An Efficient Graph Representation for Image Retrieval based on Color Composition

† YIANNIS XIROUHAKIS, ‡ ANDREAS TIRAKIS and † ANASTASIOS DELOPOULOS

†Department of Electrical and Computer Engineering
National Technical University of Athens
9 Heroon Polytechneiou str., 157 73 Zographou
GREECE
email: {jxiro,adelo}@image.ntua.gr

‡ New Technologies Department
INTRACOM SA
Markopoulou Ave., 190 02 Peania, Attika
GREECE
email: atir@intranet.gr

Abstract: - Content-based video/image indexing and retrieval is of paramount importance in modern multimedia applications mainly due to the large amount of information involved. Querying by visual example, where the user provides an example image, and querying by sketch, employing user-constructed sketches, are important options in such systems. Prototype systems providing content-based image query and retrieval capabilities have been reported in literature, enabling efficient browsing in multimedia databases. In these systems, content information is modeled in terms of color, texture and shape attributes. However, existing techniques on image similarity retrieval, based on general color attributes, as well as localized color, have generally proved not to yield adequate results. Present work focuses on the extraction of image features in terms of their color composition information using a graph-based approach. This approach inherently captures local color information, and provides improved results in both the case of seeking images similar to an example image and the case of images containing a part similar to the image at hand. Experiments have been included to verify the efficiency of the algorithm. The proposed technique is evaluated in the framework of the Esprit project under development DiVAN.

Key-Words: - content-based indexing and retrieval, color composition, color segmentation, graph representation

1 Introduction

Our society is more and more based on information gathering, manipulation and dissemination, there is a strong need that the audio-visual information (traditionally been unavailable outside the organizations where it was produced and exploited) is organized in a way enabling wide-area management, exchange and collaboration. Distributed digital libraries providing content-based indexing and supporting content-based query and retrieval are of particular interest to TV broadcasters and audio-visual archives owners. Such organizations may exploit digital libraries both for internal purposes (i.e. to organize their own audio-visual archives with advanced video annotation and retrieval mechanisms improving their operations), and for selling material via online services to external customers. Such systems should be also open to accommodate standards like MPEG-7 that refers to metadata, as it is the case with the DiVAN digital library [1,8].

Systems, providing content-based image query and retrieval capabilities have been recently reported in literature, including VIRAGE, QBIC, Photobook, and VisualSEEk among others [3,4,5,6]. A portion of these systems, like VIRAGE and QBIC have been implemented and are now in the stage of evaluation as commercial packages. More recently, there has been an increasing interest towards the extension of similar capabilities for audiovisual information in general, including audio and video [1]. In all the aforementioned systems, content information is modeled in terms of dominant colors, texture, color and texture composition, as well as shape attributes. Several approaches have been also proposed in recent literature for semantic information extraction in terms of shape modeling and classification [2,7].

Image color attribute is probably the most popular one, as far as content-based retrieval is concerned. This is mainly due to the fact that color information is in general more easily distinguished and perceived by the human eye. Moreover, distinct users generally agree on the color attributes of an image, whereas the interpretation of semantic higher-level information is a rather subjective task. However, it is often reported that features extracted using dominant color information or color statistics (e.g. histogram distribution) do not yield adequate results. A popular statement on this fact is that the French flag is often mistakenly matched to the US one, since it is true that images of similar statistical color properties (general color information) may have different color composition information. The color composition attribute has been employed in order to overcome this inconvenience. Nevertheless, the popular quad-tree decomposition approach has proved to be inadequate as well, since the French flag is distinguished from the US one, but cannot be generally detected when it appears in images under different scale or in different position.

In this work, a new method is introduced for the extraction of color composition information from images and video sequences. The proposed method is based on locating dominant color objects in given images and comparing the latter in terms of the contained color objects and their connectivity properties. This approach appears to be more powerful than the traditional color composition algorithms, since it enables detection of color objects under different scale or in different positions in the image.

The proposed method emphasizes on efficient image segmentation into dominant color areas and modeling of the extracted information (segments size, segment color and type of connectivity) in terms of a simple graph. The graph nodes represent the extracted image regions, whereas graph links represent the type of the connection between the two respective nodes (areas). As a quantitative measure of connectivity, the boundary length is employed. The proposed graph approach has the immediate advantage that guidelines for efficient graph construction and parsing are provided in literature. At the same time, color composition information modeled in terms of flexible graphs is not restricted by color objects' position or scale, but relies on the objects size and connectivity properties.

