Medical Image Analysis
Lecture 08

Image Registration II

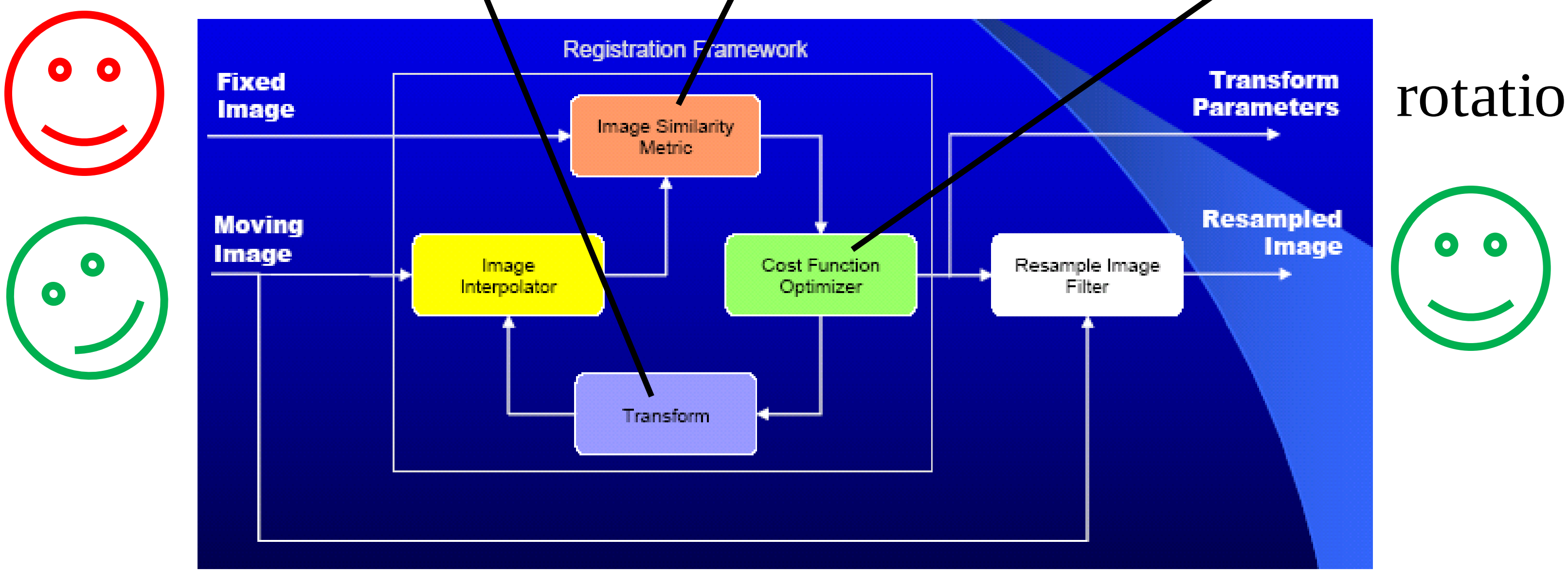
Martin Urschler

Overview - Image Registration

- What is it?
- Motivation
- Formal Definition & Categorization
- Rigid Registration Methods
  - Feature Based: Procrustes Alignment
  - Surface Based: Iterative Closest Point
  - Intensity Based

Definition of Image Registration

- Find a mapping h(x) aligning an image („moving“) with a second image („fixed“) such that a defined similarity measure is maximized.



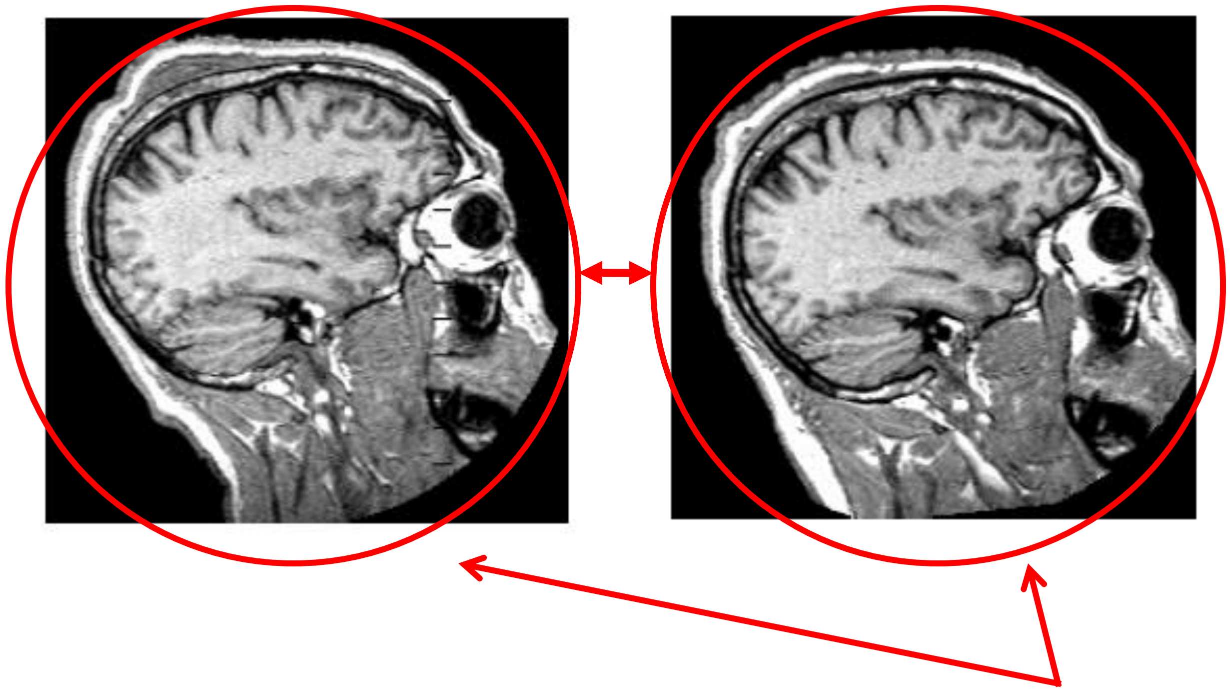
rotation R

Registration - Features vs. Voxels

Feature Based Approach

Voxel Based Approach

Registration - Features vs. Voxels



Feature Based Approach

Voxel Based Approach

Registration based on voxel similarity

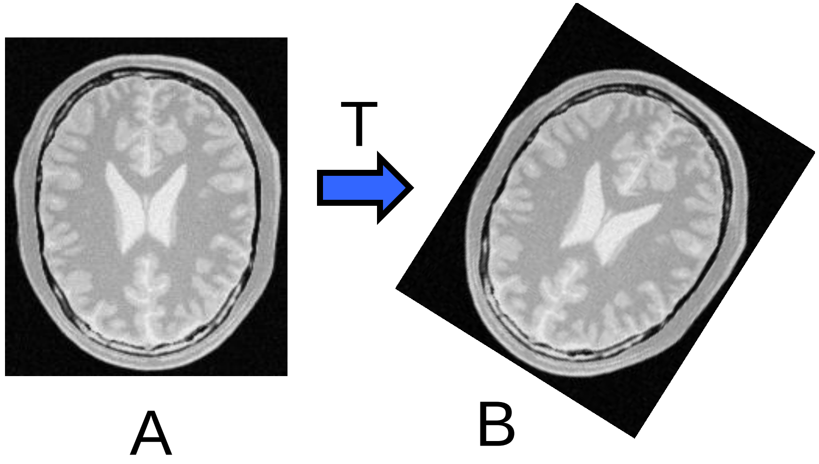
- Registration based on geometrical features
  - requires the extraction of points, lines or surfaces
  - is affected by feature localization
- Registration based on voxel similarity measures
  - uses some measure derived directly from the intensity of the image voxels
  - assumes some form of relationship between the image intensities of both images if in registration
  - does not require any feature extraction

Registration based on voxel similarity

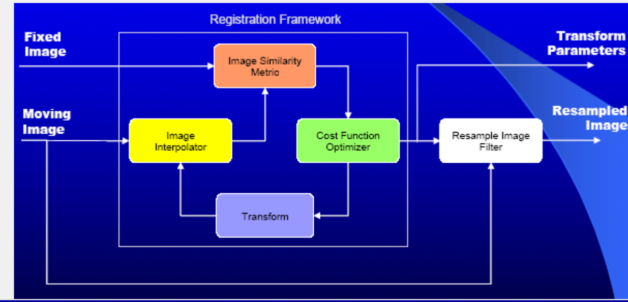
- Registration based on geometric features is independent of the modalities from which the features have been derived, if features are visible
- Registration based on voxel similarity measures must distinguish between
  - Single-modality registration:
    - CT-CT, MR-MR, PET-PET, etc
  - Multi-modality registration
    - MR-CT, MR-PET, CT-PET, etc

Registration based on voxel similarity

- Optimal transformation T is determined iteratively by minimizing a voxel-based *dissimilarity* measure C.
- Voxel-dissimilarity measure C is a function of
  - image A
    - target, reference, fixed
  - image B
    - source, moving
  - transformation T
    - rigid, affine, nonlinear



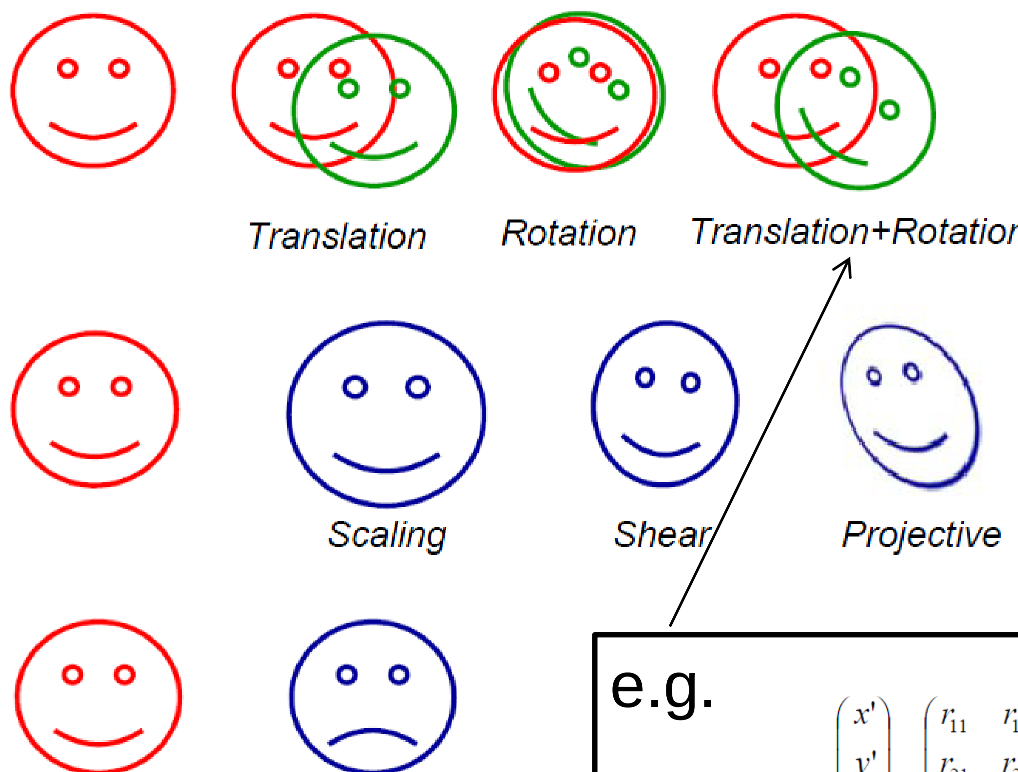
Transformations



Rigid

Affine & Projective

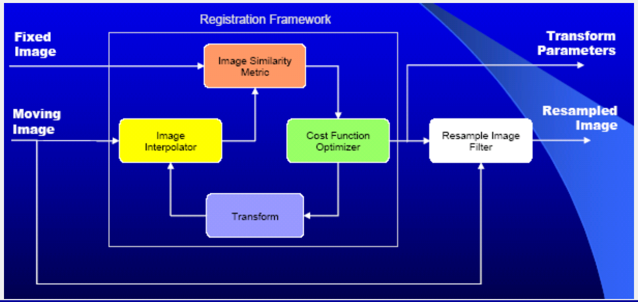
Non-rigid
- Elastic
- Splines
- Fluid
- …



e.g.

$$T_{rigid}(x,y,z)=\begin{pmatrix}x'\\y'\\z'\\1\end{pmatrix}=\begin{pmatrix}r_{11}&r_{12}&r_{13}&t_x\\r_{21}&r_{22}&r_{23}&t_y\\r_{31}&r_{32}&r_{33}&t_z\\0&0&0&1\end{pmatrix}\begin{pmatrix}x\\y\\z\\1\end{pmatrix}$$

Similarity Metrics



- Used to measure the similarity between two images
- Depends on the type of images to register

- Similarity assumptions:
  - Identity: Single-modality, only differ by Gaussian noise
  - Linear: Single-modality, differ by constant intensity
  - Information Theoretic / Probabilistic:
    Multi-modality, intensity changing, related by some statistical or functional relationship

Single-modal image registration

- Sum of Squared Differences (SSD)

$$C(x;\mathrm{T}) = \frac{1}{N}\sum_{i}(I_A(x_i) - I_B(\mathrm{T}(x_i)))^2$$

For all voxels i

  - assumes an *identity* relationship between image intensities in both images
  - optimal measure if the difference between both images is Gaussian noise
  - sensitive to outliers -> use Sum of Absolute Differences (SAD)!

