Object Image Retrieval by Shape Content in Complex Scenes Using Geometric Constraints

Agnés Borràs and Josep Lladós*

Computer Vision Center - Dept. Informàtica, UAB Bellaterra 08193, Spain
{agnesba,josep}@cvc.uab.es
http://www.cvc.uab.es

Abstract. This paper presents an image retrieval system based on 2D shape information. Query shape objects and database images are represented by polygonal approximations of their contours. Afterwards they are encoded, using geometric features, in terms of predefined structures. Shapes are then located in database images by a voting procedure on the spatial domain. Then an alignment matching provides a probability value to rank de database image in the retrieval result. The method allows to detect a query object in database images even when they contain complex scenes. Also the shape matching tolerates partial occlusions and affine transformations as translation, rotation or scaling.

1 Introduction

The goal of Content-Based Image Retrieval (CBIR) is to find all images in a given database that contain certain visual features specified by the user. The reviews of Huang [1] and Forsyth [2] expose a wide variety of feature representations and image retrieval strategies. This work is focused on the development of a CBIR system where the image classification is done according to the shape information. Given the image of an object and a database containing images of complex scenes, the system is able to retrieve those images that likely contain an instance of the object.

In the literature we can find a great variety of shape representation approaches. Some relevant surveys are those of Veltkamp [4], Safar [5], Zhang [3] or Loncarnic[11]. Some retrieval strategies represent the shape taking the information of the whole image. This fact allows to obtain a compact representation that works efficiently in a retrieval application. This is the case of approaches that use shape descriptors such as the shape context [6], the Fourier coefficients or the ART descriptor of the standard MPEG7 [7]. Although these strategies provide relevant results they are not suitable for retrieving objects in complex scenes. In this case it is essential to apply a structural approach that permits to detect a shape as a part of the entire information of an image. Structural approaches on shape representation and matching often use graph based strategies [8][9][10].

* This work has been partially supported by the project TIC2003-09291 and the grant 2002FI-00724.

J.S. Marques et al. (Eds.): IbPRIA 2005, LNCS 3522, pp. 325–332, 2005.

In our work we have chosen a shape representation based on the boundary information. This fact allows us to deal with open shapes and sketches. Query shape objects and database images are represented by polygonal approximations of their contours. Afterwards they are encoded, using geometric features, in terms of predefined structures. Shapes are then located in database images by a voting procedure on the spatial domain. Then an alignment matching provides a probability value to rank de database image in the retrieval result. The method allows detecting a query object in database images even when they contain complex scenes. Also the shape matching tolerates partial occlusions and affine transformations of translation, rotation or scaling.

The stages of our method guide the structure of this paper. In the section 2 we present the shape modelling strategy and we detail the representation of an image. Then, in the section 3 we proceed to define how the modelling information is used in the detection process. Finally, in sections 4 and 5 we present the obtained results and the conclusions of our work.

2 Shape Modelling

A CBIR system analyses the similarity of a given image against a set of images contained in a database. The retrieval needs a previous step where the shape information is modelled. The modelling is applied in the same manner to the query image and to the database images, so we generalize the explanation for any image I.

2.1 Geometric Features of the Shape Elements

To model a shape we use the boundary information polygonally approximated in terms of segments. We assign each one a reference orientation, thus we refer them as vectors instead of segments.

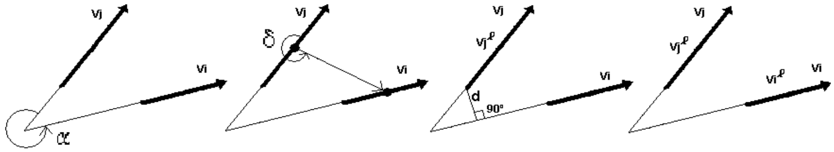
Definition 1. *We define the boundary information BI of an image I as the basic data used to model a shape. Then we denote VI the collection of N vectors v that conform the polygonal approximation of BI.*

In Figure 3 we can see graphically the steps of the retrieval system. Then, from an image I we can see the BI and VI information. The vectors that compose a shape are identified in a unique way with a set of features called *absolute features.*

Definition 2. *$AF(v)$ is defined as an attribute function that, given an image vector v assigns its length, angle and coordinates as absolute features. They are denoted $AF^l(v), AF^\alpha(v)$ and $AF^{(x,y)}(v)$ respectively.*

Notice that the absolute features contain the data that describe the scale, rotation and translation of a vector in an image. On the contrary, we want our system to be invariant to these three affine transformations. This way, instead of working directly from the absolute features we consider them pairwise and extract what we call *relative features*. For the sake of simplicity, we will denote v_{ij} the vector pair (v_i, v_j).