Experimental results have been included to illustrate the algorithm's performance.

2 Image Color Segmentation

The problem of matching two images in terms of color composition immediately addresses image color segmentation or spatial color representation schemes (for example quad-tree decomposition). In this work, the former approach is adopted. In fact, color representation using quad-trees or similar schemes bears the advantage of fast implementation in both tree construction and matching (based on the quad-tree theory and applications). However, it can be seen that even a slight change in position or orientation of objects in the scene results in undesired matching results. Maybe the most characteristic example is the case of the moving camera in a video clip, where the respective change in the background leads to completely different quad-tree color representation even for successive frames.

Before performing image segmentation, we decide on the number of color bins employed in the algorithm. If the system utilizes already a general color similarity criterion (on the basis, for example, of color histograms), the number of color bins can be taken equal to the one set for general color similarity. Image segmentation is performed in the YCrCb domain. In the general case, b_Y, b_{Cr}, and b_{Cb} color bins are utilized resulting in $b = b_Y\, b_{Cr}\, b_{Cb}$ color levels. Thus, after obtaining the YCrCb counterparts of the RGB images, each pixel color value is quantized in one of the b possible levels.

A simple recursive color segmentation technique is utilized in order to segment the image in distinct color regions. Alternatively, any color segmentation technique may be utilized at the expense of time or accuracy depending on the particular system desired functionality. For example, in the case of a web-based on-line system, where the user has the ability to submit an example picture and perform queries by visual example, the recursive techniques (or even simpler ones) are preferable. On the contrary, when no on-line feature vector extraction is needed, a more accurate and time-comsuming scheme could be employed, for example algorithms based on independently growing color regions.

Even in the case where the number of color bins b is chosen to be relatively small, color segmentation results in a large number of color regions. The latter may often represent no additional information, since they correspond to noisy image areas or to insignificant (not dominant) color objects. For example, in an image containing a truck on the foreground, small gatherings of flowers on the background can be deemed as 'insignificant', since generally they are not even observable at a first glance.

In this sense, as a second step, a region merging scheme is applied. In our approach, the choice of regions to be merged relies on their relative size. For instance, an area can be regarded as 'noise' or better 'insignificant' when its size does not exceed a certain threshold a in relation to the total image size (e.g. a=1%). The detected regions are merged to one of their neighboring ones. As the most appropriate one is considered the neighboring region with the larger relative size and the closest color (in terms of color bins' distance). In this work, the former criterion was utilized, whereas the latter was taken into consideration in order not to merge all small regions to the background. The color of the resulting regions after merging was considered to be the color of the larger region (before merging). It must be noted at this point that as color information hereon can be considered the index of the matrix entry containing the centers of the color bins (indexed YCrCb).

3 Graph Representation

On the basis of the image segmentation and merging results, a number of dominant color regions are extracted in each image. The extracted regions are available as gatherings of image pixels as illustrated in Figure 1.

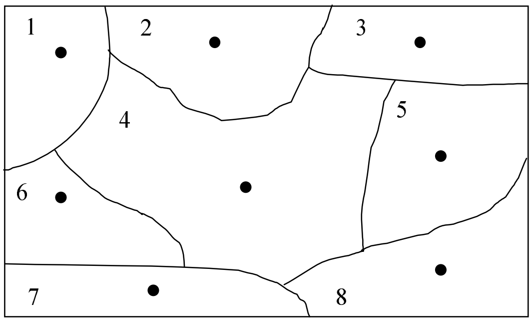

Fig. 1: Color segmentation and region merging results

Using the regions obtained, we formulate a graph representation as follows. The centers of the regions are supposed to correspond to the graph nodes, whereas for regions sharing common boundaries respective links are placed on the graph. In this sense, the graph corresponding to the areas depicted in Figure 1, is illustrated in Figure 2.

As far as the numbering of the graph nodes is concerned, we choose that the latter are numbered in descending relative area order. In this way, the first node corresponds to the background (or at least to the larger portion of it) and generally the dominant color objects are included among the first four or five nodes, depending on the thresholds applied to the merging scheme. In this sense, the region marked with the number '4' in Figure 1 would most likely be the first node.

