Articulation-Invariant Representation of Non-planar Shapes

Raghuraman Gopalan, Pavan Turaga, and Rama Chellappa

Dept. of ECE, University of Maryland, College Park, MD 20742 USA
{raghuram,pturaga,rama}@umiacs.umd.edu

Abstract. *Given a set of points corresponding to a 2D projection of a non-planar shape, we would like to obtain a representation invariant to articulations (under no self-occlusions). It is a challenging problem since we need to account for the changes in 2D shape due to 3D articulations, viewpoint variations, as well as the varying effects of imaging process on different regions of the shape due to its non-planarity. By modeling an articulating shape as a combination of approximate convex parts connected by non-convex junctions, we propose to preserve distances between a pair of points by (i) estimating the parts of the shape through approximate convex decomposition, by introducing a robust measure of convexity and (ii) performing part-wise affine normalization by assuming a weak perspective camera model, and then relating the points using the inner distance which is insensitive to planar articulations. We demonstrate the effectiveness of our representation on a dataset with non-planar articulations, and on standard shape retrieval datasets like MPEG-7.*

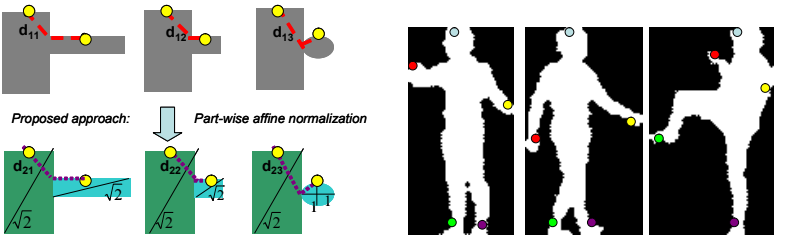
Keywords: Shape representation, articulations, convex decomposition.

1 Introduction

Understanding objects undergoing articulations is of fundamental importance in computer vision. For instance, human actions and hand movements are some common articulations we encounter in daily life, and it is henceforth interesting to know how different 'points' or 'regions' of such objects transform under these conditions. This is also useful for vision applications like, inferring the pose of an object, effective modeling of activities using the transformation of parts, and for human computer interaction in general.

Representation and matching of articulating shapes is a well-studied problem, and the existing approaches can be classified into two main categories namely, those based on appearance-related cues of the object (eg. [1]), and those using shape information which can be contours or silhouettes or voxel-sets (eg. [2–4]). Our work corresponds to the latter category, wherein we represent an object by a set of points constituting its silhouette. Although there are lots of work ([5–7]) on deformation invariant 'matching' of shapes, there is relatively less work on 'representing' a shape invariant to articulations, eg. [2, 8, 9]. Among the above-mentioned efforts only [2] deals with 2D shapes and their representation

K. Daniilidis, P. Maragos, N. Paragios (Eds.): ECCV 2010, Part III, LNCS 6313, pp. 286–299, 2010.



(a) (b)

Fig. 1. (a): Comparing distances across 2D projections of non-planar articulating shapes. (L-R) Shape 1 and 2 belong to the same 3D object, whereas shape 3 is from a different one. For a pair of points with same spatial configuration (yellow dots), Top: Inner distance [2] yields $\|d_{11} - d_{12}\|_2 > \|d_{12} - d_{13}\|_2$, whereas our method (bottom) gives $\|d_{21} - d_{22}\|_2 < \|d_{22} - d_{23}\|_2$. (b) Keypoints with similar shape description obtained from our method. Points were picked in the first frame, and their 'nearest neighbors' are displayed in other two frames. No holistic shape matching was done, emphasizing the importance of a shape representation. *(All figures are best viewed in color).*

mainly addresses planar articulations. However, most articulating shapes, such as a human, are non-planar in nature and there has been very little effort focusing on this problem. This leads us to the question we are addressing in this work.

Given a set of points corresponding to a 2D projection of an articulating shape, how to derive a representation that is invariant/insensitive to articulations, when there is no self-occlusion? An example where this question is relevant is shown in Figure 1, along with results from our proposed shape representation. Such situations also arise when multiple cameras are observing a scene containing non-planar objects, where the projection of a particular 'region' of an object will depend on its relative orientation with the cameras. Accommodating for such variations, in addition to articulations (for which, each object can have different degrees of freedom) makes this a very hard problem.

Contributions: Under the assumption that a 3D articulating object can be expressed as a combination of rigid convex parts connected by non-rigid junctions that are highly non-convex, and there exists a set of viewpoints producing 2D shapes with all parts of the object visible; given one such instance of the 2D shape, we are interested in obtaining an invariant representation across articulations and view changes. We address this problem by,

1. Finding the parts of a 2D articulating shape through approximate convex decomposition, by introducing a robust area-based measure of convexity.
2. Performing part-wise affine normalization to compensate for imaging effects, under a weak perspective camera model, and relating the points using inner distance to achieve articulation invariance (upto a data-dependent error).

