IPSJ Transactions on Computer Vision and Applications Vol. 2 224–234 (Dec. 2010)

Research Paper

IS-Match: Partial Shape Matching by Efficiently Solving an Order Preserving Assignment Problem

Michael Donoser,[†1] Hayko Riemenschneider[†1] and Horst Bischof[†1]

This paper introduces a novel efficient partial shape matching method named *IS-Match*. We use sampled points from the silhouette as a shape representation. The sampled points can be ordered which in turn allows to formulate the matching step as an order-preserving assignment problem. We propose an angle descriptor between shape chords combining the advantages of global and local shape description. An efficient integral image based implementation of the matching step is introduced which allows detecting partial matches an order of magnitude faster than comparable methods. We further show how the proposed algorithm is used to calculate a global optimal Pareto frontier to define a partial similarity measure between shapes. Shape retrieval experiments on standard shape datasets like MPEG-7 prove that state-of-the-art results are achieved at reduced computational costs.

1. Introduction

Shape matching is a well investigated problem in computer vision and has versatile applications as e.g., in object detection [1)–3)], image retrieval [4)], object tracking [5)] or action recognition [6)]. The most important part of designing a shape matcher is the choice of the shape representation which has a significant effect on the matching step. Shapes have for example been represented by curves [7)], medial axes [8)], shock structures [9)] or sampled points [10)].

In general current shape matching algorithms can be divided into two categories: global and local approaches. Global methods compare the overall shapes of the input objects by defining a global matching cost and an optimization algorithm for finding the lowest cost match. One of the most popular methods for global shape matching is the shape context proposed by Belongie, et al. [10)]. Their algorithm uses randomly sampled points as shape representation and is based on a robust shape descriptor – the shape context – which allows to formulate the matching step as a correspondence problem. The shape context is the basis for different extensions considering geodesic distances as proposed by Ling and Jacobs [11)] or the point ordering as shown by Scott and Nowak [12)].

While such global matching methods work well on most of the standard shape retrieval datasets, they cannot handle strong articulation, part deformations or occlusions. For example, shape context is a global descriptor and local articulations influence the description of every sampled point. To reduce this effect larger histogram bins are used further away from the point. Although this reduces the problem, e.g., occlusions still lead to matching errors as it is illustrated in **Fig. 1** for the shape context based COPAP framework [12)].

These problems are handled well by purely local matching methods as e.g., proposed by Chen, et al. [13)], which accurately measure local similarity, but in contrast fail to provide a strong global description for robust shape alignment. In this work, we try to bridge the gap between the two worlds by combining their advantages.

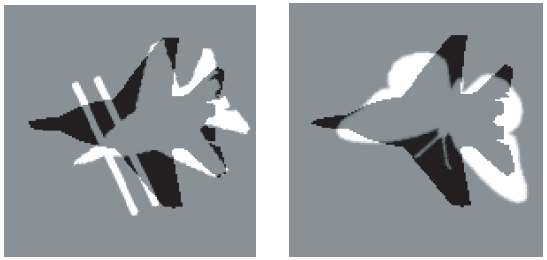
We propose a novel shape matching method denoted as *IS-Match* (Integral Shape Match). We use sampled points along the silhouette as representation and exploit the ordering of the points to formulate matching as order-preserving assignment problem. We introduce a chord angle descriptor which combines local and global information and is invariant to similarity transformations. An integral image based matching algorithm detects partial matches with low computational complexity. The method returns a set of partial matches and therefore also allows matching between differently articulated shapes.

Our main contributions are: (1) a chord angle based descriptor combining local and global information invariant to similarity transformations (2) an efficient integral image based matching scheme where matching in practice takes only a few milliseconds and (3) the calculation of a global optimal Pareto frontier for measuring partial similarity.

The outline of the paper is as follows. Section 2 describes the partial shape matching concept named *IS-Match* in detail. Section 3 presents a comprehensive evaluation of the proposed algorithm for shape retrieval experiments on common

†1 Institute for Computer Graphics and Vision, Graz University of Technology

(a) Reference 1 (b) Candidate 1 (c) Candidate 2



(d) Similarity: 44.19 (e) Similarity: 44.08

Fig. 1 Shape matching methods using global descriptors like COPAP [12] are not able to handle occlusions or strong articulations. The same similarity value for the partially occluded shape and a totally different one is returned.

datasets like MPEG-7 and an application for the task of unsupervised clustering of shapes. All evaluations prove that state-of-the-art results are achieved at reduced computational costs.

2. Partial Shape Matching: *IS-Match*

Our shape matching algorithm named *IS-Match* takes two shapes as input and returns detected partial matches and a similarity score as result. Section 2.1 describes the shape representation used, which is an ordered sequence of points sampled from the silhouette. Section 2.2 introduces a chord angle based descriptor invariant to similarity transformations. In Section 2.3 an efficient integral image based algorithm for matching the descriptor matrices to each other is outlined, which allows detecting subparts of the contours that possess high similarity with low computational complexity. Section 2.4 describes how a global optimal Pareto frontier is calculated and the corresponding Salukwadze distance is returned as measure for partial similarity. Finally, Section 2.5 analyzes the required computational complexity for single shape matches.

2.1 Shape Representation

The first step of our method is to represent the shapes by a sequence of points sampled from the contour. There are two different variants for point sampling: (a) sampling the same number of points from the contours or (b) equidistant sampling, i.e., fixing the contour length between sampled points. The type of sampling significantly influences the invariance properties of our method. Based on equidistant sampling occlusions as e.g., shown in Fig. 1 can be handled but then only shapes at the same scale are correctly matched. By sampling the same number of points our method becomes invariant to similarity transformations, but strong occlusions cannot be handled anymore. In this paper we focus on the equidistant sampling for the task of shape retrieval on single scale datasets. Nevertheless all subsequent parts of the method are defined in a manner independent of the sampling type. Therefore, we can switch the sampling type without requiring any modifications of the method. Please note that in other applications, as e.g., shape based tracking [14] or recognizing actions by matching to pose prototypes [6], a different sampling strategy might be preferred.

