

Image Retrieval by Color Regions

A. Del Bimbo, M. Mugnaini, P. Pala, F. Turco, L. Verzucoli

Dipartimento di Sistemi e Informatica
Università di Firenze, 50139 Firenze, Italy

Abstract. *The minor expressiveness of text with respect to visual features doesn't allow to fully exploit capabilities of human memory. For this reason retrieval based on visual contents has been identified as the means to provide an effective retrieval. The relative relevance of visual elements depends on user's subjectivity (which is not known in advance) and on the context of application. To cope with this problem, the description of image visual contents should exploit different levels of detail in image representation. In the following we present the PICASSO system, which supports image indexing and retrieval based on shapes, colors, and spatial relationships. The system exploits a pyramidal representation of the images which are analyzed at different levels of resolution. Multiple descriptions of image properties are created at different levels of resolution thus allowing effective retrieval through specific queries as well as imprecise ones. Retrieval results for a sample database of pictorial images are reported, with efficiency and effectiveness measures.*

1 Introduction

Image databases are now currently employed in an eclectic range of different areas such as entertainment, art history, advertising, medicine and industry among others. In all these contexts, a key problem regards the modality of easy and efficient access of image contents. Querying must be natural and friendly, and retrieved items must be *relevant* to user's expectation. The main problem with image databases is that, unlike textual and numerical databases, the visual content has no ordering rules, unless textual labels are associated with images. Conventional attempts to cast visual features into textual keywords [2] has been far recognized to be inadequate in indexing pictures. The minor expressiveness of text with respect to visual features doesn't allow to fully exploit capabilities of human memory. Items retrieved through a textual query could not be relevant at all for user's expectation.

This is the reason why retrieval based on visual contents has been identified as the means to overcome this modal clash.

The visual content of an image should be intended as what humans remember after looking at it for a couple of minutes (some relevant shapes, the colors, some textured surfaces and the arrangement of visual elements on the image plane). The relative relevance of visual elements depends on user's subjectivity (which is not known in advance) and on the context of application. Spatial arrangement

is relevant to a doctor examining a heart section, while color arrangement is important when looking at paintings, but will be better remembered by an expert than by an occasional user. Generally speaking, an efficient system for visual image retrieval should be able to:

- i*) provide a query paradigm that allows users to naturally specify *selective* queries as well as *imprecise* ones;
- ii*) implement techniques for finding the best similarity measure according to the application context of the query.

Visual querying by-example for pictorial data is the interaction paradigm that exploits human natural capabilities in picture analysis and interpretation. The query by example allows the user to express queries in a natural way, permitting imprecision and incompleteness of expression, by letting users draw a sketch of their memory of the image (colored shapes arranged in a specified pattern). It must not be expected that the first system answer be already satisfactory, rather a continuous interaction must be foreseen, through which the original query of the user can be refined or changed on the basis of the retrieved pictures.

Color-based Image Retrieval Experiences. A number of experiences have been reported which implement visual retrieval by contents from image databases by addressing different facets of the informative contents of pictorial data. With the increasing availability of devices supporting acquisition and visualization of color images, a growing attention is being focused on accounting for chromatic features in the retrieval process.

Image retrieval based on color has been formerly proposed by Swain and Ballard in [12]. There, image contents are described by global color histograms and queries are expressed by means of example images. Retrieval is performed by evaluating the similarity between the global color histograms of user provided examples and stored images.

The QBIC database system [11], [3] allows user composed queries, but still evaluates similarity in terms of global properties of color histograms. This limit is partially overcome in [10], by partitioning images into blocks, each associated with its own local histogram.

In [13] an iterative technique is proposed to identify image regions which can potentially represent a given object, based on color features. Swain and Ballard's Color Histogram Intersection is used to evaluate the match between image regions and query objects.

In [7], both color and shape features are used for retrieval. In this system the query is formulated through an example image and retrieval is accomplished by a similarity measure computed on the basis of global color histogram and image edges.

In the following we present the PICASSO system, developed at the Visual Information Processing Lab of the University of Florence, Italy.

PICASSO collects within a unified framework previous experiences carried out at the same Lab [4], [5], with new facilities for image indexing and retrieval

based on shapes, colors, and spatial relationships. In particular, the system supports retrieval by global color similarity (not discussed in this paper) and retrieval by color similarity of image regions. Shape, size, and position of color regions are also considered in the retrieval process according to the query request.

2 Hierarchical Color Image Segmentation

Image segmentation is the partition of an image into a set of non-overlapped regions, homogeneous with respect to some criteria, whose union covers the entire image. Segmentation techniques for color images have been proposed by several authors [6], [9], [8], [1].