Definition 3. *Given $AF(v_i)$ and $AF(v_j)$ we denote $RF(v_{ij})$ the attribute function that computes the relative features for the vector pair v_{ij} These features are the relative distance, the relative angle, the relative size, and the medium relative angle. We denote them $RF^d(v_{ij}), RF^\alpha(v_{ij}), RF^l(v_{ij})$ and $RF^\delta(v_{ij})$ respectively.*



$$RF^\alpha(v_{ij}) = \alpha \qquad RF^\delta(v_{ij}) = \delta \qquad RF^d(v_{ij}) = \frac{d}{AF^l(v_i)} \qquad RF^l(v_{ij}) = \frac{AF^l(v_j)}{AF^l(v_i)}$$

Fig. 1. Computation of the relative features

The relative features of two vectors can be seen as basic shape descriptors (the Figure 1 shows its computation graphically). In a higher abstraction level, a shape is described in term of primitives that combine these low level descriptors.

2.2 Labelling of the Image Elements with the Primitive Structures

Many shape recognition strategies search for particular line arrangements according to perceptual grouping of salient features [13][12][15]. In our case, we describe the relationships of perpendicularity, parallelism and co linearity due to several predefined structures that we call primitives.

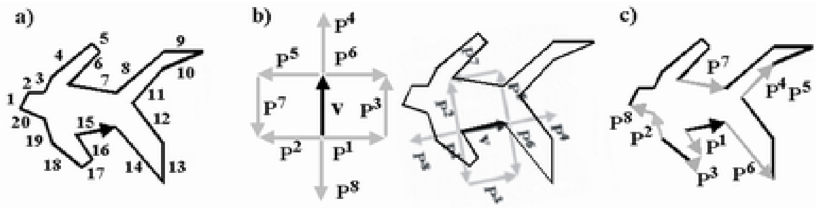
Definition 4. *A primitive P is a particular arrangement of two vectors (w_a, w_r). We denote $\mathcal{P}$ a collection of primitives $\{P^z\}$ $z = 1..N^{\mathcal{P}}$*

Figure 3 shows the set of primitive types we consider due to the arrangement of the vectors w_a and the reference vector w_r (shown in black). The aim of the shape modelling consists in obtaining a local description of the shape around each image vector. The idea is easily shown in the Figure 2. We identify the reference vectors w_r of every primitive on an image vector v_i. Then for every vector w_a we have to find the most similar image vector v_j. We choose v_j to be the most similar if it minimizes a given distance function DRF.

Definition 5. *Given the relative features of an image vector pair, $RF(v_{ij})$, and the relative features of a primitive vector pair, $RF(w_{ra})$, let us denote DRF^α, DRF^δ, DRF^d, DRF^l the distance values for each relative feature and DRF the function that globally quantifies the distance from v_{ij} to w_{ra}.*

$$DRF(RF(v_{ij}), RF(w_{ra})) = max(DRF^\alpha, DRF^\delta, DRF^d, DRF^l)$$

The election of the greatest distance for each feature assures that DRF provides a certain degree of similarity only when all the features accomplish at least this similarity degree. The general idea to compute every DRF value consists in find the difference between the relative feature values and normalize the result by the maximum feasible variation. Every shape element v_i is assigned an attribute vector composed by the distances of the relative features in relation to every primitive. This information is arranged in a table LI where the rows are referred to the vectors and the columns are referred to the primitives (see Figure 2). In other words, every line of LI can be understood as the deformation that the primitive structures have to suffer to fit locally the shape around v_i.



	P^1				...	P^z			
	DRF^α	DRF^δ	DRF^l	DRF^d		DRF^α	DRF^δ	DRF^l	DRF^d
v_1									
v_2									
:									
v_{NI}									

Fig. 2. a) Image vectors b) Arrangement of the primitive vectors w_a due to the identification of v_{15} with every w_r c) The labelling process searches the most similar vectors to every w_a (ex: w_a of belonging P^1 is identified with v_{16}). The modelled information is indexed L[i][z] in the table

3 Shape Detection

Given a query image I_1 we evaluate the retrieval of a database image I_2 in a two step process.

3.1 Location of the Shape: The Voting Process

The shape detection involves a voting procedure in the spatial domain that uses the modelled information of both images and a reference point rp in the image I_1. The evidence combination methods that share these characteristics are typically those based in the generalised Hough transform [16][17].

