IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 424-429, 2000.

Shape Descriptors for Non-rigid Shapes with a Single Closed Contour

Longin Jan Latecki and Rolf Lakämper and Ulrich Eckhardt
Dept. of Applied Mathematics
University of Hamburg
Bundesstr. 55, 20146 Hamburg, Germany
{latecki, lakaemper, eckhardt}@math.uni-hamburg.de

Abstract

The Core Experiment CE-Shape-1 for shape descriptors performed for the MPEG-7 standard gave a unique opportunity to compare various shape descriptors for non-rigid shapes with a single closed contour. There are two main differences with respect to other comparison results reported in the literature: (1) For each shape descriptor, the experiments were carried out by an institute that is in favor of this descriptor. This implies that the parameters for each system were optimally determined and the implementations were throughly tested. (2) It was possible to compare the performance of shape descriptors based on totally different mathematical approaches. A more theoretical comparison of these descriptors seems to be extremely hard. In this paper we report on the MPEG-7 Core Experiment CE-Shape-1.

1. Introduction

Shape descriptors for comparing silhouettes of 2D objects in order to determine their similarity are important and useful for applications such as database retrieval. This importance is, for example, justified by the fact that the MPEG-7 group will incorporate such shape descriptors into the forthcoming MPEG-7 standard. Since the 2D objects are projections of 3D objects their silhouettes may change due to:

- change of a view point with respect to objects,
- non-rigid object motion (e.g., people walking or fish swimming),
- noise (e.g., digitization and segmentation noise).

The goal of the Core Experiment CE-Shape-1 is to evaluate the performance of 2D shape descriptors under such conditions. The shapes were restricted to simple pre-segmented shapes defined by their outer closed contours. Some examples are given in Figure 1. The main requirement was that the shape descriptors should be robust to small non-rigid deformations due to (1), (2), or (3). In addition the descriptors should be scale and rotation invariant.

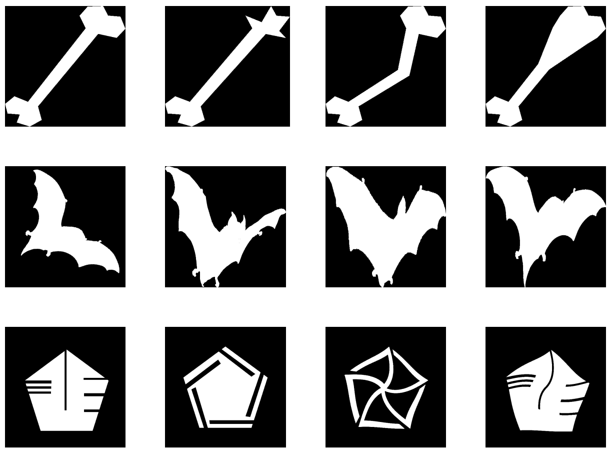

Figure 1. Some shapes used in part B of CE-Shape-1. Shapes in each row belong to the same class.

Now we list and shortly describe the shape descriptors which were tested in the Core Experiment CE-Shape-1. The shape descriptors can be divided into three main categories:

1. *contour based descriptors:* the contour of a given object is mapped to some representation from which a shape descriptor is derived,
2. *image based descriptors:* the computation of a shape descriptor is based on summing up pixel values in a digital image containing the silhouette of a given object; the shape descriptor is a vector of a certain number of parameters derived this way,

3. *skeleton based descriptors:* after a skeleton is computed, it is mapped to a tree structure that forms the shape descriptor; the shape similarity is computed by some tree-matching algorithm.

To the first category belong the following three descriptors:

P320 presented by Mitsubishi Electric ITE-VIL, based on the curvature scale-space [14, 15].

P567 presented by Heinrich Hertz Institute in Berlin, based on a wavelet representation of object contours [3]. The wavelet representation used in [3] is shown to outperform the Fourier descriptors.

P298 presented by the authors in cooperation with Siemens Munich, based on the best possible correspondence of visual parts [10, 12].

To the second category belong the following two descriptors:

P687 presented by Hanyang University, based on Zernike moments. As it is theoretically supported and experimentally verified in [8], Zernike moments significantly outperform regular moments and moment invariants.

P517 presented by Hyundai Electronics Industries, based on multilayer eigenvectors. Pixel values (0's and 1's) from certain image regions form matrices whose eigenvectors determine the form descriptor. This is the only shape descriptor to which there do not exist references in the literature.

