Shape-Based Image Retrieval Using Shape Matrix

C. Sheng, Y. Xin

***Abstract* —**Retrieval image by shape similarity, given a template shape is particularly challenging, owning to the difficulty to derive a similarity measurement that closely conforms to the common perception of similarity by humans. In this paper, a new method for the representation and comparison of shapes is present which is based on the shape matrix and snake model. It is scaling, rotation, translation invariant. And it can retrieve the shape images with some missing or occluded parts. In the method, the deformation spent by the template to match the shape images and the matching degree is used to evaluate the similarity between them.

***Keywords*—**shape representation, shape matching, shape matrix, deformation

I. INTRODUCTION

More and more images have been generated in digital from around the world. Searching for images using shape features has attracted much attention. There are many shape representation and description techniques in the literature. Reference [1] and [2] have thoroughly discussed representation and sets of criteria for the evaluation of shape. Reference [3] proposed the following three classifications. The first classification is based on the use of shape boundary points as opposed to the interior of the shape. The second is the algorithms can be made on the basis of whether the result of analysis is numeric or non-numeric and the third is whether or not the representation preserves the information such that reconstruction of the shape is possible from its representation. However, shape representation and description is a difficult task. This is because when a 3-D real world object is projected onto a 2-D image plane, one dimension of object information is lost. As a result, the shape extracted from the image only partially represents the projected object. To make the problem even more complex, shape is often corrupted with noise, defects, arbitrary distortion and occlusion.

Manuscript received October 15, 2004. This work partially supported by China Natural Science Foundation (30170264) and National Basic Research Program of China (2003CB716104).

Chen. Sheng is with the Institute of Image processing and pattern Recognition, Shanghai Jiaotong University, Shanghai 200030, China (corresponding author to provide phone: 0086-21-62932190; fax: 0086-21-62932035; e-mail: chnshn@hotmail.com).

Yang. Xin is with the Institute of Image processing and pattern Recognition, Shanghai Jiaotong University, Shanghai 200030, China (e-mail: yangxin@sjtu.edu.cn).

With regards to the shape representation technique, the Fourier descriptors have been used as representation of shape for several years. In [4], [5], shapes were represented using a Fourier expansion of the function of their tangent angle and their arc length. The lower-order Fourier coefficients were then used to represent the shape. Reference [6] determined points of high curvature of a shape and represented them in polar form. These methods were invariant to translation and scale. However, the Fourier descriptor has several shortcomings in shape representation. One of the disadvantages is that it cannot provide multi-resolution representation of a shape. Reference [7] uses a polar transformation of the shape points about the geometric center of object, the distinctive vertices of the shape are extracted and used as comparative parameters to minimize the difference of shape distance from the center. But it was not designed to be tolerant to occlusion. Reference [8] has proposed the use of elastic matching for shape based image retrieval. According to this approach, a deformed template is generated as the sum of the original template and a warping deformation. The similarity between the original template and shape of the object in image is measured by minimizing a compound function which is similar to our method, however it is usually stable to scaling and rotation in a restricted range and does not hold for full similarity and for the more general affine transformation.

In this paper, a new shape comparison method is introduced which propose a different measurement of shape similarity, based on shape matrix which is invariant and unique under rigid motions. Combined with the snake model, the original template shape is deformed to adjust itself to the shape images. The deformation spent by the original template to match the shape images and the matching degree is used to evaluate the similarity between them. It is scaling, rotation, and translation invariant. In addition, the method can recover the missing part or remove the occluded part in the shapes. The reminder of this paper is organized as follows. In section 2, the approach is introduced and then, giving the similarity measurement in section 3. At last, experiments and discussing are given in section 4 and 5.

II. THE APPROACH OF SHAPE REPRESENTATION AND MATCHING

In this approach, suppose we have a one-dimensional original shape template.

$$C_{temp}(s) = (x(s), y(s))(s \in [0,1]) \tag{1}$$

Where: s is the parameter of length along the shape.

Let Ω be a bounded open subset of R^2, with $\partial\Omega$ its boundary. We have an image $u_0 : \overline{\Omega} \to R$. So our purpose is to search for a contour $C_{deformed}$ with a shape similar to the original template $C_{temp}(s)$.

The template must warp taking into account two opposite requirements. First, it must match the edge of shape in images as closely as possible and the second requirement should be take into account is the deformation of the template shape.

