Efficient Content-Based Image Retrieval based on color homogeneous objects segmentation and their spatial relationship characterization

Youssef CHAHIR[†] and Liming CHEN[‡]

[†] *Laboratoire HEUDIASYC UMR CNRS 6599*
Université de Technologie de Compiègne
Centre de Recherches de Royallieu
BP 20529, 60205 Compiègne Cedex, France
E-mail: ychahir@hds.utc.fr

[‡] *Département Mathématiques et Informatique*
Ecole Centrale de Lyon
36, av. Guy de Collongue
BP 163 - 69131 ECULLY Cedex, France
E-mail : liming.chen@ec-lyon.fr

Abstract

In this paper we introduce several techniques which characterize color homogeneous objects and their spatial relationships for a more precise and efficient content-based image searching. We first present a region growing technique for efficient color homogeneous objects segmentation, and then we extend the 2-D string to express spatial signatures for an accurate description of spatial relationships of objects within an image. Several optimizations, including dominant color histogram clustering, have also been proposed to an efficient search engine implementation. The experimental results that we have drawn so far show that our content-based image searching techniques give a high precision while keeping a very good recall rate.

Keywords

Content-based image indexing/retrieval, region growing technique, spatial relationship

1. Introduction

Image has become an important component of today's information system due to the explosive growth of multimedia data as recent advances in storage (DVD) and network (ATM) enable it. Content-based image retrieval (CBIR) is a key technology for easy visual data management as this is now possible with alphanumeric data in large relational databases. In the recent several years, CBIR has received a lot of attention in the research community [i,ii], the main problems with image content-based storage and retrieval concern how and which kinds of image features to extract and how to integrate very different features to satisfy user's queries.

As description of semantic features within an image requires preliminary object segmentation which is a very hard task, the first work for CBIR relies on global visual features, such as color histogram, which can be easily extracted from an image [iii,iv]. If such an approach is simple for effective implementation, and has led to interesting results, its drawback is also well known. Recent work trends to integrate spatial information to different visual features such as color histogram, texture and shapes [v]. In this paper, we focus on content-based visual information searching. We first propose a new region growing based technique to segment color coherent objects from images. Then, we extend 2-D string and incorporate other geometrical features in order to get a precise spatial signature describing the spatial relationships of objects within an image. Finally, several optimizations are introduced for an efficient search engine, which fully explores the spatial relationships as described in images' spatial signatures. The experimental results that we have drawn so far show that our content-based image searching techniques give a high precision while keeping a very good recall rate.

The rest of the paper is organized as follows. Section 2 presents our region growing technique for segmenting color homogeneous objects within an image; Section 3 extends the classical 2-D strings for a precise characterization of objects' spatial relationships; Section 4 introduces our search engine implementation, which is followed by some experimental results. In conclusion, we summarize our work and compare our work to other works.

2. Region growing based homogeneous object Segmentation

For the purpose of content-based image research, the first step is to index images by segmenting and localizing significant objects which are coherent regions according to a visual feature, the color, in the case of our paper. However, the technique we have developed can apply with other visual features such as gray level, texture, etc. In our previous work, we used quadtree based split and merge algorithm to visual homogeneous objects extraction [vi,vii]. In this work, we consider an inverse approach, the bottom up one, which consists of region growing by pixel aggregation. In the following, we first define neighborhood of pixels, which identifies geometrical neighborhood of pixels; then, we introduce color homogeneity, which is used to aggregate neighbor pixels having a close color. Finally, we present our region-growing algorithm for color homogeneous object extraction.

2.1. Neighborhood of Pixels

Let $p(i,j)$ and $q(m,n)$ be two pixels in an image. The distance between p and q is defined as:

$d(p,q)=Max\{|i-m|,|j-n|\}$

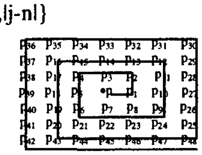


Figure1: The geometrical structure of region growing (N-neighbors)

Let ρ be an integer. We define ρ-neighbors of a pixel $p(i,j)$ to be the all pixels which are situated at ρ pixels from p, i.e.,

$v_\rho(p)=\{\ \forall q \,/\, d_\rho(p,q) \leq \rho\ \}$

As illustrated by the following figure 1, we have:

$v_1(p)=\{p_1,p_2,p_3,p_4,p_5,p_6,p_7,p_8\}$, and the cardinal of $v_1(p)$ is 8 ;

