Introduction to Deformable Registration:

Deformation Models, Similarity Metrics & Optimization Methods

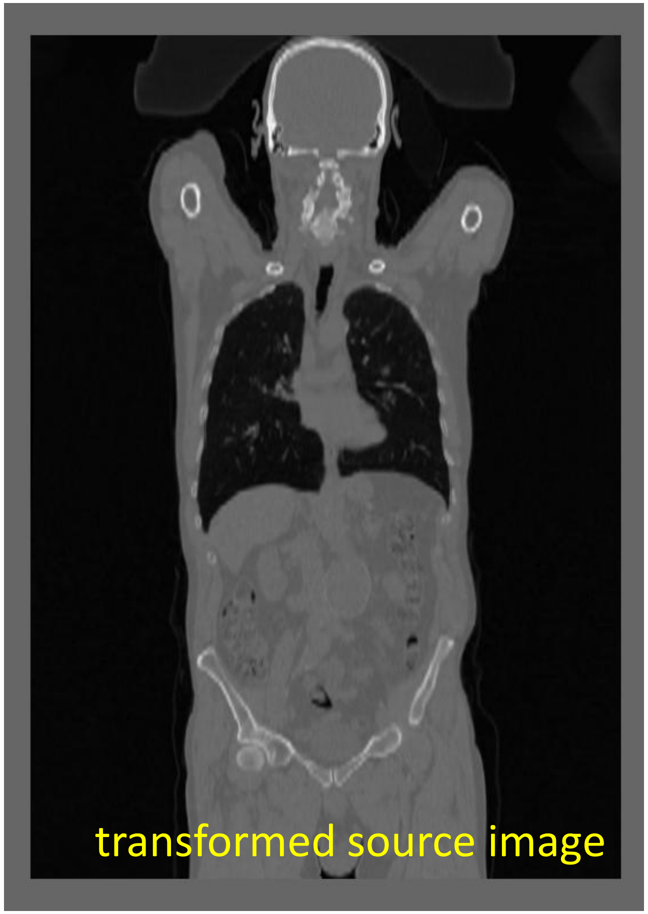
Nikos Paragios

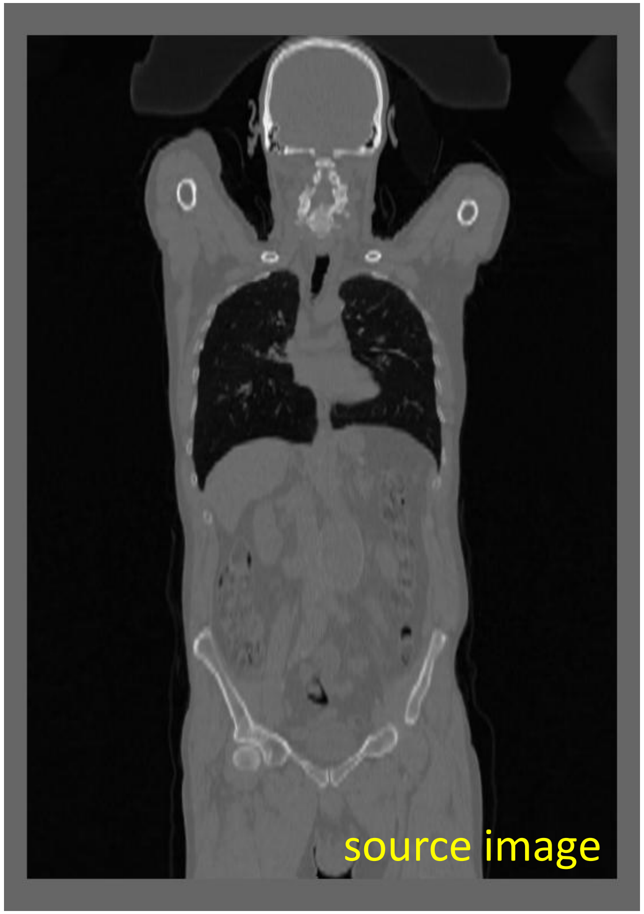
Slides Courtesy:
A. Sotiras (UPENN), C. Wachinger (MIT) & D. Zikic (MSFT)



Image Registration

Spatially align two input images, by computing the spatial transformation, such that the transformed source image matches the target image



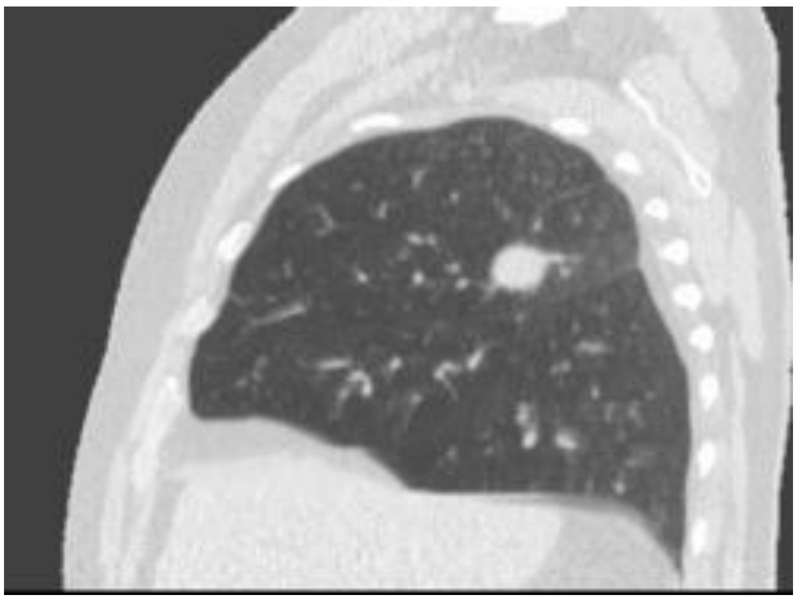


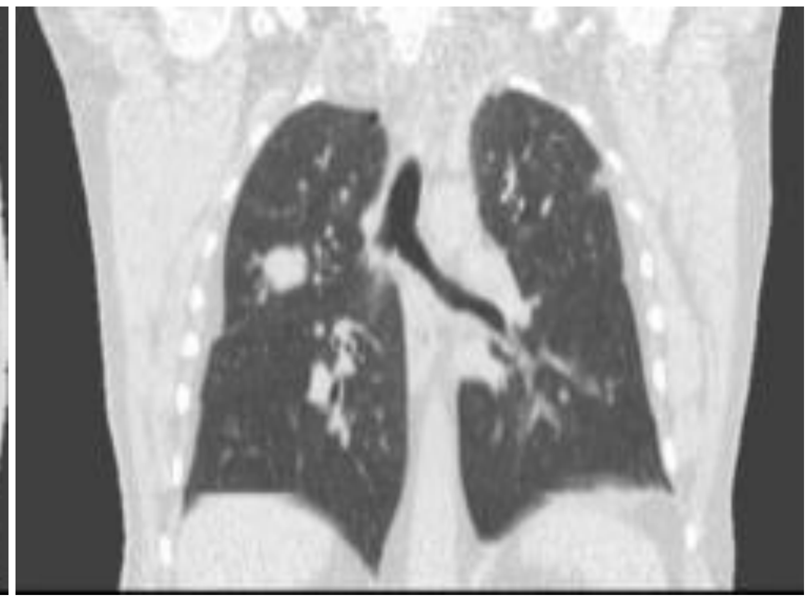
Reasons for Deformable Registration

- Subjects move
  (alignment of temporal series, patient positioning)
- Subjects change
  (longitudinal studies, pre- / post-treatment images)
- Subjects differ
  (creation of atlases, segmentation transfer)

Subjects Move

(alignment of temporal series)





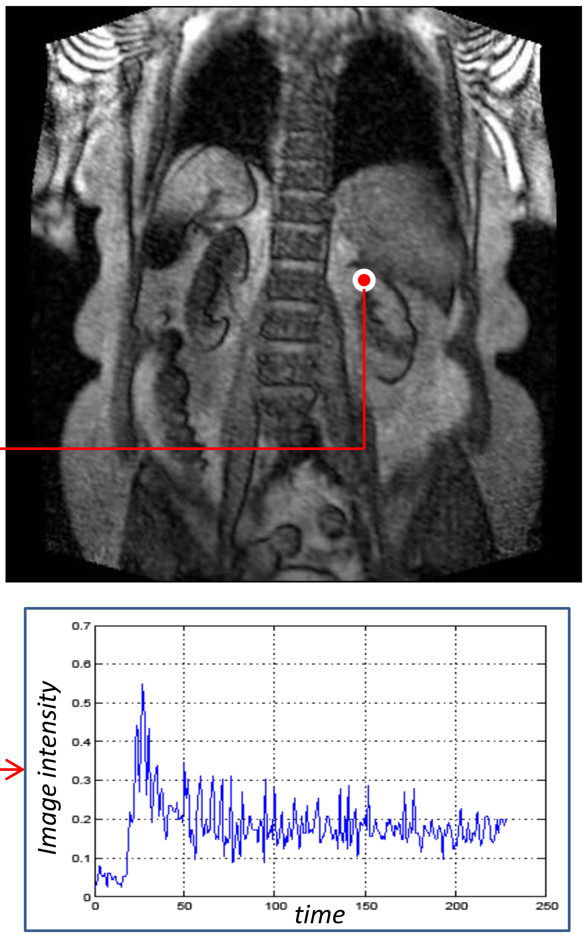
Animated images from the webpage of

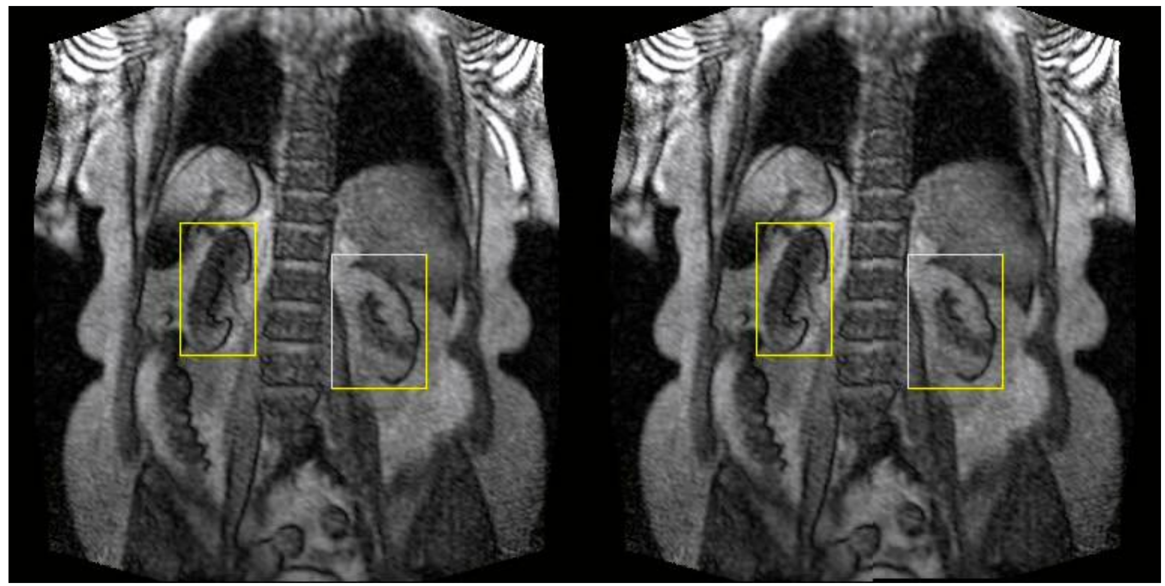
The POPI-model, a Point-validated Pixel-based Breathing Thorax Model

http://www.creatis.insa-lyon.fr/rio/popi-model

Vandemeulebroucke, J., Sarrut, D. and Clarysse, P.. *The POPI-model, a point-validated pixel-based breathing thorax model.* In *XVth International Conference on the Use of Computers in Radiation Therapy (ICCR)*, 2007.

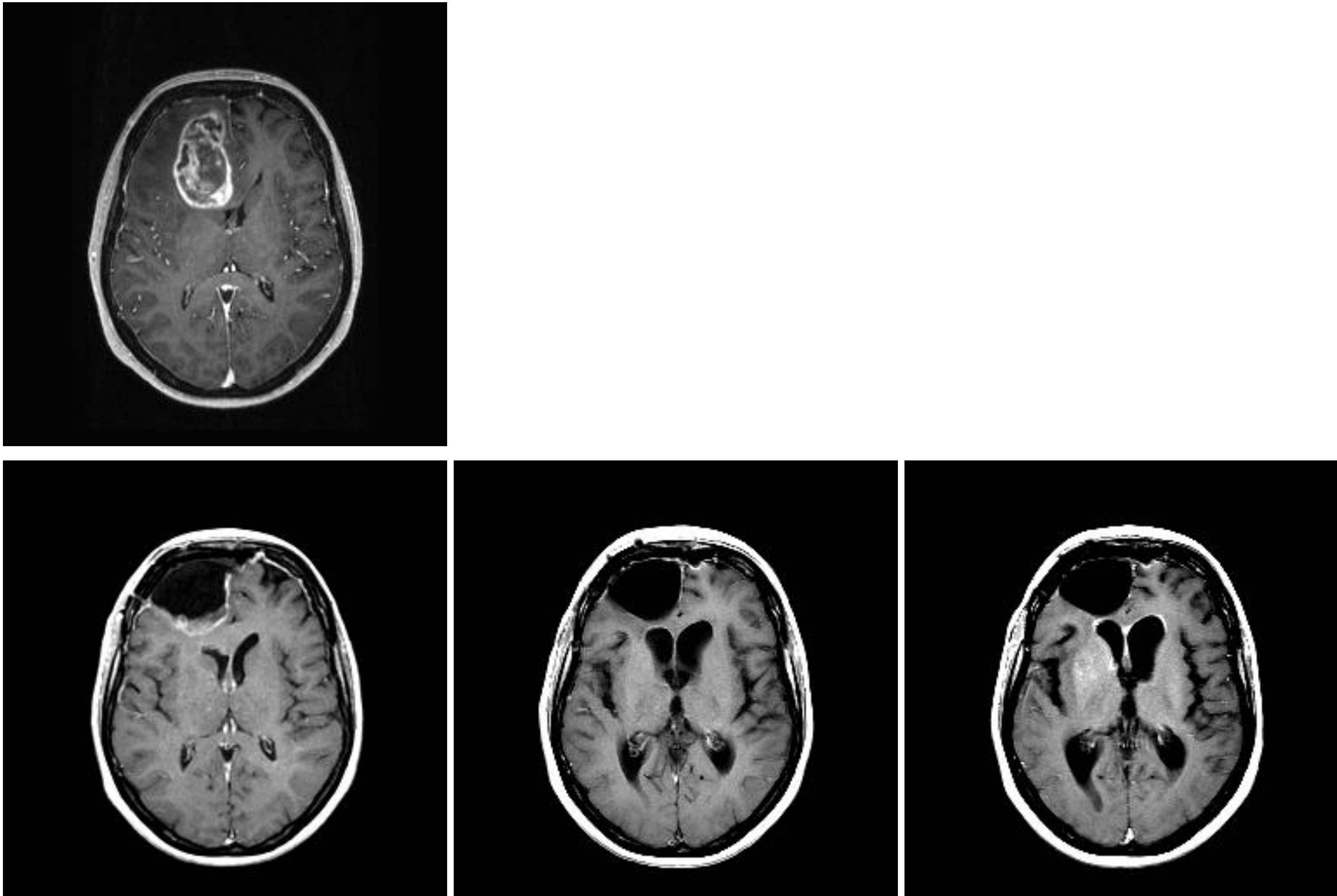
Subjects Move





D. Zikic, S. Sourbron, X. Feng, H. J. Michaely, A. Khamene, N. Navab
Automatic Alignment of Renal DCE-MRI Image Series for Improvement of Quantitative Tracer Kinetic Studies. SPIE Medical Imaging, 2008.

Subjects change over time

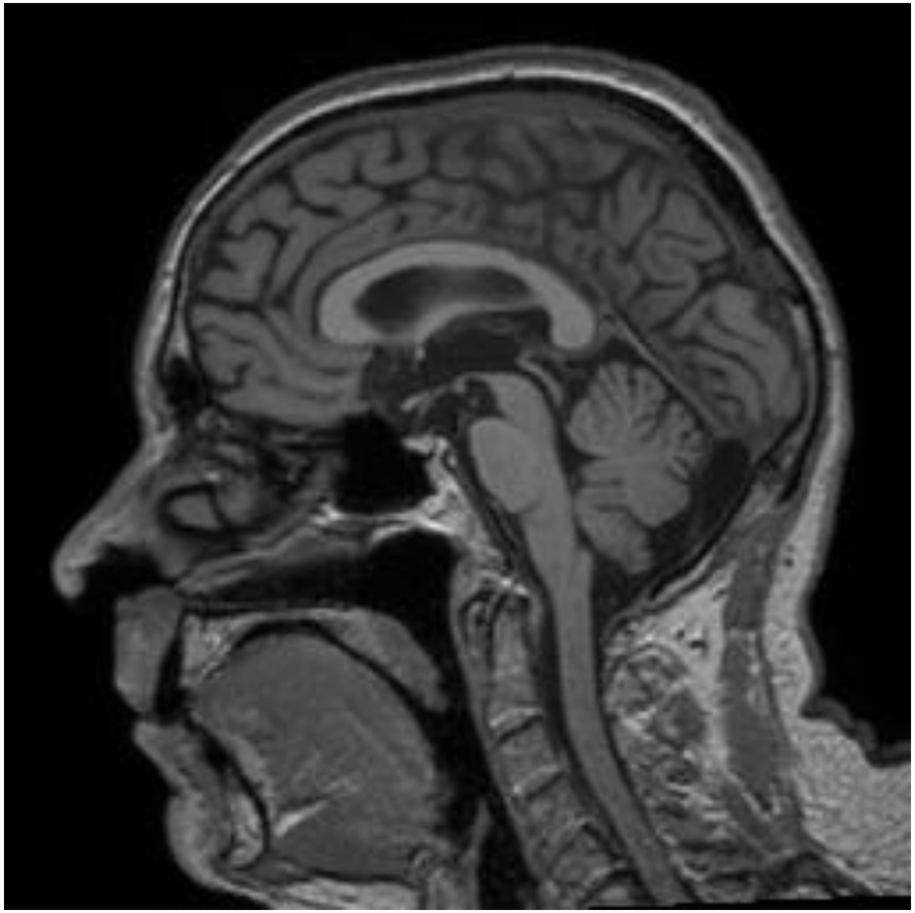


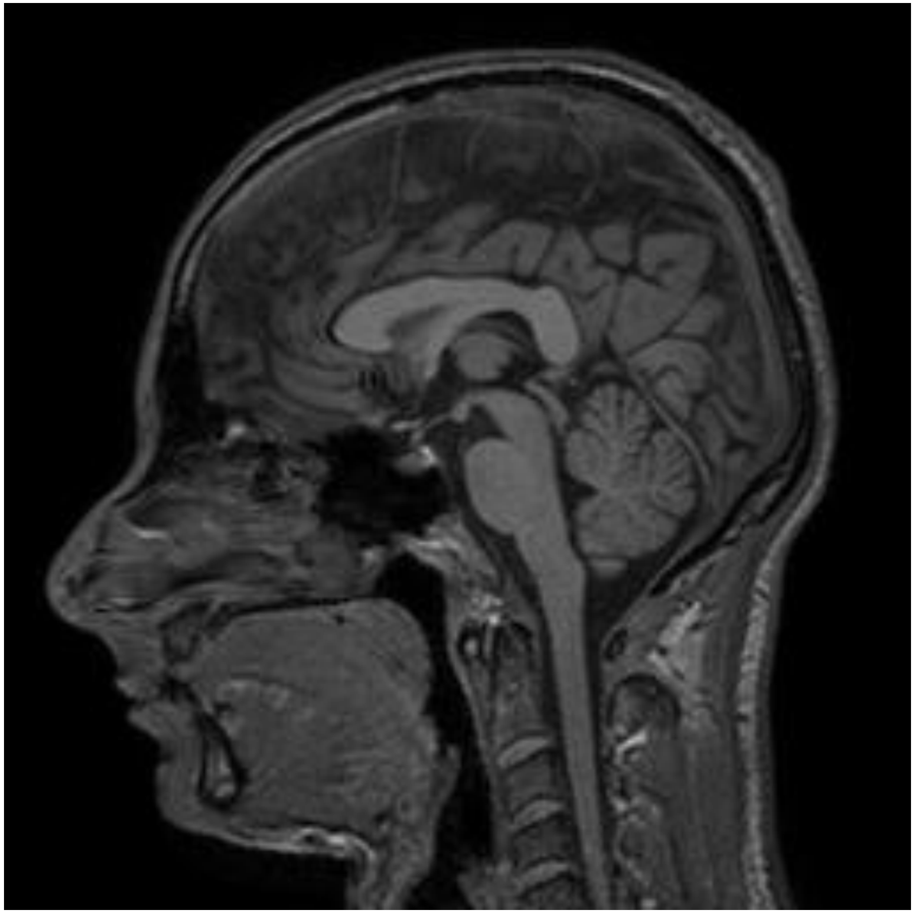
Glioma patient before surgery

Follow up scans after surgery

Subjects Differ

IXI Database
http://biomedic.doc.ic.ac.uk/brain-development/

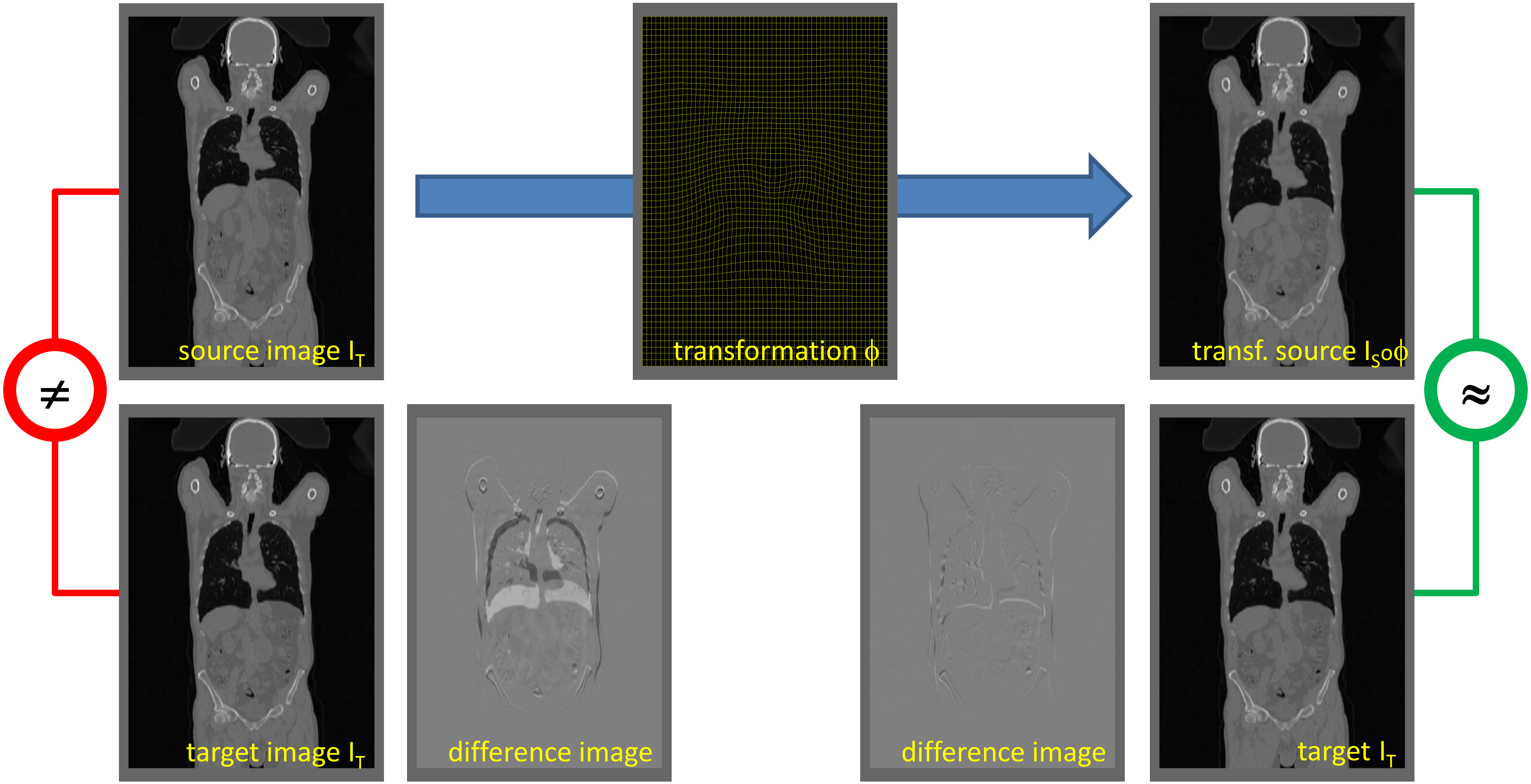




Application Example:
Non-Linear registration of brain MRI for Segmentation Propagation

[Rohlfing 2004],[Warfield 2004],[Heckemann 2006],[Klein 2009],
[Multi-Atlas Labeling Workshop at MICCAI 2012]

Intensity-based Image Registration



Compute deformation ϕ, such that the transformed *source* $I_S \circ \phi$ matches *target* I_T by minimizing the image-based difference measure E_D.

Some Basic Classes of Registration Methods

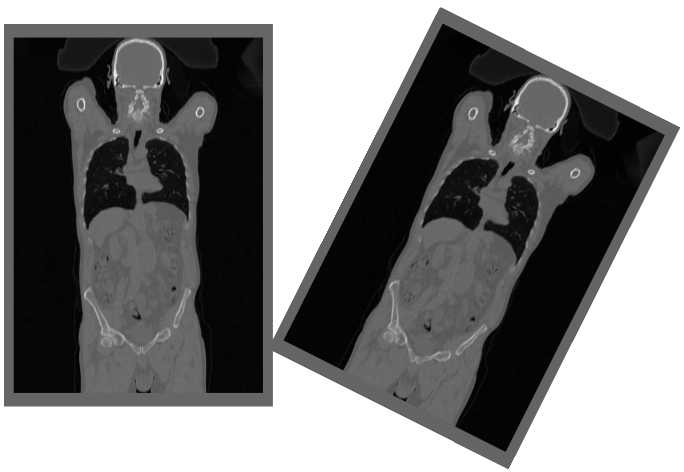
Feature-based Registration

extraction & matching of specific spatial features

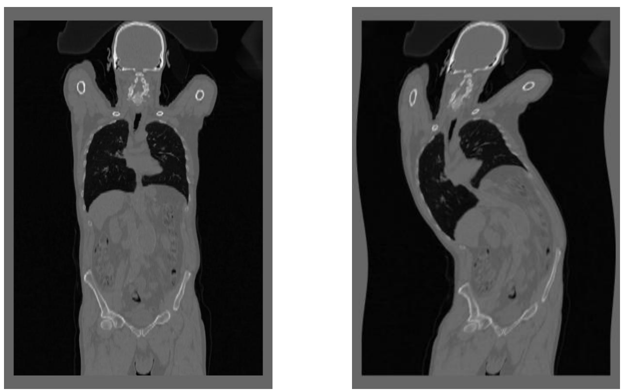
Intensity-based Registration

image-based difference measure

Linear/Rigid Registration



Deformable Registration



Intensity-based Deformable Registration as Energy Minimization

$$\phi' = \arg\min_{\phi} \left[E_{\mathrm{D}}(I_{\mathrm{S}} \circ \phi, I_{\mathrm{T}}) + \lambda E_{\mathrm{R}}(\phi) \right]$$

$\phi : \mathbb{R}^d \to \mathbb{R}^d$

$I : \mathbb{R}^d \to \mathbb{R}$

Transformation ϕ
can assumed as element of:
- Can be modeled as elemet of a Hilbert space (L^2, Sobolev space) or group/manifold (group of diffeomorphisms)
- Has to be parametrized for digital representation (B-Spline FFDs, DCT, RBFs)

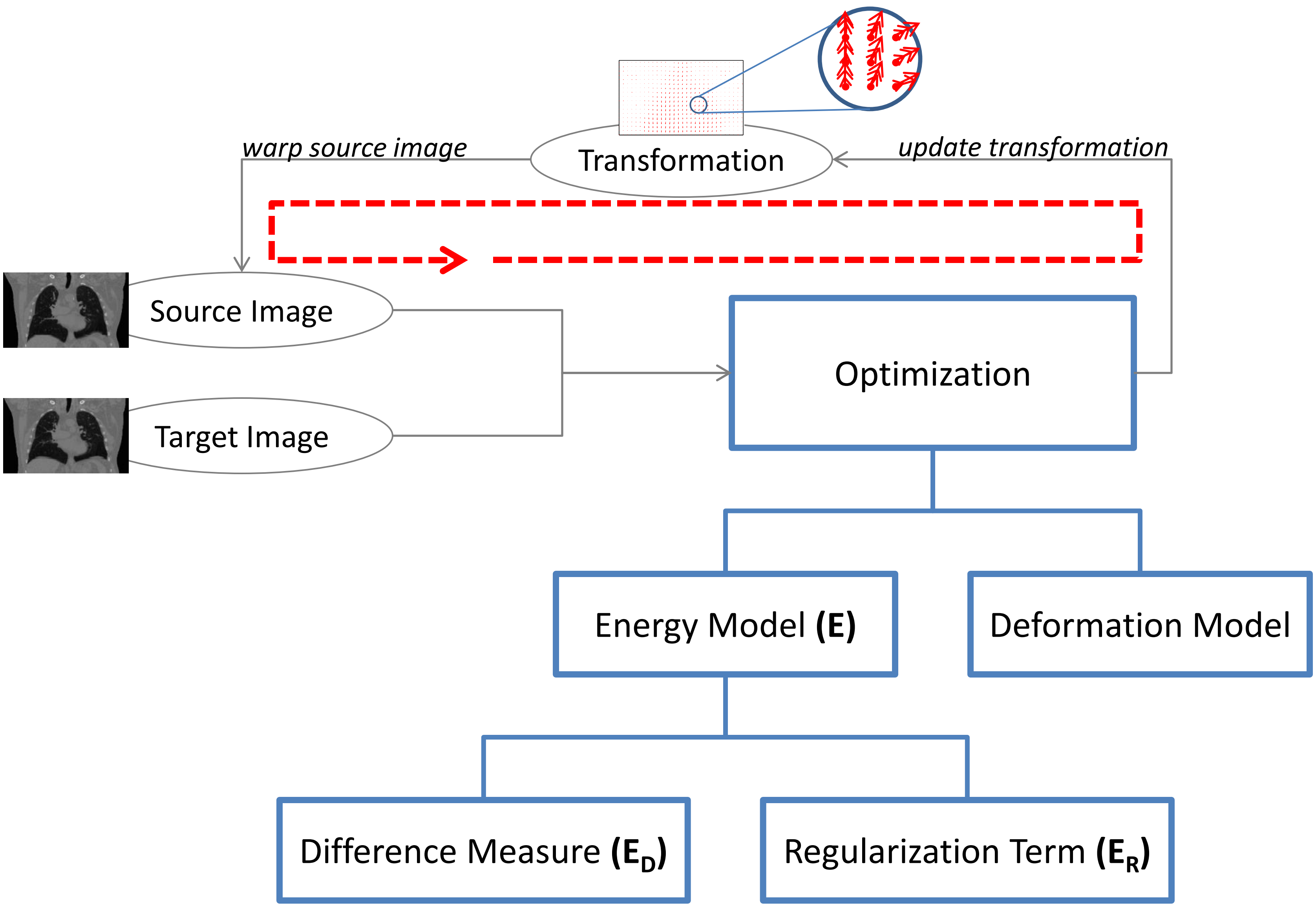
Difference Measure between:
- Target image I_T
- Warped source image $I_S \circ \phi$

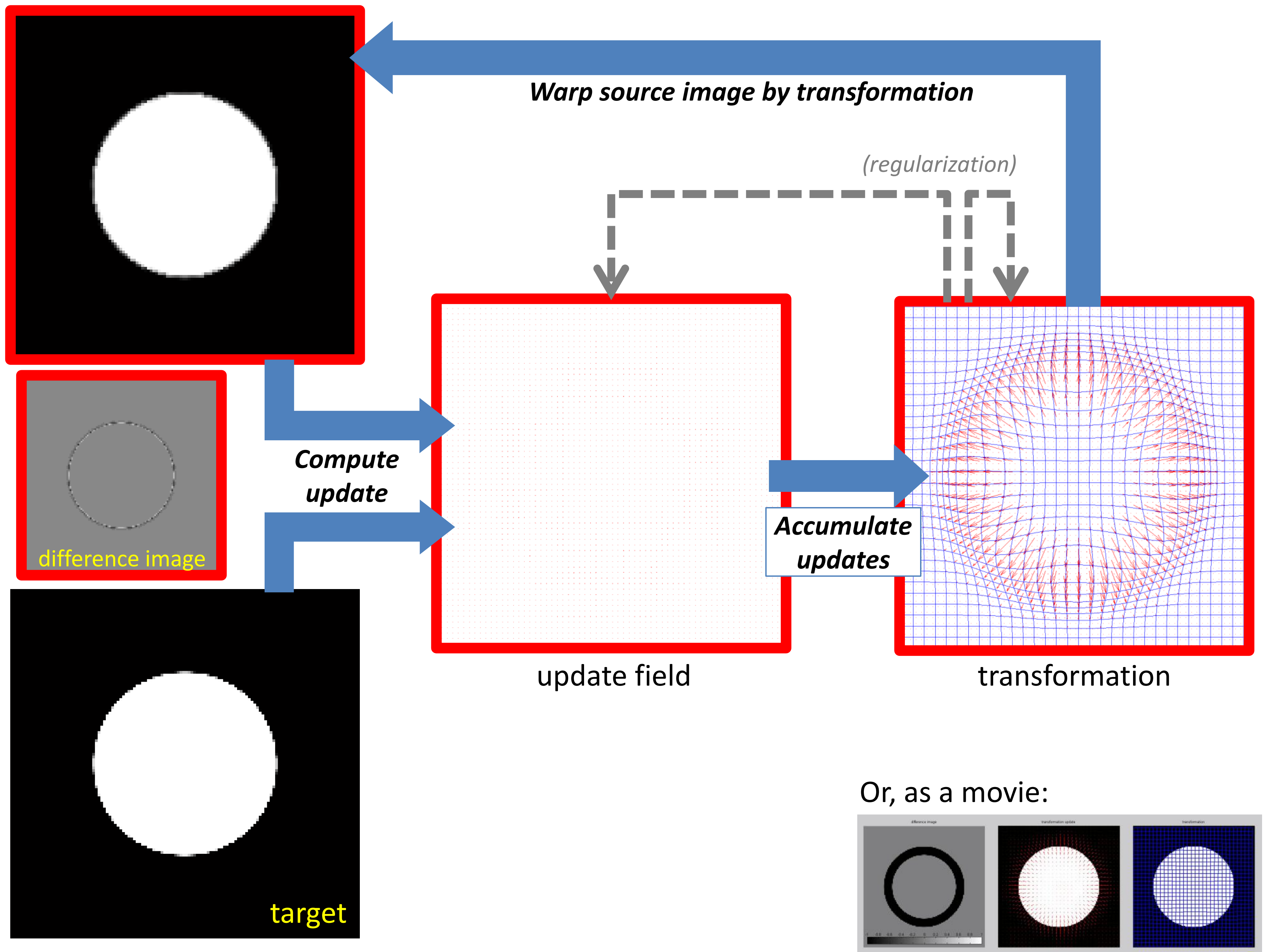
Examples:
- Sum of squared differences (SSD)
- Sum of absolute differences (SAD)
- Correlation Coefficient (CC)
- Correlation Ratio (CR)
- Mutual Information (MI)

Regularization term:
- Models the behaviour of underlying elastic model (internal energy)
- Incorporates prior knowledge
- can be required to constrain problem

Examples:
- Diffusion (1st-order) ((in-)homogeneous, (an-)isotropic)
- Curvature/Bend. Energy (2nd-order)
- Linear Elasticity

Deformable Registration: General Framework





Warp source image by transformation
(regularization)
difference image
Compute update
Accumulate updates
update field
transformation
target
Or, as a movie:

PART I

Deformation models

Deformation Modelling

$$\phi' = \arg\min_{\phi_p} \Big[\mu E_{\mathrm{D}}(I_{\mathrm{S}} \circ \phi_p, I_{\mathrm{T}}) + \lambda E_{\mathrm{R}}(\phi) \Big]$$

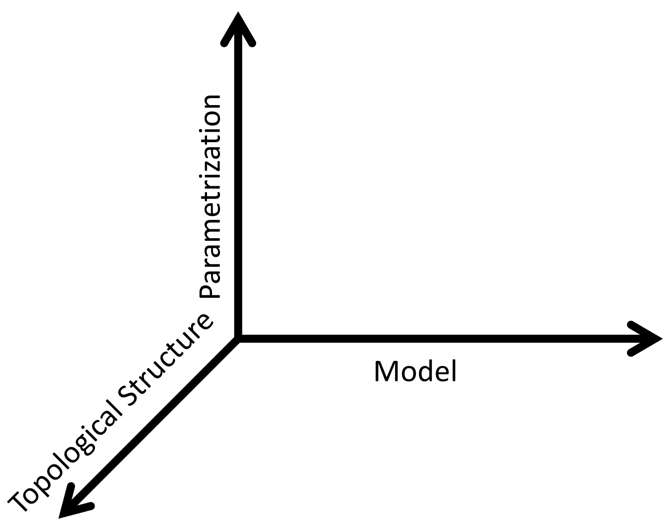
$\phi \in H$

$\phi : \mathbb{R}^d \to \mathbb{R}^d$

$\phi : (x, p) \mapsto \phi_p(x)$

$\phi : \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}^d$

$\phi_p \approx \phi$



Deformation Model
(which theoretical model should govern the process)

Elastic
(regularization term)

Fluid
(deformation space)

Deformation Parametrization
(how to represent a deformation on a computer)
(approximation to the theoretical model)

"Non-Parametric" Dense
(actually highly parametric)

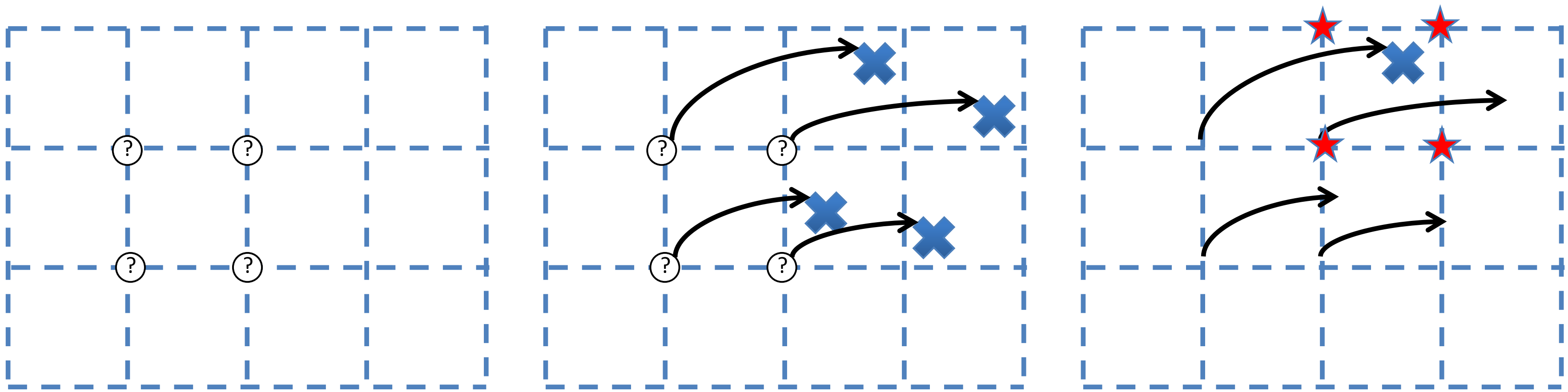
Parametric
Free-form Deformations
Finite Element Model
Cosine/Sine/fourier Transformations

Deformation Topology/Structure
(how to combine deformations)

Vector Space
(deformations are added)

Group Structure
(deformations are composed)

Deformation process



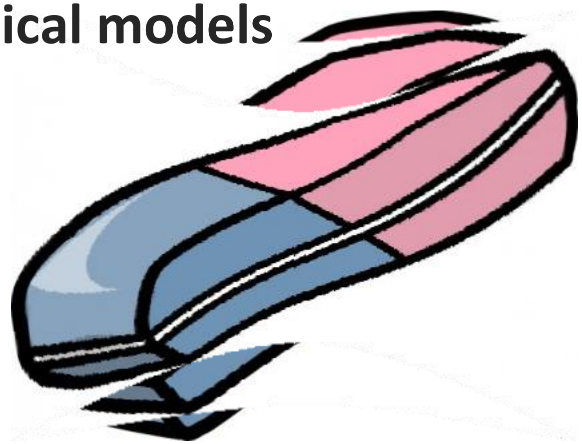
What is the intensity of $I_D(x_1)$ where x_1 is a pixel location in the deformed image I_D?

Pull from $I_S(x+u(x))$ but $x+u(x)$ is not a pixel in the source image

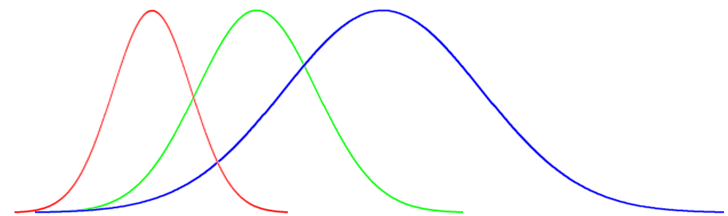
So interpolate $I_S(x+u(x))$ by considering neighboring pixels

Classification of Deformation Models

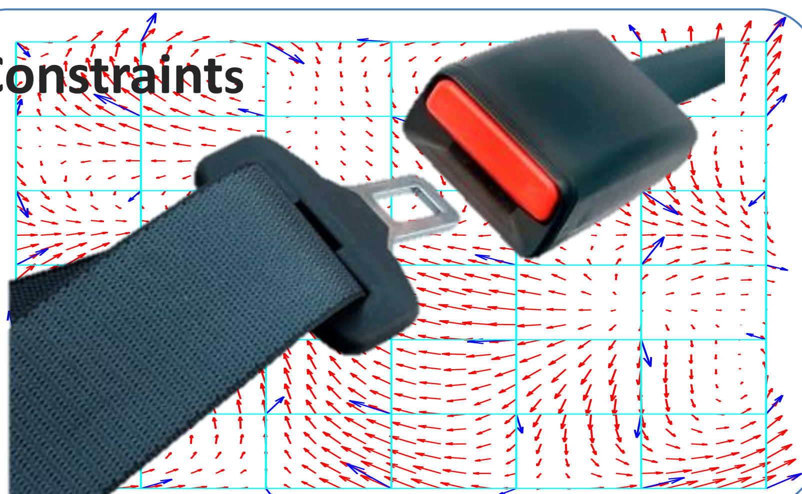
Physical models



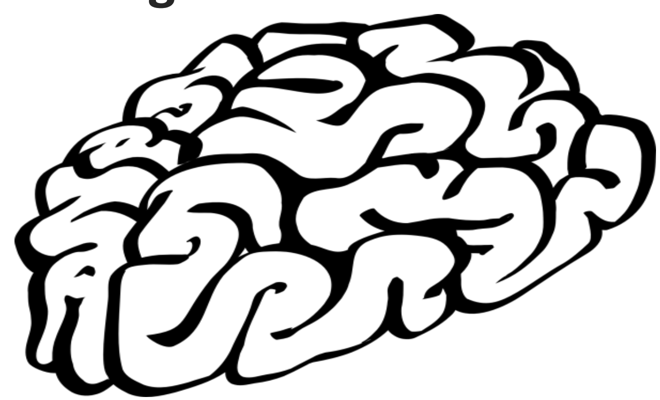
Basis function expansions



Constraints

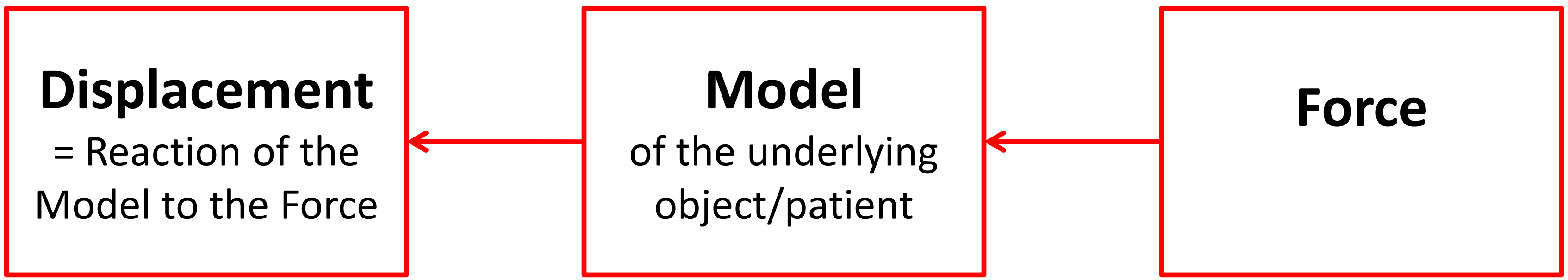


Knowledge-based



Physical Models

- Simulation point of view

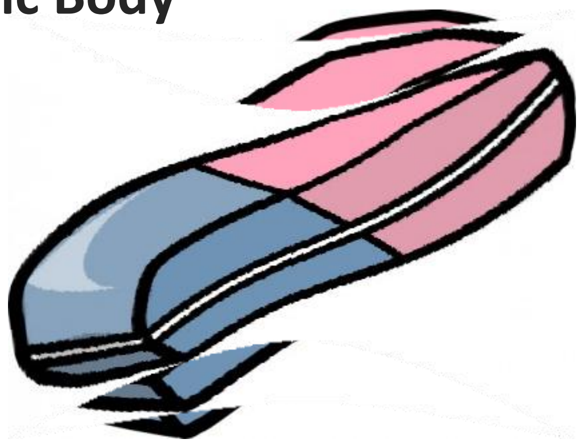


$$u_{t+\tau} = \left(\mathrm{Id} + \tau\lambda\nabla E_R\right)^{-1} \underbrace{\left(-\tau\nabla E_{\mathrm{D}}(u_t)\right)}_{f}$$

- Variety of Forces:
  - Image similarity
  - Distances between landmarks
- Variety of Models
  - Finite Element
  - Mass Spring

Physical Models

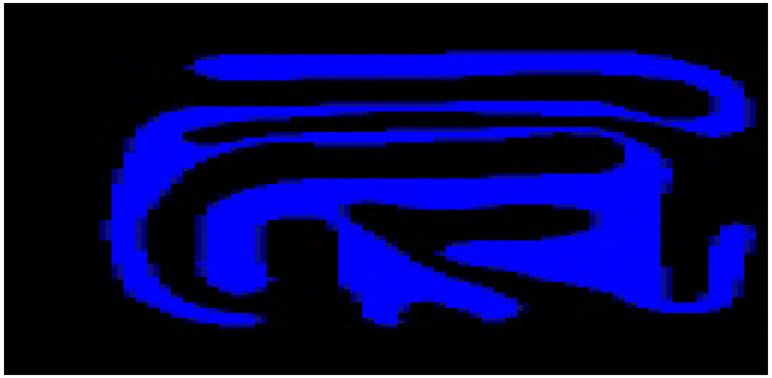
Elastic Body



Viscous Fluid

Diffusion

Curvature

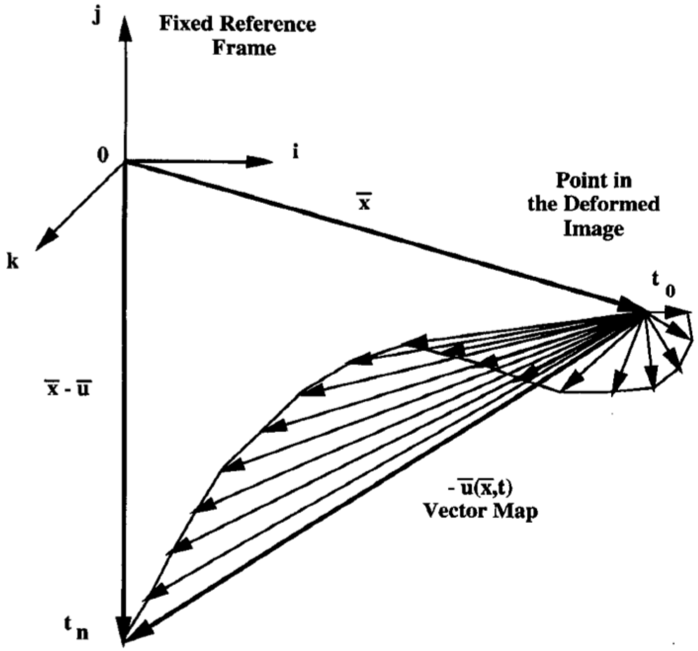


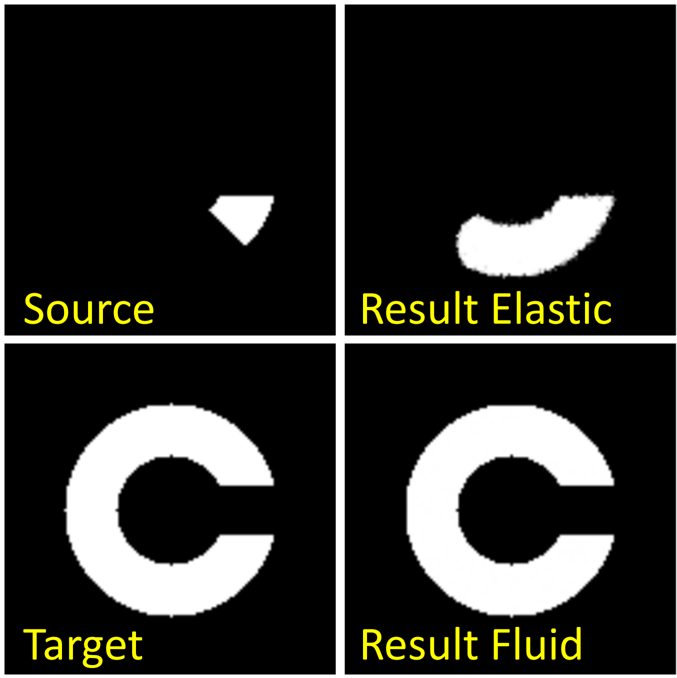
Physical Models

Viscous fluid flow

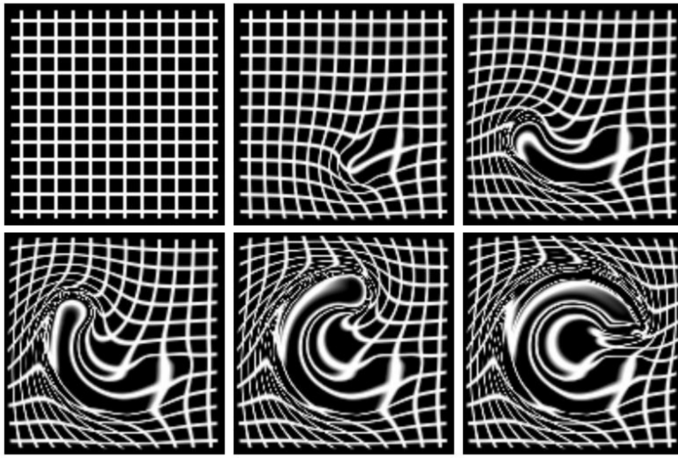
- Navier-Stokes PDE:

$$\mu_f \nabla^2 \mathbf{v} + (\mu_f + \lambda_f)\nabla(\nabla \cdot \mathbf{v}) + \mathbf{F} = 0$$





Images from [Christensen 94]



"Time progression of the fluid trans-formation applied to a rectangular grid"

Physical Models

Viscous fluid flow

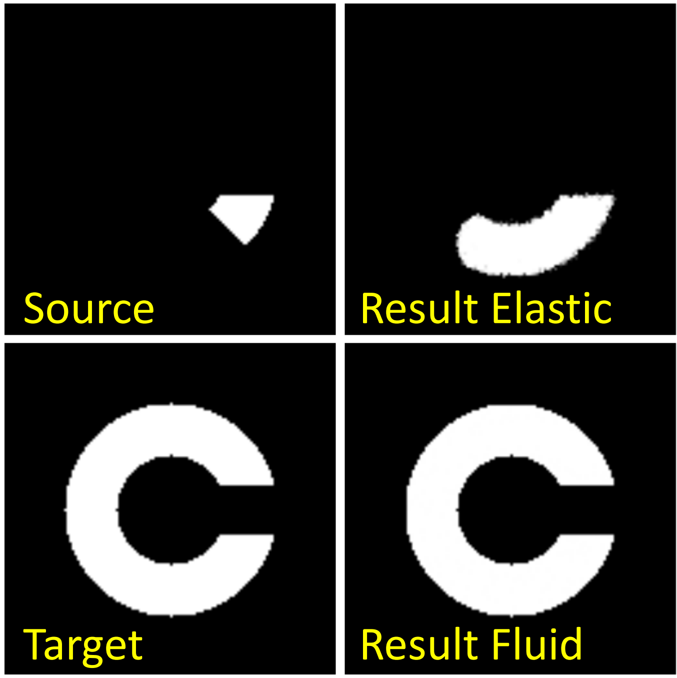
- Navier-Stokes PDE:

$$\mu_f \nabla^2 \mathbf{v} + (\mu_f + \lambda_f)\nabla(\nabla \cdot \mathbf{v}) + \mathbf{F} = 0$$

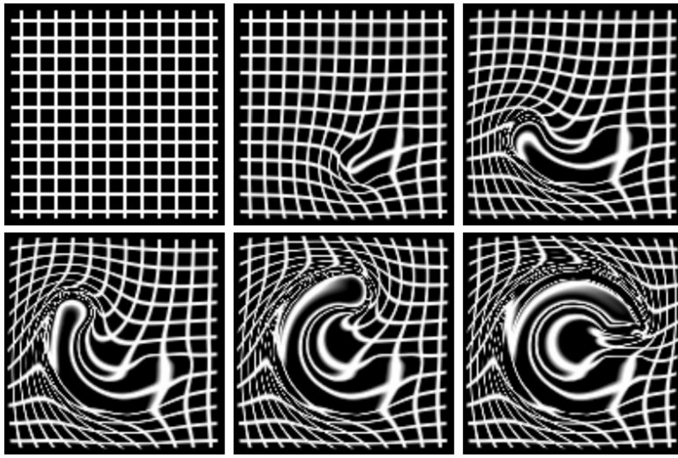
Fluid type registration = regularization of updates
(UPDATES = change of displacement = VELOCITIES)

Challenges

- Avoid folding of field
- No transport in homogeneous regions



Images from [Christensen 94]



“Time progression of the fluid trans-formation applied to a rectangular grid“

Physical Models

Curvature

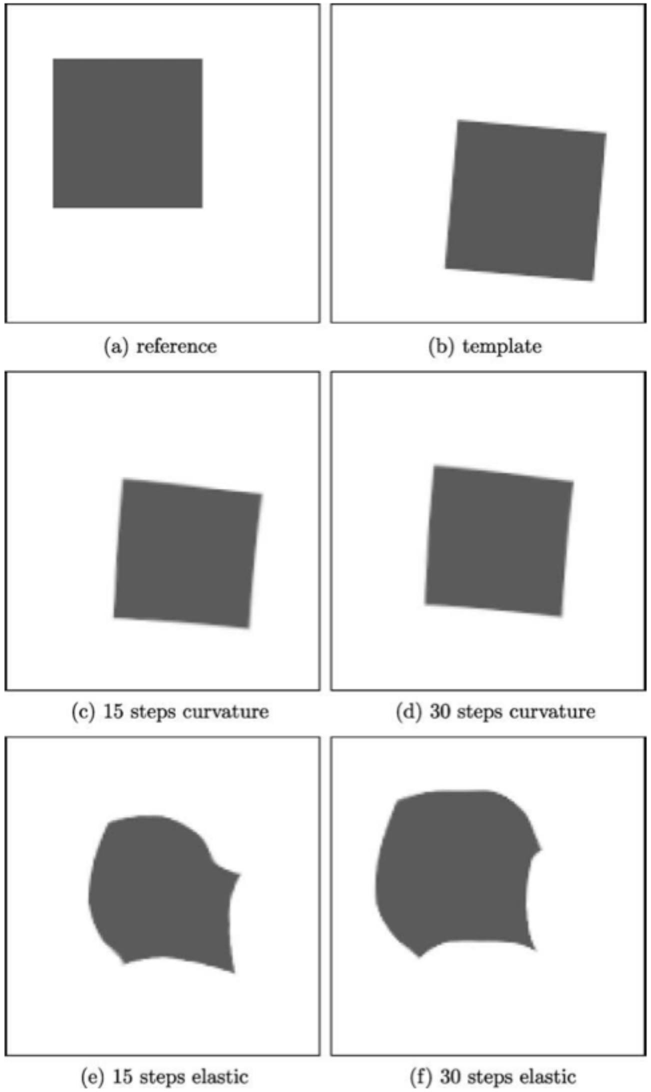
- Differential equation

$$\Delta^2 \mathbf{u} + \mathbf{F} = 0.$$

$$E_R = \sum_{i=1}^{d} \int_{\Omega} (\Delta u_i)^2 dx$$

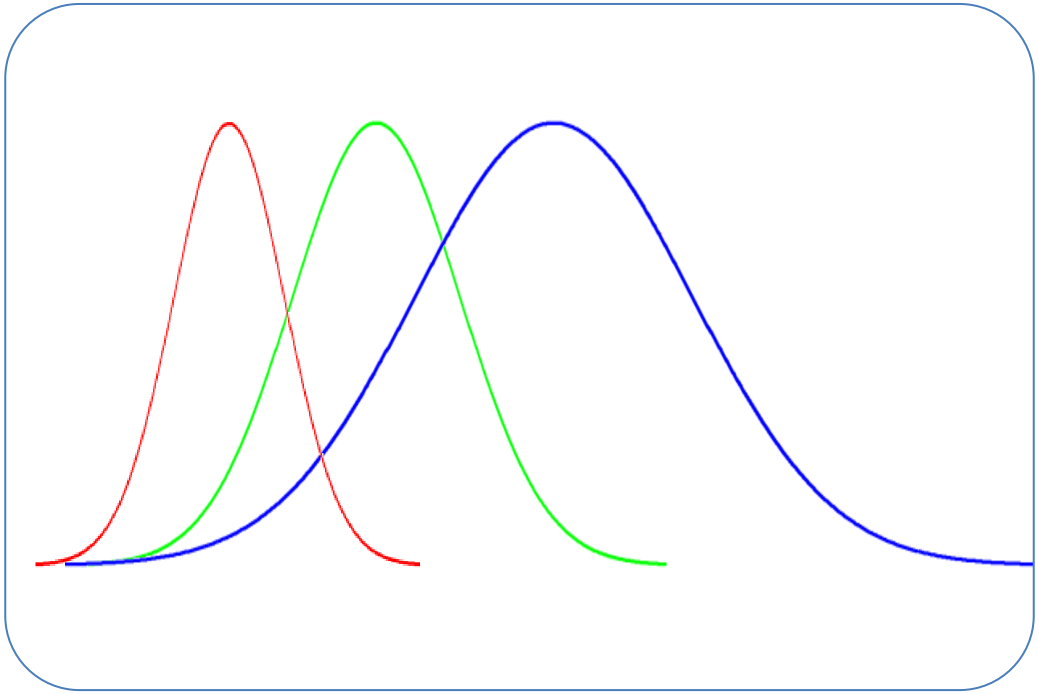
Features

- Does not penalize affine linear transformations
- Affine pre-registration may not be necessary



Images from [Fisher and Modersitzki 04]

Basis Function Expansion



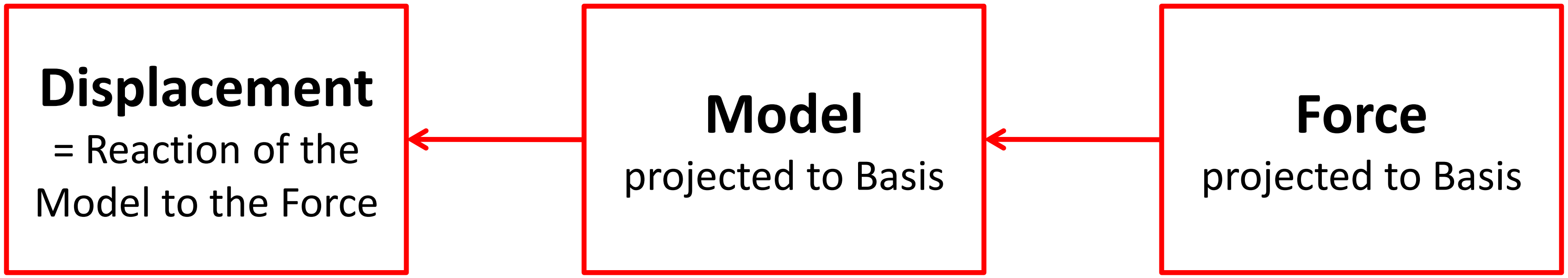
Basis Function Expansion

- Motivated from function interpolation and approximation theory
- Transformation as a linear basis expansion in R^d, where B_k is a basis function

$$u_p(x) = \sum_k p_k B_k(x)$$

- Few degrees of freedom
  - Efficiency
- Implicit smoothness of the field

Basis function expansion + Model

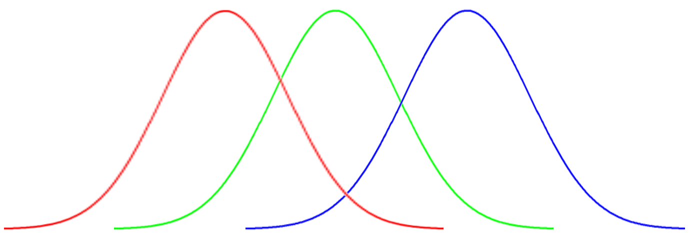


$$u_{t+\tau} = (Id + \tau\lambda\nabla_p E_R)^{-1} \; -\tau\nabla_p E_D(u_t(p))$$

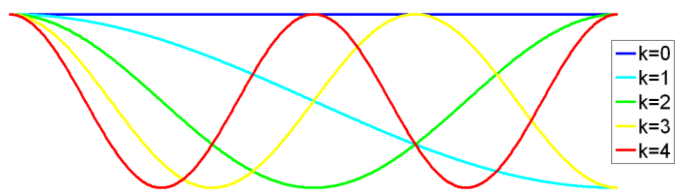
- The model is also projected to the basis
  - Smaller system
  - May result in a simplification of the problem

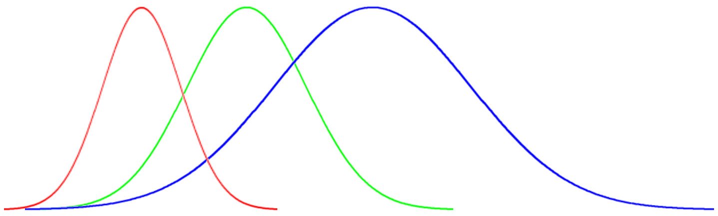
Basis function expansion – **Shape** of B_k's

- Same shape of all B_k's
  B_k's are translated versions of B:
  $B_k(x)=B(x-c_k)$
  - Free-form deformation (FFD) B-Splines



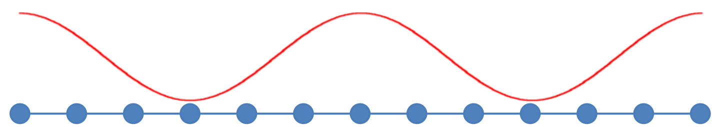
- Different shape of B_k's
  - Fourier/Cosine Bases
  - RBFs with different parameters (e.g. Gaussians with different variance)



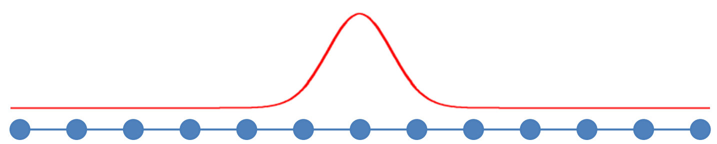


Basis function expansion – **Support** of $\boldsymbol{B_k}$'s

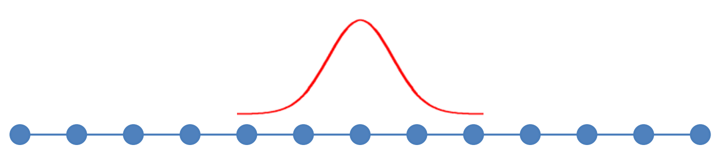
- Global Support
  - Fourier/Cosine Bases



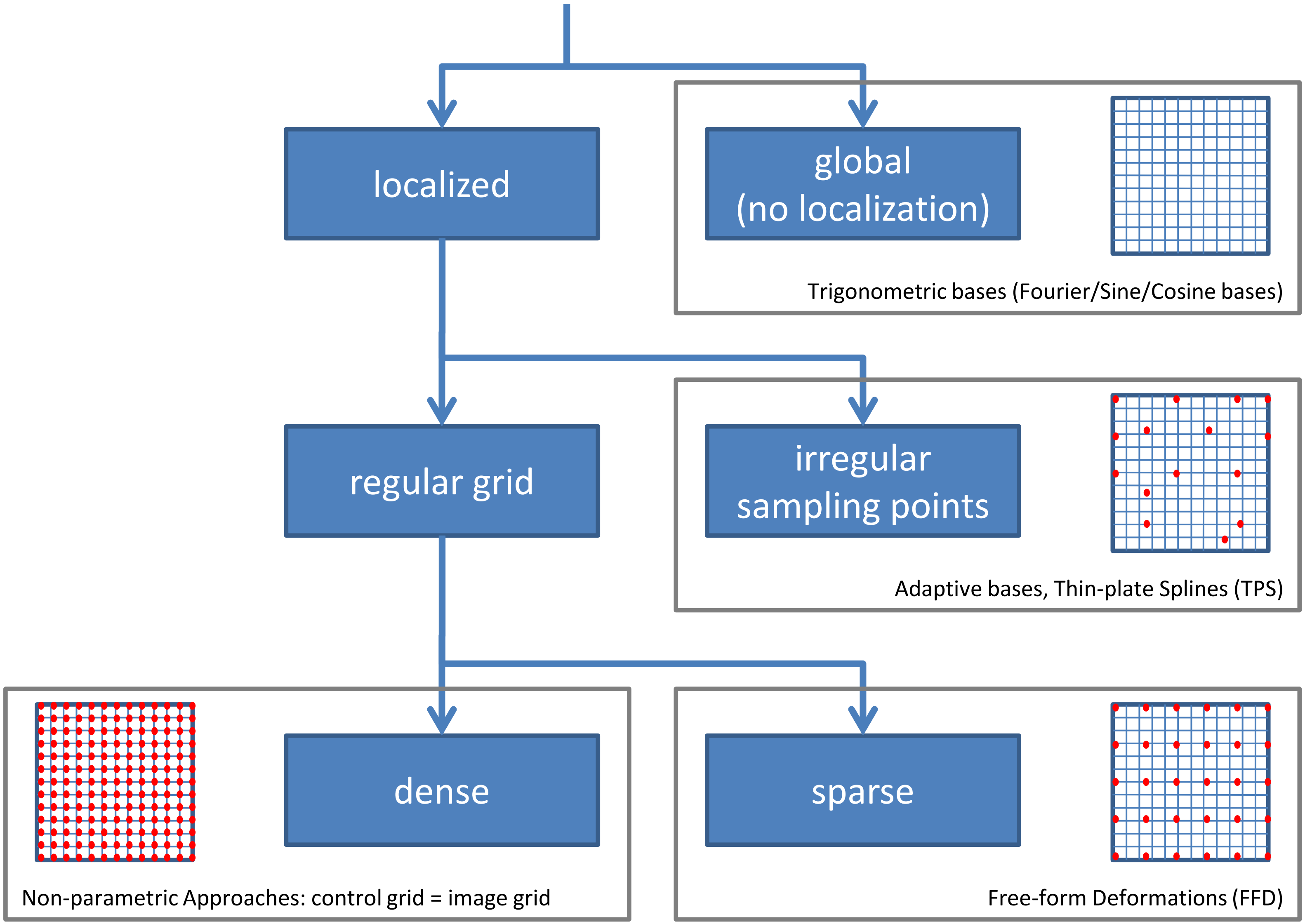
  - Radial basis functions RBFs (*e.g.,* Thin-plate Splines (TPS))
  - Gaussians (in theory)



- Compact Support
  - B-Splines
  - Some RBFs
  - Gaussians (in practice)

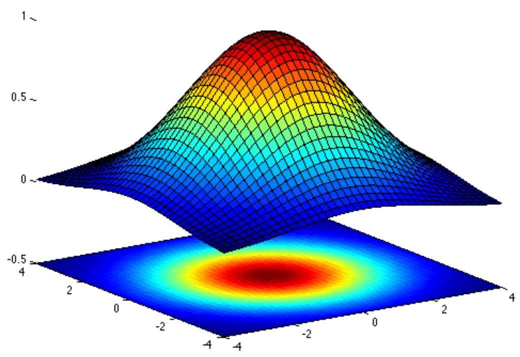


Basis function expansion – **Localization** of $\boldsymbol{B_k}$'s

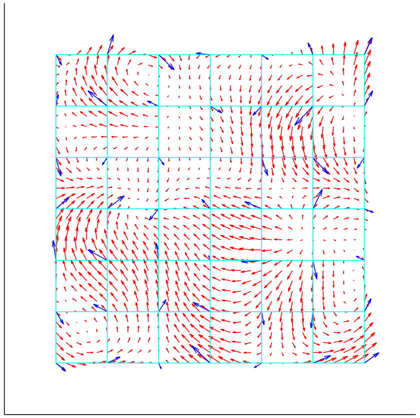


Basis Function Expansion

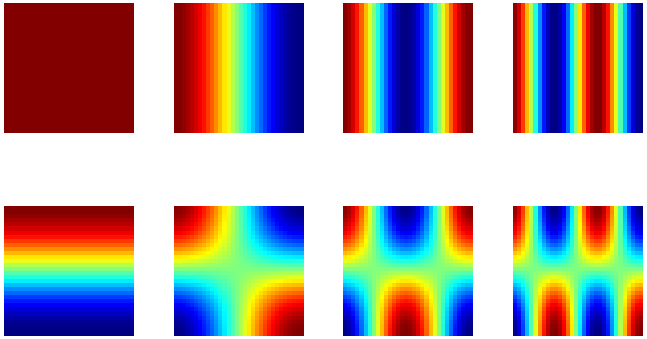
Radial Basis Functions



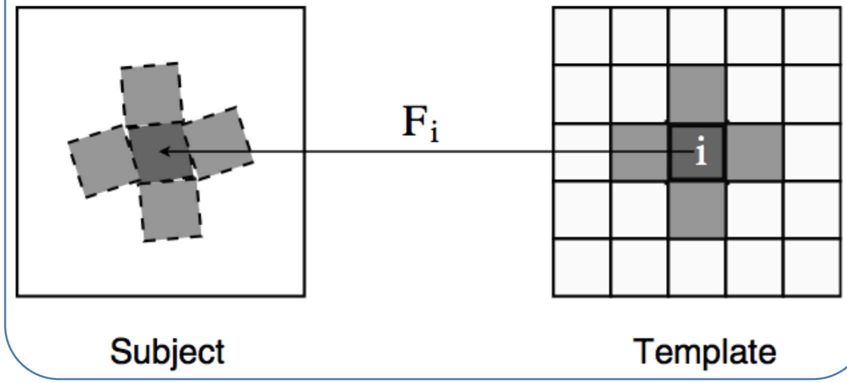
Free-Form Deformations



Basis from signal processing



Locally Affine Deformations



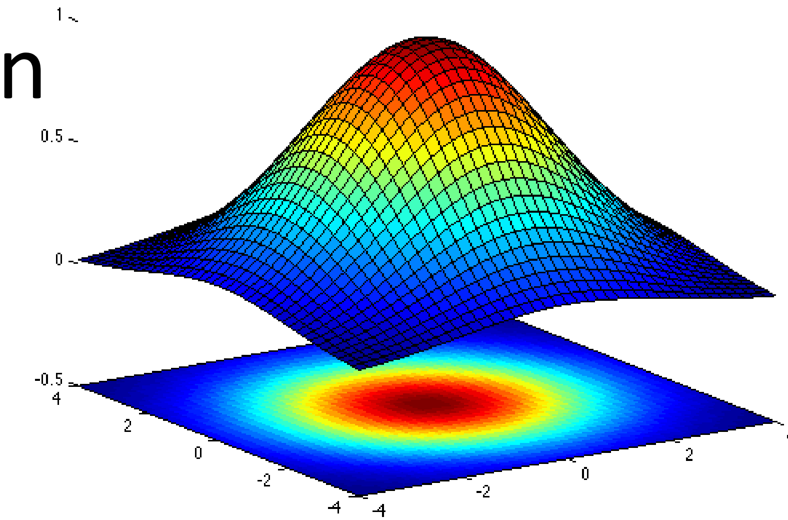
Basis Function Expansion

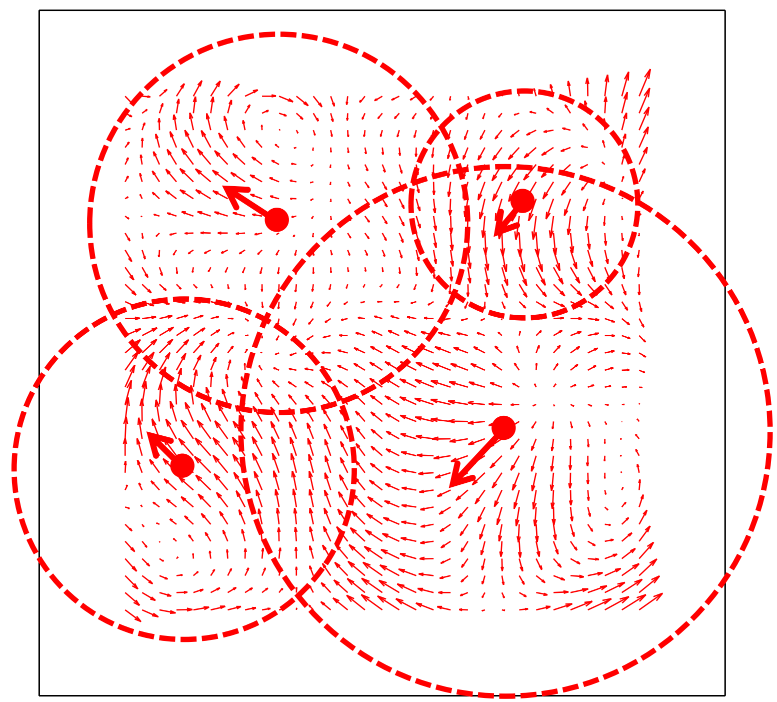
Radial basis functions

$$u(x) = \sum_k p_k B(\|x - x_k\|)$$

p_k : are estimated by solving set of linear equations

x_k : basis function center or landmark





Features

- Global support
- Tend asymptotically to zero
- Positive definite functions
  - ✓ Closed form solution
  - ✓ Solvable for all possible sets of landmarks that are not coplanar

Basis function expansion – Radial basis functions

Thin Plate Splines [Bookstein 91]

Interpolating splines : $u(x_k) = q_k$

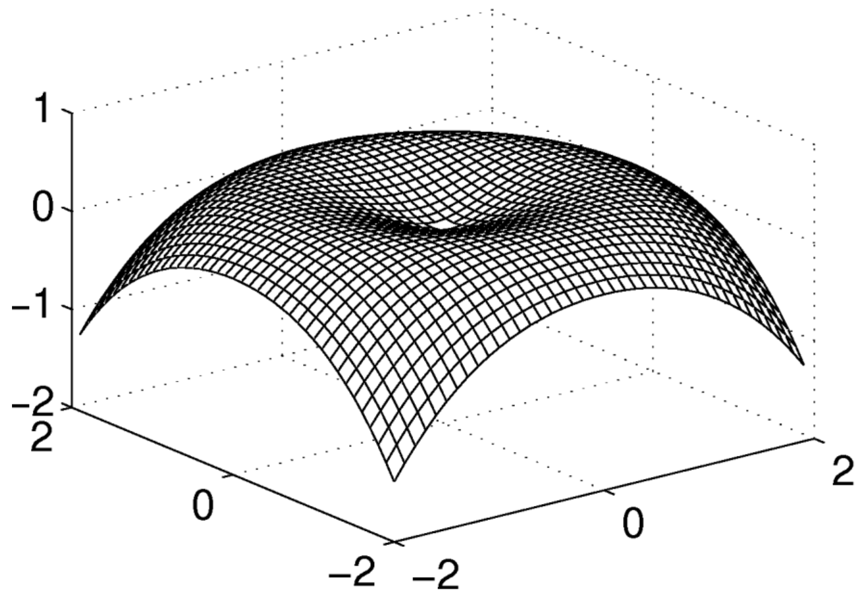
$$u(x) = Ax + B + \sum_k p_k B(\|x - x_k\|)$$

A and B define an affine transformation

In 2D, $B(r) = -r^2 \ln r^2$

Features

- ✓ Minimize bending energy
- ✓ Arbitrary landmark positions
- ✖ Global support
  - ➢ Important number of landmarks to recover local deformations
- ✖ Not topology preserving
- ✖ High computational demands when number of landmarks increase



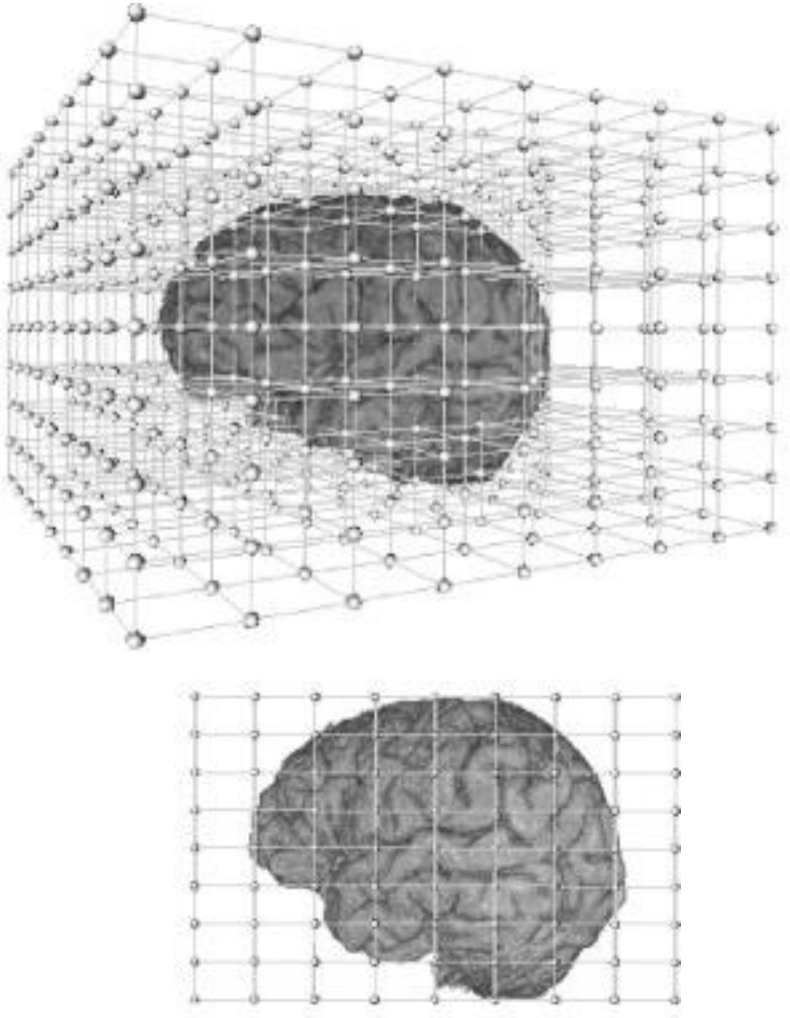
$$E = \iint \left(\left(\frac{\partial^2 u}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 u}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2 u}{\partial y^2}\right)^2 \right) dx dy$$

$$\underbrace{\qquad\qquad}_{E_B}$$

Basis function expansion

Cubic B-Splines Free-Form Deformation (FFD)

- Computer graphics technique for 3D object modeling [Sederberg 86]
- Parameterization by a grid of control points
- Object is deformed by manipulating the control points



A dataset is initially embedded in a uniform lattice of control points: (top) 3D view and (bottom) parallel projection. (Images from [Merhof 07])

Basis function expansion

Cubic B-Splines Free-Form Deformation (FFD)

Tensor product of corresponding 1-D cubic *B*-splines

$$u(x) = \sum_{l=0}^{3}\sum_{m=0}^{3} B_l(\mu_x) B_m(\mu_y) d_{i+l,j+m}$$

where

$$\mu_x = x/\delta - \lfloor x/\delta \rfloor,\ \mu_y = y/\delta - \lfloor y/\delta \rfloor,$$

$$i = \lfloor x/\delta \rfloor - 1,\ j = \lfloor y/\delta \rfloor - 1$$

and

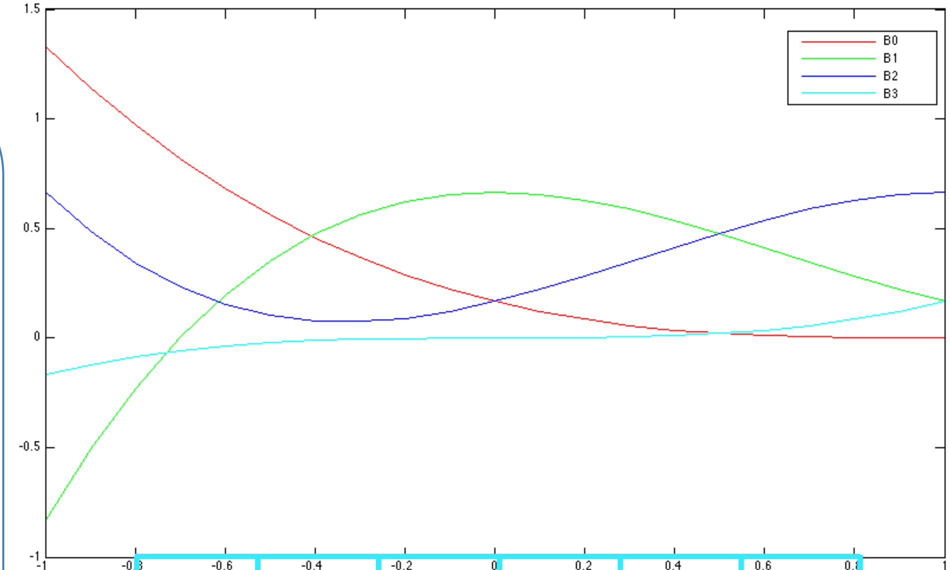
$$B_0(s) = \frac{1}{6}(1-s)^3$$

$$B_1(s) = \left(3s^3 - 6s^2 + 4\right)/6$$

$$B_2(s) = \left(-3s^3 + 3s^2 + 3s + 1\right)/6$$

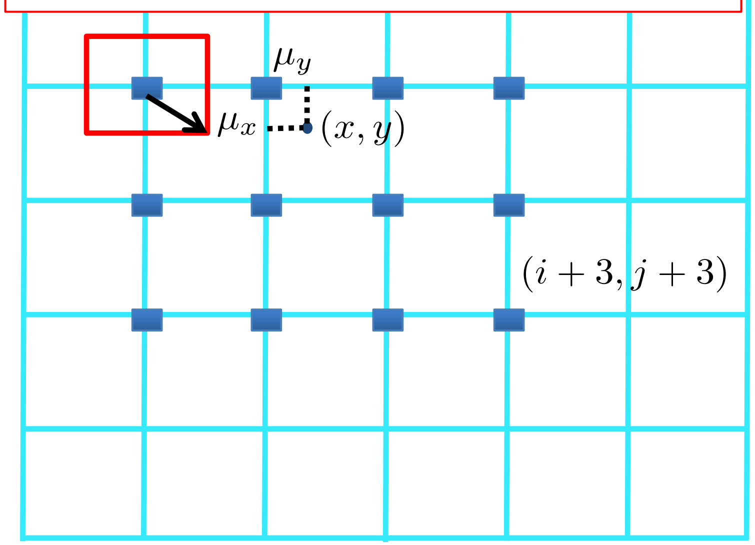
$$B_3(s) = s^3/6$$

B-spline basis functions



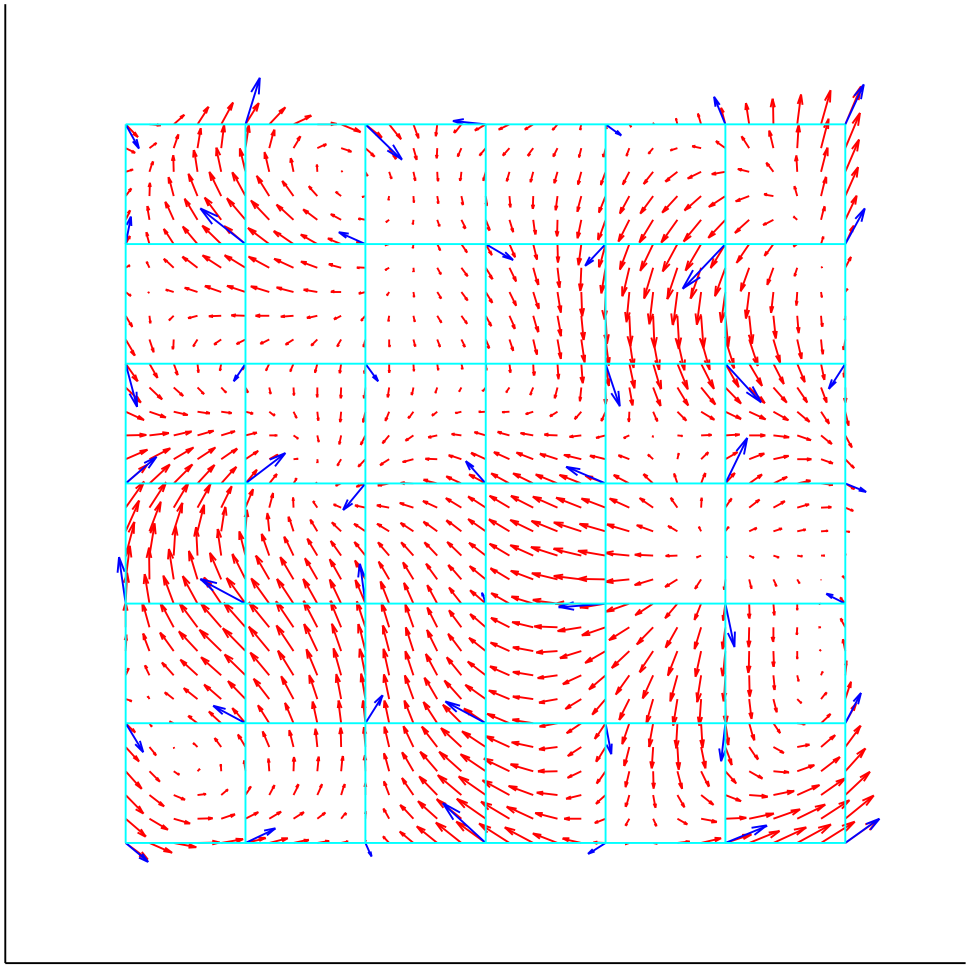
Max displacemt < 0.4 Grid spacing

➔ diffeomorphic transformation



Basis function expansion

Cubic B-Splines Free-Form Deformation (FFD) - Example



Basis function expansion

Wavelet: [Amit 94, Wu 00, Geffen 03]

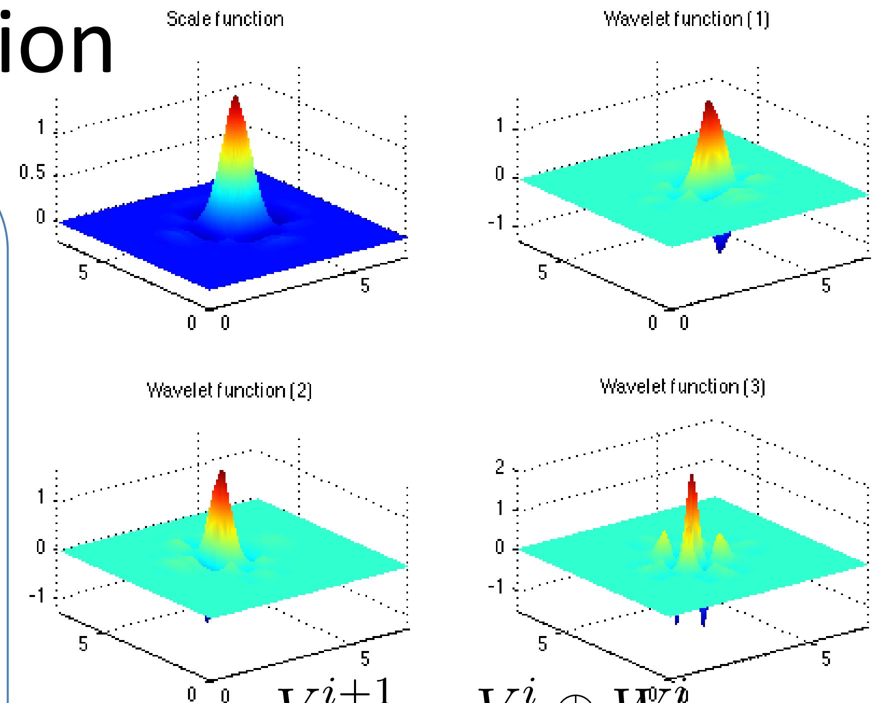
$$u(x) = \sum_{n,k,l} \langle u(x), \phi_{n,k,l} \rangle \phi_{n,k,l} + \sum_{i} \sum_{n,k,l} \langle u(x), \psi^{i}_{n,k,l} \rangle \psi^{i}_{n,k,l}$$

where i = {H, V, D} and for separable scaling and wavelet functions:

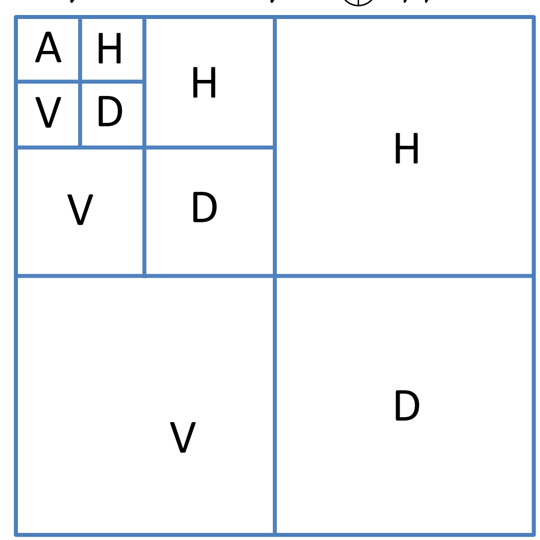
$$\phi_{n,k,l} = \phi_{n,k}(x)\phi_{n,l}(y)$$
$$\psi^{H}_{n,k,l} = \psi_{n,k}(x)\phi_{n,l}(y)$$
$$\psi^{V}_{n,k,l} = \phi_{n,k}(x)\psi_{n,l}(y)$$
$$\psi^{D}_{n,k,l} = \psi_{n,k}(x)\psi_{n,l}(y)$$

<u>Features</u>

1. Local support
   - Recover local changes



$$V^{i+1} = V^{i} \oplus W^{i}$$



Basis function expansion

Locally affine [Collins 97, Hellier 01, Pitiot 06, Zhang 06, Commowick 08]

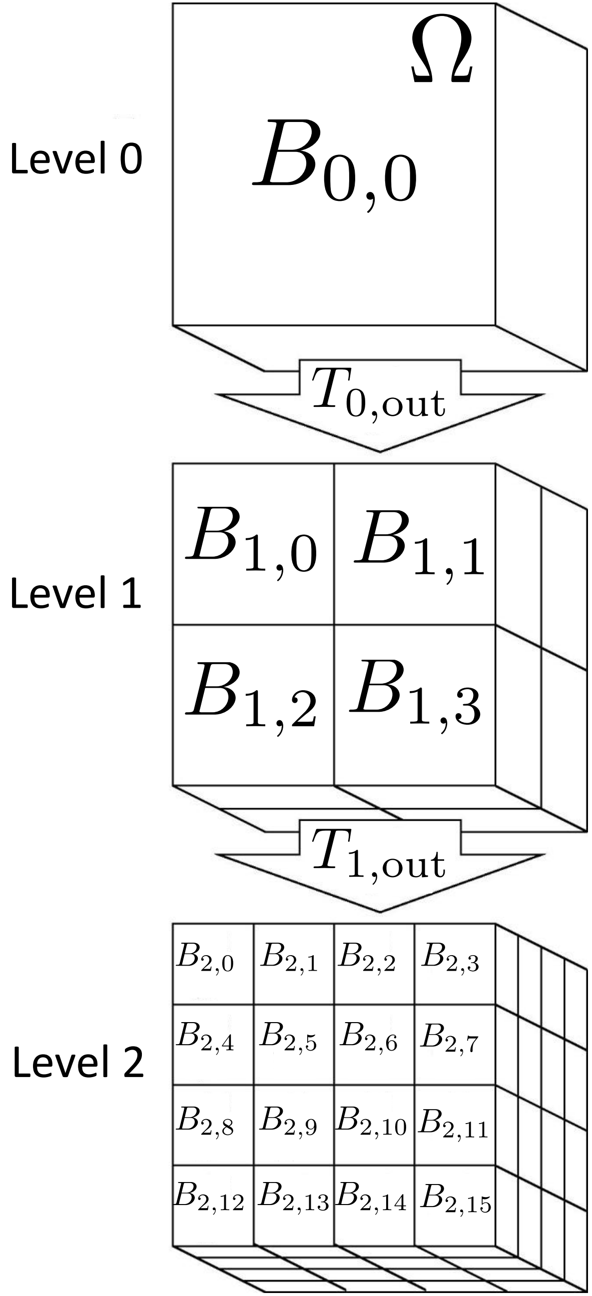
$$u(x) = \sum_k p_k A_k(x)$$

1. Partition image to triangles or tetrahedra
2. Nodes are parameters of transformation

Features

✓ Efficiency

✖ Lack of smoothness in regions boundaries

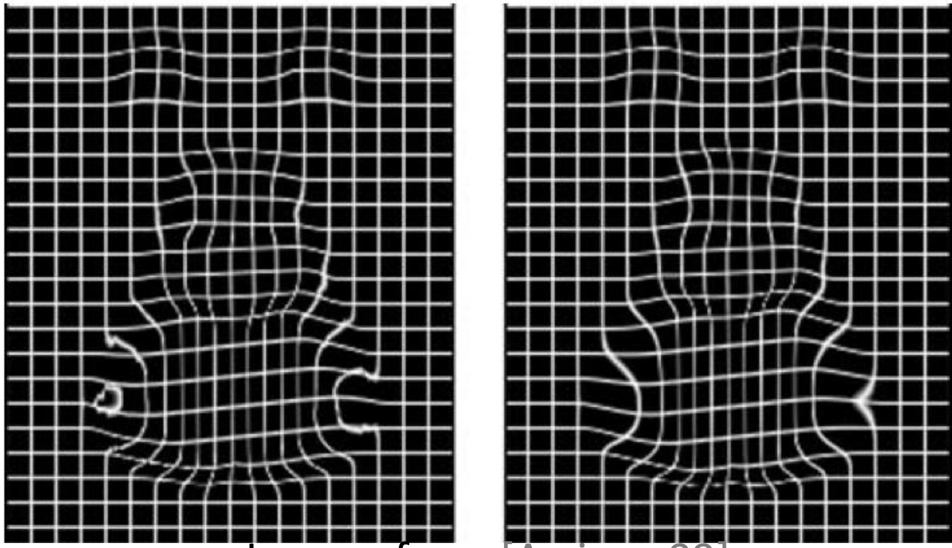


Images from [Buerger 11]

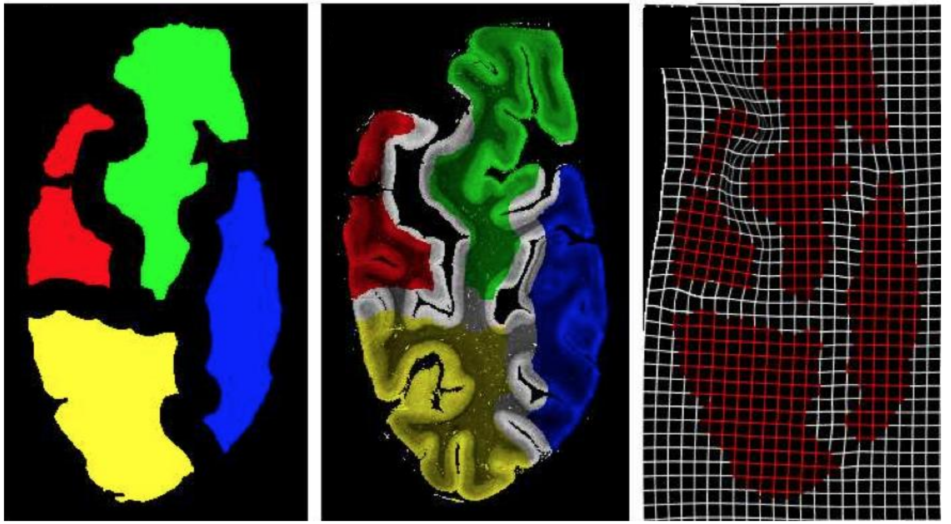
Basis function expansion

Locally affine

- Direct fusion efficient but not invertible in general
- Non overlapping parts – hybrid affine/non-linear interpolation scheme [Pitiot 06]
- Poly-affine [Arsigny 09]
  - ODEs
  - Diffeomorphic



Images from [Arsigny 09]

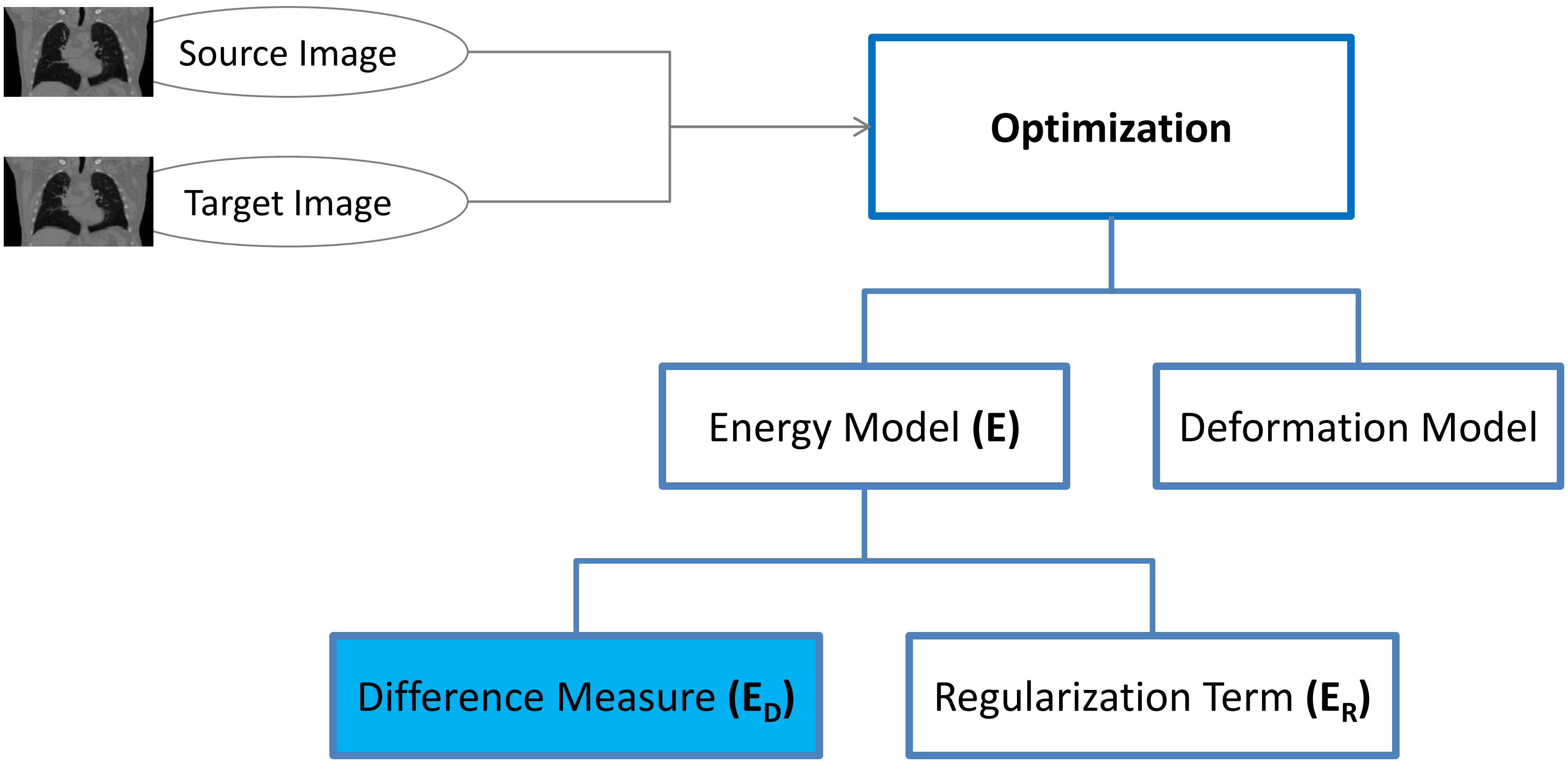


Images from [Pitiot 06]

PART II

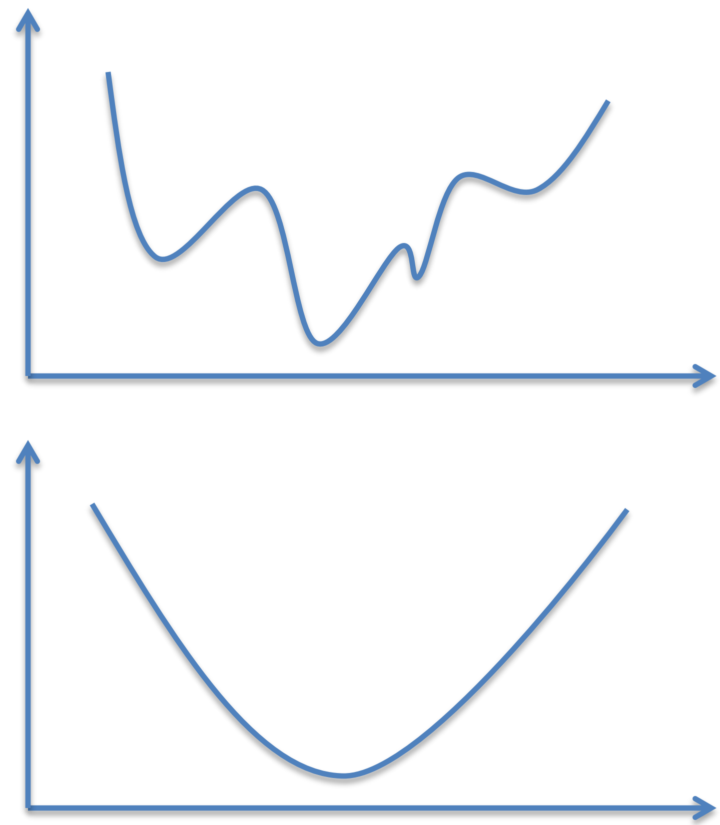
Similarity Metrics

Deformable Registration: General Framework



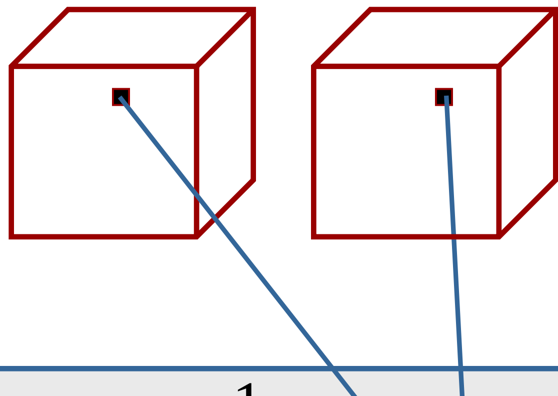
Requirements on similarity measure

- Extremum for correctly aligned images
- Smooth, best convex
- Differentiable
- Fast computation



Difference Measures

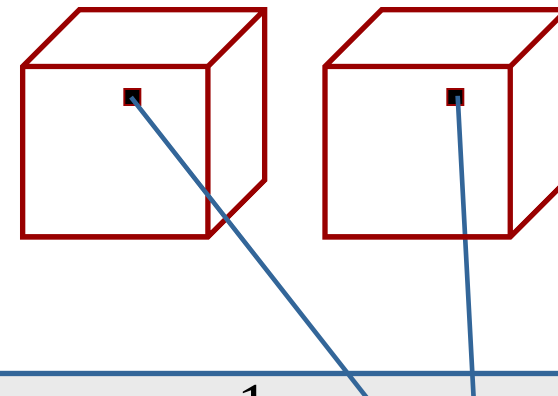
Volume X Volume Y($\underline{\phi}$)



$$SSD = \frac{1}{N}\sum_i (x_i - y_i)^2$$

Sum of Squared Differences

Volume X Volume Y($\underline{\phi}$)

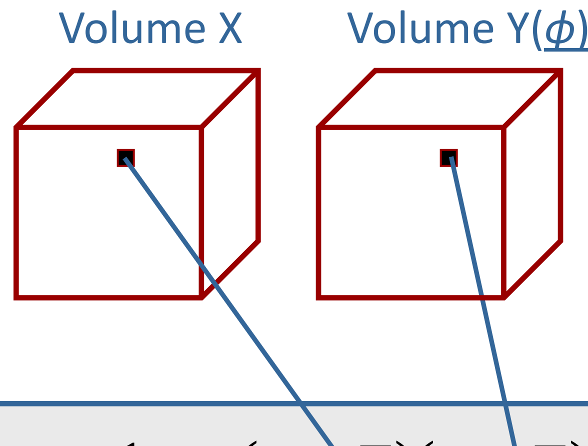


$$SAD = \frac{1}{N}\sum_i |x_i - y_i|$$

Sum of Absolute Differences:

Less sensitive on large intensity differences than SSD

Normalized Cross Correlaiton (NCC)



$\bar{x}$: Mean

σ_x : Standard deviation

N : Number of pixels

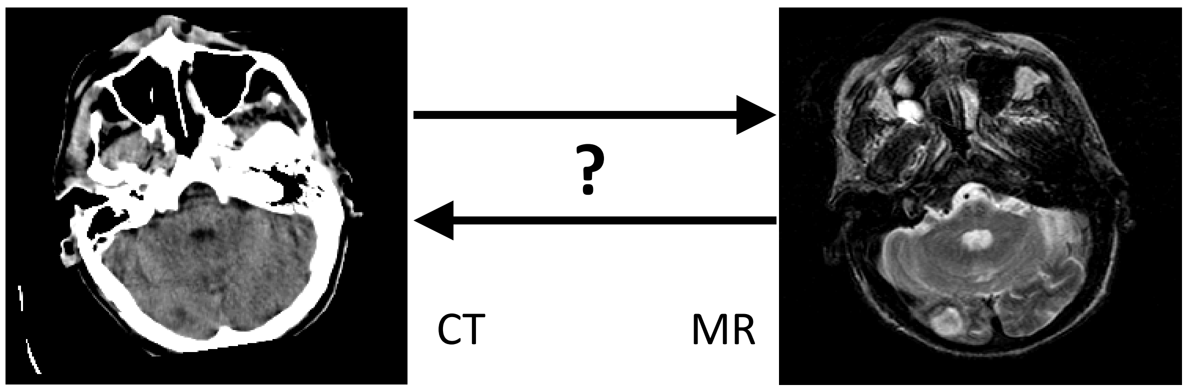
$$NCC = \frac{1}{N}\sum_i \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y}$$

Normalized Cross Correlation:

Expresses the linear relationship between voxel intensities in the two volumes

Multi-Modal Registration

- More complex intensity relationship



- Approaches:
  - Simulate one modality from the other one
  - Apply sophisticated similarity measures

Information Theoretic Approach

Histogram calculation

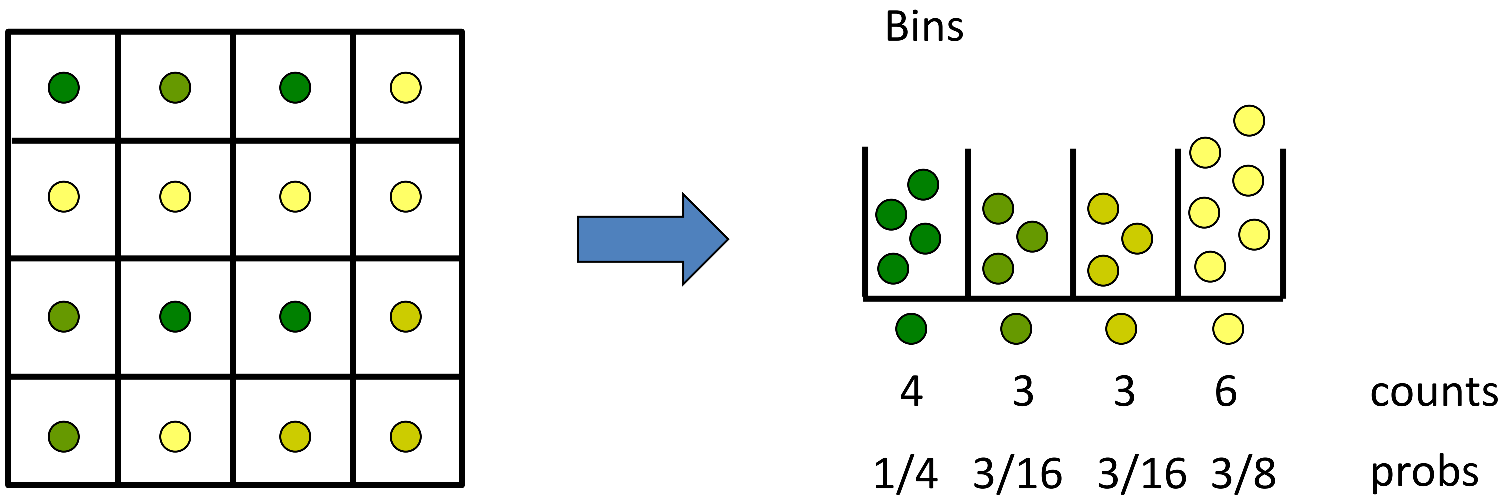
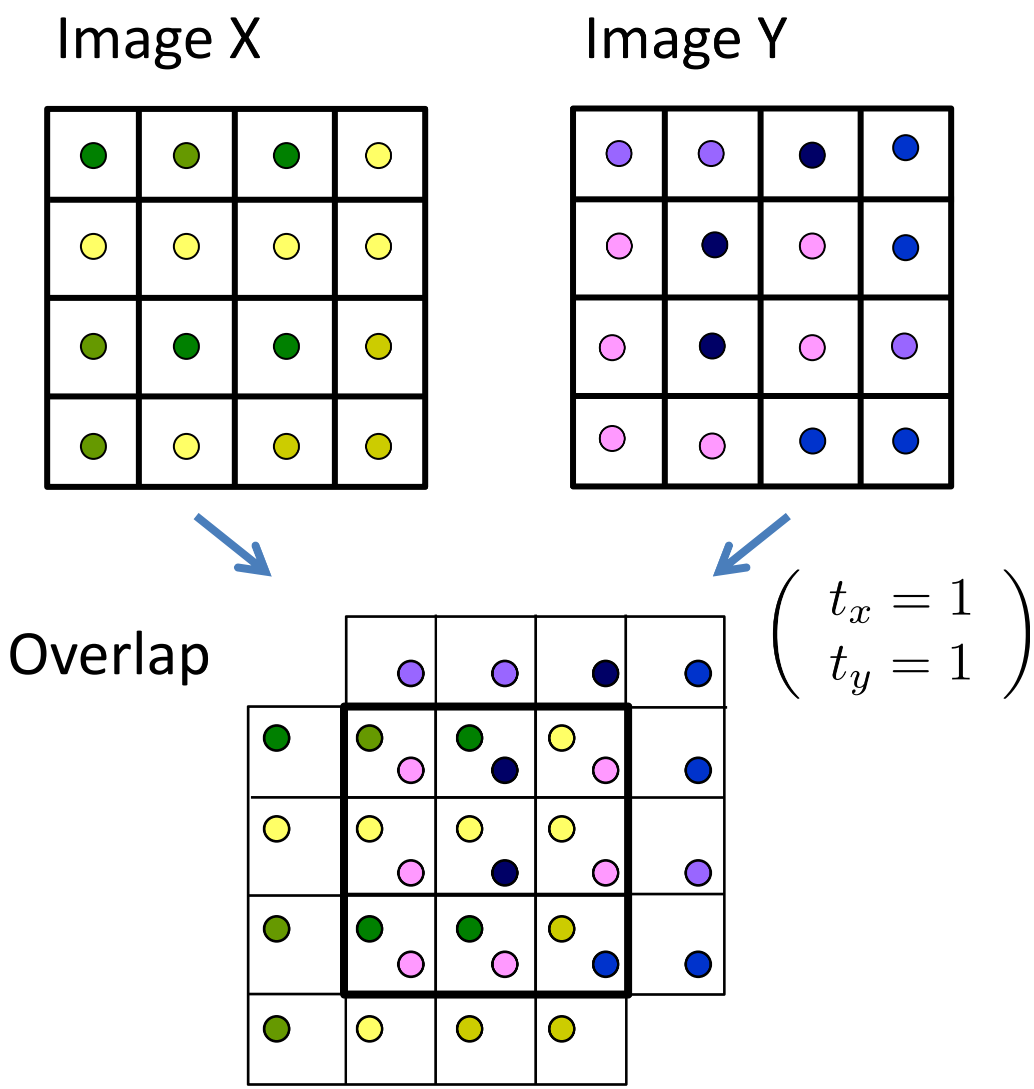
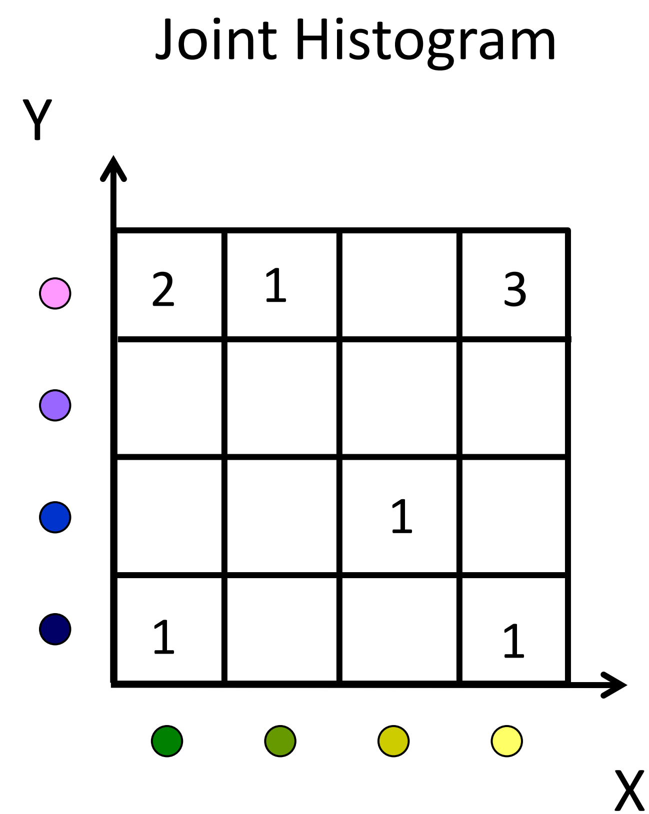


Image Histogram

Joint histogram calculation

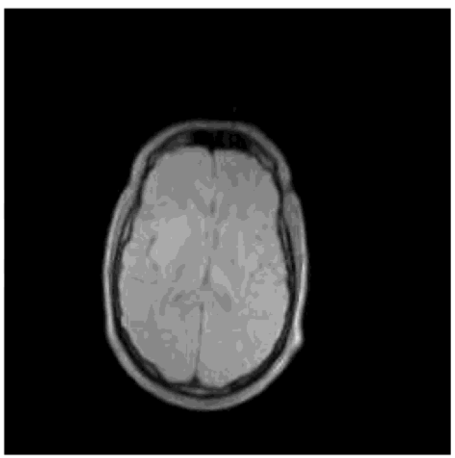




Joint Histogram

- Histogram for images from different modalities

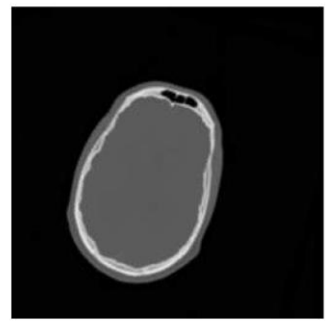
Source Image

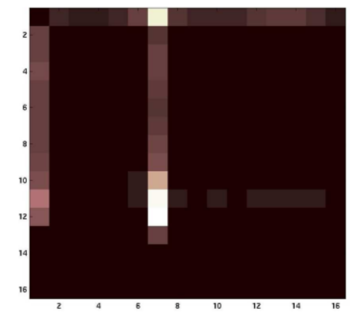


Target Image

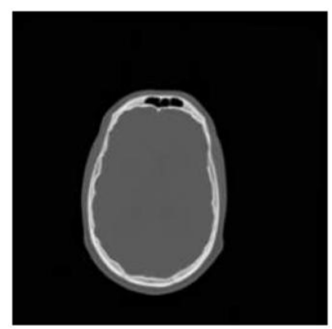
Joint Histogram

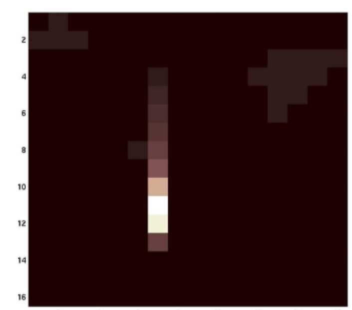
Not Aligned





Aligned



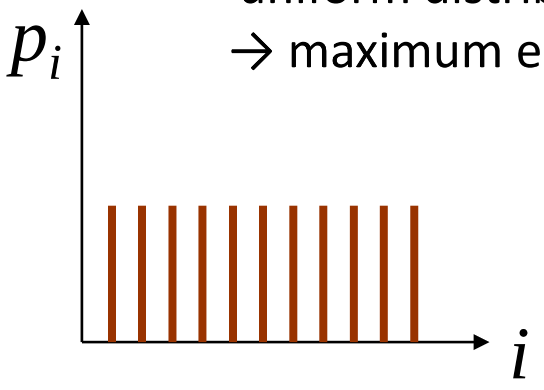


Entropy

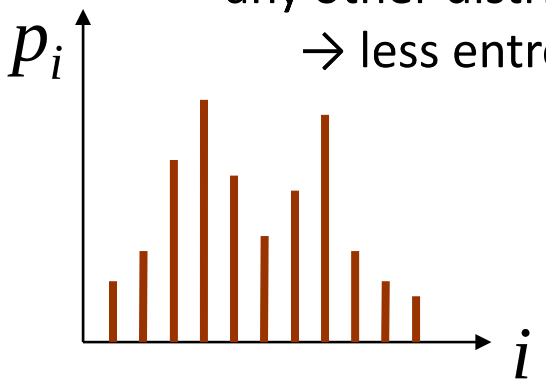
Shannon Entropy, developed in the 1940s
(communication theory)

$$H = -\sum_i p_i \log p_i$$

uniform distribution
→ maximum entropy



any other distribution
→ less entropy



Mutual Information (MI)

$$MI(X,Y) = H(X) + H(Y) - H(X,Y)$$
$$= \sum_i \sum_j p_{xy}(i,j) \log \frac{p_{xy}(i,j)}{p_x(i)\, p_y(j)}$$

- Maximized if X and Y are perfectly aligned
- *H(X)* and *H(Y)* help to make the measure more robust
- Maximization of mutual information leads to minimization of joint entropy

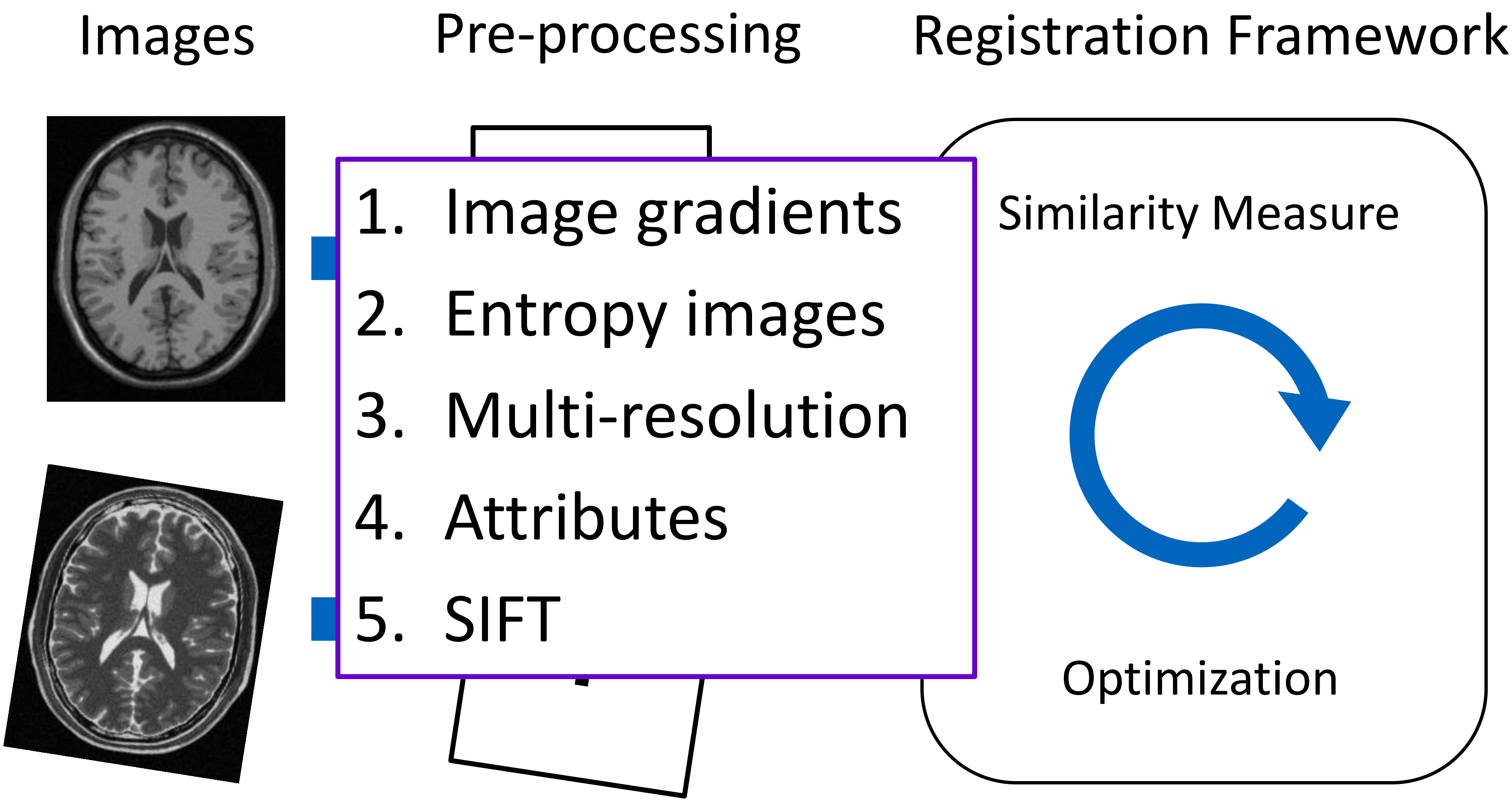
Historical Note

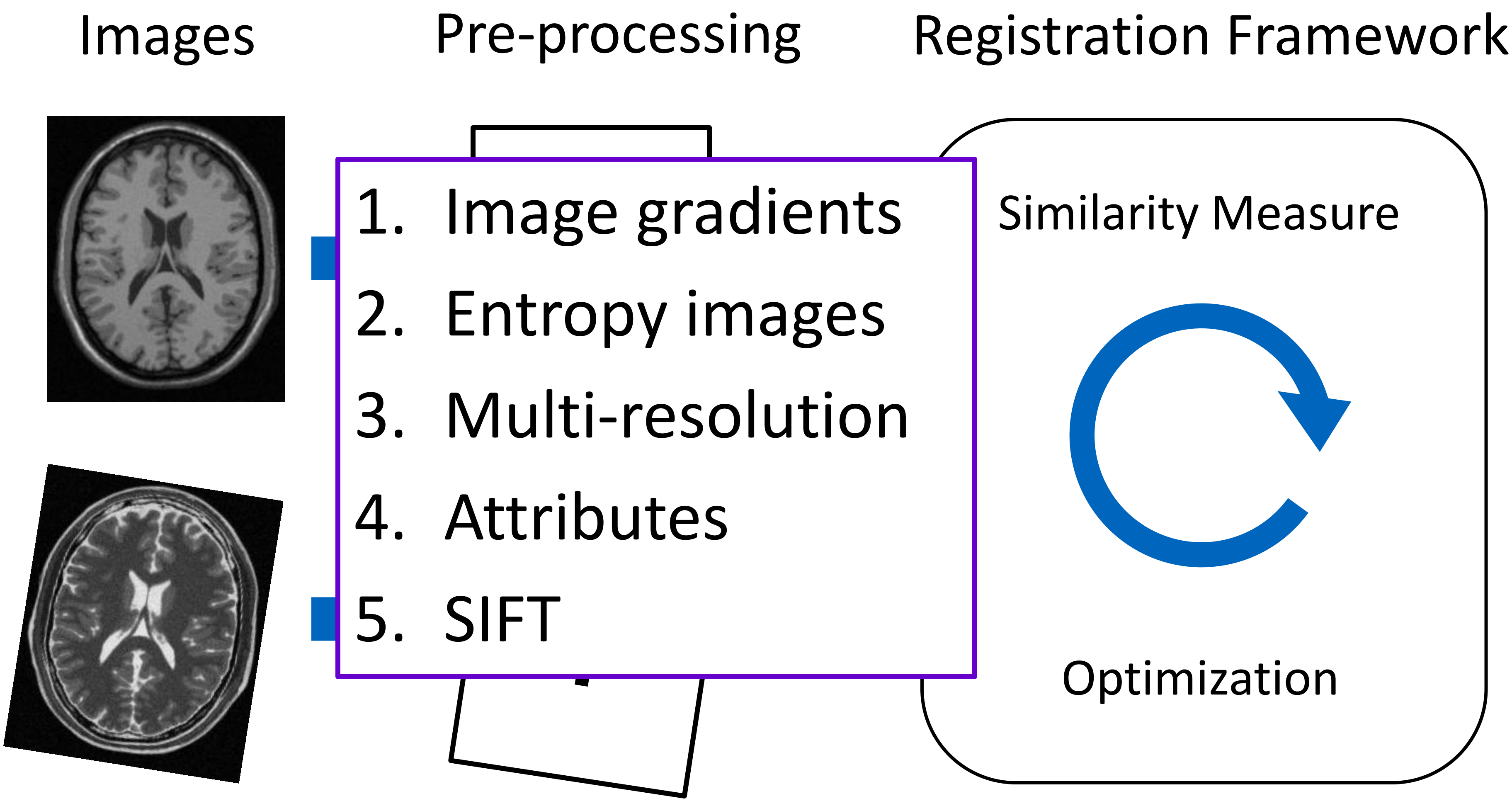
- Minimum Entropy Registration
  - Collignon A., Vandermeulen, D., Suetens, P., and Marchal, G. *3D multi-modality medical image registration using feature space clustering*. CVRMED April 1995.
- Maximum Mutual Information Registration
  - Viola, P. and Wells, W. *Alignment by maximization of mutual information*. In Proceedings of the 5th International Conference of Computer Vision, June 20 – 23, 1995.
  - Collignon A, Maes F, Delaere D, Vandermeulen D, Suetens P, Marchal G, *Automated multi-modality image registration based on information theory*. IPMI June 26, 1995.
  - Viola, P. *Alignment by maximization of Mutual Information*. MIT PhD Thesis, June 1995.

Improvements to MI

- Normalization of MI
- Density estimation
  - Parzen window
  - Partial volume distribution
  - Uniform volume histogram
  - NP windows
- Spatial information

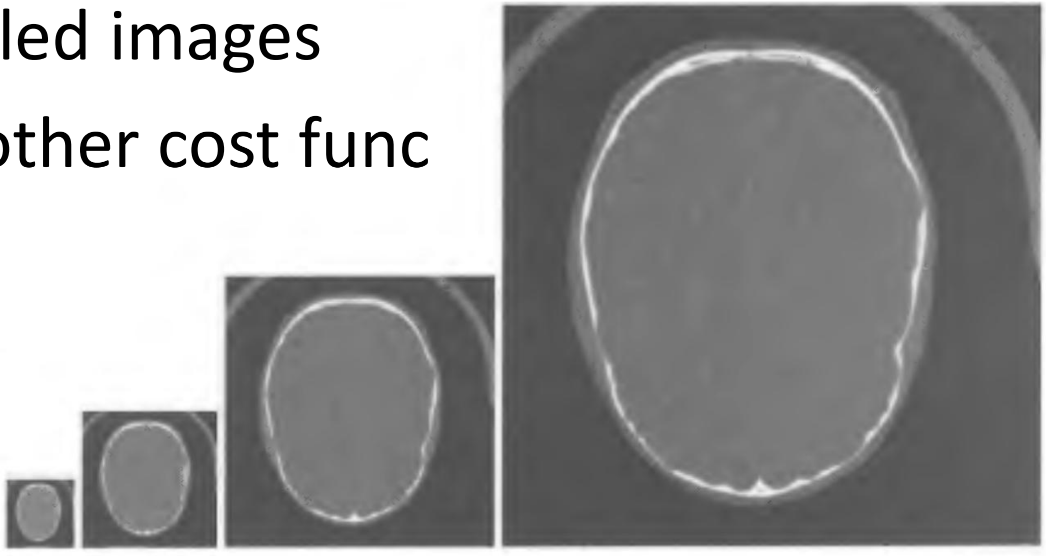
- Tutorial at MICCAI 2009:
  Information theoretic similarity measures for image registration and segmentation: Maes, Wells, Pluim
  http://ubimon.doc.ic.ac.uk/MICCAI09/a1882.html

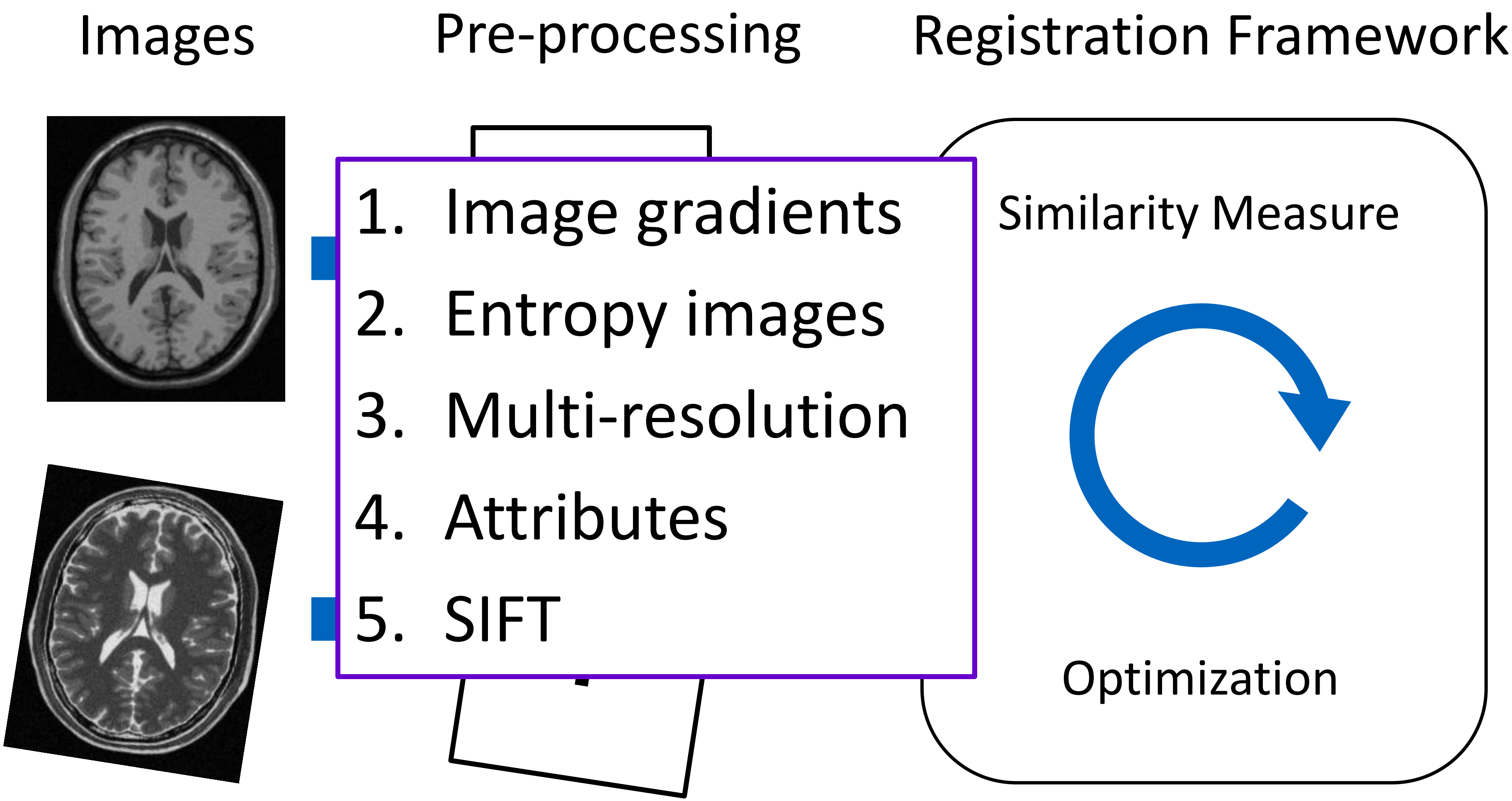

Images
Pre-processing
Registration Framework
1. Image gradients
2. Entropy images
3. Multi-resolution
4. Attributes
5. SIFT
Similarity Measure
Optimization

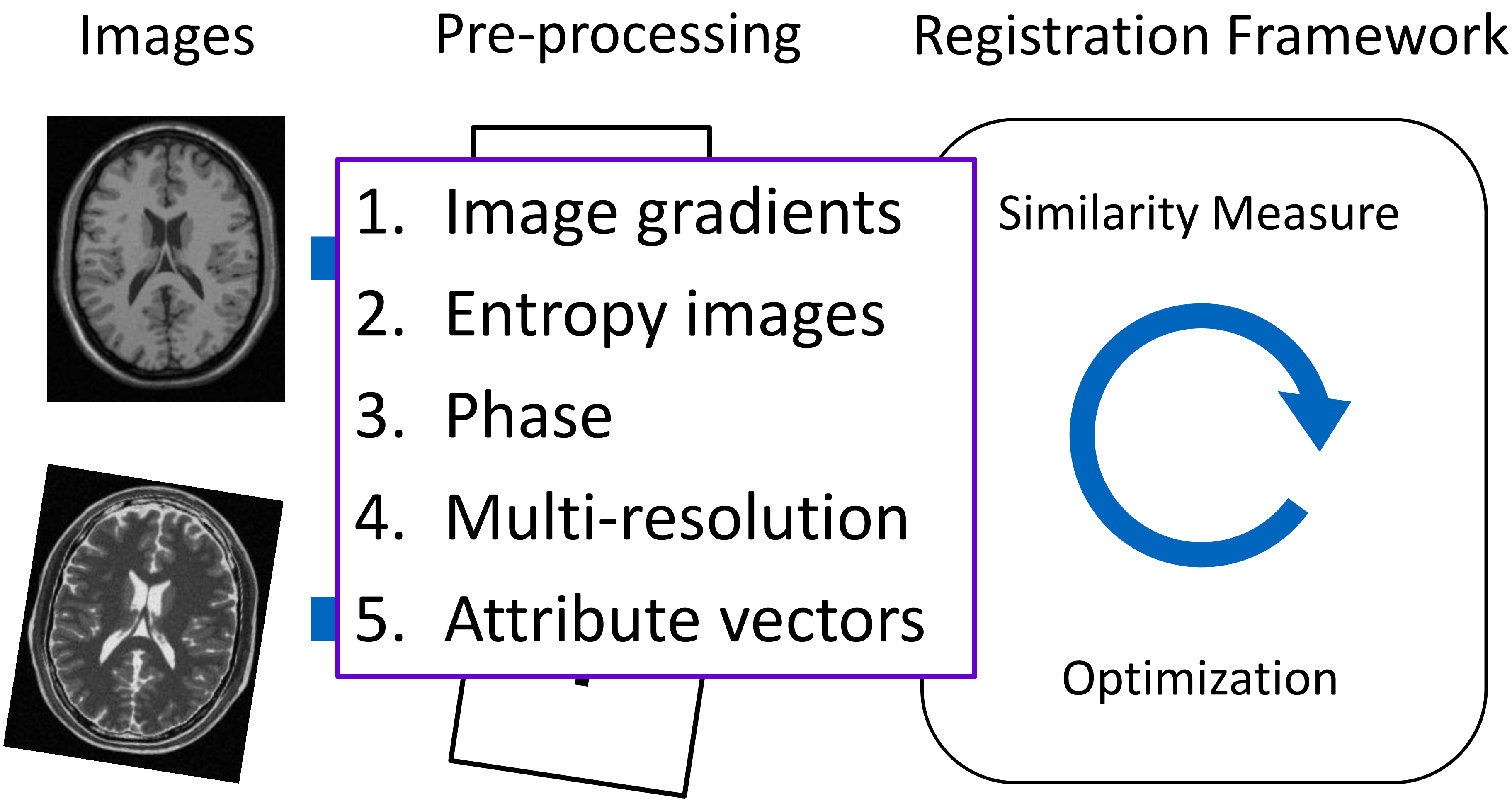

Images
Pre-processing
Registration Framework
1. Image gradients
2. Entropy images
3. Multi-resolution
4. Attributes
5. SIFT
Similarity Measure
Optimization

Multi-Resolution Registration

- Perform registration on multiple resolutions
  1. Smooth
  2. Downsample
- Advantages:
  - Speed: down-sampled images
  - Convergence: smoother cost func



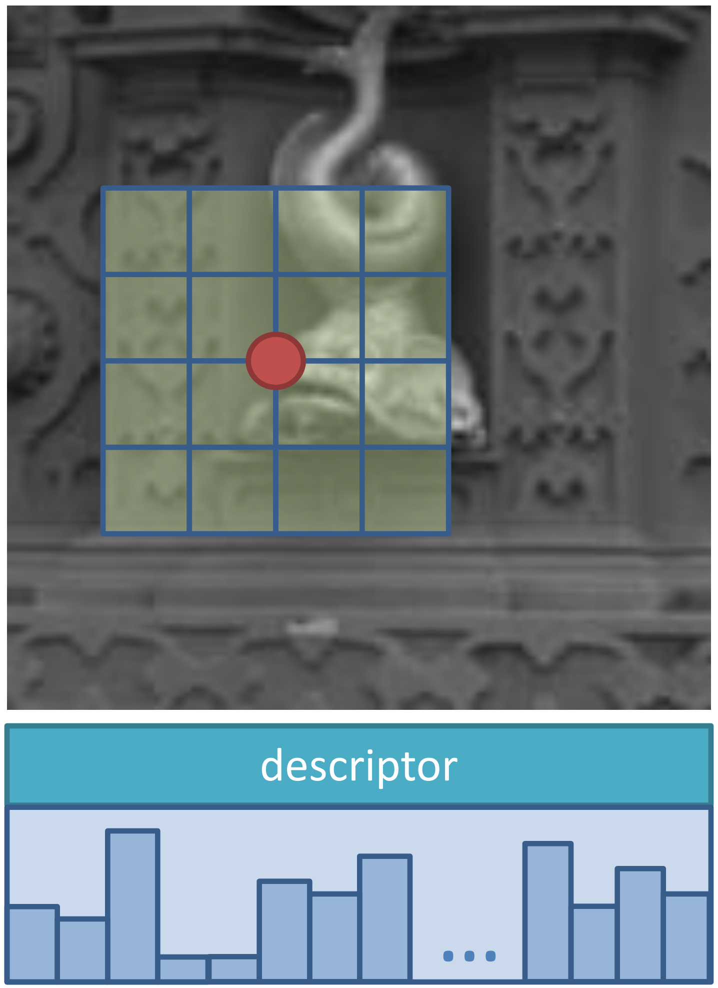

Images
Pre-processing
Registration Framework
1. Image gradients
2. Entropy images
3. Multi-resolution
4. Attributes
5. SIFT
Similarity Measure
Optimization

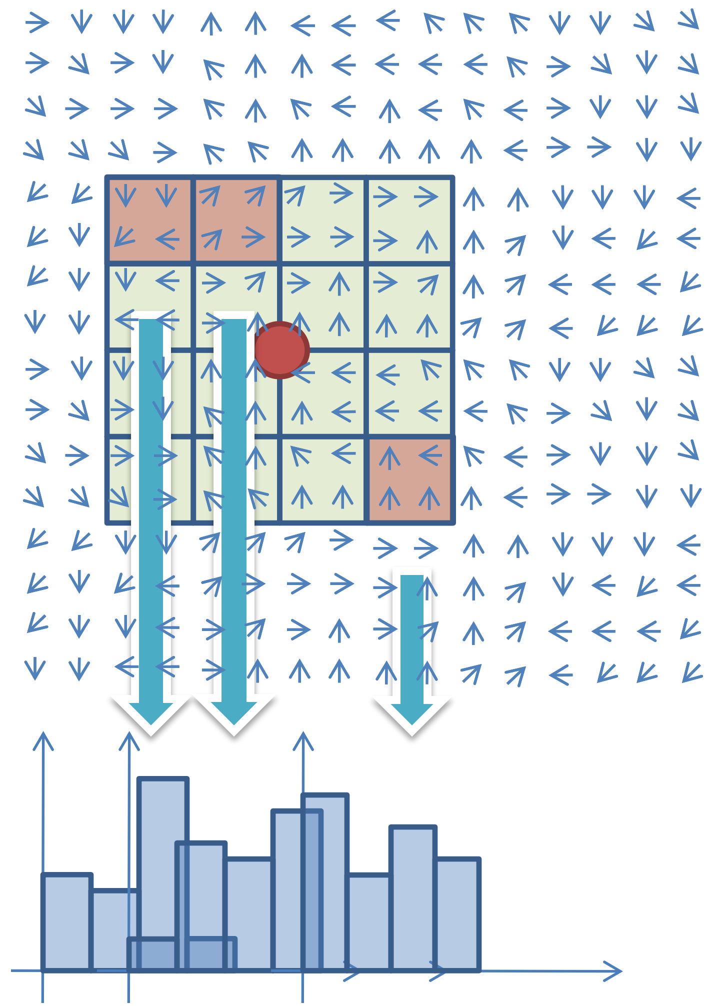


Images
Pre-processing
Registration Framework
1. Image gradients
2. Entropy images
3. Phase
4. Multi-resolution
5. Attribute vectors
Similarity Measure
Optimization

SIFT

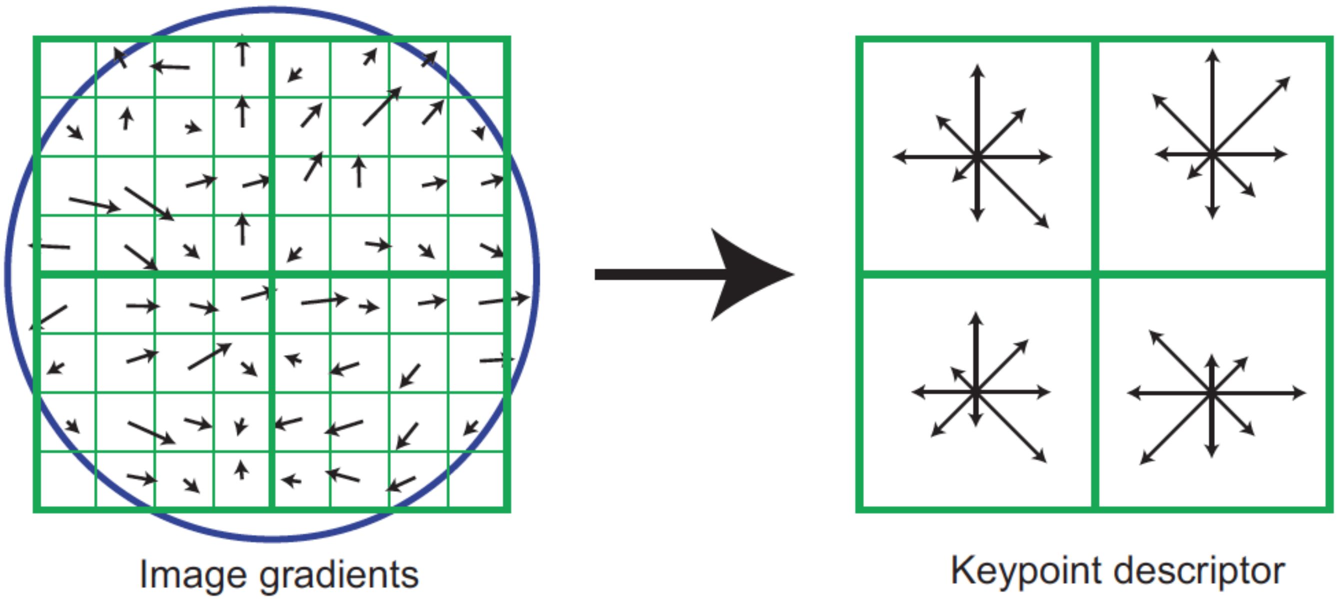
Scale Invariant Feature Transform





Lowe, Distinctive image features from scale-invariant keypoints, IJCV, 2004
Tola et al., A Fast Local Descriptor for Dense Matching

SIFT features



2 x 2 x 8 = 32

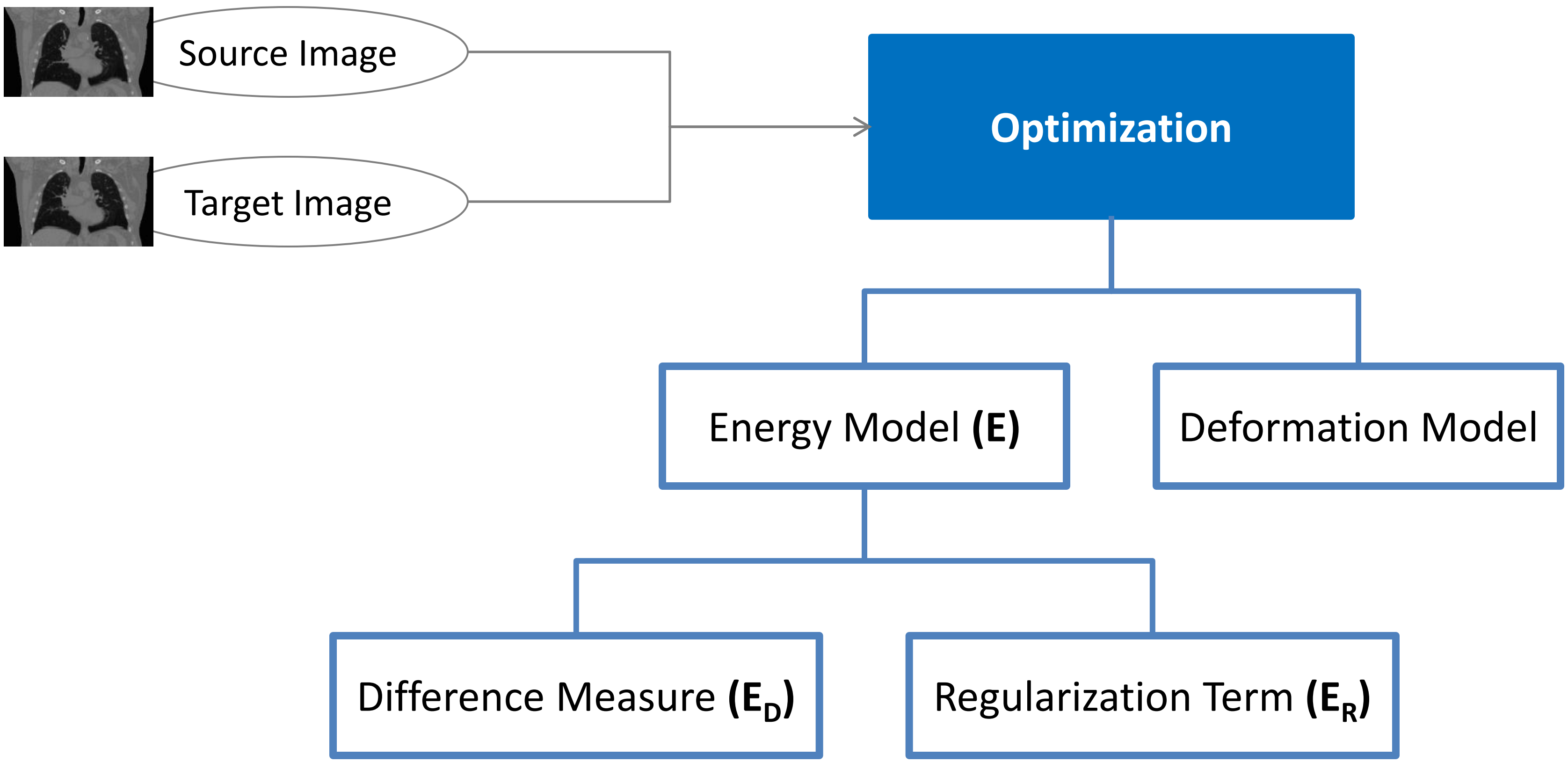
4 x 4 x 8 = 128

Matt's work for 3D SIFT

PART III

Optimization Methods

Deformable Registration: General Framework



Structure of General Energy Formulation

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

$$E(\phi) = \int_\Omega \|e_D(u)\| \, \mathrm{d}x + \lambda \int_\Omega \|e_R(u)\| \, \mathrm{d}x$$

Depends on deformation through image I_S

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

→ Non-linearity

- In many cases, the error term for the regularization is linear in the displacement: $e_R(u) = (Lu)^*Lu = u^*L^*Lu$
- the linear operator is mostly a differential operator (e.g. $L=\nabla$, $L=\Delta$, ...)

Non-linearity prohibits closed-form solutions

→ Local optimization problem

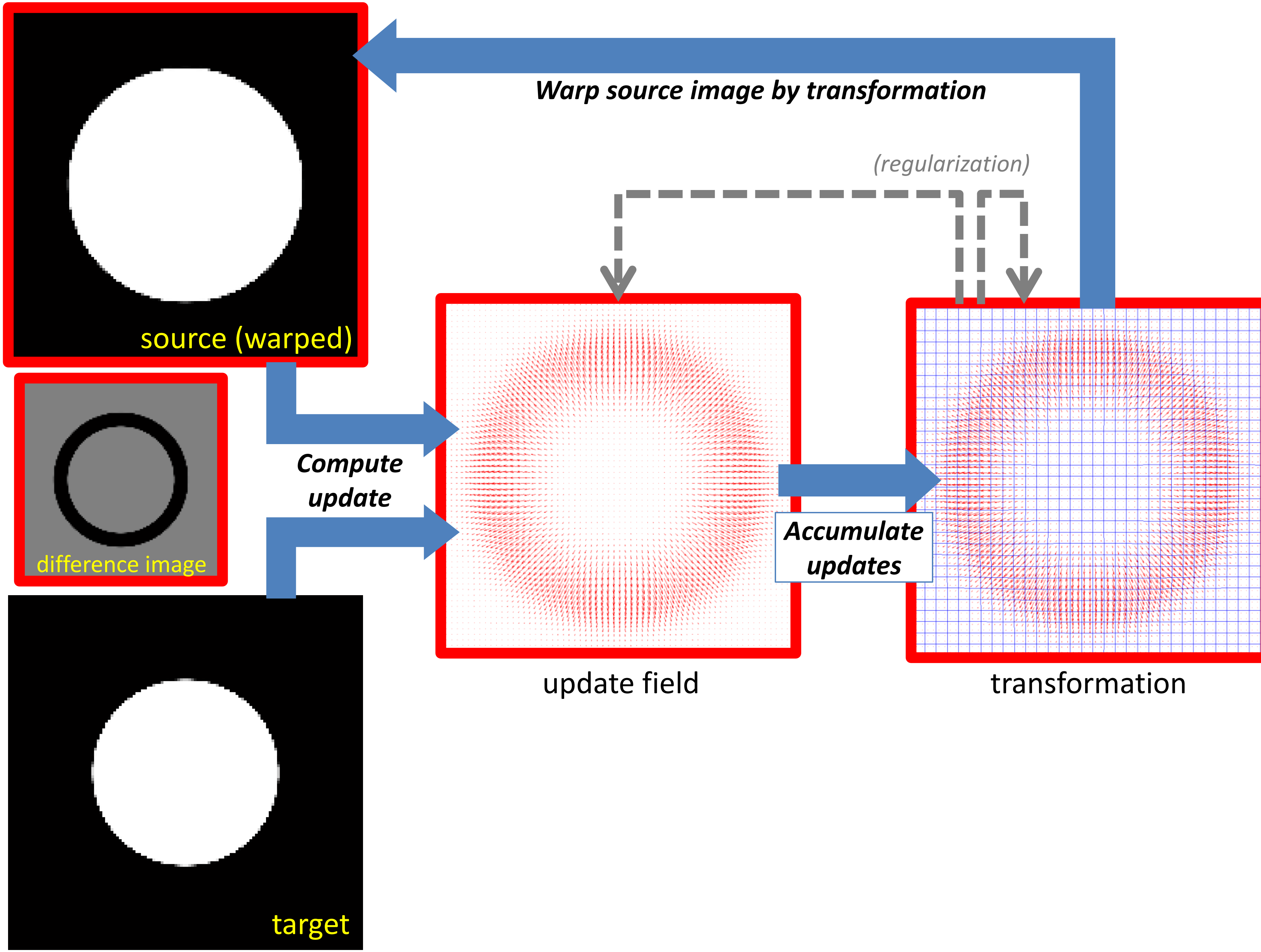
→ Solutions = local optima

→ Methods: Iterative (accumulation of updates)

Registration

=

Iterative Accumulation of Displacement Updates



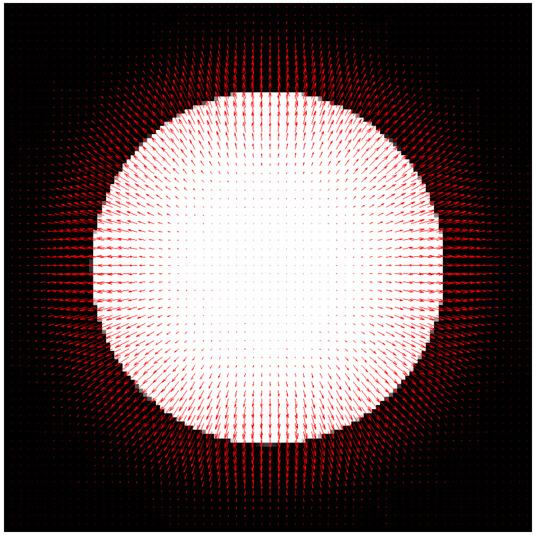
Warp source image by transformation
(regularization)
source (warped)
difference image
Compute update
Accumulate updates
update field
transformation
target

General Update Structure: Smoothed noisy update estimate

$$v \propto K\left(\nabla E_{\mathrm{D}}(\phi) + \lambda \nabla E_{\mathrm{R}}(\phi)\right)$$

Smoothing operator, e.g. Gaussian
Depends on:

- Deformation space (L2, Sobolev space, ...)
- Parametrization (B-Spline FFD, ...)
- Optimization scheme



$/ \quad K$

Registration

=

Iterative Accumulation of **SMOOTHED** Displacement Updates

How to determine the updates u_i?

→ different optimization schmes

1. Gradient-based Optimization

$$v = -\tau \ \mathrm{some\_function}\Big(\nabla E(u)\Big)$$

2. Gradient-free/Discrete Optimization

$$v = \mathrm{some\_other\_function}\Big(E_D, E_R, u\Big)$$

Steepest Gradient Descent

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

Starting with initial ϕ_0, repeat until convergence:

$$v = -\nabla E(\phi)$$ *// compute update based on gradient of energy*

$$\phi = \phi + \tau v$$ *// apply the update*

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

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

→ derivative of difference measure
→ derivative of regularization term

Derivative of the Regularization Term

General formulation of the derivative for a regularization term with a quadratic form:

- General regularization term:
- Assume:
  - Squared L2 norm: $E_R(u) = \frac{1}{2}\int_\Omega e(u)^2 = \frac{1}{2}\langle e(u), e(u)\rangle$
  - Error term is linear in u: $\langle e_R(u), e_R(u)\rangle = \langle Lu, Lu\rangle = \langle u, L^* L\ u\rangle$

$$E_R(u) = \int_\Omega \|e_R(u)\|\ \mathrm{d}x$$

$$E_R(u) = \frac{1}{2}\langle u, L^* Lu\rangle$$

- Then, the derivative reads:

$$\frac{\mathrm{d}E_R(u)}{\mathrm{d}u} = L^* Lu$$

For **diffusion regularization**, we get:

$$E_R(u) = \frac{1}{2}\int_\Omega \sum_{i=1}^{d} \|\nabla u_d(x)\|^2 \mathrm{d}x \quad \Rightarrow \quad \frac{\mathrm{d}E_R(u)}{\mathrm{d}u} = \nabla^* \nabla u = -\Delta u$$

EXAMPLE: Steepest Gradient Descent

$\phi = \mathrm{Id} + u$

Derivative of the Difference Measure

General formulation of the derivative of the difference measure:

$$\frac{\partial E_D(I_T, I_S(\phi))}{\partial u} = \underbrace{\frac{\partial E_D(I_T, I_S(\phi))}{\partial I_S(\phi)}}_{W(I_T, I_S(\phi))} \underbrace{\frac{\partial I_S(\phi)}{\partial \phi}}_{(\nabla I_S)(\phi)} \underbrace{\cancel{\frac{\partial \phi}{\partial u}}}_{\mathrm{Id}}$$

Point-wise evaluation at x∈Ω: **point-wise rescaling of the warped gradient of *I*$_s$**

$$\underbrace{\frac{\partial E_D(I_T, I_S(\phi))}{\partial u}(x)}_{\in \mathbb{R}^d} = \underbrace{W(I_T, I_S(\phi))(x)}_{\in \mathbb{R}} \underbrace{(\nabla I_S)(\phi(x))}_{\in \mathbb{R}^d}$$

For the SSD we get:

$$E_D = \frac{1}{2}\int_\Omega (I_T(x) - I_S(\phi(x)))^2 \mathrm{d}x \Rightarrow \frac{\partial E_D}{\partial u} = -(I_T - I_S(\phi))\, \nabla I_S(\phi)$$

Summary: Steepest Gradient Descent

Starting with initial ϕ_0, repeat until convergence:

$$
\begin{aligned}
v &= -\nabla E(\phi) \\
&= -\nabla E_D(\phi) - \lambda \nabla E_R(\phi)
\end{aligned}
$$

// compute update based on gradient of energy

$$\phi = \phi + \tau v$$

// apply the update

Gradient-based Optimization Methods

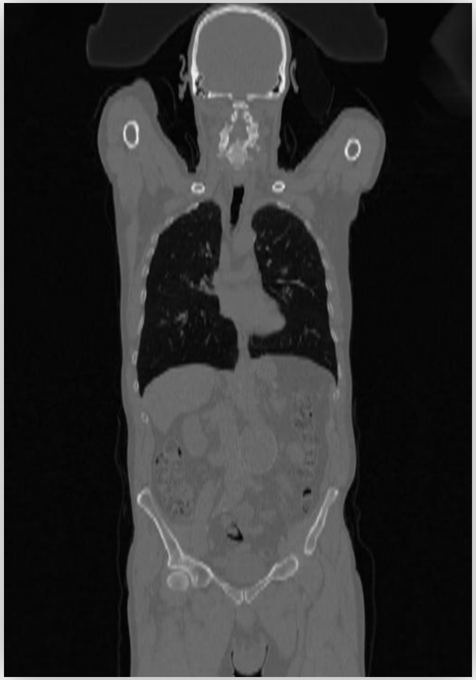
Method	Update Rule	Comment
Steepest Descent	$v = -\tau \ \mathrm{Id}^{-1} \ \nabla E$	+ Simple implementation + Only gradient required − Numerical instable: requires small time steps → many iterations needed
PDE-inspired Semi-implicit Discretization	$v = -\tau \ \left(\mathrm{Id} + \tau\lambda\nabla E_R\right)^{-1} \ \nabla E$	+ Numerically stable also for large time steps + Linear operator determined by regularization → difference measure easily exchangable − Poor convergence speed
Gauß-Newton	$v = -\tau \ \left(J_e^\top J_e\right)^{-1} \ \nabla E$	+ Numarically stable also for large time steps + Good convergence speed − Linear operator depends on both, the regularization *and* the difference term − ***J_e*** must be sparse/small for efficient treatment
Levenberg-Marquardt	$v = -\tau \ \left(\lambda\mathrm{Id} + J_e^\top J_e\right)^{-1} \ \nabla E$	**A mixture of Steepest Gradient Descent and Gauß-Newton**
L-BFGS & Conjugate Gradient	$v = -\tau \ P^{-1} \ \nabla E$ $P^{-1} = \mathrm{function}(\nabla E_{t<T})$	+ Require only gradient evaluations + Good convergence speed − Convergence depends on exact time-step requirements
Preconditioned Gradient Descent (Quasi-Newton)	$v = -\tau \ P^{-1} \ \nabla E$ With P approximating the Hessian of E, e.g.: • Jacobi preconditioning • For def. Registration: [Zikic 2010]	Most general formulation of the above. Properties depend heavily on choice of P. *"Finding a good preconditioner (...) is often viewed as a combination of art and science."* Y. Saad, Iterative Methods for Sparse Linear Systems

General Optimization References

BOOK:
Numerical Optimization
Jorge Nocedal and Stephen J. Wright
Springer Series in Operations Research

Techreport:
Madsen, K., Nielsen, H. and Tingleff, O.
Methods for Non-linear Least Squares Problems
2004

Intuition: What makes gradient-based optimization efficient for deformable registration?



Source Image

I_S

Steepest Descent Update

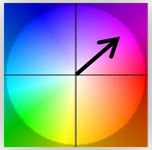
∇E

Gauss-Newton Update

$\left(J^\top J\right)^{-1} \nabla E$

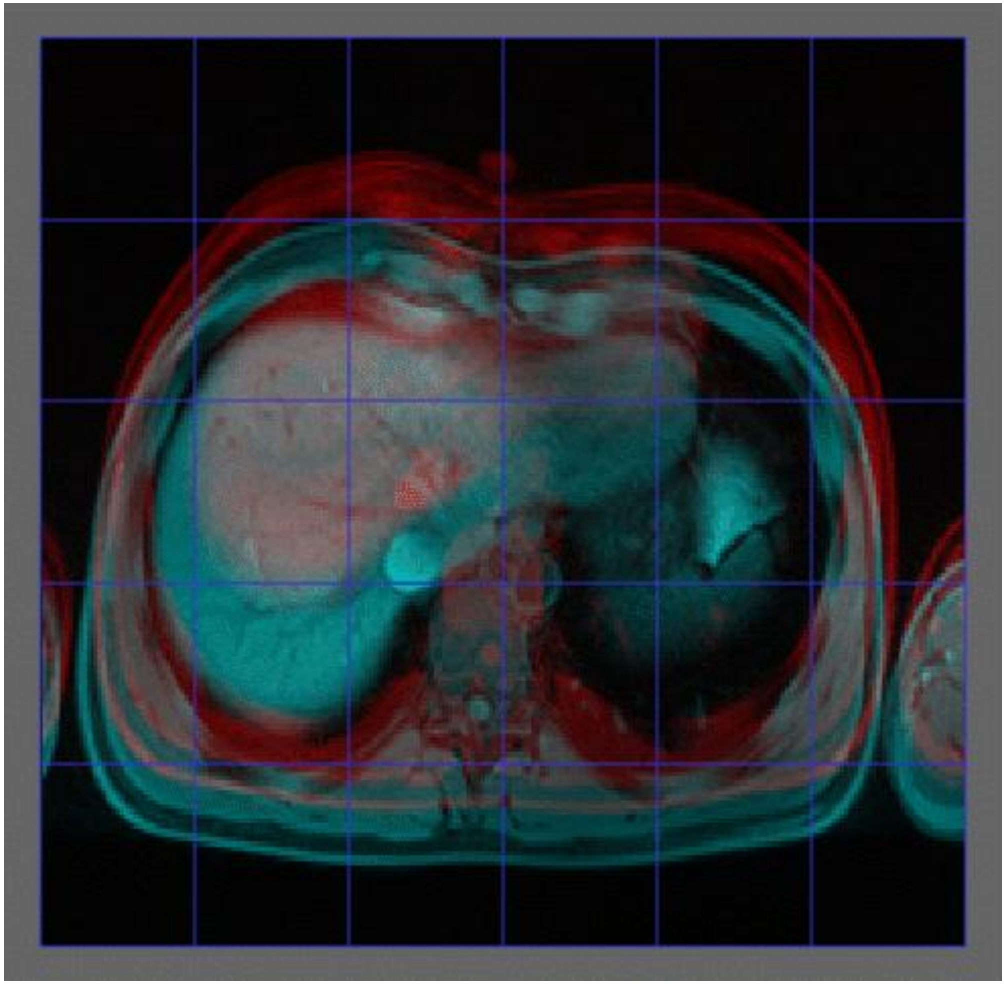
Update is not dominated by largest intensity gradients in input image only.

→ Gauss-Newton method on SSD does not suffer from “Local Gradient Bias“. [Zikic 2011]



Deformable Registration by Discrete Optimization

Low-dimensional deformation model (B-Spline FFD)



[Glocker 2008]

Deformable Registration by Discrete Optimization

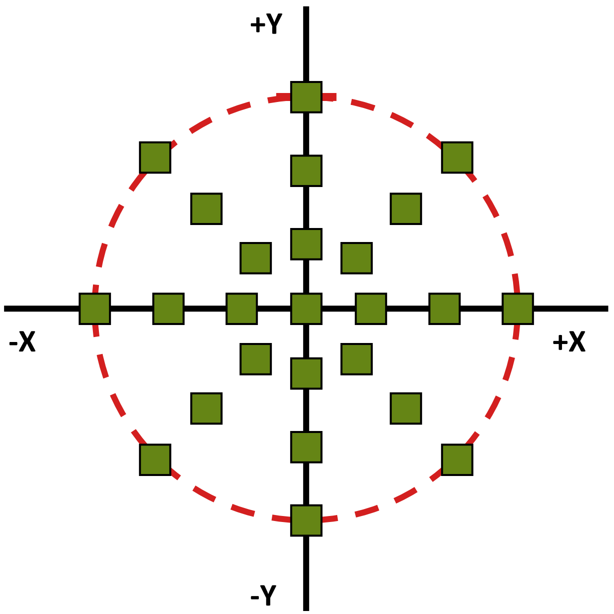
Low-dimensional deformation model (B-Spline FFD)

Update computation:

1. For each control point CP_i

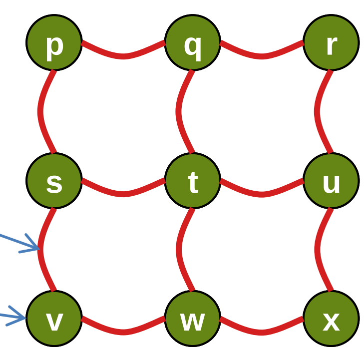
   For a discrete number of displacements d^{l_p}
   evaluate approximative change in similarity measure

$$V_p(l_p) = \underbrace{\int_\Omega \hat{\eta}(\mathbf{x}) \left(I_\mathrm{T}(x) - I_\mathrm{S}(x + d^{l_p})\right)^2 \,\mathrm{d}x}_{\text{or any other local image metric}}$$



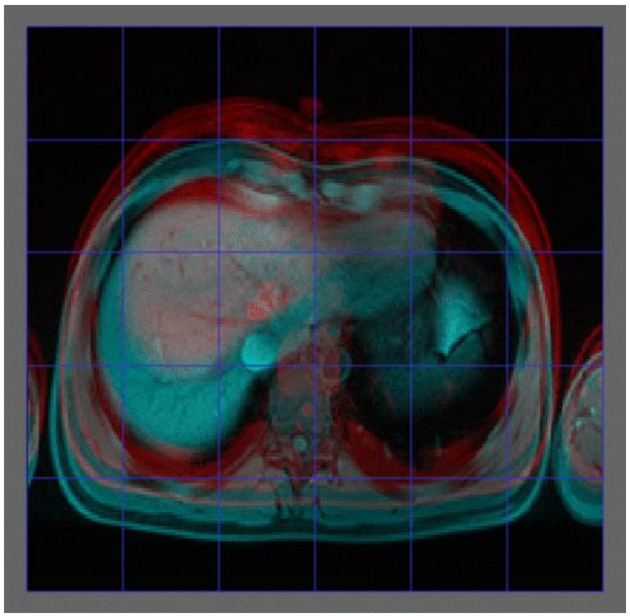
2. Compute approximately optimal combination of the pre-computed displacements w.r.t. chosen regularization with fast and accurate discrete optimization techniques

$$E_\mathrm{mrf}(\mathbf{l}) = \sum_{p \in G} V_p(l_p) + \sum_{(p,q) \in N} V_{pq}(l_p, l_q)$$



[Glocker 2008]

Deformable Registration by Discrete Optimization



Properties:

- No derivative computation required
- Similar efficiency for any difference measure
- Larger/non-local search range for each CP
  → increased capture range
- Computes only local versions of difference measures
- fast

that’s almost it

Intensity-based Deformable Registration as Energy Minimization

$$\phi' = \arg\min_{\phi} \left[E_{\mathrm{D}}(I_{\mathrm{S}} \circ \phi, I_{\mathrm{T}}) + \lambda E_{\mathrm{R}}(\phi) \right]$$

$\phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$

$I : \mathbb{R}^d \rightarrow \mathbb{R}$

Transformation ϕ
can assumed as element of:

- Can be modeled as elemet of a Hilbert space (L^2, Sobolev space) or group/manifold (group of diffeomorphisms)
- Has to be parametrized for digital representation (B-Spline FFDs, DCT, RBFs)

Difference Measure between:

- Target image I_T
- Warped source image $I_s \circ \phi$

Examples:

- Sum of squared differences (SSD)
- Sum of absolute differences (SAD)
- Correlation Coefficient (CC)
- Correlation Ratio (CR)
- Mutual Information (MI)

Regularization term:

- Models the behaviour of underlying elastic model (internal energy)
- Incorporates prior knowledge
- can be required to constrain problem

Examples:

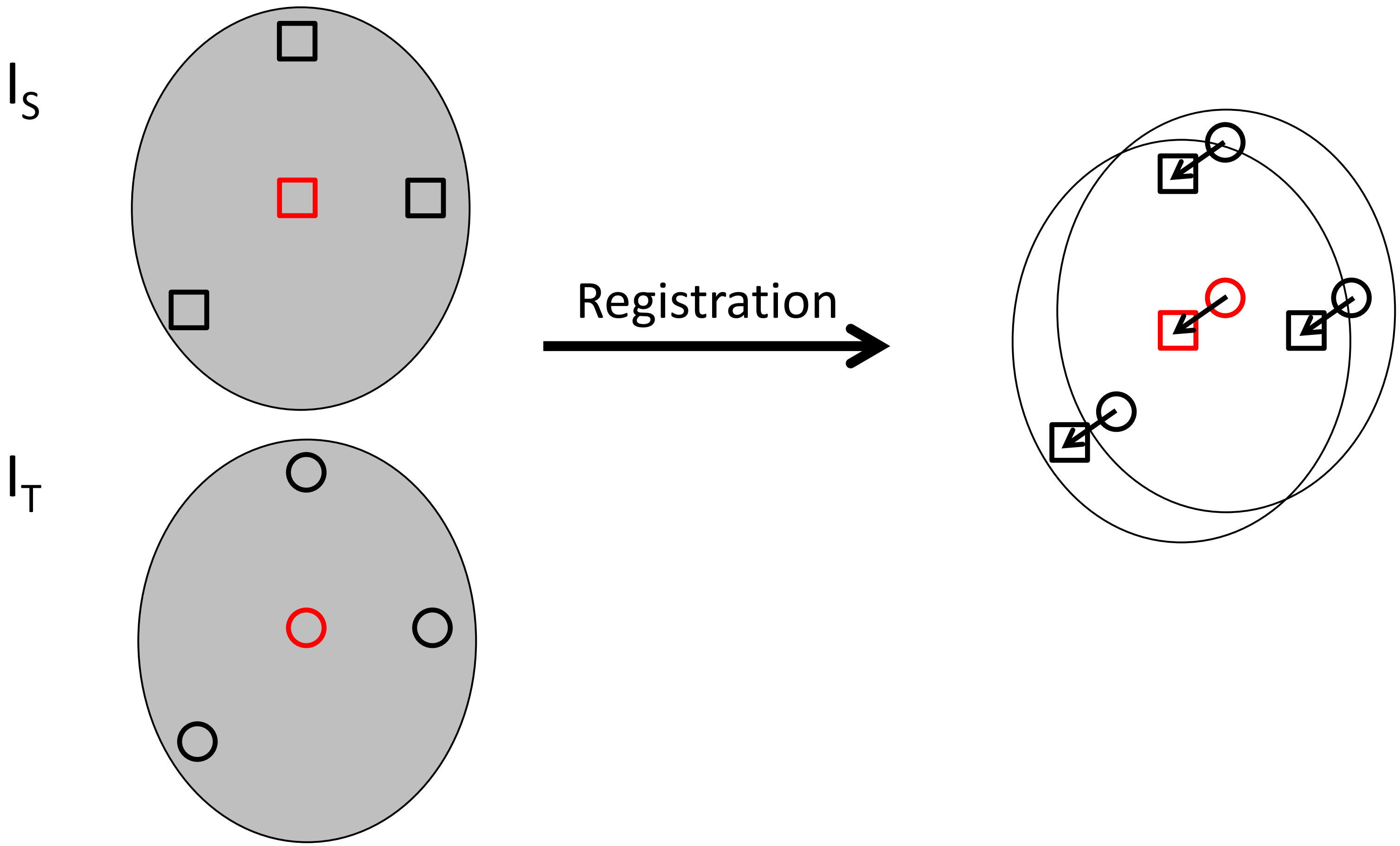
- Diffusion (1st-order) ((in-)homogeneous, (an-)isotropic)
- Curvature/Bend. Energy (2nd-order)
- Linear Elasticity

What is the best registration algorithm ?

How do we validate a registration algorithm ?

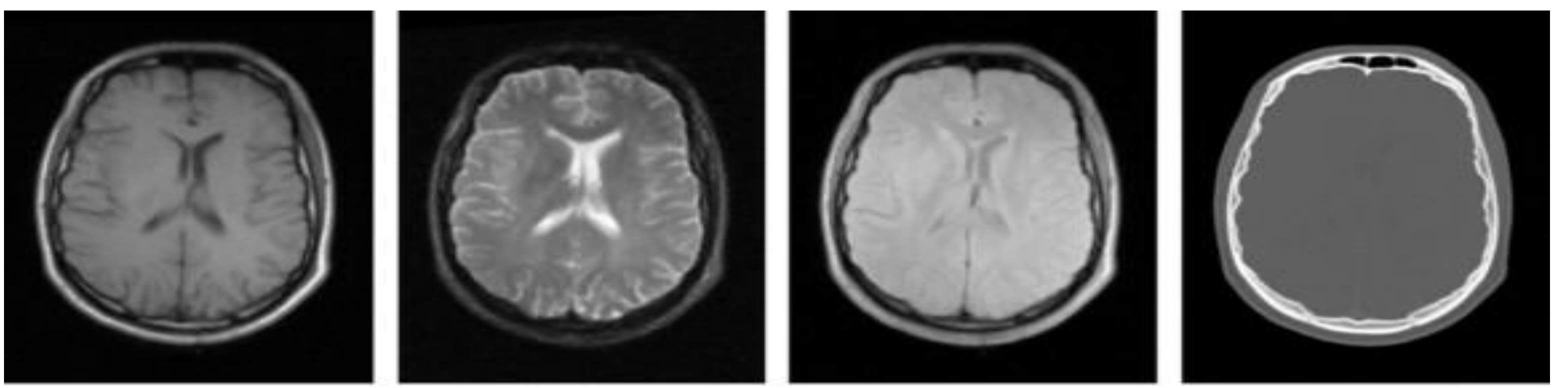
Rigid Registration Validation

- This is easy !
- Correspondences for at least 3 noncollinear landmarks
  - Sufficient to determine the error at any point



Rigid Registration Validation

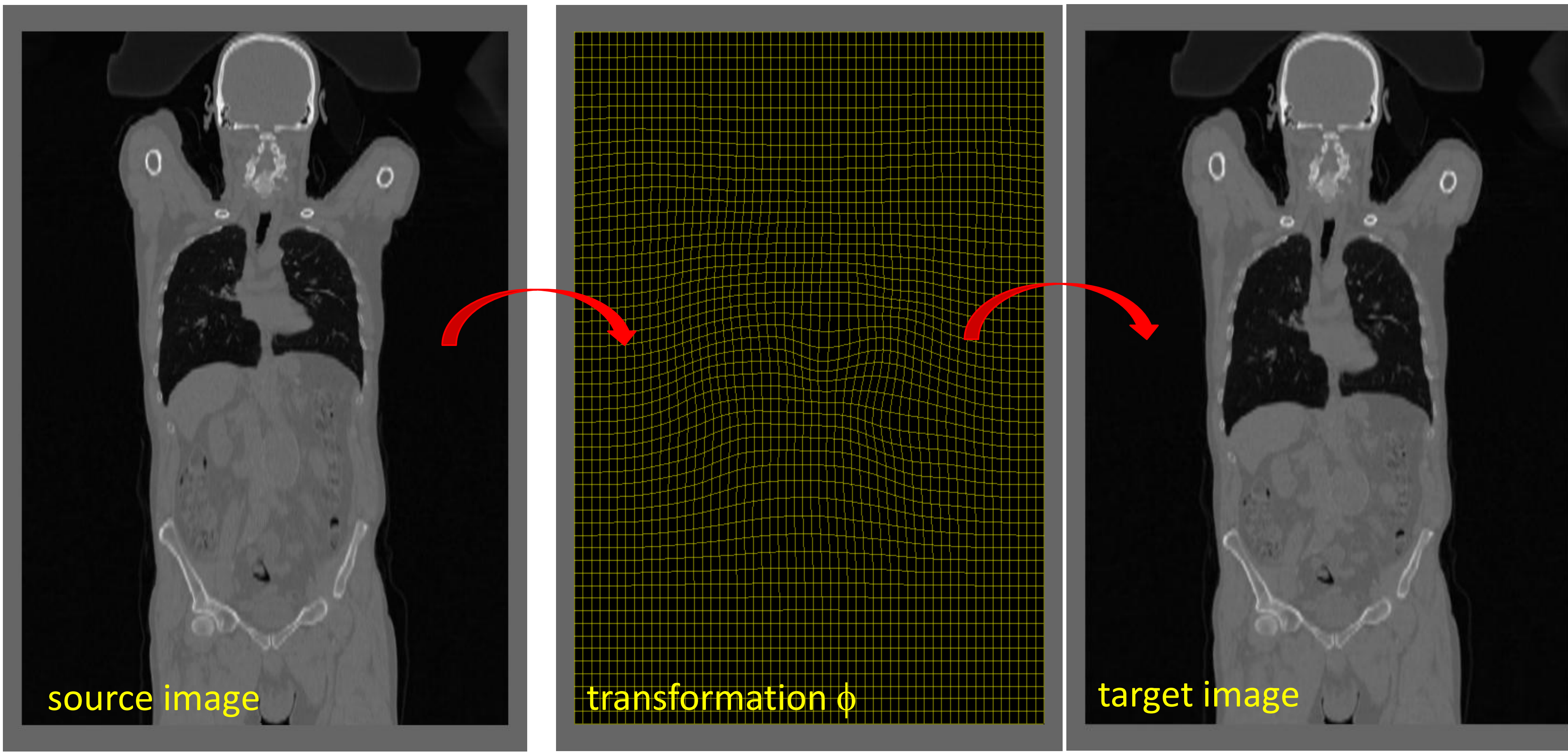
- Gold standard database available : http://www.insight-journal.org/rire/index.php
- Gold standard created by registration using marker-based registration
- Evaluation using 10 clinically relevant points
- CT-MR and PET-MR registration



Images from RIRE dataset (T1, T2, PD, CT).

Non-Rigid Registration Validation

- This is difficult !

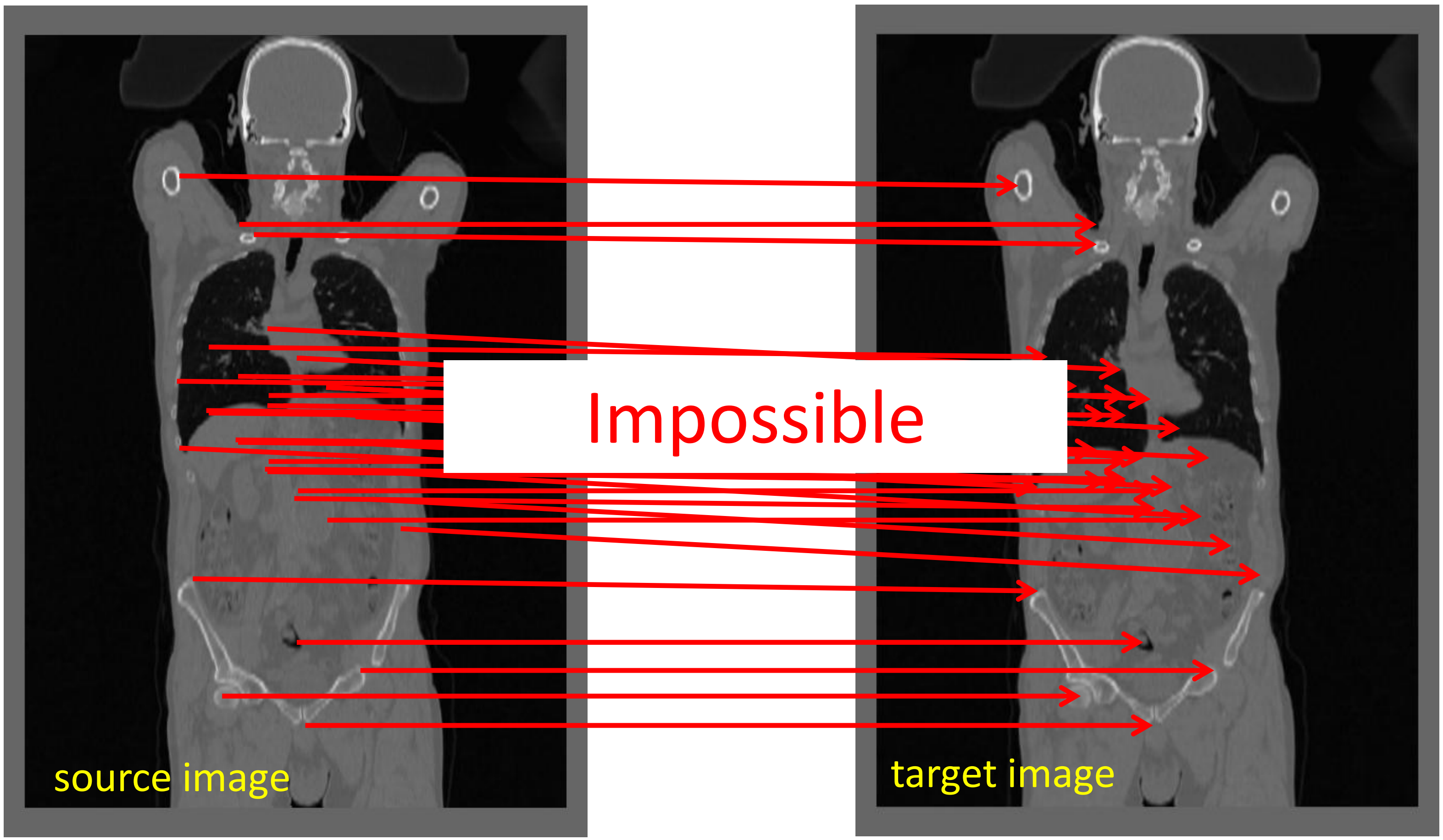


Non-Rigid Registration Validation

- This is difficult : desired ϕ is unknown

Non-Rigid Registration Validation

- Manually specify full transform ?



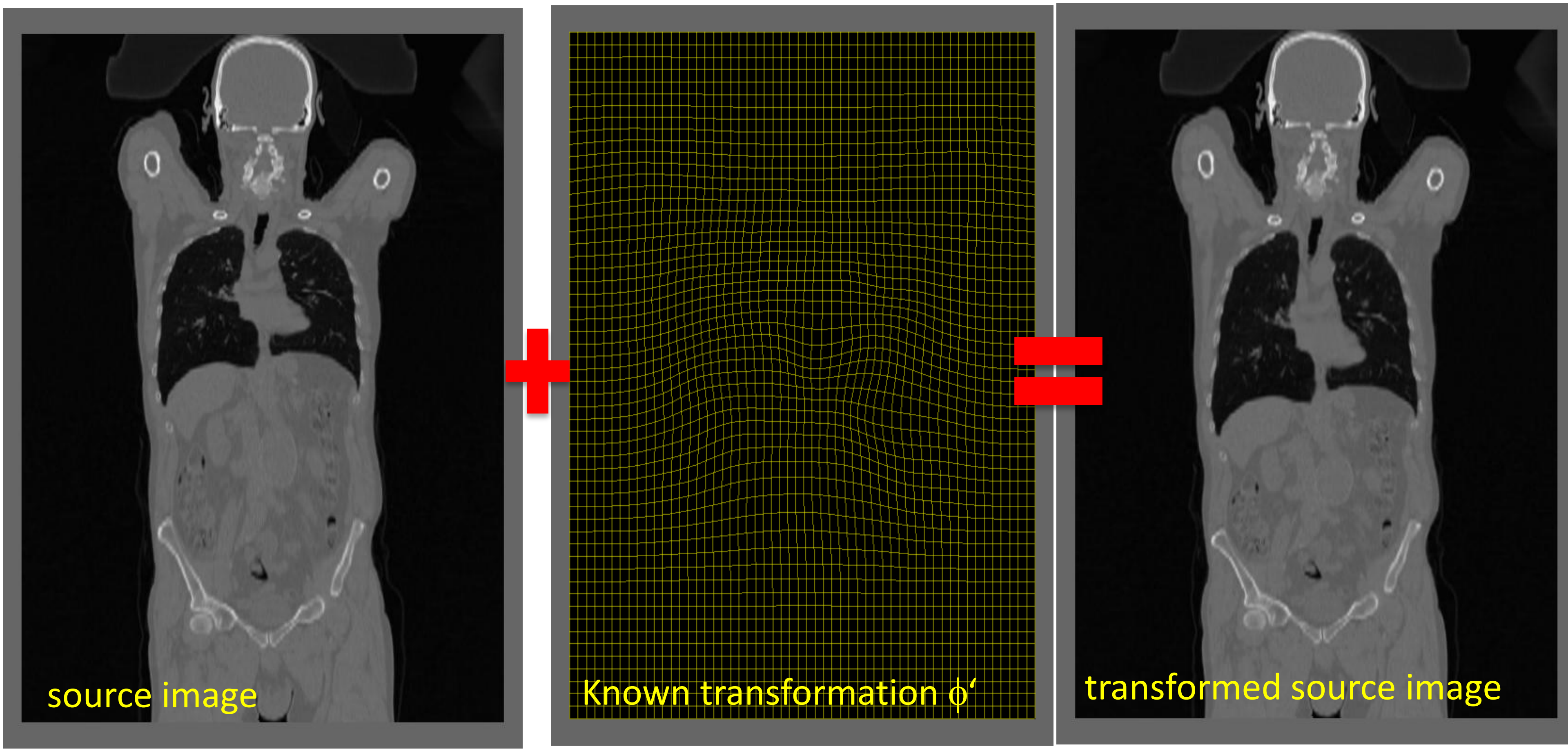
Rigid Registration Validation

- This is difficult !
- Lack of gold standard
  - Unknown desired transformation
  - Manual specification of full transform is impossible

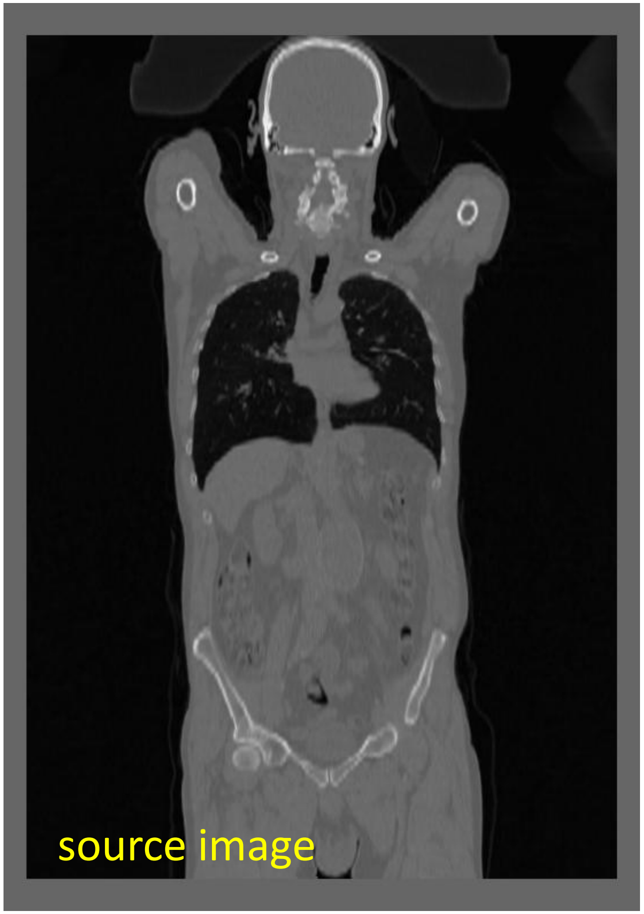
➥ Create synthetic transformations

➥ Use of surrogate measures

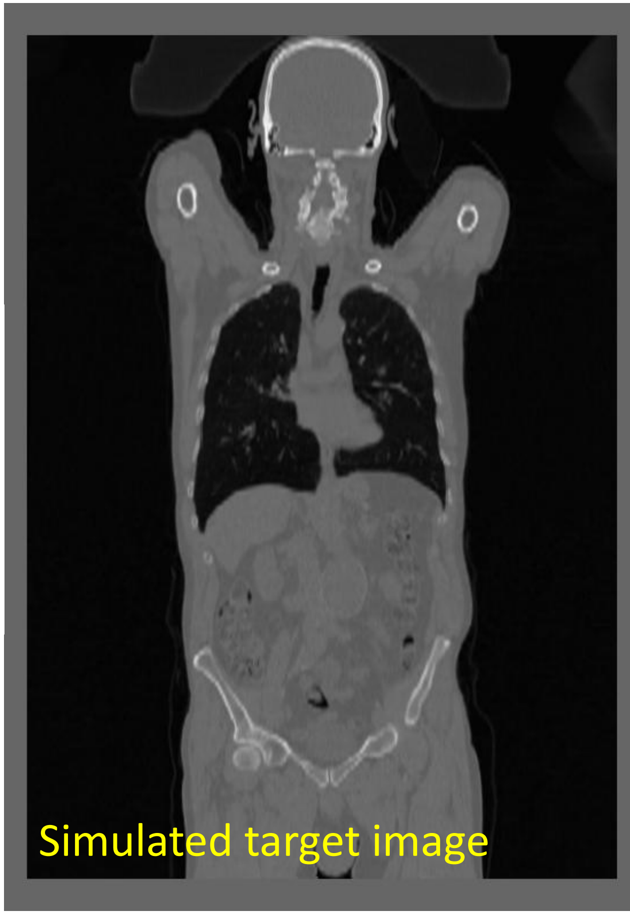
Synthetic Scenario



Synthetic Scenario



Compare transformation ϕ with ϕ'



Synthetic Scenario

1. Use of Biomechanical models to obtain deformation field
   – Only for intra-subject registration
   – Accuracy of model influences the study

2. Apply a known deformation
   – Images are not independent
   – Bias introduced by the method that estimated/created the deformation

Use of surrogate measures

- Region-Of-Interest overlap
- Intensity variance
- Inverse Consistency Error

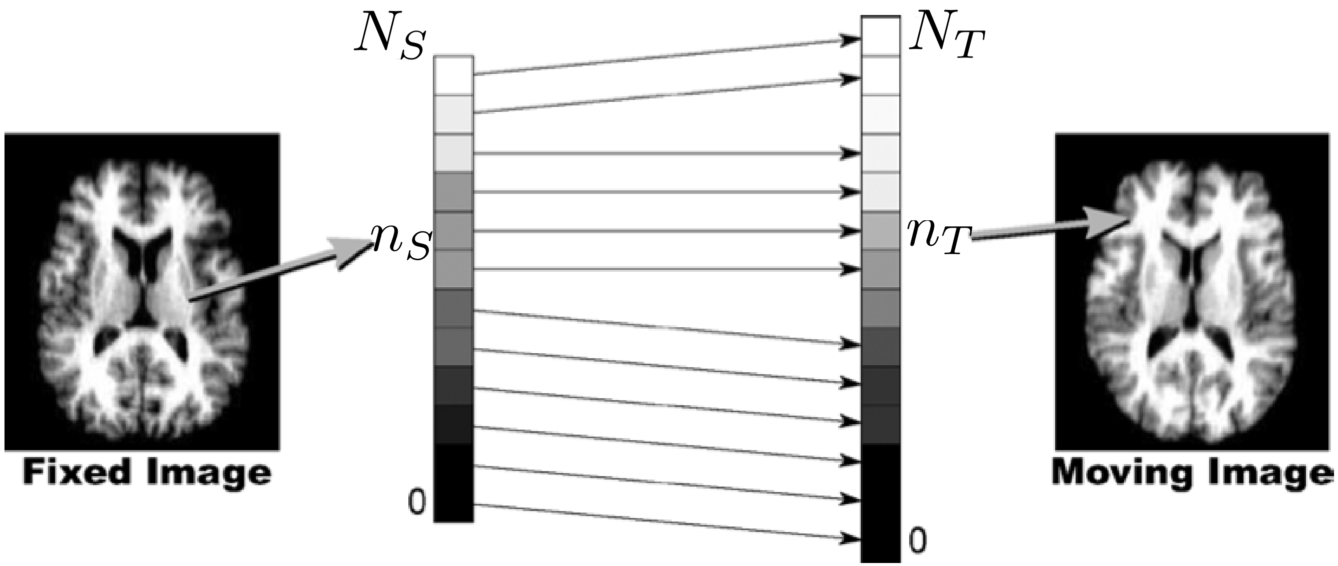
$$\phi_{ST} \circ \phi_{TS} = I$$

- Transitivity Error

$$\phi_{AB} \circ \phi_{BC} \circ \phi_{CA} = I$$

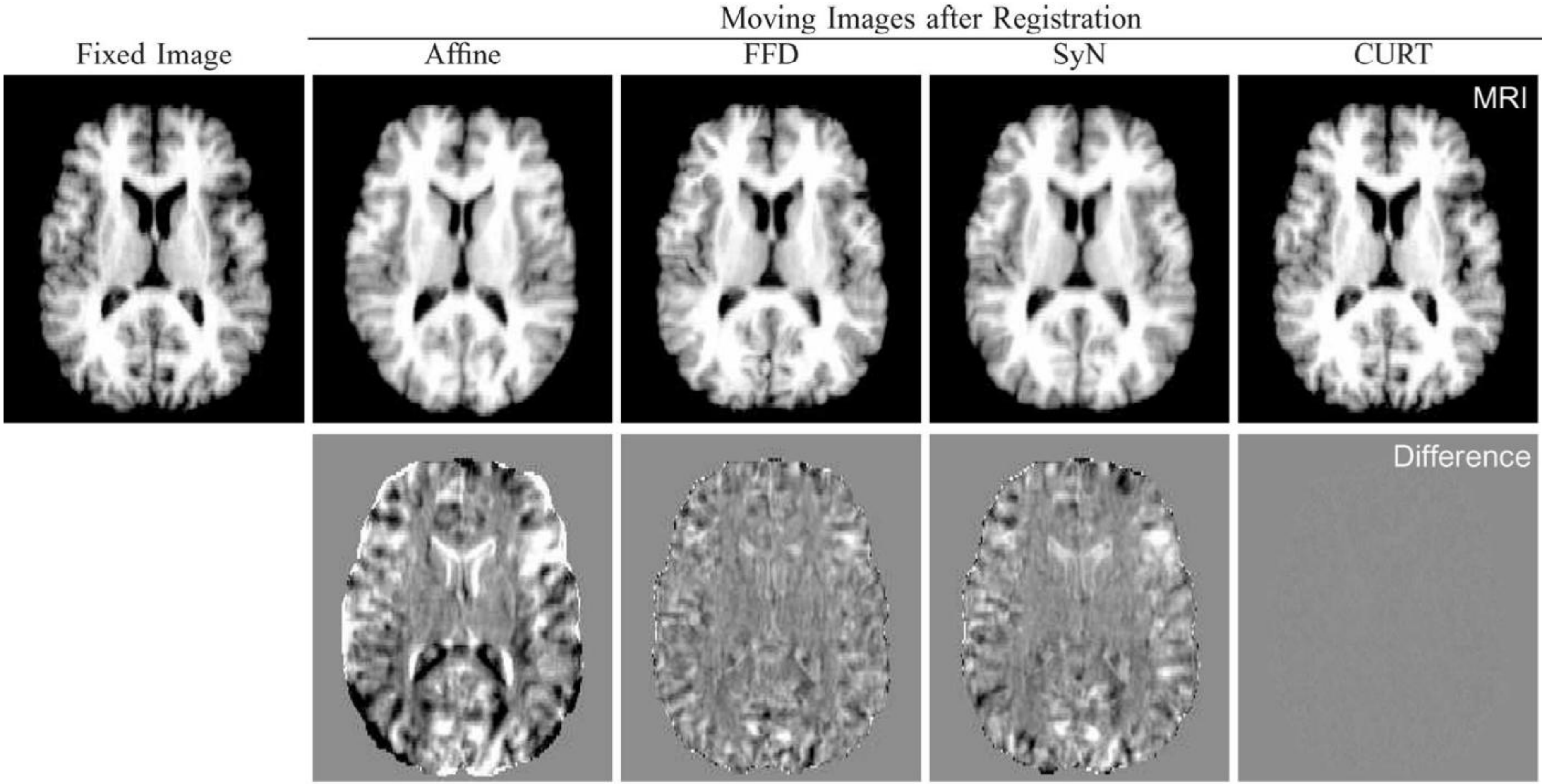
How good are these measures? [Rohlfing 12]

- Completely Useless Registration Tool (CURT)



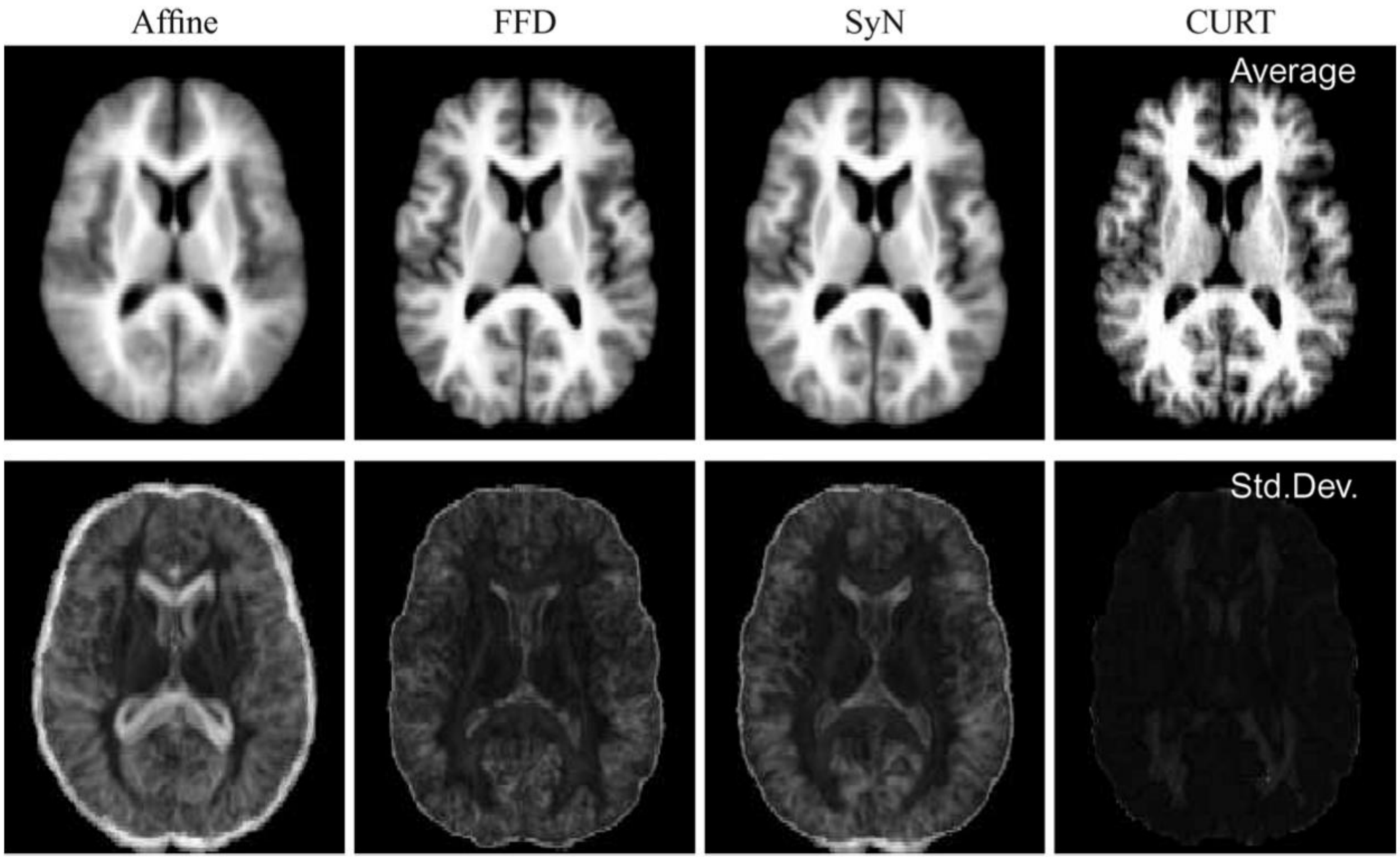
$$n_S = \lfloor N_S \frac{n_T}{N_T} \rfloor$$

Intensity Similarity

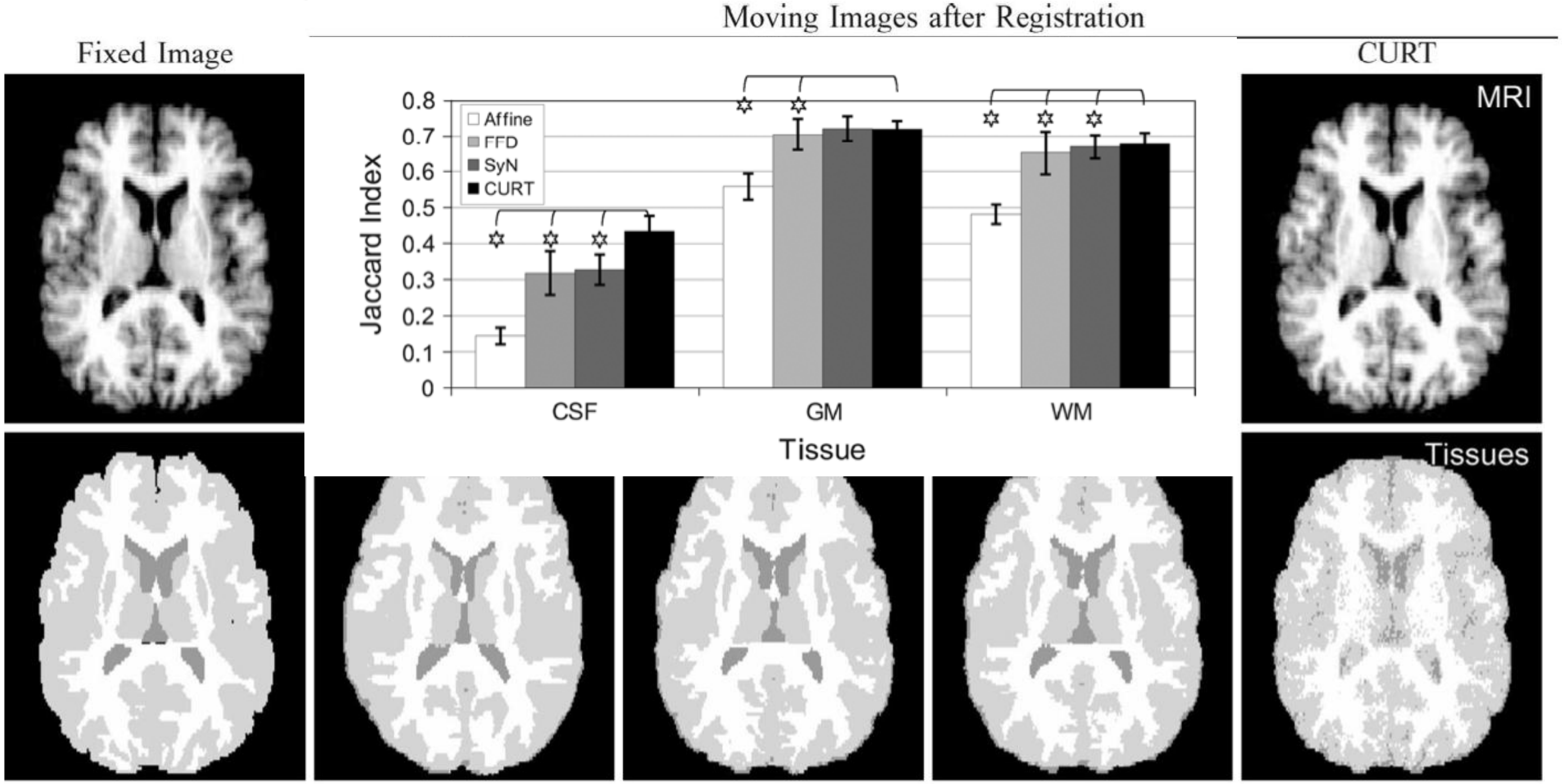


CURT outperformed the other registration methods when considering: i) RMS image difference, ii) NCC image correlation and iii) NMI image similarity !!

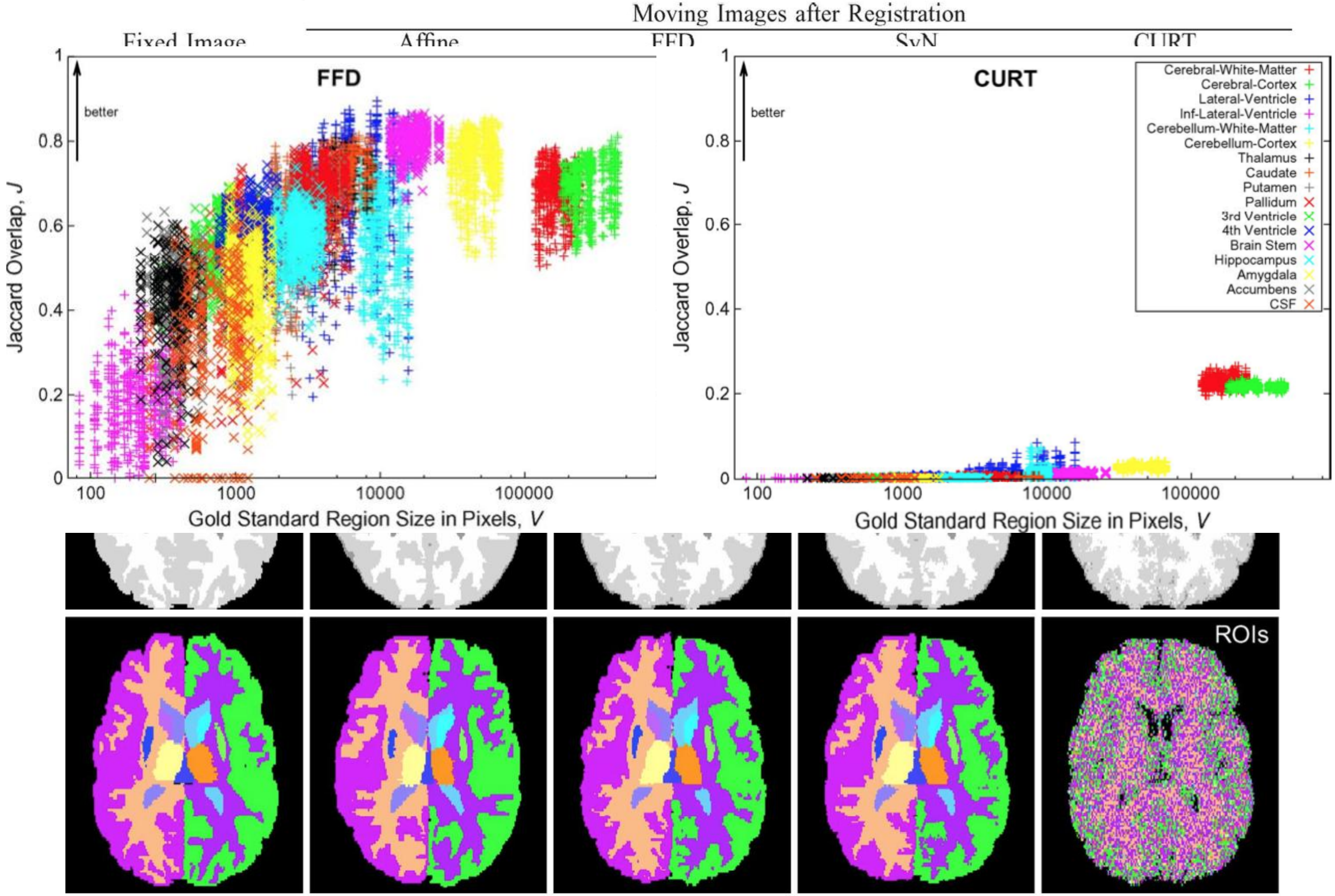
Intensity Similarity



Region-Of-Interest Overlap



Region-Of-Interest Overlap



What is the best registration algorithm ?

What is the best registration algorithm ?

Theorem: For every algorithm there is a dataset where it will outperform all others !

Future

Combination of metrics: it is inevitable that non of the existing metrics can work in the general setting, therefore the answer should come from their combination

Metrics learned from data/examples: progress of machine learning have made possible learning correlations between data, and therefore define appropriate metrics should be able to learned from examples

Introduction of anatomical constraints: anatomy is not taken into account until recently when defining appropriate regularization terms, which normally should impose deformation consistency that is constrained from the anatomy.

References

Aristeidis Sotiras, Christos Davatzikos, Nikos Paragios: Deformable Medical Image Registration: A Survey. IEEE Trans. Med. Imaging 32(7): 1153-1190 (2013)

http://cvn.ecp.fr/teaching/biomed/2013/sotiras-wachinger-zikic.zip

Slides Courtesy:
A. Sotiras (UPENN), C. Wachinger (MIT) & D. Zikic (MSFT)