Statistical linear models in Procrustes shape space

Matias Bossa, Salvador Olmos

Statistical linear models in Procrustes shape space

Matias N. Bossa and Salvador Olmos

Communications Technologies Group (GTC), Aragon Institute of Engineering Research (I3A), Zaragoza University, Spain, {bossa,olmos}@unizar.es*

Abstract. The configuration matrix of a set of labeled landmarks is one the most used shape representations. However, it is well-known that the configuration matrix is not invariant under translation, scaling and rotation. This problem is revisited in this work where a local tangent shape space characterization at a reference shape is obtained as the null space of the subspace spanned by the reference shape and the set of translation and rotation generators. This local linear description of the shape space allows us to compute mean and variance of shapes as well as apply classical mutivariate statistical techniques such as Principal Component Analysis. Our proposal is compared with previous approaches, such as the seminal work [1] and more recents works [2] and [3].

1 Introduction

Shape analysis is concerned with the study of the geometrical descriptors that are invariant to position, size and orientation. Shape analysis has proven to be very useful in several tasks of computer vision and medical imaging, such as segmentation of anatomical structures or detecting and quantifying shape differences driven by pathology.

Several shape descriptors have been proposed in the literature. Many authors use a set of landmarks on the shape boundary as relevant geometric features [1, 2]. A recent comprehensive survey on shape analysis with landmarks can be found in [4]. However it is well known that the configuration matrix is not invariant under translation, scaling and rotation.

When using a set of labeled landmarks, one way to achieve a shape description invariant to position, size and orientation is by means of Procrustes alignment. A complete analysis of Procrustes shape space was done in [5, 6], where a deep understanding of the topology of Riemannian manifolds is used.

In this work we propose a procedure to obtain a local shape characterization, by using a tangent space projection and a Riemannian exponential mapping. This shape description allows us to apply classical multivariate statistical tools in shape space with accuracy. For example we applied Principal Component

* This work was funded by grants TEC2005-07801-C03:02 from CICYT, PI04/1795 and PI05/2006 from FIS, and PIP113/05 from DGA. The work of M. Bossa is supported by DGA under the FPI grant B097/2004.

Analysis (PCA) on several characterizations of shape space and compared their performance.

2 Shape characterization and distances

Shape is usually defined as the geometrical information of an object that is invariant under a similarity transformation, i.e. location, translation and scale [7]. In this work we will focus on a particular shape description, that is the set of k labeled points in $\mathbb{R}^m$. The coordinates of these points can be arranged in a $k \times m$ configuration matrix $\mathbf{X}$, or equivalently on a $km \times 1$ configuration vector $\mathbf{x} = \mathrm{vec}(\mathbf{X})$. We will use the notation $\mathbf{z}_i$ to denote the $k \times 1$ vector containing the ith column of any $k \times m$ matrix $\mathbf{Z}$, and $\mathbf{z}^j$ to denote different $km \times 1$ vectors indexed by j.

A configuration matrix $\mathbf{X}$ is not a proper shape descriptor, because it is not pose invariant. For any similarity transformation, i.e. $s \in \mathbb{R}^+$, $\mathbf{R} \in \mathbf{SO}(m)$ (the special orthogonal group) and $\mathbf{t} \in \mathbb{R}^m$, the configuration given by $s\mathbf{X}\mathbf{R} + \mathbf{1}_k\mathbf{t}^{\mathrm{T}}$ describes the same shape than $\mathbf{X}$, where $\mathbf{1}_k$ is the $k \times 1$ vector $[1\ 1\ \ldots\ 1]^{\mathrm{T}}$.

In order to get a shape descriptor invariant under scale and translation several constraints on $\mathbf{X}$ can be used. For example, if $\mathbf{X}$ is forced to be a unity norm matrix $\|\mathbf{X}\|^2 = \mathrm{tr}(\mathbf{X}\mathbf{X}^{\mathrm{T}}) = 1$ (or equivalently $\mathbf{x}^{\mathrm{T}}\mathbf{x} = 1$), and to have null centroid $\mathbf{1}_k^{\mathrm{T}}\mathbf{X} = \mathbf{0}_m^{\mathrm{T}}$, then scaling and translation effects are removed from $\mathbf{X}$.[1] As the set of unitary vectors $\mathbf{x}^{\mathrm{T}}\mathbf{x} = 1$ corresponds to the unit sphere $\mathbb{S}^{mk-1}$, the condition of having null centroid is equivalent to intersect $\mathbb{S}^{mk-1}$ with the null space of the matrix $[(\mathbf{1}_k^{\mathrm{T}}\ \mathbf{0}_k^{\mathrm{T}}\ \ldots\ \mathbf{0}_k^{\mathrm{T}})^{\mathrm{T}}, (\mathbf{0}_k^{\mathrm{T}}\ \mathbf{1}_k^{\mathrm{T}}\ \ldots\ \mathbf{0}_k^{\mathrm{T}})^{\mathrm{T}}, \ldots, (\mathbf{0}_k^{\mathrm{T}}\ \mathbf{0}_k^{\mathrm{T}}\ \ldots\ \mathbf{1}_k^{\mathrm{T}})^{\mathrm{T}}]$. The intersection corresponds again to a unit sphere, $\mathbb{S}^{m(k-1)-1}$, named pre-shape space S_m^k.