In order to make the template match the edges in images, in this paper, the snake model [9] is introduced. It is to minimize the following energy functional

$$E_{snake}(u_0, C_{deformed}) = E_{\text{int}}(C_{deformed}) + E_{ext}(u_0, C_{deformed}) \quad (2)$$

Where: $C_{deformed}$ is the deformed template shape, E_{int} is the internal energy that controls the smoothness of the deformed template.

$$E_{\text{int}}(C_{deformed}) = \alpha \int_0^1 \left|C'_{deformed}(s)\right|^2 ds + \beta \int_0^1 \left|C''_{deformed}(s)\right| ds \quad (3)$$

On the other hand, E_{ext} is external energy that attracts of the deformed template evolving to the shape in images. Function g depending on the gradient of the image u_0.

$$E_{ext}(u_0, C_{deformed}) = \lambda \int_0^1 g\left(\left|\nabla u_0 (C_{deformed}(s))\right|^2\right) ds \quad (4)$$

When the deformed template shape lies entirely on the object's edge where the gradient is maximum, the energy E_{ext} is minimum.

Secondly, it is need another energy to measure the similarity between the original template C_{temp} and $C_{deformed}$, also called deformation by us.

We define the original template shape C_{temp} to be approximated as a vector containing a sequential set of points: $P = [p_1, p_2, \cdots, p_n]$

where $p_i = (x_i, y_i) \in \{(x, y) : x, y = 1, 2, \cdots, N\}$, and then given an arbitrary reference point g, we define another vector $U = [u_1, u_2, ..., u_n]$ where $u_i = p_i - g$. Same to the deformed template shape $C_{deformed}$, let $V = [v_1, v_2, \cdots, v_n]$ be a sequential set of displacement where $v_i = q_i - g$. $Q = [q_1, q_2, ..., q_n]$ is the sequential set of points of $C_{deformed}$.

Every u_i can be expressed as a linear combination of two linearly independent vectors. The two neighboring points are suitable as its basis:

$$u_i = \alpha_i u_{(i-1) \bmod n} + \beta_i u_{(i+1) \bmod n} \quad (5)$$

Collecting and rearranging similar equations for all i, we can obtain the shape equation.

$$AU^T = 0 \quad (6)$$

Where: A is a $n \times n$ shape matrix that contains the necessary information to describe a shape.

$$A = \begin{bmatrix} 1 & -\beta_1 & 0 & \dots & 0 & -\alpha_1 \\ -\alpha_2 & 1 & -\beta_2 & 0 & 0 & \dots \\ 0 & -\alpha_3 & 1 & -\beta_3 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -\alpha_{n-1} & 1 & -\beta_{n-1} \\ -\beta_n & 0 & \dots & 0 & -\alpha_n & 1 \end{bmatrix}$$

The original template shape can be generated from A, if any two points on shape are given. On the other hand, if the original template shape is given, we can compute the shape matrix A. It can be proved that when the original template shape transforms by rotation, scaling, and translation, the matrix A is invariant. So we can define the deformation from original template to the deformed template as following:

$$E_{deform} = \left(AV^T\right)^T \left(AV^T\right) \quad (7)$$

When there is no deformation, the energy (7) is zero. It is translation, scaling, and rotation invariant.

In order to discover the similarity between the original template shape and the shape images, we must to set some constraints on deformation. So, our goal is to minimizing the following energy:

$$E = E_{deform} + \lambda E_{snake} \quad (8)$$

In Fig.1, an original template shape (a) and an original image (b) are shown. (c), (d) are the matching result and deformation.

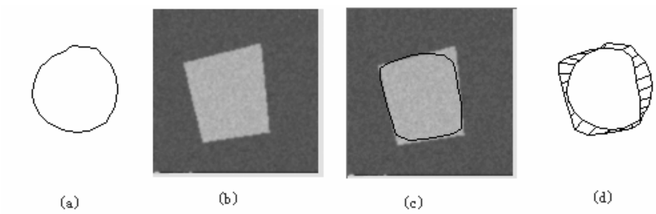

Figure.1 matching and deformation

III. SIMILARITY MEASURE

After the deformed template shape reaches convergence over a shape image. We need to measure how much the shape in image is similar to the original template. And it is a fuzzy concept. In order to measure it, the first we need to think about is overlapping between the deformed template and the shape image. We use the snake energy in (2) to measure the matching degree. The second need to be thought about is the deformation from the original template to the deformed template. In our approach we use only the energy in (7) to measure the similarity. It is in contrast to [8] where the deformation energy of the complex template shape to match the shape image is fairly the same as it of the noncomplex shape and it usually needs an additional criterion which measures the complexity of shape. While in the proposed approach, the matching of complex shape usually needs high deformation. Table.1 shows that the deformation of the horse template over the horse shape image is higher than that of the circular template over the coffee-pot shape image in Fig.2.

