Deep learning for medical imaging

Olivier Colliot, PhD
Research Director at CNRS
Co-Head of the ARAMIS Lab – www.aramislab.fr
PRAIRIE – Paris Artificial Intelligence Research Institute

Maria Vakalopoulou, PhD
Assistant Professor at CentraleSupelec
Center for Visual Computing

Master 2 - MVA

Part 8 – Registration

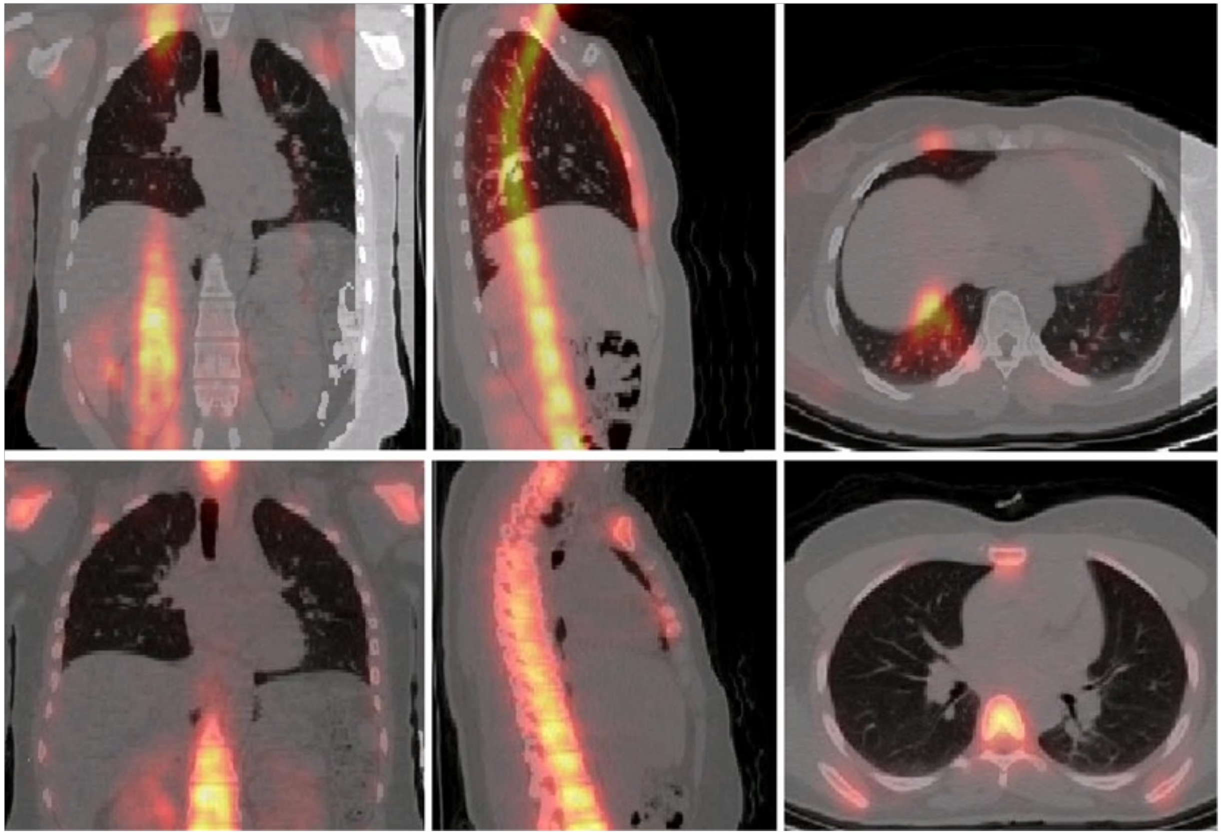
(*"Recalage"* in French)

Part 8 – Registration

8.1 Introduction

Introduction

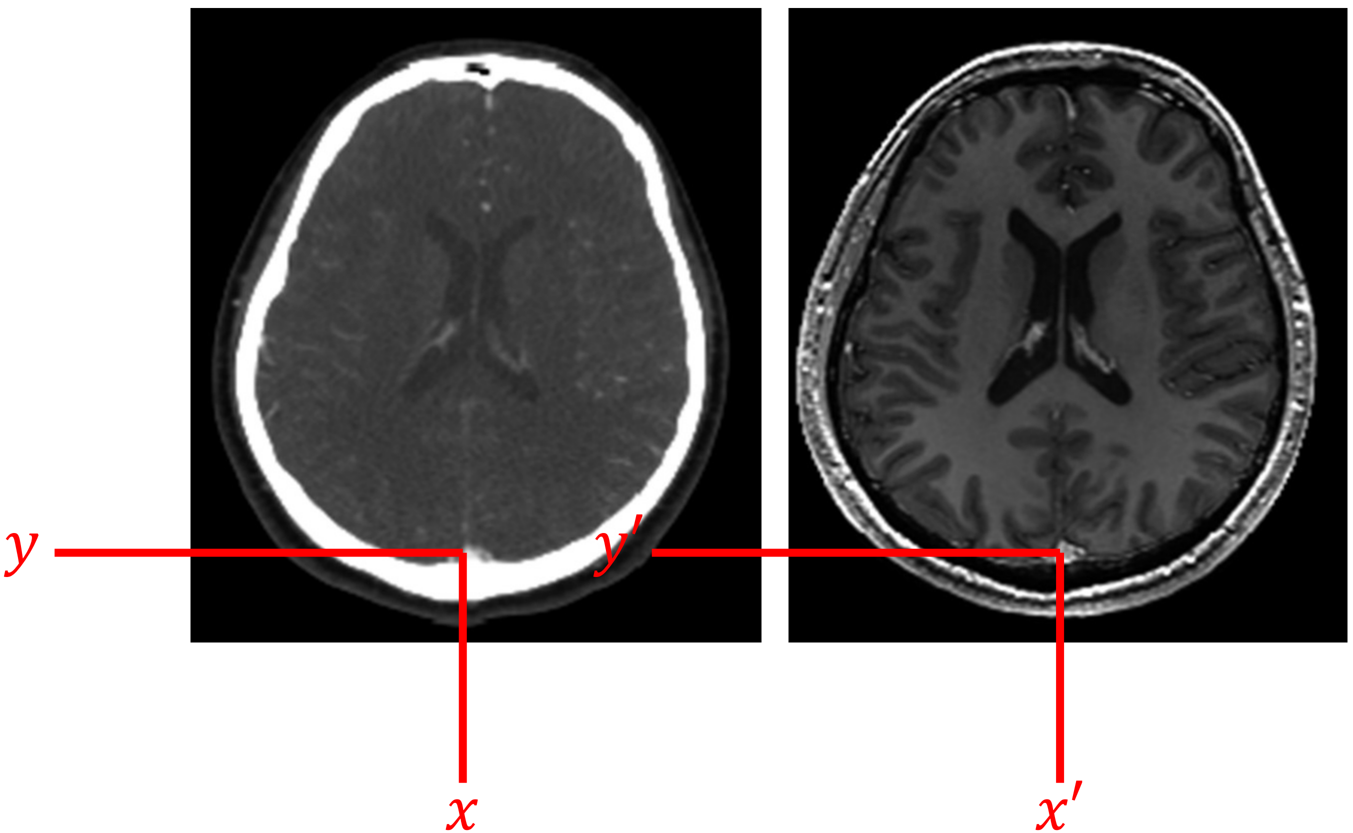
- **Definition: put two images in spatial correspondence**



[Image Source Tang et al]

Introduction

- **Definition: put two images in spatial correspondence**

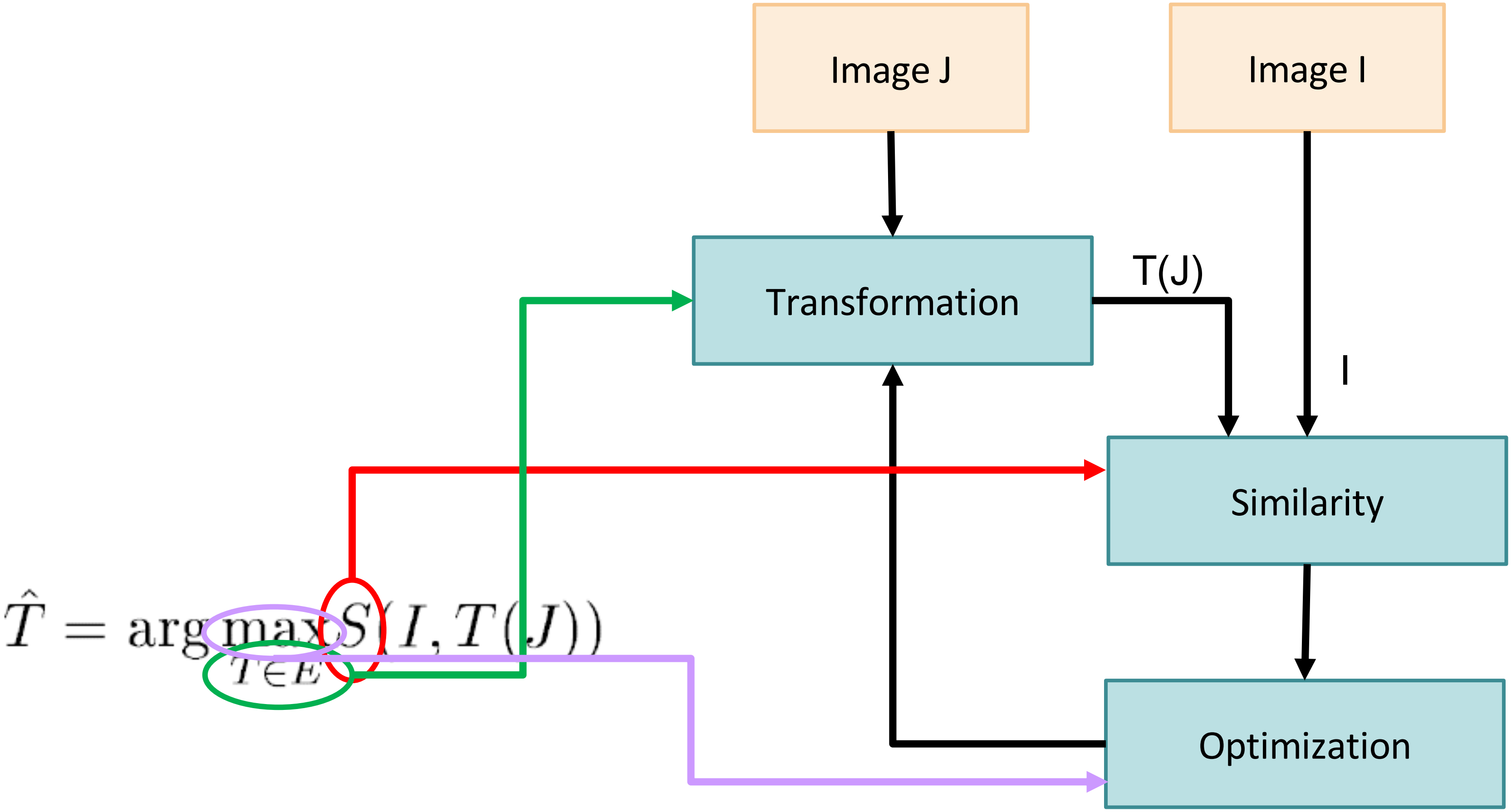


Find T such that

$$(x', y') = T(x, y)$$

[Image Source Jiang et al, Medical Physics, 2019]

Introduction

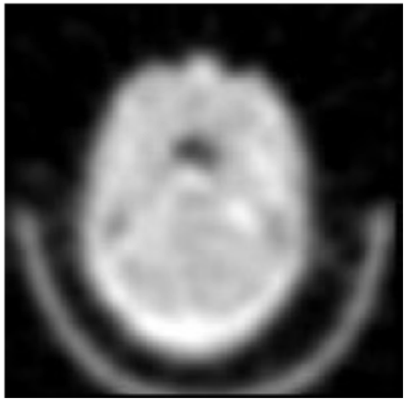


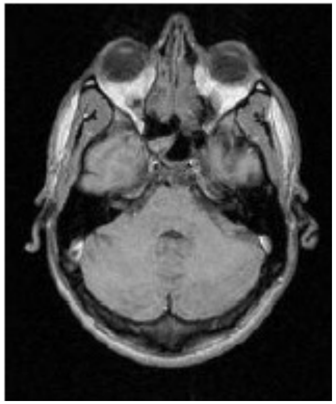
$$\hat{T} = \arg\max_{T \in E} S(I, T(J))$$

Introduction

- Intra-subject, inter-modality

Brain imaging

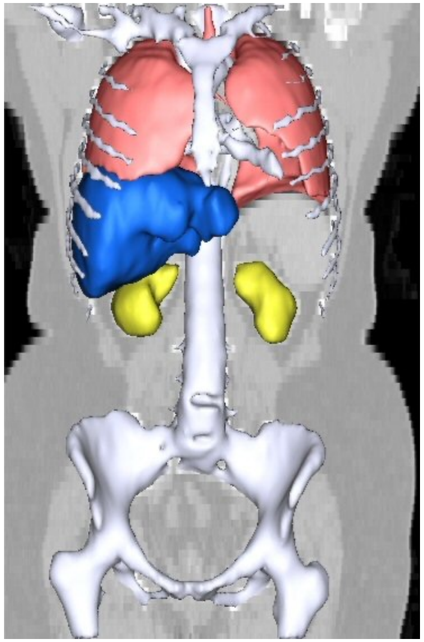




PET MRI

[Mangin, 1995]

Thoracico-adbominal imaging



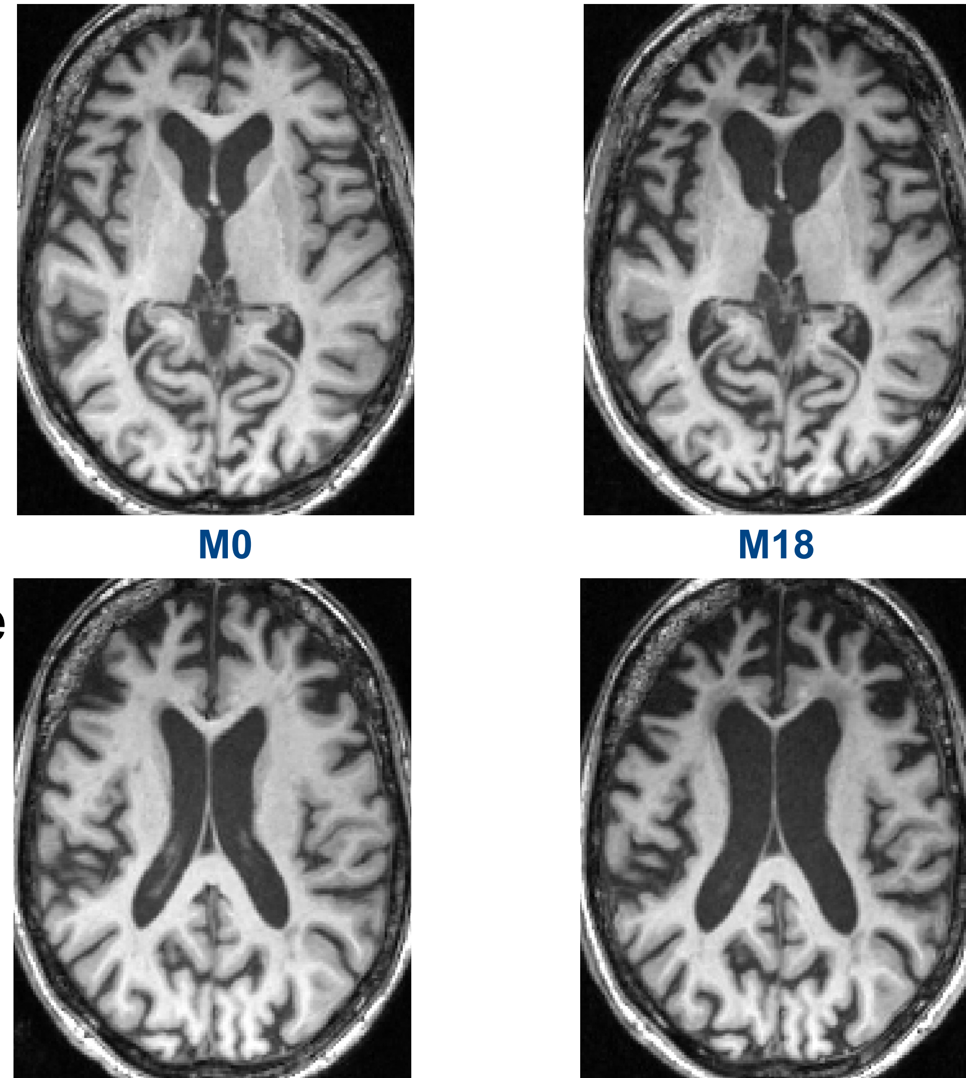
CT PET

[Camara et al,, 2007]

Introduction

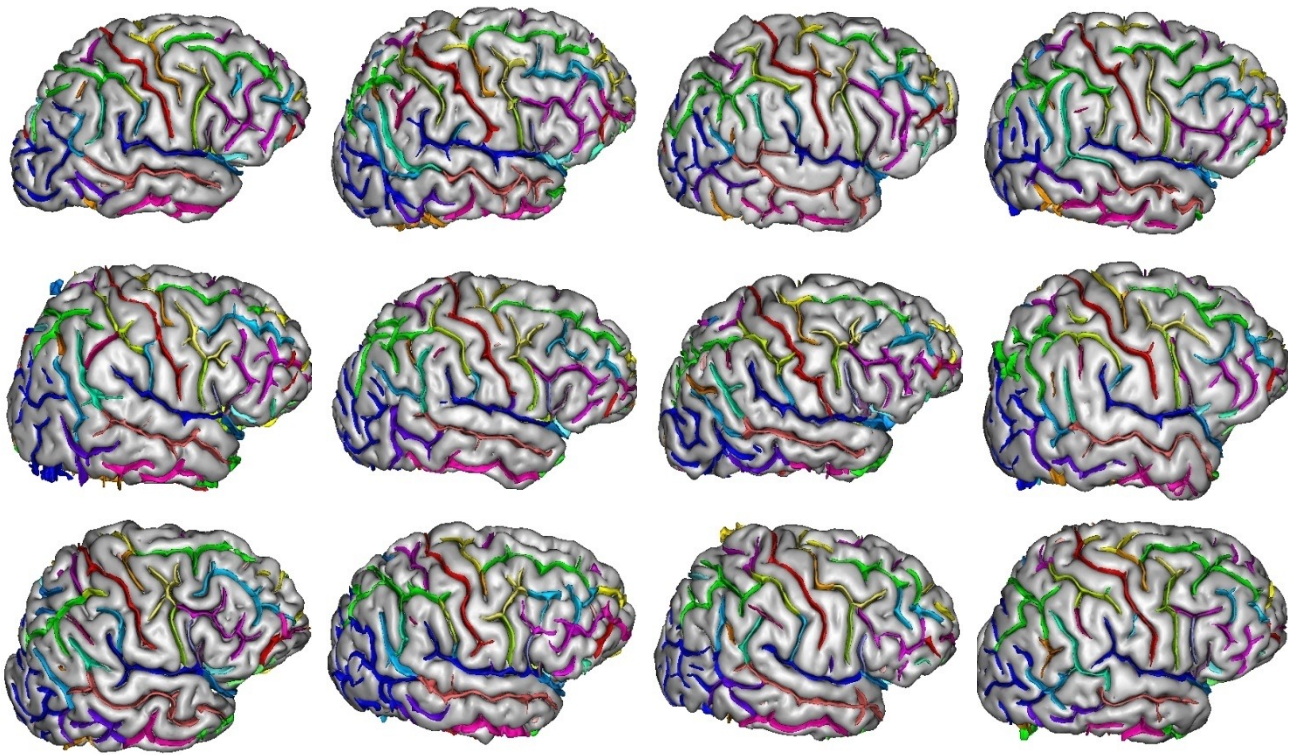
- Intra-subjet, intra-modality, longitudinal

Same patient seen at different time points (e.g. after 18 months)



Introduction

- Inter-subjet, intra-modality



(From Mangin)

Introduction

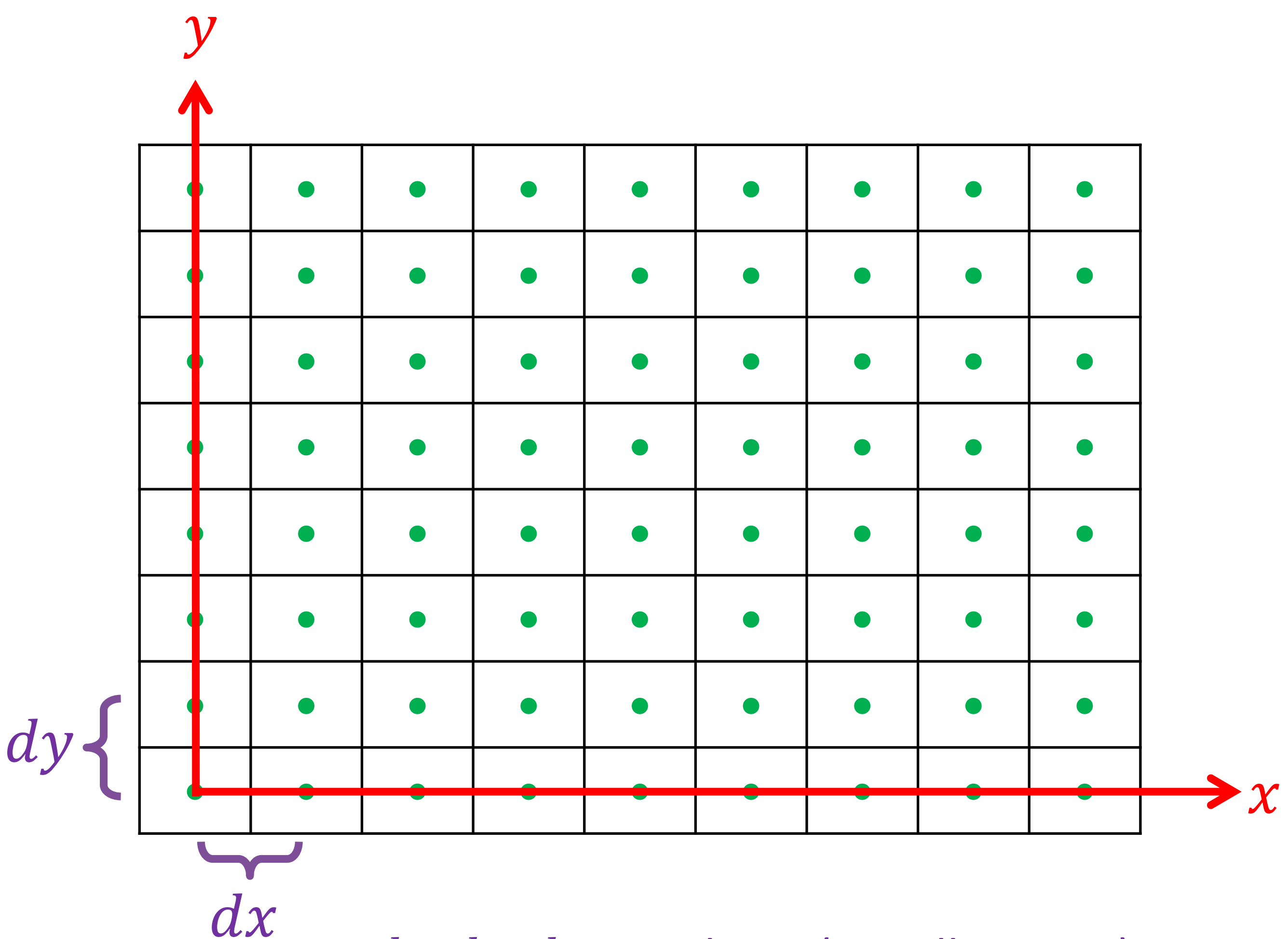
Although deep learning based registration methods have been developed, **they are not (yet) the standard** for this task (unlike, for example, for segmentation)

