SINGLE AND MULTIPLE REGIONS QUERY WITH ABSOLUTE SPATIAL LOCALIZATION

Liana Stanescu-University of Craiova, e-mail:stanescu@nt.comp-craiova.ro
Dumitru Dan Burdescu-University of Craiova, e-mail:burdescu@topedge.com
Gabriel Muresan – Continental Teves AG&Co OHG,e-mail: gabriel.muresan@contiteves.com

Abstract: This paper presents the parallel strategy for processing the single and multiple regions query with absolute localization. The efficiency of the algorithm is tested with the help of some experiments effectuated on two imagistic collections: one collection with images from nature and the second collection with synthetic images.

Keywords:*content-based image retrieval, extraction of color region, quering by color and spatial information, multimedia databases*

1. INTRODUCTION

The objective of the content–based visual query is to search and to retrieve, in an efficient manner those images from the database that are the most appropriate to the image supplied by the user as query [1]. The content-based visual query differs from a usual query by the fact that it implies search of similitude. It is made a distinction between the two modes of the content-based query usage: content–based image query and content – based region query [1]. In the case of the content–based image query, the comparison is done at the level of the whole image, and in the second case, the images are compared on the base of their regions [1]. The content-based region query can be improved by adding the spatial information to the query. There are two types of spatial query: relative and absolute. The strongest system for images retrieval is the one that allows queries in which there are specified both visual characteristics and spatial properties for the desired images.

2. OBJECTIVES DEFINITION

In case of the absolute spatial localization region query, the query is done taking into consideration the fixed position of the region in image. For finding matches between two regions, the region query q and the target query t, there are considered more sizes [1],[2],[3],[4] such that:

1.The distance that indicates colour similitude.
2.The spatial distance between the centroids of the regions.
3.The distance between the areas of the regions.
4.The distance between minimum rectangles that bound the regions.

It is assumed that the colour regions are detected with the back-projection over binary colour sets algorithm[1],[2],[3],[4] and the attributes of the regions are stored into a table named RegiuniBP. For each detected region is stored the unique identifier of the region, the identifier of the image which contains the region and the coordinates of the left upper corner and right bottom corner of the minimal rectangle that bounds the colour region. An original implementation of the back-projection over binary colour sets algorithm is presented in [5].

Considering that the automate process of detection regions from images is realised according to back-projection over colour sets algorithm, the colour information of every region is memorised as a binary colour set [1],[5]. In these circumstances, the colour similitude of two regions can be calculated with **square distance between colour sets** or with **Hamming distance between colour sets**.

For example, the square distance between binary sets s_q and s_t is [1]:

$$D^s_{q,t} = \sum_{m_0=0}^{M-1} \sum_{m_1=0}^{M-1} (s_q[m_0] - s_t[m_0])\, a_{m0,m1}\, (s_q[m_1]-s_t[m_1])$$

One of the spatial location attributes of the regions is represented by the centroid of the region (x,y) [1]. Although it has not been directly memorised the coordinates (x,y) of the centroid in **RegiuniBP** table, which memorises information about found regions by applying the back-projection over colour sets algorithm, those coordinates can be easily determined, having memorised the coordinates (minX,maxY) of the left upper corner, respectively the coordinates (minX,maxY) of the right bottom corner of the minimum rectangle which bounds the region.The spatial distance between the centroids of the regions is given by the Euclidian distance [1]:

$$D^c_{q,t} = [(x_q-x_t)^2 + (y_q-y_t)^2]^{1/2}$$

An other important distance of the regions is their area, meaning the number of colour pixels contained by the region. The distance between the areas of two regions q and t is given by the absolute distance [1]:

$$D^a_{q,t} = | aria_q - aria_t |$$

The dimensions of the minimum rectangle that bounds the region represent importance in comparison of region spatial extent. The distance between the minimum rectangles having the height (tall) and the width (wide) that bounds the regions q and t is given by [1]:

$$D^d_{q,t} = [(wide_q-wide_t)^2 + (tall_q-tall_t)^2]^{1/2}$$

In these circumstances, the total distance between two regions q and t is given by [1]:

$$D_{tot} = \alpha_1 D^s_{q,t} + \alpha_2 D^c_{q,t} + \alpha_3 D^a_{q,t} + \alpha_4 D^d_{q,t}$$

where α_i represents relative weights assigned by the user to each attribute. For example, the user may give a greater value to α_1 parameter, meaning to the colour feature of the region regarding centroid localisation, or to the minimum rectangle which bounds it.