$v_2(p)= v_1(p)\cup\{p_9,p_{10},\dots p_{24}\}$, and the cardinal of $v_2(p)$ is 24 ;

0-7695-0253-9/99 $10.00 © 1999 IEEE

$v_3(p) = v_2(p)(\{p_{25}, p_{26}, \ldots p_{48}\})$, and the cardinal of $v_3(p)$ is 48

Generally, for ρ=n, we have $v_n(p) = v_{n-1}(p)(\{p_x, p_{x+1}, \ldots p_{8*i}\})$, and the associated cardinal is $\sum_{i=1}^{n} 8 * i$

2.2. Color Homogeneity

Let C_1 and C_2 be the colors of pixels p and q. The distance between C_1 and C_2 is defined in RGB space by the Euclidean distance:

$$d_E(C_1, C_2) = \sqrt{(R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2}$$

$C_1=(R_1,G_1,B_1)$, and $C_2=(R_2,G_2,B_2)$. This distance allows defining neighborhood of two colors. In our implementation, we also proceed to a further color clustering for efficiency consideration. Consider a palette of n colors, for instance 512 as this is usually used in many systems. The cross correlation between any couple of two colors of the palette is performed in order to identify the perceptual similarity of these two colors. The result is then sorted to obtain a clustering of close colors. A color representative then represents each class of color issue from such an operation.

The advantage of performing the clustering of colors in a given palette is to reduce the dimension of the image's color vector without losing the search precision. Thus, it helps decrease the computational complexity in the retrieval process.

2.3. Color homogeneous objects extraction by region growing

A color homogeneous object in an image is a coherent area according to the color neighborhood as defined in the previous section. The region growing process consists of gathering neighbor pixels from a starting point on the basis of color homogeneity. More precisely, the process starts from a pixel, and tries to determine whether neighboring pixels belong to the same region of the starting pixel. This process eventually leads to grow a homogeneous region until no more neighboring pixels can be added to the same region on the basis of the color homogeneity.

The following structure characterizes the attributes of a coherent object: a region identifier, its color representative, its Minimum Bounding Rectangle (MBR) , its barycenter, and a set of points .

```
Struct OBJECT {
        int num_reg;
        int color;
        int xi, xs, yi, ys; /* MBR*/
        int x_centroid,y_centroid;
        VectorOfPoint list_pt;
}
```

The figure 2 depicts the region growing procedure. This function builds a region growing by pixel aggregation and returns its characteristics. The process starts with pixel p=(x,y) and continues to grow in the rectangle $R[x_i,y_i,x_s,y_s]$ taking account level neighborhood of pixels until no more pixels can be added to the region. In order to extract all color homogeneous objects, the region growing process has to be repeated for all unvisited pixels in an image.

3. Image spatial signature by Object Spatial relationship characterization

```
OBJECT
Reg_Growing(num_reg,xi,yi,xs,ys,p,level)
begin
 point * neighbor;
 OBJECT Zo;
 VectorOfPoint Tampon,Region;

 Color=p.color
 sx_centroid=sy_centroid=nb_p=0;
 xi=xs=p.i;yi=ys=p.j;
 while (Tampon.empty()=true) do
 begin
   visit[p.i][p.j]=true;
       /* Minimum Bounding Rectangle */
   if (p.i<xi) xi=p.i;if (p.j<yi) yi=p.j;
   if (p.i>xs) xs=p.i;if (p.j>ys) ys=p.j;
       /* Barycenter */
   sx_centroid +=p.i; sy_centroid +=p.j;
   nb_p++;
   Region.push_back(p);
   p=Make_List(Tampon,neigbors,p.color)
   if (Tampon.empty()=false) then
      p= Tampon.back();  Tampon.pop_back();

   endif
 End
 /* Object characteristics */
 Zo.list_pt =Region;
 Zo.xi=xi;Zo.yi=yi;Zo.xs=xs;Zo.ys=ys;
 Zo.num_reg=num_reg; Zo.couleur=color;
 Zo.x_centroid = sx_centroid /nb_p;
 Zo.y_centroid = sy_centroid /nb_p;
 return(Zone);
end
```

