Properties and Performance of Shape Similarity Measures

Remco C. Veltkamp[1] and Longin Jan Latecki[2]

[1] Dept. Computing Science, Utrecht University
Padualaan 14, 3584 CH Utrecht, The Netherlands
Remco.Veltkamp@cs.uu.nl
[2] Dept. of Computer and Information Sciences, Temple University
Philadelphia, PA 19094, USA
latecki@temple.edu

Abstract. This paper gives an overview of shape dissimilarity measure properties, such as metric and robustness properties, and of retrieval performance measures. Fifteen shape similarity measures are shortly described and compared. Their retrieval results on the MPEG-7 Core Experiment CE-Shape-1 test set as reported in the literature and obtained by a reimplementation are compared and discussed.

1 Introduction

Large image databases are used in an extraordinary number of multimedia applications in fields such as entertainment, business, art, engineering, and science. Retrieving images by their content, as opposed to external features, has become an important operation. A fundamental ingredient for content-based image retrieval is the technique used for comparing images. It is known that human observers judge images as similar if they show similar objects. Therefore, similarity of objects in images is a necessary component of any useful image similarity measure. One of the predominant features that determine similarity of objects is shape similarity.

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 Motion Picture Expert Group (MPEG), a working group of ISO/IEC (see `http://www.chiariglione.org/mpeg/`) has defined the MPGE-7 standard for description and search of audio and visual content. A region based and a contour based shape similarity method are part of the standard. The data set created by the MPEG-7 committee for evaluation of shape similarity measures (Bober et al. (1999), Latecki, Lakaemper and Eckhardt (2000)) offers an excellent possibility for objective experimental comparison of the existing approaches evaluated based on the retrieval rate. The shapes were restricted to simple pre-segmented shapes defined by their outer closed contours. The goal of the MPEG-7 Core Experiment CE-Shape-1 was to evaluate the performance of 2D shape descriptors under change of

a view point with respect to objects, non-rigid object motion, and noise. In addition, the descriptors should be scale and rotation invariant.

2 Properties

In this section we list a number of possible properties of similarity measures. Whether or not specific properties are desirable will depend on the particular application, sometimes a property will be useful, sometimes it will be undesirable. A shape dissimilarity measure, or distance function, on a collection of shapes S is a function $d : S \times S \to \mathbb{R}$. The following conditions apply to all the shapes A, B, or C in S.

1 (Nonnegativity) $d(A, B) \geq 0$.

2 (Identity) $d(A, A) = 0$ for all shapes A.

3 (Uniqueness) $d(A, B) = 0$ implies $A = B$.

4 (Strong triangle inequality) $d(A, B) + d(A, C) \geq d(B, C)$.

Nonnegativity (1) is implied by (2) and (4). A distance function satisfying (2), (3), and (4) is called a metric. If a function satisfies only (2) and (4), then it is called a semimetric. Symmetry (see below) follows from (4). A more common formulation of the triangle inequality is the following:

5 (Triangle inequality) $d(A, B) + d(B, C) \geq d(A, C)$.

Properties (2) and (5) do not imply symmetry.

Similarity measures for partial matching, giving a small distance $d(A, B)$ if a part of A matches a part of B, in general do not obey the triangle inequality. A counterexample is the following: the distance from a man to a centaur is small, the distance from a centaur to a horse is small, but the distance from a man to a horse is large, so $d(man, centaur) + d(centaur, horse) \geq d(man, horse)$ does not hold. It therefore makes sense to formulate an even weaker form:

6 (Relaxed triangle inequality) $c(d(A, B) + d(B, C)) \geq d(A, C)$, for some constant $c \geq 1$.

7 (Symmetry) $d(A, B) = d(B, A)$.

Symmetry is not always wanted. Indeed, human perception does not always find that shape A is equally similar to B, as B is to A. In particular, a variant A of prototype B is often found more similar to B than vice versa.

8 (Invariance) d is invariant under a chosen group of transformations G if for all $g \in G$, $d(g(A), g(B)) = d(A, B)$.

For object recognition, it is often desirable that the similarity measure is invariant under affine transformations.