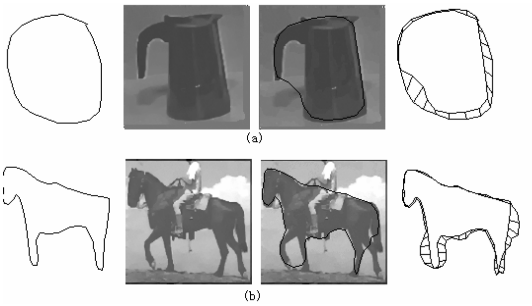

Figure.2 matching and deformation

TABLE.1 ENERGY OF DEFORMATION AND MATCHING FOR FIG.2

Energy	Energy(deform)	Energy(matching)
(a) jug	1.71	3.35
(b) horse	2.22	5.56

The template shapes are usually in an arbitrary scale and rotation and have an unknown relation to the shape in image. While the similarity measurement in this paper is scaling, rotation, translation invariant. In Fig3, the five-tip star template is made warp over a set of rotated star images that are also scaled up or down. From Table.2, it can be noticed that the deformation energy of star template matching the affined images are fairly equal.

Figure.3 original template of a five-tips star and the matched images that are rotated or scaled up and down

TABLE.2 THE ENERGY OF DEFORMATION AND MATCHING FOR STAR SHAPE RETRIEVAL IN FIG.3

Star	Energy(deform)	Energy(matching)
(a) original	0.60	1.14
(b) rotated 20°	0.62	1.21
(c) rotated 30°	0.57	1.00
(d) rotated 40°	0.61	1.40
(e) scaled down	0.70	1.31
(f) scaled up	0.68	1.42
(g) rotated and scaled	0.70	1.29

IV. EXPERIMENT RESULTS

In order to demonstrate the performance of our approach, the method was performed on the test images used in [7] and [10]. It includes nine categories with 11 shapes in each category shown in Fig.4. The results were compared to those obtained from [7] and [10].

In Fig.5, the 11 retrieved results to each template are list ranked by the similarity measurement.

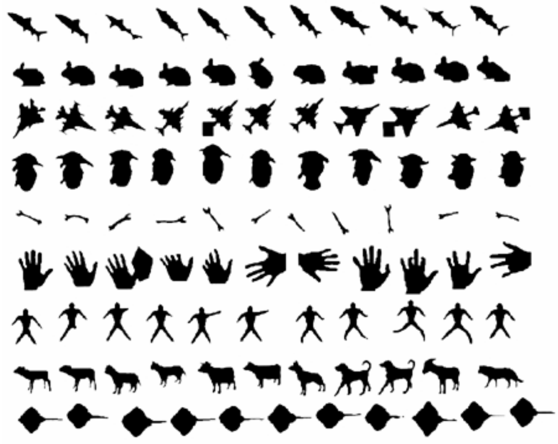

Figure.4 images for test

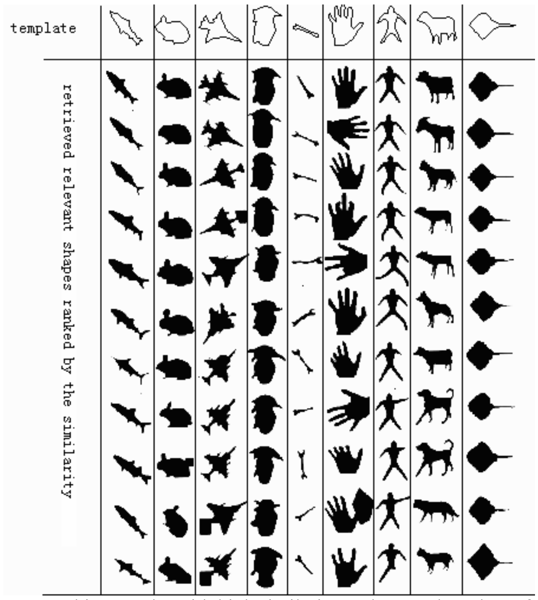

Figure.5 matching results with high similarity to the template shape for the original images in Fig.4

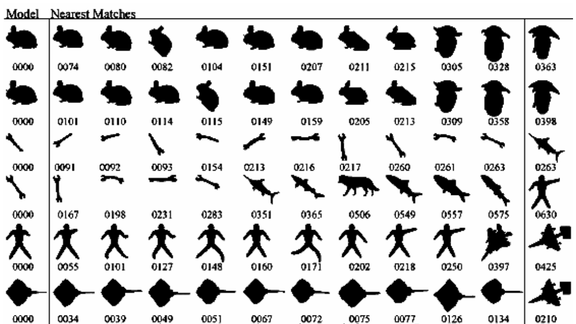

Figure.6 results selected from [7]

It is clear that the approach is tolerant to occlusion although there are some missing or occluded parts in some categories such as humans, rabbits and hands shapes. While in [7], in the categories of human, rabbits and hands shapes with missing or occluded parts were simply not considered as similar. Some of the results obtained from [7] are shown in Fig.6. The % average retrieval precision of [7], [10] and our method are shown respectively in Fig.7. Precision is defined as the ratio of the number of retrieved relevant shapes to the total number of retrieved shapes. Recall is defined as the ratio of the number of retrieved shapes to the total number of relevant shapes in the whole database. From Fig.7, we can see that our method have more precision than [7] and [10].