The shape space Σ_m^k is the set of equivalence classes of S_m^k under the action of $\mathbf{SO}(m)$. The mapping that takes a configuration vector $\mathbf{x} \in S_m^k$ to shape space is $x = \pi(\mathbf{x}) : S_m^k \to \Sigma_m^k$. All the elements in the equivalence class of $\mathbf{x}$, also called fiber, are the set $\{\pi^{-1}(\pi(\mathbf{x}))\}$.

The distance between shapes can be defined in several ways, depending on the problem at hand. The Riemannian distance in pre-shape space $\rho(\mathbf{x}, \mathbf{y})$ is the length of the shortest curve segment in $\mathbb{S}^{m(k-1)-1}$ that connects $\mathbf{x}$ and $\mathbf{y}$ (great-circle): $\rho(\mathbf{x}, \mathbf{y}) = 2\arcsin(\frac{1}{2}\|\mathbf{x} - \mathbf{y}\|)$. However, in this work we are interested in the Procrustes distance, i.e. the distance in shape space, defined as $d(\pi(\mathbf{x}), \pi(\mathbf{y})) = \inf_{\mathbf{R} \in \mathbf{SO}(m)} \rho(\mathbf{x}, \mathbf{R}\mathbf{y})$, with $\mathbf{x}, \mathbf{y} \in S_m^k$.

2.1 Local tangent parametrisation of shape space

Given a reference configuration vector $\boldsymbol{\mu} \in S_m^k$ and $\pi(\boldsymbol{\mu}) \in \Sigma_m^k$ its corresponding shape, shapes in a finite neighbourhood of $\pi(\boldsymbol{\mu})$ can be characterized by variations in S_m^k from $\boldsymbol{\mu}$.

[1] Another way to remove translation can be obtained by multiplying $\mathbf{X}$ by the $(k-1) \times k$ Helmert sub-matrix [2].

The tangent space at $\boldsymbol{\mu}$ of pre-shape space, $T_{\boldsymbol{\mu}}S_m^k$, is a linear space that can be used as a linear approximation, as well as a parametrization of S_m^k, via the exponential-logarithm mapping. We will see that a linear subspace of $T_{\boldsymbol{\mu}}S_m^k$ can be found that is also the tangent space at $\pi(\boldsymbol{\mu})$ of Σ_m^k.

Any vector $\mathbf{x} \in S_m^k$ close to the reference shape $\boldsymbol{\mu} \in S_m^k$, can be written as

$$\mathbf{x} = \cos(\theta)\boldsymbol{\mu} + \sin(\theta)\mathbf{v} \tag{1}$$

with $\mathbf{x}^{\mathrm{T}}\mathbf{x} = \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} = \mathbf{v}^{\mathrm{T}}\mathbf{v} = 1$, and $\boldsymbol{\mu}^{\mathrm{T}}\mathbf{v} = 0$. As $\boldsymbol{\mu}$ and $\mathbf{x}$ are centered, $\mathbf{x}^{\mathrm{T}}\mathbf{t}^i = \boldsymbol{\mu}^{\mathrm{T}}\mathbf{t}^i = 0$, where the pure translation vectors $\{\mathbf{t}^i\}_{i=1}^m$ are given by $\mathbf{t}_i^i = \mathbf{1}_k$ and $\mathbf{t}_j^i = \mathbf{0}_k$ for $i \neq j$. It can be checked that also $\mathbf{v}^{\mathrm{T}}\mathbf{t}^i = 0$.

The tangent space at $\boldsymbol{\mu}$, $T_{\boldsymbol{\mu}}S_m^k$, is formed by the set of vectors $\mathbf{u} = \theta\mathbf{v}$. In order to extend classical statistics on the shape space a mapping between the manifold and its tangent space that preserves distances and angles is required. The exponential map

$$\exp_{\boldsymbol{\mu}}(\mathbf{u}) = \cos(u)\boldsymbol{\mu} + \sin(u)\hat{\mathbf{u}} \tag{2}$$