3. PARALLEL STRATEGY OF QUERY EXECUTION

In computing the distance between the two regions q and t, that implies the calculation of four component distances $D^s_{q,t}$, $D^c_{q,t}$, $D^a_{q,t}$, $D^d_{q,t}$, may be approached two strategies: a **parallel strategy** and a **pipeline one** [1].
For finding the set of regions that matches the best with region query $\mathbf{Q(f_q,(x_q,y_q),aria_q,(w_q,h_q))}$, the parallel strategy presumes the computation of four separate queries, one for colour feature, another for location, the third for area and the fourth for spatial extent.
Parallel strategy of region query execution with absolute spatial localization presumes the following steps:
1. there are selected all t target regions from RegiuniBP table for which the dissimilitude measure between colour features $\mathbf{f_t}$ and colour features of region query $\mathbf{f_q}$ is under a threshold **p1**. Using relational algebra operators, it may be written:
$R1= \sigma_{D^s_{q,t}<=p1}$ (RegiuniBP)
2. there are selected all t target regions from RegiuniBP table for which the distance between the centroid of the target region $\mathbf{(x_t,y_t)}$ and the centroid of the region query $\mathbf{(x_q,y_q)}$ is under a threshold **p2**, or otherwise:
$R2= \sigma_{D^c_{q,t}<=p2}$ (RegiuniBP)
3. there are selected all t target regions from RegiuniBP table for which the distance between the area of the target region $\mathbf{aria_t}$ and the area of region query $\mathbf{aria_q}$ is under threshold **p3**, or otherwise:
$R3= \sigma_{D^a_{q,t}<=p3}$ (RegiuniBP)
4. there are selected all t target regions from RegiuniBP table for which the distance between its spatial extent $\mathbf{(w_t,h_t)}$ and spatial extent of the region query $\mathbf{(w_q,h_q)}$ is under a threshold **p4**, or otherwise:
$R4= \sigma_{D^d_{q,t}<=p4}$ (RegiuniBP)
5. for obtaining a set of matching images, it is realised the join of relations obtained at steps 1-4, meaning:
$IMGC=R1 *_{idregion} R2 *_{idregion} R3 *_{idregion} R4$
6. the regions that match the best with region query are the ones for which total distance $\mathbf{D_{tot}}$ has minimum value, meaning:
$IMGF=\min_{Dtot}(IMGC)$

4. EXPERIMENTS FOR SINGLE REGION QUERY

For testing the efficiency of the algorithms for a query after a single region with absolute spatial localization, there have been made some experiments over two types of image collections, such that: the image collection from nature and the second collection that contains images synthetically created, with different geometric figures, of various colours and that have been rotated in various angles and placed in different locations in the image.In the computed tests it has been utilised the region query parallel strategy of execution with absolute localization.
The structure of **RegiuniBP** table has been filled with fields that memorise the computed distances, such that:
d11 – memorises the distance between the centroids of the region query and target region meaning $\mathbf{D^c_{q,t}}$
d12 – memorises the distance between spatial extent (minimum rectangles) of the region query and target region, meaning $\mathbf{D^d_{q,t}}$
d13 – memorises the distance between the areas of the region query and target region, meaning $\mathbf{D^a_{q,t}}$
d14 - memorises dissimilitude between the colour set of the region query and target region, meaning $\mathbf{D^s_{q,t}}$
d1 – memorises total distance **Dtot** between region query and target region.
Experiment 1 : The region after which it has been made the absolute spatial query, is the red circle, bounded in figure 1 by a white mark. The image identifier has the value 32 (id_image=32), and the unique identifier of the region is 312 (idregion=312).

