Medical image registration

Outline

Introduction

- What is image registration?
- Motivation and main applications

Problem formulation

- Mathematical definition
- General framework

Main components

- *Features* : which information to use in the registration
- *Similarity metrics* : measure how similar two images are
- *Transforms* : deformation model to transform one image into another
- *Optimizers* : algorithm to estimate the transformation
- *Interpolators* : how to compute common coordinates from different images

Image registration

- Registration is the **process of finding the transformation** (T) that puts different images (**f** and **g**) into spatial correspondence

- Example

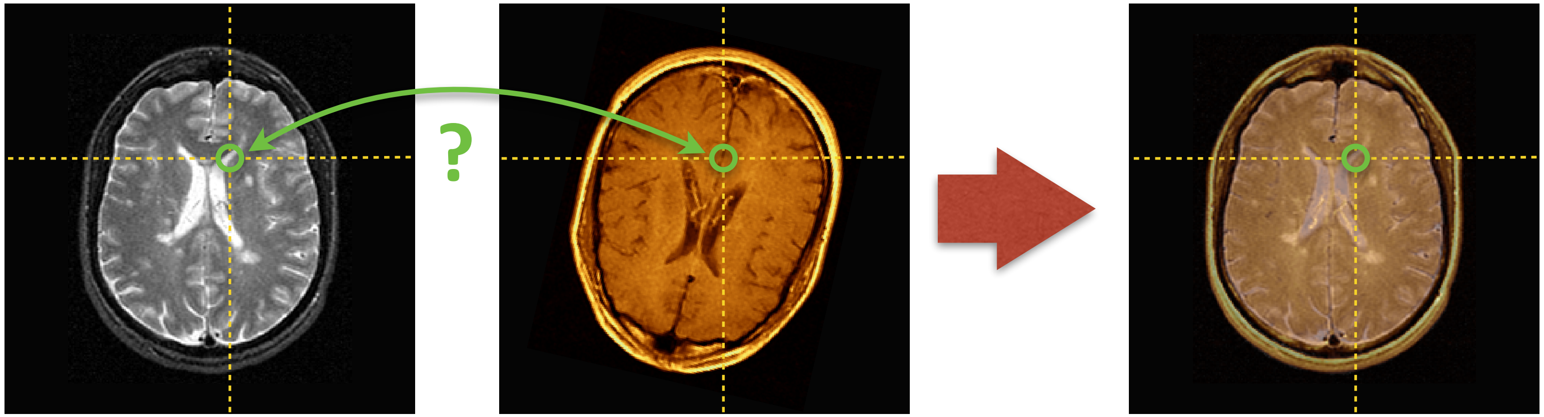
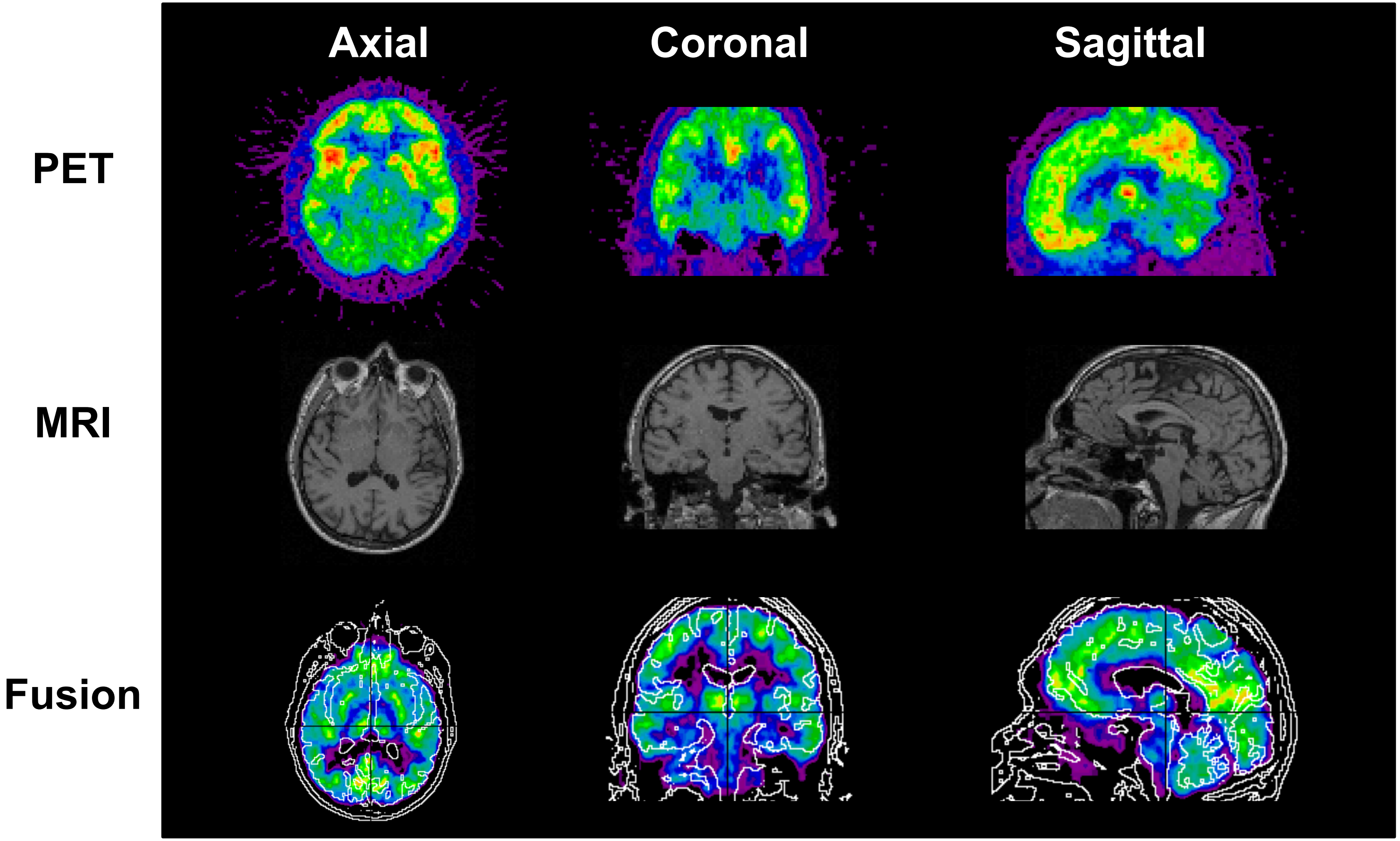


image **f** image **g** **g** registered to **f**

Why do we need to register medical images? (1/7)

Improve diagnosis

- Combining information from multiple imaging modalities

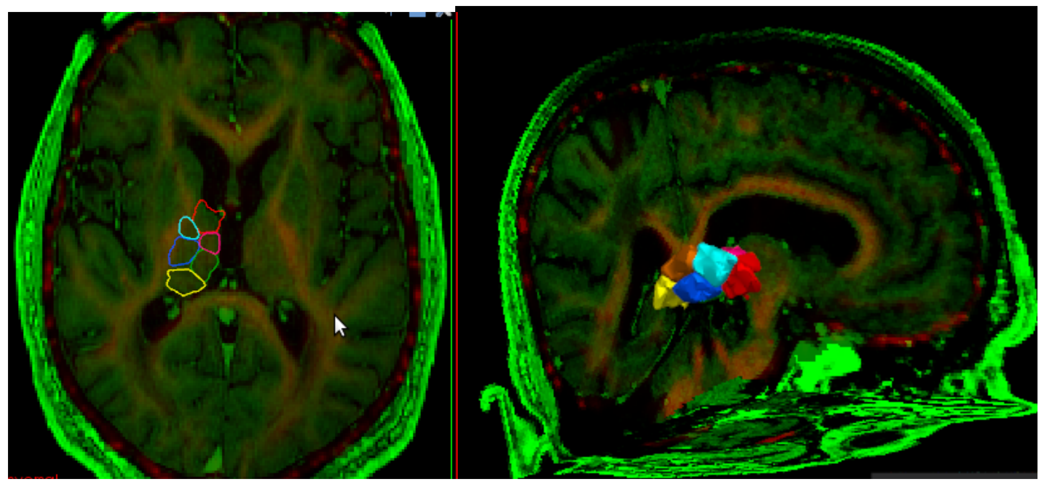


Why do we need to register medical images? (2/7)

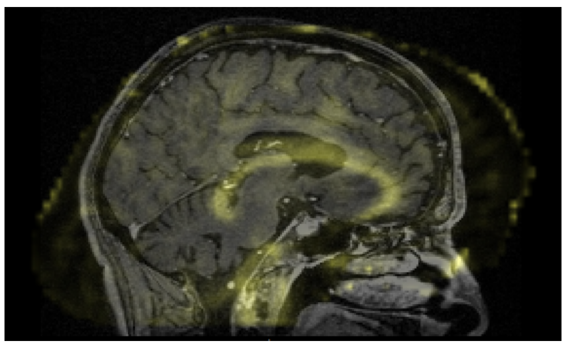
Image guided surgery or radiotherapy

- Image guided surgery or radiotherapy

After registration



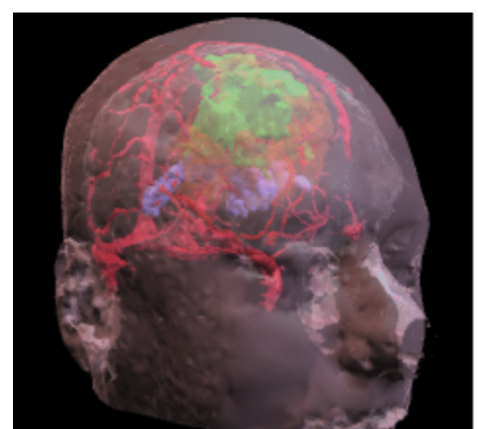
Before registration

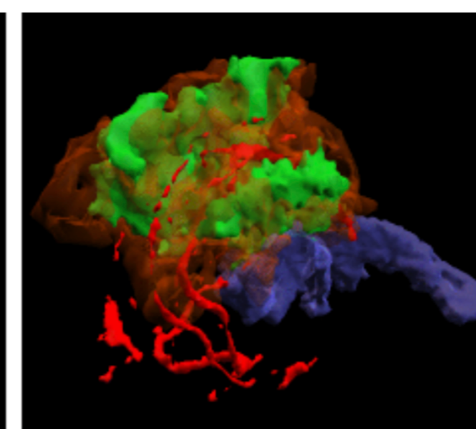


VIM targeting for therapy of movement disorders

- **T1w** : thalamus segmentation/delineation
- **DWI** : clustering of thalamus nuclei

(PhD thesis of E. Najdenovska @ EPFL)

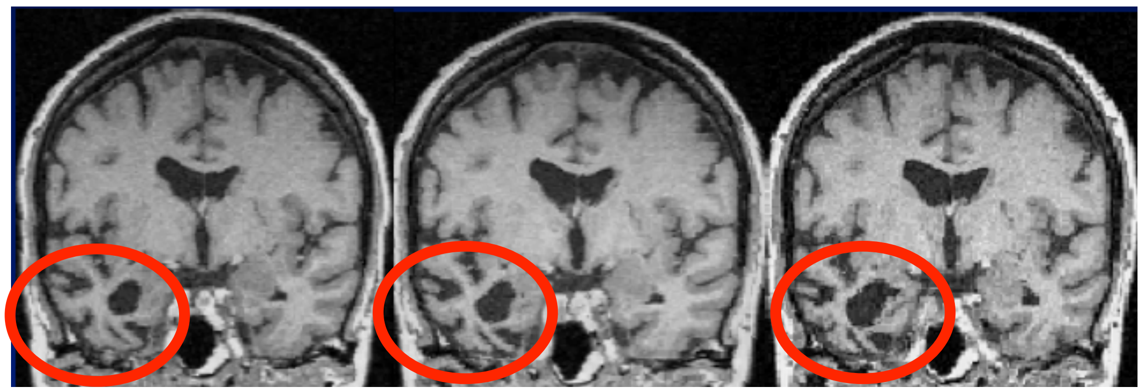




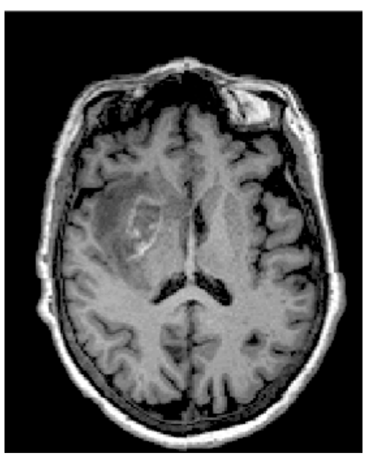
Why do we need to register medical images? (3/7)

