Combining multiple image descriptions for browsing and retrieval

Youssef Chahir
University of Caen
GREYC – URA CNRS
Department of Computer Sciences
Campus II - Bat. Science III
14032 Caen Cedex, France
Phone : 33 2 31 56 73 94
Fax : 33 2 31 56 73 30
E-mail : chahir@info.unicaen.fr

Saadallah Jomni
University of Bourgogne
LE2I
Department of I.E.M.
9, Av. Alain Savary, B.P 47870
21078 Dijon Cedex, France
Phone : 33 3 80 39 58 93
Fax : 33 3 80 39 58 92
E-mail : jomni@khali.u-bourgogne.fr

ABSTRACT

Retrieving images from large collections using image content is an important problem, in this multimedia age. A quick content-based visual access to the stored image is capital for efficient navigation through image collections. In this paper we introduce several techniques which characterize color homogeneous object and their spatial relationships for efficient content-based image retrieval. We present a region growing technique for efficient color homogeneous objects segmentation and extend the 2-D string to an accurate description of spatial information and relationships. In order to improve content-based image retrieval, our method emphasizes several objectives, such as: automated extraction of localize coherent regions and visual features, development of techniques for fast indexing and retrieval, and querying by both features and spatial information coupled with a symbolic level of image representation. We present our flexible image retrieval system and we give some experimental results.

Keywords: Visual information retrieval, progressive image retrieval, Indexing, Color homogeneous object, Spatial relationships.

1. INTRODUCTION

Image retrieval concerns developing tools for efficient and intelligent browsing of image data. In classical databases systems, such as an on-line library catalog, we search using keywords, authors names or book titles for example. Similarly, in the case of image data, generic attributes, including color, shape and texture, are used for search (Faloutsos et al., 1993). However in last decade, the content-based image retrieval subject has received enormous attention from the computer vision community (Gudivada et al., 1995) (Niblack et al., 1996) (Pentland et al., 1994). Image retrieval systems typically gather and administer large collections of images, and allow their users to retrieve a set of images that are relevant to a given query. The relevance of an image in a collection, to a given query, is automatically computed based on the degree of match between the set of features present in the image, and that present in the query (Manmatha et al., 1998).

The quick visual access to the stored data is essential for efficient navigation through image collections. At the time of visual access to image database, the users have very different objectives. Users' interests may vary from person to another (Rich, 1983). In order to satisfy the information needs of users, it is of vital importance to effectively and efficiently adapt the retrieval process to each user (Minka, 1996).

In order to achieve better quality search results, it is more judicious and interesting to permit multiple access strategies based on different image descriptions. In this paper, we suggest an integrated approach, based on combination of different image representations, and according to individual user preferences, for more flexible image retrieval process. Especially, the user can specify objects in image query, and ask the system more similar one. Based on the spatial relations of visual objects, the System searches the database and returns a list of images ranked by decreasing similarity to the image query.

The rest of the paper is organized as follows. Section 2 gives an overview of image representations. Section 3 presents the different modes of visual access to image database. Section 4 presents our approach for image retrieval. This approach combines different levels of image representations. Concluding remarks are given in section 5.

2. BACKGROUND

To be retrieved by the system, the images should be represented. Several different representations have been proposed for image retrieval systems. These representations can be classified according to the considered level (Berrut et al., 1995):

- *The symbolic level:* at this level, an image is represented by a set of keywords. This representation describes the semantic content of the image. RIVAGE (Halin, 1989) uses this approach for indexing and retrieving images.

- *The spatial level:* deals with the set of objects contained in an image and their features, their absolute and/or relative positions etc.

- *The physical level:* is related to the self-representation of the image (pixel level). Here, the image is indexed by its physical features such as color, shape, etc. contained in image.

The representation levels can be used separately or combined together. For example QBIC (Faloutsos et al., 1993) deals with the symbolic and the physical level; our system deals with the three levels. The image retrieval process is based on the image representation; the following section presents the different access modes to image databases.

