IImproving Shape retrieval by Spectral Matching and Meta Similarity

Amir Egozi, Yosi Keller, Hugo Guterman

***Abstract*—We propose two computational approaches for improving the retrieval of planar shapes. First, we suggest a geometrically motivated quadratic similarity measure, that is optimized by way of spectral relaxation of a quadratic assignment. By utilizing state-of-the-art shape descriptors and a pairwise serialization constraint, we derive a formulation that is resilient to boundary noise, articulations and non-rigid deformations. This allows both shape matching and retrieval. We also introduce a shape meta-similarity measure that agglomerates pairwise shape similarities and improves the retrieval accuracy. When applied to the MPEG-7 shape dataset in conjunction with the proposed geometric matching scheme, we obtained a retrieval rate of 92.5%.**

I. Introduction

The analysis of shape similarity is of major importance in computer vision applications, as it captures a fundamental attribute of visual objects. Shapes were found to be instrumental in a diverse range of applications, such as: content-based image retrieval [6], tracking [20] and medical imaging [28], to name a few.

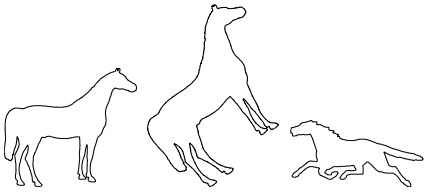


Fig. 1. Three instances of a shape. *a* and *b* are related by a rigid deformation, while c is an articulated replica of *a*.

In the context of this work, a shape is given by a discrete curve, an ordered set of n points

$$S = \{s_1, s_2, \ldots, s_n\}, \tag{1}$$

embedded in $\mathbb{R}^2$. We aim to derive a geometric shape similarity measure

$$\Psi = \Psi\left(S_1, S_2\right), \quad \Psi \in \mathbb{R}^+, \tag{2}$$

that quantifies the similarity between the shapes S_1 and S_2.

As shapes are often the silhouettes of visual objects, a shape similarity measure should be made resilient to shape distortions resulting from projections, as well as to partial alignments due to occlusions. For instance, in tracking [39] and medical imaging applications [28], an object might undergo non-rigid deformations due to articulation and elastic motion. Figure 1 shows three shapes related by rigid and non-rigid deformations, and one would expect a similarity measure to categorize these shapes as belonging to the same class.

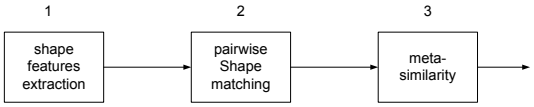


Fig. 2. Shape similarity estimation workflow: 1. Extraction of shape features. 2. Computing pairwise shape similarity using the local features. 3. Shape similarity refinement by agglomerating pairwise shape similarities.

Figure 2 depicts the workflow of contemporary shape retrievals schemes that follows the seminal work of Belongie *et al.*[4]. Given two input shapes S_1 and S_2, we compute two corresponding sets of local shape descriptors (LSDs) [4], [25], [44] in the first step. In the second step the shapes are matched by optimizing a similarity measure. The optimal matching score is then used as the similarity between the shapes.

Recent results [45], [46] introduced the third step of *meta-similarity*, where the pairwise shape similarity (the result of step #2) is improved by utilizing the manifold-like structure of sets of similar shapes. Pairwise shape similarities can be agglomerated by graph based algorithms to derive improved meta-similarities.

Different algorithmic components can be applied in each of the three steps in Fig. 2. Recent works [25], [43] extended the original shape context LSD [4], while others improved the matching phase by improving the similarity measure between LSDs [26].

In this work we provide two core contributions:

First, We present a shape matching scheme (step two in Fig. 2) that combines LSDs and global shape constraints using pairwise geometric similarities. It provides a general framework that can utilize *any* LSD, and is based on optimizing geometric pairwise affinities by solving a quadratic assignment problem. As the solution to the resulting computational problem is *np*-hard, it is approximated by spectral relaxation.

The use of LSDs allows us to prune the space of potential assignments per point, and base the affinity measure on purely geometric measures. We then introduce a serialization constraint to derive a more accurate, sparser affinity matrix. When matching a feature from a query shape to features in a reference shape, our scheme does not require exact LSD correspondence, but only that the true corresponding LSD is

Department of Electrical Engineering, Ben Gurion University. agozi@ee.bgu.ac.il.
School of Engineering, Bar Ilan University, Israel. yosi.keller@gmail.com.
Department of Electrical Engineering, Ben Gurion University. hugo@ee.bgu.ac.il.

one of the $K-$nearest neighbors (K-NN) LSDs. The resulting shape retrieval approach improves the shapes retrieval rate and outperforms previous schemes based on the *same* LSDs.

Second, we introduce a novel *meta-similarity* approach (step three in Fig. 2) that agglomerates pairwise shape similarities computed by *any* scheme to derive a novel *meta-shape-similarity* (MSS).

This paper is organized as follows: Section II reviews previous results in shape matching and retrieval, while Section III introduces the geometric matching scheme. Our meta-shape-similarity is introduced in Section IV. In Section V we verify the proposed schemes by applying them to shape matching and shape retrieval using the MPEG7, Kimia and Articulated shape databases. Concluding remarks and future extensions are discussed in Section VI.

II. Background

A gamut of methods have been proposed for computing shape similarity. These approaches can be classified as either global or local schemes and both adhere to the paradigm presented in Fig. 2. Global methods consider the shape as a whole, representing it by a single global descriptor, such as the shape's area and aspect ratio [29], Fourier-based descriptors [22], and invariant moments [3]. Shape similarity is then given by the difference of those global descriptors.

A global descriptor based on a hierarchical representation of a shape was suggested by Felzenszwalb and Schwartz [14]. A shape is represented by the set of its contours. Using a tree structure, adjacent contours are combined to form matchings between longer segments and full shapes. The authors report improved recognition rates compared to [4], [25].