Because the proposed shape matching method focuses on analyzing silhouettes, as e.g., required in the areas of segmentation, detection or tracking, the sampled points can be ordered in a sequence which is necessary for the subsequent descriptor calculation and the matching step. Thus, any input shape is represented by the sequence of points $P_1 \dots P_N$, where N is the number of sampled points.

2.2 Shape Descriptor

The descriptor constitutes the basis for matching a point P_i of the reference shape to a point Q_j of the candidate shape. We formulate the subsequent matching step presented in Section 2.3 as an order-preserving assignment problem. Therefore, the descriptor should exploit the available point ordering information. In comparison, the shape context descriptor loses all the ordering information due to the histogram binning and for that reason does not consider the influence of the local neighborhood on single point matches.

We propose a descriptor inspired by chord distributions. A chord is a line joining two points of a region boundary, and the distribution of their lengths

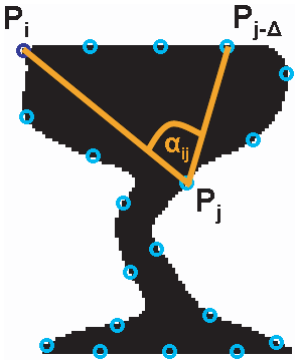


Fig. 2 Our shape descriptor is based on calculating N angles for each sampled point of the shape. In this case P_i is the reference point and the calculation of the angle α_{ij} to the point P_j with $\Delta = 3$ is shown.

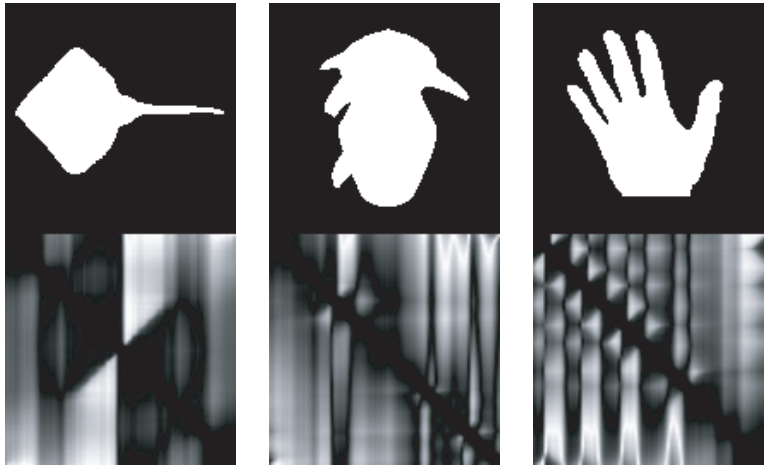
and angles was used as shape descriptor before, as e.g., by Cootes, et al. [15]. Our descriptor uses such chords, but instead of building distribution histograms, we use the relative orientations between specifically chosen chords.

Our descriptor is based on angles α_{ij} which describe the relative spatial arrangement of the sampled points. An angle α_{ij} is calculated between a chord $\overline{P_iP_j}$ from a reference point P_i to another sampled point P_j and a chord $\overline{P_jP_{j-\Delta}}$ from P_j to $P_{j-\Delta}$ by

$$\alpha_{ij} = \measuredangle\left(\overline{P_iP_j}, \overline{P_jP_{j-\Delta}}\right), \tag{1}$$

where $\measuredangle(\ldots)$ denotes the angle in the range of 0 to Pi between the two chords and $P_{j-\Delta}$ is the point that comes Δ positions before P_j in the sequence as is illustrated in **Fig. 2**. Since angles are preserved by a similarity transformation, this descriptor is invariant to translation, rotation and scale.

In the same manner N different angles $\alpha_{i1} \ldots \alpha_{iN}$ can be calculated for one selected reference point P_i. Additionally, each of the sampled points can be chosen as reference point and therefore a $N \times N$ matrix A defined as



(a) Shape 01 (b) Shape 02 (c) Shape 03

Fig. 3 Visualizations of distinct chord angle based shape descriptors. Bright areas indicate parts of the silhouettes which significantly deviate from straight lines.

$$\mathrm{A} = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1N} \\ \vdots & \ddots & \vdots \\ \alpha_{N1} & \cdots & \alpha_{NN} \end{pmatrix} \tag{2}$$

can be used to redundantly describe the entire shape. Obviously, elements on the main diagonal $\alpha_{11}, \cdots, \alpha_{NN}$ are always zero. This descriptor matrix is not symmetric because it considers relative orientations. Please note, that such a shape descriptor implicitly includes local information (close to the main diagonal) and global information (further away from the diagonal). **Figure 3** shows different descriptor matrices for selected shapes. As can be seen the obtained descriptors are highly discriminative where values close to zero (dark areas) represent a locally linear structure, whereas bright areas indicate parts of the silhouettes which significantly deviate from straight lines.

The descriptor depends on which point is chosen as the first point of the sequence. For example the descriptor matrix A shown before changes to

$$\mathrm{A}_{(k)} = \begin{pmatrix} \alpha_{kk} & \dots & \alpha_{k1} & \dots & \alpha_{k(k-1)} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ \alpha_{1k} & \dots & \alpha_{11} & \dots & \alpha_{1(k-1)} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ \alpha_{(k-1)k} & \dots & \alpha_{(k-1)1} & \dots & \alpha_{(k-1)(k-1)} \end{pmatrix} \tag{3}$$

if the k-th point is set as the first point of the sequence. Because only closed boundaries are considered, these two matrices $\mathrm{A}_{(k)}$ and A are directly related by a circular shift. Matrix A can be obtained by shifting the $\mathrm{A}_{(k)}$ matrix $k-1$ rows up and $k-1$ columns to the left. This is an important property for the efficient descriptor matching algorithm presented in the next section. Please note further, that this property only holds for closed boundaries which is also the focus of the rest of the paper and the experimental evaluation.