3. VISUAL ACCESS TO IMAGE DATABASES

The most important task to be supported by an image retrieval system is to allow the user to search for information by visually accessing an image database. This can be done in three modes, which can be used separately or combined:

i. **Free mode**, or *search by specification*. It may be expressed by a combination of keywords that describe semantic information (such as date of imaging, author, etc.), image attributes (such as color), etc. For example, *retrieve all images taken after September 10, 1998, which show a tree that flourishes in March or September*. It is an information precise request.

ii. **Guided mode**, or *search by example*. In this mode, the system extracts from an example image one or more query parameters (color, shape) and searches for images that have similar parameters. Search results are given as ranked list sorted by decreasing similarity. In this case, the query is not precisely specified. It is an information large request.

iii. **Navigational mode**, or *browsing*. In this case, the user can navigate through proposed pages until desired image (or images). Thus, it is an exploratory mode, where the user will discover the content of image database.

To satisfy the information needs of users, it is of great importance for navigation methods to adapt effectively and efficiently the retrieval process to each user. In the following section we present our system for image retrieval. This system deals with all different levels of image descriptions, and permits all visual access cited above. Especially, the user can, not only precise some keywords that describe semantic information (such as date of imaging, author, etc.), image attributes (such as table, person), but also specify some visual objects in image query, and ask the system more similar one. Based on the spatial relations of visual objects, the System searches the database and returns a list of images ranked by decreasing similarity to the image query.

4. AN IMAGE RETRIEVAL SYSTEM FOR FLEXIBLE DATABASES BROWSING AND RETRIEVAL

1. Spatial Relationships between visual Objects:

Each visual object O is enclosed by its MBR (Minimum Bounding Rectangle). Two pairs of coordinates of the start and the end represent it, one for ordinates axis and other for abscises axis. They define respectively the O_x and O_y intervals.

The spatial relationships are derived from those of Allen relationships (Allen, 1983). The thirteen existing relationships between two intervals can be grouped into seven basic relationships (Before, Meet, Overlap, Start, During, Finish and Equal). These basic relationships are represented by operators (Huang, 1997) as depicted Table 1.

In content-based image retrieval domain, we are interested in flexible search, which takes into account some margin of error from the user and/or the segmentation method. For this reason, we grouped these seven relationships into two basic classes named C_BEFORE and C_DURING.

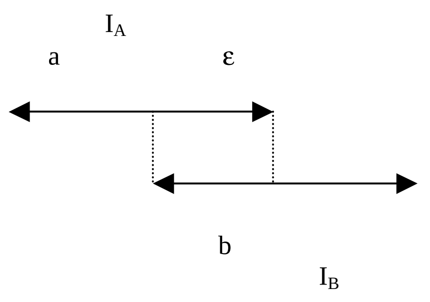
Relationships	Notation	Condition
Before	A<B	End(A)<begin (B)
Meet	A\B	End(A)=begin(B)
During	A%B	Begin (A)< begin (B); end (A)> end (B)
Finish	A[B	Begin (A)< begin (B); end (A)= end (B)
Start	A]B	Begin (A)= begin (B); end (A)> end (B)
Equal	A=B	Begin (A)= begin (B); end (A)= end (B)
Overlap	A/B	Begin (A)< begin (B)< end (A)< end (B)

Table 1 – Definition of spatial operators

Class	Relationship	Notation	Condition
C_BEFORE		Av	end(A)≤ begin(B)
	Before	A<B	end (A)< begin (B)
	Meet	A\B	end (A)= begin (B)
C_DURING		P	begin (A)≤ begin (B); end (A)≥ end (B)
	During	A%B	begin (A)< begin (B); end (A)> end (B)
	Finish	A[B	begin (A)< begin (B); end (A)= end (B)
	Start	A]B	begin (A)= begin (B); end (A)> end (B)
	Equal	A=B	begin (A)= begin (B); end (A)= end (B)
C_BEFORE or C_DURING			
	Overlap	A/B	begin (A)< begin (B)< end (A)< end (B)

Table 2 – Classes of spatial relationships

We grouped, as depicted Table 2, the two relationships Before and Meet in a class C_BEFORE, then the four relationships During, Finish, Start and Equal in a class C_DURING. The remaining relationship Overlap is a transition situation between the class C_BEFORE and C_DURING. In fact, in some cases it can be assimilated to C_BEFORE and in other to C_DURING. For us, we fixed the following rule:
Let's consider two intervals I_A and I_B of two objects A and B:



If $I_A \leq I_B$ then
If $a \geq \varepsilon$ then Overlap $\in$ C_BEFORE
else Overlap $\in$ C_DURING
End
Else
If $b \leq \varepsilon$ then Overlap $\in$ C_BEFORE
else Overlap $\in$ C_DURING
End
End;

Figure 1 - "Overlap" transition between the two classes C_BEFORE and C_DURING

Thus, by combining of these relationships, we can characterize topological relationships as Left, Right, Below, Under, Include and Contain, as well as direction spatial relationships like East, North-East, North, North-West, South-West, South and South-West. The Table 3 recaps the definition of these different basic relationships.