The number r of regions which are produced in the segmentation process determines the level of precision of the image partitionment. This level must be set so as to find a trade-off between computational complexity and adherence of representation. The value of r can be determined adaptively on the the basis of statistical analysis of color distributions in the sampled color space. However, in the absence of specific assumptions about images being stored, and the level of precision which will be requested, at storage time, the optimal level of precision cannot be defined.

In the PICASSO system, this hurdle is circumvented by creating multiple descriptions of each datum, each achieving a different level of precision, and by deferring the solution of the trade-off between complexity and adherence to the retrieval time. The system exploits a pyramidal segmentation of the images which are analyzed at different levels of resolution. At the lowest level of the pyramid, each region corresponds to a pixel in the image. Each region at level n represents a cluster of adjacent regions at level $n - 1$. A *region energy* E_R is associated with each region R , E_R being of the form:

$$E_R = \alpha \frac{1}{A(R)} + \beta D(R) + \gamma \sum_{R_i \in N(R)} \frac{1}{D(R \cup R_i)} \quad (1)$$

where α , β and γ are positive real weights, $A(R)$ the area of region R , $D(R)$ a measure of the distance in the color space among pixels belonging to region R , and $N(R)$ the set of regions adjacent to R . The *image energy* is defined as the sum of the region energies:

$$E = \sum_R E_R \quad (2)$$

Image segmentation is performed by minimizing the image energy and trying to achieve the best compromise between three different requirements. As first, since we want to capture the most relevant color groups in the image, it is requested that the segmented image is not be composed of too many regions: first term of Eq.1. As second, regions have to be uniform in color (i.e. the color distance between pixels of the same region should be small): second term of Eq.1.

As third requirement, adjacent regions should be characterized by different colors (otherwise they should be merged): third term of Eq.1.

Image segmentation is performed by iteratively updating region clusters at each resolution level, separately. Starting from the lowest level of the pyramid two adjacent regions are searched whose merging would decrease the image energy (2). If no such regions are found, the current segmentation is considered as the valid segmentation for the lowest level. Next, values of α , β and γ are changed in order to move the minimum of the image energy toward a configuration with a lower number of regions. This procedure is recursively applied until the coarsest resolution is reached, where the entire image is represented by a single region.

At the end of the segmentation process a color image \mathcal{I} is associated with N segmented images \mathcal{I}_n $n = 1, \dots, N$, obtained, by recursively decreasing the number r of allowed image regions. Images are described using multi-layered graph data structures similar to that sketched in Fig. 1. Each region R_k in image \mathcal{I}_n at layer n of the pyramid is a node $V_{n,k}$ of the graph \mathcal{G} . Node $V_{n,k}$ is connected through *intra-level links* to those nodes which represent the neighboring color regions in the image description \mathcal{I}_n . $V_{n,k}$ is connected through *inter-level links* to its sons nodes in layer $n - 1$, representing those regions in image \mathcal{I}_{n-1} that have been derived from region R_k , when the resolution was increased. The overall multi-layered graph \mathcal{G} can be regarded as multi-resolution index of chromatic content of image \mathcal{I} , which stores the positional information extracted through the segmentation process.

2.1 Region description

Description of color regions represented by the graph nodes, includes features such as the region average color, spatial location, area, and shape.

In order to support effective retrieval by content of paintings, a color space must be chosen such that close distances in the color space correspond to close distances for the user perception. This condition, which is not exhibited by the RGB space, has been accomplished with the adoption of the CIE $L^*u^*v^*$ (*extended chromaticity*) space. Moreover, a set of reference colors must be chosen, such that their number is low enough to keep low the computational effort, and that, in the meantime, colors represent the ideal centers of color clusters, according to the common human perception.

To this end, in the PICASSO system, 128 *reference colors* are used to define a quantization of the ($L^*u^*v^*$) space into 128 color clusters. A binary 128-dimensional *color vector* is associated with each region, each entry of the vector corresponding to 1 if the reference color is present (according to a fixed threshold) in the region, 0 otherwise.

At the coarsest resolution, the image is represented by a single region with a color vector retaining the global color characteristics for the whole image. At finer resolutions, the number of regions r in the image space grows. This yields an increasing localization of chromatic properties into smaller and more cohesive regions. This localization process induces an inheritance property of color vectors. The color vector of each region is computed as the union of the color vectors

associated with child regions in the pyramid. Fig. 2 shows different steps of segmentation and corresponding incremental descriptions associated with each of them, for a sample painting. It can be noticed that the color vector becomes more *selective* as the level of resolution increases, since each region includes the color representation of its children.

Region *position* is measured through the absolute position of the region centroid. Region *area* is measured through the ratio between the number of region pixels and image pixels. Region *shape* is represented through its best approximating ellipse: the ratio between the major and minor axes, and the major axis orientation are used as shape features.