Definition 6. *Let us define a vote m_{ijO} as the evaluation of the local matching of the vector v_i belonging to I_1 with the vector v_j belonging to I_2 in the orientation O (where O=1 means equal orientation and O=0 means opposite orientation).*

Given two modelled images, the process generates $N_1 x N_2 x 2$ votes. The votes are used to construct an image map M of the shape location.

Definition 7. *Let us name M an image with the same dimensions as I_2 that acts as probability map of the locations of the shape I_1 inside I_2.*

Every vote has a specific location $L(m_{ijO})$ in the map M. This location is found by the transformation of the reference point rp when we match v_i and v_j. Moreover, every vote has a weight $H(m_{ijO})$ that is computed from the information $LI_1[i]$ and $LI_2[j]$. This information describes the distortions of the shapes around the vectors v_i and v_j respect to the primitives. The query shape is detected when the local distortions are similar to those of the database shape. This way, the vote weight $H(m_{ijO})$ will have a high value if $LI_1[i]$ and $LI_2[j]$ are similar. When the vote evaluates v_i on v_j with the same orientation (O=0) we analyse pairwise the information of $L[i]$ and $L[j]$ for each primitive. Otherwise, we compare the information related to the primitives that have the same characteristics but opposite orientation (P^1 with P^5, P^2 with P^6, and so on). A vote with a high weight enforces the probability of finding the query shape in the location $L(m_{ijO})$. The map M, viewed in a 3D representation, shows as peaks the locations where the query shape is more probably located (see the example Figure 3). Then we proceed to validate the shape detection on those positions such $M(x, y)$ exceeds a certain confidence value Thr_M.

3.2 Retrieval Evaluation: The Alignment Process

The generated votes are accumulated with freedom of scale a rotation as peaks in M. Then we validate its coherence using an alignment process on the original contour information of both images, BI_1 and BI_2. Given a vote m_{ijO}, the alignment points are defined by the initial and final coordinates of v_i and v_j. We evaluate the matching by combining the spatial distance between the contours with and the angular information of the normal vectors on shape boundary [17][18]. Then, we use the maximum alignment result of all the peaks in M as a measure to rank the database images in the retrieval process (see Figure 3).

4 Results

We have test our CBIR system with three experiments. The first one is composed by 75 database images and 5 query images of logos. The second deals with 48 database images and 3 possible queries of traffic signs [19]. Finally we have used another set of 30 images and 6 image queries of tin cans. For every query image we have computed the rate of database images that contain the searched object and that have been retrieved in the first n positions (being n the total amount of database images where the query shape can be found). The obtained results for the three tests are 97%, 80%, and 65%. The images of the Figure 4 show the performance of the shape retrieval against transformations of rotation and scale. The traffic sign test shows the robustness of the algorithm against a great amount of information related to the scene. Furthermore, the first image of the third test shows the tolerance of the system against partial occlusion. We have to stand