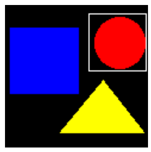


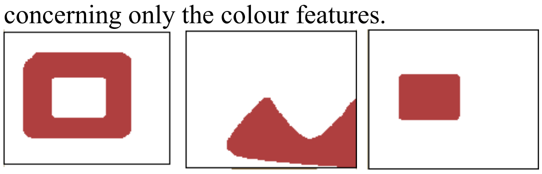
Figure 1: The image that contains the region after which it is made the absolute localization query.

Firstly there are selected all the target regions t from RegiuniBP table for which the measure of dissimilitude between their colour features $\mathbf{f_t}$ and the colour features of region query $\mathbf{f_q}$ is under threshold **p1** to which is given value zero. After this query there will be selected all the red regions. In the table from figure 2, there are some returned records after execution of the appropriate Select command.

id_image	Idregion	D14
30	11	0
29	15	0
31	36	0
2	48	0
3	49	0
4	54	0
…………………………………………		
………………………		

Figure 2: The record sequence from the RegiuniBP table for which the field d14=0 corresponding to target regions having the same colour as the region query.

In figure 3 there are some examples of returned regions concerning only the colour features.



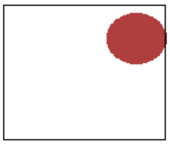
Id_image=30 Idregion=13
Id_image=29 Idregion=15
Id_image=31 Idregion=36

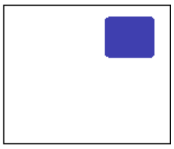
Figure 3: The sequence of regions for which the field d14=0, meaning that they have the same colour as the region query.

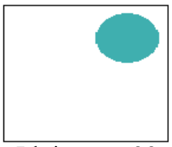
There are selected all t target regions from RegiuniBP table for which the distance between the centroid of the target region (x_t,y_t) and the centroid of region query (x_q,y_q) is under a threshold p2=35.In the table from figure 4, there are some returned records after the execution of the appropriate Select command.

id_image	Idregion	d11
32	312	0
6	67	1.5
22	141	2
23	145	2.5
19	127	3.807887
31	38	5.024938
……………………………………………… ………………		

Figure 4: The record sequence from the RegiuniBP table for which the field d11<=35 (distance between centroids of the region query and target region is under threshold 35).


Id_image=32 Idregion=312


Id_image=6 Idregion=67


Id_image=22 Idregion=141

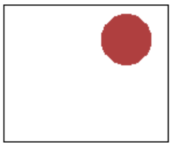

Id_image=19 Idregion=127

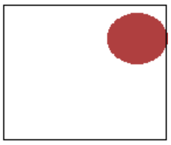
Figure 5: The sequence of regions for which the field d11<=35, meaning that they have approximately the same spatial localization of the centroid.

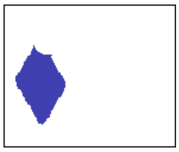
In figure 5 there are some examples of retrieved regions in case of region query considering only the spatial localization of the centroid. It may be observed that there have been returned regions of different shapes and colours, but which there are almost in the upper right corner of the image.

It is selected from the set of regions RegiuniBP, the target regions for which the distance between spatial extent of the target region (w_t,h_t) and that of the region query (w_q,h_q) is under a threshold p4=10.In the table from figure 6, there appear some of the returned records after executing the appropriate Select command.In figure 7 there are some examples of retrieved regions in case of the region query considering only the spatial extent of the region query. It is noticed that there have been returned the regions of different colours, but having approximately the same dimension, meaning approximately the same minimum rectangle that bounds them.

id_image	Idregion	d12
32	312	0
15	107	1
25	153	1
24	149	1
14	102	2
15	108	2.236068
21	136	2.236068
21	133	2.828427
2	47	3.162278
…………………………………………… ………………		

Figure 6: The record sequence from the RegiuniBP table for which the field d12<=10 (the distance between spatial extents of the region query and target region is under threshold 10).


Id_image=32 Idregion=312


Id_image=15 Idregion=107


Id_image=25 Idregion=153

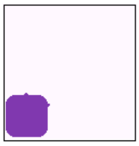

Id_image=24 Idregion=149

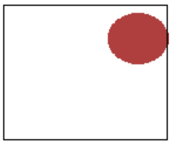
Figure 7: The sequence of regions for which the field d12<=10, meaning that they have approximately the same spatial extent.