2.3 Matching Algorithm

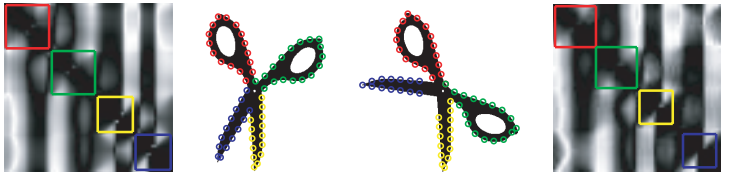
To find a partial match between two given shape contours R_1 and R_2 the corresponding descriptor matrices A_1 with size $M \times M$ and A_2 with size $N \times N$ are compared. For notational simplicity we assume that $M \leq N$.

The aim of shape matching is to identify parts of the two shapes that are similar to each other. In terms of comparing the two descriptor matrices this equals to finding $l \times l$ sized blocks starting at the main diagonal elements $\mathrm{A}_1(\mathrm{s},\mathrm{s})$ and $\mathrm{A}_2(\mathrm{t},\mathrm{t})$ of the two descriptor matrices which yield a small average angular difference value $\Delta(s,t,l)$ between 0 and Pi defined by

$$\Delta(s,t,l) = \frac{1}{l^2} \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} [\mathrm{A}_1(s+i, s+j) - \mathrm{A}_2(t+i, t+j)]^2 \tag{4}$$

between them. This equation is valid due to the previously explained property that a different starting point leads to a circular shift of the descriptor matrix (see Eq. (3)). To find such blocks, all different matching possibilities and chain lenghts have to be considered and the brute-force method becomes inefficient for larger number of points. Therefore, different authors [16] proposed approximations where for example only every n-th point is considered as starting point.

We propose an algorithmic optimization to overcome the limitations of the brute-force approach, which is based on an adaption of the Summed-Area-Table (SAT) approach to calculate all the descriptor differences $\Delta(s,t,l)$. The SAT concept was originally proposed for texture mapping and brought back to the computer vision community by Viola and Jones [17] as integral image. The integral image concept allows to calculate rectangle image features like the sum of all pixel values for any scale and any location in constant time. Details on efficient calculation of integral images can be found, e.g., in Ref. 18).



(a) Desc. Closed (b) Closed (c) Open (d) Desc. Open

Fig. 4 Articulation invariance is handled by returning a set of partially matching boundary fragments. Corresponding fragments are depicted by the same color.

For calculating the similarity scores for all possible configuration triplets $\{s,t,l\}$ in the most efficient way N integral images $Int^1 \dots Int^N$ each of size $M \times M$ are built for N descriptor difference matrices Θ^n defined by

$$\Theta^n = \mathrm{A}_1(1:M, 1:M) - \mathrm{A}_2(n:n+M-1, n:n+M-1)\ , \tag{5}$$

where $\mathrm{A}_1(1:M,1:M)$ is the cropped $M \times M$ square submatrix of A_1 including all elements of the first to the M-th row and the first to the M-th column. The difference matrices Θ^n represent the N possibilities to match the point sequences onto each other. Based on these N integral images $Int^1 \dots Int^N$ the difference values $\Delta(s,t,l)$ can be calculated for every block of any size starting at any point on the diagonal in constant time.

As a final result all matching triplets $\{s,t,l\}$ which provide a difference value $\Delta(s,t,l)$ below a fixed threshold are returned. Obviously, the detected matches may overlap. Thus, the final result is obtained by merging the different matches providing a set of connected point correspondences. This ability of matching parts between two input shapes allows to handle articulations, as it is illustrated in **Fig. 4**, where for the two scissors four matched fragments are returned.

2.4 Shape Similarity

It is important to provide a reasonable similarity measure in addition to the identified matching point sequences, e.g., for tasks like shape retrieval. Commonly, a combination of descriptor difference, matched shape distances like the Procrustes distance and bending energy of an estimated transformation like a Thin Plate Spline is used. Since we focus on partial similarity evaluation we adapt a measure described by Bronstein, et al. [19]. They proposed to use the Pareto-framework for quantitative interpretation of partial similarity. They define two quantities: partiality $\lambda(X', Y')$, which describes the length of the parts in terms of percentage covered of the entire silhouette between 0 and 1 (the higher the value the smaller the part) and dissimilarity $\varepsilon(X', Y')$, which measures the dissimilarity between the parts, where X' and Y' are two contour parts of the shape (the smaller the value the more similar the parts are). Please note, that we do not normalize the descriptor differences obtained in Eq. (4), which might range from zero to Pi. A pair $\Phi(X^*, Y^*) = (\lambda(X^*, Y^*), \varepsilon(X^*, Y^*))$ of partiality and dissimilarity values, fulfilling the criterion of lowest dissimilarity for the given partiality, defines a Pareto optimum. All Pareto optimums can be visualized as a curve, referred to as the set-valued Pareto frontier.

Since finding the Pareto frontier is a combinatorial problem in the discrete case, mostly rough approximations are used as final distance measure. Our matching algorithm automatically evaluates all possible matches for all possible lengths. Therefore, by focusing on the discretization defined by our point sampling, we can calculate a global optimal Pareto frontier, by returning the minimum descriptor difference for all partialities.