3 Color Image Retrieval

The PICASSO system supports visual retrieval by example of images with one or more colored regions. The user is allowed to draw region contours and arrange them in any positions on the interface blackboard. Each region is then filled with the appropriate color, which is selected from a color palette on a proper color window interface. Queries can regard region color or a combination of color and region attributes such as position, area, elongation, or orientation.

In the system, at database population time, raw images are segmented, and the pyramidal representation is derived. To avoid retrieval through a sequential scanning of the whole database, a *color index file* is built, storing the color vectors associated with the highest node of each pyramid structure.

Once the query has been issued, the color index file is analyzed to select a set of candidate images containing regions with the same colors as the query. A fast selection of non interesting images is obtained by filtering out those images where a query color is not present in the color vector. Then the pyramid structure of each candidate image is analyzed to find the best matching region for each region in the query. A matching score M between a query region QR and an image region R is computed as the weighted sum of the distances between all the features selected.

$$M = w_{col.} * d_{col.}(QR, R) + w_{loc.} * d_{loc.}(QR, R) + w_{area} * d_{area}(QR, R) + w_{shape} * d_{shape}(QR, R) \quad (3)$$

A similarity coefficient for the whole image is derived as $S = \sum_{i=1}^N M_i$, being N the number of query regions and M_i the match score of each region.

Once all the images in the database have been processed, they are sorted depending on the value of S , and presented to the user.

In Fig. 3, an example of image retrieval by multiple color regions is shown. In this example, on the left side canvas, the user has sketched two regions, to search for landscape pictures with a blue sky over a green patch. The system output is shown on the right side canvas. Such figure shows the effects of the multi-resolution segmentation of the images: images are retrieved even if their green

patch is not represented as a uniform homogeneous region but is obtained as a mosaic of different color patches. Nevertheless, from the point of view of user's perception, the presence of such distributed color patches provides a sensation similar to that provided by uniform green patches.

Performance figures obtained from our experiments are reported below in terms of *precision* \mathcal{P} and *recall* \mathcal{R} . Both parameters are a function of the total number of retrieved items (T_r), which is a user tunable parameter. They are interdependent, in the sense that one cannot be increased without decreasing the other. Parameters have been estimated through a set tests performed on a database of about 400 images, using random arrangements of color regions as queries. In Tab.1, they are shown the values of \mathcal{R} and \mathcal{P} for $T_r = 20$ (the number of output images usually requested in our tests).

References

1. T. Uchiyama and M.A. Arbib. Color image segmentation using competitive learning. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, (16)12:1197–1206, Dec. 1994.
2. R. K. Srihari. Automatic indexing and content-based retrieval of captioned images. *IEEE Computer*, 28(9):49–56, Sept. 1995.
3. W. Cody. Querying multimedia data for multiple repositories by content: The GARLIC project. In *Proc. on Visual Data Base Systems III*, Lausanne, 1995.
4. A. Del Bimbo and P. Pala. Visual image retrieval by elastic matching of user sketches. Accepted for publication in *IEEE Trans. on Pattern Analysis and Machine Intelligence*.
5. A. Del Bimbo, P. Pala. Image Indexing Using Shape Based Visual Features *International Conference on Pattern Recognition*, Wien, August 1996.
6. C.L. Huang, T.Y. Cheng and C.C. Chen. Color images' segmentation using scale space filter and Markov random field. *Pattern Recognition*, (25)10:1217–1229, 1992.
7. A. K. Jain and A. Vailaya. Image retrieval using color and shape. *Pattern Recognition*, 29(8):1233–1244, Aug. 1996.
8. Y.W. Lim and S.U. Lee. On the color image segmentation algorithm based on the thresholding and the fuzzy *c*-means techniques. *Pattern Recognition* (23)9:935–952, 1990.
9. J. Liu and Y.H. Yang. Multiresolution color image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, (16)7:689–700, 1994.
10. A. Nagasaka and Y. Tanaka. Automatic video indexing and full video search for object appearances. In *IFIP Trans., Visual Database Systems II*, pages 113–127, Knuth, Wegner (Eds.), 1992. Elsevier.
11. W.Niblack et alii, "The QBIC Project: Querying Images by Content Using Color, Texture and Shape". Res.Report 9203, IBM Res.Div. Almaden Res.Center, Feb.1993.
12. M.J.Swain, D.H.Gallard, "Color Indexing". *Int.Journal of Computer Vision*, Vol.7, No.1, 1991.
13. V. Vinod, H. Murase, "Focussed Retrieval of Color Images". To appear on *Pattern Recognition*