Sum of Squared Differences (SSD)

Matlab!

intensity reg-
istration test

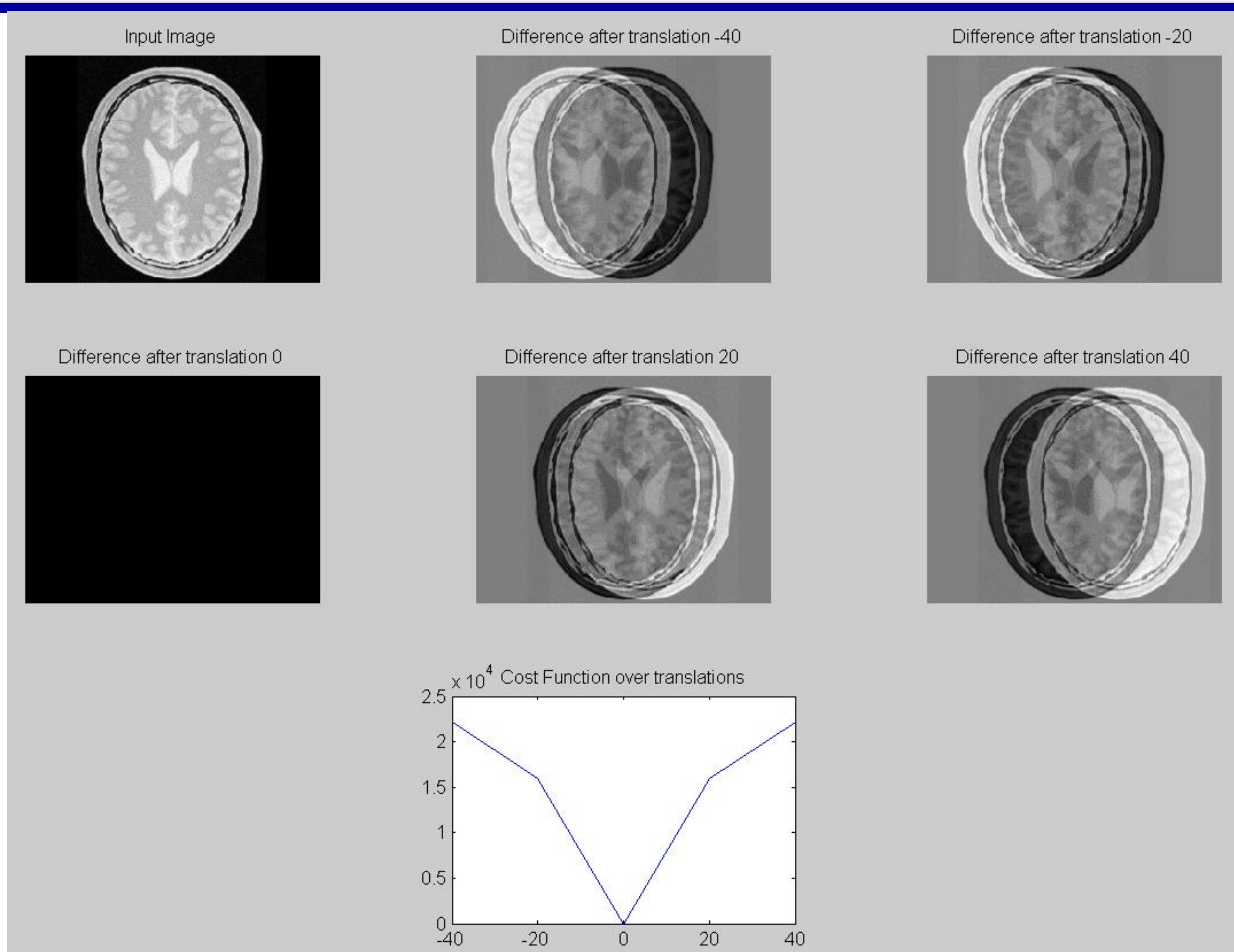
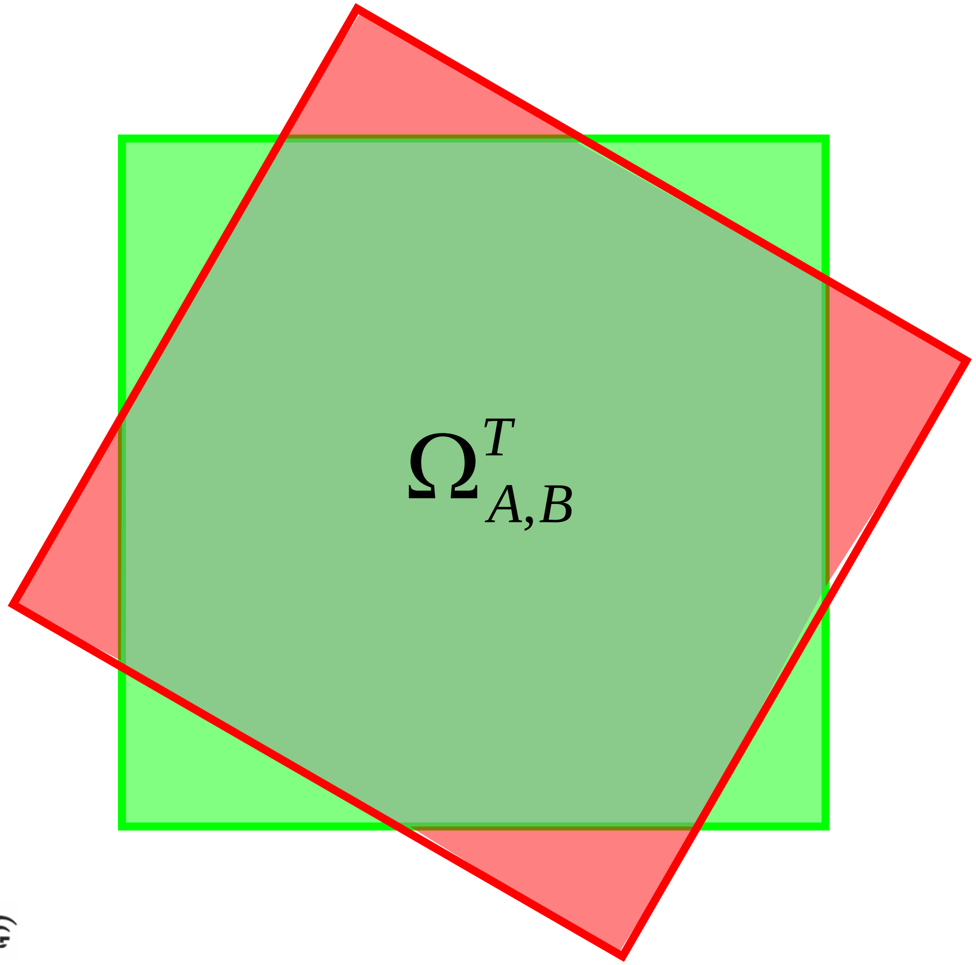


Image overlap



Ω_A Image A only

Ω_B^T Image B^T only

Overlap between image A and B^T

Single-modal image registration

- Normalized Cross Correlation (CC)

$$CC(x;\mathrm{T}) = \frac{\sum\left[\left(I_A(x) - \mu_A\right)\left(I_B(\mathrm{T}(x)) - \mu_B\right)\right]}{\sqrt{\sum\left(I_A(x) - \mu_A\right)^2 \sum\left(I_B(\mathrm{T}(x)) - \mu_B\right)^2}}$$

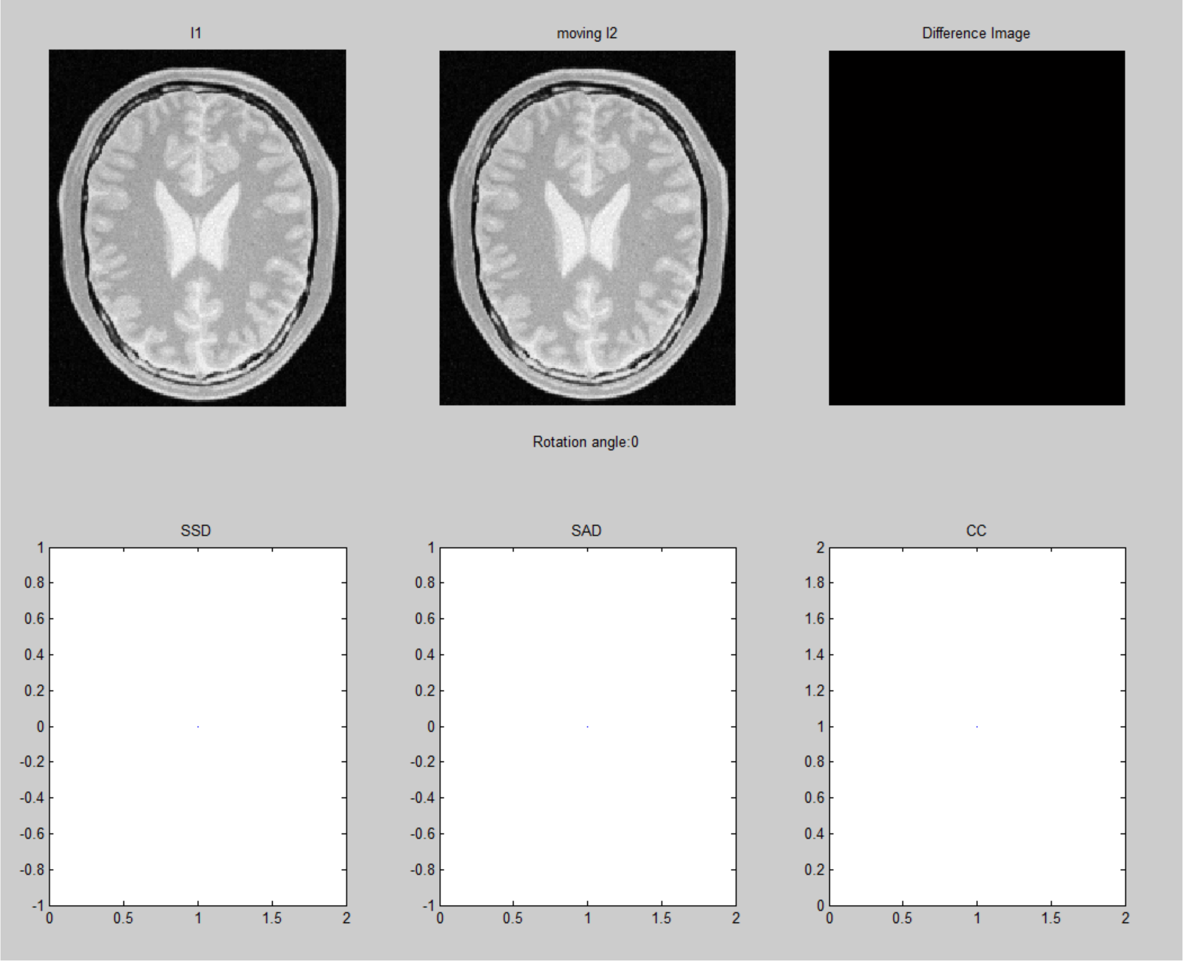
  - μ_A average intensity in image A
  - μ_B average intensity in image B
  - assumes a *linear* relationship between image intensities ($I_B = a*I_A+b$)
  - useful if images have different intensity windows

Metrics

Matlab!