Study **disease progression**

- Monitoring changes in size, shape, position or image intensity over time

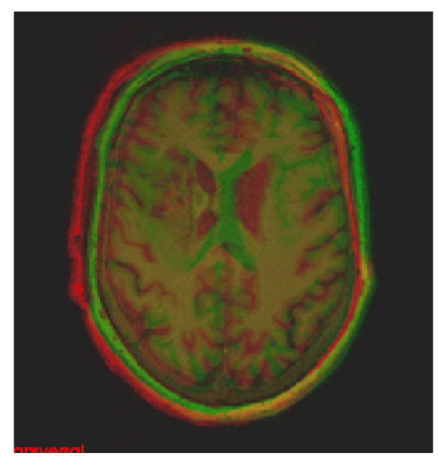


2000 2001 2002

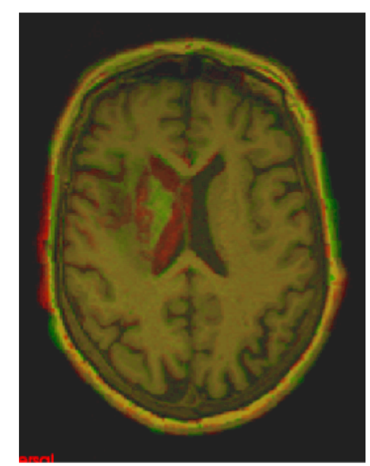


Time 1

Time 2

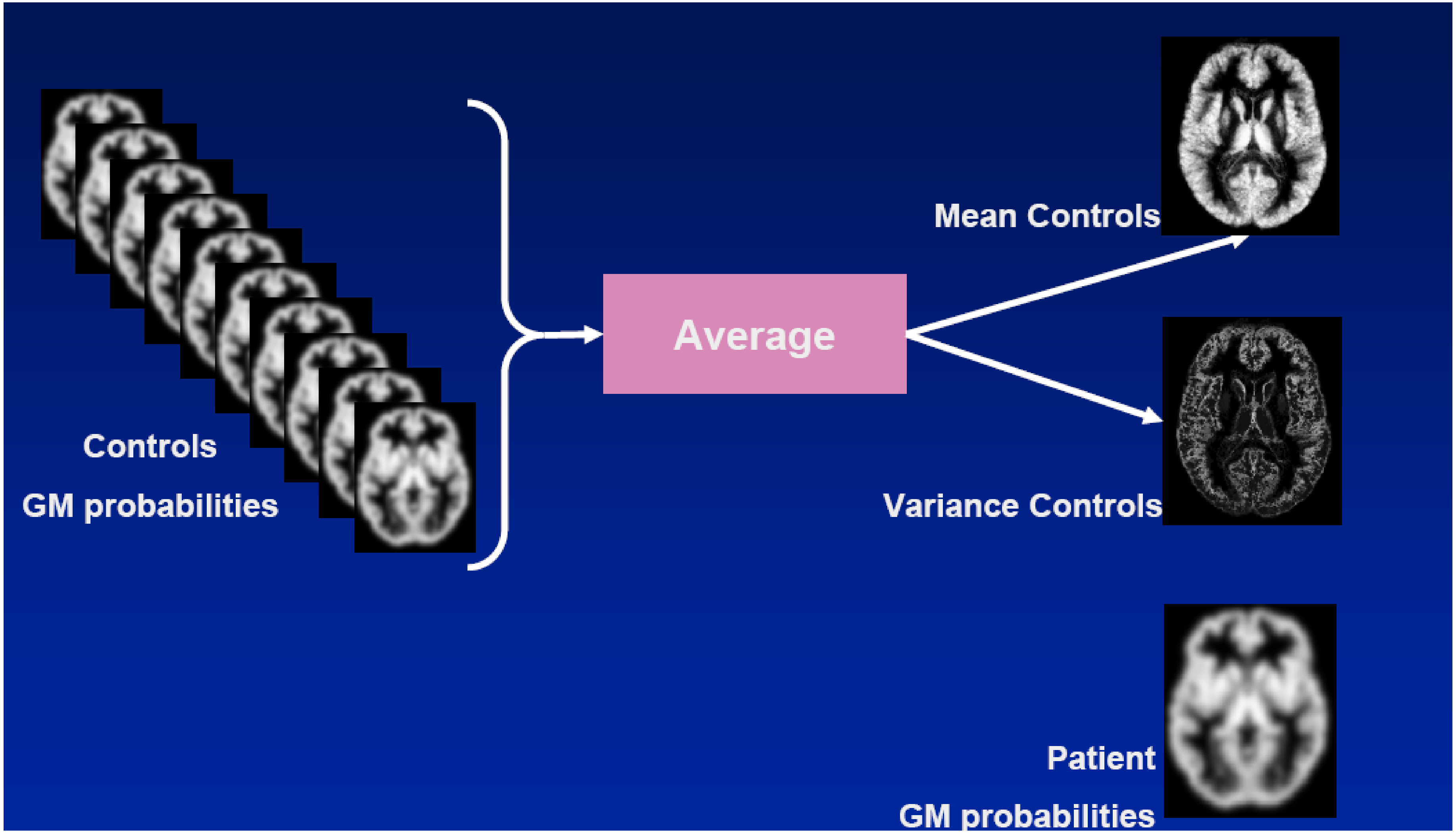


Before registration



After registration

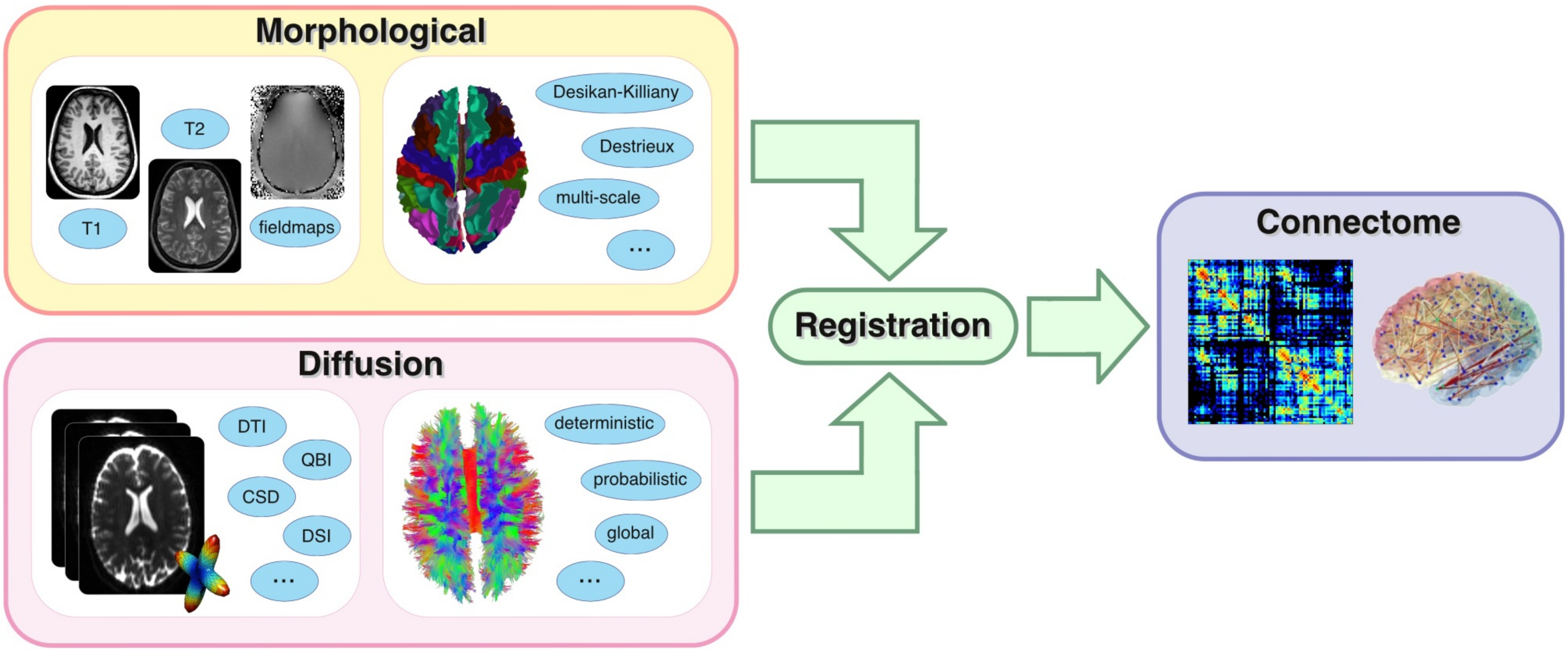
- **Patient comparison** (group studies) or **atlas construction**
  - Relating one individual's anatomy to a standardized atlas or group of subjects



Why do we need to register medical images? (5/7)

Estimating **brain connectivity** from diffusion MRI

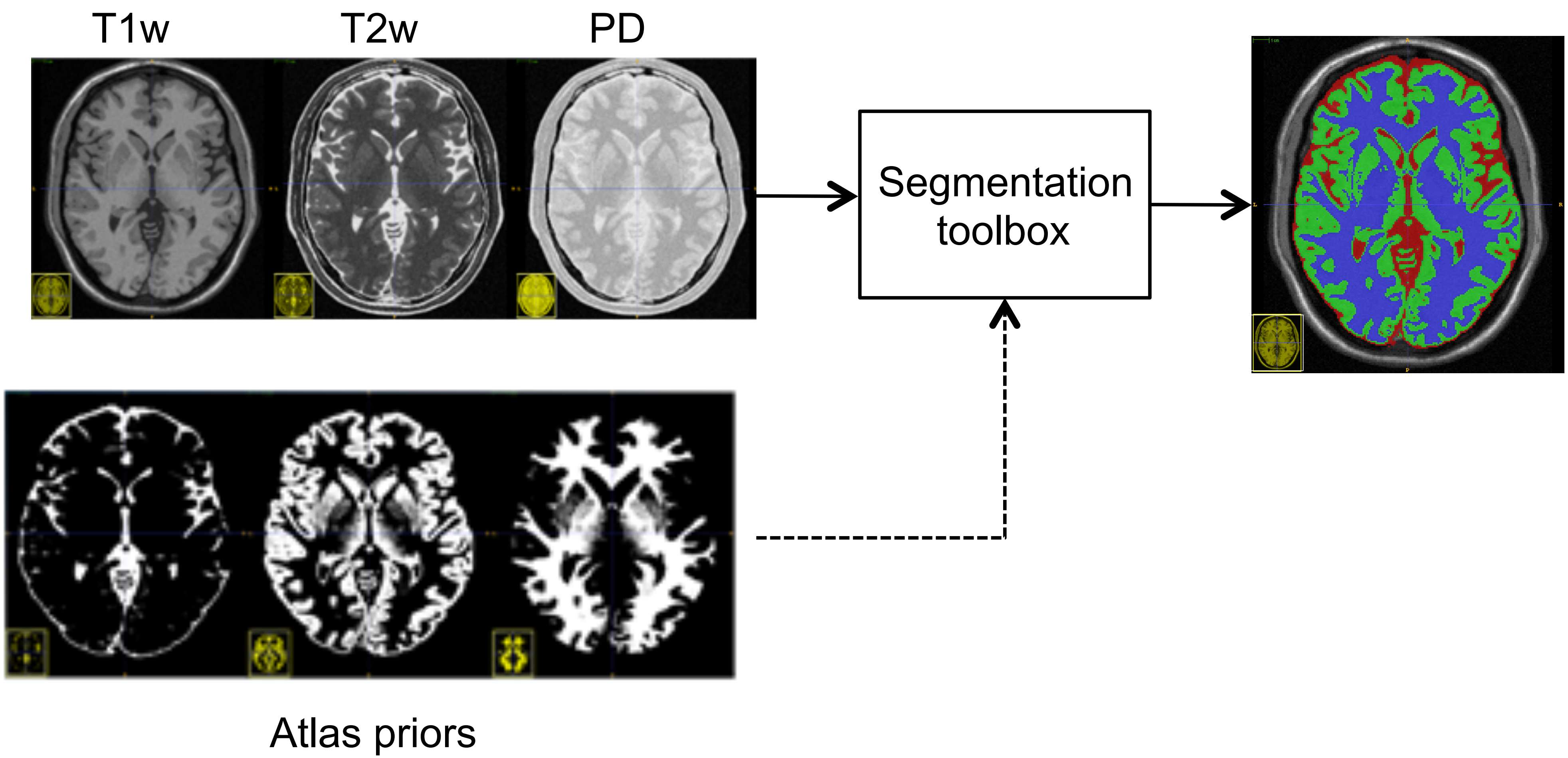
- Estimate *fiber bundles* from diffusion MRI, i.e. DWI
- Define *cortical segmentation* from structural MRI, e.g. T1w



http://hardi.epfl.ch

Multi-spectral **segmentation**

- Use more than one modality to improve the segmentation of brain anatomy

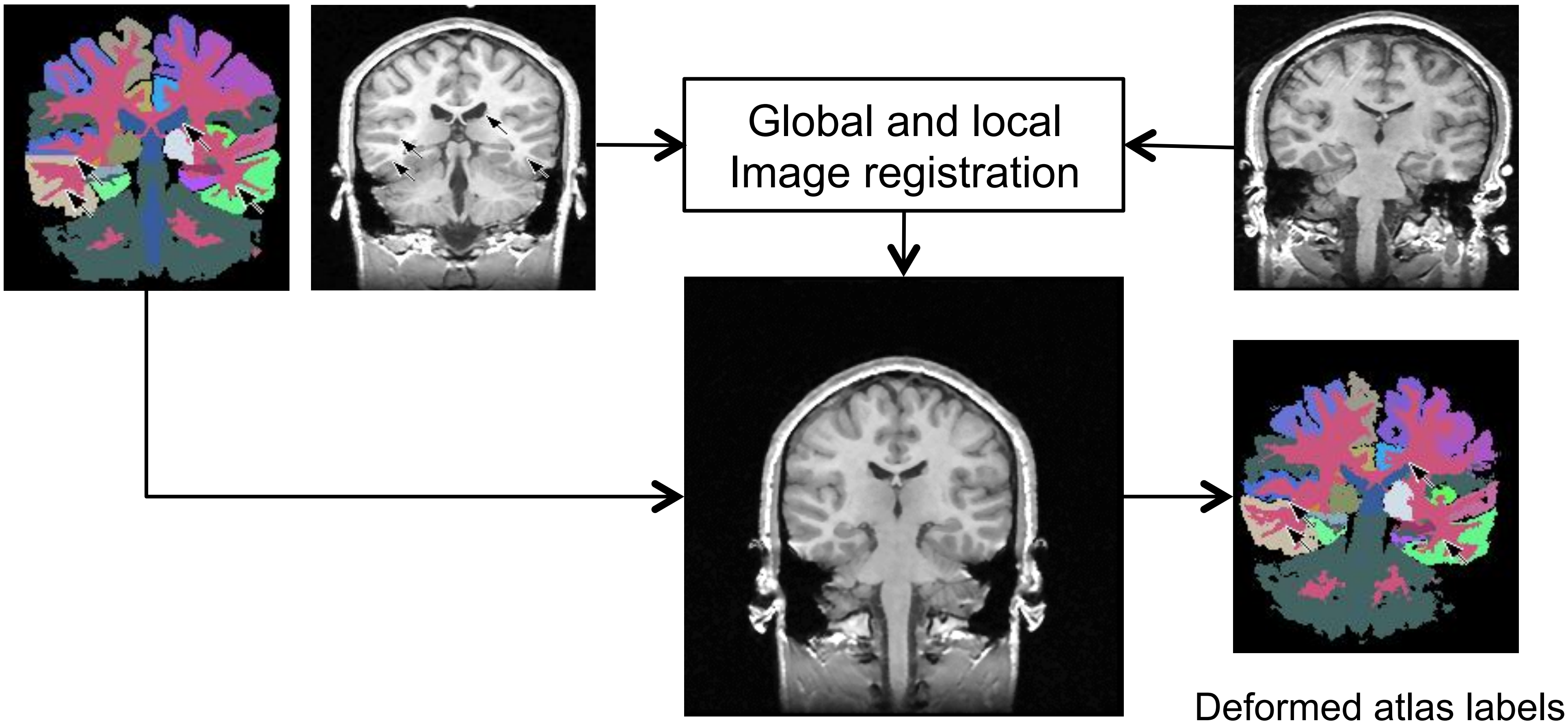


(PhD thesis of O. Esteban @ Madrid+EPFL)