To the third category belongs one descriptor:

DAG presented by Mitsubishi Electric and Princeton University[1]. It assigns a DAG (Directed Acyclic Graph) ordered tree to an object skeleton. The shape similarity measure is then based on the similarity of corresponding trees that is based on the matching algorithm for DAG ordered trees presented in [13]. The idea of representing shapes by their skeletons in Computer Vision goes back to Blum [1]. Siddiqi et al. [17] also convert object skeletons to a tree representation and use a tree-matching algorithm to determine the shape similarity.

In Section 2 deals with the CE-Shape-1. In Section 3, we derive some important conclusions form the experimental results.

[1]This descriptor has not been assigned any identification number by the MPEG-7 group.

2. Results of the MPEG-7 Core Experiment CE-Shape-1

In this section we shortly describe the Core Experiment CE-Shape-1 and present the experimental results. The experiment was performed as specified in the MPEG-7 document [7]. The presented results are based on [2] and on [4].

The core experiment was divided into three parts with the following main objectives:

A: robustness to scaling (A1) and rotation (A2)

B: performance of the similarity-based retrieval

C: robustness to changes caused by no-rigid motion

Part A can be regarded as a necessary condition that every shape descriptor should satisfy. The main part is part B, where a set of semantically classified images with a ground truth is used. Part C can be viewed as a special case of part B. Here also the performance of the similarity-based retrieval is tested, but only the deformation due to no-rigid motion is considered. Only one query is used for part C.

The recall retrieval performance is tested, where *recall* is the ratio of the number of the retrieved relevant shapes to the number of the relevant shapes in the database. There were together 3450 shapes used in the three parts.

2.1. Part A: Robustness to Scaling and Rotation

A-1 Robustness to Scaling The database includes 420 shapes; 70 basic shapes and 5 derived shapes from each basic shape by scaling digital images with factors 2, 0.3, 0.25, 0.2, and 0.1. Each of the 420 images was used as a query image. A number of correct matches was computed in the top 6 retrieved images. Thus, the best possible result is 2520 matches.

A-2 Robustness to Rotation The database includes 420 shapes: the 70 basic shapes are the same as in part A-1 and 5 derived shapes from each basic shape by rotation (in digital domain) with angles: 9, 36, 45 (composed of 9 and 36 degree rotations), 90 and 150 degrees. Each of the 420 images was used as a query image. A number of correct matches was computed in the top 6 retrieved images. Thus, the best result is 2520 matches.

In Figure 2, a table with results for Part A is presented. Since the best possible performance for A-1 was about 93% as shown below, all shape descriptors except DAG performed nearly optimal.

About 17 shapes obtained by scaling 0.1 are too small. For example, Figure 3 shows for each object in the first row

	P298	P320	P517	P567	P687	DAG
A1	88.65	89.76	92.42	88.04	92.54	no results
A2	100.00	99.37	100.00	97.46	99.60	no results
Part A	94.33	94.57	96.21	92.75	96.07	85

	P298	P320	P517	P567	P687	DAG
Part B	76.45	75.44	70.33	67.76	70.22	60

	P298	P320	P517	P567	P687	DAG
Part C	92.0	96.0	88.0	93.0	94.5	83.0

Figure 2. Results for Parts A, B, and C.

its scaled version in the second row obtained by scaling with factor 0.1. The scaled objects in the second row are more similar to the basic shapes in the third row than to the shapes they originate from (first row). Therefore, it is impossible to obtain correct matches in about
170 cases = 17 too small shapes $\times$ (5 errors for each too small shape used as a query + 5 errors for 5 derived shapes used as a query).
Thus, the best possible result in A-1 is about 93%.

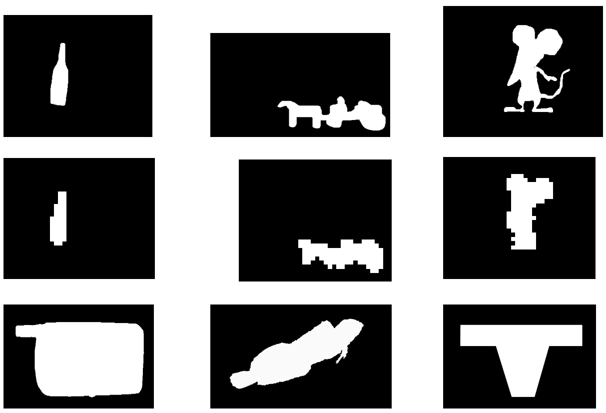

Figure 3. Example shapes used in part A of CE-Shape-1. Below each object in the first row, its scaled version with factor 0.1 is shown, second row. For each column, the shapes in the third row seem to be more similar to the scaled objects in the middle row than the ones in the first row.

