Invariance to Affine-Permutation Distortions

Liang-Yan Gui  
Carnegie Mellon University  
lgui@andrew.cmu.edu

David A. Sepiashvili  
Independent Consulting  
dasepi@qanalyst.org

José M. F. Moura  
Carnegie Mellon University  
moura@andrew.cmu.edu

Abstract

An object imaged from various viewpoints appears very different. Hence, effective shape representation of objects becomes central in many applications of computer vision. We consider affine and permutation distortions. We derive the affine-permutation shape space that extends, to include permutation distortions, the affine only shape space (the Grassmannian). We compute the affine-permutation shape space metric, the sample mean of multiple shapes, the geodesic defined by two shapes, and a canonical representative for a shape equivalence class. We illustrate our approach in several applications including clustering and morphing of shapes of different objects along a geodesic path. The experimental results on key benchmark datasets demonstrate the effectiveness of our framework.

1. Introduction

Shape is an important characteristic of an object and shape analysis has wide applications in medical imaging [13, 12], document analysis [18, 17], neuroscience [10, 1], and many other computer vision problems [5]. A major challenge lies with the large distortions in imaged objects that result from wide variations in viewpoint. In many practical applications, these variations are well approximated by affine distortions [8], like when modeling a pinhole camera as an affine camera when the camera center is far from the object and the object is rigid. In this paper, we represent shapes by the collective of the landmarks on the boundary and in the interior of the image of an object, when available. We consider that the ordering of the pixels is unknown, and so they can be shuffled or permuted between different configurations of the same object; in other words, the images are obtained under affine and permutation distortions.

Our work develops an affine-permutation shape space that is invariant to both affine and permutation distortions. We define a shape similarity measure that is invariant to these distortions to quantify the difference between two affine-permutation distorted shapes. We propose to calculate this similarity measure, and we find a canonical representative for each distinct object. We use this canonical representative to solve for the geodesic connecting two shapes and the Karcher mean given multiple shapes—representing possibly different affine-permutation distorted objects.

2. Shape Space: A Group Perspective

We adopt here the group representations in [16, 6, 15].

Configuration Space. As described by Ha and Moura [6], the configuration of a rigid object consists of \( N \) landmarks (pixels) \( \{p_k\}_{k=1,...,N} \) on a 2D plane \( \mathbb{R}^2 \). Given a reference coordinate system, the location of point \( p_k \) is specified by a pair of coordinates \( (x_k, y_k) \). We represent the configuration by an \( N \times 2 \) matrix \( X \), the configuration matrix. The collection of all configurations of \( N \) points on a 2D image plane is defined as the configuration space \( \mathcal{X} \). Here, \( \mathcal{X} \) is the Euclidean space \( \mathbb{R}^{N \times 2} \). We exclude matrices with rank \( = 1 \) since they correspond to degenerate shapes, i.e., a single point or a straight line.

We consider the affine distorted configuration \( \hat{X} \) and the affine-permutation distorted configuration \( \tilde{X} \) of configuration \( X \). Two affine distorted configurations \( X \) and \( \tilde{X} \) belong to the same equivalence class \( [X]_A \). The affine shape of \( X \) is the corresponding equivalence class. The quotient space \( \mathcal{X}/A \), where \( A \) defines the affine group action [14], collects all possible equivalence classes and defines the affine shape space. Similarly, the equivalence class \( [X]_{AP} \) defines the affine-permutation shape of \( X \) under both affine and permutation distortions. The quotient space \( \mathcal{X}/AP \) defines the affine-permutation shape space.

Affine transformations account for unknown translation, reflection, rotation, scaling, and skewing distortions between two configurations. We sequentially factor out these different distortions.

