Similarity-Based Object Retrieval Using Appearance and Geometric Feature Combination

Agnés Borràs and Josep Lladós*

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

Abstract. This work presents a content-based image retrieval system of general purpose that deals with cluttered scenes containing a given query object. The system is flexible enough to handle with a single image of an object despite its rotation, translation and scale variations. The image content is divided in parts that are described with a combination of features based on geometrical and color properties. The idea behind the feature combination is to benefit from a fuzzy similarity computation that provides robustness and tolerance to the retrieval process. The features can be independently computed and the image parts can be easily indexed by using a table structure on every feature value. Finally a process inspired in the alignment strategies is used to check the coherence of the object parts found in a scene. Our work presents a system of easy implementation that uses an open set of features and can suit a wide variety of applications.

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. When these features refer not to the whole image but a subpart, we deal with a problem known as Similarity-Based Object Retrieval (SBOR). Some authors consider two main approaches on the SBOR problem: data-independent and data-dependent [3]. In the data-independent approach images are coarsely divided into rectangular regions where a searched object is mean to be found. Images are indexed from the feature vectors that had been computed using the whole information of the image regions. This fact represents the main advantage on the data-independent systems because classical strategies of CBIR can then be applied to characterize the image from its parts [1] [2]. Otherwise, they involve the hard restriction of dealing with query objects that must fit a rectangular piece of the scene [4] [5]. To overcome this limitation data-dependent approaches deal directly with the particular content of each image. The strategy consists in detecting a set

* This work has been partially supported by the grant UABSCH2006-02.

J. Martí et al. (Eds.): IbPRIA 2007, Part II, LNCS 4478, pp. 33–39, 2007.
© Springer-Verlag Berlin Heidelberg 2007

of invariants from which to decompose the image content in a set of regions. Then, local descriptions of these regions are extracted and represented by feature vectors. Two of the most popular approaches on detecting image invariants are the use of the Harris corner detector and the use of the DoG (Difference of Gaussians) operator. Data-dependent strategies are based on the evidence that a query object is likely to be found in a scene if the feature vectors that describe its parts can be matched in the scene. Even though this criterion represents a useful filter in the retrieval solution, it is not robust enough when the target object constitutes a small portion of the whole scene. To avoid the incorporation of false positives in the query result the system has to check the structural coherence of the object parts found in a scene. This testing process can be performed with techniques as diverse as Hough-like voting strategies [8] or correspondence algorithms such as RANSAC [7]

We present a SBOR system of general purpose that given the image of an object is able to retrieve those cluttered database images that likely contain an instance of this object. The retrieval strategy is based on a data-dependent approach to be flexible enough to handle with a single instance of an object despite its rotation, translation and scale variations. Finally a process inspired in the alignment strategies is used to check the spatial disposition of the object parts. The main contribution of our work is centered in the selection and treatment of the image descriptors. The selection of the image descriptors has to be understood as a compromise between the discriminant power for the content indexing and the tolerance in the similarity matching. Some authors [9] discriminate between the descriptors based on the signal image information [8] and those based on the geometrical properties [6]. In one hand, signal-based descriptors stand out to be very precise and discriminant and, in the other hand, geometrically-based ones provide a suitable encoding of the object structure. Consequently, we propose to use a combination of simple features of both groups instead of using a sophisticated description compacted in a single feature. This way, the feature combination allows a fuzzy computation of the similarity values and provides robustness and tolerance to the retrieval process.

In the next section of this paper we describe the region extraction process and we give a general view of the database features organization. In section 3 we present the two main stages of the object detection strategy: the local matching and the global matching. Section 4 contains some results and finally, in the section 5, we expose the conclusions of this work.