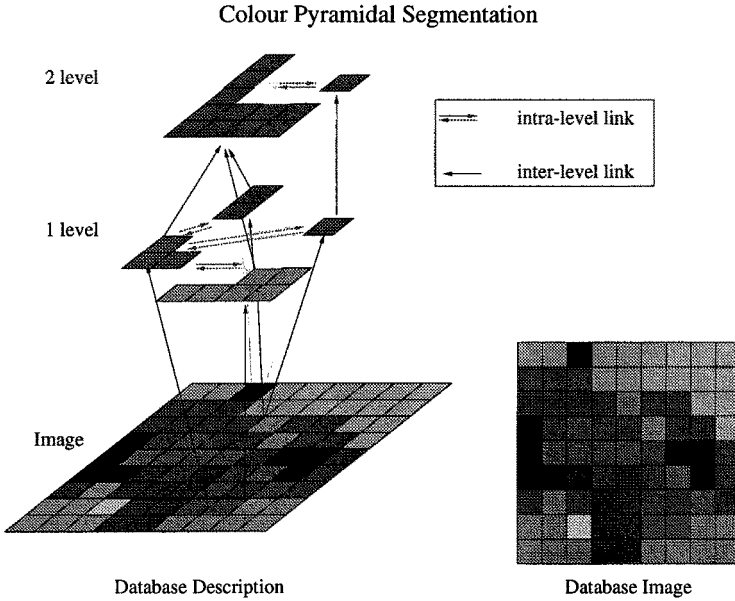


Fig. 1. Image description

	\mathcal{P}	\mathcal{R}
<i>Color</i>	92.1%	86.4%
<i>Color Area</i>	80.2%	76.8%
<i>Color Position</i>	92.8%	88.3%
<i>Color Area Position</i>	87.4%	82.1%

Table 1. Values of Precision and Recall for $T_r = 20$. Values are shown for 4 kinds of search and suggest that the best performance is provided for search combining the color and position features.

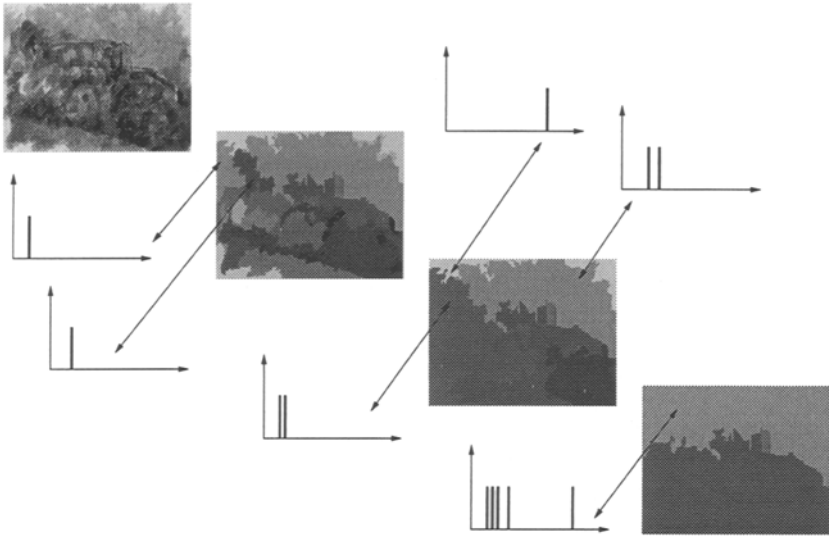


Fig. 2. Segmentation on a painting. The original image is shown with three images representing its segmentation at three different resolutions. Also shown are the binary color vectors of some regions. The color vector of each region is computed as the union of the color vectors associated with its children.

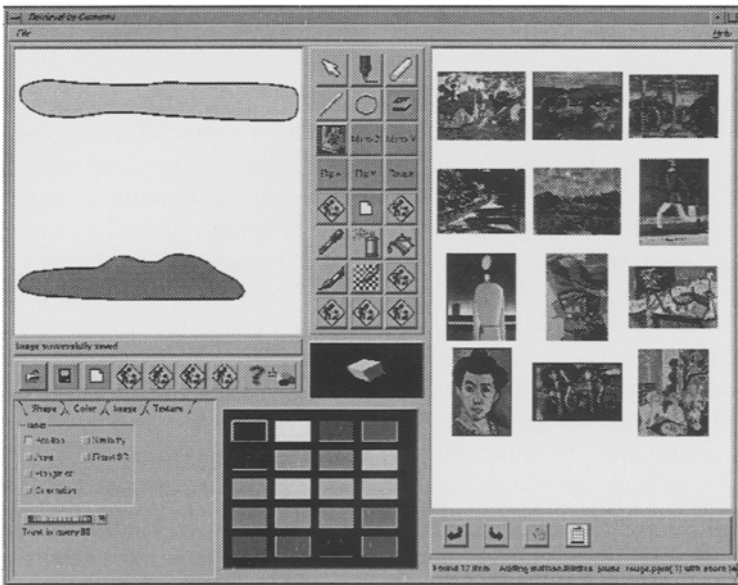


Fig. 3. Retrieved images of landscape paintings.