In order to gain robustness, some schemes aim to quantify intrinsic geometry by computing geodesic distances. Ben Hamza and Krim [18], [19] derived a global shape descriptor for shapes and triangulated surfaces that represents an object by a histogram of the geodesic distance computed between all points on its perimeter or surface. The shape similarity is then given by an information theoretic measure.

Elad and Kimmel [12], [13] used geodesic distances to drive deformation-invariant descriptors of three-dimensional non-rigid shapes. Their approach computed a global descriptor (a canonical form) of the intrinsic geometry of a shape. The descriptor was computed by applying multidimensional scaling (MDS) to the geodesic distances between the vertices of a three-dimensional shape. Their approach was later extended by Bronstein *et al.*to expression-invariant three-dimensional face recognition [8]. The alignment is achieved by minimizing the Gromov-Hausdorff (GH) distance, which was first applied to shape matching by Memoli and Sapiro [30]. A relaxed iterative solution of the GH minimization problem, denoted generalized MDS was suggested by Bronstein *et al.*[9], and was applied by them to planner shape matching [7].

A spectral graph approach to global shape descriptors was introduced by Reuter *et al.*[33] by representing shapes as planar graphs, and analyzing the spectral properties of the Laplace-Beltrami operator of these graphs. The eigenvalues of the discrete Laplace-Beltrami were used as a descriptor and referred to as *shape DNA*.

Alternatively, global shape similarity can be computed without the use of descriptors by aligning the shapes, using either the Hough transform [2], Hausdorff distance [21], or the Iterative Closest Point (ICP) algorithm [10]. In such approaches the LSD is merely the coordinates of each point. As such schemes optimize an objective function, such as the maximal number of elements in a bin, for the Hough transform, or the minimum of the Hausdorff distance, this optimal value can be used as a similarity measure.

In contrast, Local Methods represent a shape by a set of local shape descriptors (LSDs) $\{\phi_i\}_{i=1}^n$, each corresponding to a certain part of the shape or simply to a sample point, $s_i \in S$, along its contour. Thus, computing the similarity measure, requires matching the corresponding local features across the different shapes, and integrating local similarities into a global score. A common practice is to optimize an objective function that comprises of local similarities, with respect to the assignment function, and use its optimum as the similarity score.

As for local shape features and descriptors, curvature is a fundamental perceptual shape feature [15] that is invariant to rigid transformations [37]. Yet, due to discretization issues and sensitivity to noise, it is difficult to compute accurately for discrete shapes [47]. To that aim, Manay *et al.*[27] proposed to compute the curvature using an integral measure, related to the area of intersection of the curve and a circle centralized at the point and the interior of the shape.

A LSD based on an integral curvature measure was presented by Xu *et al.*[44]. They define a shape descriptor denoted as *contour flexibility* that represents the deformable potential at each point along a closed curve. This approach allows to define both local and global descriptors, and a shape similarity measure. Shapes can then be matched by jointly optimizing local and global similarities by Dynamic Programming (DP).

The *shape context* (SC) local shape descriptor was introduced in the seminal work of Belongie *et al.*[4]. Given a set of n points $\{s_i\}_{i=1}^n$, the SC of each point s_i, is the histogram of the relative distances and orientations between the point s_i and the remaining $n-1$ points on S. A logarithmic radial scale is used to make the descriptor more sensitive to the positions of close points, rather than to those further away. The shape contexts is invariant to translations as it consists of relative distances, and can be made scale invariant by normalizing the relative distances. In order to attain rotational invariance, Belongie *et al.*[4] suggest to measure the angle between the edge connecting the two points and the tangent at the sample point. Thus, each SC descriptor characterizes a single point s_i, and the entire shape S is characterized by a set of n descriptors $\{\phi_i\}_{i=1}^n$. The distance between shape descriptors is computed by histogram distance measures, such as the χ^2 statistic [4], EMD distance [26], or the L_2 norm.

The SC was successfully applied to object detection and recognition tasks [31]. In addition, several extensions were proposed. Wang *et al.*[43] blur the SC histogram along the angular axis in order to handle slight shape deformation. Thayananthan *et al.*[40] incorporate edge orientation and a continuity constraint to match image features in cluttered scenes.

The *inner-distance shape contexts* (IDSC) introduced by Ling and Jacobs [25] is a valuable extension of the SC. It replaces the Euclidean distances used by the SC with the *inner-distance*. The inner-distance is a geodesic distance $d_G(s_i, s_j)$, given by the length of the shortest path connecting $s_i, s_j \in S$, where the path is completely contained within S. The inner-distance is an intrinsic geometric feature, robust to articulations and boundary noise.

Given two shapes S_1 and S_2 and their sets of local descriptors, $\left\{\phi_i^1\right\}_{i=1}^n$ and $\left\{\phi_i^2\right\}_{i=1}^n$, respectively, computing a similarity measure $\Psi(S_1, S_2)$, requires the matching of the corresponding points in S_1 and S_2. Let $f : S_1 \rightarrow S_2$ be the assignment function that matches $s \in S_1$ to $f(s) \in S_2$ [1]. It is common [4], [25] to define $\Psi(S_1, S_2)$ as

$$\Psi(S_1, S_2; f) = \sum_{s \in S_1} C(\phi_s^1, \phi_{f(s)}^2), \tag{3}$$

where $c_{ij} = C(\phi_{s_i}, \phi_{s_j})$ is the cost of matching $s_i \in S_1$ to $s_j \in S_2$ based on some local features discrepancy measure (L_1, L_2, χ^2). Hence, the assignment function is the one that optimizes the similarity,

$$\widehat{f} = \arg\min_f \Psi(S_1, S_2; f). \tag{4}$$

This optimization problem is known as the *Linear Assignment Problem* (LAP), and can be optimally solved by the Hungarian algorithm [32] or approximated by DP [4].