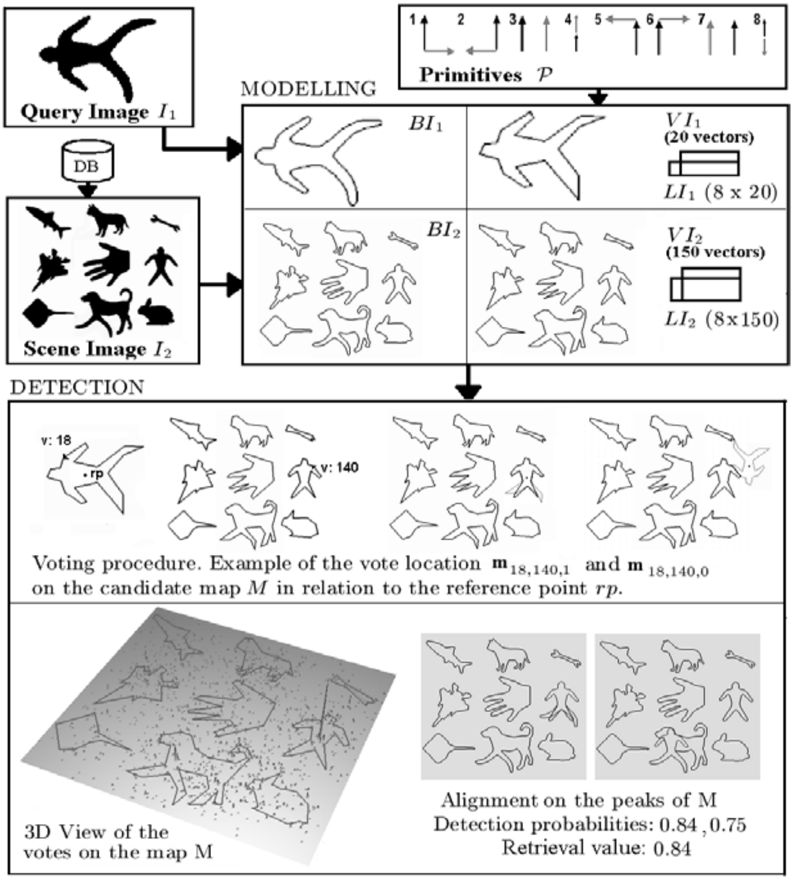


Fig. 3. Example of the algorithm steps

out that the object retrieval in real scenes, such as the tin can example, adds the difficulty of dealing with the effects such as specular reflections, shades or slightly modifications of the viewpoint. These effects are captured by the contour information and affect directly to the vector structures and to the alignment result. Finally we have experimentally set at 0.75 the confidence threshold on the retrieval results as a compromise between the presision and the recall.

5 Conclusions

We have developed a CBIR system of 2D objects by shape content that is capable to deal with databases of complex scenes. The system is modularised in two blocs: the shape modelling and the shape detection. The independence of the two parts allows to precompute the shape representation for any database image. The algorithm has been tested in real scenes to evaluate the robustness of the object location against transformations of scale, rotation and partial occlusions.