Relationship	**Notation**	**Condition**
	Basic Relationships	
Left	A G B	A_x Av B_x
Below	A B B	A_y Av B_y
Contain	A ⊃B	A_x P B_x ; A_y P B_y
South	A ↓ B	{A_x P B_x or B_x P A_x};A B B
West	A ← B	A G B; {A_y P B_y or B_y P A_y}
South-West	A ↙ B	A G B ; A B B
North-West	A ↖ B	A G B ; B B A
	Inverted Relationships	
Right	A D B ≡A -G B	B_x Av A_x
Above	A H B ≡A -B B	B_y Av A_y
Include	A ⊂ B ≡ A ⊃ B	B_x P A_x ; B_y P A_y
North	A ↑ B ≡ A -↓ B	{A_x P B_x or B_x P A_x};B B A
East	A → B ≡ A -← B	B G A; {A_y P B_y or B_y P A_y}
North-East	A ↗ B ≡ A -↙ B	B G A ; B B A
South-East	A ↘ B ≡ A -↖ B	B G A ; A B B

Table 3 - Definition of the basic spatial relationships

As shown in Table 3, there are some basic relationships and their inverted relationships. For example, by opposition, A ‘in the Right of’ B is equivalent to B ‘in the Left of’ B. similarly for the relationships Under, South, West, NorthEast, and SouthEast. The figure 2 shows the image cutting up in areas and characterizes the eight direction spatial relationships.

NW	**N**	**NE**
W		**E**
SW	**S**	**SE**

Figure 2 – Cutting up in areas and spatial relationships

A spatial signature of image can be obtained from the spatial relationships between objects of image. The spatial arrangement between objects of image can or cannot be described precisely. The image representation by a spatial signature permits not only to improve the efficiency and the quality of content-based retrieval, but also to execute partial query on image parts and to allow the user to refine his query according to his needs.

After segmentation in homogeneous objects by our method, based on the region growing (Chahir et al., 1999). All the relationships between all pairs of objects are extracted. All spatial relationships are saved in a graph called Spatial Relationships Graph (SRG). In this graph, each node represent an object in image whereas each arch represents the spatial relationships between two nodes (objects) of the extremities (Chahir et al., 2000).
Ever implementation, each object will be represented by its characteristic color. For practical reasons, we are interested in dominant visual objects of image. One Visual Object is an object with area or MBR is superior to a given threshold. Then the SRG will have a limited number of nodes, maximum of 20 in our application.
We presented in Figure 3 an original image taken, from INA[1] images collection, followed by its segmented one obtained by our segmentation method. The figure depicts a sample of segmented image sorted by interesting different coherent regions. The Visual Objects are given in decreasing order of importance in image and are semantically explained in Table 4.

