Statistical shape analysis

Pietro Gori

Enseignant-chercheur
Equipe IMAGES - Télécom Paris
pietro.gori@telecom-paris.fr

Summary

Statistical shape analysis

Definition

Statistical shape analysis deals with the study of the geometrical properties of a set of shapes using statistical methods. It is based on:

- Define a **computational model** to mathematically represent an object
- Define a **metric** (i.e. distance) between shapes
- Estimate the **mean shape** of a set of objects
- Estimate the shape **variability** of an ensemble of objects

Main applications

- Quantify shape differences between two groups of objects (i.e. healthy and pathological anatomical structures)
- Estimate the number of clusters within a set of objects
- Estimate of an average object, usually called atlas or template, that is used to compare different groups of objects

Introduction

Definition: shape

Shape is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object [1]

Computational models

Several computational models exist in the Literature to mathematically represent the geometry of anatomical structures:

- landmarks
- cloud of points
- fourier series
- m-reps
- currents
- varifolds ...

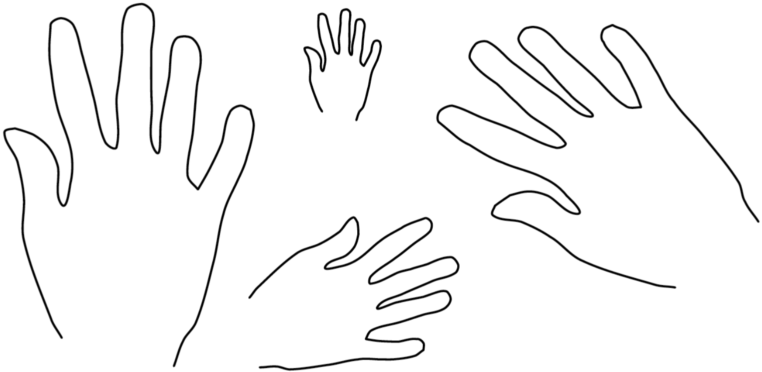


Figure 1: Four objects representing a hand with the same shape. Taken from [5]

Introduction

- In this lecture, we will focus on anatomical labelled (i.e. ordered) landmarks
- Given a set of N anatomical structures $\{S_i\}_{i=1,...,N}$ each one of them labelled with a configuration of M ordered landmarks $X_i = [x_{i1}^T; x_{i2}^T; ...; x_{iM}^T]$ where, $x_{ij}^T \in \mathbb{R}^2$ is the j-th landmark of the i-th structure, we aim at estimate the **average shape** $\bar{X}$ of the group and its **shape variability**.

Introduction

- The first step of our analysis is to remove the “location, scale and rotational effects” from the mathematical representations of our objects.
- In this way, the configurations of landmarks will describe the shape of each object
- In order to do that, we use a technique called **Generalized Procrustes Analysis (GPA)**

Summary

Generalized Procrustes Analysis (GPA)

- GPA involves translating, rotating and uniformly scaling every configuration in order to superimpose (i.e. align) all configurations among each other. This means minimizing:

$$
\begin{aligned}
s_i^*, R_i^*, t_i^* &= \underset{s_i, R_i, t_i}{\arg\min} \frac{1}{N} \sum_{i=1}^{N} \sum_{k=i+1}^{N} ||(s_i X_i R_i + \mathbb{1}_M t_i^T) \\
&\quad - (s_k X_k R_k + \mathbb{1}_M t_k^T)||_F = \qquad (1) \\
&= \underset{s_i, R_i, t_i}{\arg\min} \sum_{i=1}^{N} ||(s_i X_i R_i + \mathbb{1}_M t_i^T) - \bar{X}||_F
\end{aligned}
$$

- where $\bar{X} = \frac{1}{N} \sum_{k=1}^{N} (s_k X_k R_k + \mathbb{1}_M t_k^T)$, $\mathbb{1}_M$ is a column vector $[M \times 1]$ of ones, s_i is a scalar, R_i is a rotation (orthogonal) matrix $[2 \times 2]$ and t_i is a translation vector $[2 \times 1]$.

Generalized Procrustes Analysis (GPA)

- GPA involves translating, rotating and uniformly scaling every configuration in order to superimpose (i.e. align) all configurations among each other. This means minimizing:

$$s_i^*, R_i^*, t_i^* = \underset{s_i, R_i, t_i}{\arg\min} \sum_{i=1}^{N} ||(s_i X_i R_i + \mathbb{1}_M t_i^T) - \bar{X}||_F \tag{2}$$

- **IMPORTANT**: We are not interested in the values of the parameters s_i^*, R_i^*, t_i^*. They are considered as nuisance parameters. We are interested in the **Procrustes residuals**: $r_i = (s_i X_i R_i + \mathbb{1}_M t_i^T) - \bar{X}$. They are used to analyse differences in shape.

