Region Queries without Segmentation for Image Retrieval by Content

Jamal Malki, Nozha Boujemaa, Chahab Nastar, Alexandre Winter

INRIA / Rocquencourt, BP 105, 78153 Le Chesnay, France
Nozha.Boujemaa@inria.fr

Abstract. Content-based image retrieval is today ubiquitous in computer vision. Most systems use the query-by-example approach, performing queries such as "show me more images that look like this one". Most often, the user is more specifically interested in specifying an object (or region) and in retrieving more images with similar objects (or regions), as opposed to similar images as a whole. This paper deals with that problem, called region querying. We suggest a method that uses a multiresolution quadtree representation of the images and thus avoids the hard problem of region segmentation. Several experimental results are presented in real-world databases.

1 Introduction

Surfimage is a Content-Based Image Retrieval (CBIR) system developed at INRIA since 1996. Its specificity is the capacity to deal with both categories of image databases:

- For image databases with *ground truth*, the system should be as *efficient* as possible on the specific application. Examples include face recognition or medical image retrieval. A quantitative evaluation of the system can then be reported in terms of recognition rate, precision/recall graph, etc.
- For image databases where *no ground truth* is available, the system should be *flexible*, since the notion of perceptual similarity is subjective and context-dependent. Smart browsing, query refinement, multiple queries, and partial search on user-defined regions are among the desirable features of the system. Applications include stock photography and the World Wide Web.

`Surfimage` uses the query-by-example approach for retrieving images and integrates advanced features such as image signature combination, multiple queries, query refinement, and partial queries [1,2]. We focus on the latter problem in this paper.

Indeed, the user is most often interested in performing a query on an object (or region), rather than on the whole image. The goal of the system is then to retrieve those images in the database that contain similar objects. This observation motivates recent research on spatially-localized features and region matching. Methods range

from histogram computation without segmentation [3-5], approximate region segmentation [6-9], graphical and spatial structure of the image [10-12].
In this paper, we suggest a multiresolution quadtree approach for performing region queries. It represents a simple and efficient way to select image parts for retreival. Endeed, when manual object segmentation through all database images is too hard or when images are too complex for performing automatic segmentation, quadtrees image subdivision is an appropriate solution for partial queries. Image signatures are then computed systematically on each subimage. A dedicated similarity measure is computed, allowing to retrieve images with similar regions. The similarity measure defines the invariance properties of the query (e.g. "find similar objects anywhere in the image" vs. "find similar objects in the same image location"). Section 2 details feature computation in the multiresolution quadtrees. Section 3 discusses user queries and the associated inducted invariance. In section 4 we present comparative experimental results. We draw the conclusions in section 5.

2. Multiresolution quadtrees

Several problems occur with region (or partial) matching. The first one is segmentation: accurate segmentation of an image into regions in the general case is quite difficult. Another problem is invariance: is the user interested in finding other occurrences the object (or region) with the same position, orientation, and scale, or do they require invariance against these transformations?
In order to deal with partial queries, we use a multiresolution quadtree representation (similar to [13]) that localizes image features in structured regions (fig. 1). This approach avoids image segmentation and offers effective alternatives for the invariance problem.

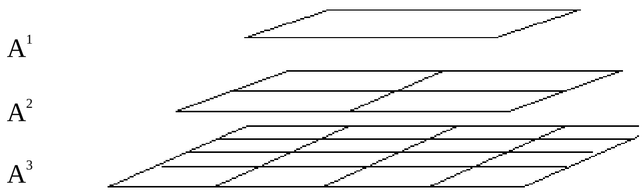

Fig. 1. Computation of image signatures on the multiresolution quadtree representation of the original image for 3 levels

Let A the original image and p the number of levels: For a given level i ($1 \leq i \leq p$) A is divided into n_i parts such that, for all levels, A is represented by a total number of N subimages:

$$A = \left(A^i_{n_i}\right)_{n_i}, \quad \sum_{i=1}^{p} n_i = N, \quad n_i = 4^{i-1}, \quad 1 \leq i \leq p \qquad (1)$$

2.1. Feature computation

We use a large selection of features offered by the Surfimage system [1]. Examples include color, shape and texture, mostly captured via histograms. The presentation of these commonly used features is not the scope of this paper -for further details, see [14, 15] or [1].