After reviewing the prior work in Section 2, we formally define the problem in Section 3. We then present our proposed method in Section 4 by providing detailed analysis on the model assumptions. We evaluate our shape descriptor

in Section 5 through experiments for articulation invariance on a dataset with non-planar shapes, including both intra-class and inter-class studies, and for standard 2D shape retrieval using the MPEG-7 [10] dataset. Section 6 concludes the paper.

2 Related Work

Representation and matching of shapes described by a set of N-dimensional points has been extensively studied, and the survey paper by Veltkamp and Hagedoorn [11] provides a good overview of the early approaches. More recently, there have been advances in matching two non-rigid shapes across deformations. For instance, Felzenszwalb and Schwartz [6] used a hierarchical representation of the shape boundary in an elastic matching framework for comparing a pair of shapes. Yang et al [12] used a locally constrained diffusion process to relate the influence of other shapes in measuring similarity between a pair of shapes. Registering non-rigidly deforming shapes has also been addressed by [7] and [13]. Mateus et al [4] studied the problem of articulation invariant matching of shapes represented as voxel-sets, by reducing the problem into a maximal subgraph isomorphism. There are also efforts, for instance by Bronstein et al [14], on explaining partial similarity between the shapes.

Though there has been considerable progress in defining shape similarity metrics and matching algorithms, finding representations invariant to a class of non-rigid transformations has not been addressed extensively. This is critical for shape analysis because, rather than spending more efforts in matching, we stand to gain if the representation by itself has certain desirable properties. Some works towards this end are as follows. Elad and Kimmel [8] construct a bending invariant signature for isometric surfaces by forming an embedding of the surface that approximates geodesic distances by Euclidean distances. Rustamov [9] came up with a deformation invariant representation of surfaces by using eigenfunctions of the Laplace-Beltrami operator. However in this work, we are specifically interested in articulation insensitive representation of 3D shapes with the knowledge of its 2D projection alone. A key paper that addresses this particular problem is that of Ling and Jacobs [2]. They propose the inner distance, which is the length of the shortest path between a pair of points interior to the shape boundary, as an invariant descriptor of articulations when restricted to a set of translations and rotations of object parts. But such an assumption is applicable only for planar shapes, or when the shape is viewed using an ideal orthographic camera. Since neither of these two settings hold true in most real world scenarios, representing a 2D projection of a 3D non-planar shape invariant to articulations becomes an important problem, which we formalize in the following section.

3 Problem Formulation

An articulating shape $X \subset \mathbb{R}^3$ containing n parts, $\{P_i\}_{i=1}^n$, together with a set of Q junctions, can be written as $X = \{\bigcup_{i=1}^n P_i\} \bigcup \{\bigcup_{i \neq j,\ 1 \leq i,j \leq n} Q_{ij}\}$, where

1. $\forall i, 1 \leq i \leq n$, $P_i \subset \mathbb{R}^3$ is connected and closed, and $P_i \bigcap P_j = \phi, \forall i \neq j, 1 \leq i, j \leq n$
2. $\forall i \neq j, 1 \leq i, j \leq n, Q_{ij} \subset \mathbb{R}^3$, connected and closed, is the junction between P_i and P_j. If there is no junction between P_i and P_j, then $Q_{ij} = \phi$. Otherwise, $Q_{ij} \bigcap P_i \neq \phi$, $Q_{ij} \bigcap P_j \neq \phi$. Further, the volume of Q_{ij} is assumed to be small when compared to that of P_i[1].

Let $A(.)$ be the set of articulations of X, wherein $A(P_i) \in E(3)$ belong to the rigid 3D Euclidean group, and $A(Q_{ij})$ belong to any non-rigid deformation. Further, let V be the set of viewpoints, and $M \subset (A \times V)$ denote the set of conditions such that the 2D projection of X, say $S \subset \mathbb{R}^2$, has all parts visible; i.e. $S_k = \{\bigcup_{i=1}^{n} p_{ik}\} \bigcup \{\bigcup_{i \neq j,\ 1 \leq i,j \leq n} q_{ijk}\}, \forall k = 1\ to\ M$, where $p_{ik} \subset \mathbb{R}^2$ and $q_{ijk} \subset \mathbb{R}^2$ are the corresponding 2D projections of P_i and Q_{ij} respectively. The problem we are interested now is, given an instance of S, say S_1, how to obtain a representation $\tilde{R}(.)$ such that,

$$\tilde{R}(S_1) = \tilde{R}(S_k),\ \forall k = 1\ to\ M \tag{1}$$

4 Proposed Method

In pursuit of (1), we make the following assumptions. (i) X has approximate convex parts P_i that are piece-wise planar, and (ii) X is imaged using a weak-perspective (scaled orthographic) camera to produce $\{S_k\}_{k=1}^{M}$. Let each S_k be represented by a set of t points $\{u_{lk}\}_{l=1}^{t}$. Given two such points $u_{1k}, u_{2k} \in S_k$, we would now like to obtain a distance D such that