Finally, to get a single value measuring the overall similarity between two provided shapes, the so-called Salukwadze distance d_s is calculated based on the Pareto frontier by

$$d_s = \inf_{(X^*, Y^*)} |\Phi(X^*, Y^*)|_1 , \tag{6}$$

where $|\ldots|_1$ is the L1-norm of the vector. Therefore, $d_s(X, Y)$ measures the distance from the utopia $(0, 0)$ to the closest point on the Pareto frontier. The Salukwadze distance is then returned as the shape matching similarity score. **Figure 5** illustrates the calculation of the global optimal Pareto frontier and Salukwadze distance.

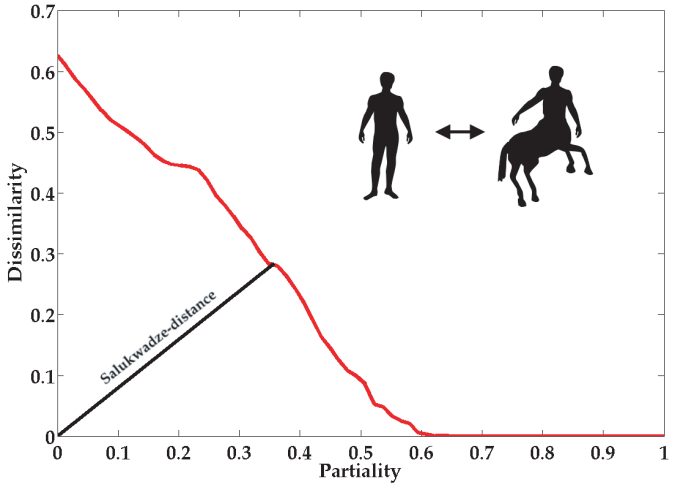


Fig. 5 *IS-Match* returns similarities for all possible matches at all fragment lengths which allows calculating a global optimal Pareto frontier. The Salukwadze distance is returned as partial similarity score.

2.5 Computational Complexity

An exhaustive search over all possible matches for all possible lengths has a complexity of $O(2^{n+t})$. Our proposed approach based on integral image analysis enables matching in $O(nm)$ time, where n and t are the number of sampled points on the two input shapes. We implemented our method in C, which enables shape matching on a desktop PC within milliseconds.

For comparison, **Table 1** summarizes complexities and runtimes of current state-of-the-art shape matching methods. As it is shown in Section 3, only 30 sampled points are required to provide close to state-of-the-art shape retrieval results, which is possible within only 3 ms. Please note, that the runtimes may vary due to differences in implementations and machine configurations. But as can be seen in general *IS-Match* outperforms state-of-the-art concerning computational complexity and actual runtime. To the best of our knowledge this constitutes the fastest method for combinatorial matching of 2D shapes published so far.

Table 1 Comparison of computational complexity and runtime in milliseconds for a single match. Please note, that as it is shown in Fig. 6 our algorithm only requires 30 points to achieve competitive results on reference datasets.

Method	N	Complexity	Runtime
Felzenszwalb [20]	100	$O(t^3k^3)$	500 ms
Scott [12]	100	$O(mnl)$	450 ms
IDSC [11]	100	$O(t^2n)$	310 ms
SC [10]	100	$O(t^2n)$	200 ms
Schmidt [21]	200	$O(t^2log(t))$	X
Brendel and Todorovic [22]	100	$O(nm)$	X
IS-Match	**30**	$O(nm)$	**3 ms**

3. Experimental Evaluation

To evaluate the overall quality of *IS-Match*, we first analyzed the influence of the number of sampled points and different parametrizations on the performance of shape retrieval on a common dataset in Section 3.1. The evaluation shows that only approximately 30 sampled points are required to achieve promising results, where a single match only requires 3 ms of computation time outperforming all other shape matching algorithms by an order of magnitude. Section 3.2 shows results on the largest and currently most important benchmark for evaluating shape matching algorithms, the MPEG-7 dataset. Finally, Section 3.3 analyzes an application of *IS-Match* for the task of shape clustering.

3.1 Performance Analysis

To evaluate the influence of the number of sampled points and different parameterizations we applied *IS-Match* for the task of shape retrieval on two frequently used datasets from Sharvit, et al. [23]. The first dataset consists of 25 images of 6 different classes. Each shape of the dataset was matched against every other shape of the dataset and the global optimal Salukwadze distance as described in Section 2.4 was calculated for every comparison. Then for every reference image all the other shapes were ranked by increasing similarity value. To evaluate the retrieval performance the number of correct first-, second- and third ranked matches that belong to the right class was counted. In all the experiments Δ was set to 5, but experimental evaluations with different parameterizations revealed

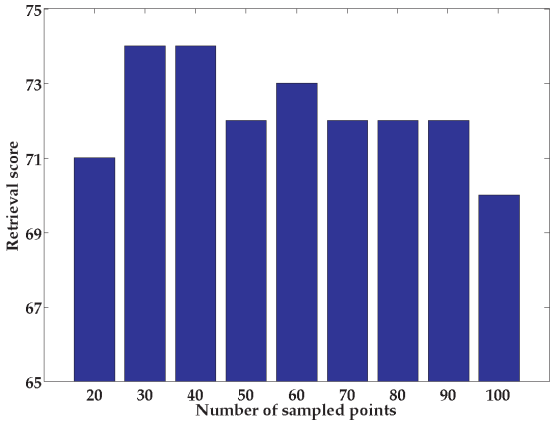


Fig. 6 Retrieval results in dependence of number of sampled points on dataset of [23] consisting of 25 shapes of 6 different classes. Maximum achievable score is 75.

Table 2 Comparison of retrieval rates on dataset of [23] consisting of 25 shapes of 6 different classes. The number of correct first-, second- and third ranked matches is shown.