The features are computed on each subimage of the quad tree representation, yielding higher dimensional feature vectors. If we consider the p-level representation, the following feature vector is computed:

$$h=\left(h_{n_i}^{i}\right), \quad 1\le i\le p\ , \quad n_i=4^{i-1} \tag{2}$$

where each $\left(h_{n_i}^{i}\right)$ is defined as the feature *h* computed on the sub-image $A_{n_i}^{i}$. The set of features $\left(h_{n_i}^{i}\right)$ for a fixed *i* corresponds to the features of the *i*-th level of the representation. For instance, if the original single feature *h* is a 64-bin histogram, the dimensionality of the 3-level quad-tree feature is (1+4+16)*64 = 1344. Eventually, note that, when using histograms, we normalize the feature vectors so that they sum up to 1 -they thus represent distributions of that feature in the corresponding sub-image.

2.2. Feature combination

Combination of different features has been a recent focus of image retrieval [2]. The main problem is how do we combine ‘‘apples and oranges”, i.e. features that have different number of components, different scales etc.? We have experimented with several combination methods: voting, gymnastics-rule, etc… [16]. Among these methods, one seems to be the most appropriate and is described in the following.

Under Gaussian assumption, *the normalized linear combination* method uses the estimated mean μ_i and standard deviation σ_i of the distance measure *d* for each feature *i*, providing the normalized distance:

$$d\,'(x^{(i)},y^{(i)}) = \frac{d\,(x^{(i)},y^{(i)}) - (\mu_i - 3\sigma_i)}{6\sigma_i} \tag{3}$$

Where $x^{(i)}$ and $y^{(i)}$ are the vector signatures of images *X* and *Y* within feature *i*. The new distance measure *d’* will essentially have its values in [0...1]. The combined distance between *X* and *Y* is then:

$$D(Q,X) = \sum_{i=1} \rho(d\,'(x^{(i)},y^{(i)})) - \sum_{i} \rho(-\alpha_i) \tag{4}$$

where $\alpha_i = \frac{\mu_i - 3\sigma_i}{6\sigma_i}$, ρ is an increasing function (e.g. $\rho(x)=x$, $\rho(x)=x^3$ etc.).

3. User queries

3.1. Performing a partial query

For performing a region query, the user has to specify the regions of interest (RoI) at each level. Note that the RoIs do not have to be connex (fig. 2). For a given level, we build the corresponding bounding box, defined as the smallest subimage containing the RoIs (see fig.2). This bounding box (Bb) is able to catch the relative geometric positions of the different RoIs, and will be used for retrieving images .

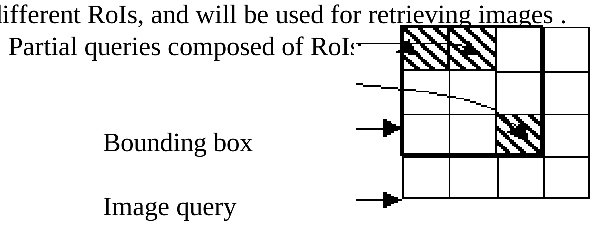

Fig. 2. Bounding box of partial query at the third level

In connection to the RoIs, the user has to specify which metric (fixed, location invariant) they are using. Note that most of the image signatures that we use are histograms, yielding most often rotation, translation and scale invariance within a subimage. The main invariances that the user can specify are thus: no location invariance (e.g. a face on top left, a hand on bottom left), no relative location encoding (a face and a hand anywhere in the image), location invariance (a face and a hand with specified relative positions).

3.2. Similarity metric and invariance issues

The previous invariances have to be translated into similarity metrics (or equivalently, distances). We have dealt with the ”no invariance" case in our previous work [16]. The ”loose relative location” case is an alternative which can be easily derived. We describe the location invariance hereafter; the method encodes the relative locations of the RoIs and is translation invariant (fig. 3).

