Optimal partial shape similarity

Longin Jan Latecki[a,*], Rolf Lakaemper[a], Diedrich Wolter[b]

[a]*Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19094, USA*
[b]*FB 3 - Cognitive Systems, Universität Bremen, Germany*

Abstract

Humans are able to recognize objects in the presence of significant amounts of occlusion and changes in the view angle. In human and robot vision, these conditions are normal situations and not exceptions. In digital images one more problem occurs due to unstable outcomes of the segmentation algorithms. Thus, a normal case is that a given shape is only partially visible, and the visible part is distorted. To our knowledge there does not exist a shape representation and similarity approach that could work under these conditions. However, such an approach is necessary to solve the object recognition problem. The main contribution of this paper is the definition of an optimal partial shape similarity measure that works under these conditions. In particular, the presented novel approach to shape-based object recognition works even if only a small part of a given object is visible and the visible part is significantly distorted, assuming the visible part is distinctive.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Shape similarity; Visual parts; Shape representation; Object recognition

1. Introduction and motivation

When trying to compare natural shapes, one is seldom successful in finding an identical match. How can shape similarity then be defined in accordance to human perception? Our intuitive definition of similarity emphasizes the contribution of common features of the objects compared to those distinguishing them. As an example, imagine the shape similarity of a centaur to a human and a horse; although there is no global similarity between the particular similar parts: the upper body makes the centaur similar to a human, while the lower body makes it similar to a horse. As a result, we interpret a centaur to be similar to a human as well as to a horse.

This example suggests that one needs to concentrate on parts that are common and disregard parts where objects differ. Humans judge two objects as being similar if they have common parts that are similar and that are significant for the shape of both objects. The question arises how to identify the common parts. A conceptually simple

procedure to answer this question is to try to remove certain parts and see whether the objects become more similar without them. This leads to a first approximation of the proposed definition of shape similarity:

The *optimal partial similarity* of Q to T is the similarity of Q to modified T, denoted as T_Q^*, where T_Q^* is T with all parts that make T distinct from Q removed.

We need to state clearly that there are two different shape similarity concepts in this definition. The first one, which we will call 'global shape similarity', is the measure that compares Q to modified T (without the process of modification). All shape similarity measures presented in the literature define this global shape similarity. However, it is not what humans understand under shape similarity. The second concept is the defined concept of optimal partial shape similarity. It abstracts from distinct parts before comparing the modified shapes using the global shape similarity. Hence it is much closer to human perception. Let s be any global shape similarity measure, more formally, s is a distance measure (the smaller its values, the greater is the similarity of compared shapes). Now we restate the proposed definition with the conceptual confusion removed:

* Corresponding author.
E-mail addresses: latecki@temp.edu (L.J. Latecki), lakamper@temple.edu (R. Lakaemper), dwolter@informatik.uni-bremen.de (D. Wolter).

The *optimal partial similarity* of Q to T, $ops(Q, T)$, is the global minimum of global similarities $s(Q, T - S_Q)$ taken over all parts S_Q of T.

A subpart S_Q^* of T that yields the global minimum identifies the parts that makes T most dissimilar to Q. Consequently, a subpart $T_Q^* = T - S_Q^*$ of T is the part of T that makes T most similar to Q. We call T_Q^* the optimal subpart of T with respect to Q. Thus, *ops* measures how similar to Q part of T is that makes T most similar to Q or, more simply stated, how much of the shape of Q is contained in T. Clearly, additionally we could introduce a penalty for the removed parts. To keep the proposed definition as simple as possible, and since this definition in its simple form is sufficient for many applications, we will consider its extensions later in the paper (Section 7). Observe also that this definition is not symmetric, i.e. the query part Q is used in its entirety. This implies that the query part should be carefully selected. It should be sufficiently distinctive to allow us to distinguish the shape of the object we want to retrieve from other objects. It should also be descriptive in that it should match well the part of the object we are looking for. The process of query-part selection is similar to keyword selection in text-based searches, where a selected keyword must be distinctive and descriptive. The main contribution of this paper, then, can be made clear using the analogy of text-based retrieval. Existing shape similarity measures force the user to submit a description of the entire object shape (i.e. the whole object contour) as the query, which corresponds to submitting the whole sentence as the query, making it very unlikely to find a good match. The proposed partial similarity allows the user to submit key parts of the shape (parts of object contours) as queries the same way keywords are used in text-based searches.

Although we focus here on the shape of 2D objects, our approach is also related to shape recognition of 3D objects. The shape of 3D objects can be recognized using their 2D projections [10,22]. The key observation is that the contour variation in the projection of a single visual part is significantly smaller than the variation in the projection of the whole 3D object. Moreover, a significant occlusion makes the recognition of the whole contour impossible, while a part of contour can still be recognized as long as it is visible.

2. Technical introduction

We use boundary representation of shape. Since we work with the boundaries of objects in digital images, we can assume without loss of generality that shape is represented by polylines (polygonal curves) that form their boundaries. Given a query part Q (that is a visually significant part) and a target object T, the goal is to find part P of T that is the most similar to the query Q (Fig. 1). An initial, simple thing to do is to try out all subparts P of T and compare them to Q. Since we work with polylines, and each polyline can be
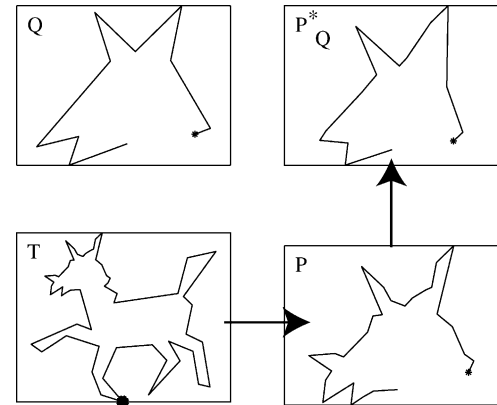