Part 8 – Registration

8.2 Coordinate systems

Coordinate systems

Image coordinate system

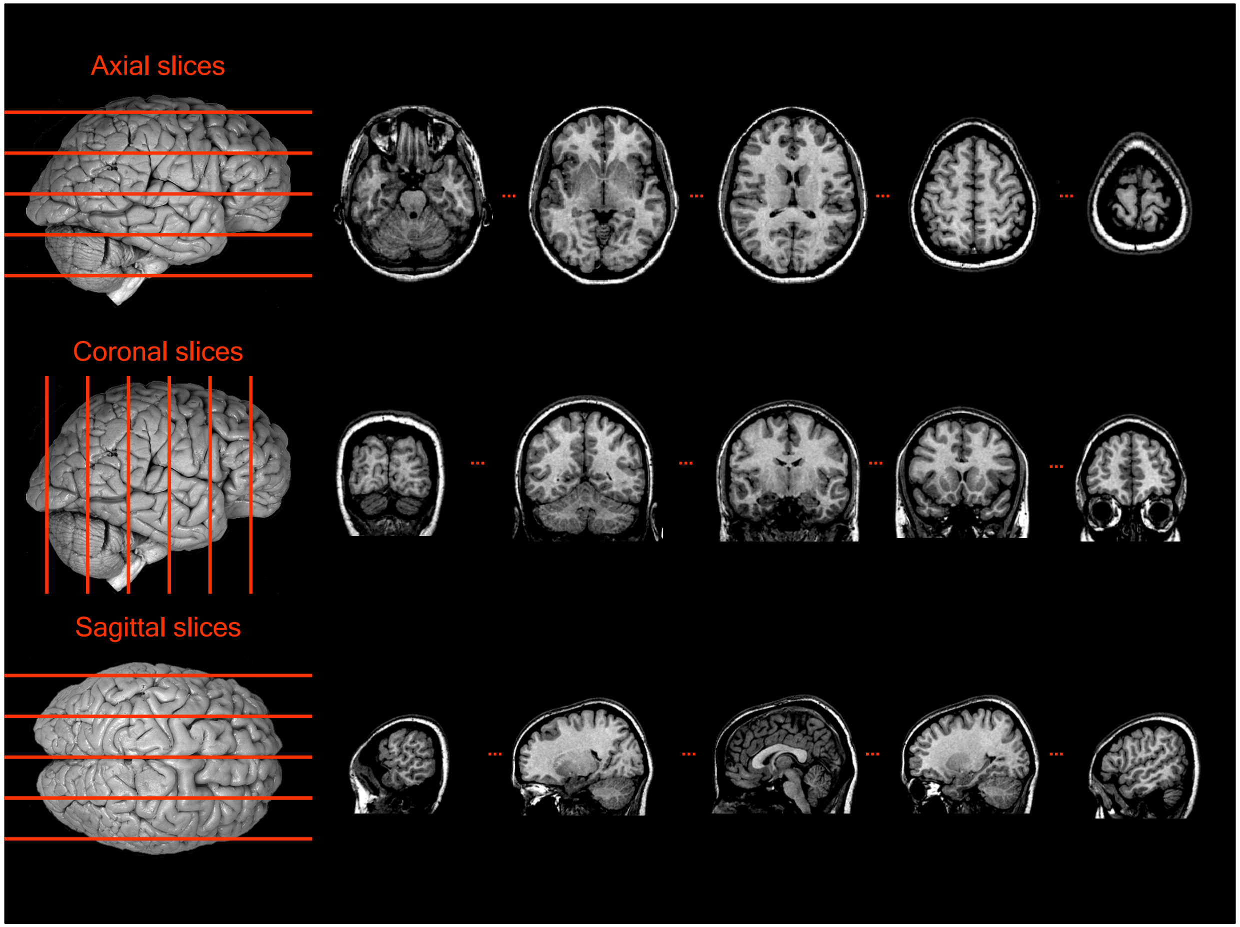


dx, dy, dz: voxel size (in millimeters)

Medical images often have non-isotropic voxels/pixels:

$dx \neq dy \neq dz$

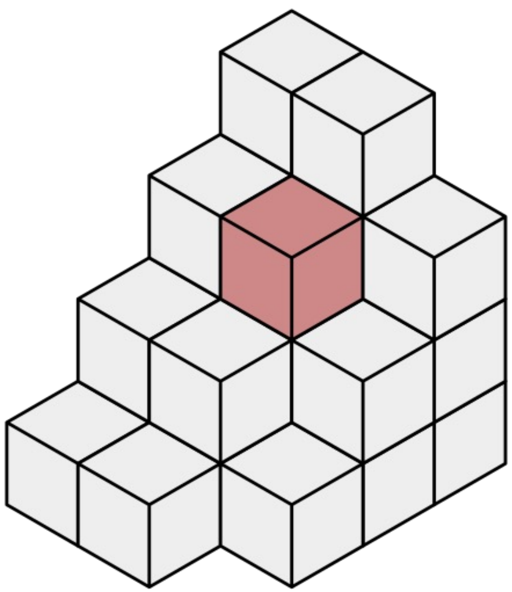
3D images

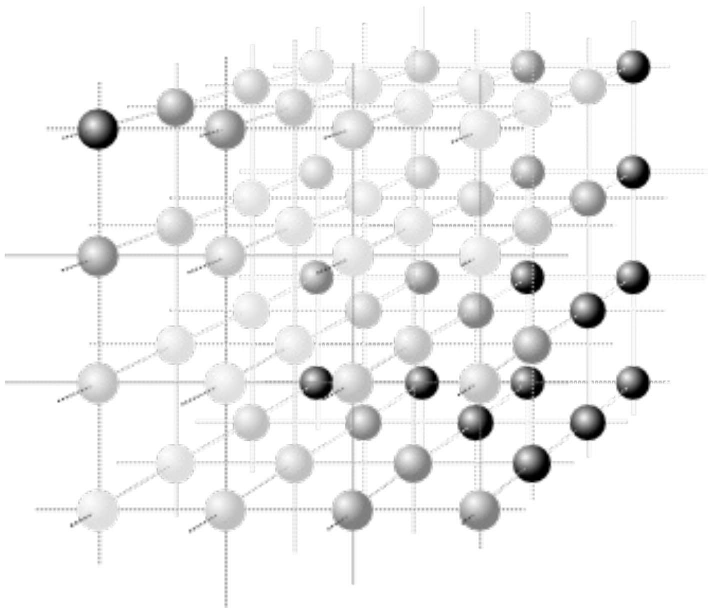


Remember:
many medical images are 3D

Some of the course's equations will be given in 2D but real applications are most often in 3D

3D images





Voxel=volume element

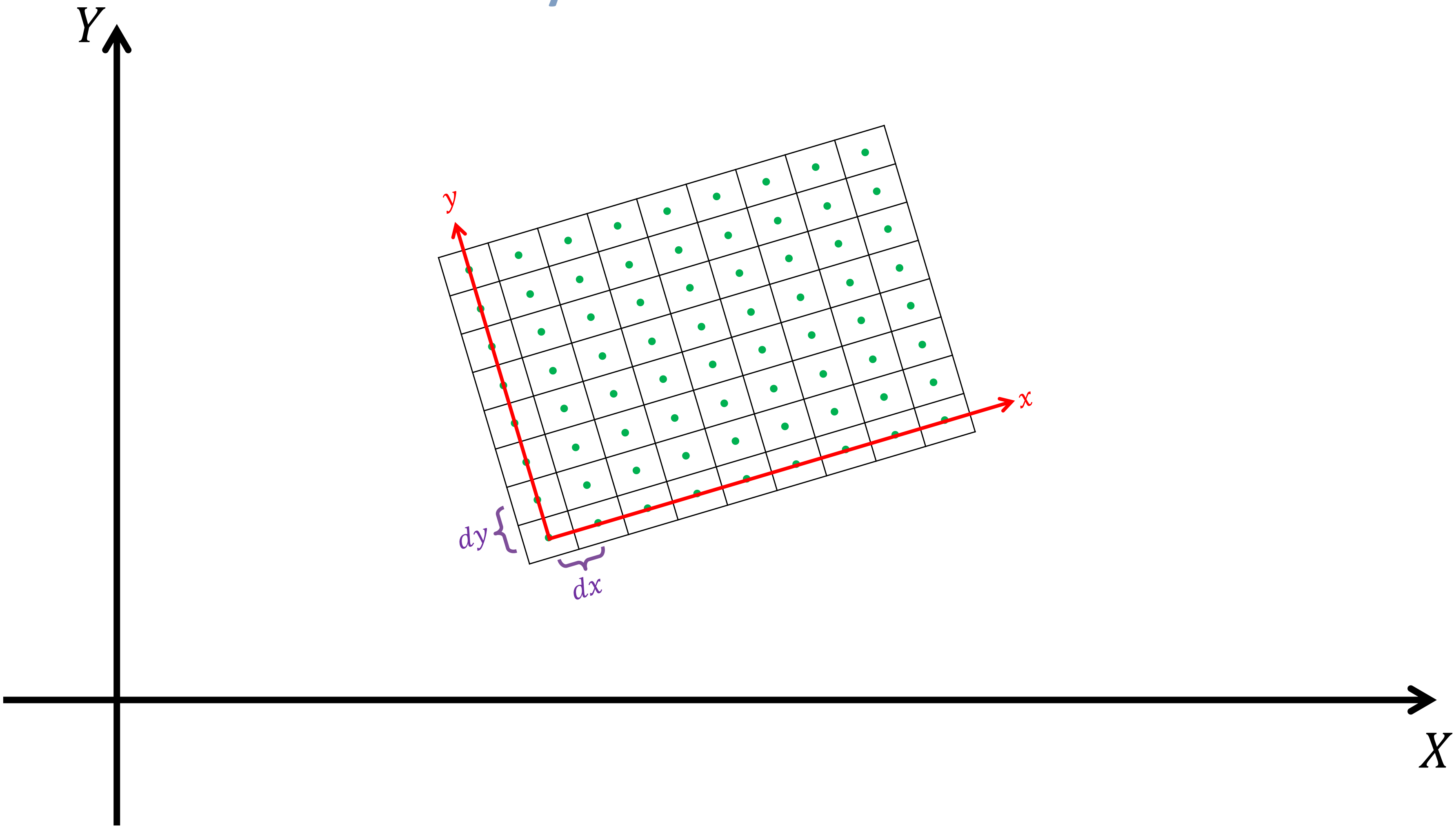
Image source: https://en.wikipedia.org/wiki/Voxel

Remember: many medical images are 3D

Some of the course's equations will be given in 2D but real applications are most often in 3D

Coordinate systems

World coordinate system

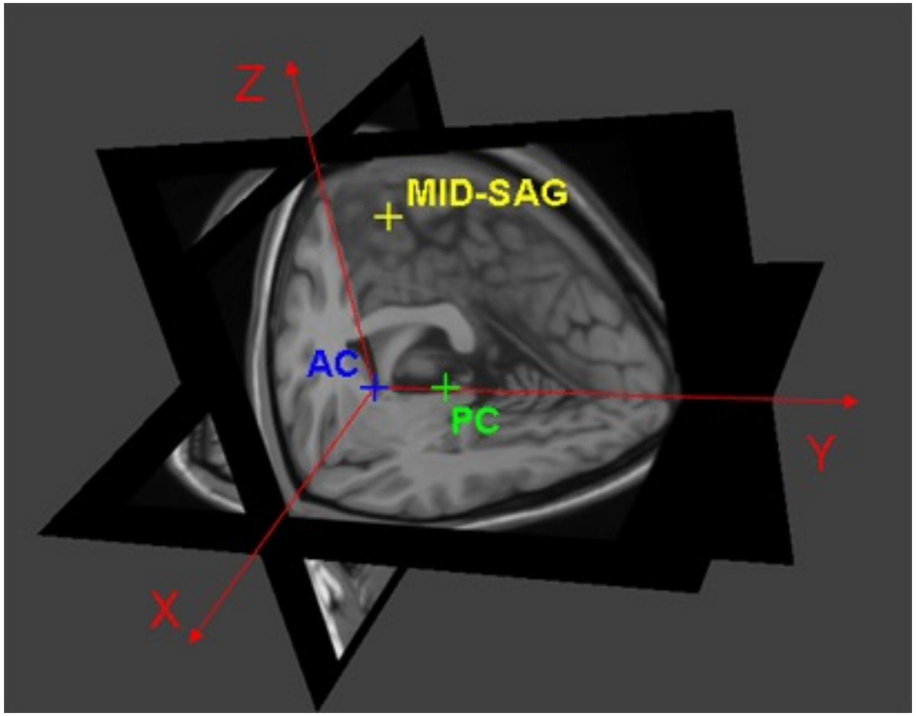


Coordinate systems

World coordinate system

Example: Talairach coordinate system for the brain

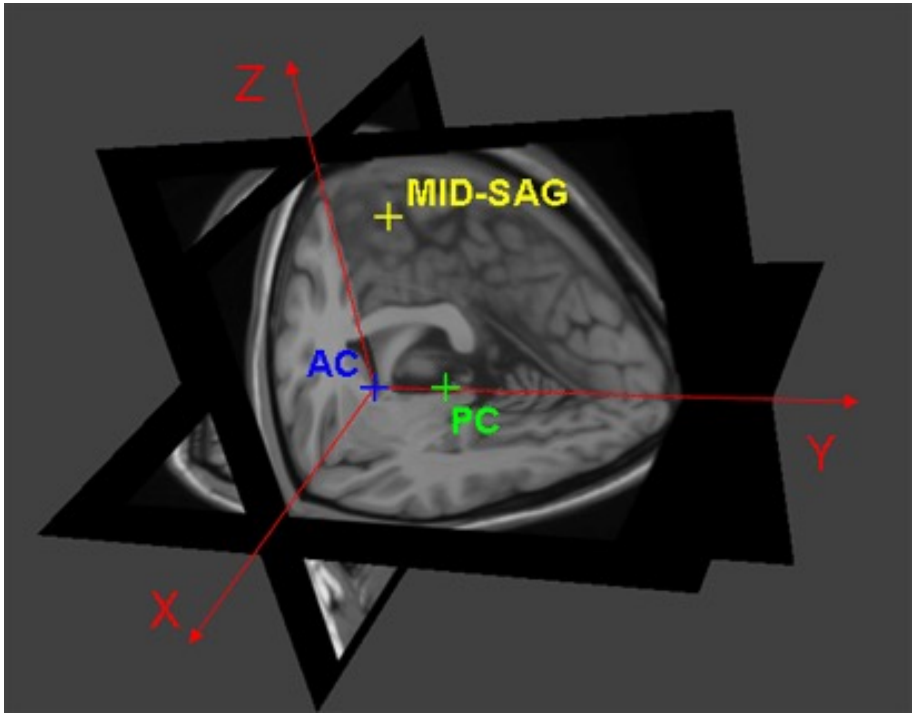
- The **origin is at a specific anatomical point** (the anterior commissure – AC)
- The Y axis is defined as passing through the anterior commissure (AC) and the posterior commissure) and oriented towards the front of the head
- The Z axis is defined as passing through the mid-sagittal plane (MID-SAG), orthogonal to Y and oriented towards the top of the head
- The X axis is oriented from left to right
- The scales of the axis is defined through a linear transformation (actually piecewise linear)

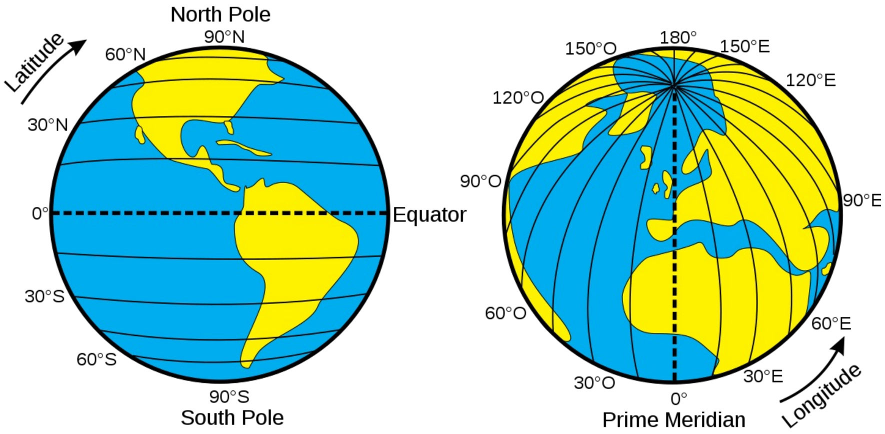


Introduction

World coordinate system

Example: Talairach coordinate system for the brain



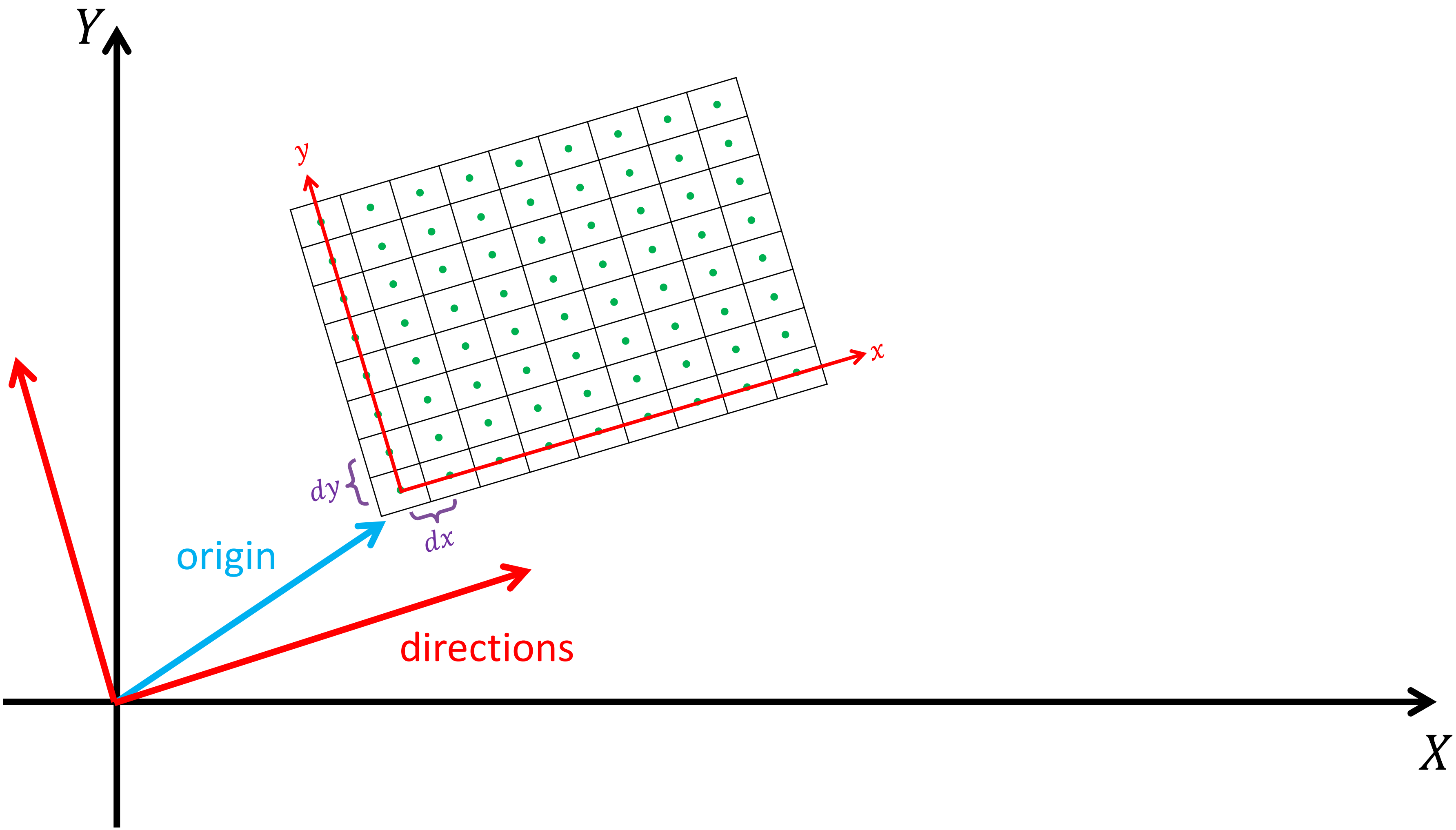


Can locate places in the brain as you would do on Earth…

Coordinate systems

World coordinate system

Inspired from B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London



Coordinate systems

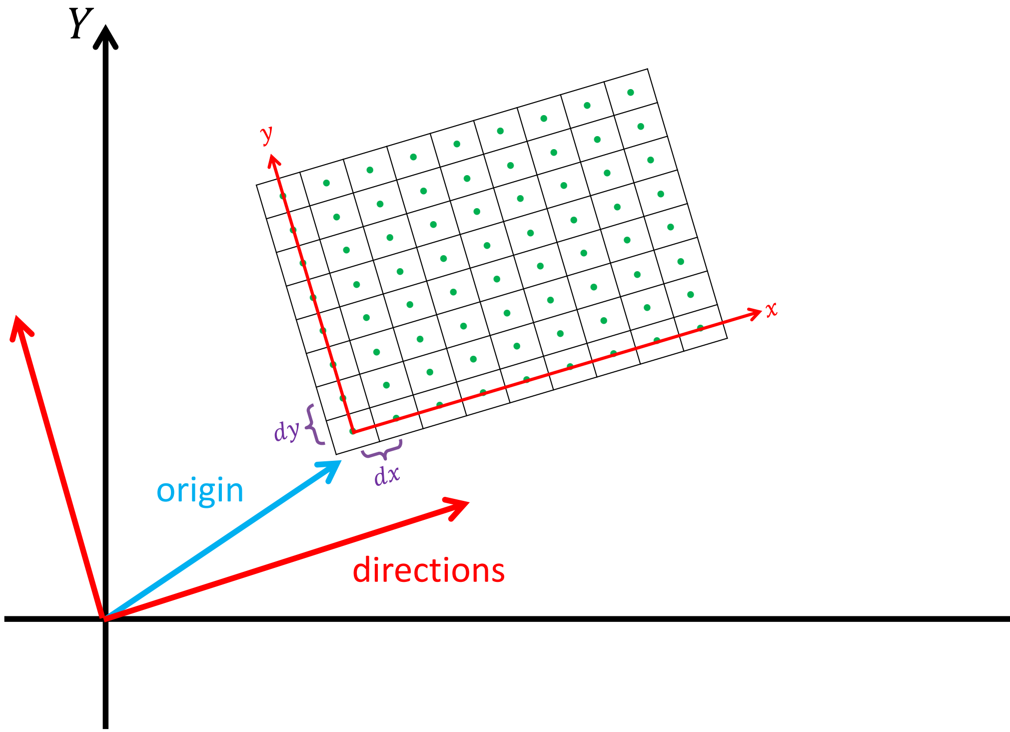
Image to world

Inspired from B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

$$\begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & ox \\ 0 & 1 & oy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} rxx & ryx & 0 \\ rxy & ryy & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} dx & 0 & 0 \\ 0 & dy & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{T_{ItW}} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