It is selected from the set of regions RegiuniBP, the target regions for which the distance between its area $aria_t$ and that of the region query $aria_q$ is under a threshold p3=150, meaning that differs for the most 150 pixels. In the table from figure 8, there appear some of the returned records after executing the appropriate Select command.

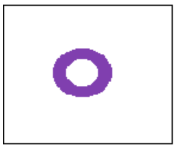
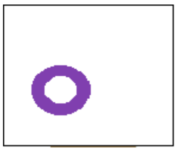
id_image	Idregion	d13
32	312	0
16	114	2
24	148	4
17	119	20
16	112	31
31	38	43
10	273	55
11	94	55
……………………………………… …………………		

Figure 8: The record sequence from the RegiuniBP table for which the field d13<=150 (the distance between areas of the region query and target region is under a threshold 150).

In figure 9 there are some examples of retrieved regions in case of the region query considering the area of the region query.

Id_image=32 Idregion=312 | Id_image=16 Idregion=114

Id_image=24 Idregion=148 | Id_image=17 Idregion=119

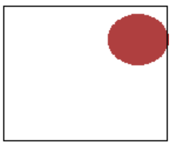
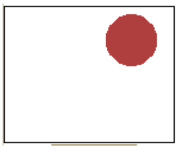
Figure 9: The sequence of regions for which the field d13<=150, meaning that their areas (number of pixels) differs for the most 150 pixels.

It may be observed that there have been returned the regions of different colours or shapes, but with approximately the same number of pixels.
In the table from figure 10, there appear some of the returned records after executing the joining of the previous obtained relations. The join is realised after the field **idregion**. The records are returned in ascend order of the distance d1, which represents total distance between the query region and the target region.
In the presented example it has been considered that all the four query criteria present the same interest, and the values α_1, α_2, α_3, α_4 take values equal to 100.

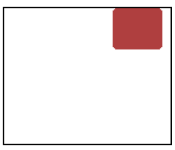
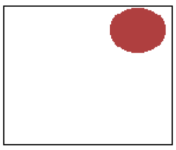
id_image	idregion	d1
33	312	0
23	145	107.8852
9	81	148.3129
7	69	163.089
4	54	177.8019

Figure 10: The record sequence from RegiuniBP table resulted after the joining of the four previous obtained relations. It represents the final result of the absolute localization region query

In figure 11 there are the regions corresponding to the records from figure 10.

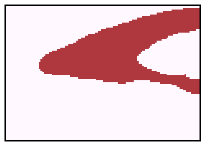
Id_image=33 Idregion=312 | Id_image=23 Idregion=145

Id_image=9 Idregion=81 | Id_image=7 Idregion=69

Figure 11: The target regions returned after the query in the marked region in fig. 1, considering the absolute localization and the weights α_1=100 ; α_2=100; α_3=100 and α_4=100.
Because at this absolute localization region query all the four distances have mattered equally, the returned regions have the same colour as the region query, are localized in the right upper corner and have approximately the same dimension and concerning the number of pixels (area) and the minimum rectangle that bounds them. It may be conclude that the result is very good.

Experiment 2: It is applied over the image collection from nature. In figure 12 appears the image with identifier 25 from which is extracted the region with identifier 750. In this case the values for parameters α_1, α_2, α_3, α_4 are: α_1=100 ; α_2=30; α_3=30 and α_4=0, which means that it is assigned a high importance to the region's colour, it is not important the spatial extent, and the localization of the centroid and the number of pixels are important in a percent of 30%.In the table from figure 13, there are some records after the execution of the absolute spatial localization region query algorithm.

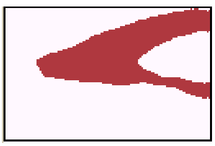
Id_imagine=25 Idregiune=750

Figura 12: The image that contains the region after which it is made the absolute localization query

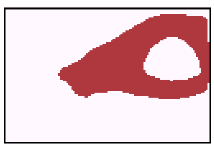
id_imagine	Idregiune	d1
25	750	0
39	598	14.1
36	588	42.15
38	595	61.92991
35	613	117.1801
33	533	142.4109
30	434	156.4934

Figura 13: The sequence of regions that represents the final result of the absolute localization region query for experiment number 2.

