

## **MEDICAL IMAGE COMPUTING (CAP 5937)**

**LECTURE 15:** Medical Image Registration I (Introduction)

**Dr. Ulas Bagci**

HEC 221, Center for Research in Computer  
Vision (CRCV), University of Central Florida  
(UCF), Orlando, FL 32814.

[bagci@ucf.edu](mailto:bagci@ucf.edu) or [bagci@crcv.ucf.edu](mailto:bagci@crcv.ucf.edu)

# Outline

- Motivation
- Registration basics
- Rigid registration
- Non-rigid registration
- Example Applications



# Modalities in Medical Imaging

- **Mono-modality:**

- **Mono-modality:**
  - ✓ A series of same modality images (CT/CT, MR/MR, Mammogram pairs,...).
  - ✓ Images may be acquired weeks or months apart; taken from different viewpoints.
  - ✓ Aligning images in order to detect subtle changes in intensity or shape

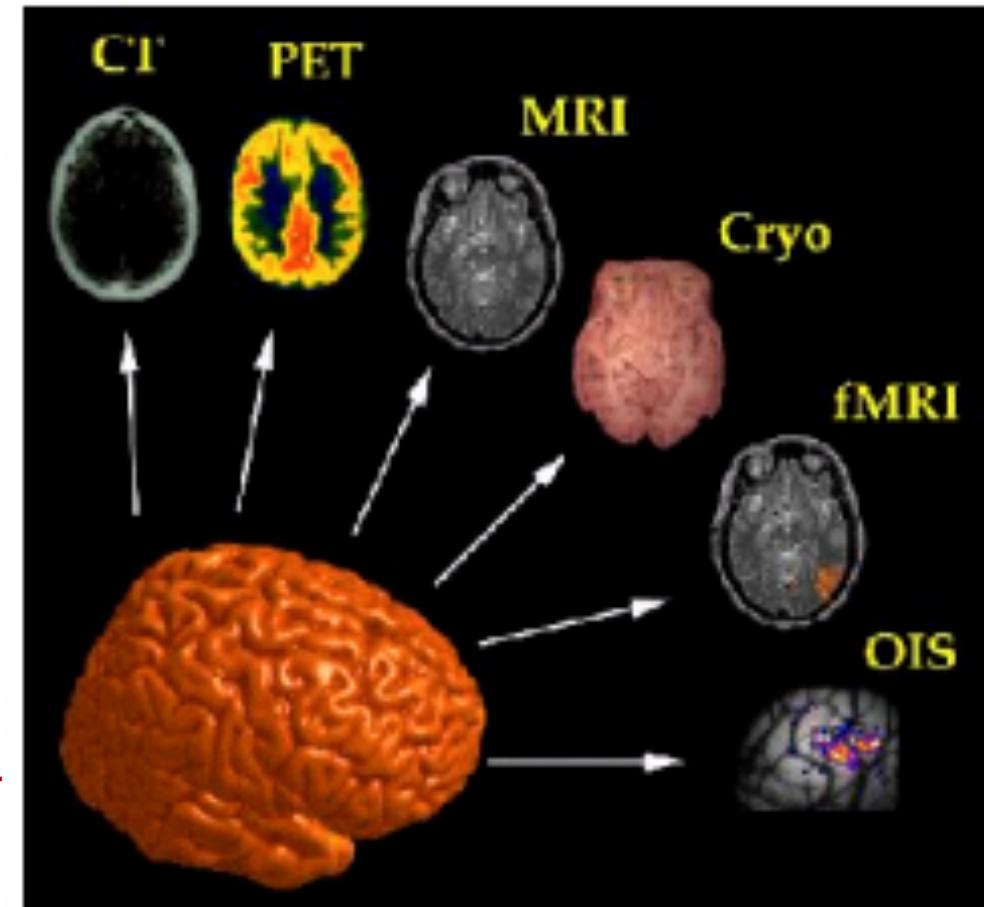
- **Multi-modality:**

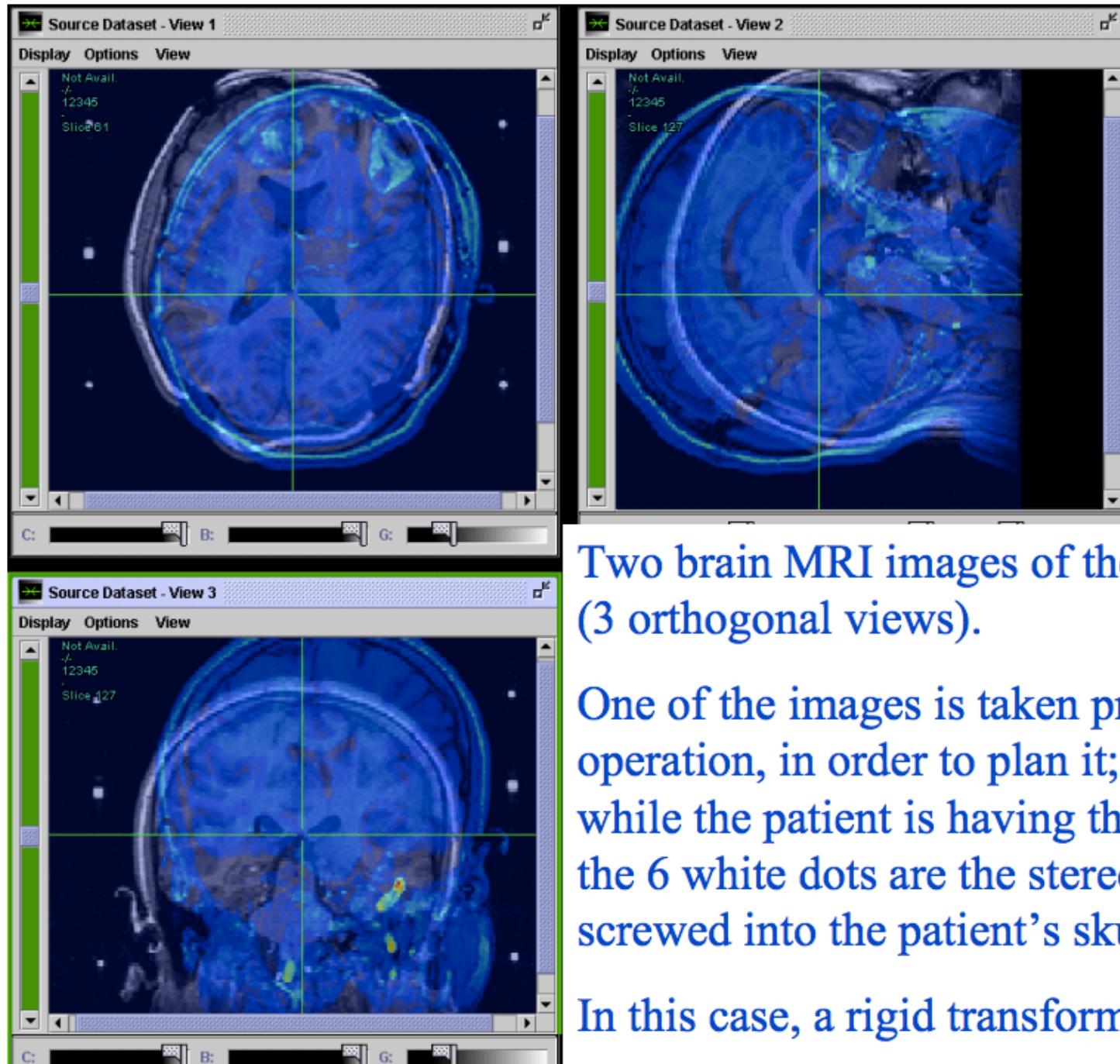
- **Multi-modality:**
  - ✓ Complementary anatomic and functional information from multiple modalities can be obtained for the precise diagnosis and treatment.
  - ✓ Examples: PET and SPECT (low resolution, functional information) need MR or CT (high resolution, anatomical information) to get structure information.



## In other words,...

- CT, MRI, PET and Ultrasound produce 3D volume images
- Typically  $256 \times 256 \times 256 = 16,777,216$  image **voxels**.
- Combining modalities (**inter modality**) gives extra information.
- Repeated imaging over time same modality, e.g. MRI, (**intra modality**) equally important.
- Have to spatially register the images.





## Before Registration

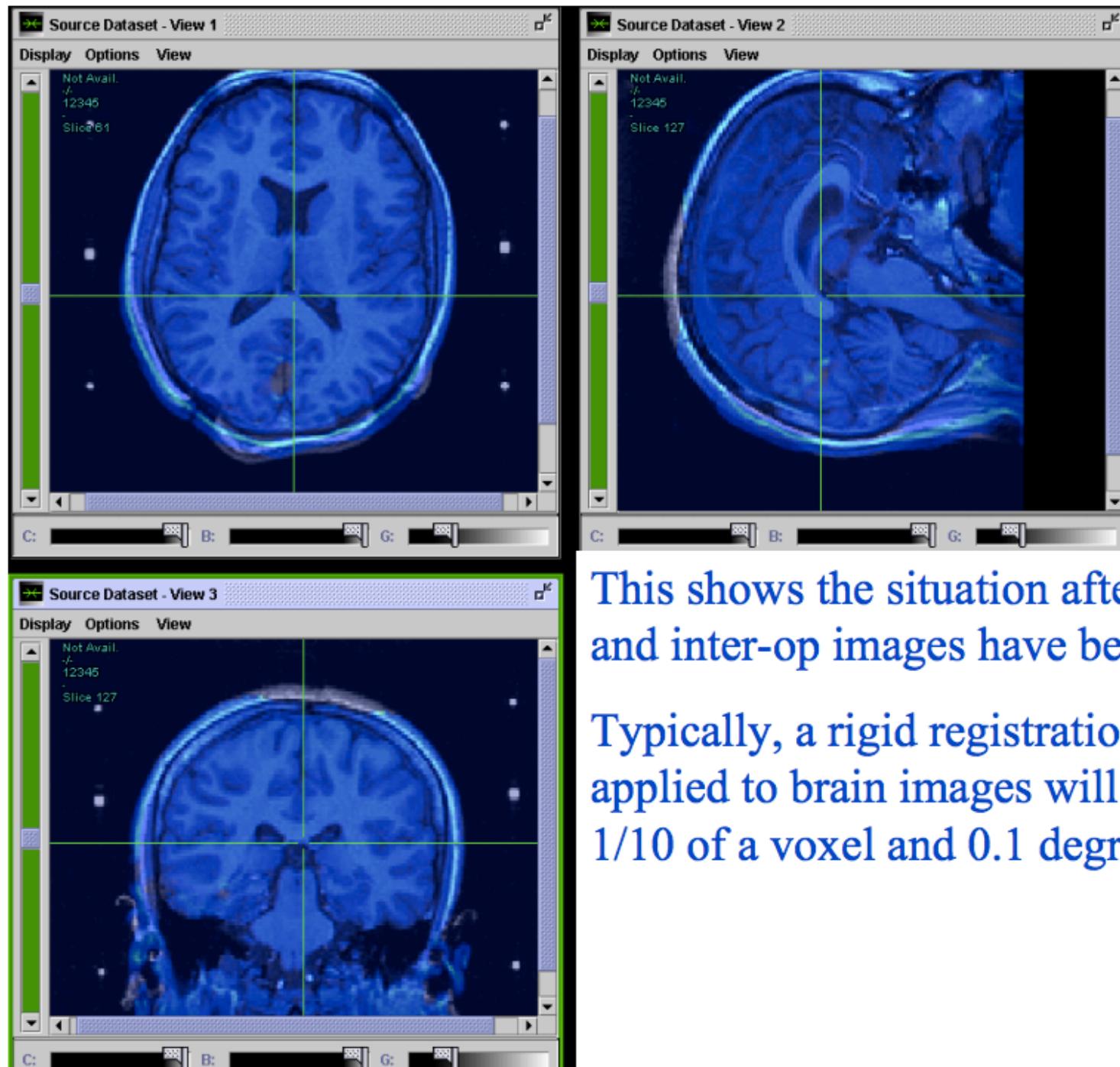
Two brain MRI images of the same patient (3 orthogonal views).

One of the images is taken prior to the operation, in order to plan it; the second while the patient is having the operation: the 6 white dots are the stereotactic frame screwed into the patient's skull.