$$D(u_{1k}, u_{2k}) = c, \forall k = 1\ to\ M \tag{2}$$

where c is a constant, using which a representation $\tilde{R}(.)$ satisfying (1) can be obtained. Now to preserve distances D across non-planar articulations, we need to account for (atleast) two sources of variations. First, we compensate for changes in the 2D shape S due to changes in viewpoint V and due to the varying effect of imaging process on different regions of a non-planar X, by performing separate affine normalization to each part $p_{ik} \in S_k$. Let T denote the transformation that maps each part p_{ik} to p'_{ik}. Inherently, every point $u_{lk} \in S_k$ gets transformed as $T(u_{lk}) \rightarrow u'_{lk}$, where the transformation parameters depend on the part to which each point belongs. Next, to account for changes in S_k due to articulations A, we relate the two points $u'_{1k}, u'_{2k} \in S_k$ using the inner distance ID [2] which is unchanged under planar articulations. Essentially, we can write (2) as

$$D(u_{1k}, u_{2k}) = ID(u'_{1k}, u'_{2k}), \forall k = 1\ to\ M \tag{3}$$

which, ideally, can be used to construct $\tilde{R}$ (1). But, in general,

$$D(u_{1k}, u_{2k}) = c + \epsilon_k, \forall k = 1\ to\ M \tag{4}$$

[1] A glossary of symbols used in this paper is given in the supplementary material.

where,

$$\epsilon_k = \epsilon_{P_k} + \epsilon_{D_k} + \epsilon_{S_k}, \forall k = 1 \; to \; M \tag{5}$$

is an error that depends on the data S_k. ϵ_{P_k} arises due to the weak perspective approximation of a real-world full-perspective camera. ϵ_{D_k} denotes the error in the inner distance when the path between two points, u_{1k} and u_{2k}, crosses the junctions $q_{ijk} \in S_k$; this happens because the shape change of q_{ijk}, caused by an arbitrary deformation of the 3D junction Q_{ij}, can not be approximated by an affine normalization. But this error is generally negligible since the junctions q_{ijk} are smaller than the parts p_{ik}. ϵ_{S_k} is caused due to changes in the shape of a part p_{ik}, while imaging its original piece-wise planar 3D part P_i that has different shapes across its planes. An illustration is given in Figure 2(a).

Under these assumptions, we propose the following method to solve for (1). By modeling an articulating shape $S \subset \mathbb{R}^2$ as a combination of approximate convex parts p_i connected by non-convex junctions q_{ij}, we

1. Determine the parts of the shape by performing approximate convex decomposition with a robust measure of convexity.
2. Affine normalize the parts, and relate the points in the shape using inner distance to build a shape context descriptor.

We provide the details in the following sub-sections.

4.1 Approximate Convex Decomposition

Convexity has been used as a natural cue to identify 'parts' of an object [15]. An illustration is given in Figure 2(b), where the object consists of two approximate convex parts p_1 and p_2, connected by a non-convex junction q_{12}. Since exact convex decomposition is NP-hard for shapes with holes [16], there are many approximate solutions proposed in the literature (eg. [17]). An important component of this problem is a well-defined measure of convexity for which there are two broad categories of approaches namely, contour-based and area-based. Each has its own merits and limitations, and there are works addressing such issues (eg. [18–20]). But the fundamental problems, that of the intolerance of contour-based measures to small boundary deformations, and the insensitivity of area-based measures to deep (but thin) protrusions of the boundary, have not been addressed satisfactorily.

4.1.1 A New Area-Based Measure of Convexity

In this work, we focus on the problem with existing area-based measures. We start from the basic definition of convexity. Given t points constituting an N-dimensional shape S', the shape is said to be convex if the set of lines connecting all pairs of points lie completely within S'. This definition, in itself, has been used for convex decompositions with considerable success (eg. [21, 22]). What we are interested here is to see if a robust measure of convexity can be built upon it.

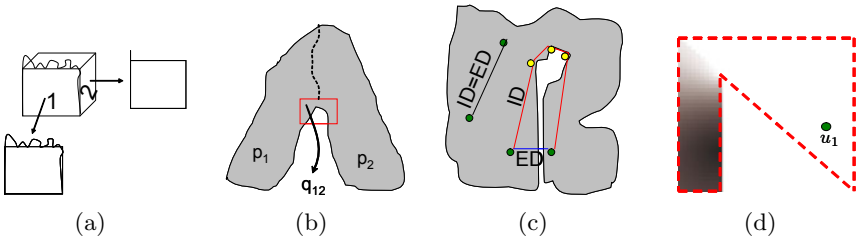


Fig. 2. (a): Error ϵ_{S_k} (5) illustrated by 2D projections, p_{ik}, with the camera parallel to planes 1 and 2. (b): Our model of an articulating object with two approximate convex parts p_1 and p_2, connected by a non-convex junction q_{12}. (c): Variation between ID and ED for a pair of points (green dots). $ID - ED$ is large for non-convex points, with the yellow dots indicating junction regions. (d): Information conveyed by (6) on the potential convex neighbors of u_l. The shape is enclosed by dashed red line. Color of other points u_m is given by $\frac{ED(u_l,u_m)}{ID(u_l,u_m)}$, with value 1 (white) for convex neighbors and tending towards 0 (black) for non-convex neighbors.

