the case for the colors "green" and "blue" which are represented by trapezoidal fuzzy subsets.

We have used in this work the same fuzzy representation of colors, except that we state that two functions representing two successive colors shall have their intersection point value equal to 1/2. It means when h corresponds to an intersection point it can be assigned to both colors with the same weight.

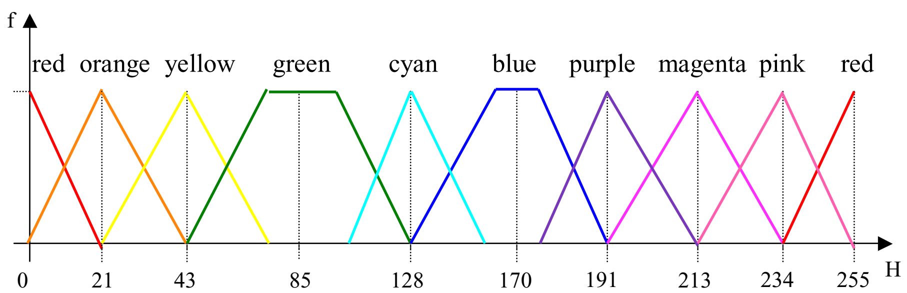

Figure 3. The dimension H.

A trapezoidal fuzzy subset is usually denoted (a, b, α, β) (cf. figure 4) and when the kernel is reduced to one point, it is a triangular subset denoted by (a, α, β) since $a = b$ [17].

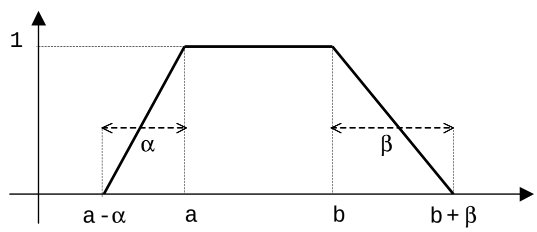

Figure 4. Trapezoidal fuzzy subset.

Let us now define the membership function of any color t :

$$\forall t \in \mathcal{T}\,,\ f_t(h) = \begin{cases} 1 & \mathit{if} \quad h \geq a \\ & \wedge \quad h \leq b \\ 0 & \mathit{if} \quad h \leq a - \alpha \\ & \wedge \quad h \geq b + \beta \\ \dfrac{h - (a - \alpha)}{\alpha} & \mathit{if} \quad h > a - \alpha \\ & \wedge \quad h < a \\ \dfrac{(b + \beta) - h}{\beta} & \mathit{if} \quad h > b \\ & \wedge \quad h < b + \beta \end{cases}$$

For example, for $t = \mathit{orange}$ we have a triangular subset with $(a = 21, \alpha = 21, \beta = 22)$:

$$f_{orange}(h) = \begin{cases} 0 & \mathit{if} \quad h \geq 43 \\ \dfrac{h}{21} & \mathit{if} \quad h < 21 \\ \dfrac{43 - h}{22} & \mathit{if} \quad h \geq 21 \end{cases}$$

For $t = \mathit{green}$ we have a trapezoidal subset with $(a = 75, \alpha = 22, b = 95, \beta = 33)$:

$$f_{green}(h) = \begin{cases} 1 & \mathit{if} \quad h \geq 75 \\ & \wedge \quad h \leq 95 \\ 0 & \mathit{if} \quad h \leq 43 \\ & \wedge \quad h \geq 128 \\ \dfrac{h - 43}{22} & \mathit{if} \quad h > 43 \\ & \wedge \quad h < 75 \\ \dfrac{128 - h}{33} & \mathit{if} \quad h > 95 \\ & \wedge \quad h < 128 \end{cases}$$

To complete the modelization, it is necessary to take into account the two other dimensions (L and S). Each colorimetric qualifier is associated to one or both dimension(s). To facilitate the process, each dimension interval is divided into three sub-intervals: *low value*, *average value* and *strong value*. Thus, we obtain six "one dimension-dependent" qualifiers and nine "two dimension-dependent" qualifiers [18] denoted by Q.

$\mathcal{Q} = \{$ *somber, dark, deep, gray, medium, bright, pale, light, luminous* $\}$.

Figure 5 shows the nine "two dimension-dependent" qualifiers.

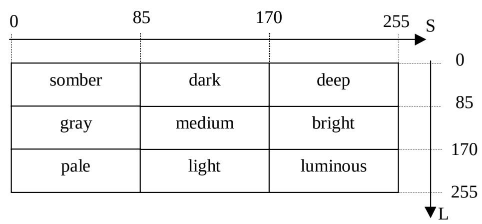

Figure 5. Fundamental color qualifiers.

Each qualifier of $\mathcal{Q}$ is associated to a membership function varying between 0 and 1 ($\tilde{f}_q$ with $q \in \mathcal{Q}$).

As for the hues, the intersection point value of these functions is also supposed equal to 1/2 (cf. figure 7). Thus, every function is represented through the 3 dimension-set $(a, b, c, d, \alpha, \beta, \gamma, \delta)$ (cf. figure 6).

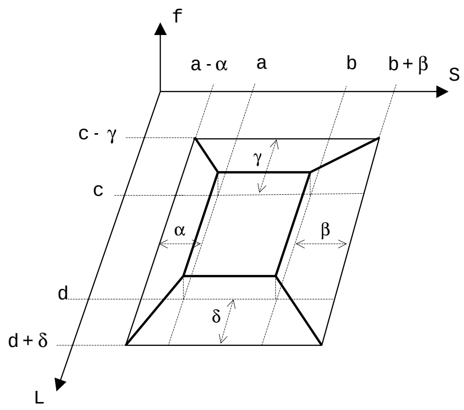

Figure 6. Trapezoidal 3-D fuzzy subset.

The membership function of any qualifier q is defined below :

$$\forall q \in \mathcal{Q},\ \tilde{f}_q(l,s) = \begin{cases} 1 & if \quad a \leq s \leq b \\ & \wedge \quad c \leq l \leq d \\ 0 & if \quad s \leq a-\alpha \vee s \geq b+\beta \\ & \vee \quad l \leq c-\gamma \vee l \geq d+\delta \\ \dfrac{l-(c-\gamma)}{\gamma} & if \quad c-\gamma < l < c \\ & \wedge \quad \alpha l - \gamma s \leq \alpha c - \gamma a \\ & \wedge \quad \beta l + \gamma s \leq \beta c + \gamma b \\ \dfrac{(d+\delta)-l}{\delta} & if \quad d < l < d+\delta \\ & \wedge \quad \beta l - \delta s > \beta d - \delta b \\ & \wedge \quad \alpha l + \delta s > \alpha d + \delta a \\ \dfrac{s-(a-\alpha)}{\alpha} & if \quad a-\alpha < s < a \\ & \wedge \quad \alpha l - \gamma s > \alpha c - \gamma a \\ & \wedge \quad \alpha l + \delta s \leq \alpha d + \delta a \\ \dfrac{(b+\beta)-s}{\beta} & if \quad b < s < b+\beta \\ & \wedge \quad \beta l + \gamma s > \beta c + \gamma b \\ & \wedge \quad \beta l - \delta s \leq \beta d - \delta b \end{cases}$$

For example, for $q = somber$ we have $(a = \alpha = 0, b = 43, \beta = 84, c = \gamma = 0, d = 43, \delta = 84)$:

$$\tilde{f}_{somber}(l,s) = \begin{cases} 1 & \textit{if} \quad s \leq 43 \\ & \wedge \quad l \leq 43 \\ 0 & \textit{if} \quad s \geq 127 \\ & \vee \quad l \geq 127 \\ \dfrac{127-l}{84} & \textit{if} \quad 43 < l < 127 \\ & \wedge \quad l > s \\ \dfrac{127-s}{84} & \textit{if} \quad 43 < s < 127 \\ & \wedge \quad l \leq s \end{cases}$$

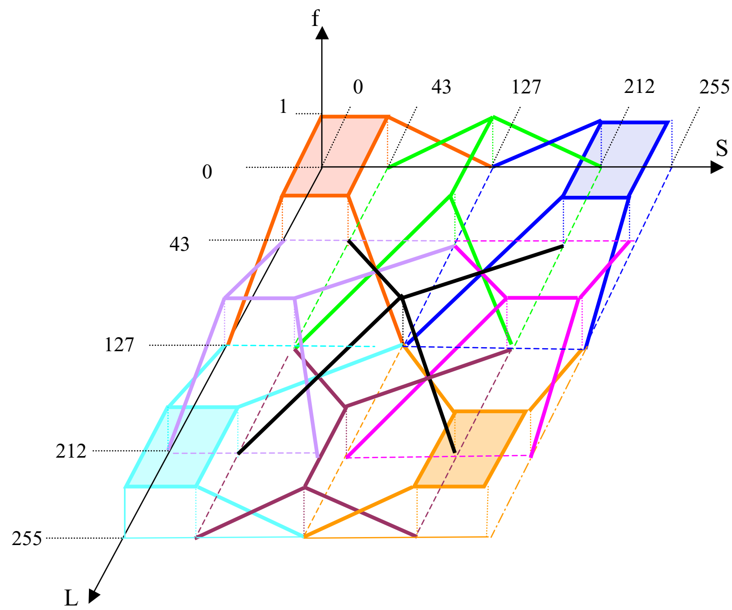

Figure 7. Dimensions L and S.

Of course, black, gray and white colors were also taken into account, i.e. were associated to fuzzy membership functions: f_{black}, f_{white} and f_{gray}. It is to note that these "colors" (we shall call them *non-colors*) are completely defined through the spaces L and S because they don't contain any hue (h is undefined). When l becomes very low, the color turns to white. When l becomes very high, the color turns to black. When s becomes very low, the color turns to gray. Inside the gray color we define three qualifiers: *dark, medium* and *light* that are associated to fuzzy membership functions: $\tilde{f}_{dark}$, $\tilde{f}_{medium}$ and $\tilde{f}_{light}$ (cf. figure 8).

For example, for $t = black$ we have $(a = \alpha = 0, b = 255, \beta = 0, c = \gamma = 0, d = 15, \delta = 10)$:

$$f_{black}(l) = \begin{cases} 1 & \textit{if} \quad l \leq 15 \\ 0 & \textit{if} \quad l \geq 25 \\ \dfrac{25-l}{10} & \textit{if} \quad 15 < l < 25 \end{cases}$$

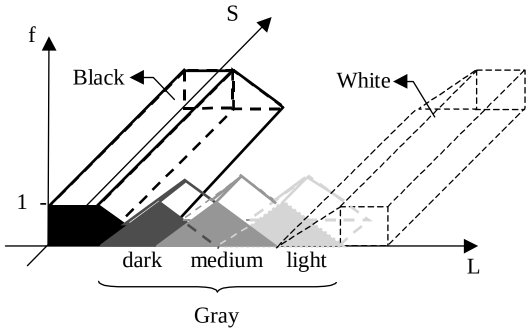

Figure 8. Black, gray and white

4. Image profiling

We just saw the step of data fuzzification (colors and qualifiers) and the step of classes partitioning. These two steps were required before the image processing itself. The aim is now to construct an image profile containing interesting features of the image. These features are related to the nine colors, three non-colors and the nine qualifiers which are now partitioned in fuzzy classes (cf. figure 9). For each pixel of the image we can determine the values taken by the various membership functions of the categories. For each category the value obtained corresponds to the ratio between the sum, on all the pixels of the image, of the membership functions values and the number of pixels, which gives a quantity between 0 and 1. This quantity is the membership degree of an image to the given class.

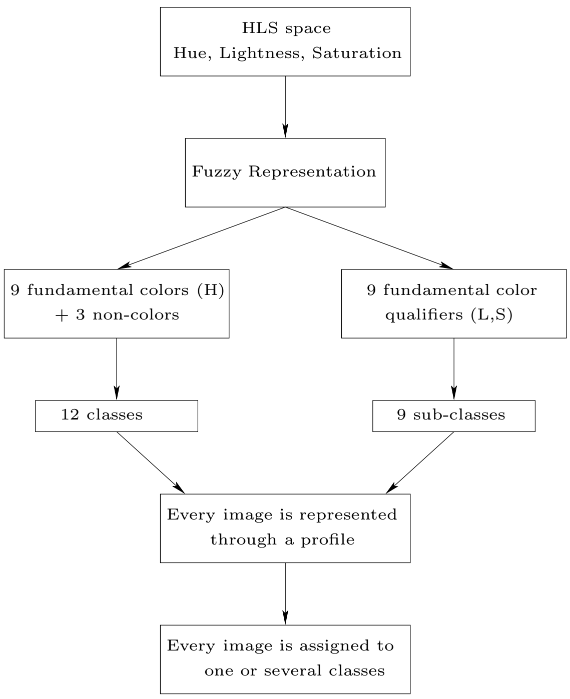

Figure 9. Image processing.

The membership degree of an image to a certain class is defined as follows:

Let I be an image and $\mathcal{P}$ be the set representing the pixels of I. Each element p of the set $\mathcal{P}$ is defined by its color coordinates (h_p, l_p, s_p). p can be one pixel or a set of pixels. We can calculate the functions $f_t(h_p)$, $\tilde{f}_q(l_p, s_p)$ for $t \in \mathcal{T}$ and $q \in \mathcal{Q}$. Let F_t and $\widetilde{F}_{t,q}$ be the following functions, representing the membership degree of I to the classes t and (t, q):

- $\forall t \in \mathcal{T},\ F_t(I) = \dfrac{\sum_{p\in\mathcal{P}} f_t(h_p)}{|\mathcal{P}|}$

- $\forall (t,q) \in \mathcal{T} \times \mathcal{Q}, \widetilde{F}_{t,q}(I) = \dfrac{\sum_{p\in\mathcal{P}} \tilde{f}_q(l_p, s_p) \times g_t(h_p)}{|\mathcal{P}|}$

  with $g_t(h_p) = \begin{cases} 1 & \textit{if } f_t(h_p) \neq 0 \\ 0 & \textit{else} \end{cases}$

Every image is defined by a profile of 96 elements ($|T| + |T \times Q| + |\{black, white, gray\}| + |\{gray\} \times \{dark, medium, light\}| = 9 + 81 + 3 + 3$). A profile can be presented as follows: $[\,F_t(I),\ \widetilde{F}_{t,q}(I)]$

Figure 10 shows the profile of an image and the relationships between the various membership functions. We denote $\mathcal{Q}_1 = \{dark, medium, light\}$, $\mathcal{Q}_2 = \mathcal{Q} \setminus \mathcal{Q}_1$ and $NC = \{black, white, gray\}$, with $q_1 \in \mathcal{Q}_1$, $q_2 \in \mathcal{Q}_2$, $c \in \mathcal{T} \cup \{gray\}$ and $nc \in NC \setminus \{gray\}$, NC standing for "non-colors".

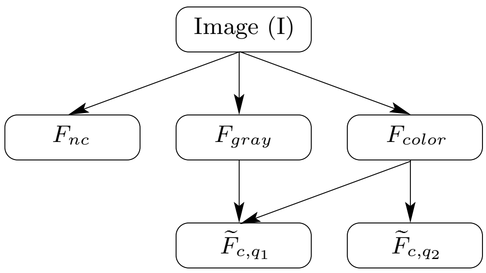

Figure 10. Profile with fuzzy membership functions of an image.

An image can be assigned to several classes, there are 96 classes, 12 principal : C_t with $t \in \mathcal{T} \cup NC$, and 84 subclasses which correspond to a refinement of the research: $\widetilde{C}_{t,q}$ with $(t,q) \in \mathcal{T} \times \mathcal{Q} \cup \{gray\} \times \mathcal{Q}_1$.

As shown in figure 10 the classes can be represented through a tree with father-son relationship, the classes C_t with $t \in \mathcal{T}$ can be considered as fathers and the classes $\widetilde{C}_{t,q}$ with $(t,q) \in \mathcal{T} \times \mathcal{Q}$ as their sons.

Let us denote:

- $F^*(I) = \max_{t\in T}(F_t(I))$

- $\widetilde{F}^*_t(I) = \max_{q\in Q}(\widetilde{F}_{t,q}(I))$ $\forall t \in T$, and for $t = gray$, $q \in \mathcal{Q}_1$

An image I will be assigned to:

- the classes C_t if $F_t(I) \geq F^*(I) - \lambda$, $\forall t \in \mathcal{T} \cup NC$,

  with λ a tolerance threshold.

- the classes $\widetilde{C}_{t,q}$ if $F_t(I) \geq F^*(I) - \lambda$ and $\widetilde{F}_{t,q}(I) \geq \widetilde{F}_t^*(I) - \lambda$, $\forall (t,q) \in \mathcal{T} \times \mathcal{Q} \cup \{gray\} \times \mathcal{Q}_1$.

Thus, an image can be assigned to several classes, and it is forbidden to assign an image to a subclass if it is not already assigned to its father class.

5. Improving the processing

We know that the position of the pixels and/or the digitization and the compression sometimes alter the perception and the nature of the color image. This aspect is not taken into account in the presented processing.

In fact we have identified two main problems that shall be discussed:

1. Pixels around the edges are frequently aberrant (cf. figure 11) [19].

Figure 11. Edge detection and “noise pixels” around an edge

2. We can have aberrant pixels due to the digitization or compression in uniform perceived zones (cf. figure 12).

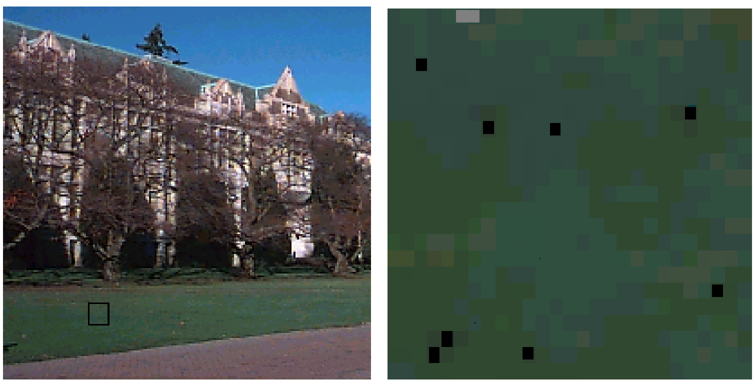

Figure 12. Uniform zone

So we must eliminate all the "edge-pixels" before the processing. The edge detection method we apply is the "canny algorithm" [20] which uses the gradient method. Gradient methods detect the edges by looking for the maximum and minimum in the first derivative of the image, while Laplacian methods search for zero crossings in the second derivative.

There are at least three possibilities to determine all the uniform zones thanks to an edge detection, and then assign to each one a dominant color (cf. figure 13).

- We can compute a sub-zone representing the center of gravity C of the zone and affect as the dominant color of the zone the average color of the pixels belonging to C (cf. figure 13(a)).
- We can choose to randomize two or three sub-zones. The dominant color of the zone is the average of the average colors (cf. figure 13(b)).
- We can randomize a subset of pixels in each zone and the average is computed directly on the pixels. I.e. the dominant color is the average color of the subset (cf. figure 13(c)).

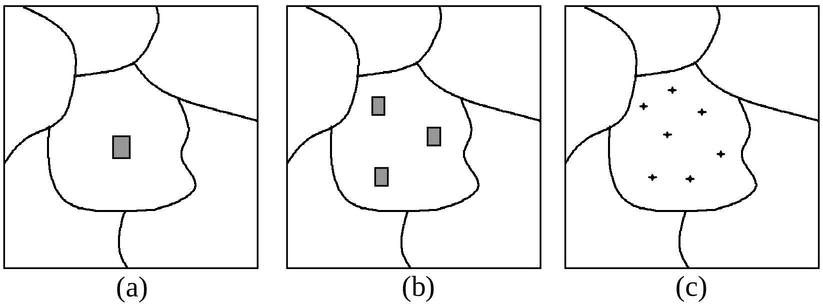

Figure 13. Three different ways to determine the color of a zone

6. The application

6.1. Software architecture

Each image is represented through its previously defined profile and these data are stored in a database (cf. figure 14). It helps us to optimize the exploitation of these information and it permits to make the processing faster. We associate a table for each kind of classes, one representing the 9+3 principal classes (table "Image-hue") and another for the subclasses (table "Image-qualifier"). Two other tables ("Hue", "Qualifier") will facilitate any modification of the sets $\mathcal{T}$ and $\mathcal{Q}$.

The software is divided into two sections, the first one corresponds to the processing (i.e. construction of the image profile) and the insertion of the images in the database while the second one corresponds to the exploitation of this database through requests with linguistic terms (cf. figure 15).

The image processing aims at building its profile. In the first section, one can select and display the image to be inserted in the database. Once the image is inserted the software displays all the stored images.

In the second section, the user of the software has the possibility of carrying out research on two levels. The first one corresponds to the nine fundamental colors (dimension H), the second one to the nine color qualifiers.

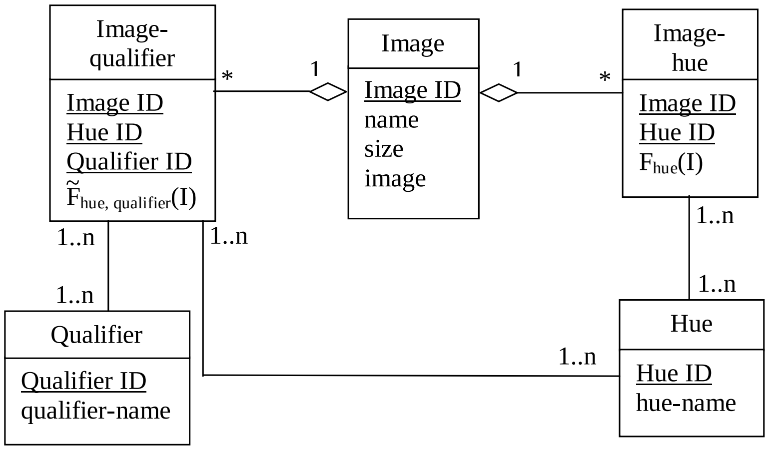

Figure 14. Class diagram of the database.

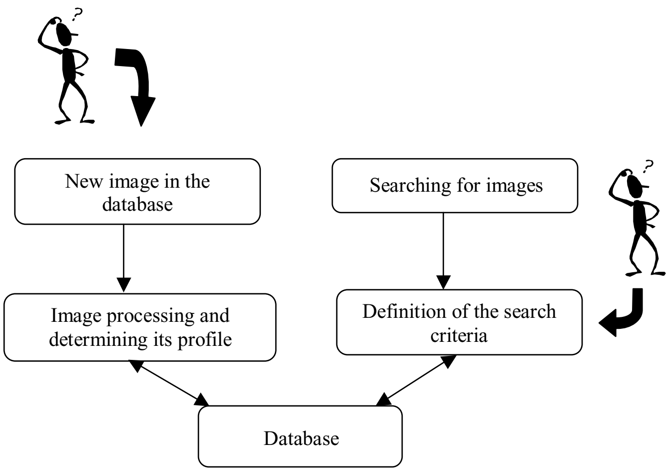

Figure 15. Human-computer interaction.

Indeed, once the Hue is selected, the user has the possibility to refine his request by specifying a color qualifier. For that, it is enough for him to choose one proposed in the list, or to click on the corresponding zone in the image.

Two other kinds of requests are handled: the first one allows us to retrieve B&W and gray-level images (*in this case, H is not considered*), and the second one allows us to retrieve images with more than one dominant color. One-color requests can be successively added (composed) to obtain a multi-color request. For example, figure 16 shows the images whose dominant color is “luminous blue” (left) and the images whose dominant colors are “luminous blue” and “somber green” (right).

The classification obtained with this software is useful at, at least, two levels. In the first level, a classification by dominant color helps to find more easily an image, since color is one of the features that the human brain remembers the most [21]. In the second level, that should be called a “meta–level”, this classification may become more meaningful since it allows a discrimination such as harmonious/non–harmonious images, indoor/outdoor images ... i.e. a discrimination based on color exclusively. Defining a

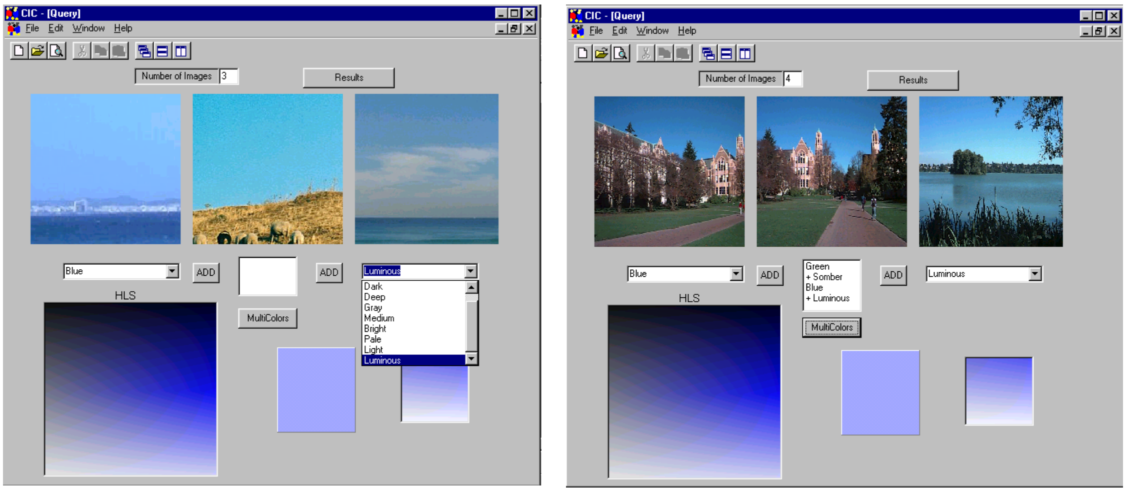

Figure 16. Multi-color query.

set of usual harmonies, such as *2–color*, *3–color*, *mono–color* ... harmonies, it is easy to determine whether an harmony is more or less present in the image or if it is not.

6.2. Results

For the validation of our software we considered two approaches. The first one uses a color image database containing a description of the images by keywords and the second one compares our sofware to an existing online image retrieval system based on the image content (STRICT) [8].

1. First of all, the choice of image database is important to validate the software. We have used a color image database containing a description of the images by keywords, like "the_color": the ImageBank[1]. Two rates were calculated, the first one represents the number of images selected by our software and fitting the query, i.e. the number of images "well-classified" (according to ImageBank), and the second one, the number of images fitting the query but not selected by our software. The former must be high and the latter low.

   On a total of 1000 images we have obtained the following results: 89.7% for the first rate and 16.4% for the second rate (cf. table 1).

# images	# "well-classified" images	# images not selected
1000	897	164
Rates	89.7 %	16.4 %

   Table 1. The validation results

   The reason why 103 images are not "well-classified" and 164 images are "not selected" is due to the subjectivity of the classification of ImageBank (IB) experts. Our software permits to associate more dominant colors to an image than the IB does. Indeed, IB experts associate usually at most two color-keywords. Moreover, the naming of the colors is very subjective: for example, an expert may

[1] `http://creative.gettyimages.com/imagebank/`

call a *dark pink* what another expert may call a *magenta*. This perception also depends a lot on the neighboring colors in the image: e.g. a *yellow* on a black background may not be perceived the same way than a *yellow* on a white background. This problem of perception is handled in section 6.3.

2. We also used an existing image retrieval system based on the image content (STRICT[2]) to validate our software. This system retrieves images according to fuzzy similarity measures of an image query [8].

   We used a free database (Washington-GreenLake), and selected an image image query (cf. figure 17, Image (1)). This image has two dominant colors blue and green. The STRICT system retrieves 20 images according to this image query with a decreasing similarity degree (cf. figure 17).

Figure 17. Images selected by STRICT.

18 of these images are considered as blue and 16 as green by our software (cf. table 2).

Colors	Blue	Green
Images not selected by our software	(14), (18)	(10), (13), (15), (17)

Table 2. Comparison with STRICT

[2]`http://strict.lip6.fr`

Using our software, six images don't fit the STRICT selection. Indeed, the image (14) is considered as "cyan" in our software, which is not "blue". But, as in the first validation approach, this can also be handled through the learning process (cf. 6.3). For the other (five) images the aim of STRICT system is to retrieve images similar to an image query according to the region content. That is why our software doesn't agree with STRICT in these cases.

6.3. Towards a learning process

The handling of the perception would be an interesting improvement of our software. Indeed, the presented method is pixel-dependent while the retrieval must be perception-dependent. We know that each user has his own perception of colors according to his eye sensibility [22].

As in [23], we propose an intelligent learning process for a colorimetric application. We memorize the "meaning" of linguistic expressions associated to the colors (cf. figure 3), according to a certain learning process. The use of a graph where the nodes represent the modified colors and the branches the linguistic modifiers (such as "much more", "a bit less", "a few less", etc.) permits to store the updated knowledge and its associated nuance [23].

For example, the linguistic expression "much more" can be learnt as being "a bit more" in the process if the user wishes to change "much more" into "a bit more".

The number of chosen colors can be changed and it is possible to delete or insert a given color. The definition of the fundamental qualifiers (cf. figure 5) can also be changed.

One can also alter the different membership functions. A recent version of the software proposes the user to modify the function parameters through linguistic modifiers used in [23], "green a little bit more bluish" instead of the initial green.

7. Conclusion

The main originality of our work lies in the proposition of a color image profile according to the fuzzy representation of colors.

A data mining process must include at least two parts: Knowledge Discovery and Assignation Procedure. Knowledge Discovery is a very important step because it allows the representation of the objects to be handled, i.e. to be classified or to be retrieved.

In order to define the dominant color of an image, we assume that we must summarize all pixels according to their values. In our work, these values are originally represented by fuzzy membership functions of nine fundamental colors, three *non-colors* and nine color qualifiers. This procedure permits to define the color image by 21 features. So we introduce by this structure a new representation of knowledge: a color image is represented by a vector, each image feature corresponds to a vector component.

The Assignation Procedure determines in a fuzzy way the class(es) to assign each image. The main idea of our model is to "fuzzify the pixels" in order to represent the image features. The results are very satisfactory for several databases. The learning phase related in the improvements is in test phase.

References

[1] L. Foulloy, Du contrôle symbolique des processus : démarche, outils, exemples, Ph.D. Thesis, Université Paris XI, September 1990.

[2] M. Hammami, L. Chen, D. Zighed and Q. Song, "Définition d'un modèle de peau et son utilisation pour la classification des images", Proceedings of MediaNet'2002, pp. 187-198, Sousse, Tunisia, June 2002.

[3] P. Hong, T. Qi and T. S. Huang, "Incorporate support vector machines to content-based image retrieval with relevance feedback", Proceedings of IEEE International Conference on Image Processing, Vancouver, Canada, 2000.

[4] B. Le Saux, Classification non exclusive et personnalisation par apprentissage : Application à la navigation dans les bases d'images, Ph.D. Thesis, INRIA, France, July 2003.

[5] A. Barla, F. Odone and A. Verri, "Histogram intersection kernel for image classification", Proceedings of the International Conference on Image Processing (ICIP03), Barcelona, september 2003.

[6] V. Vapnik, *Statistical learning theory*, John Wiley and Sons, New York, 1998.

[7] Y. Chen and J. Z. Wang, "A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 9, pp. 1252-1267, 2002.

[8] J.F. Omhover, M. Detyniecki and B. Bouchon-Meunier, "A Region-Similarity-Based Image Retrieval System", Proceedings of IPMU'04, Perugia, Italy, July 2004.

[9] C. Vertan and N. Boujemaa, "Embedding Fuzzy Logic in Content Based Image Retrieval", Proceedings of the 19th International Meeting of the North American Fuzzy Information Processing Society NAFIPS 2000 Atlanta, July 2000.

[10] J. Z. Wang and Y. Du, "Scalable Integrated Region-Based Image Retrieval Using IRM and Statistical Clustering", Proceedings of the ACM and IEEE Joint Conference on Digital Libraries, pp. 268-277, Roanoke, VA, June 2001.

[11] M. J. Swain, D. H. Ballard, "Color indexing", IJCV, vol. 7, no. 1, pp. 11-32, 1999.

[12] J. P Couwenbergh, *Guide complet et pratique de la couleur*, Paris, Eyrolles, 2003.

[13] J. Roire, *Les noms des couleurs*, Pour la science, Hors série, no. 27, 2000.

[14] F. Herrera and L. Martínez, "A model based on lingustic two-tuples for dealing with multigranularity hierarchical linguistic contexts in multiexpert decision-making", IEEE Transactions on Systems, Man and Cybernetics, Part B, 31(2), pp. 227-234, 2001.

[15] I. Truck, Approches symbolique et floue des modificateurs linguistiques et leur lien avec l'agrégation, Ph.D. Thesis, Université de Reims Champagne-Ardenne, France, December 2002.

[16] I. Truck, H. Akdag and A. Borgi, "A Symbolic Approach for Colorimetric Alterations", Proceedings of EUSFLAT 2001, 105-108, Leicester, England, September 2001.

[17] B. Bouchon-Meunier, *La Logique Floue et ses Applications*, Addison-Wesley, 1995.

[18] I. Truck, H. Akdag and A. Borgi, "Using Fuzzy modifiers in Colorimetry", Proceedings of the 5th World Multiconference on Systemics, Cybernetics and Informatics, SCI 2001, pp. 472-477, Orlando, Florida, USA, 2001.

[19] S. Bourghorbel, N. Boujemaa and C. Vertan, “Histogram-based color signatures for image indexing”, Proceedings of Information processing and management of uncertainty in knowledge-based systems, IPMU, 2002.

[20] J. Canny, “A Computational Approach to Edge Detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 6, Nov 1986.

[21] S. Shoemaker, “Colors, Subjective Relations, and Qualia”, Philosophical Issues, no. 7, pp. 55-66, 1996.

[22] C. Boust, H. Chahine, F. Viénot, H. Brettel, M. Ben Chouikha and G. Alquié, “Color Correction Judgements of Digital Images by Experts and Naive Observers”, Proceedings of PICS 2003, pp. 4-9, Rochester, NY, USA, 2003.

[23] I. Truck and H. Akdag, “Supervised Learning using Modifiers: Application in Colorimetrics”, Proceedings of the ACS/IEEE International Conference on Computer Systems and Applications (AICCSA’03), pp. 116 (7 pages), Tunisia, July 2003.