Approaching Shape Matching with the Local/Global Space

Affiliation and mailing address:
Christoph Rasche
Laboratorul de Analiza si Prelucrarea Imaginilor
Universitatea Politehnica din Bucuresti
Bucuresti 061071, RO
email: rasche15@gmail.com

Abstract

A shape matching approach is introduced, which is based on a novel curve description, namely a (local/global) amplitude space. Two matching principles are tested with this description. First, a point-based (correspondence) matching is carried out with the entire amplitude space, for which the MPG7 retrieval score is 78.74%. Second, a segment-based matching with abstracted boundary segments is introduced, with the goal to move away from the typical constraints of point-based matching. Those segments are obtained by analyzing the local/global space. The retrieval score for this type of matching is 70.48% and although it is lower than the former, it can be applied to gray-scale images. When the two matching metrics are combined, a retrieval score of 84.80% is obtained, which is near top-performing, reported methods. Using an optimization method for the distance matrix, the score can be driven up to 95.01% (2nd best reported so far). The particular advantage of the presented approach is that it allows part interpretation (irrespective of the matching type).

Keywords: arbitrary shape, retrieval, classification, radial description, MPG7

1 Introduction

Shape matching systems for arbitrary shape retrieval use either point-based and/or segment-based matching methods. By point-based is meant the exhaustive point-to-point correspondence matching. Point-based matching is the prevailing matching principle and can be exploited for either aligning and/or for discriminating between two shapes, e.g. (Belongie et al., 2002; Sebastian et al., 2003; Grigorescu and Petkov, 2003; Mokhtarian and Bober, 2003; Jalba et al., 2006; Adamek and O'Connor, 2004; Ling and Jacobs, 2007; Daliri and Torre, 2008; Xie et al., 2008; Xu et al., 2009; Gopalan et al., 2010; Raftopoulos and Kollias, 2011). Although powerful descriptions exist meanwhile for this type of matching, there are a number of short-comings associated with it:

1) Lack or limited robustness to fragmentation: the systems often require that the shape is a closed curve and they can therefore not be easily applied to gray-scale images, where shapes appear mostly fragmented due to 'noise'.

2) Lack of part interpretability: the systems often cannot identify shape parts and their relations, which however would be essential for manipulating the shapes (e.g. a robot trying to understand shapes for interacting with them).

3) Long matching duration: point-to-point measurements are inherently time-consuming; their algorithmic complexity is square $O(N^2)$, and if a closed curve is tested, it is of cubic complexity $O(N^3)$ (N=number points).

There are attempts to alleviate those problems. For instance, Ghosh and Petkov (Ghosh and Petkov, 2005) have pleaded for testing shape recognition systems with fragmented shapes and presented the incomplete-contour representation (ICR) test; their own solution appears to be robust, but remains a point-based recognition system. Schmidt et al. (Schmidt et al., 2007) have developed an optimized point-based matching procedure, specifically a faster minimization method to find the proper global alignment of two shapes . Or one can use feature points (landmarks) to reduce the number of alignment matchings. Still, those attempts cannot truly shake off the above limitations.

Segment-based matching systems extract boundary segments, which are individually matched amongst two shapes (or a shape and a representation in case of classification). This is in principle a step toward escaping the shortcomings of point-based matching, but many segment-based matching systems remain footed in the former. For instance, Latecki and Lakamper introduced a curve evolution which provides a good degree of part interpretability; their segment similarity measure is based on relating curvepoints (Latecki and Lakamper, 2000). Felzenszwalb and Schwartz's system is robust to fragmentation and was shown to work on contours obtained from gray-scale images (Felzenszwalb and Schwartz, 2007); their curve comparison also remains a point-to-point matching. Daliri and Torre's shape matching system uses point-to-point matching only for aligning two shapes, after which then the shape is divided into segments of equal lengths, which in turn are transformed into a symbolic representation by classifying the segments into 32 types made of four discretized radii and eight angles (Daliri and Torre, 2008). McNeill and Vijayakumar's system intentionally refrains from feature extraction and thus is not suitable for part interpretation (McNeill and Vijayakumar, 2006).

Some segment-based systems are free from any point matching. The recognition system by Attalla and Siy uses a polygonal description, specifically arcs of equal length (Attalla and Siy, 2005). For each arc a small number of attributes is determined: the arc's chord length, its degree of curvature and its distance from the shape center. The latter attribute represents a radial description of the shape. The system however also relies on the shape being closed. And because segments are of equal length, a part interpretation is not possible and the shape description lies in some sense between point- and segment-based matching. The complexity of the system is $O(N)$ only and thus clearly less complex than any point matching system; it is possibly the most efficient shape recognition system with respect to the speed/accuracy tradeoff.

Some of the mentioned systems are summarized in table 1. The noted complexity is only a principal complexity and sometimes estimated by us, because not all studies report an explicit algorithmic estimation. A binary robustness 'rating' is given, whereby a plus sign indicates when a system has been shown to perform also on shapes in contour images. (Explanations for the part rating are given in section 3). Clearly, all the systems have their advantages and disadvantages.