Figure 2: Region Growing Function

Object spatial-relationship identification aims at characterizing spatial arrangement relationships among objects, giving the image spatial signature. Including such spatial information in the image representation will not only improve the retrieval quality but also enables users to query an image database based on spatial arrangement of sub-images. The major approaches for describing spatial relationships between objects propose to use a single content descriptor such as 2D-string and its variants [viii,ix]. The major drawback of such a representation is its fuzziness, as its associated spatial operators are not sufficient to give a complete description of the spatial relationships that can exist among objects. For instance, when an object is included in another one and these two objects have the same barycenter, the 2-D string representation can not express this particular spatial relationship. The figure 3 illustrates such fuzziness

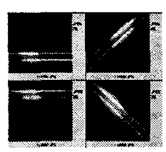


Figure 3: Images with the same 2-D string but they are spatially different

This consideration has led us to extend 2-D string representation with other features, which have been extracted during the object segmentation process. The best representation of spatial relationship among objects would take into account all points of the contour of each object, but such a method would greatly increase the computational complexity as well as the storage requirements.

To fully capture the spatial relationship among objects in an image, we have proposed 2-D-R++ string. First, MBR is used to frame objects in an image. Each object is then localized by a Peano key [6], which gives the spatial position of the center of gravity, associated with the MBR. The basic 2-D string is extended to include for each object its MBR, its barycenter, its relative area and a chain code, which gives more details on the object. An image is also divided into 4 quadrants : 0 (SW), 1 (NW), 2 (SE) and 3 (NE).

Our approach consists of adding a set of new symbols which give information on spatial positions of objects, thus enabling a more precise description of spatial relationships such as "overlap inverse", "contain", "meet", "begin", "end", "overlap", "equal", "and less than", etc. Once these spatial relationships represented, the spatial relationship based image retrieval problem is brought back to a classical 2-D sequence matching. The set of MBRs associated with all characteristics are indexed using an R-tree [x], providing a dynamic structure for rectangles indexing.

As each object in a given picture is framed by a MBR, it has two pairs of begin and end, one for the x-axis and the other for the y-axis. The following features are added to the spatial operators [xi] which are recalled in table 1 :

Notations	Conditions
A<B	end(A)<begin(B)
A=B	begin(A)=begin(B), end(A)=end(B)
A\B	end(A)=begin(B)
A%B	begin(A)<begin(B), end(A)>end(B)
A]B	begin(A)=begin(B), end(A)>end(B)
A[B	begin(A)<begin(B), end(A)=end(B)
A/B	begin(A)<begin(B)< end(A)<end(B)

Table 1: The definition of spatial operators

♦ Each object R with size s is denoted as R_s, where s is the total pixels of the object.

♦ An object R is denoted $_uR$, where u is combination between the quadrants that occupies R in the increasing order.

♦ Operator "<" is extended to include the distance d between two objects A and B, denoted as $A <_d B$, where d=begin(B)-end(A)

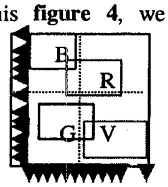
♦ Operator "%" is extended with the two distances (d,d') between objects A and B, denoted as $A \,\%_{d,d'}\, B$, where d=begin(B)-begin(A) and d'=end(A)-end(B)

♦ Operator "]" is enriched with the distance d between objects A and B, denoted as $A \,]_d\, B$, where d=end(A)-end(B)

♦ Operator "[" is extended with distance d between objects A and B, denoted as $A \,[_d\, B$, where d=begin(B)-begin(A)

♦ Operator "\" is parameterized with distances d_1,d_2,d_3 between objects A and B, denoted as $A \setminus_{d1,d2,d3} B$, where d_1=begin(B)-begin(A),d_2=end(A)-begin(B), d_3=end(B)-end(A)

Example: Applying these extensions to this **figure 4**, we illustrate the way we obtain the corresponding 2-D-R++ string describing the spatial relationship among its objects. The image in figure 4 is represented by the following color homogeneous objects : gray rectangle (G), green one (V), blue one (B) and red one (R). According to our previous definition, we obtain the following 2-D-R++-String : ($1B_{35} < 0G_{20} < 0G_{20} < 1B_{35} < 3R_{30} < 1B_{35} < 2V_{24} < 0G_{20} < 2V_{24} < 3R_{30} < 3R_{30} < 2V_{24}$, $2V_{24} < 2V_{24} < 0G_{20} < 0G_{20} < 2V_{24} < 0G_{20} < 3R_{30} < 3R_{30} < 1B_{35} < 3R_{30} < 1B_{35} < 1B_{35}$)



Where $0G_{20}$ indicates that the gray object is located at quadrant 0, and has a relative area 20. The first component of the string is the projection of all objects on X-axis while the second on Y-axis. For instance, $0G_{20} < 1B_{35}$ in the first component indicates that the gray object centroid is at the left of the blue one.