Equation 4 can also be solved by graph based schemes. Building upon the planarity of shapes, Schmidt *et al*.proposed a particular formulation of the graph cuts discrete optimization scheme [36] that is adapted for planar graphs. This improves previous results in graph based shape matching [35]

A different class of similarity functions is based on pairwise matching

$$\Psi(S_1, S_2; f) = \sum_{s_i \in S_1} \sum_{s_j \in S_1} \Omega(d_{S_1}(s_i, s_j), d_{S_2}(f(s_i), f(s_j))), \tag{5}$$

where $d_S(\cdot, \cdot)$ is a distance measure, and $\Omega(\cdot, \cdot)$ is an affinity kernel such as a Gaussian. Pairwise matching implies that the optimal assignment f^* is the one that best preserves the distances between pairs of points in both sets. Each distance is measured within a particular set and thus relates to its geometry. In contrast, Eq. 3 maximizes a cross-set distance, based on LSDs, and is therefore non-geometric.

Optimizing Eq. 5 is denoted as the *Quadratic Assignment Problem* (QAP). Its solution is known to be *np*-hard, and can be derived by linear [6] or spectral [24] relaxations. Other approaches include graduated assignment [17], softassign [16] and semidefinite programming [34]. QAP was applied to image retrieval by Berg *et al*.[6], by matching sets of salient pixels. Those were not detected on an object's contour, as edge pixels are difficult to localize. Torresani *et al*.[41] proposed a graph matching scheme for solving the QAP denoted as *dual decomposition*, that was shown to be more robust than spectral relaxation.

[1] A common constraint is that the matching function be one-to-one but not necessarily onto.

A novel graph based approach to the refinement of shape retrieval was presented by Yang *et al*.in [45]. This is a *meta-algorithm* (step #3 in Fig. 2), a high-level machine learning algorithm that refines an existing pairwise shape retrieval scheme, by propagating similarities using a given pairwise shape affinity matrix $A_{N \times N}$. Yang *et al*.utilized the pairwise shape similarity computed by Ling and Jacobs [25] to form a K-NN graph, where shapes are the nodes, and the affinities form the edges. The similarities are propagated over the graph (hence the term *Graph Transduction*), utilizing the clique-like structure, one expects to find in classes of similar objects

$$\begin{aligned} \mathbf{y}_0(i) &= 0, \ \forall i \neq i_0, \ \mathbf{y}_0(i_0) = 1 \\ \mathbf{y}_{n+1} &= M\mathbf{y}_n \\ \mathbf{y}_{n+1}(i_0) &= 1 \end{aligned} \tag{6}$$

where M is a row normalized replica of the A, i_0 is the index of the shape whose similarity is being propagated, and $\mathbf{y}_n$ is the vector of meta-similarities. Using the IDSC LSDs and DP for matching, this approach achieves a retrieval rate of 91% on the MPEG-7 data set.

A meta-algorithm based on spectral embeddings [11] was proposed by Yang *et al*.[46]. In their approach the initial pairwise shape similarities are used to embed the set of shapes, and shape similarities are measured by the L_2 distance between the embedding coordinates. Due to the noisy pairwise similarities and the lack of manifold structure of the shape datasets, computing their embedding using a straight forward implementation of diffusion maps is shown to be inaccurate and numerically unstable. Hence, Yang *et al*.propose two valuable extensions to the implementation of diffusion maps. First, they densify the shape dataset by adding *ghost points* that are additional data points represented by their interpolated affinity entries. This improves the numerical stability of the manifold representation. Second, they refine the pairwise affinities by computing a *Locally Constrained Diffusion Process*. Their approach provides a retrieval rate of 93.32% on the MPEG7 data set.

These two works on meta-similarity [45], [46] are of particular interest as our meta-similarity scheme (Section IV) also utilizes the underlying graph structure to improve K-NN retrieval.

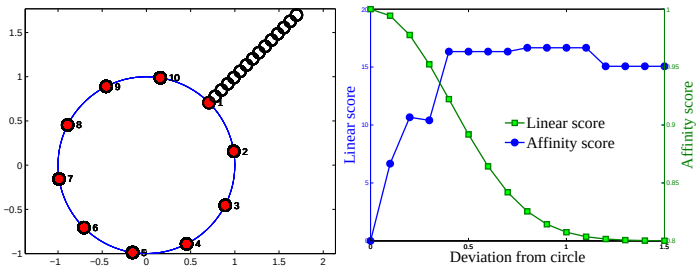
III. Geometric shape matching

In this section we submit that the quadratic similarity measure in Eq. 5 has particular implications to shape matching. It encodes and maximizes isometric matchings between shapes, which constitute a pure geometric similarity, rather than the similarity of shape descriptors. Those are often histograms of distances [4], [25] and lack a clear geometric interpretation. The optimization of this measure corresponds to the *second* step in Fig. 2. Thus, it can be applied with *any* LSD. In this work we use state-of-the-art LSDs that allow the comparison to previous works using different matching schemes.

To motivate our geometric approach we present a schematic example in Fig. 3, where we match the circle shown in Fig. 3a to itself. The circle is given by ten points and we compute the IDSC LSDs for all of them. Given the true assignments, we

compute the optimal scores of the linear (Eq. 3) and geometric (Eq. 5) assignments. The computation is repeated as we move one of the points in Fig. 3a away from the circle and match it to the original circle.

Moving the point distorts its IDSC descriptor significantly, while the IDSCs of the other points (consisting of the histogram of 10 points overall) will remain almost the same. Yet, the assignment error of the DP based scheme, depicted in Fig. 3b increases significantly, due to the assignment error of the single point that was moved. In contrast, the geometrical similarity degrades gracefully.