We make the following observation. Given two points $u_1, u_2 \in S'$, let $ID(u_1, u_2)$ denote the inner distance between them, and $ED(u_1, u_2)$ denote their Euclidean distance. For a convex S', $ID = ED$ for any given pair of points, whereas for a non-convex S' this is not the case, as shown in Figure 2(c). We can see that, unlike the Euclidean distance, the inner distance inherently captures the shape's boundary and hence is sensitive to deep protrusions along it. Whereas, the difference between ID and ED is not much for minor boundary deformations. Using this property, which significantly alleviates the core issue of the existing area-based convexity measures, we propose a new measure of convexity as follows

$$1 - \frac{1}{(t^2 - t)} \sum_{u_l \in S'} \sum_{u_m \in S', m \neq l} \left(1 - \frac{ED(u_l, u_m)}{ID(u_l, u_m)}\right) \tag{6}$$

where t is the number of points in S', and $1 \leq l, m \leq t$. For a perfectly convex object, this measure will have a value one. We evaluate the robustness of this measure in Section 5.3, and discuss how it conforms to the properties that a convexity measure should satisfy in the supplementary material.

4.1.2 An Algorithm to Obtain Approximate Convex Segments

We now use (6) to segment an articulating shape S into approximate convex regions p_i. We first study if $\frac{ED(u_1,u_2)}{ID(u_1,u_2)}$, in addition to saying whether points u_1 and u_2 belong to a convex region, can shed more information on the potential 'convex neighbors' of a particular point u_1. We proceed by considering a 2D shape S'_1 having two convex regions, shown in Figure 2(d), and measure how $\frac{ED(u_1,.)}{ID(u_1,.)}$ from u_1 to all other $t-1$ points in S'_1 vary. We observe that for those points lying in the same convex region as u_1 this term has a value one, whereas its value decreases for points that lie deeper into the other convex region. Hence

(6) also gives a sense of ordering of convex neighbors around any specific point of interest. This is a very desirable property. Based on this, we formulate the problem of segmenting an articulating shape $S \subset \mathbb{R}^2$ as,

$$\min_{n,p_i} \sum_{i=1}^{n} \sum_{u_l \in p_i} \sum_{u_m \in p_i, u_l \neq u_m} \Big(1 - \frac{ED(u_l, u_m)}{ID(u_l, u_m)}\Big) \tag{7}$$

where $1 \leq l, m \leq t$, n is the desired number of convex parts, and p_i are the corresponding convex regions. We then obtain approximate convex decomposition of S by posing this problem in a Normalized cuts framework [23] and relating all points belonging to S using the information conveyed by (6). The details are provided in Algorithm 1, which is applicable for any N-dimensional shape S'.

Given a set of points t corresponding to an N-dimensional articulating shape S' (which can be a contour or silhouette or voxel-sets, for instance), an estimate $n(> 0)$ of the number of convex parts, and the desired convexity (a number between 0 and 1) for the parts,

(i) Connect every pair of points $(u_l, u_m) \in S'$ with the following edge weight

$$w_{u_l u_m} = exp^{-(\#junctions(u_l,u_m))} * exp^{\frac{-\|1-\frac{ED(u_l,u_m)}{ID(u_l,u_m)}\|_2^2}{\sigma_I^2}} * \begin{cases} exp^{\frac{-\|ID(u_l,u_m)\|_2^2}{\sigma_X^2}} & if \| ID(u_l,u_m) - ED(u_l,u_m) \|_2 \leq T_2 \\ 0 & otherwise \end{cases} \tag{8}$$

(ii) Do: Number of segments from $n - \eta$ to $n + \eta$ (to account for possible errors in junction estimates, see Figure 3(a) for example)

(iii) Perform segmentation using Normalized cuts [23]

(iv) Until: The resulting segments satisfy the desired convexity (6).

Algorithm 1. Algorithm for segmenting an N-dimensional shape into approximate convex parts

Estimate of the Number of Parts: We automatically determine the potential number of parts n using the information contained in (6). We do this by identifying junctions $q_{ij}, i \neq j, 1 \leq i, j \leq n$, which are the regions of high non-convexity. For those pair of points with $ID \neq ED$, we analyze the shortest path SP using which their inner distance is computed. This SP is a collection of line segments, and its intermediate vertice(s) represent points, which by the definition of inner distance [2], bridge two potentially non-convex regions. This is illustrated in Figure 2(c) (see the yellow dots). We then spatially cluster all such points using a sliding window along the contour, since there can be many points around the same junction. Let the total number of detected junctions be n_j. The initial estimate of the number of parts n is then obtained by $n = n_j + 1$, since a junction should connect at least two parts.

