

# Introduction to image registration Landmarks and Statistical Shape Analysis

Pietro Gori

Enseignant-chercheur  
Equipe IMAGES - Télécom Paris  
[pietro.gori@telecom-paris.fr](mailto:pietro.gori@telecom-paris.fr)



# Plan

## 1 Landmark based registration

- Affine registration
- Procrustes superimposition
- Non-linear registration (small displacement)
- Non-linear registration (diffeomorphism)

## 2 Statistical Shape Analysis

- Generalized Procrustes Analysis (GPA)
- Tangent space projection
- Shape variations
- Atlas constructions
- Atlases-Templates

# Summary

## 1 Landmark based registration

- Affine registration
- Procrustes superimposition
- Non-linear registration (small displacement)
- Non-linear registration (diffeomorphism)

## 2 Statistical Shape Analysis

- Generalized Procrustes Analysis (GPA)
- Tangent space projection
- Shape variations
- Atlas constructions
- Atlases-Templates

# Anatomical landmarks

Definition: anatomical landmark

An anatomical landmark is a point precisely defined onto an anatomical structure which establishes a correspondence among the population of homologous anatomical objects

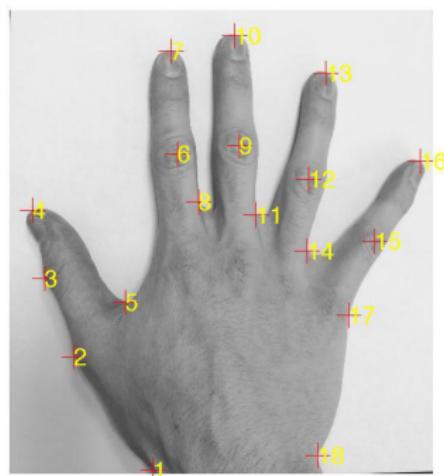
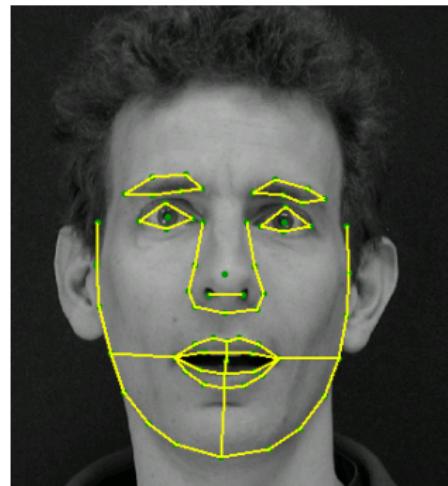


Figure 1: Example of manually labeled landmarks

# Anatomical landmarks

- Given a set of  $N$  landmarks  $\mathbf{i}$  defined on  $I$  and  $N$  corresponding landmarks  $\mathbf{j}$  defined on  $J$ , we seek to minimize:

$$\theta^* = \arg \min_{\theta} \sum_{p=1}^N \|\mathbf{T}(i_p) - j_p\|^2 \quad (1)$$

- where  $\mathbf{T}(i_p)$  means that we apply the deformation  $\mathbf{T}$  to the  $p$ -th landmark  $i_p$
- We suppose that all landmarks belong to  $\mathbb{R}^2$  (it would be similar for  $\mathbb{R}^3$ )
- The metric is the Euclidean norm (or Frobenius when using matrices)

# Summary

## 1 Landmark based registration

- Affine registration
- Procrustes superimposition
- Non-linear registration (small displacement)
- Non-linear registration (diffeomorphism)

## 2 Statistical Shape Analysis

- Generalized Procrustes Analysis (GPA)
- Tangent space projection
- Shape variations
- Atlas constructions
- Atlases-Templates

# Landmark based affine registration

- We define:

$$\mathbf{T}(i_p) = Ai_p + t \quad \forall p \in [1, N] \quad (2)$$

- Thus,  $\theta = A, t$

$$A^*, t^* = \arg \min_{A, t} f(A, t) = \sum_{p=1}^N \|Ai_p + t - j_p\|^2 \quad (3)$$

- From which it results:

$$\frac{\partial f}{\partial t} = 2 \sum_{p=1}^N (Ai_p + t - j_p) = 0 \rightarrow t^* = \bar{j} - A\bar{i} \quad (4)$$

- We notice that if we center the data (i.e.  $\tilde{i}_p = i_p - \bar{i}$  and  $\tilde{j}_p = j_p - \bar{j}$ ) then  $t^* = 0$ . The criterion thus becomes  $f = \sum_{p=1}^N \|A\tilde{i}_p - \tilde{j}_p\|^2$

## Landmark based affine registration

- Now we differentiate wrt  $A$  :

$$\begin{aligned}\frac{\partial f}{\partial A} &= \sum_{p=1}^N \frac{\partial ||A\tilde{i}_p||^2}{\partial A} - 2 \frac{\partial \langle A\tilde{i}_p, \tilde{j}_p \rangle}{\partial A} \\ &= 2 \sum_{p=1}^N A\tilde{i}_p \tilde{i}_p^T - \tilde{j}_p \tilde{i}_p^T = 2 \sum_{p=1}^N (A\tilde{i}_p - \tilde{j}_p) \tilde{i}_p^T = 0\end{aligned}\tag{5}$$

- It results:

$$A^* = \left( \sum_{p=1}^N \tilde{j}_p \tilde{i}_p^T \right) \left( \sum_{p=1}^N \tilde{i}_p \tilde{i}_p^T \right)^{-1}\tag{6}$$

The matrix  $\left( \sum_{p=1}^N \tilde{i}_p \tilde{i}_p^T \right)$  is invertible if the landmarks are not all aligned on a straight line.

# Landmark based affine registration

- From a computational point of view, it is easier to use homogeneous coordinates:

$$\mathbf{T}^* = \arg \min_{\mathbf{T}} \|\mathbf{x}\mathbf{T} - \mathbf{y}\|_F^2 \quad (7)$$

- where we define

$$\underbrace{\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 \\ & & \vdots & & & \\ x_N & y_N & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1 & y_1 & 1 \\ & & & \vdots & & \\ 0 & 0 & 0 & x_N & y_N & 1 \end{bmatrix}}_{\mathbf{x}} - \underbrace{\begin{bmatrix} A_{11} \\ A_{12} \\ t_1 \\ A_{21} \\ A_{22} \\ t_2 \end{bmatrix}}_{\mathbf{T}} = \underbrace{\begin{bmatrix} u_1 \\ \vdots \\ u_N \\ v_1 \\ \vdots \\ v_N \end{bmatrix}}_{\mathbf{y}} \quad (8)$$

$$\mathbf{T}^* = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{y} \quad (9)$$

# Summary

## 1 Landmark based registration

- Affine registration
- **Procrustes superimposition**
- Non-linear registration (small displacement)
- Non-linear registration (diffeomorphism)

## 2 Statistical Shape Analysis

- Generalized Procrustes Analysis (GPA)
- Tangent space projection
- Shape variations
- Atlas constructions
- Atlases-Templates

## Procrustes superimposition (similarity transformation)

- Using the same notation as Schönemann ( $R \rightarrow R^T$ ), we seek to minimize:

$$(s, R, t)^* = \arg \min_{s, R, t} \sum_{p=1}^N \|sR^T i_p + t - j_p\|_2^2 \quad (10)$$

- where  $s$  is a uniform scaling factor (scalar) and  $R$  is a rotation matrix. The translation vector  $t$  is, as before, equal to  $t^* = \bar{j} - sR^T \bar{i}$ . Thus, by centering the data  $(X_c, Y_c)$ , we obtain:

$$(s, R)^* = \arg \min_{s, R} f(s, R) = \sum_{p=1}^N \|sR^T \tilde{i}_p - \tilde{j}_p\|_2^2 = \|sX_c R - Y_c\|_F^2 \quad (11)$$

- Remember that:  $\sum_{p=1}^N \|\tilde{i}_p - \tilde{j}_p\|_2^2 = \|X_c - Y_c\|_F^2$  where  $X = [\tilde{i}_1^T; \tilde{i}_2^T; \dots; \tilde{i}_N^T]$  and  $\|X\|_F^2 = \text{Tr}(X^T X)$