Algorithm	Top 1	Top 2	Top 3	Sum
Sharvit, et al. [23]	23/25	21/25	20/25	64
Gdalyahu, et al. [24]	25/25	21/25	19/25	65
Belongie, et al. [10]	25/25	24/25	22/25	71
Scott and Nowak [12]	25/25	24/25	23/25	72
Biswas, et al. [25]	25/25	25/25	23/25	73
Ling and Jacobs [11]	25/25	24/25	25/25	74
IS-Match	**25/25**	**25/25**	**24/25**	**74**

that changing Δ only has a small effect on shape retrieval performance.

Figure 6 illustrates the performance of our algorithm on this dataset, where the sum over all correct first-, second- and third ranked matches is shown. Therefore, the best achievable performance value is 75. We present results of *IS-Match* in dependence of the number of sampled points in a range from 20 to 100 sampled points. As can be seen by sampling 30 points we achieve the highest score of 25/25, 25/25, 24/25 which represents state-of-the-art for this dataset as it is shown in **Table 2**.

The second analyzed dataset [23] has 9 classes each consisting of 11 images. We

Table 3 Comparison of retrieval rates on KIMIA 99 dataset [23] consisting of 99 shapes of 9 different classes.

Algorithm	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
Belongie, et al. [10]	97	91	88	85	84	77	75	66	56	37
Kuijper and Olson [8]	99	96	93	93	87	82	75	71	57	56
Biswas, et al. [25]	99	97	98	96	97	97	96	91	83	75
Ling and Jacobs [11]	99	99	99	98	98	97	97	98	94	79
Sebatian, et al. [9]	99	99	99	98	98	98	96	95	94	86
Felzenszwalb, et al. [20]	99	99	99	99	99	99	99	97	93	86
Our method	**99**	**99**	**99**	**99**	**98**	**92**	**90**	**85**	**82**	**75**

again performed an all vs. all comparison and ranked the similarity values for each reference shape. To analyze the ranking properties of our approach we checked if the 10 closest matches to each reference image are in the same category. As outcome of the results on the first dataset we always sampled 30 points from the boundary for matching. **Table 3** compares the results to other algorithms. We again achieve competitive results, e.g., the top 4 ranks in our results were always correct.

To summarize, our algorithm achieves state-of-the-art results on the two datasets with only 30 sampled points, which allows to perform a single match within 3 ms, outperforming the running time of all other shape matching algorithms by an order of magnitude. But please note, that since our method heavily depends on the ordering of the sampled silhouette points, it might be quite sensitive to noise. This is not a problem for the clean shapes of the standard shape matching datasets but might be an issue for application in different real-world scenarios.

3.2 Shape Retrieval on MPEG-7 Dataset

We further applied *IS-Match* to the MPEG-7 silhouette dataset [26] which is currently the most popular dataset for shape matching evaluation. The dataset consists of 70 shape categories, where each category is represented by 20 different images with high intra-class variability. The parametrization of our algorithm is based on the results shown in the previous section. The overall shape matching performance was evaluated by calculating the so-called bullseye rating, in which each image is used as reference and compared to all of the other images. The

Table 4 Comparison of retrieval rates and estimated overall runtimes in hours (!) for calculating the full $N \times N$ similarity matrix on MPEG-7 dataset consisting of 1,400 images showing 70 different classes.

Algorithm	Mokht. [27]	Belongie [10]	Scott [12]	Ling [28]	Felz. [20]	**IS-Match**
Score	75.44%	76.51%	82.46%	86.56%	87.70%	**84.79%**
Runtime	X	54 h	122 h	84 h	136 h	1 h

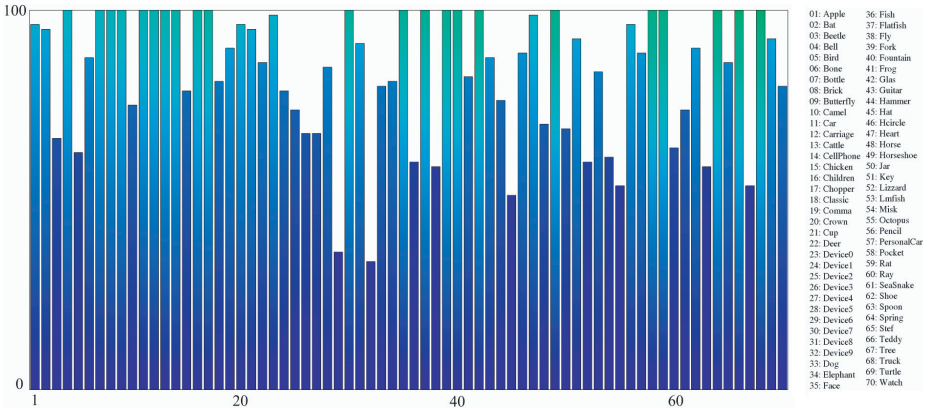


Fig. 7 Shape matching result for the MPEG-7 silhouette dataset. For each of the 70 different classes the bullseye score, i.e., the number of correct matches in the 40 first ranked shapes, is shown. A total bullseye score of 84.79% is achieved.

mean percentage of correct images in the top 40 matches (the 40 images with the highest shape similarity scores) is taken as bullseye rating.

The measured bullseye rating for *IS-Match* was 84.79% and is compared to state-of-the-art algorithms in **Table 4**. As can be the seen the score is close to the best ever achieved by Felzenszwalb, et al. [20] of 87.70%. But please note that [20] uses a much more complex descriptor and requires about 500 ms per match. Therefore, analyzing the entire dataset takes approximately 136 hours for Ref. 20), while with *IS-Match* all similarity scores are provided within a single hour (!). **Figure 7** further illustrates the bullseye-scores for all the 70 different classes independently, revealing that most of the classes are retrieved well. Only two classes (Device 6 and Device 9) have a retrieval score below 50%.