[1] INA: Institut National de l’Audiovisuel

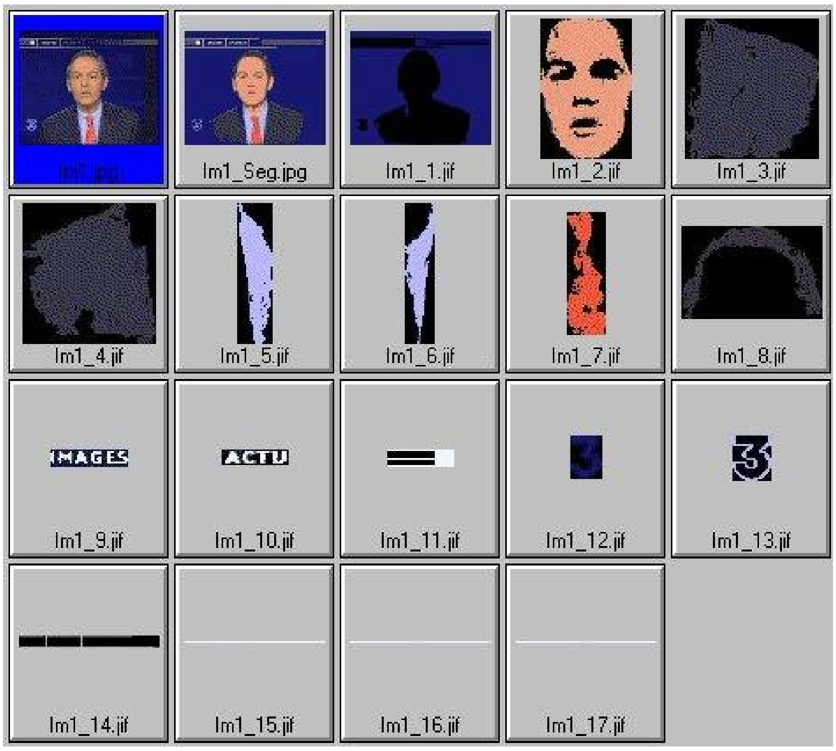


Figure 3 - Example of image with its dominant regions

Region	Significance	Region	Significance
Im1_1.jif	Image's bottom	Im1_10.jif	Text1
Im1_2.jif	Face	Im1_11.jif	Frame's part
Im1_3.jif	Jacket's right part	Im1_12.jif	"3" in blew
Im1_4.jif	Jacket's left part	Im1_13.jif	Edge of "3"
Im1_5.jif	Tie's left part	Im1_14.jif	Frame of text
Im1_6.jif	Tie's right part	Im1_15.jif	Line 1
Im1_7.jif	Tie	Im1_16.jif	Line 2
Im1_8.jif	Head	Im1_17.jif	Line 3
Im1_9.jif	Text 1		

Table 4: Regions and their different significance

To memorize all spatial relationships between dominant objects of image, we use one table to recap the spatial relationships graph SRG. The associated table of spatial relationships to the example given above is presented by the following Table:

Γ	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	R_{11}	R_{12}	R_{13}	R_{14}	R_{15}
R_1	-	⊃	⊃	⊃	⊃	⊃	⊃	⊃	⊃	⊃	⊃	⊃	⊃	⊃	⊃
R_2	⊂	-	↑	↗	↖	↗	↑	↓	↘	↓	↘	↗	↗	↘	↙
R_3	⊂	↓	-	→	→	→	→	↘	↘	↘	↘	→	→	↘	↓
R_4	⊂	↙	←	-	←	←	←	↙	↓	↓	↘	→	→	↓	↙
R_5	⊂	↘	←	→	-	←	←	↓	↓	↘	↘	→	→	↓	↙
R_6	⊂	↙	←	→	→	-	→	↓	↘	↘	↘	→	→	↘	↙
R_7	⊂	↓	←	→	→	←	-	↓	↘	↘	↘	→	→	↓	↙
R_8	⊂	↑	↖	↗	↑	↑	↑	-	↘	↘	↘	↗	↗	↓	↙
R_9	⊂	↖	↖	↑	↑	↖	↖	↖	-	→	→	↗	↗	⊂	←
R_{10}	⊂	↑	↖	↑	↖	↖	↖	↖	←	-	→	↗	↗	⊂	←
R_{11}	⊂	↖	↖	↖	↖	↖	↖	↖	←	←	-	↑	↑	⊂	←
R_{12}	⊂	↙	←	←	←	←	↙	↙	↙	↙	↓	-	⊂	↓	↙
R_{13}	⊂	↙	←	←	←	←	↙	↙	↙	↙	↓	⊃	-	↓	↙
R_{14}	⊂	↖	↖	↑	↑	↖	↑	↑	⊃	⊃	⊃	↑	↑	-	←
R_{15}	⊂	↗	↑	↗	↗	↗	↗	↗	→	→	→	↗	↗	→	-

Table 5 - This table recaps a spatial relationships graph

Generally, to each couple of objects (U, V) we associate one spatial relationship Γ. This is not symmetric relation. To indicate the direction of one inverted relation, we put the sign -, so U Γ V ≡ V -Γ U. In fact, we just memorize the basic relationships.
The route of one SRG composed by n nodes require S operations, who's calculated by the following formula:

$$S = (n-1) + (n-2) + \ldots + 1 = \sum_{i=1}^{n} (n-i) = \sum_{i=1}^{n-1} i = \frac{n(n-1)}{2}$$

For an identified image ImId, the regions information are memorized in REGION relation as depicted the Table 6. In this Table, we stock for each region the representative color, the MBR coordinates, the region size, and just as its code string which represents the spatial dispersion of the object in the image. The figure 4 depicts the coding of four quadrants.

