

Image Retrieval by Multidimensional Elastic Matching

P. Pala, S. Santini

University of Florence, Italy
University of California, San Diego, USA

Abstract. *Effective image retrieval by content from database requires that visual image properties are used instead of textual labels to recover pictorial data. Retrieval by image similarity given a template image is particularly challenging. The difficulty is to derive a similarity measure that combines shape, grey level patterns and texture in a way that closely conforms to human perception. In this paper a system is presented which supports retrieval by image similarity based on elastic template matching. The template can be both a 1D template modeling the contour of an object, and a 2D template modeling a part of an image with a significant grey level pattern. The retrieval process is obtained as a continuous interaction by which the original query of the user can be refined or changed on the basis of the results provided by the system.*

1 Introduction

Image databases are now currently employed in an eclectic range of different areas such as entertainment, art history, advertising, medicine and industry among others. In all these contexts, a key problem regards the modality of easy and efficient access of image contents. The minor expressiveness of text with respect to visual features doesn't allow to fully exploit capabilities of human memory. Items retrieved through a textual query could not be relevant at all for user's expectation. 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 accessing the database. A number of techniques have appeared in the literature which deal with visual querying by example of single images; different approaches depend on what facets of pictorial data are taken into account. Indexing and querying based on picture color distribution or object texture organization has also been proposed. In [6], image contents are described by global color histograms and queries are expressed by means of example images. Retrieval is performed by evaluating the similarity between the global color histograms of user provided examples and stored images. The QBIC database system [4], [2] allows user composed queries, but still evaluates similarity in terms of global properties of color histograms. In [1], both color and shape features are used for retrieval. In this system the query is formulated through an example image and retrieval is accomplished by a similarity measure computed on the basis of global color histogram and image edges.

Retrieval by content based on similarity between imaged object shapes and user-drawn sketches has been proposed in [4], [3], [9]. 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 users feel as a shape.

Specifying image content by the shape of some object in the image is a natural way for a user to represent image content. However there are cases in which there is no natural specification of a portion of an image in terms of contours. Search by matching portions of images is a very natural way to pose a query in interactive image databases. This type of query-by-example requires a good and robust way to determine the similarity between images. Image similarity is commonly measured using either *feature-based* or *em* statistical methods: Feature-based methods extract a limited number of features from the two images to be compared, and then use some suitable metric in the feature space \mathcal{F} as a measure of similarity.

Statistical methods consider the pixels as a realization of a particular multivariate stochastic process, usually described by a Gibbs distribution of the form $P(I) = \exp(-E(I))$, and try to determine the probability that the same parameters of the energy function E gave rise to the two images [5].

Both these methods suffers from serious drawbacks when applied to image databases. In fact feature-based methods can compare images only as long as the comparison criterion is captured by the features. Statistical methods are tractable only if certain locality assumptions are made on the random process that generates the images. Because of this, it is impossible to model accurately long-range effects like geometric distortions.

In this paper, we present an elastic based approach to measure image similarity. Given two images (namely a *template* and a *target* image) that we want to compare, one of the two (e.g. the template) is considered as an elastic body. The template is stretched in order to achieve the best match with the target in terms of difference in grey level of corresponding pixels. The measure of match and the amount of *energy* used to warp the template are used to derive a similarity measure between the two images.

2 Image matching: 2D elastic models

The idea which we present to measure the similarity between two images is equivalent to the solution of an elastic deformation problem in which one of the two images—which is considered as an elastic body—is deformed under the action of a force field generated by the difference between the two images, and its own elastic reaction to deformation. When equilibrium is reached between the two forces, we have the optimal deformation.

The theory behind the approach has been worked out in several places, under different assumptions [7]. Call *template* (T) the image we use to make the query, and *target image* (I) the image we want to compare the template against.