The regions corresponding to the records from the above table appear in figure 14.

Id_imagine=39 Idregiune=598

Id_imagine=36 Idregiune=588

Id_imagine=38 Idregiune=595

Figura 14: The target regions returned after the query in the marked region in figure 11, considering the absolute localization and the weights α_1=100 ; α_2=30; α_3=30; α_4=0

It may be conclude that the result is very good. There are returned only the red colour regions, of various shapes because it was not important the minimal rectangle. The spatial localization of the target regions is a bit larger because this weight was only 30%. Also, the areas of the target regions differ of the original because this weight was 30%. The regions with identifier 598,588 and 595 are very similar with query region and as a result the calculated distance d1 has small values in this cases, which confirm the efficiency of the algorithm.

5. MULTIPLE REGIONS QUERY – ABSOLUTE LOCALIZATION

In case of multiple regions query with absolute localization it proceeds in the following way:

1. For each query region Q_i, it is applied the algorithm presented in the paragraph 3, that is the algorithm for the query region with absolute spatial localization. Let be $\{R_i\}$ the set of regions resulted from the region query based on the region Q_i.
2. It is realized the natural join of the regions $\{R_i\}$ resulted at step 1, resulting the target candidate images
3. The image or the candidate images that are most appropriate to the query image are images that minimize the weighting sum of the distances between the target and query regions.

For example, if we have to resolve a query specified in this way: let retrieve the images that contain three regions that are most appropriate to the image query Q (Q^A, Q^B, Q^C). Each region Q^i is characterized by the attributes already specified in the paragraph 2, $(f^i_q,(x^i_q,y^i_q),aria^i_q,(w^i_q,h^i_q))$, namely: color characteristics, the position of the centroid, the area or the number of pixels and the spatial extent, namely the minimum bounding rectangle. The query is resolved in the following way:

1.There are retrieved the target regions that are most appropriate to the region Q^A.

$R_A \leftarrow$ Single_ region_query (Q^A)

2.There are retrieved the target regions that are most appropriate to the region Q^B.

$R_B \leftarrow$ Single_ region_query (Q^B)

3.There are retrieved the target regions that are most appropriate to the region Q^C.

$R_C \leftarrow$ Single_ region_query (Q^C)

4.For obtaining a common set of images there is realized a join of the relations obtained to the steps 1-3, namely:

$IMGC=R_A *_{id\_imagine} R_B *_{id\_imagine} R_C$

5.The images that are most appropriate to the query image Q are the images for which the total distance D_{tot} has the minimum value, namely:

$IMGF=\min_{Dtot}(IMGC)$ where

$D_{tot}=\beta_1 d_{QA,T}+\beta_2 d_{QB,T}+\beta_3 d_{QC,T}$

The values β_1, β_2, β_3 represent the relative weights assigned to each region, by the user. For example, the user may accord a greater value to the β_1 parameter, namely he is interested to retrieve the region Q^A than to retrieve the other regions.

6. EXPERIMENTS FOR THE MULTIPLE REGIONS QUERY WITH ABSOLUTE SPATIAL LOCALIZATION

For verifying the efficiency of the proposed algorithm for the multiple regions query with absolute spatial localization, we consider from the collection of synthetic images, the image from figure 15 in which we framed with white color the two regions on which the search is based.