with $u = \|\mathbf{u}\|$ and $\hat{\mathbf{u}} = \mathbf{u}/u$, generates geodesics with initial velocity $\mathbf{u}$, and accomplishes that u is the Riemannian distance between $\boldsymbol{\mu} = \exp_{\boldsymbol{\mu}}(\mathbf{0})$ and $\exp_{\boldsymbol{\mu}}(\mathbf{u})$, i.e. $\rho(\exp_{\boldsymbol{\mu}}(\mathbf{0}), \exp_{\boldsymbol{\mu}}(\mathbf{u})) = u$. The angle between geodesics is the angle between their corresponding initial velocities. The logarithm map is the inverse of the exponential map: $\log_{\boldsymbol{\mu}}(\mathbf{x}) = \mathbf{u}$, where $\hat{\mathbf{u}} = \mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\mu}^T\mathbf{x})/\|\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\mu}^T\mathbf{x})\|$ and $\|\mathbf{u}\| = 2\arcsin(1/2\|\mathbf{x} - \boldsymbol{\mu}\|)$, therefore:

$$\rho(\boldsymbol{\mu}, \mathbf{x}) = \|\log_{\boldsymbol{\mu}}(\mathbf{x})\| \tag{3}$$

In order to get rid of rotations, the variations from $\boldsymbol{\mu}$ should be orthogonal to the fibers $\pi^{-1}(\boldsymbol{\mu})$. Let $\boldsymbol{\Omega}^{i,j}(\theta) \in \mathbf{SO}(m)$, $i < j \leq m$, be a rotation matrix in the plane $\{i, j\}$ with rotation angle θ. It can be seen that the fiber $\mathbf{R}^{i,j}(\theta)\boldsymbol{\mu}$, with $\mathbf{R}^{i,j}(\theta) = (\mathbf{I}_k \otimes \boldsymbol{\Omega}^{i,j}(\theta))$, is a curve in S_m^k but not a geodesic for $m > 2$. It was shown in [5, 8, 9] that if one moves away from $\boldsymbol{\mu}$ along geodesics orthogonally to fibers (no matters fibers are not geodesics), a Riemannian submersion from S_m^k onto Σ_m^k is obtained. Roughly, this means that identifying parts of Σ_m^k with submanifolds of S_m^k orthogonal to fibers, is a smooth mapping with the same metric. This allows us to work in S_m^k as if it were Σ_m^k, in a neighbourhood of $\boldsymbol{\mu}$. The tangent directions to fibers $\mathbf{R}^{i,j}(\theta)\boldsymbol{\mu}$ at $\boldsymbol{\mu}$ are given by their derivatives $\mathbf{r}^{i,j}$: with $\mathbf{r}_i^{i,j} = \boldsymbol{\mu}_j$, $\mathbf{r}_j^{i,j} = -\boldsymbol{\mu}_i$, and $\mathbf{r}_l^{i,j} = \mathbf{0}_k$ for $i, j \neq l$. The same result can be found in [6], from a more theoretical point of view. Any configuration $\mathbf{x}$ Procrustes aligned to $\boldsymbol{\mu}$, fulfils the following linear constraints: $\mathbf{x}^{\mathrm{T}}\mathbf{r}^{i,j} = 0$. This is also true for $\boldsymbol{\mu}$, then $\mathbf{v}^{\mathrm{T}}\mathbf{r}^{i,j} = 0$.

Fig. 1 shows schematically the pre-shape space S_m^k as a 2-sphere. The fiber $\mathbf{R}^{i,j}(\theta)\boldsymbol{\mu}$ is a small circle labeled as $\mathbf{R}\boldsymbol{\mu}$, and its tangent direction as $\mathbf{r}$. The shape space Σ_m^k is obtained by intersecting the pre-shape sphere with a plane orthogonal to $\mathbf{r}$.

Summing up all the constraints on $\mathbf{v}$ can be written as $\mathbf{v}^{\mathrm{T}}\mathbf{N} = \mathbf{0}_{km}$, with $\mathbf{N} = (\boldsymbol{\mu}, \mathbf{t}^1, \mathbf{t}^2, \ldots, \mathbf{t}^m, \mathbf{r}^{1,2}, \mathbf{r}^{1,3}, \ldots, \mathbf{r}^{(m-1),m})$. The intersection of the pre-shape

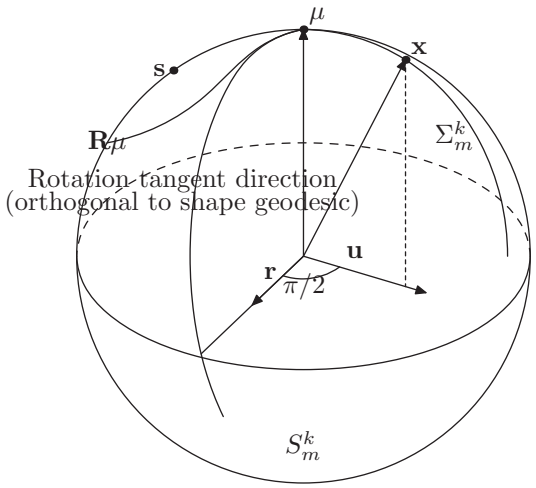

Fig. 1. Schematic representation of pre-shape space S_m^k. Shape space Σ_m^k is identified with the submanifold orthogonal to fiber $\mathbf{R}$ in a neighbourhood of reference shape .

sphere $\mathbb{S}^{m(k-1)-1}$ and the hyperplane defined by the null space of $\{\mathbf{r}^{i,j}\}$ is again a smaller dimension sphere, in this case $\mathbb{S}^{m(k-1-(m-1)/2)-1}$. Therefore, shape space, although being a manifold with singularities for $m > 2$, it behaves like a sphere in a finite neighbourhood of any nonsingular reference shape $\boldsymbol{\mu}$, with the metric defined in (3), i.e. $d(\pi(\mathbf{x}), \pi(\mathbf{y})) = \rho(\mathbf{x}, \mathbf{y})$.