(a) (b)

Fig. 3. Geometric Vs. Local Shape Descriptors (LSD) based matching. The geometric scheme provides shape affinity, while LSDs matching provides an assignment error measure.

Let S_1 and S_2 be two shapes, consisting of n_1 and n_2 points, respectively. The matching $f : S_1 \rightarrow S_2$ is the one that maximizes the pairwise shape similarity measure Ψ, given in Eq. 5. We represent the assignment function f by an assignment matrix $X \in \{0,1\}^{n_1 \times n_2}$, where an element $x_{ii'} = 1$ indicates that $s_i \in S_1$ is matched to $s_{i'} \in S_2$. The matrix X is vectorized row-wise into $\mathbf{x} \in \{0,1\}^{n_1 n_2}$. Using these notations, Eq. 5 can be cast as:

$$\Psi(S_1, S_2; f) = \mathbf{x}^T A\, \mathbf{x}, \quad \mathbf{x} \in \{0,1\}^{n_1 n_2} \tag{7}$$

where $A \in \mathbb{R}^{(n_1 n_2)\times(n_1 n_2)}$ is the pairwise affinity matrix

$$A\left((i-1)\cdot n_1 + i', (j-1)\cdot n_2 + j'\right) = \Omega\left(d_{S_1}(s_i, s_j), d_{S_2}(s_{i'}, s_{j'})\right), \quad i, j = 1..n_1, i', j' = 1..n_2. \tag{8}$$

The affinity Ω is given by the kernel

$$\Omega\left(d_1, d_2\right) = \exp\left\{-\frac{1}{\sigma}\frac{\|d_1 - d_2\|}{\min\left(d_1, d_2\right)}\right\}, \tag{9}$$

where σ is predefined. The division by $\min\left(d_1, d_2\right)$ allows to handle discrepancies for both small and large values of d_1, d_2. As the same articulation will result in larger distance deviation $\|d_1 - d_2\|$ for larger distances than for smaller ones.

We propose to use $d_{G_k}(s_i, s_j)$, the inner geodesic distance between the points $(s_i, s_j) \in S_k$, as the pairwise geometric measure d_{S_k}, as it yields an affinity measure Ψ that optimizes intrinsic geometrical consistency, rather than shape descriptors consistency [25], [4].

Yet, matching models of size $n_1, n_2 = O\left(100\right)$ might prove intractable as the matrix A is of dimensions $n_1 n_2 \times n_1 n_2$, and would require solving huge linear programming problems or spectral decompositions. To resolve that, we propose to utilize LSDs to reduce the number of potential assignments of each point in S_1. For each point $s_i \in S_1$ we retain the K points $\{s_j\}_{j=1}^{K} \in S_2$ whose LSDs are the K-NN features to $s_i \in S_1$. This reduces the dimensions of A to $n_1 K \times n_1 K$, where K is typically $1-5$. For instance, given two shapes consisting of 100 points each, using $K = 5$, results in an affinity matrix A of size 500×500 rather than $10^4 \times 10^4$.

The *numerical values* of the LSDs are only used to determine the set of K-NN LSDs. Once this set is determined, our scheme only uses the geometric pairwise measure. In contrast, DP based schemes [25], [4] minimize the discrepancy in LSDs and thus depend on their values.

We sparsify the pairwise affinity matrix A by inducing a *serialization constraint*. The serialization constraint is depicted in Figure 4, where given a pair of point assignment $(s_i, s_{i'})$ and $(s_j, s_{j'})$ we set $A\left((i-1)\cdot n_1 + i', (j-1)\cdot n_2 + j'\right) = 0$ for

$$\left|\Delta\left(s_i, s_j\right) - \Delta\left(s_{i'}, s_{j'}\right)\right| > \Delta_{\max}, \tag{10}$$

where $\Delta\left(s_i, s_j\right)$ and $\Delta\left(s_{i'}, s_{j'}\right)$ are the arclengths of the curves (s_i, s_i) and $(s_{i'}, s_{j'})$, respectively. Ideally, for pure isometric matching, $\Delta_{\max}$ could be set to zero, but as we aim to handle articulations, it is set to a few percents of the total arclength. The optimal assignment $\mathbf{x}^*$ is given by

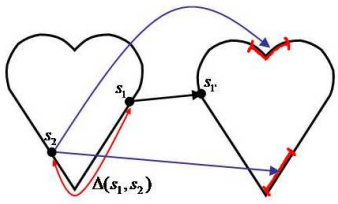


Fig. 4. The serialization constraint. Given one assignment $s_1 \rightarrow s_{1'}$ the image of the point s_2 is constrained by the arclength distance between s_1 and s_2 (denoted by $\Delta(s_1, s_2)$. The red segments are the possible assignments for s_2.

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} \mathbf{x}^T A\, \mathbf{x}, \quad \mathbf{x} \in \{0,1\}^{n_1 K}. \tag{11}$$

But, as this assignment problem is known to be *np*-hard it is approximately solved by spectral relaxation

$$\mathbf{z}^* = \arg\max_{\mathbf{z}} \frac{\mathbf{z}^T A \mathbf{z}}{\mathbf{z}^T \mathbf{z}}, \quad \mathbf{z} \in \mathbb{R}^{n_1 K}. \tag{12}$$

Hence, $\mathbf{z}^*$ is given by the eigenvector corresponding to the leading eigenvalue of A. Given the relaxed solution $\mathbf{z}^*$, we apply a discretization procedure to recover the binary solution $\mathbf{x}^*$. The discretization aims to maximize the sum of the chosen entries in $\mathbf{z}^*$, while enforcing the matching constraints overlooked by the spectral relaxation

$$\mathbf{x}^* = \arg\max_{x} \left(\mathbf{x}^T \mathbf{z}^*\right). \tag{13}$$

Equation 13 corresponds to a rectangular linear assignment problem of dimensions $n_1 \times K$ that is solved by an extension of the Hungarian algorithm to rectangular matrices [32].

