

Articulation Prior in an Axial Representation

Aykut Erdem, Erkut Erdem, and Sibel Tari

Middle East Technical University, Department of Computer Engineering, Ankara,
TR-06531, TURKEY,

{aykut,erkut}@ceng.metu.edu.tr, stari@metu.edu.tr

Abstract. Local symmetry axis based schemes have been used for generic shape recognition as they lead to articulation insensitive representations. Despite their strengths, purely syntactic level of axial representations precludes the possibility of distinguishing a likely articulation from an unlikely one. In order to overcome this weakness, syntax should be combined with pragmatics and/or semantics. As a solution we propose a novel articulation space which enables inferences on the likelihood of possible articulations. Articulation priors can be constructed directly from examples (pragmatics) or set externally (semantics). We incorporate articulation priors to a skeletal matching scheme to arrive at an enriched axial representation which is sensitive to unlikely articulations but insensitive to likely ones.

1 Introduction

Generic shape recognition demands representations which are robust with respect to local deformations and articulations as well as transformations which arise due to viewpoint changes. Different representations and matching schemes which are invariant or insensitive to articulations and bendings were presented in the literature [1–9]. The schemes based on local symmetry axis have been particularly considered due to their ability to capture the shape structure. Despite their strengths, presemantic and purely syntactic representations used in these techniques fail to distinguish a likely articulation from an unlikely one. Such a situation is demonstrated in Figure 1: Instances of two shapes share a similar change in local symmetry structure, but based on our past knowledge, the shape given in Figure 1(d) should not be matched with the one given in Figure 1(c) and not be considered as a fork. Therefore, a mechanism to represent possible variations on the symmetry branches and infer them from data becomes essential.

In this work, we propose *articulation space* as a new representation space in which similar articulations and bendings yield closer coordinates. A point in the articulation space is deprived of shape information and global deformation such as affine transformation and represents only a local posture assumed by a section of the shape. We demonstrate that it is possible to build articulation priors and make inferences on this space. These inferences can be combined with spatial layout models such as skeletal matching to arrive at a shape recognition scheme which is insensitive to likely articulations and sensitive to unlikely ones, thus articulation space provides a mechanism to combine syntax with pragmatics.

An important feature of our method is that the part information is utilized without explicit computation of parts or joints. Even though we start with the coarsest representation with one level of part hierarchy representing just the limbs of a main body, existence of ignored joints such as an elbow leads to special patterns in the articulation space.

Main inspiration for our work comes from two sources. First one is the landmark based shape analysis of Kendall [10] and Bookstein [11], which has been

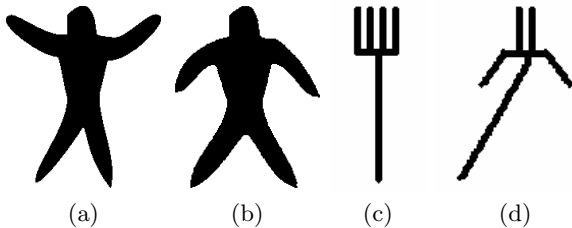


Fig. 1. Human silhouettes (a) and (b) and the shapes (c) and (d) share a similar change in local symmetry structure. These cases should be distinguished at a higher semantic level which takes into account the fact that the human arm should have much larger variability than the parts of a fork.

adopted by Perona et. al. [12] to design a recognition scheme by considering the relative spatial arrangement of shape sections. Our goal is completely different from that of Perona et.al. [12]. They use Kendall’s idea to filter out global transformations in order to capture shape. On the other hand we filter out the shape information in addition to global transformations to capture the articulations.

Second inspiration comes from a recent local symmetry formulation of Aslan et.al. [8] which computes, in a naive sense, the coarsest representation of a shape with one level of hierarchy. Main appeal of this representation is that it allows very robust and easy reference point extraction and makes the presented ideas practically possible.

Previously in [4] deformations on two primitive shapes (the circle and the worm) are considered. A linear shape space is constructed from examples with the help of principle component analysis. Principal component analysis is also used in [13] to learn a linear shape space which captures the global deformations of the medial axis.

Articulation priors are considered particularly in applied problems involving motion and tracking [14]. Pose configurations are represented as data determined manifolds embedded in high dimensional measurement spaces [14, 15].