Suppose the comparison is done by deforming the template and that the deformation in the point x_1, x_2 of the template is $u(x_1, x_2) = (u_1(x_1, x_2), u_2(x_1, x_2))$. The function $u = (u_1, u_2)$ models the deformation of the template, providing the new coordinates of the point (x_1, x_2) according to the deformation u . To discover similarity between the template and the target image, we must set some constraint on deformation. In our approach the optimal deformation is obtained as a compromise between two opposite requirements. As first, we want to maximize the match between the deformed template and the target image. That is we want to minimize the distance between the two images:

$$\mathcal{D}(T, I, u) = \int \int_S (T(u_1(x_1, x_2), u_2(x_1, x_2)) - I(x_1, x_2))^2 d(x_1, x_2)$$

As second, we want to minimize the deformation energy of the template image. For a function $f : \mathbf{R}^2 \rightarrow \mathbf{R}$ the total amount of bending of the surface $(x_1, x_2, f(x_1, x_2))$ can be measured as:

$$\mathcal{J}(f) = \int \int_S ((f_{x_1 x_1})^2 + 2(f_{x_1 x_2})^2 + (f_{x_2 x_2})^2) d(x_1, x_2) \quad (1)$$

The deformation energy of the template image is therefore measured as:

$$\mathcal{J}(u_1) + \mathcal{J}(u_2)$$

A compromise between these two opposite requirements is thus achieved by minimizing the compound functional:

$$\mathcal{F} = \mu(\mathcal{J}(u_1) + \mathcal{J}(u_2)) + \mathcal{D}(T, I, u) \quad (2)$$

where μ is a parameter that determines the “stiffness” of the template. The higher μ the less the template will warp.

2.1 Numerical Solution

We approximate the deformation u through a linear combination of radial functions defined on a regular rectangular grid of N points $\bar{p}_1, \dots, \bar{p}_N$:

$$u(\bar{x}, \mathbf{C}) = \sum_{i=1}^N \bar{c}^i g(\|\bar{x} - \bar{p}_i\|) \quad \text{with} \quad \bar{x} = (x_1, x_2) \quad (3)$$

being $g : \mathbf{R}^+ \rightarrow \mathbf{R}$ the radial basis function, $\|\cdot\|$ the usual Euclidean norm on \mathbf{R}^2 , and $\mathbf{C} = \{\bar{c}^1, \dots, \bar{c}^N\}$ with $\bar{c}_i = (c_1^i, c_2^i)$ the array of 2D coefficients.

The minimization of the functional $\mathcal{F}(\mathbf{C})$, obtained substituting (3) in (2), with respect to the variables $(c_1^1, \dots, c_1^N, c_2^1, \dots, c_2^N)$ is achieved through a gradient descent technique. Variables are determined iteratively according to:

$$\mathbf{C}(k+1) = \mathbf{C}(k) - \epsilon \nabla \mathcal{F}(\mathbf{C})$$

where

$$\nabla \mathcal{F}(\mathbf{C}) = \left(\frac{\partial \mathcal{F}}{\partial c_1^1}, \dots, \frac{\partial \mathcal{F}}{\partial c_1^N}, \frac{\partial \mathcal{F}}{\partial c_2^1}, \dots, \frac{\partial \mathcal{F}}{\partial c_2^N} \right)$$

and

$$\frac{\partial \mathcal{F}}{\partial c_1^k} = \frac{\partial \mathcal{J}(u_1)}{\partial c_1^k} + \frac{\partial \mathcal{D}(T, I, u)}{\partial c_1^k} \quad (4)$$

$$\frac{\partial \mathcal{F}}{\partial c_2^k} = \frac{\partial \mathcal{J}(u_2)}{\partial c_2^k} + \frac{\partial \mathcal{D}(T, I, u)}{\partial c_2^k} \quad (5)$$

From Eq.(3) it follows that:

$$\begin{aligned} \mathcal{J}(u_1) = & \iint_S \left[\left(\sum_i c_1^i \frac{\partial^2 g(\|\bar{x} - \bar{p}_i\|)}{\partial x_1^2} \right)^2 + \right. \\ & 2 \left(\sum_i c_1^i \frac{\partial^2 g(\|\bar{x} - \bar{p}_i\|)}{\partial x_1 \partial x_2} \right)^2 + \\ & \left. \left(\sum_i c_1^i \frac{\partial^2 g(\|\bar{x} - \bar{p}_i\|)}{\partial x_2^2} \right)^2 \right] d(x_1, x_2) \end{aligned}$$

and similarly for $\mathcal{J}(u_2)$.