2.2. Part B: Similarity-based Retrieval

This is the main part of the Core Experiment CE-Shape-1. The total number of images in the database is 1400: 70 classes of various shapes, each class with 20 images. Each image was used as a query, and the number of similar images (which belong to the same class) was counted in the top 40 matches (bulls-eye test). Since the maximum number of correct matches for a single query image is 20, the total number of correct matches is 28000. Some example shapes are shown in Figure 1, where the shapes in the same row belong to the same class. In Figure 2 a table with results for part B is presented.

The 100% retrieval rate was again not possible, since some classes contain objects whose shape is significantly different so that it is not possible to group them into the same class using only shape knowledge. For example, see the third row in Figure 1 and the first and the second rows in Figure 6. In the third row in Figure 6, we give examples of two spoon shapes that are more similar to shapes in different classes than to themselves.

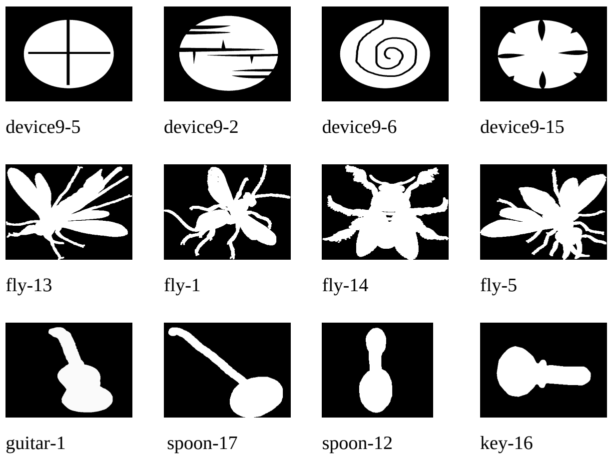

Figure 6. Example shapes used in part B of CE-Shape-1. The shapes with the same name prefix belong to the same class.

	P298	P320	P517	P567	P687	DAG
Total Score$_1$	87.59	88.67	84.85	84.50	86.93	76

Figure 4. Total Score$_1$ = average over the three parts of the experiment.

	P298	P320	P517	P567	P687	DAG
Total Score$_2$	83.16	82.62	80.04	77.14	79.92	69.38

Figure 5. Total Score$_2$ = average over the number of queries.

2.3. Part C: Motion and non-rigid deformations

Part C adds a single retrieval experiment to part B. The database for part C is composed off 200 frames of a short video clip with a bream fish swimming plus a database of marine animals with 1100 shapes. Fish bream-000 was used as a query (see Figure 7), and the number of bream shapes in the top 200 shapes was counted. Thus, the maximal number of possible matches was 200. In Figure 2 a table with results for Part C is presented.

Since about 14 bream fish do not have similar shape to bream-000, i.e., some of the 1100 marine animals are more similar to bream-000 than the 14 bream fish, the best possible result is 93% (186/200). For example, in Figure 7 the four kk shapes (in the second row) taken from the 1100 marine animals are more similar to bream-000 than the bream shapes 113, 119, and 121 (in the first row).

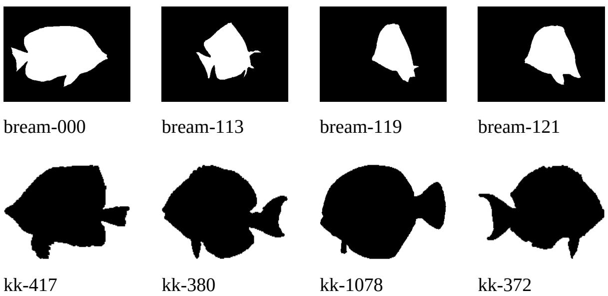

Figure 7. Example shapes used in part C of CE-Shape-1. The shapes with kk prefix (in the second row) are more similar to bream-000 than the bream shapes.

2.4. Average Performance in the Core Experiment CE-Shape-1

Average performance with the average over the three parts, i.e., **Total Score$_1$** = $\frac{1}{3}A+\frac{1}{3}B+\frac{1}{3}C$, is given in Figure 4. Average performance with the average over the number of queries, i.e.,

$$\mathbf{TotalScore_2} = \frac{840}{2241}A + \frac{1400}{2241}B + \frac{1}{2241}C,$$

where 2241 is the total number of queries in parts A, B, and C, is given in Figure 5.

