Research of Similar Images Based on Global Descriptors and Multiple Clustering

Omar Bencharef [#1], Brahim Jarmouni[*2], Ali Souissi[*3]
Higher School of Technology-Essaouira , Cadi Ayyad University, Marrakech, Morocco
[1] bencharef98@gmail.com
* GAN, LMA, Faculty of Sciences, Mohammed Fifth University-Agdal, Rabat, Morocco
[2]brahim_jarmouni@yahoo.fr
[3]souissi@fsr.ac.ma

***Abstract*— The development of search engine for similar images stays yet a scientific challenge. Since every image is represented by one or more vectors, the research stage becomes very hard. To raise these difficulties, we suggest inserting a search engine based on multiple representations of images where every image will be represented by three vectors that are color, texture and shape. In the research phase, we applied on the first one clustering on every level of representation for all images stored on the database. This has enabled us to obtain fifteen classes of colors, ten classes of texture and twenty classes of shape. Finally we compute the descriptors associated to the research image and we detect the classes of membership for each level of the representation. Note that the proposed method allowed us to reduce the number of images to check to 7% on average of the database.**

Keywords- Search engine, classes of color, texture and shape, clustering, Global descriptors.

I. Introduction

The researchers in the field of computer vision settle the problem of automatic images indexing by their content. This technique palliates problems posed by the textual research and allows improving the typical applications and contributes to the emergence of new applications in various domains. The idea of making the access to the images data easier is not new, research techniques of images were developed with this effect since the end of the seventies. Among these approaches we find the techniques of research for images based on text description known under the name “Text-based Image Retrieval” (TBIR) [1, 2, 3, 4, 5] which is the oldest approach used until our days. Each image is manually annotated by a set of keywords describing their contents, and also using a management system database to manage the images.

Through text descriptors, images can be organized hierarchically according to themes or semantics, in order to facilitate navigation and research in the database. However, since the automatic generation of text descriptors for a set of images is not feasible, most of these systems require manual annotation of the images [1, 6].

What we propose in this paper is to integrate a search engine for similar images, based on the multiple representations of images and clustering based on K-means algorithm.

On the representation phase, each image is symbolized by three vectors, which represent the visual characteristics of low level, which are the color, by using the compressed histograms of the colors and the texture represented by the algorithm of compression entitled: Segmentation based Fractal Texture Analysis (SFTA) [7] and finally the shape based on Zernike moments.

For the phase of classification, the fact that each image is represented by a dependent vector and not by isolated values, without forgetting the enormous number of image existing on the web, makes the research of the similar images a painful mission compared to the classical research methods. To do this, after having calculated the three vectors of representations for each image of database (DB), we separately applied to each level of representation the clustering algorithm, which automatically divides the images into several classes, for each level of representation (Fig.1).