Fig. 1. We are not only able to find the best matching part P of the target object T for a given query part Q but also to modify P to P_Q^*, composed of only the features of P that are similar to Q. The algorithm introduced in Section 4 automatically computed the result shown.

viewed as an ordered set of vertices, this goal can be easily achieved by comparison of all subsets of vertices of T to Q using some global shape similarity measure s and then selecting as P the subset of vertices that yields a global minimum of s. However, this strategy leads to combinatorial explosion that occurs since we consider all subsets of vertices of T. Clearly, the combinatorial explosion can easily be avoided if only all connected subsets P of vertices of T are considered, which reduces the complexity to $O(n^2)$, where n is the number of vertices of T. This strategy is known as sliding window with variable size. However, it does not produce desired results. The problem is that the subpart P of the target shape T that correctly corresponds to Q has an additional shape feature (the horn on the snout), Fig. 1. Since the horn is a significant shape feature of P (viewed as a single shape), it makes P significantly dissimilar to Q. Therefore, any global shape similarity measure applied to Q and P yields a value indicating that Q and P are dissimilar. Consequently, a sliding window approach with any global shape similarity measure may fail to find P as the most similar subpart of T to query Q.

In contrast, the proposed optimal partial similarity never directly compares Q and P, instead we first simplify P in the context of Q to obtain P_Q^*, and then compare Q and P_Q^* using a global shape similarity measure. Hence the partial shape similarity provides a solution to the problem of additional shape features without running into the problem of combinatorial explosion. The simplification of T to P followed by the simplification of P to P_Q^* is obtained by a single process that recursively removes carefully selected vertices of T. In each step a vertex that makes T most distinct from Q is selected for removal. This process is computed by the algorithm introduced in Section 4. The outcome is shown by two arrows in Fig. 1. After removing a first set of vertices, T is transformed into P (horizontal arrow). The second set of removed vertices transforms P to P_Q^* (vertical arrow). Thus, what remains is $P_Q^* = T_Q^*$, a modified T that is composed only of the features of T that make T similar to Q.

Observe that not all removed features of P to obtain P_Q^* can be viewed as noise, in particular the horn on the snout. Note that the length of the most similar part P_Q^* of T to the query Q, in this example, is only about 20% of the length of T.

3. Relevant work

We use the term visual parts more loosely than parts of visual form as defined in [12,13]. In the context of this paper, visual parts mean significant parts of object contours that are good candidates for identifying the objects. Our definition includes parts of visual form, but we also allow a composition of visual parts to be called a visual part, e.g. the contour of a hand is a visual part of the human body, according to our definition. In any case, an important feature of our approach is that we use parts of objects as query shapes. According to Siddiqi et al. [26], part-based representations allow for robust object recognition and play an important role in theories of object categorization and classification. There is also strong evidence for part-based representations in human vision, see e.g. [12,13,25,26].

There is a huge variety of shape descriptors, most of them requiring the presence of the whole shape, some of them tolerating minor missing or distorted parts, e.g. by occlusion. Even feature-based approaches, although potentially being based on local features, require the presence of most of the object to compute the statistics of the features. This statement applies to all shape descriptors presented in the recent overview articles [17,18], as well as to the new shape descriptors presented in [4,11]. There exist feature-based approaches that allow for object recognition under occlusion and/or perspective distortion. However, in all these approaches only a relatively small part of the object may be occluded. To these approaches belongs probabilistic model of spatial arrangement of local features [8,27,28], where the object shape is modeled as spatial arrangement of point features. Due to a sophisticated Gaussian mixture model and the correspondence computation of point configurations that employs an EM algorithm, this approach works even if a small part of feature points is not present. This means that a small part of the object may be occluded. In contrast the presented approach works even if only a small part of the object is visible.

The process of simplification of target shape T in the context of query part Q is a context sensitive extension of discrete curve evolution (DCE) [15,16]. We simplify polyline T by removing the set of vertices whose removal yields the highest gain in the similarity value of the simplified T to query Q. This way we check how much of the shape of Q is contained in the shape of P. This is in contrast to all deformation energy approaches (Basri et al. [3], Sebastian et al. [24]), where a query shape is deformed until it becomes a target shape and the amount of deformation defines the similarity value. In our case the whole process is driven by a shape similarity measure, but

does not define the shape similarity value. The further differences are that we allow only for shape simplification, and we simplify until a global minimum of similarity is reached.

Deformation energy approaches use penalties for length differences of matched curves. This makes it impossible to recognize the same shape at different scales. The proposed approach is scale invariant, since before query Q is compared to simplified part $T-S_Q$ of target T both curves are normalized to length one without any penalty. Observe that scaling both curves to the same length alone, does not solve the problem of different scales when both curves are not very similar. The simplification of T combined with scaling to the same length provides a solution to the problem of different scales.

In contrast to early AI and CV approaches, we do not regard objects to be composed of primitive parts nor a specific shape vocabulary like generalized cylinders, super-quadrics, or geons [5–7,23]. The visual parts we are interested in are not built of any primitive elements. Therefore, we have no restriction on shape that may be complex.

Our approach applies also to the recognition of 3D objects. There is substantial cognitive evidence that recognition of 3D objects can be based on 2D projections, and planar shape similarity measure can be used in this context (Cyr and Kimia [10]).

4. Computation of the optimal partial shape similarity

Given a query polyline Q and a target polyline T, we face two related goals: (1) to localize part P of target polyline T that is most similar to Q, and (2) to simplify P to query polyline Q such that the simplified version of P is most similar to Q. Both goals will be achieved by a single process of simplification of T in the context of Q described in this section. To achieve these goals we need a global shape similarity measure of high quality, which we call s, that can be applied to compare two polylines. We use an improved version of the cognitively motivated shape similarity measure introduced in [16], which we describe in Section 6.