Distance Optimization. Recently, efforts went into postprocessing (optimizing) the distance matrix (that is generated for pairwise retrieval). The postprocessing is a type of unsupervised learning, that seeks to decrease the distances between similar shapes - and thus within-class distances -, and to increase the distance between dissimilar classes - and thus extra-class distances. This learning comes in different forms, e.g. (Yang et al., 2009; Kontschieder et al., 2010; Bai et al., 2010; Ling et al., 2010). We tested two optimization algorithms. One is the page-rank related algorithm by Bai et al. (Bai et al., 2010), which improved Ling and Jacobs' popular inner-distance description (Ling and Jacobs, 2007) by 6.21 percent. Bai et al.'s method is simple to implement as it essentially consists of a single recursive equation, and it appears to be applicable to different tasks (Bai et al., 2010). The other optimization algorithm is based on a modified mutual graph method, by Kontschieder et al. (Kontschieder et al., 2010). Their method is more complex (and thus we employed their software package[1]), but improved the inner-distance description by 8 percent. It appears more efficient for the task of distance matrix optimization and reports also much shorter optimization duration. We later distinguish between *geometric* and *optimized* distance matrix or retrieval performance, with the former as obtained without any learning (optimization), and the latter with learning (see also table 1 for comparison of scores).

Presented Approach The shape description used here is based on a (multi-resolution) local/global analysis (Rasche, 2010), in which no modification of the contour occurs - as opposed to the curvature-scale space, which is generated by lowpass filtering the contour and hence creating a fine/coarse scale (Mokhtarian and Bober, 2003). Contrasting the two methods in a nutshell: in the curvature-scale space the tangential curvature for each curvepoint at different fine/coarse scales is determined, whereas in the local/global analysis the amplitude for a neighborhood (a window centered on each curvepoint) is determined, and that for different window sizes (local/global scale). Because the contour is not modified in this analysis, the description is very rich. Some more explanations about this analysis are given later, but a full treatment is impossible here. One type of matching presented in this study, is based solely on this space (section 2) and is thus a typical point-based matching. Another type of matching is based on abstracted segments, which were obtained from the space after partitioning and geometric parameterization of the segments (section 3). Segment-based matching is free of the above mentioned constraints (of the point-based systems): it does not rely on the shape being closed, the matching duration is shorter and it allows part interpretation. For both matching types, space and segment matching, a retrieval task on the MPG7 collection is carried out, including an optimization of the distance matrix. Finally, we also combine the two methods, which gives us the best retrieval results on the MPG7 collection (section 4).

2 Space Matching

The type of point-based matching that we carried out, is most similar to the matching system by Adamek and O'Connor and also analogous to curvature-scale space matching (Adamek and O'Connor, 2004; Mokhtarian and Bober, 2003). Adamek and O'Connor create a fine/coarse scale

[1] `http://vh.icg.tugraz.at/index.php?content=topics/beyondshape.php`

T	Perf.	Complexity	Unit	Principal	Parts	R	Authors
Optimized	95.96%	$O(N^3)$	pt	ID,Aspect	+		Ling etal,10
	95.01%	$O(N^3)$	pt/seg	LGS/seg vec.			**combined [proposed]**
	93.40%	$O(N^3)$	pt	ID,mod/mut Gph	+		Kontschieder etal,09
	93.32%	$O(N^3)$	pt	ID,LCDP,uns GP	+		Yang etal,09
	91.61%	$O(N^3)$	pt	ID			Bai etal,10
	91.03%	$O(N^3)$	pt	LGS	+++		**space matching**
	85.08%	$O(N)$	seg	segment vectors	+++	+	**segment matching**
Geometric Only	89.31%	$O(N^3)$	pt	contour flexib.			Xu etal,09
	87.70%	$O(N^3)$	pt/seg	hier. multi-res.		+	Felzenszwalb&Schwartz,07
	85.40%	$O(N^3)$	pt	inner dist. (ID)	+		Ling&Jacobs,07
	84.93%	$O(N^3)$	pt	MCC			Adamek&OConnor,04
	84.80%	$O(N^3)$	pt/seg	LGS/seg vec.			**combined [proposed]**
	84.33%	$O(N)$	seg	equal-length segs			Attalla&Siy,05
	81.12%	$O(N^3)$	pt	CSS	++		Mokhtarian&Bober,03
	78.74%	$O(N^3)$	pt	LGS	+++		**space matching**
	78.38%	$O(N^3)$	pt	distance sets		+	Grigorescu&Petkov,03
	76.45%	$O(N)/O(N^2)$	seg/pt	parts	+		Latecki&Lakamper,00
	70.48%	$O(N)$	seg	segment vectors	+++	+	**segment matching**