Why do we need to register medical images? (7/7)

Atlas-based **segmentation**

- Use an accurate atlas to define one subject's anatomy

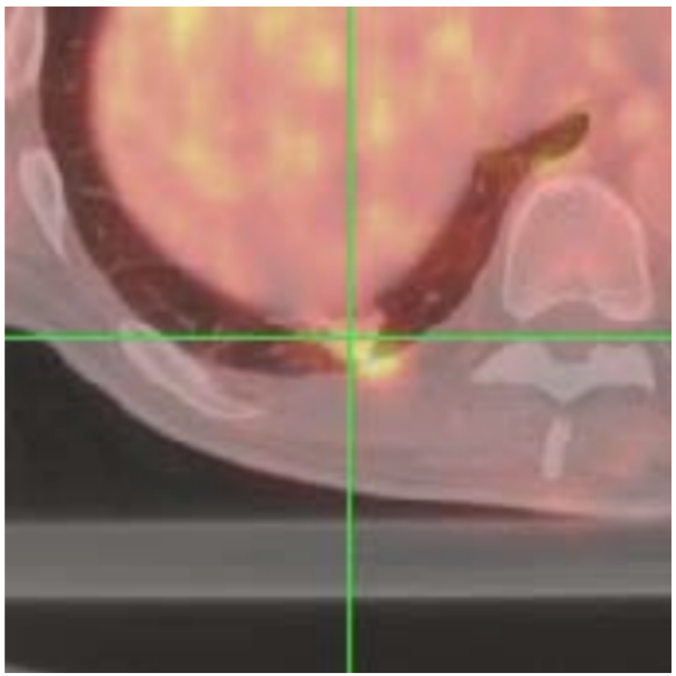


(PhD thesis of S. Gorthi @ EPFL)

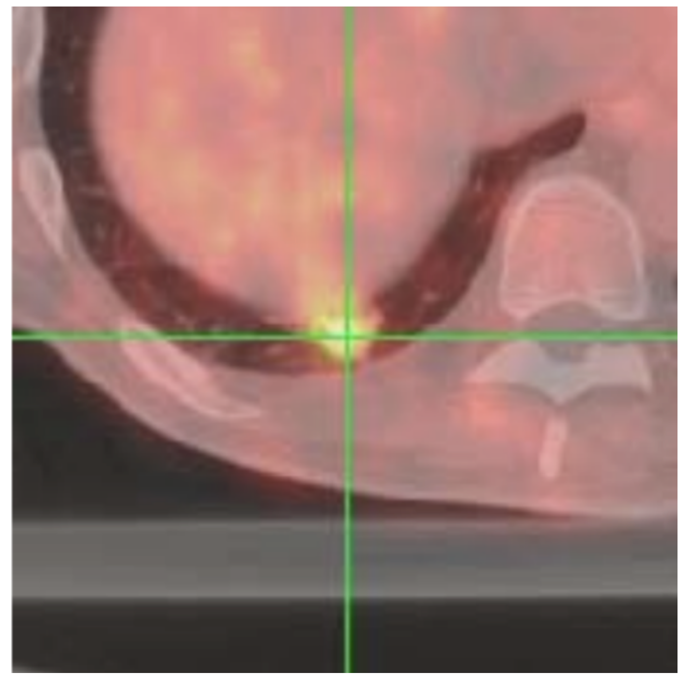
Registration is very important

- In **medical imaging**, registration is particularly important
  - **Example**: *PET-MRI* registration to study tumor location

Registration algorithm 1



Registration algorithm 2

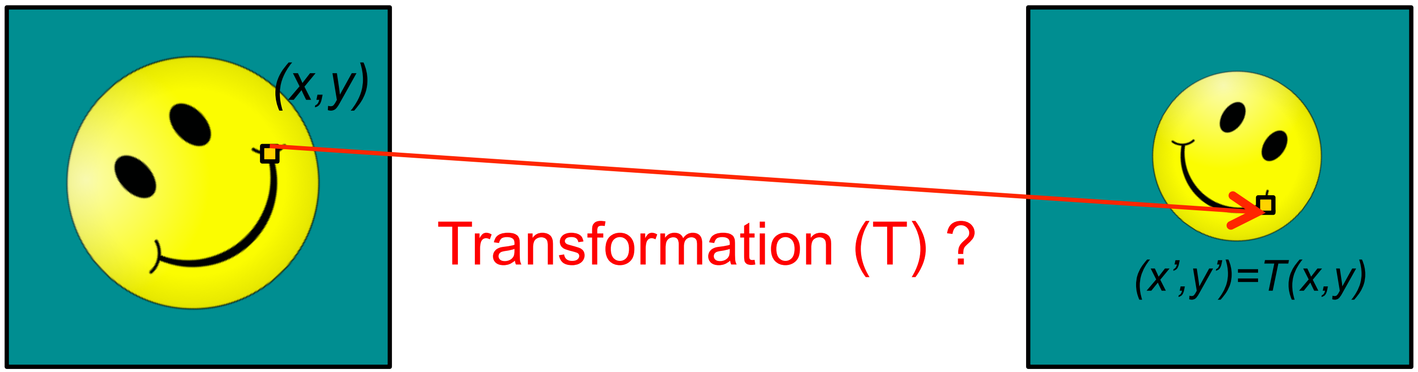


  - Is the tumor in the lung only?
  - Algorithm #2 looks more plausible:
    are you ready to risk your software against getting sued?

Mathematical formulation (1/2)

Registration is an **alignment problem**

- "...*find the spatial transformation* that maps points from one image B to the corresponding points in another image A..."



Mathematical formulation (1/2)

Registration is an **alignment problem**

- "…*find the spatial transformation* that maps points from one image B to the corresponding points in another image A…"

Usually solved as **energy minimization problem** (or maximization)

optimal transformation

$$\mathcal{T}^* = \underset{\mathcal{T}\in\mathbb{T}}{\operatorname{argmin}}\, d(A, B^{\mathcal{T}})$$

images

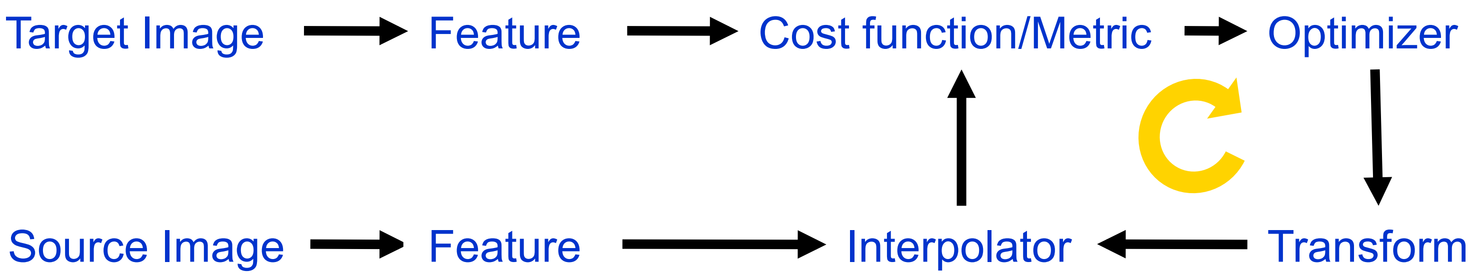
space of all possible transformations

similarity/dissimilarity between the images

Notation

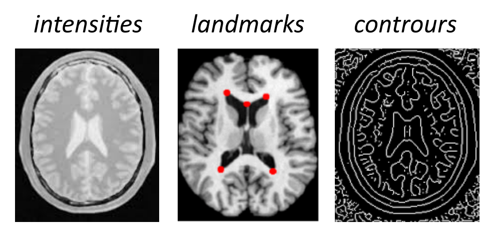
$A : \mathbf{x}_A \in \Omega_A \mapsto A(\mathbf{x_A})$	Intensity of image A at location $\mathbf{x}$
$\mathbf{T} : \mathbf{x}_B \mapsto \mathbf{x}_A \iff \mathbf{T}(\mathbf{x}_B) = \mathbf{x}_A$	Transforms a position $\mathbf{x}$ from one image to another
$\mathcal{T}$	Transforms an image (both coordinates $\mathbf{x}$ and intensities)
$B^{\mathcal{T}}$	Image B transformed
$\Omega^{T}_{A,B} = \{\mathbf{x}_A \in \Omega_A \vert \mathbf{T}^{-1}(\mathbf{x}_A) \in \Omega_B\}$	Overlap domain after transformation $\mathbf{T}$

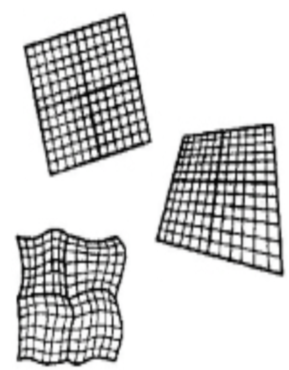
General framework

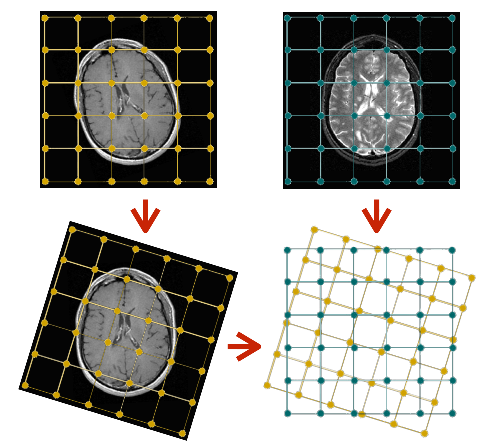


The main actors

- **Feature**
  - Which information to use for driving the registration
- **Similarity metric**
  - Measures of how similar the features are in the two images
- **Interpolator**
  - How to compute similarity metrics from different grids
- **Transform**
  - The deformation model to transform one image into another
- **Optimizer**
  - The optimization algorithm to estimate the transformation

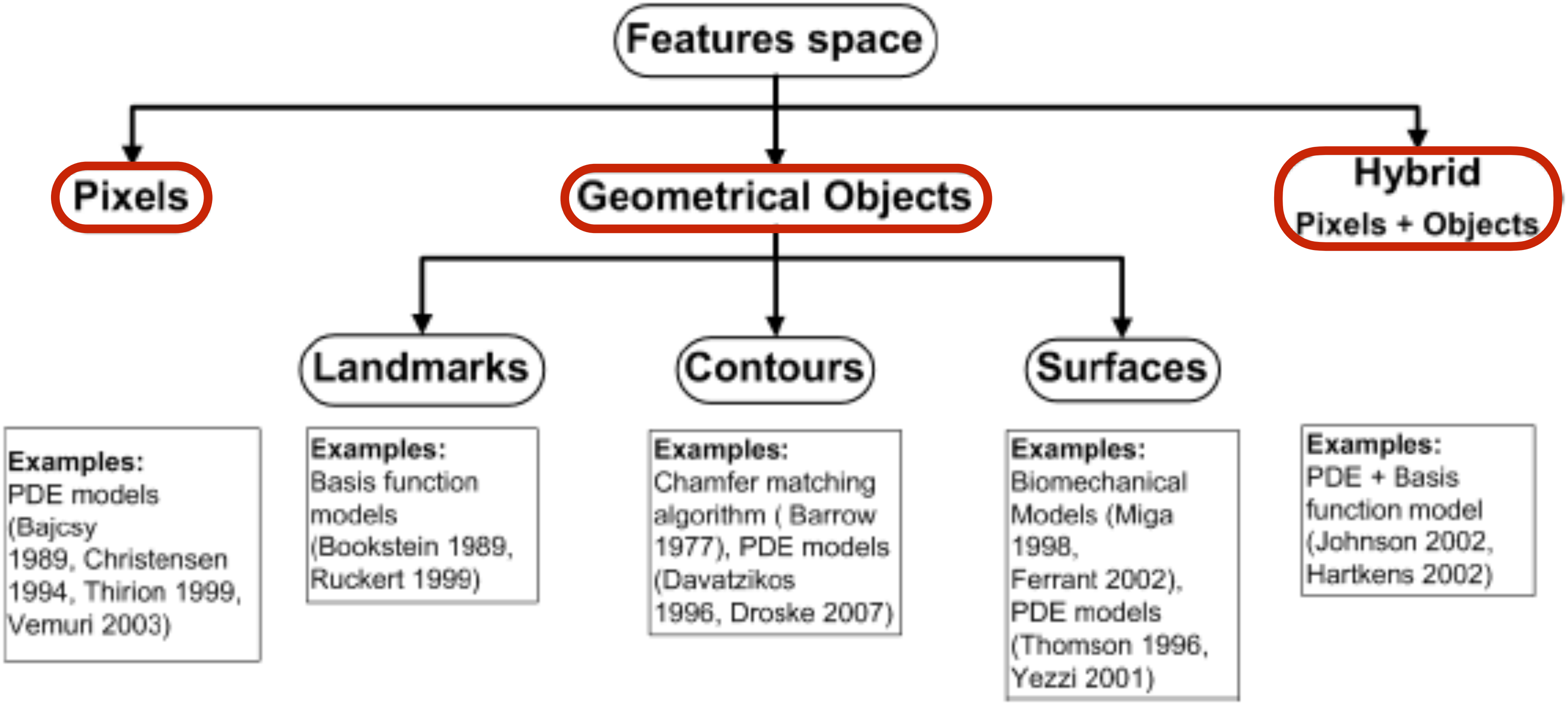






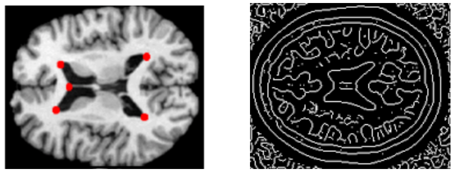
Two main approaches

- **Feature based**: use corresponding points or features in the images to align them
- **Intensity based**: operate directly on the image intensities