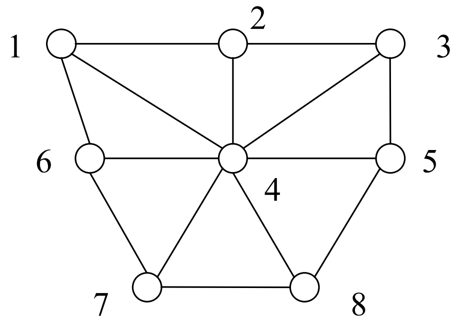

Fig. 2: Graph representation corresponding to the extracted areas of Figure 1.

Each node, contains the following information:

- the color index of the region (i.e. one of the b possible color indices/bins), and
- the area size of the corresponding region.

At this point, it must be noted that information such as centrality or elongation where deliberately excluded from this work since they correspond most likely to the shape attribute rather than to the color composition one. However, the extracted region information could well be utilized for such purposes [5].

The graph links represent the type of connectivity between image regions. In our formulation, there are four types of connectivity between two nodes i, j:

- 'simple' connectivity when the node (region) i shares a common boundary with node j,
- 'containing' connectivity when the region i contains region j,
- 'contained' connectivity when i is contained in j, and
- 'no' connectivity for disconnected regions.

In fact, for node interconnections both the the type of connectivity and the length of the boundary are stored in the feature vector. The latter is computer as the number of pixels that comprise the boundary. In particular, for each node, the links corresponding to larger boundary values, are deemed as more important. In this sense, its neighboring nodes are ordered in descending boundary length.

As a matter of fact, the obtained graph representation excludes information related either to the position of a region inside the image, or to its shape. The exact position of the color object in the image is not of any particular importance, since on the opposite case, we would not expect the algorithm to match similar objects on different positions on the pictures. That information would better suit in cases that we wanted, for example, to choose based on the relative similarity among pictures where the color object was already found to be included. In the case of object shape information such a task not only would correspond to another attribute of the image (the shape attribute), but it would also be very sensitive to affine object transformations (i.e. scale, rotation, translation). The exclusion of such information makes the algorithm immune to color object transformations (for shape modeling invariant to affine transformations see [7] and references therein). In addition, such a task would be considerably sensitive to (expected) faults in the image segmentation scheme.

The retrieval tool based on the proposed approach will be able to detect whether a portion of the given image is contained in one of the existing images, since the graph representation of the user-selected portion, will be a sub-graph of the graphs corresponding to images containing it. Taking into account the above, we may indicate that this approach accommodates two types of matching: matching images as whole entities and matching between specific regions (or objects) belonging to them.

4 Graph Matching Using Similarity Metrics

Along the lines of the methodology described in the previous sections, a graph representation is extracted for each input image. In fact, this graph represents the extracted feature vector corresponding to the particular image (at least, the portion of the feature vector corresponding to the color composition attribute). Image retrieval is then performed on the basis of the corresponding feature vectors using some similarity measures/function.

The particular similarity function employed for the comparison of the obtained graph representations may vary w.r.t. to the desired results. In other words, the graph representation adopted does not dictate the choice of the function. In this work, we implemented a similarity function by assigning 'similarity' bonus marks based on the following criteria:

- Node color similarity
- Node area similarity
- Node connectivity (number and type of links)

As mentioned in the previous section, each node contains the color information of the respective color area, that is the index of the color bin it corresponds to. Color similarity between nodes belonging to different graphs can be implemented using a color distance function. It must be noted here that the same color bins were used in the extraction of the feature vector for all images imported in the system, so that color indices correspond to the same actual colors for any extracted graph. For this purpose, color indices are ordered w.r.t. their actual color similarity. In this way, given color information for two nodes, the color similarity function returns the distance of the two indices in the sequence of indices, normalized w.r.t. the maximum bin distance.

Area similarity is treated on the basis of the percentage of the area covered by the region corresponding to the particular node over the total image area. Comparing these percentages for the nodes under comparison, one may decide on their similarity. However, the latter may not be suitable for our purposes, since identical nodes (regions) in different scale will be reported as 'different'. Thus, in this work, area size information is utilized only in terms of relevant importance, i.e. a region is deemed as more important than another, if the area covered by it, is larger in comparison to the other's. In this sense, the similarity function for the area attribute yields a better match when the nodes are equally important in their images. However, it must be noted that areas of different importance in their images are not deemed as not matched, but as worse matched. This issue will be clarified in the sequel.