A polyline T can be defined as an ordered set of vertices $T = \{t_1, \ldots, t_n\}$. Our goal is to find and remove a subset S_Q^* of vertices of T so that the polyline $T_Q^* = T - S_Q^*$ is the most similar subpolyline of T to Q. Thus, we find S_Q^* as argument of the global minimum

$$S_Q^* = \mathrm{argmin}\{s(Q, T - S_Q) : S_Q \subseteq T\}$$

and the *optimal partial similarity* between Q and T is defined by

$$ops(Q, T) = \min\{s(Q, T - S_Q) : S_Q \subseteq T\}.$$

The length of both polylines Q and $T-S_Q$ is normalized to one before $s(Q, T-S_Q)$ is computed.

Observe again that there are fundamental differences of our approach to the deformation energy approaches ([3], [24]). First, we only permit the simplification of a given shape, i.e. we do not allow for arbitrary deformation. Second we do not measure the cost of deforming a shape, but instead the shape similarity after deformation. A nice feature of this definition is that we are always guaranteed to obtain a global minimum of our shape similarity (therefore we are parameter free), but its computation may lead to combinatorial explosion.

Therefore, we now introduce a suboptimal algorithm to compute *ops*. It is suboptimal in the sense that we select in an optimal way a single vertex to be removed in each step, but the subset of all removed vertices may not be optimal. First, we recursively generate a sequence of polylines

$$T = T^n, T^{n-1}, ..., T^2$$

in which T^{k-1} is obtained by removing a single vertex from T^k such that

$$T_Q^{k-1} = \text{argmin}\{s(Q, T^k - \{x\}) : x \in T^k\}.$$

Then we compute the global minimum of similarities between Q and T_Q^k:

$$ps(Q, T) = \min\{s(Q, T_Q^k) : k = 2, ..., n\}.$$

and

$$T_Q^* = \text{argmin}\{s(Q, T_Q^k) : k = 2, ..., n\}.$$

The length of both polylines Q and T_Q^k is normalized to one before $s(Q, T_Q^k)$ is computed.

We call the obtained similarity measure *ps* partial similarity to stress the fact that it is not globally optimal any more. The computation of the global minimum over all possible vertex subsets of T is substituted by a global minimum over the sequence of optimally removed single vertices.

An important property of the defined partial similarities *ops* and *ps* is the fact that they are invariant to scale differences between Q and T, since we normalize the length of polylines compared by s. Observe that if Q and T are at different scales, making their arc length equal, does not solve the problem of different scales. The reason is that actually Q is only similar to a subpolyline T_Q^* of T. Hence making Q and T_Q^* to be of the same length solves the problem of different scales. This is what happens during the proposed computation of *ops* and *ps*, because when global similarity measure s is used to compare query Q to simplified target T both curves are scaled to length one. To summarize simplification of T combined with scaling to the same length provides a solution to the problem of different scales.

The vertex removal process has complexity $O(n^2)$, where n is the number of vertices of T. This does not account for complexity of the global shape similarity measure s that is used in each step (defined in Section 6). Since s can be computed in $O(n \log(n))$, the total complexity of partial shape matching is $O(n^3 \log(n))$.

5. Experimental illustration

The best matching part T_Q^* of T to query Q shown in Fig. 1 was computed by the proposed algorithm for partial similarity *ps*. A few simplification stages of P that led to identification of T_Q^* are shown in Fig. 2. Only a subset of all simplified versions of P is depicted. The polylines shown are marked with letters A to F, where A is part P from Fig. 1. A plot of the values of $s(Q, T_q^k)$ is shown in Fig. 3. The global minimum of the similarity values is obtained for $k = 15$, which is for polyline $E = P_Q^* = T_Q^*$.

As stated before, not all removed features from polyline $P = A$ to obtain polyline E can be viewed as noise, in particular the horn on the snout. This feature can only be removed in the context of the query Q. Since the horn is a significant shape feature of P (viewed as a single object), it makes P significantly dissimilar to Q. Therefore, any classical shape similarity measure yields a value that indicates that Q and P are dissimilar.

In addition to several successful experiments as the one presented in Figs. 1–3, we also evaluated the retrieval rate of our optimal shape similarity measure on the dataset created by the MPEG-7 committee for evaluation of shape similarity measures [9,17]. 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. 4 shows four shapes from the same class.

The goal of this experiment is to prove the ability of our approach to correctly retrieve the search objects when only a single part is given as query part Q. As in the case of a single keyword query in the text-based search, the query part Q must be a descriptive feature that on the one hand is common to all searched shapes and on the other hand must be able to discriminate the search shapes from shapes of other classes. For this experiment a sufficiently descriptive
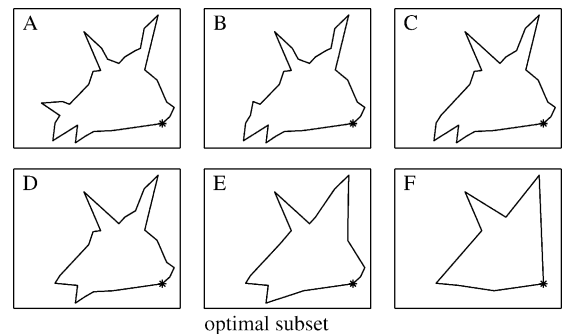

optimal subset

Fig. 2. Illustration of the computation of partial similarity $ps(Q, A)$, where A is part P from Fig. 1. We see a few simplified versions of A. The global minimum is obtained for $P_Q^* = E$.
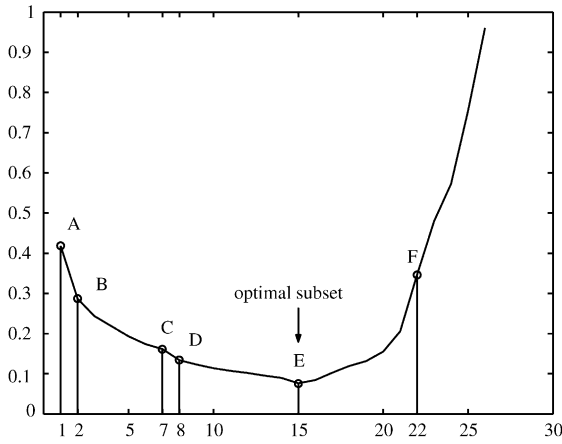
Fig. 3. The plot of the values of the similarity measure $s(Q, T_q^k)$. The global minimum is obtained for $k=15$, which is for polyline $E = P_Q^* = T_Q^*$ (shown in Fig. 2).