Once obtained a 2-D-R++-string representation, one can build a complete weighted graph to summarize and identify spatial orientation between any couple of objects [xii]. In such a graph, each node represents an object while each vertex a spatial relationship. Figure 5 gives the corresponding graph of the image in figure 4

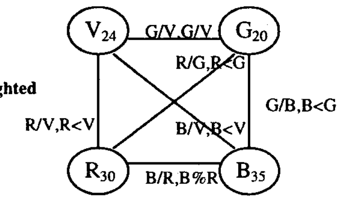


Figure 5: Complete Weighted Graph

4. Search engine Implementation

For searching images into the database, a user gives a query image, which may be selected from the database. This query image is then compared with all images of the database, first on the basis of global color histogram to perform a first filtering, then on the basis of spatial signatures. For instance, on a base of 10 000 images, this operation usually selects down a set of 100 to 200 candidate images. Here we assume that the input of the search engine is the spatial signature of the query image. If the query image comes from the database of images, we already have its spatial signature calculated during its insertion into the base. When the query image is not selected from the database, we first apply our region growing procedure to segment color homogeneous object and determine its spatial signature as explained in the previous two paragraphs.

4.1. Color histogram based filtering

The color histogram is the color distribution in an image, thus is a global visual feature. Its major advantages are its simplicity in computation, and its insensitivity to image rotation and translation. In order to allow comparison of images, we first normalize color histogram, then a similarity is computed between two images, using a similarity metric.

A color histogram is normalized by the formula :

$$H^N(im_n, i) = \frac{H(im_n, i)}{\sum H(im_n, i)}$$

where $H(im_n, i)$ is the histogram of an image im_n, where the index i represents a histogram bin. A similarity distance between two color histograms, $H(im_n)$ and $H(im_m)$, each consisting of n bins, is quantified by the following metric:

$$d_E^2(H(im_n),H(im_m)) = \sum^n \sum^n a_{ij}(H(im_n,i) - H(im_m,i))(H(im_n,j) - H(im_m,j))$$

where the matrix a_{ij} represents the similarity between the colors corresponding to bins i and j, respectively. This matrix needs to be determined from human visual perception studies. Notice that if a_{ij} is the identity matrix, then this measure becomes Euclidean distance.

4.2. Color dominant based optimization

Actually only a small number of colors, formed by the dominant colors of an image, is used to construct an approximate representation of color distribution in our implementation. This simplification does not alter the precision of search result, since only dominant colors are most visually significant, and capture the human attention, as observe Gong et al. [xiii]. Our current implementation makes use of this observation, and limits color classes to the first fifty dominant classes of colors, which correspond to the fifty-histogram bins containing the maximum of pixels. Furthermore, in our experiments, a color histogram bin is ignored if the number of pixels is less than 1% of the image.

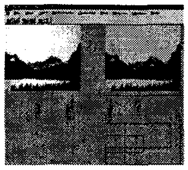


Figure 6: Image optimized segmentation

Figure 6 gives an illustration of the dominant colors based optimization applied to a real image on the left. The image on the right is obtained in keeping only the dominant colors, and the image below gives the MBRs of all extracted objects.

4.3. Spatial signature based images comparison

The color histogram based filtering selects from the database a set of candidate images as results to the query image. The further step consists of using spatial signatures to compute sorted lists of these candidate images as compared to the query image.

The spatial signature based images comparison process has as input a color object table which specifies pour each candidate image and each color the number of homogeneous objects having the corresponding color in that image, as illustrated in Table 2.

	0	1	...	N-1	Votes
Image 1	2	3		2	
Image 2	x_2	y_2		z_2	
...					
Image M	x_m	y_m	...	z_m	

Table 2 General table of objects in images and its votes

For instance, the first line of the table specifies that candidate image 1 has 2 objects of color 0, 3 objects of color 1, ..., and 2 objects of color N-1. Now the query image is also decomposed by the region growing process, and thus has also a mask (x, y,..., z) which specifies that it has x objects of color 0, y objects of color 1, ..., and z objects of color N-1. This mask is compared to each line of the color object table. Each time that a pair of two color objects exists both in the mask and the corresponding line, all spatial relationships between the two colors objects in the query image are compared to those in the candidate image of the corresponding line. A vote is incremented each time there is a matching between the corresponding spatial relationship.