Node connectivity is the most important similarity measure in our approach. Practically, the same object appearing in two images under different scale and position can only be recognized (apart from shape information) in terms of the connectivity of its dominant color regions. For example, a red car can be represented as a graph consisting of 3 nodes: a dominant red node connected with two smaller black ones, the wheels. The same graph pattern will appear under any scale or translation of the car. In this sense, two nodes belonging to different images are deemed as similarly connected when connected to equal number of nodes. However, it can be seen that color segmentation schemes may result in different number of connections for similar nodes due to small attached areas belonging to the background or even to other irrelevant objects (e.g. the car driver). For this reason, the nodes connected to the one at hand are ordered in descending length of common boundary. In other words, some connections are supposed to be more stable than others, in the sense that the nodes (regions) they link together have a larger common boundary. Thus, when comparing two nodes in terms of their connectivity, and the former has n_1 links whereas the latter only n_2, they are similar in terms of connectivity, when the rest n_1-n_2 links of the former are relatively insignificant.

In fact, the employed similarity function is a bonus-assignment mechanism taking into account the above three criteria. In the experiments included in this work, the implemented algorithm given two graphs F, G in steps can be analyzed as follows:

1. Compare the dominant nodes (in terms of size) of F and G in terms of color similarity. If the color similarity criterion yields small resemblance, exit comparison.
2. For every two similar dominant nodes compare the number of corresponding links. If these are different, investigate the possibility of rejecting the less significant links (in terms of common boundary). If not, exit comparison.
3. For every connection, compare neighbouring nodes in terms of color.
4. Repeat steps 2 and 3 recursively.

The recursive application of the algorithm results in gathering a number of bonus marks w.r.t. the output of the three employed criteria. The exact bonus quantities added in each step are determined by the system designer and w.r.t. the types of images imported in the system. It is though recommended that the more generic the employed images are, the less strict must be the criteria of resemblance. In a similar manner, when more color bins are utilized for increased region detail, the less strict must be the criterion of color similarity.

5 Experimental Results

The proposed approach has been tested over a large number of both simple and complex images. Particular attention was paid on whether images with similar color histogram distribution can be distinguished, on the basis of their difference in color composition employing the proposed algorithm. In this section, algorithm's performance is illustrated using a small set of representative images with relatively close color histogram distributions, containing large objects of similar color properties, whereas different position and shape properties (see Figure 3). Note that the color counterparts of the depicted images where employed in the algorithm. In all photos, the main foreground objects (cars, motor and football players) are of the same color (red).

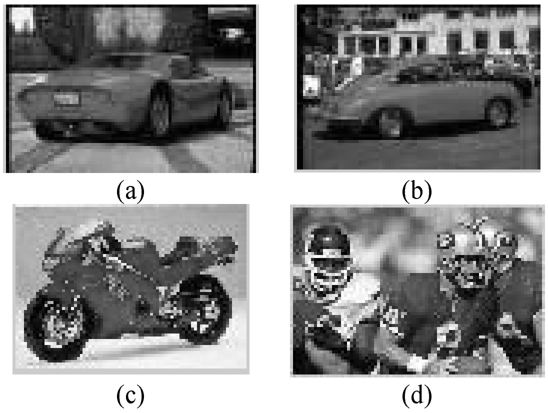

Fig. 3: The employed color images

As a first step, the YCrCb counterparts of the RGB images were obtained. Then the YCrCb components were reduced to the system default number of bins. In this example, the former were reduced to b_Y=2, b_{Cr}=3 and b_{Cb}=3 bins respectively resulting into b=18 bins /colors (the original images contained 256 RGB colors).

The resulting images were then segmented into distinct color regions using a recursive color segmentation algorithm. In the sequel, region merging was applied, where we considered an area to be 'insignificant' when its size did not exceed a=0.5% compared to the image total area. The obtained results are illustrated in Figure 4.

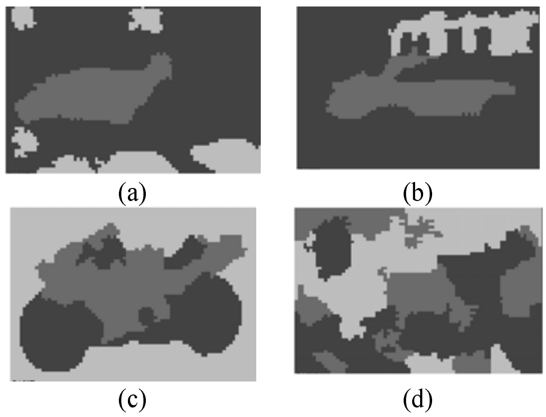

Fig. 4: Color segmentation and region merging results

At this point, the effectiveness of the algorithm becomes clear. The proposed color segmentation approach results in well-defined color areas, especially in the case of distinct foreground color objects.