SimilarityMetrics

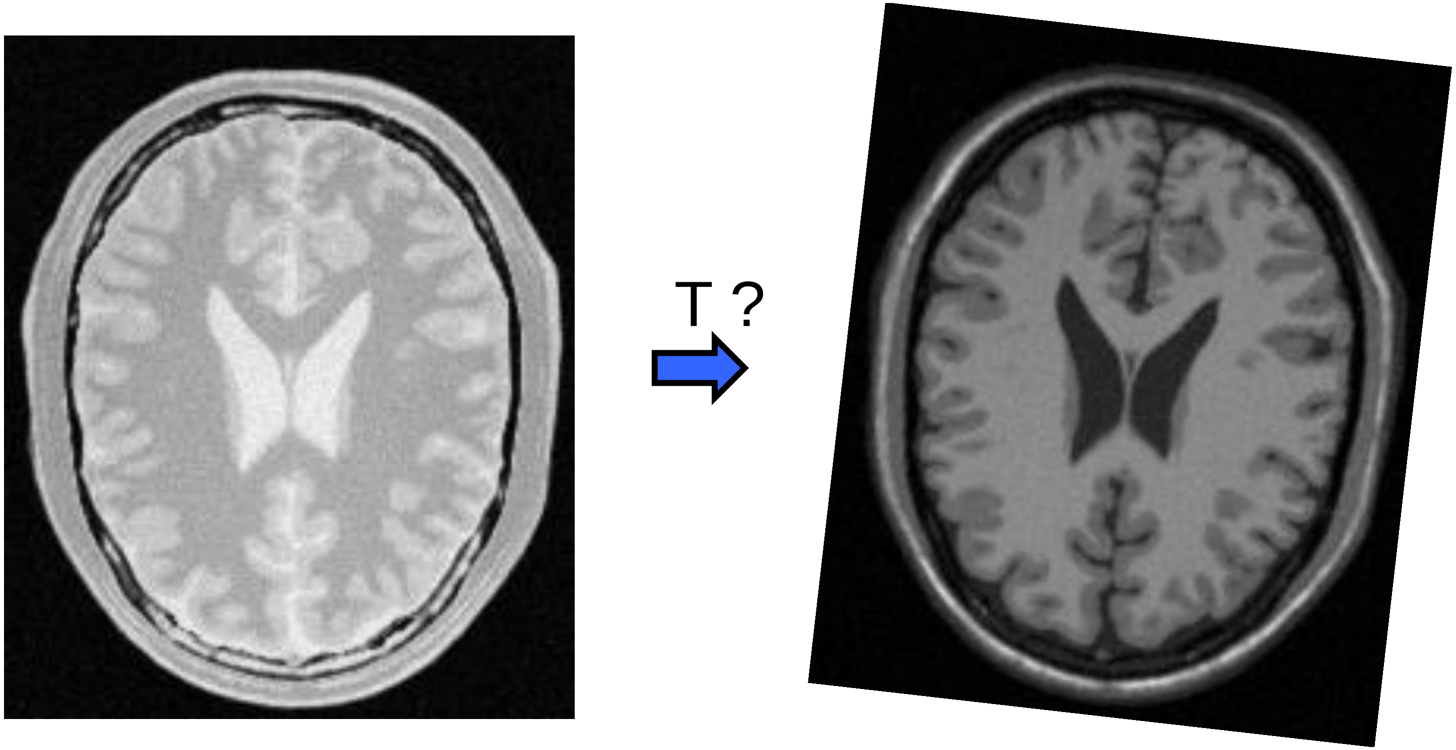
Movie



Registration basis: Image intensity

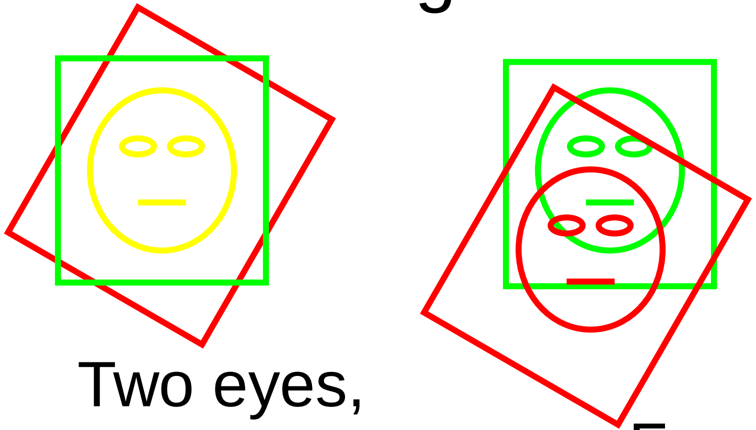
- Single-modal image registration
  - Image intensities are related by some simple function
    - identity: Use SSD or SAD
    - linear: Use CC
- Multi-modal image registration
  - Image intensities are related by some unknown function or statistical relationship
  - Relationship between intensities is not known a-priori
  - Relationship between intensities can be viewed by inspecting a 2D histogram or co-occurrence matrix

Multi-modality image registration



Multi-modality image registration

- Information theoretic point of view
- Look at two images combined



Two eyes,
a mouth

Four eyes,
two mouthes

Same as *reducing amount of information* *in combined image!*

- Registration tries to maximize the amount of shared information in two images!

How do we measure information?

- Given:
  - a set of distinct symbols $\{v_1, v_2, \ldots, v_n\}$
  - events of randomly drawing a symbol v from this set
  - a probability distribution p over the set
- Measure of information:
  - Metaphor: measure the „degree of surprise“ of a random event
  - How much information do we gain, after a certain v has been drawn from the set?

How do we measure information?

$$\{v_1, v_2, \ldots, v_n\}$$

- Drawing v_i from the set is done with probability p_i
- A highly improbable event surprises us, while an almost certain event gains us not much information!

- -> Measure of information (gain) depends on p!
- Define measure of information h:
  - For two unrelated events x,y the information gain from observing both is h(x,y) = h(x) + h(y)

How do we measure information?

$$\left\{v_1, v_2, \ldots, v_n\right\}$$

- Two unrelated events x,y also are statistically independent! $p(x,y) = p(x)*p(y)$
- The only possible definition for h using p is: $h(x) = -\log_2 p(x)$
- Example: coin flip $\{v_1\text{: head}, v_2\text{: tails}\}$
  - $p(x=v_1) = p(x=v_2) = 0.5$
  - $h(x=v_1) = -\log_2 p(x) = -\log_2 0.5 = 1$
  - $h(x=v_2) = 1$

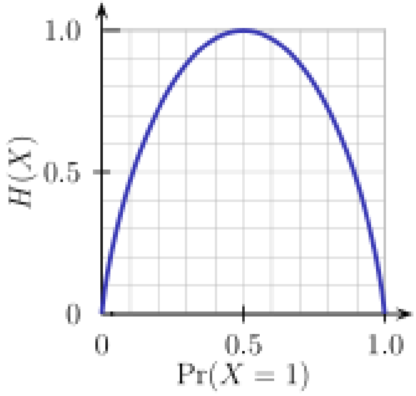
Measuring information

- Now suppose we want to transmit value of random variable x over a communication channel:
  - Average amount of information transmitted is the expectation of h with respect to p

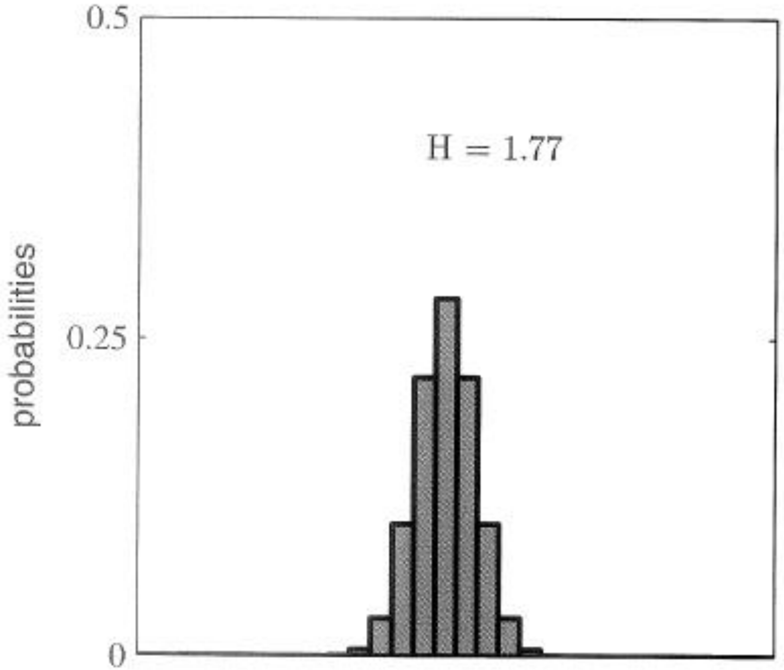
$$H(X) = \sum_x p(x)\left(-\log_2 p(x)\right)$$

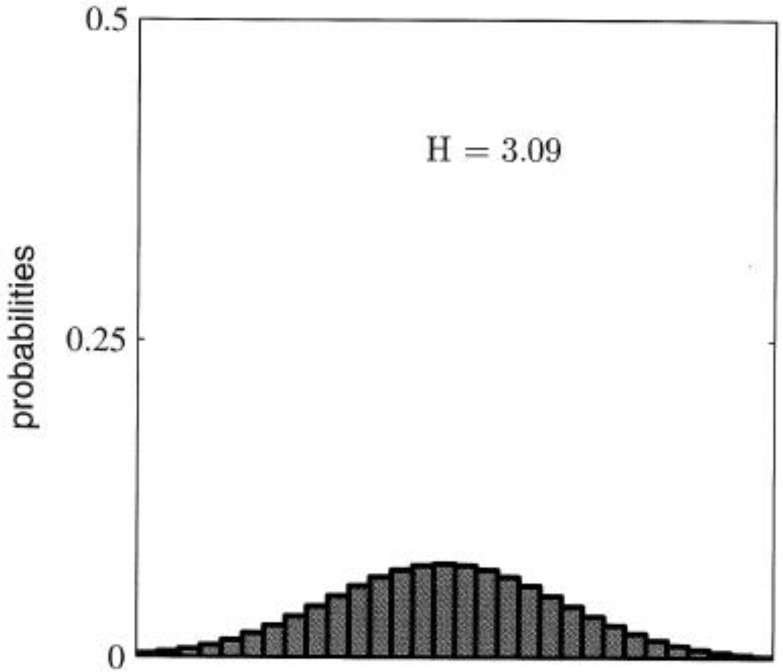
Quantity known as entropy!