World to image

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = T_{ItW}^{-1} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}$$



Part 8 – Registration

8.3 Transformation models

Transformation models

Linear

identity

rigid
rotation+translation

similarity
rigid+uniform scaling

affine
rigid+nonuniform scaling+shear

Non-linear

deformable

degrees of freedom / parameters

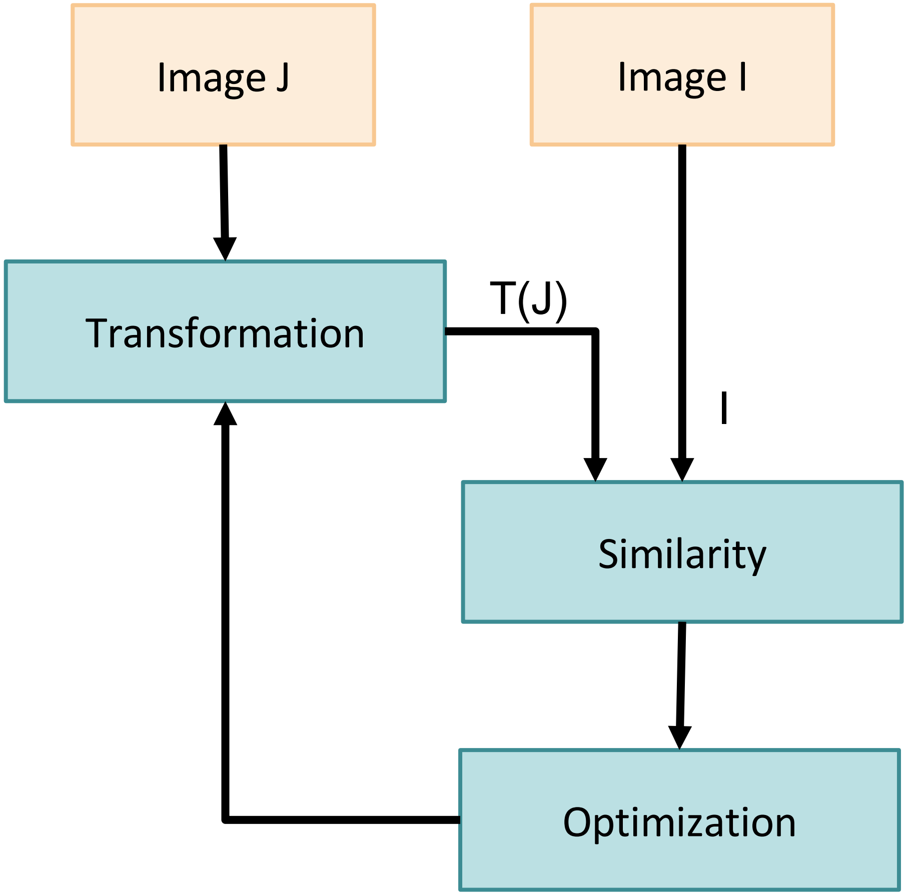
0 3/6 4/7 6/12 thousands/millions

2D **3D**

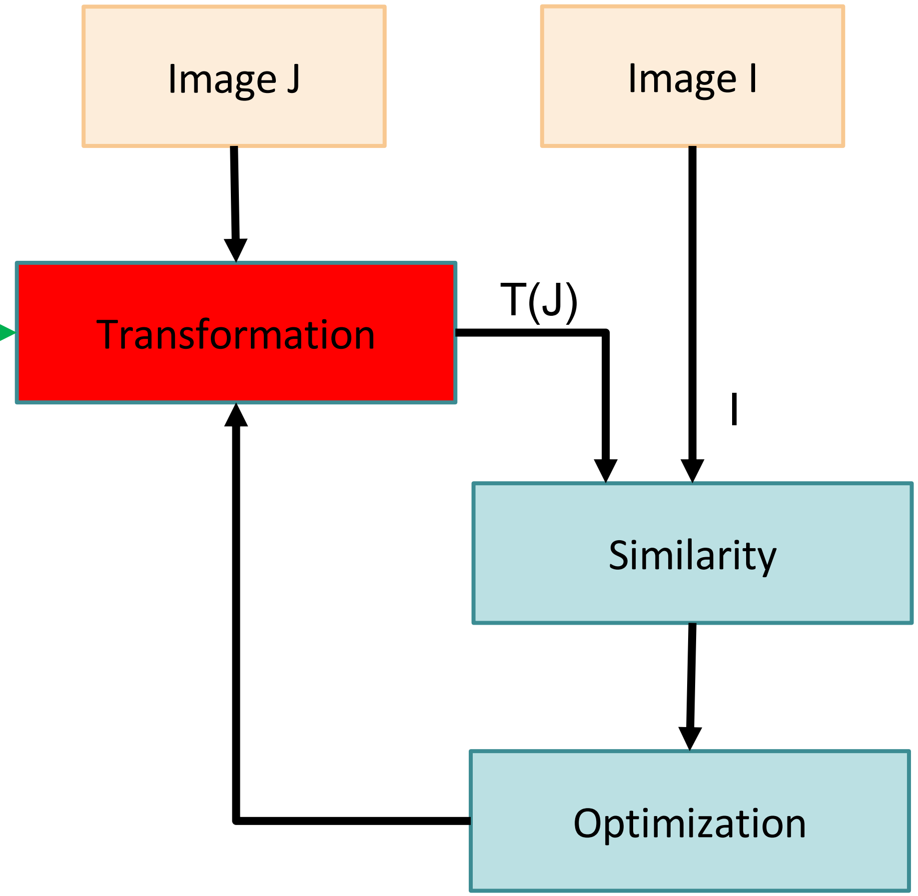
Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Part 8 – Registration

8.3 Transformation models
8.3.1 Linear transformations



$$\hat{T} = \arg\max_{T \in E} S(I, T(J))$$

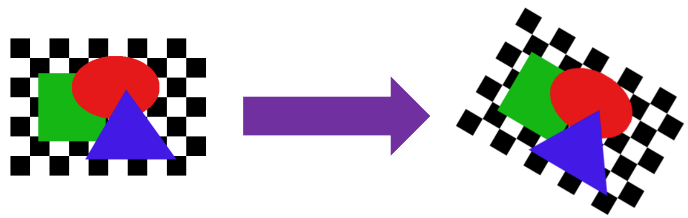


$$\hat{T} = \arg\max_{T \in E} S(I, T(J))$$

Linear transformations

Parameterisation of linear transformations (using homogeneous coordinates)

Rigid transformation



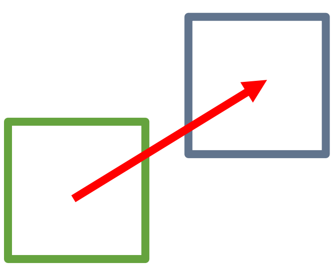
Translation

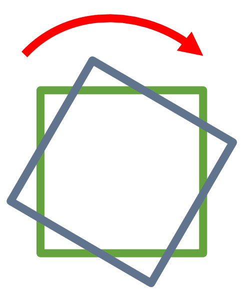
$$T = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$$

Rotation

$$R = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

3 degrees of freedom / parameters





$$A = R\,T$$

Account for the position of the patient

Inspired from: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Linear transformations

Parameterisation of linear transformations (using homogeneous coordinates)

Similarity

Translation

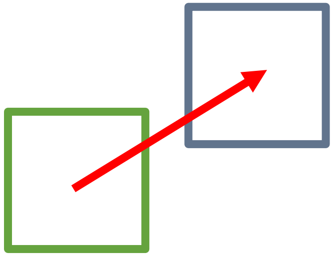
$$T = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$$

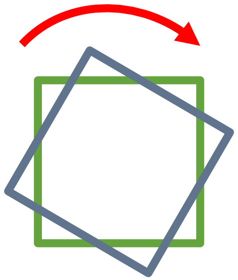
Rotation

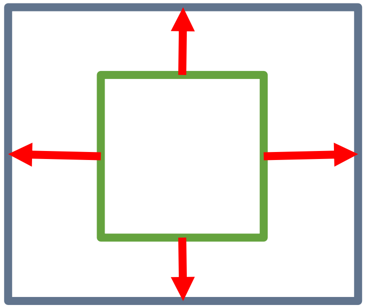
$$R = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Uniform scaling

$$S = \begin{bmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{bmatrix}$$







Account for the position of the patient and body size (e.g. head size)

Inspired from: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Linear transformations

Parameterisation of linear transformation (using homogeneous coordinates)

2D affine: translation (2), rotation (1), scaling (2), shear(1) = 6 DOF

Translation

$$T = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$$

Rotation

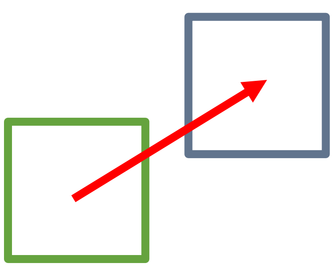
$$R = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

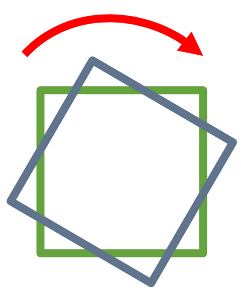
Scaling

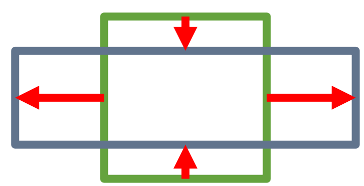
$$S = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Shearing

$$H = \begin{bmatrix} \cos\omega & \sin\omega & 0 \\ -\sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{bmatrix}$$







?

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Linear transformations

Parameterisation of linear transformation (using homogeneous coordinates)

2D affine: translation (2), rotation (1), scaling (2), shear(1) = 6 DOF

Translation

$$T = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$$

Rotation

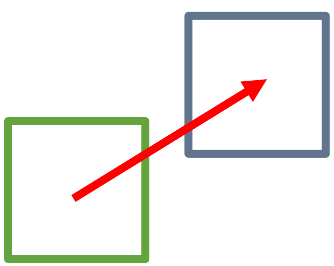
$$R = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

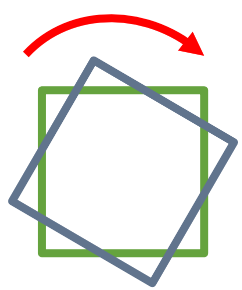
Scaling

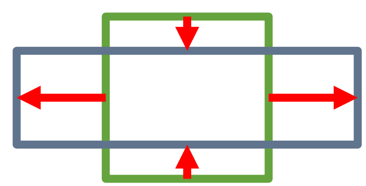
$$S = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Shearing

$$H = \begin{bmatrix} \cos\omega & \sin\omega & 0 \\ -\sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{bmatrix}$$







$$H\,S\,H^{-1}$$

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Linear transformations

Parameterisation of linear transformation (using homogeneous coordinates)

2D affine: translation (2), rotation (1), scaling (2), shear(1) = 6 DOF

Translation

$$T = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$$

Rotation

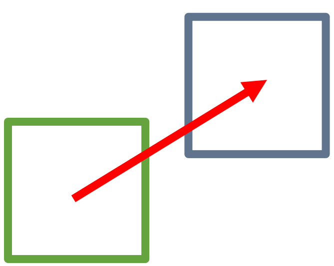
$$R = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

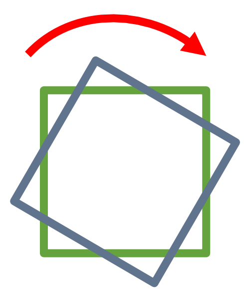
Scaling

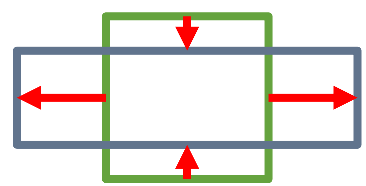
$$S = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

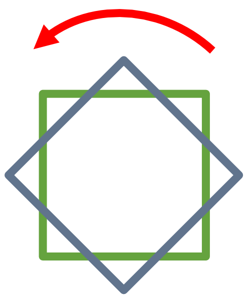
Shearing

$$H = \begin{bmatrix} \cos\omega & \sin\omega & 0 \\ -\sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{bmatrix}$$









Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Linear transformations

Parameterisation of linear transformation (using homogeneous coordinates)

2D affine: translation (2), rotation (1), scaling (2), shear(1) = 6 DOF

Translation

$$T = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$$

Rotation

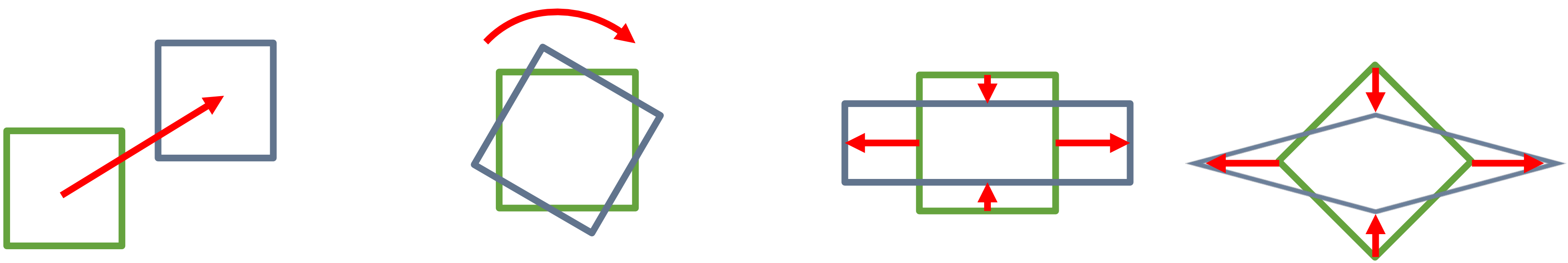
$$R = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Scaling

$$S = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Shearing

$$H = \begin{bmatrix} \cos\omega & \sin\omega & 0 \\ -\sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

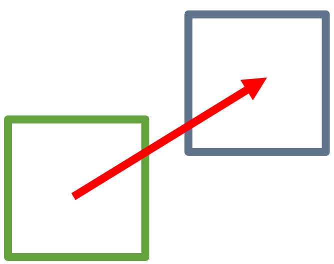
Linear transformations

Parameterisation of linear transformation (using homogeneous coordinates)

2D affine: translation (2), rotation (1), scaling (2), shear(1) = 6 DOF

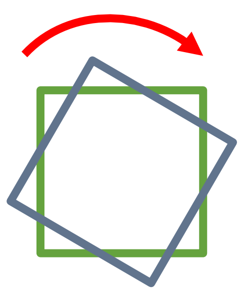
Translation

$$T = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$$



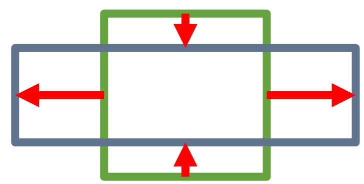
Rotation

$$R = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



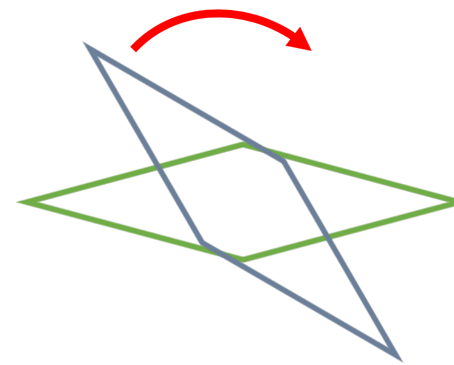
Scaling

$$S = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



Shearing

$$H = \begin{bmatrix} \cos\omega & \sin\omega & 0 \\ -\sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

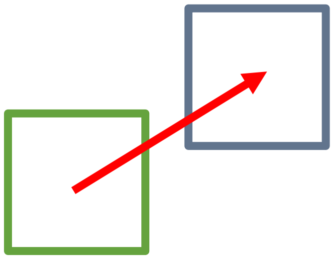
Linear transformations

Parameterisation of linear transformation (using homogeneous coordinates)

2D affine: translation (2), rotation (1), scaling (2), shear(1) = 6 DOF

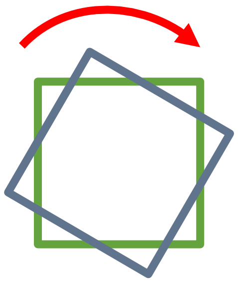
Translation

$$T = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$$



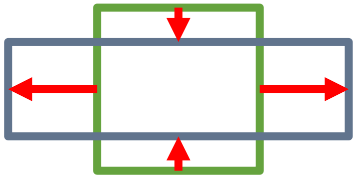
Rotation

$$R = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



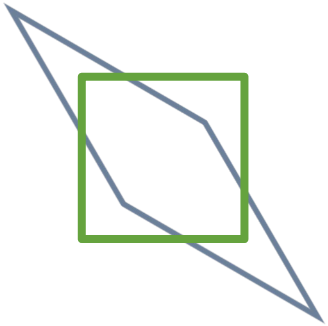
Scaling

$$S = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



Shearing

$$H = \begin{bmatrix} \cos\omega & \sin\omega & 0 \\ -\sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Linear transformations

Parameterisation of linear transformation (using homogeneous coordinates)

2D affine: translation (2), rotation (1), scaling (2), shear(1) = 6 DOF

Translation

$$T = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$$

Rotation

$$R = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Scaling

$$S = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Shearing

$$H = \begin{bmatrix} \cos\omega & \sin\omega & 0 \\ -\sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$A = H\,S\,H^{-1} R\,T = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix}$$

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

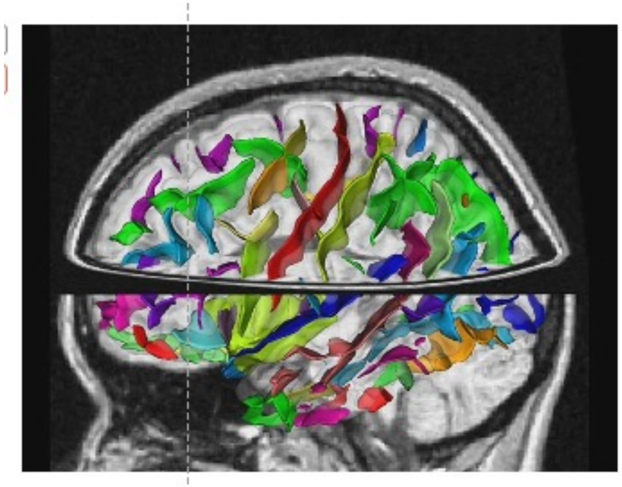
Limitations of linear transformations

Inter-subject registration

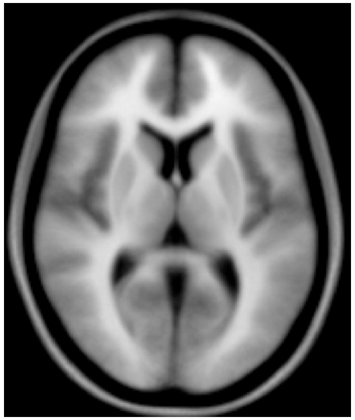
Anatomical variability is too large to be captured by linear transformations

Single subject MRI

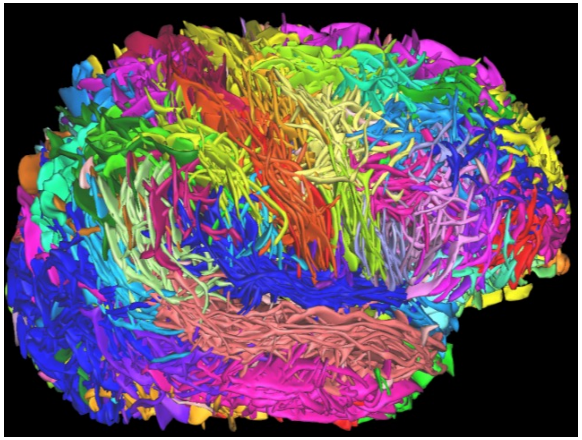
Single subject cortical sulci



Average of 305 subjects aligned with a linear registration



Superimposition of the sulci of 20 subjects after linear registration



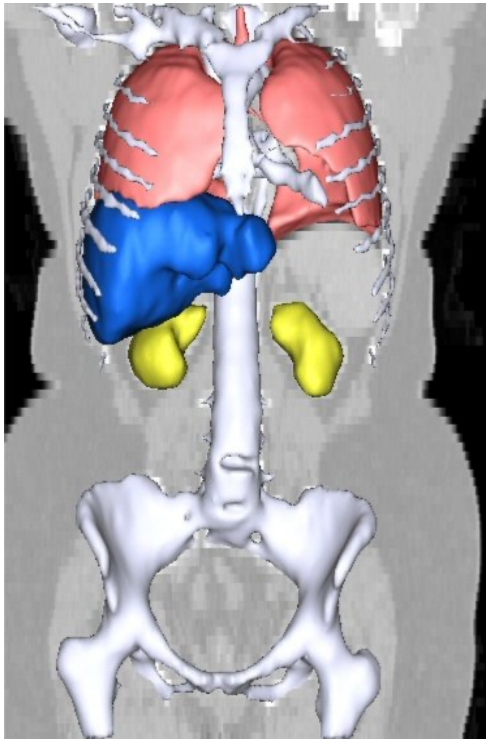
[Evans et al, 1993]

[from Mangin et al]

Limitations of linear transformations

Intra-subject registration

Some organs may have a very different shape depending on the modality (e.g. if they have moved)