Fig. 5. Top: the 21 most similar shapes retrieved from the MPEG-7 shape database for the query part Q (shown on top). Bottom: the subparts of the top shapes that are similar to Q.

part Q is given by the cross-like top part of the shape 'fountain-01', shown in Fig. 5, top left. Every shape of the MPEG dataset was compared to Q using the partial similarity measure $ps(Q,T_i)$, for $i=1,\dots,1400$. The top section of Fig. 5 shows the first 21 shapes that contain parts that are the most similar to Q. The bottom section shows the corresponding parts, i.e. the optimal subsets $T_{iQ}^* \subseteq T_i$. The algorithm found the desired 20 shapes of class 'fountain$-$01' in the top 21 results. The only misclassification (object 'device0$-$16', rank 19, marked with '*') contains a part visually similar to Q. Note that this result was achieved with the query being a single part only.

We performed a similar experiment on a small database composed of 27 shapes shown in Fig. 6. The shapes are grouped into 9 classes with 3 shapes each. The set of our query parts is shown in Fig. 7; two query parts for each of the 9 classes were selected. We measured the retrieval rate following the MPEG-7 Bulls-Eye test [9]: the number of objects from the same class that are contained among the first N most similar objects to the query part, where N is equal to double the number of objects in the given class ($N=6$ in our test). The overall retrieval rate for selected

query parts (Fig. 7) is 100%. When complete object contours of the objects in the database are submitted as queries and we use the global shape similarity s, the retrieval rate is 91.95%, which is worse than the one for query by parts. This experiment demonstrates the gain obtained by using the proposed partial shape similarity for parts as queries. The result of this test is particularly significant, since the partial shape similarity is based on the global shape similarity s. There are two points that must be addressed here. Clearly, the excellent performance of the partial shape similarity depends on the selection of query
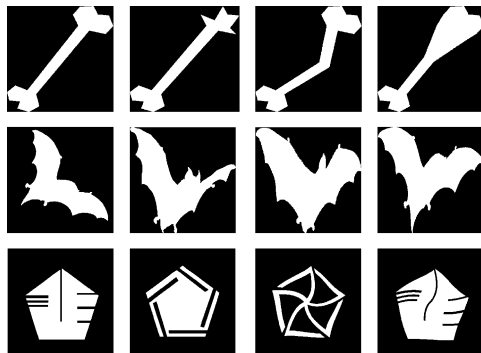

Fig. 4. Some shapes used in part B of MPEG-7 core experiment CE-Shape-1. Shapes in each row belong to the same class, i.e. we see in the first row four different shapes (out of 20) of class 'bone'.
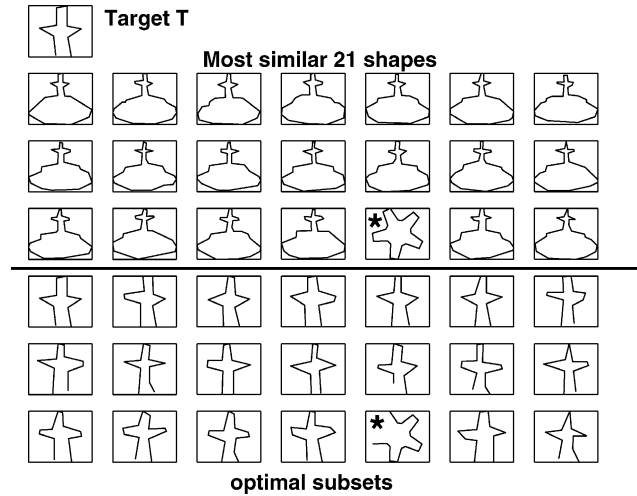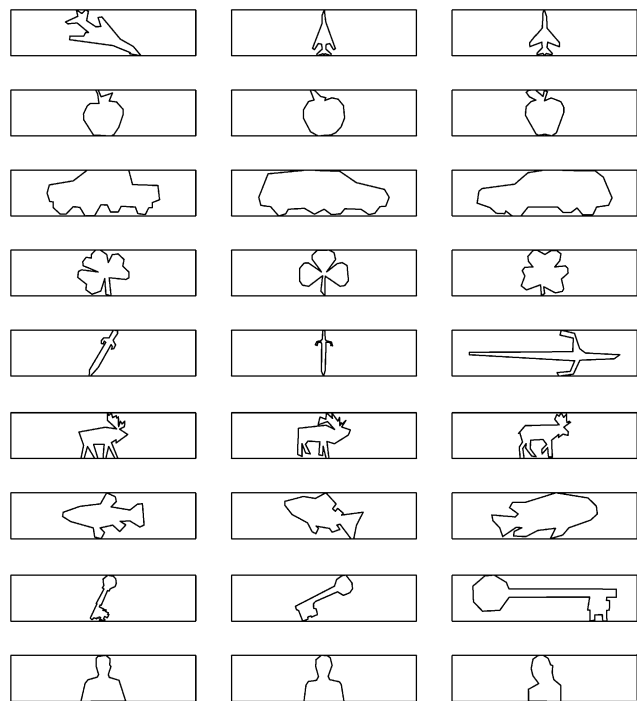

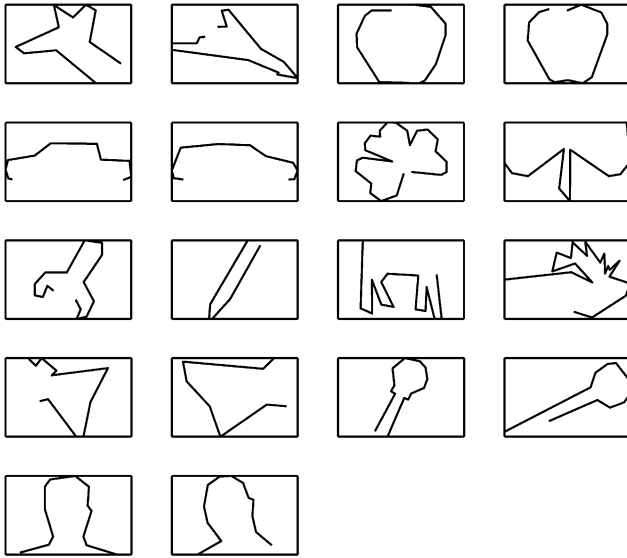Fig. 6. Our small test database composed of 27 objects grouped into 9 classes (shown in rows).