The paper is organized as follows: In Section 2 we review local symmetry representation presented in [8] and discuss how candidate deformable shape sections are identified from it. In Section 3 we explain representation of articulations in a special coordinate system which is used to form the articulation space. In Section 4 we explore the structure of articulation space. In Section 5 we give some illustrative examples demonstrating some inferences on this space. Finally, we conclude with the summary and future work.

2 Disconnected Skeleton and Deformable Sections

Consider a shape whose boundary is Γ . Let v be the solution of the following differential equation:

$$\nabla^2 v = \frac{v}{\rho^2}, \quad v|_{\Gamma} = 1 \tag{1}$$

Successive level curves of v are gradually smooth versions of Γ and the amount of smoothing increases with increasing ρ . v function is closely related to the Ambrosio-Tortorelli approximation [16] of Mumford-Shah segmentation functional [17]. In [18] this connection is explored and v function is proposed as a multiscale alternative to distance transform which can be directly computed from unsegmented real images. Recently, Aslan et. al. [8] considered the limit case behavior of Equation (1) as $\rho \rightarrow \infty$ for silhouettes. By this way, in a naive sense, the coarsest scale at which the shape is perceived as a single blob is selected. Resulting surface v has similar characteristics to the distance surface used in [19] and permits to a very stable shape analysis by capturing the ribbon-like sections in the form of a *disconnected skeleton* – a set of disconnected protrusion and indentation branches. Each protrusion branch emanates from a boundary

protrusion and ends at an interior point possibly locating a joint, therefore it indicates a section of the shape which may or may not bend. Conceptually, such a branch measures the deviation of the shape from a circle [20, 8, 21, 18]. Some shapes and their coarsest level representations are shown in Figure 2. Notice that the approach computes only 1-level of hierarchy – limbs of the main body.

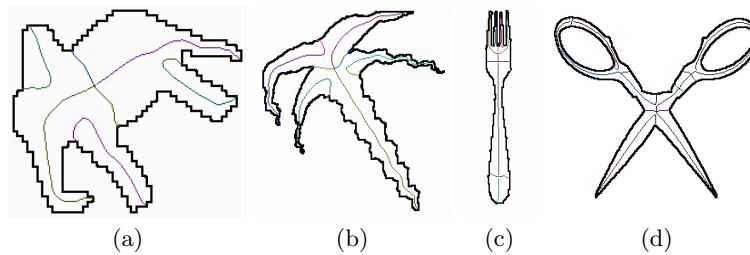


Fig. 2. Some shapes and their disconnected skeletons computed using the method presented in [8]. Shapes with (a)-(b) noisy boundary (c) thin structures (d) holes.

3 Local Affine-Invariant (\mathcal{LA}) Coordinates

In the disconnected skeleton representation, each protrusion branch is neighbored by two indentation branches [18, 22]. The end point of a protrusion branch and the two indentations fix a local coordinate frame in which the start point (tip) of the protrusion moves freely.

These four points define three vectors starting from the end point of the protrusion branch and ending respectively at the tips of the two indentations and the protrusion. The third vector can be represented as a linear combination of the remaining two. Some vector combinations are shown in Figure 3. Note that forming vector combinations is an automatic procedure and does not require explicit part computation.

When these vectors are transformed to standard bases, every configuration can be represented by only a single point (Figure 4). We call this point as Local Affine-Invariant (\mathcal{LA}) coordinate. In this coordinate system each point denotes

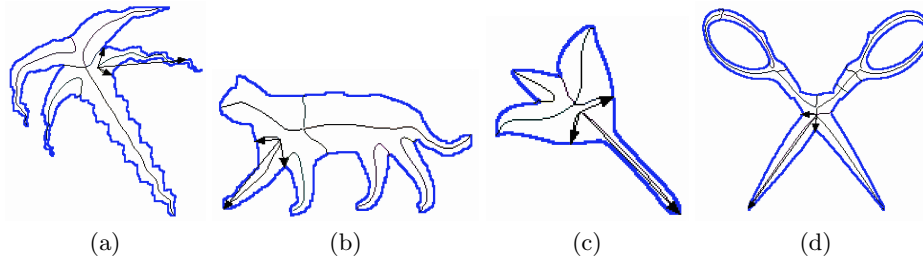


Fig. 3. Vector combinations (extracted automatically) of some sections of (a) palm tree (b) cat (c) flower (d) scissors.

“local” pose of a shape section which may or may not allow to articulate or bend.

