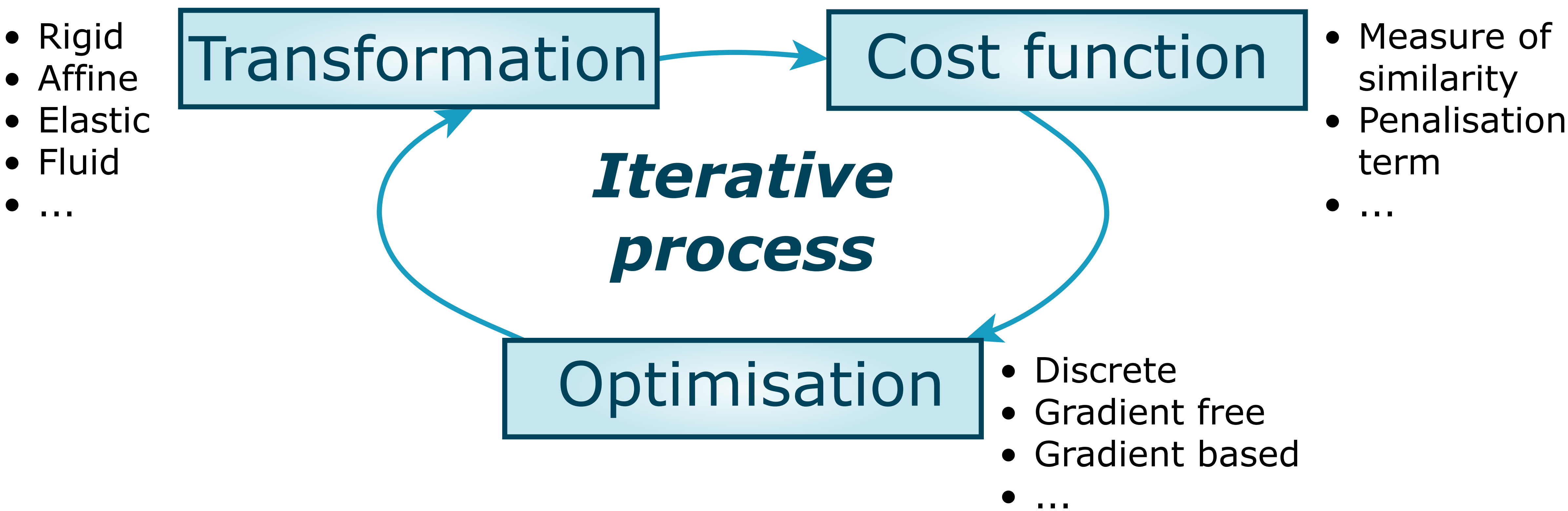
Medical Image Registration
Measures of similarity

Marc Modat

Centre for Medical Image Computing
Dementia Research Centre
University College London

Medical image registration

- Overall scheme



Outline

- SSD for Gaussian noise
- Normalised Cross correlation for linear relationships
- Joint entropy, the dispersion in the joint image histogram
- Mutual information
- Normalised mutual information widely used.
- Feature based alignment

Sum Squared Differences

- SSD is given below

$$SSD = \sum_{n}^{N} (A_n - B_n(\mathbf{u}))^2$$

- Minimising overall SSD is equivalent to minimising over arbitrary sub-regions, i.e. pixels.

- Thus on a pixel level in each dimension, we only need to compare the nearest neighbours. Thus the gradient is given by:

$$\nabla SSD_n = -2(A_n - B_n(\mathbf{u}))\nabla B_n(\mathbf{u})$$

Sum Squared Differences

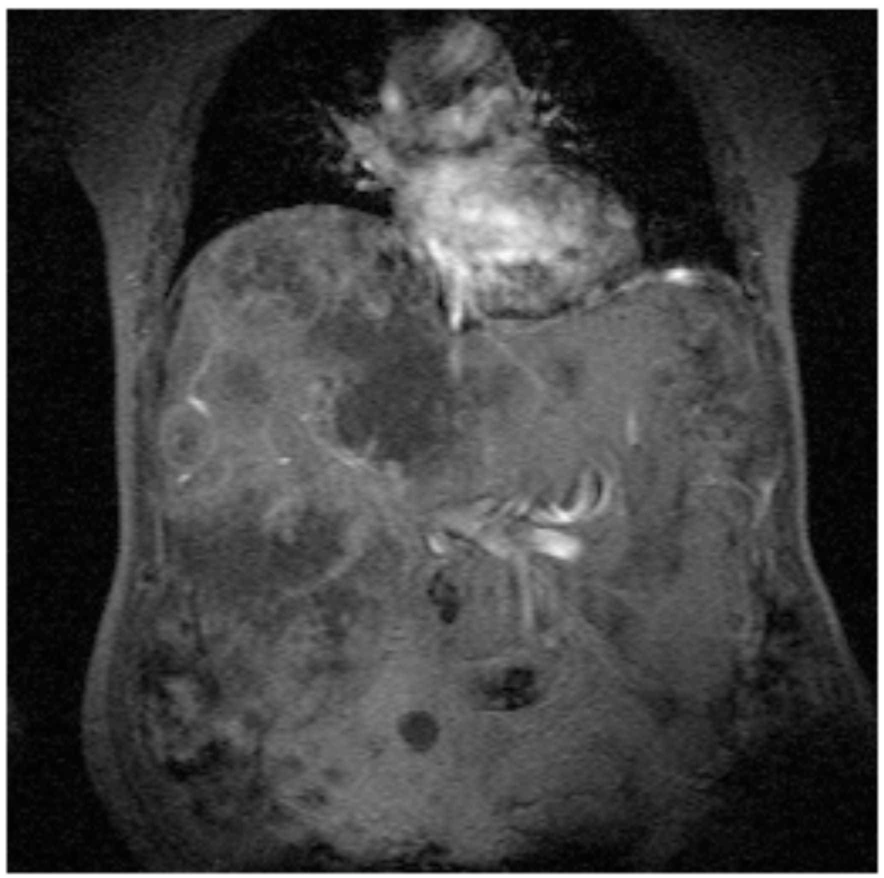


Image A

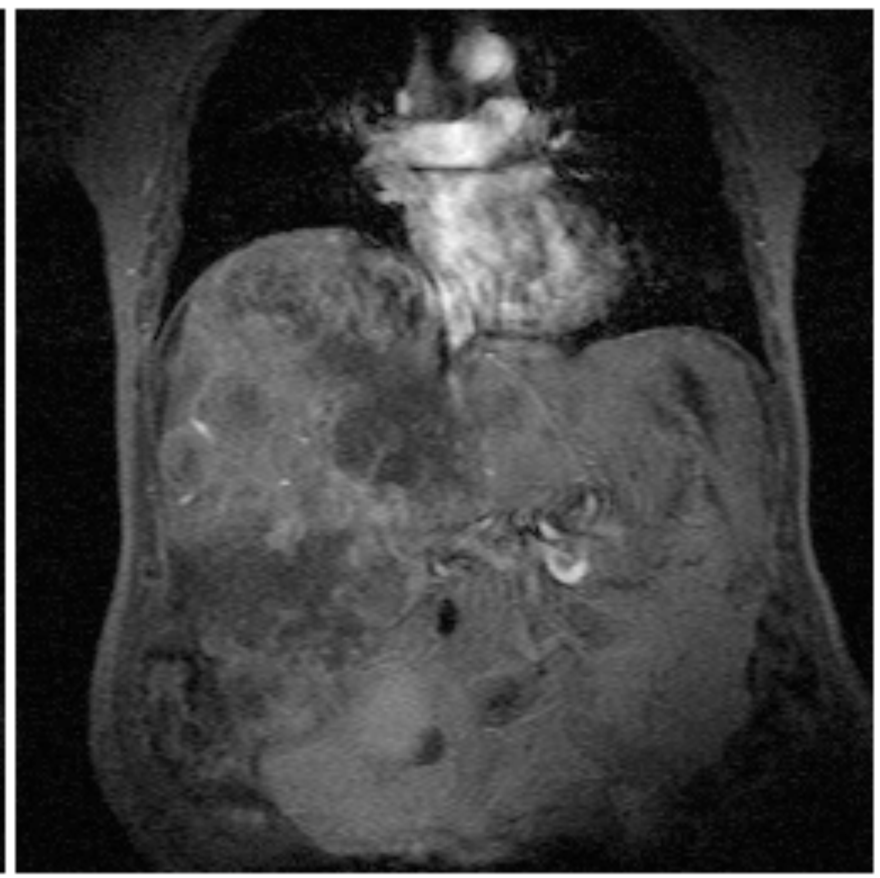
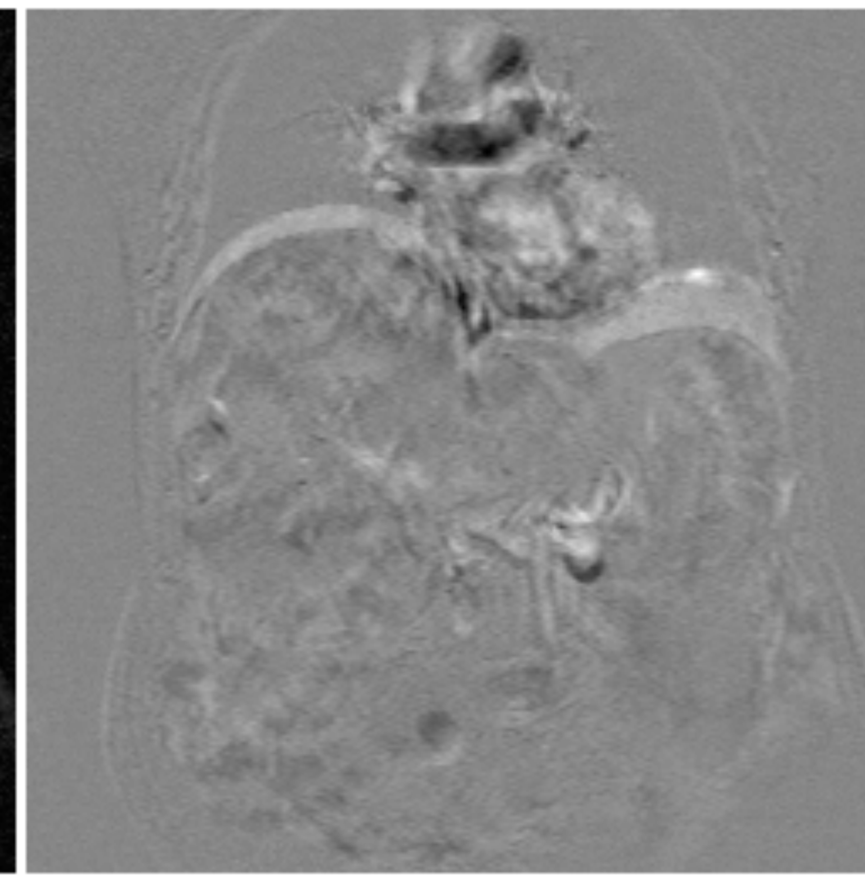


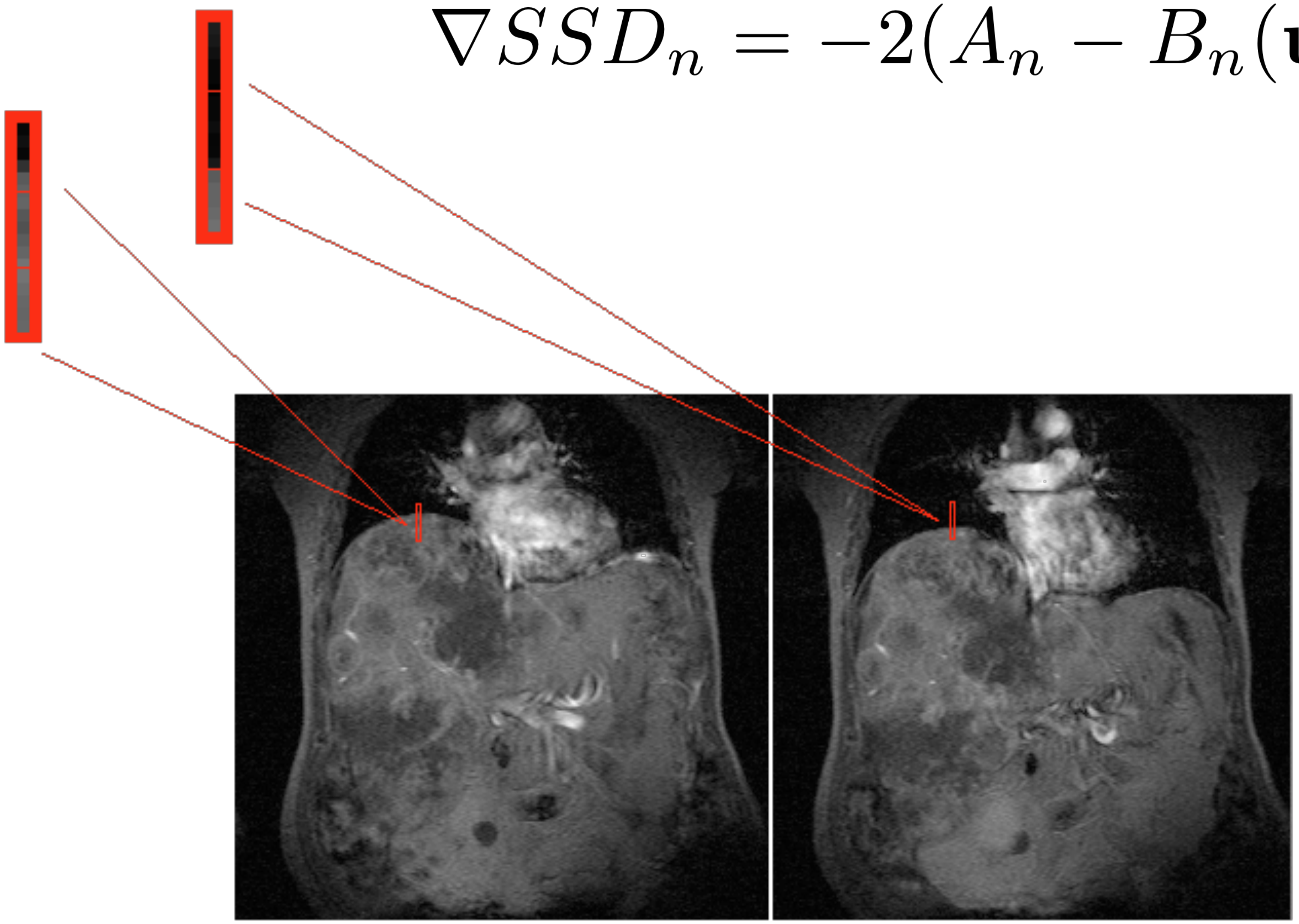
Image B



A-B

Sum Squared Differences

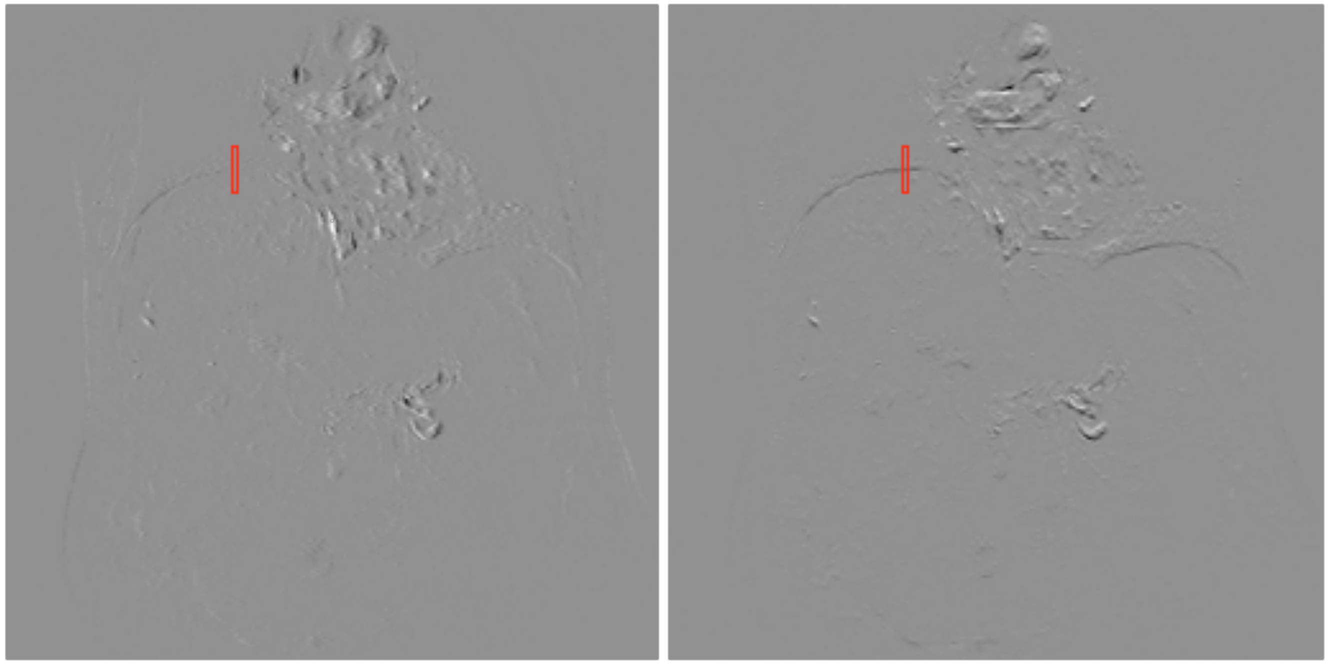
$$\nabla SSD_n = -2(A_n - B_n(\mathbf{u}))\nabla B_n(\mathbf{u})$$