CT

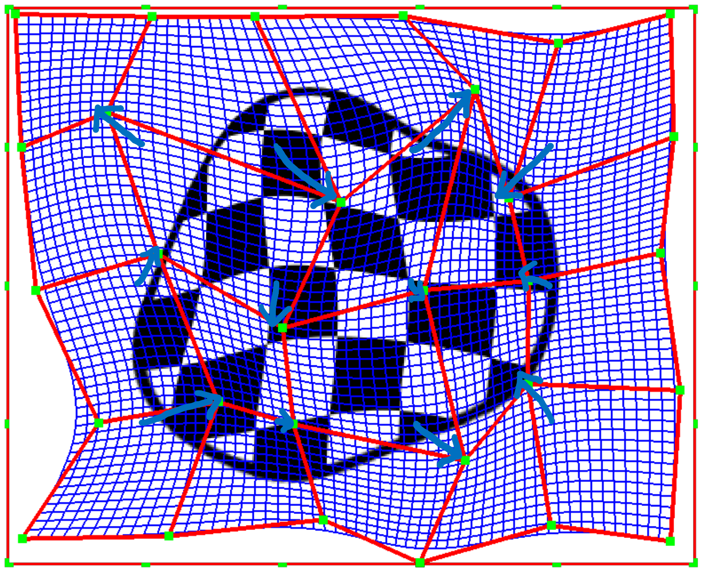
PET

[Camara et al, IEEE TMI, 2007]

Part 8 – Registration

8.3 Transformation models
8.3.2 Non-linear transformations

Free-form deformations (FFD)



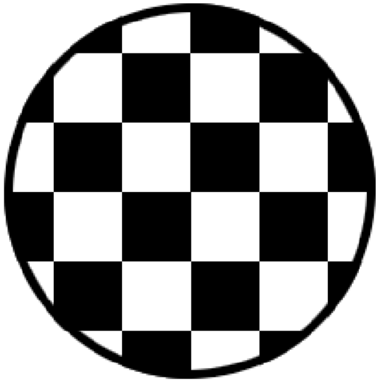
Principle:
Deform an object by manipulating an underlying mesh of control points

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Reference: Rueckert et al, IEEE TMI, 1999

Free-form deformations (FFD)

Object to be deformed



Principle:

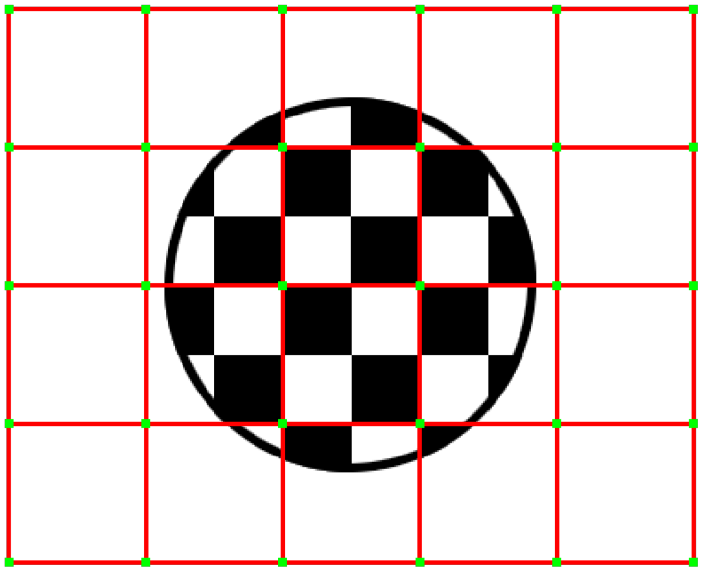
Deform an object by manipulating an underlying mesh of control points

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Reference: Rueckert et al, IEEE TMI, 1999

Free-form deformations (FFD)

Underlying mesh of control points



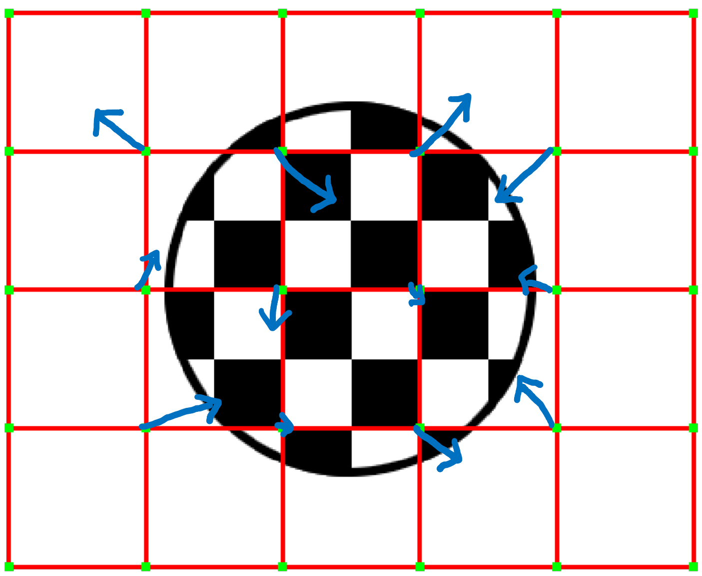
Principle:
Deform an object by manipulating an underlying mesh of control points

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Reference: Rueckert et al, IEEE TMI, 1999

Free-form deformations (FFD)

Move the control points



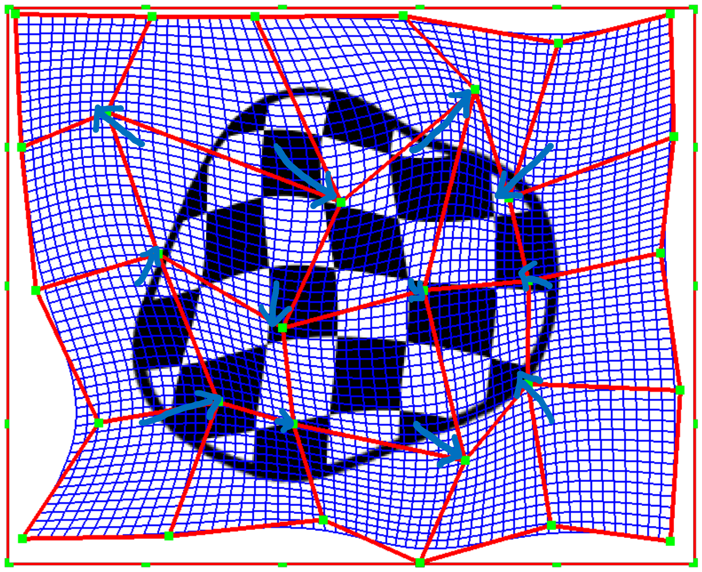
Principle:
Deform an object by manipulating an underlying mesh of control points

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Reference: Rueckert et al, IEEE TMI, 1999

Free-form deformations (FFD)

Deform the control point accordingly
Deform the underlying image grid (interpolation)



Principle:
Deform an object by manipulating an underlying mesh of control points

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Reference: Rueckert et al, IEEE TMI, 1999

Free-form deformations (FFD)

The deformation is defined with **B-splines**

Let Φ denote a $n_x \times n_y \times n_z$ mesh of control points $\phi_{i,j,k}$ with uniform spacing δ. Then, the FFD can be written as the 3 -D tensor product of the familiar 1 -D cubic B-splines

$$
\begin{aligned}
&\mathbf{T}_{\text{local}}(x, y, z) \\
&\qquad = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_l(u) B_m(v) B_n(w) \phi_{i+l, j+m, k+n}
\end{aligned}
$$

$$
\begin{aligned}
&\text{where } i = \lfloor x/n_x \rfloor - 1, j = \lfloor y/n_y \rfloor - 1, k = \lfloor z/n_z \rfloor - 1 \\
&u = x/n_x - \lfloor x/n_x \rfloor, v = y/n_y - \lfloor y/n_y \rfloor, w = z/n_z - \lfloor z/n_z \rfloor
\end{aligned}
$$

and where B_l represents the l th basis function of the B-spline

$$
\begin{aligned}
B_0(u) &= (1-u)^3/6 \\
B_1(u) &= \left(3u^3 - 6u^2 + 4\right)/6 \\
B_2(u) &= \left(-3u^3 + 3u^2 + 3u + 1\right)/6 \\
B_3(u) &= u^3/6
\end{aligned}
$$

Reference: Rueckert et al, IEEE TMI, 1999

Free-form deformations (FFD)

Number of control points

More control points means a more flexible transformation

Reference: Rueckert et al, IEEE TMI, 1999

Free-form deformations (FFD)

In order to produce reasonably realistic deformations, one needs to **add a smoothness constraint** on the transformation

$$\begin{aligned}\mathcal{C}_{\text{smooth}} = \frac{1}{V}\int_0^X\int_0^Y\int_0^Z\Bigg[&\left(\frac{\partial^2\mathbf{T}}{\partial x^2}\right)^2 + \left(\frac{\partial^2\mathbf{T}}{\partial y^2}\right)^2 \\ &+ \left(\frac{\partial^2\mathbf{T}}{\partial z^2}\right)^2 + 2\left(\frac{\partial^2\mathbf{T}}{\partial xy}\right)^2 + 2\left(\frac{\partial^2\mathbf{T}}{\partial xz}\right)^2 \\ &+2\left(\frac{\partial^2\mathbf{T}}{\partial yz}\right)^2\Bigg]dxdydz\end{aligned}$$

The **objective function** is a compromise between similarity of images and smoothness of the transformation

$$\mathcal{C}_{\text{similarity}}\left(I, \mathbf{T}(I)\right) + \lambda\mathcal{C}_{\text{smooth}}\left(\mathbf{T}\right)$$

Reference: Rueckert et al, IEEE TMI, 1999

Elastic deformations

The transformation is defined by a displacement field u over the whole image

$$\phi_u(x) = x + u(x)$$

where x is a point in $\mathbb{R}^2$ (or $\mathbb{R}^3$ for a 3D image)

The registration problem is then

$$\hat{u} = \arg\max_{u \in H} S(I, \phi_u(J)) + Reg(u)$$

Problem: no garantee that the transformation is invertible. In particular, invertibility would be only obtained for very small displacement fields

LDDMM (Large Deformation Diffeomorphic Metric Mapping)

Diffeomorphisms: invertible, differentiable deformations which inverse is also differentiable

LDDMM: mathematical framework for diffeomorphic registration

- principled approach
- can be used for morphometry through analysis of the computed deformations

References: Trouvé, IJCV, 1998; Dupuis et al, 1998; Beg et al, IJCV, 2005

LDDMM (Large Deformation Diffeomorphic Metric Mapping)

Intuition: One can consider ϕ as the concatenation of small deformations ϕ_{u_i}.

$\phi = \Phi_n$ where the Φ_k are recursively defined by $\Phi_0 = Id$ and $\Phi_{k+1} = \phi_{u_{k+1}} \circ \Phi_k$

LDDMM: The deformation ϕ is defined as the integration of a time-dependent velocity vector field $v(.,t)$:

$$\begin{cases} \frac{\partial \phi_v}{\partial t}(x,t) = v(\phi_v(x,t),t) & \forall t \in [0,1] \\ \phi_v(x,0) = x & \forall x \in \mathbb{R}^3 \end{cases}$$

Position of x at time t

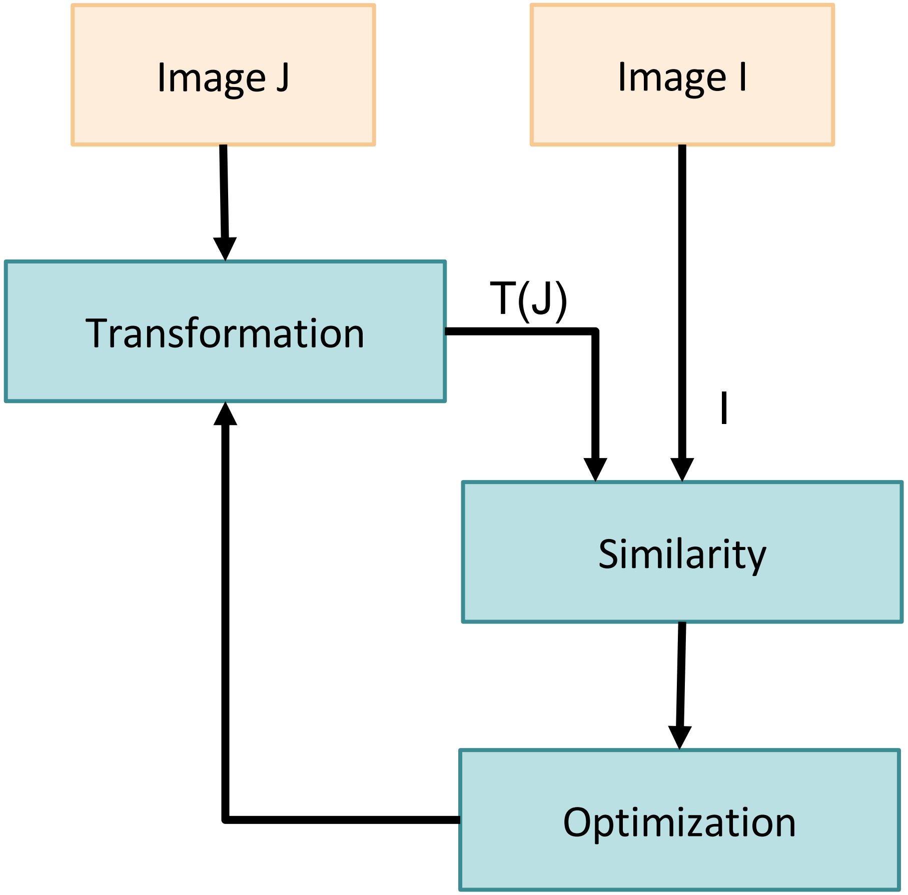
The resulting transformation is $\phi = \phi_v(x,1)$

One can define a right-invariant distance on the group of diffeomorphisms: $D(Id,\phi) = \inf\left\{ \int_0^1 \|v(\cdot,t)\|_V dt; v \in L^2([0,1],V), \phi_v = \phi \right\}$

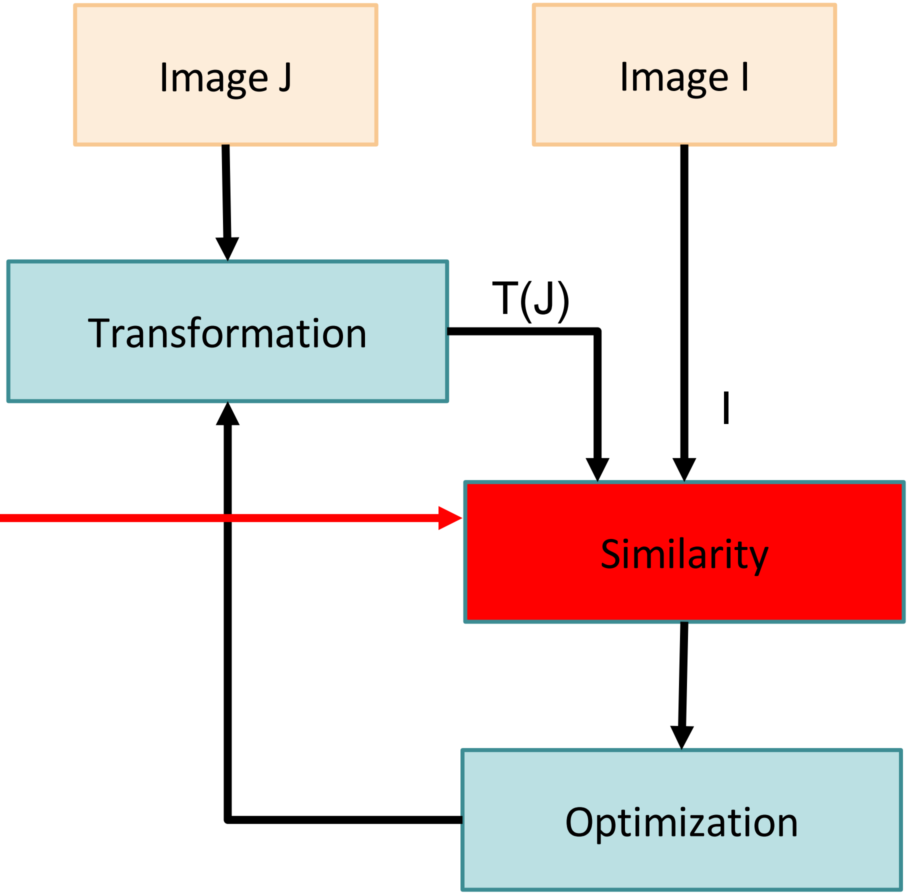
References: Trouvé, IJCV, 1998; Dupuis et al, 1998; Beg et al, IJCV, 2005

Part 8 – Registration

8.4 Similarity measures



$$\hat{T} = \arg\max_{T \in E} S(I, T(J))$$



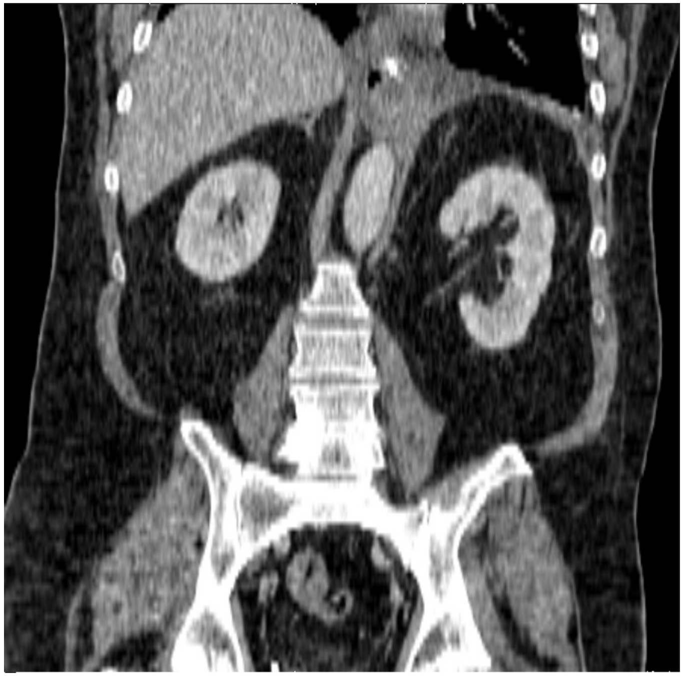
$$\hat{T} = \arg\max_{T \in E} S(I, T(J))$$

Similarity measures

Aim: register source (J) onto target (I)