With this knowledge, we define the edge weight between a pair of points in (8) where the first two terms collectively convey how possibly can two points lie in the same convex region, and the third term denotes their spatial proximity. T_2, σ_I and σ_X are thresholds chosen experimentally. T_2 governs when two nodes need to be connected, and is picked as the mean of $ID(u_l, u_m) - ED(u_l, u_m), 1 \leq l, m \leq t$. σ_I and σ_X are both set a value of 5. We chose $\eta = 2$ and desired convexity of 0.85 in all our experiments. Sample segmentation results of our algorithm on silhouettes and voxel data are given in Figure 3.

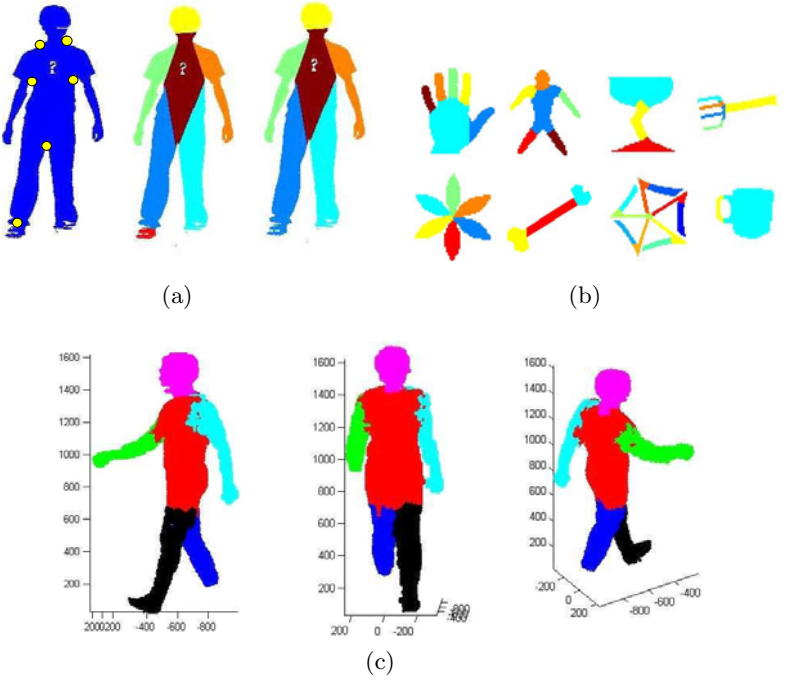


Fig. 3. (a): Result of the segmentation algorithm (Section 4.1.2) on a 2D shape. Junction detection (yellow dots), initial segmentation, followed by the refined segmentation using the desired convexity (=0.85 here) as the user input. (b) Results on shapes from Brown [5] (Top row) and MPEG-7 [10] (Bottom row) datasets. (c): Segmenting a shape represented by voxel-sets using the same algorithm

4.2 Shape Representation Invariant to Non-planar Articulations

We now have an approximate convex decomposition of the articulating shape $S \subset \mathbb{R}^2$, i.e. $S = \{\bigcup_{i=1}^{n} p_i\} \bigcup \{\bigcup_{i \neq j,\ 1 \leq i,j \leq n} q_{ij}\}$. Given a set of M 2D projections of the 3D articulating shape X, $\{S_k\}_{k=1}^{M}$ with all n parts visible, we want to find a representation $\tilde{R}$ that satisfies (1). As before, let $\{u_{lk}\}_{l=1}^{t}$ be the number of points constituting each S_k. Let $u_{1k}, u_{2k} \in S_k$, be two such points. We now compute a distance $D(u_{1k}, u_{2k})$ satisfying (2) using a two step process,

4.2.1 Affine Normalization

To compensate for the change in shape of S_k due to the varying effect of the imaging process on different parts of the non-planar X and due to the changes in viewpoint V, we first perform part-wise affine normalization. This essentially amounts to finding a transformation T such that,

$$T(p_{ik}) \rightarrow p'_{ik} \tag{9}$$

where T fits a minimal enclosing parallelogram [24] to each p_{ik} and transforms it to a unit square. Hence this accounts for the affine effects that include, shear, scale, rotation and translation. This is under the assumption that the original 3D object X has piece-wise planar parts P_i for which, the corresponding 2D part $p_{ik} \in S_k$ can be approximated to be produced by a weak perspective camera.