# Procrustes superimposition (similarity transformation)

- We minimize wrt  $s$ :

$$\begin{aligned} f(s, R) &\propto s^2 \|X_c R\|_F^2 - 2s \langle X_c R, Y_c \rangle_F \\ \frac{\partial f(s, R)}{\partial s} &= 2s \|X_c\|_F^2 - 2 \langle X_c R, Y_c \rangle_F \\ s^* &= \frac{\langle X_c R, Y_c \rangle_F}{\|X_c\|_F^2} \end{aligned} \tag{12}$$

- where we use the fact that  $R^T R = R R^T = \mathbb{I}$ . Substituting into  $f$ :

$$\begin{aligned} R^* &= \arg \min_R f(R) = -\frac{(\langle X_c R, Y_c \rangle_F)^2}{\|X_c\|_F^2} \\ &= \arg \max_R |\langle X_c R, Y_c \rangle_F| = |\langle R, X_c^T Y_c \rangle_F| \\ &= \arg \max_R |\langle R, U \Sigma V^T \rangle_F| = |\langle U^T R V, \Sigma \rangle_F| = |\langle Z, \Sigma \rangle_F| \end{aligned} \tag{13}$$

- where we use the SVD decomposition  $X_c^T Y_c = U \Sigma V^T$  and the definition of the trace  $\langle X_c R, Y_c \rangle_F = \text{Tr}(R^T X_c^T Y_c) = \langle R, X_c^T Y_c \rangle_F$

## Procrustes superimposition (similarity transformation)

- Notice that  $Z = U^T RV$  is an orthogonal matrix since it is the product of orthogonal matrices. Thus  $Z^T Z = \mathbb{I}$  and  $z_j^T z_j = 1$ . It follows that  $z_{ij} \leq 1$ .

$$\begin{aligned} R^* &= \arg \max_R |\langle Z, \Sigma \rangle_F| = |\text{Tr}(\Sigma^T Z)| = \\ &= |\text{Tr} \left( \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \right) | = \sum_{d=1}^2 \sigma_d z_{dd} \leq \sum_{d=1}^2 \sigma_d \end{aligned} \quad (14)$$

- The maximum is obtained when  $z_{dd} = 1 \quad \forall d$ , which means when:

$$Z = U^T RV = I \rightarrow R^* = UV^T \quad (15)$$

- In order to be sure that  $R$  is a rotation matrix ( $\det(R) = 1$ ), we compute [5]:

$$R^* = U \begin{bmatrix} 1 & 0 \\ 0 & \det(UV^T) \end{bmatrix} V^T = USV^T \quad (16)$$

# Procrustes superimposition (similarity transformation)

- To recap [5]:

$$R^* = USV^T$$

$$s^* = \frac{\langle R, Y_c X_c^T \rangle_F}{\|X_c\|_F^2} = \frac{\text{Tr}(S\Sigma)}{\|X_c\|_F^2} \quad (17)$$

$$t^* = \bar{j} - \frac{1}{N} \sum_{p=1}^N s R i_p$$

# Summary

## 1 Landmark based registration

- Affine registration
- Procrustes superimposition
- **Non-linear registration (small displacement)**
- Non-linear registration (diffeomorphism)

## 2 Statistical Shape Analysis

- Generalized Procrustes Analysis (GPA)
- Tangent space projection
- Shape variations
- Atlas constructions
- Atlases-Templates

## Non-linear registration (small displacement)

- We define the deformation of a pixel at location  $z = (x, y)$  as:

$$\mathbf{T}(z) = z + v(z) \quad \text{with} \quad v(z) = \sum_{p=1}^N K(z, i_p) \alpha_p \quad (18)$$

- where  $K(z, i_p)$  is a kernel, for instance  $K(z, i_p) = \exp\left(-\frac{\|z - i_p\|_2^2}{\lambda^2}\right)$  and  $\alpha_p$  is a 2D vector which need to be estimated.
- The displacement at any point  $z$  depends on the displacement of the neighbor landmarks

## Non-linear registration (small displacement)

- We minimize

$$\begin{aligned}\boldsymbol{\alpha}^* &= \arg \min_{\boldsymbol{\alpha}} f(\boldsymbol{\alpha}; \lambda) = \sum_{p=1}^N \|i_p + v(i_p) - j_p\|_2^2 \\ &= \sum_{p=1}^N \|i_p + \left( \sum_{d=1}^N K(i_p, i_d) \alpha_d \right) - j_p\|_2^2 \\ &= \|\mathbf{i} + \mathbf{K}\boldsymbol{\alpha} - \mathbf{j}\|_F^2\end{aligned}\tag{19}$$

- where  $\mathbf{K} = \begin{bmatrix} 1 & K(i_1, i_2) & \dots & K(i_1, i_N) \\ K(i_2, i_1) & 1 & \dots & K(i_2, i_N) \\ \dots & \dots & \dots & \dots \\ K(i_N, i_1) & K(i_N, i_2) & \dots & 1 \end{bmatrix}$  and  
 $\boldsymbol{\alpha} = [\alpha_1^T; \dots; \alpha_N^T]$

## Non-linear registration (small displacement)

- By differentiating wrt  $\alpha$ :

$$\frac{\partial \|\mathbf{i} + \mathbf{K}\alpha - \mathbf{j}\|_F^2}{\partial \alpha} = 2\mathbf{K}^T(\mathbf{i} + \mathbf{K}\alpha - \mathbf{j}) = 0 \quad (20)$$
$$\alpha^* = \mathbf{K}^{-1}(\mathbf{j} - \mathbf{i})$$

- The matrix  $\mathbf{K}$  might not always be invertible (if  $\lambda$  is too big for instance). We need to regularize it. A possible solution is to use a Tikhonov matrix such as  $\alpha^T \mathbf{K} \alpha$ , thus obtaining:

$$\alpha^* = \arg \min_{\alpha} f(\alpha; \lambda, \gamma) = \|\mathbf{i} + \mathbf{K}\alpha - \mathbf{j}\|_F^2 + \gamma \alpha^T \mathbf{K} \alpha \quad (21)$$

$$\frac{\partial f(\alpha; \lambda, \gamma)}{\partial \alpha} = 2\mathbf{K}^T(\mathbf{i} + \mathbf{K}\alpha - \mathbf{j}) + 2\gamma \mathbf{K} \alpha = 0 \quad (22)$$
$$\alpha^* = (\mathbf{K} + \gamma \mathbb{I})^{-1}(\mathbf{j} - \mathbf{i})$$

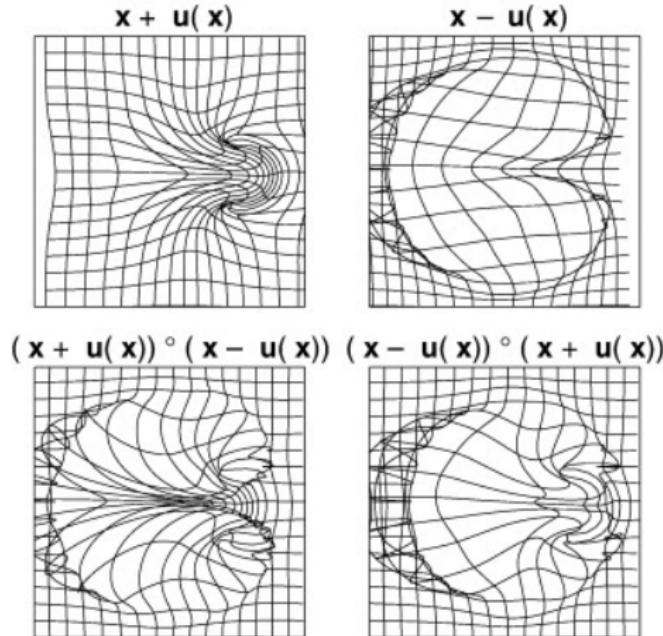
## Non-linear registration (small displacement)

- Why only small displacement ? → We could approximate the inverse of  $\mathbf{T}(z)$  as  $\mathbf{T}^{-1}(z') = z' - v(z')$  obtaining:

$$\mathbf{T}(\mathbf{T}^{-1}(z')) = z' - v(z') + v(z' - v(z')) \neq z' \quad (23)$$

- The error is small only if  $v(z' - v(z')) - v(z')$  is small, which is the case only when the displacement is small !
- We might have intersections, holes or tearing in area where the displacement is large

# Small-displacement registration



**Figure 2:** First row: forward and inverse transformation. The first one is a one-to-one mapping whereas the second one presents intersections. Second row: composition of transformations. They should be the identity transforms. Image taken from [7].