Source J

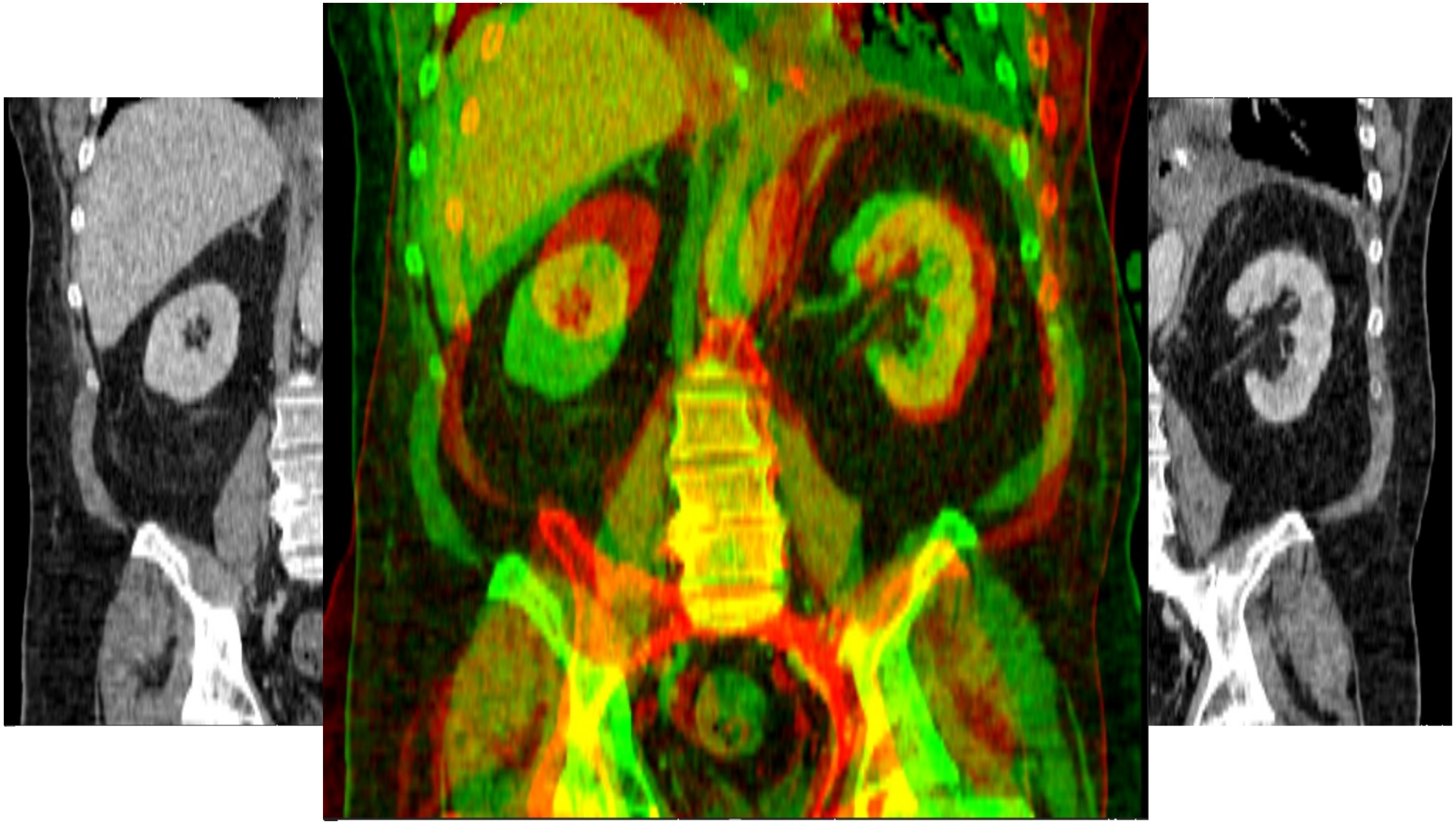
Source I



Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Similarity measures

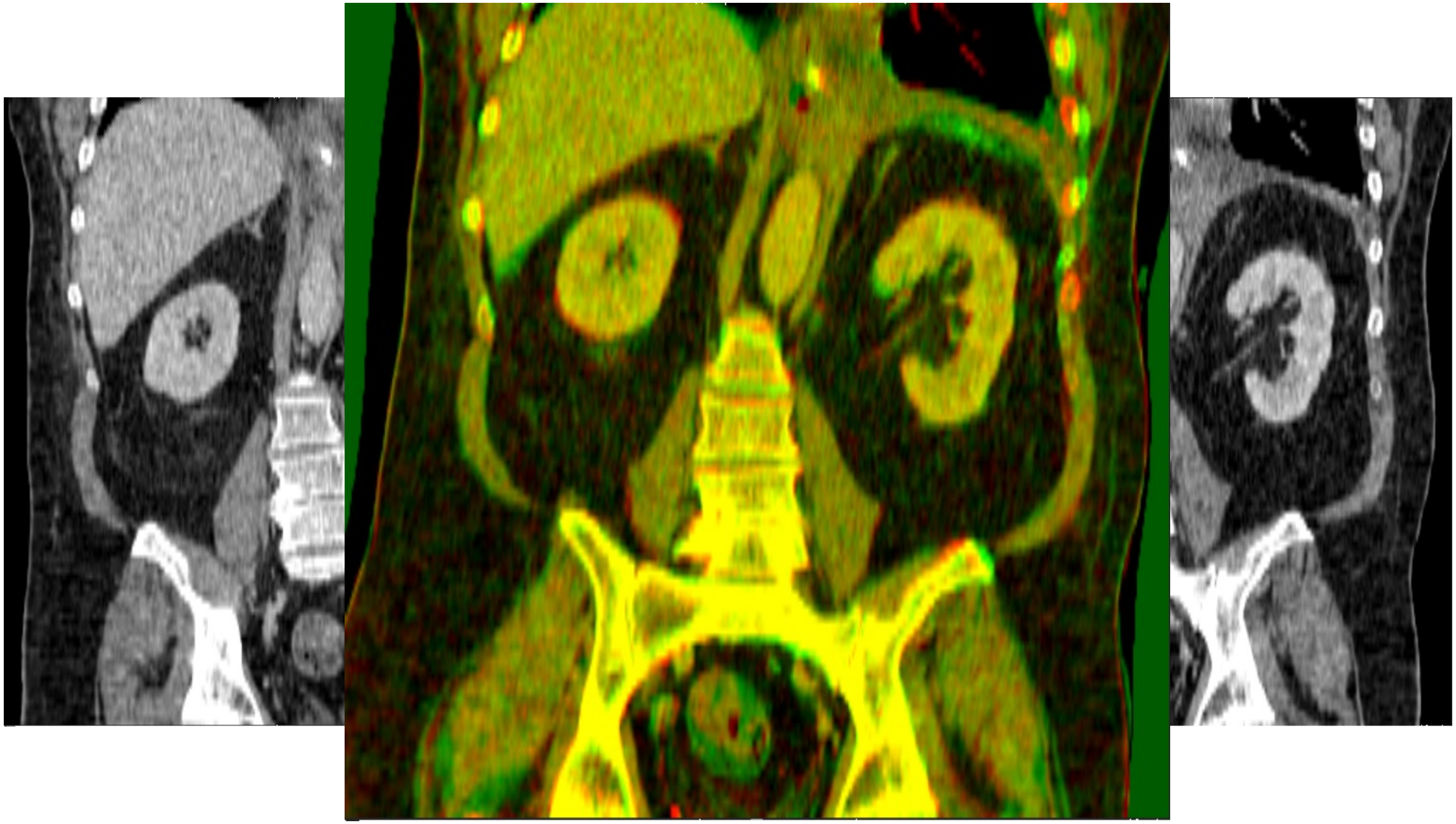
Superimposition before registration



Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Similarity measures

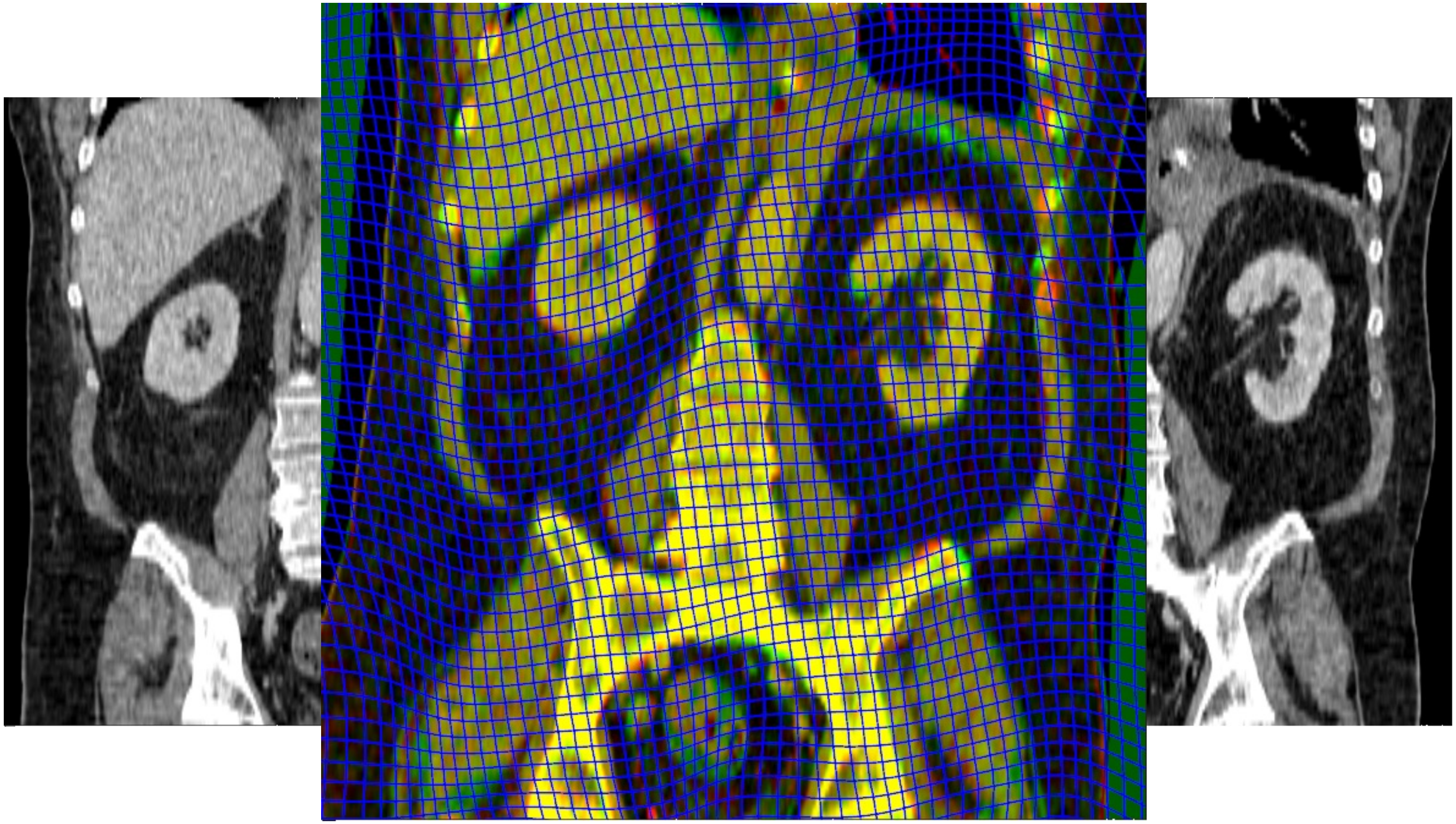
Superimposition after registration (superimposition of I and T(J)



Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Similarity measures

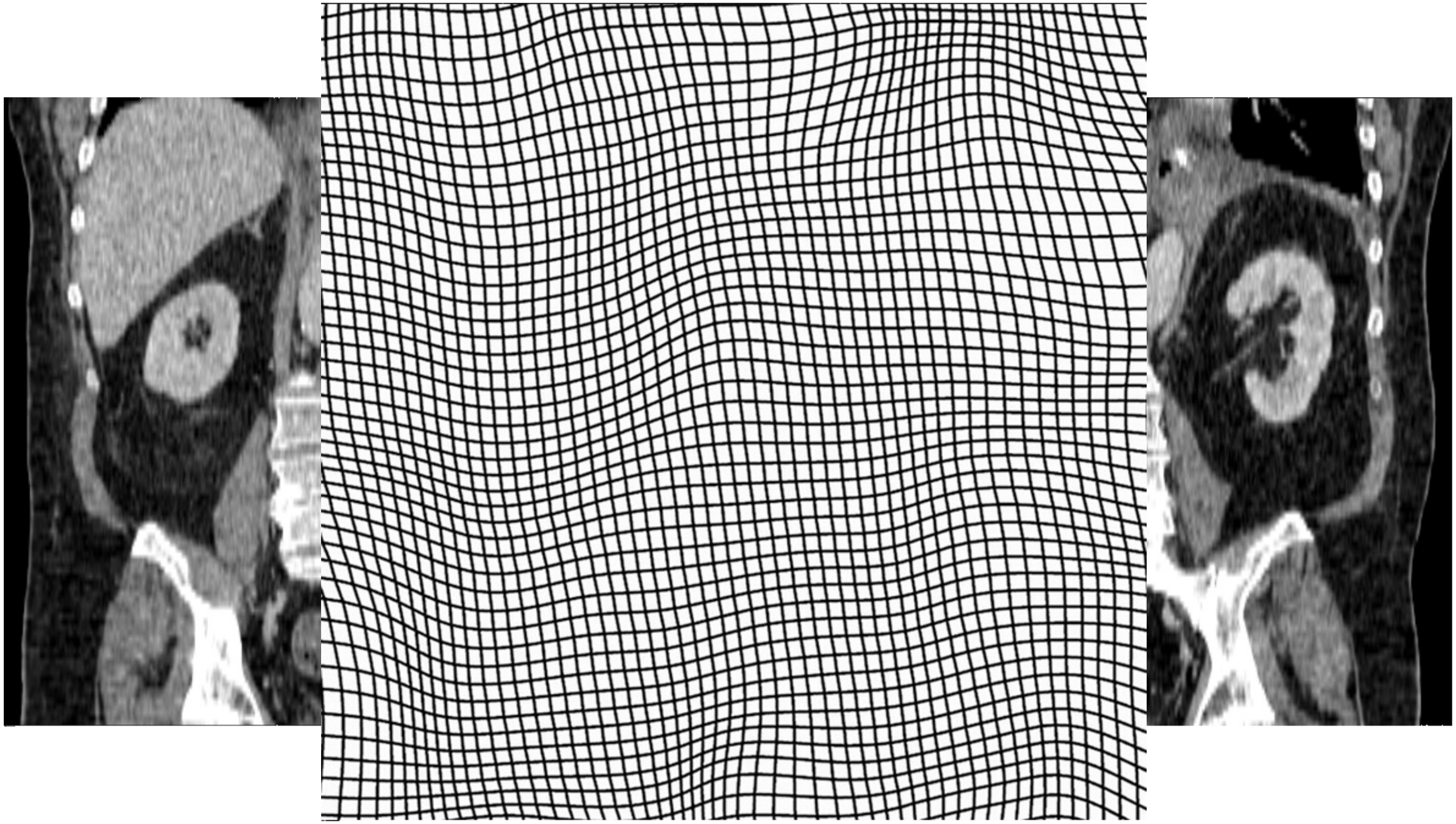
Transformation



Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

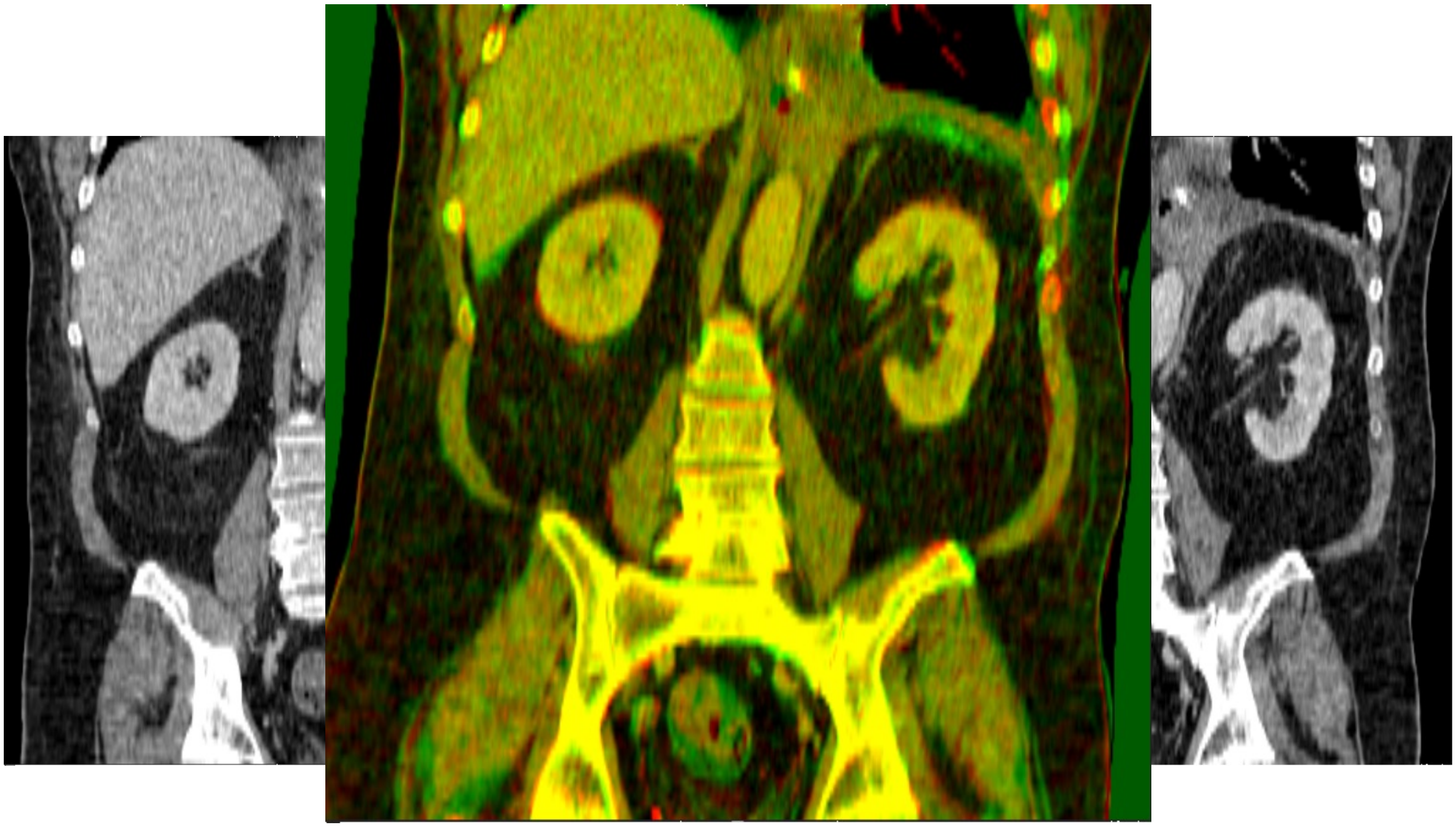
Similarity measures

Transformation



Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Similarity measures



Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Similarity measures

Superimposition before registration

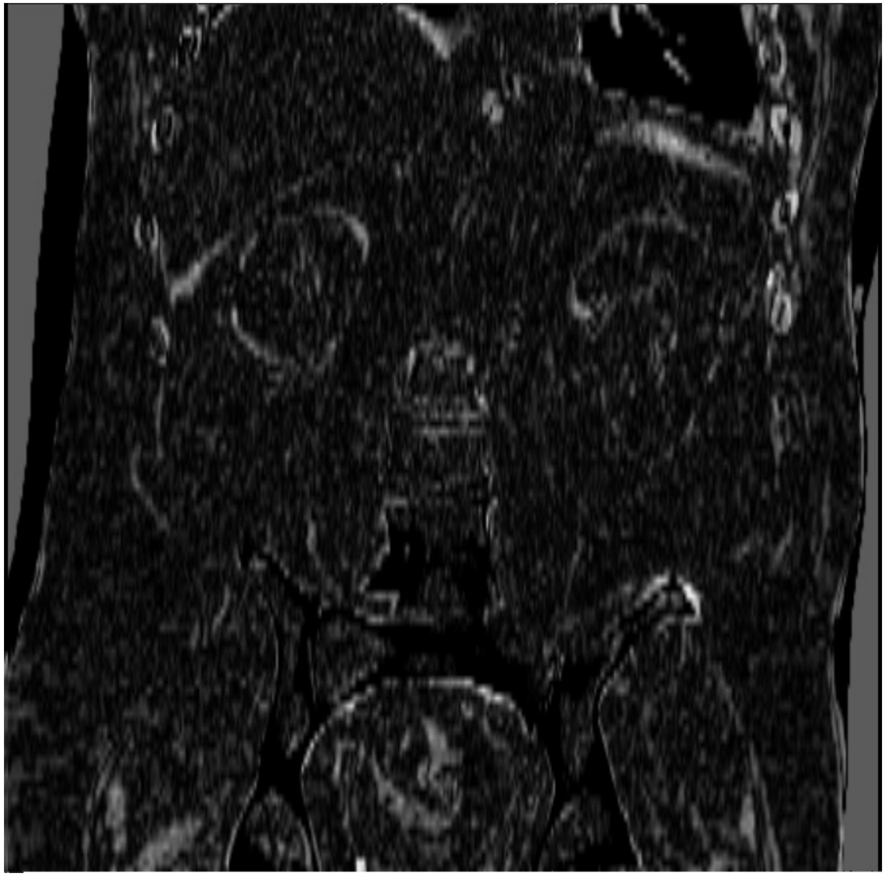
Superimposition after registration

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Similarity measures

Difference before registration |I-J|

Difference after registration |I-T(J)|



-|I-T(J)| is a possible similarity measure

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Similarity measures

$$\hat{T} = \arg\max_{T \in E} S(I, T(J))$$

Similarity measure

Equivalently, we can define a dissimilarity measure

$$D\big(I, T(J)\big) = -S(I, T(J))$$

Similarity measures

Mono-modal registration

- Image intensities are related by a (simple) function

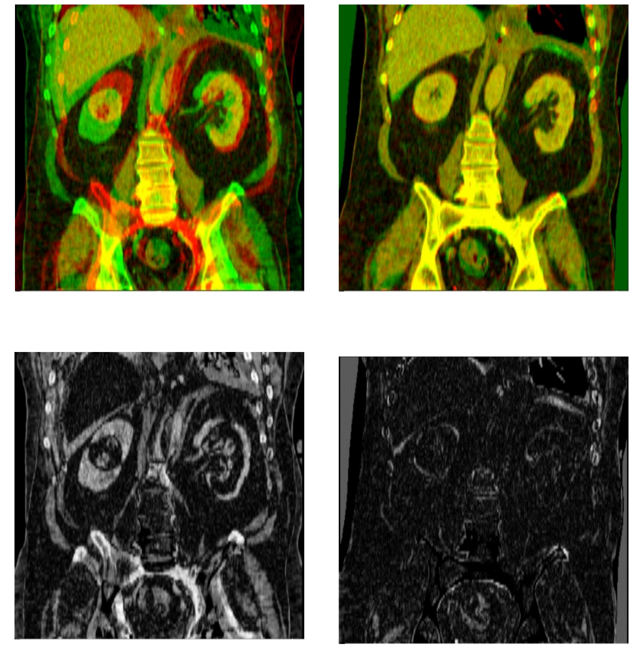
Multi-modal registration

- Image intensities are related by a complex function or statistical relationship

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

(Dis)similarity measures

Intensity differences



Sum of squared differences (SSD)

$$D_{SSD}(I,T(J)) = \frac{1}{N}\sum_{i=1}^{N}(I(x_i) - T(J(x_i)))^2$$

Sum of absolute differences (SAD)

$$D_{SAD}(I,T(J)) = \frac{1}{N}\sum_{i=1}^{N}|I(x_i\ )-\mathrm{T}(J(x\_i))|$$

Assumption: identity relationship between intensity distributions

Application:

- mono-modal registration (e.g. CT-CT)
- can even be to simple for some mono-modal registrations (e.g. MR-MR)

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

(Dis)similarity measures

Correlation coefficient (CC)

$$D_{CC}(I,T(J)) = -\frac{\overbrace{\frac{1}{N}\sum_{i=1}^{N}(I(x_i)-\mu_I)\big(T(J(x_i))-\mu_J\big)}^{Cov(I,T(J))}}{\underbrace{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(I(x_i)-\mu_I)^2}}_{\sigma_I}\underbrace{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\big(T(J(x_i))-\mu_J\big)^2}}_{\sigma_{T(J)}}}$$

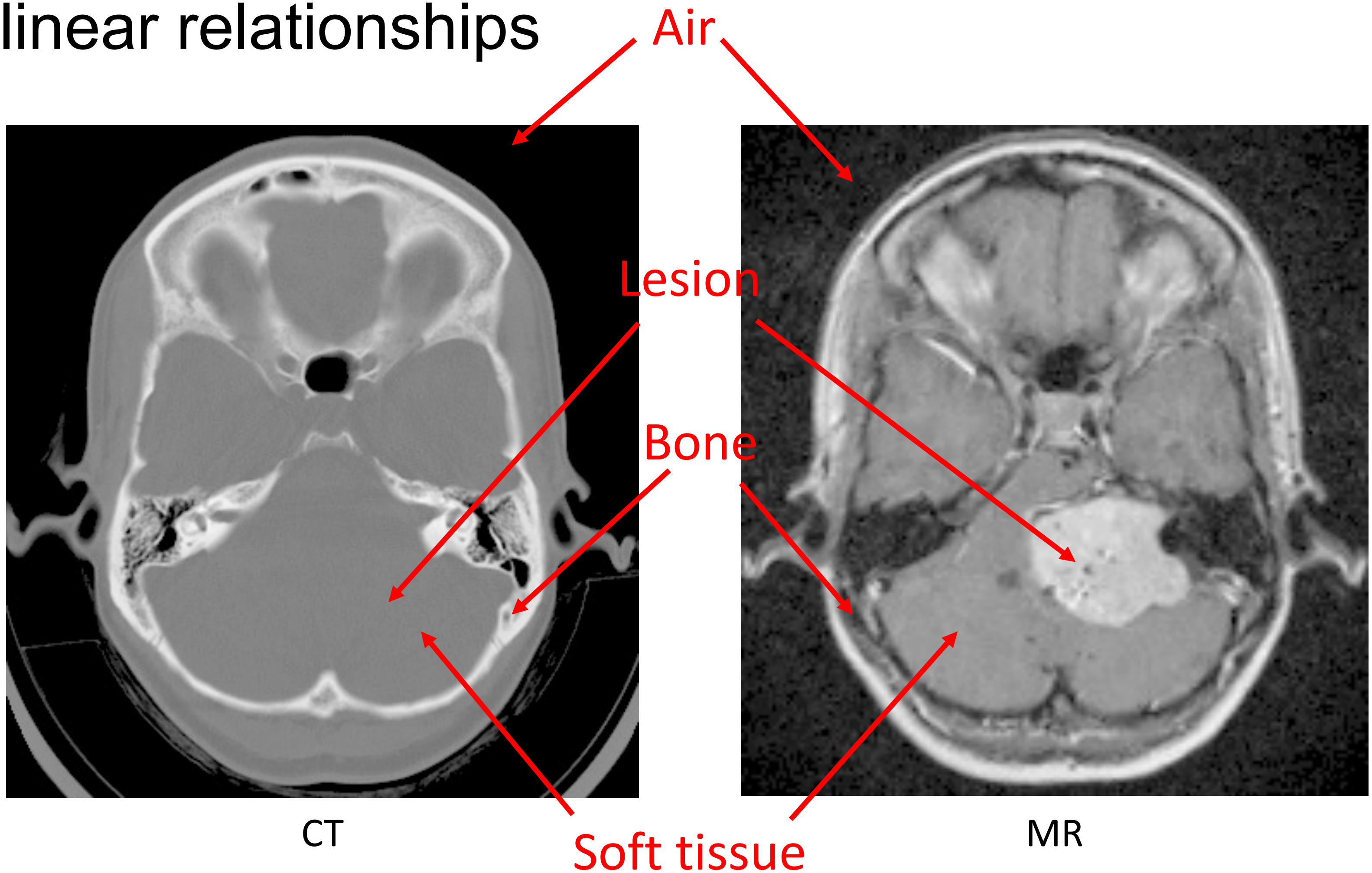
Assumption: linear relationship between intensity distributions