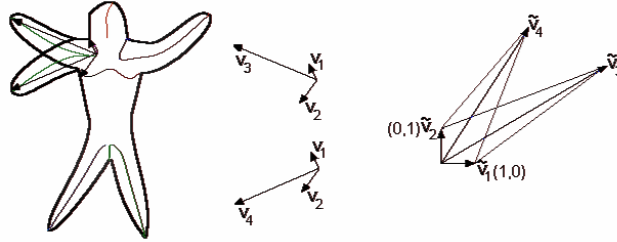


Fig. 4. Articulation of a section can be defined by means of a single point when a local coordinate frame is fixed.

4 Articulation Space

In geometric point of view, shape is defined as the geometric information that remains when location, scale and rotational effects are filtered out [10]. On the other hand, it is the shape information which has to be filtered out in order to make articulations explicit. \mathcal{LA} coordinate representations do not carry any shape information.

Notice that the four points used in the construction of vector combinations of a section’s pose define a quadrangle. Therefore it is possible to associate each \mathcal{LA} coordinate with a set of affine related quadrangles or equivalently a canonical quadrangle represented in \mathcal{LA} coordinates. The collection of such quadrangles in \mathcal{LA} coordinates may be considered as an articulation space (Figure 5). Notice the qualitative similarity of this space to Kendall shape space [10].

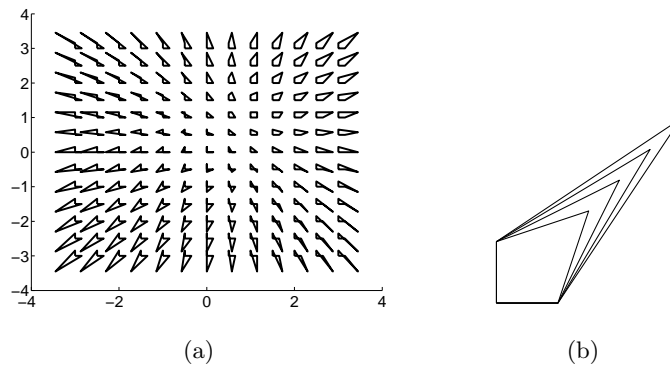


Fig. 5. Articulation space. (a) Each point in the articulation space can be associated with a quadrangle (b) four quadrangles that fall on $x = y$ line in the articulation space.

In articulation space, similar articulations or bendings yield closer coordinates. Consider the two human silhouettes shown in Figure 6. Since the left arms have similar postures, the corresponding articulations are represented by two nearby points in \mathcal{LA} coordinates. On the contrary, \mathcal{LA} coordinates of right arms are far distant from each other. Notice that a horizontal arm will be on $x = y$ line, whose polar representation is $(l, \pi/4)$ where l is the dimensionless arm length.

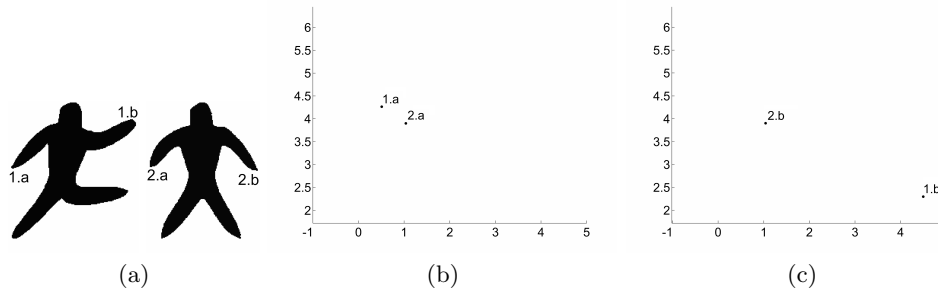


Fig. 6. \mathcal{LA} coordinates. (a) two human silhouettes with different postures (b) \mathcal{LA} coordinates of the left arms (c) \mathcal{LA} coordinates of the right arms.

Assuming that the arm is a single rigid body, possible coordinates should fall into a circle whose radius is l since the size information is already filtered out. One may think that quadrangles that lie on a constant angle line in the articulation space (such as any two quadrangles shown in Figure 5(b)) can not both belong to the same shape section and may come to a conclusion that the whole space is not utilized and the articulations lie on a 1-D manifold. This is not the case. Figure 7(d) shows three different postures of human arm in a single image consisting of two rigid body movements of arm (Figure 7(a)-(b)) and a case where a bending occurs (Figure 7(c)). The corresponding \mathcal{LA} coordinates of left arms (which are determined from the disconnected skeleton representations computed from extracted silhouettes) are given in Figure 7(e). Notice that due to initially ignored joints such as elbows, \mathcal{LA} coordinates of a shape section may not always lie on a circular arc.

