Image Retrieval by Elastic Matching of User Sketches

A. Del Bimbo[1] and P. Pala[2]

[1] Dipartimento di Elettronica per l'Automazione,
Università degli Studi di Brescia, via Branze, 38, 25133 Brescia, Italy
[2] Dipartimento di Sistemi e Informatica,
Università degli Studi di Firenze, via S. Marta 3, 50139 Firenze, Italy

Abstract. Image retrieval by contents from database is a major research subject in advanced multimedia systems. The intrinsic visuality associated with pictorial data suggests the use of iconic indexes and visual techniques to perform retrieval effectively. In this paper we present a method for image retrieval based on sketches of object shapes. In our method, the shape drawn by the user is deformed to match as well as possible the objects in the images. The degree of match achieved, and the elastic deformation energy spent to achieve such a match are used as a measure of the similarity between the template and the image object. The elastic matching is integrated with arrangements to provide for scale invariance, to take into account rotations and spatial relationships between objects for multiple-object queries.

1 Introduction

The intrinsic visuality of the information contents associated with pictorial data advises against the use of indexing and retrieval based on textual keywords as traditionally used in text documents. Iconic indexes have been proposed in [8] to effectively support image retrieval by contents. Iconic indexes are in the form of symbolic descriptions of pictorial data or pictorial data relationships but may also include the actual values of object features, or be in the form of abstract images taking the salient features of the original image. The use of iconic indexes naturally fits with the accomplishment of image retrieval according to visual querying by-example. Once iconic indexes have been built, the user reproduces on the screen an approximated visual representation of pictorial contents of images to be retrieved. Retrieval is reduced to the matching of the user visual representation against iconic indexes in the database. Visual queries by-example for pictorial data exploit human natural capabilities in picture analysis and interpretation and largely reduce the cognitive effort of the user in the access to the database. A number of techniques have appeared in the literature which deal with iconic indexing and visual querying by example of single images; different approaches in performing queries are related to the type of facets of pictorial data that are taken into account. Indexing and querying based on spatial relationships have been proposed in [2], [1], [4], [3]. Spatial relationships are represented symbolically through *2D strings* according to the *symbolic projection* approach.

2D-strings encode the positional relationships between the projections of the objects on two reference coordinate axes. Indexing and querying based on picture color distribution or object texture organization has been proposed in [7] and [6], respectively: images are requested that contain object colors and textures similar to those selected from a menu; matching is performed by comparing color histograms or the Euclidean distance in the texture space. Retrieval by-contents based on similarity between imaged object shapes and user-drawn sketches has been proposed by a few authors [6], [5]. Unlike indexing and retrieval by colors or textures, or spatial relationships, here the problem is complicated by the fact that shape does not have a mathematical definition that exactly matches what the user feels as a shape. In this paper, we present a system for image retrieval-by-contents based on shape matching with elastic deformations. In our system the user sketches a drawing on the computer screen and this is deformed to adapt to shapes of objects in the images. For queries including multiple objects, spatial relationships between the different objects are taken into account.

2 The elastic approach to shape matching

Suppose we have a one-dimensional template, modeled by a first order spline $\tau : \mathbf{R} \mapsto \mathbf{R}^2$. We will always assume that the template is parametrized with respect to arclength, and normalized so as to result of length 1. We have an image $I : \mathbf{R}^2 \mapsto [0,1]$ – we suppose the luminance at every point normalized in $[0,1]$ – that we search for a contour with a shape *similar* to that of τ. To make a robust match even in the presence of deformations, we must allow the template to warp. If $\theta : \mathbf{R} \mapsto \mathbf{R}^2$ is the deformation, then the two components of the deformed template (also parametrized with respect to arclength) are given by: $\phi_j(s) = \tau_j(s) + \theta_j(s)$. The template must warp taking into account two opposite requirements. First, it must follow as closely as possible the *edges* of the image. The second requirement to take into account is the deformation of the template. Allowing arbitrary deformations every template matches every image, and results in a mathematically ill-posed problem. In order to discover similarity between the original shape of the template and the shape of the edge areas on the image, we must set some constraint on deformation. In this way our goal is to minimize the compound functional:

$$\int_0^1 \alpha \left[\left(\frac{d\theta_x}{ds} \right)^2 + \left(\frac{d\theta_y}{ds} \right)^2 \right] + \beta \left[\left(\frac{d^2\theta_x}{ds^2} \right)^2 + \left(\frac{d^2\theta_y}{ds^2} \right)^2 \right] - \left| \nabla I \left(\phi(s) \right) \right|^2 ds. \quad (1)$$