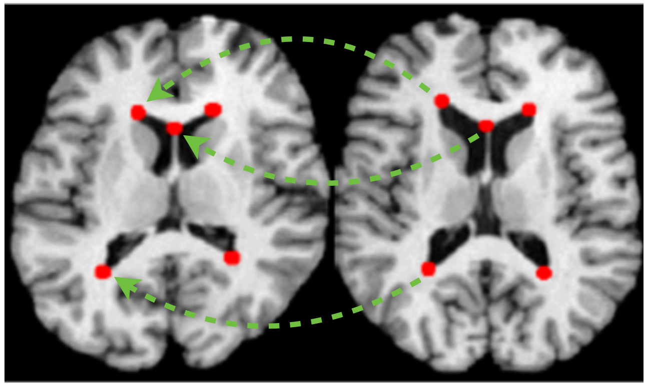
(PhD thesis of V. Duay @ EPFL)

Feature based approach



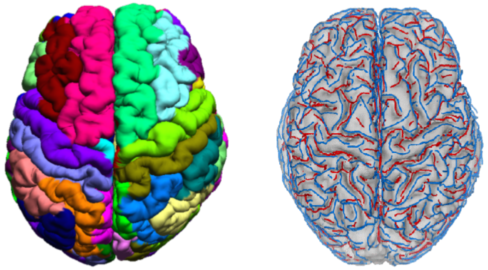
- Extract corresponding features from both images
- Compute transformation $\mathbf{T}$ by minimizing some "*measure of distance*" between them

Example: landmarks



- Identify "*fiducial markers*" on the images
  - Internal anatomical structures, e.g. anterior commissure
  - Pins/markers fixed to the patient, e.g. skin markers
- *Compute the centroid* of each point cloud
  - Difference between centroids = *translation* that must be applied
- *Rotate* one point-set until the *distance between corresponding points* is minimized
  - Iterative Closest Point (ICP) algorithm

Can be extended to other features

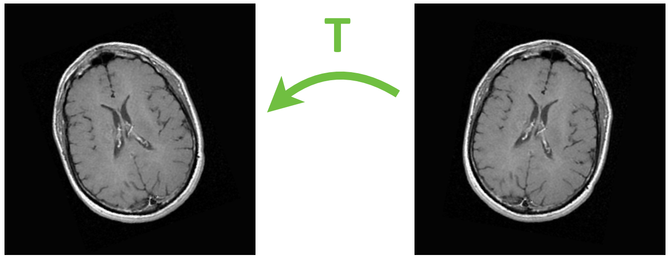


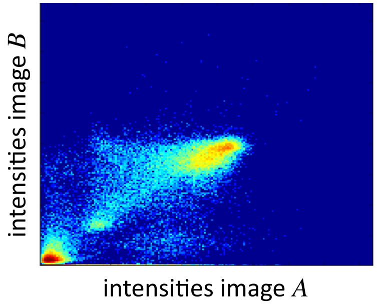
- e.g. *surfaces* ("*Head and Hat*") or *contours* ("Crest Lines")
- **Critical** : define a good *similarity metric* for that feature

I - Features of interest (3/3)

Intensity based approach

- Use the intensities in the two images alone
  - No need to delineate corresponding structures
- Like having "features = pixels"
- Transformation $\mathbf{T}$ computed by *comparing intensity patterns* in both images via "**pixel similarity metrics**"
- These are based on the **joint histogram**



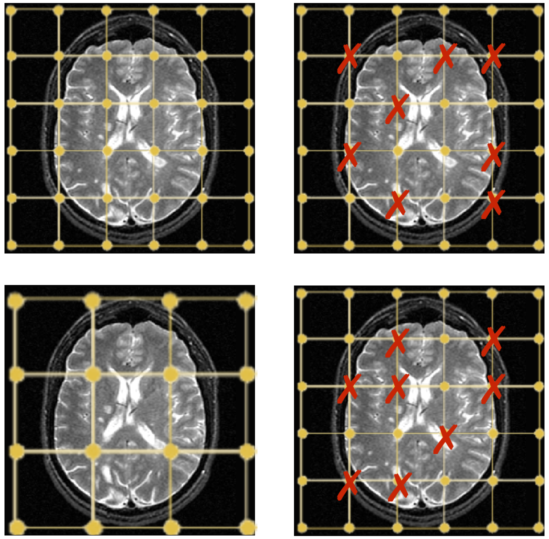


NB: we will focus on this approach

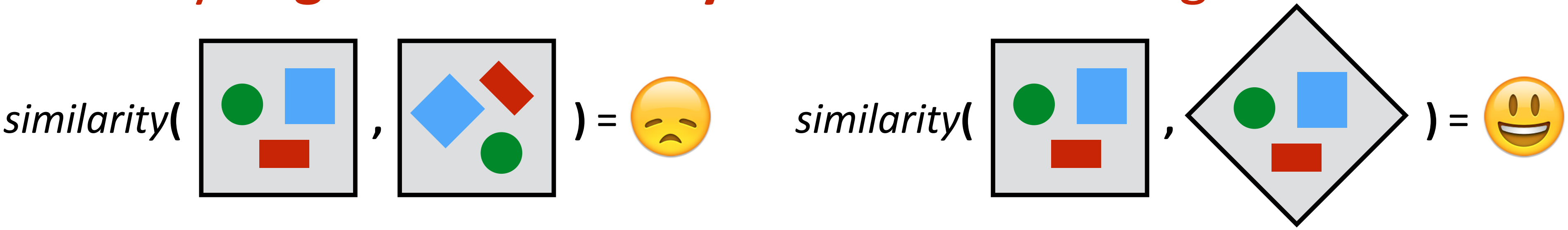
(it's the most used in medical imaging)

Image **sampling strategy**

- **Full sampling**
  - Similarity metrics computed on *all voxels of the image*
- **Subset sampling**
  - In general, it is not necessary to evaluate all voxels
  - Examples : **subsampled regular grid**, **random locations** …

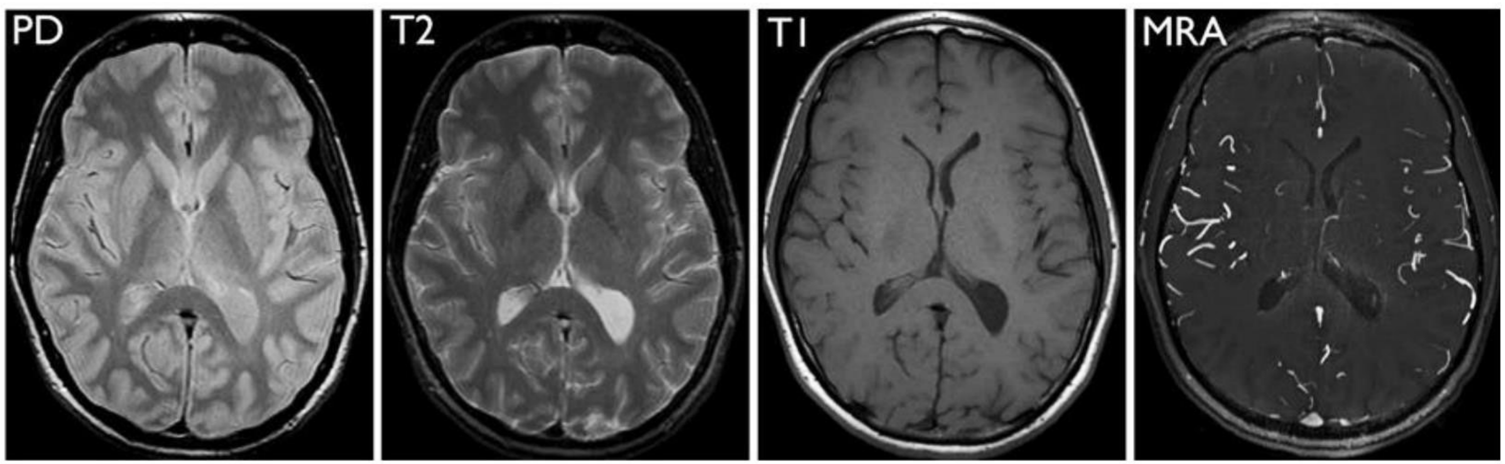


Quantify **degree of similarity** between two images



Example

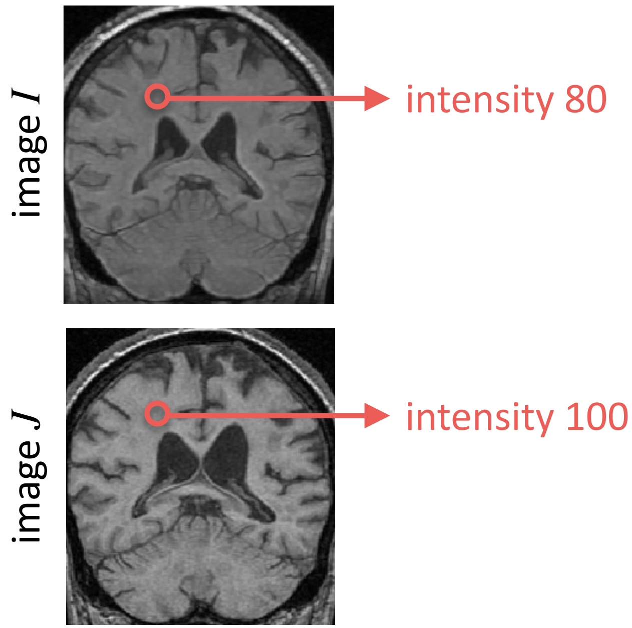
- Same subject/session, but images from different modalities look different

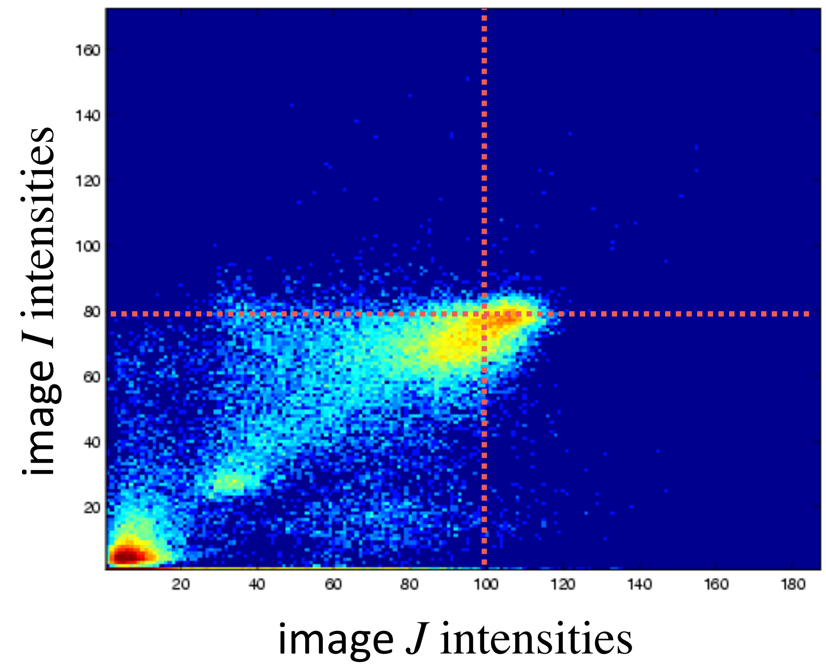


- How to construct a metric to realize they are all the "same object"?
- $\sum_{\mathbf{x}\in\Omega_A} |A(\mathbf{x}) - B^{\mathcal{T}}(\mathbf{x})|$ would be very high in any case. Any other idea?