Degenerate Cases

Representation with respect to \mathcal{LA} coordinate system deteriorates when the points defining the coordinate system are nearly colinear (see head section in Figure 8(a)). Even a very small change in the location of the fourth point leads to a significant change in the \mathcal{LA} coordinate. Another degenerate situation is encountered when the second indentation is not close enough (see leg sections in Figure 8(a)). In this case, the length of one of the vectors defining the coordinate

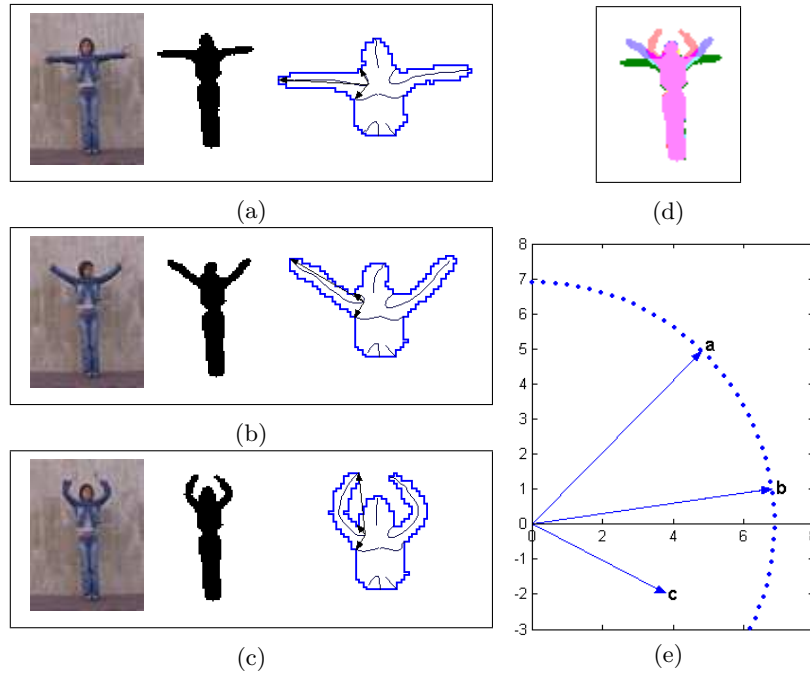


Fig. 7. Articulations and bendings in the articulation space. (a)-(c) three different postures of a human figure (taken from `ira_wave2` video sequence from action-silhouette database of [19]), the corresponding binary silhouettes and disconnected skeletons extracted from upper body portions (d) these three postures combined (e) \mathcal{LA} coordinate representations of left arms in the articulation space.

frame becomes too large and \mathcal{LA} coordinates fails to capture the variations of the tip of protrusion. Note that this latter degenerate case is a side effect of the symmetry extraction and may be alleviated by modifying it.

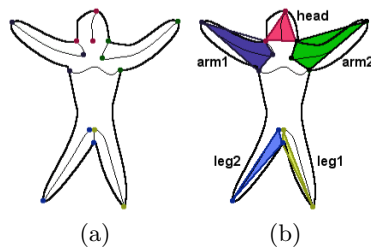


Fig. 8. Deformable sections of a human shape. (a) start and end points of protrusion and indentation branches (b) quadrangle or triangle representations of deformable sections.

In these degenerate situations, it is possible to define the local frame using only two points and the coordinate representation becomes equivalent to that of Bookstein used for analyzing landmark data [11]. At such object sections the set of quadrangles are replaced by a set of triangles.

5 Inferences in Articulation Space

Collection of possible postures defines a subset of the articulation space (static view) or a trajectory in the articulation space (dynamic view). In this section, we adopt the static view and discuss very basic inferences that can be made in the articulation space. We restrict our discussion to a set of twenty human shapes with different postures (Figure 9). The main reason for selecting this data set is that the sections as captured by protrusions and their movements are intuitive and one can judge relative closeness of two different postures (arms up/arms down). Secondly, human figure provides a rich data set since each figure contains five flexible sections to cover all possible situations that may arise in terms of degeneracies.