  - Back to coin flip:
    H(x) = 0.5 * h(x=v_1) + 0.5 * h(x=v_2) = 1 bit
  - What happens with non-equal probabilities?
  - „Unfair Coin“: p(x=v_1) = 0.3, p(x=v_2) = 0.7
    H(x) = 0.3 * -$\log_2$(0.3) + 0.7 * -$\log_2$(0.7) = 0.88 bit



Distributions and Entropy

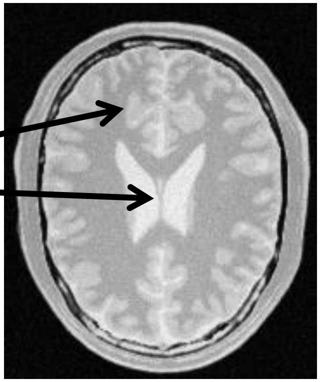




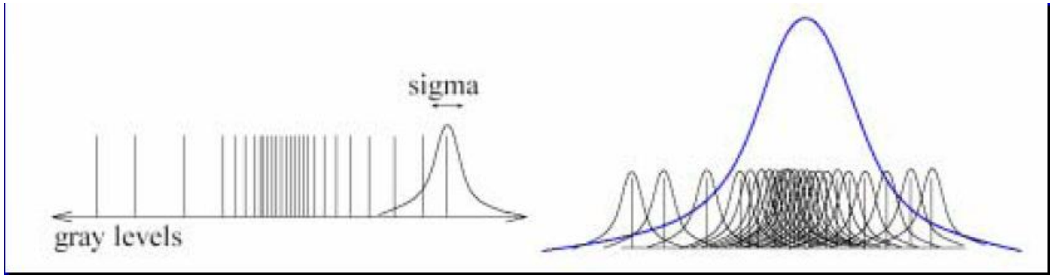
Q: How does a prob. distribution have to look like for H=0?
How for H=max?

Images as probability distributions

- Images can be viewed as (pixel-wise) grey level probability distributions
  - marginal probability p(a) of any pixel having intensity a

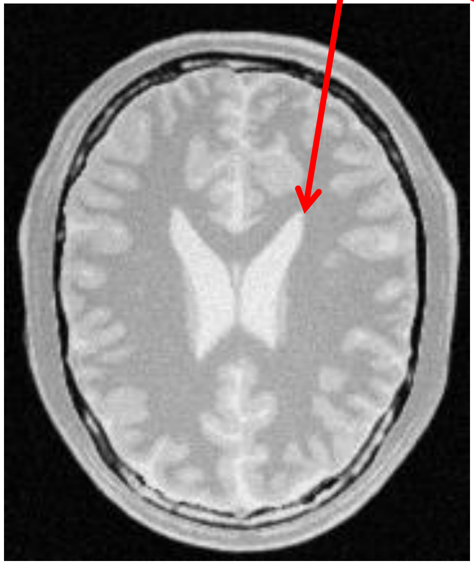


  - Probability distribution of an image can be estimated using histograms
    - Requires binning!
    - 32 to 256 bins
    - May be refined by Parzen Windowing



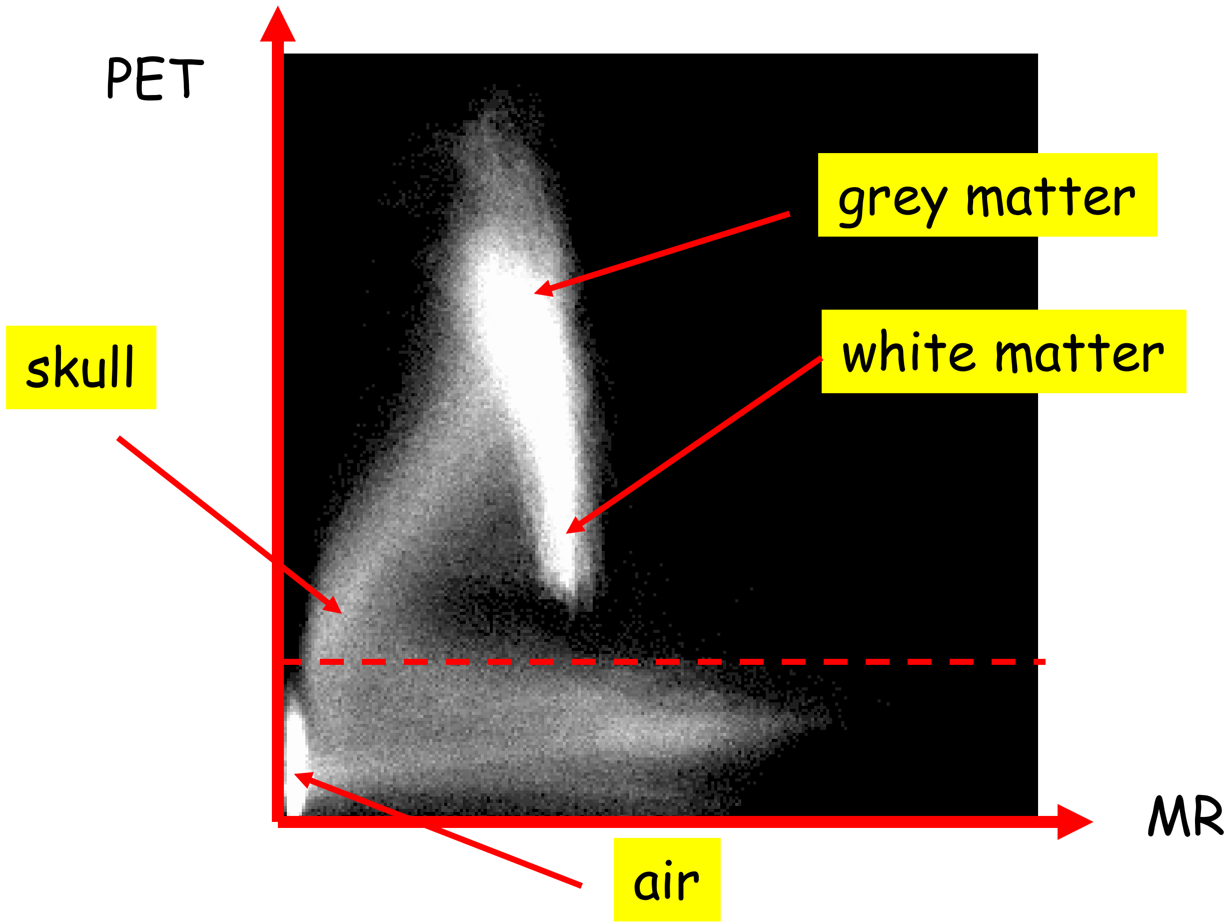
Images as probability distributions

- Joint probability p(a, b) of a pixel having intensity a in one image and intensity b in another image (corresponding locations!)

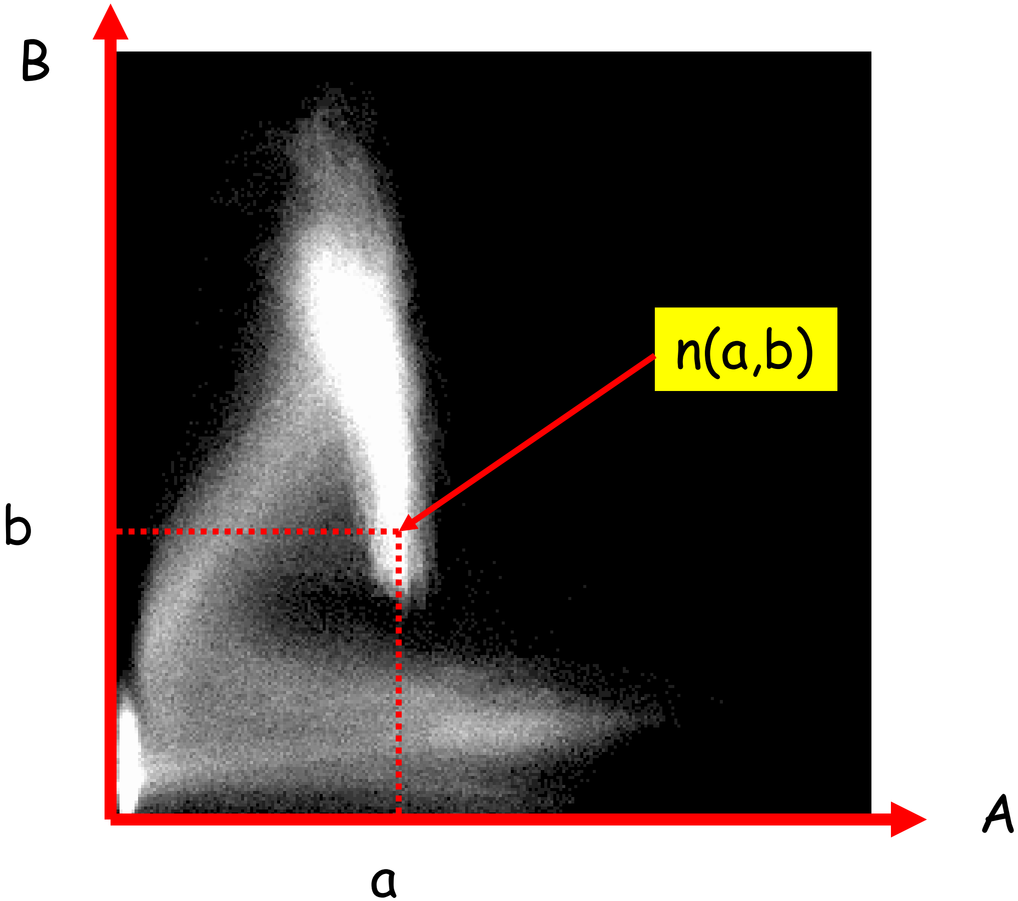


Joint Histogram

2D Joint Histograms



2D Joint Histograms



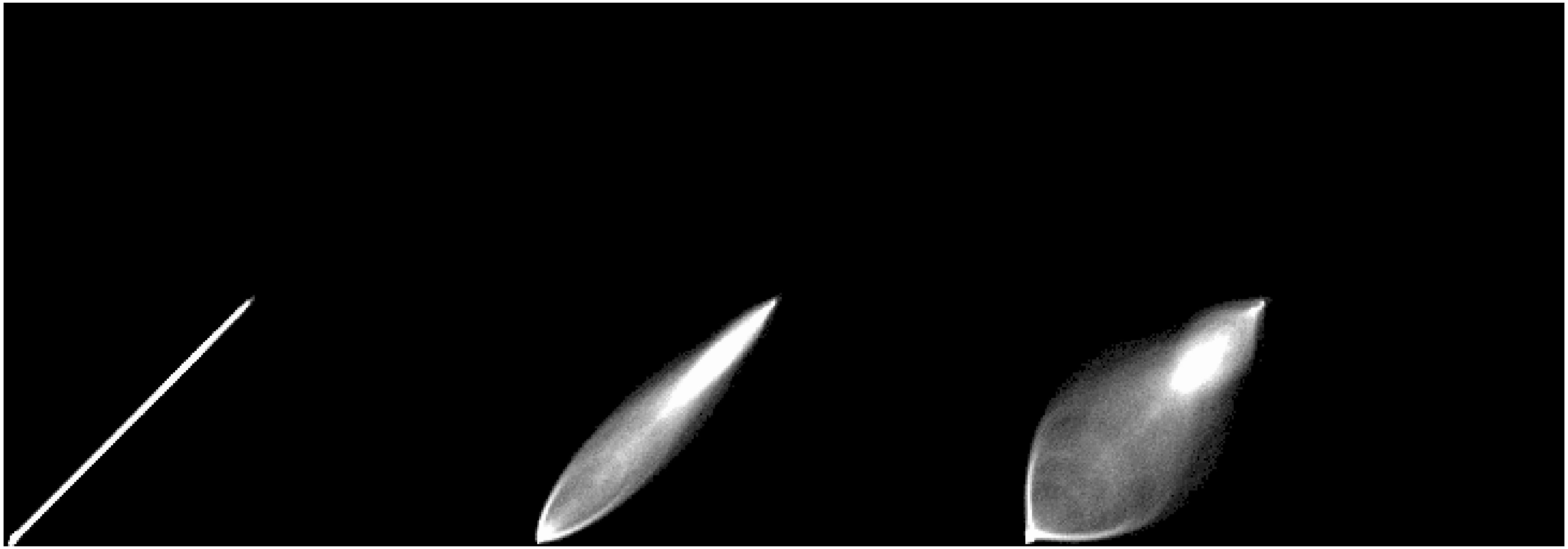
How often does it occur that intensities a and b are located at the same location in the two images?