The final result is obtained by sorting candidate images in decreasing order on the basis of the votes obtained by each candidate image.

5. Experimental results

Current implementation of our program runs on the PC platform. The system has been implemented using Visual C++ 4.0 Standard Template Library. The database contains more than 1200 images.

5.1. dominant color histogram based searching

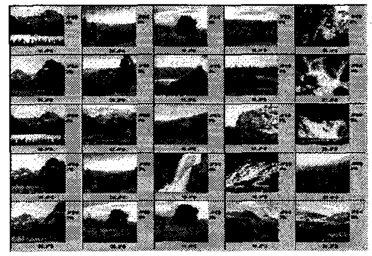


Figure 7: Query result based on color histogram versus dominant color

The first query is to retrieve images based on color histogram. The search process has been made by comparison between histograms of images composed of 50 dominant clusters of color. The figure 7 illustrates some results of retrieved images (25 first similar images). The query image is the one on the top and left corner. The retrieved images are sorted by the column. The result provides us a basis of images whose dominant color distribution is the closest to the query image.

5.2. Spatial signature based searching

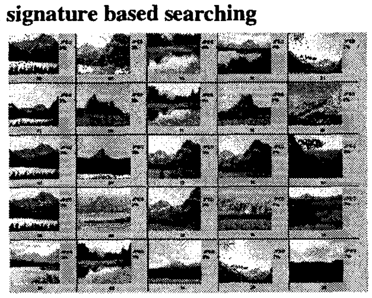


Figure 8: Query result based on spatial arrangement

In this query, we classify the images obtained during the first selection, by their similitude degree, which is represented by its vote. Images which have a maximum of votes are those which respect the spatial arrangement between the objects in the query image. For example, in figure 8, the first 4 images following the query image are those that represent the best disposition of regions as compared to the initial query image, i.e. sky above mountain which is above water.

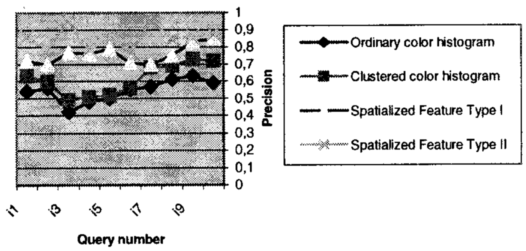
5.3. Recall and precision assessment of the searching engine

To retrieval effectiveness assessment, we calculated two rates in terms of *recall* and precision [xiv]. Recall that the rate of precision is defined as the ratio of relevant images in the answer on the retrieved images, while the rate of recall is defined as the ratio of the relevant images in the answer on all relevant images in the database. Let α the set of retrieved images as response to the query, and β the set of relevant images in the database that are considered to be relevant to the query. Then we have :

Precision = $(\alpha \cap \beta) / \alpha$, and Recall = $(\alpha \cap \beta) / \beta$

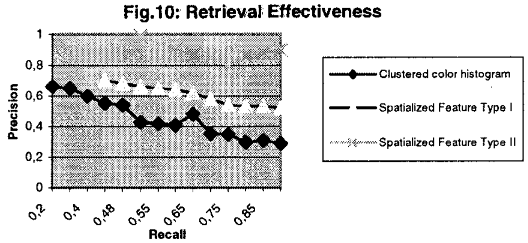
Figure 9 shows several curves, which compare the precision rate of 4 image searching engines. 10 arbitrary query images were used to draw these curves.

Fig.9: Precision comparison



The blue curve corresponds to the precision rate when only color histogram is used to retrieve the most similar images from the base as compared to query images ; Red curve is the precision rate when clustered color histogram is used for the retrieval ; Yellow curve measures the precision rate when spatial relationships among color homogeneous objects are expressed in 2-D-R string (2-D string together with MBR) and used by the search engine ; and finally the green curve gives the precision rate when our 2-D-R++ strings are used to represent spatial relationships of objects and to answer the query. We can see on the curves that the worst precision rate is obtained when only color histogram is used by the searching engine, while the best precision rate, above 0,8 for all 10 query images, is given by the 2-D-R++ string based searching. The use of clustered color histogram improves the color histogram based search, while taking spatial relationships into account gives further improvement on the rate of precision.