Application: (mainly) mono-modal registration (e.g. MR-MR)

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

(Dis)similarity measures

Non linear relationships

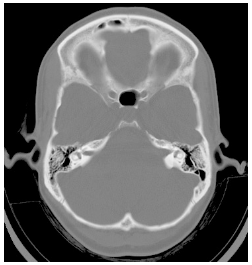


CT MR

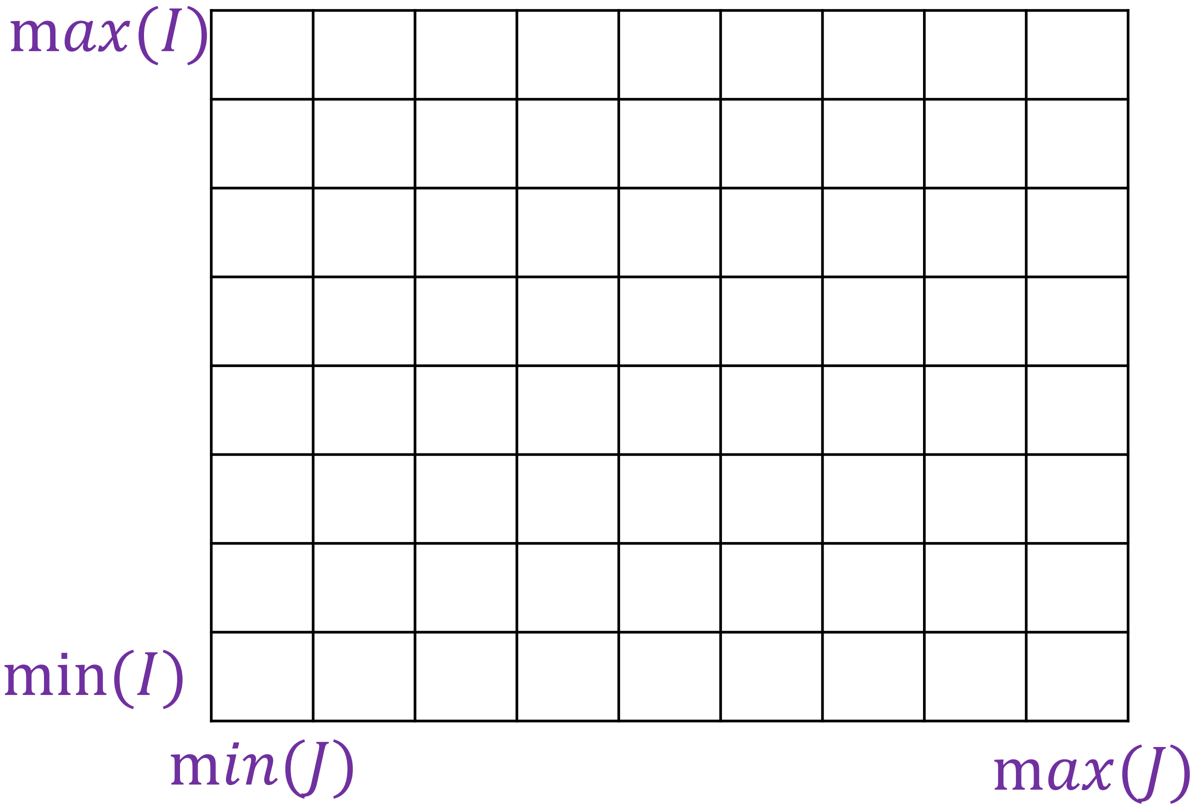
Adapted from: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

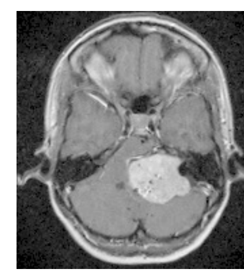
(Dis)similarity measures

2D intensity histograms



I

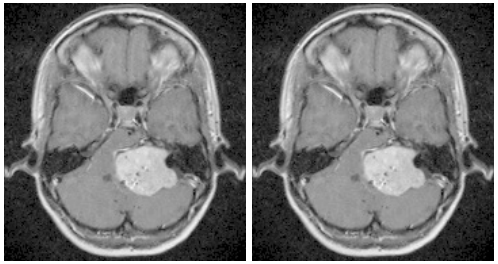


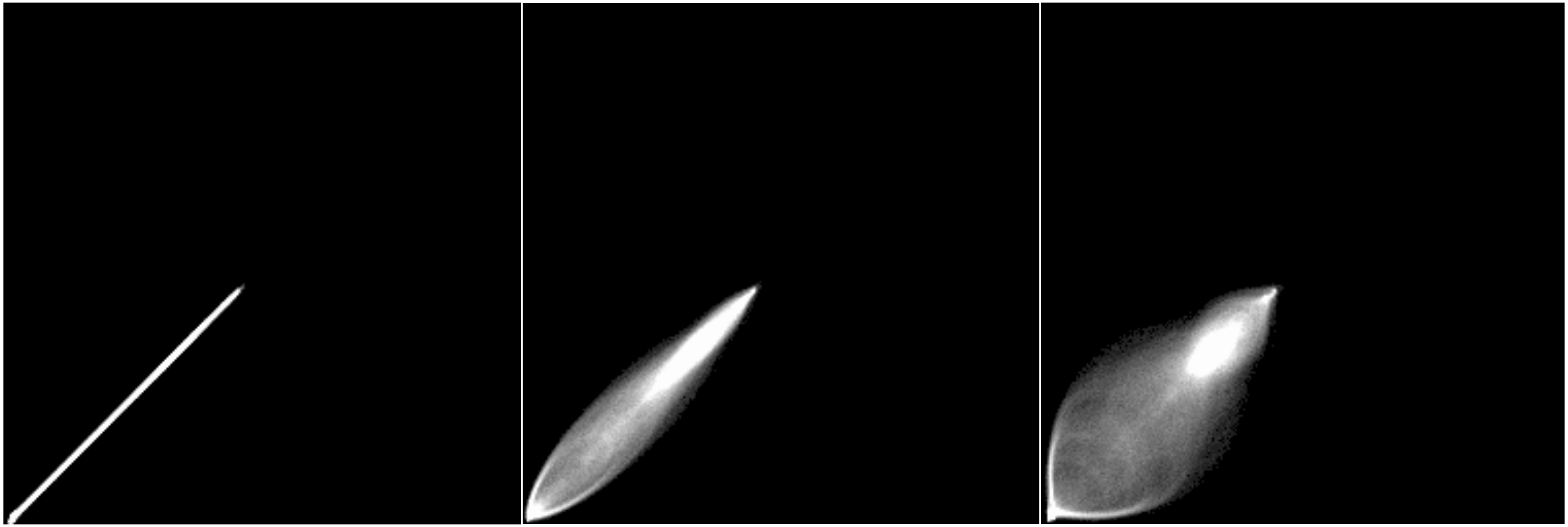


J

(Dis)similarity measures

MR/MR



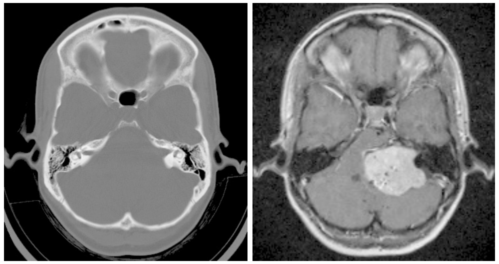


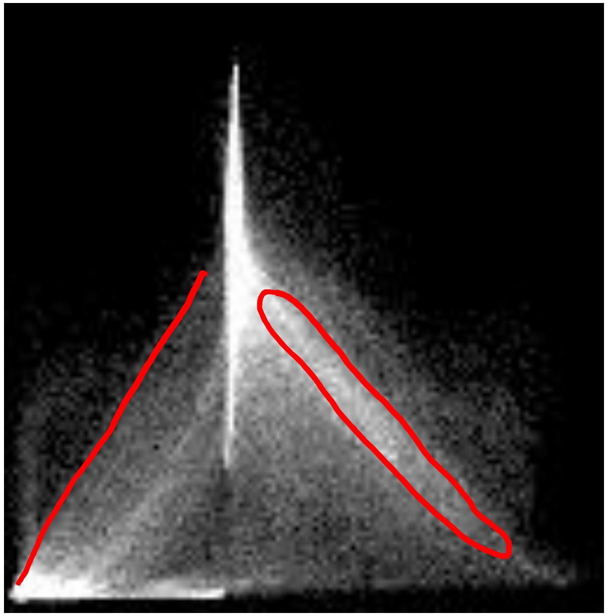
registered misregistered by 2mm misregistered by 5mm

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

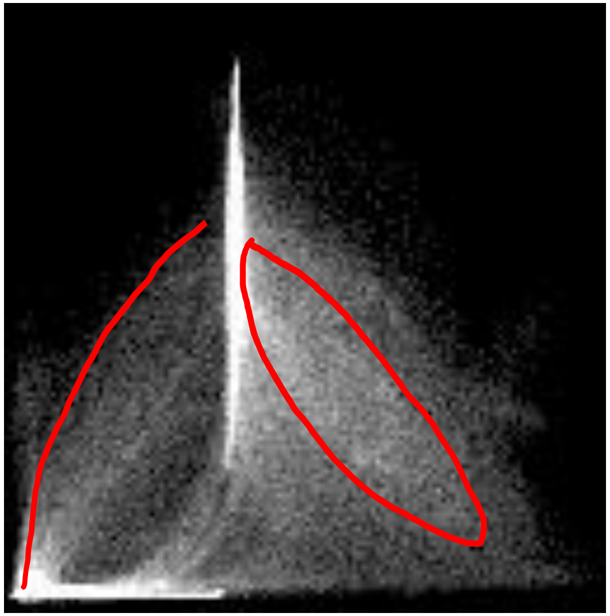
(Dis)similarity measures

CT/MR

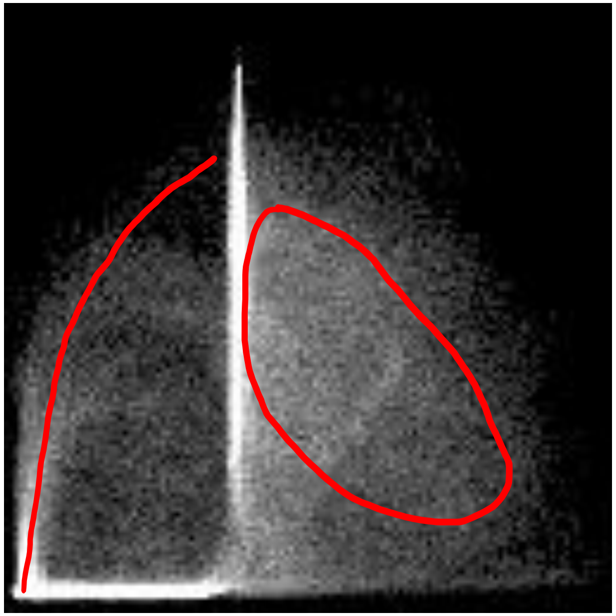




registered



misregistered by 2mm

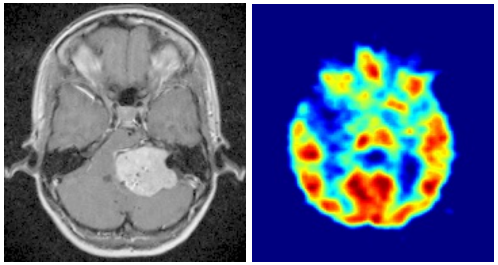


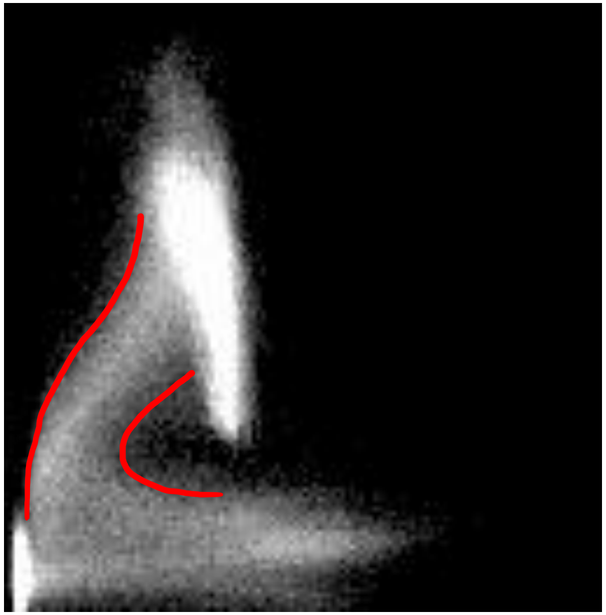
misregistered by 5mm

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

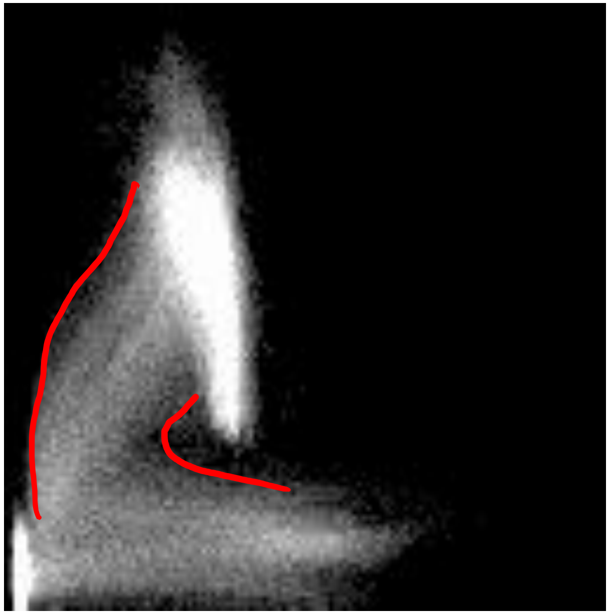
(Dis)similarity measures

MR/PET

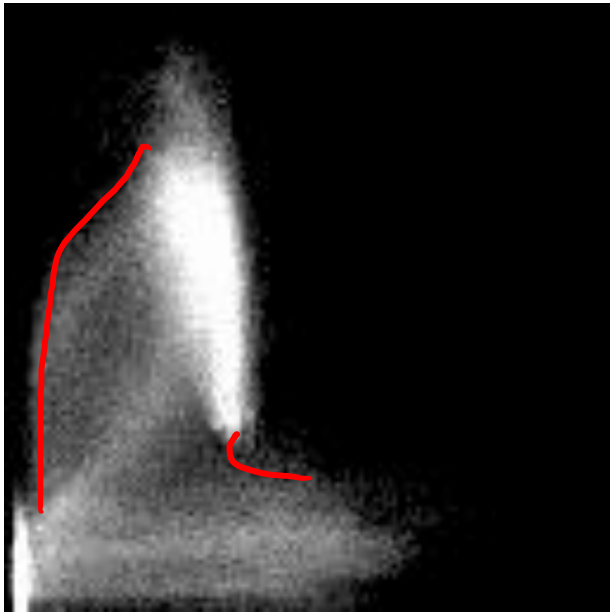




registered



misregistered by 2mm



misregistered by 5mm

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

(Dis)similarity measures

Intensity distributions

$$p(i,j) = \frac{h(i,j)}{N}$$ – counts in histogram

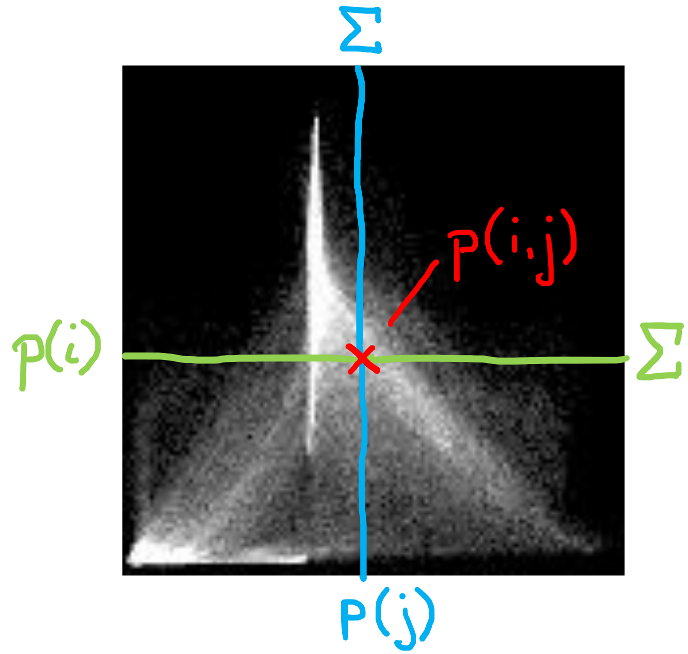
joint probability of an image point having a value i in image I and value j in image J

$$p(i) = \sum_j p(i,j)$$

marginal probability of an image point having a value i in image I

$$p(j) = \sum_i p(i,j)$$

marginal probability of an image point having a value j in image J



(Dis)similarity measures

Shannon entropy

$$H(I) = -\sum_i p(i) \log p(i)$$

amount of information contained in image I

Joint entropy

$$H(I,J) = -\sum_i \sum_j p(i,j) \log p(i,j)$$

amount of information contained in the combined image I, J

Could be used for registration…
But there are some drawbacks

(Dis)similarity measures

Mutual information [Viola et al. 1995, Collignon et al, 1995]

$$MI(I,J) = H(I) + H(J) - H(I,J)$$

describes how well one image can be explained by another image

This can be rewritten in terms of marginal and joint probabilities

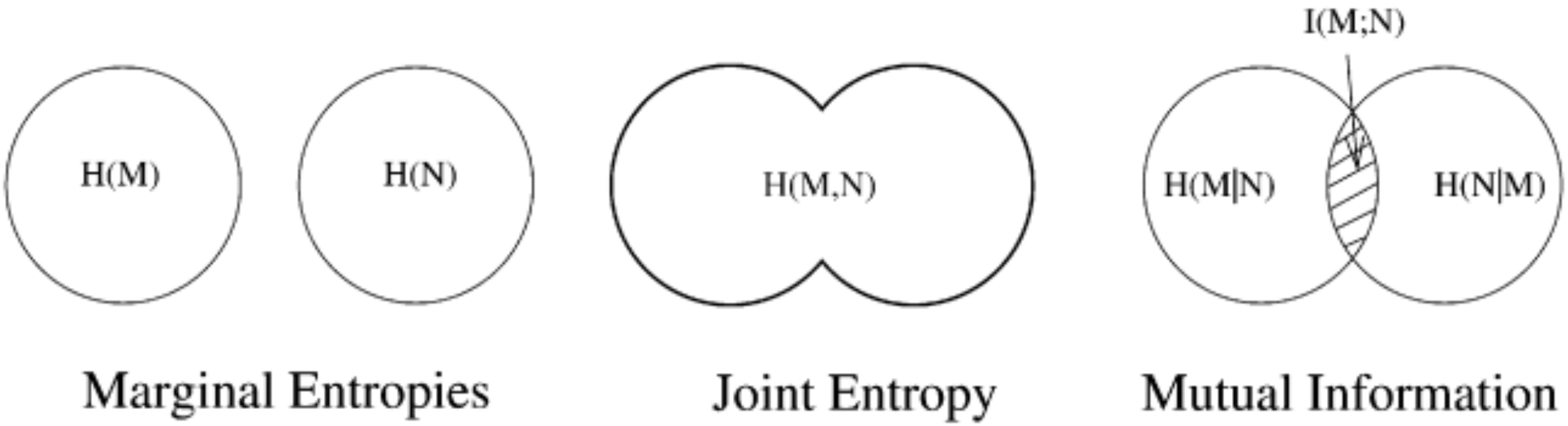
$$MI(I,J) = -\sum_{i}\sum_{j} p(i,j)\log\frac{p(i,j)}{p(i)\,p(j)}$$

The dissimilarity measure is then defined as

$$D_{MI}(I,T(J)) = -MI(I,T(J))$$

(Dis)similarity measures

Mutual information vs Joint entropy



By minimising joint entropy we are simply trying to find the overlap which contains least information, not necessarily the most corresponding information. What we need to do is to relate changes in the value of the joint entropy H(M, N) to the marginal entropies of the two images H(M) and H(N) derived from their region of overlap.