The following properties are about robustness, a form of continuity. They state that a small change in the shapes lead to small changes in the dissimilarity value. For shapes defined in $\mathbb{R}^2$ we can require that an arbitrary small change in shape leads to an arbitrary small in distance, but for shapes in $\mathbb{Z}^2$ (raster images), the smallest change in distance value can be some fixed value larger than zero. We therefore speak of an 'attainable $\epsilon > 0$'.

9 (Deformation robustness) For each attainable $\epsilon > 0$, there is an open set F of homeomorphisms sufficiently close to the identity, such that $d(f(A), A) < \epsilon$ for all $f \in F$.

10 (Noise robustness) For shapes in $\mathbb{R}^2$, noise is an extra region anywhere in the plane, and robustness can be defined as: for each $x \in (\mathbb{R}^2 - A)$, and each attainable $\epsilon > 0$, an open neighborhood U of x exists such that for all B, $B - U = A - U$ implies $d(A, B) < \epsilon$. When we consider contours, we interpret noise as an extra region attached to any location on the contour, and define robustness similarly.

3 Performance

First we shortly describe the settings of the MPEG-7 Core Experiment CE-Shape-1. The core experiment was divided into part A: robustness to scaling (A1) and rotation (A2), part B: performance of the similarity-based retrieval, and part C: robustness to changes caused by non-rigid motion.

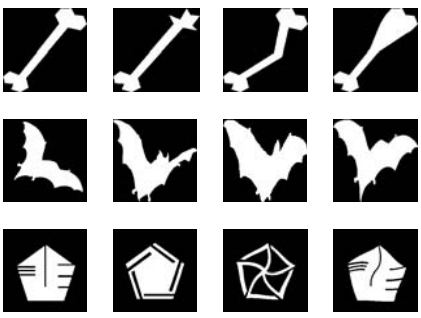

Fig. 1. Some shapes used in MPEG-7 Core Experiment CE-Shape-1 part B.

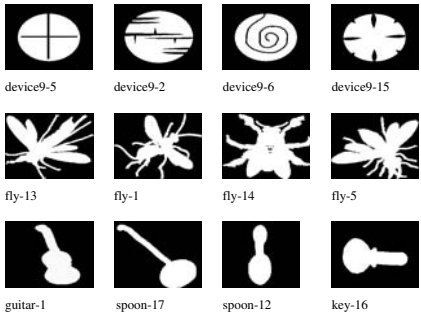

Fig. 2. The shapes with the same name prefix belong to the same class.

Part A can be regarded as a useful 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 non-rigid motion is considered. Only one query shape is used for part C.

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 data set therefore consists of 1400 shapes. For example, each row in Fig. 1 shows four shapes from the same class.

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 (Bober et al. (1999)). Each image was used as a query, and the retrieval rate is expressed by the so called Bull's Eye Percentage (BEP): the fraction 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 shape variations within the same classes make that no shape similarity measure achieves a 100% retrieval rate. E.g., see the third row in Fig. 1 and the first and the second rows in Fig. 2. The third row shows spoons that are more similar to shapes in different classes than to themselves.

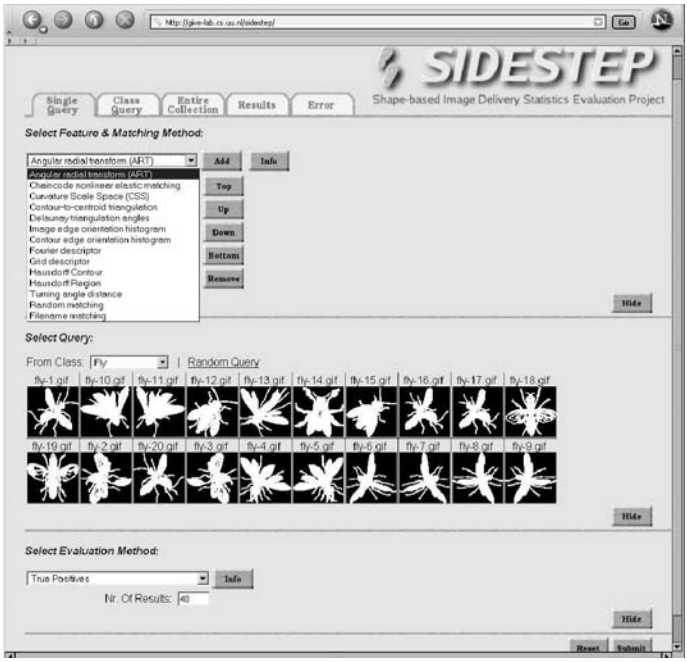

Fig. 3. SIDESTEP interface.