3.3 Shape Clustering

Experiments on the reference shape retrieval datasets revealed that the global optimal Salukwadze distance provided by *IS-Match* captures a meaningful notion of shape dissimilarity. Therefore, we applied it for clustering shapes, which has versatile applications in computer vision. For example shape clusters are used to improve the efficiency of current object detection methods [29),30)] by providing a hierarchical shape structure which allows to perform detection in an efficient coarse-to-fine approach. It enables automatic labeling in image datasets by outlining existing groups and relations between them [31)]. Finding clusters of similar shapes also facilitates the unsupervised clustering of object categories.

Shape clustering methods were recently presented by Schmidt, et al. [21)] who clustered 40 shapes of 4 classes using dynamic time warping matching and k-means clustering. Yankov and Keogh [31)] clustered shapes for grouping together objects in large collections by a manifold clustering approach.

We use a two-step approach for shape clustering. First, we build a pairwise similarity matrix by comparing every possible combination of shapes in the input dataset. Second, we apply a pairwise or proximity-based clustering method on the similarity matrix to find the clusters. The shape similarity measure is not necessarily metric (unsymmetric, violation of the triangle inequality) which has to be taken into account by the clustering method. While most classical methods for pairwise clustering [32)–34)] only consider symmetric similarity matrices, recent methods as e.g., affinity propagation clustering [35)] also work in non-metric spaces.

To prove the quality of the provided similarity scores of *IS-Match* and to find the best suited clustering algorithm we evaluated all combinations between three shape matching algorithms (Shape Context [10)], COPAP [12)] and *IS-Match*) and three clustering algorithms (k-center clustering, hierarchical agglomerative clustering and affinity propagation clustering).

While k-center clustering and hierarchical agglomerative clustering are rather common algorithms, affinity propagation [36)] was just recently proposed by Dueck and Frey [36)] and only first applications in computer vision for multiple view segmentation [37)] and image categorization [35)] were presented. Affinity propagation enables clustering of data points analyzing a pairwise similarity matrix. It is based on iteratively exchanging messages between the data points until a good solution emerges. While most clustering methods only keep track of some candidate exemplars during search, affinity propagation considers all data points as candidates. Affinity propagation is also able to handle missing data points and non-metric similarity measures. Therefore, its perfectly suited for shape clustering, because in general shape similarities do not lie in a metric space.

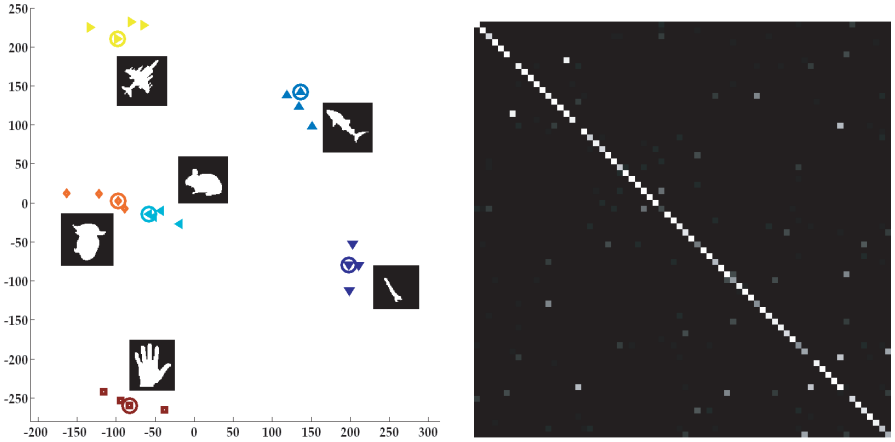
Table 5 Comparison of combinations of shape matching and clustering methods on the KIMIA 25 [23)] dataset consisting of 25 shapes for 6 classes. The F-values and the mutual information scores are shown.

F-Value / MI	K-Center	Agglom. Clustering	Affinity Propagation
Shape Context	0.52/0.66	0.56/0.70	0.77/0.76
COPAP	0.65/0.76	0.72/0.84	1.00/0.97
IS-Match	**0.70/0.84**	**1.00/1.00**	**1.00/1.00**

We first analyzed shape clustering results for all possible combinations of shape matchers and clustering methods on the dataset [23)] already described in Section 3.1 consisting of 25 images of 6 different classes. We used the default parametrization for the two reference shape matching algorithms. For *IS-Match* we again use the parameters defined in Section 3.1. Furthermore, for both k-center clustering and hierarchical agglomerative clustering we set the number of clusters to the true value, whereas for affinity propagation all preference parameters were set to the median of the similarity values, i.e., affinity propagation finds the number of clusters by itself. The k-center algorithm was repeated 10,000 times to cope with the random initialization.

Table 5 summarizes the corresponding clustering results for the six possible combinations of shape matchers and clustering algorithms. Clustering quality is measured by the F-Value analyzing precision and recall and the information theory based mutual information (MI) value. Therefore, in both cases the higher the value the better the corresponding clustering result. As can be seen *IS-Match* strongly outperforms the shape context and COPAP method in terms of clustering quality. Furthermore, affinity propagation clustering leads to better results compared to k-center and hierarchical agglomerative clustering and additionally does not require to fix the number of clusters.

In order to be able to visualize the results of shape clustering we projected the shapes to a two-dimensional space by applying non-metric multi-dimensional



(a) KIMIA-25 (b) MPEG-7

Fig. 8 Clustering results: a) Two-dimensional multi-dimensional scaling visualization of perfect clustering results on KIMIA 25 dataset [23]. Different colors reflect the different clusters identified by affinity propagation. b) Confusion matrix for clustering on MPEG-7 shape dataset consisting of 1,400 shapes for 70 classes. Please note that this result is achieved by affinity propagation without specifying the number of real clusters.

scaling (MDS) to the calculated shape similarity matrix. Because MDS requires a symmetric similarity matrix, we use the minimum distance of the shape pair comparisons. **Figure 8** (a) illustrates the clustering result of *IS-Match* and affinity propagation with color coded cluster assignment, where we get a perfect clustering result (without specifying the number of clusters).