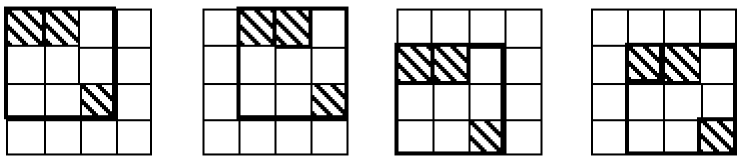

Fig. 3. Authorized translations of partial queries bounding box at level 3

Location invariance is built in as follows. Let R_Q the set of Nr partial queries selected by the user at the level *l*: $R_Q = \bigcup_{\beta=0}^{Nr-1} Q_{l_\beta} = \{Q_{l_0}, \dots Q_{l_{NR-1}}\}$

Let M the Bb of $\left(\bigcup_{i=0}^{Nr-1} \{Q_i\}\right)$, Nt is the number of authorized translations. The similarity measure between a query image Q and any image X in the database is given by:

$$d(Q, X) = \sum_{j=0}^{Nt-1} d\left(M(Q), M_j(X)\right) \tag{5}$$

$M_j(X) = \bigcup_{\alpha_j=0}^{Nr-1} Q_{l\alpha_j}$ is the sub-image obtained after a translation *j*. Note that:

$$d\left(M(Q), M_j(X)\right) = \sum_{\beta=0}^{Nr-1} \sum_{\alpha_j=0}^{Tm} \delta_{\alpha_j \beta}\; d\left(Q_{l_\beta}, X_{l\alpha_j}\right) \tag{6}$$

Where δ is the Kronecker symbol.

3.3. Redundancy

We note that the quadtree representation of image features is redundant, especially with respect to histogram computation. More precisely, a histogram feature computed at any node of the tree is equal to the sum of the histogram features of its 4 sons (regardless of the normalization procedure). The quadtree representation is thus redundant. However, experimentally, this redundancy is effective and useful: indeed, the lower levels tend to specify the holistic features of the image, thus allowing to restrict the search (e.g. I am looking for red roses, but in landscape scenes only). Note also that the contribution of each subimage to the global metric being equivalent, the lower levels participate less to the overall metric. Examples are shown in section 4. The user selects the first level to focus on an image category (e.g. a landscape), and specifies the query regions in a higher level.

4. Retrieval Results

We present results on our homebrew *bigdatabase* (Fig.4), which was built by merging the MIT *Vistex* database of textures, the *BTphoto* database of city and country scenes, a homebrew *paintings* database, and the *homeface* database of people in the lab. The total number of images in *bigdatabase* is 3670.

We use a kd-tree structure to optimize the search process. Note that when spatial relations between partial queries are not preserved, retrieval results are different from those displayed in figure 5, 6 and 7.

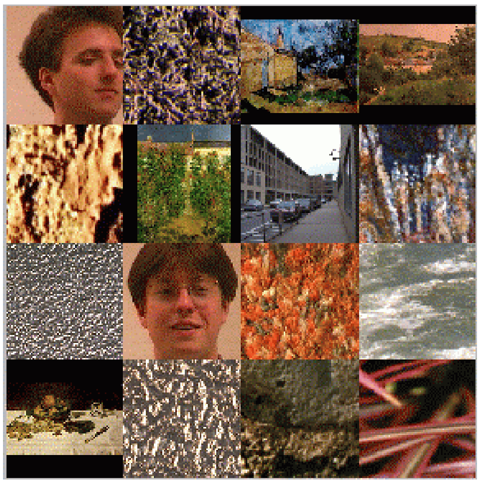

Fig4. A sample of *bigdatabase* consisting in 3670 heterogeneous images

Fig.5 Retrieval of the top left image. Retrieved images are from top left to bottom right in order of best match. Query image at level 1 and partial queries at level 3 are also presented. We search for selections every where in the database images