In this case, a rigid transform suffices



## After Registration

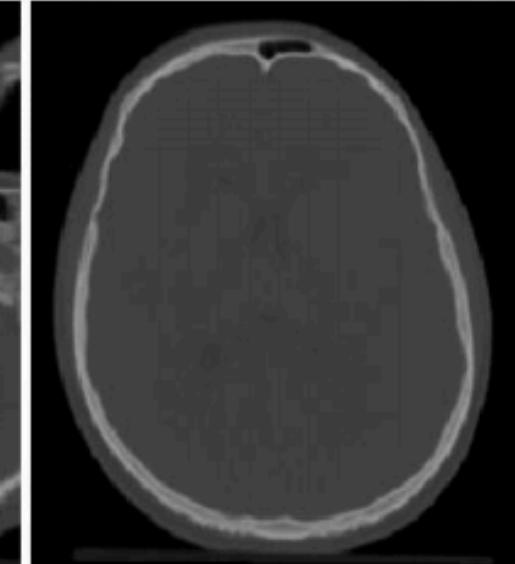
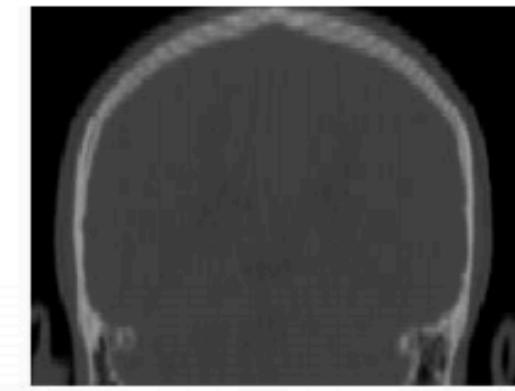
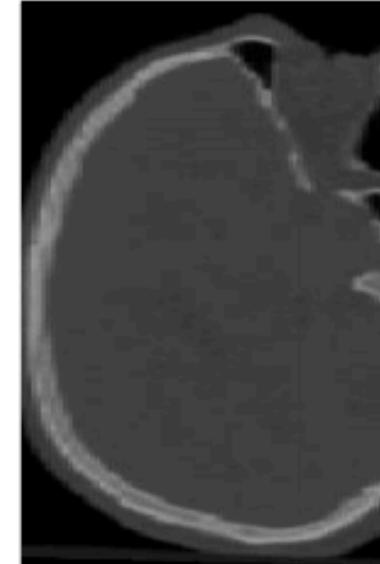
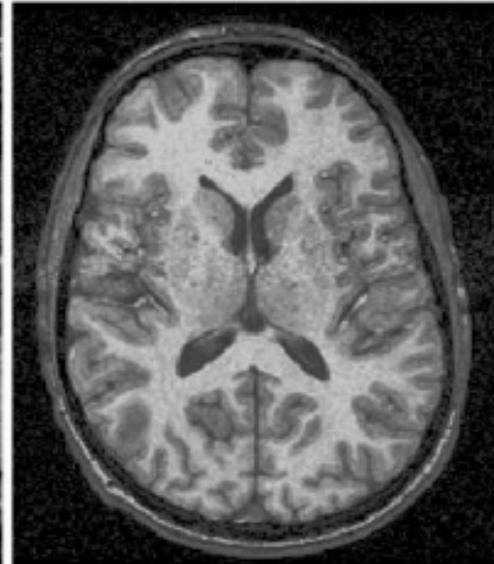
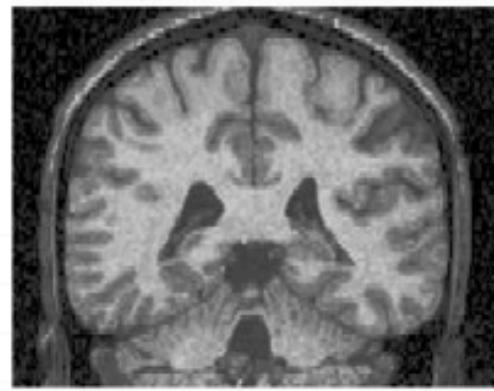
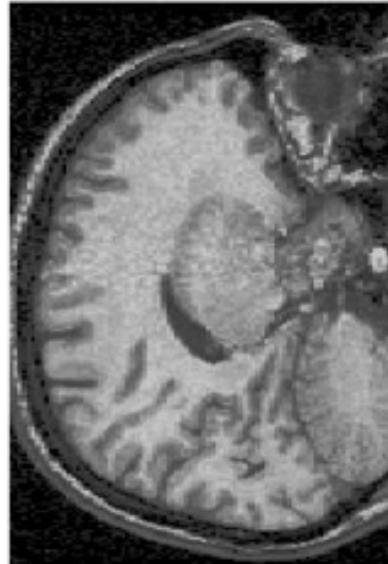


This shows the situation after the pre-op and inter-op images have been aligned.

Typically, a rigid registration algorithm applied to brain images will be accurate to 1/10 of a voxel and 0.1 degrees of rotation



# Example: rigid CT/MR registration





# Multiple Fusion Algorithms

- **Rigid fusion** – no compensation for motion or patient position
- **Deformable fusion** – crucial when structures have changed position or shape between or during scans due to voluntary or physiological motion or imperfect scanning protocols

Rigid fusion (fig 1) can be ambiguous- the active growth identified on PET might be either one of two CT lesions. However, deformable fusion (fig 2) identifies the PET activity with the anterior lesion on CT



Figure 1



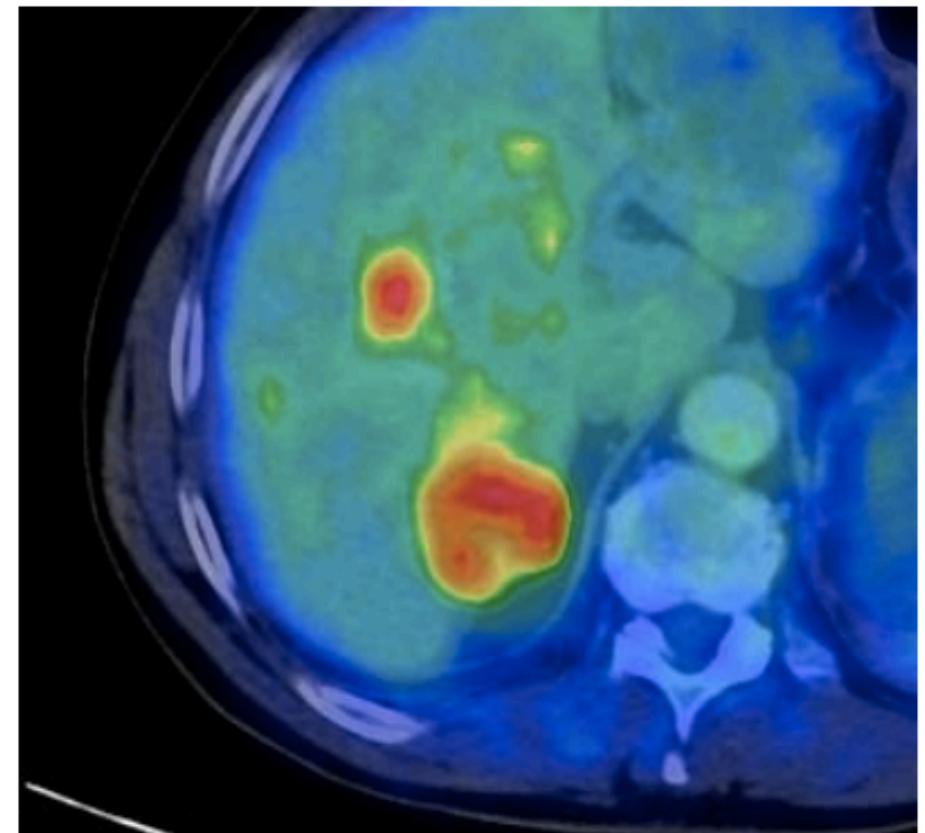
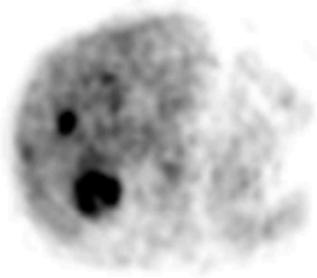
Figure 2

## Fusion of information = registration plus combination in a single representation: PET/CT

CT



PET



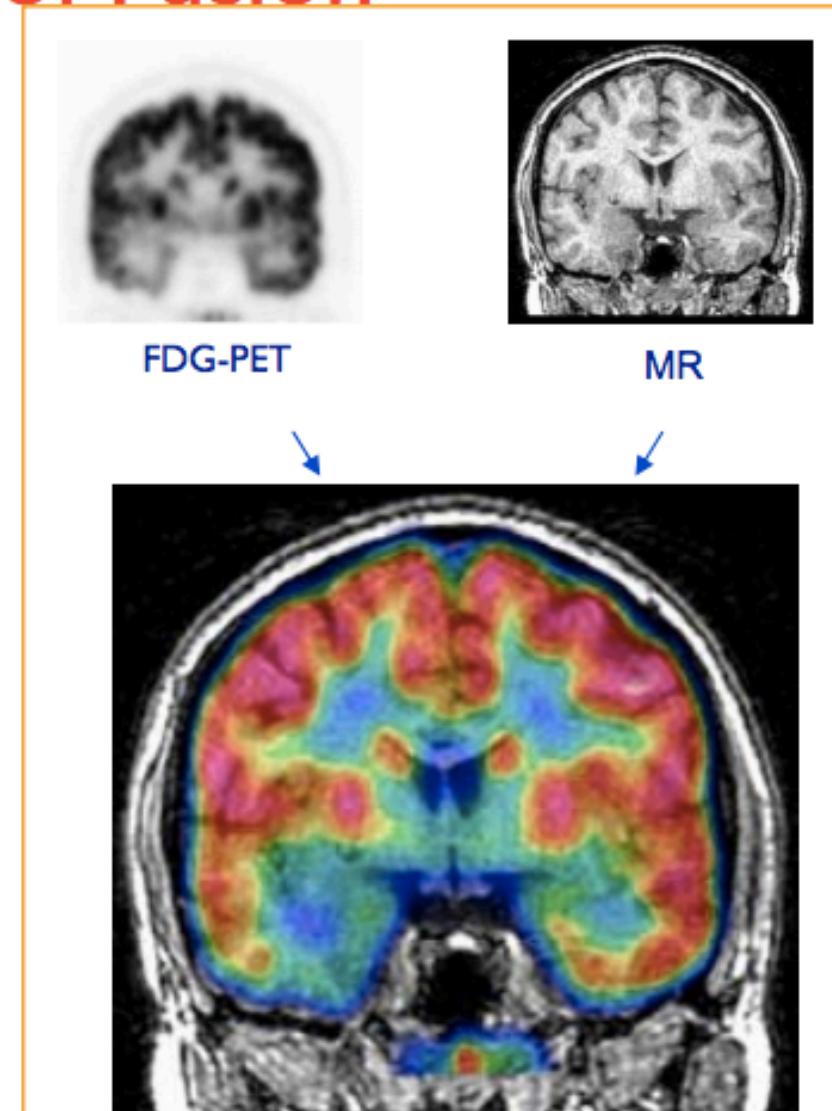
Deformable fusion- PET shows increased metabolism in lesions identified on CT, consistent with active tumour growth rather than necrosis post-radiotherapy



# Many Clinical Applications of Fusion

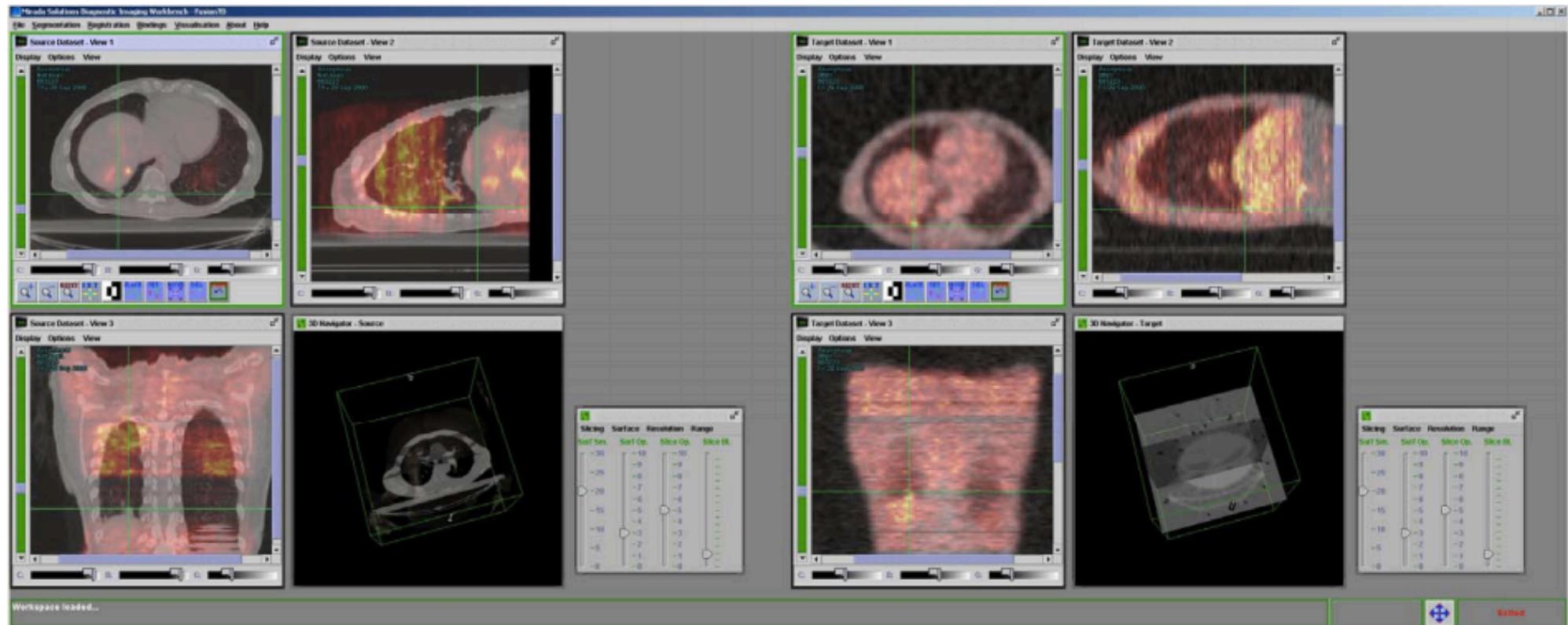
- Cancer staging
- Biopsy planning
- Radiotherapy treatment planning
- Quantitative assessment of treatment response
- Pre-surgical assessment of other conditions e.g. epilepsy
- As an effective communication tool when reporting to clinical meetings, referring physicians or to patients
- Whenever multiple data sources may be better assessed together

*PET data identifies a region of hypometabolism due to epilepsy. Fusion with MR localises the damage to the anterior and medial areas of the right temporal gyrus*