-> n(a,b)

2D Joint Histograms

Histograms for increasing joint entropy

MR/MR

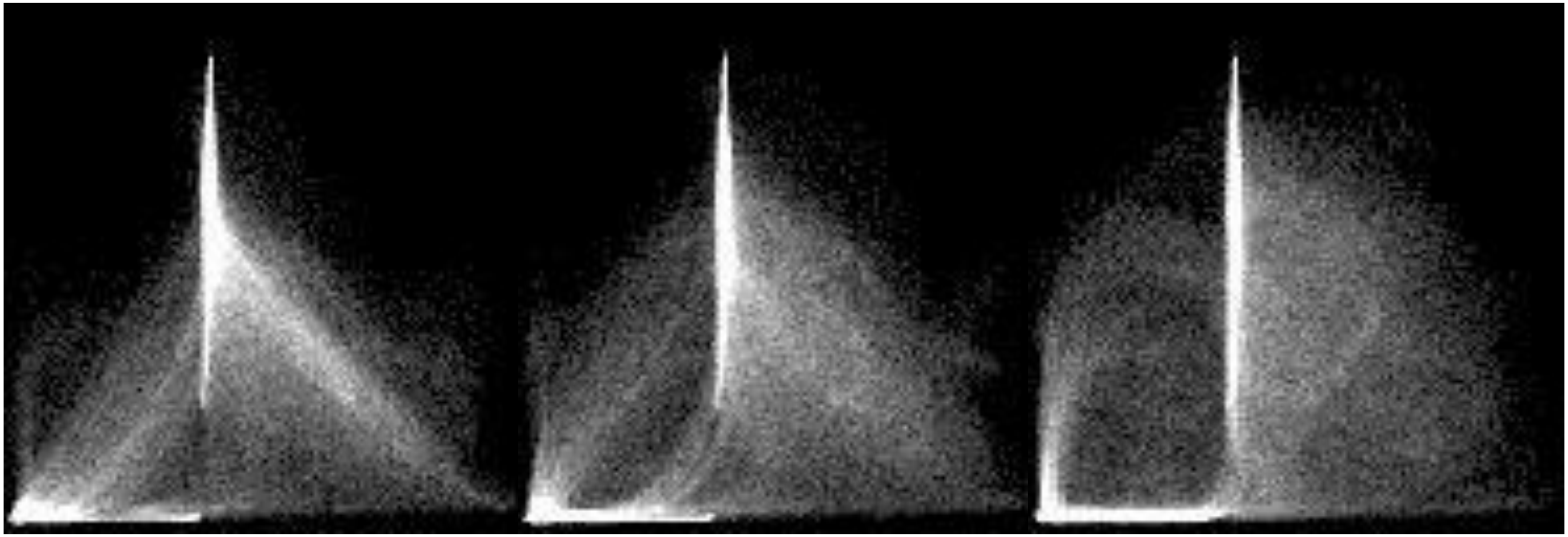


registered   misregistered by 2mm   misregistered by 5mm

2D Joint Histograms

MR/CT

Histograms for increasing joint entropy

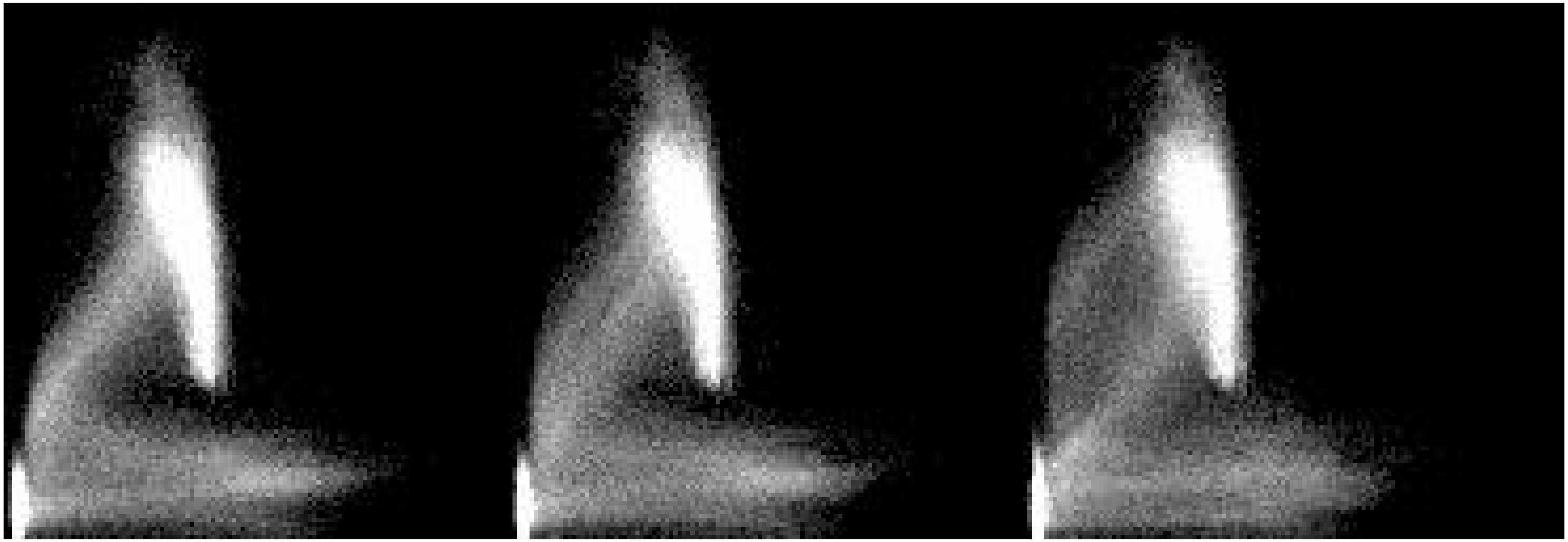


registered misregistered by 2mm misregistered by 5mm

2D Joint Histograms

MR/PET

Histograms for increasing joint entropy



registered   misregistered by 2mm   misregistered by 5mm

Images as probability distributions

- Frequency of corresponding intensity pairs can be interpreted in terms of probabilities

$$p(a,b) = \frac{n(a,b)}{N}$$

is the joint probability of a voxel having greyvalue a in the first image and greyvalue b in the second image

$$p(a) = \sum_b p(a,b)$$

is the marginal probability of a voxel in the first image having greyvalue a

$$p(b) = \sum_a p(a,b)$$

is the marginal probability of a voxel in the second image having greyvalue b

TU Graz

Voxel similarity based on information theory

- Entropy (Shannon-Wiener)

$$H(A) = -\sum_a p(a) \log_2 p(a)$$

describes the average amount of information in image A.

  - The information content of an image is maximal (in the information theoretic sense) if all intensities have equal probability.
  - The information content of an image is minimal (in the information theoretic sense) if one intensity a has a probability of one, i.e. $p(a) = 1$.

Voxel similarity based on information theory

- Joint Entropy (Hill et al., 1994)

$$H(A,B) = -\sum_a \sum_b p(a,b) \log_2 p(a,b)$$

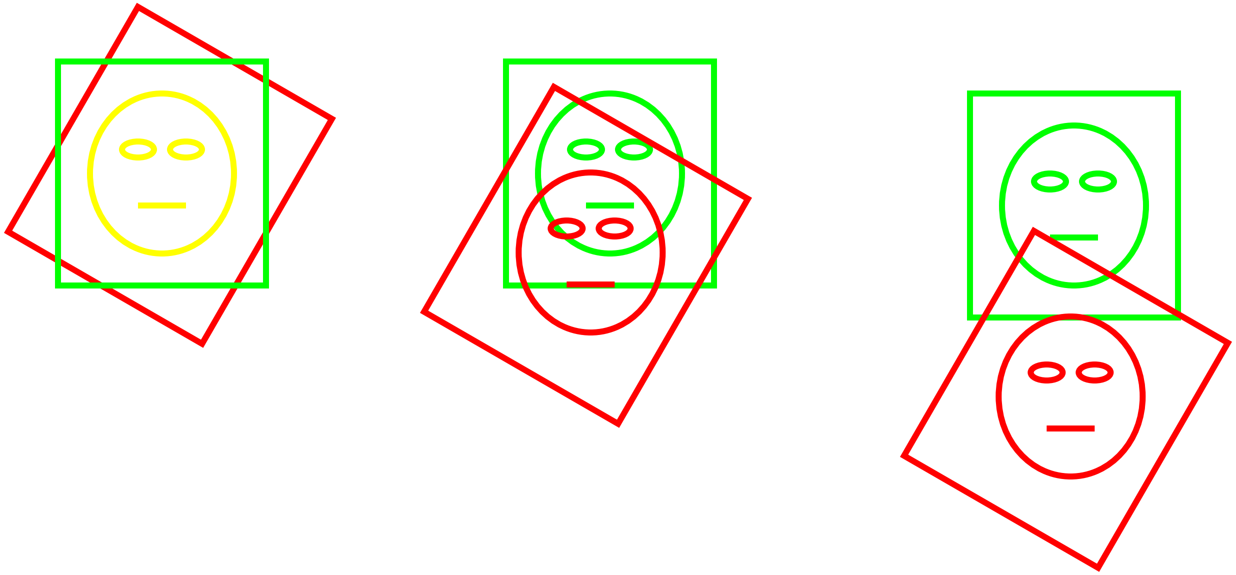
  describes the average amount of information in the combined images A and B.
  - If A and B are totally unrelated, the joint entropy will be the sum of the entropies of A and B
  - If A and B are related, the joint entropy will be smaller, i.e.

$$H(A,B) \leq H(A) + H(B)$$

- Registration can be achieved my minimizing the joint entropy between both images