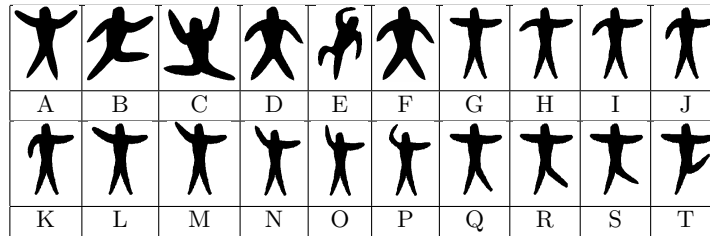


Fig. 9. Set of 20 human silhouettes used in the experiments.

Once the deformable shape sections are extracted from the training set and mapped to \mathcal{LA} coordinates, the distributions of the points can serve as prior knowledge about possible degrees of articulation in each section. These distributions are modeled as Gaussians.

The collected statistics about part articulations for the shape set $\mathcal{S}_1 = \{A, \dots, E\}$ is illustrated in Figure 10. Ellipses are drawn at 2σ . The largest ellipse corresponds to the distribution of *arm1* coordinates where the postural variability is the highest whereas the very small ellipse shown in the square window corresponds to the distribution of the *head* coordinates practically having no variability. Individual plots are provided in Figure 11 (note that the scales are not equal). One can observe that similar articulations in a part are represented with nearby points in the articulation space. For example, the articulations of *arm1* for shapes *B* and *D* and the articulations of *leg2* of shapes *A* and *B* are close to each other (Figure 11(b) and (e)).

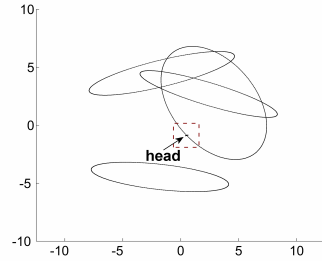


Fig. 10. Collected statistics of each part for shape set $\mathcal{S}_1 = \{A, \dots, E\}$ in the articulation space. The ellipses are drawn at 2σ . The largest one corresponds to *arm1* and the small dot corresponds to *head*.

When we consider only *arm1* and use the set $\mathcal{S}_2 = \{A, \dots, P\}$, the distribution of this articulated section becomes as shown in Figure 12. Table 1 shows the similarity of shapes based on the articulations of *arm1*. Similarity scores are computed with the function $sim(\mathbf{x}, \mathbf{y}) = \frac{1}{1 + \frac{d^2(\mathbf{x}, \mathbf{y})}{\theta^2}}$ where θ is the soft threshold and d is the Mahalanobis distance between \mathbf{x} and \mathbf{y} measured using the estimated covariance matrix. Observe that the similar configurations have relatively high similarity scores.

Table 1. Similarity scores of *arm1* for the shape set $\mathcal{S}_2 = \{A, \dots, P\}$.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
A	1.0000	0.1583	0.2487	0.1861	0.0842	0.2355	0.7549	0.4335	0.3373	0.1973	0.1253	0.9277	0.8243	0.6217	0.4677	0.1332
B	×	1.0000	0.1612	0.9448	0.1277	0.7966	0.2470	0.1802	0.4549	0.8999	0.8728	0.1603	0.1623	0.1442	0.1030	0.1128
C	×	×	1.0000	0.1917	0.2619	0.2118	0.3041	0.1238	0.1752	0.1708	0.1471	0.2025	0.3745	0.4680	0.3584	0.5967
D	×	×	×	1.0000	0.1385	0.9192	0.3004	0.1991	0.5188	0.9308	0.7655	0.1863	0.1939	0.1716	0.1197	0.1285
E	×	×	×	×	1.0000	0.1308	0.1060	0.0595	0.0914	0.1165	0.1399	0.0755	0.1064	0.1186	0.0993	0.4154
F	×	×	×	×	×	1.0000	0.3938	0.2502	0.6725	0.9272	0.5810	0.2370	0.2413	0.2076	0.1410	0.1331
G	×	×	×	×	×	×	1.0000	0.4316	0.5286	0.3164	0.1885	0.7068	0.7326	0.5577	0.3522	0.1556
H	×	×	×	×	×	×	×	1.0000	0.4787	0.2359	0.1365	0.5644	0.2938	0.2195	0.1701	0.0764
I	×	×	×	×	×	×	×	×	1.0000	0.6329	0.3148	0.3717	0.2943	0.2316	0.1590	0.1055
J	×	×	×	×	×	×	×	×	×	1.0000	0.6617	0.2031	0.1960	0.1686	0.1181	0.1128
K	×	×	×	×	×	×	×	×	×	×	1.0000	0.1252	0.1316	0.1205	0.0883	0.1108
L	×	×	×	×	×	×	×	×	×	×	×	1.0000	0.6504	0.4720	0.3605	0.1129
M	×	×	×	×	×	×	×	×	×	×	×	×	1.0000	0.8988	0.6382	0.1852
N	×	×	×	×	×	×	×	×	×	×	×	×	×	1.0000	0.7903	0.2287
O	×	×	×	×	×	×	×	×	×	×	×	×	×	×	1.0000	0.2028
P	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	1.0000