Therefore, given two shapes S_1 and S_2 their similarity is given by the optimal quadratic assignment score

$$\Psi\left(S_1, S_2\right) = \mathbf{x}^{*T} A \mathbf{x}^{*}. \tag{14}$$

The similarity measure $\Psi\left(S_1, S_2\right)$ is asymmetric, $\Psi\left(S_1, S_2\right) \neq \Psi\left(S_2, S_1\right)$, hence we symmetrize it by

$$\Psi_S\left(S_1, S_2\right) = \Psi_S\left(S_2, S_1\right) = \Psi\left(S_1, S_2\right) + \Psi\left(S_2, S_1\right) \tag{15}$$

The proposed shape matching scheme can by applied with any combination of shape descriptors and geometric distance measures. In practice, we found the Inner-distance shape context [25] and inner-shape geodesic distances to provide the most accurate results. The similarity between shape descriptors was computed by the L_1 norm.

IV. Meta shape-similarity

In this section we propose a novel unsupervised shape retrieval scheme that utilizes and improves upon a given set of precomputed *pairwise* shape similarities $A_{N\times N}$. The input similarity matrix A can be computed by *any* pairwise similarly measure, such as [4], [25], [45] or the geometric scheme proposed in Section III. Thus, as it quantifies shape similarity *indirectly* we denote it *meta-shape-similarity* (MSS), which corresponds to the *third* step in Fig. 2.

The focal point of the proposed MSS approach is to characterizes a given shape by its similarity to its K-NN shapes. Our premise is that similar shapes form tightly connected shape subsets, and we aim to quantify the similarity between those subsets. The subsets can be represented by constructing a K-NN similarity graph for each shape.

We represent the K-NN similarity graph corresponding to a shape S_i by a vector Λ_i we denote a meta-descriptor

$$\Lambda_i = \begin{cases} 0 & S_t \notin K-NN \\ \dfrac{A_{i,t}}{\sum\limits_{S_t \in K-NN} A_{i,t}} & S_t \in K-NN \end{cases}. \tag{16}$$

Λ_i is a vector of length N that contains K non-zeros entries, each corresponding to a particular NN, and the meta-similarity $A^M_{i,j}$ is given by

$$A^M_{i,j} = \left\|\Lambda_i - \Lambda_j\right\|_{L_1} \tag{17}$$

This is elucidated in Fig. 5. S_3 is the query shape and we depict its graph vicinity along the set of $N = 8$ shapes. Using $K = 3$ NN, vertices #1, #2, and #4 are true similar shapes as they are tightly connected. In contrast, vertex #6 is an outlier although the given similarity measure Ψ_S considers S_3 and S_6 similar. In this example we get $\Lambda_3 = \{0, 0.7, 0, 0.8, 0, 0.6, 0, 0\}$ and $\Lambda_6 = \{0, 0, 0.6, 0, 0, 0, 0.7, 0.8\}$ and $A^M_{3,6} = 4.2$ while $A^M_{3,5} = 1.9$.

Compared to Graph-Transduction [45] the proposed scheme does not propagate similarities, but aims to compare local graph structures as an intrinsic similarity measure.

Evaluating the proposed meta-similarity using Eq. 17 requires $O\left(N\right)$ evaluations. Although the L_1 is computationally efficient, we consider two extensions suitable for large-scale shape databases. Learning from the evolution of image retrieval research, it seems that the scale of multimedia databases (images/shapes) is increasing exponentially. Hence, efficient large-scale retrieval schemes are a necessity. Our first proposal is to only apply the MSS to the K-NN shapes retrieved by the given pairwise similarity. The second, is based on Λ_i being a descriptor in $\mathbb{R}^n$. Hence, it can be retrieved by efficient NN schemes, such as KD-trees [5] and LSH [1].

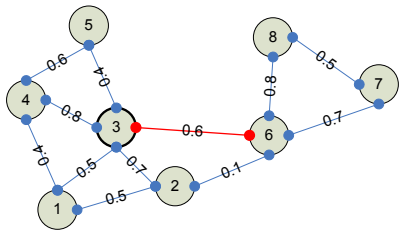


Fig. 5. Graph based similarity. The query shape is represented by Vertex #3. The edges correspond to the pairwise shape similarities. Vertex #6 is an outlier.

V. Experimental results

In this section we experimentally verify the performance of the proposed schemes by applying them to shape matching and retrieval. As for notations, we denote GM the Geometric Matching, DP the Dynamic Programming and MSS as the proposed meta-shape-similarity scheme. Shape retrieval is tested using the MPEG7, Kimia and Articulated shape databases. These were used in previous works, and can be considered a de facto standard in shape retrieval.

Following the three-phase paradigm of shape retrieval in Fig. 2, we conducted experiments in which we compared algorithmic setups that differed in a single algorithmic step. As the focal point of this work are steps two and three, any LSD could have been a viable choice. We used the SC and IDSC and the inner distance measure [25] as a pairwise geometric similarity measure. Those were chosen as their Matlab implementations are publicly available, and those particular implementations were used in previous state-of-the-art works [25], [26], providing a baseline for comparison.

A. MPEG-7 shape retrieval

We applied our schemes to the widely used MPEG7 CE-shape-1 database [23]. This is a challenging data set that measures the performance of shape retrieval algorithms in the presence of rigid, as well as of non-rigid deformations. The database consists of 70 classes of objects, each containing 20 different silhouette images. It is customary to measure the retrieval rate by the "Bull's eye test" [23]. For every shape in the database, we measure its similarity to all of the 1400 shapes, and count the number of retrieved shapes within its class, in the top 40 matches. The score of the test is averaged over all 70 classes. As the SC and IDSC are not reflectionally-invariant, each query shape is compared to a reference shape and to its horizontally flipped replica

$$\overline{\Psi}(S_1, S_2) = max\{\Psi_S(S_1, S_2), \Psi_S(S_1, S_2^F)\}, \tag{18}$$

where S_2^F is the horizontal flipped replica of S_2.