According to this, it can be derived that:

$$\begin{aligned} \frac{\partial \mathcal{J}(u_1)}{\partial c_1^k} = & \iint_S \left[2 \left(\sum_i c_1^i \frac{\partial^2 g(\|\bar{x} - \bar{p}_i\|)}{\partial x_1^2} \right) \frac{\partial^2 g(\|\bar{x} - \bar{p}_k\|)}{\partial x_1^2} + \right. \\ & 4 \left(\sum_i c_1^i \frac{\partial^2 g(\|\bar{x} - \bar{p}_i\|)}{\partial x_1 \partial x_2} \right) \frac{\partial^2 g(\|\bar{x} - \bar{p}_k\|)}{\partial x_1 \partial x_2} + \\ & \left. 2 \left(\sum_i c_1^i \frac{\partial^2 g(\|\bar{x} - \bar{p}_i\|)}{\partial x_2^2} \right) \frac{\partial^2 g(\|\bar{x} - \bar{p}_k\|)}{\partial x_2^2} \right] d(x_1, x_2) \end{aligned}$$

That is, by taking the summations out of the integrals:

$$\frac{\partial \mathcal{J}(u_1)}{\partial c_1^k} = \sum_i c_1^i \gamma_{i,k}$$

$$\frac{\partial \mathcal{J}(u_2)}{\partial c_2^k} = \sum_i c_2^i \gamma_{i,k}$$

with

$$\begin{aligned} \gamma_{i,k} = & \iint_S \left[2 \frac{\partial^2 g(\|\bar{x} - \bar{p}_i\|)}{\partial x_1^2} \frac{\partial^2 g(\|\bar{x} - \bar{p}_k\|)}{\partial x_1^2} + \right. \\ & 4 \frac{\partial^2 g(\|\bar{x} - \bar{p}_i\|)}{\partial x_1 \partial x_2} \frac{\partial^2 g(\|\bar{x} - \bar{p}_k\|)}{\partial x_1 \partial x_2} + \\ & \left. 2 \frac{\partial^2 g(\|\bar{x} - \bar{p}_i\|)}{\partial x_2^2} \frac{\partial^2 g(\|\bar{x} - \bar{p}_k\|)}{\partial x_2^2} \right] d(x_1, x_2) \end{aligned}$$

According to this, Eqs.(4),(5) can be written as:

$$\frac{\partial \mathcal{F}}{\partial c_1^k} = \mu \sum_i c_1^i \gamma_{i,k} + \frac{\partial \mathcal{D}(T, I, u)}{\partial c_1^k} \quad (6)$$

$$\frac{\partial \mathcal{F}}{\partial c_2^k} = \mu \sum_i c_2^i \gamma_{i,k} + \frac{\partial \mathcal{D}(T, I, u)}{\partial c_2^k} \quad (7)$$

The first term of Eqs. (6) and (7) depend on μ and model the elasticity of the template, that is the behavior of every point of the template to move in the same direction as its neighbors. The higher μ , the less the template can warp. The value of μ is initially low and is increased during the deformation process. In this way at the beginning of the deformation process a precise adaptation is achieved essentially for those parts which are similar in both the template and the image, while approximate adaptation is determined for those which are not equally represented in the two. In the second stage, the increased value of μ compels the template to regularize its deformation without loss of match.

At the end of the deformation process, the value of \mathcal{F} provides a measure of the similarity between the template and the target image. As an example in Fig.1 they are shown the deformations of a template image representing a bottle with a particular grey level pattern over images of bottles. The grey level pattern of the template image exhibit a remarkable vertical striature which is present, even if with different significance, in some other bottles. For each deformation the template and the target images as shown, with the deformation subjected by a regular grid of points (on the right). It can be noticed that the deformation keeps low values in the first three cases, showing the fact that even if the grey levels of the target and the template image are different, nevertheless the elastic approach allows to capture the similar grey level pattern of the two images. Differently, in the fourth case, the template is subjected to a considerable deformation, highlighting a structural difference between the grey level pattern of the template and target images.

3 Image Retrieval System

The elastic based approach to image similarity has been coupled with a system which supports content-based image retrieval by shape and spatial relationships ([9]) into a unified framework to provide retrieval modalities based both on shapes, spatial relationships and image patterns.

The system interface allows the user to select the preferred type of search. To draw a contour the user has simply to enter a curve. In the case of 2D search the user picks an image from the database eventually selecting it from a set of images which were retrieved by the system in a preceding search. The user can select just a portion of an image and interactively clean the selected region from those details which he is not interested in. In Fig. 3(a)(b) it is shown an example of the interactive editing of a selected image.