(Studholme et al, Pattern Recognition, 1999)

(Dis)similarity measures

Normalised mutual information [Studholme et al. 1999]

$$NMI(I,J) = \frac{H(I) + H(J)}{H(I,J)}$$

is independent of the amount of overlap between images

The dissimilarity measure is then defined as

$$D_{NMI}(I \circ T, J) = -NMI(I \circ T, J)$$

Assumption: statistical relationship between intensity distributions
Application: (mainly) multi-modal registration (e.g. CT-MR)

(Dis)similarity measures

Normalized mutual information vs mutual information

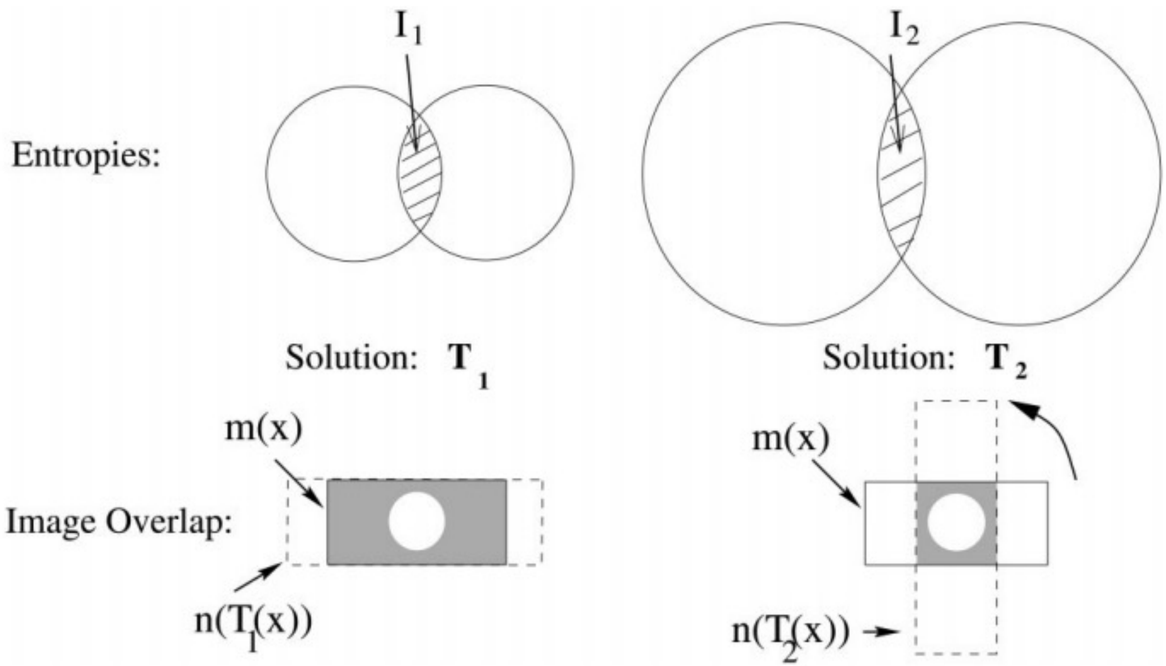
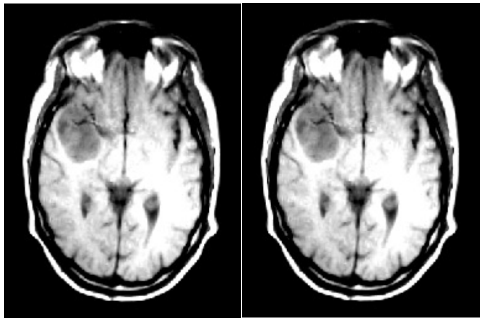


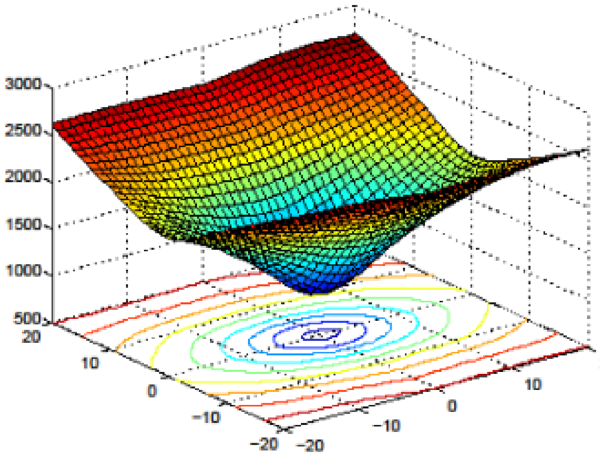
Fig. 5. As we vary rotational alignment (T_1 and T_2) of two images ($m(x)$ and $n(y)$) of a circle we would expect any measure not to favour a particular alignment. Here though direct measures of information such as the joint entropy and mutual information are influenced by the change in image statistics, as the proportion of background in the region of overlap varies. Overall estimate T_2 has most similar foreground and background probability and therefore greatest uncertainty and information content.

(Studholme et al, Pattern Recognition, 1999)

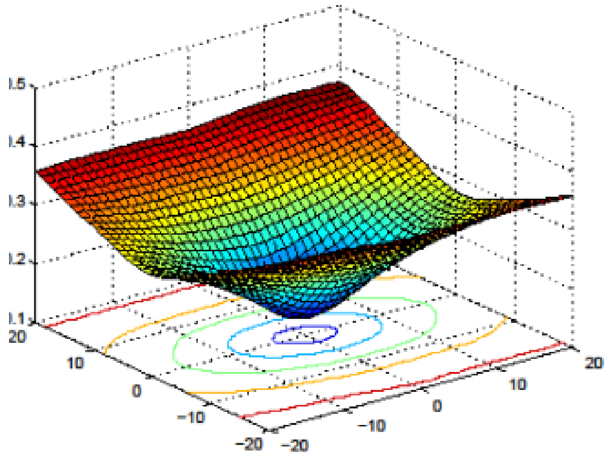
(Dis)similarity measures

Translation experiment: mono-modal

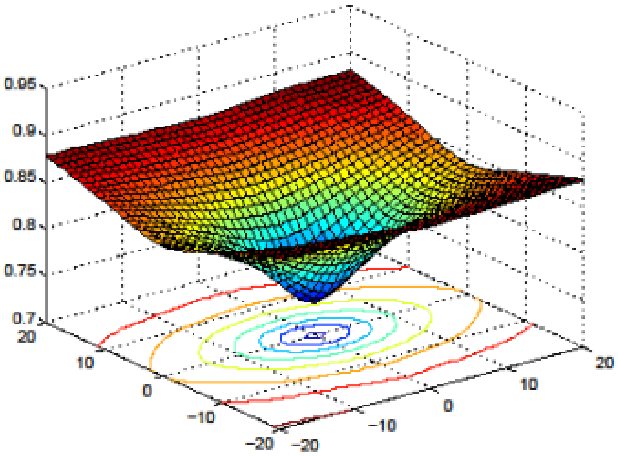




D_{SSD}



D_{CC}

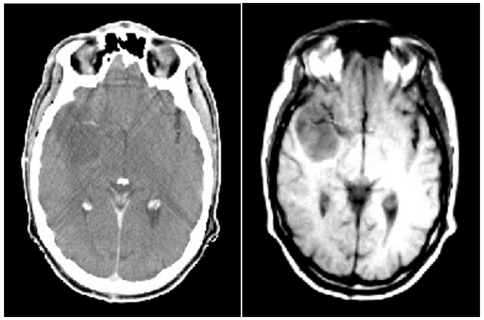


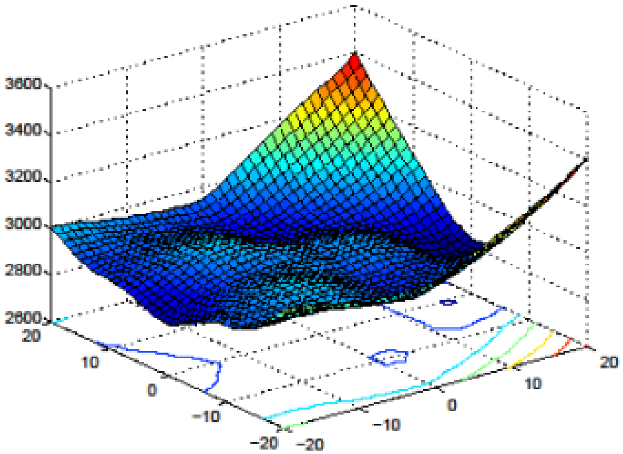
D_{NMI}

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

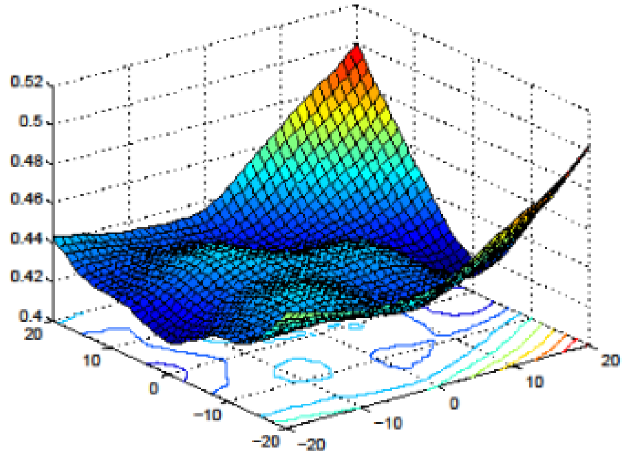
(Dis)similarity measures

Translation experiment: multi-modal

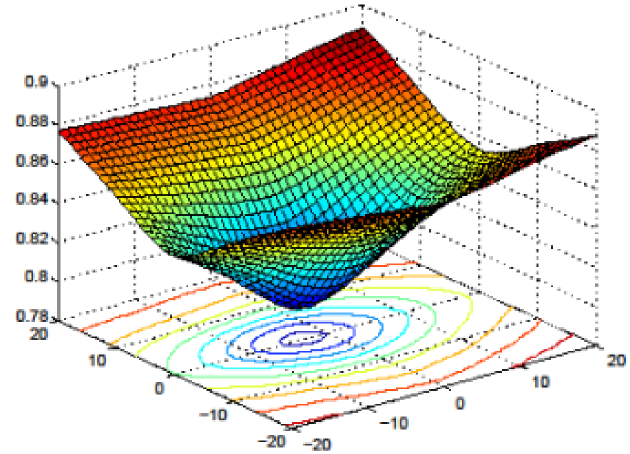




D_{SSD}



D_{CC}

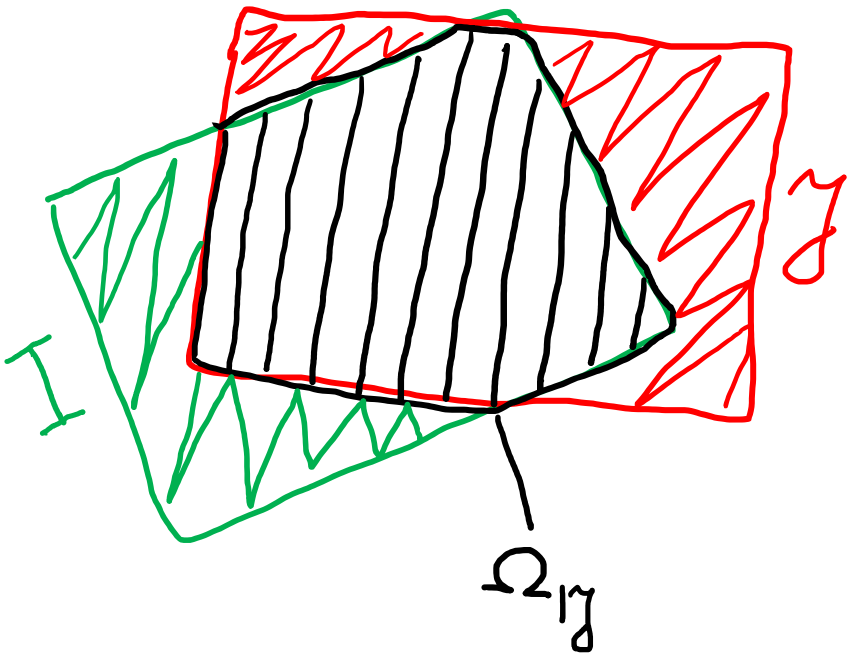


D_{NMI}

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

(Dis)similarity measures

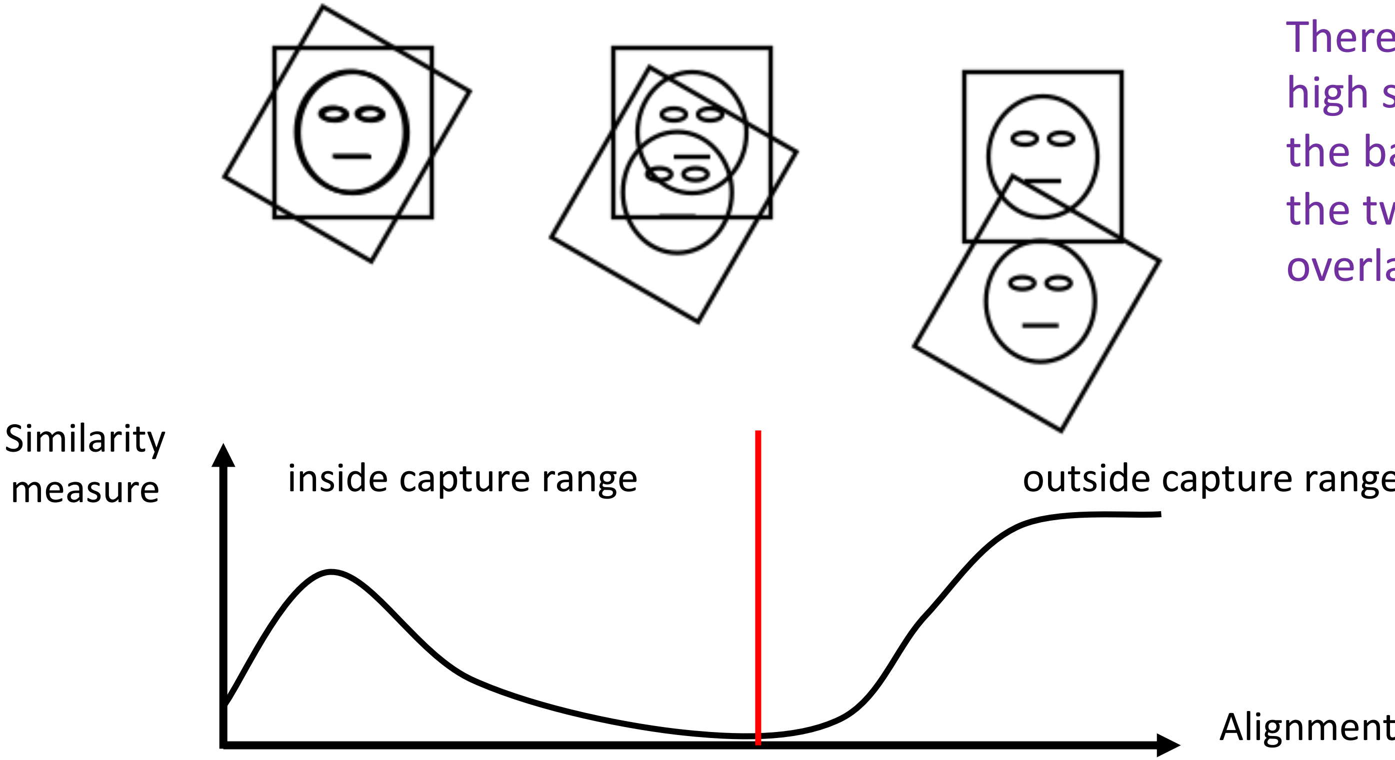
(Dis)similarity measures are evaluated in the overlapping region of the two images



Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

(Dis)similarity measures

Capture Range



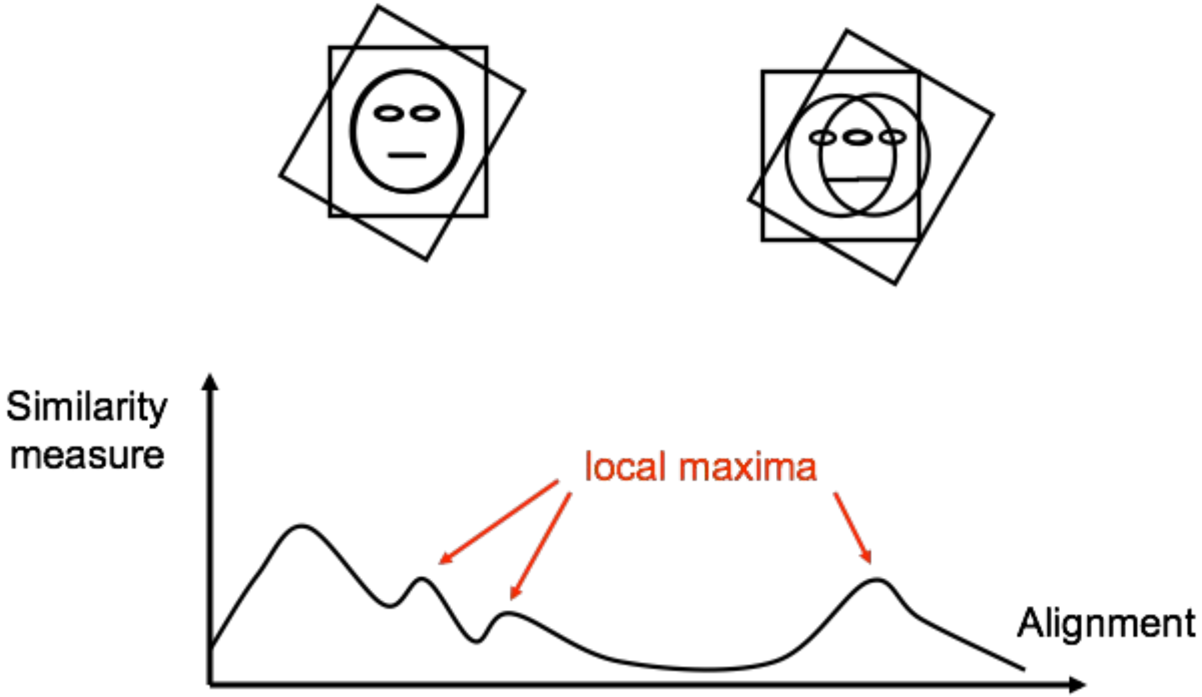
There can be a very high similarity if only the background of the two images overlap

We need to initialize the registration within the capture range
(see after for some solutions)

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

(Dis)similarity measures

Local Optima

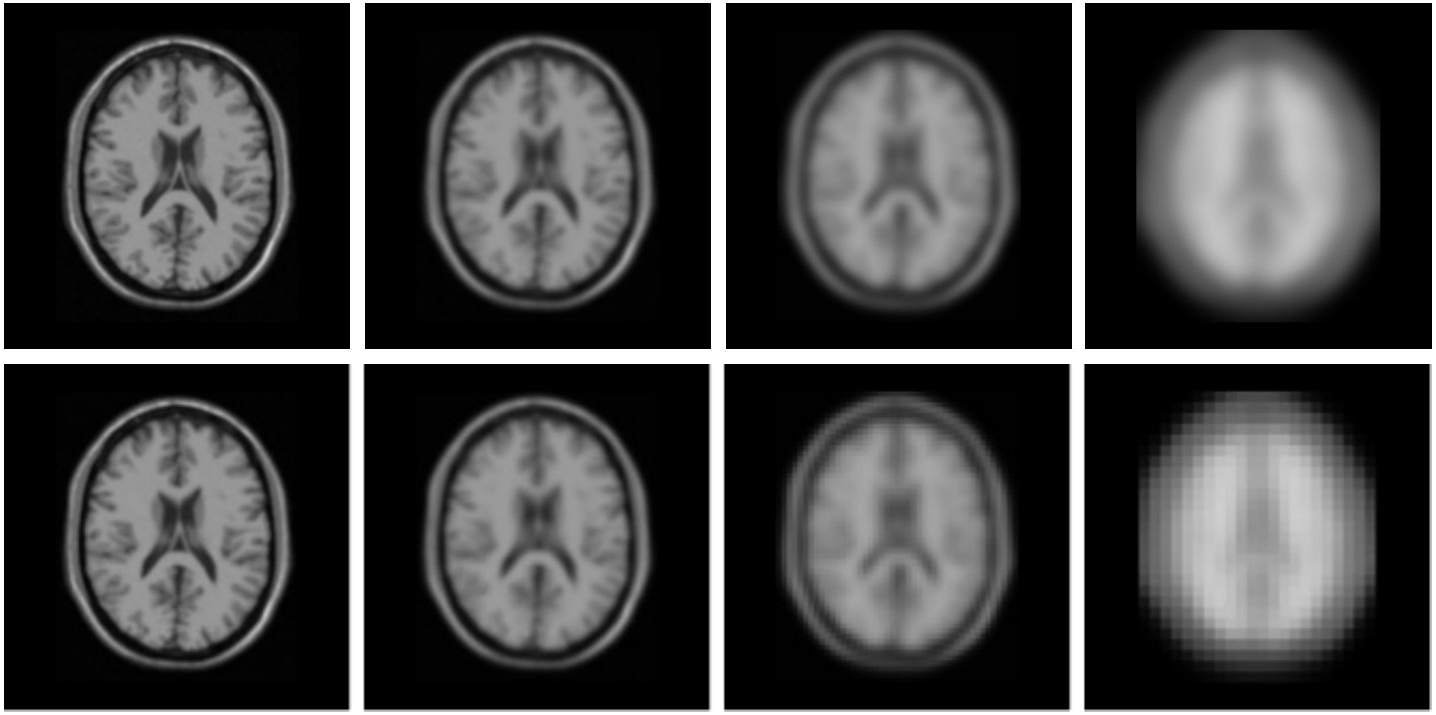


Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

(Dis)similarity measures

Multi-scale, hierarchical Registration

- Successively increase degrees of freedom
- Gaussian image pyramids

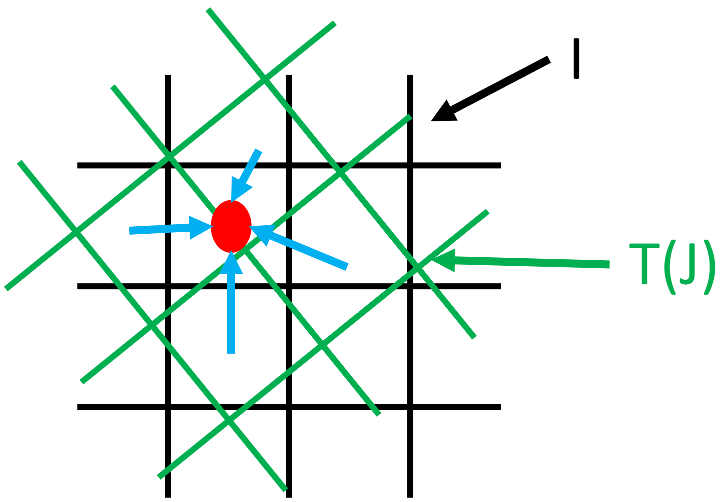


Part 8 – Registration

8.5 Other technical aspects

Interpolation

$$D_{SSD}(I, T(J)) = \frac{1}{N}\sum_{i=1}^{N}(I(x_i) - T(J(x_i)))^2$$



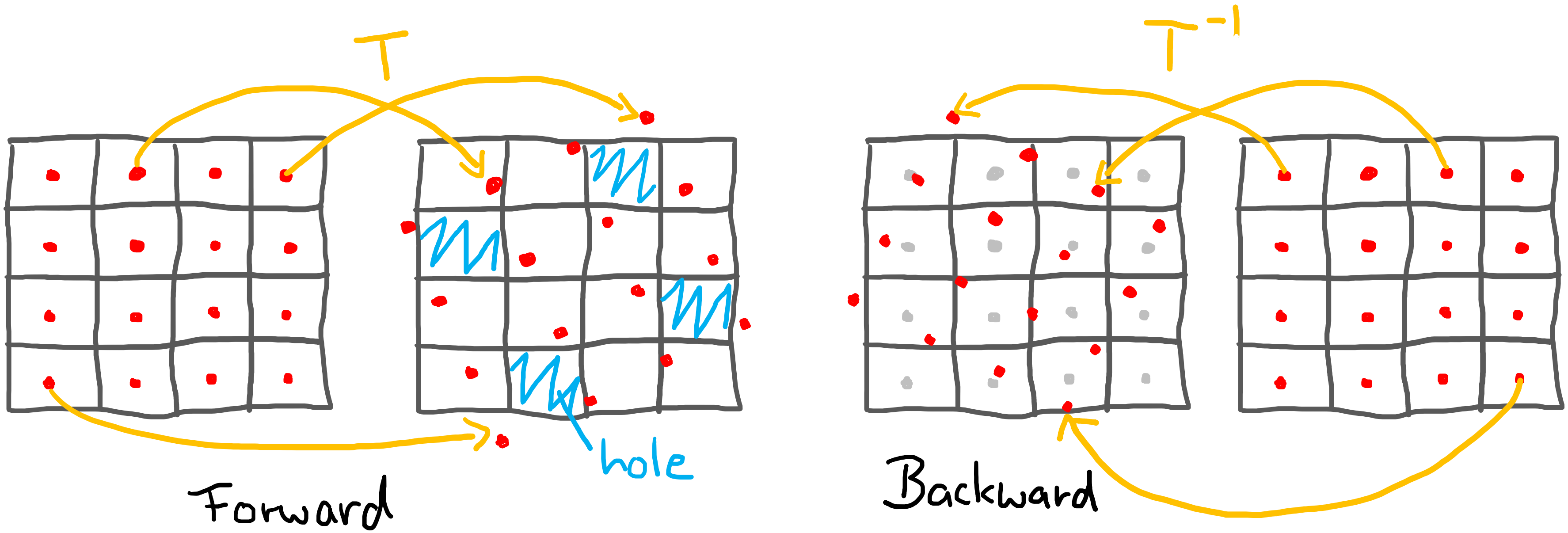
Resampling

Forward warping

- Cycle through original image, and transport intensities forward
- Holes can occur!