Table 1: Retrieval systems, their Bull's Eye score (for the MPG7 collection) and their characteristics. **T**ype: Geometric only (no optimization); Optimized: distance matrix tuned by learning algorithm. **Unit**: pt=point, seg=segment. **Principal**: segs=segments, LGS=local/global space (this study), MCC: multi-scale convexity/concavity, CSS=curvature scale space. ID=inner distance (incl. shape context); see text for more abstractions. **Parts**: part interpretability. **R**obustness: tested on gray-scale images (gray-scale applicability).

space from which they derive convexity/concavity (later collectively called bulginess) by determining the spatial relation between low-pass filtered curves. Shape similarity is determined by matching the signed inter-curve distance spaces and by exploiting dynamic programming to find the minimum distance. The method presented here matches the amplitude scale space. To avoid cubic matching complexity we reduce the exhaustive alignment search by using the segments' midpoints as keypoints (feature points). Those segments can be obtained from the partitioning process (mentioned later), but are not per se used in the following space matching method.

2.1 Implementation

For a shape contour $\mathfrak{c}(v)$, with arc length variable v, 100 equally spaced points were selected. For a given curvepoint, a window (local neighborhood) of length ω along the arclength variable v is selected and the amplitude of the selected segment determined. The window is shifted through the curve, and the amplitude is taken for different sizes, hence creating an amplitude-scale space $\mathbf{B}(\omega, v)$. The amplitude is taken only for windows, in which the segment appears as an arc, meaning all segment points need to lie on either side of the segment's chord; otherwise the amplitude is set to 0. 10 window sizes were selected, ranging from 5 to 79 pixels with an increment of ca. $\sqrt{2}$ (see also figures 1 and 2 in supplementary material).

For matching, the space was regarded as a 1000-dimensional vector; the difference between two shapes i and j was implemented with the Manhattan distance: $d^{\mathbf{B}}_{ij} = \sum_{\omega,v} |\mathbf{B}^i - \mathbf{B}^j|$. A partitioned shape offered in average 20 keypoints (that is segments, see also section 3; this partitioning comes at little temporal cost in comparison to the generation of the entire amplitude space.) Due to various types of variability (structural, aliasing), the keypoints are not at exactly the same locations (for shape instances of one class), causing a decrease in performance in comparison to exhaustive point-based matching. To compensate for this variability, the neighboring points around the keypoints were used for alignment as well: by dilating 2 pixels, the average number of keypoints increased to 30 (some segments overlap due to extraction at different scales), which returned a reasonable approximation to an (estimated) full matching. The algorithmic complexity is thus less than cubic, $O(< N^3)$. The resulting distance matrix between two shapes is of size $m \times n$, with m and n being the number of keypoints for each shape. A dynamic programming approach did not make sense and we simply took the minimum value of the entire distance matrix. Matching duration with 30 keypoints (along v) was ca. 80 milliseconds for a pair of shapes on a 2.66 GHz Intel Pentium.

2.2 Evaluation

For each shape, the remaining 1399 shapes were ordered according to their increasing distance. Retrieval accuracy was measured with the Bull's eye score, which counts the number of same-class instances within the first 40 most similar shapes (including self-similarity). This count is determined for all 1400 shape retrievals, summed and divided by the maximal possible count, namely 28000 (1400*20; (Latecki et al., 2000)).

The geometric Bull's eye score is 78.74 percent, see 'space matching' in table 1. Increasing the resolution to 200 points for instance, only marginally increased the performance by ca. 1-2 percent.

The optimized score with the modified-mutual graph method is 91.03 percent (a gain of 12.29), with kNN=12 for local scaling normalization, kNN=6 and c=3.5 for the graph. The class-individual gain can be up to 50 percent; it is slightly negative for classes and larger for another (figure 1, upper right).

The optimized score with the page-ranking related algorithm is 88.36 percent (a gain of 9.62), with kNN=15 and α=0.4 for local scaling normalization. As results with this method were consistently lower by a few percent than the modified-mutual graph method - as can be assumed from the comparison of the two methods on the inner-distance method - they are no longer reported here and only results with the modified-mutual graph method are given from now on.

2.3 Discussion

The geometric retrieval accuracy lies in the range of other respectable shape recognition systems, but lies significantly under Adamek and O'Connor's system performance (84.93%), probably because their system elegantly exploits concavity and convexity information (Adamek and O'Connor, 2004). We also attempted to include such information into the amplitude space, by for instance using signed values in $\mathbf{B}$, but that reduced performance significantly.

The increase in resolution (to gain higher performance) is not worth the square increase in matching duration. The observation that 100 points provide a sufficient recognition performance was also observed in Attalla and Siy's study (Attalla and Siy, 2005).

3 Segment Matching

In this study we particularly explore a radial description of segments. A radial formulation is powerful because it provides a high degree of coherence - akin to a template - despite its relatively generic form of 'structural description'. The radial formulation is probably the main reason why Attalla and Siy's system performs so well in comparison to point-based matching systems.