Fig. 7. The set of 18 query parts grouped into 9 classes corresponding to the 9 classes in the database in Fig. 6.

parts. However, it is easier for humans submitting the queries to sketch a query part than to sketch the whole contour, and the successful experience with key words shows that their selection is not a problem for humans. The second point is whether a different global shape similarity measure would do a better job. The performance of several global shape similarity measures on the MPEG-7 Shape1 data set [9] (queries consist of the whole shapes) strongly indicates that no global shape similarity measure is capable of yielding 100% retrieval rate due to variations in global shape. This is justified by the fact that our underlying global shape similarity measure *s* has a retrieval rate of 76.45% on the MPEG-7 data set. The best know retrieval rate for the MPEG-7 data Shape1 set is 83.19% [1]. The global shape similarity measure selected for the MPEG-7 standard has a retrieval rate of 81.12% [21]. It is based on the global measure introduced by Mokhtarian and Mackworth in [20].

The best six published retrieval rates for global shape measures (and consequently, for queries being the whole shapes) for the MPEG-7 data set are shown in Fig. 8. Further other retrieval results, all with retrieval rates lower than the rates in Fig. 8, for various global shape similarity approaches are recorded in the report on the original MPEG-7 Shape1 experiment in [9].

6. Global shape similarity measure

A global shape similarity measure is the underlying basis of the partial similarity. There exist numerous approaches to define the similarity between polygonal curves, some of which we mentioned in Section 3. However, since this measure is used to find the optimal window to drive the simplification process, and finally to provide the numerical value representing the partial similarity of arbitrary complex natural shapes, it must be robust to non-uniform contour deformations. We use an improved version of a *visual-part based shape similarity measure*(VPS) introduced in [16]. It yielded excellent performance on the MPEG-7 experiments reported in the previous section, which can also be verified by querying our online shape database at [14]. Like all global shape similarity measures VPS requires that the whole object is present as input. Since it is contour based, this means that the complete contour is given. (Although our experimental results show that our measure performs well in the presence of minor occlusions, or to be more precise: if some significant parts of the contour are absent and other significant parts are present, then the results are in accordance with our intuition.)

To compute VPS between two polygonal curves, we establish the best possible correspondence of maximal convex arcs. To achieve this, we first decompose the polygonal curves into maximal convex subarcs. Note that a convex subarc of an object contour may be either convex or concave with respect to the object area. Since a simple one-to-one comparison of maximal convex arcs of two polygonal curves is of little use, due to the fact that the curves may consist of a different number of such arcs and even similar shapes may differ in small features, we allow for 1-to-1, 1-to-many, and many-to-1 correspondences of the maximal convex arcs. The main idea here is that we have at least on one of the contours a maximal convex arc that corresponds to a part of the other contour composed of adjacent maximal convex arcs. In this context the corresponding parts of contours can be identified with visual object parts. Two example correspondences obtained by our approach are shown in Fig. 9.

The main advantage of VPS is that it is based on correspondence of visual parts. A correspondence of visual parts is defined as a sequence of pairs of parts of contours *A* and *B*

$$C = ((p_{1A}, p_{1B}), \ldots, (p_{nA}, p_{nB}))$$

such that $A = P_{1A} \cup \ldots \cup p_{nA}$ and $B = p_{1B} \cup \ldots \cup p_{nB}$, the order of parts is respected and at lease one part in each pair $d(p_{iA}, p_{iB})$ is a convex arc. The global shape similarity *s* is defined as the global minimum of the sum of distances $d_{arcs}(p_{iA}, p_{iB})$ over all possible correspondences, where polygonal similarity measure d_{arcs} is defined below. VPS uses dynamic programming to find the optimal

Retrieval	76.45	76.51	78.18	78.38	81.12	83.19
Reference	[20]	[4]	[28]	[14]	[25]	[1]
Authors	Latecki ...	Belongie ...	Sebastian ...	Grigorescu ...	Mokhtarian ...	Attalla

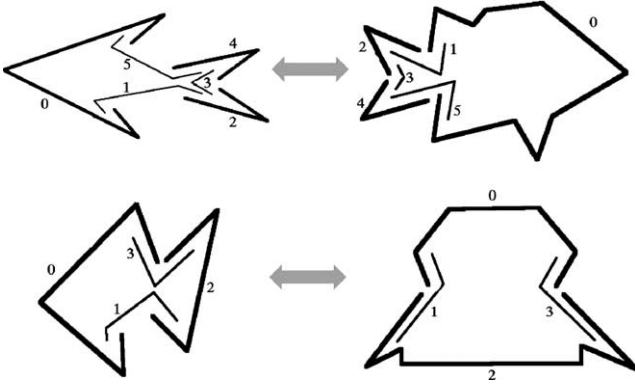Fig. 8. The best six published retrieval rates in percents on the MPEG-7 shape-1 part B dataset.

Fig. 9. Correspondence of visual parts as computed by our global shape similarity measure. Corresponding arcs are labeled with the same numbers.

correspondence. Fig. 10 shows an example of corresponding parts computed by VPS.