Algorithm	Performance	Algorithm	Performance
Visual Parts [23]	76.45%	Shape-tree [14]	87.7%
Shape Contexts [4]	76.51%	GM+IDSC	**87.47%**
Curve Edit Distance [37]	78.17%	GM+SC	**88.11%**
Generative models [42]	80.03%	Contour Flexibility [44]	89.31%
MDS+SC+DP [25]	84.35%	Graph Transduction [45]	91%
Planar Graph cuts [36]	85%	Locally Constrained Diffusion [46]	93.32%
IDSC+DP [25]	85.40%	GM+SC+Meta Descriptor	**92.51%**
IDSC+DP+EMD-L_1[26]	86.56%	GM+IDSC+Meta Descriptor	**91.46%**

TABLE I
THE RETRIEVAL RATES OF DIFFERENT ALGORITHMS FOR THE MPEG7 CE-SHAPE-1 DATABASE. THESE RESULTS WERE REPORTED IN THE CORRESPONDING REFERENCES.

Table I compares the retrieval rates achieved by the proposed approach with those of state-of-the-art shape retrieval schemes. As we were unable to retrieve the code for some of them, we present the results reported by their authors. We compared the proposed GM and state-of-the-art DP schemes by utilizing the exact same set of LSDs (step #1). The GM outperformed the DP based schemes and the Shape-tree approach.

It is only inferior to the results reported by Xu's [44], in a work that introduced a novel *local shape descriptor* (step #1) and used DP for matching. We were unable to retrieve Xu's code to run it alongside the GM, in order to present a true comparison using the same LSDs.

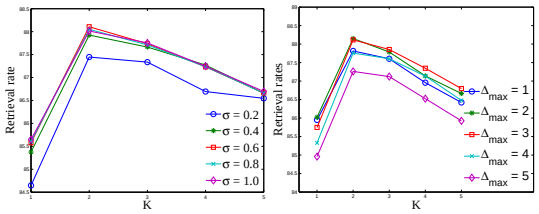
We also compared the proposed MSS scheme (Section IV) to Yang's Graph Transduction [45] and Locally Constrained Diffusion [46]. Both are meta-similarity algorithms corresponding to step #3. The MSS achieved a retrieval rate of 92.51% and 91.46% for the SC and IDSC LSDs, respectively and is only inferior to the Locally Constrained Diffusion approach [46].

We applied the MSS using the publicly available Matlab LSH implementation by Greg Shakhnarovich. We did not limit the search to the K-NN shapes and searched over the entire set of 1400 shapes. On average, the LSH used 15 L_1 computations to retrieve each shape without allowing any misses, compared to the 1400 computations required for exhaustive search. This implies that the MSS is applicable to large scale shape databases.

1) Sensitivity analysis: We conducted further experiments to analyze the sensitivity of our approaches to different parameters values. Figure 6a shows the robustness of the retrieval rate with respect to the affinity bandwidth σ. For low values of $\sigma = 0.2$, the affinity measure becomes susceptible to variations in the geodesic measure. Yet for a wide range of σ values we achieve an accuracy of 88.1%. In Fig. 6b we show the sensitivity to the serialization constraint $\Delta_{\max}$, where the optimal values are $\Delta_{\max} = 2, 3$. Increasing $\Delta_{\max}$ beyond those values, reduces the outlier rejection effect.

We studied the influence of variations in $\Delta_{\max}$ on the accuracy and sparsity of the pairwise affinity matrix in Fig. 7. The serialization constraint is given as a percent of arclength. For $\Delta_{\max} = 1\%$ the sparsity is maximized, while the retrieval rate is maximized for $\Delta_{\max} = 2\%$. We attribute that to the non-rigid articulations characterizing some of the shape classes.

The MSS algorithm depends on a single parameter that is the number of nearest neighbors K used to construct the meta-descriptor Λ_i. We ran the MPEG7 database trials with different K values. The optimal retrieval rate is achieved for a range of $K = 12 - 20$, implying that the MSS is robust to its choice.



(a) (b)

Fig. 6. The retrieval rate of the Geometric Matching. We present the retrieval rate as a function of the number of K-NN. (a) The retrieval rate Vs. the kernel bandwidth σ. We used $\Delta_{max} = 1$ for all the depicted results. (b) The retrieval rate Vs. the serialization threshold $\Delta_{\max}$. We used $\Delta_{max} = 1$ for all the depicted results we used $\sigma = 1$.

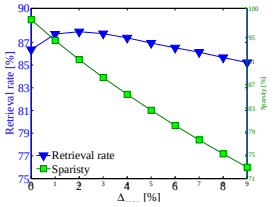


Fig. 7. The average retrieval rate and sparsity of the affinity matrices computed for the shapes in the MPEG7 database.

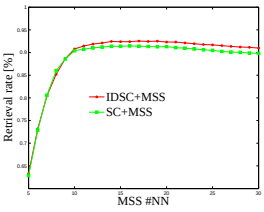


Fig. 8. The MSS retrieval rate for the MPEG7 database. The best retrieval rate is stable over a wide range of nearest neighbors number K.