Joint histogram

$$H_{I,J}(i,j) = \mathrm{Card}\left\{(x,y) | I(x,y) = i \text{ and } J(x,y) = j\right\}$$





Notes

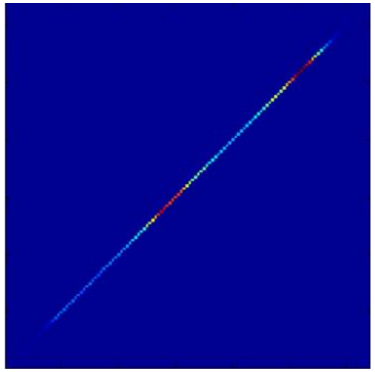
- I and J must have the *same dimensions, e.g* $M \times N$ (**NB**: in this context J = $B^{\mathcal{T}}$)
- If I and J have *intensities* in [0...255]
  - $size(\ H_{I,J}\ )$ = 256×256 and $sum(\ H_{I,J}\ ) = M \cdot N$

II - Similarity measures (1/5)

Examples

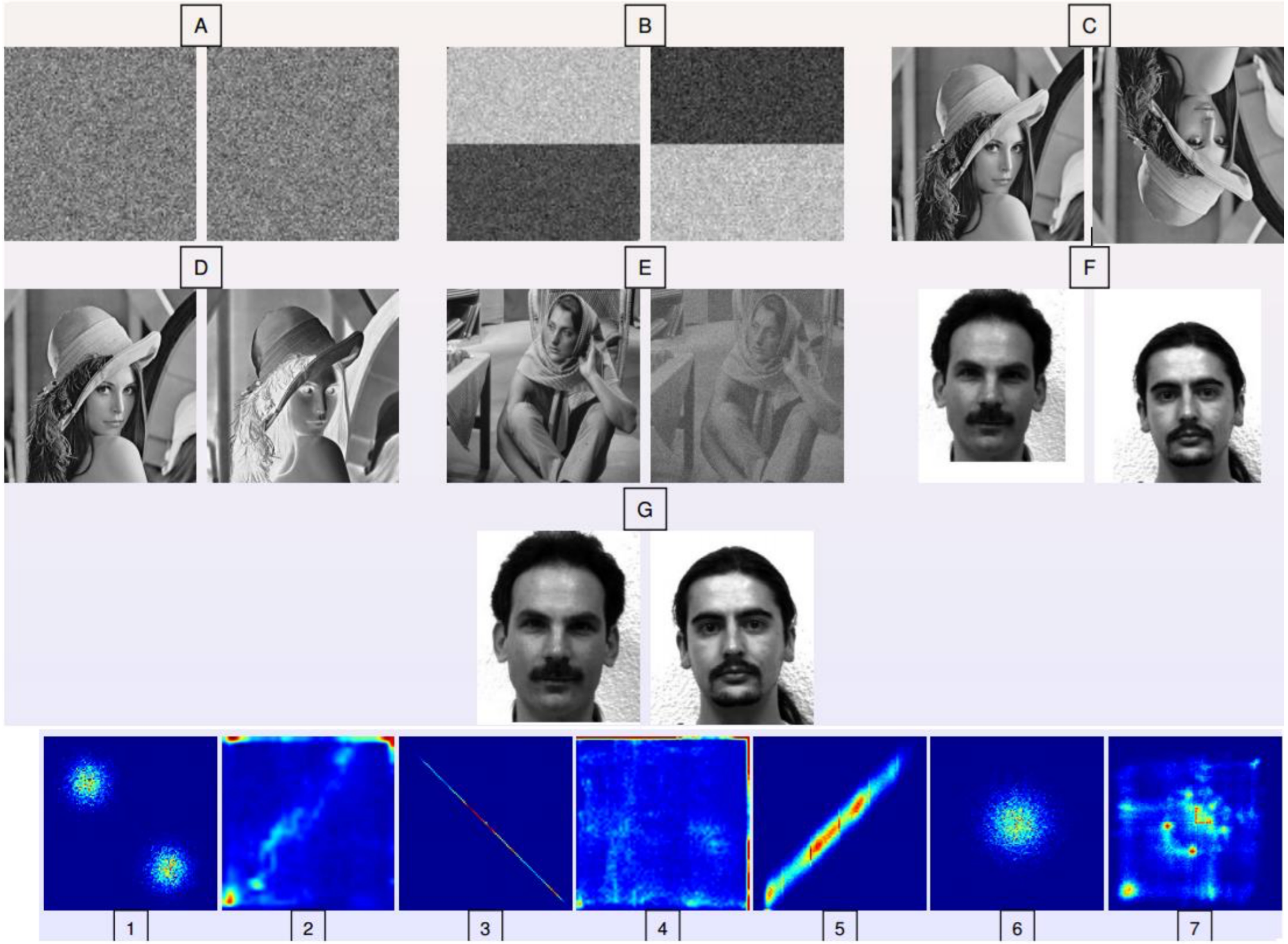
image I

image J=I



$H_{I,J}$

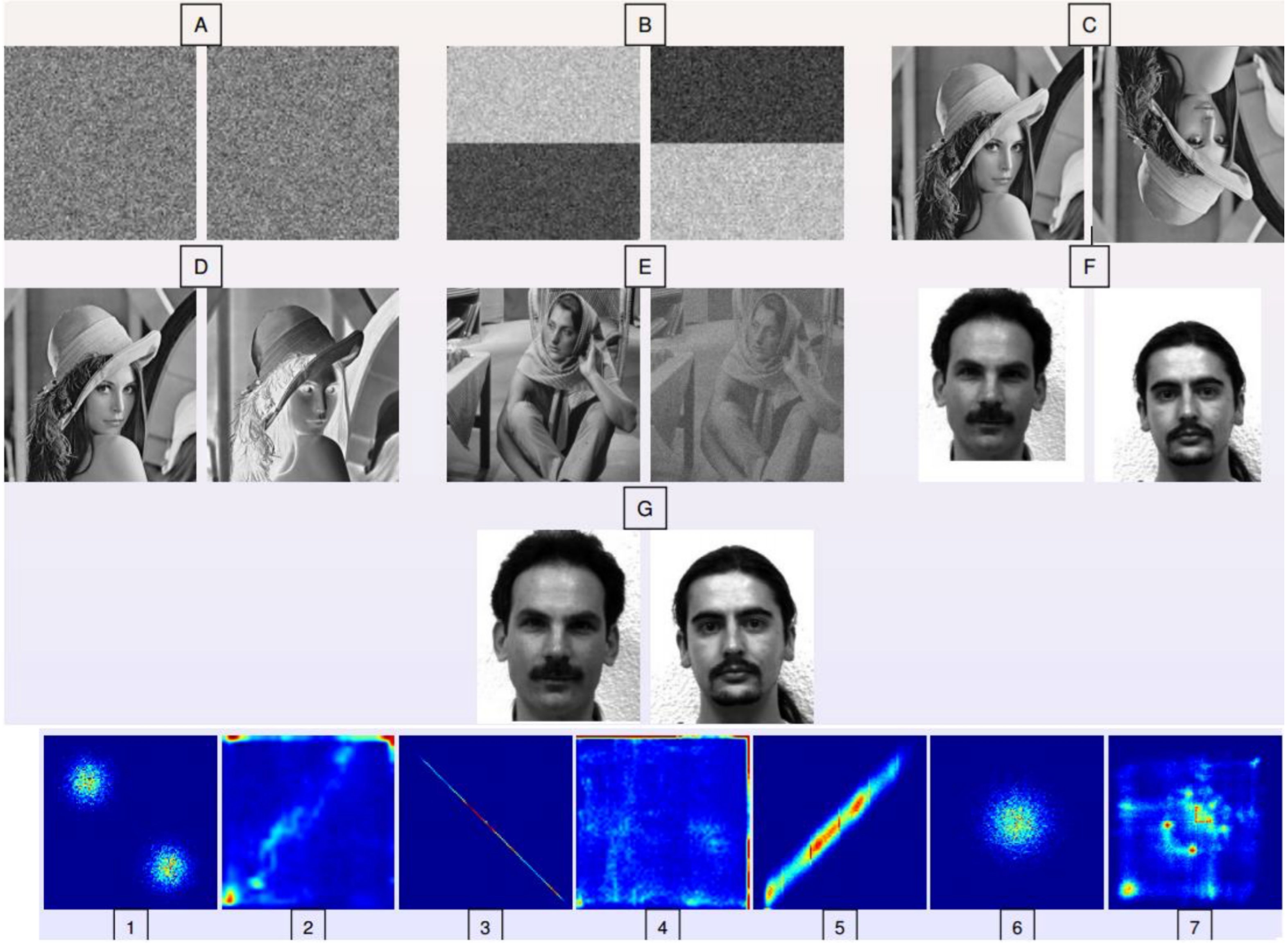
Examples



II - Similarity measures (1/5)

Examples

A→6, B→1, C→7, D→3, E→5, F→4 , G→2



II - Similarity measures (2/5)

Minimizing intensity differences

- **Sum of squared differences** (SSD)

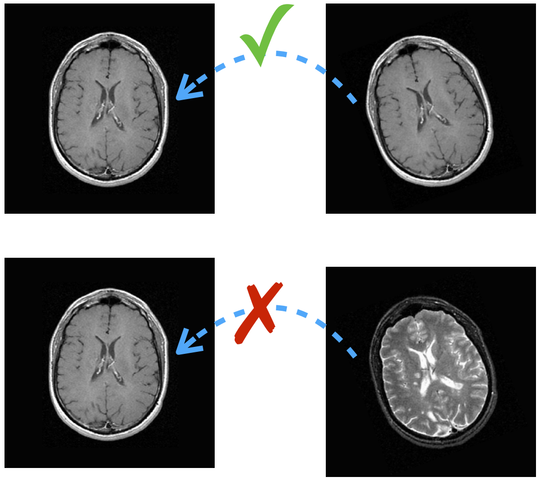
$$\mathrm{SSD} = \sum_{\mathbf{x}_A \in \Omega^{\mathcal{T}}_{A,B}} \left| A(\mathbf{x}_A) - B^{\mathcal{T}}(\mathbf{x}_A) \right|^2$$

- SSD *very sensitive* to few voxels with very different intensities between images
  - e.g. contrast agent is injected between two acquisitions
- **Sum of absolute differences** (SAD) reduces the effect of these outliers

$$\mathrm{SAD} = \sum_{\mathbf{x}_A \in \Omega^{\mathcal{T}}_{A,B}} \left| A(\mathbf{x}_A) - B^{\mathcal{T}}(\mathbf{x}_A) \right|$$

Notes

- Computed from $H_{I,J}$: $\mathrm{SSD} = \sum_{i,j} H(i,j) \cdot (i-j)^2$
- SSD/SAD can be used only when “images are the same”
  - Same *modality*, same *contrast*, same *scaling, same visible details*…
  - …but, in practice, this is never the case
  - *Implicit assumption*: after registration the images differ only by Gaussian noise

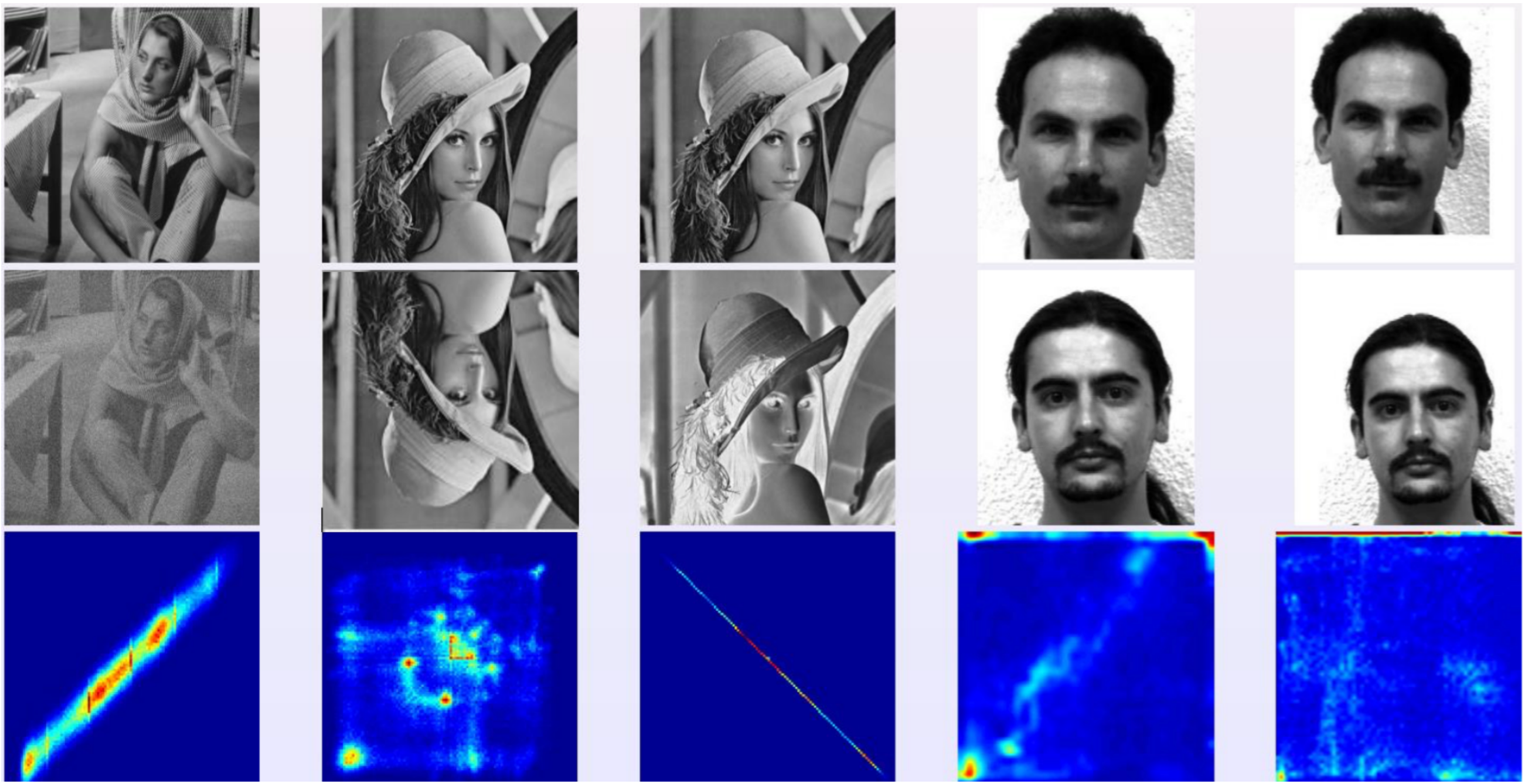


SSD examples

I

J

$H_{I,J}$



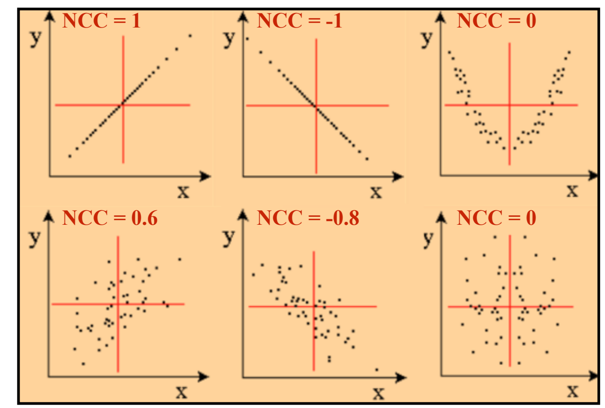
SSD	652	10^9	10^{10}	5600	13400

Correlation approach

- Use a *slightly less strict assumption*
  - We don't try to have $B^{\mathcal{T}} = A$ at registration
  - We require only a relationship of the form $B^{\mathcal{T}} = \alpha A + \beta$ (linear)