Fig.10: Retrieval Effectiveness



Consider figure 10, with criteria of precision and recall rates, it respectively compares searches based on clustered histogram (S1), spatial relationships given by simple 2-D-R string (S2), and spatial features expressed by 2-D-R++-string (S3). As we can see on the figure, the search S1 is not good as compared to precision rate neither to recall rate. When it reaches its best precision rate (little bite above 0,6), it gives on the contrary the worst recall rate. Inversely, when S1 trends to give an interesting recall rate, its loses in precision rate. While S2 improves the result of S1 by the use of spatial relationships, S3, fully exploring objects' spatial relationships, gives the best result both on recall and precision rates. We can see that S3 always gives a precision rate above 0,8, while keeping the recall rate above 0,55 in the worst case.

7. Conclusion

In this paper we introduced several techniques which characterize color homogeneous objects and their spatial relationships for a efficient content-based image searching. We presented a region growing technique for efficient color homogeneous objects segmentation, and extended the 2-D string to accurate description of spatial information and relationships. Several optimizations, including dominant color histogram clustering, have also been proposed to an efficient search engine implementation. The experimental results that we have drawn so far show that our content-based image searching techniques give a high precision while keeping a very good recall rate.

Region growing is a simple and sophisticated method for image segmentation, which was proposed in other context than the content-based image retrieval [xv]. Usually the performance of region growing methods suffers from noise in the image and therefore the segmentation result becomes inaccurate. This fact led us to apply the region growing method to an image divided into quadrants and base our pixel aggregation on dominant colors with 3-neighbor pixels. On the characteristics of a pixel and its neighborhood, several criteria linked to spatial similarity can be used to decide whether a pixel belongs to a homogeneous region. These criteria can be defined from local, or global considerations. In our approach we have used the Neighborhood Homogeneity Criterion of color homogeneity which corresponds to a local comparison between adjacent pixels and its neighbors. Our future work includes the extension of color homogeneous object segmentation to simple shapes, and textures.

8. References

[1] V.N. Gudivada et al., Guest editors, Special issue on Content-Based Image Retrieval Systems, Computer, Vol. 28(9), Sept. 1995

[2] A.D. Narasimhalu, Guest Editor, Special Issue on Content-Based Retrieval , ACM Multimedia Systems, Vol.3(1), Feb.1995

[3] M. Flickner et al. Query by image and video content: The QBIC system. IEEE Computer, 28(9):23-32, September 1995

[4] Alex Pentland et al. "Photobook: Content-based manipulation of image databases. Int.Journal of Comp.Vision, 18(3): 233-254, 1996

[5] Smith J.R., Chang S.F. "Visual SEEK: a fully automated content-based image query system", ACM Multimedia, Nov. 96

[6] Y. Chahir , L. Chen. "Peano key rediscovery for content-based retrieval of images", SPIE, M.S.A S. II, pp. 172-181, November 1997

[7] Y. Chahir, L. Chen. Multi-Spatialized Visual Features Based image retrieval. NTICF' 98, November 98, Rouen France

[8] S.K. Chang et al.. Iconic indexing by 2D Strings. IEEE Transaction on Pattern Analysis and Machine Intelligence, p(3):413-428, 1987

[9] S-Y Lee and F.-J.Hsu ,"Spatial reasoning and similarity retrieval of images using 2D C String knowledge representation", Pattern. Recognition, 25(3):305-318,1992

[10] A. Guttman. "R-trees: A dynamic index structure for spatial searching". In ACM Proc. Int. Conf. Manag. Data , p 47-57, June 1984

[11] P. Huang, "Indexing pictures by key objects for large-scale image databases", Pat. Rec., Vol.30, N. 7 p1229-1237, 1997

[12] V. Gudivada and V. Raghavan,"Design and evaluation of algorithms for image retrieval by spatial similarity", ACM Trans. on Inf. Syst., 13(1): 115-144, April 1995

[13] Y. Gong et al., "An image database system with content capturing and fast image indexing abilities", Proc. Int. Conf. Multimedia Computing and Systems, Boston, Massachusetts, 121-130 1994

[14] C.J. van Rijsbergen, "Retrieval Effectiveness, Information Retrieval Experiment in K.S.Jones", pp.32-43, MA,1981

[15] R. Haralik, L.Shapiro, "Survey: Image Segmentation Techniques". CVGIP, 96, pp.100-132