An example of shape based retrieval of pictures with query refinement by image pattern is shown in Figs. 2-3. The database includes images from a Morandi's catalogue representing bottles with other still life objects of different aspect and in different combinations. Fig. 2(a) shows the sketch drawn by the user roughly representing the contour of a rounded body bottle. Results of the similarity retrieval through 1D elastic matching are shown in Fig. 2(b).

Query refinement is shown in Figs. 3. Among the retrieved images the user is now interested in selecting those images which include a rounded bottle, like that depicted in the first ranked image (top left image), in the same position in the picture, and with a similar grey level pattern. Interactive editing of this image is shown in Fig. 3(a)(b). Results of querying by image patterns through 2D elastic matching are shown in Fig. 3(c). The template image and the last matched image of the database are shown in the top left part of the screen with their representations at coarse resolution (below). Results are shown in the right part of the screen in a smaller size, showing the final ranking of the 2D elastic match.

References

1. A. K. Jain and A. Vailaya. Image retrieval using color and shape. *Pattern Recognition*, 29(8):1233-1244, Aug. 1996.
2. W. Cody. Querying multimedia data for multiple repositories by content: The GARLIC project. In *Proc. on Visual Data Base Systems III*, Lausanne, 1995.
3. 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,
4. 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.
5. S. Geman, D. Geman "Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images". *IEEE Transactions on Pattern Analysis and Machine* Vol.6, No.6, November 1984.
6. M.J.Swain, D.H.Gallard, "Color Indexing". *Int.Journal of Computer Vision*, Vol.7, No.1, 1991.
7. D.Terzopoulos, D.Metaxas "Dynamic 3D Models with Local and Global Deformations: Deformable Superquadratics". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.13, No.7, pp. 703-714, 1991.
8. R.W.Picard, "A Society of Models for Video and Image Libraries". Tech.Report 352, MIT Media Lab Perceptual Computing BM Res.Div. Almaden Res.Center, Feb.1993.
9. A. Del Bimbo, P. Pala "Visual Image Retrieval by Elastic Matching of User Sketches". *IEEE Transactions on PAMI*, February 1997.

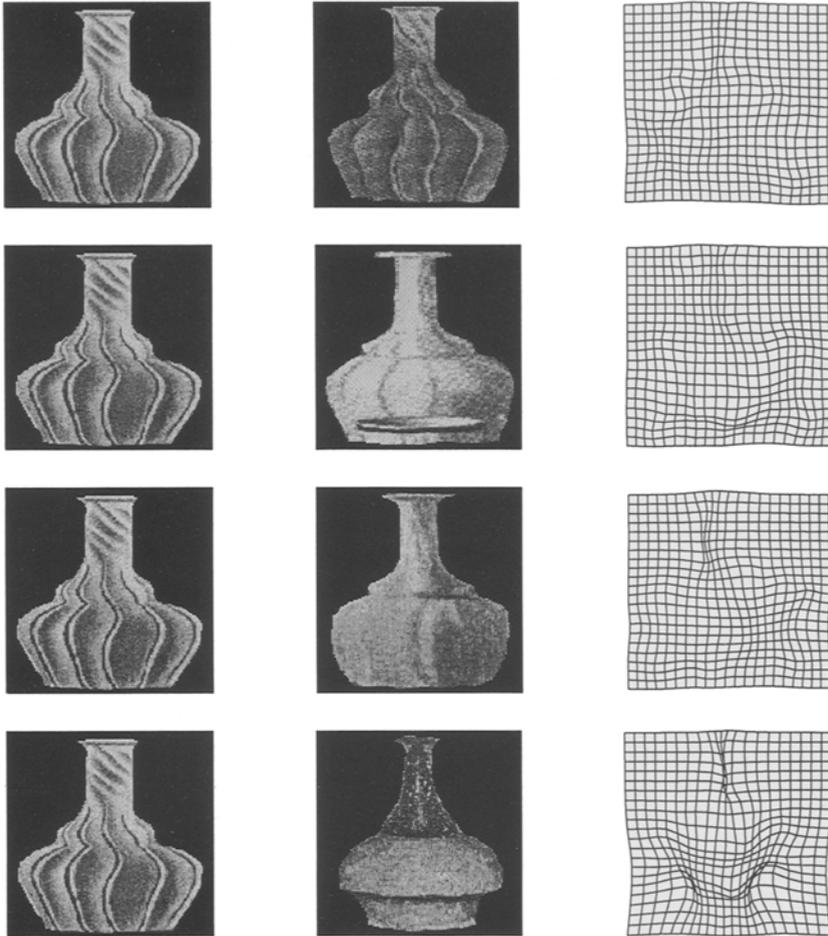


Fig. 1. Elastic deformation of a bottle image over different bottles. In the first three cases, the bottle grey level pattern is similar to that exhibited by the template bottle. In the last case the two grey level patterns are different. The deformation of a regular grid is shown on the right part of the figure. In the first three cases the deformation of the grid keeps smooth whereas in the last case it is subjected to a considerable change.