2 Information Modelling

Our retrieval system consists in a data-dependent approach where the image parts are obtained from the polygonal approximation of the contour information. Let us name I an image and v a vector belonging to its polygonal approximation. Every vector has associated an influence area from which the image content is decomposed in parts. These parts are denoted p and are characterized by a set of independent features F. Thus, a set of tables, one for each feature type, provides

an easy system to store and index the image parts. Let us denote T^k the table structure that stores the image information for a certain feature type F^k. The lines of a table are referred to the values comprised in its feature range and the columns are referred to the image parts. A table describes the image content using binary information: a cell $T^k(x, y)$ is set to 1 if the image part p_x has the value y for the feature F^k. Figure 2 exemplifies the extraction of the image parts and feature storage structure.

We distinguish between two kind of features used by our system: the local features F_L and the global features F_G. The local features allow to obtain an independent description of the image parts. Otherwise, the global features are used to establish the relations between these parts and describe their translation, rotation and scale with respect to the whole image. We use a total amount of 14 features distributed in 4 global features and 10 local features (6 based on the signal information and 4 based on the geometric properties). Figure 1 shows them graphically.

In the next section we expose how the features are used in the retrieval process: the local features identify the presence of the image parts and the global ones assure their structural coherence.

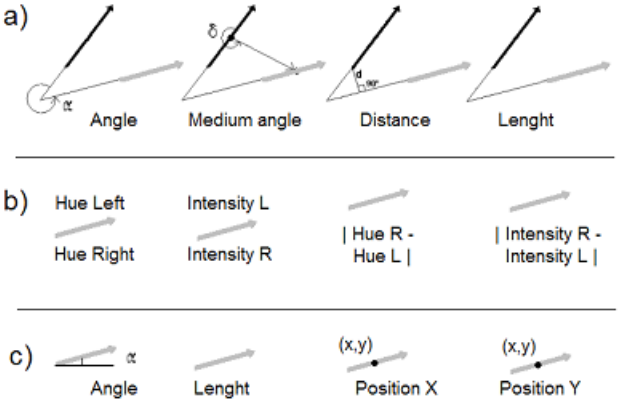

Fig. 1. Image features $F = \{F^k\}$ a) Local features based on the geometric properties of the vectors belonging to the influence area b) Local features based on the signal values sampled on the left and right side along the vector c) Global features

3 Retrieval Process

The retrieval process is divided in two main stages: the retrieval of the query object parts and the analysis of their structural distribution.

3.1 Object Part Identification

Let us name p_i^M a part of the query object and p_j^E a part of a scene. To evaluate the similarity between these two image parts we formulate a query on the

database information for each value of the local features F_L. Let us name $F_i^{M,k}$ the value of the feature k belonging to the query object part p_i^M. Instead of retrieving only the scene parts that match exactly the value $F_i^{M,k}$ we use a similarity function FS that deals with a wider range of solutions. The similarity computation consists in a ramp function that evaluates de difference of the feature values respect to a given tolerance ϵ^k. The result varies in the range of 0 to 1 where 1 means maximum similarity.

$$FS(\epsilon^k, F_i^{M,k}, F_j^{E,k}) = \begin{cases} 0 & \text{if } d > \epsilon^k \\ 1 - \frac{d}{\epsilon^k} & \text{otherwise} \end{cases}$$

$$\text{where} \quad d = |F_i^{M,k} - F_j^{E,k}|$$

The matching between a part of a query object and a part of a scene is evaluated by the mean of the similarity values for all the local features $LFS_{i,j}^{M,E}$. Then, the matching of p_i^M in the whole scene is denoted $ILFS_i^{M,E}$ and it is obtained by the maximum similarity of all the possible comparisons.

$$LFS_{i,j}^{M,E} = \frac{\sum FS(\epsilon^k, F_i^{M,k}, F_j^{E,k})}{\#F_L} \quad | \quad k \in F_L$$