Basic similarity of polylines d_{arcs} is defined using their tangent function representation. Tangent function, also called turning function, is a multi-valued step function mapping a polyline into the interval $[0, 2\pi)$ by representing angular directions of line-segments only. Furthermore, arc lengths are normalized to one prior to mapping into tangent space. This representation was previously used in computer vision, in particular, in Ref. [2]. Denoting the tangent function by Tg, the similarity gets defined as follows:

$$d_{\text{arcs}}(C, D) = \left(\int_0^1 (Tg_C(s) - Tg_D(s) + \Theta(C, D))^2 \, ds \right)$$

$$\times \max \left\{ \frac{l(C)}{l(D)}, \frac{l(D)}{l(C)} \right\},$$

where $l(C)$ denotes the arc length of C and the integral is taken over the arc length s. The constant $\Theta(C, D)$ is chosen to minimize the integral (it accounts for different orientation of curves) and is given by

$$\Theta(C, D) = \int_0^1 (Tg_C(s) - Tg_D(s))^2 ds.$$
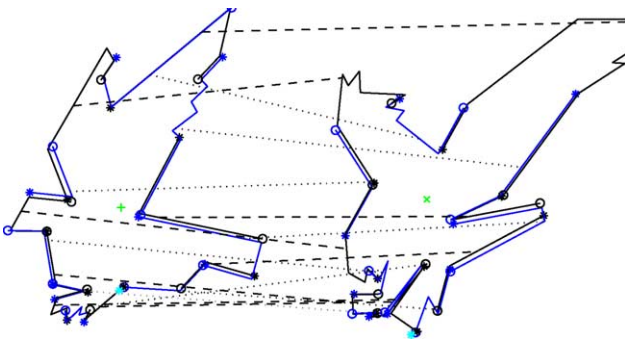

Fig. 10. Our global shape similarity measure is able to compute an intuitive correspondence of visual parts.

7. General case of the optimal shape similarity

The aforementioned experiments and definitions did not introduce any penalty for removing vertices of T in our partial similarity measure ps. An introduction of penalty makes sense when some estimation of the noise model is known as we will illustrate in Section 8. The penalty can be easily added to obtain an *optimal partial similarity measure with penalty* as linear combination of the similarity between query Q and simplified T and the similarity between simplified T and T:

$$opsp(T, P) = \min\{\alpha\, s(Q, T') + \beta\, s(T', T) : T' \subseteq T\},$$

where $\alpha + \beta = 1$ are weights. The penalty $s(T', T)$ measures how much T must be simplified in order to become more similar to query Q. The weights depend on importance of both terms and are application dependent. As it was the case for computation of optimal partial shape similarity, we can use a suboptimal method to compute the optimal partial similarity measure with penalty. We will call the obtained measure *partial similarity measure with penalty* and denote it as psp. The same applies to the extensions defined in the reminder of this section.

Our definition of ops in Section 4 assumes that query part T does not contain any significant distortions. This assumption is true for many applications, in particular when Q belongs to a set of learned visual parts, which are used for the shape-based image retrieval. However, there exist applications in which both Q and T should be simplified (e.g. when both T and P may be corrupted by noise). Therefore, it makes sense to define an *optimal partial symmetric similarity measure with penalty*

$$opssp(Q, T) = \min\{\alpha\, s(Q', T') + \beta\, s(T', T) + \gamma\, s(Q', Q)$$

$$: T' \subseteq T \wedge Q' \subseteq Q\}$$

where $\alpha + \beta + \gamma = 1$ are weights.

The proposed $opssp$ constitutes out of two parts: a global shape similarity measure of simplified shapes and a simplification measure modeling the noise plausibility. Computation of optimal similarity is an iterative process. An initial similarity of contours is computed. As long as a simplification on either one of the contours involved exists such that it reduces the dissimilarity added to the overall sum of simplifications carried out, the simplification is performed. The resulting similarity plus the overall sum of simplifications yields the resulting similarity. The reason for this process being symmetrical is due to the fact that only vertices may be removed and not introduced. So, if a shape's feature is missing in one contour due to noise, the corresponding feature may be removed from the original contour in order to yield a better similarity. The challenging task here is appropriate selection of weights.

In the above definition of $opssp$ we use our global shape similarity measure in both parts. When the extent of noise can be estimated, we may use a different measure for the

second part (that computes the cost of simplification). The estimated noise level allows for deriving a measure judging the likelihood that an individual vertex was caused by noise. Thus, simplifications of shape by removal of vertices may be measured according to noise plausibility. A simplification is performed whenever the noise plausibility is lower than a gain in shape similarity. Therefore, when comparing two shapes, exactly the differing shape features caused by noise can be removed. This leads us to the following definition of an *optimal partial symmetric similarity measure with simplification measure* that differs from *opssp* in that we use a relevance measure r to measure the amount of simplification of two shapes in the second part:

$$opsss(Q, T) = \min\{\alpha\, s(Q', T') + \beta\, r(T', T) + \gamma\, r(Q', Q)$$

$$: T' \subseteq T \wedge Q' \subseteq Q\}$$

where $\alpha + \beta + \gamma = 1$ are weights.

Expressing a simplification measure r that accounts for noise in the context of the target shape T allows us to differentiate between simplifications that just cancel out noise and those that would remove shape features. For example, if we remove a given vertex that was displaced by noise, the penalty r for its removal will be compensated in the decreased value of the similarity measure s. In the next section we show experimental results that demonstrate the usefulness of the *opsss* measure.
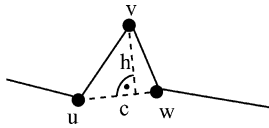
8. Experimental results for optimal partial symmetric similarity measures

Within this section we present results from two different domains, namely shape retrieval and scan matching as used in robot mapping. In both domains the application of the optimal symmetric similarity measure yields improvements as compared to classical shape similarity measures.

8.1. Recognizing digitally scaled shapes

Suppose, shapes (contours) extracted from images at varying sizes should be recognized The relative amount of noise, e.g. caused by segmentation, varies with the image's size. Single pixels may be negligible in large images, while providing valuable shape information in tiny images. The key idea is based on the observation that often the noise level can be estimated. Let us assume that contours can be extracted with a noise of ± 1 pixel. This is the case, for instance, when scaling bitmap images digitally. We introduce a term resolution r', which is said to be the distance a vertex may be translated such that the translation is still plausible. One possible simplification measure judging the likelihood for a contour vertex to be caused by noise is based on area.