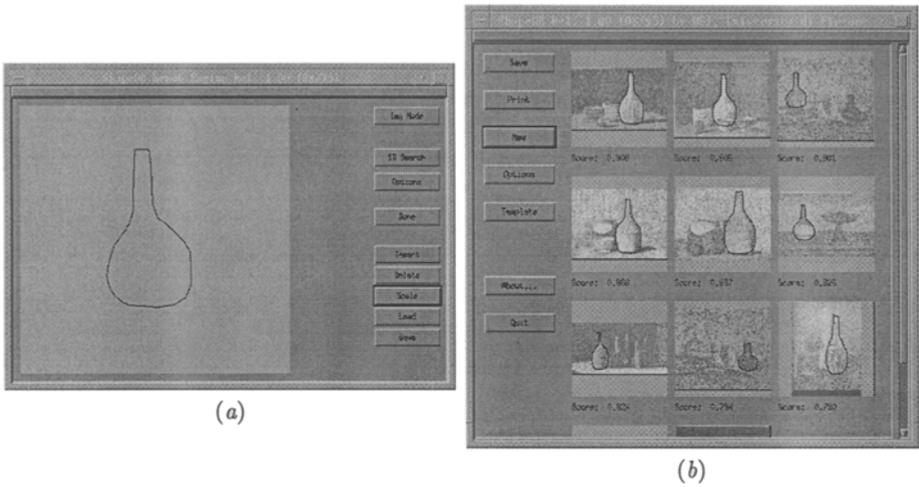


Fig. 2. (a) User-drawn sketch representing the shape of a rounded-body bottle. (b) Retrieved images with highest similarity ranks.

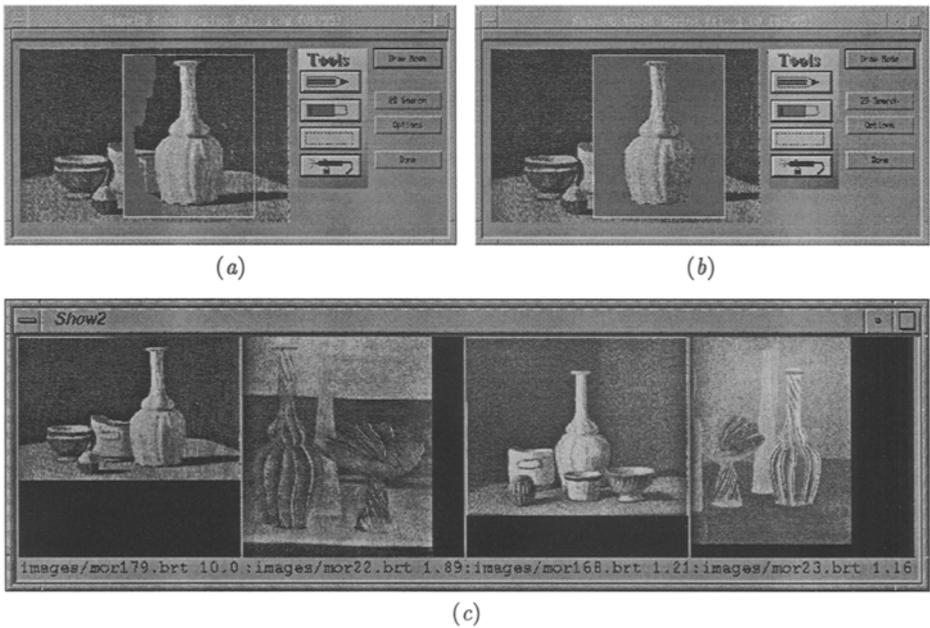


Fig. 3. (a): Image with similarity score 0.908 output by the system is used as a new query by image. Editing allows to eliminate details in the image that are not significant for elastic matching. (b): Edited image used for the query by image. (c): Best ranked images according to image pattern similarity.