B. The Articulated data set

The Articulated data set was introduced by Ling and Jacobs [25] for testing the IDSC descriptor. It consists of 40 shapes, divided into 8 classes. Within each class, shapes are subject to

different non-rigid articulation. We followed the experimental setup used in [25], [26]. All schemes used IDSC descriptors with $n_r = 5$ and $n_\theta = 12$, and each shape was made of $n = 200$ samples, as this is the setup used in the previous works whose results we cite. The results are reported in Table II. For each shape, the 40 most similar shapes are retrieved from the data set, and we count the number of first to 40th matches that correspond to the correct object. Thus, for the 40 shapes we can correctly retrieve 40 shapes at most in each position. The GM slightly outperformed the EMD-L_1 results [26]. We attribute that to the shape variations present in this database, which are tailor-made for the IDSC shape descriptor, making the performance of the matching scheme less significant.

	Top1	Top2	Top3	Top4
EMD-L_1[26]	39/40	39/40	34/40	32/40
GM	40/40	40/40	35/40	23/40

TABLE II
RETRIEVAL RATES FOR THE ARTICULATED DATA SET [25].

C. The Kimia silhouettes database

The Kimia database was introduced in [38]. It consists of 99 images from 9 classes. In this experiment, each curve was made of $n = 300$ points and the SC was computed using $n_r = 8$ and $n_\theta = 12$. The affinity measure was computed using $\sigma = 0.5$, $K = 3$ nearest neighbors and $\Delta_{max} = 10$. The retrieval results are reported in Table III. As in the Articulated data set, for each shape we retrieve the ten most similar shapes from the database and count the number of true retrievals at each position. The proposed schemes outperformed the DP based schemes and are on par with the result of Felzenszwalb and Schwartz [14] that are state-of-the-art.

D. Timing results

The average timings for a single shape-to-shape matchings are reported in Table IV. The LSDs were precomputed, and the timing measurements were taken on a 2.0GHz Quad computer. The different schemes were implemented in Matlab using the MEX implementations of the DP kindly provided by the authors of [25]. The majority of the running time of the GM is spend on the retrieval of the K nearest LSDs as it is implemented via exhaustive search. We also compared the running times of the MSS and Graph Transduction [45], for which we used its publicly available Matlab implementation [2]. It should be noted that this implementation is unoptimized, so this comparison qualitative rather than quantitative.

DP	GM	GM+EMD-L_1	Graph Transduction[45]	MSS K=13
14ms	22ms	1.4s	1s	4.6ms

TABLE IV
TIMING RESULTS FOR COMPUTING THE SIMILARITY SCORE OF TWO SHAPES. EACH SHAPE IS MADE OF 100 DESCRIPTORS AND THE GM APPROACHES USED $K = 3$ NEAREST NEIGHBORS.

[2]http://happyyxw.googlepages.com/xingweiyang

VI. CONCLUSIONS

In this work we proposed a general geometric framework for the matching and retrieval of planar shapes. Our scheme agglomerates a pairwise geometric measure by approximating the solution of a quadratic assignment problem using an efficient spectral computational approach. It is able to utilize any combination of local shape descriptor and geometric similarity measures. We suggested a serialization constraint to sparsify the pairwise assignment matrix, and improve the retrieval rate. By applying it to the MPEG7 CE-Shape-1, Kimia silhouettes and Articulated shapes data sets, we show that the proposed algorithm outperforms previous schemes based on descriptors matching and linear assignment.

We also introduced a shape meta-similarity approach that agglomerates pairwise *shape* similarity to reject outliers and improve the retrieval accuracy. This achieves an overall retrieval rate of 92.5% when applied to the MPEG-7 shape dataset and is applicable to large scale datasets.

Finally, we submit that a new large scale shape database is needed. The retrieval rate of our algorithm on the MPEG7 database, as well as other recent schemes, are close to saturation, and recent gains in retrieval rates are becoming less and less significant. Judging by recent innovations in image retrieval, shape retrieval schemes should be made applicable to large scale shape databases of $O\left(10^5\right) - O\left(10^6\right)$ shapes. In our opinion this is the next phase in shape retrieval research evolution.

REFERENCES

[1] A. Andoni and P. Indyk. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. *Commun. ACM*, 51(1):117–122, 2008.

[2] D. H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981.

[3] S. O. Belkasim, M. Shridhar, and M. Ahmadi. Pattern recognition with moment invariants: A comparative study and new results. *Pattern Recognition*, 24(12):1117–1138, 1991.

[4] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.

[5] J. L. Bentley. Multidimensional binary search trees used for associative searching. *Commun. ACM*, 18(9):509–517, 1975.

[6] A. C. Berg, T. L. Berg, and J. Malik. Shape matching and object recognition using low distortion correspondences. In *CVPR*, pages 26–33, 2005.

[7] A. M. Bronstein, M. M. Bronstein, A. M. Bruckstein, and R. Kimmel. Analysis of two-dimensional non-rigid shapes. *Int. J. Comput. Vision*, 78(1):67–88, 2008.

[8] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Three-dimensional face recognition. *Int. J. Comput. Vision*, 64(1):5–30, 2005.

[9] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. *Proc. National Academy of Sciences (PNAS)*, 103(5):1168–1172, January 2006.

[10] Y. Chen and G. G. Medioni. Object modeling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155, 1992.

[11] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5 – 30, 2006. Diffusion Maps and Wavelets.

[12] A. Elad and R. Kimmel. Bending invariant representations for surfaces. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:168, 2001.

[13] A. Elad and R. Kimmel. On bending invariant signatures for surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1285–1295, 2003.

[14] P. F. Felzenszwalb and J. D. Schwartz. Hierarchical matching of deformable shapes. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:1–8, 2007.