4.2.2 Articulation Invariance

Let u'_{1k}, u'_{2k} be the transformed point locations after (9). As a result of T, we can approximate the changes in S_k due to 3D articulations A, by representing them as articulations in a plane. Hence, we relate the points u'_{1k}, u'_{2k} using inner distance (ID) and inner angle (IA) [2] that are preserved under planar articulations. We then build a shape context descriptor [25] for each point u'_{lk}, which is a histogram h_{lk} in log-polar space, relating the point u'_{lk} with all other $(t-1)$ points as follows

$$h_{lk}(z) = \#\{u'_{mk}, m \neq l, 1 \leq m \leq t : ID(u'_{lk}, u'_{mk}) \times IA(u'_{lk}, u'_{mk}) \in bin(z)\} \tag{10}$$

where z is the number of bins. We now construct the representation $\tilde{R}(S_k) = [h_{1k}\ h_{2k}\ .\ .\ h_{tk}]$ that satisfies (1) under the model assumptions of Section 4.

5 Experiments

We performed two categories of experiments to evaluate our shape descriptor (10). The first category measures its insensitivity to articulations of non-planar shapes on an internally collected dataset[2], since there is no standard dataset for this problem. Whereas, the next category evaluates its performance on 2D shape retrieval tasks on the benchmark MPEG-7 [10] dataset. We then validated the robustness of our convexity measure (6) on the dataset of Rahtu et al [20].

For all these experiments, given a shape $S \subset \mathbb{R}^2$, we model it as $S = \{\bigcup_{i=1}^{n} p_i\} \bigcup \{\bigcup_{i \neq j,\ 1 \leq i,j \leq n} q_{ij}\}$. We then sample 100 points along its contour, by enforcing equal number of points to be sampled uniformly from each affine normalized part p'_i. Then to compute the histogram (10), we used 12 distance bins and 5 angular bins, thereby resulting in total number of bins $z = 60$. The whole process, for a single shape, takes about 5 seconds on a standard 2GHz processor.

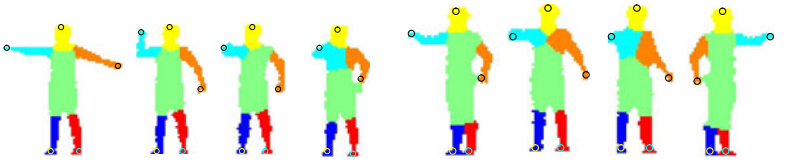
[2] The dataset is available at
`www.umiacs.umd.edu/~raghuram/Datasets/NonPlanarArt.zip`

5.1 Non-planar Articulations

We did two experiments, one to measure the variations in (10) across intra-class articulations, and the other to recognize five different articulating objects.

5.1.1 Intra-class Articulations

We collected data of an articulating human, observed from four cameras, with the hands undergoing significant out-of-plane motion. The silhouettes, shown in Figure 4, were obtained by performing background subtraction, where the parts p_i of the shape (from Section 4.1) along with some points having similar representation (10) are identified by color-codes.



(a) (b)

Fig. 4. Dataset with non-planar articulations: Intra-class variations of an articulating human. (a): A set of actions observed from a single camera. (b): A same action observed from 4 cameras. The regions obtained from segmentation (Section 4.1) along with the points having similar shape representation (Section 4.2), are color-coded

We divided the dataset of around 1000 silhouettes, into an unoccluded part of about 150 silhouettes (where there is no self-occlusion of the human) and an occluded part, and compared our representation (10) with the inner distance shape context (IDSC) [2] that is insensitive to articulations when the shape is planar. We chose to compare with this method since, it addresses articulation invariance in 2D shapes from the 'representation' aspect rather than matching. We used dynamic programming to obtain point correspondences between two shapes. Given in Table 1 are the mean and standard deviations of the difference (in L_2 sense) of the descriptions (10) of the matched points. We do this for every pair of shapes in our dataset, with and without occlusion.

It can be seen that the matching cost for our descriptor is significantly less for the unoccluded pair of shapes, and is noticeably lower than [2] for the occluded pair too. This, in a way, signifies that our model assumptions (Section 4) is a good approximation to the problem of representing a shape invariant to non-planar articulations (Section 3).

5.1.2 Inter-class Variations

We now analyze how our representation (10) can be used for recognition across the 2D shapes produced by different 3D non-planar articulating objects. We collected silhouettes of five different objects, a human and four robots, performing articulations observed from different viewpoints. There were ten instances per

Table 1. Shape matching costs on the dataset with an articulating human. The cost for our descriptor is around one-tenth of that of [2].

Method	Matching cost (mean ± standard deviation)	
	Without occlusion	With occlusion
IDSC [2]	0.48 ± 0.21	3.45 ± 1.63
Ours	0.025 ± 0.0012	0.46 ± 0.11

subject, with significant occlusion, leading to fifty shapes in total as shown in Figure 5. We compared our algorithm with IDSC in both a leave-one-out recognition setting by computing the Top-1 recognition rate, and also in a validation setting using the Bulls-eye test that counts how many of the 10 possible correct matches are present in the top 20 nearest shapes (for each of the 50 shapes). We report the results in Table 2. It can be seen that our descriptor, in addition to handling non-planar articulations, can distinguish different shapes. This validates the main motivation behind our work (Figure 1). The errors in recognition are mostly due to occlusions, which our model can not account for. It is an interesting future work to see how to relax our assumptions to address the more general problem stated in Section 3.

