Retrieval Results for Shape Similarity on the MPEG-7 Data Set

Longin Jan Latecki

Dept. of Computer and Information Sciences
Temple University
Philadelphia, PA 19094, USA
latecki@temple.edu

Abstract. The test data used in the *MPEG-7 Core Experiment CE-Shape-1* provides a unique opportunity to compare various shape descriptors for non-rigid shapes with a single closed contour. There are two main advantage of such results: (1) It is 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. (2) For each shape descriptor, the experiments are 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 thoroughly tested. In this contribution we give an overview of the recent retrieval results obtained on the MPEG-7 Core Experiment CE-Shape-1.

There exist a large variety of approaches to define shape similarity measures of planar shapes, some of which are listed in the references. Since an objective comparison of their qualities seems to be impossible, experimental comparison is needed. The dataset created by the MPEG-7 committee for evaluation of shape similarity measures [9, 20] offers an excellent possibility for objective experimental comparison of the existing approaches evaluated based on the retrieval rate.

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 reflected in the fact that the MPEG-7 group incorporated such shape descriptors into the 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 MPEG-7 Core Experiment CE-Shape-1 was 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. 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.

Two kinds of shape descriptors are included in the MPEG-7 standard. In addition to the descriptors for shapes defined by their outer closed contours, on which we focus in this contribution, region based shape descriptors are also included. Such descriptors treat shape as binary textures. Thus, the input shape does not need to be defined by a single connected bitmap. Although the region based descriptors are more universal, since they apply to shapes composed of several components, they perform significantly worse when shapes are given by a single connected component. The only region-based shape descriptor included in the MPEG-7 standard is based on ART (Angular Radial Transform) coefficients [16, 29].

First we shortly describe the settings of the MPEG-7 Core Experiment CE-Shape-1. 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 shape is used for part C.

Since it is the essential part, we focus our attention on the performance evaluation of shape descriptors in experiments established in Part B of the MPEG-7 CE-Shape-1 data set [9]. The test set consists of 70 different classes of shapes, each class containing 20 similar objects, usually (heavily) distorted versions of a single base shape. The whole dataset therefore consists of 1400 shapes. For example, each row in Fig. 1 shows four shapes from the same class.

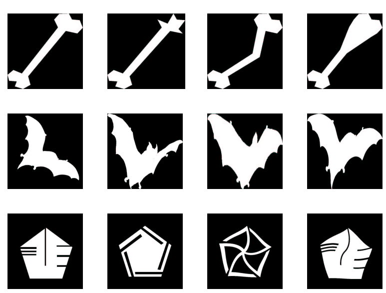

Fig. 1. Some shapes used in part B of MPEG-7 Core Experiment CE-Shape-1. Shapes in each row belong to the same class, i.e., we see in the first row four different shapes (out of 20) of class 'bone'.

Each image was used as a query, and the retrieval rate is tested using so called bulls-eye test: for each query image, we count the number of images that belong to the same class in the top 40 matches. Since the maximum number of correct matches for a single query image is 20, the total number of correct matches is 28000.

Strong variations of shapes in the same classes indicate that no shape similarity measure is capable of yielding 100% retrieval rate. For example, see the third row in Figure 1 and the first and the second rows in Figure 2. In the third row in Figure 2, we give examples of two spoon shapes that are more similar to shapes in different classes than to themselves.

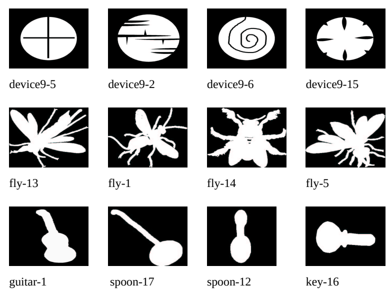

Fig. 2. Example shapes used in part B of CE-Shape-1. The shapes with the same name prefix belong to the same class.