$$ILFS_i^{M,E} = max\{LFS_{i,j}^{M,E}\} \quad \forall j \in p_j^E$$

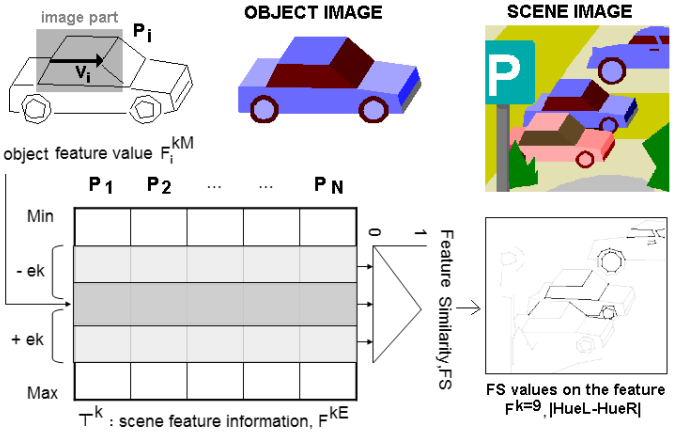

Fig. 2. The figure shows the vectorization of an object image and an image part example (shaded region) belonging to one of its vectors. A binary table contains the feature information of the scene image. An example shows the similarity results FS for every scene vector respect to the object value F_i^k (where k=9 for the feature $|HueR-HueL|$). We represent in black the scene vectors with maximum similarity.

Figure 2 shows an example of similarity computation FS for a single feature F_k. Moreover, Figure 3 illustrates the combination values $ILFS$ between the signal features and the geometric ones.

Some retrieval systems select those database images that present the highest accumulation of the local part identification similarities $ILFS$. This single criterion does not check the coherence of the spatial arrangement of the object parts. Thus, a large amount of false positives can be introduced in the retrieval solution. To solve this problem, our proposal introduces a final phase where the global structure is tested for the local matching pairs with highest score.

3.2 Checking of the Structural Arrangement of the Object Parts

Given a vector of the model object image v_q^M and a vector of the scene image v_r^E we can define an alignment of both image contents by computing the affine geometric transformations to map v_q^M on v_r^E in the orientation O (the same or opposite). These geometric transforms can consist in changes of scale, rotation, and translation. As we have introduced in the section 2 the features that describe the image content in relation to the whole image aspect are identified as global features F_G. This way, the alignment transformations only affect to the global features of the query object.

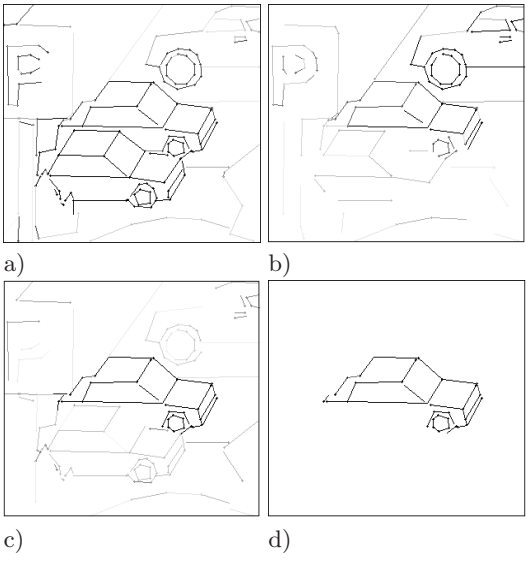

Fig. 3. a) Similarity values $ILFS$ of the object parts (black means maximum similarity) using signal-based features b) using geometrical features c) using both feature groups d) Vectors representing the best matching solution, maximum $IGFS$ value, for both feature groups. Notice the collaboration between the signal based-based features that match the cars by their color and the geometrical-based ones that match them by shape.

Let us name F' the modified global feature values of the query object according to a given vector alignment. The object similitude is computed using the same strategy as the local one but adding a hard restriction to the global feature values. To preserve the spatial disposition of the object parts is necessary the similarity of all the features values to be accomplished. The following function, GFS, describes the calculus of the similarity between an object part p_i^M and a scene part p_j^E given a fixed alignment.

$$GFS_{i,j,(q,r,O)}^{M,E} = min\{LFS_{i,j}^{M,E}, FS(\epsilon^k, F_i'^{M,k}, F_j^{E,k})\}$$

$$\forall k \in F_G$$

Then the similarity between the query object and the scene image correspond to the best result provided by the function $IGFS$ on the checked alignments.

$$IGFS_{(q,r,O)}^{M,E} = max\{GFS_{i,j,(q,r,O)}^{M,E}\}$$

$$\mid LFS_{i,j}^{M,E} > Thr, \quad \forall i \in p_i^M, \quad \forall j \in p_j^E$$

The computed value is used in the retrieval process to rank the solutions of given query. The example of the Figure 3 d) shows the object detection solution as the scene vectors with maximum $IGFS$ value.