We have explicitly removed the m degrees of freedom of translation, the $m(m-1)/2$ ones of rotation and the one of size. If $\mathbf{M}$ denotes the matrix containing a basis of the null space of $\mathbf{N}$, any shape can be written as

$$\mathbf{x} = \cos(b)\boldsymbol{\mu} + \sin(b)\mathbf{M}\hat{\mathbf{b}}, \tag{4}$$

with $\hat{\mathbf{b}} = \mathbf{b}/b$, $b = \|\mathbf{b}\|$ and $\mathbf{b}$ an arbitrary vector of $\mathbb{R}^{m(k-1-(m-1)/2)-1}$. The parameter vector $\mathbf{b}$ is a linear, non-redundant representation of shape space, that preserve distances to the reference shape $\boldsymbol{\mu}$ and angles between geodesics. Therefore it can be used as an approximate and nice characterization of the shape space. Note that $\mathbf{M}$ only depends on $\boldsymbol{\mu}$, and it is not required a training set to compute $\mathbf{M}$.

The above conditions are valid at least within a ball around $\boldsymbol{\mu}$ with radius smaller than the distance from $\boldsymbol{\mu}$ to the closest singularity, which is usually far away enough in real problems. The singularities (labeled as $\mathbf{s}$ in Fig. 1) arise when $\text{rank}(\mathbf{X}) < m-1$, . For $m = 3$, this means that all the landmarks are located over a line, and for $m = 2$ there are no singularities. Fig. 2 shows a simple example of a geodesic passing through a singularity. It can be seen on bottom left panel that shapes further away than a singularity accomplish $d(\pi(\mathbf{x}), \pi(\mathbf{y})) \neq \rho(\mathbf{x}, \mathbf{y})$.

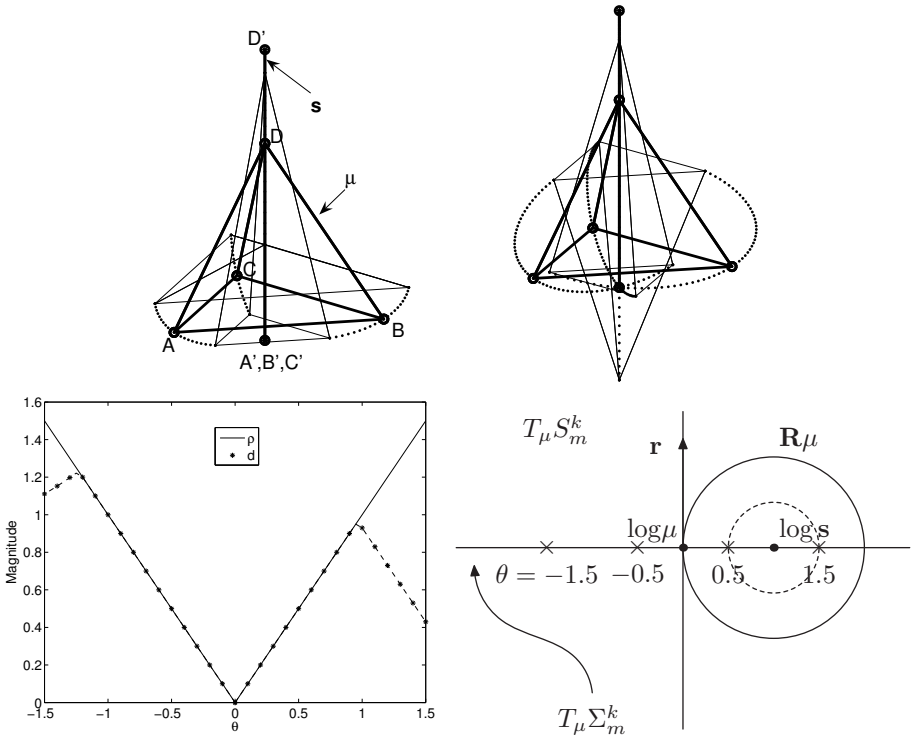

Fig. 2. Up: The reference shape (a tetrahedron) is denoted as and its closest singular shape as **s** (vertical line). The Procrustes distance between and **s** is about 1. The vertexes ABC collapse at A'B'C'. Geodesics from towards **s** are illustrated with dotted lines. On the left upper panel, θ runs in the interval (-0.5,0.5), and on the right (-1.5,1.5). Note that the configuration obtained for $\theta = 1.5$ corresponds to an inverted tetrahedron, which is not Procrustes aligned to . Bottom left: Values of the pre-shape Riemannian distance ρ (solid line) and Procrustes distance d (dashed line) from for running values of θ. Bottom right: schematic view of the tangent space T S_m^k.