The best actually reported retrieval rate for the MPEG-7 CE-Shape-1 data set is 83.19% [1]. The only descriptor for shapes defined by their outer closed contours [24] that is included in the MPEG-7 standard has retrieval rate of 81.12%. The best six published retrieval rates are shown in Fig. 3. The retrieval results of shape similarity descriptors that took part in the MPEG-7 competition are recorded in the report on the original MPEG-7 CE-Shape-1 experiment in [9]. Except the two descriptors presented in [24] and [19], the retrieval rates of other MPEG-7 competing descriptors were lower than the rates shown in Fig. 3.

Retrieval	76.45	76.51	78.18	78.38	81.12	83.19
Reference	[19]	[4]	[26]	[13]	[24]	[1]
Authors	Latecki ...	Belongie ...	Sebastian ...	Grigorescu ...	Mokhtarian ...	Attalla

Fig. 3. The best six published retrieval rates in percents on the MPEG-7 CE-Shape-1 Part B dataset.

We now shortly describe each shape descriptor listed in Fig. 3

[19] The shape similarity in [19] is based on an optimal correspondence of contour parts of both compared shapes. The correspondence is restricted so that at least one of element in a corresponding pair is a maximal convex contour part. Since the correspondence is computed on contours simplified by a discrete curve evolution [18], the maximal convex contour parts represent visually significant shape parts. Thus, the computation of shape similarity is based on an optimal correspondence of visual parts [18]. This correspondence is computed using dynamic programming.

[4] A shape representation in [4] is first built for each contour point using a statistics of other contour points 'seen' by this point in quantized angular and distance intervals. The obtained view of a single point is represented as a 2D histogram matrix. To compute a distance between two contours the correspondence of contour points is established that minimizes the distances of corresponding matrices.

[26] The distance between two contours in [26] is expressed as the minimal amount of deformation needed to transform one contour to the other.

[13] Similar to [4], the shape representation in [13] is first built for each contour point as a distance set, which is a set of distances from a given contour points to its N nearest neighbors. Thus, in contrast to [4] no angular information but only local distance information is obtained. A shape is represented as a set of distance sets at sample points on the shape. The distance between two shapes is expressed as the cost of a cheapest correspondence relation of the sets of distance sets.

[24] The shape representation in [24] is based on curvature scale-space introduced in [23]. This shape descriptor is included in the MPEG-7 standard. First simplified contours are obtained by the scale-space curve evolution (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) curve [23]. The positions of the maxima on the CSS curve yield the shape descriptor. These positions when projected on the simplified object contours give the positions of the mid points of the maximal convex 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 CSS curves.

[1] Surprisingly, the shape similarity measure with actually best performance is based on a simple geometric shape representation. First sample points are placed on the object contour with equal arc length distances. The contour is represented as sequence of planar distances between consecutive sample points and the sequence of angles between the shape centroid and the sample points. The shape distance of two contours is based on the distance of the associated sequences.

To our best knowledge only one shape descriptor led to a development of a shape-based image search engine. It is [19]. The reader is invited to try it out by following the Internet link [17].