To compare the performance of similarity measures, we built the framework SIDESTEP – Shape-based Image Delivery Statistics Evaluation Project, `http://give-lab.cs.uu.nl/sidestep/`. Performance measures such as the

number of true/false positives, true/false negative, specificity, precision, recall, negative predicted value, relative error, k-th tier, total performance, and Bull's Eye Percentage can be evaluated for a single query, over a whole class, or over a whole collection, see Fig. 3.

4 Shape similarity measures

In this section we list several known shape similarity measures and summarize some properties and their performance in Table 1 on the MPEG-7 CE-Shape-1 part B data set. The discussion of the results follows in Section 5.

Shape context (Belongie, Malik and Puzicha (2002)) is a method that first builds a shape representation for each contour point, using 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.

Image edge orientation histogram (Jain and Vailaya (1996)) is built by applying an edge detector to the image, then going over all pixels that lie on an edge, and histogramming the local tangent orientation.

Hausdorff distance on region is computed in the following way. First a normalization of the orientation is done by computing the principal axes of all region pixels, and then rotating the image so that the major axis is aligned with the positive x-axis, and the minor axis with the positive y-axis. The scale is normalized by scaling the major axes all to the same length, and the y-axes proportionally. Then the Hausdorff distance between the sets A and B of region pixels is computed: the maximum of all distances of a pixel from A to B, and distances of a pixel from B to A. The Hausdorff distance has been used for shape retrieval (see for example Cohen (1995)), but we are not aware of experimental results on the MPEG-7 Core Experiment CE-Shape-1 test set reported in the literature.

Hausdorff distance on contour is computed in the same way, except that it is based on set of all contour pixels instead of region pixels.

Grid descriptor (Lu and Sajjanhar (1999)) overlays the image with a coarse grid, and assigns a '1' to a grid cell when at least 15% of the cell is covered by the object, and a '0' otherwise. The resulting binary string is then normalized for scale and rotation. Two grid descriptors are compared by counting the number of different bits.

Fourier descriptors are the normalized coefficients of the Fourier transformation, typically applied to a 'signal' derived from samples from the contour, such as the coordinates represented by complex numbers. Experiments have shown that the centroid distance function, the distance from the contour to the centroid, is a signal that works better than many others (Zhang and Lu (2003)).

Distance set correspondence (Grigorescu and Petkov (2003)) is similar to shape contexts, but consists for each contour point of a set of distances to N nearest neighbors. Thus, in contrast to shape contexts, no angular information but only local distance information is obtained. The distance between two shapes is expressed as the cost of a cheapest correspondence relation of the sets of distance sets.

Delaunay triangulation angles are used for shape retrieval in Tao and Grosky (1999) by selecting high curvature points on the contour, and making a Delaunay triangulation on these points. Then a histogram is made of the two largest interior angles of each of the triangles in the triangulation. The distance between two shapes is then simply the L_2-distances between the histograms.

Deformation effort (Sebastian, Klien and Kimia (2003)) is expressed as the minimal deformation effort needed to transform one contour into the other.

Curvature scale space (CSS) (Mokhtarian and Bober (2003)) is included in the MPEG-7 standard. First simplified contours are obtained by convolution with a Gaussian kernel. The arclength position of inflection points (x-axis) on contours on every scale (y-axis) forms so called Curvature Scale Space (CSS) curve. 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.

Convex parts correspondence (Latecki and Lakaemper (2000)) 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 (Latecki and Lakaemper (1999)), the maximal convex contour parts represent visually significant shape parts. This correspondence is computed using dynamic programming.

Contour-to-centroid triangulation (Attalla and Siy (2005)) first picks the farthest point from the centroid of the shape and use it as the start point of segmenting the contour. It then divides the contour into n equal length arcs, where n can be between 10 and 75, and considers the triangles connecting the endpoints of these arcs with the centroid. It builds a shape descriptor by going clockwise over all triangles, and taking the left interior contour angle, the length of the left side to the centroid, and the ratio contour segment length to contour arc length. To match two descriptors, the triangle parameters are compared to the correspond triangle of the other descriptor, as well as to its left and right neighbor, thereby achieving some form of elastic matching.