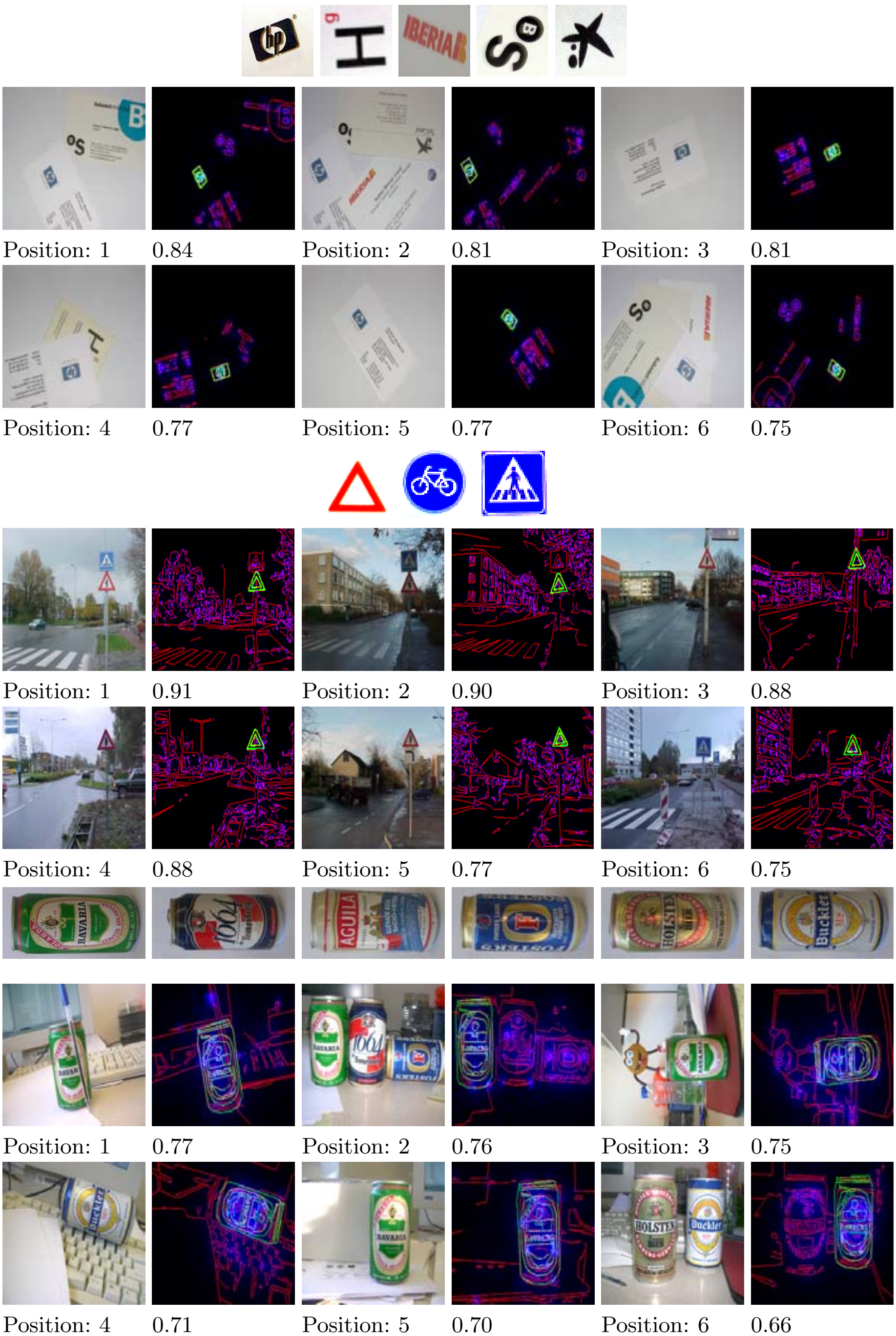


Fig. 4. Retrieval examples: Traffic signs, logos and beer cans. The first line corresponds to the query images. We can see the first 6 results on the leftmost query image, the vectorized image that shows the location solution, and the retrieval value

References

1. Huang, T., Rui, Y.: Image retrieval: Past, present, and future. International Symposium on Multimedia Information Processing, 1997
2. Forsyth, D.A., Malik, J., Fleck, M.M., Greenspan, H., Leung, T.K., Belongie, S., Carson, C., Bregler, C.: Finding Pictures of Objects in Large Collections of Images, Object Representation in Computer Vision (1996) 335-360
3. Zhang, D., Lu, G.: Review of shape representation and description techniques, PR Volume 37, No 1, 1-19, Jan. 2004
4. Veltkamp, R., Hagedoorn, M.: State-of-the-art in shape matching, UU-CS-1999-27, Utrecht University, the Netherlands, 1999
5. Safar, M., Shahabi C., Sun, X.: Image Retrieval by Shape: A Comparative Study, IEEE International Conference on Multimedia and Expo (I), 141-144, 2000
6. Belongie, S., Malik, J.:Matching with shape context, IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL) 2000
7. B.S. Manjunath, P. Salembier, T. Sikora, editors: Introduction to MPEG-7: Multimedia Content Description Interface. Wiley, 2002
8. Huet, B., Cross, A., Hancock, E.R.: Shape Retrieval by Inexact Graph Matching, ICMCS, Volume 1, 772-776, 1999
9. Lourens T., Rolf P. Würtz: Object Recognition by matching symbolic edge graphs, Proc. on the Third Asian Conf.on CV, Hong Kong, 8 - 11, 1998
10. Bunke, H.: Inexact Graph Matching for Structural Pattern Recognition, PRL, Volume 1, No 4, "245-253", 1983
11. Loncaric, S.: A survey of shape analysis techniques, PR, Volume 31, No 8, 983-1001, 1998
12. Stein, F., Medioni, G: Structural indexing: Efficient 2D object recognition, IEEE Transactions on PAMI, Volume 14 , Issue 12 , 1198 - 1204, Dec. 1992
13. Eakins, JP., Shields, K., Boardman, J.: ARTISAN – a shape retrieval system based on boundary family indexing, Storage and Retrieval for Still Image and Video Databases IV. Proceedings SPIE 2670, 17-28, 1996
14. Hu, J., Pavlidis, T.: A hierarchical approach to efficient curvilinear object searching, Computer Vision and Image Understanding Volume 63, Issue 2 , 208-220, 1996
15. Etemadi, A. , Schmidt, J.P., Matas, G., Illingworth, J., Kittler, J.V.: Low-Level Grouping of Straight Line Segments, BMVC91, 119-126, 1991
16. Lee, H.M., Kittler, J.V., Wong, K.C.: Generalised Hough Transform in Object Recognition, ICPR92, 285-289, 1992
17. Gonzalez-Linares, J.M, Guil, N., Zapata, E.L.: An Efficient 2D Deformable Objects Detection and Location Algorithm, PR, Volume 36, Issue 11, 2543-2556, Nov. 2003
18. Huttenlocher, D.P., Ullman, S.: Object Recognition Using Alignment IUW, Volume 87, 370-380
19. Grigorescu, C., Ptekov, N.: Distance Sets for Shape Filters and Shape Recognition, IEEE Transactions on Image Processing, Volume 12, No 10, 1274-1286, Oct. 2003