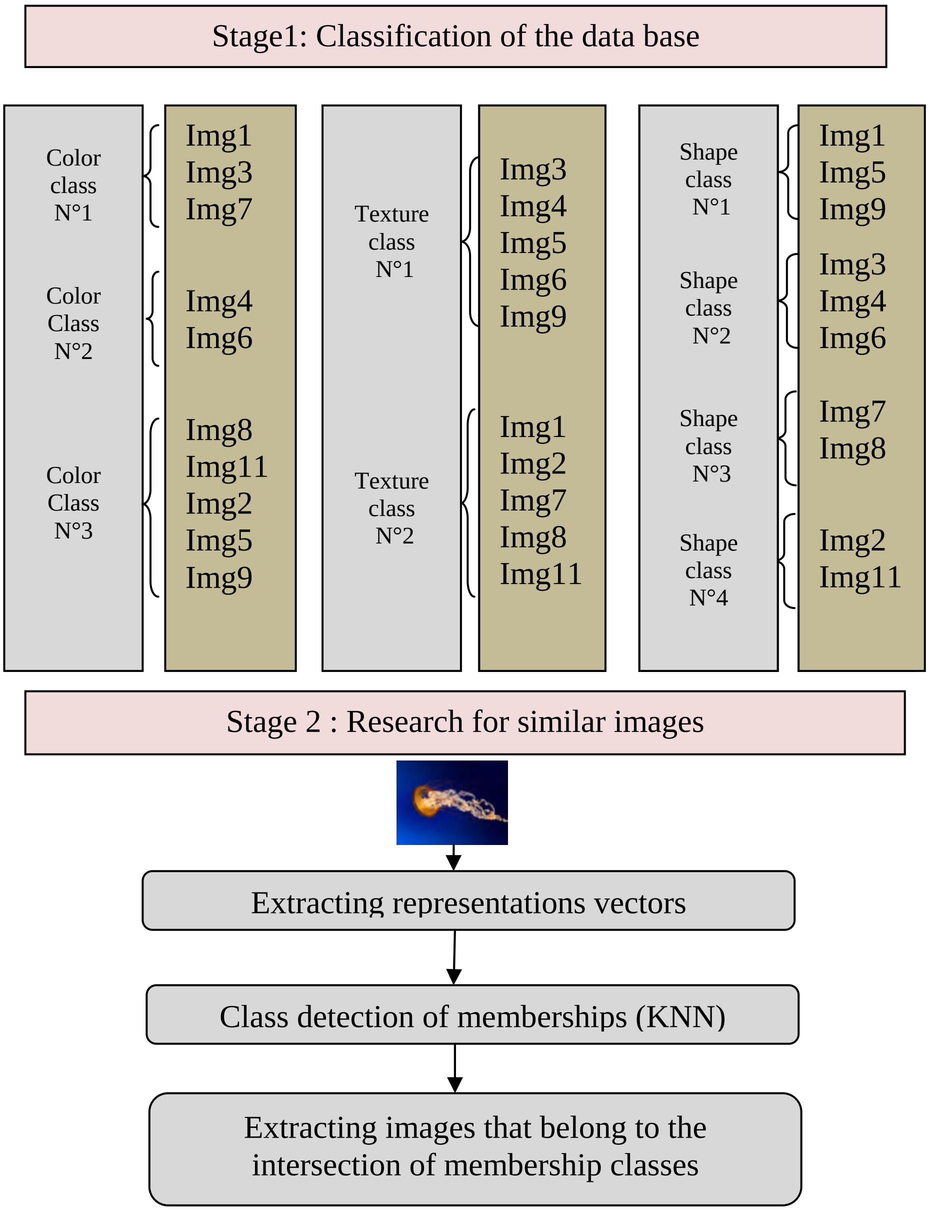

Fig.1: The proposed system for similar images research.

For example, an image which belongs to the color class N= 3, texture class N= 1 and shape class N= 1, in Fig.1, the result will be the images 5 and 9.

II. THE EXTRACTION OF CHARACTERISTICS

It consists of a set of computing, that is mathematical or semantic, to extract reduced vectors, but able to represent the contents of the image. In this section we will present the following three types of descriptors representing respectively color, texture and global shape.

A. Color descriptor

The color is the first descriptor that is used to research for images. There are some studies that have shown that the color is the most effective descriptor [8, 9, 10, 11]. There exist several descriptors of colors, in our case we worked with the histogram of condensed color which is based on the separation of three levels of color (Red, Green, Blue). We consider the following algorithm:

Algorithm: compressed color histogram algorithm.

Require: Image.