1	**3**
0	**2**

Fig. 4 : Coding of 4 quadrants

The key of one region is composed by the concatenation of image number followed by the region's importance. The homogeneous regions of image are ordered by decreasing area. For example, the identified region by the key 110101 contained in image 1101, is characterized by:

- its color representation (127)
- the region size represents 35 % of this image
- the MBR's points coordinates are (10, 20) and (210, 200)

The region is dispersed through the four quadrants in this order: 3, 2, 1 and 0.

REGID = IMID ∪ NUMREG	COLOR	SIZE	MBR	CODE
110101 :	127	35 %	(10,12,210,200)	3210

Table 6 - The structure of relationships REGION and its attributes

For each image, we also memorize, in one relation named IMAGE, the spatial relationships between its regions. Each region is represented by its representative color. We represent the spatial relationships between dominant visual objects by their importance order.

Let's take the example presented below, and let's suppose that the number 1101 identify image, and we are limited to the 15 first dominant regions. These regions will be identified by 1101**01**, 101**02**, …1101**15**. The region 110101 is the most important followed by the region 110102, etc.

In particular, each line of Table corresponds to one image and codes the set of spatial relationships of dominant regions of this image. We memorize just the spatial relationships, which implicitly assume that the comparison is done always in this order:
The first region with the second, and then with the third, and then with the fourth, etc.
For example, the first line of Table 7, memorizes the spatial relationships of the 15 dominant regions of image 1101 with the following string:

"14⊃↑↗↖↗↑↓↘↓↘2↗↘↙4→4↘2→↘↓3←↙2↓↘2→↓↙2←2↓2↘2→↓↙→↓3↘2→↘↙↓3↘2→↓↙3↘2↗↓↙2→2↗⊂←→2↗⊂←2↑⊂←⊂↓↙↓↙←".

This string recaps the different spatial relationships.

If we would like constructing a spatial relationships graph from this table, we first have to know the number of nodes (i.e. of dominant regions). Knowing the number of spatial relationships between regions of one image, it is possible to deduct its dominant region number. In our example, if we count the number of spatial relationships, we find N_O=105 operators in the string. This expresses that the image is made up of 15 dominant regions (N_R=15).
Let's remark that the representation of n nodes necessitates S spatial operations where,

$$S = \frac{n(n-1)}{2}$$

$$N_O = S = n\,(n-1)/2$$

By formula development, we obtain n2 - n= 2NR or $n^2 - n - 2N_O = 0$. So the dominant regions number N_R is:

$$N_R = \frac{1+\sqrt{1+8N_O}}{2}$$

We deduce that the 14 first relations are from type "⊃".

10101 ⊃ 10102 , ... 10101 ⊃ 10115

10102 ↑ 10103, 10102↗10103, ….10102↙10115
….

Until 10114←10115

IMID	Spatial Relationships between regions
1101	14⊃↑↗↖↗↑↓↘↓↘2↗↘↙4→4↘2→↘↓3←↙2↓↘2→↓↙2←2↓2↘2→↓↙ →↓3↘2→↘↙↓3↘2→↓↙3↘2↗↓↙2→2↗⊂←→2↗⊂←2↑⊂←⊂↓↙↓↙←
1102	…
…	
1120	

Table 7 – Table of spatial relationships associated to images

For performance reasons, this table of spatial relationships was implemented by 'inverted file' structure as depicted Table 8. In fact, for each basic spatial relationship, we have a separated file into two columns. Each region is represented by its representative color, so the lines of Table indicate all the spatial relationships between the two extracted representative colors of all images from the database. For example, the line from Table 6 indicates that the spatial relationships Object of color 12 'Above' Object of color 15, was identified 12 times in image 11, then 2 times in image 111 and 3 times in opposite direction. In other words the spatial relationships 'Below', then 4 times in image 151, and finally one time for the spatial relationships 'Below' in image 189.