The partitioning and abstraction process is relatively complex and a detailed description is provided somewhere else (Rasche, 2010); here it is only reviewed (subsection 3.1). The abstracted segments are then related to each other using a simple radial description (subsection 3.2). How the lists of segments are matched is described in subsection 3.3 (including an estimation of the algorithmic complexity), followed by an implementation and an evaluation subsection.

3.1 Partitioning, Segment Abstraction

The amplitude space is searched for consistent arcs, that is arcs that appear also on adjacent scales (for more local or global window sizes). This process reliably detects arcs corresponding to human interpretation; it does not correspond to high-curvature detection, but to detection of segments between high curvatures. It was shown that this partitioning is relatively homologous across class instances in the MPG7 collection (under review; but see also figure 2 for some classes that are difficult to partition). It is therefore rated as highest in table 1, with three plus symbols, followed by the curvature scale space (two plus symbols), followed by Latecki and Lakampers polygonal decomposition and the inner distance analysis (one plus each). The partitioned segments are often non-disjoint; they can be overlapping or nested, with the latter being the result of an extraction at different (local/global) scales.

Each segment is described by the following geometric attributes: orientation (o), the total arc length ($l = l_c$), degree of curvature (b) (amplitude of arc segment), angularity (e) (1st derivative), (arc) flatness (f), irregularity (i), circularity (ζ) and transition (t). Those parameters form an 8-dimensional vector: $\mathbf{a}(o, l, b, e, f, i, \zeta, t)$. In the following subsection, three more dimensions will be added, which express the a

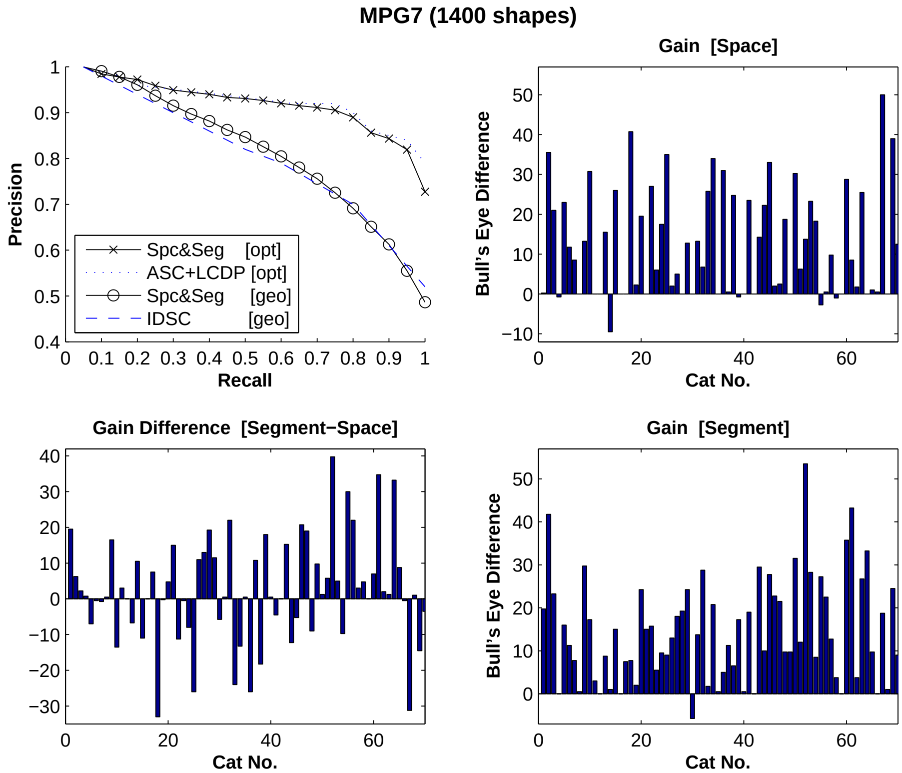

Figure 1: Optimization Analysis. Upper left: Precision/Recall curves for geometric and optimized combination of space and segment matching; IDSC: inner-distance (& shape context) by (Ling&Jacobs,2007); ASC+LCDP by (Ling et al., 2010). Class-individual gain for space matching (lower left) and the difference in gain between space and segment matching (lower right).

segment's radial relation to a pole. The average size of the representation consists thus ca. 220 real values (20 segments times 11 dimension values).

That the segments are sometimes nested, makes it challenging to exploit the dynamic programming procedure for lowest-cost path detection. Nesting reduces the degree of order and the lowest-cost path appears rather as a 'sparsely marked trail' at best. The search for such a path was therefore not attempted.

3.2 Relating Segments

A radial description requires the choice of a pole (center). The simplest way to obtain a pole is to take the mean of all shape pixel coordinates. For each segment the distance (radius) and the angular 'orientation' to the pole is determined using the segments' midpoints. For orientation we separate between curved segments (arcs) and straight segments. For curved segments we determine the degree of bulginess, which represents convexity and concavity. For straight segments, we determine a tangentiality value.