Next we consider the shape set $\mathcal{S}_3 = \{A, C, E, L, M, N, O, P\}$ containing only the shapes having their *arm1*s up (Figure 13(a)). In this particular case, the past experience is incomplete, therefore when a human shape whose *arm1* has a different posture is encountered, it must be considered as impossible. The Mahalanobis distances from *arm1* of shapes *J*(arm down) and *G* (horizontal arm) to the distribution reflect this fact with the values 5.0084 and 1.7907 respectively.

We can expand our knowledge about *arm1* by inserting the instance *F* where *arm1* is down. New distribution covers the cases where *arm1* is down (Figure13(b)). As expected, the distances of configurations for shapes *J* and *G* are reduced to 2.6379 and 0.9231 respectively.

Similar inferences are also valid for the degenerate cases. When the articulation distribution of *leg1* for the shape set $\mathcal{S}_4 = \{A, C, D, E, F, G, Q, R, S\}$ is considered, the articulation of shape *T* can be regarded as impossible (see Figure 14) since the Mahalanobis distance from it to the distribution is very high (6.2898) compared to the others.

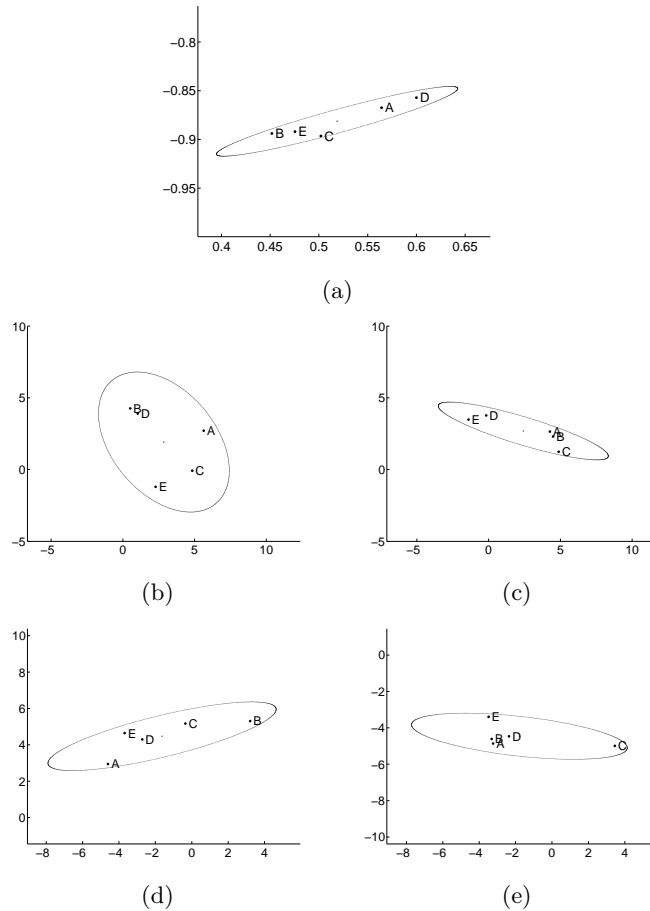


Fig. 11. For the shape set $\mathcal{S}_1 = \{A, \dots, E\}$, distributions of (a) *head* (b) *arm1* (c) *arm2* (d) *leg1* and (e) *leg2* in $\mathcal{L}\mathcal{A}$ coordinates.

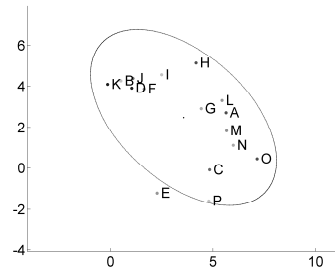


Fig. 12. Articulation distribution of *arm1* for the shape set $\mathcal{S}_2 = \{A, \dots, P\}$. Observe that in \mathcal{LA} coordinates *arm1* of shape *G* (straight arm posture) is close to $x = y$ line.