Fig6: Second example of partial queries on *bigdatabase*

Fig 7 Retrieved face images from *bigdatabase* with both level 1 and 3

Fig 8 Retrieved face images from *bigdatabase* with only level 1. System returns just face images among texture, landscape images and painting

5. Conclusion

To address the issue of region matching, we avoided the difficult problem of image segmentation using a multiresolution quadtree representation of the image. Our experience shows that we need two levels: level 1 to pick image category and level 3 to specify more precise details in a same category. This approach provides flexibility

for general and heterogeneous image databases retrieval without using specialized features such for face retrieval as shown in fig.7 and fig.8. Various general features are computed and combined together on each subimage. Dedicated similarity metrics allow for inferring invariance properties -in particular location invariance.
Experimental results show that the method is effective for local queries in image databases. They also emphasize the importance of preserving spatial organization of local features to achieve good retrieval results.

6. References

1. Nastar C., Mitschke M., Meilhac C. and Boujemaa N.: Surfimage: a Flexible Content-Based Image Retrieval System", ACM Multimedia'98", Bristol, September, 1998
2. Nastar C., Mitschke M., Meilhac C.: Efficient Query Refinement for Image Retrieval, Computer Vision and Pattern Recognition (CVPR'98), Santa Barbara, June, 1998
3. Michaeal J. Swain and D. H. Ballard: Color indexing, IJCV,vol.7, n°1, pp.11-32, 1991
4. John R. Smith and Shih-Fu Chang,Tools and Techniques for Color Image Retrieval,Proceedings of Storage & Retrieval for Image and VideoDatabases I, San Jose, CA, USA, February, pp. 426-437, 1996,
5. Jing Huang and Ravi Kumar and Mandar Mitra and Wei-Jing Zhuand Ramin Zabih,Image Indexing Using Color Correlograms,CVPR, Puerto Rico, June, pp.762-768, 1997
6. Sergio D. Servetto and Yong Rui and Kannan Ramchandran and Thomas S. Huang,A: Region -Based Representation of Images in MARS, Special Issue on Multimedia Signal Processing, Journal on VLSI Signal Processing, Oct, 1998
7. Belongie S. and Carson C. and Greenspan H. and Malik J.: Color- and Texture-Based Image Segmentation Using EM and Its Application to Content-Based Image Retrieval", Proceedings of the Sixth International Conference on Computer Vision(ICCV'98), Bombay, January,1998
8.Cinque L., Lecca F., Levialdi S., and Tanimoto S. : Retrieval of images using rich image descriptions, Proceedings of the ICPR, 1998
9. Neill W. Campbell, William P. J. Mackeown and Barry T. Thomas and Tom Troscianko: Interpreting image databases by region classification,Pattern Recognition (Special Edition on Image Databases), vol. 30,n° 4, Apr,pp. 555--563, 1997
10. Simone Santini and Ramesh Jain: The Graphical Specification of Similarity Queries,Journal of Visual Languages and Computing,1997
11. Del Bimbo and Vicario E. : Using Weighted Spatial Relationships in Retrieval by Visual Contents, IEEE workshop on Image and Video Libraries",Santa Barbara,June,1998
12. Soffer A., Samet H. and Zotkin D. : Pictorial Query trees for query specification in image databases,Proceedings of the ICPR,1998
13. Vellaikal A. and Kuo C.: Joint Spatial-Spectral Indexing of JPEG Compressed Data for Image Retrieval, Int'l Conf. on Image Proc.", Lausanne, 1996
14. Pentland A., Picard R. and Sclaroff S. : Photobook: Tools for Content-Based Manipulation of Image Databases", Storage and Retrieval of Image and Video Databases II,vol. 2185, San Jose, 1994
15. Jain A. and Vailaya A. : Image Retrieval Using Color and Shape: Pattern Recognition,vol.29,n°8, 1996
16. Nastar C., Mitschke M., Meilhac C., Boujemaa N., Bernard H. and Mautref M.: Retrieving Images by Content: the Surfimage System, Multimedia Informations Systems'98,Istanbul, September,1998