Backward warping

- Cycle through the new image grid, and do backward look up in original imag



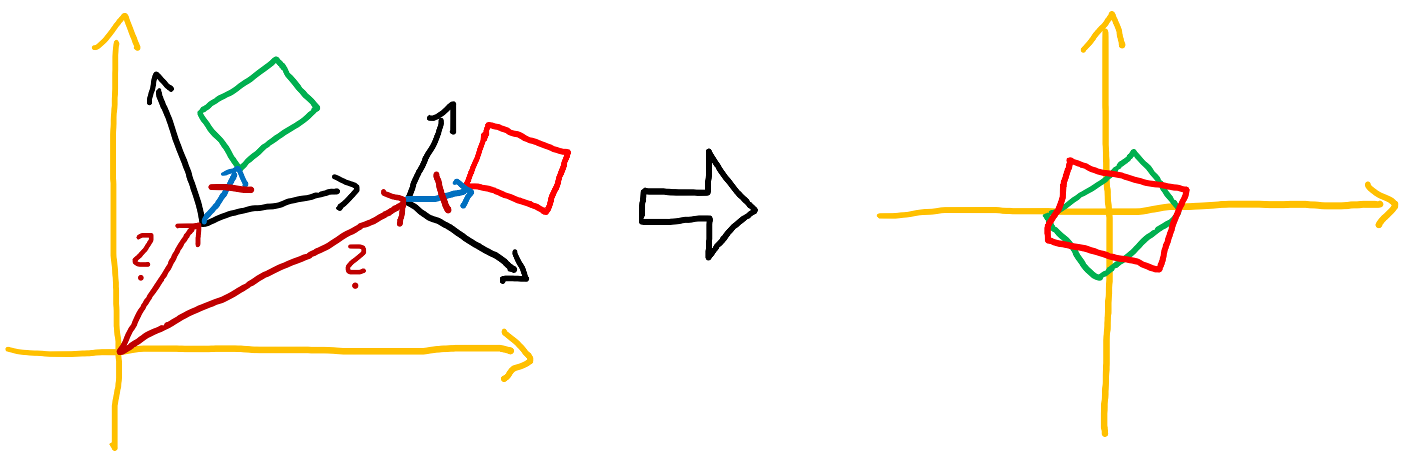
Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Initialisation

Initialisation is critical for intensity-based registration

Two common heuristics to deal with initialisation:

1) Ignore image specific origin information, align image centres



Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Initialisation

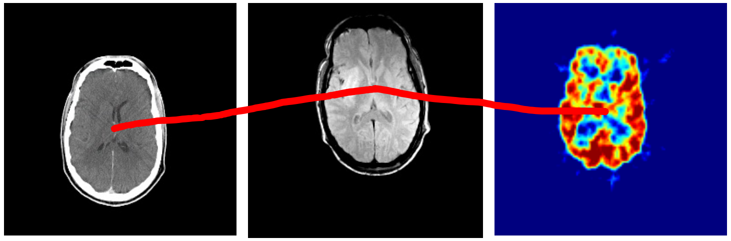
Initialisation is critical for intensity-based registration

Two common heuristics to deal with initialisation:

2) Align the centres of intensity masses

$$c_I = \frac{1}{Z} \sum_{i=1}^{N} (I(x_i) - \min(I)) \cdot T_{ItW} x_i$$

with $Z = \sum_{i=1}^{N} (I(x_i) - \min(I))$

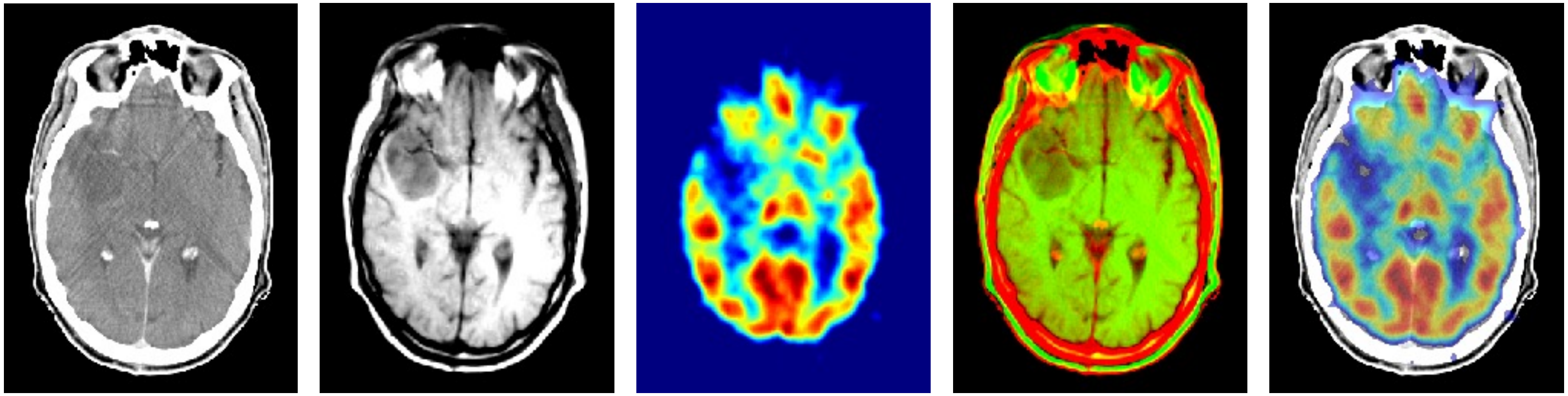


Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Part 8 – Registration

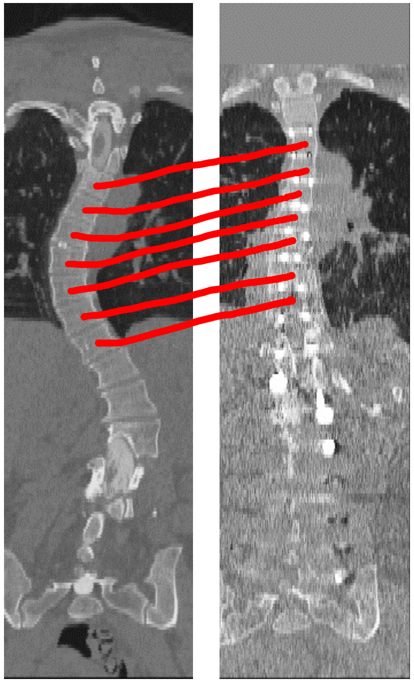
8.6 Applications

Multimodal image fusion



Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

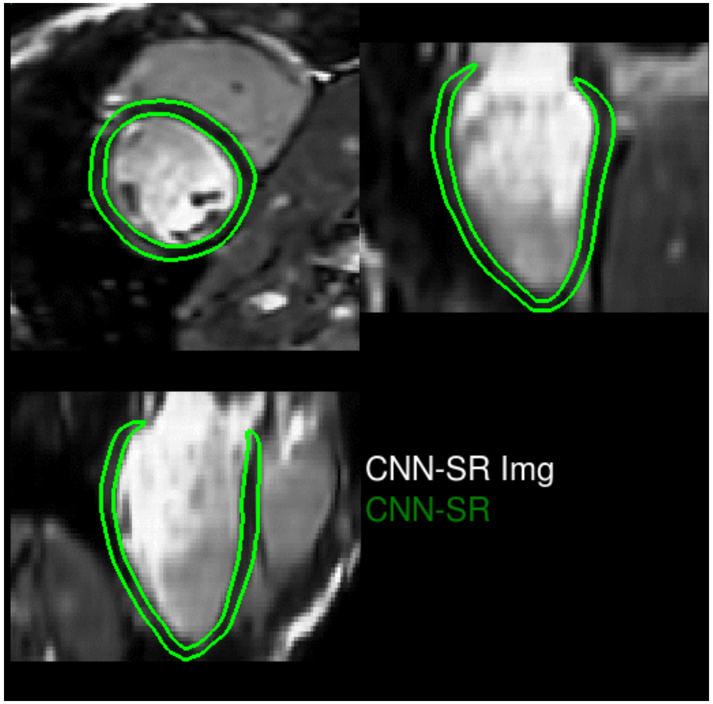
Pre and post-op (surgical operation)



Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Motion

Cardiac Motion

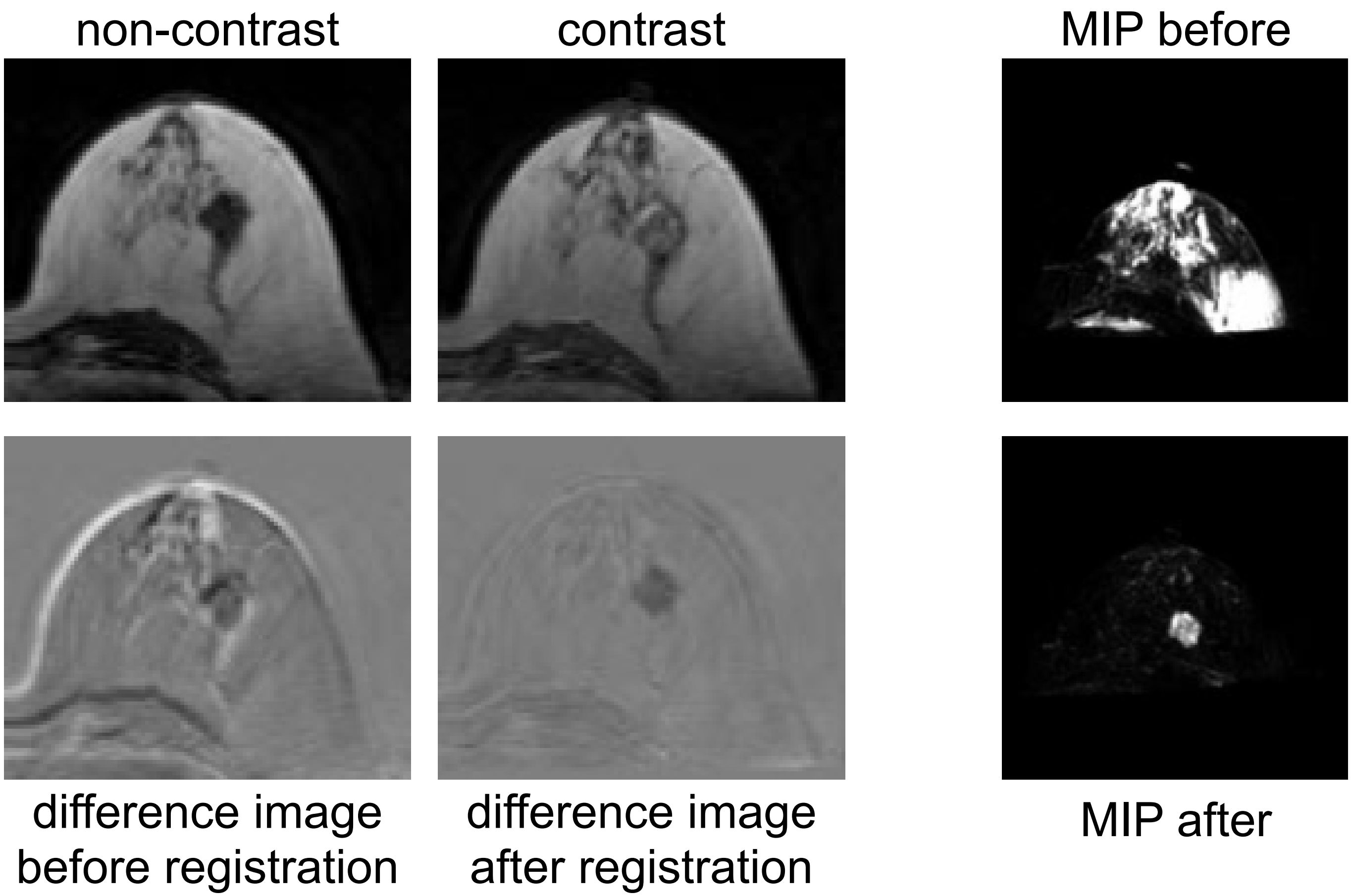


Source: Oktay et al. MICCAI 2016

Respiratory Motion

Contrast and non-contrast images

Contrasted and non-contrasted images [Rueckert et al. 1999]

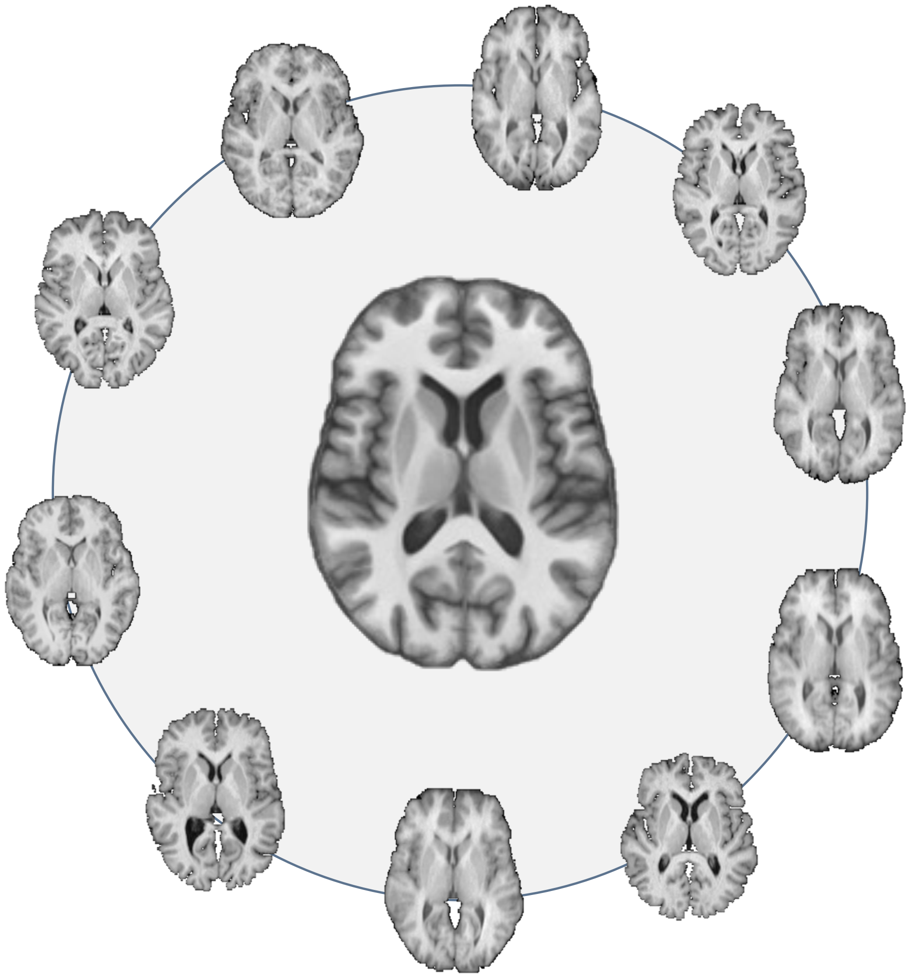


Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Atlas construction

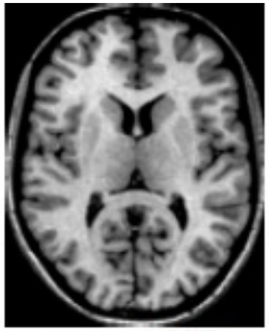
Atlas construction

- Builds an “average” of a group of participants
- This avoids chosing an arbitrary patient as a target when registering N (N>2) patients together
- The “average” is called an atlas or template
- Two options when doing inter-subject registration (with N>2)
  - Build an atlas from the group of patients
  - Use a predefined atlas (built from another population)



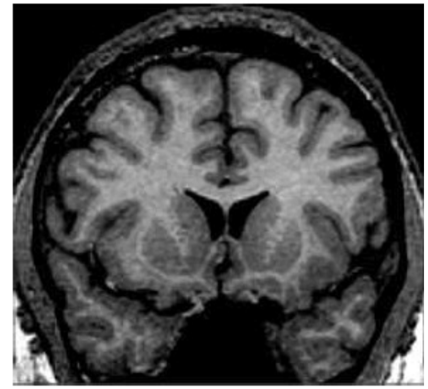
Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Segmentation based on registration

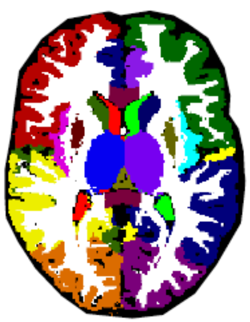


Reference MRI

Registration: compute transformation ϕ

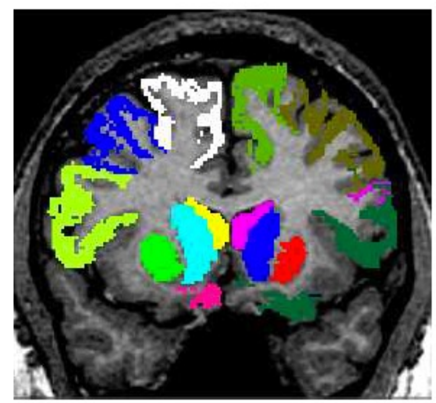


MRI to segment



Reference segmentation

Apply transformation ϕ

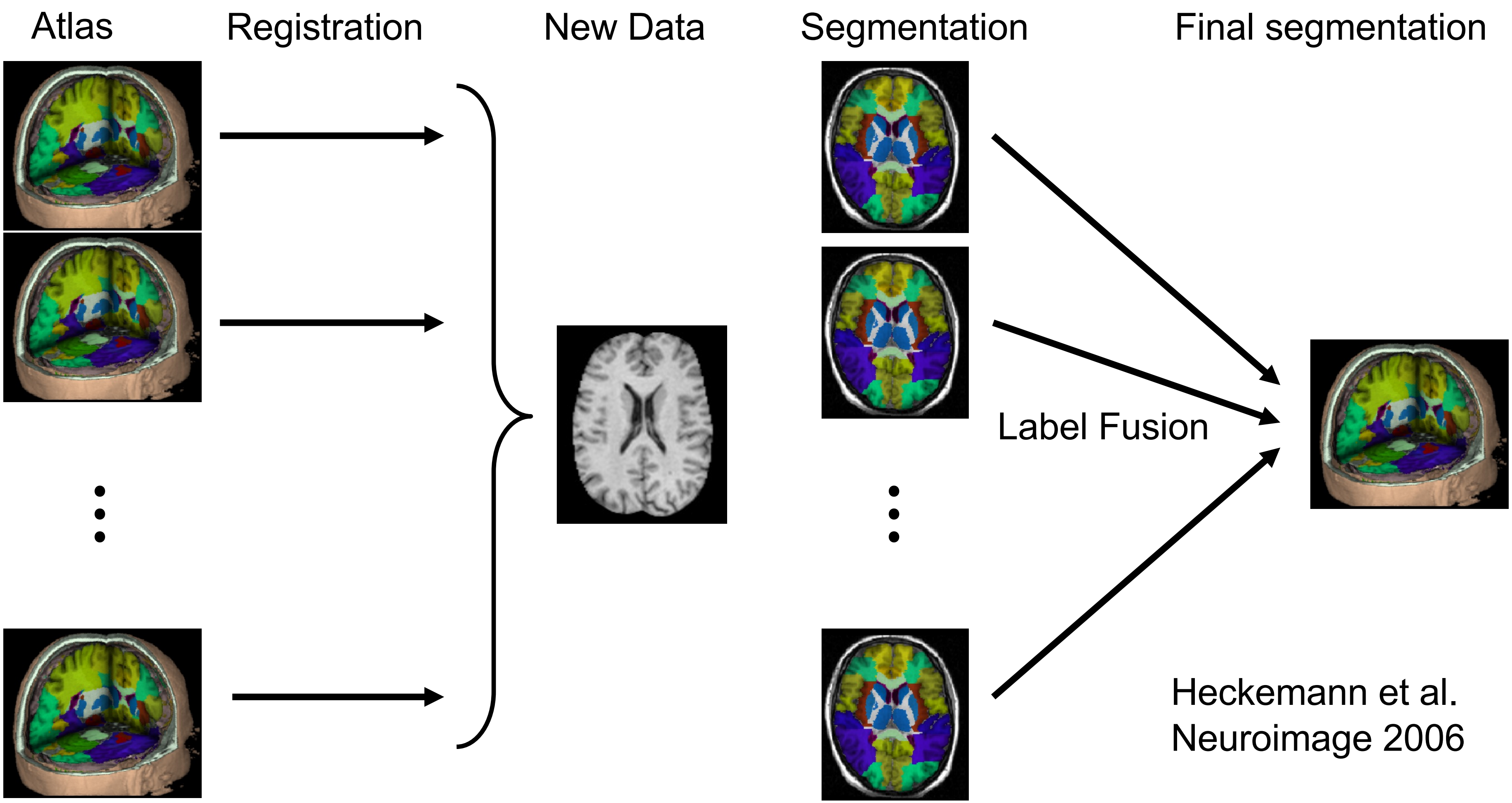


Obtained segmentation

Collins et al, 1999

With a single reference

Segmentation based on registration



With a multiple references (better handling of anatomical variability)

Source: B. Glocker/D. Rueckert – Course on Machine Learning for Imaging – Imperial College London

Segmentation based on registration

- Multi-atlas registration-based segmentation
  - Was the state of the art before deep learning
  - Is still a competitive method, thanks to its precision and robustness

Part 8 – Registration

8.7 Some classical registration softwares

Software

- Often, you may need to register your data prior to doing machine learning
- For instance, **it is quite common to perform a linear registration before applying deep learning algorithms**
  - This often helps the training
- For this, there are some robust and well-validated freely available software developped by the community

Software

- ANTs - http://stnava.github.io/ANTs/
  - Provides different types of deformation models and similarity metrics
    - Linear transformation
    - Non-linear transformation using B-splines
    - Non-linear transformation using diffeomorphisms

- FSL – https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL
  - Very commonly used for linear registration
  - FIRST – linear registration tool
    - https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLIRT