A New Content Based Image Retrieval Method Based on a Sketch-Driven Interpretation of Line Segments

Marco Anelli[2] and **Alessandro Micarelli**[1,2] and **Enver Sangineto**[1]

1. Centro di Ricerca in Matematica Pura e Applicata (CRMPA), Sezione "Roma Tre".
Via della Vasca Navale 79, 00146 Roma, Italia. sanginet@dia.uniroma3.it
2. Dipartimento di Informatica e Automazione, AI Lab, Università degli Studi "Roma Tre".
Via della Vasca Navale 79, 00146 Roma, Italia. {anelli,micarel}@dia.uniroma3.it

Abstract

We present a new method for image retrieval by shape similarity able to deal with real images with not uniform background and possible touching/occluding objects. First of all we perform a sketch-driven segmentation of the scene by means of a Deformation Tolerant version of the Generalized Hough Transform (DTGHT). Using the DTGHT we select in the image some candidate segments to be matched with the user sketch. The candidate segments are then matched with the sketch checking the consistency of the corresponding shapes.

1 Motivations and Goals

In this paper we present a new method for Content Based Image Retrieval (CBIR) based on the analysis of the object shapes. In [Anelli *et al.*, 2002] we have proposed a CBIR technique based on a Deformation Tolerant version of the well-known Generalized Hough Transform (DTGHT). In this paper we start from that work and we go further with the objective to augment the system precision. We introduce some segmentation rules corresponding to Gestalt principles (see, for example, [Ullman, 1996] and [Koffka, 1935]) exploiting the continuity properties of the edge segments and a more accurate matching method which compares the searched shape with the image lines. We have obtained an acceptance/rejection rule able to classify a data base image as *relevant* or *not relevant* with respect to the user's query.

The novelties of our approach with respect to similar systems is its capability to deal with every kind of real images, without caring of light conditions, with not uniform backgrounds and possible occluding objects.

2 Pre-Processing and Sketch Localization

Each image of the system data base is off-line pre-processed in order to extract an edge map and to reduce the noise. We perform a standard edge extraction and thinning process using the Canny edge detector with Sobel 3×3 masks [Shapiro and Stockman, 2001]. After this standard pre-processing, we use two *salience filters* to erase those thick textured areas which deteriorate the retrieval process (for details, see [Anelli *et al.*, 2002]). From now on we call I the image's edge map after these pre-processing phases.

We are now interested in merging the pixels in I with their neighbors in order to build line segments. We define a *segment* s as an ordered list of points belonging to I: $s =< p_1, p_2, ...p_{n_s} >$ such that each p_i $(i = 1, ..., n_s)$ is adjacent to p_{i-1} and p_{i+1} in a 8-connected interpretation of pixel adjacency. Moreover, p_1 and p_{n_s} (the segment end-points) can alternatively be such that: (1) they have only one adjacent point; or (2) they are a junction with another segment, i. e., they have more than 2 adjacent points. All the other points p_i $(i = 2, ..., n_s - 1)$ have exactly 2 adjacent points. This segmentation process is performed because pixels that are part of the same continuous segment are more likely to belong to the same object.

On-line, the system accepts a user drawn query representing the shape she/he is interested to find in the data base. The sketch S is processed in order to extract the R-Table T [Ballard, 1981]. We use the following algorithm:

Sketch Representation Construction(S)
1 Compute the centroid p_r of S.
2 For each point p_k $(k = 1, ..., m)$ of S,set:
$T[k] := p_r - p_k$
where $m = |S|$ (m is the cardinality of S).

Once the R-Table is computed, we compare such sketch representation with the previously processed data base images with the aim to localize a region in each image which maximizes the likelihood of the presence of a shape similar to the sketch. This phase is important because we do not manually select in the images the interesting objects from their background like the most working CBIR systems presently do. We adopt a Deformation Tolerant version of the Generalized Hough Transform (DTGHT) [Anelli *et al.*, 2002].

The main difference with the Generalized Hough Transform (GHT)[Ballard, 1981] concerns the voting phase. Here we increment all those accumulator's positions belonging to a bi-dimensional squared range centered in p_r (called the *voting window* $W(p_r)$), for each edge point p of I and each vector $v = p_r - p$ in T. This is efficiently achieved using a dynamic programming technique and splitting the voting process in two sequential phases: the *exact voting phase* and the *spreading phase* (for details, see [Anelli *et al.*, 2002]). If M is the point with the highest score in the final Hough accumulator (A) after the spreading phase, then we take it as the

most likely position for the center of mass of a possible shape similar to S in I. In other words, if such a shape exists in I, M is its most likely centroid. Experimental results presented in [Anelli *et al.*, 2002] confirm this assumption. Nevertheless, differently from the original GHT, in which the score of the accumulator in its maximum *is equivalent to the number of points of S in I*, with a DTGHT approach, due to the augmented voting area, it is sometimes possible that a thick and random distribution of points and segments in I can produce a high value $A[M]$ not corresponding to a shape really similar to S. In order to reject such cases, in the next section we perform some tests on the segments distribution around M.