Pairs of colors of objects	Images Identifications
12,15	11(+12)111(-3+2)151(+4)189(-1)
123 ,145	11(+5)45(-2)456(+13)
…	
C_i,C_j	I_x(u), I_y(v), …, I_z(w)

Table 8 - Structure of inverted file associated to spatial relationships

2. 'By vote' search algorithm:

The problem here, is to find the most similar images from image database, to a given image query. We apply the vote principle.
Each given image query is segmented into dominant regions. After that, their visual characteristics and the spatial relationships between those regions are extracted. After that, the matter is to verify for each spatial relationship between two regions denoted (C_i, Cj), if it is also present in one candidate image. In this case, we increase one counter who represents the vote of the image with the number of its frequency. Like this, at the end of the process, the images with the maximum votes are considered as the most similar to the given image query. The 'by vote' search algorithm is presented in the following procedure.

```
Procedure Vote ( )
Begin
for each spatial relationship Γ between two homogeneous regions (Ci,Ck) from image query do
begin
        Open File Γ;
        if    Ci < Ck then  couple = (Ci,Ck);  sign = '-'
                      else  couple = (Ck,Ci) ; sign = '+' ;
        end;
        Search the line containing (Ci,Ck) ;
        Repeat
                if sign = '-' then
                            If image I is with –V then
                                      Vote[I] = Vote[I]+V
                          else
                            If image I is with +W then
                                      Vote[I] = Vote[I]+W ;
                End;
        Until eol ;
Close File Γ;
End;
Sort the votes of images in decreasing order;
End;
```

5. CONCLUSION AND FUTURE WORKS

Image retrieval has emerged as one of the most active research areas in this last decade. In this paper, we proposed an integrated approach for describing and retrieving image. This approach permits not only the separately access modes to images collections, but also combines the different image representations, for flexible image search and efficient navigation through image databases. In addition, we are studying new algorithms for query refinement, known as a relevance feedback. At the time of visual access to image database, the users have very different objectives. In order to satisfy the information needs of users, it is of capital importance to effectively and efficiently adapt the retrieval process to each user. In the visual information retrieval process, the user's position is central. Thus, in this iterative process, the user should indicate the relevance or irrelevance of retrieved images. This promising approach, successfully used for text retrieval, has focused on adaptively refining a user's initial query to more accurately, and progressively, select the desired items. Thus, in this way, the user and the system will collaborate together for convergence to the best items.

6. REFERENCES

1. Allen J.F. *Maintaining knowledge about temporal intervals*, Communication of ACM, 26(11): pp. 832-843, 1983
2. Berrut, C.; Mulhem, P.; and Bouchon, P. *Modelling and Indexing Medical Images: the RIME approach*. Proceedings of HIM'95, Konstanz, Germany, pp. 105-115, April 1995.
3. Chahir Y. and Chen L., *Efficient Content-Based Image Retrieval based on color homogenous objects segmentation and their spatial relationship characterization*, IEEE Multimedia Systems'99, International Conference on Multimedia Computing and Systems ,June 7-11, 1999, Florence, ITALY
4. Chahir Y. and Chen L. , *Searching Images on the basis of color homogeneous objects and their spatial relationship*, Journal of Visual Communication and Image Representation 11 , pp. 302-326 , 2000
5. Faloutsos, C.; Filcker, M.; Niblack, D.; Petrovic, D.; Equitz, W.; Barber, R. *Efficient and effective querying by image content*. IBM Almaden Research Center, San Jose (USA), Research report, August 1993.
6. Gudivada, V. N. and Raghavan, V. V. *Conent-based image retrieval systems*. IEEE Magazine, September 1995.
7. Gupta, A.; Weymouth, T.; and Jain, R. *Semantic queries with pictures: The VIMSYS model*. Proceedings of VLDB'91, Barcelona (Spane), pp. 69-79, September 1991.
8. Halin, G. *Apprentissage pour la recherche interactive d'images: processus EXPRIM et prototype RIVAGE*. Thèse de l'université de Nancy I, Nancy (France), 1989.
9. Manmatha R.; Ravela S; and Chitti Y. *On computing local and global similarity in images*. Proceedings of SPIE conf. On Human and Electronic Imaging III, 1998.
10. Minka, T.P. *An image database browser that learns from user interaction,* Master's thesis, MIT, Cambridge, MA, Februray 1996.
11. Niblack, W. et al., *the QBIC project: Query image by content using color, texture and shape*. Storage and retrieval for image and vodeo databases, pp. 173-187, San Jose, 1996, SPIE.
12. Pentland, A.; Picard, R.W., and Sclaroff, S. *Photobok: Tools for content-based manipulation of image databases.* Proceedings of Storage and Retrieval for image and Video Databases II, volume 2(185), pp. 34-47, Bellingham, wash, 1994, SPIE.
13. Rich, E. *Users are individuals: individualizing user models*. Int. J. Man-Machine Studies 18, pp. 199-214, 1983.