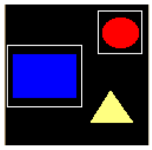


Figure 15: The image that contains the two regions on which it is realized the query with absolute localization

In the query the weights β_1, β_2 are equal, namely we are interested in the same way of the retrieval of the two regions.For each region the weights α_i have the values: α_1=100%; α_2=100%;α_3=0 and α_4=0, meaning that we are only interested by the color and the localization of the centroids of each region.Realizing the query on the blue region in conformity with the algorithm presented in the section 3, there are resulted the records from the table in the figure 16. The table includes the target regions identifiers in the ascending order of the total distance with respect to the query region (the blue square)

id_imagine	Idregiune	d1
33	595	0
34	598	0
26	561	9.124144
5	432	15.70032
10	462	17.6706
16	498	25.10976
14	485	25.73907
38	609	25.81182
13	479	26.04323
35	602	26.57537
40	616	28.29311
30	585	28.57009
19	519	30.41381
36	605	30.46309
39	612	31.53569
29	582	33.13608

1	411	33.13608
23	543	33.76759

Figure 16: The sequence of records from the table RegiuniBP that represents the result of the query with absolute spatial localization in case in which the query region is the blue rectangle.

Realizing the query on the red color region the records from the table from figure 17 were resulted. The table includes the identifiers of the target regions in the ascending order of the total distance with respect to the query region (the red circle).

id_imagine	Idregiune	d1
33	593	0
23	541	2.5
38	607	3.905125
39	611	9.300538
8	447	10.92017
7	441	10.96586
9	453	12.09339
40	615	16.56804
35	600	18.02776
15	488	24.12986
4	423	26.19637
18	509	27.35416
24	546	30.64719
25	553	31.81588
14	482	34.07712

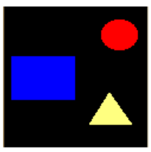
Figure 17: The sequence of records from the table RegiuniBP that represents the result of the query with absolute spatial localization in case in which the query region is the red circle.

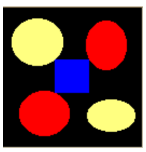
The natural join between the two relations established previously leads to the records from the table in the figure 18 that represent the identifiers of the images that include the both query regions.

id_imagine	D1
33	0
23	2.5
38	3.905125
39	9.300538
8	10.92017
40	16.56804
35	18.02776

Figure 18: The sequence of records from the table RegiuniBP that represents the result of the query with absolute spatial localization in case of both query regions

Some of the images corresponding to the records from the above table appear in the figure 19.


Id_imagine=33


Id_imagine=23


Id_imagine=38

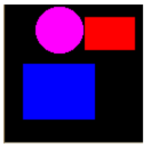

Id_imagine=39

Figure 19: The target images returned by the query on the two marked regions from figure 15.

Taking into account that at this multiple region query we take into consideration the color attribute and the absolute position of the centroids of the regions, the result includes these images that contain a region of red color situated in the upper right corner and a region of blue color in the left corner and bellow to the centroid of the image. Because of the fact that we did not take into consideration the number of pixels and the spatial extent of the query regions, the target regions have diverse forms and sizes.

7. CONCLUSIONS

In the case of region query with absolute localization it was tested the efficiency of an algorithm which assumes the parallel execution of the query. There are considered more sizes, such that:the distance that indicates colour similitude, the spatial distance between the centroids of the regions, the distance between the areas of the region, the distance between minimum rectangles that bound the regions. The weights of these sizes can be established by the user depending on the characteristics of the retrieved desired regions. The results of this algorithm were satisfactorily both in nature and synthetic images with different values of the four weights, because the relevant images for each query are returned in the top of the results.

BIBLIOGRAPHY

1. J. R. Smith. *Integrated Spatial and Feature Image Systems: Retrieval, Compression and Analysis.* Ph.D. thesis, Graduate School of Arts and Sciences, Columbia University, February, 1997.
2. J.R.Smith, Shih-Fu Chang. *Local Color and Texture Extraction and Spatial Query.*IEEE International Conference on Image Processing 1996.
3. J.R.Smith, Shih-Fu Chang. *SaFe: A General Framework for Integrated Spatial and Feature Image Search. IEEE Signal Processing Society 1997 Workshop on Multimedia Signal Processing June 23-25 1997, Princetown, New Jersey, USA*
4. J.R.Smith, Shih-Fu Chang. *Automated Image retrieval Using Color and Texture. IEEE Transactions On Pattern Analysis and Machine Intelligence, November 1996*
5. L.Stanescu, C. Cojocareanu. *The Colour Set Back-Projection Method for Colour Image Retrieval. Sintes 10 , Craiova 25-26 mai 2000*