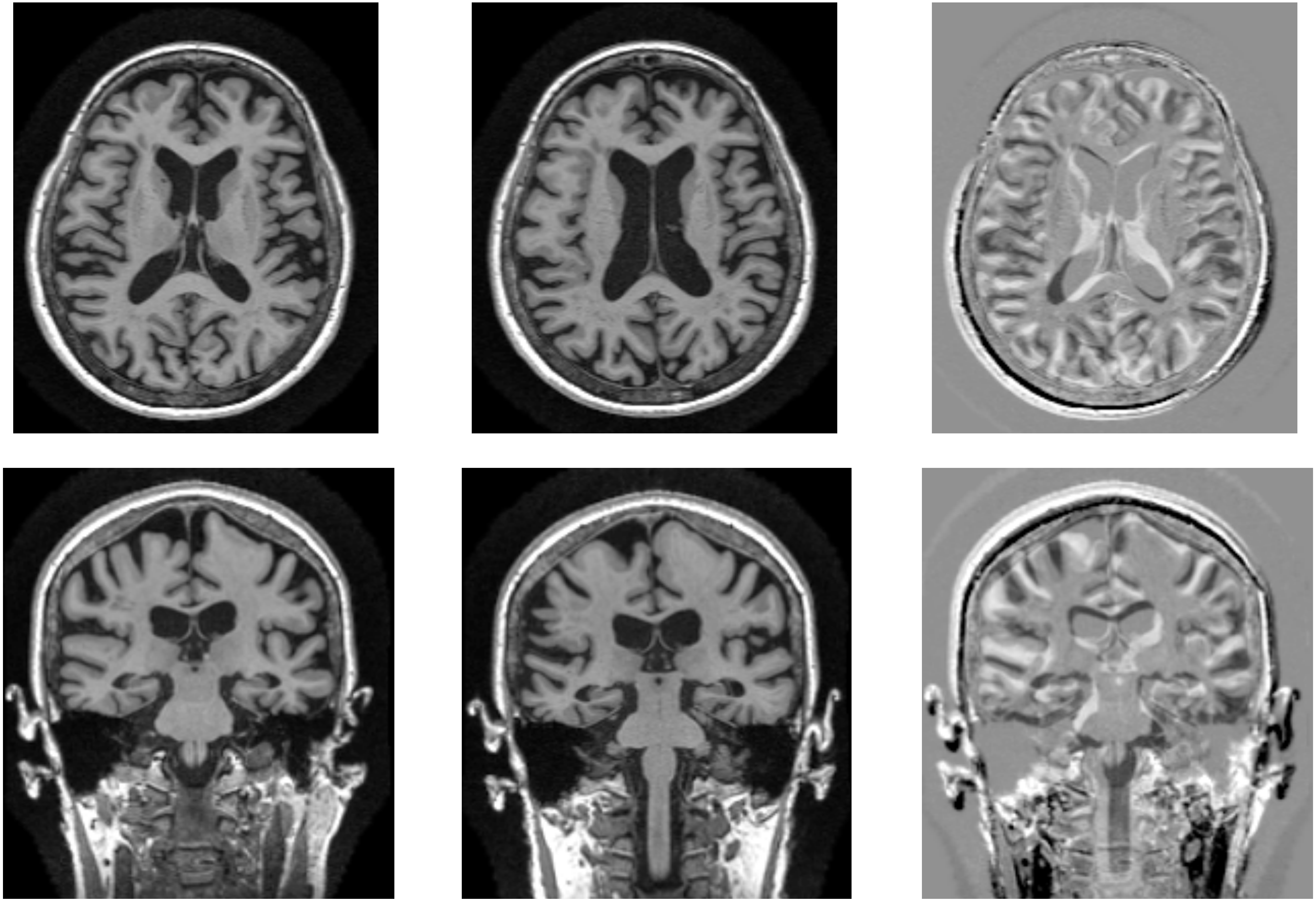
Sum Squared Differences

$$\nabla SSD_n = -2(A_n - B_n(\mathbf{u}))\nabla B_n(\mathbf{u})$$

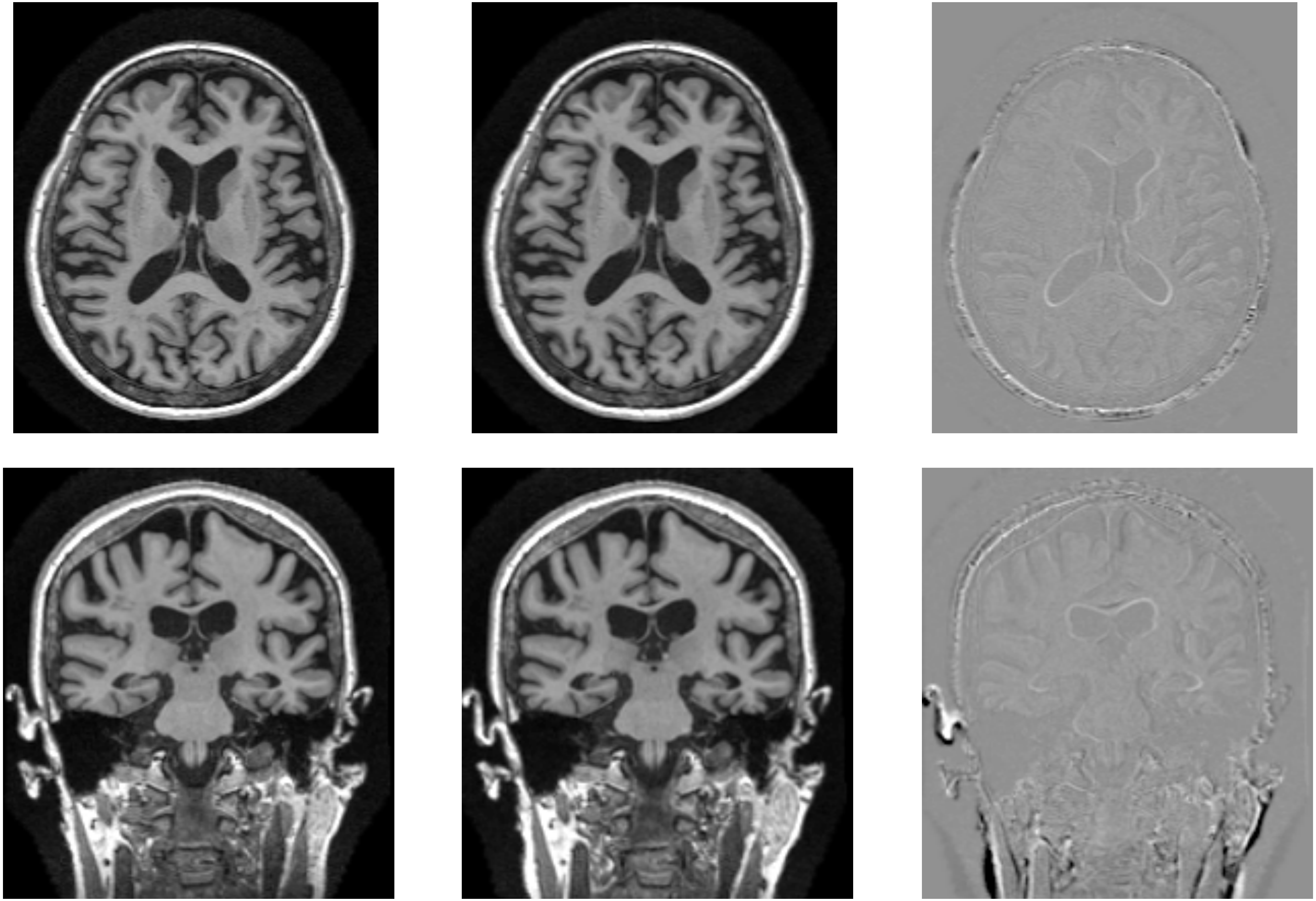
$\frac{\delta \mathrm{SSD}}{\delta x}$



$\frac{\delta \mathrm{SSD}}{\delta y}$

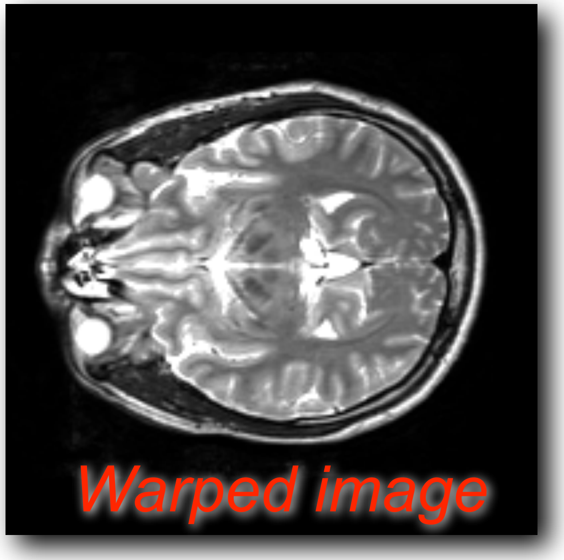


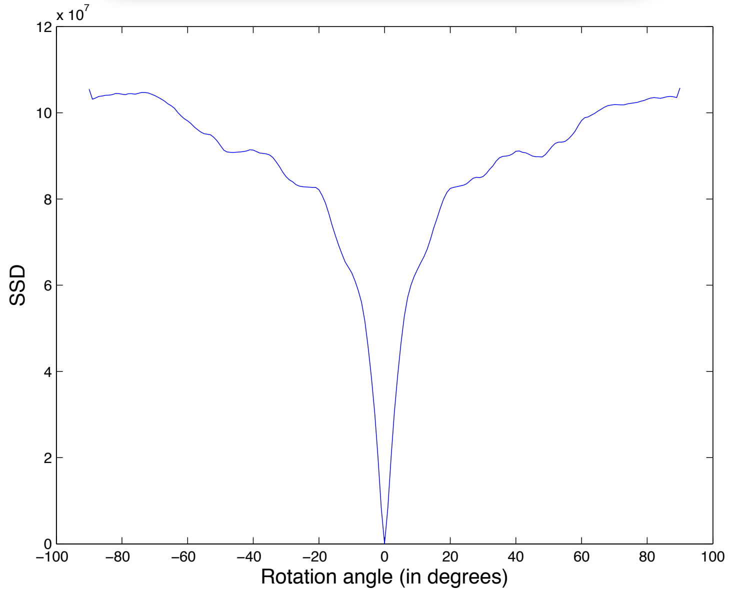
Baseline scan | Follow-up scan | Difference image

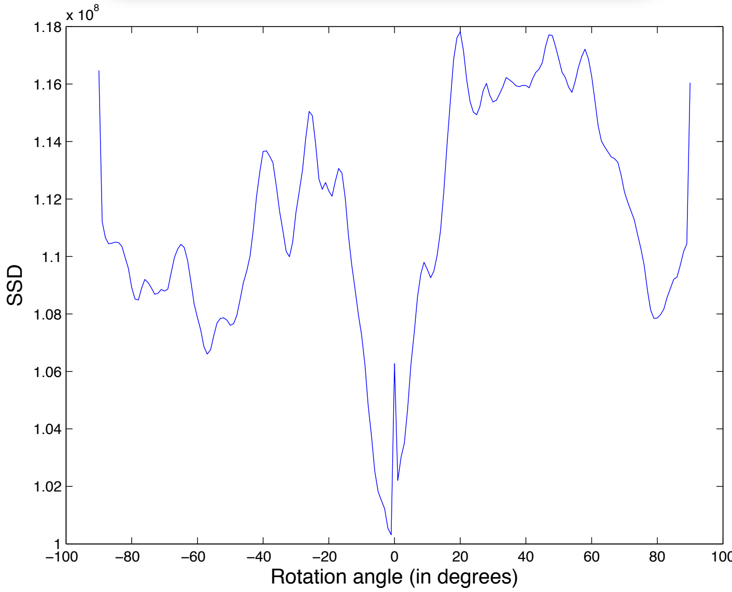


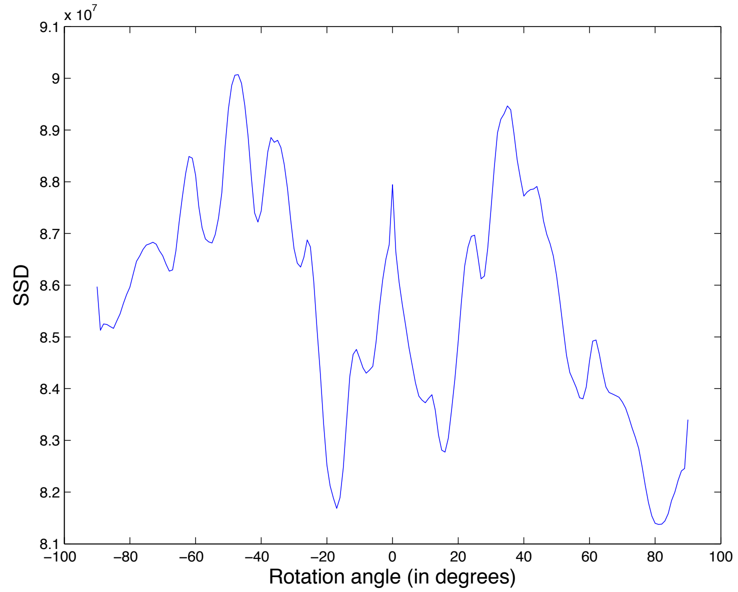
Baseline scan | Follow-up scan | Difference image

Sum Squared Differences









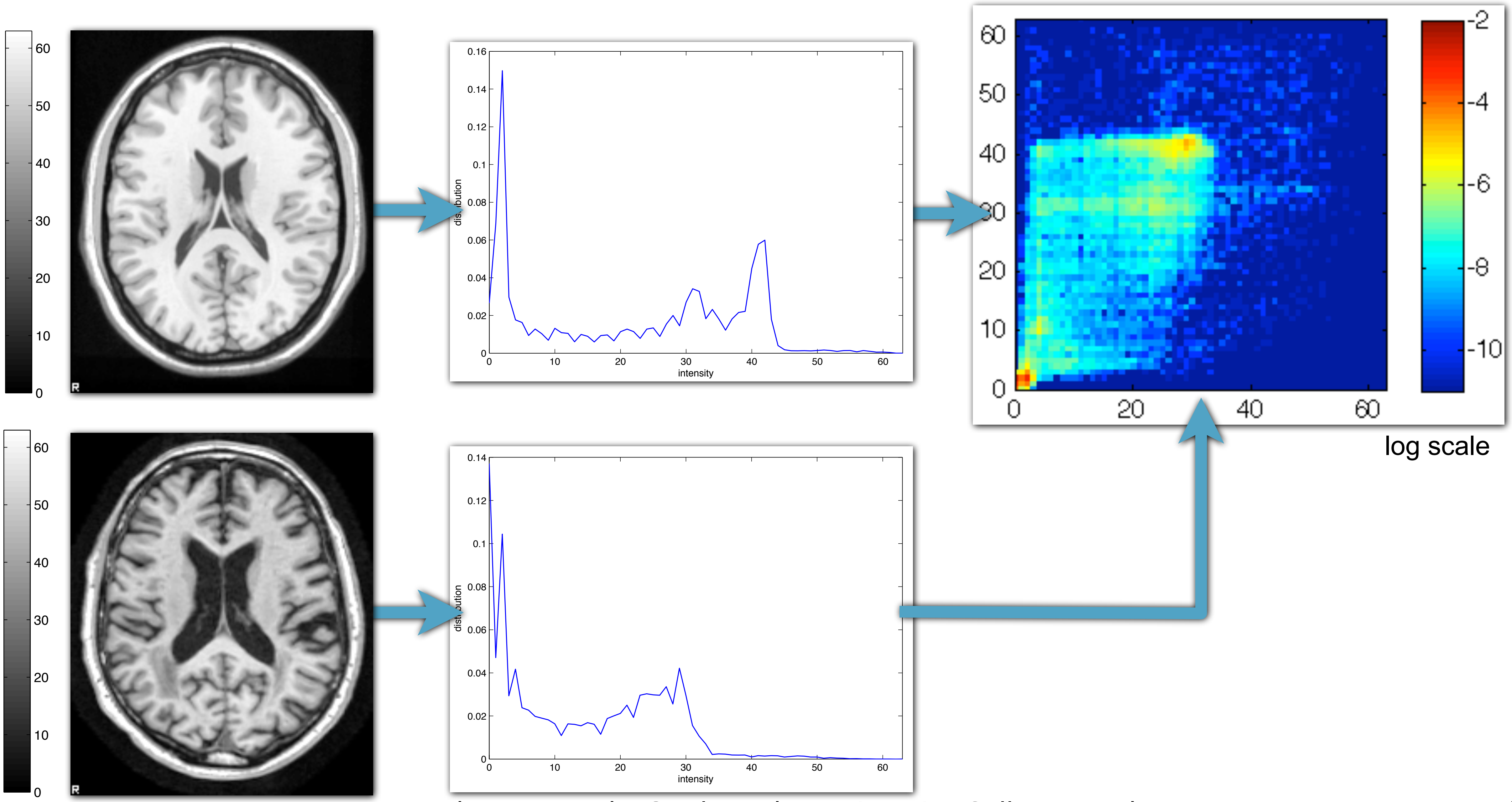
Normalised Cross Correlation

- Linear relationship between the images intensities

$$NCC = \frac{1}{N\sigma_A\sigma_B}\sum_n^N \left[(A_n - \bar{A})(B_n(\mathbf{u}) - \bar{B}(\mathbf{u}))\right]$$

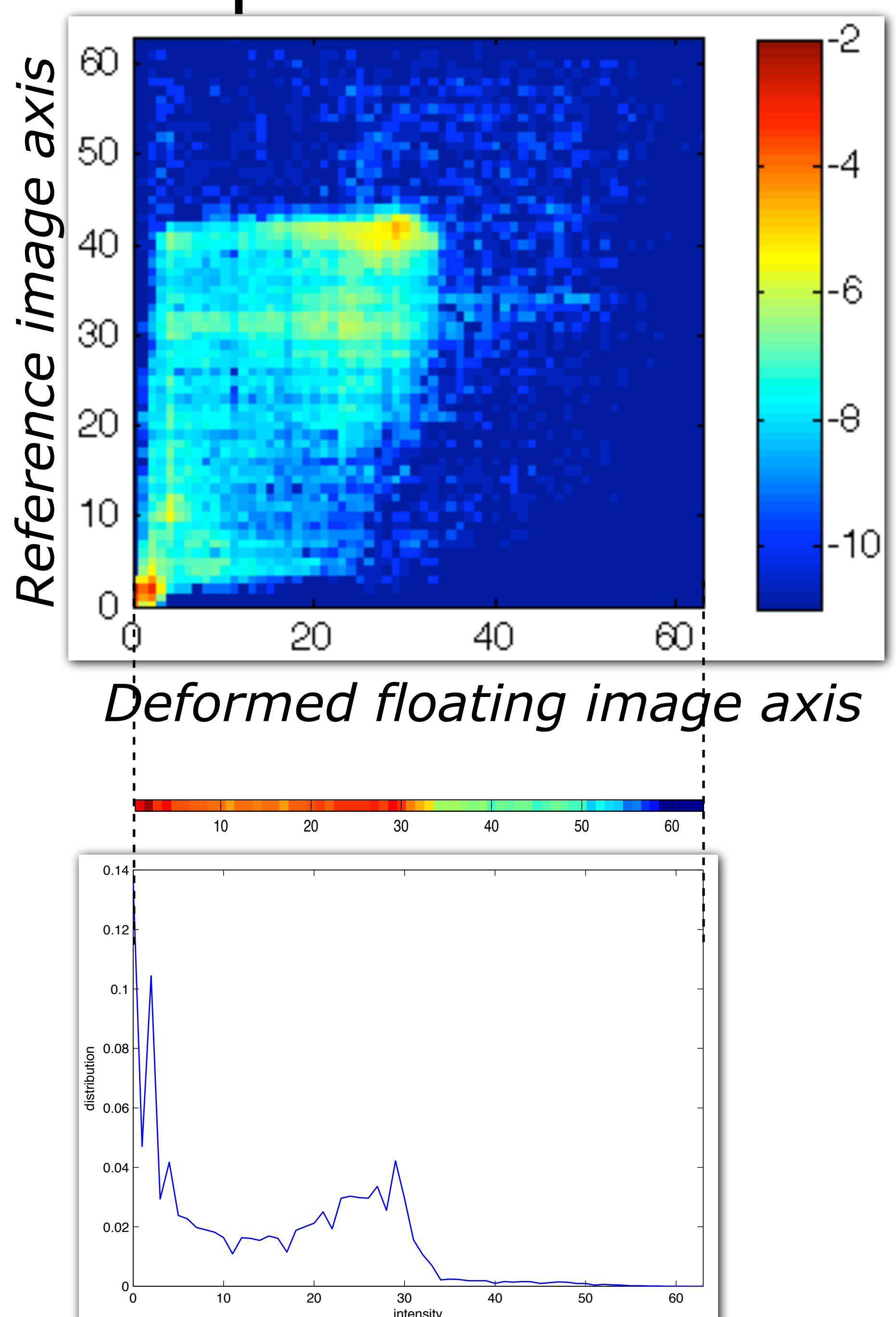
Joint histogram

- Joint histogram
  - intensity paired-distribution of two images



Marginal and joint entropies

- Computation from the joint histogram



  - Joint entropy

$$H(R,F(\mathbf{T})) = -\sum_{r=0}^{bin-1}\sum_{f=0}^{bin-1} p(r,f)\times log(p(r,f))$$

  - Reference image entropy

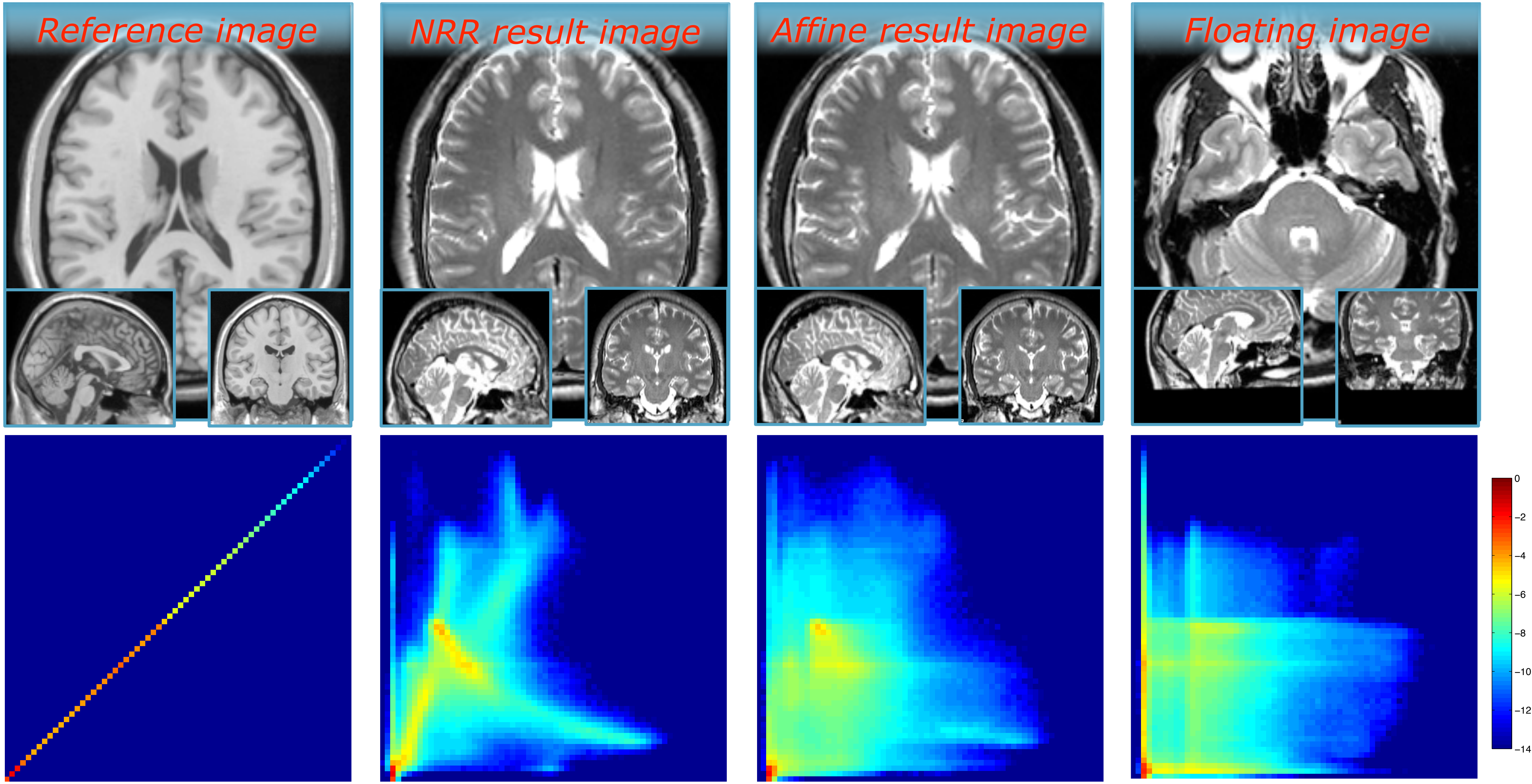
$$H(R) = -\sum_{r=0}^{bin-1} p(r)\times log(p(r))$$

  - Deformed floating image entropy

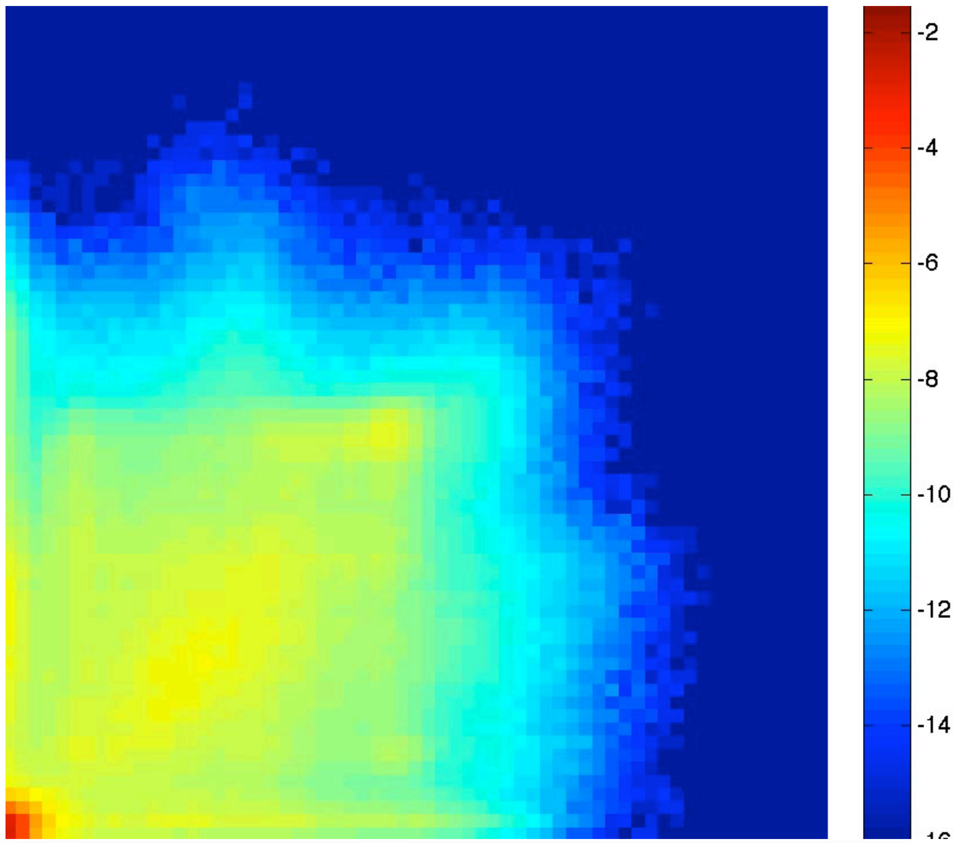
$$H(F(\mathbf{T})) = -\sum_{f=0}^{bin-1} p(f)\times log(p(f))$$

Joint Histogram - Example

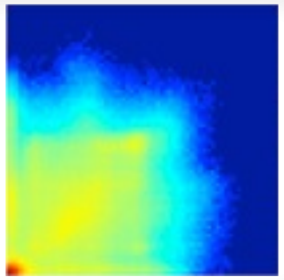
- Example of multi-modal application

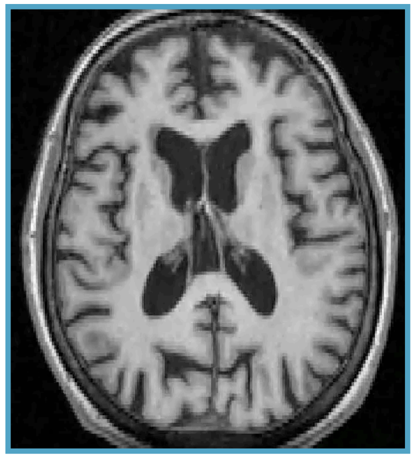


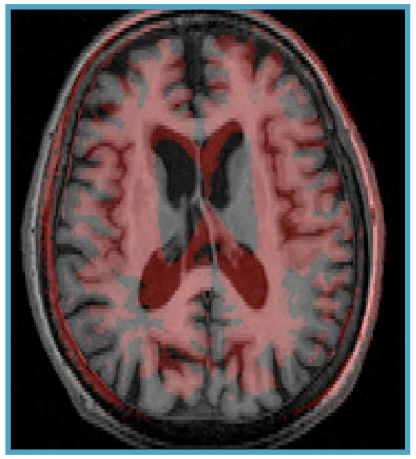
Joint Histogram - Example



Initial joint histogram

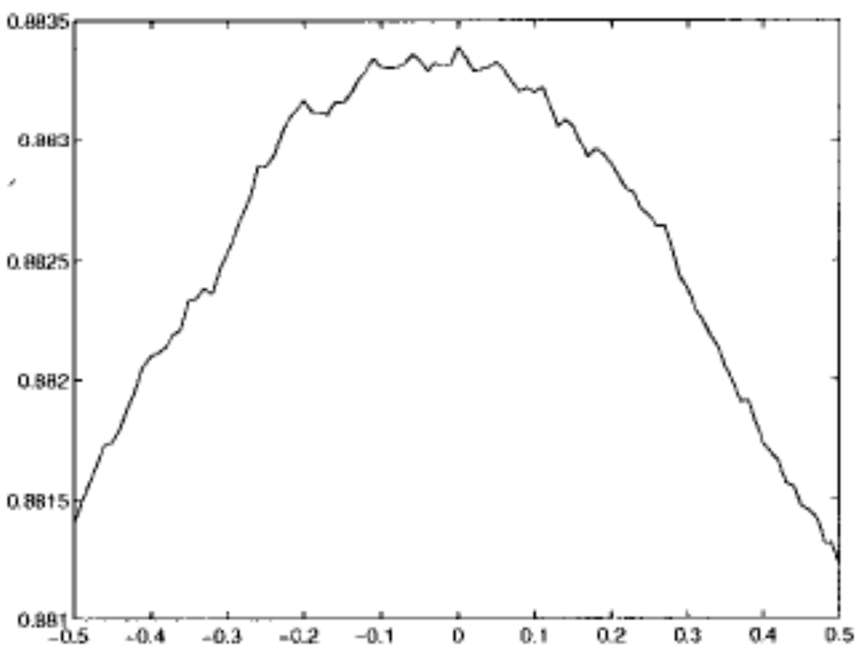


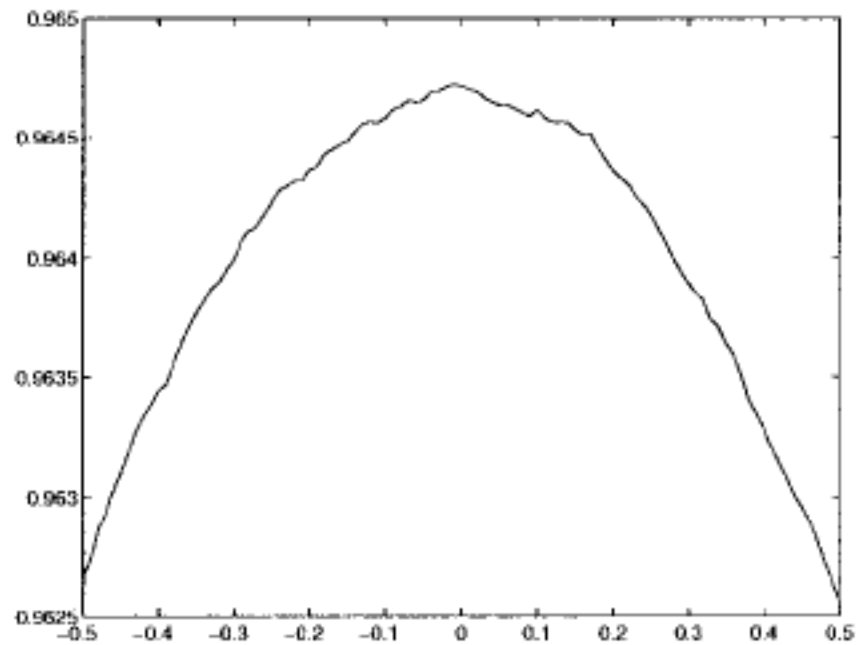


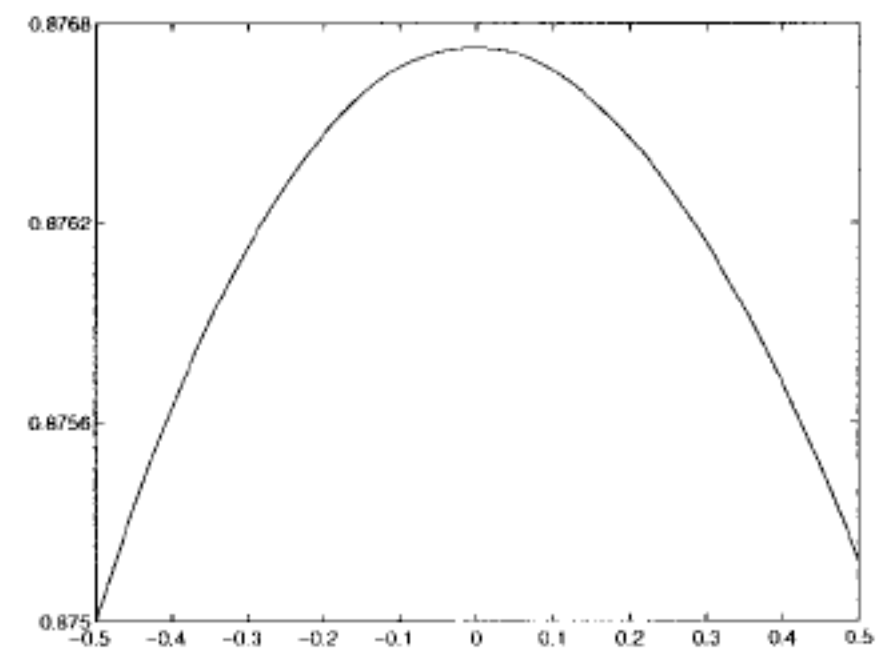


How to fill a joint histogram

- Floating image interpolation
- Nearest neighbour
- Partial volume
- Parzen windowing

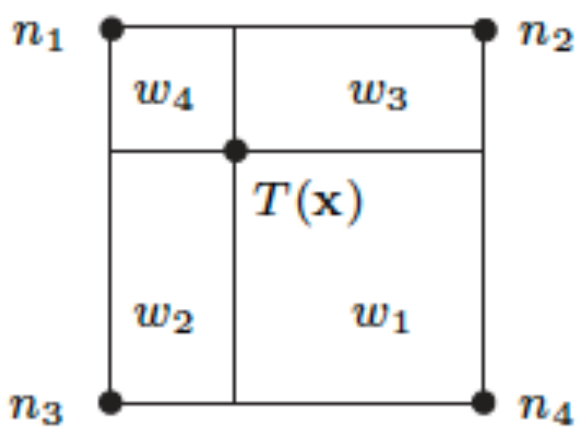






(Generalised) Partial Volume

- Interpolate the intensity value to nearest neighbours



- Is it better than the image interpolation techniques?

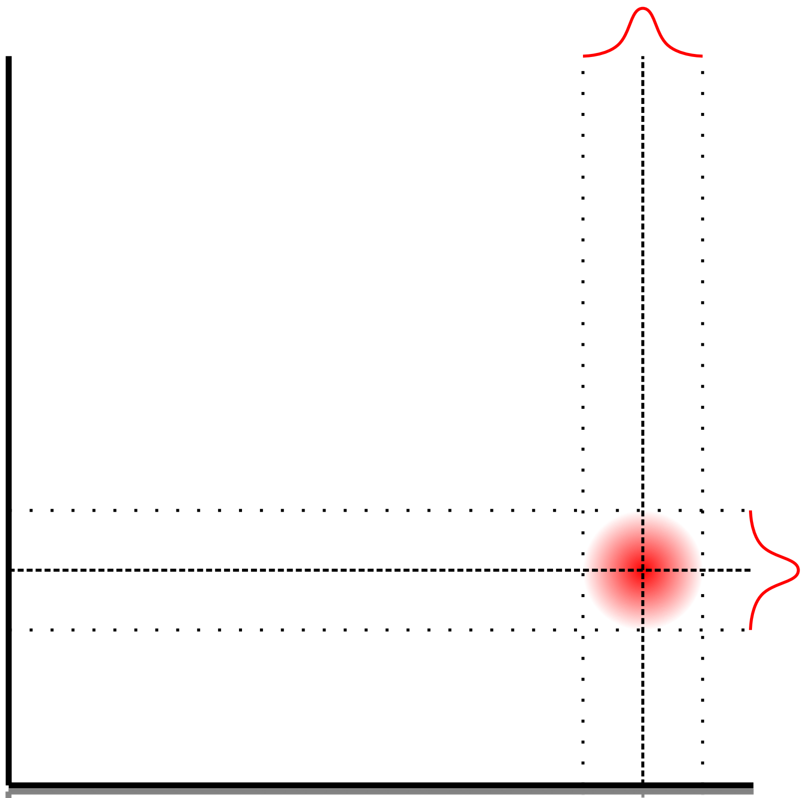
- It has been "shown" as more accurate
- Smoother estimates of derivatives can be estimated

- Lead to Generalised Partial Volume

Parzen Windowing

- A different method to fill a joint histogram H

$$\mathbf{H}(r,f) = \sum_{\vec{x}\in\Omega} \beta_r^3(R(\vec{x});r)\beta_f^3(F(\mathbf{T}(\vec{x}));f)$$



- Is it better than the (Generalised) Partial Volume technique?

- It does generate "new" source intensity
- It has been "shown" as more accurate
- Smooth derivatives can be found

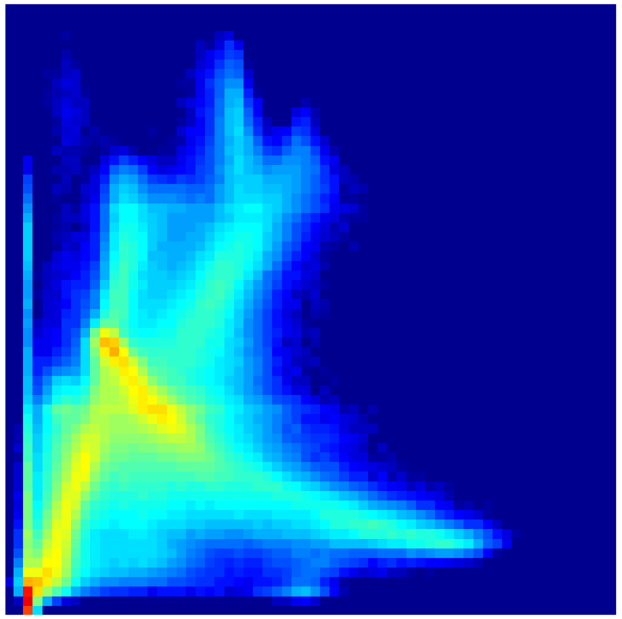
Information theoretic similarity measures

- Joint Entropy
- Kullback Liebler divergence
- Mutual Information
- Normalised Mutual Information and Symmetric Uncertainty

Information theoretic similarity measures

- Joint Entropy

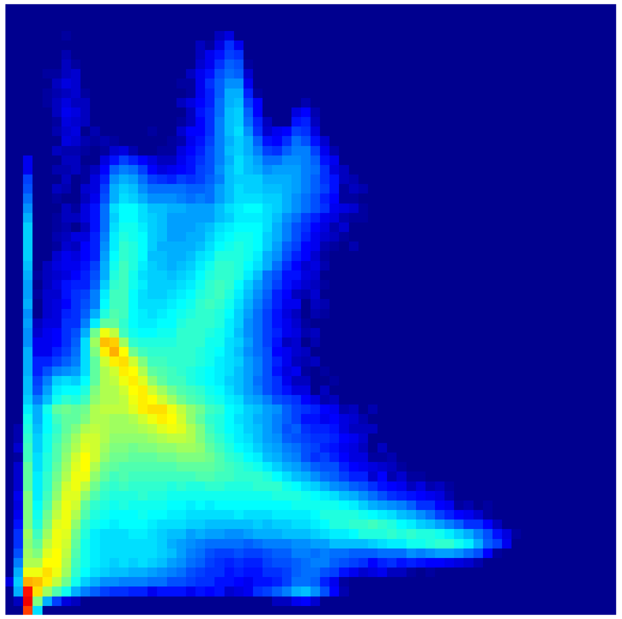
$$H_{AB} = -\sum_{ij} p_{AB}(i,j) \log p_{AB}(i,j)$$



Information theoretic similarity measures

- Kullback-Leibler Divergence

$$KLD_{AB} = \sum_i p_A(i) \log \frac{p_A(i)}{p_B(i)}$$

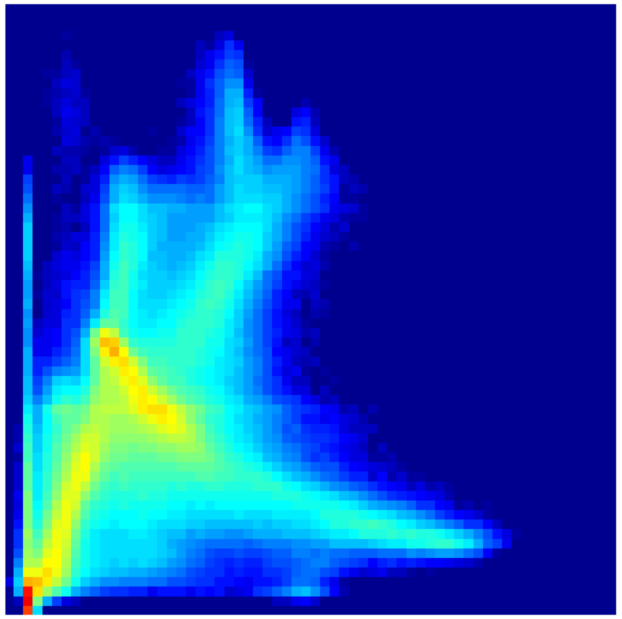


Information theoretic similarity measures

- Mutual Information

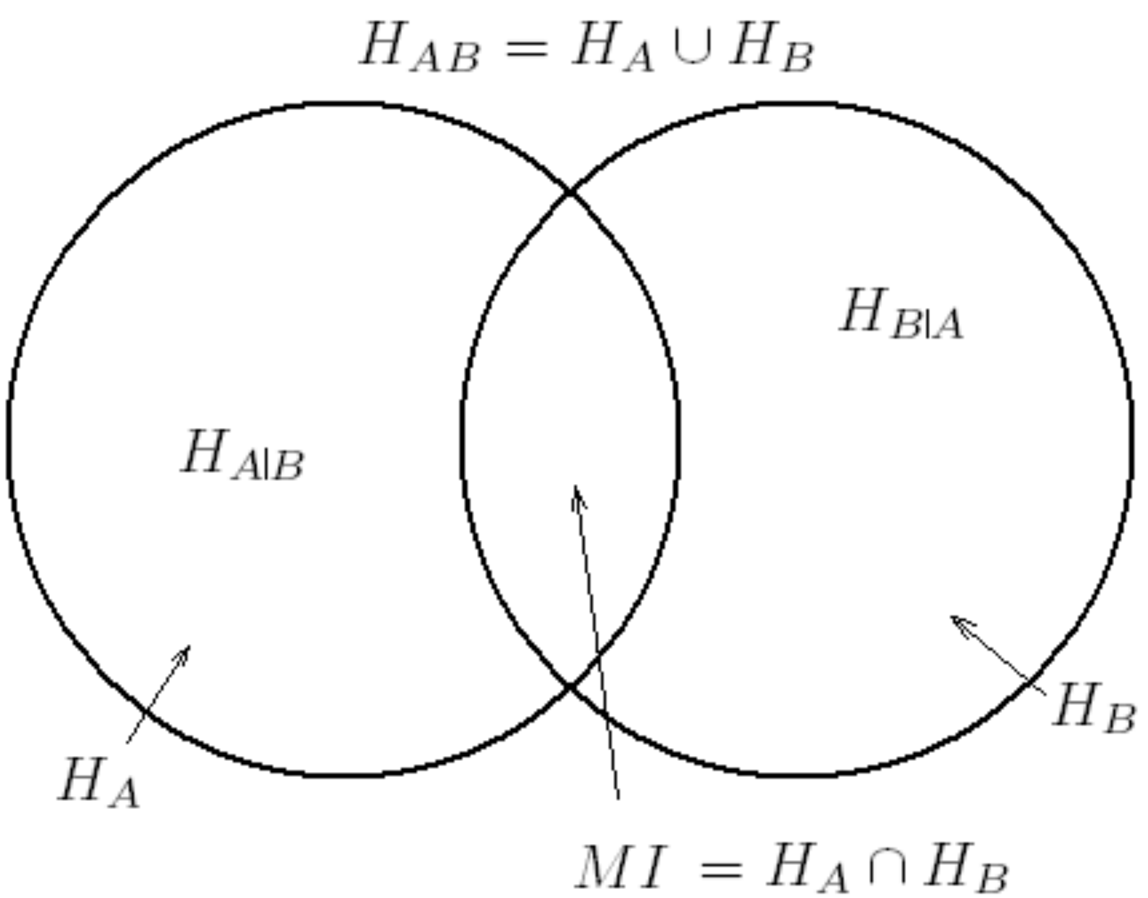
$$MI_{AB} = \sum_{ij} p_{AB}(i,j) \log \frac{p_{AB}(i,j)}{p_A(i) p_B(j)}$$

$$MI_{AB} = H_A + H_B - H_{AB}$$



Information theoretic similarity measures

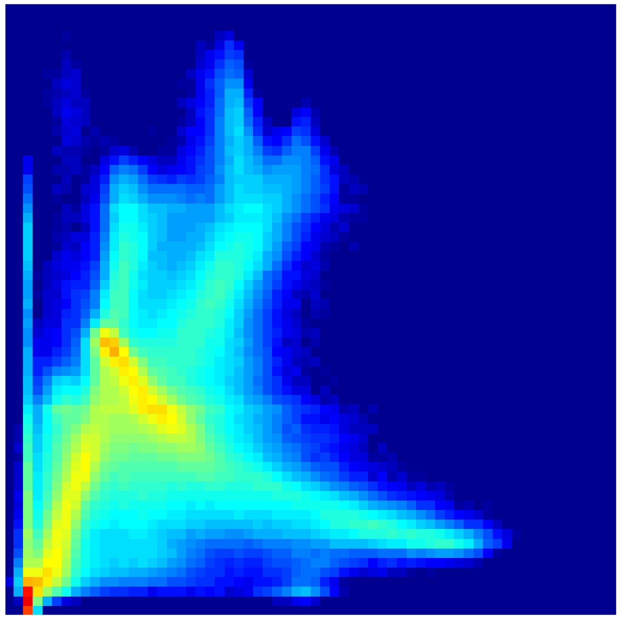
- Mutual Information



Information theoretic similarity measures

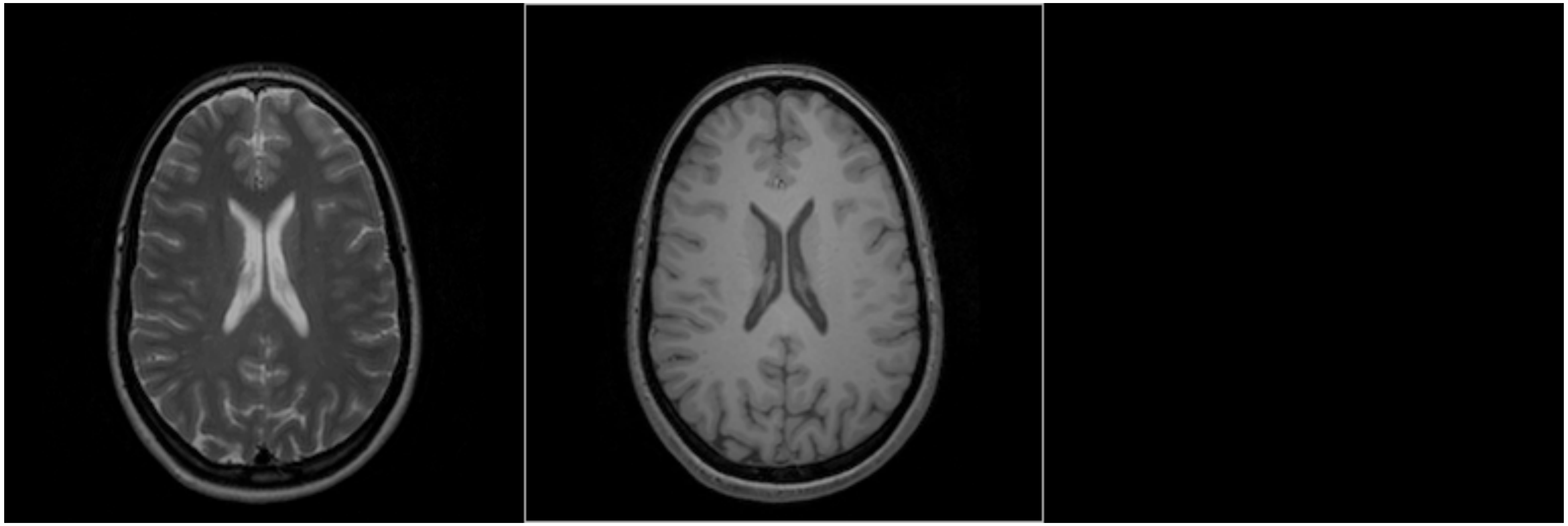
- Normalised Mutual Information

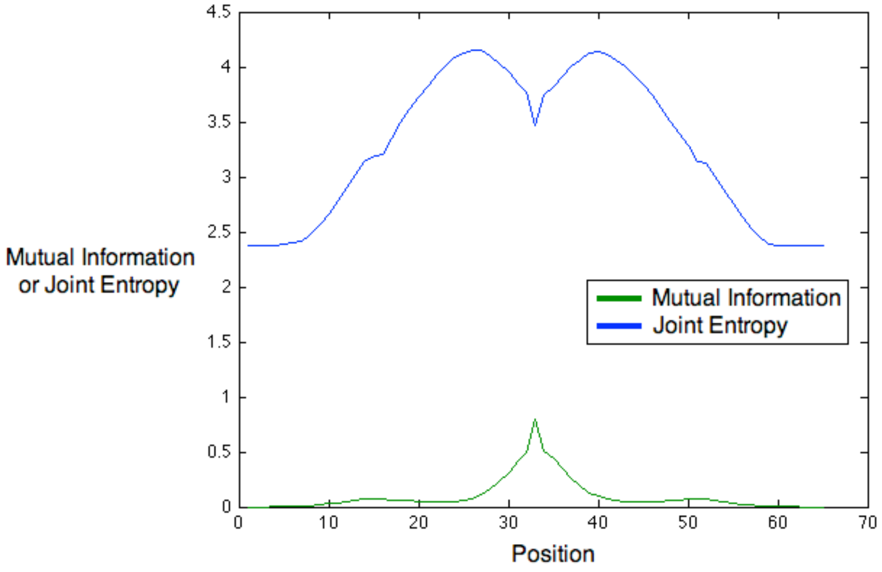
$$NMI_{AB} = \frac{H_A + H_B}{H_{AB}}$$



Joint Entropy vs. Mutual Information

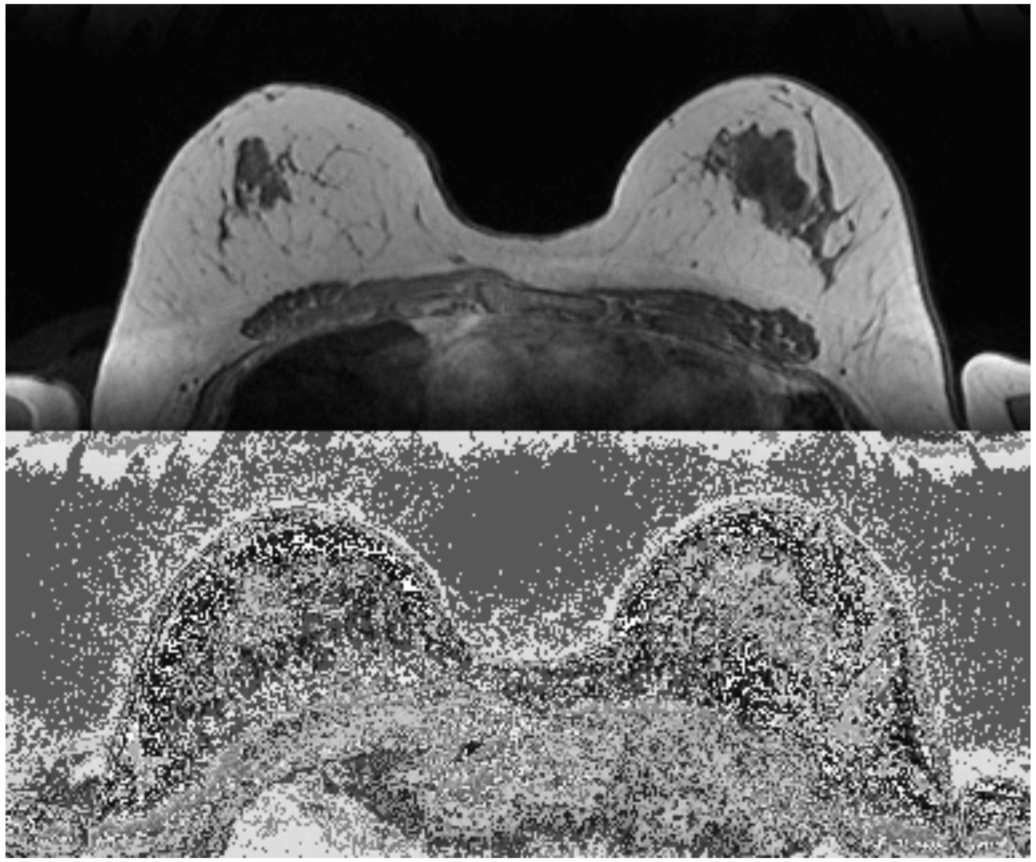
- However …

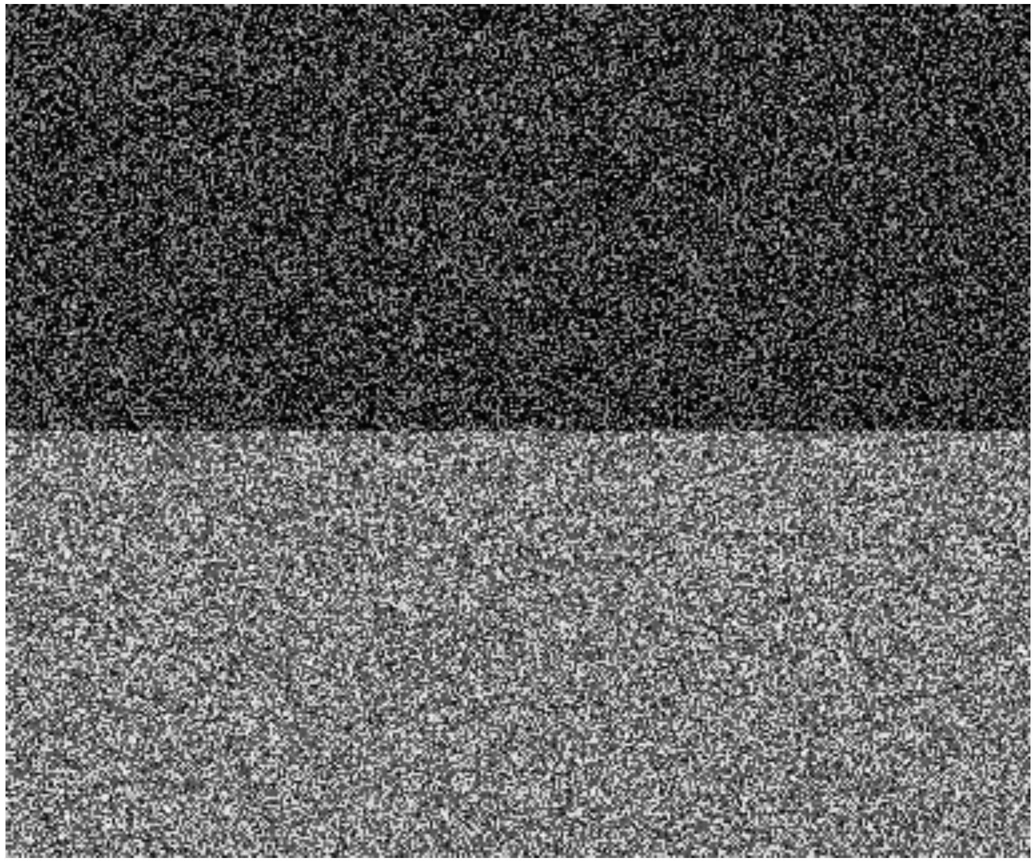




Normalised Mutual Information

- MI / NMI has some known drawbacks ...
  - Global measure - prone to local minima
  - No use of spatial information





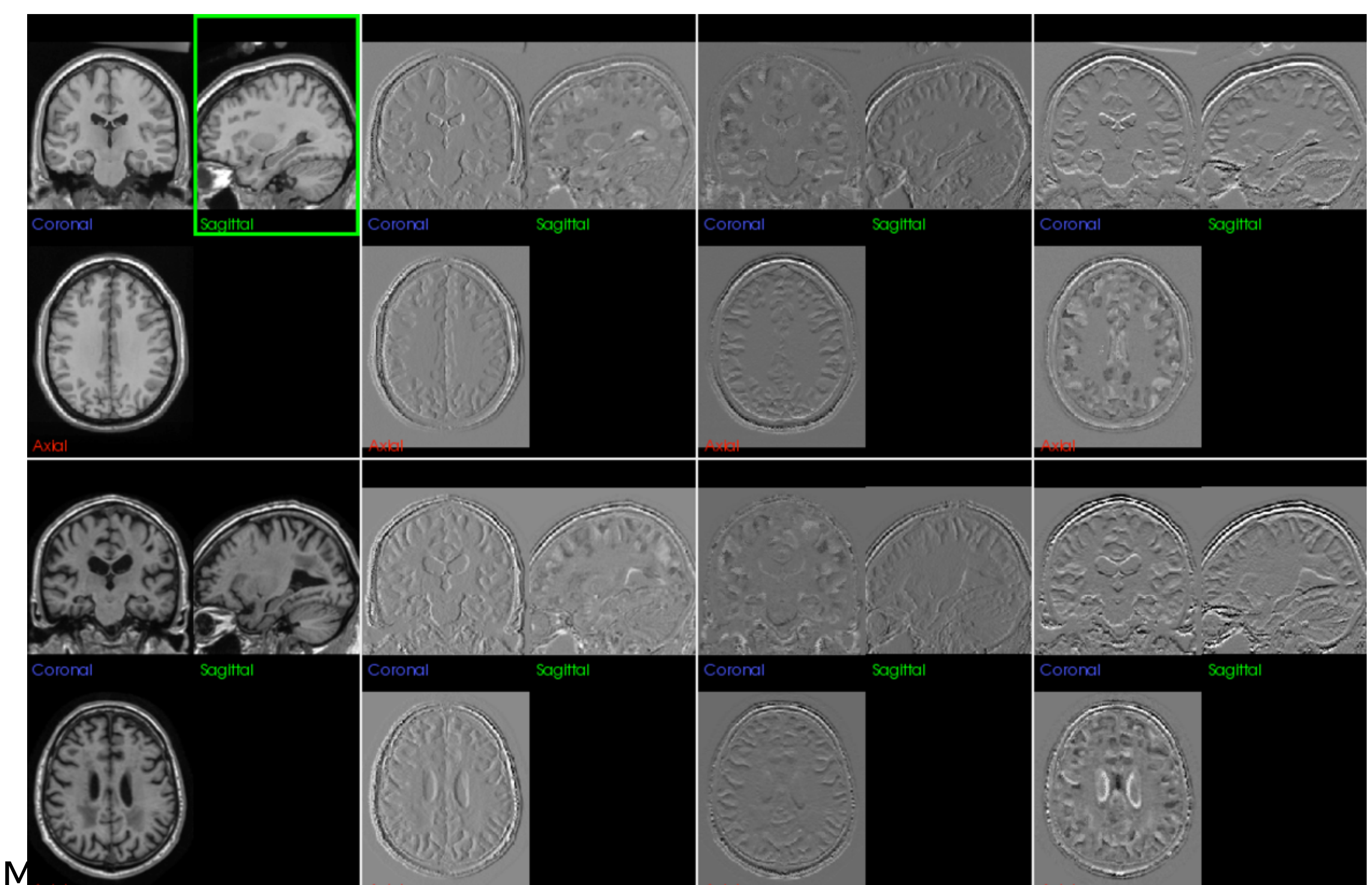
Locally normalised image similarity

- Locally normalised sum-of-squared difference
- Locally normalised cross correlation
- Locally normalised entropy

- Dependent on patch size and shape.
- Robust to global intensity variation
- Computationally more expensive

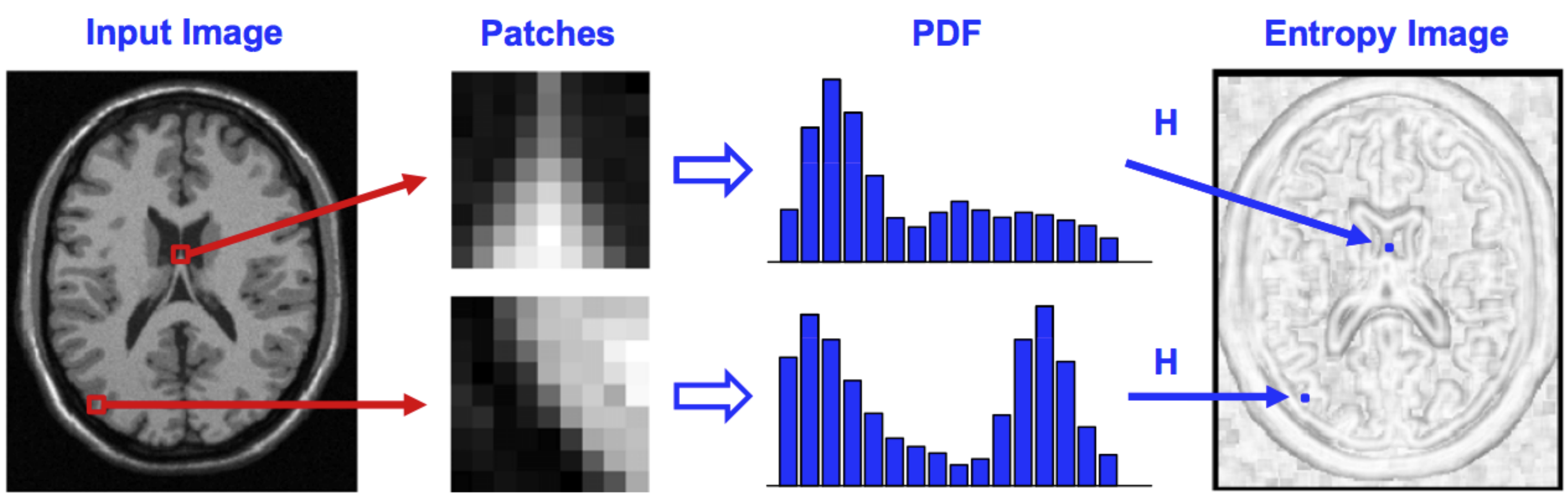
Feature-based measure of similarity

- **<u>Gradient orientation</u>** [Pluim et al., 2000]
- Local entropy [Wachinger and Navab, 2012]
- Hierarchical attribute [Shen and Davatzikos,2003] (HAMMER)
- Gabor attribute [Ou *et al.*, 2010] (DRAMMS)
- SIFT
- MIND



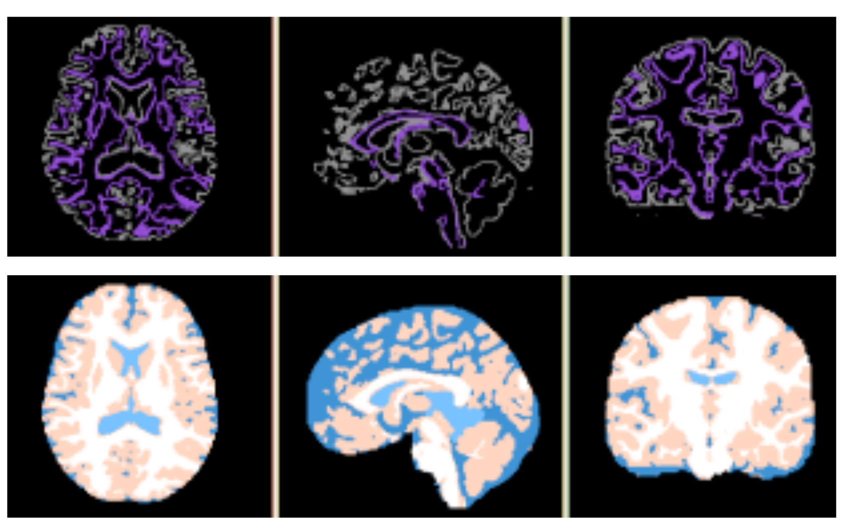
Feature-based measure of similarity

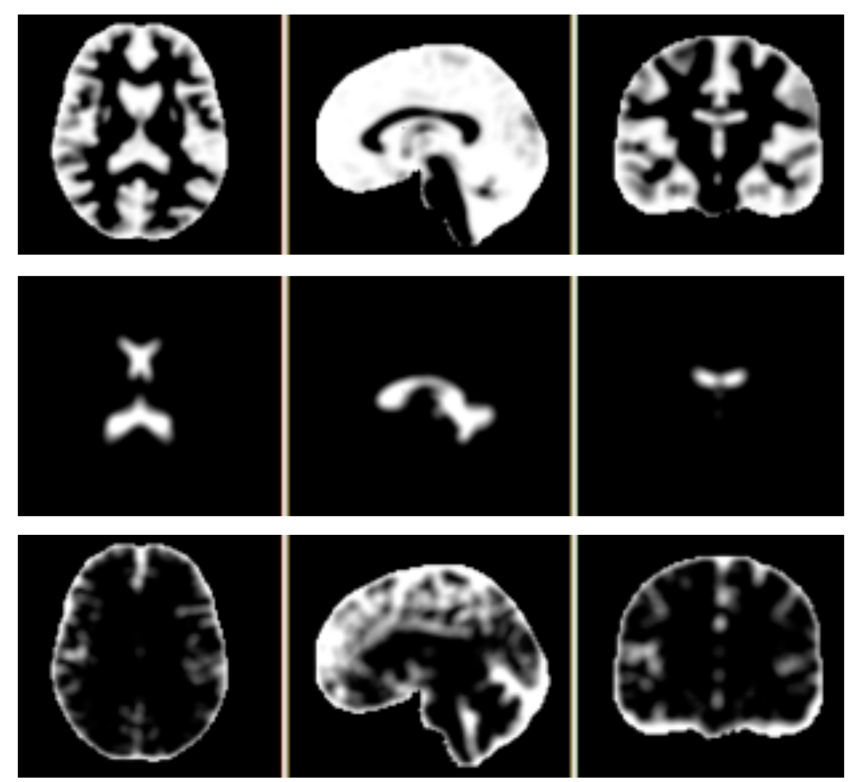
- Gradient orientation [Pluim et al., 2000]
- **Local entropy** [Wachinger and Navab, 2012]
- Hierarchical attribute [Shen and Davatzikos,2003] (HAMMER)
- Gabor attribute [Ou *et al.*, 2010] (DRAMMS)
- SIFT [David Lowe *et al.*, 1999]
- MIND



Feature-based measure of similarity

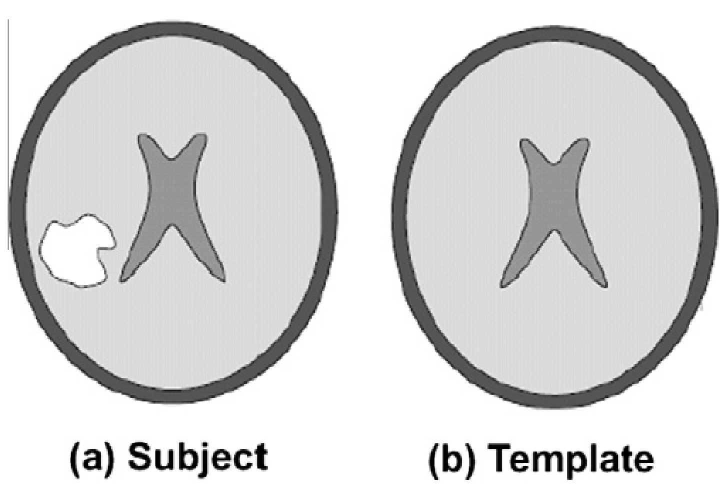
- Gradient orientation [Pluim et al., 2000]
- Local entropy [Wachinger and Navab, 2012]
- **<u>Hierarchical attribute</u>** [Shen and Davatzikos,2003] (HAMMER)
- Gabor attribute [Ou *et al.*, 2010] (DRAMMS)
- SIFT [David Lowe *et al.,* 1999]
- MIND [Heinrich 2012]

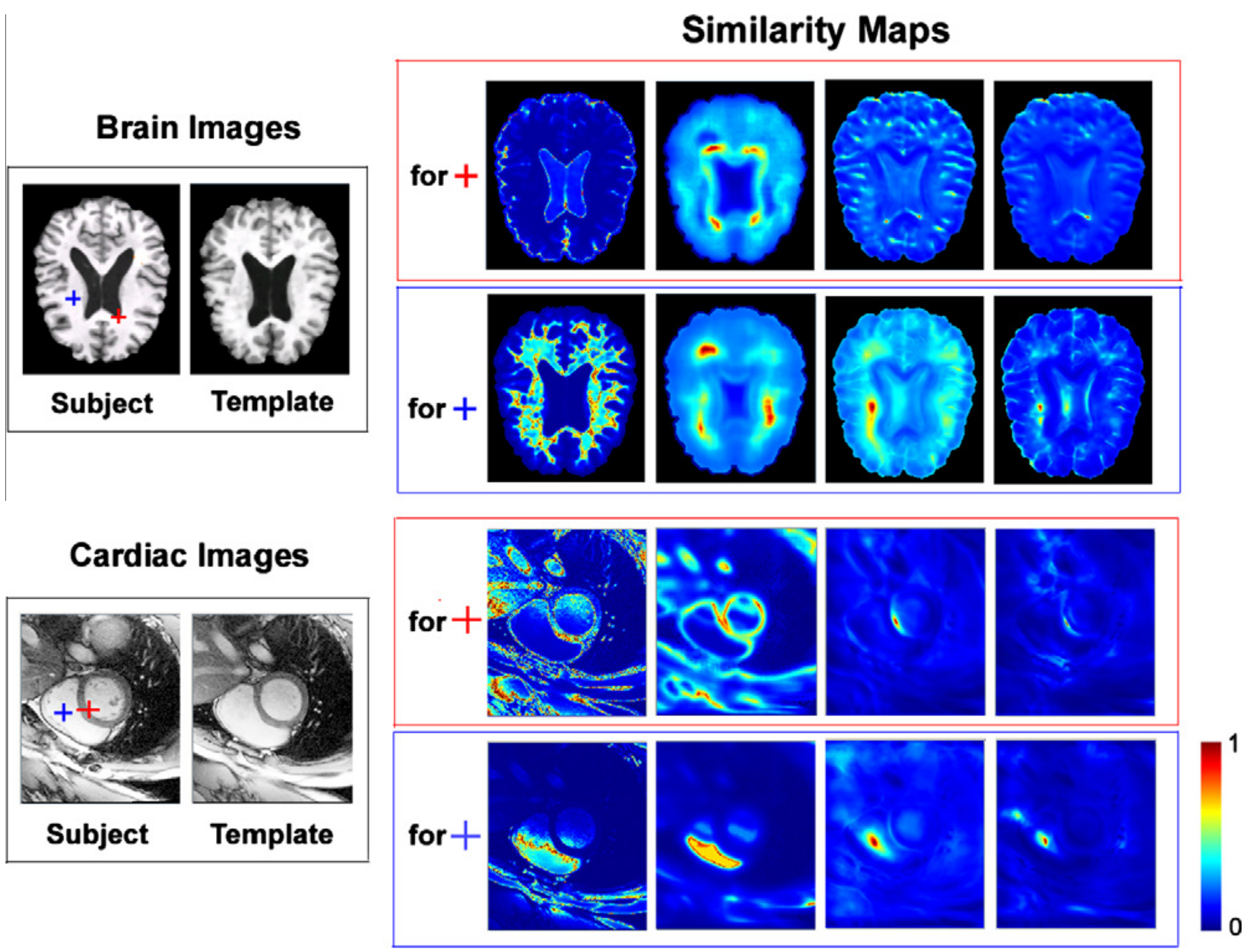




Feature-based measure of similarity

- Gradient orientation [Pluim et al., 2000]
- Local entropy [Wachinger and Navab, 2012]
- Hierarchical attribute [Shen and Davatzikos,2003] (HAMMER)
- **<u>Gabor attribute</u>** [Ou *et al.*, 2010] (DRAMMS)
- SIFT [David Lowe *et al.,* 1999]
- MIND [Heinrich 2012]



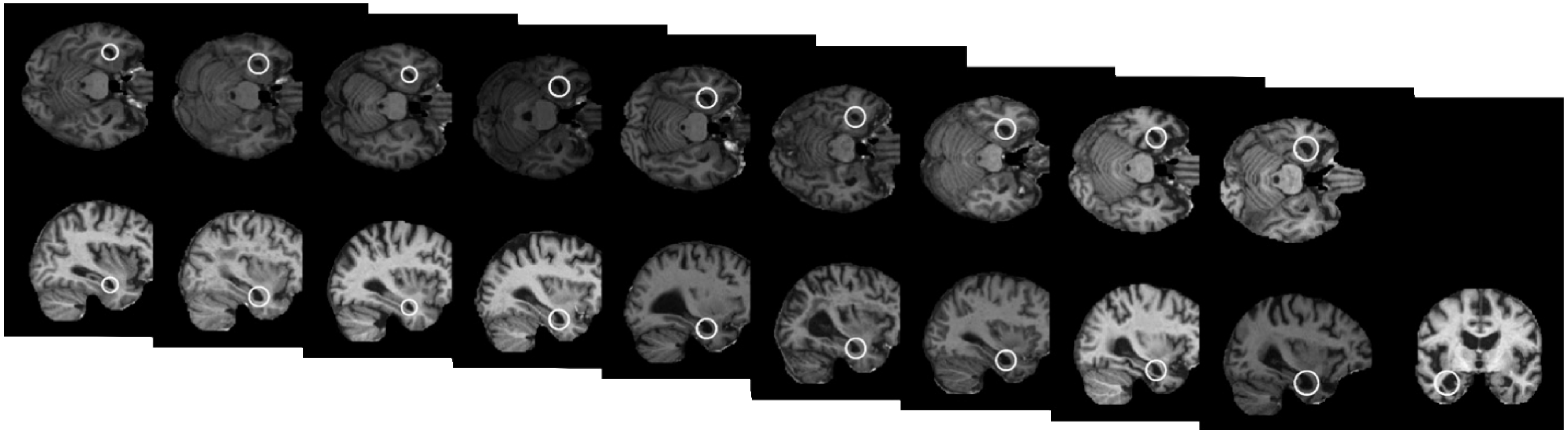


Feature-based measure of similarity

- Gradient orientation [Pluim et al., 2000]
- Local entropy [Wachinger and Navab, 2012]
- Hierarchical attribute [Shen and Davatzikos,2003] (HAMMER)
- Gabor attribute [Ou *et al.*, 2010] (DRAMMS)
- **SIFT** [David Lowe *et al.,* 1999]
- MIND [Heinrich 2012]

Feature-based measure of similarity

- Gradient orientation [Pluim et al., 2000]
- Local entropy [Wachinger and Navab, 2012]
- Hierarchical attribute [Shen and Davatzikos,2003] (HAMMER)
- Gabor attribute [Ou *et al.*, 2010] (DRAMMS)
- **SIFT** [David Lowe *et al.,* 1999]
- MIND [Heinrich 2012]

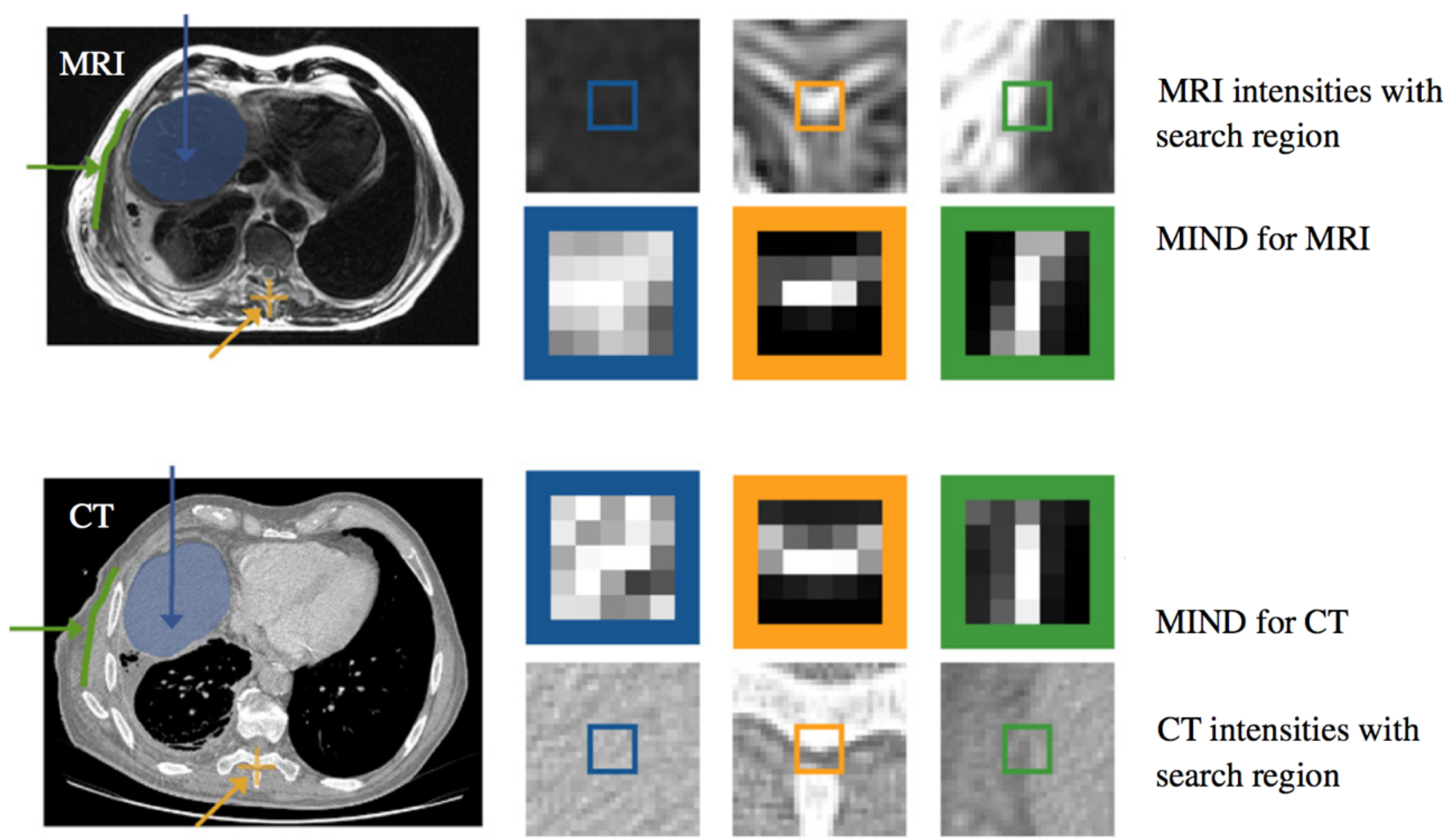


MIND

- **MIND**: **M**odality **I**ndependent **N**eighbo~~u~~rhood **D**escriptor
  - Inspired from de-noising [Buades *et al.*, 2005]
  - Previously applied to medical images [Coupe *et al.*, 2008]

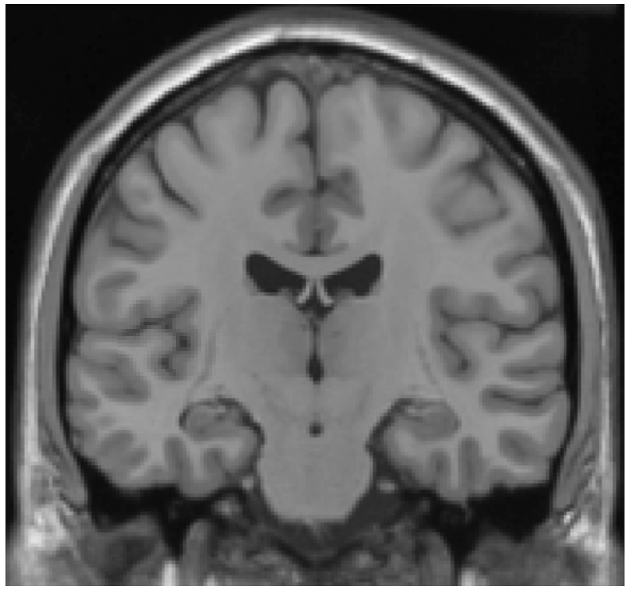
MIND

- **MIND**: **M**odality **I**ndependent **N**eighbo~~u~~rhood **D**escriptor
  - Uses the concept of patches
  - Uses the concept of local similarity



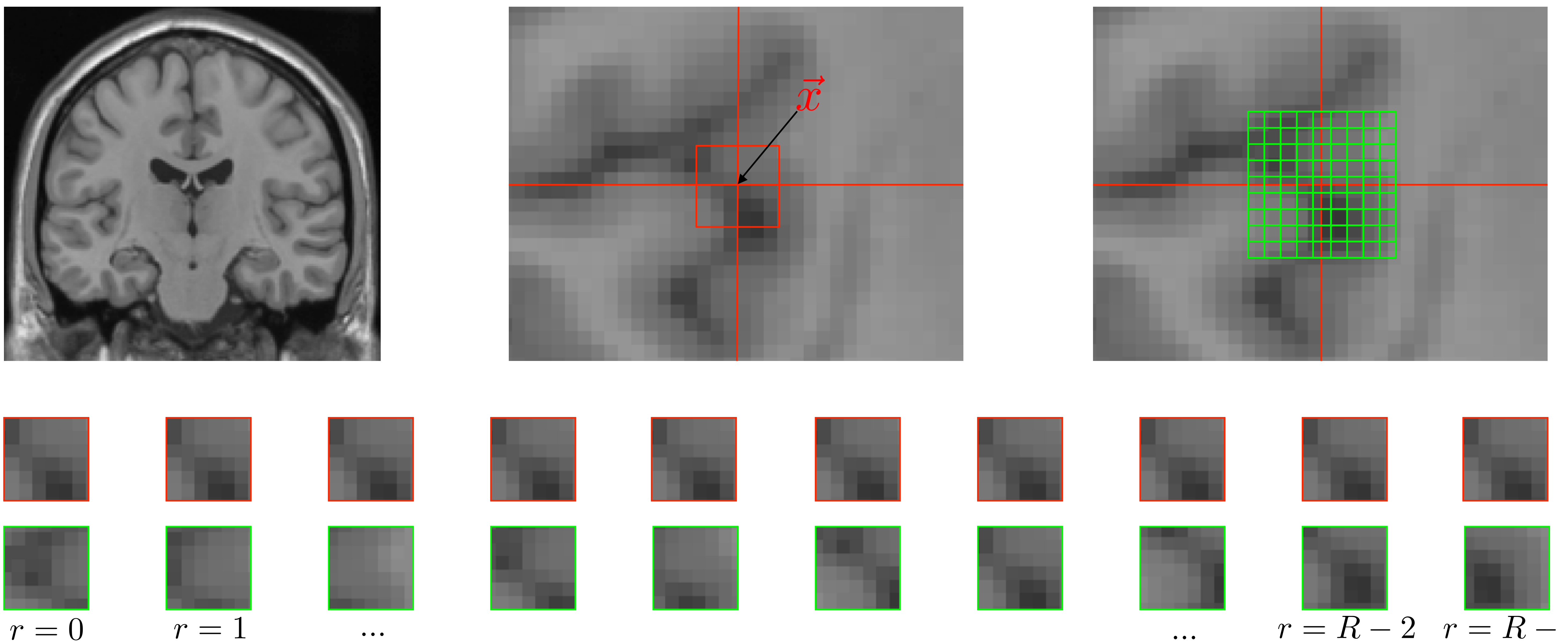
MIND

- **MIND**: How does it work?



MIND

- **MIND**: How does it work?



$$D_p(I, \vec{x}, \vec{x} + \vec{r})$$

MIND

- **MIND**: How does it work?

$$D_p(I, \vec{x}, \vec{x}+r) = \sum_{p \in P} \left(I\left(\vec{x}+p\right) - I(\vec{x}+r+p)\right)^2$$

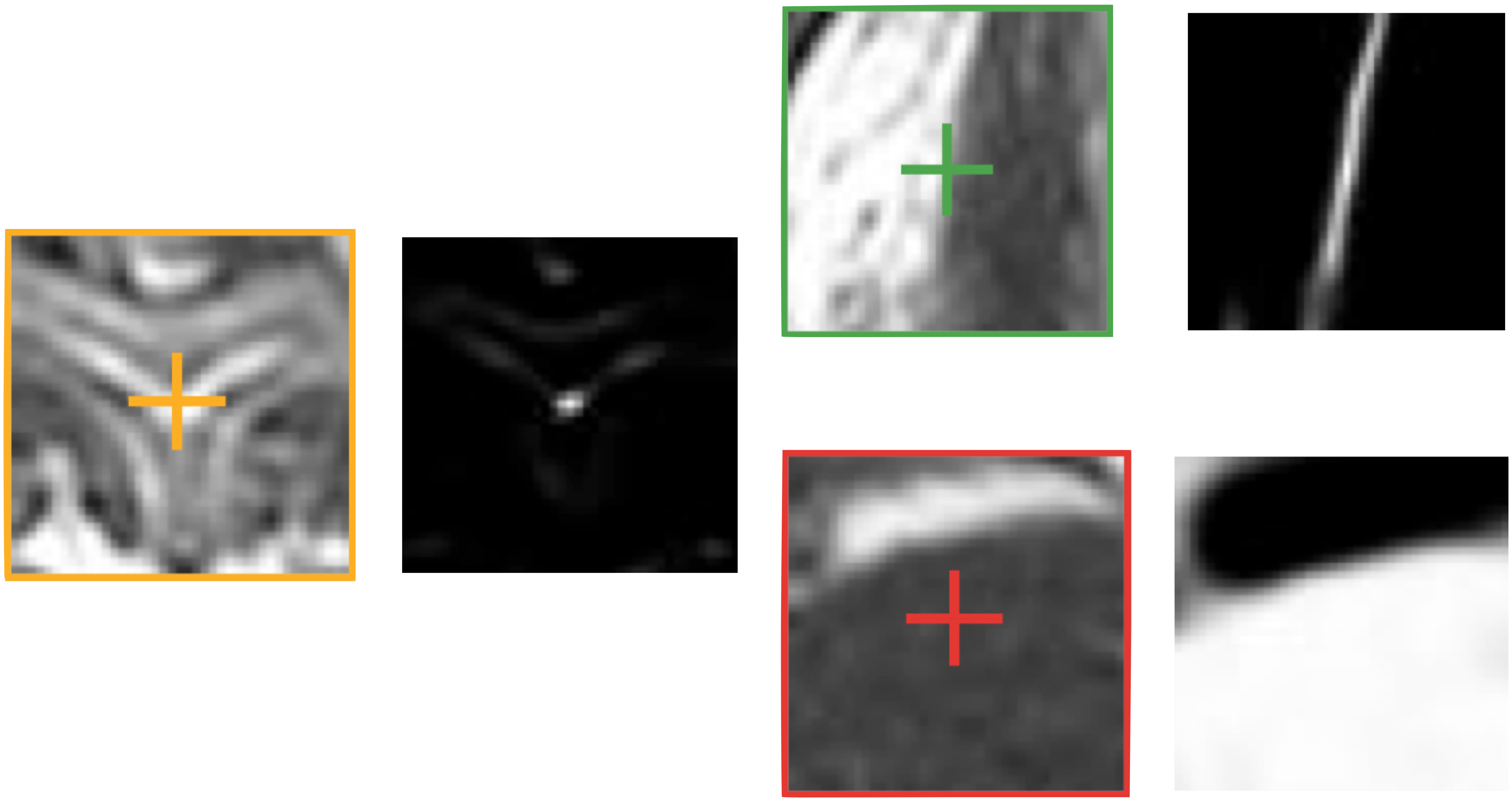
$$\mathrm{MIND}\left(I, \vec{x}, r\right) = \frac{1}{n} \exp\left(\frac{D_p\left(I, \vec{x}, \vec{x}+r\right)}{V\left(I, \vec{x}\right)}\right)$$

$$V(I, \vec{x}) = \frac{1}{6} \sum_{n \in \mathcal{N}} D_p(I, \vec{x}, \vec{x}+n)$$

  - for each voxel, we obtain a vector of R scalar
  - 2 parameters to set: *R* and *P*

MIND

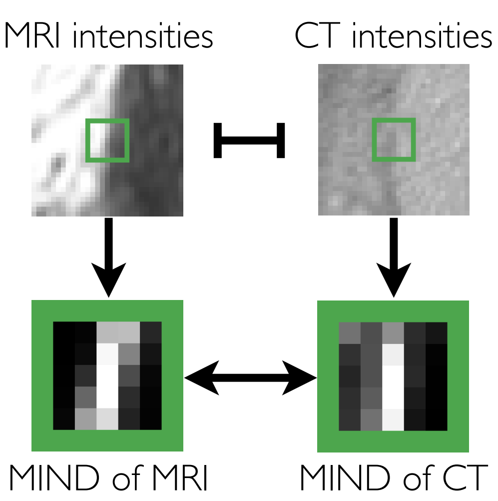
- **MIND**: Examples



MIND

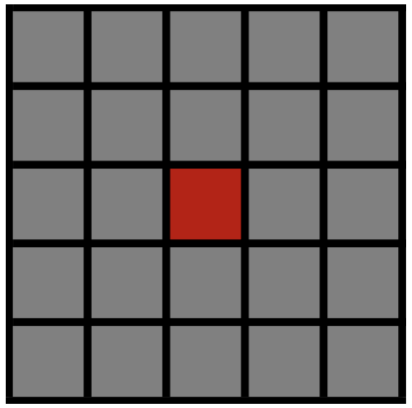
- The MIND vectors are computed at all voxel position and for both images

$$S(I,J,\vec{x}) = \frac{1}{R}\sum_{r\in R}\left|\mathrm{MIND}\left(I,\vec{x},r\right) - \mathrm{MIND}\left(J,\vec{x},r\right)\right|$$

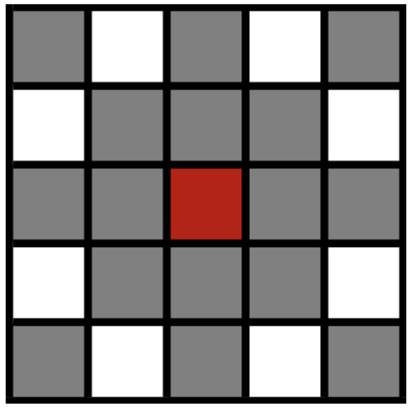


MIND - Further details

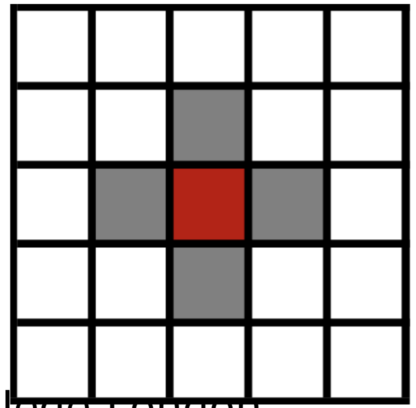
- Spatial search region
  - three ways to define a neighborhood have been tested
    - Dense sampling



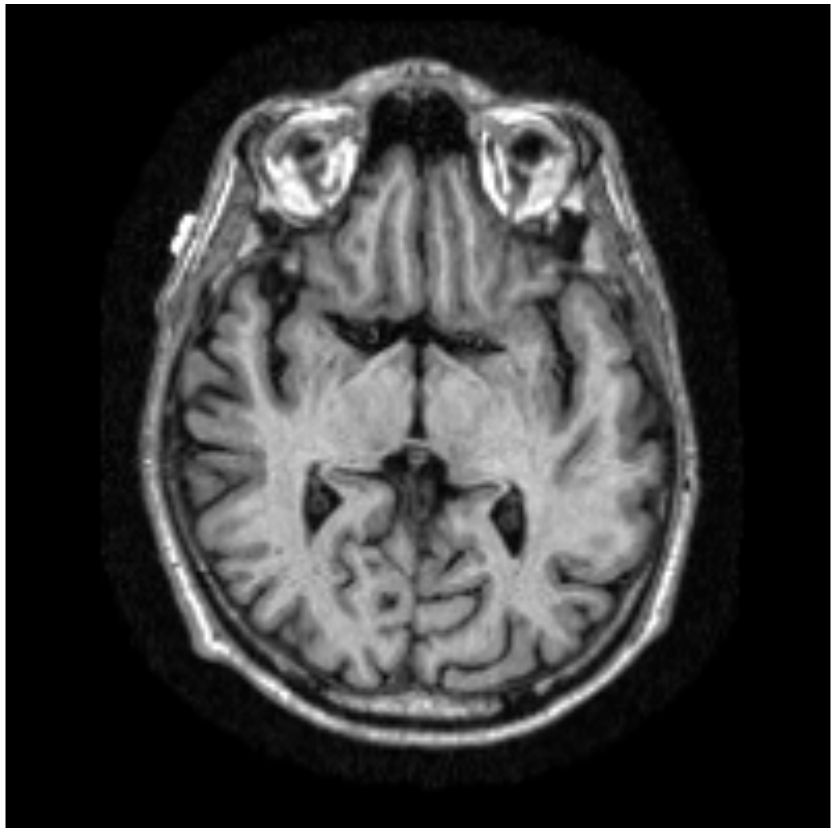
    - Sparse sampling (45º rays)

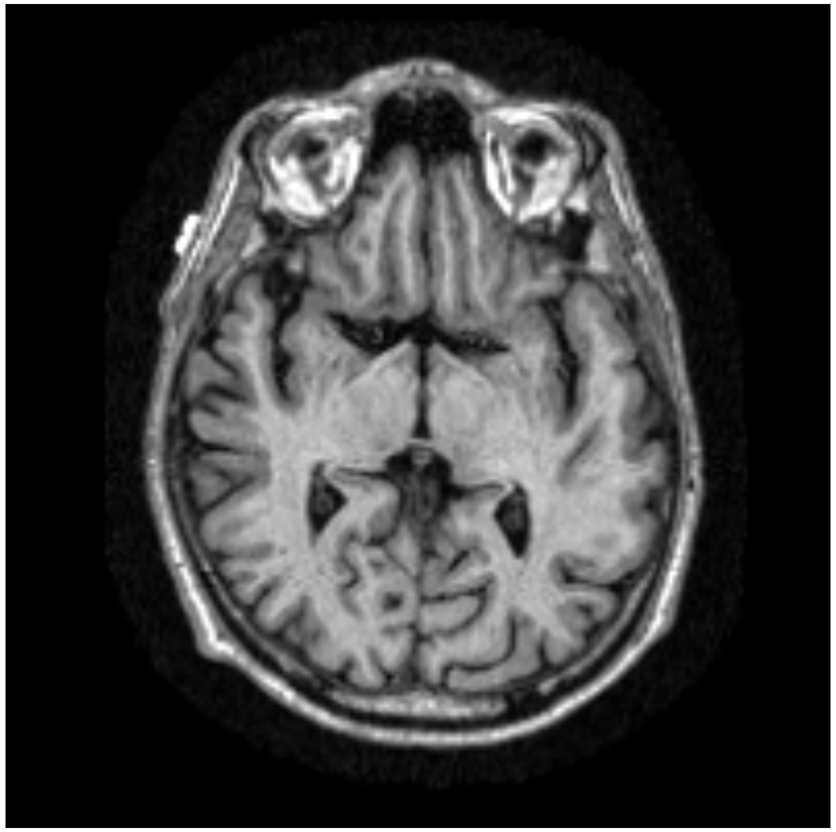


    - Six-neighborhood

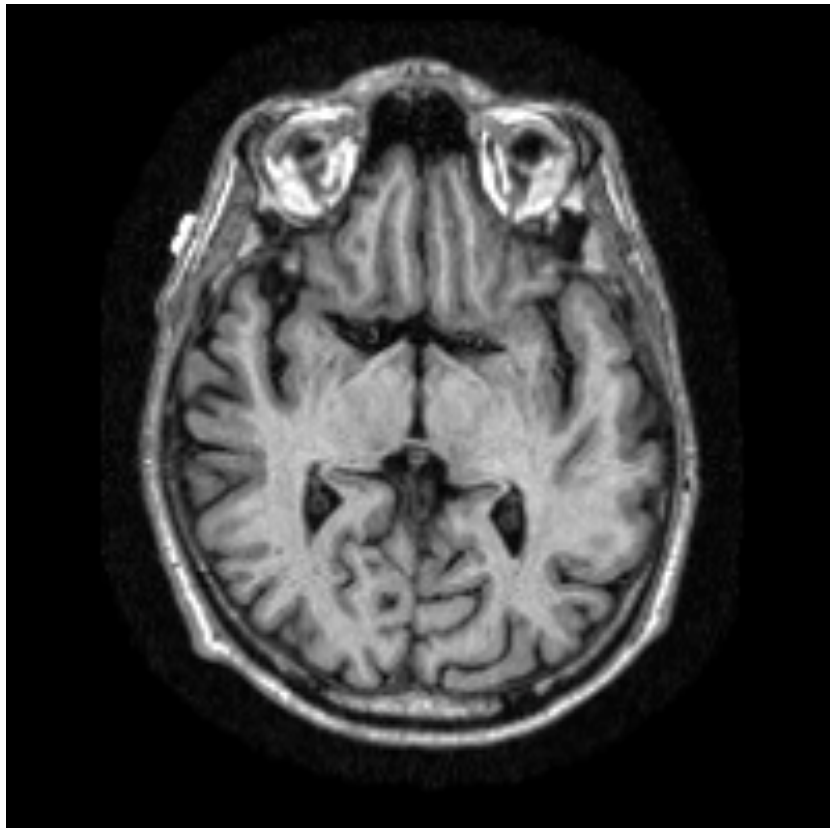


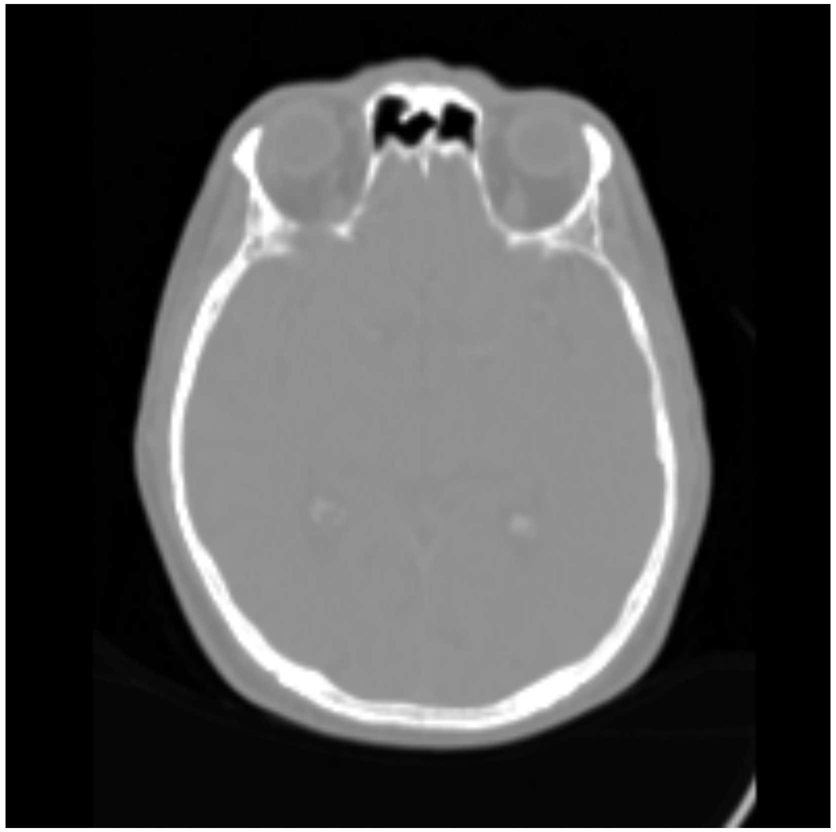
Which similarity measure?



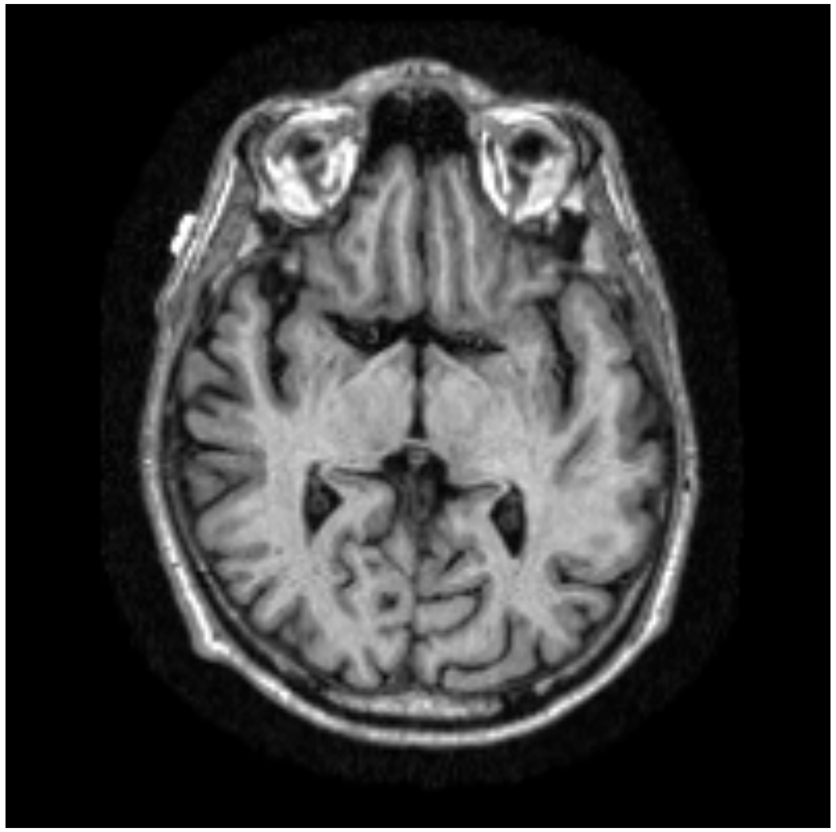


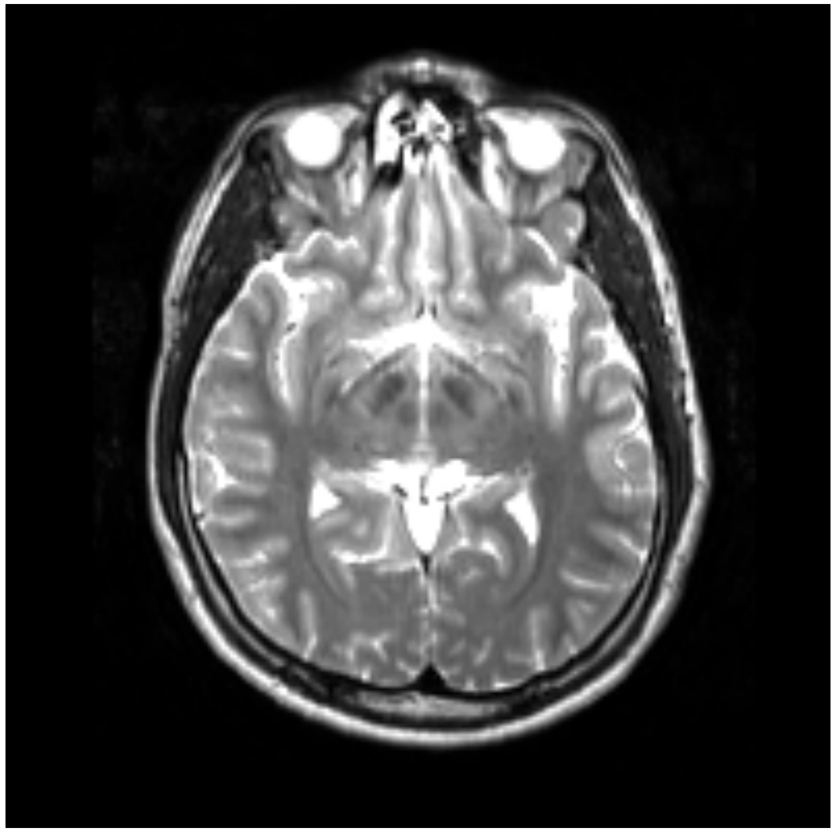
Which similarity measure?



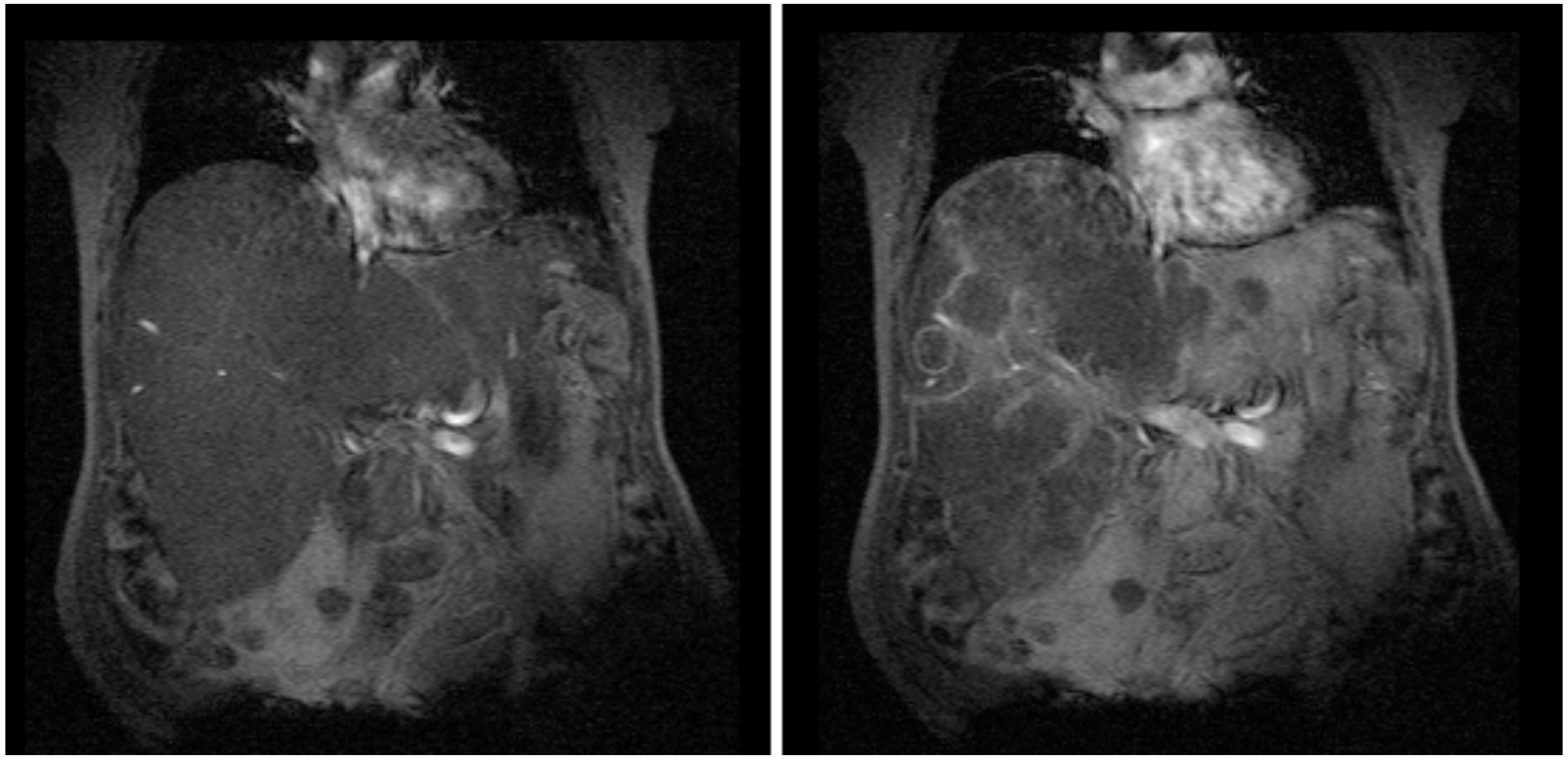


Which similarity measure?

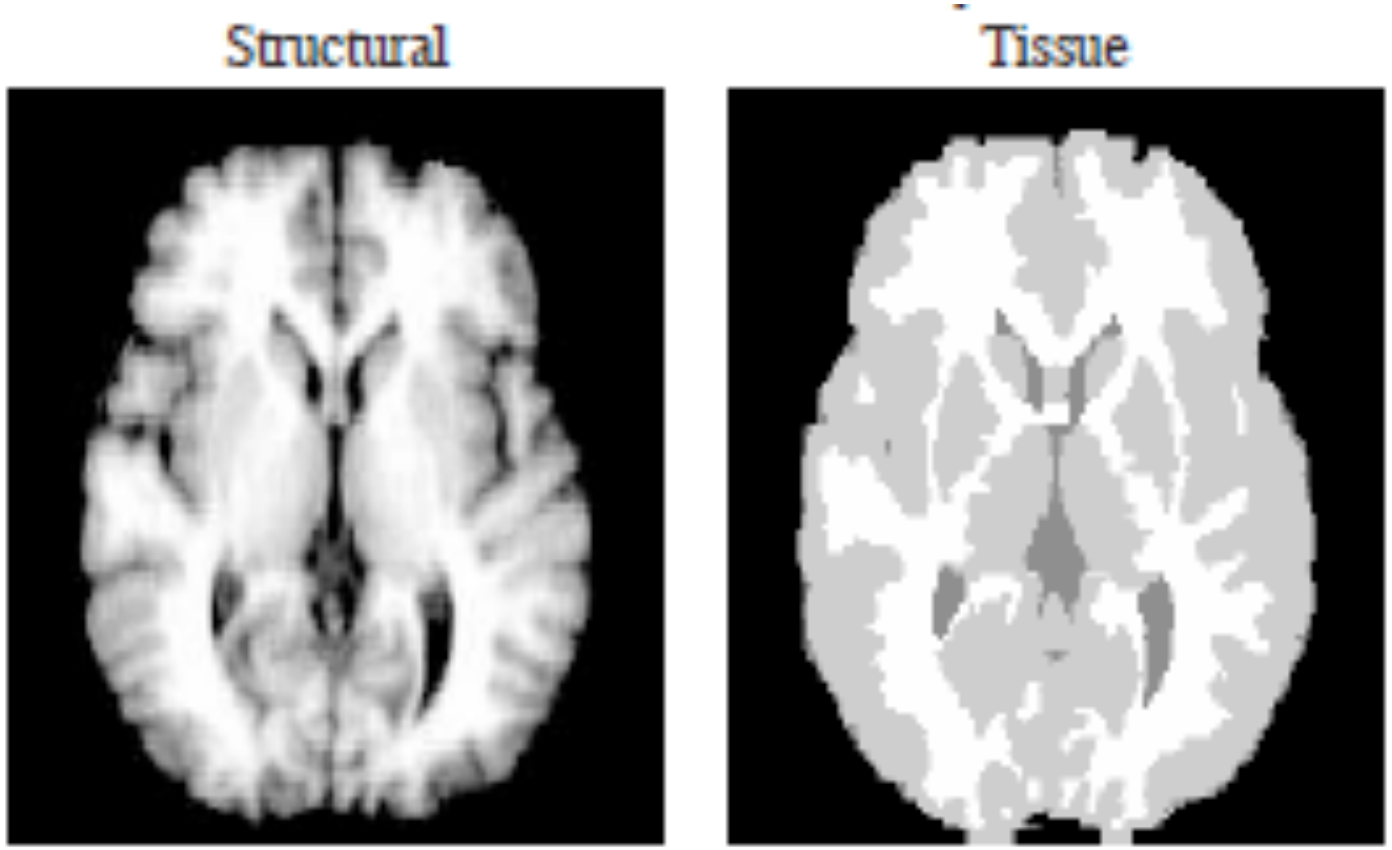




Which similarity measure?



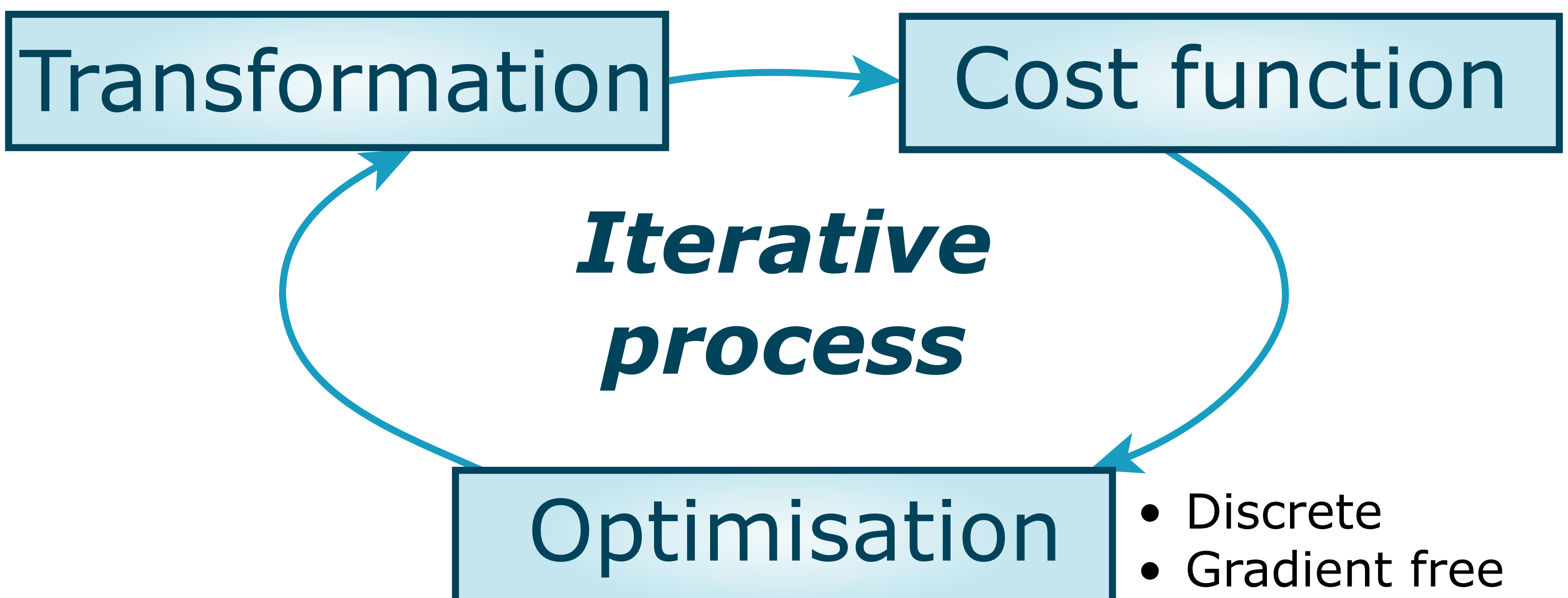
Which similarity measure?



Cost function

- Need for regularisation (explicit or implicit)

- Rigid
- Affine
- Elastic
- Fluid
- …

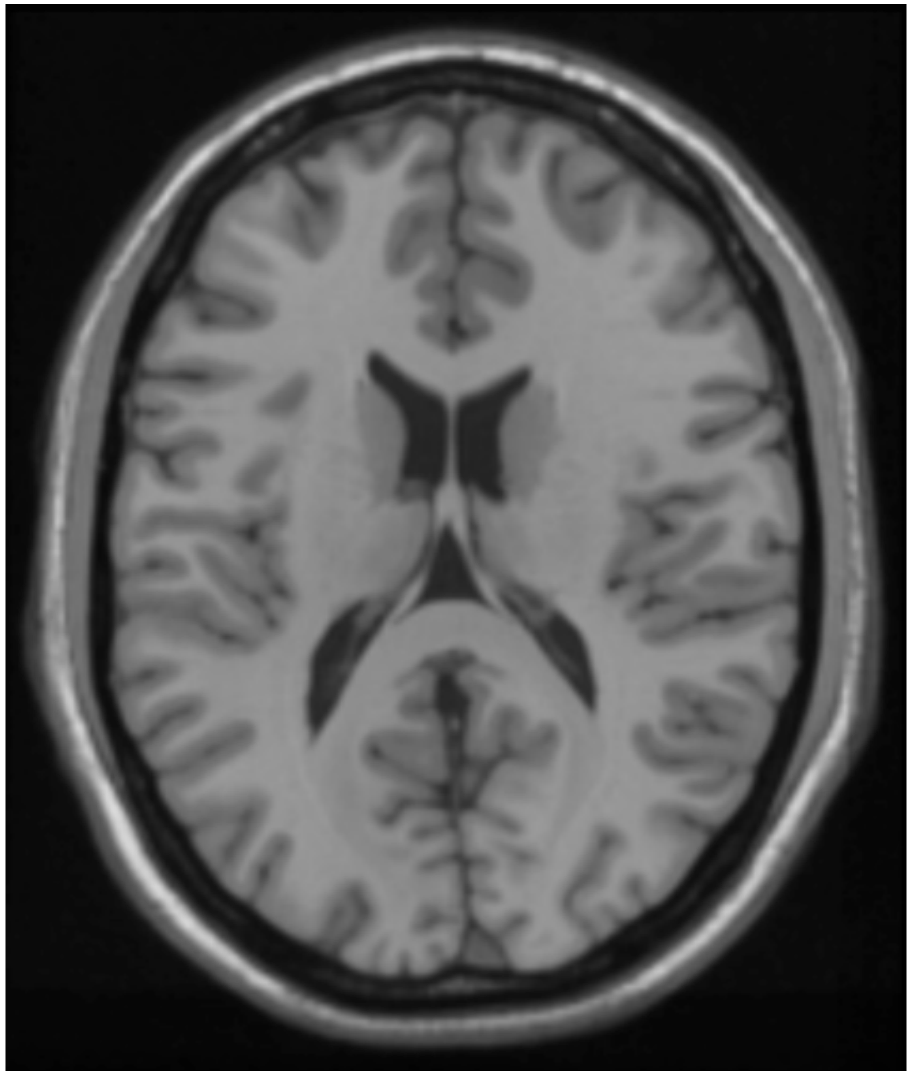


- Measure of similarity
- Penalisation term
- …

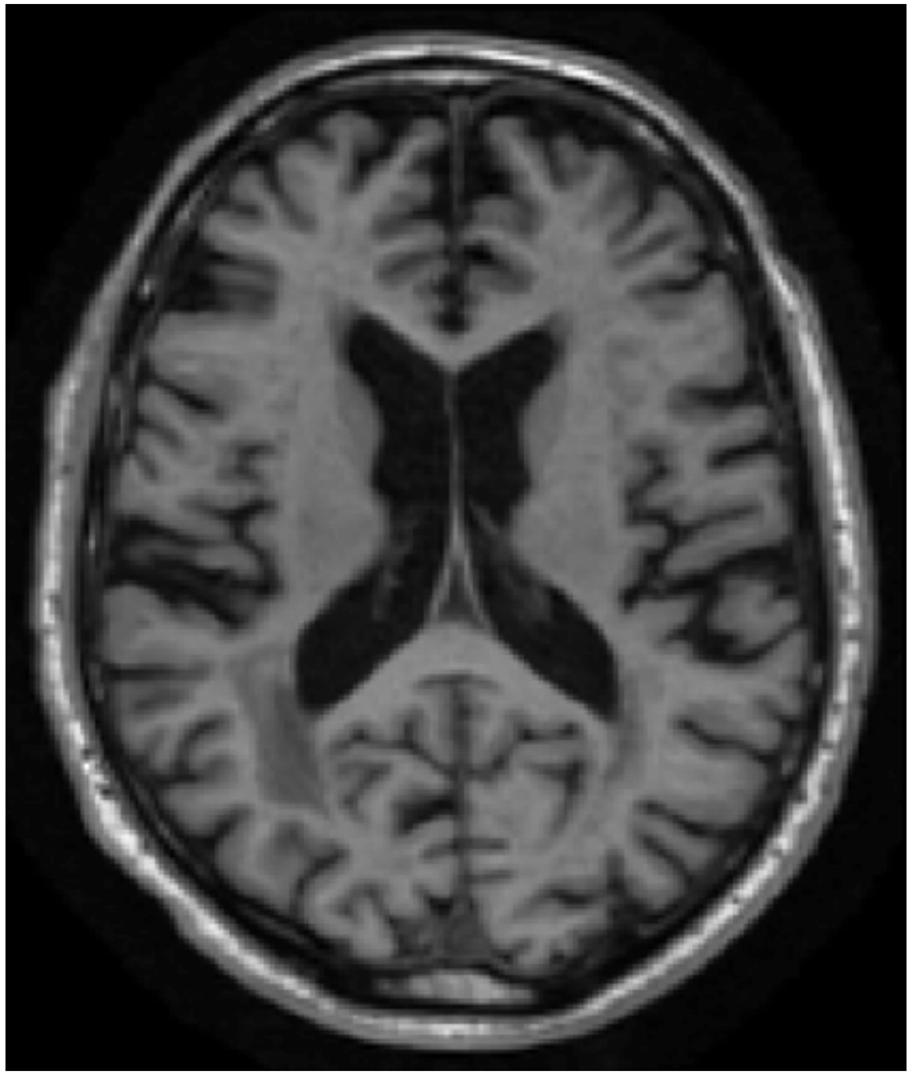
- Discrete
- Gradient free
- Gradient based
- …

Cost function

- Need for regularisation (explicit or implicit)



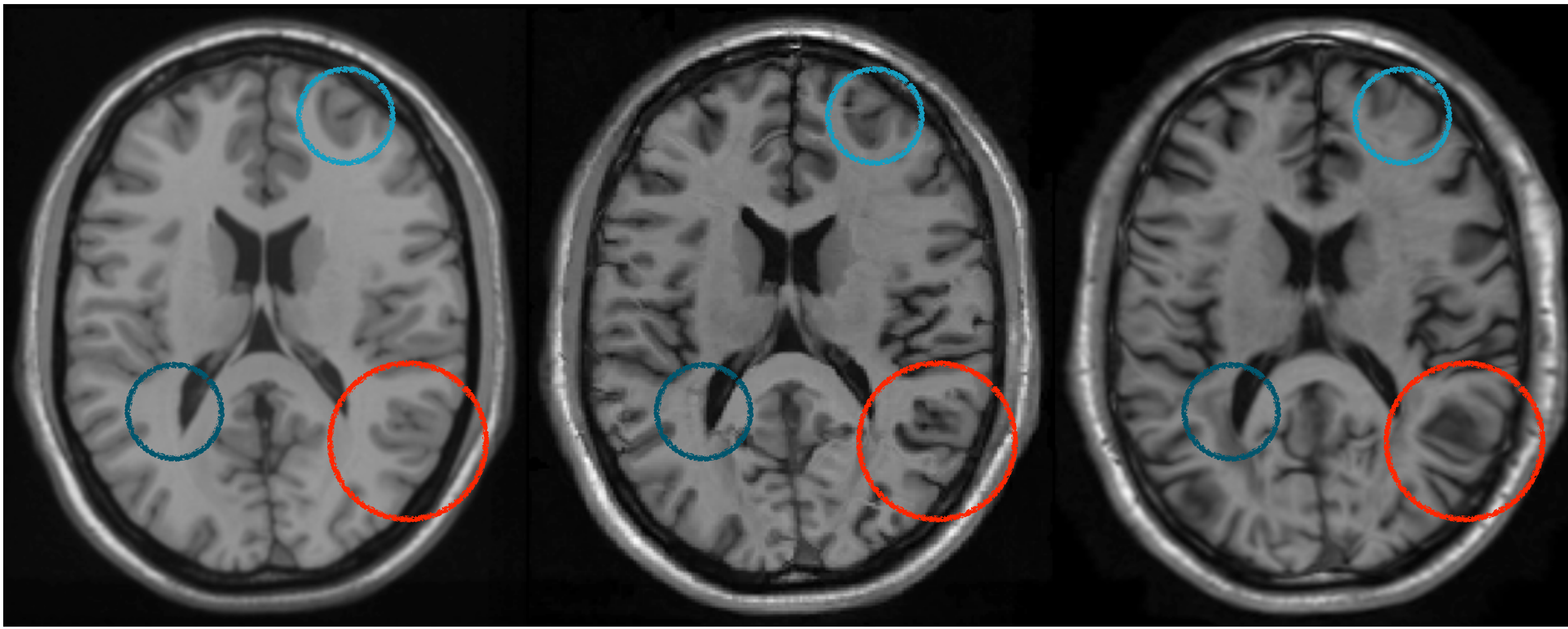
Reference Image



Floating Image

Cost function

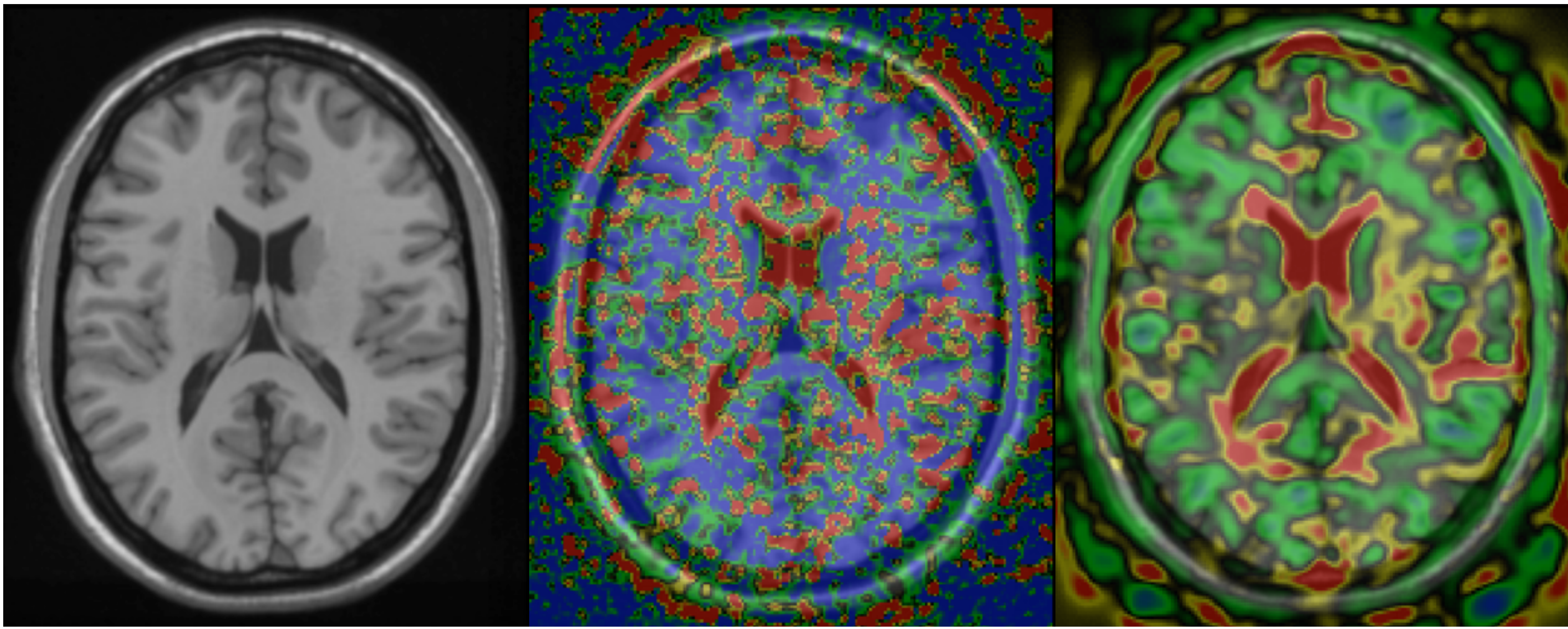
- Need for regularisation (explicit or implicit)



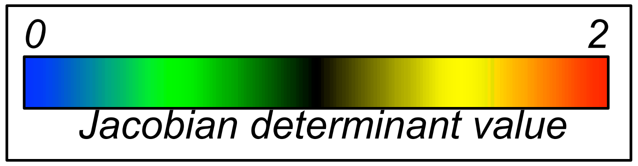
Reference Image *Warped Image 1* *Warped Image 2*

Cost function

- Need for regularisation (explicit or implicit)

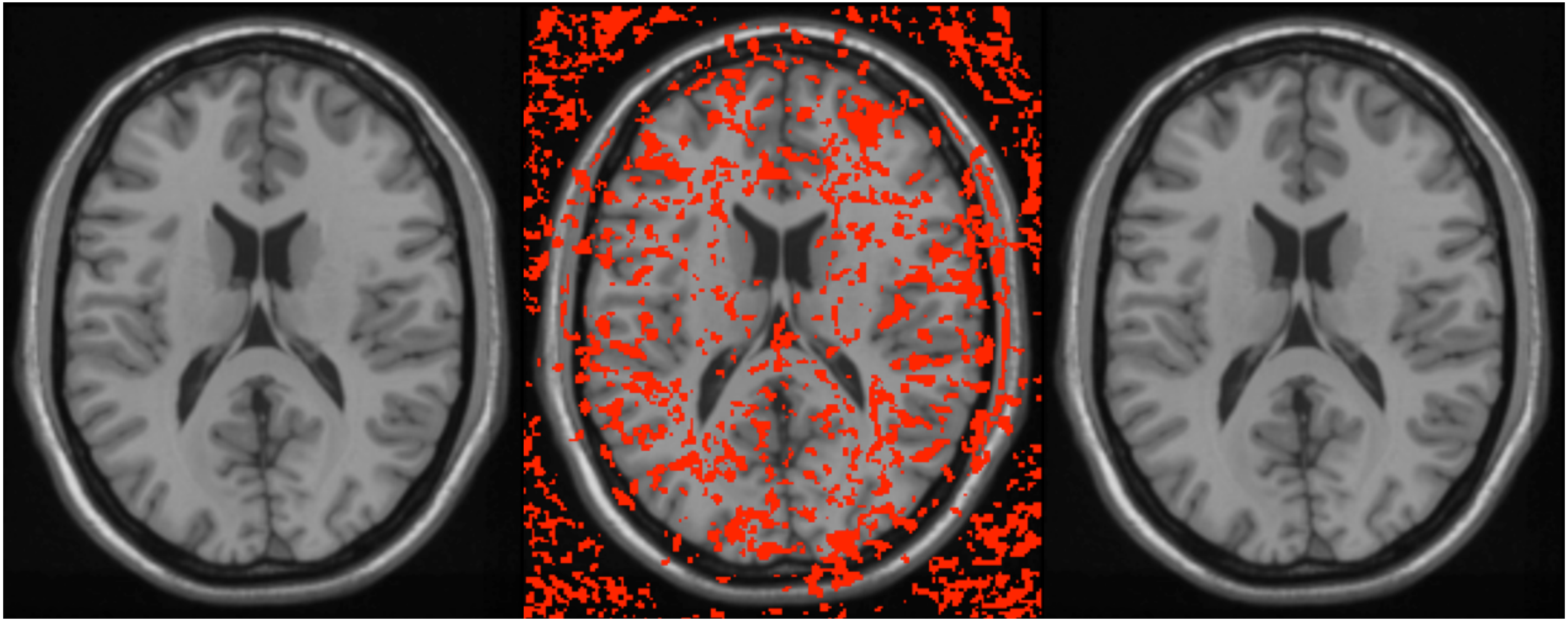


Reference Image *Jacobian map 1* *Jacobian map 2*



Cost function

- Need for regularisation (explicit or implicit)



Reference Image *Folded area 1* *Folded area 2*

Regularisation

- Can be implicit or explicit:
  - Embedded with the transformation model
  - Penalty term

- What's next:
  - Transformation: from rigid to non-linear