The quantity depending on the first derivative is a measure of how the template τ has been *locally stretched* by the deformation θ, while the quantity depending on the second derivative is an approximate measure of the energy spent to *locally bend* the template. The match between the deformed template and the edges in the image can be measured through the third integral of the compound functional (1). Using the gradient descending technique, the solution θ can be approximated as a third order spline. To support template deformation the edge

image is blurred with a gaussian filter. The elastic energy depends only on the first and second derivatives of the deformation θ. This allows to do not penalize discontinuities and sharp angles that are already present in the template, but to penalize only the degree by which we depart from those discontinuities or angles. Also, since the energy depends only on the derivatives of θ, pure translation of the template, for which θ is constant, does not result in additive cost. This makes our scheme inherently translation invariant.

2.1 Template Matching

After a template reached convergence over an image shape, we need to measure how much the two are *similar*. Again, the similarity is a fuzzy concept, and to measure it we need to take into account a number of things. A first thing to take into account is, of course, the degree of overlapping between the deformed template and the gradient of the image. This can easily be measured as:

$$\mathcal{M} = \int_0^1 [\nabla I(\tau(s) + \theta(s))]^2 \, ds \tag{2}$$

Another factor to consider is how much the template had to warp to achieve that match. We use two different measures of deformation, corresponding to the two energy terms used in the compound functional (1): *Strain energy* ($\mathcal{S}$) and *Bend energy* ($\mathcal{B}$) defined as:

$$\mathcal{S} = \int_0^1 \left(\frac{d\theta_j}{ds} \right)^2 ds \qquad \mathcal{B} = \int_0^1 \left(\frac{d^2\theta_j}{ds^2} \right)^2 ds \tag{3}$$

Coefficients $\mathcal{M}$, $\mathcal{S}$, $\mathcal{B}$ alone are not enough to operate a good discrimination between different shapes. In our approach constraints are imposed by considering the changes of the number $\mathcal{C}$ of zeroes of the curvature function. So another factor to take into account is the variation of the number of zeroes of the curvature function during the deformation process; that is $\mathcal{C} - \mathcal{NC}$. All these 5 parameters $(\mathcal{M}, \mathcal{S}, \mathcal{B}, \mathcal{C}, \mathcal{C} - \mathcal{NC})$ are classified by a back-propagation neural network suitable trained. The neural classifier gives one output value ranging from 0 to 1, which represents the similarity between the shape in the image and the template.

3 Spatial Relationships

Another important source of information about the scene represented in an image is the spatial disposition of objects. In our approach each object is represented by its minimum enclosing rectangle (MER); we use spatial relationships between rectangles both as a mean to filter the database and as a mean to make a more precise multi-objects query. We do this by a slight modification of a method developed in [1]. Projection of such rectangle on the two coordinate axes determine begin and end boundaries of the object. In this way each object is represented by four boundaries: begin (bb) and end (eb) boundaries in the x-axis direction and begin (bb) and end (eb) boundaries in the y-axis direction.

Boundaries can then be sorted introducing two precedence operators: "$<$" (left-right, below-above) and "$=$" (same location). All possible relations between two objects may be ranked in five categories based on relations among object's boundaries:

$$\begin{aligned}
A \textit{ disjoint } B &= \{[bb_x(A) < eb_x(B)] \vee [bb_x(B) < eb_x(A)] \vee \\
&\quad [bb_y(A) < eb_y(B)] \vee [bb_y(B) < eb_y(A)]\} \\
A \textit{ meet } B &= \{[eb_x(A) = bb_x(B)] \vee [bb_x(A) = eb_x(B)] \vee \\
&\quad [eb_y(B) = bb_y(A)] \vee [bb_y(A) = eb_y(B)]\} \\
&\quad \wedge \tilde{}(A \textit{ disjoint } B) \\
A \textit{ contain } B &= \{[eb_x(A) \leq eb_x(B)] \wedge [bb_x(B) \leq bb_x(A)] \wedge \\
&\quad [eb_y(A) \leq eb_y(B)] \wedge [bb_y(B) \leq bb_y(A)]\} \\
A \textit{ inside } B &= \{[eb_x(B) \leq eb_x(A)] \wedge [bb_x(A) \leq bb_x(B)] \wedge \\
&\quad [eb_y(B) \leq eb_y(A)] \wedge [bb_y(A) \leq bb_y(B)]\} \\
A \textit{ partly overlap } B &= \tilde{}(A \textit{ disjoint } B) \wedge \tilde{}(A \textit{ meet } B) \\
&\quad \wedge \tilde{}(A \textit{ contain } B) \wedge \tilde{}(A \textit{ inside } B)
\end{aligned}$$