Normalized Configuration Space. We consider translations, scaling, and skewing distortions and define normalized configurations that are invariant to these distortions. Given a configuration \( X \), its normalized configuration \( Y \) satisfies three conditions: (1) \( Y \) and \( X \) are in the same equivalence class (for affine distortion, \( Y \in [X]_A \); for affine-permutation distortion, \( Y \in [X]_{AP} \)); (2) The center of mass is at the origin: \( \text{mean}(Y) = 0 \), where \( \text{mean}(Y) \)
is defined as a $1 \times 2$ row vector containing the mean values of the columns of $Y$; (3) $Y$ is an orthonormal matrix: $Y^TY = I$, where $I$ is the $2 \times 2$ identity matrix.

The collection of the normalized configurations forms the normalized configuration space, i.e., a Stiefel manifold, denoted by $G\mathcal{F}(2,N)$ [4]. To find the normalized configuration of $X$, we first center it by removing translation $X_c = X - \text{mean}(X)$, where $1$ is an $N \times 1$ vector of ones, such that $\text{mean}(X_c) = 0$ where $0$ is a $1 \times 2$ zero vector, and $\otimes$ is the Kronecker product. We then Normalize it by computing the compact singular value decomposition (SVD) of $X_c$: $X_c = USVT$. The normalized configuration $Y$ is defined as $Y \equiv UV^T$. The normalized configuration $Y$ is invariant to translation, scaling, and skewing distortions. $Y$ is an $N \times 2$ orthonormal matrix, mapped to a point on the Stiefel manifold. Next, we construct the affine shape space and the affine-permutation shape space as quotient spaces of $G\mathcal{F}(2,N)$.

### Affine Shape Space

If two configurations $X$ and $\hat{X}$ are affine distortions of each other, their corresponding normalized configurations $Y$, $\hat{Y}$ are related by $\hat{Y} = YV$, where $V$ is a $2 \times 2$ orthogonal matrix. That is, after centering and normalizing steps, only orientation ambiguity remains. Let $O_2$ represent the orthogonal group consisting of the $2 \times 2$ orthogonal matrices (rotations and reflections) and define the cost of $O_2$ with respect to $Y$ as

$$[Y]_O = \{YV : V \in O_2\}. \quad (1)$$

The equivalence class $[Y]_O$ is invariant to orientation transformations since both $Y$ and $\hat{Y}$ belong to $[Y]_O$. The orthogonal equivalence relation denoted by $\equiv_O$ partitions the Stiefel manifold $G\mathcal{F}(2,N)$. The resulting quotient space $G\mathcal{F}(2,N)/O_2$ is the affine shape space $X/A$, which is the Grassmannian manifold $Gr_r(2,N)$ [7]. Each affine shape $[Y]_O$ is mapped to a point on $Gr_r(2,N)$. Any matrix belonging to the equivalence class $[Y]_O$ stores the point $p$ numerically and is a matrix representation of $p$.

### Affine-Permutation Shape Space

For the affine-permutation distortions, after the centering and normalizing steps, orientation and permutation ambiguities remain. Let $P_N$ be the set of $N \times N$ permutation matrices. $P_N$ together with matrix multiplication forms the permutation group $P_N$. The double coset of $Y$ by $O_2$ and $P_N$ is

$$[Y]_OP = \{PYV : Y \in G\mathcal{F}(2,N), P \in P_N, V \in O_2\}. \quad (2)$$

This double coset is invariant to the affine and permutation distortions; it is the affine-permutation shape $[X]_{AP}$. Any matrix belonging to the equivalence class $[Y]_OP$ is a matrix representation of this affine-permutation shape. We also present a canonical representative for each affine-permutation shape in Section 3. The quotient space of the affine shape space $Gr_r(2,N)$ by $P_N$ is the affine-permutation shape space. We use $GS(2,N)$ to represent the affine-permutation shape space.

### 3. Affine-Permutation Shape Space

#### Distance Definition between Two Affine-Permutation Shapes

The definition of distance refers to the geometry of the affine-permutation space. It should be invariant to both affine and permutation distortions. We first define a distance in the affine-permutation shape space $GS(2,N)$ [15] and then provide an efficient method for its computation.

