Multi-Level Image Segmentation and Object Representation for Content Based Image Retrieval

Pınar Duygulu and Fatoş Yarman-Vural
Department of Computer Engineering, Middle East Technical University, Ankara, Turkey, Tr-06531

ABSTRACT

Due to the increasing demand and offer of the technology, the next generation of the image file formats will be more likely to store and retrieve images based on their semantic content. Thus, an image should be segmented into "meaningful" regions, each of which corresponds to an object and/or background. In this study, we propose a scheme for multi-level image segmentation, based on a simple descriptor, called "the closest color in the same neighborhood". The proposed scheme generates a stack of images without using any segmentation threshold. The stack of images is hierarchically ordered in a uniformity tree. The uniformity tree is, then, associated with a semantic tree, which is built by the user for content based representation. The experiments indicate superior results for retrieving images, which consist of few objects and a background.

Keywords: Content Based Image Retrieval, Multi-level Image Segmentation, Color Descriptor

1. INTRODUCTION

Efficient image retrieval in digital visual database systems has been of great interest over the last decade. Recent developments in digital imaging technology, broadband networking and digital storage devices have set the stage for the generation, transmittal, manipulation and storage of large numbers of digital images and documents. In order to access these images automatically and on demand requires the ability to segment, index, store, and retrieve visual information effectively and efficiently and as such offers unprecedented challenges in the development of these technologies. These challenges have generated significant interest in the development of content based image indexing and retrieval algorithms and systems.

Recently, prototype systems became available that allow the user to utilize simple textual descriptions, or complex visual features such as color, texture, shape and/or motion to search large databases (see:,[1] ,[2] [3] and[4] for recent surveys on image and video indexing and retrieval technologies). Available systems include : Trademark image database system,[6] QBIC,[7] Photobook,[8] VisualSEEK[9] and WebSEEK,[10] Chabot,[11] Illustra/Virage,[12] FourEyes,[13] MARS,[14] RetrievalWare[15] and Netra.[16]

In everyday life, humans are accustomed to utilizing high level concepts, like objects, people, places, etc., to help us navigate through our daily quests. While these concepts come naturally to a human observer, they pose a significant challenge to computer systems that are attempting to perform content based image indexing and retrieval in an automatic fashion. Researchers have utilized various features such as color, shape, texture, and motion in an attempt to develop a "semantic" level of understanding for image content. Although the progress in this area has been steady and forthcoming with many papers published on a yearly basis, the research is still in its infancy stages with many breakthroughs yet to be made.

If an image query system has questions, related to the quantitative measures of the image, like "get me the images with high content of red", the problem is relatively easy and can be solved by histogram descriptors. However, answering a query of "get me the picture of flowers" requires more sophisticated methods to resolve the query. It is almost impossible to represent a large class of images by using a generic set of image descriptors. A set of descriptors, which is very efficient in a small set of images, may go totally astray in another set. Therefore, one needs to answer the question of "to what level of abstraction we use a generic representation of images using color, texture and/or shape descriptors?" It is evident that for higher abstraction levels, application dependent tools are required. Now, the state of art in the current research is to find a set of descriptors and related description schemes, which are generic, robust and capable to describe a large class of images in highest possible level of abstraction.

Color is one of the widely used descriptors, since it provides important semantic information and easily recognizable by the machines. Use of the average color value is popular.[1] However, it represents, only, the rough properties of the image. Dividing an image into blocks and computing the average value can give local information. However, it fails, when there is scaling or translation. Swain and Ballard[19] proposed an image retrieval technique based on color histograms. However, the histogram gives information only about the general color content and does not account for the spatial arrangement which actually plays a major role in the user's perception of the image contents. The problem can be handled by partitioning the image into blocks and using local histograms of these blocks.[20] However, this static partitioning still does not give clues about the spatial arrangements in the image.

Generally, an image database user is interested in objects that captures the semantic information of the image, which cannot be obtained by the low level features. In order to get the semantic information, the image should be segmented into "meaningful" regions. These regions should be "homogeneous" with respect to a set of descriptors and their related similarity measures. Then, a tool for application dependent abstractions is to be developed for object recognition. It is well known that segmentation problem itself is application dependent. Defining a general threshold for finding the homogeneous regions of objects is not possible. In one application one needs to find very fine details, whereas in another, one tries to find large object boundaries. These two queries require conflicting approaches in finding an appropriate threshold for identifying the homogeneous regions.

In this study, we propose a scheme for content-based retrieval of images, using a hierarchical segmentation algorithm based on a color descriptor. The proposed scheme generates a stack of images with "homogeneous regions" which is ordered in a uniformity tree. At the bottom leave of the uniformity tree, there are the pixels of the same colors in the same neighborhood, in $\mathbf{L^*a^*b^*}$ color space. At the root of the tree the image is reduced to "one-color" plane. Each layer corresponds to a segmented image, in the stack. The more we approach to the bottom layer, the more details we reserve in the image. It is evident that, if we inquire a large region with minimal texture, we look for the nodes in higher levels. As we inquire small textured regions, we look for the lower levels.