• **Radius** (r): is the distance between pole and segment's midpoint normalized by the image size.

• **Tangentiality** (τ): is the (smaller) angle of the intersection of the straight segment and the radial line connecting the shape center ($\mathsf{p_c}$) and the segment's midpoint (gray dashed in figure 3a). Maximal tangentiality is therefore a right angle (segments no. 1 and 3 in figure 3); minimal tangentiality (τ=0) occurs if the segment's chord aligns (parallel) with the radial line (segment no. 2). A segment is considered straight if its curvature value is below a tolerance T_b^τ: $\exists\ \tau_i \in [0, \frac{\pi}{2}]\ \mathbf{a}_i(b) < T_b^\tau$.

• **Bulginess** (β): is the angular difference between two direction angles (both $\in [0, 2\pi]$). One directional angle is the face angle ϕ obtained from the ray pointing from a segment's midpoint to its own halfpoint (the midpoint of a segment's chord) - the dotted arrow pointing 'north' from a segment's midpoint in figure 3b. The other directional angle is the ray pointing from the pole to the segment's midpoint (dashed in figure). The angular difference lies in the interval $[0, \pi]$; convex segments have a bulginess value in the

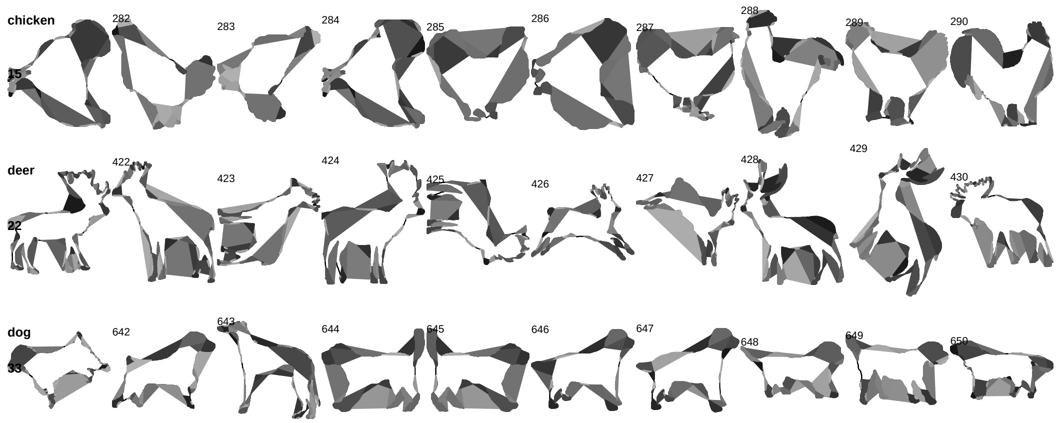

Figure 2: Partitioned arcs segments in MPG7 shapes for three classes. A shaded region outlines an arc; the darkness of the shading corresponds to a consistency measure (dark=consistent).

interval $[0, \frac{\pi}{2})$, concave segments a value in the interval $(\frac{\pi}{2}, \pi]$, a value $\beta = \frac{\pi}{2}$ means neither convex nor or concave (segment 2 in figure 3b). Maximal convexity and concavity are the interval's endpoints (0 and π, respectively, segments 1 and 3 in figure 3b). The bulginess angle is determined if the segment was detected as an arc ($\Phi^{\beta}(l_{\mathfrak{c}}) > 0$): $\exists\, \beta_i \in [0, \pi]\ \Phi^{\beta}(l_{\mathfrak{c}}) > 0$.

Due to the presence of structural variability caused by noise (e.g. deformation) and the ever-present ambiguity of arc/straight distinction, the two radial angles tangentiality and bulginess are taken from overlapping intervals of the curvature value range (hence the two different tolerances).

Adding the above three dimensions to the previously mentioned vector, the vector is now: $\mathbf{a}(o, l, b, e, f, i, \zeta, t, r, \tau, \beta)$.

3.3 List Matching, Complexity

List Matching To match the list vectors of two shapes, the segment vectors are pairwise matched to form a similarity matrix from which the segments' best matches are selected and integrated. The similarity measure is called *congruence* measure in this study, as no scale (size) independence is included (the MPG7 collection shows little if any intra-class size variations).