**Definition 1.** The distance between two points $s_0$, $s_r \in GS(2,N)$, which are represented by $[Y_0]_OP$ and $[Y_r]_OP$, is defined by

$$d_{GS(2,N)}([Y_0]_OP, [Y_r]_OP) = \min_{P_0, P_r \in P_N} d_{Gr(2,N)}([Y_0]_OP, P_0 \cdot [Y_2]_OP), \quad (3)$$

$$= \min_{P \in P_N} d_{Gr(2,N)}([Y_0]_OP, P \cdot [Y_r]_OP). \quad (4)$$

where $P_0$ and $P_r$ are $N \times N$ permutation matrices from the permutation set $P_N$. The second equation is because $P_0$ and $P_r$ are not independent, and we fix one of them as reference without loss of generality.

#### Distance Computation

The definition of distance involves a combinatorial minimization problem over the set of $N$-dimensional permutations. It can be solved by

$$d_{GS(2,N)}([Y_0]_OP, [Y_r]_OP) = \min_{P \in P_N} \sqrt{2} \sqrt{\text{trace}(E_r^T)},$$

where $E_r$ is a diagonal matrix $E_r = \text{acosm}(C_r)$, and $C_r$ is obtained by an SVD $Y_0^T \cdot (PY_r) = Y_0C_rV_r^T$. Let cost function $J(P) = \text{trace}(E_r^T)$ and consider $P = \arg\min_{P \in P_N} J(P)$.

Since permutation matrices are doubly-stochastic and orthogonal, $P_N$ is a subset of the set of doubly stochastic matrices $D_N$. According to Birkhoff’s theorem [9], the optimum of a concave cost function over all doubly stochastic matrices $D \in D_N$ is obtained if and only if $D$ is a permutation matrix. So, we relax the constraint to $P = \arg\min_{P \in D_N} J(P)$. This problem is still challenging since the double stochastic matrix $P$ affects the three SVD components. Typical methods, i.e., finding the derivative, are intractable. However, an approximation to the distance is still feasible when casting it as a linear cost function. We approximate this distance by the Frobenius Norm:

$$d_{GS(2,N)}([Y_0]_OP, [Y_r]_OP) = \min_{P \in P_N, O \in O_2} \|Y_0 - P \cdot Y_r \cdot O\|_F. \quad (5)$$

Thus, we define a new approximate cost function

$$J_L(P) = \|Y_0 - P \cdot Y_r \cdot O\|_F = -\text{trace}(P \cdot (Y_r \cdot O)^T Y_0^T), \quad (6)$$

whose gradient is $\nabla J_L(P) = -Y_0 \cdot (Y_r \cdot O)^T$. To simplify the minimization, we further use the vectorization notations vec$(\cdot)$ to formulate a linear approximate cost function. By letting $w = \text{vec}(D)$, Eq. (6) and (3) become

$$J_L(w) = -\text{vec}(Y_0 \cdot (Y_r \cdot O)^T)^T \cdot w, \quad (7)$$

$$\nabla J_L(w) = -\text{vec}(Y_0 \cdot (Y_r \cdot O)^T). \quad (8)$$
As to the constraints, considering that a doubly-stochastic matrix is a square matrix with nonnegative numbers and that each of its rows and columns sums to 1, its vector form $w$ should satisfy the following conditions:

$$(1^T \otimes I) \cdot w = 1, (I \otimes 1^T) \cdot w = 1, w \geq 0,$$  \hspace{1cm} (9)

where $\otimes$ denotes the Kronecker product, $1$ is an $N \times 1$ vector, and $I$ is an $N \times N$ identity matrix.

Since two variable factors exist, orientation and permutation, the optimization problem can be solved by an alternating minimization. We fix the orientation matrix $O$ as one possible value first, then optimize the objective function (7) over permutation, and finally obtain the minimization of the distance in the current setting. Throughout the distances over different orientation matrices, we select the minimum as the desired distance $d_{\mathcal{GS}(2,N)}([Y_0]_{OP}, [Y_r]_{OP})$ together with the corresponding parameters of orientation $O$ and permutation vector $w$. We recover the permutation matrix $P$ from $w$. When applying this algorithm to two points in the shape space, we determine the orientation and permutation between them simultaneously.