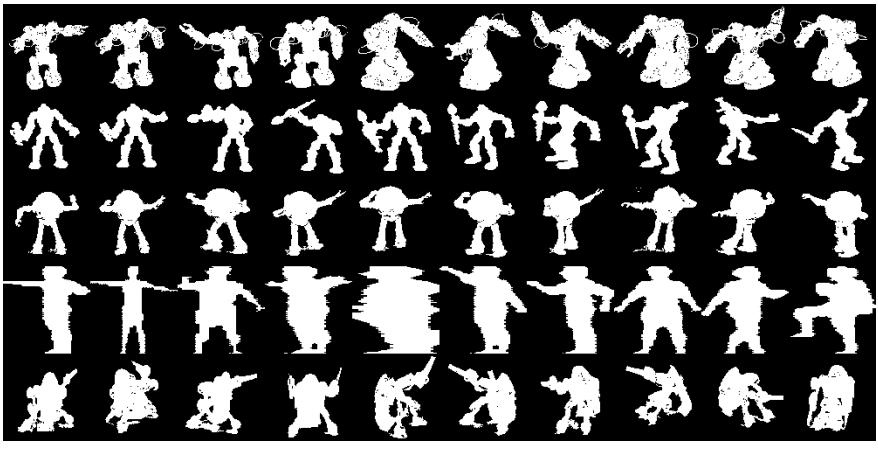


Fig. 5. Dataset of non-planar articulations of different subjects. Four robots and human, with a total of 50 shapes.

Table 2. Recognition across inter-class non-planar articulations

Method	Top-1 Recognition rate (in %)	BullsEye score (in %)
IDSC [2]	58	39.4
Ours	80	63.8

5.2 Shape Retrieval

We then evaluated our descriptor for 2D shape retrieval[3] tasks to study its ability in handling general shape deformations, in addition to pure articulations. We used the benchmark MPEG-7 dataset [10], which contains 70 different shape classes with 20 instances per class. This is a challenging dataset with significant intra-class shape deformations. Some example shapes are given in Figure 3(b). The recognition rate is calculated using the Bulls-Eye test by finding the top 40 closest matches for each test shape, and computing how many of the twenty possible correct matches are present in it. The retrieval rates are given in Table 3, and we compare with the most recent and other representative methods.

Almost all shapes in this dataset are planar. So the least we would expect is to perform as well as [2], since but for handling non-planar articulations our representation resembles IDSC. The improvement using our representation is mainly due to cases where the shapes have distinct part structure, and when the variations in the parts are different. A part-driven, holistic shape descriptor can capture such variations better. It is interesting to see that we perform better than methods like [12, 26] that use sophisticated matching methods by seeing how different shapes in the dataset influence the matching cost of a pair of shapes. Hence through this study, we would like to highlight the importance of a good underlying shape representation.

Table 3. Retrieval results on MPEG-7 dataset [10]

Algorithm	BullsEye score (in %)
SC+TPS [25]	76.51
Generative models [27]	80.03
IDSC [2]	85.40
Shape-tree [6]	87.70
Label Propagation [26]	91.00
Locally constrained diffusion [12]	93.32
Ours	93.67

5.3 Experiment on the Convexity Measure

Finally, we performed an experiment to evaluate our convexity measure (6) by comparing it with the recent work by Rahtu et al [20]. Since there is no standard dataset for this task, we provide results on their dataset in Figure 6. We make two observations. 1) For similar shapes (text in red and blue), the variation in our convexity measure is much smaller than that of [20]. This reinforces the insensitivity of our measure to intra-class variations of the shape, which is very desirable. 2) It can also been seen that our convexity measure is very sensitive to lengthy disconnected parts (text in green). This is mainly because, we compute pair-wise variations in ID and ED for all points in the shape, which will be high in such cases. These results, intuitively, are more meaningful than that of [20].

[3] Evaluations on the Brown dataset [5] and some illustrations on incorrect retrievals are provided in the supplementary material.

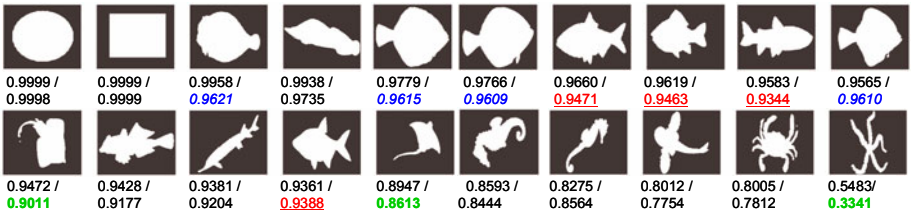


Fig. 6. Performance of our convexity measure on the dataset of [20]. Given at bottom of each shape are the convexity measures of [20] followed by ours (6). Our measure is insensitive to intra-class shape variations (text in red and *blue*), and is more sensitive when a part of the shape is disconnected from other parts (text in **green**).