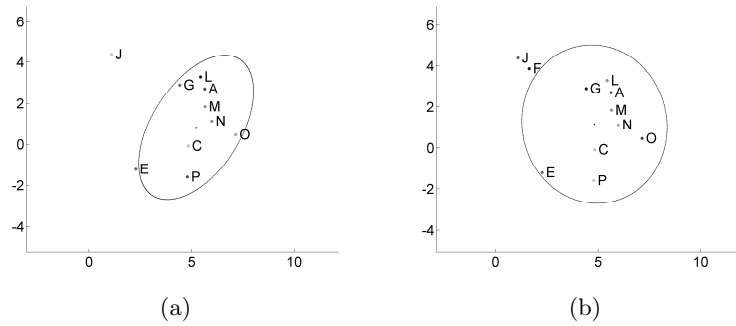


Fig. 13. Articulation distributions of *arm1* for the shape sets (a) $\mathcal{S}_3 = \{A, C, E, L, M, N, O, P\}$ (b) $\mathcal{S}'_3 = \{A, C, E, F, L, M, N, O, P\}$. Notice the change in the distribution when shape *F* (*arm1* down) is added.

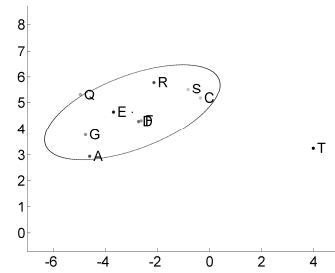


Fig. 14. Articulation distribution of *leg1* for the shape set $\mathcal{S}_4 = \{A, C, D, E, F, G, Q, R, S\}$. The articulation of shape *T* is far distant from the distribution.

5.1 An Application: Articulation Prior as a Feedback to Skeletal Matching

We now utilize the developed ideas to form a computational procedure that will serve as a feedback mechanism for shape recognition. The shape recognition framework presented in [8] matches skeletal branches of a reference shape and a query shape to determine a total matching score between two shapes. This score is computed by the weighted sum of the similarity scores of the matched branch pairs (the normalized length of the branches are used as weights).

We re-evaluate this matching score by taking into account the past experiences about observed articulations. Since movement of the tip of the protrusion branch in a shape section changes the \mathcal{LA} coordinate of that section, we use the similarity function values between shape sections as extra weights only for the positive symmetry branch pairs. By this way, the framework can distinguish a likely articulation from an unlikely one. In the experiments, we have evaluated the similarity between shapes with and without feedback on the recognition framework of [8].

Consider the matching between shapes A and T (Figure 15). In the syntactic level these two shapes are found to be similar with a score of 0.8259. However when we take into consideration the prior knowledge about articulations obtained from the set $\mathcal{S}_5 = \{A, \dots, S\}$, this score is reduced to 0.4578. This updated matching score reflects the difference in the posture of *leg1s* of the given shapes.

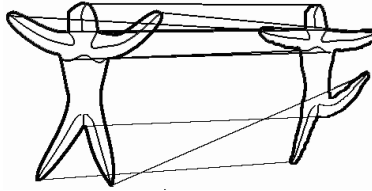


Fig. 15. Matching of two human silhouettes with matching scores 0.8259 without feedback and 0.4578 with feedback.

Figure 16 illustrates the effect of feedback within shape classification with some query examples when the prior experience is expressed with the shape set \mathcal{S}_5 . For each query five best matches with and without feedback are listed. See how the five best matches to shape G are re-ordered. Also notice the drastic change in the best match list of shape A .

6 Summary and Future Work

We defined articulation as what remains after shape information as well as global transformations are filtered out. Based on this notion we presented a 2D space in which similar articulations yield closer coordinates. Using illustrative examples




















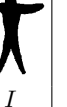

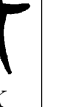
Query	5 best matches without feedback					5 best matches with feedback				
 A	 F 0.9749	 D 0.9621	 B 0.9396	 C 0.9172	 R 0.8899	 L 0.5815	 M 0.5591	 G 0.5504	 B 0.5318	 R 0.5152
 G	 M 0.9990	 I 0.9962	 K 0.9949	 J 0.9942	 L 0.9912	 M 0.9450	 L 0.9321	 I 0.9009	 J 0.8690	 K 0.8494

Fig. 16. Some query results with and without feedback.

we demonstrated that it is possible to build articulation priors and incorporate them to a skeletal matching scheme to arrive at an enriched skeletal representation.