Generalized Procrustes Analysis (GPA)

$$s_i^*, R_i^*, t_i^* = \underset{s_i, R_i, t_i}{\arg\min} \sum_{i=1}^{N} ||(s_i X_i R_i + \mathbb{1}_M t_i^T) - \bar{X}||_F \tag{3}$$

- Note that, if you do not impose any constraints, there might be a trivial solution. Do you see it ?

Generalized Procrustes Analysis (GPA)

$$s_i^*, R_i^*, t_i^* = \underset{s_i, R_i, t_i}{\arg\min} \sum_{i=1}^{N} ||(s_i X_i R_i + \mathbb{1}_M t_i^T) - \bar{X}||_F \quad (3)$$

- Note that, if you do not impose any constraints, there might be a trivial solution. Do you see it ?
- Focus on the s_i, what happens if all s_i are close to 0 ?

Generalized Procrustes Analysis (GPA)

$$s_i^*, R_i^*, t_i^* = \underset{s_i, R_i, t_i}{\arg\min} \sum_{i=1}^{N} ||(s_i X_i R_i + \mathbb{1}_M t_i^T) - \bar{X}||_F \tag{3}$$

- Note that, if you do not impose any constraints, there might be a trivial solution. Do you see it ?
- Focus on the s_i, what happens if all s_i are close to 0 ?
- A possible (and popular) solution is to constraint the centroid size of the average configuration of landmarks $S(\bar{X}) = \sqrt{\sum_{k=1}^{M} \sum_{d=1}^{2} (x_{kd} - \bar{x}_d)^2} = 1$ where x_{kd} is the (k, d)th entry of $\bar{X}$ and $\bar{x}_d = \frac{1}{M} \sum_{k=1}^{M} x_{kd}$ which is equal to $||C\bar{X}||_F$ with $C = I_M - \frac{1}{M} \mathbb{1}_M \mathbb{1}_M^T$

Generalized Procrustes Analysis (GPA)

- GPA can also be embedded in a Gaussian (generative) model. We assume that:

$$X_i = \alpha_i(\bar{X} + E_i)\Omega_i + \mathbb{1}_M \omega_i^T \tag{4}$$

- where $\text{vec}(E_i) \sim \mathcal{N}(0, \mathbb{I}_{2M})$. Now, calling $(\alpha_i, \Omega_i, \omega_i) = (\frac{1}{s_i}, R_i^T, -\frac{1}{s_i} R_i t_i)$ we can rewrite the previous equation as:

$$\bar{X} + E_i = s_i X_i R_i + \mathbb{1}_M t_i^T \tag{5}$$

- Thus, considering s_i, R_i and t_i as nuisance and non-random variables, it follows that: $\text{vec}(s_i X_i R_i + \mathbb{1}_M t_i^T) \sim \mathcal{N}(\bar{X}, \mathbb{I}_{2M})$. It can be shown that a Maximum Likelihood estimation is equivalent to Eq. 1.

Generalized Procrustes Analysis (GPA)

$$s_i^*, R_i^*, t_i^* = \underset{s_i, R_i, t_i}{\arg\min} \sum_{i=1}^{N} ||(s_i X_i R_i + \mathbb{1}_M t_i^T) - \bar{X}||_F \qquad (6)$$

- How do we minimize this cost function ? We can use an iterative method where we alternate the estimation of s_i^*, R_i^*, t_i^* and $\bar{X}$.
  1. Choose an initial estimate $\bar{X}_0$ of the mean configuration and normalize $\bar{X}_0$ such that $S(\bar{X}_0) = 1$
  2. Align all configurations X_i to the mean configuration $\bar{X}_0$
  3. Re-estimate the mean of the configurations $\bar{X}_1$
  4. Align $\bar{X}_1$ to $\bar{X}_0$ and normalize $\bar{X}_1$ such that $S(\bar{X}_1) = 1$
  5. If $\sqrt{||\bar{X}_0 - \bar{X}_1||_F} \geq \tau$ set $\bar{X}_0 = \bar{X}_1$ and return to step 2
- A usual pre-processing is to translate each configuration X_i such that its centroid is equal to 0