3 Multivariate statistics on shape space

Classical multivariate statistical techniques such as PCA have been applied to some linear approximations of shape space. PCA involves estimating the mean and the modes of variation from a training set.

The mean $\boldsymbol{\mu}_E$ of a set of vectors $\mathbf{x}^n \in S_m^k$ is computed as an iterative process including averaging and alignment. At each iteration, classical averaging is performed on landmark coordinates and the vectors are aligned to the normalized average. The final result $\boldsymbol{\mu}_E$ is the normalized average after convergence. $\boldsymbol{\mu}_E$ provides the shape that minimizes the Euclidean distance $\sum_n \|\boldsymbol{\mu} - \mathbf{x}^n\|^2$, but not the Procrustes distance $\sum_n d^2(\boldsymbol{\mu}, \mathbf{x}^n)$ that we are interested in.

An alternative procedure to compute the mean of a set of elements in Riemmanian manifolds was proposed in [10]. The method consists on iteratively estimating the mean as $\boldsymbol{\mu}^{j+1} = \exp(1/n\Sigma_n \log_{\boldsymbol{\mu}^j}(\mathbf{x}^n))$. We used this method with $\boldsymbol{\mu}_E$ as initialization. This iterative procedure has a fast convergence.

Modes of variation are computed in the tangent space at $\boldsymbol{\mu}$, by means of SVD of the residues $\mathbf{u}^n = \log_{\boldsymbol{\mu}}(\mathbf{x}^n)$. The principal geodesic directions are given by the singular vectors $\mathbf{p}^k$. Singular vectors are collected in the matrix $\mathbf{P}$. The singular values σ^k account for their relative importance. The procedure of computing PCA in the tangent space of a Riemannian manifold is called Principal Geodesic Analysis (PGA) [10, 11]. Principal geodesics are $\mathbf{y}^k(b_k) = \exp_{\boldsymbol{\mu}}(\mathbf{p}^k b_k) = \cos(b_k)\boldsymbol{\mu} + \sin(b_k)\mathbf{p}^k$, where b_k measures the deviation from the mean. An arbitrary vector of the PGA model is:

$$\mathbf{y}(\mathbf{b}) = \exp_{\boldsymbol{\mu}}(\mathbf{P}\mathbf{b}) = \cos(b)\boldsymbol{\mu} + \sin(b)\mathbf{P}\hat{\mathbf{b}}, \tag{5}$$

where $\hat{\mathbf{b}} = \mathbf{b}/b$, and $b = \|\mathbf{b}\|$.

3.1 Discussion of previous approaches

Previous approaches proposed for PCA on shape space failed on finding the best linear approximation of the shape space. Cootes' proposal [1] is one of the most referenced approaches. Shape was modeled as

$$\mathbf{y}_C(\mathbf{b}) = \boldsymbol{\mu} + \mathbf{P}_C\mathbf{b}, \tag{6}$$

where $\mathbf{P}_C$ is the matrix with the principal components of the residues $\tilde{\mathbf{x}}_C^n = \mathbf{x}^n - \boldsymbol{\mu}$, and $\mathbf{b}$ is a parameter vector. However, the spherical topology of the shape space was disregarded. Shape model in (6), labeled as PCA_C in this work, has two drawbacks: firstly, shape instances $\mathbf{y}_C$ don't have unity norm. Secondly, the shape Procrustes distance between an instance and $\boldsymbol{\mu}$ is neither proportional to $\|\mathbf{b}\|$, nor symmetric under a sign change of $\mathbf{b}$, i.e. $d(\boldsymbol{\mu}, \mathbf{y}_C(\mathbf{b})) \neq d(\boldsymbol{\mu}, \mathbf{y}_C(-\mathbf{b}))$. The reason is that $\|\mathbf{P}_C^{\mathrm{T}}\boldsymbol{\mu}\| \neq 0$ because the residues are not orthogonal to the mean. Accordingly the maximum number of degrees of freedom is $m(k-1-(m-1)/2)$ (i.e. one dimension larger than actual Procrustes shape space). Splitting $\mathbf{b}$ into two orthogonal parts, $\mathbf{b} = \mathbf{b}_{\boldsymbol{\mu}} + \mathbf{b}_{\perp\boldsymbol{\mu}}$, such that $\boldsymbol{\mu}^{\mathrm{T}}(\mathbf{P}_C\mathbf{b}_{\perp\boldsymbol{\mu}}) = 0$, the asymmetry in norm can be shown as