To make the system's description coherent with that operated by our visual perception, four orientation parameters O_1, O_2, O_3, O_4 are introduced that represent relations as NORTH-SOUTH EAST-WEST. While in [1] these four parameters are defined based on precedence relationships among object's boundaries, we found that a more appropriate definition of these parameters for the relation of object A with respect to object B considers the position of centroid of A with respect to boundaries of object B. The spatial relation of object A with respect to object B is represented by a symbolic 5-tuple: $R(A,B) = [C, O_1, O_2, O_3, O_4]$, where C is the category (*disjoint*, *meet*, *contain*, *inside*, *partly overlap*) which the spatial relation belongs to. In order to speed up the spatial matching process a binary codeword *signature file* is associated with each image encoding mutual spatial relationships, and matching is performed through a hash function.

4 Image Retrieval System

Based on shapes and spatial relations representations previously discussed, a system has been developed for image retrieval by contents. In the databases raw images are passed through a Canny edge detector and blurred through a gaussian filter. To reduce the computational effort in the retrieval phase, objects which are considered interesting for retrieval purposes are bounded with their minimum enclosing rectangle. Templates are deformed over the shapes included in these rectangles. Based upon the technique proposed in Sect. 3, spatial relationships among objects are analyzed and recorded in a symbolic *description file* associated to the image. Then, a *signature file* is built which is used as an index for fast access to spatial relationships informations. In this way to each raw image of the database the following structures are associated: *edge image*, *image description file*, *signature file*. The system can be requested to retrieve images representing one or more shapes. A query is composed by drawing a sketch of

one or more shapes on a graphic screen. If the sketch is composed of N templates, the system first searches an image where N objects are represented in the same spatial relation of drawn templates. Once it has been found, the system warps each template over the shape located in the same relative position in the image, computing a coefficient $S_i \in [0, 1]$, (as the output of the neural classifier) to measure similarity between the shape and the template. A similarity coefficient for the whole image is derived as $S = \sum_{i=1}^{N} S_i$. Once all the images in the database have been processed, they are sorted depending on the value of S and presented to the user. The technique explained in the previous sections has been applied for retrieval of images by sketch from a database of 20th century pictures from the Morandi catalogue, in the context of a joint project with industry for artwork musem database management systems. Fig. 1 shows the sketch of a Morandi's bottle, with a roughly rounded body sketched on the blackboard. Retrieval results are shown in Fig. 4 where the six more similar bottles are presented. The deformed template is shown superimposed over the original image. Pictures retrieved are ordered by decreasing similarity rank; rounded body bottles are ranked in the first positions, followed by those with more strained shape. A query for a different bottle shape over the same database and results obtained ordered according to their similarity rankings are shown in Figs. 2, 5. Examples of retrieval based on shapes and relative spatial relationships are shown in Figs. 3, 6. The user is allowed to draw object sketches and arrange their positions in the blackboard. Images are analyzed for matching of spatial relations between object including rectangles; only those rectangles which have passed the sieve are hence subjected to the template elastic deformation process.