For every given image, on the basis of the extracted color areas, a graph is constructed as explained in Section 3. In the construction of the respective graph, which in fact corresponds to the extracted feature vector, the size and the color of each area, as well as the common boundary between areas are taken into account. For example, for the image depicted in Figure 3(a), the resulting graph is depicted in Figure 5. Letters 'B', 'R' and 'G' indicate color index information for each node and letter 'c' indicates that the connection is of type 'containing-contained'.

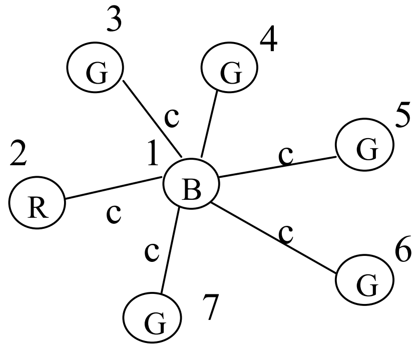

Fig. 5: Graph representation for image (a)

In the example, the application of the aforementioned methodology for the extracted feature vectors matching is rather trivial, since the resulting graphs are relatively simple. Considering that image (a) is given as input to the system where images (b), (c), (d) are already stored, image (b) is returned as the better match, followed by images (c) and (d). In this case, the results can be verified just by observing the extracted graphs. The graph for image (b) is similar to the one corresponding to (a), except from the fact that there are less 'G'-nodes. The graph corresponding to (c) would indicate that the larger object (background) differs from the one in (a) and the 'R'-node is adjacent to four 'B'-nodes, however there is still a dominant red object contained in the background node. Finally, the graph for (d) is by all means different to (a).

The algorithm's performance was tested in a number of images, both simple and complex with satisfactory results.

6 Conclusions

In this work, a system for content-based image retrieval based on the color composition attribute is introduced. This approach fundamentally differs from the traditional quad-tree decomposition techniques, resulting in a more generic and efficient representation of color content. The algorithm's efficiency was tested using natural images with satisfactory results. The proposed technique is currently under evaluation for inclusion in the final prototype of the Esprit project under development DiVAN.

Acknowledgements:

This work was partially supported by both the Esprit Project DiVAN (EP-24956) and the National Program ΥΠΕΡ'97: "Efficient Content-Based Image and Video Query and Retrieval in Multimedia Systems" of the General Secretariat of Research & Development of Greece.

References:

[1] A. Tirakis, P. Katalagarianos, M. Papathomas and C. Hamilakis, "Distributed audio-Visual Archives Network," accepted for publication at *IEEE International Conference on Multimedia Computing and Systems (ICMCS'99)*, Florence, Italy, 1999.

[2] Z.Huang and F.S.Cohen, "Affine-invariant B-Spline Moments for Curve Matching," *IEEE Trans. Image Processing*, Vol.5, No.10. October 1996.

[3] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by Image and Video Content: The QBIC System," *Computer*, vol. 28, pp. 23-32, September 1995.

[4] A. Pentland, R.W. Picard and S. Sclaroff, "Photobook: Content-Based Manipulation of Image Databases," *SPIE Storage and Retrieval Image and Video Databases II*, no. 2185, San Jose, February 1994.

[5] A. Gupta and R. Jain, "Visual Information Retrieval," *Communications of the ACM*, vol. 40, no. 5, May 1997.

[6] Y. Xirouhakis, G. Votsis, K. Karpouzis, and S. Kollias, "Efficient Browsing in Multimedia Databases using Intelligent Agents and Content-Based Retrieval Schemes", *Workshop on Multimedia Signal Processing (IEEE)*, Los Angeles, California, USA, December 1998.

[7] Y. Xirouhakis, Y. Avrithis and S.Kollias, "Image Retrieval and Classification Using Affine Invariant B-Spline Representation and Neural Networks," *Proc. IEE Colloquium Neural Nets and Multimedia,* London, UK, October 1998.

[8] http://divan.intranet.gr