References

1. E. Attalla. *Shape Based Digital Image Similarity Retrieval.* Ph.D. Thesis, Wayne State University, Detroit, 2004.
2. M. Arkin, L. P. Chew, D. P. Huttenlocher, K. Kedem, and J. S. B. Mitchell. An efficiently computable metric for comparing polygonal shapes. *IEEE Trans. PAMI*, 13:209–206, 1991.
3. R. Basri, L. Costa, D. Geiger, and D. Jacobs. Determining the Similarity of Deformable Shapes. *Vision Research* 38, p. 2365–2385, 1998.
4. S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE PAMI* 24, p. 509 Center number 16 1811 705–522, 2002.
5. L. Biderman. Human image understanding: Recent research and a theory. *Computer Vision, Graphics, and Image Processing* 32, 29-73, 1985.
6. T. Binford. Visual Perception by Computer. *IEEE Conf. on Systems and Control*, 1971.
7. R. Brooks. Symbolic Reasoning Among 3D Models and 2D Images. *Artificial Intelligence* 17, p. 285-348, 1981.
8. M. C. Burl and P. Perona. Recognition of Planar Object Classes. *CVPR*, 1996.
9. M. Bober, J. D. Kim, H. K. Kim, Y. S. Kim, W.-Y. Kim, and K. Muller. Summary of the results in shape descriptor core experiment. *MPEG-7, ISO/IEC JTC1/SC29/WG11/ MPEG99/M4869*, Vancouver, July 1999.
10. L. da F. Costa and R. M. Cesar. *Shape Analysis and Classification. Theory and Practice*, CRC Press, Boca Raton, 2001.
11. C. M. Cyr and B. B. Kimia. 3D Object Recognition Using Shape Similarity-Based Aspect Graph. *ICCV*, 2001.
12. D. Huttenlocher, G. Klanderman, and W. Rucklidge. Comparing Images Using the Hausdorff Distance, *IEEE Trans. PAMI* 15, p. 850-863, 1993.
13. C. Grigorescu and N. Petkov. Distance Sets for Shape Filters and Shape Recognition. *IEEE Trans. Image Processing* 12(9), 2003.
14. D. D. Hoffman and W. A. Richards, Parts of Recognition, *Cognition* 18, 65–96, 1984.
15. D. D. Hoffman and M. Singh. Salience of Visual Parts. *Cognition* 63, p. 29-78, 1997.
16. A. Khotanzan and Y. H. Hong. Invariant image recognition by Zernike moments. *IEEE Trans. PAMI* 12, p. 489–497, 1990.
17. R. Lakaemper and L. J. Latecki. Shape Search Engine. `http://knight.cis.temple.edu/~shape/`
18. L. J. Latecki and R. Lakaemper: Convexity Rule for Shape Decomposition Based on Discrete Contour Evolution. *Computer Vision and Image Understanding (CVIU)* 73, 441-454, 1999.
19. L. J. Latecki and R. Lakaemper: Shape Similarity Measure Based on Correspondence of Visual Parts. *PAMI* 22, p. 1185-1190, 2000.
20. L. J. Latecki, R. Lakaemper, and U.Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. *CVPR*, p. 424-429, 2000.
21. L. J. Latecki, A. Gross, and R. Melter (eds.): Special Issue on Shape Representation and Similarity for Image Databases. *Pattern Recognition*, 35(1), 2002.
22. L. J. Latecki, R. Lakaemper, and D. Wolter. Shape Similarity and Visual Parts. *DGCI*, Naples, Italy, November 2003.

23. F. Mokhtarian and A. K. Mackworth. A Theory of Multiscale, Curvature-Based Shape Representation for Planar Curves. *IEEE Trans. PAMI* 14, p. 789–805, 1992.
24. F. Mokhtarian and M. Bober. *Curvature Scale Space Representation: Theory, Applications and MPEG-7 Standardization.* Kluwer Academic, 2003.
25. K. Mueller, J.-R. Ohm, J. Cooper, and M. Bober. Results of 2d/3d shape core experiments ms-4. In *ISO/IEC/JTC1/ SC29/WG11.* MPEG/M6190, July 2000.
26. T. B. Sebastian, P. Klien, and B. B. Kimia. On aligning curves. *PAMI*, 25, p. 116-125, 2003.
27. K. Siddiqi, and B. B. Kimia. Parts of visual form: Computational aspects. *PAMI*, 17, p. 239-251, 1995.
28. K. Siddiqi, K. J. Tresness, and B. B. Kimia. Parts of visual form: Ecological and psychophysical aspects. *Perception*, 25, p. 399-424, 1996.
29. B. S. M. P. Salembier and T. Sikora, editors. *Introduction to MPEG-7: Multimedia Content Description Interface.* JohnWiley and Sons, 2002.