Contour edge orientation histogram are built by going over all pixels that lie on object contours, and histogramming the local tangent orientation. It is the same as the 'image edge orientation histogram', but then restricted to pixels that lie on the object contour.

Chaincode nonlinear elastic matching (Cortelazzo et al. (1994)) represents shape in images as a hierarchy of contours, encoded as a chaincode string: characters '0' to '7' for the eight directions travelling along the contour. Two images are compared by string matching these chaincode strings. Various different string matching methods are possible, we have taken the 'nonlinear elastic matching' method.

Angular radial transform (ART) is a 2-D complex transform defined on a unit disk in polar coordinates. A number of normalized coefficients form the feature vector. The distance between two such descriptors is simply the L_1 distance. It is a region-based descriptor, taking into account all pixels describing the shape of an object in an image, making it robust to noise. It is the region-based descriptor included in the MPEG-7 standard (Salembier and Sikora (2002)).

Table 1. Performances and properties of similarity measures.

method	unique	deform	noise	BEP reported	BEP reimpl
Shape context	+	+	+	76.51	
Image edge orientation histogram	−	+	+		41
Hausdorff region	+	+	−		56
Hausdorff contour	+	+	+		53
Grid descriptor	−	+	+		61
Distance set correspondence	+	+	+	78.38	
Fourier descriptor	−	+	+		46
Delaunay triangulation angles	−	−	−		47
Deformation effort	+	+	+	78.18	
Curvature scale space	−	+	+	81.12	52
Convex parts correspondence	−	+	+	76.45	∼
Contour-to-centroid triangulation	−	−	−	84.33	79
Contour edge orientation histogram	−	+	+		41
Chaincode nonlinear elastic matching	+	+	+		56
Angular radial transform	+	+	+	70.22	53

5 Discussion

The Angular radial transform, the grid descriptor, the 'Hausdorff region', and image edge orientation histogram are region based methods, all others

work only for shapes defined by a single contour. Naturally, the region based methods can also be applied to contour shapes.

Even though invariance under transformations is not always a property of the base distance, such as the Hausdorff distance, it can be easily obtained by a normalization of the shape or image, as many of the methods do.

Table 1 tabulates a number of properties and performances of the similarity measures listed in section 4. The column 'unique' indicates whether (+) or not (−) the method satisfies the uniqueness property, 'deform' indicates deformation robustness, 'noise' indicates robustness with respect to noise, 'BEP reported' lists the Bull's Eye Percentage reported in the literature, 'BEP reimpl' lists the BEP of the reimplementations (performed by master students) plugged into SIDESTEP. The symbol ∼ indicates that the method is of one of the authors.

Methods that are based on sampling, histogramming, or other reduction of shape information do not satisfy the uniqueness property: by throwing away information, the distance between two shapes can get zero even though they are different.

The methods that are based on angles, such as the 'Contour-to-centroid triangulation' and 'Delaunay triangulation angles' methods, are not robust to deformation and noise, because a small change in the shape can lead to a large change in the triangulation.

The Hausdorff distance on arbitrary sets is not robust to noise (an extra region anywhere in the plane), and therefore also not for regions. However, for contours, we interpret noise as an extra point attached to any contour location. As a result the Hausdorff distance on contours is robust to noise.

Fourier descriptors have been reported to perform better than CSS (Zhang and Lu (2003)), but the comparison has not been done in terms of the Bull's Eye Percentage.

It is remarkable that the 'Contour-to-triangulation' does not satisfy, theoretically, uniqueness and robustness properties, while in practice it performs so well. This is explained by the fact that the method does not satisfy the property for *all* shapes, while the performance is measured only on a limited set of shapes, where apparently the counterexamples that prevent the method from obeying the property simply don't occur.

The difference between the Bull's Eye Percentages of the method as reported in the literature and the performances of the reimplement methods is significant. Our conjecture is that this is caused by the following. Firstly, several methods are not trivial to implement, and are inherently complex. Secondly, the description in the literature is often not sufficiently detailed to allow a straightforward implementation. Thirdly, fine tuning and engineering has a large impact on the performance for a specific data set. It would be good for the scientific community if the reported test results are made reproducible and verifiable by publishing data sets and software along with the articles.