Voxel similarity based on information theory

- Interpretation of Joint Entropy

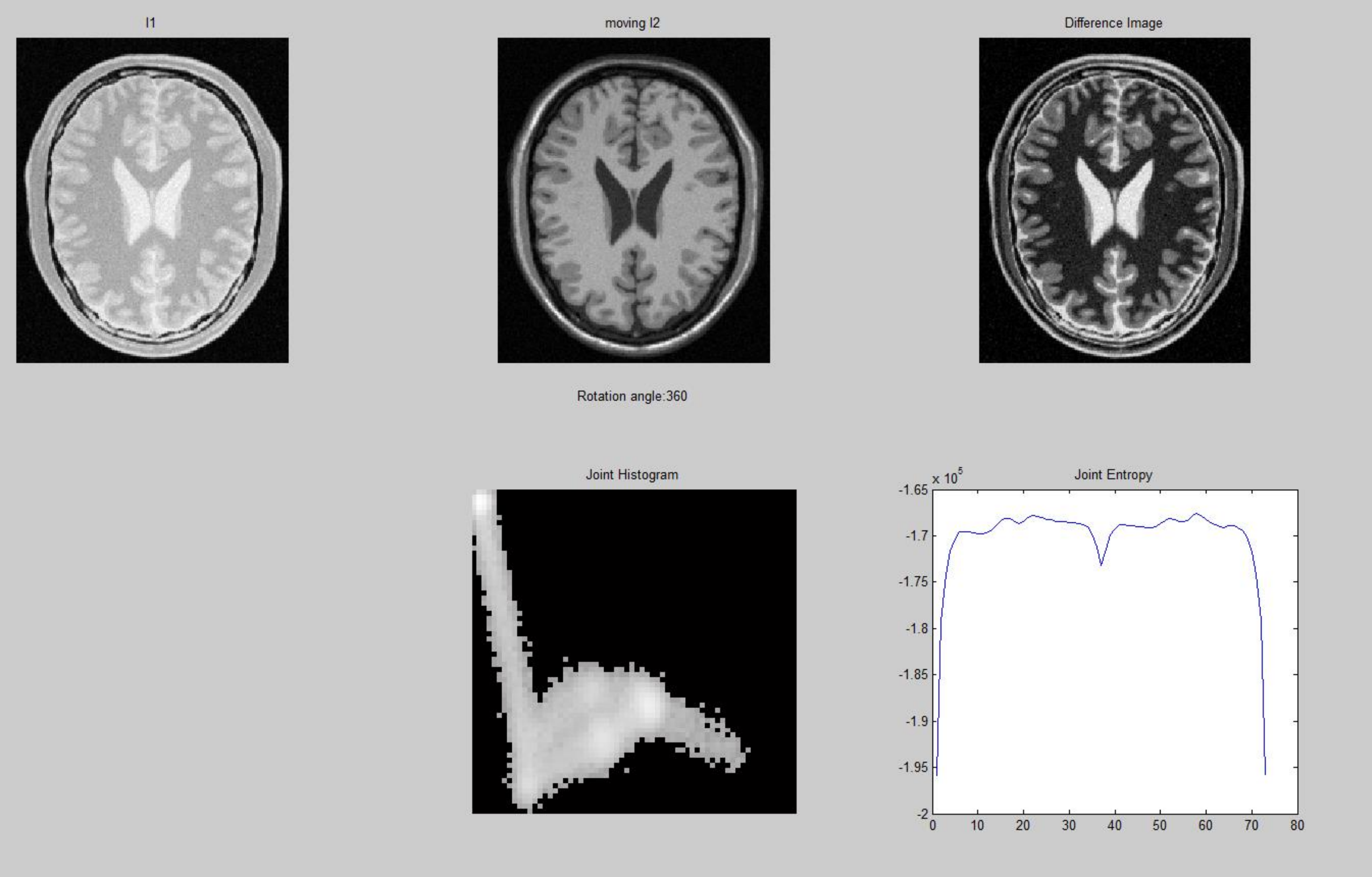


low joint entropy

high joint entropy

Joint Entropy Metric

Matlab!
joint_histogram_
movie



Voxel similarity based on information theory

- Joint Entropy is highly sensitive to the overlap of the two images
- Mutual Information (MI, Viola et al., 1995 and Collignon et al., 1995)

$$I(A,B) = H(A) + H(B) - H(A,B)$$

  describes how well one image can be explained by another image (i.e. the reduction of uncertainty about A when knowing B).
- Mutual Information can be expressed in terms of marginal and joint probability distributions:

$$I(A,B) = \sum_a \sum_b p(a,b) \log_2 \frac{p(a,b)}{p(a)\,p(b)}$$

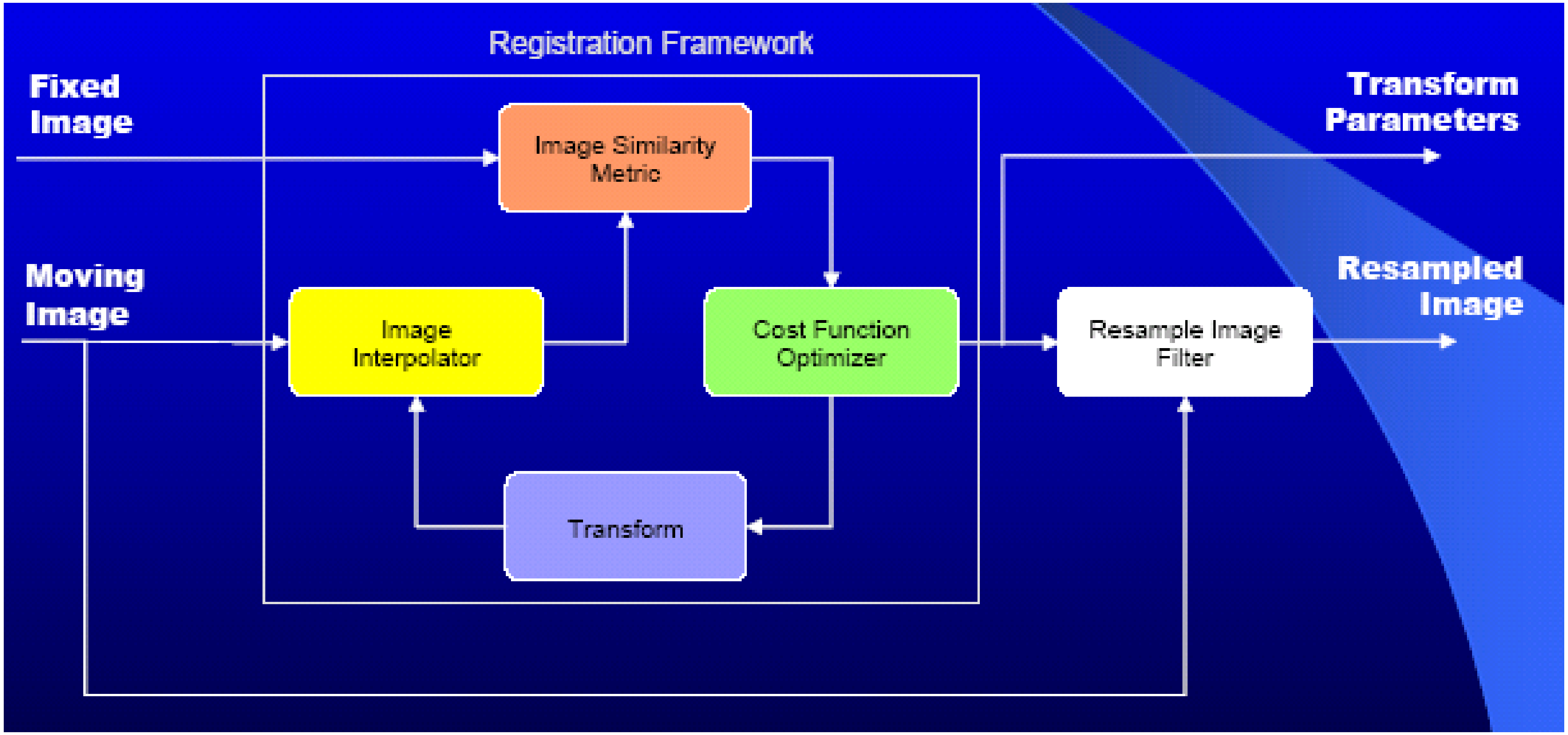
Voxel similarity based on information theory

- Mutual Information is still to a certain degree sensitive to the overlap of the two images
- Normalized Mutual Information (Studholme et al, 1999)

$$I(A,B) = \frac{H(A)+H(B)}{H(A,B)}$$

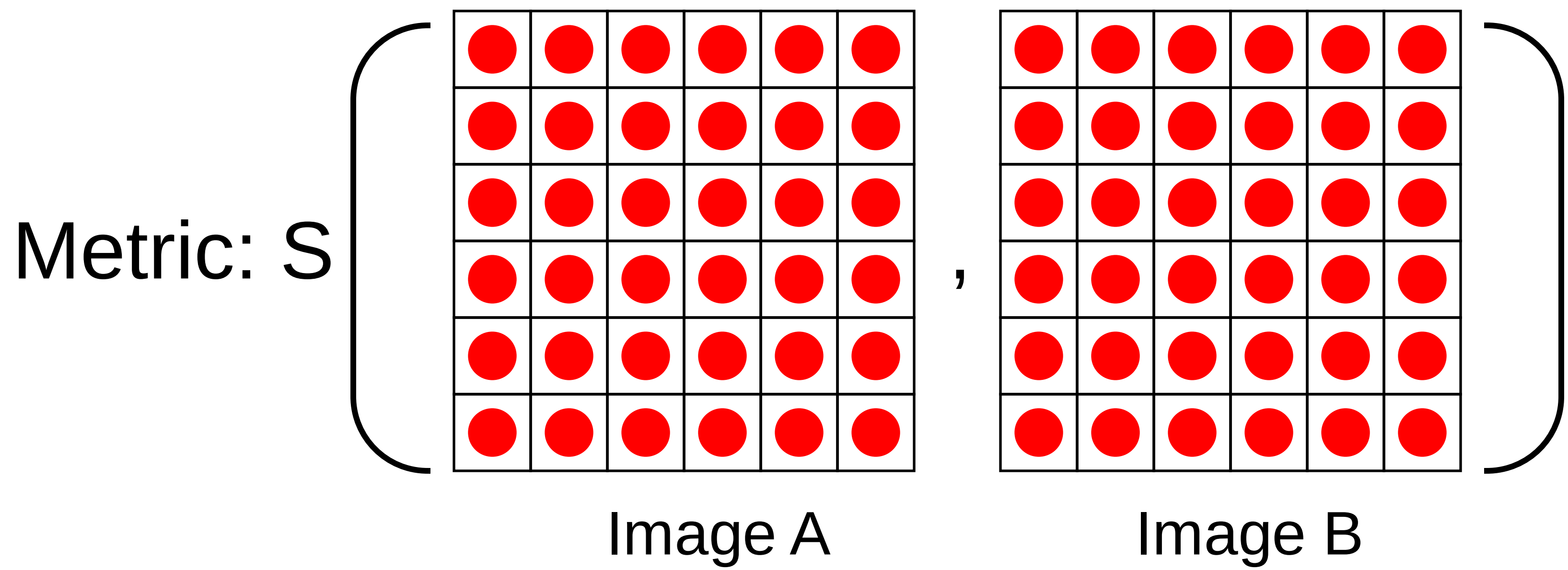
  can be shown to be independent of the amount of overlap between images.
- Registration can be achieved by maximizing (Normalized) Mutual Information between both images (i.e. a similarity measure)

Registration in practice



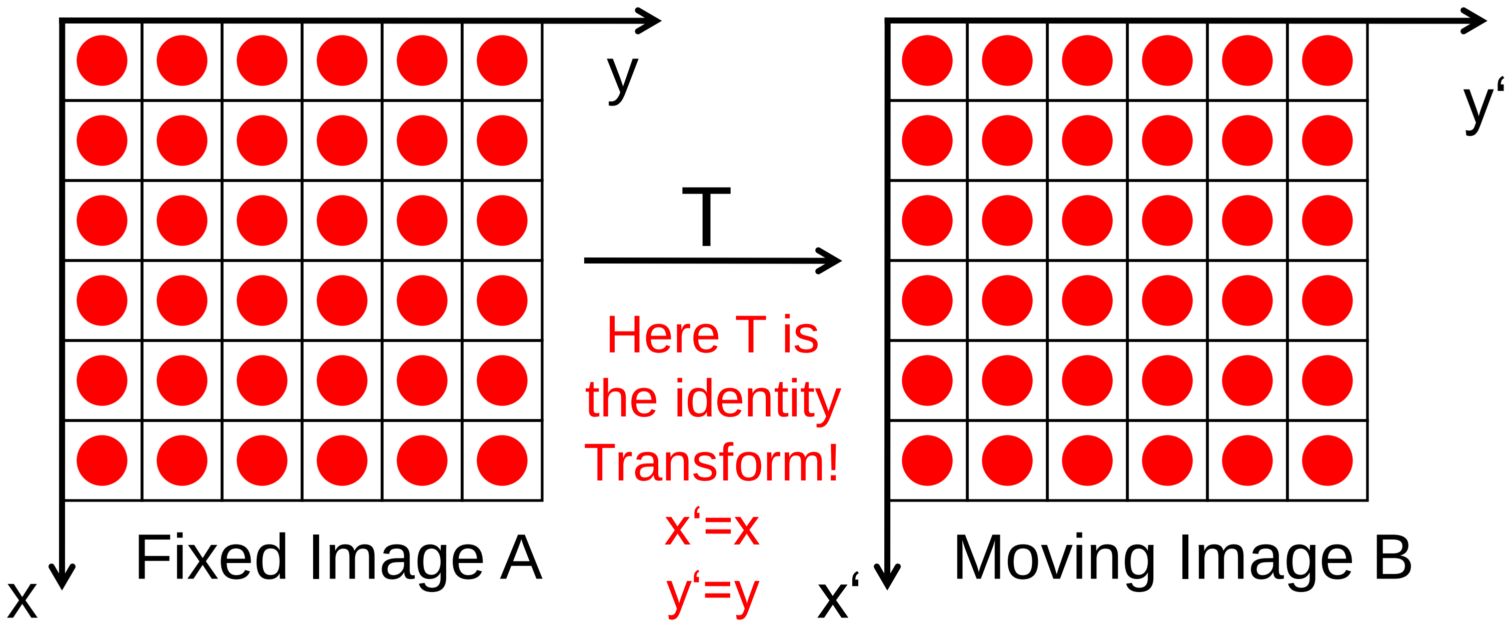
Registration Comparing Images

- (Dis-)Similarity Metric (S) computation needs discrete representations of two images