- **Cross-Correlation** (CC)

$$\mathrm{CC} = \sum_{\mathbf{x}_A \in \Omega^{\mathcal{T}}_{A,B}} A(\mathbf{x}_A) \cdot B^{\mathcal{T}}(\mathbf{x}_A)$$

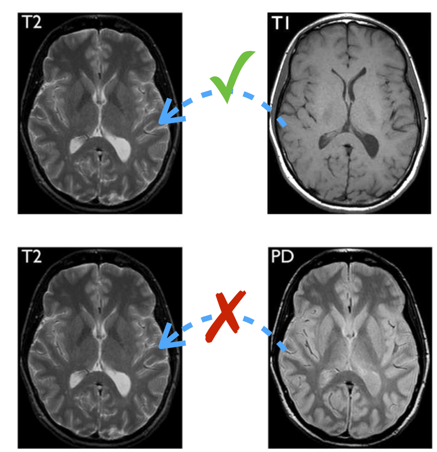


- **Normalized Cross-Correlation** (NCC)

$$\mathrm{NCC} = \frac{\sum_{\mathbf{x}_A \in \Omega^{\mathcal{T}}_{A,B}} (A(\mathbf{x}_A) - \bar{A}) \cdot (B^{\mathcal{T}}(\mathbf{x}_A) - \bar{B})}{\sqrt{\sum_{\mathbf{x}_A \in \Omega^{\mathcal{T}}_{A,B}} (A(\mathbf{x}_A) - \bar{A})^2} \cdot \sqrt{\sum_{\mathbf{x}_A \in \Omega^{\mathcal{T}}_{A,B}} (B^{\mathcal{T}}(\mathbf{x}_A) - \bar{B})^2}}$$

Notes

- NCC(I,J)∈[-1,1] $\forall I,J$.   NCC(I,J)=0 ➜ no correlation
- Can be computed from $H_{I,J}$
- Have to be *maximized*
- Model *contrast differences*, only if linearly dependent



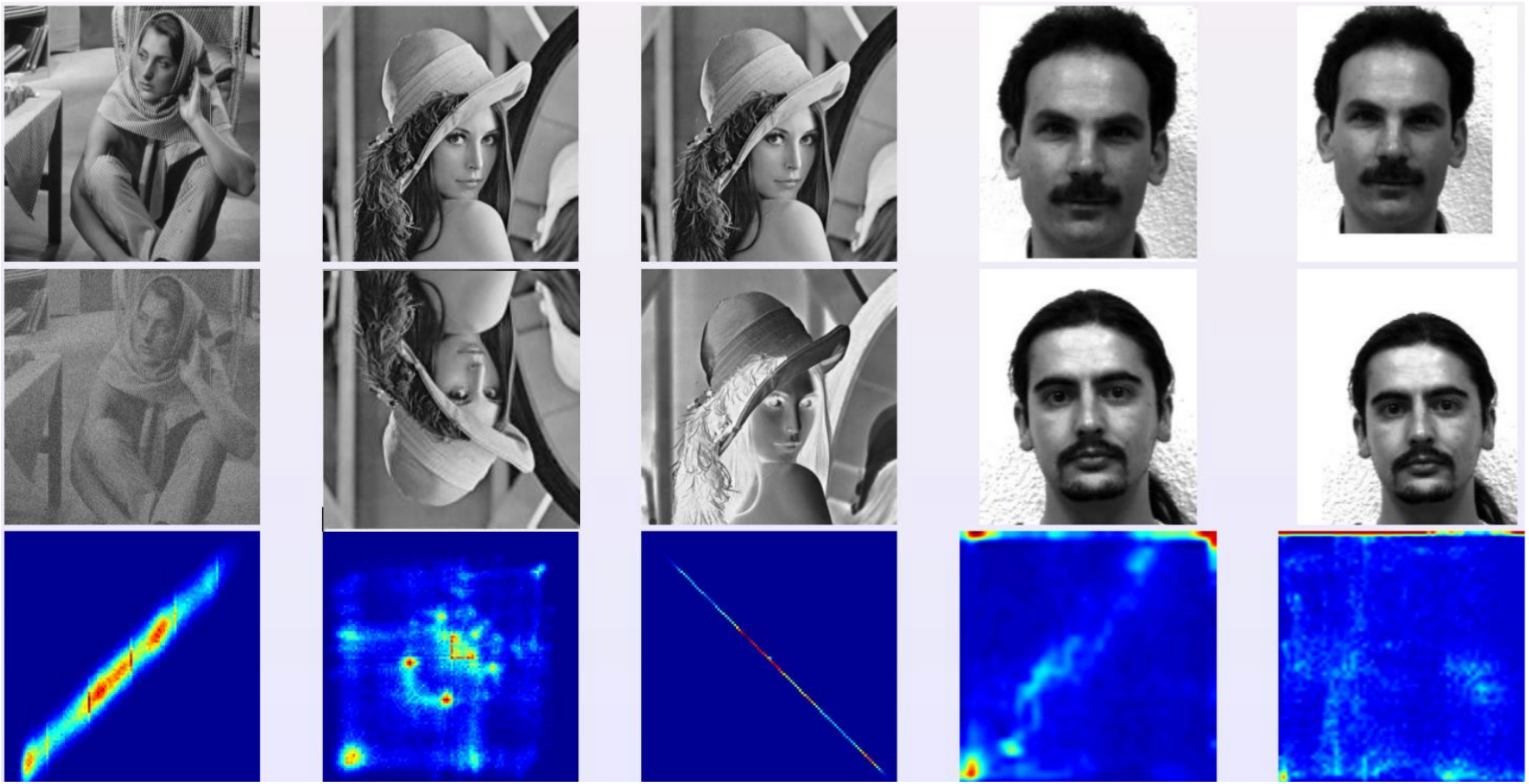
II - Similarity measures (3/5)

SSD vs NCC

I

J

$H_{I,J}$



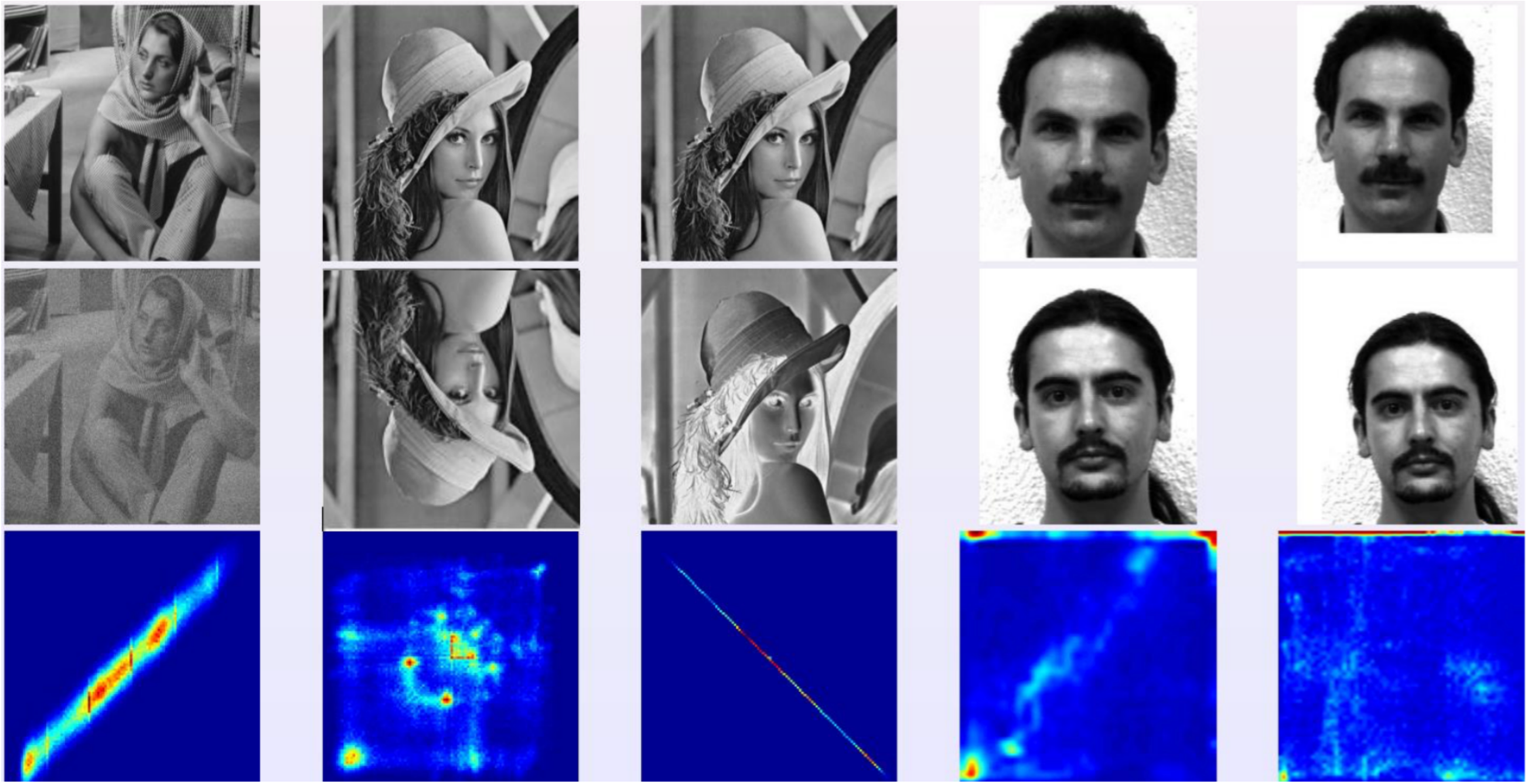
SSD	652	10^9	10^{10}	5600	13400

SSD vs NCC

I

J

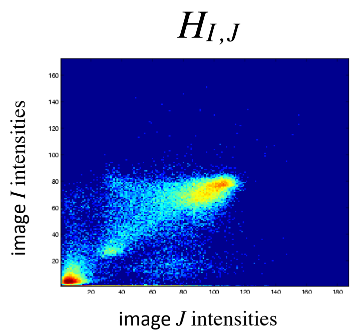
$H_{I,J}$



NCC	0.95	-0.16	-1	0.54	0.14

Statistical approach

- $H_{I,J}(i,j)$ = "*probability* that a randomly chosen pixel has *intensity i in the image I* and *intensity j in the image J*"
- Suggests the use of **statistical/information theory** techniques



Uncertainty and information

- When we say <u>something obvious</u> *(e.g. tomorrow the sun will rise)* it's **not interesting**, there's **no information/uncertainty** in it
- When <u>something unlikely</u> happens *(e.g. tomorrow a meteor will hit the Earth)* it's **very interesting**, it's a **important information**

Entropy is *measure of uncertainty* of a system

$$\mathrm{H} = -\sum_i p_i \log p_i$$

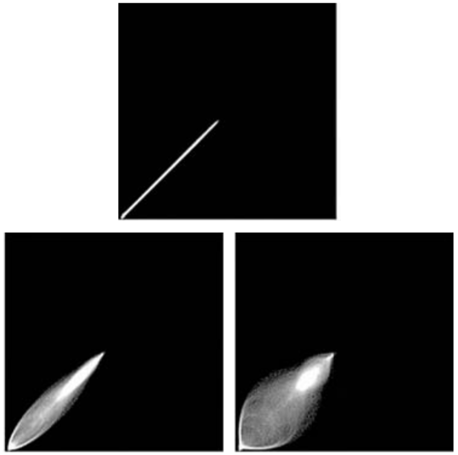
Developed by **Claude Shannon** (*"father of information theory"*)

- Set of n symbols with probability of occurrence $p_1, p_2, \ldots, p_n$
  - All symbols have equal probability ➜ max uncertainty/information ➜ H **is max**
  - One has probability 1, the rest 0 ➜ no uncertainty/information ➜ H **is min**

Entropy for **image registration**

- Two images to align, so *two symbols* at each pixel
- $H_{I,J}(i,j)$ = **joint probability distribution** of images A and B (let's call it p_{AB})
- **Joint entropy** measures the information in the two images combined:

$$\mathrm{H}(A,B) = -\sum_a \sum_b p_{AB}(a,b)\ \log p_{AB}(a,b)$$



Registration seen as seeking to *reduce the amount of information* in the combined image