Therefore the triangular area defined by three consecutive points in shape contour is computed.

$$\tilde{r}(u, v, w) = \tfrac{1}{2}\left(\tfrac{h}{r'} \cdot c\right)^2$$

The cost function $\tilde{r}$ induces the simplification measure r used in *opsss*. To prove the performance of *opsss*, a test on recognizing digitally scaled contours is performed. Recognition is possible when the similarity between the sealed query Q and the original contour C_1 is higher than between Q and another contour C_2. Therefore, we define a relative similarity.

$$S'(Q, C_1) = 100 \frac{\text{Sim}(Q, C_1)}{S(Q, C_1) + S(Q, C_2)} [\%]$$

Values above 50% represent a successful recognition (the original contour is the most similar one), values below 50% represent a confusion of contours.

In a test, we selected two shapes from the MPEG-7 dataset. Then, query shapes are generated by reducing the original shapes' resolution. The queries are within a range of resolution from 10×10 pixel up to 100×100 pixel. Whereas a query corresponding to the highest resolution shows basically no distortion, the query obtained from a 10×10 resolution is heavily distorted. Two original contours C_1 and C_2 together with their example images with reduced resolution are depicted in Fig. 11. For each query (of reduced resolution) the relative similarity to its original contour is computed. The computation of shape similarity is performed using the global shape similarity measure s, marked with black columns in Fig. 11, and the novel *opsss* measure, marked with gray columns. Examining the retrieval rates shown in Fig. 11, it can be observed that *opsss* outperforms the classical shape similarity. *opsss* is able to recognize scaled contours of both queries correctly, whereas the classical similarity is not able to recognize shapes at a resolution of less than 80×80 pixel (cp. Fig. 11(d)). This example shows that *opsss* can significantly improve shape retrieval in contexts where the noise level can be estimated. We will discuss another example in the context of robot mapping.

8.2. Matching shapes extracted from laser range finder data

In the field of robot mapping, mobile robots (typically equipped with a laser range finder as sensor) must construct an internal spatial representation, a map of the environment, from sensory data. A key technique here is to match two sensor readings against each other. Range information obtained from a laser range finder can be interpreted as a description of the robot's surroundings by means of polylines that represent obstacle boundaries. Therefore, the matching may be formulated as a shape matching (cp. [19]).
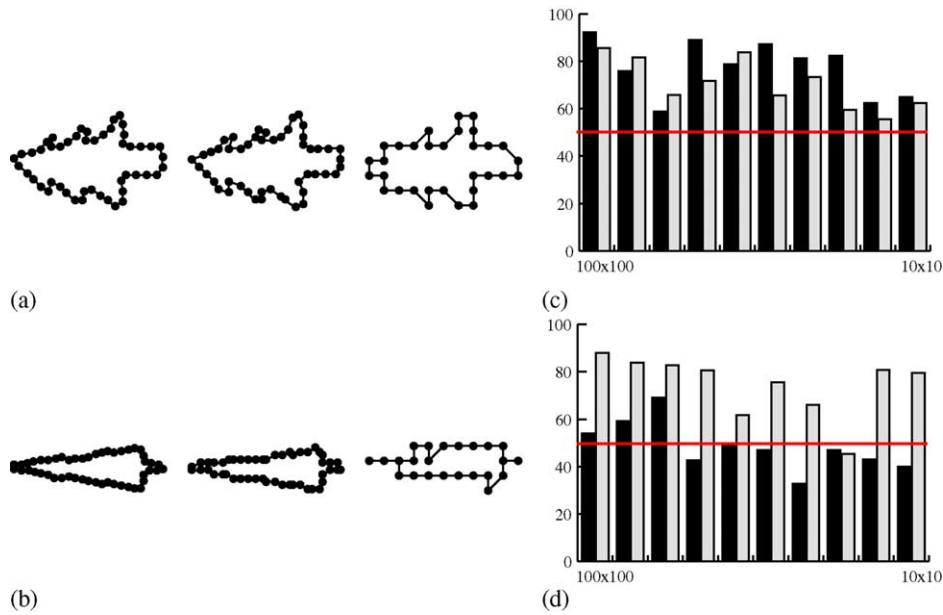
Fig. 11. (a),(b): Two shapes from the MPEG-7 database as extracted from their images with resolutions of 100×100, 40×40, and 10×10 pixel each. (c) Relative similarity of matching the shape depicted in (a) against the shapes in (a) and (b). Each column represents the relative similarity for a query at the resolutions 100×100, 90×90,…, down to 10×10 pixel. Black columns denote the results of global similarity measure *s*, gray columns the results of *opsss*. The higher the columns, the more reliable the retrieval. (d) The same as (c), but for retrieval of shape (b).

Polylines extracted from laser range finder data suffer from sensor noise that cannot be removed easily. This is due to the relative size of the noise which changes with the amount of shape information present. To be more precise, the typical noise of the laser range finder used in our experiments is about ±2 cm, which is just about the size of a door frame in a wall or table legs. Thus, removing noise up to the magnitude of ±2 cm would also remove shape features. However, loosing any shape feature cannot be afforded, as shapes perceived in typical office environments are poor salient features. To overcome this problem we apply *opsss* in a similar way as explained above. Expressing a simplification measure that accounts for noise allows to differentiate between simplifications that just cancel out noise and those that would remove shape features. For example, if we remove a given vertex that was displaced by noise, the penalty *r* for its removal will be compensated in the decreased value of the similarity measure *s*. As the cost for removing a vertex likely to be introduced by noise is very low, shape similarity can still be detected reliably, even when the shapes involved are—relative to the overall shape information—strongly distorted. An example is presented in Fig. 12, where two exemplary polylines are shown. Computing the similarity of these polylines by means of standard shape similarity yields a similarity value of more than 12, which means for this application that they are dissimilar. Since while computing the *opsss* measure we remove vertices that make them dissimilar, the computed similarity value is just about 1 which means for this application that they are similar.
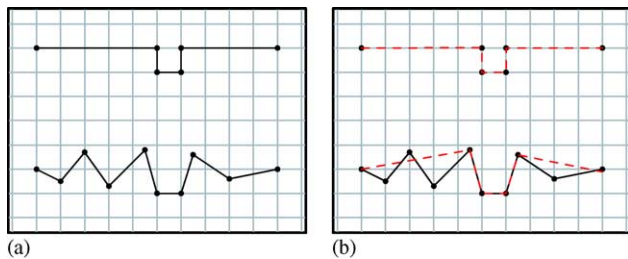
9. Conclusions