3. Interpretation

As can be seen in Figure 2, all shape descriptors except DAG passed the necessary test in part A, i.e., they are sufficiently robust to scaling and rotation. Keeping in mind that the best possible performance for A-1 was about 93%, the five shape descriptors performed nearly optimal with respect to scaling. The descriptors P298, P320, P517, and P687 correctly retrieved over 99% rotated images in A-2. The performance of P567 (wavelet representation of object contours) was slightly less robust to rotation.

The low average result in part A of the DAG descriptor indicates that it is neither robust to scaling nor to rotation. This result seems to give a clear experimental verification of the known fact that the computation of skeletons in digital images is not robust. It seems that none of the existing approaches to compute skeletons in digital images would provide better results on the given data set.

We discuss the results of the main part of the Core Experiment CE-Shape-1 now. As can be seen in part B of Figure 2 two shape descriptors P298 and P320 significantly outperform the other four, and are, therefore, the most useful to search for similar objects obtained by non-rigid transformations. Both descriptors have also the best total scores in Figures 4 and 5.

For both similarity measures used with descriptors P298 and P320, while computing the similarity values for two objects, a best possible correspondence of the maximal convex/concave arcs contained in object contours is established. The maximal convex/concave arcs are not taken from the original contours, but from their simplified versions. Significant maximal convex arcs on simplified contours correspond to significant parts of visual form [9], whose importance and relevance for object recognition is verified by numerous cognitive experiments [5, 6, 16, 18].

For P298 a single simplified contour is used as a shape descriptor. This contour is optimally determined by a novel process of contour simplification called discrete curve evolution in [11]. This process allows to reduce the influence of noise without changing the overall appearance of objects. To compute the similarity measure between two shapes, the best possible correspondence of maximal convex/concave contour arcs is established. Finally the similarity between corresponding parts is computed using their tangent functions and aggregated to a single output similarity value [12].

For P320 simplified contours are obtained by a classical scale-space curve evolution based on contour smoothing by convolution with a Gauss function. The arclength position of inflection points (x-axis) on contours on every scale (y-axis) forms so called Curvature Scale Space (CCS) [15], see Figure 8. The positions of the maxima of CSS yield the shape descriptor P320. These positions projected on the simplified object contours give the positions of the mid points of the maximal convex/concave arcs obtained during the curve evolution. The shape similarity measure between two shapes is computed by relating the positions of the maxima of the corresponding CSSs.

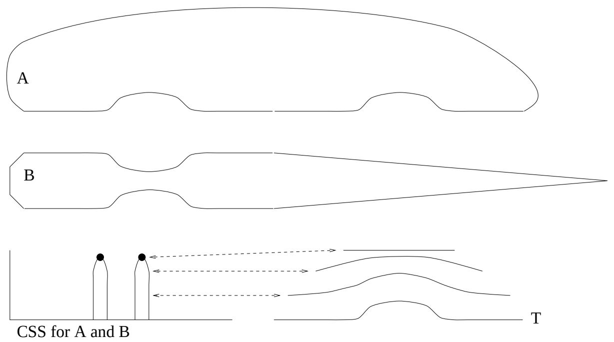

Figure 8. Both shapes A and B have the same positions of the maxima of CSS.

Observe that both shapes A and B in Figure 8 are assigned very similar shape descriptors by P320. Both shapes A and B contain two copies of the same piece T (shown at the bottom). Since only piece T contains inflection points, the CSS functions of both shapes are nonzero only for the T parts of their boundaries. Therefore, their CSS representations (maxima of the CSS functions) are very similar.

This cannot happen for P298, since the mapping to the representation space (the tangent space) is one-to-one, i.e., the original polygon can be uniquely reconstructed given the length and the tangent directions of its line segments.

Observe that all convex objects are identical for the shape descriptor P320, since it is based on inflection points, and there are no inflection points on the contour of a convex object. This implies that a triangle, square, or circle cannot be distinguished using this descriptor.

4. Conclusions

We report here on performance of six shape descriptors that accomplished the Core Experiment Shape-1 for the MPEG-7 standard. Two shape descriptors P298 (correspondence of visual parts) and P320 (curvature scale-space) significantly outperform the other four, and are, therefore, the most useful to search for similar objects obtained by non-rigid transformations. Both descriptors base the computation of their similarity measures on the best possible correspondence of maximal convex/concave arcs contained in the simplified versions of boundary contours. The simplified boundary contours are obtained by two different processes of curve evolutions. Both shape descriptors are cognitively motivated, since the maximal convex arcs represent visual parts of objects, whose important role in the human visual perception is verified by numerous experiments.