Ensure: matrix $S = [C_r\ C_g\ C_b]$

$C_r \leftarrow$ Image(:, :,1) ▷ extraction of the first color component(Read)

$C_g \leftarrow$ Image(:, :,2) ▷ extraction of the second color component(Read)

$C_b \leftarrow$ Image(:, :,3) ▷ extraction of the third color component(Read)

$[l, h] \leftarrow$ size(image)

for $i \leftarrow 1, l$ **do**

  for $i \leftarrow 1, h$ **do**

    one compute the occurrence of each value of pixel between 0 and 255

  end for

end for

▷ one divides the beaches of pixels on 16 fields and one compute the occurrence of each beach

$$C_r(1) \leftarrow \frac{1}{16} * \sum_{n=1}^{16} hr(n,1)$$

$$C_g(1) \leftarrow \frac{1}{16} * \sum_{n=1}^{16} hg(n,1)$$

$$C_b(1) \leftarrow \frac{1}{16} * \sum_{n=1}^{16} hb(n,1)$$

for $k \leftarrow 1, l5$ **do**

$$C_r(k+1) \leftarrow \frac{1}{16} * \sum_{n=16*k}^{16*k+1} hr(n,1)$$ ▷ hr is the red histogram

$$C_g(k+1) \leftarrow \frac{1}{16} * \sum_{n=16*k}^{16*k+1} hg(n,1)$$ ▷ hg is the green histogram

$$C_b(k+1) \leftarrow \frac{1}{16} * \sum_{n=16*k}^{16*k+1} hb(n,1)$$ ▷ hb is the blue histogram

end for

Fig.2 shows examples of the histogram of color for some image of the Data base.

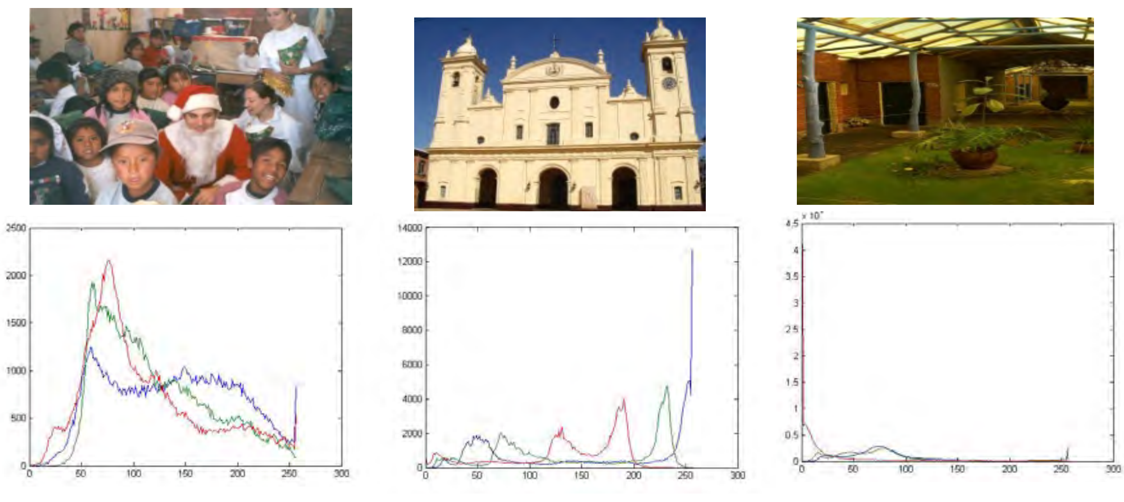

Fig.2 : Projection color histograms of three images from the database.

B. The texture descriptor

A texture is characterized by the repetition of a reason or some elements. There exist several methods to analyze texture. The approach of Tamura [12] described possible textures according to concepts which correspond to human visual perception with six suggested properties: granularity, contrast, prevalent direction, linearity, regularity and roughness. Each one of these parameters is measured to establish a vector of texture. In our case, we used the method of Segmentation of images by the analysis of textures based on the projective fractals SFTA (Segmentation-based Fractal Texture Analysis) [13]. This technique allows synthesizing images very close to the reality. In the analysis of texture, the fractal dimension, which is a measure of the degree of irregularity of an object, describes a certain property of the texture. The fractal model is essentially based on the estimation by spatial methods of the fractal dimension of the surface representing the levels of grey of the image. We consider the following algorithm:

Algorithm: SFTA extraction algorithm.

Require: Grayscale image I and number of thresholds n_t.

Ensure: Feature vector V_{SFTA}.