Algorithm	Top 1	Top 2	Top 3	Top 4	Top 5	Top 6	Top 7	Top 8	Top 9	Top 10
DP+IDSC [25]	99	99	99	98	98	97	97	98	94	79
SC [4]	97	91	88	85	84	77	75	66	56	37
Shock edit [38]	99	99	99	98	98	98	96	95	94	86
Shape-tree [14]	99	99	99	99	99	99	99	97	93	86
GM	**99**	**99**	**99**	**99**	**99**	**99**	**99**	**97**	**93**	**86**

TABLE III

RETRIEVAL RATES FOR THE KIMIA 99 DATA SET. THE TERM *Top N* REFERS TO THE NUMBER OF CORRECTLY RETRIEVED SHAPES, RANKED AS THE N'TH NEAREST NEIGHBOR, SUMMED OVER ALL 40 SHAPES. THE RETRIEVAL RESULTS USING THE SHAPE CONTEXTS DESCRIPTOR ARE DUE TO SEBASTIAN *et al*LET@TOKENEONEDOT[38].

[15] M. A. Fischler and H. C. Wolf. Locating perceptually salient points on planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):113–129, 1994.

[16] S. Gold and A. Rangarajan. Softmax to softassign: neural network algorithms for combinatorial optimization. *J. Artif. Neural Netw.*, 2(4):381–399, 1995.

[17] S. Gold and A. Rangarajan. A graduated assignment algorithm for graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(4):377–388, 1996.

[18] A. Hamza and H. Krim. Probabilistic shape descriptor for triangulated surfaces. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 1, pages I–1041–4, Sept. 2005.

[19] A. Hamza and H. Krim. Geodesic matching of triangulated surfaces. *Image Processing, IEEE Transactions on*, 15(8):2249–2258, Aug. 2006.

[20] I. Haritaoglu, D. Harwood, and L. S. Davis. W4s: A real-time system for detecting and tracking people in $2\frac{1}{2}$D. In *ECCV*, pages 877–892. 1998.

[21] D. Huttenlocher, G. Klanderman, and W. Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, 1993.

[22] H. Kauppinen, T. Seppanen, and M. Pietikainen. An experimental comparison of autoregressive and fourier-based descriptors in 2d shape classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):201–207, 1995.

[23] L. J. Latecki, R. Lakamper, and T. Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. In *CVPR*, volume 1, pages 424–429 vol.1, 2000.

[24] M. Leordeanu and M. Hebert. A spectral technique for correspondence problems using pairwise constraints. In *ICCV*, volume 2, pages 1482 – 1489, 2005.

[25] H. Ling and D. W. Jacobs. Shape classification using the inner-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):286–299, 2007.

[26] H. Ling and K. Okada. An efficient earth mover's distance algorithm for robust histogram comparison. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(5):840–853, May 2007.

[27] S. Manay, D. Cremers, B.-W. Hong, A. Yezzi, and S. Soatto. Integral invariants for shape matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1602–1618, 2006.

[28] T. Mcinerney and D. Terzopoulos. Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 1(2):91–108, 1996.

[29] B. M. Mehtre, M. S. Kankanhalli, and W. F. Lee. Shape measures for content based image retrieval: A comparison. *Information Processing & Management*, 33(3):319–337, 1997.

[30] F. Mémoli and G. Sapiro. Comparing point clouds. In *SGP '04: Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 32–40, New York, NY, USA, 2004. ACM.

[31] G. Mori and J. Malik. Recovering 3d human body configurations using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(7):1052–1062, 2006.

[32] J. Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32–38, 1957.

[33] M. Reuter, F.-E. Wolter, M. Shenton, and M. Niethammer. Laplace-beltrami eigenvalues and topological features of eigenfunctions for statistical shape analysis. *Computer-Aided Design*, doi:10.1016/j.cad.2009.02.007, 2009.

[34] C. Schellewald and C. Schnörr. Probabilistic subgraph matching based on convex relaxation. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 171–186. 2005.

[35] F. R. Schmidt, D. Farin, and D. Cremers. Fast matching of planar shapes in sub-cubic runtime. In *IEEE International Conference on Computer Vision (ICCV)*, Rio de Janeiro, Brazil, October 2007.

[36] F. R. Schmidt, E. Toeppe, and D. Cremers. Efficient planar graph cuts with applications in computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, Florida, June 2009.

[37] T. B. Sebastian, P. N. Klein, and B. B. Kimia. On aligning curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(1):116–125, 2003.

[38] T. B. Sebastian, P. N. Klein, and B. B. Kimia. Recognition of shapes by editing their shock graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):550–571, 2004.

[39] J. Sullivan and S. Carlsson. Recognizing and tracking human action. In *ECCV*, pages 629–644. 2002.

[40] A. Thayananthan, B. Stenger, P. H. S. Torr, and R. Cipolla. Shape context and chamfer matching in cluttered scenes. *CVPR*, 2003.

[41] L. Torresani, V. Kolmogorov, and C. Rother. Feature correspondence via graph matching: Models and global optimization. In *ECCV '08: Proceedings of the 10th European Conference on Computer Vision*, pages 596–609, Berlin, Heidelberg, 2008. Springer-Verlag.

[42] Z. Tu and A. Yuille. Shape matching and recognition using generative models and informative features. In *ECCV*, pages 195–209. 2004.

[43] L. Wang, J. Shi, G. Song, and I.-F. Shen. Object detection combining recognition and segmentation. In *ACCV*, pages 189–199, 2007.

[44] C. Xu, J. Liu, and X. Tang. 2d shape matching by contour flexibility. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(1):180–186, 2009.

[45] X. Yang, X. Bai, L. J. Latecki, and Z. Tu. Improving shape retrieval by learning graph transduction. In *European Conference on Computer Vision*, volume 4, pages 788–801, 2008.

[46] X. Yang, S. Koknar-Tezel, and L. J. Latecki. Locally Constrained Diffusion Process on Locally Densified Distance Spaces with Applications to Shape Retrieval. In *IEEE Computer Society Conference on Computer vision and Pattern Recognition 2009(CVPR 2009)*, June 2009.

[47] L. Yu and C. Dyer. Perception-based 2d shape modeling by curvature shaping. In *Visual Form 2001*, pages 272–282. 2001.