Alignment of two shapes (Procrustes superimposition)

- We assume that all configurations X_i have been centred (i.e. $x_{ij} = x_{ij} - \frac{1}{M}\sum_{j=1}^{M} x_{ij} = x_{ij} - \bar{x}_i$)
- Remember that $t_i^* = \bar{x} - \frac{1}{M}\sum_{j=1}^{M} s^* R^* x_{ij} = \frac{1}{M}(\bar{X}\mathbb{1}_M) - \frac{1}{M}(s_i^* R_i^{*T} X_i^T \mathbb{1}_M)$, where $\bar{x}$ is the centroid of $\bar{X}$. Thus, if all configurations have been previously centred, all t^* are equal to 0.
- For each configuration i, we need to minimize the cost function: $\arg\min_{s,R} ||sXR - \bar{X}||_F$. From the previous lecture, it follows that:

$$
\begin{aligned}
R^* &= USV^T \\
s^* &= \frac{\langle R, X^T\bar{X}\rangle_F}{||X||_F} = \frac{\mathrm{Tr}(S\Sigma)}{||X||_F}
\end{aligned} \tag{7}
$$

- where we employ the SVD decomposition $X^T\bar{X} = U\Sigma V^T$ and $S = \begin{bmatrix} 1 & 0 \\ 0 & \det(UV^T) \end{bmatrix}$

Tangent space projection

- One could use the Procrustes residuals $r_i = (s_i X_i R_i + \mathbb{1}_M t_i^T) - \bar{X}$ to describe the shape of each configuration X_i with respect to the average (reference or consensus) $\bar{X}$
- However, after alignment and normalization, all configurations lie on a $2M$-dimensional hyper sphere. The actual curved distance between two configurations ρ is not the linear distance $D\rho$ used to calculate the Procrustes residuals

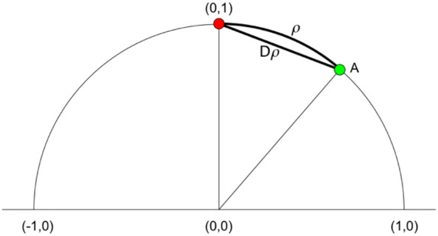


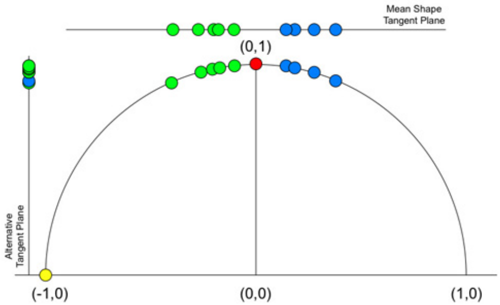
Figure taken from [6]

Tangent space projection

- The shape space is a curved manifold. We should use geodesic distances and not Euclidean distances.
- Another solution is to project all configurations onto an hyper plane that is **tangent** to the hyper sphere at a point. In this way, we can use the (linear) Euclidean distances on the hyper plane and not the true geodesic distances on the hyper sphere.
- Which point should we choose ?

Tangent space projection

- The shape space is a curved manifold. We should use geodesic distances and not Euclidean distances.
- Another solution is to project all configurations onto an hyper plane that is **tangent** to the hyper sphere at a point. In this way, we can use the (linear) Euclidean distances on the hyper plane and not the true geodesic distances on the hyper sphere.
- Which point should we choose ? The one that reduces the distortion of the projection: **the mean shape** !



Stereographic tangent space projection

- There are several projection schemes. Here, we will describe the stereographic one. In any case, the closer the configurations to the mean shape, the smaller the distortions.

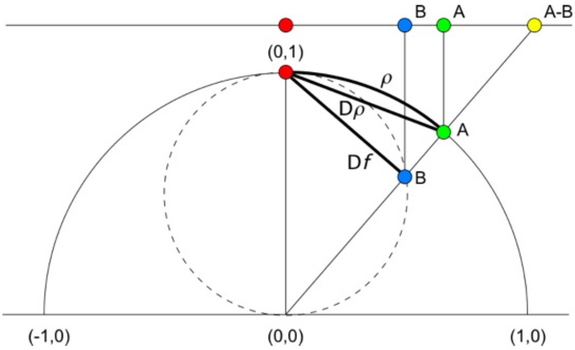


Figure 3: Different kind of projection. Yellow point A-B represents the stereographic projection. Figure taken from [6]