Since the results of the experiments are verified using the human visual perception (the ground truth was manually determined by humans), a cognitively motivated computation of the similarity values seems to be essential. In this context, a cognitive motivation seems to be far more important than a motivation in physics for P687 (Zernike moments), or in signal theory for P567 (wavelet contour descriptor).

Clearly, the cognitive motivation is not the only criterium. The robustness of the computation is also extremely important. Since the computation of the similarity measure for the shape descriptor DAG (DAG ordered trees) is based on object skeletons, it has a nice cognitive motivation. However, until now there does not exist any computation of object skeletons that is robust to scaling, rotation, and noise in the digital domain.

On the other hand, a careful attention to the robustness is paid by the extraction of descriptors P298 and P320, and during the computation of their similarity measures.

References

[1] H. Blum. Biological shape and visual science. *Journal of Theor. Biol.*, 38:205–287, 1973.

[2] M. Bober, J. D. Kim, H. K. Kim, Y. S. Kim, W.-Y. Kim, and K. Müller. Summary of the results in shape descriptor core experiment. *MPEG-7*, ISO/IEC JTC1/SC29/WG11/MPEG99/M4869, Vancouver, July 1999.

[3] G. Chuang and C.-C. Kuo. Wavelet descriptor of planar curves: Theory and applications. *IEEE Trans. on Image Processing*, 5:56–70, 1996.

[4] J. Heuer, A. Kaup, U. Eckhardt, L. J. Latecki, and R. Lakämper. Region localization in MPEG-7 with good coding and retrieval performance. *MPEG-7*, ISO/IEC JTC1/SC29/WG11/MPEG99/M5417, Maui, December 1999.

[5] D. D. Hoffman and W. A. Richards. Parts of recognition. *Cognition*, 18:65–96, 1984.

[6] D. D. Hoffman and M. Singh. Salience of visual parts. *Cognition*, 63:29–78, 1997.

[7] S. Jeannin and M. Bober. Description of core experiments for MPEG-7 motion/shape. *MPEG-7*, ISO/IEC JTC1/SC29/WG11/MPEG99/N2690, Seoul, March 1999.

[8] A. Khotanzan and Y. H. Hong. Invariant image recognition by zernike moments. *IEEE Trans. PAMI*, 12:489–497, 1990.

[9] L. J. Latecki and R. Lakämper. Convexity rule for shape decomposition based on discrete contour evolution. *Computer Vision and Image Understanding*, 73:441–454, 1999.

[10] L. J. Latecki and R. Lakämper. Contour-based shape similarity. In D. P. Huijsmans and A. W. M. Smeulders, editors, *Proc. of Int. Conf. on Visual Information Systems*, volume LNCS 1614, pages 617–624, Amsterdam, June 1999.

[11] L. J. Latecki and R. Lakämper. Polygon evolution by vertex deletion. In M. Nielsen, P. Johansen, O. Olsen, and J. Weickert, editors, *Scale-Space Theories in Computer Vision. Proc. of Int. Conf. on Scale-Space'99*, volume LNCS 1682, Corfu, Greece, September 1999.

[12] L. J. Latecki and R. Lakämper. Shape similarity measure based on correspondence of visual parts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, to appear.

[13] I.-J. Lin and S. Y. Kung. Coding and comparison of dags as a novel neural structure with application to on-line hadwritten recognition. *IEEE Trans. Signal Processing*, 1996.

[14] F. Mokhtarian, S. Abbasi, and J. Kittler. Efficient and robust retrieval by shape content through curvature scale space. In A. W. M. Smeulders and R. Jain, editors, *Image Databases and Multi-Media Search*, pages 51–58. World Scientific Publishing, Singapore, 1997.

[15] F. Mokhtarian and A. K. Mackworth. A theory of multiscale, curvature-based shape representation for planar curves. *IEEE Trans. PAMI*, 14:789–805, 1992.

[16] K. Siddiqi and B. B. Kimia. Parts of visual form: Computational aspects. *IEEE Trans. PAMI*, 17:239–251, 1995.

[17] K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Zucker. Shock graphs and shape matching. *Int. J. of Computer Vision*, to appear; http://www.cim.mcgill.ca/ siddiqi/journal.html.

[18] K. Siddiqi, K. J. Tresness, and B. B. Kimia. Parts of visual form: Psychophysical aspects. *Perception*, 25:399–424, 1996.