Finally, to demonstrate the high quality and efficiency of the shape clustering method, we applied it to the MPEG-7 dataset. To be able to cluster the entire MPEG-7 dataset (1,400 shapes) the full similarity matrix, containing matching results for 1,960,000 shape comparisons (non-symmetric scores), has to be calculated. This is possible using *IS-Match* in only 50 minutes (!) on a single-core desktop PC. Performing clustering on such datasets based on shape context or COPAP is not possible in reasonable time. Figure 8 (b) shows the confusion matrix of the shape clustering result on MPEG-7 demonstrating excellent clustering quality. The corresponding F-Value is 0.81 and the mutual information value is 0.90. Please note, that these results are achieved without pre-specifying the number of clusters, since affinity propagation finds a reasonable number (91) using default parameters in an autonomous manner.

4. Conclusion

This paper introduced a partial shape matching method denoted as *IS-Match*. A chord angle based descriptor is presented which in combination with an efficient matching step allows detecting subparts of two shapes that possess high similarity. We proposed a fast integral image based implementation which enables matching two shapes within a few milliseconds. Shape retrieval and clustering experiments on common datasets like the MPEG-7 silhouette dataset proved that promising results are achieved at reduced computational costs. Due to the efficiency of the proposed algorithm it is also suited for real-time applications as e.g., in action recognition by matching human silhouettes to reference prototypes or for tracking applications, which will be the focus of future work.

Acknowledgments This work was supported by the Austrian Research Promotion Agency (FFG) project FIT-IT CityFit (815971/14472-GLE/ROD) and the Austrian Science Fund (FWF) under the doctoral program Confluence of Vision and Graphics W1209.

References

1) Felzenszwalb, P.F.: Representation and detection of deformable shapes, *Proc. Conf. Computer Vision and Pattern Recognition*, Vol.1, pp.102–108 (2003).
2) Ferrari, V., Tuytelaars, T. and Gool, L.V.: Object detection by contour segment networks, *Proc. European Conf. Computer Vision*, Vol.3, pp.14–28 (2006).
3) Opelt, A., Pinz, A. and Zisserman, A.: A boundary-fragment-model for object detection, *Proc. European Conf. Computer Vision*, Vol.2, pp.575–588 (2006).
4) Mori, G., Belongie, S. and Malik, H.: Shape contexts enable efficient retrieval of similar shapes, *Proc. Conf. Computer Vision and Pattern Recognition*, pp.723–730 (2001).
5) Wang, H. and Oliensis, J.: Shape matching by segmentation averaging, *Proc. European Conf. Computer Vision*, pp.562–575 (2008).
6) Weinland, D. and Boyer, E.: Action Recognition using exemplar-based embedding, *Proc. Conf. Computer Vision and Pattern Recognition* (2008).
7) Sebastian, T.B., Klein, P.N. and Kimia, B.B.: On aligning curves, *Trans. Pattern Analysis and Mach. Intelligence*, Vol.25, No.1, pp.116–125 (2003).
8) Kuijper, A. and Olsen, O.F.: Describing and matching 2D shapes by their points