Given two lists of segments, $\mathbf{a}_i$ and $\mathbf{a}_j$ from shapes A and B, the pair-wise (metric) similarities $\mathrm{sim}(\mathbf{a}, \mathbf{a}')$ of the individual vectors are taken, resulting in a $[n \times m]$ similarity matrix, $\mathbf{S} = [\mathrm{sim}_{ij}]$, with n and m being the list lengths. The similarity measure sim can be further refined by a weight vector $\mathbf{v}(i)$, whose components correspond to the significance of the attributes. Next, the maximum with respect to each shape is taken, returning a congruence vector $\mathbf{g}$,

$$\mathbf{g}_A(i) = \max_j \mathbf{S}[i, j], \tag{1}$$

$$\mathbf{g}_B(j) = \max_i \mathbf{S}[i, j]. \tag{2}$$

A weighted sum of the individual components is taken, with the weights $\mathbf{w}$ corresponding to the segments' significances and acting as a normalizer. The final congruence value is a multiplication of the weighted sum of both shapes:

$$\mathrm{cong} = \mathbf{g}'_A \mathbf{w}_A \times \mathbf{g}'_B \mathbf{w}_B. \tag{3}$$

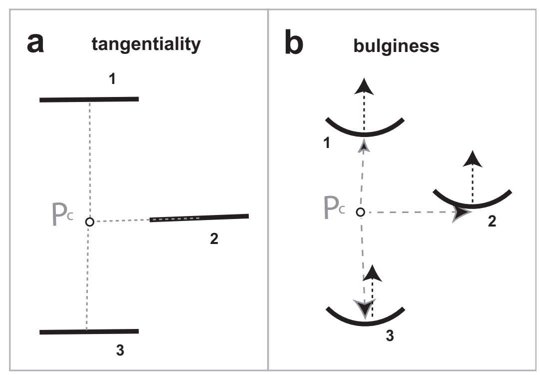

Figure 3: Two angles for a radial description. **a.** Tangentiality angle for straight segments: segments 1 and 3 have the same tangentiality and are maximal (π); segment 2 shows minimal tangentiality. **b.** Bulginess (directional) angle for curved (bent) segments; segment 1 is maximally concave; segment 2 shows no bulginess - it is neither convex nor concave; segment 3 is maximally convex.

The complete congruence measure is summarized as $\text{cong}_{kl}(\mathbf{a}_i, \mathbf{a}_j, \mathbf{v}, \mathbf{w}_A, \mathbf{w}_B)$, with $\mathbf{v}$ as the attribute weight vector with dimensionality equal the number of attributes (parameters), $\mathbf{w}$ as the segment weights with dimensionality corresponding to the list lengths (n and m), and k and l two shapes.

Complexity An informal complexity estimate is provided. The most time consuming part is the generation of the local/global (scale) space for partitioning and abstraction, which is analogously complex as creating the curvature scale space, thus $O(N \log N)$.

As the partitioned segments are merely a list of parameters and selected coordinates (end- and midpoints), the segment-matching duration is 2.8 milliseconds only and thus negligible as compared to the duration of generating the local/global space (next paragraph).

3.4 Implementation and Evaluation

No subsampling occurred, meaning the full size of the silhouette was employed. The curve partitioning and abstraction process takes ca. 400ms on a 2.66 GHz Intel Pentium (implemented in Matlab, whereby the amplitude signature is generated exploiting matrix operations, but the use of different windows is not). T_b^τ was set to 0.4. The similarity metric for two vectors (sim) is determined using a Gaussian radial-basis function, whose variance for the individual dimension is set to their overall variance of the entire collection (vector $\mathbf{v}$).

The geometric retrieval score is at ca. 68 percent for unitary attribute weights. A simple heuristic search was employed to tune the attribute weights $\mathbf{v}$, which increased the performance to 70.48 percent (see 'segment matching' in table 1). Exploiting distance optimization (modified mutual graph), the score rises to 85.08 percent, a gain of 14.60 percent (with kNN=20 for local scaling normalization, kNN=6 and c=2.5 for the graph). The class-individual gain in Bull's eye score can be more than 50 percent, and is negative for only one class (figure 1, lower right). To analyze more specifically the differences in class-individual gain for the two matching methods, we subtract the two sets of gain values (pairwise) and observe that there can be significant differences for some classes (plot in lower left in figure 1).

3.5 Discussion

The geometric retrieval score is rather low, but this type of matching is also the first of its kind: only abstracted (parameterized) segments are used and no particular dependence on the closeness of the shape exists. Thus, the description as such is applicable to gray-scale images; an earlier version of this approach was already successfully applied to gray-scale images, yet an actual radial description was not tested yet.

One reason for the relatively low performance may be that the segments have been aligned only radially with respect to a pole, but that may be still an over-simplified structural description. Further improvement may be obtained by grouping segments.

4 Combined Matching

The class-individual gain analysis in figure 1 shows that the two matching methods favor different classes and that encourages us to combine the matching methods. Since the processes of segment partitioning, abstraction and matching occur relatively rapid, the combination of the two methods comes at little additional temporal cost; however, it comes at moderate increase in storage size, as now in addition to the 1000-dimensional vector, also a list of vectors is kept (an average of 1220 real values per shape).

4.1 Implementation and Evaluation

The combination occurs by adding the two distance measures for each method for a pair of shapes ($d_{comb} = d_{ij}^{\mathbf{B}} + \text{cong}_{ij}$). The geometric score is 84.80 percent and is thus in the range of other top-performing systems (see 'combined' in table 1). The precision-recall curve for this retrieval is shown in the upper left of figure 1: it is practically equal to the one for the inner distance description (IDSC).