Even though we adopted the coarsest structural definition in the form of limbs of a main body, articulation space analysis reveals initially ignored real joints such as an elbow. Hence in a dynamic view, collection of possible articulations can be viewed as a trajectory which may provide insight related to action. This issue needs further exploration.

Furthermore, our method do not require explicit extraction of deformable sections. Note that the sections are extracted automatically using the protrusion branches, and as such, they do not necessarily correspond to real parts. Consider an ellipse shape with two protrusions which do not actually represent deformable sections. Articulation space analysis for a collection of ellipses will reveal this fact.

Our method does not require perfect silhouettes, but it can not directly work on real images. However, due the connection of the underlying skeleton extraction to Ambrosio-Tortorelli approximation of Mumford-Shah segmentation functional [18] we plan to investigate the possibility of extending our method to real images.

Acknowledgments This work is partially supported by research grant TUBITAK 105E154 and TUBITAK-BAYG through PhD scholarships to Erkut Erdem and Aykut Erdem .

References

1. Ling, H., Jacobs, D.W.: Using the inner-distance for classification of articulated shapes. In: CVPR. (2005) 719–726
2. Basri, R., Costa, L., Geiger, D., Jacobs, D.: Determining the similarity of deformable objects. Vision Research **38** (1998) 2,365–2,385

3. Zhang, J., Siddiqi, K., Macrini, D., Shokoufandeh, A., Dickinson, S.: Retrieving articulated 3-d models using medial surfaces and their graph spectra. In: EMM-CVPR. (2005) 285–300
4. Zhu, S.C., Yuille, A.L.: Forms: a flexible object recognition and modeling system. *Int. J. Comput. Vision* **20** (1996) 187–212
5. Liu, T., Geiger, D.: Approximate tree matching and shape similarity. In: ICCV. (1998) 983–1001
6. Siddiqi, K., Shokoufandeh, A., Dickinson, S.J., Zucker, S.W.: Shock graphs and shape matching. *Int. J. Comput. Vision* **35** (1999) 13–32
7. Sebastian, T.B., Klein, P.N., Kimia, B.B.: Recognition of shapes by editing their shock graphs. *IEEE Trans. Pattern Anal. Mach. Intell.* **26** (2004) 550–571
8. Aslan, C., Tari, S.: An axis-based representation for recognition. In: ICCV. (2005) 1339–1346
9. Marr, D., Nishihara, H.K.: Representation and recognition of the spatial organization of three dimensional structure. In: *Proceedings of the Royal Society of London B. Volume 200.* (1978) 269–294
10. Kendall, D.G., Barden, D., Carne, T.K., Le, H.: *Shape and shape theory.* John Wiley and Sons (1999)
11. Bookstein, F.L.: *Morphometric tools for landmark data—geometry and biology.* Cambridge Univ. Press (1991)
12. Burl, M.C., Leung, T.K., Perona, P.: Recognition of planar object classes. In: CVPR. (1996) 223–230
13. Styner, M.: Combined boundary-medial shape description of variable biological objects. PhD thesis, Computer Science (2001)
14. Wu, Y., Lin, J., Huang, T.S.: Analyzing and capturing articulated hand motion in image sequences. *IEEE Trans. Pattern Anal. Mach. Intell.* **27** (2005) 1910–1922
15. Brand, M.: Shadow puppetry. In: ICCV. (1999) 1237–1244
16. Ambrosio, L., Tortorelli, V.: On the approximation of functionals depending on jumps by elliptic functionals via Γ -convergence. *Commun. Pure Appl. Math.* **43** (1990) 999–1036
17. Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. *Commun. Pure Appl. Math.* **42** (1989) 577–685
18. Tari, S., Shah, J., Pien, H.: Extraction of shape skeletons from grayscale images. *CVIU* **66** (1997) 133–146
19. Blank, M., Gorelick, L., Shechtman, E., Irani, M., Basri, R.: Actions as space-time shapes. In: ICCV. (2005) 1395–1402
20. Leyton, M.: A process-grammar for shape. *Artificial Intelligence* **34** (1988) 213–247
21. Kimia, B.B., Tannenbaum, A.R., Zucker, S.W.: Shapes, shocks, and deformations i: the components of two-dimensional shape and the reaction-diffusion space. *Int. J. Comput. Vision* **15** (1995) 189–224
22. Aslan, C.: Disconnected skeletons for shape recognition. Master’s thesis, Department of Computer Engineering, Middle East Technical University (2005)