**Canonical Representative in $\mathcal{GS}(2, N)$**. It is crucial to find a unique canonical representative for each point in the affine-permutation shape space $\mathcal{GS}(2, N)$. We start by fixing one element in one equivalence class and define the corresponding element in every other equivalence class as the canonical representative.

The procedure is the following: given two points $[Y_0]_{OP}$ and $[Y_r]_{OP}$ with their matrix representations $Y_0$ and $Y_r$, fix one as reference, such as $Y_0 = Y_0$; then apply the above algorithm to find the permutation $P$ and orientation distortion $O$ between them; and, finally, transform the other point $Y_r$, to its corresponding canonical representative

$$\hat{Y}_r = P \cdot Y_r \cdot O.$$  \hspace{1cm} (10)

With the canonical representative for the configurations, we can compute the Karcher mean of the affine-permutation shapes and compute the geodesic between two affine-permutation shapes.

### 4. Experimental Results

**Datasets**. We use three datasets [19, 11, 20] for our experiments, which are widely used in shape analysis: the UCI repository [19], MPEG-7 [11], and MCD [20].

**Canonical Representatives in Affine-Permutation Shape Space**. We find the canonical representative for an affine-permutation shape with the MCD dataset in Fig. 1. Each row illustrates an example. Each row of the left most column visualizes the matrix representation of a shape $[Y_0]_{OP}$ that is taken to be reference $\hat{Y}_0$. Each row of the second column visualizes one possible matrix representations $Y_r$ of a shape $[Y_r]_{OP}$. Each row of the third column shows the calculated distance, whose minimum is the affine-permutation distance between the shapes in the first two columns, under rotation or reflection by angles in $[-\pi, \pi]$. For a fixed angle $i$, $i \in [-\pi, \pi]$, the minimum of the distance across all possible permutations is considered to be the distance $d_{\mathcal{GS}_1}(i)$ in Eq. (5) if under a rotation transformation $R_1(i)$, or the distance $d_{\mathcal{GS}_2}(i)$ in Eq. (5) if under a reflection transformation $R_2(i)$. For all possible angles $i$, the minimum of the distance $\{d_{\mathcal{GS}_1}(i), d_{\mathcal{GS}_2}(i), i \in [-\pi, \pi]\}$ is considered to be the affine-permutation distance $d_{\mathcal{GS}(2,N)}([Y_0]_{OP}, [Y_r]_{OP})$ between these two affine-permutation shapes. The orientation matrix $O$ and the permutation matrix $P$ corresponding to the affine-permutation distance are then used to compute the canonical representative $\hat{Y}_r$ with Eq. (10). Each row of the fourth column contains the canonical representative $\hat{Y}_r$ of the shape in the second column $Y_r$ with the shape in the first column as reference $\hat{Y}_0$.

We now analyze in detail Fig. 1. The color of the landmarks indicates their scanning order. For example, the red landmark is the first scanned landmark in the landmark sequence and the blue landmark is the last in the sequence.

With respect to the first row, only affine and permutation distortions exist between Fig. 1(a) and (b), since we generated Fig. 1(b) by random affine and permutation transformations of Fig. 1(a). From Fig. 1(c), we observe that, when the rotation angle $-1.04$ rad, the distance is 0 under the appropriate permutation. Fig. 1(d) is recovered from Fig. 1(b) by Eq. (10) and is exactly the same as in Fig. 1(a). This proves that our method recovers one unique canonical representative for each affine-permutation shape. With respect
to the second row, Fig. 1(e) and (f) are projective distorted with respect to each other. We generated these two affine-permutation shapes from two samples of the same category “butterfly.” From Fig. 1(g), we see that, when the rotation angle 0.56 rad, the distance achieves its minimum 0.281, which is the affine-permutation distance between Fig. 1(e) and Fig. 1(f). Fig. 1(h) is the canonical representative of Fig. 1(f) obtained by Eq. (10) and is well aligned and appropriately oriented.