Global similarity measures fail to cope with partial visibility (due to occlusion or point of view change) and not uniformly distributed noise, which is actually a normal situation for shapes extracted from digital images. Therefore, we introduce a partial similarity measure that is capable of finding the most similar part in a target object to a given query shape, even if the query shape is only a small part of the target object. The partial shape similarity measure is optimal in the sense that it is able to focus on similar shape features and to abstract from distinct features in an optimal way. We are motivated by the fact that human

Fig. 12. (a) Two exemplary polylines as could be obtained when sensing an object with a laser range finder. As the upper polyline is free of noise, tire lower one suffers from distortions in the same magnitude as the shape features present. The grid shown denotes 1 cm distances. (b) Applying the *opsss* to compare the two polylines, differing shape features are removed prior to computing shape similarity (dashed lines). The cost for removal is low, as the removed vertices are judged likely to be caused by noise.

perception of shape is based on similarity of common parts to the extent that a single, significant visual part is sufficient to recognize the whole object. For example, if you see a hand in front of a door, you expect there to be a human behind the door. For a given query part Q, the proposed partial shape similarity measure allows us not only to retrieve a target object T from a database of shapes but also to identify part P of T that is most similar to Q under the following conditions:

1. The location of P in T that corresponds best to query part Q is unknown
2. Part P is a distorted version of Q
3. Part P may be at a different scale than Q

Our experimental results verify that the introduced partial shape similarity yields excellent results under partial visibility (e.g. occlusion on the level of 80%) if visible parts are sufficiently distinctive to identify given objects. Moreover, in contexts where the amount of distortion can be estimated, the proposed measure is able to account for noise plausibility.

Acknowledgements

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, J.S.B. Mitchell, An efficiently computable metric for comparing polygonal shapes, IEEE Trans. PAMI 13 (1991) 206–209.

[3] R. Basri, L. Costa, D. Geiger, D. Jacobs, Determining the similarity of deformable shapes, Vis. Res. 38 (1998) 2365–2385.

[4] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, pp. 509 Center number 16 1811 705-522 IEEE PAMI 24 2002; pp. 509–522.

[5] L. Biderman, Human image understanding: recent research and a theory, Comput. Vis. Graph. Image Process. 32 (1985) 29–73.

[6] T. Binford, Visual perception by computer, IEEE Conf. Systems Control 1971.

[7] R. Brooks, Symbolic reasoning among 3D models and 2D images, Artif. Intell. 17 (1981) 285–348.

[8] M.C. Burl, P. Perona, Recognition of plana: object classes, CVPR 1996.

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

[10] C.M. Cyr, B.B. Kimia, 3D object recognition using shape similarity-based aspect graph, ICCV 2001.

[11] C. Grigorescu, N. Petkov, Distance sets for shape filters and shape recognition, IEEE Trans. Image Process. 12 (9) (2003).

[12] D.D. Hoffman, W.A. Richards, Parts of recognition, Cognition 18 (1984) 65–96.

[13] D.D. Hoffman, M. Singh, Salience of visual parts, Cognition 63 (1997) 29–78.

[14] R. Lakaemper, L.J. Latecki. Shape search engine. http://knight.cis.temple.edu/shape/

[15] L.J. Latecki, R. Lakaemper, Convexity rule for shape decomposition based on discrete contour evolution, Comput. Vis. Image Understanding (CVIU) 73 (1999) 441–454.

[16] L.J. Latecki, R. Lakaemper, Shape similarity measure based on correspondence of visual parts, PAMI 22 (2000) 1185–1190.

[17] L.J. Latecki, R. Lakaemper, U. Eckhardt, Shape descriptors for non-rigid shapes with a single closed contour, CVPR 2000; 424–429.

[18] L.J. Latecki, A. Gross, R. Melter (Eds.), Special issue on shape representation and similarity for image databases Patt. Recogn. 35(1) (2002).

[19] L.J. Latecki, R. Lakaemper, D. Wolter, Shape similarity and visual parts, DGCI, Naples, Italy 2003.

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

[21] F. Mokhtarian, M. Bober, Curvature Scale Space Representation: Theory, Applications and MPEG-7 Standardization, Kluwer, Dordrecht, 2003.

[22] K. Mueller, J.-R. Ohm, J. Cooper, M. Bober, Results of 2d/3d shape core experiments ms-4. In ISO/IEC/JTC1/SC29/WG11. MPEG/M6190, July 2000.

[23] A. Pentland, Recognition by parts, ICCV 1987; 612–620.

[24] T.B. Sebastian, P. Klien, B.B. Kimia, On aligning curves, PAMI 25 (2003) 116–125.

[25] K. Siddiqi, B.B. Kimia, Parts of visual form: computational aspects, PAMI 17 (1995) 239–251.

[26] K. Siddiqi, K.J. Tresness, B.B. Kimia, Parts of visual form: ecological and psychophysical aspects, Perception 25 (1996) 399–424.

[27] M. Weber, M. Welling, P. Perona, Unsupervised learning of models for recognition, ECCV 2000; 18–32.

[28] M. Weber, M. Welling, P. Perona, Towards automatic discovery of object categories, CVPR 2000; 101–109.