$T \leftarrow$ MultiLevelOtsu(I, nt)

$T_A \leftarrow \{\{t_i, t_{i+1}\} : t_i, t_{i+1} \in T, i \in [1..\|T\| - 1]\}$

$T_B \leftarrow \{\{t_i, n_l\} t_i \in T, i \in [1..\|T\| - 1]\}$

$i \leftarrow 0$

for $\{\{t_l, t_u\} : \{t_l, t_u\} \in T_A \sqcup T_B\}$ **do**

  $I_b \leftarrow$ TwoThresholdSegmentation(I, t_l, t_u)

  $\Delta(x, y) \leftarrow$ FindBorders(I_b)

  $V_{SFTA}[i] \leftarrow$ BoxCounting(Δ)

  $V_{SFTA}[i] \leftarrow$ MeanGrayLevel(I, I_b)

  $V_{SFTA}[i+1] \leftarrow$ PixelCount(I_b)

  $i \leftarrow i + 3$

end for

return V_{SFTA}

Fig.3 shows the projected values of the texture descriptor for some images of the database.

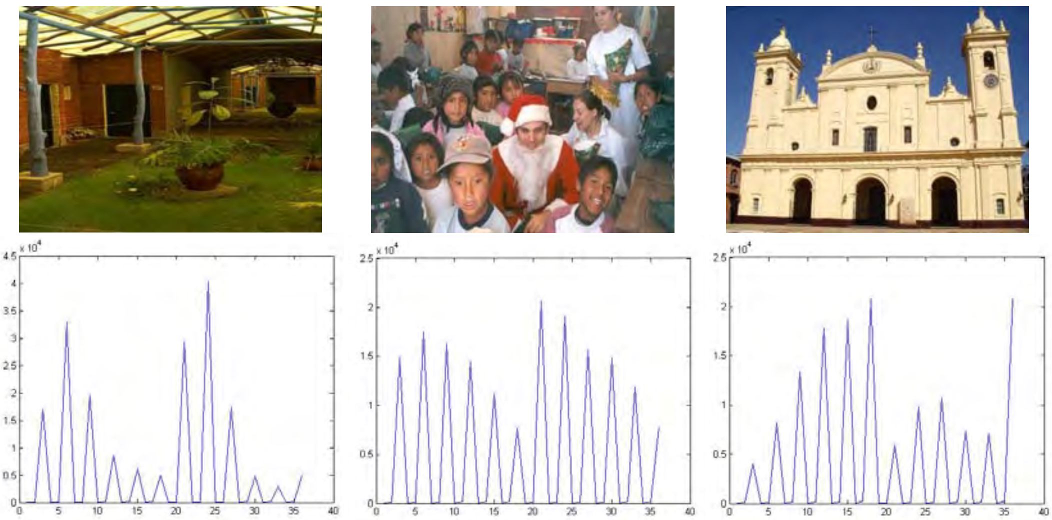

Fig.3: projection texture descriptors (SFTA).

C. Shape descriptors

The shape descriptor that we adopted is based on the time of Zernike. The interest of the computing of the moments of Zernike is to provide a compact and easy coding to compute, equipped with good properties of theoretical invariance by translation, scaling and rotation. They constitute a vector space in which the image of the form is projected [14, 15,16,17].

The geometrical moment of a function f (x, y) is the projection of this function on the space of polynomials generated by $x^p y^q, (p, q) \in \mathbb{N}^2$. Zernike introduces a set of complex polynomials which form a definite orthonormal base inside the unit circle, i.e. such as $x^2 + y^2 \leq 1$. The form of these polynomials is the following one:

$$V_{nm}^*(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho).\exp(jm\theta)$$

Where:

n : A positive integer or zero

m : An integer such that $|m| \leq n$

ρ : Length of the vector between the origin and the pixel (x, y)