$$\begin{aligned}\|\mathbf{y}_C(\mathbf{b})\| = \|\boldsymbol{\mu}(1+b_{\boldsymbol{\mu}}) + \mathbf{P}_C\mathbf{b}_{\perp\boldsymbol{\mu}}\| = \sqrt{(1+b_{\boldsymbol{\mu}})^2 + b_{\perp\boldsymbol{\mu}}^2} \neq \\ \neq \sqrt{(1-b_{\boldsymbol{\mu}})^2 + b_{\perp\boldsymbol{\mu}}^2} = \|\boldsymbol{\mu}(1-b_{\boldsymbol{\mu}}) - \mathbf{P}_C\mathbf{b}_{\perp\boldsymbol{\mu}}\| = \|\mathbf{y}_C(-\mathbf{b})\|,\end{aligned} \tag{7}$$

with $b_{\boldsymbol{\mu}} = \|\mathbf{b}_{\boldsymbol{\mu}}\|$ and $b_{\perp\boldsymbol{\mu}} = \|\mathbf{b}_{\perp\boldsymbol{\mu}}\|$. Similarly the Procrustes distance accomplishes

$$\begin{aligned}d(\boldsymbol{\mu}, \mathbf{y}_C(\mathbf{b})) = 2\arcsin\left(\frac{1}{2}\left\|\boldsymbol{\mu}\left(1 - \frac{1+b_{\boldsymbol{\mu}}}{\|\mathbf{y}_C(\mathbf{b})\|}\right) - \frac{\mathbf{P}_C\mathbf{b}_{\perp\boldsymbol{\mu}}}{\|\mathbf{y}_C(\mathbf{b})\|}\right\|\right) \neq \\ \neq 2\arcsin\left(\frac{1}{2}\left\|\boldsymbol{\mu}\left(1 - \frac{1-b_{\boldsymbol{\mu}}}{\|\mathbf{y}_C(-\mathbf{b})\|}\right) + \frac{\mathbf{P}_C\mathbf{b}_{\perp\boldsymbol{\mu}}}{\|\mathbf{y}_C(-\mathbf{b})\|}\right\|\right) = d(\boldsymbol{\mu}, \mathbf{y}_C(-\mathbf{b})).\end{aligned} \tag{8}$$

Later works, even realizing the spherical topology of shape space, made use of a linear projection instead of using the log mapping to project data on the tangent space.Therefore a distortion in the metric was introduced.

In [3] data is projected into tangent space by scaling $\mathbf{x}^n$ in such a way that the residue is perpendicular to $\boldsymbol{\mu}$, $\tilde{\mathbf{x}}_A^n = \mathbf{x}^n(\mathbf{x}^{n\mathrm{T}}\boldsymbol{\mu})^{-1} - \boldsymbol{\mu}$. In [2] the tangent coordinates were defined as the linear projection on the plane perpendicular to the mean, $\tilde{\mathbf{x}}_B^n = \mathbf{x}^n - \boldsymbol{\mu}(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}^n)$. The approximation done is $\theta \approx \tan\theta$ in the first case, while in the second is $\theta \approx \sin\theta$, being θ the Procrustes distance from $\boldsymbol{\mu}$ to $\mathbf{x}$. The corresponding shape models are

$$\mathbf{y}_A(\mathbf{b}) = \frac{\boldsymbol{\mu}}{\sqrt{1+b^2}} + \frac{\mathbf{P}_A\mathbf{b}}{\sqrt{1+b^2}} \tag{9}$$

$$\mathbf{y}_B(\mathbf{b}) = \boldsymbol{\mu}\sqrt{1-b^2} + \mathbf{P}_B\mathbf{b} \tag{10}$$

labeled as PCA_A and PCA_B respectively.

The space spanned by the principal vector is one dimension smaller than $\mathbf{P}_C$ as $\mathbf{P}_{\{A,B\}}^{\mathrm{T}}\boldsymbol{\mu} = \mathbf{0}$ and so $b_{\boldsymbol{\mu}} = 0$. Therefore, unlike the model in (6), these approaches are symmetric under a change of sign of the parameter $\mathbf{b}$. The Procrustes distance from the mean is

$$d(\boldsymbol{\mu}, \mathbf{y}_A(\mathbf{b})) = \| \arctan(b) \| = d(\boldsymbol{\mu}, \mathbf{y}_A(-\mathbf{b})) \tag{11}$$

$$d(\boldsymbol{\mu}, \mathbf{y}_B(\mathbf{b})) = \| \arcsin(b) \| = d(\boldsymbol{\mu}, \mathbf{y}_B(-\mathbf{b})). \tag{12}$$

Left panel of Fig. 3 illustrates the Procrustes distance from the mean shape versus signed b. A value of $b_{\boldsymbol{\mu}}/b$ of 0.5 was used in the PCA_C model. PGA is the only model for which a linear change of the model parameters implies a linear variation of the Procrustes distance. On one hand PCA_A model generates instances closer to the mean than PGA, especially for large values of b, which means that PCA_A is more sensitive to outliers. PCA_C model has a similar behaviour for $b_{\boldsymbol{\mu}}/b$ small. On the other hand, PCA_B generates instances further away of the mean than PGA in the valid running interval $b \in [-1, 1]$.