Applying the distance optimization method (modified mutual graph), the score rises to 93.71 percent. We also tested a combination after individual optimization, that is the optimized distance matrices for each method are added ($d_{comb} = d_{ij}^{\mathbf{B},opt} + cong_{ij}^{opt}$), then the score is marginally higher with 93.89 percent. Applying the optimization algorithm again to that combined matrix, we obtain 95.01 percent (gain of only 1.12 percent), which is the second best reported so far. The precision-recall curve for that latter combination is shown in figure 1 (upper left) and is practically equal with the one of the optimized aspect space description (dotted) (Ling et al., 2010).

This last combination thus consists of three optimizations in total, which however given the temporal efficiency of the modified mutual graph algorithm is again negligible. The downside of this combination is rather that it lacks the robustness to gray-scale

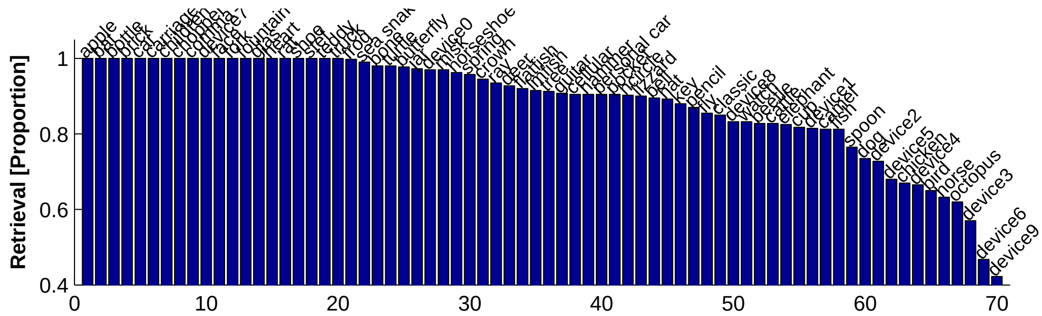

Figure 4: Ranking classes according to their proportion of correct retrieval (within the first 20 similar retrievals).

images, but its advantage of part interpretability remains.

In figure 4, the classes are ranked according to proportion of correct retrieval. Classes that show a particularly low proportion are the 'device' classes (no. 3, 4, 6 and 9), which generally are difficult to discriminate due to their intra-class variability (compare also to figure 16 in (Daliri and Torre, 2008) and figure 7 in (Adamek and O'Connor, 2004)).

We looked at the (class-wise) confusion matrix and observed that the confusions are often within the same 'super-ordinate category': e.g. large mammals are confused (elephant, camel, horse...); geometric shapes are confused (device classes); insects are confused (flies and beetles). The supplementary material shows these confusions, for both classes and individual retrievals.

4.2 Discussion

The combination of the two descriptions (space and segment) was obviously succesfull, although the improvement can not necessarily be anticipated for certain - the differences in gain in figure 1 (right column) was merely a hint. The supplementary material contains a first step toward an analysis of this improvement, by showing the wrong retrievals (confusions) for both methods (figures 5 to 8 in suppl. mat.). The analysis shows, that the confusions are sometimes the same for both methods, e.g. an apple is confused with the pocket watch, or the cup with the (sting-)ray. This indicates that in those cases the two measures may emphasize the same set of 'class characteristics'. For classes, where different confusions are present (for the two methods), the measures may act complementary, that is different sets of class characteristics contribute to a higher performance. A deeper 'combination' analysis may provide insights on how to possibly improve the benefit of combining.

5 General Discussion

The combined retrieval performance evidences the potential of the local/global analysis. Matching the entire space is relatively straightforward, but it relies on the typical constraints of a point-based (correspondence) matching. To move away from these constraints, a radial description of segments was tested which showed only mediocre geometric retrieval performance, but excels at speed (pair matching duration of 3 ms only). A strength of the local/global shape analysis is that it allows an exact part interpretation, independent of the type of matching (figure 2).

Recent studies emphasize in particular the need for invariance to articulation (part alignment variability), e.g. (Xu et al., 2009; Ling et al., 2010; Gopalan et al., 2010). In those studies, the problem was formulated as a tradeoff or balance between an increased articulation invariance and a decrease in class discriminability. In our description, this issue of balance was not particularly addressed. Instead, the presented description is a rather detailed description, in which articulation is expressed by a segment characterization - it is parameterized essentially. Thus, the potential problem that arises with such a detailed description is that it may be too detailed and lead to intra-class clusters, which deteriorates discrimination. Further analysis may give more clues about the exact nature of the representation, but presently the increase in discriminability appears to be elegantly provided by the distance optimization methods.