In the next step, we form a semantic tree, where each node is represented by an object in our query alphabet and the branches correspond to the hierarchical relations among the objects. Finally, the proposed scheme establishes the relation between the homogeneous regions of the uniformity tree and the alphabet of the semantic tree. The coarse-to-fine representation of the regions in the image is used for retrieval simply by representing each object at any level with its bounding box and corresponding color.

2. MULTI-LEVEL SEGMENTATION

In this section, we introduce a multi-level segmentation algorithm based on the "closest colors in the same neighborhood". Then, we introduce two trees: The first one is the uniformity tree, constructed by using the proposed segmentation algorithm. The second one is a semantic tree that consists of the query alphabet. These trees are used to relate the low-level representation of the image with its semantic content.

3. A SIMPLE COLOR DESCRIPTOR TO DEFINE HOMOGENEOUS REGIONS

Let the Euclidean distance, in the $\mathbf{L^*a^*b^*}$ color, space between two neighboring pixels *(k,l)* be

$$E(k,l) = \sqrt{(k_L - l_L)^2 + (k_a - l_a)^2 + (k_b - l_b)^2}.$$

Given an image, two neighboring pixels in 4-neihborhood forms a homogeneous region R, if they have the closest colors in the color palette of an image with respect to he Euclidean distance in $\mathbf{L^*a^*b^*}$ color space. Mathematically speaking,

if $E(k,l) < E(m,n)$, *for all* $(m,n) \in \eta$, *then* $(k,l) \in R$,
where η indicates all the pixels in the image in 4-neighborhood.

In other words, two neighboring pixels are merged into a homogeneous region, when the Euclidean distance between the colors of them has the smallest possible value in the color palette of the image. This descriptor is called as "the closest colors in the same neighborhood".

3.1. Multi-Level Segmentation Algorithm

By using the above metric it is possible to generate a stack of images $I_1, ..., I_M$, starting from the original image and iteratively growing the regions in the image I_j, from the homogeneous regions of image I_{j-1}. In the initial image, the neighboring pixels with the same color automatically forms the homogeneous regions. Then, as the iterations increase, while the homogeneous regions grow in size, the Euclidean distance increases for the closest colors in the same neighborhood.

Let, $\{R_{i,j}\}$, i=1,....,N, be a set of homogeneous regions in an image, in level j. Then, an image at this level can be represented as mutually exclusive and collectively exhaustive set of homogeneous regions as $I_j = \bigcup_{i=0}^{N} R_{i,j}$, for j=1,.....,M.

Notice that a homogeneous region consists of a single pixel if its neighboring pixels have colors that are not most similar in the palette. Also, note that $R_{i,j-1} \subseteq R_{i,j}$ for *j=1,......,M.* In this study the concept of homogeneity is restricted to color. However, the proposed multi-level segmentation scheme can be extended by using shape and texture descriptors in addition to the color descriptor.

The above algorithm can be described as follows:

```
Find the initial homogeneous regions
For each pair of neighboring regions k and l
   Compute the Euclidean distance E(k,l) in the L*a*b* color space and keep in a distance list
Repeat
   Find the region pair (k,l) with the minimum distance
   Merge the regions k and l
   Update the distance list
Until a single region is obtained
```

3.2. Uniformity Tree

The uniformity tree of an image consists of the homogeneous regions of the image stack produced by the multi-level segmentation algorithm. The nodes of the uniformity tree in level j, correspond to the homogeneous regions of image I_j, obtained from the I_{j-1}^{st} image by merging the closest colors in the same neighborhood at the level (j-1). The lowest level of the tree has the original image. As the level of the tree increases, homogeneous regions grow bigger and the number of homogeneous regions which corresponds to the number of nodes decreases.

The uniformity tree eliminates the need for a segmentation threshold and generates a coarse-to-fine representation of an image. At the root of the uniformity tree, the image reduces to a plain color of one region (see; Figure 1) and at each level, the number of regions increases by one. Some examples of multi-level segmentation results are shown in figures 4,5 and 6.

3.3. Semantic Tree

The semantic tree consists of an alphabet, which corresponds to the objects in the query. The content of the alphabet is generated by the user. Each entry in the alphabet may correspond to an object or a background. Suppose that the user defines an alphabet A=a_1,.....,a_n consisting of entries each of which correspond to an object of interests to the user. The alphabet A may consist of a complete or incomplete listings of the image objects, or it can be partitioned