6 Conclusion

We proposed a method to represent a 2D projection of a non-planar shape invariant to articulations, when there is no occlusion. By assuming a weak perspective camera model, we showed that a part-wise affine normalization can help preserve distances between points, upto a data-dependent error. We then studied its utility through experiments for recognition across non-planar articulations, and for general shape retrieval. It is interesting to see how our assumptions can be relaxed to address this problem in a more general setting.

Acknowledgements. This work was supported by a MURI Grant N00014-08-1-0638 from the Office of Naval Research. R.G. would like to thank Dr. Ashok Veeraraghavan for motivating the problem, and Kaushik Mitra for helpful discussions.

References

1. Zhang, J., Collins, R., Liu, Y.: Representation and Matching of Articulated Shapes. In: CVPR, pp. 342–349 (2004)
2. Ling, H., Jacobs, D.: Shape classification using the inner-distance. IEEE TPAMI 29, 286–299 (2007)
3. Bronstein, A.M., Bronstein, M.M., Bruckstein, A.M., Kimmel, R.: Matching two-dimensional articulated shapes using generalized multidimensional scaling. In: AMDO, pp. 48–57 (2006)
4. Mateus, D., Horaud, R.P., Knossow, D., Cuzzolin, F., Boyer, E.: Articulated shape matching using Laplacian eigenfunctions and unsupervised point registration. In: CVPR, pp. 1–8 (2008)
5. Sebastian, T.B., Klein, P.N., Kimia, B.B.: Recognition of Shapes by Editing Their Shock Graphs. IEEE TPAMI 26, 550–571 (2004)
6. Felzenszwalb, P.F., Schwartz, J.D.: Hierarchical matching of deformable shapes. In: CVPR, pp. 1–8 (2007)
7. Schoenemann, T., Cremers, D.: Matching non-rigidly deformable shapes across images: A globally optimal solution. In: CVPR, pp. 1–6 (2008)

8. Elad, A., Kimmel, R.: On bending invariant signatures for surfaces. IEEE TPAMI 25, 1285–1295 (2003)
9. Rustamov, R.M.: Laplace-Beltrami eigenfunctions for deformation invariant shape representation. In: Eurographics Symposium on Geometry Processing, pp. 225–233 (2007)
10. Latecki, L.J., Lakämper, R., Eckhardt, T.: Shape descriptors for non-rigid shapes with a single closed contour. In: CVPR, pp. 424–429 (2000)
11. Veltkamp, R.C., Hagedoorn, M.: State of the Art in Shape Matching. In: Principles of Visual Information Retrieval, pp. 87–119 (2001)
12. Yang, X., Köknar-Tezel, S., Latecki, L.J.: Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval. In: CVPR, pp. 357–364 (2009)
13. Wang, J., Chan, K.L.: Shape evolution for rigid and nonrigid shape registration and recovery. In: CVPR, pp. 164–171 (2009)
14. Bronstein, A.M., Bronstein, M.M., Bruckstein, A.M., Kimmel, R.: Partial similarity of objects, or how to compare a centaur to a horse. IJCV 84, 163–183 (2009)
15. Hoffman, D.D., Richards, W.: Parts of recognition. Cognition 18, 65–96 (1984)
16. Lingas, A.: The power of non-rectilinear holes. In: Colloquim on Automata, Languages and Programming, pp. 369–383 (1982)
17. Lien, J.M., Amato, N.M.: Approximate convex decomposition of polygons. In: Computational Geometry: Theory and Applications, vol. 35, pp. 100–123 (2006)
18. Rosin, P.L.: Shape partitioning by convexity. IEEE Transactions on Systems, Man, and Cybernetics, Part A 30, 202–210 (2000)
19. Zunic, J., Rosin, P.L.: A new convexity measure for polygons. IEEE TPAMI 26, 923–934 (2004)
20. Rahtu, E., Salo, M., Heikkila, J.: A new convexity measure based on a probabilistic interpretation of images. IEEE TPAMI 28, 1501–1512 (2006)
21. Shapiro, L.G., Haralick, R.M.: Decomposition of two-dimensional shapes by graph-theoretic clustering. IEEE TPAMI 1, 10–20 (1979)
22. Walker, L.L., Malik, J.: Can convexity explain how humans segment objects into parts? Journal of Vision 3, 503 (2003)
23. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE TPAMI 22, 888–905 (2000)
24. Schwarz, C., Teich, J., Vainshtein, A., Welzl, E., Evans, B.L.: Minimal enclosing parallelogram with application. In: Symposium on Computational Geometry, pp. 434–435 (1995)
25. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE TPAMI 24, 509–522 (2002)
26. Yang, X., Bai, X., Latecki, L.J., Tu, Z.: Improving shape retrieval by learning graph transduction. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 788–801. Springer, Heidelberg (2008)
27. Tu, Z., Yuille, A.L.: Shape matching and recognition-using generative models and informative features. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3023, pp. 195–209. Springer, Heidelberg (2004)