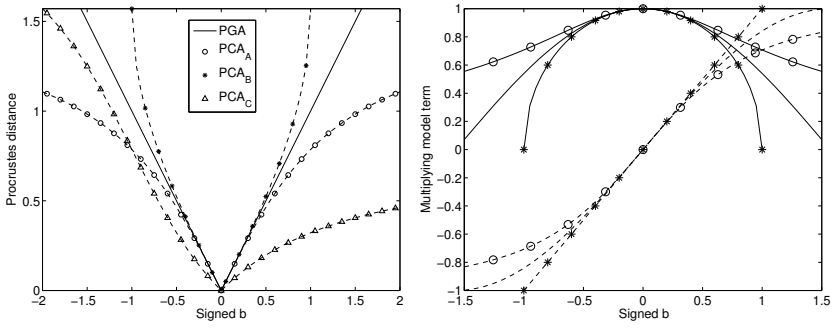

Fig. 3. Left: Procrustes distance versus signed b. Right: Model scaling terms versus b. Solid lines for mean terms and dashed lines for modes of variation terms.

There is a similarity between PGA, PCA_A and PCA_B model equations (5), (9) and (10). Mean and variation modes are orthogonal vectors multiplied by nonlinear scalar functions. The corresponding Taylor expansions of the mode of variation term differ at the 3rd order on b, being $b - 1/6b^3$ for PGA, $b - 1/2b^3$ for PCA_A and b for PCA_B. The model PCA_A underestimates the variation term compared to PGA, while the model PCA_B overestimates it, which is in agreement with the behaviour explained before on the Procrustes distance. The Taylor expansions of the mean term are different at the 4th order: $1 - 1/2x^2 + 1/24x^4$ for PGA, $1 - 1/2x^2 + 3/8x^4$ for PCA_A and $1 - 1/2x^2 - 1/8x^4$ for PGA_B. Right panel in Fig. 3 illustrates both terms for these models.

The Procrustes distance is always smaller than or equal to $\pi/2$, but it is much smaller in most of the real cases. Differences between PGA, PCA_A and PGA_B are then very small for this range of distances.

4 Illustrative example

In order to provide a simple illustrative example of the differences between previous approaches, 100 instances were randomly generated by adding noise to a reference shape (a cube in this simulation with $k = 8$). The standard deviation of the noise was set to 30% of the edge length.

Statistical shape models considered before were computed from the simulated instances. Mean shape $\boldsymbol{\mu}$ as well as the first two variation modes are illustrated in Fig. 4. There were no visible differences in the mean shape, but significant differences were found in the variation modes. The second mode of variation of PCA_C (upper right panel in Fig. 4) has a very different direction compared to PGA because this mode includes a significant part of the mean shape, introducing a relevant radial component. In contrast, PGA and $PCA_{\{A,B\}}$ do not suffer this drawback.

It can be seen in Fig. 4 that the length of PCA_A trajectories as well as PCA_C after normalization , are shorter than PGA, while PCA_B trajectories are longer. This is in agreement with the results shown in the left panel of Fig. 3. It is worthy to note that a large value of b was required to obtain visible differences.

The Procrustes distance between instances and mean shape $\boldsymbol{\mu}$ were computed for several values of the model parameters and are shown in Fig. 5. The profile of the curves in Fig. 5 resembles the theoretical ones in Fig. 3. The Procrustes distance in the second mode of variation of PCA_C is highly asymmetric due to the presence of the mean shape in that mode.

PGA is the only approach where a linear change of the model parameters involves a linear variation of the Procrustes distance from the mean. This property is really crucial for any statistical analysis performed on the linear space of the parameters, such as hypothesis testing, clustering, classification.