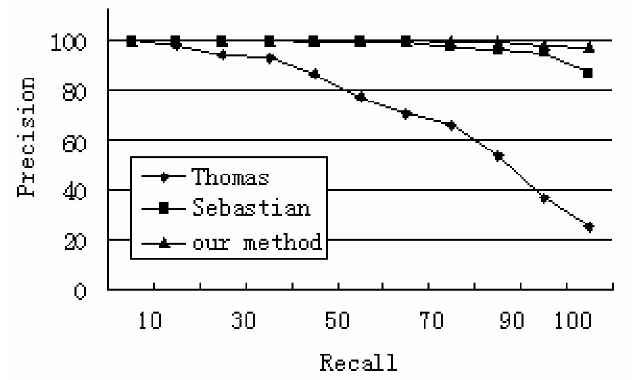

Figure.7 average retrieval precision

More significantly, in order to demonstrate scale and rotational invariance, the experiment was repeated with the images that are scaled down to 50% and rotated by 90°. From Fig.8, we can see that the matching is identical regardless of their scale and rotation.

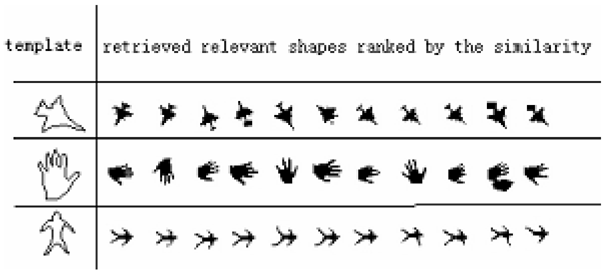

Figure.8 selected matching results with high similarity to the template shape for the images in Fig.4 that are scaled down to 50% and rotated by 90°

In the experiments, each of the images of shape is represented by 128*128 pixels. The system is implemented on a 1.6GHz intel machine and Visual C++ 6.0 environment. Every shape matching requires about 0.2s.

V. Conclusions

In this paper, a new method for shape based image retrieval is proposed. The degree of matching between the deformed template and the shape image and the energy of deformation from the original template to the deformed template are used to measure the similarity between the original template and the shape image. It allows one to handle situations in which part of the shape information is missing or occluded. And this method is scaling, rotation, and translation invariant.

Acknowledgment

This paper partially supported by China Natural Science Foundation (30170264) and National Basic Research Program of China (2003CB716104).

References

[1] D. Marr, H. Nishihara, “Representation and recognition of the spatial organization of three-dimensional shapes,” Proc. R. Soc. London B 200(1979), pp.269-194.
[2] M. Brady, “Criteria for Representations and of shape, Human and Machine Vision,” Academic Press. New York, 1993, pp. 39-84.
[3] T. Pavlidis, “A Review of Algorithms for Shape Analysis,” Computer. Graphics Image Process. 7(1978), pp.243-258.
[4] C. T. Zahn, R. Z. Roskies, “Fourier descriptors for plane closed curves,” IEEE Trans. Computer. C-21, 124(1972), pp.269-281.
[5] J. R. Bennet, J.S. McDonald, “On the measurement of curvature in a quantized environment,” IEEE Trans. Computer. 24(1975), pp.803-820.
[6] W.N. Lie, Y.C. Chen, “Shape representation and matching using the polar signature,” Proceedings of the International Computer Symposium, Tainan, Taiwan, 1986, pp. 710-718.
[7] Thomas Bernier, J. A. Landry, “A new method for representing and matching shapes of natural objects,” Pattern Recognition. 36(2003), pp.1711-1723.
[8] A. Del Bimbo, P. Pala, “Visual image retrieval by elastic matching of user sketches,” IEEE Trans. Pattern Analysis and Machine Intelligence. 19(2) (1997), pp.121-132.
[9] Kass M, Witkin A, Terzopoulos D, “Snakes: active contour models,” International Journal of Computer Vision. 1(1987), pp.321-331.
[10] T. B. Sebastian, P.N Klein, B.B. Kimia, “Recognition of shapes by editing shock graphs,” Eight IEEE International Conference on Computer Vision, ICCV 1(2001), pp. 755-762.

Chen Sheng received the MSc and BSc degrees both in computer science from the University of Xidian respectively in 2002 and 1999, he is currently a PhD candidate in the University of Shanghai Jiaotong. His research interests are in image analysis and computer vision.
Yang Xin received the PhD degree in Electronics and computer Engineering from Vrije University Brussel. He is currently a full Professor of Shanghai Jiaotong University. His research interests include medical image analysis and computer vision.