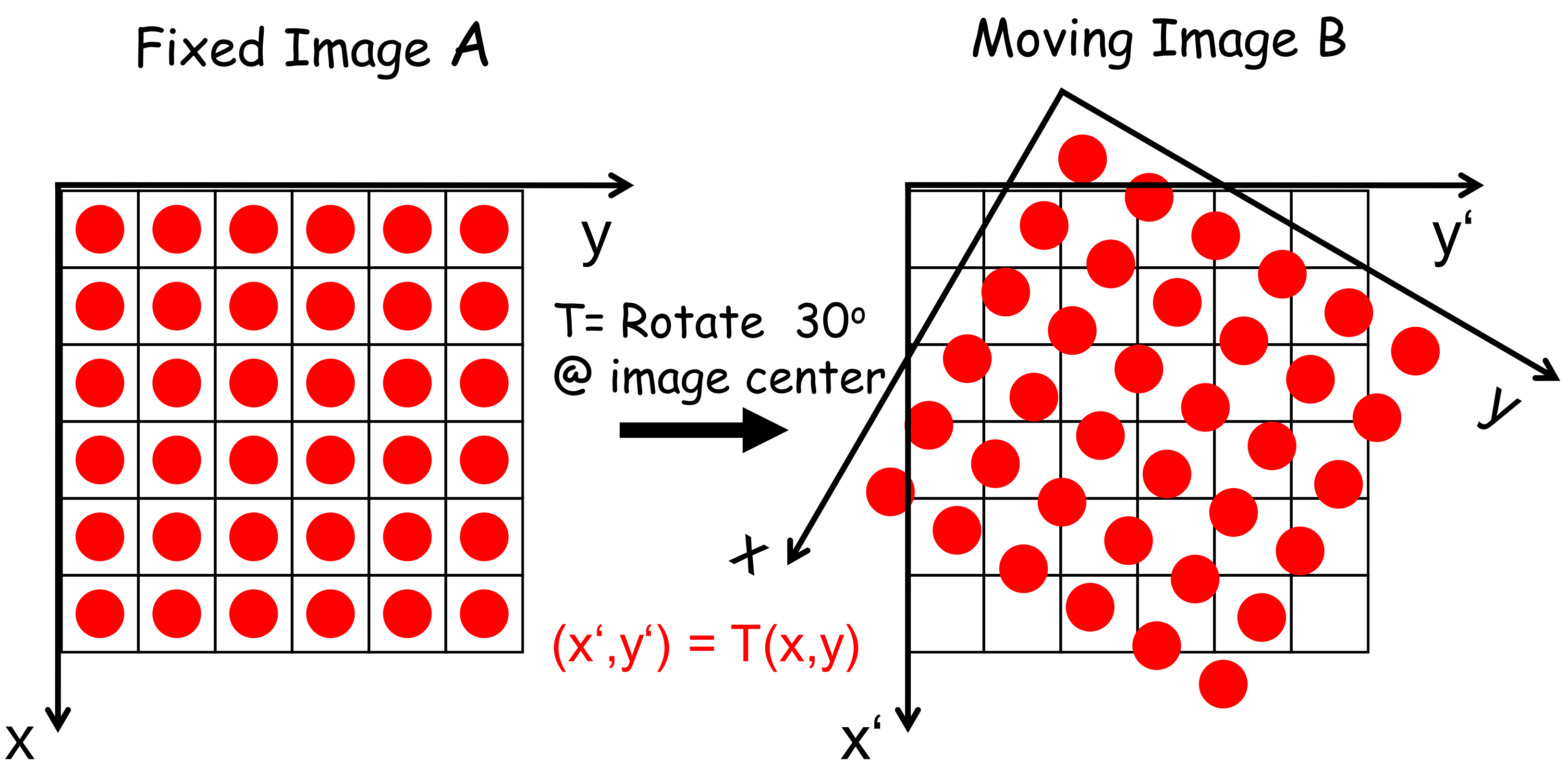


Registration Transformation

- Relation of the two coordinate systems depends on transformation T

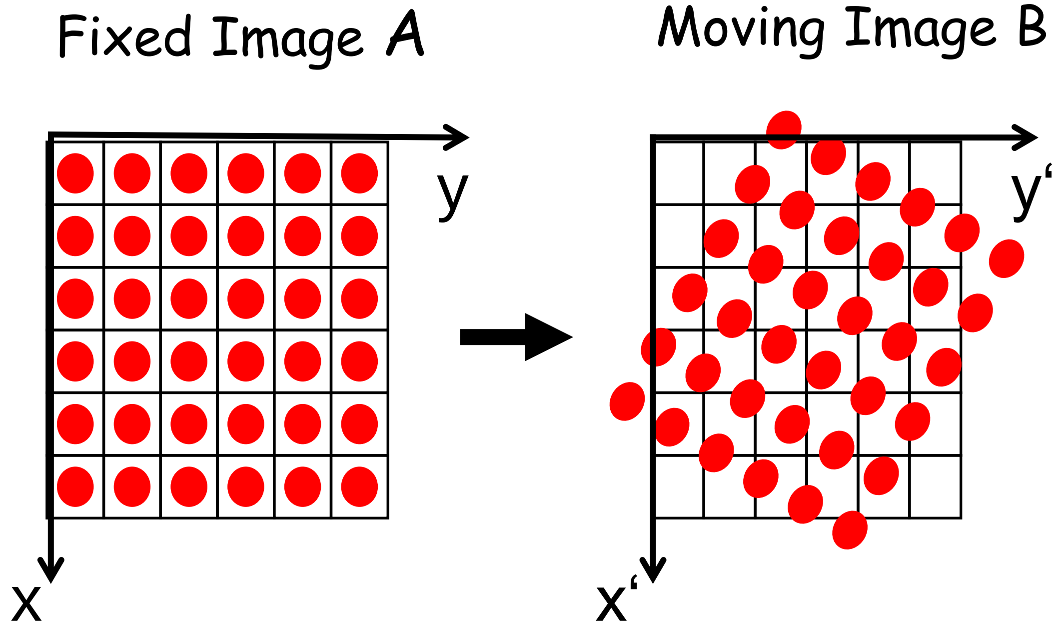


Forward transformation



- Coordinate system of image A mapped onto image B

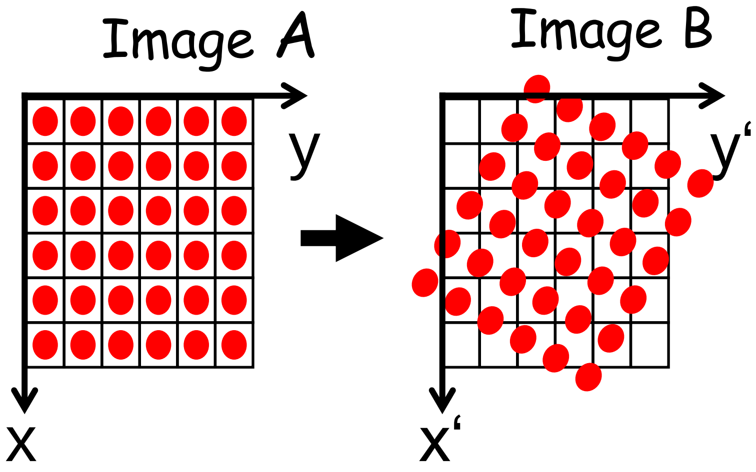
How to calculate Forward transform?



- Forwards Mapping
- Inverse Mapping

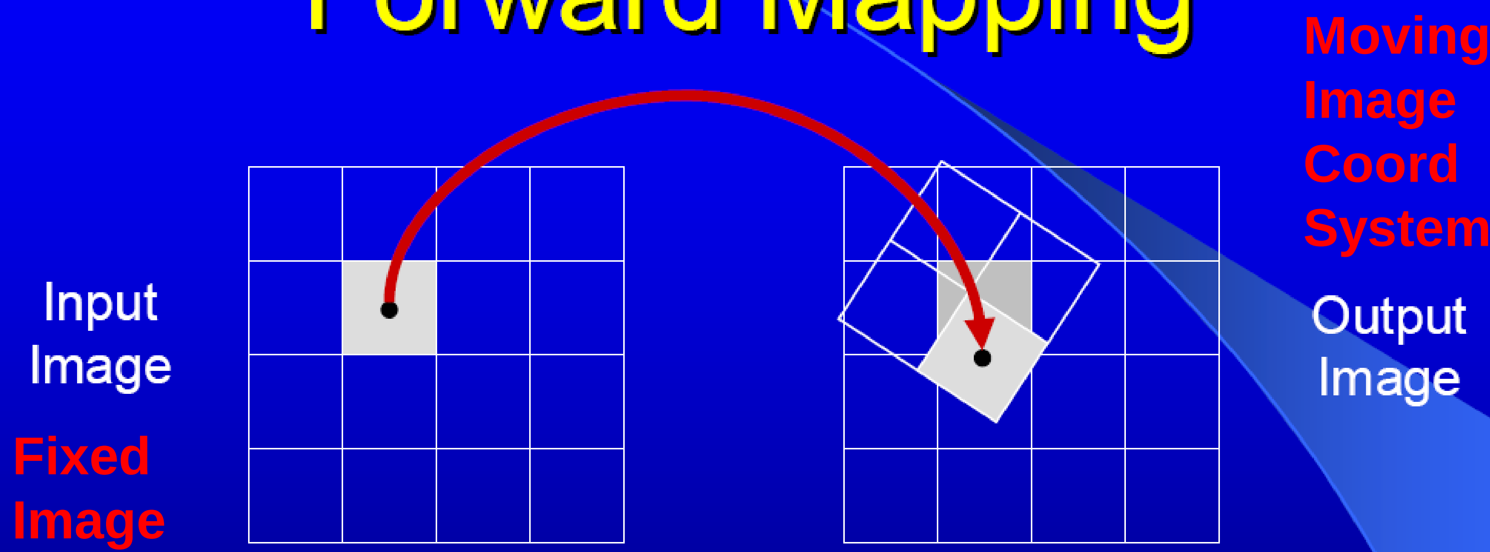
How to calculate Forward transform?