The most striking differences between the performances reported in the literature and obtained by the reimplementation are the ones that are part of the MPEG-7 standard: the Curvature Scale Space and the Angular Radial Transform. In the reimplementation of both methods we have followed closely the precise description in the ISO document (Yamada et al. (2001)), which is perhaps less tweaked towards the specific MPEG-7 Core Experiment CE-Shape-1 test set.

The time complexity of the methods often depends on the implementation choices. For example, a naive implementation of the Hausdorff distance inspects all $O(N^2)$ pairs of points, but a more efficient algorithm based on Voronoi Diagrams results in a time complexity of $O(N \log N)$, at the expense of a more complicated implementation.

Acknowledgement This research was supported by the FP6 IST projects 511572-2 PROFI and 506766 AIM@SHAPE, and by a grant NSF IIS-0534929. Thanks to Dènis de Keijzer and Geert-Jan Giezeman for their work on SIDESTEP.

References

ATTALLA, E. and SIY, P. (2005): Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching. Patt. Recogn. 38, 22292241.

BELONGIE, S., MALIK, J. and PUZICHA, J. (2002): Shape Matching and Object Recognition Using Shape Contexts. IEEE PAMI, 24(24), 509-522.

BOBER, M., KIM, J.D., KIM, H.K., KIM, Y.S., KIM, W.-Y. and MULLER, K. (1999): Summary of the results in shape descriptor core experiment. ISO/IEC JTC1/SC29/WG11/MPEG99/M4869.

COHEN, S. (1995): Measuring Point Set Similarity with the Hausdorff Distance: Theory and Applications. Ph.D thesis, Stanford University.

CORTELAZZO, G., MIAN, G.A., VEZZI, G. and ZAMPERONI, P. (1994): Trademark Shapes Description by String-Matching Techniques. Patt. Recogn. 27(8), 1005-1018.

GRIGORESCU, C. and PETKOV, N. (2003): Distance Sets for Shape Filters and Shape Recognition. IEEE Trans. Image Processing, 12(9).

JAIN, A.K. and VAILAYA, A. (1996): Image Retrieval using Color and Shape. Patt. Recogn. 29(8), 1233-1244.

LATECKI, L.J. and LAKAEMPER, R. (1999): Convexity Rule for Shape Decomposition Based on Discrete Contour Evolution. Computer Vision and Image Understanding 73, 441-454.

LATECKI, L.J. and LAKAEMPER, R. (2000): Shape Similarity Measure Based on Correspondence of Visual Parts. IEEE PAMI 22, 1185-119.

LATECKI, L.J., LAKAEMPER, R. and ECKHARDT, U. (2000): Shape descriptors for non-rigid shapes with a single closed contour. Proc. CVPR, 424-429.

LU, G. and SAJJANHAR, A. (1999): Region-based shape representation and similarity measure suitable for content-based image retrieval. Multimedia Systems, 7, 165174.

MOKHTARIAN, F. and BOBER, M. (2003): Curvature Scale Space Representation: Theory, Applications and MPEG-7 Standardization. Kluwer Academic.

SEBASTIAN, T.B., KLIEN, P. and KIMIA, B.B. (2003): On aligning curves. IEEE PAMI, 25, 116-125.

SALEMBIER, B.S.M.P. and SIKORA, T., editors (2002): Introduction to MPEG-7: Multimedia Content Description Interface. JohnWiley and Sons.

TAO, Y. and GROSKY, W.I. (1999): Delaunay trriangularion for image object indexing: a novel method for shape representation. Proc. 7th SPIE Symposium on Storage and Retrieval for Image and Video Databases, 631-642.

YAMADA, A., PICKERING, M., JEANNIN, S., CIEPLINSKI, L., OHM, J.R. and KIM, M. (2001): MPEG-7 Visual part of eXperimentation Model Version 9.0. ISO/IEC JTC1/SC29/WG11/N3914.

ZHANG, D. and LU, G. (2003): Evaluation of MPEG-7 shape descriptors against other shape descriptors. Multimedia Systems 9, 1530.