We generated Fig. 1(i), (j), (m), and (n) from two different categories, “fish” and “guitar.” Fig. 1(i) and (m) visualize the same configuration of “fish.” Fig. 1(j) and (n) are affine-permutation distorted versions of each other. For Fig. 1(i) and (j), Fig. 1(k) shows, when the rotation angle −2.442 rad, the distance achieves its minimum 0.479. Fig. 1(l) shows the canonical representative of Fig. 1(j) obtained with Eq. (10). For Fig. 1(m) and (n), Fig. 1(o) shows, when the reflection angle −0.06 rad, the distance achieves its minimum 0.479. Fig. 1(p) shows the canonical representative of Fig. 1(n) obtained with Eq. (10). Fig. 1(l) and Fig. 1(p) are exactly the same, which also proves that our method generates a unique canonical representative for an affine-permutation shape given a reference shape. The shapes of “fish” and “guitar” are visually quite different, but the landmarks are well permuted since the colors of Fig. 1(l) are ordered as the colors of Fig. 1(i), and similarly the colors of Fig. 1(p) are ordered as the colors of Fig. 1(m).

These examples show that our procedure of finding canonical representatives is effective.

**Application I: Clustering.** Clustering is a natural task to evaluate the distance and the Karcher mean. In centroid-based clustering, such as $K$-means or $K$-means++ [2] clustering, clusters are represented by a central vector, and the data memberships are determined by their distances to the cluster centers. We first solve the permutation distortions by finding the canonical representatives. We then extend $K$-means++ from the typical Euclidean space to affine-permutation shape space by replacing the corresponding concepts and calculations of the central vector and distance to the Karcher Mean and the distance in the affine-permutation shape space.

We tested the modified $K$-means++ on the UCI handwritten digit database. Each digit was resampled to $N = 100, 200, 300, 400, 500, 600$ landmarks. For our test, we assumed that the coordinates of the points are displaced by additive independent Gaussian noises. We arbitrarily selected one configuration from each class 0-9 and generated $M = 200, 400, 600, 800, 1000$ affine distortions that were disturbed by additive noise of different intensities. We generated $10 \times M$ images from 10 classes in total. We preprocessed the original configurations and obtained the canonical representatives as the input of $K$-means++. We use [3] as a baseline. This reference proposed different methods for the affine shape space construction and Karcher mean.

We compare our method and Algorithm [3] with respect to the accuracy in Fig. 2(a) and computational time in Fig. 2(b). Fig. 2 shows that our method achieves higher accuracy while being faster. This improved performance indicates that our constructed affine-permutation shape space identifies the intrinsic characteristics of the digital shapes.

**Application II: Shape Morphing on Labeled/Unlabeled Data.** Morphing produces a sequence of images that allows a smooth transition from one image (shape) into another, thus connecting different shapes. Given a starting shape and an ending shape, we find the canonical representative of the ending shape with respect to the starting shape with our method, and then use the linear solution to the logarithmic map to morph a shape into another. Morphing constructs the geodesic that bridges the two endpoints and obtains the intermediate shapes along this geodesic.

To be specific, we use in our experiments the UCI handwritten digit database, the MCD dataset, and the MPEG-7 database. With the starting shapes as references, we first permuted and rotated/reflect the end shapes to get their canonical representatives with our method and Eq. (10). We then morphed the starting shape into the ending shape. Fig. 3 demonstrates the results on three sets of shapes. The distance between any two adjacent intermediate shapes is $d/5$, given that $d$ is the distance between the two endpoints. For the top two rows, we plot the contours given by the boundary landmarks for better visualization. From Fig. 3, we see the smooth deformation along the geodesic between each pair of ending shapes, which proves our method and the linear solution are feasible.
References