# Summary

## 1 Landmark based registration

- Affine registration
- Procrustes superimposition
- Non-linear registration (small displacement)
- Non-linear registration (diffeomorphism)

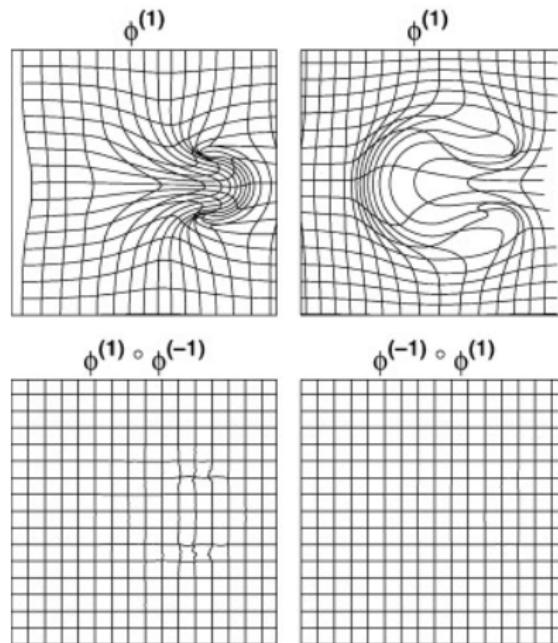
## 2 Statistical Shape Analysis

- Generalized Procrustes Analysis (GPA)
- Tangent space projection
- Shape variations
- Atlas constructions
- Atlases-Templates

## Large-displacement registration: diffeomorphism

- Instead than using small-displacement transforms we should use **diffeomorphisms**
- A diffeomorphism is a differentiable (smooth and continuous) bijective transformation (one-to-one) with differentiable inverse (i.e. nonzero Jacobian determinant)
- Using diffeomorphic transformations we can preserve the topology and spatial organization, namely no intersection, folding or shearing may occur

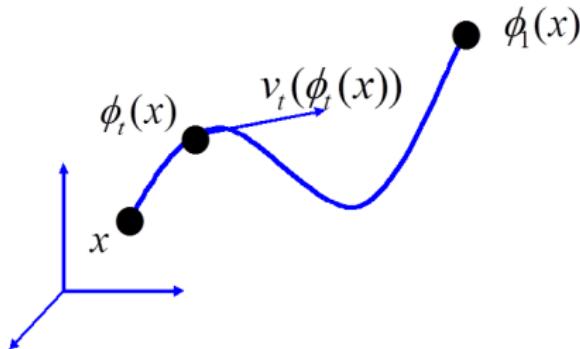
# Diffeomorphism



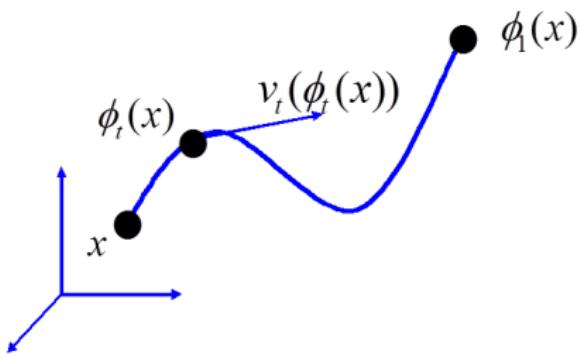
**Figure 3:** First row: forward and inverse diffeomorphic transformation (both are one-to-one). Second row: composition of forward and inverse transformations. The result is the identity transform (i.e. no deformation). Image taken from [7].

# Diffeomorphism

- One of the most used algorithm in medical imaging to create diffeomorphic deformations is called LDDMM: Large Deformation Diffeomorphic Metric Mapping
- Deformations are built by integrating a time-varying vector field  $v_t(x)$  over  $t \in [0, 1]$  where  $v_t(x)$  represents the instantaneous velocity of any point  $x$  at time  $t$  (and no more a displacement vector !)

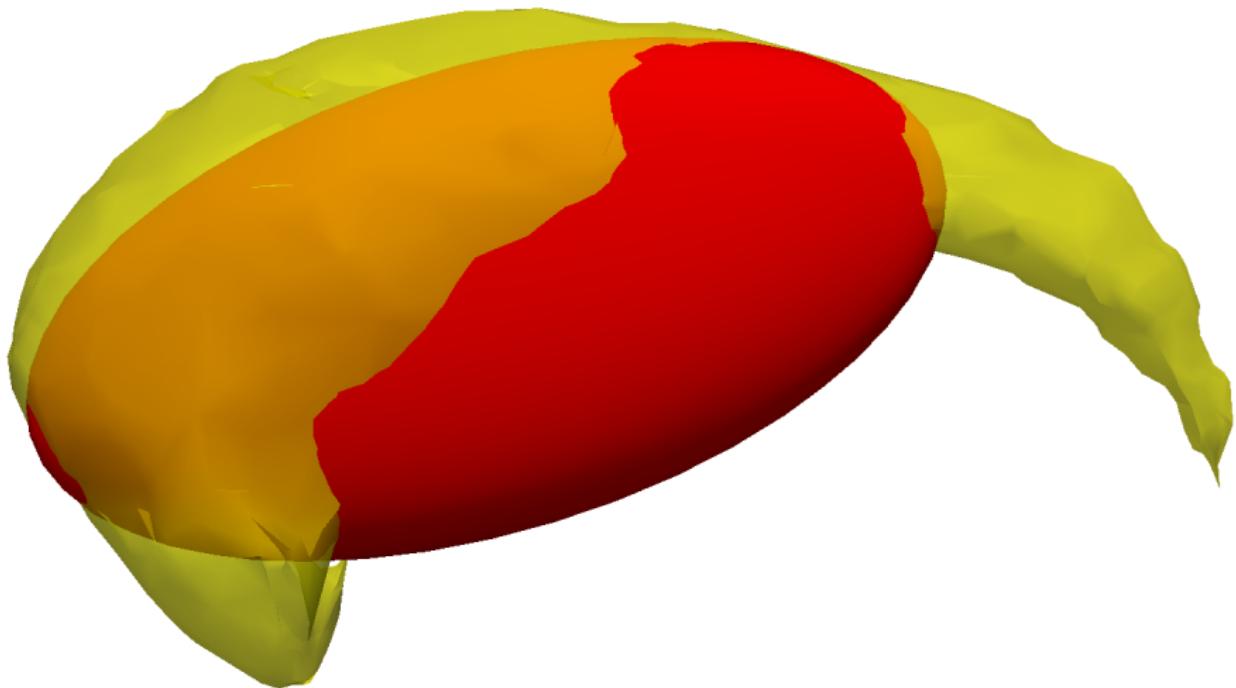


# Diffeomorphism

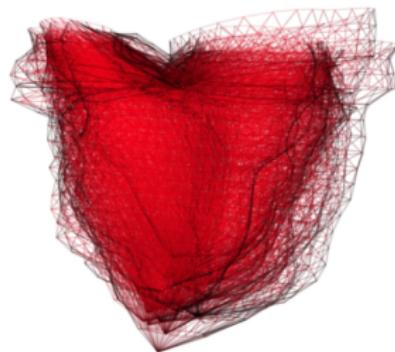


- Calling  $\phi_t(x)$  the position of a point at time  $t$  which was located in  $x$  at time  $t = 0$ , its evolution is given by:  $\frac{\partial \phi_t(x)}{\partial t} = v_t(\phi_t(x))$  with  $\phi_0(x) = x$
- Integrating  $\frac{\partial \phi_t(x)}{\partial t} = v_t(\phi_t(x))$  between  $t \in [0, 1]$  produces a flow of diffeomorphisms (if  $v$  is square integrable). The last diffeomorphism is the one we are interested into.

# Diffeomorphism



# Diffeomorphism



Rigid alignment



Non-linear registration to the template

**Figure 4:** Image taken from T. Mansi - MICCAI - 2009

# Diffeomorphism

<http://www.deformetrica.org/>

# Diffeomorphism

- The flow of diffeomorphisms produces a dense deformation of the entire 3D space. We know how to deform every point in the space.
- The last diffeomorphism is parametrised by the initial velocity  $v_0$
- From a mathematical point of view, to register a source image or mesh  $I$  to a target image or mesh  $J$  we minimize:

$$\arg \min_{v_0} D(\phi_1(I), J) + \gamma \text{Reg}(v_0) \quad (24)$$

- where  $D$  is a data term,  $\text{Reg}$  is a regularization term and  $\gamma$  their trade-off. We use an optimization scheme (e.g. gradient descent) to estimate the optimal deformation parameters  $v_0$ .

# Summary

## 1 Landmark based registration

- Affine registration
- Procrustes superimposition
- Non-linear registration (small displacement)
- Non-linear registration (diffeomorphism)

## 2 Statistical Shape Analysis

- Generalized Procrustes Analysis (GPA)
- Tangent space projection
- Shape variations
- Atlas constructions
- Atlases-Templates

# 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 ...

# Introduction

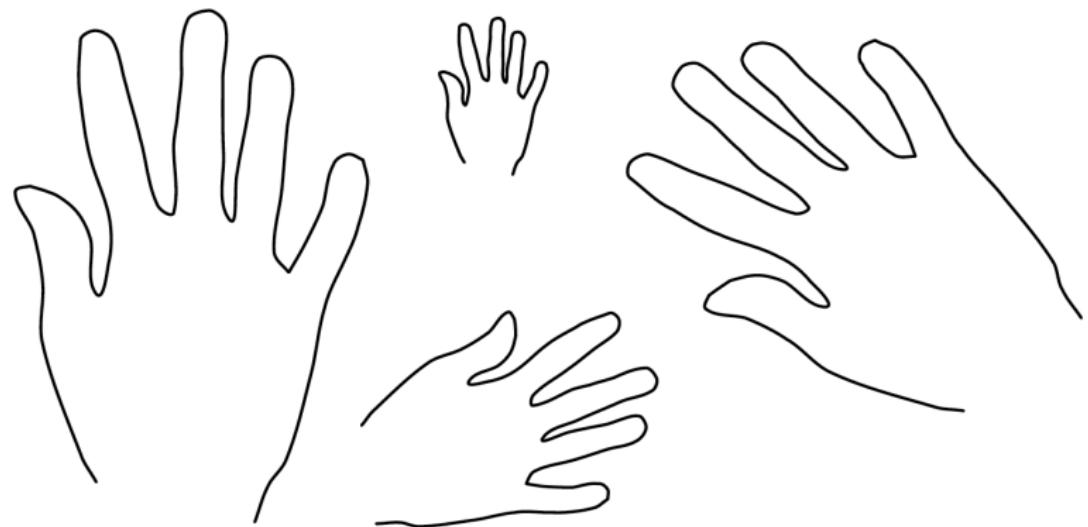


Figure 5: Four objects representing a hand with the same shape. Taken from [5]

# Introduction

- In this lecture, we will focus on anatomical labelled (i.e. ordered) landmarks
- Given a set of  $N$  anatomical structures  $\{S_i\}_{i=1,\dots,N}$  each one of them labelled with a configuration of  $M$  ordered landmarks  
 $X_i = [x_{i1}^T; x_{i2}^T; \dots; x_{iM}^T]$  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

## 1 Landmark based registration

- Affine registration
- Procrustes superimposition
- Non-linear registration (small displacement)
- Non-linear registration (diffeomorphism)

## 2 Statistical Shape Analysis

- Generalized Procrustes Analysis (GPA)
- Tangent space projection
- Shape variations
- Atlas constructions
- Atlases-Templates

# 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:

$$\begin{aligned}s_i^*, R_i^*, t_i^* &= \arg \min_{s_i, R_i, t_i} \frac{1}{N} \sum_{i=1}^N \sum_{k=i+1}^N \| (s_i X_i R_i + \mathbb{1}_M t_i^T) \\&\quad - (s_k X_k R_k + \mathbb{1}_M t_k^T) \|_F = \\&= \arg \min_{s_i, R_i, t_i} \sum_{i=1}^N \| (s_i X_i R_i + \mathbb{1}_M t_i^T) - \bar{X} \|_F\end{aligned}\tag{25}$$

- 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:

$$s_i^*, R_i^*, t_i^* = \arg \min_{s_i, R_i, t_i} \sum_{i=1}^N \|(s_i X_i R_i + \mathbb{1}_M t_i^T) - \bar{X}\|_F \quad (26)$$

- 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 aligned configurations:  $s_i X_i R_i + \mathbb{1}_M t_i^T$ . They are used to analyse variations in shape.

# Generalized Procrustes Analysis (GPA)

$$s_i^*, R_i^*, t_i^* = \arg \min_{s_i, R_i, t_i} \sum_{i=1}^N \|(s_i X_i R_i + \mathbb{1}_M t_i^T) - \bar{X}\|_F \quad (27)$$

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

# Generalized Procrustes Analysis (GPA)

$$s_i^*, R_i^*, t_i^* = \arg \min_{s_i, R_i, t_i} \sum_{i=1}^N \|(s_i X_i R_i + \mathbb{1}_M t_i^T) - \bar{X}\|_F \quad (27)$$

- 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)

$$s_i^*, R_i^*, t_i^* = \arg \min_{s_i, R_i, t_i} \sum_{i=1}^N \| (s_i X_i R_i + \mathbb{1}_M t_i^T) - \bar{X} \|_F \quad (27)$$

- 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:

$$X_i = \alpha_i(\bar{X} + E_i)\Omega_i + \mathbb{1}_M\omega_i^T \quad (28)$$

- 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:

$$\bar{X} + E_i = s_i X_i R_i + \mathbb{1}_M t_i^T \quad (29)$$

- 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. 25.

# Generalized Procrustes Analysis (GPA)

$$s_i^*, R_i^*, t_i^* = \arg \min_{s_i, R_i, t_i} \sum_{i=1}^N \| (s_i X_i R_i + \mathbb{1}_M t_i^T) - \bar{X} \|_F \quad (30)$$

- 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)

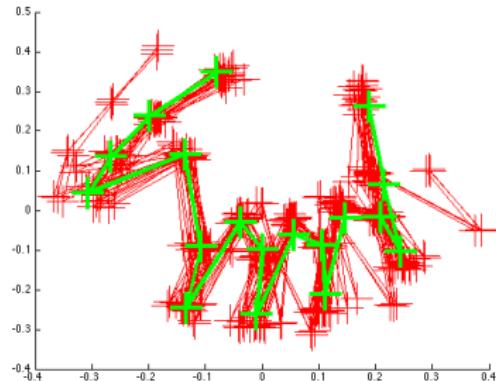
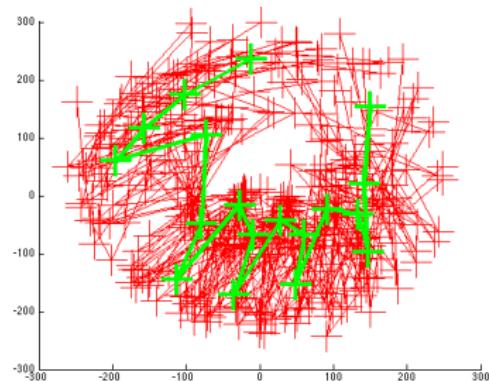
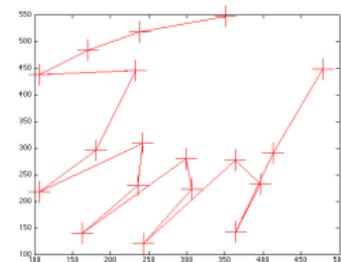
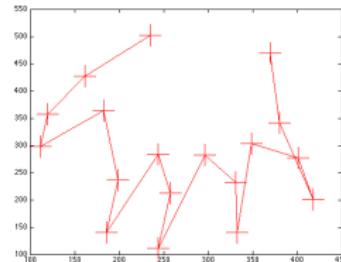
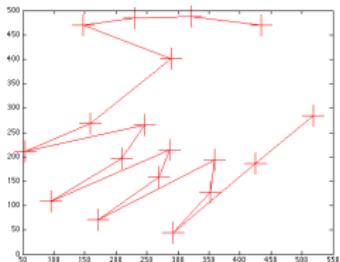
- We assume that all configurations  $X_i$  have been centred (i.e.  $x_{ij} = x_{ij} - \frac{1}{M} \sum_{j=1}^M x_{ij} = x_{ij} - \bar{x}_i$ )
- Remember that  $t_i^* = \bar{x} - \frac{1}{M} \sum_{j=1}^M s^* R^* x_{ij} = \frac{1}{M}(\bar{X} \mathbb{1}_M) - \frac{1}{M}(s_i^* R_i^{*T} X_i^T \mathbb{1}_M)$ , 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:

$$R^* = USV^T$$
$$s^* = \frac{\langle R, X^T \bar{X} \rangle_F}{\|X\|_F} = \frac{\text{Tr}(S\Sigma)}{\|X\|_F} \quad (31)$$

- where we employ the SVD decomposition  $X^T \bar{X} = U\Sigma V^T$  and

$$S = \begin{bmatrix} 1 & 0 \\ 0 & \det(UV^T) \end{bmatrix}$$

# Generalized Procrustes Analysis (GPA)



# Summary

## 1 Landmark based registration

- Affine registration
- Procrustes superimposition
- Non-linear registration (small displacement)
- Non-linear registration (diffeomorphism)

## 2 Statistical Shape Analysis

- Generalized Procrustes Analysis (GPA)
- **Tangent space projection**
- Shape variations
- Atlas constructions
- Atlases-Templates

# Tangent space projection

- Once all configurations have been aligned to a common coordinate frame filtering out similarity transformations, the left geometrical information represents the **shape** of each structure
- However, after alignment and normalization, all configurations lie on a  $2M$ -dimensional hyper sphere. The actual curved distance between two configurations  $\rho$  is not the linear distance  $D\rho$  used to calculate the Procrustes residuals

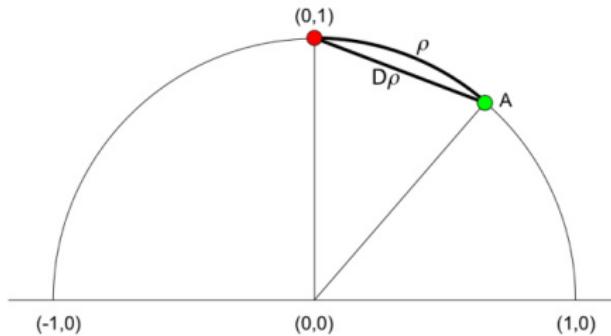


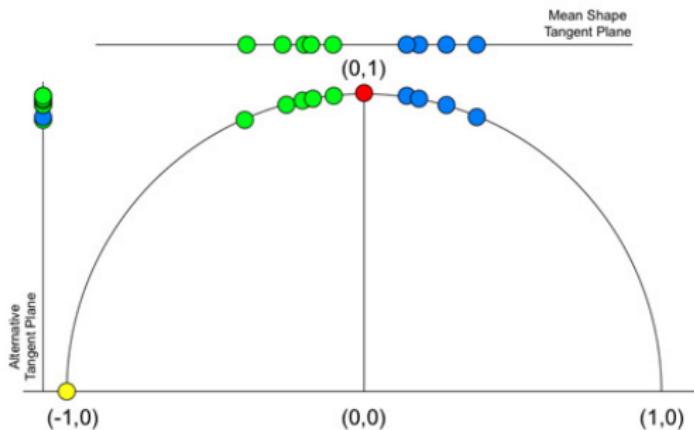
Figure taken from [6]

## 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 ?

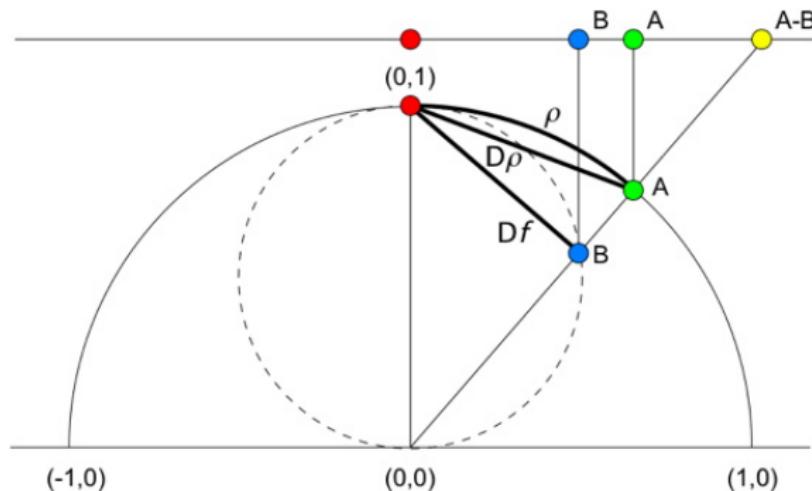
# Tangent space projection

- 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 7:** 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  $\mathbf{x} = [x_1, \dots, x_M, y_1, \dots, y_M]^T$
- We notice that:  $|\mathbf{x}_t| \cos(\theta) = |\mathbf{x}_t| \frac{\langle \mathbf{x}_t, \bar{\mathbf{x}} \rangle_2}{|\mathbf{x}_t| |\bar{\mathbf{x}}|} = |\bar{\mathbf{x}}| \rightarrow \langle \mathbf{x}_t, \bar{\mathbf{x}} \rangle_2 = |\bar{\mathbf{x}}|^2$
- Calling  $\mathbf{x}_t = \alpha \mathbf{x}$  we can rewrite:  $\langle \mathbf{x}_t, \bar{\mathbf{x}} \rangle_2 = \langle \alpha \mathbf{x}, \bar{\mathbf{x}} \rangle_2 = |\bar{\mathbf{x}}|^2$  and thus  $\alpha = \frac{|\bar{\mathbf{x}}|^2}{\langle \mathbf{x}, \bar{\mathbf{x}} \rangle_2}$
- It follows that  $\mathbf{x}_t = \alpha \mathbf{x} = \frac{|\bar{\mathbf{x}}|^2}{\langle \mathbf{x}, \bar{\mathbf{x}} \rangle_2} \mathbf{x}$

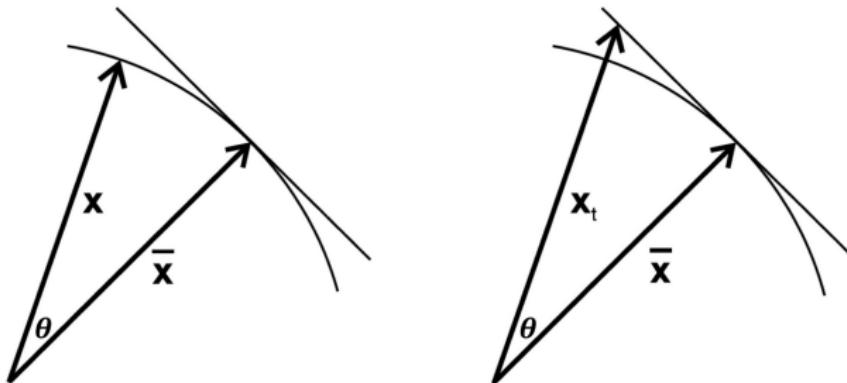


Figure 8: The vectors  $\mathbf{x}$  and  $\mathbf{x}_t$  are the configuration vectors respectively before and after projection. Figure modified from [5]

# Generalized Procrustes Analysis (GPA)

$$s_i^*, R_i^*, t_i^* = \arg \min_{s_i, R_i, t_i} \sum_{i=1}^N \| (s_i X_i R_i + \mathbb{1}_M t_i^T) - \bar{X} \|_F \quad (32)$$

- 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

## 1 Landmark based registration

- Affine registration
- Procrustes superimposition
- Non-linear registration (small displacement)
- Non-linear registration (diffeomorphism)

## 2 Statistical Shape Analysis

- Generalized Procrustes Analysis (GPA)
- Tangent space projection
- **Shape variations**
- Atlas constructions
- Atlases-Templates

## 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

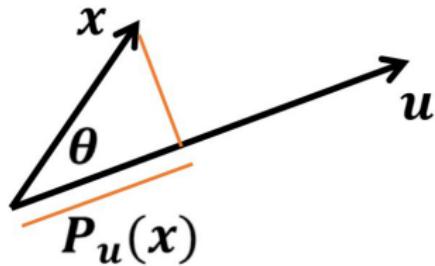
- Once all configurations have been aligned to a common coordinate frame filtering out similarity transformations, they represent the shape of each structure
- We have already seen how to measure the average shape, what about shape variability ?
- We could use **Principal Component Analysis (PCA)** onto the vectorized ( $\mathbf{x} = [x_1, \dots, x_M, y_1, \dots, y_M]^T$ ) 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  $\textcolor{blue}{x}$  onto a unit-length vector  $\textcolor{blue}{u}$  is :  $P_u(x) = (x^T u)u$
- Every configuration matrix  $\textcolor{blue}{X}_i$  of size  $[M, 2]$  is now represented as a vector  $\textcolor{blue}{x}_i = \text{vec}(X_i)$  of size  $2M$



# PCA

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

$$\text{Var}(|P_u(\{\mathbf{x}_i\})|) = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i^T \mathbf{u} - \bar{\mathbf{x}}^T \mathbf{u})^2 = \mathbf{u}^T C \mathbf{u} \quad (33)$$

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

$$\begin{aligned} u^* &= \arg \max_u u^T C u \quad \text{s.t.} \quad \|u\|_2 = 1 \\ &= \arg \max_u f(u; C) = u^T C u + \lambda(1 - u^T u) \end{aligned} \quad (34)$$

- By differentiating wrt  $\mathbf{u}$  and setting equal to 0 we obtain:

$$\frac{df}{du} = 2Cu - 2\lambda u = 0 \rightarrow Cu = \lambda u \quad (35)$$

# PCA

$$\underbrace{Cu = \lambda u}_{\text{Eigenequation}} \rightarrow u^T Cu = \lambda \rightarrow \max(u^T Cu) = \max \lambda \quad (36)$$

- In order to maximize  $u^T Cu$ , i.e. projected variance, we need to compute the eigenvalues and eigenvector of  $C$  and select the greatest eigenvalue  $\lambda$  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  $D$

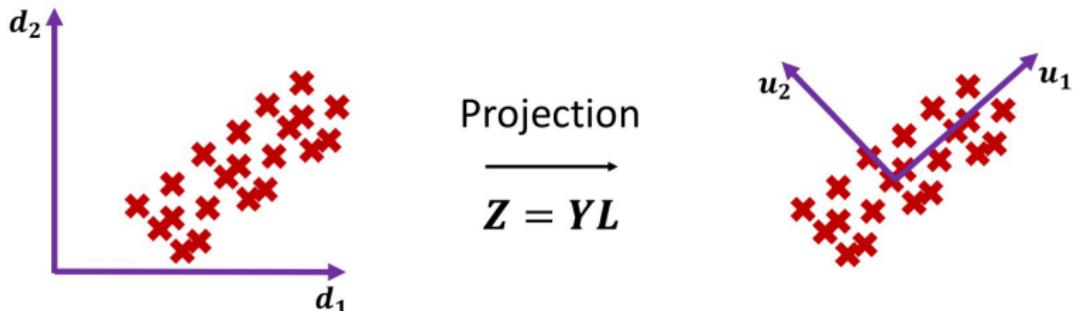
$$C = UDU^T \rightarrow CU = UD \quad (37)$$

# PCA

$$C = UDU^T \rightarrow CU = UD \quad (38)$$

- Calling  $\mathbf{Y}$  the  $[N, 2M]$  matrix where each row is a **centred** configuration  $\mathbf{x}_i$ , we can compute the sample covariance matrix as  $C = \frac{1}{N-1} \mathbf{Y}^T \mathbf{Y}$  where  $C$  is a  $[2M, 2M]$  matrix
- $\mathbf{U}$  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  $\lambda$ . It is an orthogonal matrix and thus it represents a linear transformation (either a rotation or a reflection)
- $\mathbf{D}$  is a  $[2M, 2M]$  diagonal matrix whose entries are the eigenvalues  $\lambda$  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  $\mathbf{Y}$
  - ② Use Singular Value Decomposition SVD (i.e.  $\mathbf{Y} = \mathbf{R}\Sigma\mathbf{W}^T$  and so  $\mathbf{Y}^T\mathbf{Y} = \mathbf{W}\Sigma^2\mathbf{W}^T$ )
- High-dimensional data ( $2M \gg N$ )
  - $\mathbf{C}$  has the same eigenvalues different from zero as  $\tilde{\mathbf{C}} = \frac{1}{N-1}\mathbf{Y}\mathbf{Y}^T$  which is a  $[N, N]$  matrix.
  - The eigenvectors of  $\mathbf{C}$  can be computed from the ones of  $\tilde{\mathbf{C}}$

$$\mathbf{C}\mathbf{U} = \mathbf{U}\mathbf{D} \rightarrow \tilde{\mathbf{C}}(\mathbf{Y}\mathbf{U}) = (\mathbf{Y}\mathbf{U})\mathbf{D} \rightarrow \tilde{\mathbf{C}}\tilde{\mathbf{U}} = \tilde{\mathbf{U}}\mathbf{D} \quad (39)$$

- Thus,  $\tilde{\mathbf{U}} = \mathbf{Y}\mathbf{U}$

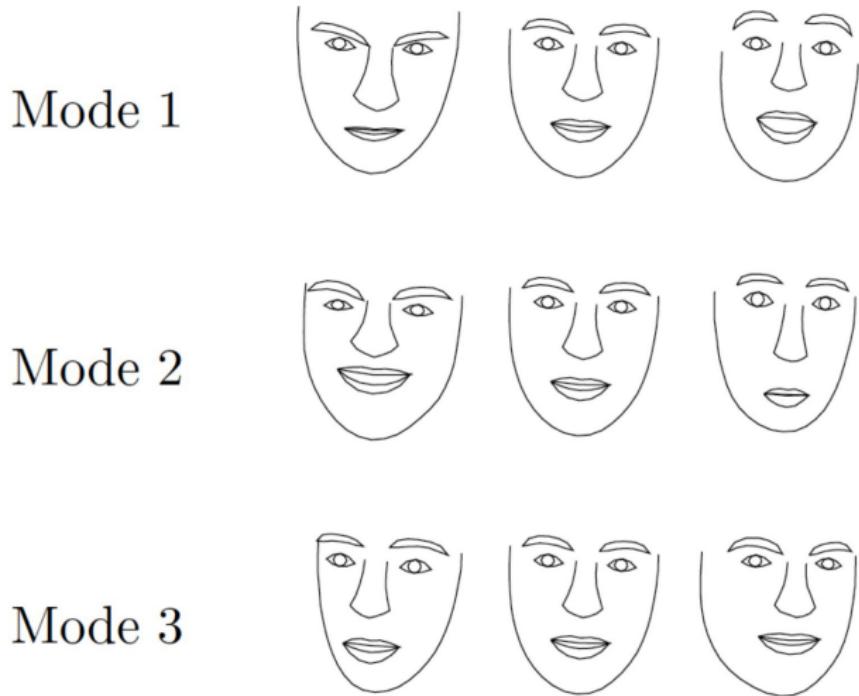
# PCA

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

$$z_i = L^T(y_i - \bar{y}) \rightarrow y_i \approx \bar{y} + Lz_i \quad (40)$$

- 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 assumptions behind PCA)
- Since  $\lambda_j$  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 9:** 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].



**Figure 10:** 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

## 1 Landmark based registration

- Affine registration
- Procrustes superimposition
- Non-linear registration (small displacement)
- Non-linear registration (diffeomorphism)

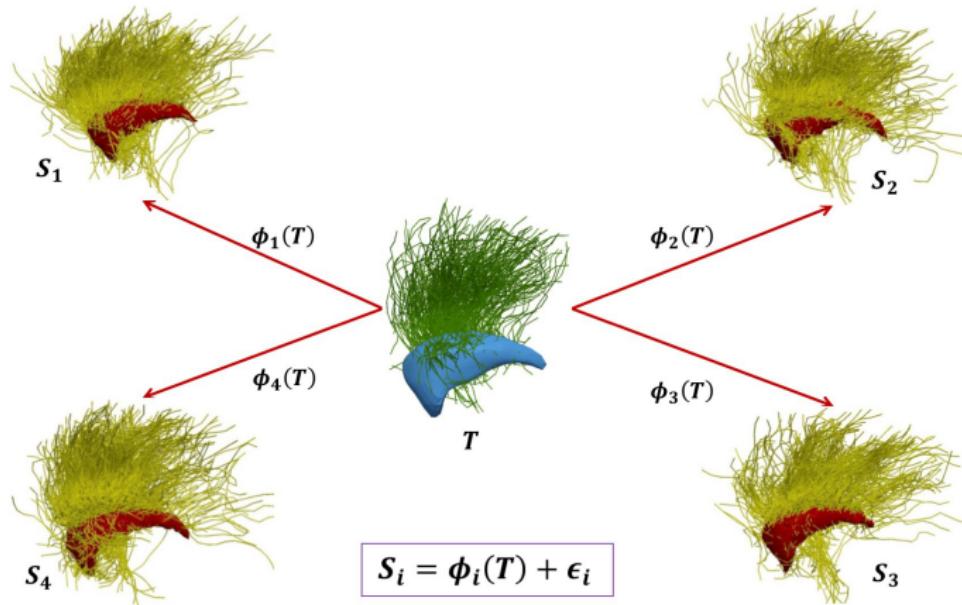
## 2 Statistical Shape Analysis

- Generalized Procrustes Analysis (GPA)
- Tangent space projection
- Shape variations
- **Atlas constructions**
- Atlases-Templates

# 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

# Atlas constructions



- 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  $\phi_i$  (i.e. diffeomorphism) of  $T$  plus a residual error  $\epsilon_i$

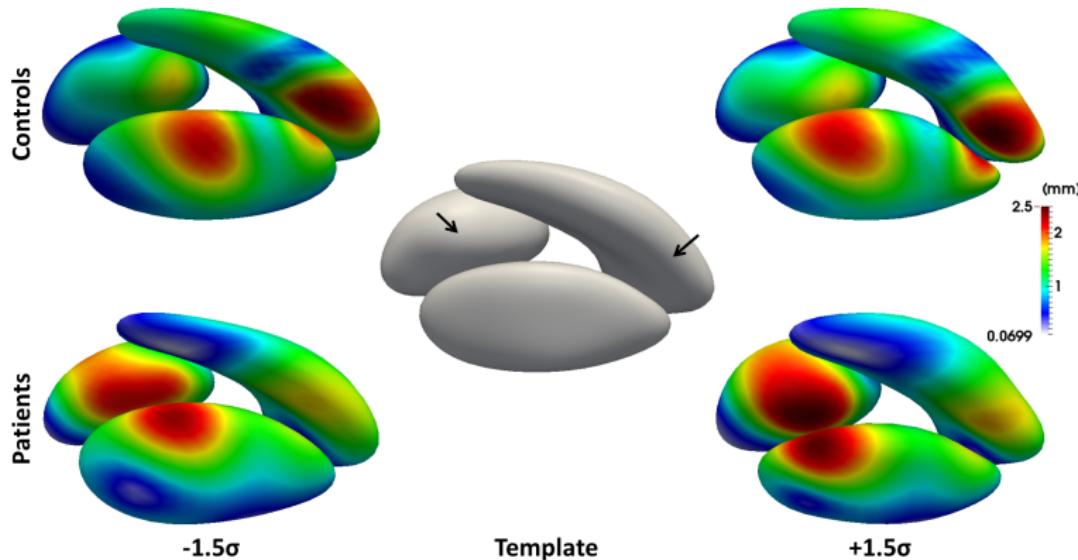
# Atlas construction

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

$$T^*, \{\alpha_i^*\} = \arg \min_{T, \alpha_i} \sum_{i=1}^N \|S_i - \phi_i^{\alpha_i}(T)\| + \gamma \text{Reg}(\phi_i^{\alpha_i}) \quad (41)$$

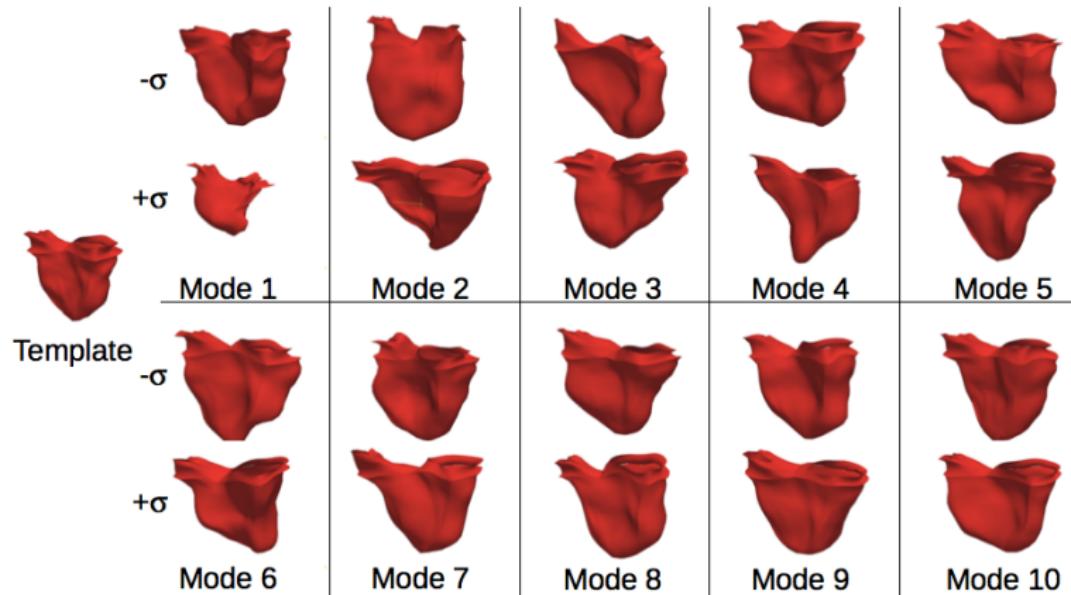
- The parameters are the average shape  $T$  and the deformation parameters  $\alpha_i$ , one for every subject  $i$ . They may be, for instance, the initial velocities  $v_0$  of the diffeomorphisms (see previous lecture)
- Once estimated them, we can use a PCA to study the shape variability within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

# Atlas construction



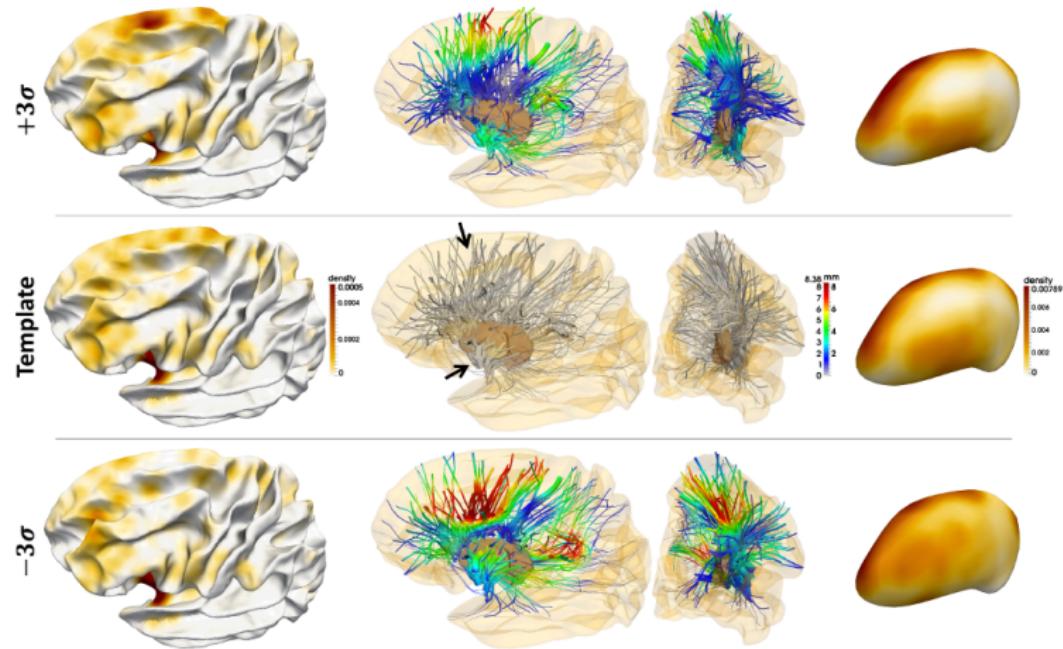
**Figure 11:** First mode of a PCA computed using the deformation parameters  $\alpha_i$  and three anatomical structures of the brain. We compare the results for two different populations. The template has been computed considering both populations together.

# Atlas construction



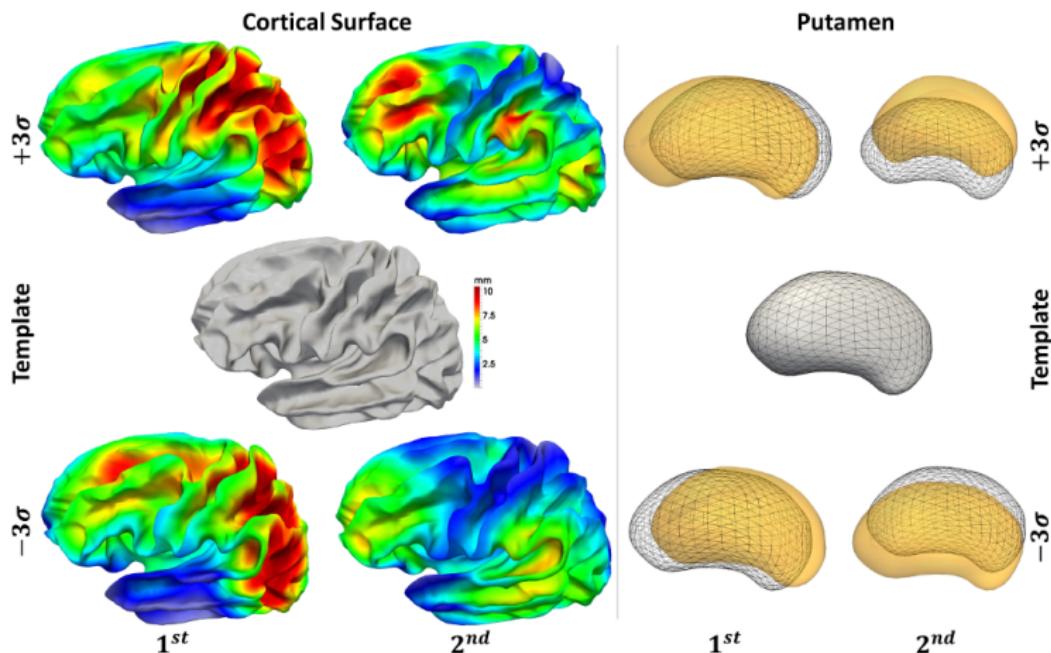
**Figure 12:** PCA deformation modes on a population of 18 patients suffering from repaired Tetralogy of Fallot. Image taken from T. Mansi - MICCAI - 2009.

# Atlas construction



**Figure 13:** Structural connectivity changes in a population composed of both controls and patients with Gilles de la Tourette syndrome.

# Atlas construction



**Figure 14:** Morphological changes in a population composed of both controls and patients with Gilles de la Tourette syndrome.

# Summary

## 1 Landmark based registration

- Affine registration
- Procrustes superimposition
- Non-linear registration (small displacement)
- Non-linear registration (diffeomorphism)

## 2 Statistical Shape Analysis

- Generalized Procrustes Analysis (GPA)
- Tangent space projection
- Shape variations
- Atlas constructions
- Atlases-Templates

# 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)

# Atlases-Templates

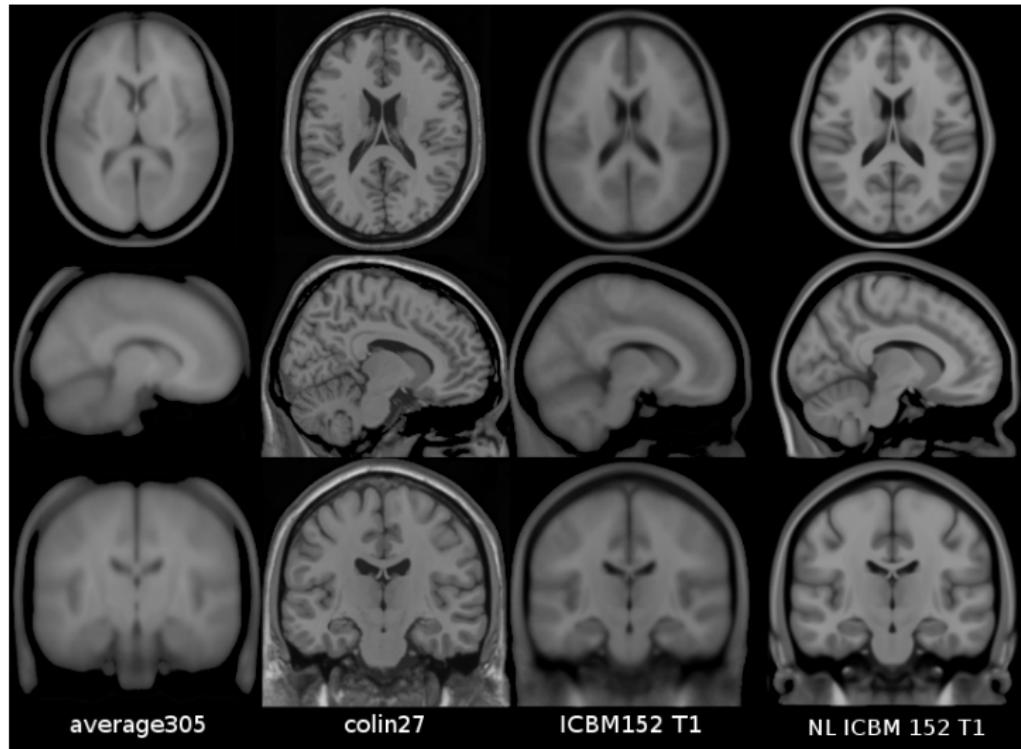


Figure 15: Examples of neuroimaging atlases from T1-w MRI

## References

- ① J. Modersitzki (2004). *Numerical methods for image registration*. Oxford university press
- ② J. V. Hajnal, D. L.G. Hill, D. J. Hawkes (2001). *Medical image registration*. CRC press
- ③ G. Wolberg (1990). Digital Image Warping. IEEE
- ④ The Matrix Cookbook
- ⑤ S. Umeyama (1991). Least-squares estimation of transformation parameters.... IEEE TPAMI
- ⑥ S. Durrleman et al. (2014). Morphometry of anatomical shape complexes . . . . NeuroImage.
- ⑦ J. Ashburner (2007). A fast diffeomorphic image registration algorithm. NeuroImage
- ⑧ J. P. W. Pluim (2003). Mutual information based registration of medical images: a survey. IEEE TMI
- ⑨ S. Baker, I. Matthews (2004). Lucas-Kanade 20 Years On: A Unifying Framework. IJCV

## 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 P.H. Schonemann (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika*
- ⑦ L. Younes (2010). Shapes and Diffeomorphisms. Springer
- ⑧ Joshi et al. (2004). Unbiased diffeomorphic atlas construction... *NeuroImage*