References

1. S.Y. Lee and F.J. Hsu. "Spatial Reasoning and Similarity Retrieval of Images using 2D-C String Knowledge Representation". *Pattern Recognition*, 25(3), 1992.
2. S.K.Chang, Q.Y.Shi, C.W.Yan, "Iconic Indexing by 2-D Strings". *IEEE Transactions on Pattern Analysis and Machine* Vol.9, No.3, July 1987.
3. S.K.Chang, C.W.Yan, D.C.Dimitroff, T.Arndt, "An Intelligent Image Database System". *IEEE Transactions on Software Engineering*, Vol.14, No.5, May 1988.
4. A.Del Bimbo, E.Vicario, D.Zingoni, "A Spatial Logic for Symbolic Description of Image Contents". to appear on *Journal on Visual Languages and Computing.*
5. K.Hirata, T.Kato, "Query by Visual Example: Content-Based Image Retrieval". In *Advances in Database Technology - EDBT'92*, A.Pirotte, C.Delobel, G.Gottlob (Eds.), Lecture Notes on Computer Science, Vol.580,
6. 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.
7. M.J.Swain, D.H.Gallard, "Color Indexing". *Int.Journal of Computer Vision*, Vol.7, No.1, 1991.
8. S.L.Tanimoto, "An Iconic/Symbolic Data Structuring Scheme". in *Pattern Recognition and Artificial Intelligence*, C.H.Chen (Ed.), New York Academic, 1976.

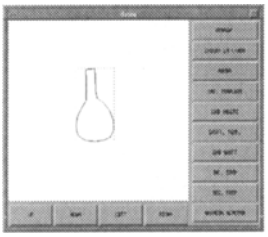

Fig. 1. A rounded body Morandi's bottle.

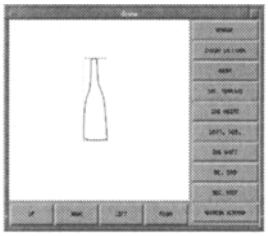

Fig. 2. Sketch made to retrieve a thin bottle.

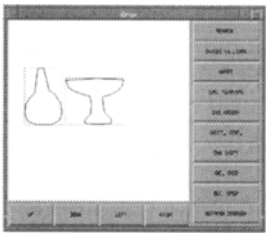

Fig. 3. Sketch made to retrieve a bottle with a fruit dish at its right hand side.

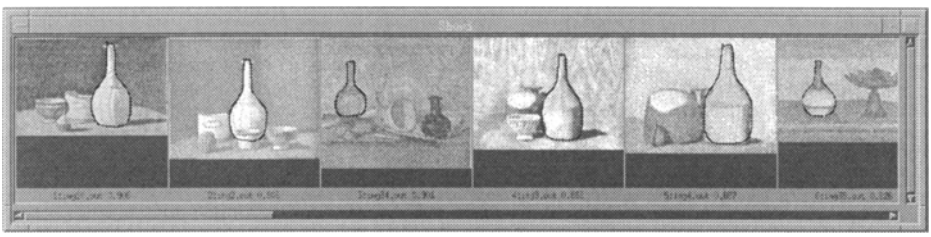

Fig. 4. Retrieval results for the sketch of Fig. 1.

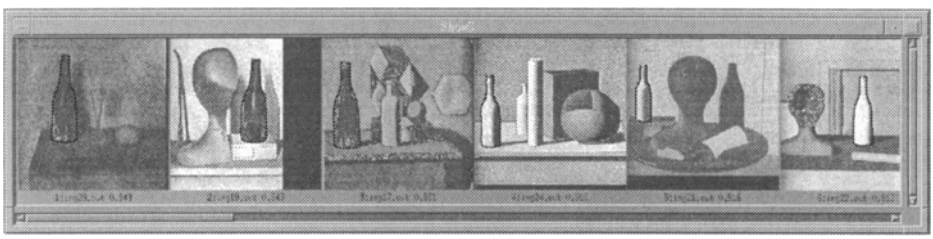

Fig. 5. Retrieval results for the sketch of Fig. 2.

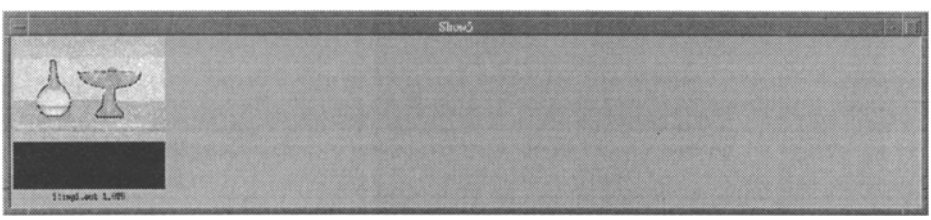

Fig. 6. Only one image in the DB matches both in shape and in spatial relation the sketch representing the bottle with the fruit dish (Fig. 3).