4 Results

We have tested the system with 72 images belonging to two databases. The first database consists in a set of 40 images of invoices that can be identified by 4 different logos. The other database is conformed of 32 scenes where 4 objects

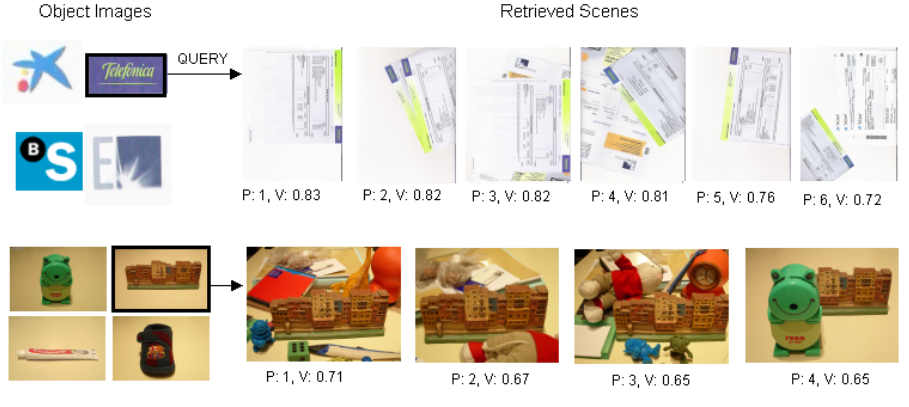

Fig. 4. Query examples on the selected objects. Every retrieved image has its position (P) and retrieval value (V).

can be found. 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 object can be found). The obtained results for both tests are 92% of success.

We have observed that the variations that mainly affect to the retrieval measure $IGFS$ are caused by illumination changes and viewpoint distortions. Nevertheless the success on the object location is maintained due to the feature combination and the effect of the query tolerance ranges. Figure 4 illustrates the results with two examples.

5 Conclusions

We have developed a SBOR system that deals with a combination of independent image features that provides a fuzzy value on the similarity comparison of the image parts. A future research line of our work is centered in the development of a process that initially analyzes the query image and adapts the similarity tolerances according to the most characteristic features of the query object. The system has proved to be robust against effects such as noise, shades, slightly modifications of the viewpoint and partial occlusions. We have tested the system with two databases of scanned documents and images of objects taken in real environments obtaining promising results.

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, pp. 335–360 (1996)
3. Fodor, I.K.: Statistical Techniques to Find Similar Objects in Images. In: Proceedings of the American Statistical Association (October 2003)
4. Luo, J., Nascimento, M.A.: Content Based Sub-Image Retrieval via Hierarchical Tree Matching. In: Proceedings of ACM MMDB 2003, New Orleans, USA, pp. 63–69, (November 2003)
5. Lewis, P.H., Martinez, K., Abas, F.S., Ahmad Fauzi, M.F., Addis, M., Lahanier, C., Stevenson, J., Chan, S.C.Y., Mike, J.B., Paul, G.: An Integrated Content and Metadata based Retrieval System for Art. IEEE Trans. on Image Proc. 13(3), 302–313 (2004)
6. Huet, B., Cross, A., Hancock, E.R.: Shape Retrieval by Inexact Graph Matching. ICMCS 1, 772–776 (1999)
7. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. Image Vision Computing 22(10), 761–767 (2004)
8. Lowe, D.G.: Object Recognition from Local Scale-Invariant Features. In: ICCV, pp. 1150–1157 (1999)
9. Lamiroy, B., Gros, P., Picard, S.: Combining Local Recognition Methods for Better Image Recognition. Vision 17(2), 1–6 (2001)