- Sharper $H_{I,J}$ ➜ lower $\mathrm{H}(A,B)$ ➜ reduced uncertainty

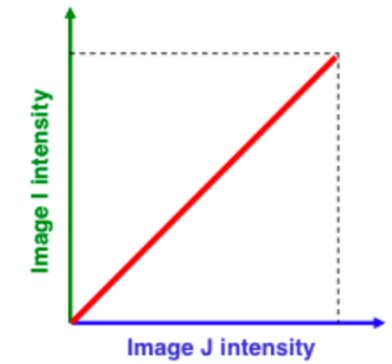
Mutual Information (MI) usually preferred

$$\mathrm{MI}(A,B) = \sum_a \sum_b\ p_{AB}(a,b)\ \log \frac{p_{AB}(a,b)}{p_A(a)\cdot p_B(b)} = \mathrm{H}(A) + \mathrm{H}(B) - \mathrm{H}(A,B)$$

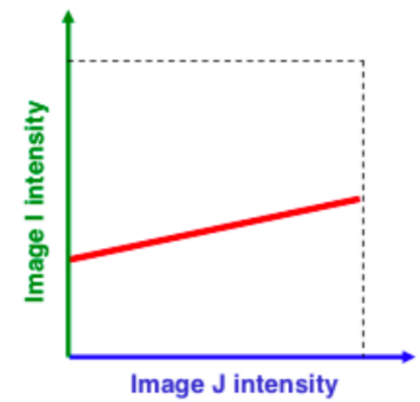
- Measures how well **one image explains the other**
- MI is *maximimum* at optimal alignment

Summary

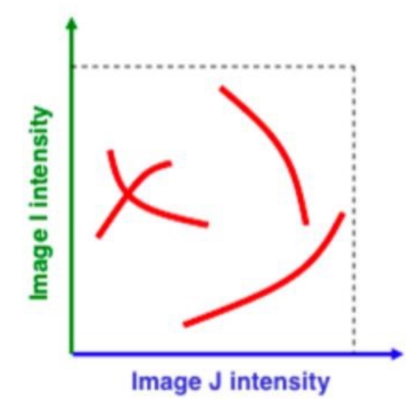
- **Minimizing intensity differences**
  - ***Sum of squared differences*** (SSD) or ***sum of absolute differences*** (SAD)
  - To be *minimized*
  - Suited for *mono-modal*, *intra-subject* registration
  - Strong assumption on intensities



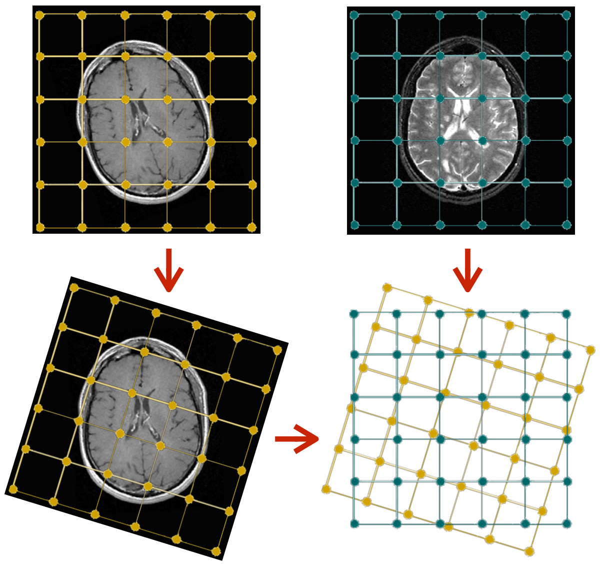
- **Correlation approach**
  - Relaxes the previous assumption, allowing linear dependence
  - ***Normalized Cross-Correlation*** (NCC)
  - To be *maximized*
  - Suited for *mono-modal*, *intra-* or *inter-subject* registration



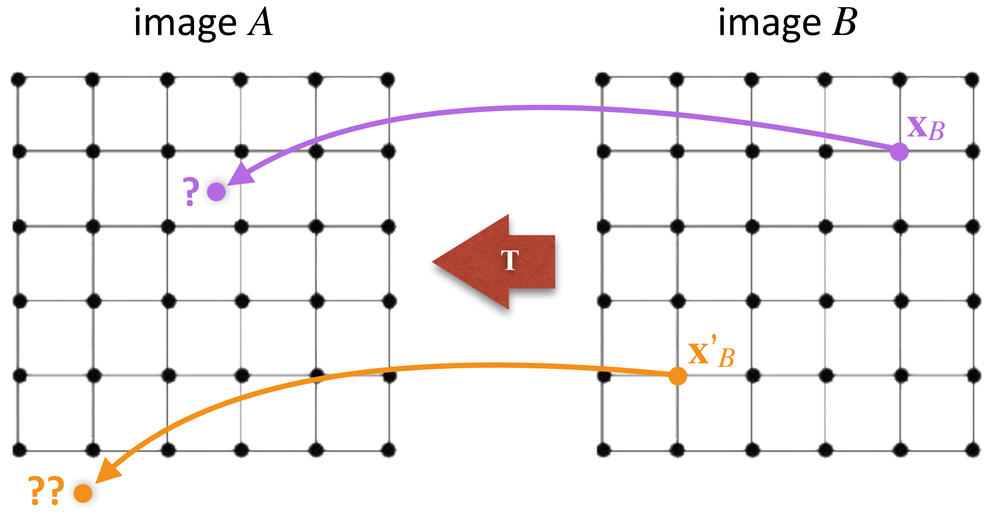
- **Statistical interpretation**
  - Weakest assumption on the relationship between intensities
  - ***Mutual Information*** (MI)
  - To be *maximized*
  - Suited for *multi-modal* registration (*intra-* and *inter-subject*)



- To compute distance/similarity $d(A, B^{\mathcal{T}})$ we need to compare features/intensities at **same locations on both images**
  - If $\mathcal{T}$ maps the pixels of B exactly at the same locations of the pixels of A, there are no problems
  - But usually the locations/grids do not match

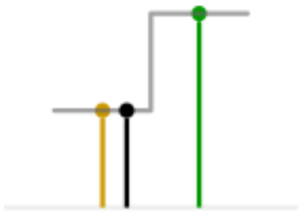


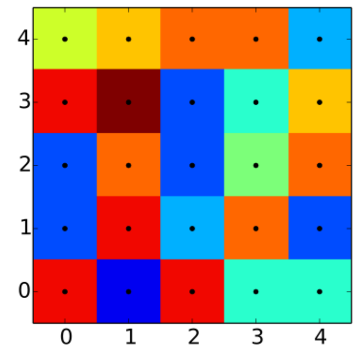
- Two cases
  - **Interpolation**
    - For the points $\mathbf{T}(\mathbf{x}_B)$ falling *inside* the grid of A (but not on the grid points themselves)
    - Value for these points needs to be estimated from the *neighboring pixels*
  - **Extrapolation**
    - For the points $\mathbf{T}(\mathbf{x}_B)$ falling *outside* the grid of A
    - Points not considered? Mirror or extend pixels?

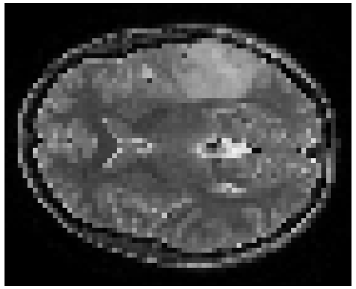


Most common choices

- **Nearest neighbor**

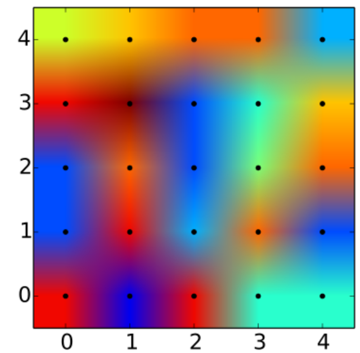
1D


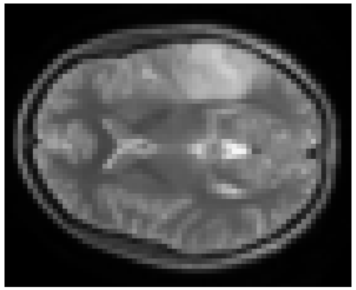
2D




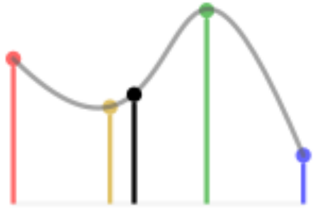
- **Linear**

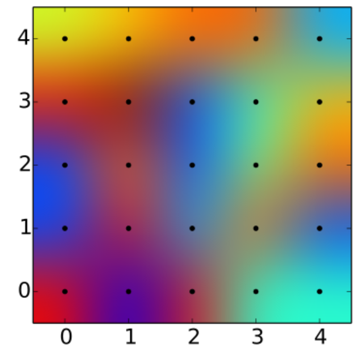
1D

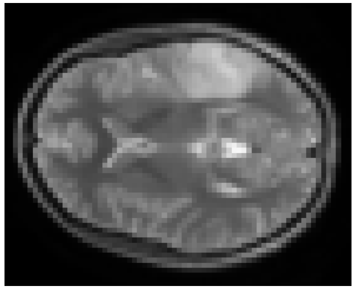

2D




- **Higher order**, e.g. cubic or B-spline

1D


2D


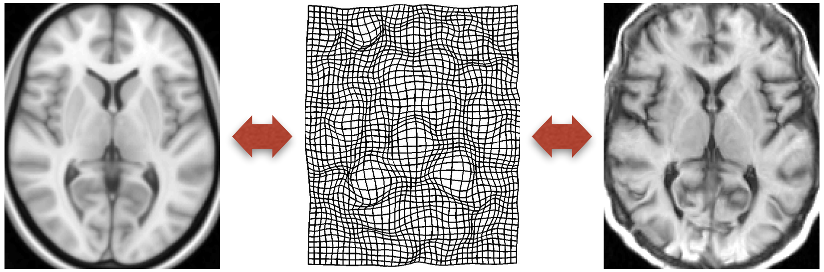


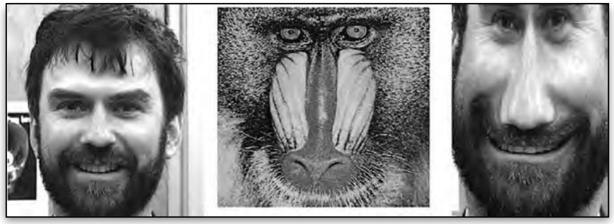
IV - Deformation models (1/2)

- Two main categories
  - **Linear** (a.k.a. *rigid*)
    - Only a limited number of *degrees of freedom* is allowed

$$\begin{pmatrix} \cos\beta\cos\gamma & \cos\alpha\sin\gamma+\sin\alpha\sin\beta\cos\gamma & \sin\alpha\sin\gamma-\cos\alpha\sin\beta\cos\gamma & t_x \\ -\cos\beta\sin\gamma & \cos\alpha\cos\gamma-\sin\alpha\sin\beta\sin\gamma & \sin\alpha\cos\gamma+\cos\alpha\sin\beta\sin\gamma & t_y \\ \sin\beta & -\sin\alpha\cos\beta & \cos\alpha\cos\beta & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

  - **Non-linear** (a.k.a. *non-rigid*)
    - Virtually any transformation/deformation is possible

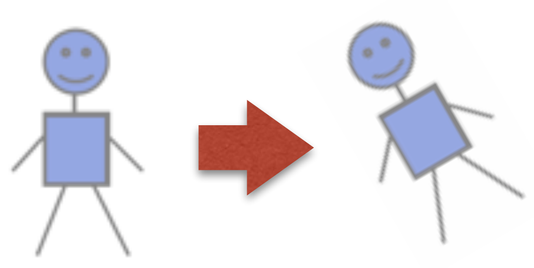




- **NB**: the **choice of the deformation model** to use depends on the application, i.e. which tissue/structure to register
  - *Bones of the skull* restrict the movement of the brain
  - *Soft tissue* tends to deform in more complicated ways

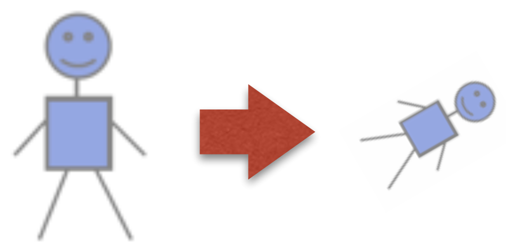
Linear transformations

▸ **Rigid :** $\mathbf{T}(\mathbf{x}) = \mathbf{R}\mathbf{x} + \mathbf{t}$



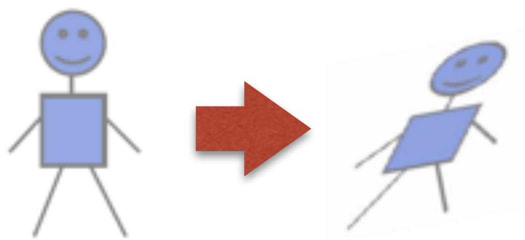
- **6 parameters** : rotation ($\mathbf{R}$) and translation ($\mathbf{t}$)
- *Invariants*: distances (isometric), curvature, angles, lines
- *Use*: same structure in a different position

▸ **Similitude :** $\mathbf{T}(\mathbf{x}) = \mathbf{s}\mathbf{R}\mathbf{x} + \mathbf{t}$



- **7 parameters**: adds a scaling factor ($\mathbf{s}$)
- *Invariants*: distance ratios, angles, line

▸ **Affine :** $\mathbf{T}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{t}$



- **12 parameters**: $\mathbf{A}$ includes stretching and shearing
- *Invariants*: lines, parallelism
- *Use*: - correct for scanner deformations/artifacts
  - find approximate alignment before nonlinear registration

IV - Deformation models (2/2)

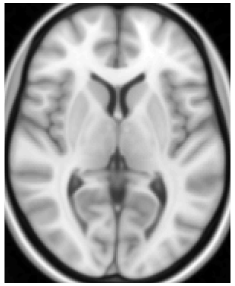
<u>Nonlinear transformations</u> required when registering:

- An image of one individual and **atlas**
- Image from **different individuals**
- **Tissue that deforms** over time


time 1

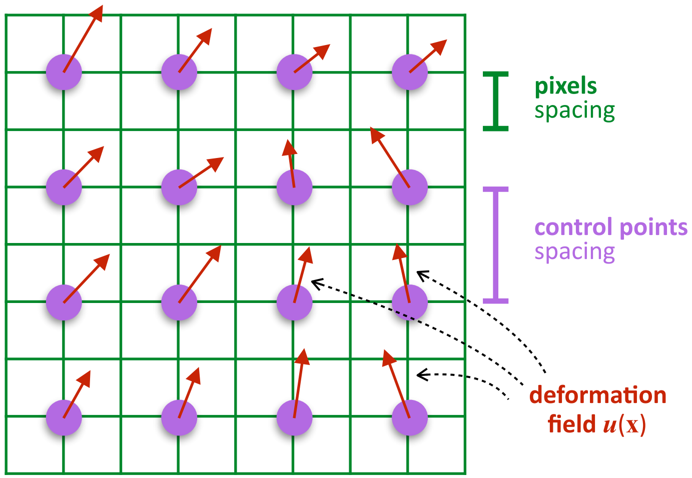

time 2

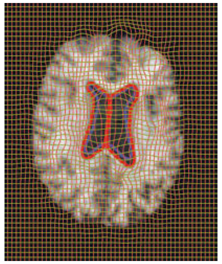

subject

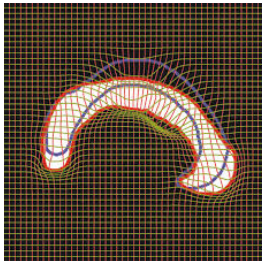

template

General approach

- **Each pixel** can virtually be moved independently
  - One displacement per pixel
  - Actual *tissue deformations* are usually more smooth/regular
- Usually **grids** of **control points** are defined
  - One displacement $\boldsymbol{u}(\mathbf{x})$ ( ) per control point ( )
  - *Smoothness constraints* are usually added to obtain "anatomically reasonable" deformations
  - Control points are *not independent*
- Several solutions **inspired by physics**
  - *Elastic, viscous fluid, optical flow, diffusion model* (demons) …

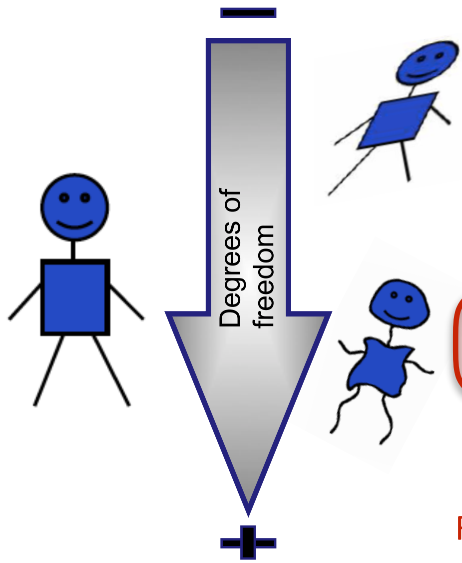






IV - Deformation models (2/2)

- **Increased complexity**: *overfitting* and *regularization*

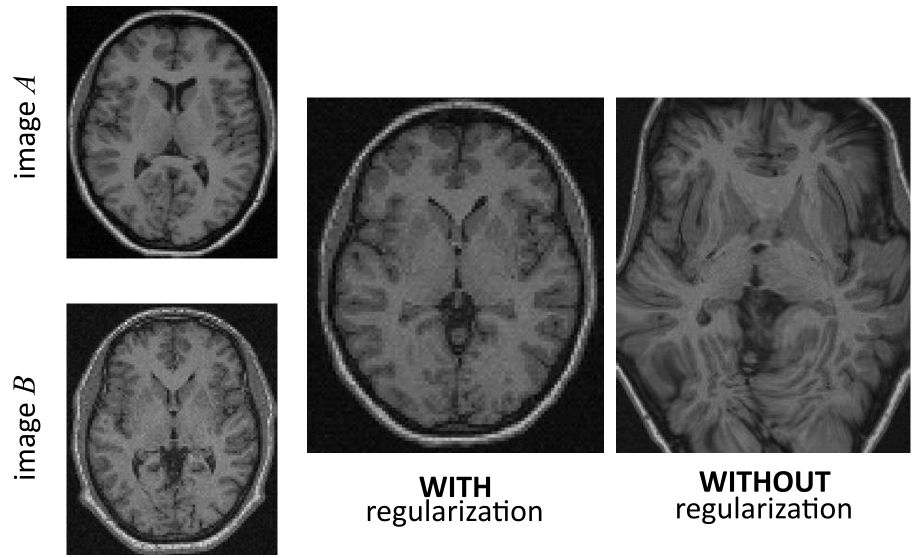


Rigid (6 parameters)
Similitude (7 parameters)
Affine (12 parameters)

Parametric (~ 100's parameters)
Non-parametric (~ 10000's parameters)

Regularization is necessary to prevent nonlinear registration to introduce **unnecessary deformations**

- Example



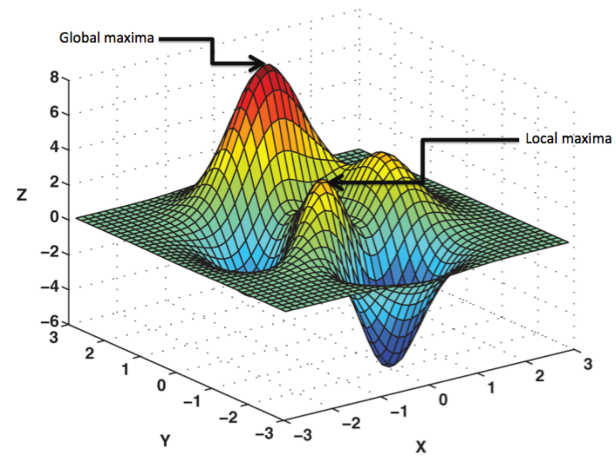
WITH regularization

WITHOUT regularization

V - Optimizers

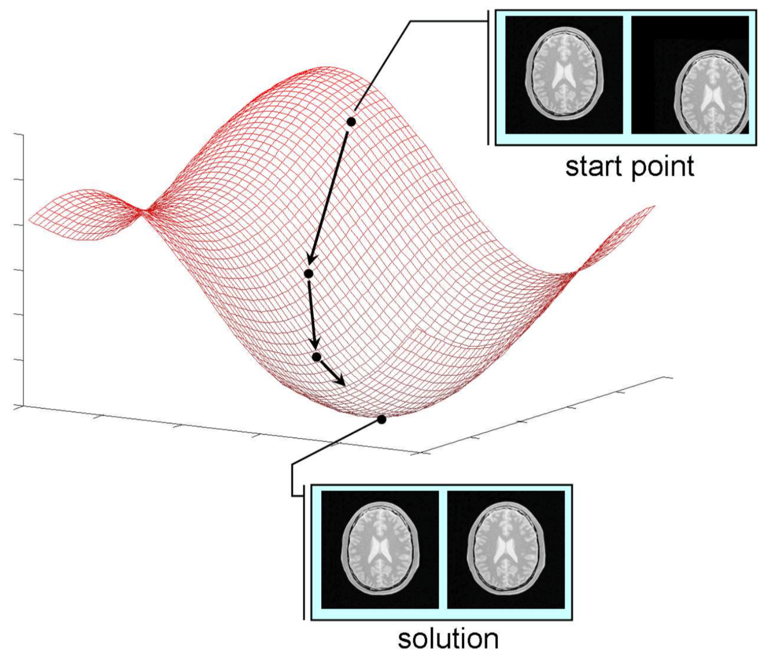
Registration is an **optimization problem**

- The search space is **high dimensional**
  (i.e. space of all possible transformations)
- The problem is **nonlinear**
  (possibly with many **local minima**)



Usually **iterative approaches** are used

- Start with *initial estimate* of transformation, $\mathbf{T}^0$
- At each iteration *t*, current estimate $\mathbf{T}^t$ is used to compute a *similarity measure* $d(A, B^{\mathcal{T}})$
- Using d, *refine the transformation* $\mathbf{T}^t \rightarrow \mathbf{T}^{t+1}$
- Continues until the convergence



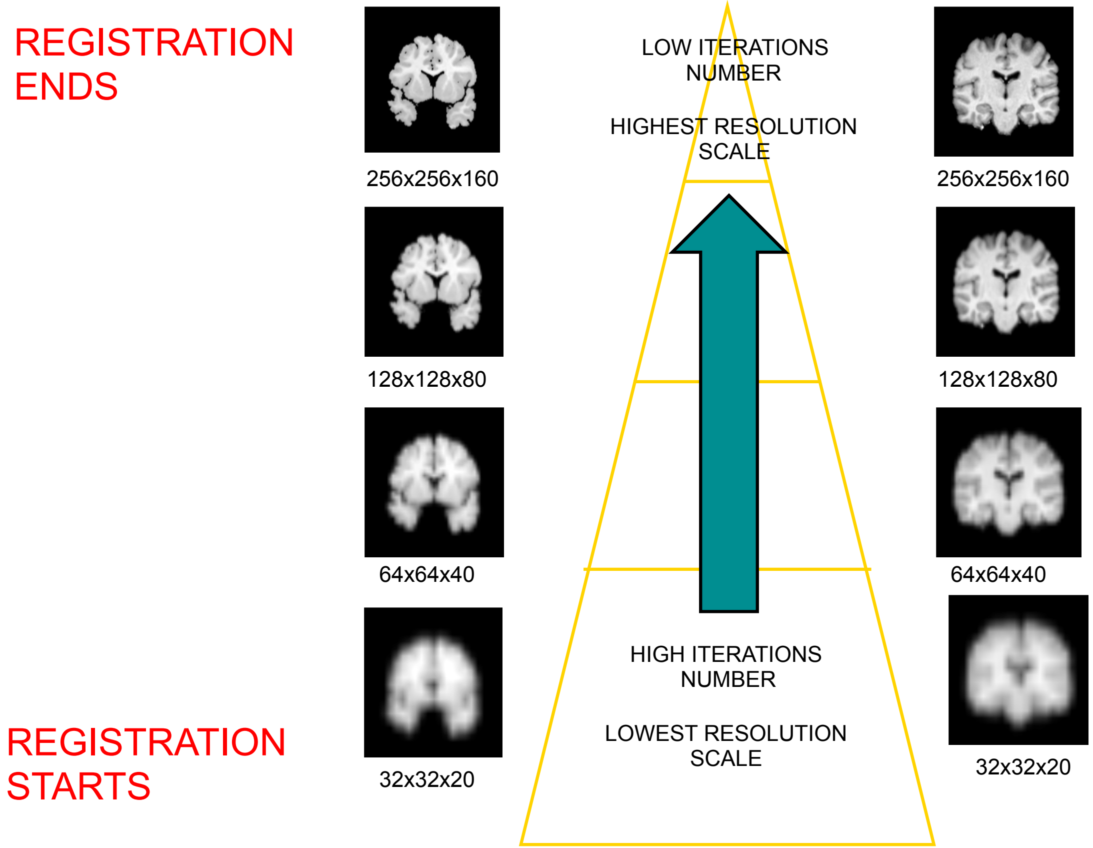
Classical algorithms

- *Gauss-Newton, (stochastic) gradient descent* etc...

Multi-scale pyramid

Strategy to improve **registration accuracy**

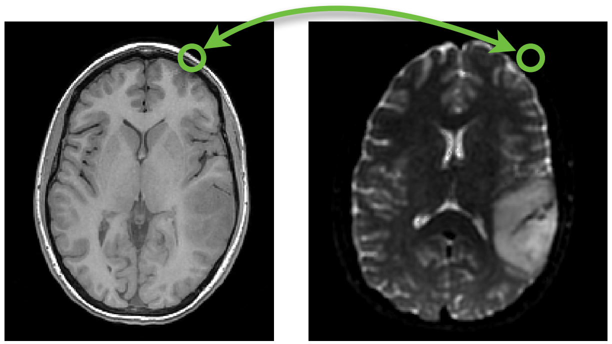
- Start the registration using **images with low complexity** → *smoothing*, *downsampling*
- At convergence, **increase the complexity/details of the images** and repeat
- This reduces the chance of falling in **local minima** (bad registration)



Use of masks

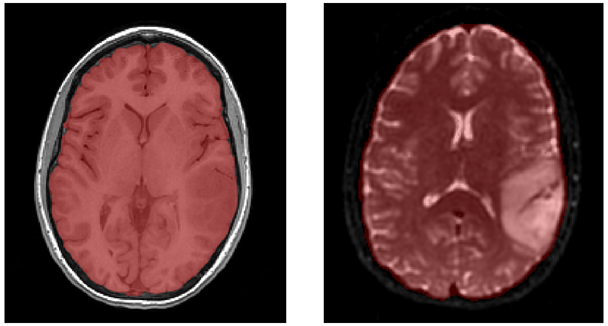
Sometimes it is desirable to **align only part of an image**

- We are interested **only on a portion or some details** of the image
- We need to **ignore parts of the images** that can confound the registration
  (e.g. *artificial edges*)



Some
anatomical details
are not visible
in both images

With a mask registration is **constrained to a region**



- A mask is a **binary image**
  - **"1"** ➡ the pixel in *considered*
  - **"0"** ➡ the pixel in *ignored*
- A **fixed image mask is usually sufficient** to focus the registration on a region, since samples are drawn from the domain of the fixed image

Available tools

- ITK.org : MITK, MedINRIA, Slicer3D, etc
- Elastix
  - Choice for our lab: **power of ITK** (all algorithms) with **simple interface**
- FSL FLIRT (linear) and FNIRT (nonlinear)
- ANTs (Advanced Normalization Tools)
- SPM
- Freesurfer
- Hammer / Glirt
- BrainVisa / Anatomist
- and many others more….