The distance optimization with the modified mutual graph method yielded higher gains in our work (12.29, 14.60 and ca. 10.0 percent for space, segment and combined matching) as opposed to the 8 percent increase in Kontschieder et al's study (Kontschieder et al., 2010). There are two possible reasons for this higher gain:

1) Starting with a lower geometric score - as in case of space and segment at least - allows more room for improvement, as opposed to a geometric score that is high already.

2) The local/global description bears a better separation between classes. As Ling et al. already assume, the crucial issue is that the geometric distance measure provides an overall good separation between perceptually dissimilar classes - even if the geometric score is only mediocre in comparison to some other systems (see their section 4 in (Ling et al., 2010)).

Future work will address how grouping operations can provide a more elaborate description and we assume that this is where the greatest potential of this framework lies. The development of such a structural description system is undoubtly more complex than many of the other point-based systems, however it will be more robust in its application.

References

Adamek, T. and O'Connor, N. (2004). A multiscale representation method for nonrigid shapes with a single closed contour. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(5):742–753.

Attalla, E. and Siy, P. (2005). Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching. *Pattern Recogn.*, 38:2229–2241.

Bai, X., Yang, X., Latecki, L. J., Liu, W., and Tu, Z. (2010). Learning context-sensitive shape similarity by graph transduction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32:861–874.

Belongie, S., Malik, J., and Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):509–522.

Daliri, M. R. and Torre, V. (2008). Robust symbolic representation for shape recognition and retrieval. *Pattern Recogn.*, 41:1782–1798.

Felzenszwalb, P. F. and Schwartz, J. D. (2007). Hierarchical matching of deformable shapes. *IEEE Conference on Computer Vision and Pattern Recognition, 17-22 June 2007, Minneapolis, USA*, pages 1–8.

Ghosh, A. and Petkov, N. (2005). Robustness of shape descriptors to incomplete contour representations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27:1793–1804.

Gopalan, R., Turaga, P., and Chellappa, R. (2010). Articulation-invariant representation of non-planar shapes. In *Proceedings of the 11th European conference on computer vision conference on Computer vision: Part III*, ECCV'10, pages 286–299, Berlin, Heidelberg. Springer-Verlag.

Grigorescu, C. and Petkov, N. (2003). Distance sets for shape filters and shape recognition. *Image Processing, IEEE Transactions on*, 12:1274 – 1286.

Jalba, A., Wilkinson, M., and Roerdink, J. (2006). Shape representation and recognition through morphological curvature scale space. *IEEE Transactions on Image Processing*, 15:331–341.

Kontschieder, P., Donoser, M., and Bischof, H. (2010). Beyond pairwise shape similarity analysis. In Zha, H., Taniguchi, R.-i., and Maybank, S., editors, *Computer Vision - ACCV 2009*, volume 5996 of *Lecture Notes in Computer Science*, pages 655–666. Springer Berlin / Heidelberg.

Latecki, L. and Lakamper, R. (2000). Shape similarity measure based on correspondence of visual parts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22:1185–1190.

Latecki, L., Lakamper, R., and Eckhardt, T. (2000). Shape descriptors for non-rigid shapes with a single closed contour. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 424 – 429.

Ling, H. and Jacobs, D. W. (2007). Shape classification using the inner-distance. *IEEE Trans. Pattern Anal. Mach. Intell*, 29:286–299.

Ling, H., Yang, X., and Latecki, L. J. (2010). Balancing deformability and discriminability for shape matching. In *Proceedings of the 11th European conference on computer vision conference on Computer vision: Part III*, ECCV'10, pages 411–424, Berlin, Heidelberg. Springer-Verlag.

McNeill, G. and Vijayakumar, S. (2006). Hierarchical procrustes matching for shape retrieval. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1*, pages 885–894, Washington, DC, USA. IEEE Computer Society.

Mokhtarian, F. and Bober, M. (2003). *Curvature Scale Space Representation: Theory, Applications, and MPEG-7 Standardization*. Kluwer Academic Publishers, Norwell, MA, USA.

Raftopoulos, K. A. and Kollias, S. D. (2011). The global-local transformation for noise resistant shape representation. *Comput. Vis. Image Underst.*, 115:1170–1186.

Rasche, C. (2010). An approach to the parameterization of structure for fast categorization. *International Journal of Computer Vision*, 87:337–356.

Schmidt, F., Farin, D., and Cremers, D. (2007). Fast matching of planar shapes in sub-cubic runtime. In *IEEE 11th International Conference on Computer Vision*.

Sebastian, T., Klein, P., and Kimia, B. (2003). On aligning curves. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(1):116–125.

Xie, J., Heng, P.-A., and Shah, M. (2008). Shape matching and modeling using skeletal context. *Pattern Recogn.*, 41:1773–1784.

Xu, C., Liu, J., and Tang, X. (2009). 2d shape matching by contour flexibility. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31:180–186.

Yang, X., Kknar-tezel, S., and Latecki, L. J. (2009). L.: Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval. In *In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR*.