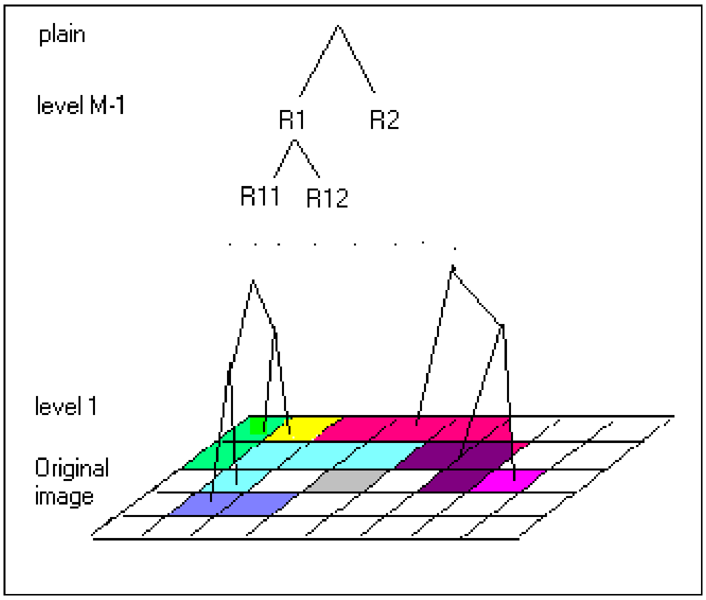

Figure 1. The uniformity tree

into parts of the objects.(see; Figure 2) The alphabet A can, also, be represented in a graph structure, if the objects of interest have structural relations. However, in this study we restrict ourselves to the tree representation.

Given the uniformity tree and the semantic tree of an image, the problem of object based segmentation reduces to associating the nodes of the semantic tree and that of the uniformity tree. For this purpose, each node of the semantic tree should be represented in terms of a set of descriptors, which will be explained in the next section.

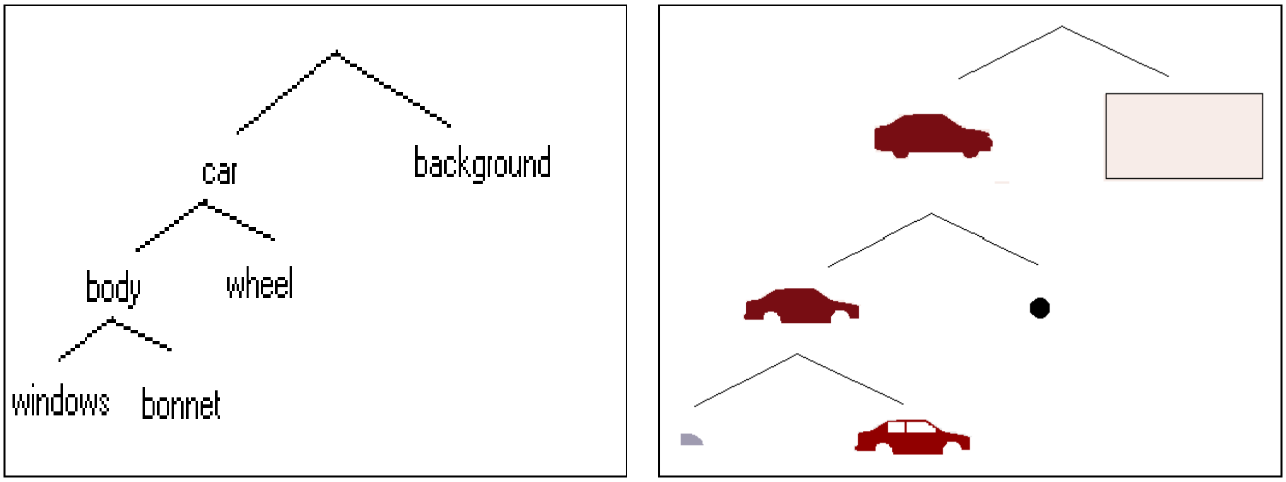

Figure 2. An example of semantic tree extracted from a car image

4. CONSTRUCTION OF SEMANTIC TREE

A simple way of constructing the semantic tree is to represent each node by a template which consists of an object and a background. For this purpose, an image with the object of the query alphabet, is segmented by the proposed multi-level segmentation algorithm and the nodes of the uniformity tree, which best represent the alphabet are selected as the nodes of the semantic tree, corresponding to the objects (see; Figure 2). Note that, the semantic tree has only few nodes, each of which corresponds to an object, whereas the uniformity tree has too many nodes of homogeneous regions, which may or may not represent an object.

Each node is represented with its color and its bounding box which is the tightest rectangular box that fully encompasses the region. This representation provide us a simple semantic tree representation, which is compatible to the uniformity tree. A query image can be easily searched in an image database by ranking the most similar "n" images The user can specify a number of query regions and assign some weigths to this regions. Each query region is searched in the uniformity trees of the images in the database in a top-down fashion. In the experiments, only the color feature is used for the similarity. This can be extended by the shape and texture features of the regions.

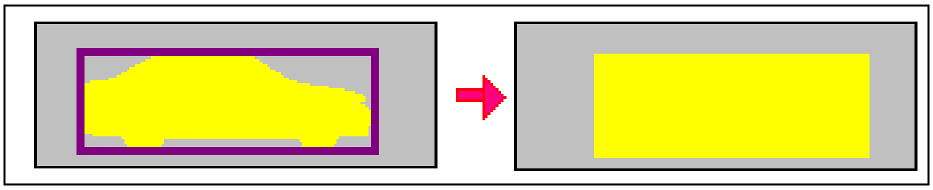

Figure 3. Bounding box representation of an image.

5. EXPERIMENTAL RESULTS

The proposed multi-level segmentation algorithm is tested on 200 raw images, downloaded from the internet containing single object-background, indoor and outdoor images generated from real photography or created artificially. The tests are performed in C programming language under UNIX environment.

The objects of all sizes can be captured by searching the segmentation regions in the uniformity tree. Depending on the application domain, one can find very fine details and as well as large objects. It is observed that, for the images with moderate texture, the proposed scheme extracts the objects close to the root of the uniformity tree. This fact can be observed from the image of the mountain shown in figure 6. As it can be seen from the selected levels of mountain, the slightly textured cloud and mountain is extracted in higher levels, whereas highly textured details, such as flowers, can be captured in relatively lower levels. The details disappear towards the root of the tree, since the proposed scheme has tendency to preserve the large regions as the level of the tree increases. The search algorithm is top-down. In other words, searching an object in the uniformity tree starts from the root. This makes the search fast. Since there is no threshold for segmentation, over-segmentation and under-segmentation is no longer a problem. The object of interest, no matter how large or small, is captured in one of the levels, in the uniformity tree. Therefore, false segmentation in one level is corrected in another level.

Figure 4 and 5 indicates simple objects with smooth backgrounds. Note that the proposed multi-level segmentation method captures the "meaningful" objects at various levels. The only disadvantage of having more than one regions for the same object is the increase in the search time.

In the second group of experiments, the nodes in the semantic tree is formed by query templates and the most similar objects are searched from the image database. In Figure 7, a single car object is given as a query template. The given region, corresponding to the body of a car, is searched in the uniformity trees of the images in the database. We restrict our search to only a few number of levels in order to speed up the query, since we are looking for a large object in our examples. The images are ranked according to their degree of similarities with the query region in color. For complex objects, a single region may not be sufficient for querying. In the examples shown in Figure 8 and 9 we use two regions corresponding to the background and the object in the image.

The experiment indicates very satisfactory results for our image database. The simple color descriptor provides us simple and robust representation for semantic tree. In order to increase the quality of the query, other descriptors, such as texture and shape information can be incorporated to the uniformity and semantic trees.

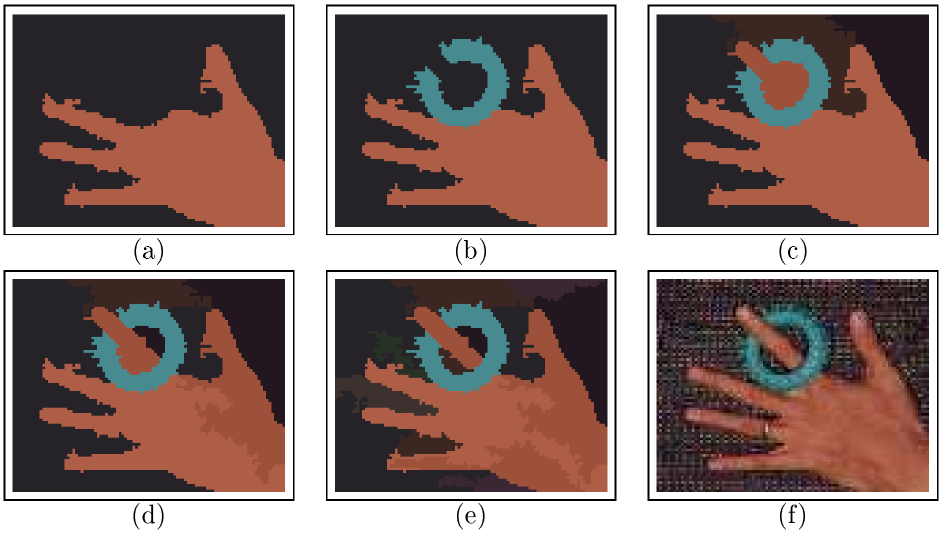

Figure 4. (a)-(e) Segmented images with 2,3,6,9,16 regions, (f) Original image.

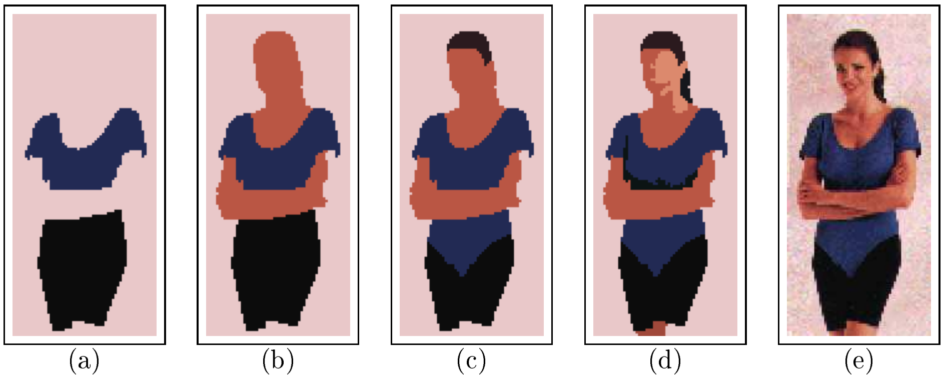

Figure 5. (a)-(d) Segmented images with 3, 5, 7 and 11 regions, (e) Original image.

6. CONCLUSION

Content-based image description for storage and retrieval, requires appropriate utilization of the color, texture and shape information related to the content of the images. The semantics of the image can be obtained by defining the objects and the relations between them that requires the segmentation of the image into "meaningful" regions. Finding a general threshold to segment the image into homogeneous regions corresponding to object is not possible. The multi-level segmentation scheme, proposed in this study, avoids the empirical selection of a threshold. Instead of a single segmentation, the concept of uniformity tree is introduced, where each node correspond to a homogeneous region according to the closest colors in the color palette at any level of the uniformity tree. As a result, a stack of segmented images is obtained with different levels of detail. The semantic tree, proposed in this study is an efficent tool to represent the object alphabet and their relations in the query. When combined with the uniformity tree, it provides a consistent comparison platforms between the objects and the segmentation regions.

This approach has many superiorities compared to classical segmentation based object recognition schemes. First of all, it avoids application dependent selection of the similarity measure and threshold. Secondly, it enables the search of an object in any size and detail. Finally, the top-down search of an object in the uniformity tree provides great efficiency in the search procedure. Experiments indicate superior results compared to classical segmentation and search algorithms for object based retrieval.

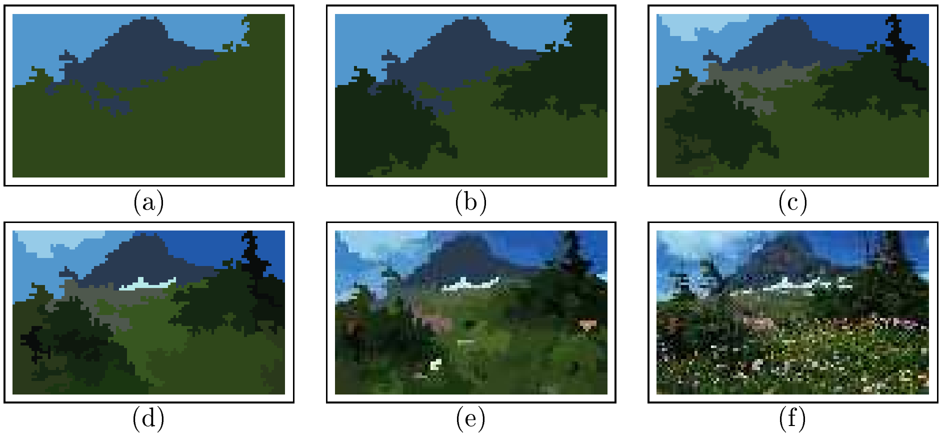

Figure 6. (a)-(e) Segmented images with 3, 5, 7, 9, 15, 60 regions, (f) Original image.

REFERENCES

1. Shapiro,L., Stockman,G., *A new Computer Vision Textbook*, Chapter8. http://www.cs.washington.edu/education/courses/576
2. Rui,Y., Huang,T.S., Chang,S., "Image Retrieval: Current Techniques, Promising Directions, and Open Issues", *Journal of Visual Communication and Image Representation*, vol. 10, no. 4, pp. 39-62, April 1999.
3. Idris,F., panchanathan,S., "Review of Image and Video Indexing Techniques", *Journal of Visual Communication and Image Representation*, vol. 8, no.2, pp. 146-166, June 1997.
4. De Marsicoi,M., Cinque,L., Levialdi,S., "Indexing Pictorial Documents By Their Content: A Survey Of Current Techniques", *Image and Vision Computing*, vol.15, no.2, pp. 119-141, February 1997.
5. Bach,J.R., Paul,S., Jain,R., "A Visual Information Management System For The Interactive Retrieval Of Faces", *IEEE Transactions on Knowledge and Data Engineering*, vol. 5, no. 4, pp. 619-628, August 1993.
6. Jain,A., Vailaya,A., "Shape Based Retrieval: A Case Study With Trademark Image Databases", *Pattern Recognition*, vol. 31, no. 9, pp. 1369-1390, 1998.
7. Faloutsos,C., Barber,R., Flickner,M., Hafner,J., Niblack,W., Petkovic,D., Equitz,W., "Efficient and Effective Querying By Image Content", *Journal of Intelligent Information Systems*, vol. 3, no. 3-4, pp. 231-262, July 1994.
8. Pentland,A., Picard,R.W., Sclaroff,S., "Photobook: Content-Based Manipulation Of Image Databases", *International Journal on Computer Vision*, vol. 18, no. 3, pp. 233-254, June 1996.
9. Smith,J.R., Chang,S.F., "VisualSEEK: A Fully Automated Content-Based Image Query System", *Proceedings of ACM Multimedia 96*, 1996.
10. Smith,J.R., Chang,S.F., "Visually Searching The Web For The Content", *IEEE Multimedia Magazine*, vol. 4, no. 3, pp. 12-20, 1997.
11. Ogle,V.E., Stonebraker,M., "Chabot: Retrieval From Relational Database Of Images", *Computer*, vol. 28, no. 9, pp. 40-48, September 1995.
12. Okon,C., "Image Recognition Meets Content Management For Authoring, Editing And More", *Advanced Imaging*, vol. 10, no. 7, July 1995.
13. Minka,T.P., Picard,R.W., "Interactive Learning Using A Society Of Models", *Pattern Recognition*, vol. 30, no. 4, pp. 565-581, April 1997.
14. Rui,Y., Huang,T.S., Ortega,M., Mehrotra,S., "Relevance Feedback: A Power Tool In Interactive Content-Based Image Retrieval", *IEEE Transactions On Circ. Sys. Video Tech.*, September 1998.
15. Dowe,J., "Content-Based Retrieval In Multimedia Imaging", *Proceedings SPIE Storage and Retrieval For Image and Video Databases*, 1993.
16. Ma,W.Y., Manjunanth,B.S., "Netra: A Toolbox For nevigating Large Image Databases", *International Conference On Image Processing*, 1997.
17. Carson,C., Thomas,M., Belongie,S., Hellerstein,J.M., Malik,J., "Blobworld: A system for region-based image indexing and retrieval", *Third Int. Conf. on Visual Information Systems*, (Visual'99), June 1999.

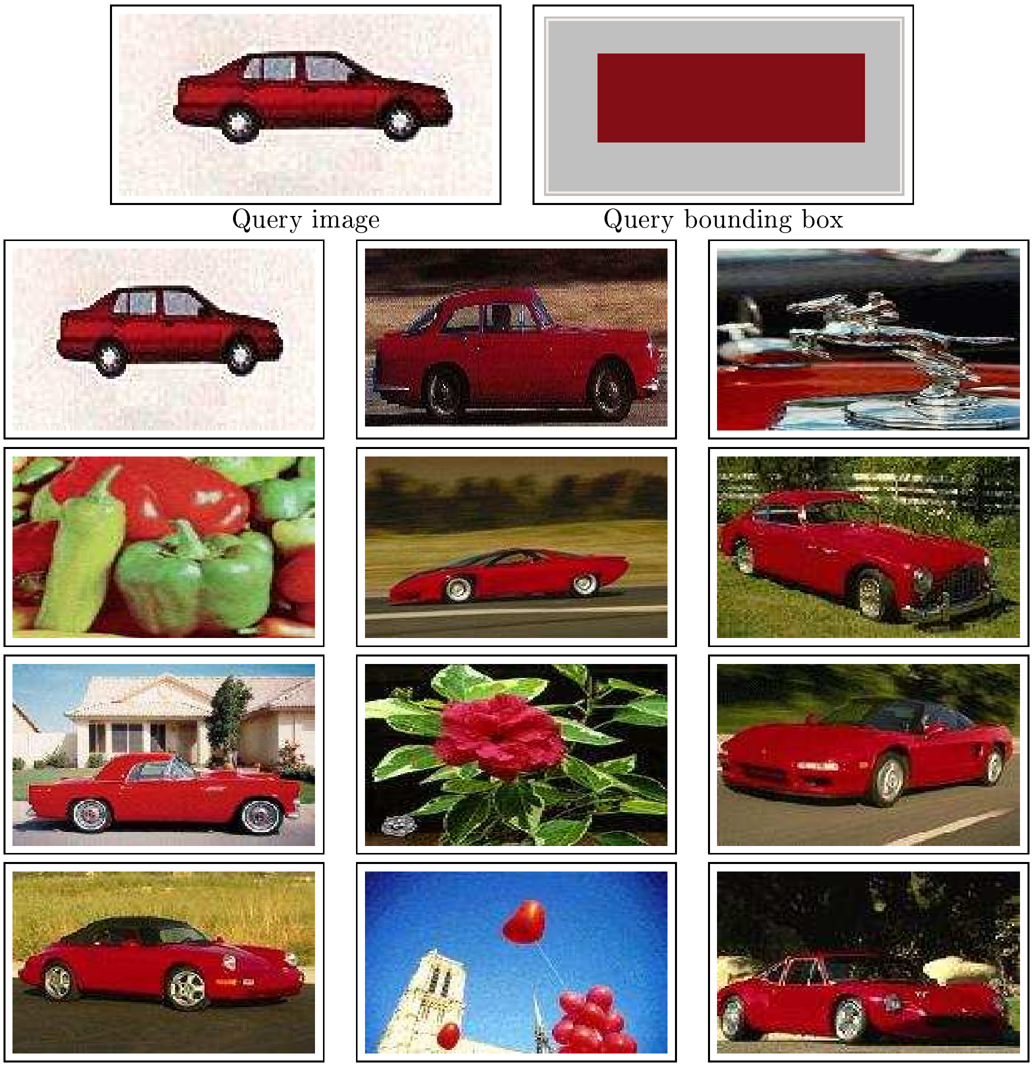

Figure 7. The query template and the closest twelve images in the database.

18. Sethi,I.K., Coman,I., Day,B., Jiang,F., Li,D., Segovia-Juarez,J., Wei,G., You,B., "Color-WISE: A system for image similarity retrieval using color", *Proceedings of SPIE, Storage and Retrieval for Image and Video Databases VI*, Vol. 3312, January 1998
19. Swain, N., Ballard, D., "Color Indexing", *International Journal of Computer Vision*, 7(1), January 1993.
20. Nagasaka, A., Tanaka, Y., "Automatic Video Indexing and Full Video Search for Object Appearances", *IFIP Trans., Visual Database Systems 2*, Knuth, Wegner, 1992, Elsevier.
21. Fuh, C.-S., Cho, S.-W., Essig, K., "Hierarchical Color Image Region Segmentation for Content Based Image Retrieval System", *IEEE Transactions on Image Processing*, vol. 9, no. 1, January 2000
22. Deng, Y., Manjunath, B.S., Shin, H., "Color Image Segmentation", *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition* (CVPR99), 1999
23. Xu,Y., Duygulu,P., Saber,E., Tekalp,M., Yarman-Vural,F.T., "Object Based Image Retrieval Based On Multi-Level Segmentation", *IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP2000), June 2000, Istanbul, Turkey
24. Duygulu,P., Carkacıoğlu,A., Yarman-Vural,F.T., "Multi-Level Object Description: Color or Texture", *First IEEE Balkan Conference on Signal Processing, Communications, Circuits, and Systems*, June 1-3, 2000, Istanbul, Turkey

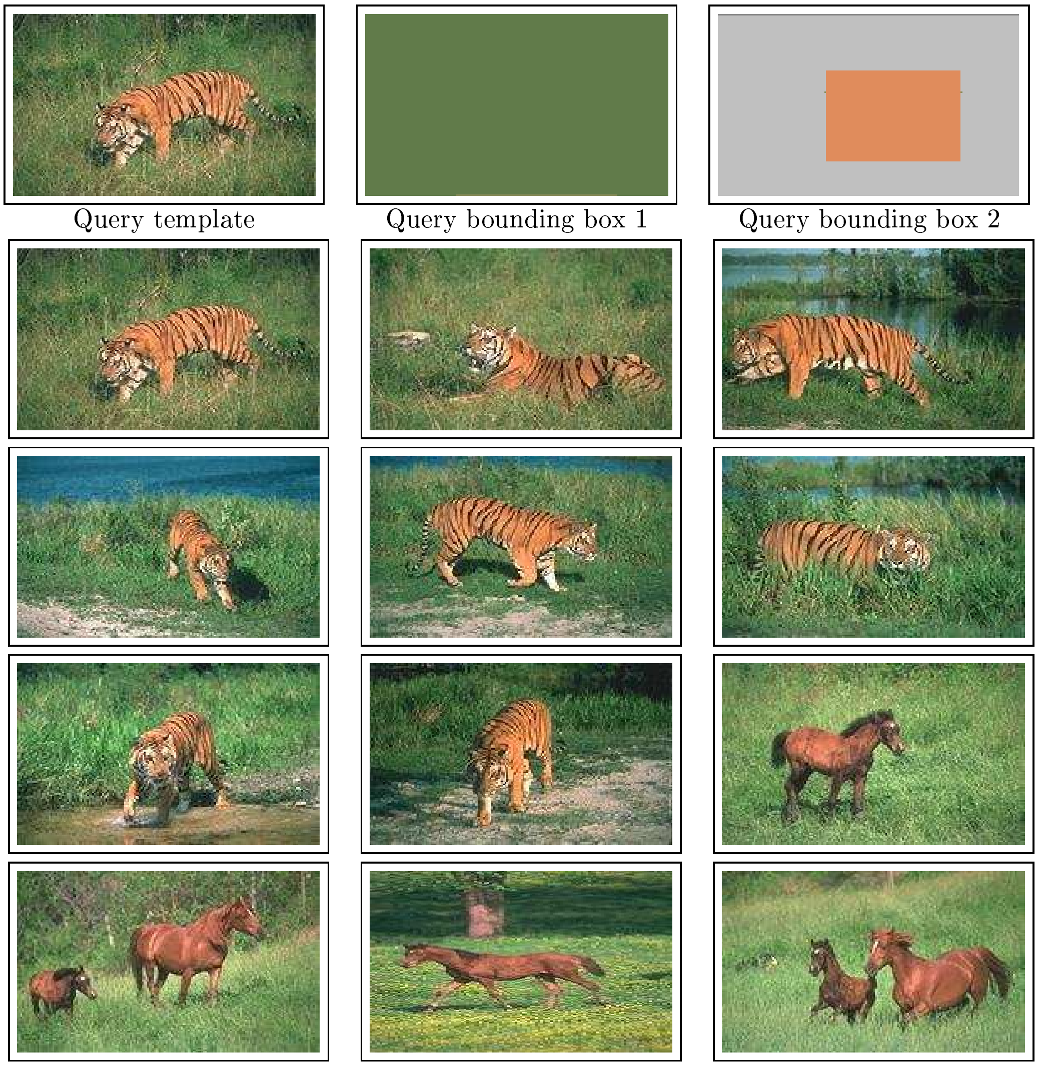

Figure 8. The query template and the closest twelve images in the database.

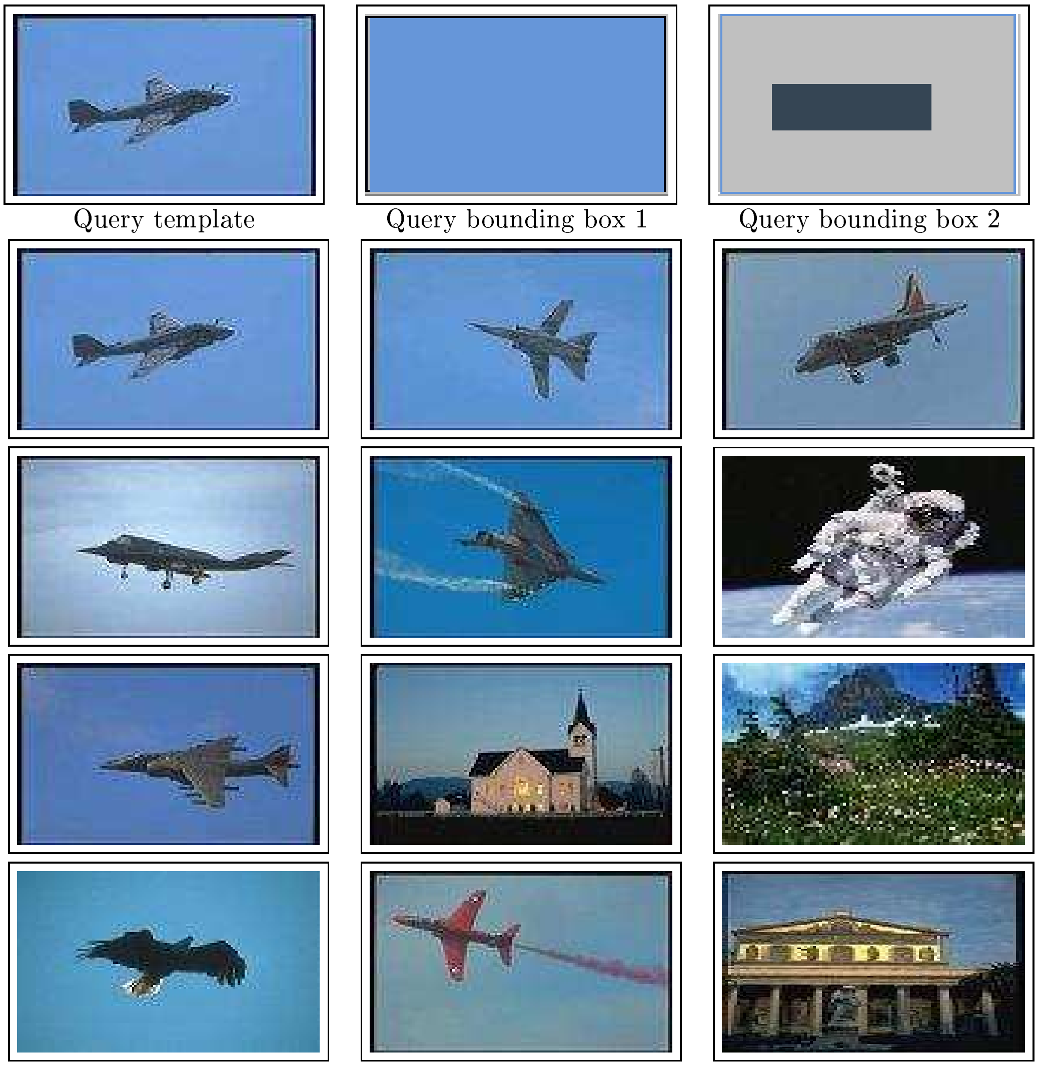

Figure 9. The query template and the closest twelve images in the database.