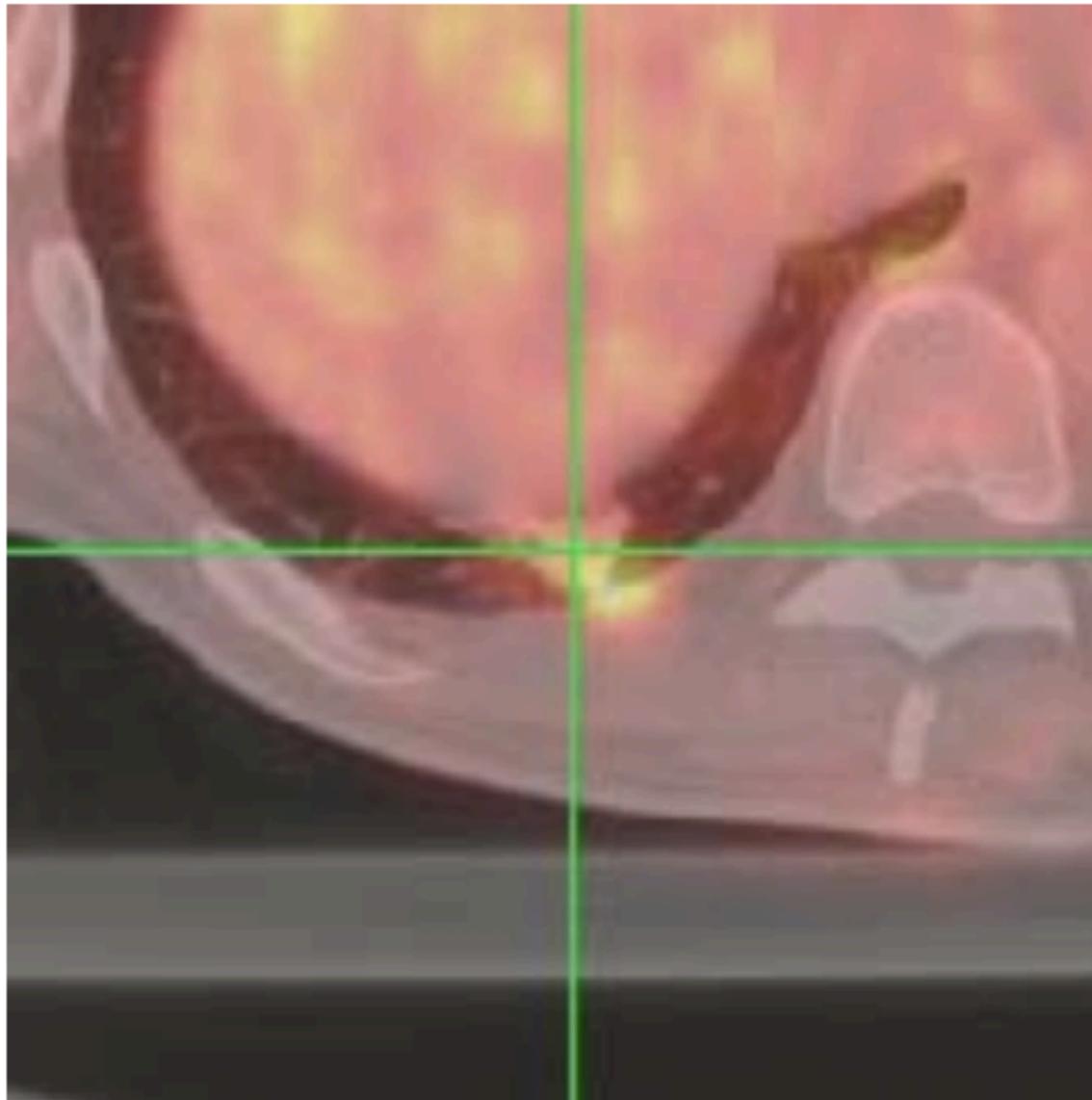
# CT – PET registration



Non-rigid registration is necessary



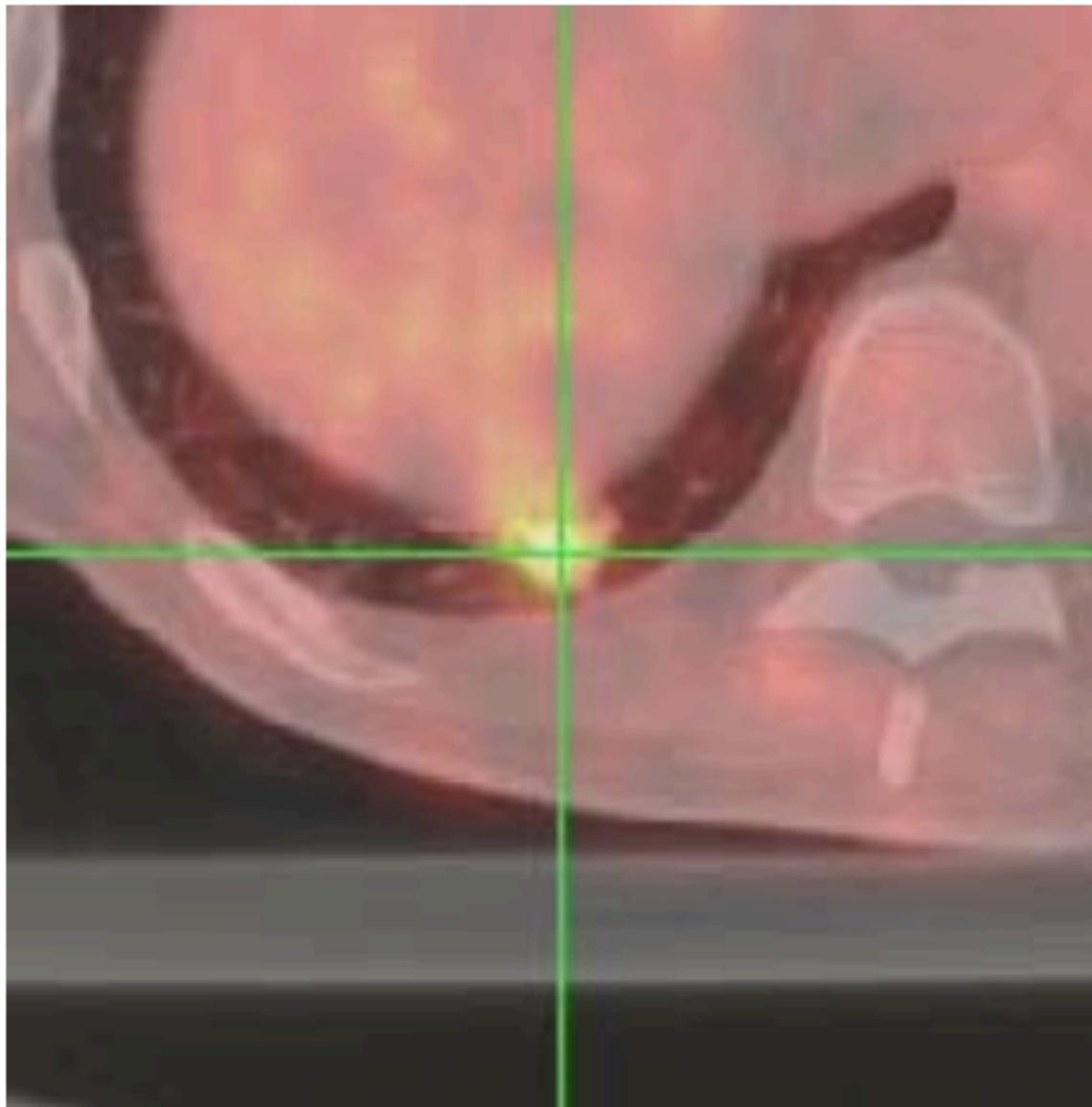
# Rigid registration poor



Is the tumour in  
the lungs or the  
stomach?



# Non-rigid registration



Looks plausible;  
but how could you  
be sure?

Are you prepared  
to risk your  
software against  
getting sued?

# Examples of image registration 2

## images of a single individual

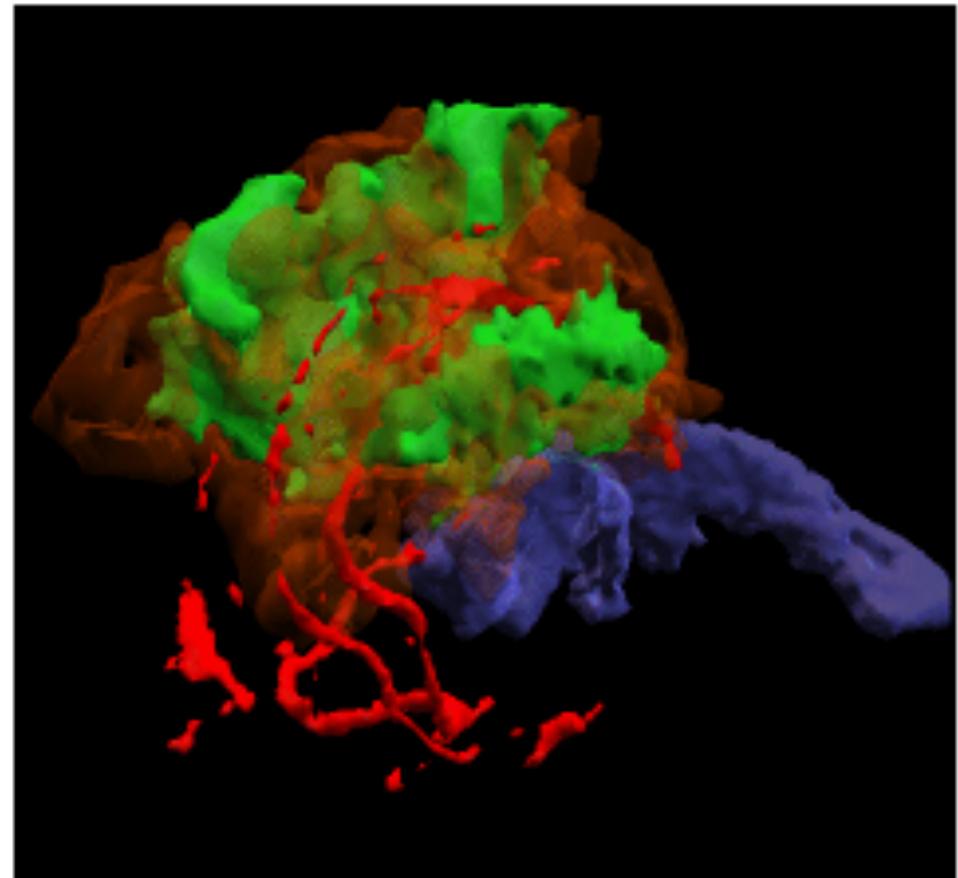
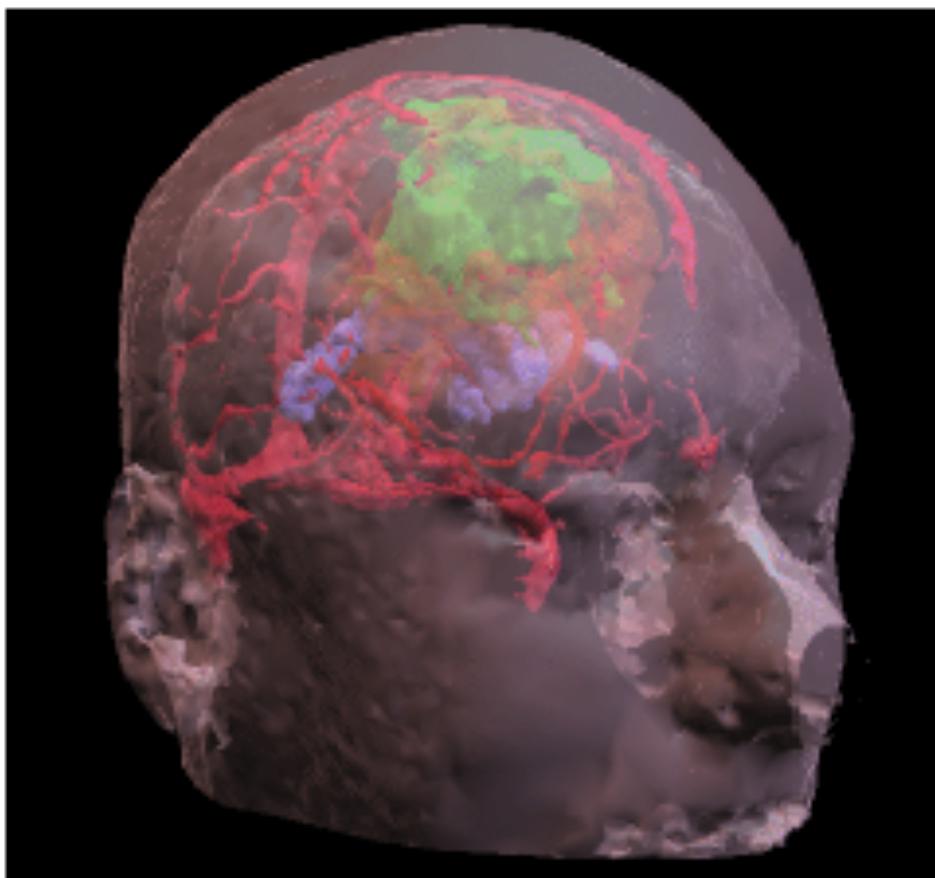
- Aligning the images from two different patients;
- Aligning the images of a subject to an atlas, or, constructing such an atlas from the images of several subjects;
- Aligning the images of patients and aligning those of normals to develop a statistical model of variation associated with a disease;
- Aligning the images from many thousands of subjects around the world as part of a clinical/drug trial

# Summary of Applications

- **Diagnosis**
  - Combining information from multiple imaging modalities
- **Studying disease progression**
  - Monitoring changes in size, shape, position or image intensity over time
- **Image guided surgery or radiotherapy**
  - Relating pre-operative images and surgical plans to the physical reality of the patient
- **Patient comparison or atlas construction**
  - Relating one individual's anatomy to a standardized atlas



# Surgery/Therapy Planning



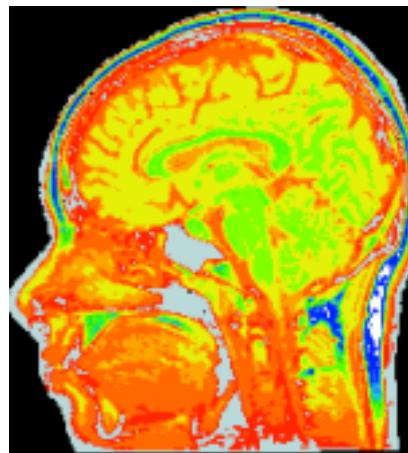


# What is Image Registration then?

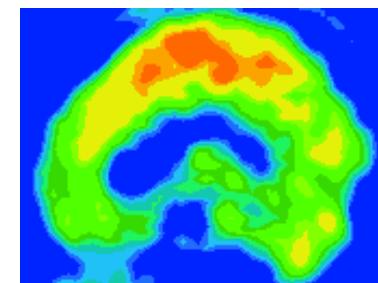
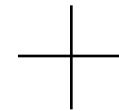


# Image Registration is a

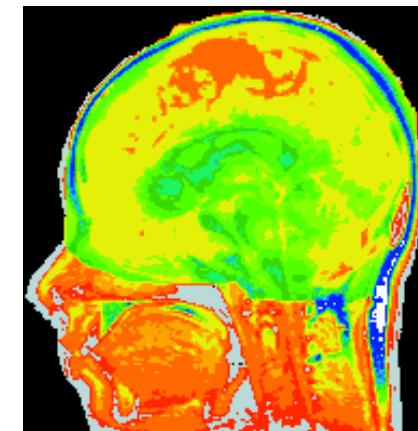
- **Spatial transform** that maps points from one image to corresponding points in another image
  - matching two images so that corresponding coordinate points in the two images correspond to the same physical region of the scene being imaged also referred to as image fusion, superimposition, matching or merge*



MR



SPECT



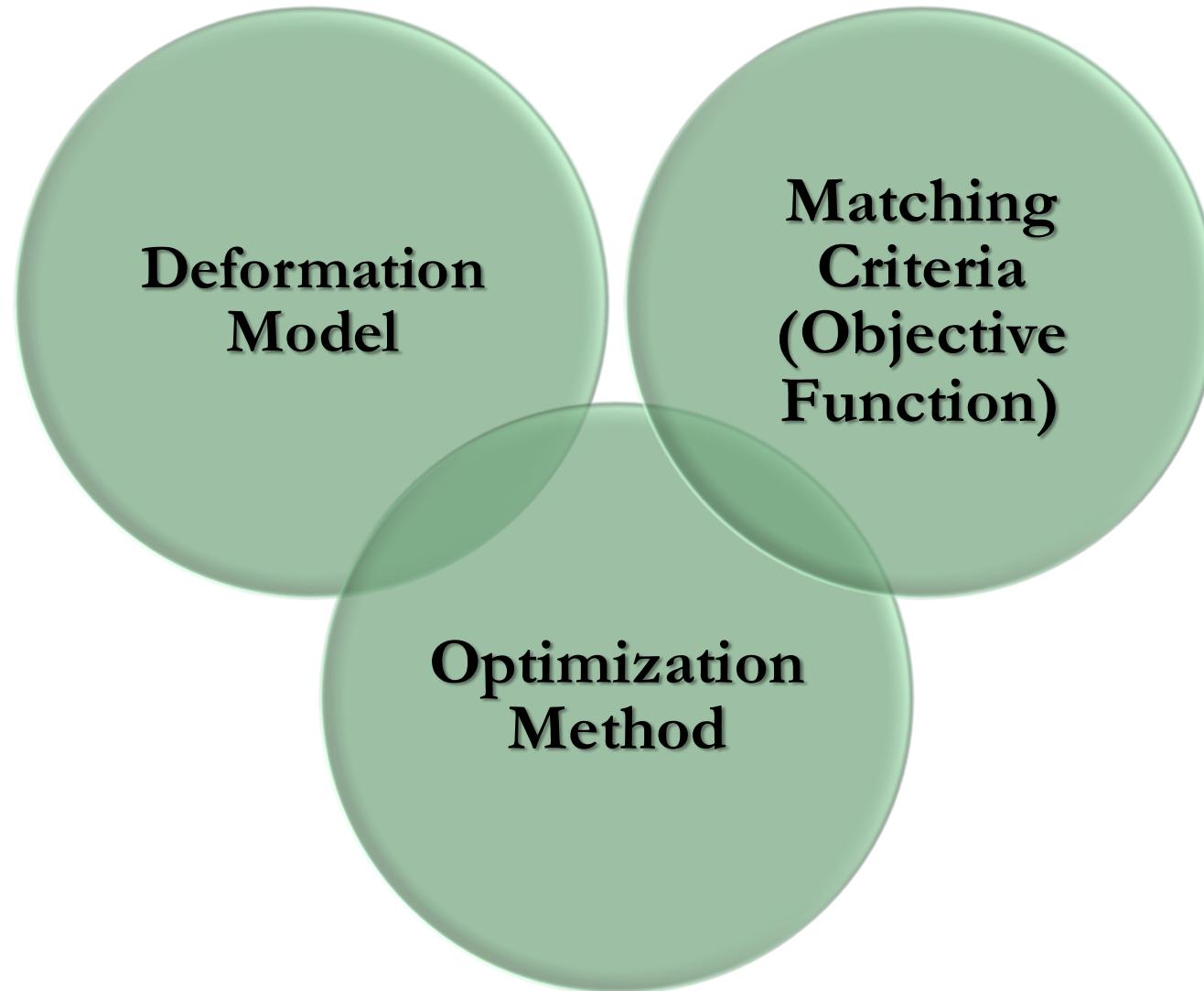
registered

# Image Registration is a

- **Spatial transform** that maps points from one image to corresponding points in another image
  - Rigid
    - Rotations and translations
  - Affine
    - Also, skew and scaling
  - Deformable
    - Free-form mapping



# Registration Framework



# Registration Taxonomy

- **Dimensionality**
  - 2D-2D, 3D-3D, 2D-3D
- **Nature of registration basis**
  - Image based
    - Extrinsic, Intrinsic
  - Non-image based
- **Nature of the transformation**
  - Rigid, Affine, Projective, Curved
- **Interaction**
  - Interactive, Semi-automatic, Automatic
- **Modalities involved**
  - Mono-modal, Multi-modal, Modality to model
- **Subject:**
  - Intrasubject
  - Intersubject
  - Atlas
- **Domain of transformation**
  - Local, global
- **Optimization procedure**
- **Object**

# Deformation Models

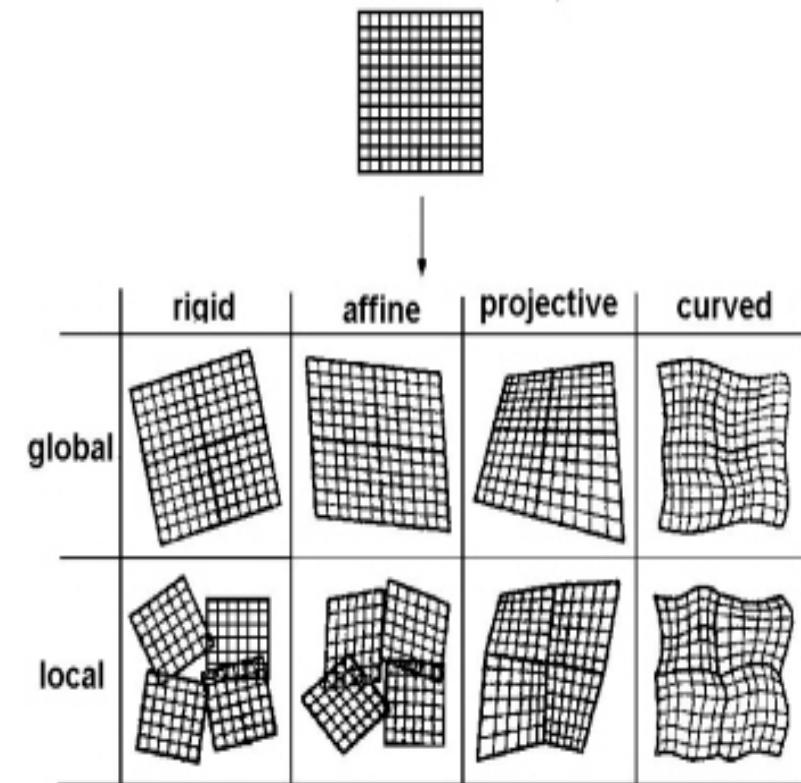
Method used to find the transformation

- **Rigid & affine**

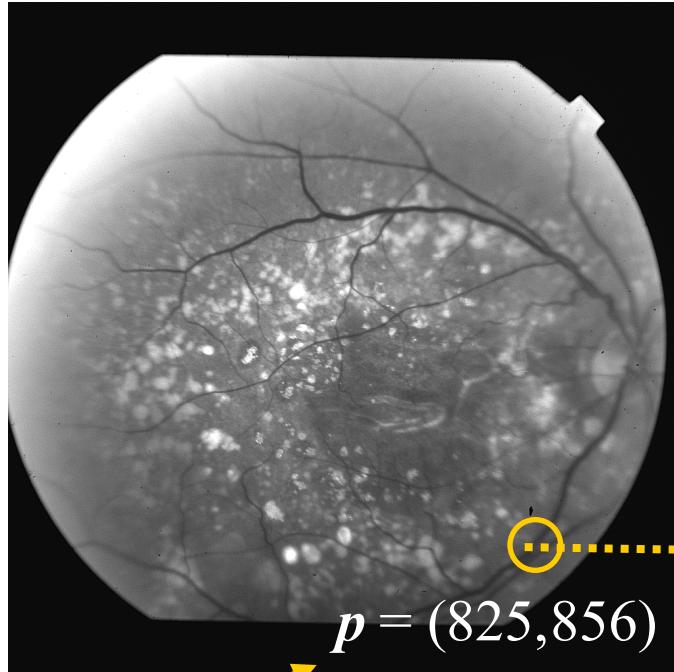
- Landmark based
- Edge based
- Voxel intensity based
- Information theory based

- **Non-rigid**

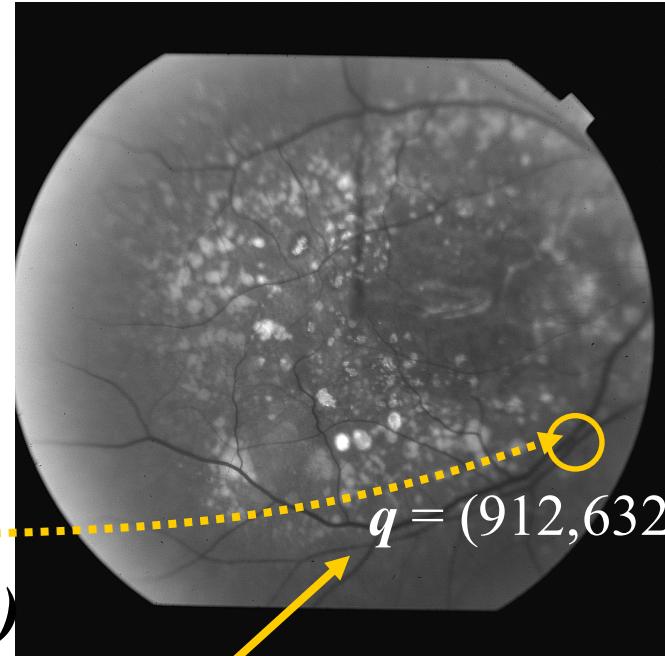
- Registration using basis functions
- Registration using splines
- Physics based
  - Elastic, Fluid, Optical flow, etc.



# Registration is an alignment problem



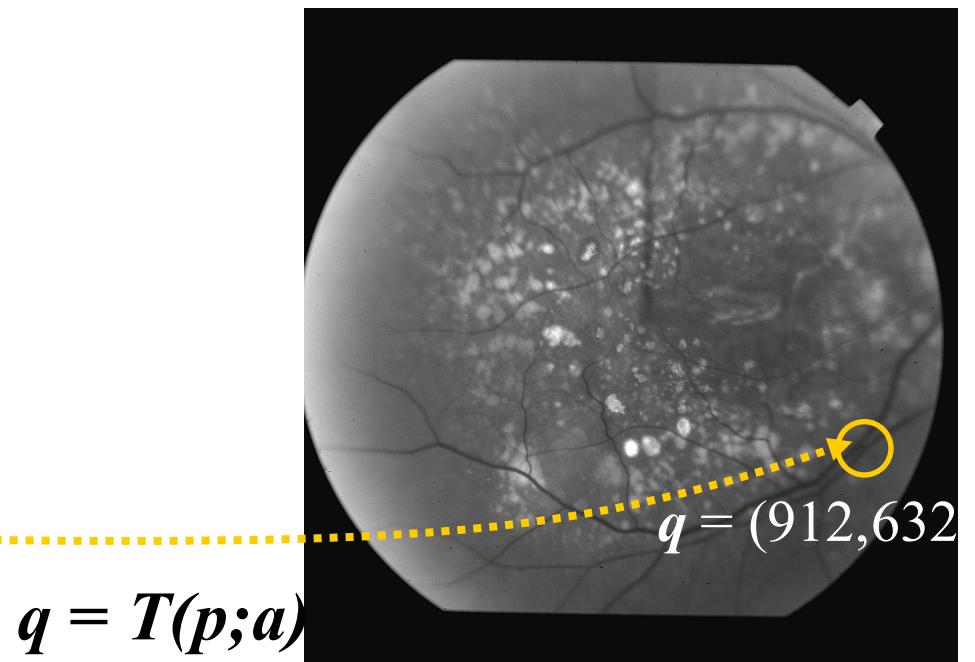
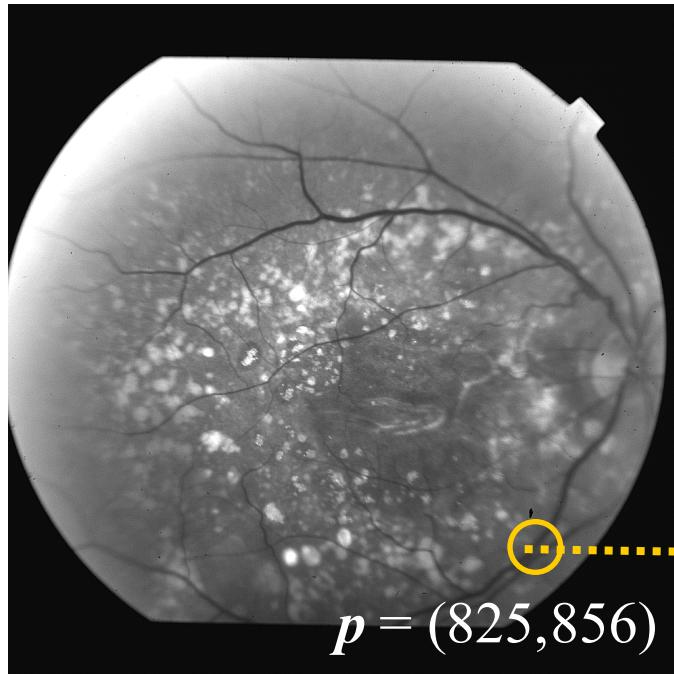
Pixel location in first image



Homologous pixel location in  
second image

Pixel location mapping function

# Registration is an alignment problem



$$\mathbf{p} = (x, y)^T$$

$$\boldsymbol{\Theta} = (s, t_x, t_y)^T$$

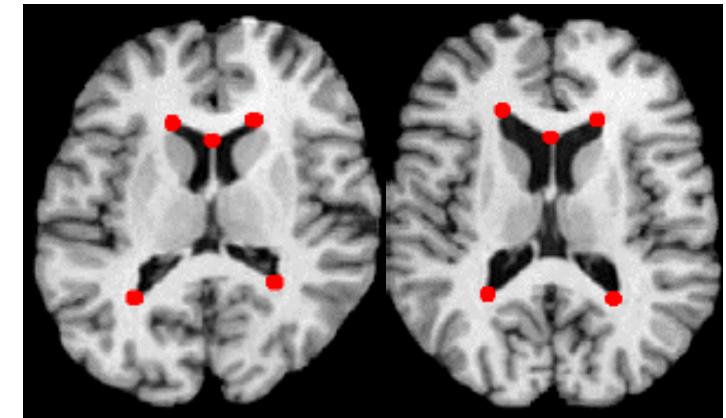
$$T(\mathbf{p}; \boldsymbol{\Theta}) = \begin{pmatrix} sx + t_x \\ sy + t_y \end{pmatrix}$$

Pixel scaling and  
translation



# Landmark Based

- Identifying corresponding points in the images and inferring the image transformation
- Types of landmarks
  - Extrinsic
    - artificial objects attached to the patient
  - Intrinsic
    - internal anatomical structures
- Computing the average or “centroid” of each set of points → translation
- Rotated this point set about the new centroid until the sum of the squared distances between each corresponding point pair is minimized

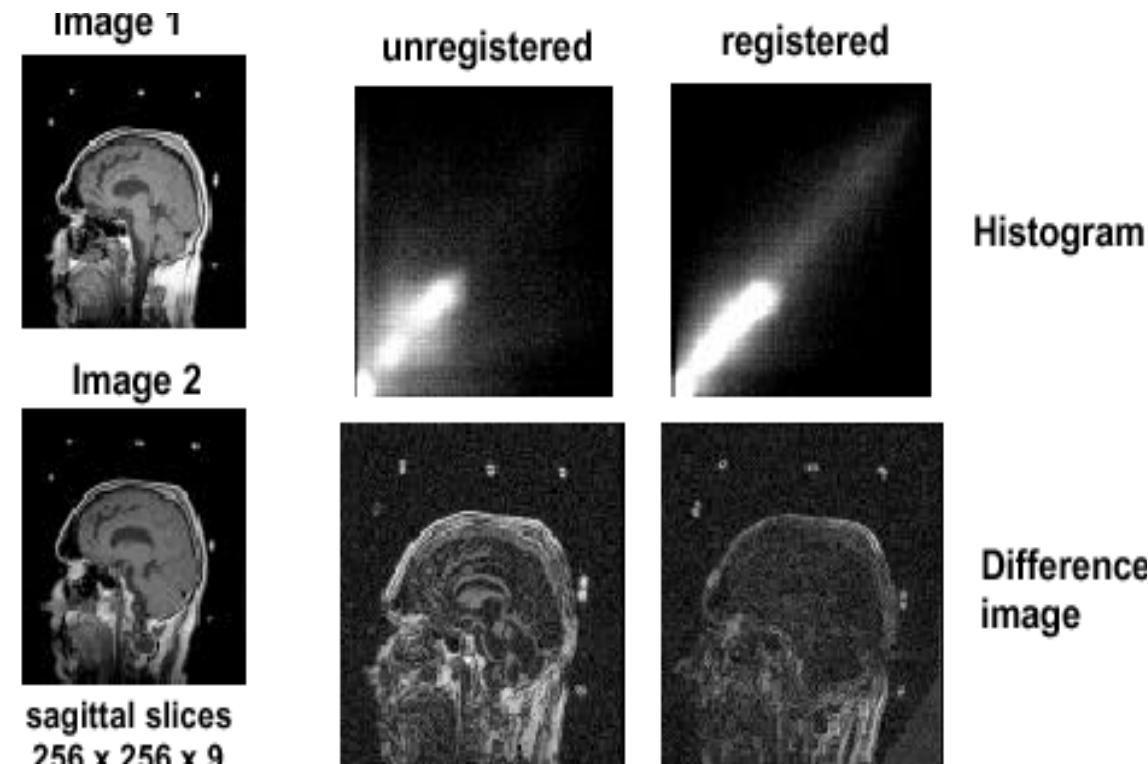


# Surface Based

- Method
  - Extracting corresponding surfaces
  - Computing the transformation by minimizing some measure of distance between the two surfaces
- Algorithms used
  - The “Head and Hat” Algorithm
  - The Iterative Closest Point Algorithm
  - *Registration using crest lines*

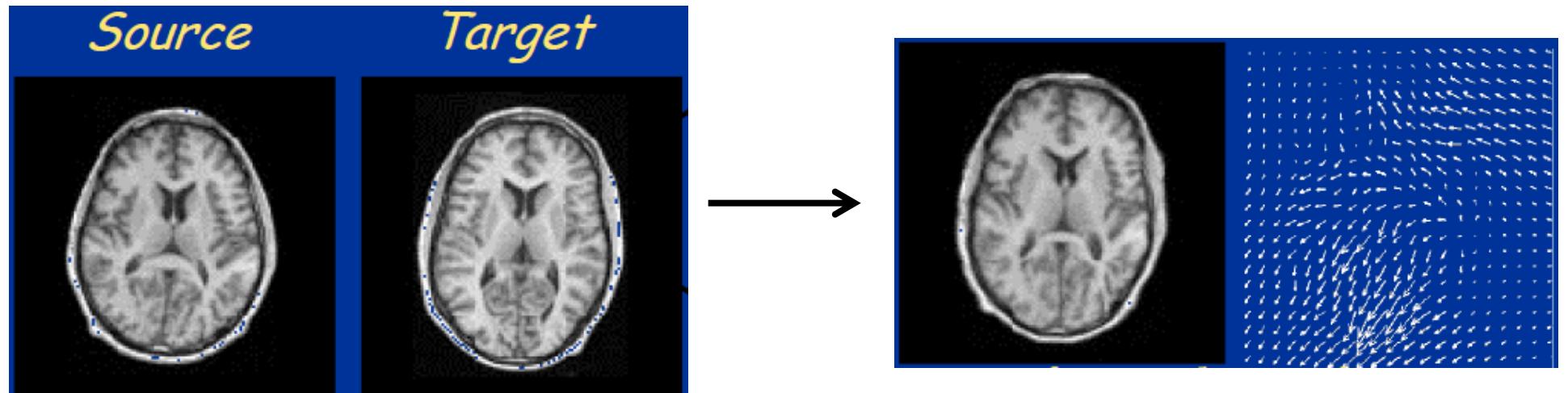
# Intensity Based

- Method
  - Calculating the registration transformation by optimizing some measure calculated directly from the voxel values in the images
- Algorithms used
  - Registration by minimizing intensity difference
  - Correlation techniques
  - Ratio image uniformity
  - Partitioned Intensity Uniformity



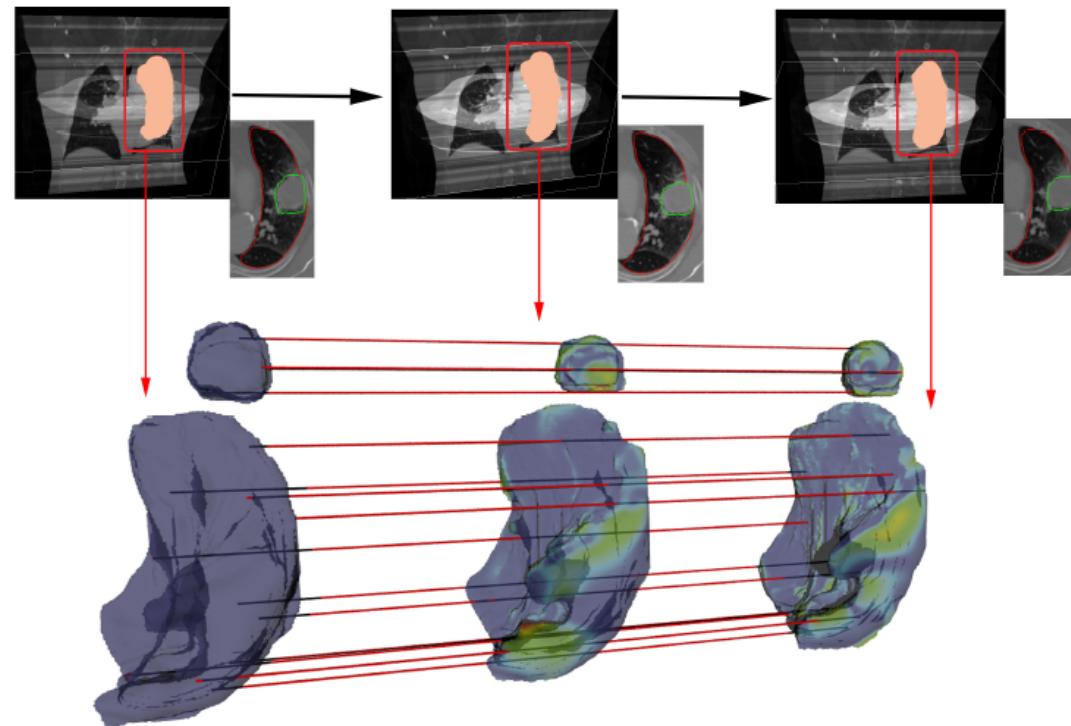
# Intensity Based

- Intensity-based methods compare intensity patterns in images via correlation metrics
- Sum of Squared Differences
- Normalized Cross-Correlation
- Mutual Information



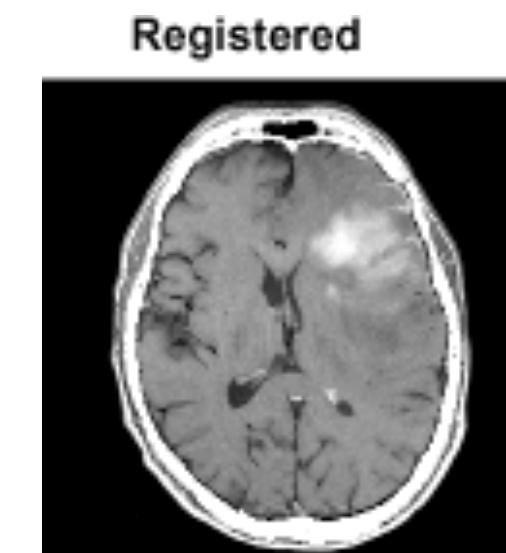
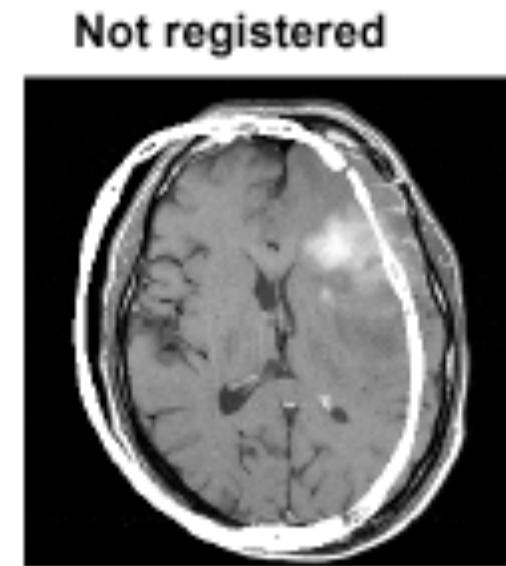
## Feature Based

- Feature-based methods find correspondence between image features such as points, lines, and contours.
- Distance between corresponding points
- Similarity metric between feature values
  - e.g. curvature-based registration



# Information Theory Based

- Image registration is considered as to maximize the amount of shared information in two images
  - reducing the amount of information in the combined image
- Algorithms used
  - **Joint entropy**
    - Joint entropy measures the amount of information in the two images combined
  - **Mutual information**
    - A measure of how well one image explains the other, and is maximized at the optimal alignment
  - **Normalized Mutual Information**





# Mutual Information

$$MI(I, J | T) = \sum_{i,j} p_{i,j} \log \frac{p_{i,j}}{p_i p_j}$$

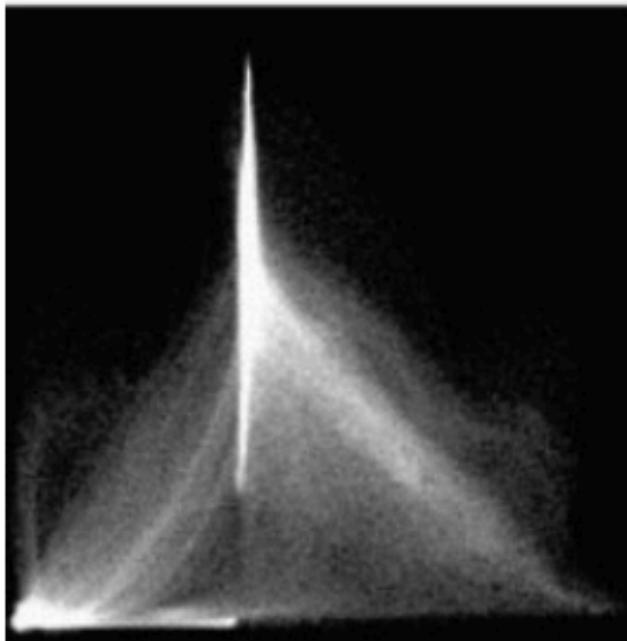
Algorithms for maximising mutual information (between intensities) have been the most popular for medical image registration to date.

There are many refinements underway ... not least using measurements of local phase instead of intensity\*

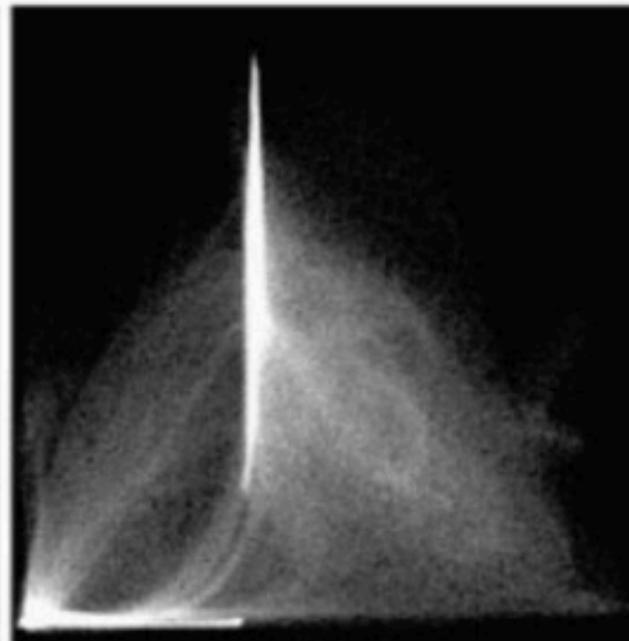
\*Mellor and Brady, Medical Image Analysis, 2004



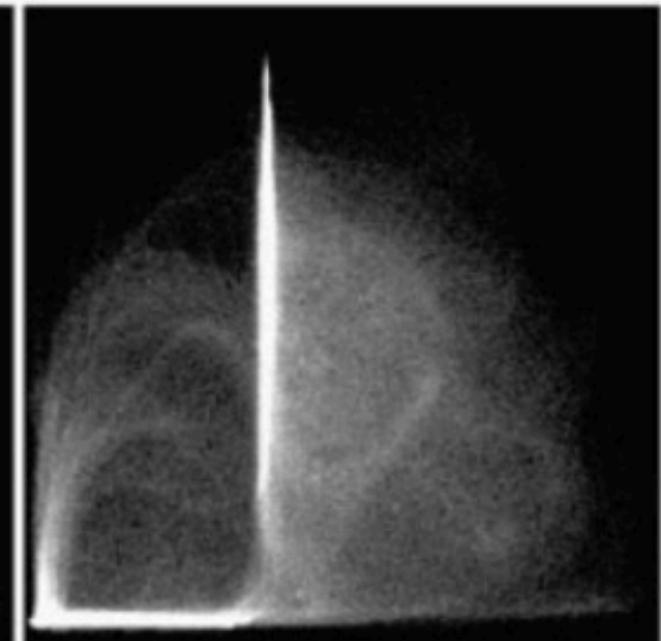
# Roger Woods' heuristic observation



Images perfectly aligned



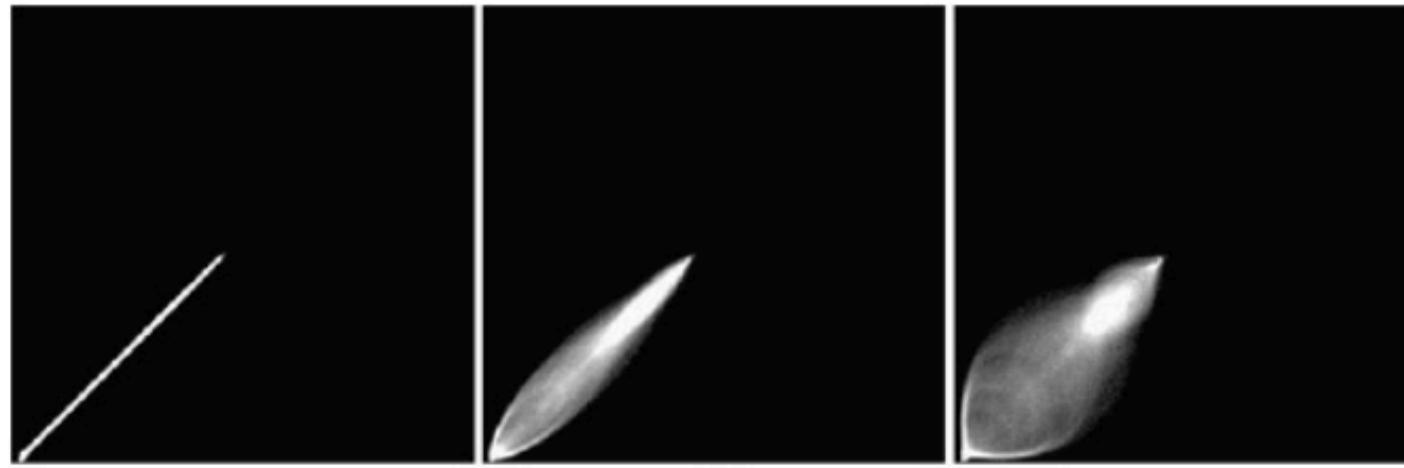
2mm displacement of one image  
to the side



5mm displacement of one image  
to the side

Heuristic observation is that when the images are aligned, the joint histogram appears “sharpest” : “Woods’ criterion”

*Why* this should be the case is still not certain!

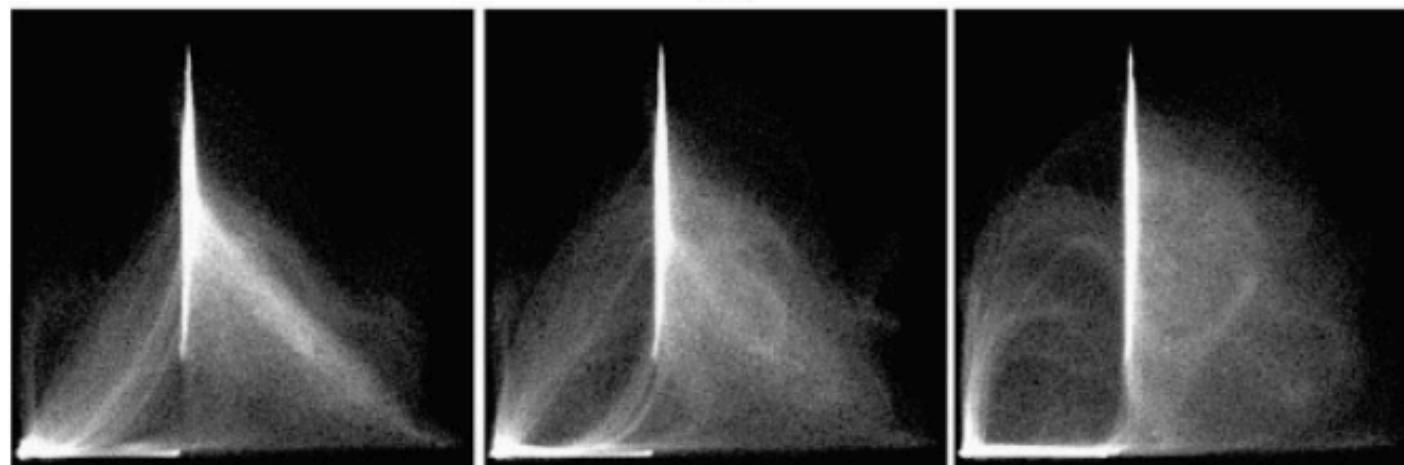


(a)

Top: MR-MR (head)

Middle: MR-CT

Bottom: MR-PET

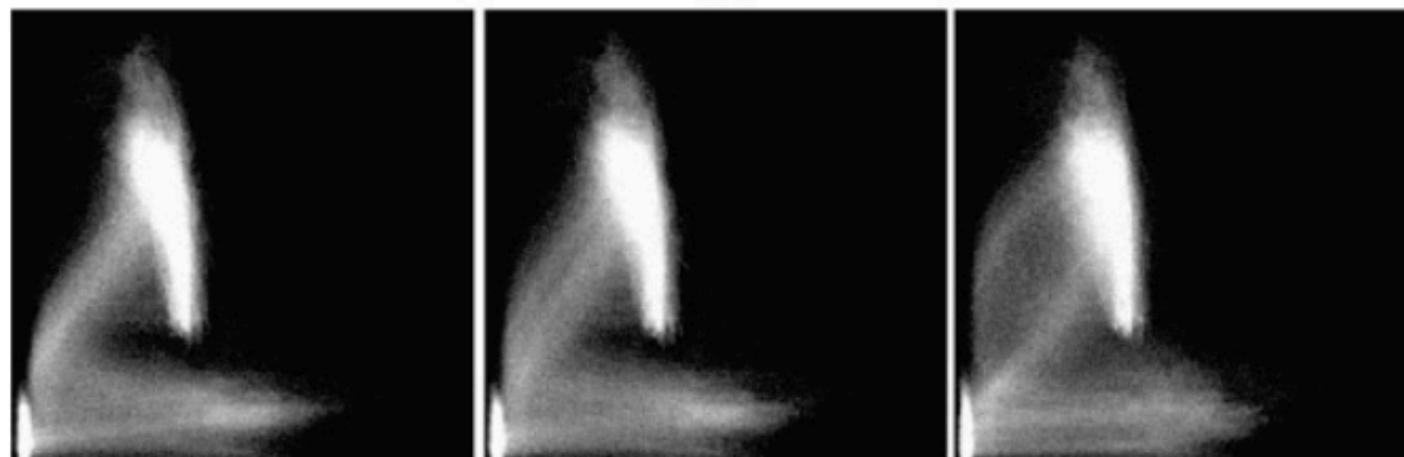


(b)

Left: aligned

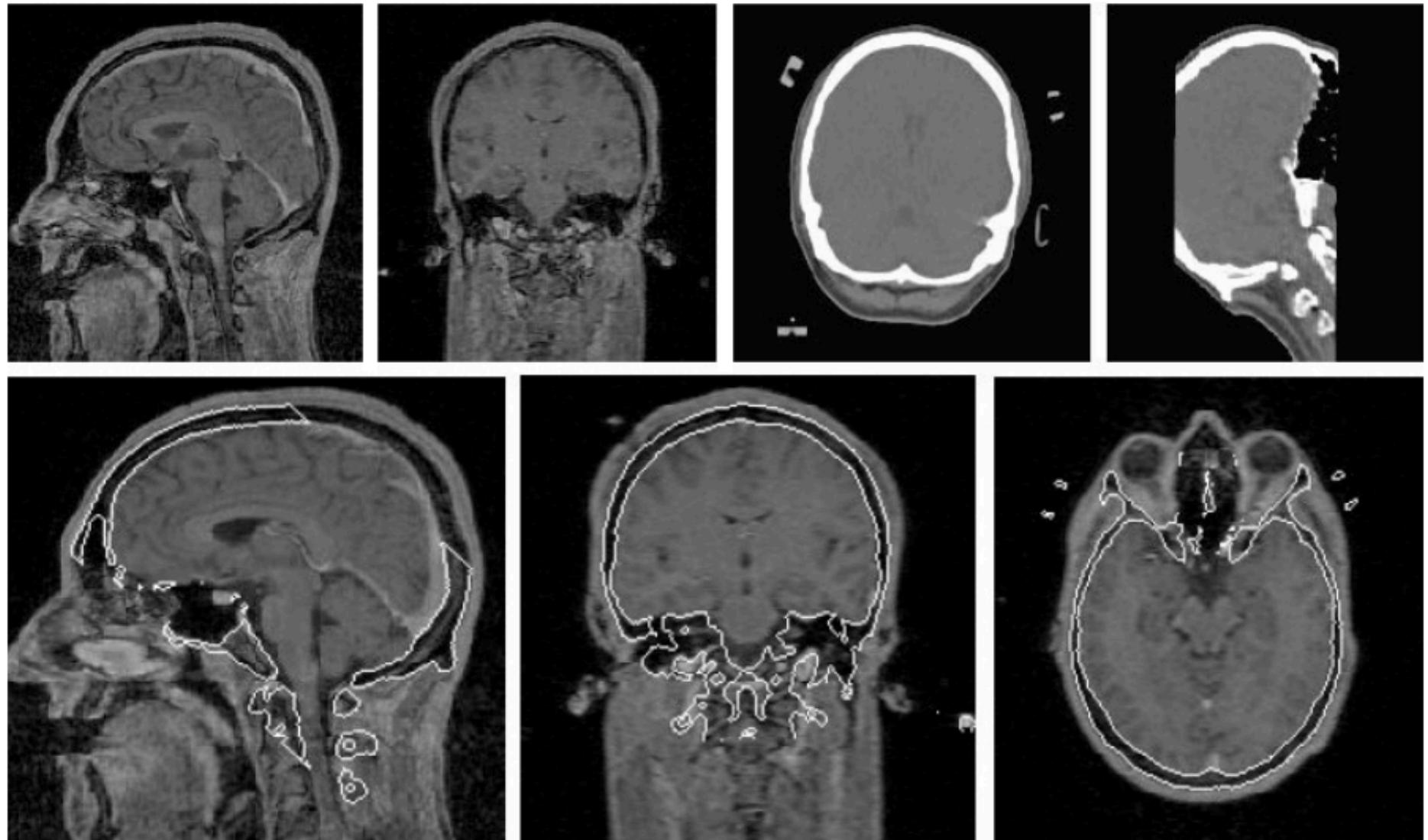
Middle: 2mm  
translation

Right: 5mm  
translation



Heuristic observation  
is that when the  
images are aligned,  
the joint histogram  
appears “sharpest”

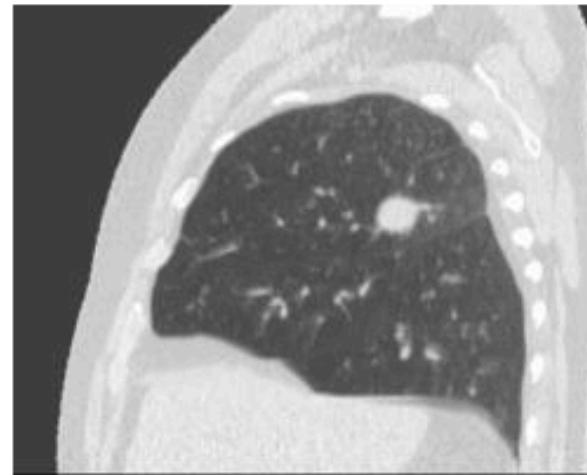
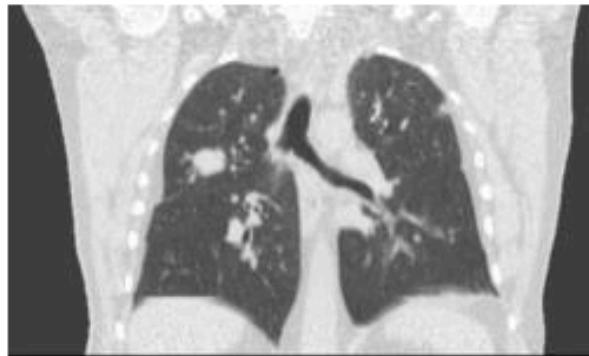
## Registration by maximising mutual information





# Non-Rigid (Deformable) Registration

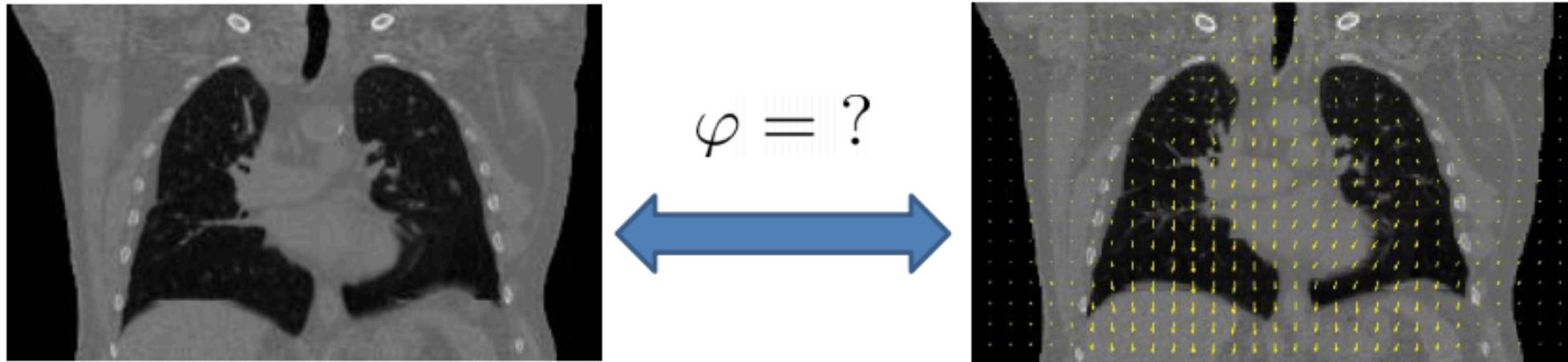
# Reasons for Deformable Registration



- Patients move (*alignment of temporal series*)
- Patients change (*pre- / post-treatment images*)
- Patients differ (*creation of atlases*)

# Deformable Registration

Computing a non-linear spatial transformation between  
***corresponding*** structures in two images



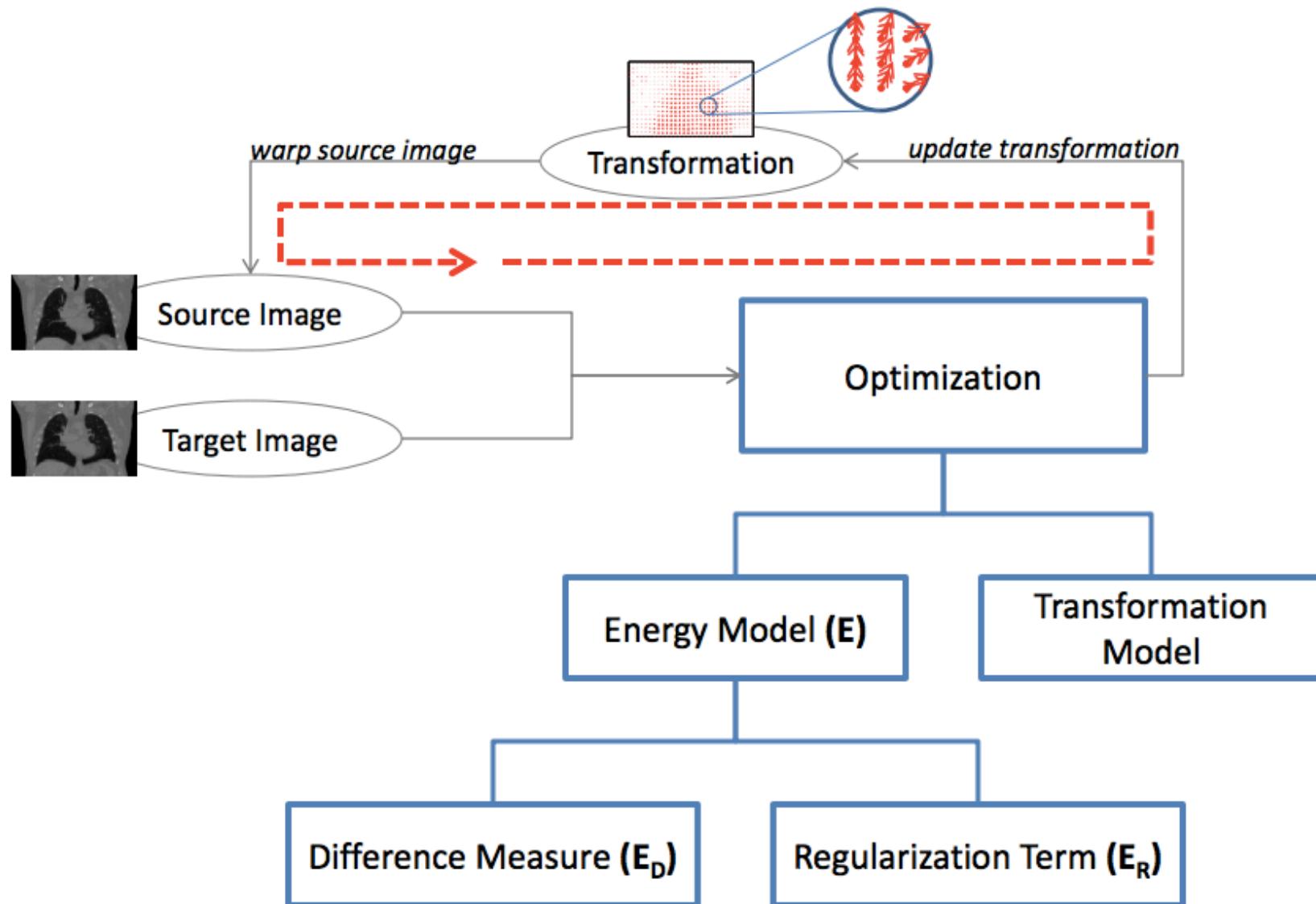
Intensity-based registration:

Minimize a difference term, based on the (pre-processed) image intensities.

No feature-based registration:

- extraction of distinct, sparsely located features
- matching of extracted features

# Deformable Registration General Framework



# Deformable Registration General Framework

Images (target  $I_T$ , source  $I_S$ ):  $I : \Omega \rightarrow \mathbb{R}$

Image domain:  $\Omega \subset \mathbb{R}^d$ ,  $d=2, 3$

$$u' = \arg \min_{u \in H} E_D(I_T, I_S(\varphi)) + \lambda E_R(u)$$

Displacements are elements of a Hilbert space  $H$ , e.g.:  $u \in L^2$

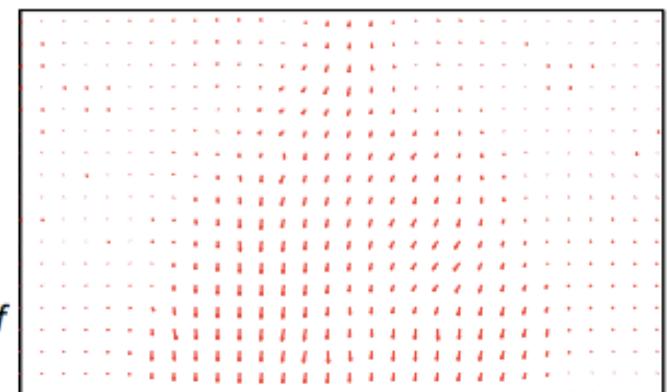
Deformation:  $\varphi = \text{Id} + u$ ,  $\varphi : \Omega \rightarrow \mathbb{R}^d$   
or point-wise:  $\varphi(x) = x + u(x)$

Displacement:

$$u : \Omega \rightarrow \mathbb{R}^d$$

e.g.  $u = [u_x, u_y, u_z]$

*difference between original position of a point  $x$  and its transformation  $\varphi(x)$*



Please note:  
Highly heterogeneous  
notation in the field



# Why do we need regularization?

# Why do we need regularization?

Since minimizing the difference measure is not enough...

- Motivation for regularization:

→ **Necessity:**

Minimization of difference measure only can be ill-posed ( #measurements < #unknowns )  
(Optical Flow community: “*Aperture Problem*”)

→ **Modeling:**

Regularization can be used to include prior knowledge, for example about underlying tissue properties

→ **Practical Reasons:**

Without regularization: High number of local minima in the energy function (→ bad for optimization)



# Regularization Strategies

$$u' = \arg \min_{u \in H} E_D(I_T, I_S(\text{Id}+u)) + \lambda E_R(u)$$

Restricting the space of deformations to a subspace with certain regularity properties:

- Parametrization reducing the number of degrees of freedom (e.g. FFD B-Splines)
- Assumption of function spaces which are more restricted than  $L^2$   
→ Sobolev spaces

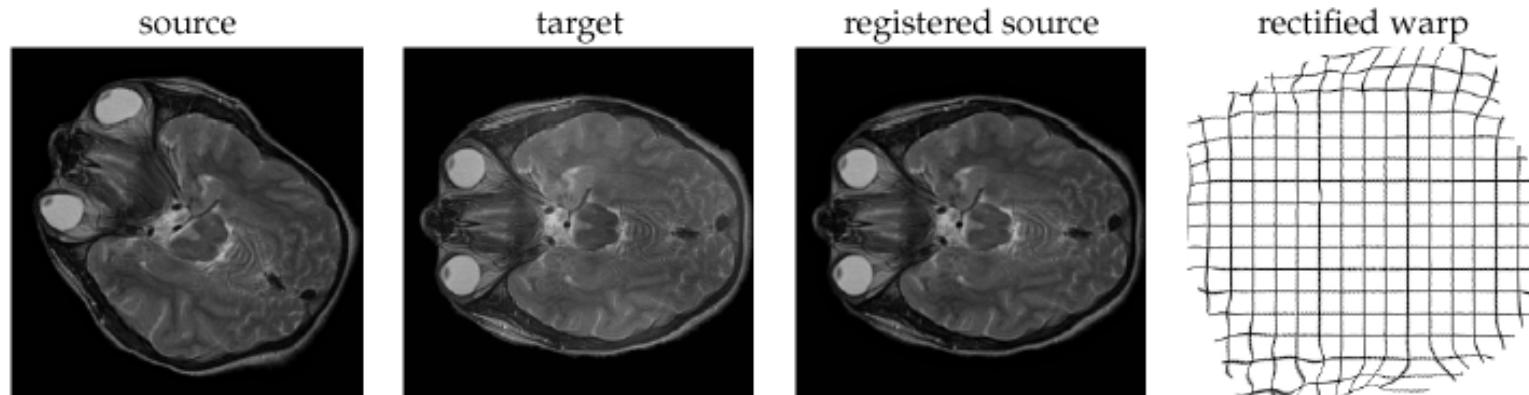
Adding a regularization term to the energy formulation, e.g.:

- Diffusion Regularization
- Curvature Regularization
- Bending Energy
- Linear Elasticity
- Volume Preservation



## Example: Elastic Registration

- Model the deformation as a physical process resembling the stretching of an elastic material
  - The physical process is governed by the internal force & external force
  - described by the Navier linear elastic partial differential equation
- The external force drives the registration process
  - The external force can be the gradient of a similarity measure
    - e.g. local correlation measure based on intensities, intensity differences or intensity features such as edge and curvature
  - Or the distance between the curves and surfaces of corresponding anatomical structures.



# Example: Free-Form Deformation (FFD)

- The general idea is to deform an image by manipulating a regular grid of control points that are distributed across the image at an arbitrary mesh resolution.
- Control points can be moved and the position of individual pixels between the control points is computed from the positions of surrounding control points.



# Other Physics Based Registration

- Fluid registration
  - The image was modeled as a highly viscous fluid
- Registration using mechanical models
  - using a three-component model to simulate the properties of rigid, elastic and fluid structures.
- Registration using optical flow



source



rotate



global scale



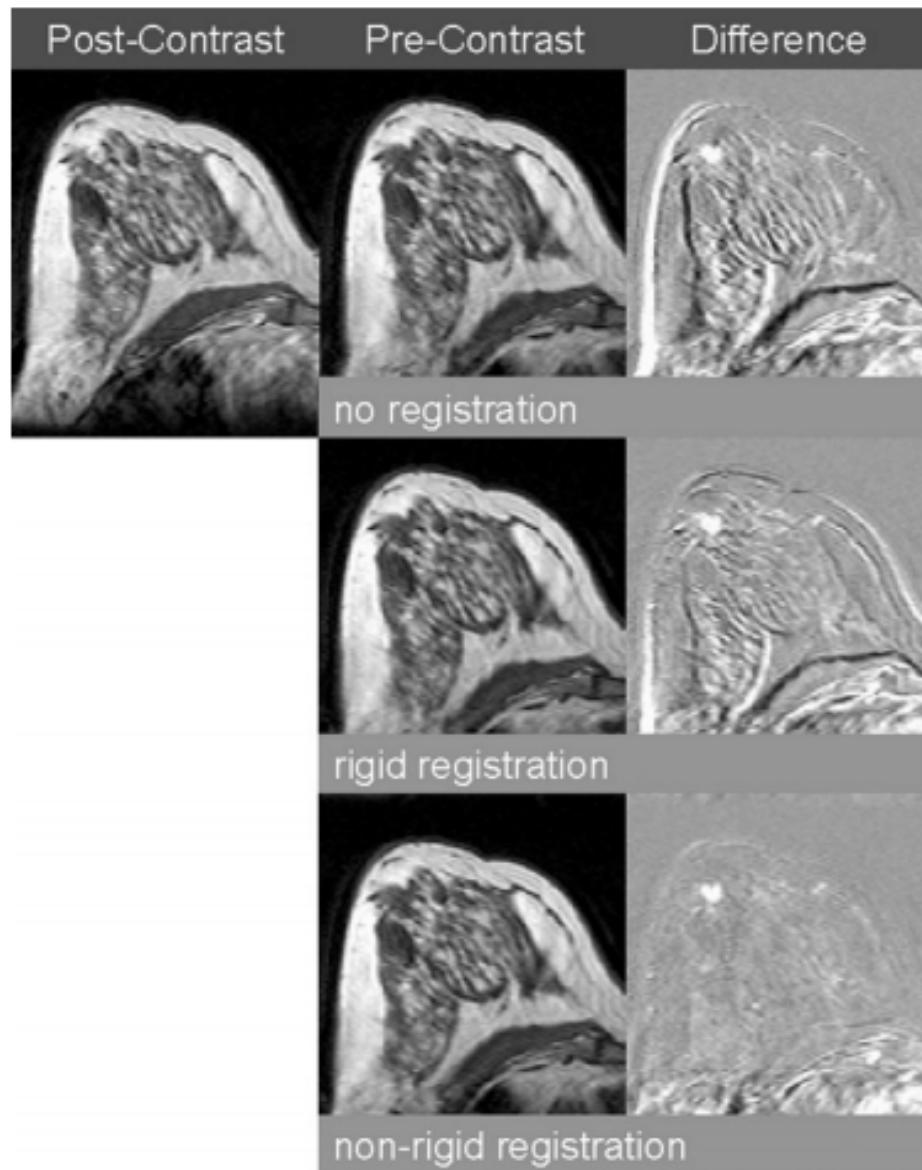
non-rigid



target



# Example



# Regularization – Classical Variational Approach

- Treatment of transformation in  $L^2$
- Explicit definition of regularization term in model

$$u' = \arg \min_{\underline{u \in L^2}} E_D(I_T, I_S(\text{Id}+u)) + \lambda E_R(u)$$

- Non-linear with respect to the displacement because of the dependence on the image function
- High-dimensional problem: e.g.  $3 \times 256^3 = 50\,331\,648$   
→ Numerous local minima
- This line of work started with [Broit 1981], [Bajcsy and Broit 1982]
- Details e.g. in [Modersitzki 2004]



# Exemplary Model Problem

$$E(u) = E_D(I_T, I_S(\varphi)) + \lambda E_R(u)$$

- Difference Measure: Sum of Square Differences (SSD)

$$E_D = \frac{1}{2} \int_{\Omega} (I_T(x) - I_S(\varphi(x)))^2 dx$$

- Regularization Term: Diffusion Regularization  
(a.k.a. Tikhonov regularization, 1st order regularization)

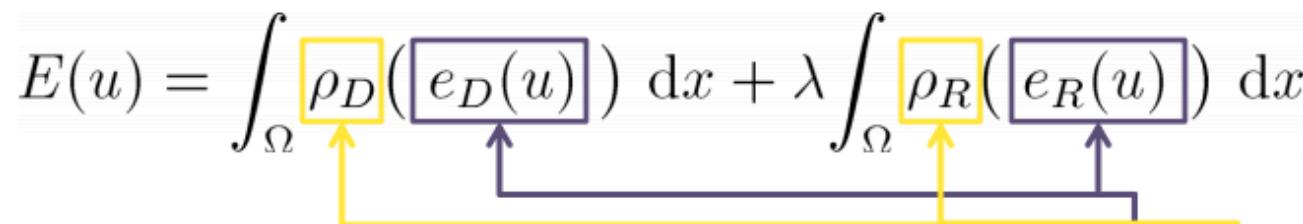
$$E_R(u) = \frac{1}{2} \int_{\Omega} \sum_{i=1}^d \|\nabla u_d(x)\|^2 dx$$

e.g.  $E_R(u) = \frac{1}{2} \int_{\Omega} \|\nabla u_x(x)\|^2 + \|\nabla u_y(x)\|^2 dx$

→ Non-linear Least-squares Energy

# General Energy Formulation

$$E(u) = E_D(I_T, I_S(\varphi)) + \lambda E_R(u)$$

$$E(u) = \int_{\Omega} \rho_D(e_D(u)) \, dx + \lambda \int_{\Omega} \rho_R(e_R(u)) \, dx$$


- Energy is defined by
  - an error term based on the displacement:  $e(u)$
  - penalty function applied to the error term:  $\rho$ 
    - assures that the error terms are positive
    - weights the error term

## Examples of penalty functions

$\rho(x) = x^2 \rightarrow$  results in application of the L<sup>2</sup> norm

$\rho(x) = |x| \rightarrow$  results in application of the L<sup>1</sup> norm

# General Energy Formulation

$$E(u) = \int_{\Omega} \rho_D(e_D(u)) \, dx + \lambda \int_{\Omega} \rho_R(e_R(u)) \, dx$$

Depends on deformation through images

$$e_D(u) \equiv e_D(I_T, I_S \circ (\text{Id} + u))$$

→ Non-linearity

- In many cases, the error term for the regularization is linear in the displacement
- Linear operator is mostly a differential operator (e.g.  $G=\nabla$ ,  $G=\Delta$ , ...)

Displacement

Source image  
( $\rightarrow$  warped source image)  
**Non-Linearity:** change of intensity  
in one point does not depend  
linearly on the displacement

Difference measure

Displacement

Computation of  
Displacement Derivatives  
(Linear Operation)

Regularization term

# Optimization

- Many registration algorithms require an iterative approach
  - an initial estimate of the transformation is gradually refined
  - In each iteration, the current estimate of the transformation is used to calculate a similarity measure
  - makes another estimate of the transformation, evaluates the similarity measure again, and continues until the algorithm converges
    - no transformation can be found that results in a better value of the similarity measure, to within a preset tolerance.



# Iterative Optimization

$$u' = \arg \min_u E(u)$$



- Start with an initial estimate  $u_0$
- Estimate a series of updates  $h_1, \dots, h_n$  such that

$$u' \approx u_0 + h_1 + \dots + h_n$$



```
do repeat:  
    h = compute_update(E, φ);  
    φ = φ + h;
```



# How to determine the updates $h_i$ ?

# How to determine the updates $h_i$ ?

- Gradient-based optimization

$$h = -\tau \text{ some\_function}(\nabla E(u))$$

- Gradient-free optimization

$$h = \text{some\_other\_function}(E_D, E_R, u)$$

# Steepest Gradient Descent

- Energy:  $E(u) = E_D(I_T, I_S(\varphi)) + \lambda E_R(u)$

Starting with initial  $\varphi_0$ , repeat until convergence:

$$h = -\tau \nabla E(\varphi)$$

// compute update based on gradient of energy

$$\varphi = \varphi + h$$

// apply the update

Only the derivative of the energy w.r.t the displacement is required:

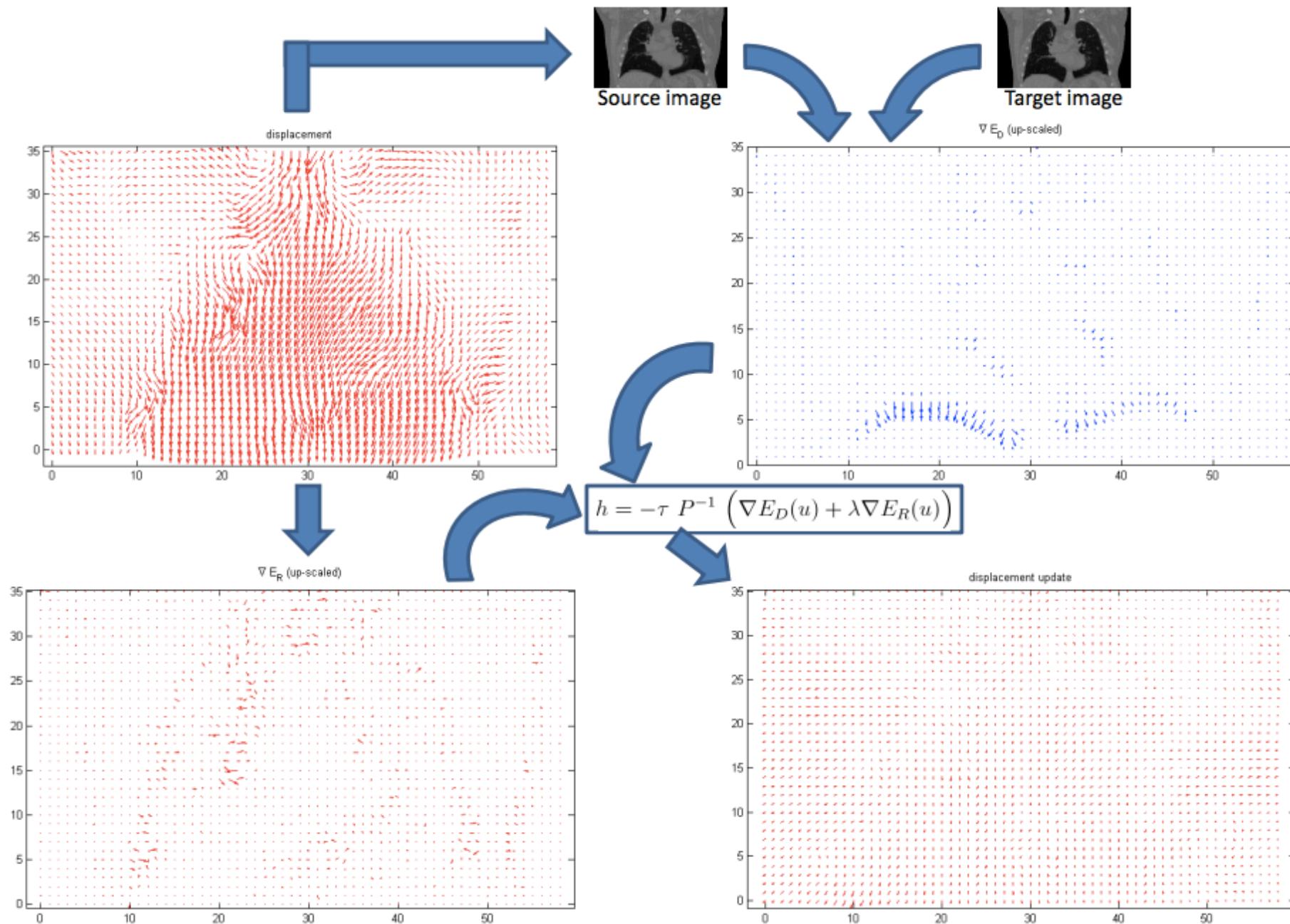
$$\nabla E(\varphi) = \nabla E_D(\varphi) + \lambda \nabla E_R(\varphi)$$

→ derivative of difference measure

→ derivative of regularization term

# Overview of Gradient Descent Based Optimization Methods

Method	Update Rule	Comment
Steepest Descent	$h = -\tau \text{ Id}^{-1} \nabla E(u)$	<ul style="list-style-type: none"> <li>+ Simple implementation</li> <li>+ Only gradient required</li> <li>- Numerical issues: requires small time steps → many iterations needed</li> </ul>
PDE inspired Semi-implicit Discretization	$h = -\tau (\text{Id} + \tau \lambda \nabla E_R)^{-1} \nabla E(u)$	<ul style="list-style-type: none"> <li>+ Numerically stable also for large time steps</li> <li>+ Linear operator determined by regularization → difference measure easily exchangeable</li> <li>- Poor convergence speed</li> </ul>
Gauß-Newton	$h = -\tau (J_e^\top J_e)^{-1} \nabla E(u)$	<ul style="list-style-type: none"> <li>+ Numerically stable also for large time steps</li> <li>+ Good convergence speed</li> <li>- Linear operator depends on both, the regularization <i>and</i> the difference term</li> <li>- Applicable only to least-squares problems</li> <li>- <math>J_e</math> must be sparse for efficient treatment</li> </ul>
Preconditioned Gradient Descent	$h = -\tau P^{-1} \nabla E(u)$ With $P$ approximating the Hessian of $E$ , e.g.: <ul style="list-style-type: none"> <li>• Jacobi preconditioning</li> <li>• BFGS</li> </ul>	Most general formulation of the above. Properties depend heavily on choice of $P$ . <p><i>"Finding a good preconditioner (...) is often viewed as a combination of art and science."</i> Y. Saad, Iterative Methods for Sparse Linear Systems</p>



# Summary

- Introduction to the medical image registration
- Transformation types
  - Rigid, affine, non-rigid
- Mono-modal, multi-modal image registration
- Similarity metric
- Regularization and Optimization
- **Next Lecture(s):** further details on the topic.

# Slide Credits and References

- **Credits to:** Jayaram K. Udupa of Univ. of Penn., MIPG
- Sir M. Brady's Lecture Notes (Oxford University)
- Darko Zikic's MICCAI 2010 Tutorial
- Bagci's CV Course 2015 Fall.
- K.D. Toennies, Guide to Medical Image Analysis,
- Handbook of Medical Imaging, Vol. 2. SPIE Press.
- Handbook of Biomedical Imaging, Paragios, Duncan, Ayache.
- Seutens,P., Medical Imaging, Cambridge Press.
- Aiming Liu, Tutorial Presentation.
- Jen Mercer, Tutorial Presentation.