θ : The angle between vector ρ and x axis in counter-clockwise direction

R_{nm} : Radial polynomial

$V^*(x, y)$: polynomial complexe projection of $f(x, y)$ on the space of the complex polynomials. Such polynomials are quite orthogonal since:

$$\int_{x^2+y^2\leq 1} \int \left[\mathbf{V}_{nm}^*(\mathbf{x},\mathbf{y})\right].\mathbf{V}_{pq}(\mathbf{x},\mathbf{y})\mathbf{dxdy} = \mathbf{1} \text{ for } (\mathbf{n},\mathbf{m}) = (\mathbf{p},\mathbf{q}) \text{ and } \mathbf{0} \text{ otherwise}$$

The geometrical moment of the Zernike is the projection of a function $f(x, y)$ describing an image space of orthogonal polynomials generated by $V_{nm}^*(\rho, \theta)$:

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x, y).V_{nm}^*(\rho, \theta)dxdy$$

And for identification of the image, the absolute values of Zernike moments are used:

$$|A_{nm}| = \sqrt{Re^2(A_{nm}) + Im^2(A_{nm})}$$

III. K-MEANS CLUSTERING

In this work, we used the algorithm K-means definite by McQueen [18, 19]. It is an algorithm of clustering more known and more used, because of its simplicity of implementation. It partitions data in K clusters. Contrary to other methods known as hierarchical, which create a structure in tree of clusters to describe the groupings, K-means creates only a single level of clusters. The algorithm returns a partition of the data, in which the objects within each cluster are as close as possible to each other as far as possible and objects other clusters. Each cluster in the partition is defined by its objects and its centroid. K-means is an iterative algorithm that minimizes the sum of distances between each object and its cluster centroid. The initial position of centroids determines the final result, so that the centroids should initially be placed as far as possible from each other in order to optimize the algorithm. K-means out exchange the objects of cluster until the sum connot decrease any more. The result is a set of compact and clearly separated clusters, provided that they have chosen the correct value K the number of clusters. The main steps of k-means algorithm are:

Algorithm: K-means algorithm.

Require: Set of N data, denoted by x. Number of desired group, denoted by k

Ensure: A partition of K groups $\{\mathcal{C}_1, \mathcal{C}_1, ...\mathcal{C}_k\}$

Begin

1) Random initialization of the $\mathcal{C}_k$ centers;

repeat

2) Assignment: create a new partition by assigning each object to the group whose center is the closed;

$$x_i \in \mathcal{C}_k \text{ si } \forall j \|x_i - \mu_k\| = \min_j \|x_i - \mu_j\|$$

With μ_k the center of the classe K;

3) Representation: Compute the associated centers to the new partition;

$$\mu_k = \frac{1}{N} \sum_{x_i \in \mathcal{C}_k} x_i$$

until the algorithm converges to a stable partition;

End

IV. Proposed Methods and Experimental Results

To test our method we used the database iaprtc12 [20] it is composed of a variety of 20000 images (fig. 4).

Fig .4: Example of images of the database iparc 12.

A. *Preparation of the training data base*

Each image represented by three vectors: a color vector VC (1, 48), a texture vector VT (1, 36) and a shape vector VF(1,21).

• The color descriptors: all images are collected in a matrix denoted MC (20000,48). (20000 lines per every picture, 48 the number of color components that we have chosen.

• The texture descriptors: all images are collected in a matrix denoted MT (20000,36).

• The shape descriptors: all images are collected in a matrix denoted MF (20000,21).

We applied the algorithm K-means to the matrix MC with k = 15 which allowed us to have 15 clusters. Similarly for the matrix MT with k = 10 (Fig. 5) and the matrix MF with k = 12.

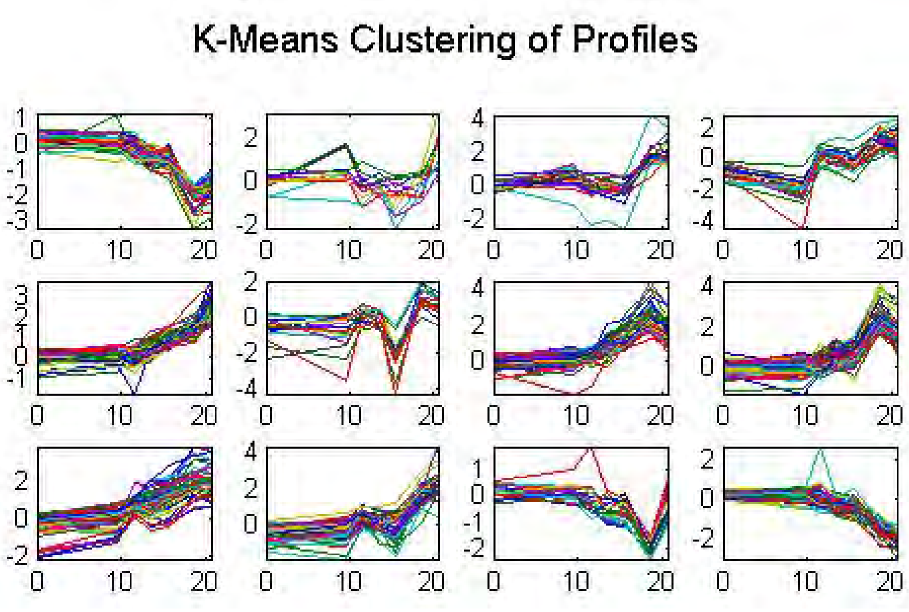

Fig .5: The K-means clustering applied to the texture matrix with k = 10

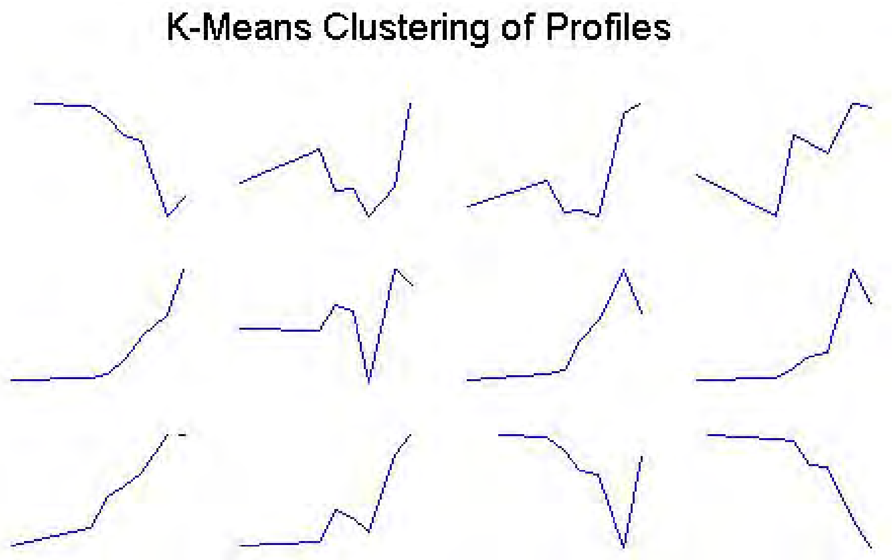

Fig. 6: The average curves of each of the 12 classes of shapes.

B. *Phase of recognition*

The researched images must undergo the same pretreatments as the images of the base of training. Firstly, we compute descriptors of the three levels of representations. Secondly, we define classes of membership by using the KNN algorithm (K nearest neighbors) .

Note that KNN is a simple algorithm that explores all available huts and classifies new huts based on similarity measure.

To estimate the output associated with a new input x, the method of the k nearest neighbors is to consider the k training samples whose entrance is closer to the new input x.

In our case we worked with the Euclidean distance to determine the membership of the wanted image to the different classes that represent the three levels of representation.

The search result is the intersection of the three classes of membership.

The system allows reducing the number of images to check, theoretically, to a maximum of 10% percent of the training set. Experimentally, the number of images to be checked varied between 3% and 9% of the training set according to the nature of the image.

The table 1 introduces a valuation of the rates of recognition of some categories of pictures of the training set.

TABLE. 1: Rate of manual evaluation of some types of images.

Category	**Group1 (Fig.7)**	**Group2 (Fig.8)**	**Group3 (Fig.9)**	**Group4 (Fig10)**	**Group5 (Fig.11)**	**Group6 (Fig.12)**
Manual evaluation of the Request	68%	58%	60%	81%	79%	64%

Fig.7: Example of images of Group1.

Fig.8: Example of images of Group 2.

Fig.9: Example of images of Group 3.

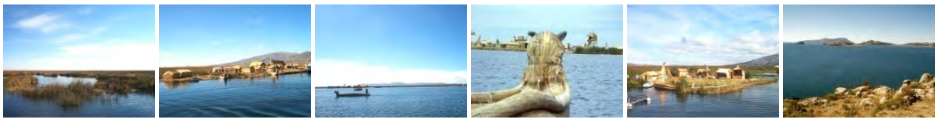

Fig.10: Example of images of Group 4.

Fig.11: Example of images of Group 5.

Fig.12: Example of images of Group 6.

The results of manual valuation stay between 81 % and 85 %, this returns in most cases to the quantity of information that exists in every image.

The use of composed query can overcome these obstacles. What we did:

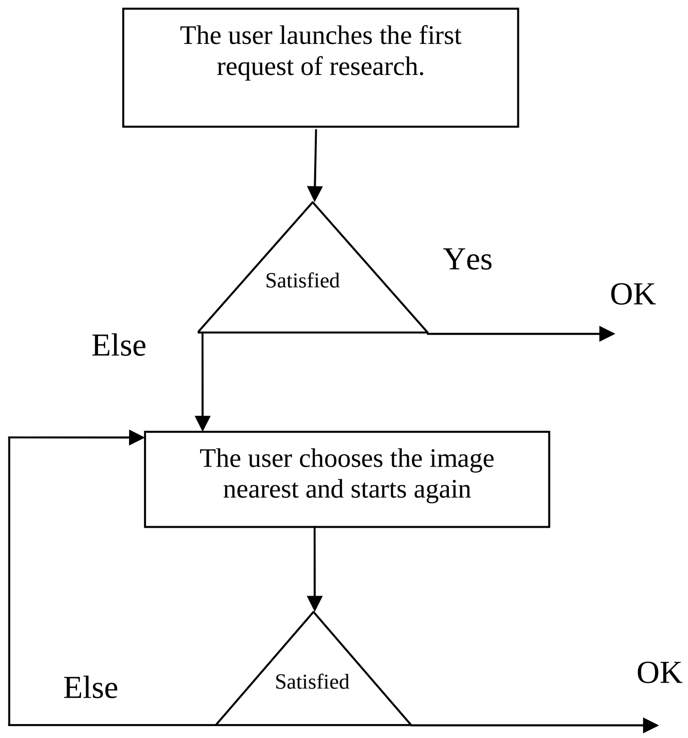

Fig. 13: Diagram of client requests

In the following table we show the rates of evaluation of a double request:

TABLE 2: Rate of manual evaluation, of some categories of images, for a double request.

Category	**Group 1 (Fig.7)**	**Group 2 (Fig.8)**	**Group 3 (Fig.9)**	**Group 4 (Fig10)**	**Group 5 (Fig.11)**	**Group 6 (Fig.12)**
Manual evaluation of the Request	74%	75%	84%	88%	87%	91%

The use of double requests allowed an improvement of recognition rate. We note for the group 6 (Single person) we found that 91% of the results meet the criteria of a direct search by a complete scanning of the KNN.

Finally, we can say that the proposed system presents the same rates of recognitions obtained by a complete sweeping, an image by image from the database, but with a time of research 10 times less.

V. CONCLUSION

In this work, we have presented a search engine for similar images, based on the multiple clustering and global descriptors of images, the proposed system has the same performance as the search engine based on full sweeping or probabilistic methods , but with a smaller execution time.

VI. ACKNOWLEDGEMENTS

The authors would like to thank the editor and the anonymous referees.

This work has been partially supported by Hydrinv Euro-Mediterranean Project, International Laboratories: LERMA and LIRIMA, IMIST, CNRST.

REFERENCES

[1] G. Salton, C. Buckley, Term weighting approaches in automatic text retrieval, Information Processing and Management, 24(5):513-523, 1988.

[2] K. Sparck Jones, C. J. Van Rijsbergen, Progress in documentation, Journal of Documentation, 32:59-75, 1976.

[3] J. J. Rocchio, Relevance feedback in information retrieval, The SMART Retrieval System, Experiments in Automatic Document Processing, pages 313-323.

[4] M. Braschler, C. Peters, Cross-Language Evaluation Forum: Objectives, Results, Achievements, Information Retrieval, 2004.

[5] J. Gobeill, H. Müller, P. Ruch, Translation by Text Categorization: Medical Image Retrieval in ImageCLEFmed 2006, Springer Lecture Notes in Computer Science (LNCS 4730), pages 706-710, 2007.

[6] P. Agouris, J. Carswell , A. Stefanidis, An environment for content-based image retrieval from large spatial databases, Journal of photogrammetry and remote sensing, Vol. 54, No. 4, pp. 263-272, 1999.

[7] W. Y. Ma and B. S. Manjunath, “Texture features and learning similarity,” Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition, pp. 425-430, San Francisco, CA, June 1996.

[8] M. B. Rao, B. P. Rao, A. Govardhan, CTDCIRS: Content based Image Retrieval System based on Dominant Color and Texture Features, International Journal of Computer Applications, Vol. 18, No.6, pp.40-46, 2011.

[9] M. Swain, D. Ballard, Color indexing , International Journal of Computer Vision, 7, pp.11–32, 1991.

[10] K. Sandeep, A.N. Rajagopalan. Human Face Detection in Cluttered Color Images Using Skin Color and Edge Information.

[11] M. H. Yang, N. Ahuja, Detecting Human Faces in Color Images. In Proc. of ICIP'98, vol 1, pp.127-130, 1998.

[12] H. Tamura, S. Mori, T. Yamawaki. Texture features corresponding to visual perception. IEEE transaction on Systems, Man, Cybernetics, vol. SMC-8, no. 6, 460-473, 1978.

[13] A. F. Costa, G. E. Humpire-Mamani, A. J. M. Traina. An Efficient Algorithm for Fractal Analysis of Textures, Conference on Graphics, vol. 1, pp. 39-46, 2012.

[14] M. Petrou , A. Kadrov. Affine invariant features from the trace transform. IEEE Transaction on Pattern Analysis and Machine Intelligence, 26(1) , pp. 30-44, 2004.

[15] F. Ghorbel. A complete invariant description for gray-level images by the harmonic analysis approach , Pattern Recognition Letters, 15, pp. 1043-1051, 1994.

[16] H. K. Kim, J.-D. Kim, D.-G. Sim, D.-I. Oh, A modified Zernike moment shape descriptor invariant to translation, rotation and scale for similarity-based image retrieval, in Proc. IEEE Int. Conf. Multimed. Expo, vol. 1, pp. 307-310, Jul. 2000.

[17] A. Khotanzad , Y. H. Hong, Invariant image recognition by Zernike moments, IEEE Trans. Pattern Anal. Mach. intell, vol. 12, no. 5,pp. 489-497 , May 1990.

[18] B. Mirkin, Clustering for Data Mining: A Data Recovery Approach, Chapman and Hall/CRC, 2005.

[19] G. Celeux, E. Diday., G. Govaert, Y. Lechevallier, H. Ralam-Bondrainy. Classification Automatique des Données. Bordas, Paris, 1989.

[20] M. Grubinger, Analysis and Evaluation of Visual Information Systems Performance. PhD thesis, Victoria University, Melbourne, Australia, 2007