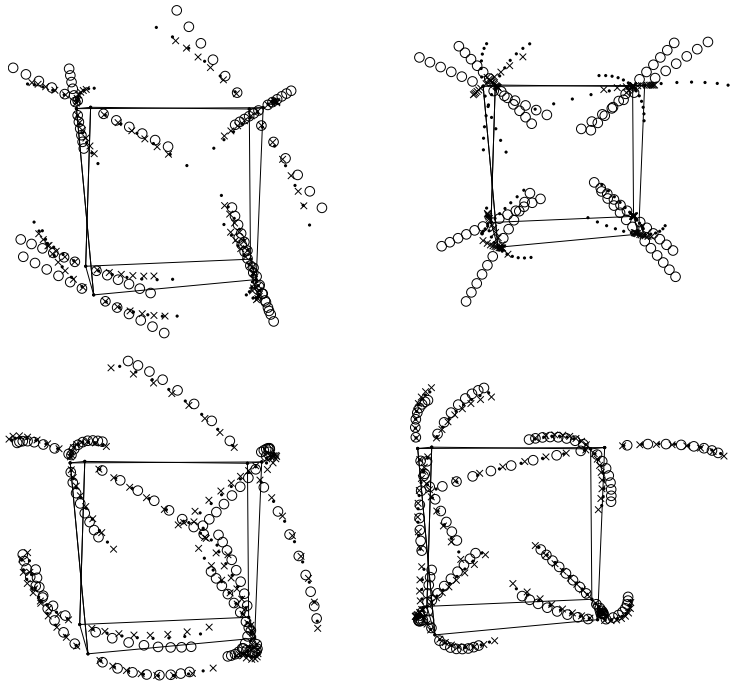

Fig. 4. Mean shape (solid line) and the first/second (left/right), modes of variation. PGA model is illustrated with dotted trajectories of geodesics. Top: Model parameter $b_{\{1,2\}}$ runs in the interval $(-3\sigma_{\{1,2\}}, 3\sigma_{\{1,2\}})$. Circles denote PCA_C straight line trajectories and crosses their corresponding normalized instances. Bottom: Model parameter $b_{\{1,2\}}$ runs in the interval $(-4\sigma_{\{1,2\}}, 4\sigma_{\{1,2\}})$. Circles denote PCA_A and crosses PCA_B trajectories.

5 Conclusions

The problem of shape description by means of a set of landmarks was revisited in this work. A local tangent shape characterization was obtained in earlier works by removing the degrees of freedom corresponding to scaling, translation and rotation. The main added value of this work is twofold: firstly, to propose a simple and intuitive way to understand this shape characterization; secondly, to apply classical multivariate statistics, such as PCA, on the shape space. The obtained shape representation is linear, non-redundant and preserves the Procrustes distance between instances and the reference shape, and also preserves angles between geodesics. Some of these properties do not hold in alternative models previously proposed in the literature. Additionally, these benefits are obtained with a minimum extra modelling effort and the formulation of the model resembles the classical ones.

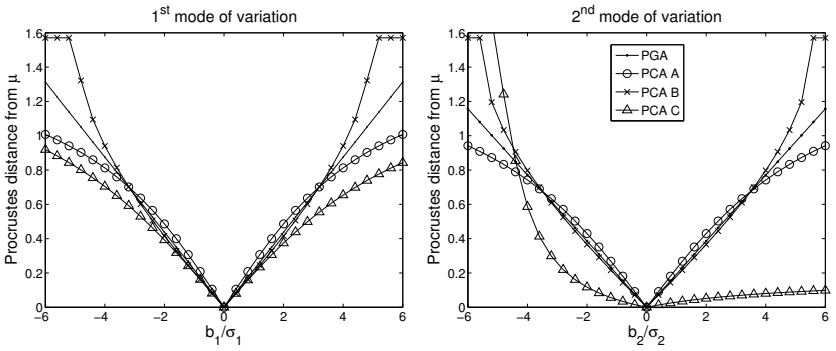

Fig. 5. Procrustes distance between generated instances $\mathbf{y}$ and reference shape .

References

1. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. Computer Vision and Image Understanding **61**(1) (1995) 38–59
2. Dryden, I., Mardia, K.: Statistical shape analysis. John Wiley (1998)
3. Fletcher, P.: Statistical Variability in Nonlinear Spaces: Application to Shape Analysis and DT-MRI. PhD thesis, Department of Computer Science, University of North Carolina (2004)
4. Heo, G., Small, C.G.: Form represention and means for landmarks: A survey and comparative study. Computer Vision and Image Understanding **102**(2) (2006) 188–203
5. Le, H., Kendall, D.: The Riemannian Structure of Euclidean Shape Spaces: A Novel Environment for Statistics. Annals of Statistics **21**(3) (1993) 1225–1271
6. Kent, J.T., Mardia, K.V.: Shape, Procrustes tangent projections and bilateral symmetry. Biometrika **88**(2) (2001) 469–485
7. Kendall, D.G.: The diffusion of shape. Adv. in Appl Prob **9** (1977) 428–430
8. Le, H.: On geodesics in euclidean shape space. J. London Math. Soc. **44**(2) (1991) 360–372
9. Le, H.: Unrolling Shape Curves. Journal of the London Mathematical Society **68**(2) (2003) 511–526
10. Pennec, X.: Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements. In Cetin, A., Akarun, L., Ertuzun, A., Gurcan, M., Yardimci, Y., eds.: Proc. of NSIP 1999. Volume 1., June 20-23, Antalya, Turkey, IEEE-EURASIP (1999) 194–198
11. Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Trans. Med. Imaging **23**(8) (2004) 995–1005