Stereographic tangent space projection

- $X = \{x_p, y_p\}$ is vectorized as $\boldsymbol{x} = [x_1, ..., x_M, y_1, ..., y_M]^T$
- We notice that: $|\boldsymbol{x}_t| \cos(\theta) = |\boldsymbol{x}_t| \frac{\langle \boldsymbol{x}_t, \bar{\boldsymbol{x}} \rangle_2}{|\boldsymbol{x}_t||\bar{\boldsymbol{x}}|} = |\bar{\boldsymbol{x}}| \rightarrow \langle \boldsymbol{x}_t, \bar{\boldsymbol{x}} \rangle_2 = |\bar{\boldsymbol{x}}|^2$
- Calling $\boldsymbol{x}_t = \alpha \boldsymbol{x}$ we can rewrite: $\langle \boldsymbol{x}_t, \bar{\boldsymbol{x}} \rangle_2 = \langle \alpha \boldsymbol{x}, \bar{\boldsymbol{x}} \rangle_2 = |\bar{\boldsymbol{x}}|^2$ and thus $\alpha = \frac{|\bar{\boldsymbol{x}}|^2}{\langle \boldsymbol{x}, \bar{\boldsymbol{x}} \rangle_2}$
- It follows that $\boldsymbol{x}_t = \alpha \boldsymbol{x} = \frac{|\bar{\boldsymbol{x}}|^2}{\langle \boldsymbol{x}, \bar{\boldsymbol{x}} \rangle_2} \boldsymbol{x}$

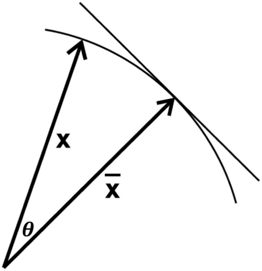


Figure 4: The vectors $\boldsymbol{x}$ and $\boldsymbol{x}_t$ are the configuration vectors respectively before and after projection. Figure modified from [5]

Generalized Procrustes Analysis (GPA)

$$s_i^*, R_i^*, t_i^* = \underset{s_i, R_i, t_i}{\arg\min} \sum_{i=1}^{N} ||(s_i X_i R_i + \mathbb{1}_M t_i^T) - \bar{X}||_F \tag{8}$$

1. Translate each configuration X_i such that its centroid is equal to 0
2. Choose an initial estimate $\bar{X}_0$ of the mean configuration (e.g. any configuration of the population) and normalize $\bar{X}_0$ such that $S(\bar{X}_0) = 1$
3. Align all configurations X_i to the mean configuration $\bar{X}_0$
4. **Project all configurations X_i into the tangent space**
5. Re-estimate the mean of the configurations $\bar{X}_1$
6. Align $\bar{X}_1$ to $\bar{X}_0$ and normalize $\bar{X}_1$ such that $S(\bar{X}_1) = 1$
7. If $\sqrt{||\bar{X}_0 - \bar{X}_1||_F} \geq \tau$ set $\bar{X}_0 = \bar{X}_1$ and return to step 3

Summary

Shape variations

- Once all configurations have been aligned to a common coordinate frame filtering out similarity transformations, they represent the shape of each structure
- We have already seen how to measure the average shape, what about shape variability ?

Shape variations

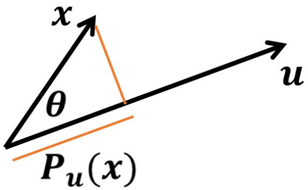
- Once all configurations have been aligned to a common coordinate frame filtering out similarity transformations, they represent the shape of each structure
- We have already seen how to measure the average shape, what about shape variability ?
- We could use **Principal Component Analysis (PCA)** onto the vectorized ($\boldsymbol{x} = [x_1, ..., x_M, y_1, ..., y_M]^T$) and aligned data to find a (small) set of orthonormal directions that explain most of the shape variability

PCA

Definition (Hotelling 1933)

PCA is an orthogonal projection of the data onto a low-dimensional linear space such that the variance of the (orthogonally) projected data is maximized

- The definition of orthogonal projection of a vector x onto a unit-length vector u is : $P_u(x) = (x^T u)u$
- Every configuration matrix X_i of size $[M, 2]$ is now represented as a vector $\boldsymbol{x}_i = \text{vec}(X_i)$ of size $2M$



- The variance of the projected data onto a $2M$-dim vector u is:

$$\text{Var}(|P_u(\{\boldsymbol{x}_i\})|) = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{x}_i^T u - \bar{\boldsymbol{x}}^T u)^2 = u^T C u \tag{9}$$

- where $C = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T$ and $\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i$
- By definition, we look for the direction u such that the variance of the projected data is maximized, thus:

$$\begin{aligned} u^* &= \arg\max_u u^T C u \quad \text{s.t.} \quad ||u||_2 = 1 \\ &= \arg\max_u f(u; C) = u^T C u + \lambda(1 - u^T u) \end{aligned} \tag{10}$$

- By differentiating wrt u and setting equal to 0 we obtain:

$$\frac{df}{du} = 2Cu - 2\lambda u = 0 \rightarrow Cu = \lambda u \tag{11}$$

$$\underbrace{Cu = \lambda u}_{Eigenequation} \rightarrow u^T C u = \lambda \rightarrow \max(u^T C u) = \max \lambda \tag{12}$$

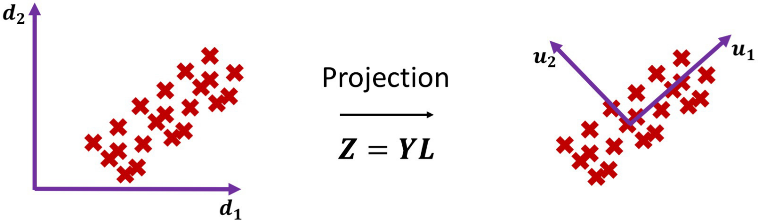
- In order to maximize $u^T C u$, i.e. projected variance, we need to compute the eigenvalues and eigenvector of C and select the greatest eigenvalue λ and the corresponding eigenvector u
- The other directions are the ones that maximize the projected variance among all possible orthogonal basis to u
- Since C is a symmetric positive-semidefinite matrix with real entries, the finite-dimensional spectral theorem asserts that:
  1. C has always $2M$ linearly independent eigenvectors mutually orthogonal U ($U^T = U^{-1}$)
  2. All eigenvalues of C are real and non-negative D

$$C = UDU^T \rightarrow CU = UD \tag{13}$$

$$C = UDU^T \rightarrow CU = UD \tag{14}$$

- Calling Y the $[N, 2M]$ matrix where each row is a **centred** configuration $\boldsymbol{x}_i$, we can compute the sample covariance matrix as $C = \frac{1}{N-1} Y^T Y$ where C is a $[2M, 2M]$ matrix
- U is a $[2M, 2M]$ matrix whose columns are the normalized right eigenvectors of C ordered such that the first column represents the eigenvector relative to the greatest eigenvalue λ. It is an orthogonal matrix and thus it represents a linear transformation (either a rotation or a reflection)
- D is a $[2M, 2M]$ diagonal matrix whose entries are the eigenvalues λ of C (decreasing order)

PCA



- The eigenvectors U represent a new orthonormal basis such that the projected data has maximal variance
- $L = U(:, 1:k)$ is a $[2M, k]$ matrix, where $k \leq 2M$, containing the *loadings*.
- Z is a $[N, k]$ matrix containing the *scores*. Its columns are called Principal Components (PC) and they are uncorrelated since their covariance matrix is diagonal (i.e. $D(1:k, 1:k)$). The first k PC explain $(\sum_{t=1}^{k} \lambda_t)/(\sum_{t=1}^{2M} \lambda_t)$ of the total variability

PCA

- How to compute PCA ?
  1. Center the data Y
  2. Use Singular Value Decomposition SVD (i.e. $Y = R\Sigma W^T$ and so $Y^T Y = W\Sigma^2 W^T$)
- High-dimensional data ($2M >> N$)
  - C has the same eigenvalues different from zero as $\tilde{C} = \frac{1}{N-1} YY^T$ which is a $[N, N]$ matrix.
  - The eigenvectors of C can be computed from the ones of $\tilde{C}$

$$CU = UD \rightarrow \tilde{C}(YU) = (YU)D \rightarrow \tilde{C}\tilde{U} = \tilde{U}D \tag{15}$$

  - Thus, $\tilde{U} = YU$

- From the previous equations, it follows that we can approximate every configuration i as:

$$z_i = L^T(y_i - \bar{y}) \rightarrow y_i \approx \bar{y} + Lz_i \tag{16}$$

- Furthermore, each eigenvector u_j describes a direction in the shape space with large variability (variance). The explained variance of u_j is $\frac{\lambda_j}{\sum_{t=1}^{2M} \lambda_t}$
- We can build a generative model to capture and see these variations: $g_j = \bar{y} \pm 3\sqrt{\lambda_j} u_j$ where g_j is the j-th **mode** and where we assume that data follow a Gaussian distribution (which is one of the assumption behind PCA)
- Since λ_j is the variance, the scalar $3\sqrt{\lambda_j}$ simply means 3 standard deviations, that is to say $99,73\%$ of the data

PCA

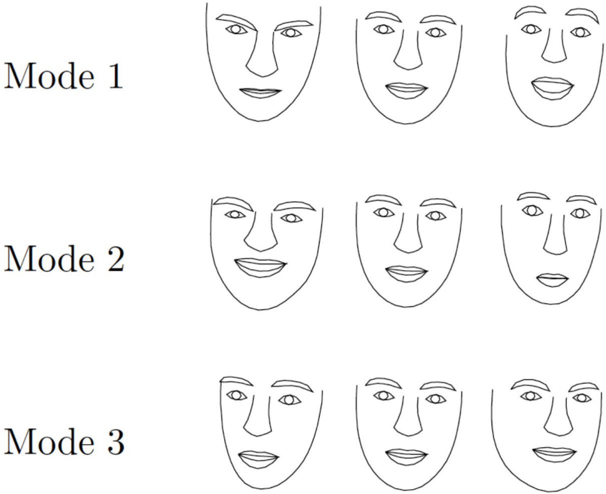


Figure 5: Average shape in the middle. First three modes at $-3\sqrt{\lambda_j}$ and $+3\sqrt{\lambda_j}$ on the left and right respectively. Taken from [3].

PCA

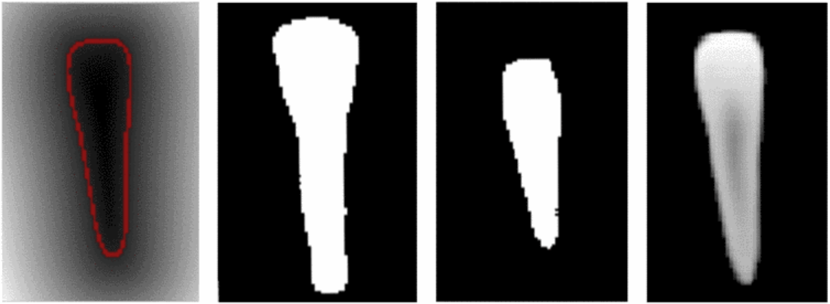
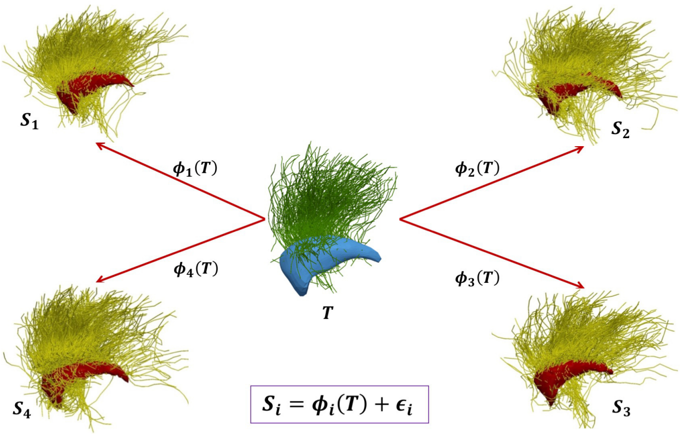


Figure 6: Statistical shape models applied to teeth segmentation. From left to right. Mean shape distance map with isolines. Two shape variations at $\pm 3\sqrt{\lambda}$ of the first mode. Intensity mean model computed by averaging the intensities of all images after being registered with B-splines towards the average image.

Summary

Diffeomorphometry

- Instead than using the Procrustes residuals to measure the shape differences between two configurations, another technique is based on **deformations**
- Once performed PA (or GPA for $N > 2$), one could deform one configuration into another one quantifying shape differences by looking at the “amount” of deformation
- By using diffeomorphism, which are smooth and invertible deformations whose inverse is also smooth, it is possible to define local non-linear deformations at every point in the space
- This allows a better alignment than affine transformations and above all we can quantify the shape differences (i.e. “amount” of deformation) at every point of the anatomical structure



- The estimate of the average shape or template T and shape variations is called **atlas construction** [8]. Every shape S_i is modelled as a deformation ϕ_i (i.e. diffeomorphism) of T plus a residual error ϵ_i

- From a mathematical point of view, we minimize a cost function of this type:

$$T^*, \{\alpha_i^*\} = \underset{T,\alpha_i}{\arg\min} \sum_{i=1}^{N} ||S_i - \phi_i^{\alpha_i}(T)|| + \gamma \text{Reg}(\phi_i^{\alpha_i}) \qquad (17)$$

- The parameters are the average shape T and the deformation parameters α_i, one for every subject i. They may be, for instance, the initial velocities v_0 of the diffeomorphisms (see previous lecture)
- Once estimated them, we can use a PCA to study the shape variability within the population as before
- The only difference is that this time the PCA is computed with the deformation parameters α_i

Atlas construction

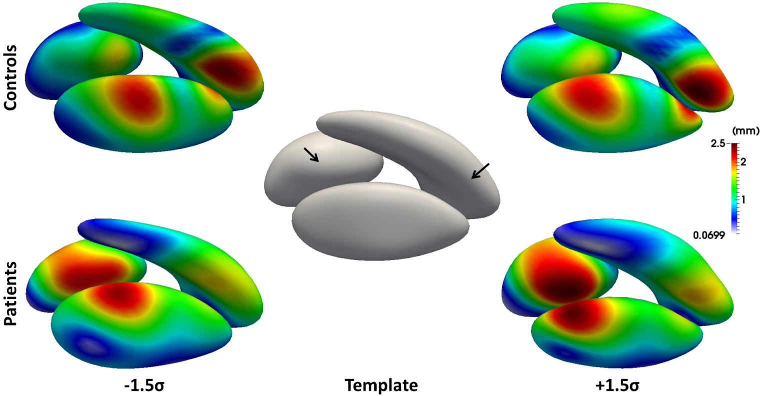


Figure 7: First mode of a PCA computed using the deformation parameters α_i and three anatomical structures of the brain. We compare the results for two different populations. The template has been computed considering both populations together.

Atlas construction

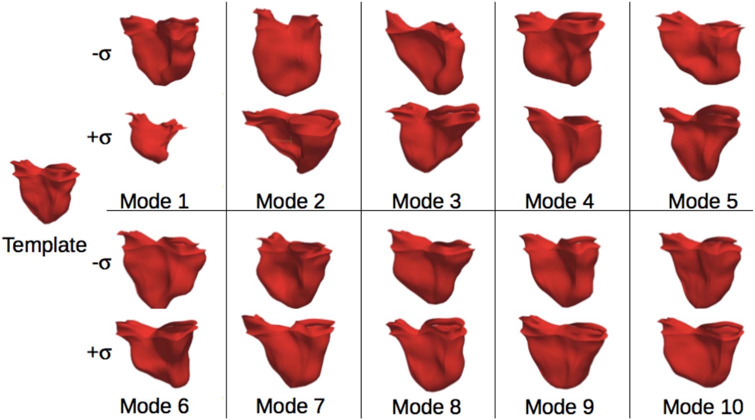


Figure 8: PCA deformation modes on a population of 18 patients suffering from repaired Tetralogy of Fallot. Image taken from T. Mansi - MICCAI - 2009.

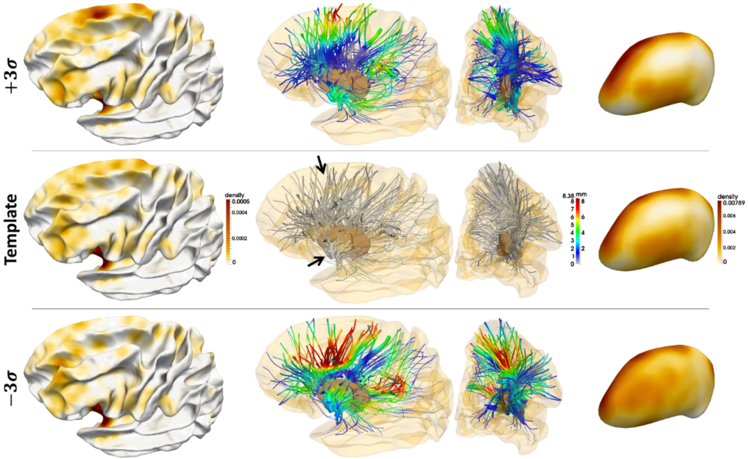


Figure 9: Structural connectivity changes in a population composed of both controls and patients with Gilles de la Tourette syndrome.

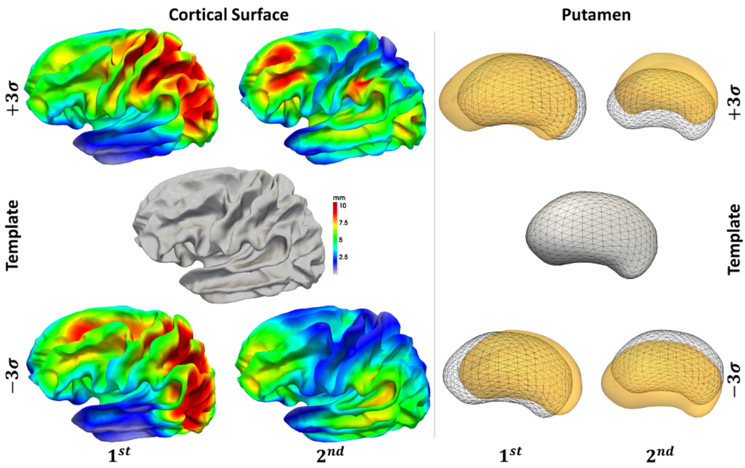


Figure 10: Morphological changes in a population composed of both controls and patients with Gilles de la Tourette syndrome.

Summary

Atlases-Templates

- Instead than estimating the average or reference shape of a group of anatomical structures, one can also use pre-computed atlases (or templates)
- An atlas gives a common coordinate system for the whole population where anatomical prior information can be mapped (labels, functional information, etc.)
- Examples of neuroimaging atlases:
  1. **Talairach atlas**: built from a single post-mortem brain (60-year-old French healthy woman). It is composed of:
     - A coordinate system to identify a particular area of the brain with respect to three anatomical landmarks (AC, PC, IH fissure)
     - A specific similarity transformation to align a brain with the atlas
     - No histological study. Inaccurate anatomical labels

2. **MNI atlases**: built from a series of MR images of healthy young adults.
   - **MNI250**: the MR scans of 250 brains were aligned to the Talairach atlas using a similarity transformation based on manually labelled landmarks. Then, the MNI241 atlas was the average of all the registered scans
   - **MNI305**: Other 55 MR images were registered to the MNI250 atlas with an automatic linear registration method. The MNI305 is the average of all 305 scans (right hand, 239 M, 66 F, 23.4 average age $\pm$ 4 years)
   - **ICBM152**: current standard MNI template. It is the average of 152 normal MRI scans matched to the MNI305 atlas using an affine transformation (9 degrees)

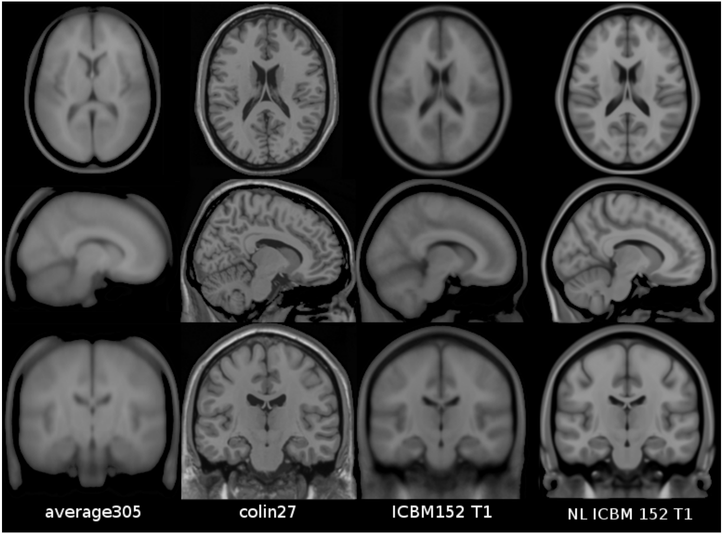


Figure 11: Examples of neuroimaging atlases from T1-w MRI

References

1. I.L. Dryden and K.V. Mardia (1998). Statistical Shape Analysis. Wiley and Sons
2. C. Goodall (1991). Procrustes Methods in the Statistical Analysis of Shape. J.R. Statist. Soc. B
3. T.F. Cootes (1995). Active Shape Models-Their Training and Application. Computer Vision and Image Understanding
4. M. Miller (1993). Mathematical textbook of deformable neuroanatomies. PNAS
5. Stegmann and Gomez (2002). A Brief Introduction to Statistical Shape Analysis.
6. N. MacLeod. PalaeoMath: Part 17 - Shape Theory
7. L. Younes (2010). Shapes and Diffeomorphisms. Springer
8. Joshi et al. (2004). Unbiased diffeomorphic atlas construction... NeuroImage