3 Verification Phase

In the following we assume that M is the point in I found as described in Section 2, $m = |S|$ and W is the voting window. The aim of the verification phase we are going to explain in this section is to verify if the disposition of the segments in I around M is really perceptually similar to the S' shape. In order to avoid false positives, first of all we check the *portion* of S which can be matched with segments in I. This is intuitively done by projecting S on I (let us call S' such projection, with M its centroid), looking for segments on I in the the neighborhood of S' and marking the portion of S *justified* by these segments. Furthermore, we need to take into account *inhibitory* segments intersecting the neighborhood of S' but not really following the shape S. In the following we give the operational definitions for justifying and inhibitory segments.

A point p belonging to I is called a *valid* point if:

$$\exists i \in [1, m] : p + T[i] \in W(M) \wedge |\phi_I(p) - \phi_S(i)| \leq \alpha, \quad (1)$$

where α is a suitable threshold and $\phi_I(p)$, $\phi_S(i)$ are, respectively, the edge orientation of the point p in I and the edge orientation of the i-th point of S. We call i a *valid hypothesis* for p. A segment s_i is a *justifying* segment if:

$$|V_i| \geq k_1 |s_i|, \quad (2)$$

where k_1 is a fixed threshold (presently $k_1 = 0.7$) and V_i is the set of all the valid points of the segment s_i. Let J be the subset of I's segments which are justyfing segments.

Finally, a segment s_i is called an *inhibitory* segment if s_i is not a justifying segment and:

$$k_3 |s_i| \geq |F_i| \geq k_2 |s_i|, \quad (3)$$

where k_3 and k_2 are two fixed thresholds (presently $k_2 = 0.2$, $k_3 = 0.4$) and F_i is the set of all the points p in s_i which satisfy:

$$\exists i \in [1, m] : p + T[i] \in W(M). \quad (4)$$

Let H be the subset of all the inhibitory segments of I. The verification test is given by the logical and of the justification test and the inhibitory test.

Justification Test(J)
1 For $k \in [1, m]$, set the local boolean vector $B[k] := false$.
2 For each $s_i \in J$ do:
3 For each point $p \in V_i$ do:
4 For each valid hypothesis k, set $B[k] := true$.
5 Return:= $|\{B[k] : B[k] = true\}| \geq k_4 m$

Table 1: Precision and recall scores for the three users. The columns show, for each user, the precision and recall values obtained for each query type.

Query	User 1		User 2		User 3		Average	
Car	0.94	0.65	0.81	1	0.91	0.81	0.88	0.82
Pistol	1	0.70	0.93	0.70	0.86	0.60	0.93	0.74
Racket	0.90	0.50	0.73	0.85	1	0.95	0.86	0.76
Watch	0.28	0.10	0.66	0.30	0.75	0.15	0.54	0.18
Bottle	0.60	0.45	0.75	0.60	0.93	0.7	0.76	0.58
Guitar	0.93	0.75	0.94	0.80	0.66	0.50	0.85	0.68

In the above algorithm k_4 is a threshold, currently fixed to 0.7. We check in this way that at least a portion k_4 of S is present in I in the right position with respect to M.

Finally, the following is the inhibitory test:
Inhibitory Test(H,J) $:= \Sigma_{s_i \in H} |s_i| < \Sigma_{s_i \in J} |s_i|$.

4 Experimental Results

We have tested our method with a data base composed of 283 images randomly taken from the Web. They show a great variety of subjects: 26 cars, 22 horses, 20 guitars, 20 pistols, 20 tennis rackets, 20 saxophones, 20 bottles, 20 watches, 15 crucifixes, 10 vases. Moreover, the data base contains different other subjects such as: animals, faces, landscapes, airplanes, boats, fruits, mushrooms, cups and so on.

No simplifying assumption has been done about images: they usually have not uniform backgrounds and often the retrieved objects are occluded by other objects. No manual segmentation has been performed on the images.

We paid attention to choose as users people not aware of the data base images' shapes. Indeed, only one of the three users in Table 1 has had the possibility to observe the repository images before his drawings (User 3). Each user has performed six queries. The drawn subjects are: a car, a pistol, a tennis racket, a watch, a bottle and a guitar. We report our results in Table 1. Average precision and recall for all the three users and all the queries are, respectively: 0.80 and 0.62.

References

[Anelli *et al.*, 2002] M. Anelli, A. Micarelli, and E. Sangineto. A deformation tolerant version of the generalized hough transform for image retrieval. In *Fifteenth European Conference on Artificial Intelligence (ECAI 2002), Lyon, France*, 2002.

[Ballard, 1981] D. H. Ballard. Generalizing the Hough Transform to detect arbitrary shapes. *Pattern Recognition*, 13. No. 2:111–122, 1981.

[Koffka, 1935] K. Koffka. *Principles of Gestalt Psychology*. New York: Harcourt, Brace and World, 1935.

[Shapiro and Stockman, 2001] L. Shapiro and G. Stockman. *Computer Vision*. Prantice hall, 2001.

[Ullman, 1996] S. Ullman. *High–level Vision. Object Recognition and Visual Cognition*. A Bradford Book. The MIT Press Cambridge, Massachusetts, 1996.