of mutual symmetry, *Proc. European Conf. Computer Vision* (2006).
9) Sebastian, T.B., Klein, P.N. and Kimia, B.B.: Recognition of shapes by editing shock graphs, *Proc. International Conf. Computer Vision*, Vol.1, pp.755–762 (2001).
10) Belongie, S., Malik, J. and Puzicha, J.: Shape matching and object recognition using shape contexts, *Trans. Pattern Analysis and Mach. Intelligence*, Vol.24, No.4, pp.509–522 (2002).
11) Ling, H. and Jacobs, D.W.: Using the inner-distance for classification of articulated shapes, *Proc. Conf. Computer Vision and Pattern Recognition*, Vol.2, pp.719–726 (2005).
12) Scott, C. and Nowak, R.: Robust contour matching via the order-preserving assignment problem, *IEEE Trans. Image Processing*, Vol.15, No.7, pp.1831–1838 (2006).
13) Chen, L., Feris, R. and Turk, M.: Efficient partial shape matching using Smith-Waterman algorithm, *Proc. NORDIA Workshop at CVPR* (2008).
14) Donoser, M. and Bischof, H.: Fast non-rigid object boundary tracking, *Proc. British Machine Vision Conf.* (2008).
15) Cootes, T., Cooper, D., Taylor, C. and Graham, J.: Trainable method of parametric shape description, *Journal of Image Vision Computing* (1992).
16) Osada, R., Funkhouser, T., Chazelle, B. and Dobkin, D.: Shape distributions, *ACM Transactions on Graphics*, Vol.21, pp.807–832 (2002).
17) Viola, P. and Jones, M.: Rapid object detection using a boosted cascade of simple features, *Proc. Conf. Computer Vision and Pattern Recognition*, pp.511–518 (2001).
18) Crow, F.C.: Summed-area tables for texture mapping, *Proc. Conf. Computer Graphics and Interactive Techniques*, pp.207–212 (1984).
19) Bronstein, A., Bronstein, M., Bruckstein, A. and Kimmel, R.: Partial similarity of objects, or how to compare a centaur to a horse, *International Journal of Computer Vision* (2008).
20) Felzenszwalb, P.F. and Schwartz, J.D.: Hierarchical matching of deformable shapes, *Proc. Conf. Computer Vision and Pattern Recognition* (2007).
21) Schmidt, F.R., Farin, D. and Cremers, D.: Fast matching of planar shapes in sub-cubic runtime, *Proc. International Conf. Computer Vision* (2007).
22) Brendel, W. and Todorovic, S.: Video object segmentation by tracking regions, *Proc. International Conf. Computer Vision* (2009).
23) Sharvit, D., Chan, J., Tek, H. and Kimia, B.: Symmetry-Based indexing of image database, *Journal of Visual Communication and Image Representation*, Vol.9, No.4 (1998).
24) Gdalyahu, Y. and Weinshall, D.: Flexible syntactic matching of curves and its application to automatic hierarchical classification of silhouettes, *Trans. Pattern Analysis and Mach. Intelligence*, Vol.21, No.2, pp.1312–1328 (1999).
25) Biswas, S., Aggarwal, G. and Chellappa, R.: Efficient indexing for articulation invariant shape matching and retrieval, *Proc. Conf. Computer Vision and Pattern Recognition* (2007).
26) Latecki, L.J., Lakämper, R. and Eckhardt, U.: Shape descriptors for non-rigid shapes with a single closed contour, *Proc. Conf. Computer Vision and Pattern Recognition*, pp.424–429 (2000).
27) Mokhtarian, F., Abbasi, S. and Kittler, J.: Efficient and robust retrieval by shape content through curvature scale space, *Proc. International Workshop on Image Databases and Multimedia Search*, pp.35–42 (1996).
28) Ling, H. and Okada, K.: An efficient earth movers distance algorithm for robust histogram comparison, *Trans. Pattern Analysis and Mach. Intelligence*, Vol.29, pp.840–853 (2007).
29) Gavrila, D.M.: A bayesian, exemplar-based approach to hierarchical shape matching., *Trans. Pattern Analysis and Mach. Intelligence*, Vol.29, pp.1408–1421 (2007).
30) Lin, Z., Davis, L.S., Doermann, D. and DeMenthon, D.: Hierarchical part-template matching for human detection and segmentation, *Proc. International Conf. Computer Vision* (2007).
31) Yankov, D. and Keogh, E.: Manifold clustering of shapes, *Proc. International Conference on Data Mining*, pp.1167–1171 (2006).
32) Hofmann, T. and Buhmann, J.: Pairwise data clustering by deterministic annealing, *Trans. Pattern Analysis and Mach. Intelligence*, Vol.19 (1997).
33) Pavan, M. and Pelillo, M.: Dominant sets and pairwise clustering, *Trans. Pattern Analysis and Mach. Intelligence*, Vol.29, No.1, pp.167–172 (2007).
34) Shi, J. and Malik, J.: Normalized cuts and image segmentation, *Trans. Pattern Analysis and Mach. Intelligence*, Vol.22, No.8, pp.888–905 (2000).
35) Dueck, D. and Frey, B.J.: Non-metric affinity propagation for unsupervised image categorization, *Proc. International Conf. Computer Vision* (2007).
36) Frey, B.J. and Dueck, D.: Clustering by passing messages between data points, *Science*, Vol.315, pp.972–976 (2007).
37) Xiao, J., Wang, J., Tan, P. and Quan, L.: Joint affinity propagation for multiple view segmentation, *Proc. International Conf. Computer Vision* (2007).

(Received February 16, 2010)
(Accepted August 9, 2010)
(Released December 15, 2010)

(Communicated by *Hongbin Zha*)

Michael Donoser was born in 1977. He received his M.S. and Ph.D. degrees with highest distinction from Graz University of Technology, Austria in 2003 and 2007 respectively. He is currently a research associate at the Institute for Computer Graphics and Vision (ICG). His research interests are on unsupervised segmentation and efficient shape matching for category recognition. He received the Josef-Krainer award in 2008 and several best paper awards, e.g., at the International Conference on Pattern Recognition (ICPR) in 2008. He is a senior member of IEEE-CS and IAPR.

Hayko Riemenschneider was born in 1980. He received his M.S. degree with highest distinction from Graz University of Technology, Austria in 2008. He is currently a Ph.D. research assistant at the Institute for Computer Graphics and Vision (ICG), where he is involved in the CityFit project within the Virtual Habitat work group. His current research interests are shape-based object representation, matching and detection, as well as symmetry and repetition analysis. He received the best paper award of the Austria Association for Pattern Recognition (AAPR) in 2009. He is a member of IEEE-CS and IAPR.

Horst Bischof was born in 1967. He received his M.S. and Ph.D. degrees in computer science from the Vienna University of Technology in 1990 and 1993, respectively. In 1998 he got his Habilitation (venia docendi) for applied computer science. Currently he is Professor at the Institute for Computer Graphics and Vision at the Technical University Graz, Austria. Horst Bischof is member of the scientific boards of the applied research centers ECV, VrVis and KNOW. He is board member of the Fraunhofer Institute for Computer Graphics Research (IGD). His research interests include object recognition, visual learning, motion and tracking, visual surveillance and biometrics, medical computer vision, and adaptive methods for computer vision where he has published more than 400 peer reviewed scientific papers. Horst Bischof was co-chairman of international conferences (ICANN, DAGM), and local organizer for ICPR 1996. He was program co-chair of ECCV 2006 and Area chair of CVPR 2007, ECCV 2008, CVPR 2009, ACCV 2009, ACCV 2010. Currently he is Associate Editor for IEEE Trans. Pattern Analysis and Machine Intelligence, Pattern Recognition, Computer and Informatics and the Journal of Universal Computer Science. Horst Bischof has received several awards among them the 29th Pattern Recognition award in 2002; the main price of the German Association for Pattern Recognition DAGM in 2007, the Best scientific paper award at the BMCV 2007 and the Best scientific paper awards at the ICPR 2008 and ICPR 2010.