- Forwards Mapping
  - Generate output image of same size as image B
  - Go over voxels of A, transform coordinate using T and put the intensity of voxel from A into transformed location in output
  - Compare output image and image B in coordinate frame of image B



Problem with Forward Mapping

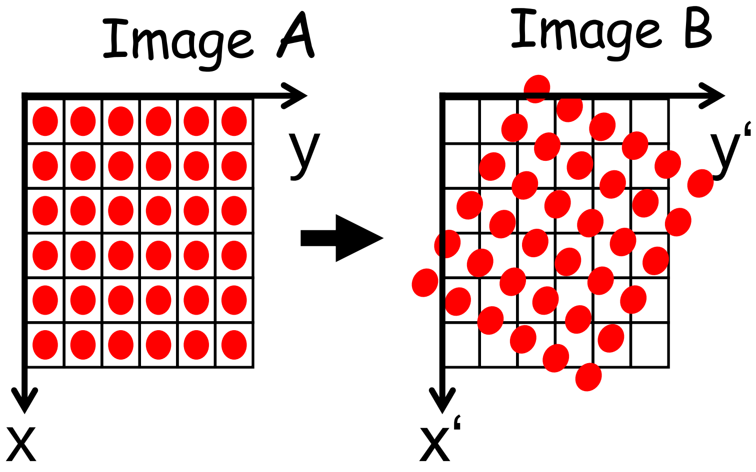
Forward Mapping



- Input image pixel is mapped onto the output image
- Output pixels with more than one hit: overlap
  - Value must be accumulated from overlapping pixels
- Output pixels with no hits: hole

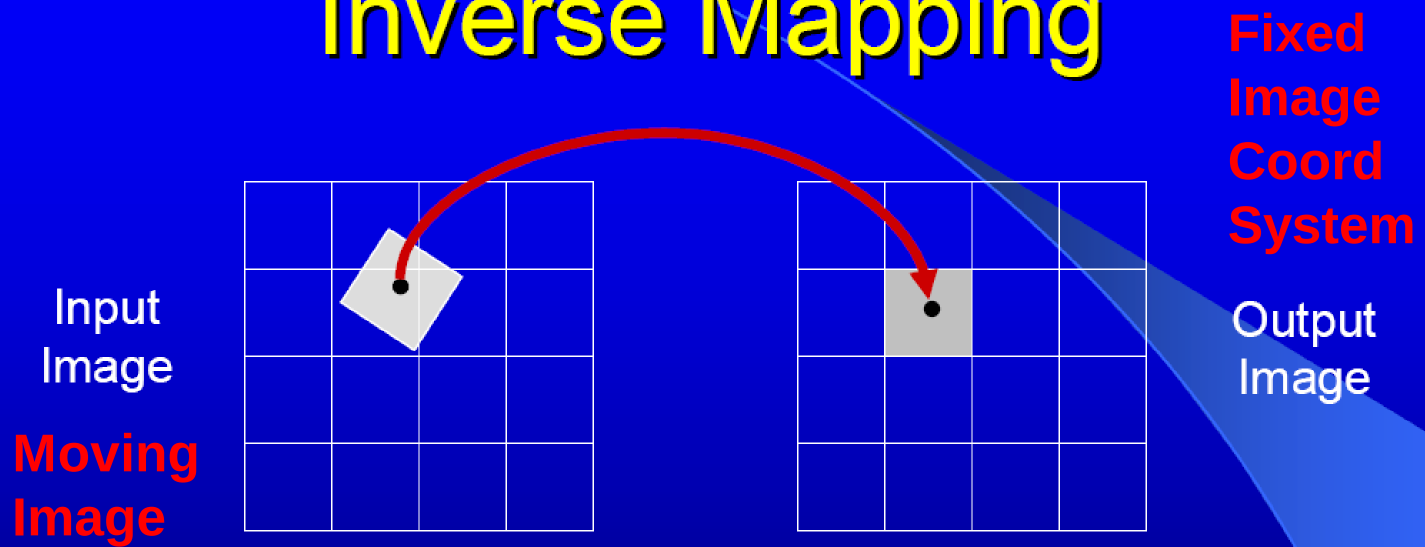
How to calculate Forward transform?

- Inverse Mapping
  - Generate output image of same size as image A
  - Go over voxels of image A, transform coordinate using T, interpolate the intensity of voxel from B and put the interpolated value into same coordinate of output image
  - Compare output image and image A in coordinate frame of image A



Backwards (Inverse) Mapping

Inverse Mapping



- Output pixels are mapped back onto the input image
- Scheme avoids any holes and overlaps in the output image because all pixels are scanned sequentially
- Output pixel value must be interpolated from a neighborhood in the input image

Registration in practice – Summary so far

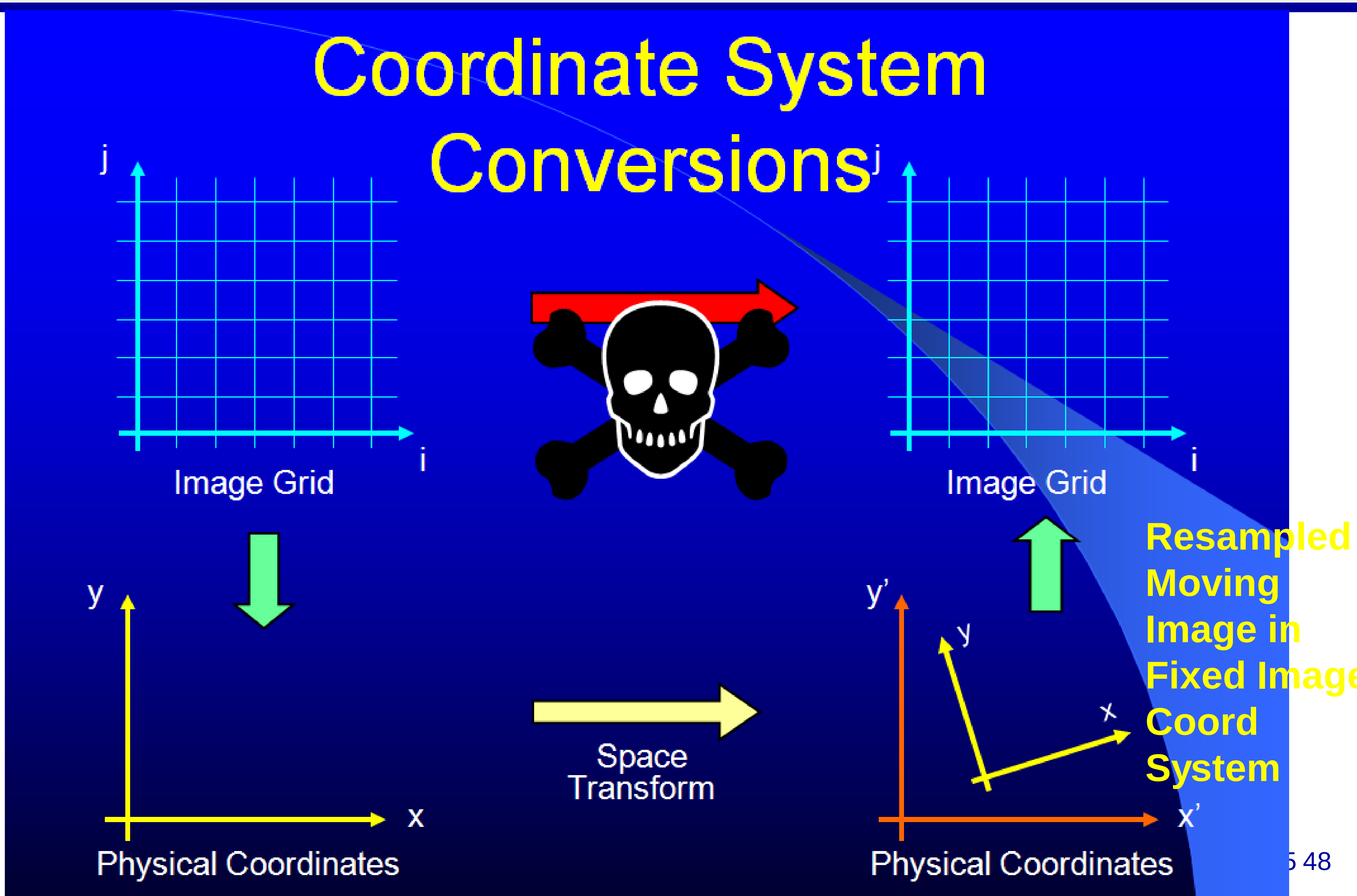
- Calculate (dis-)similarity metric in fixed image coordinate frame!
- Registration frameworks use inverse mapping to interpolate moving image
- Transforms map points from fixed image space to moving image space (forward transform)

$$\mathbf{x'} = \mathbf{T}(\mathbf{x})$$

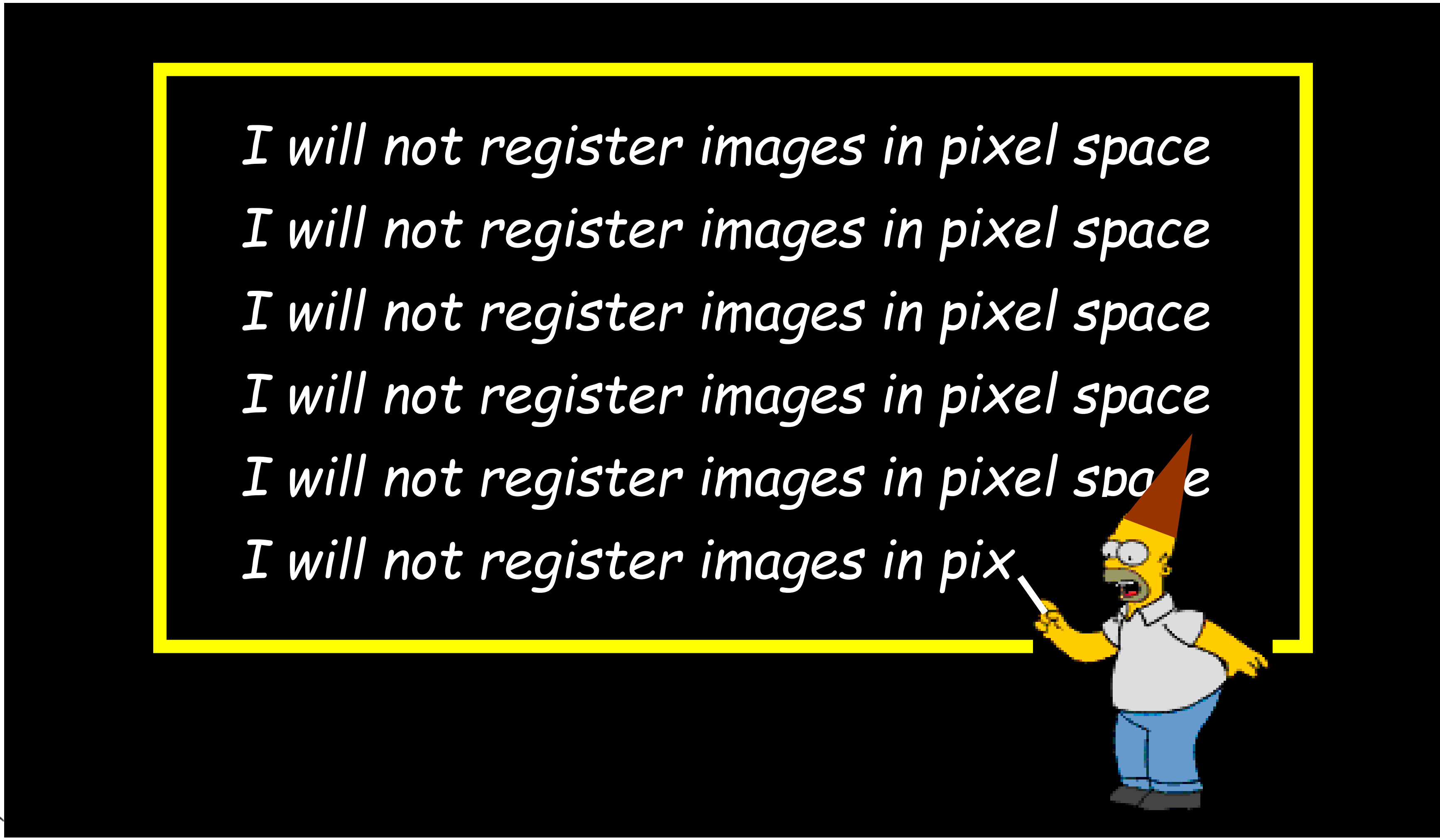
Point in moving image space

Point in fixed image space

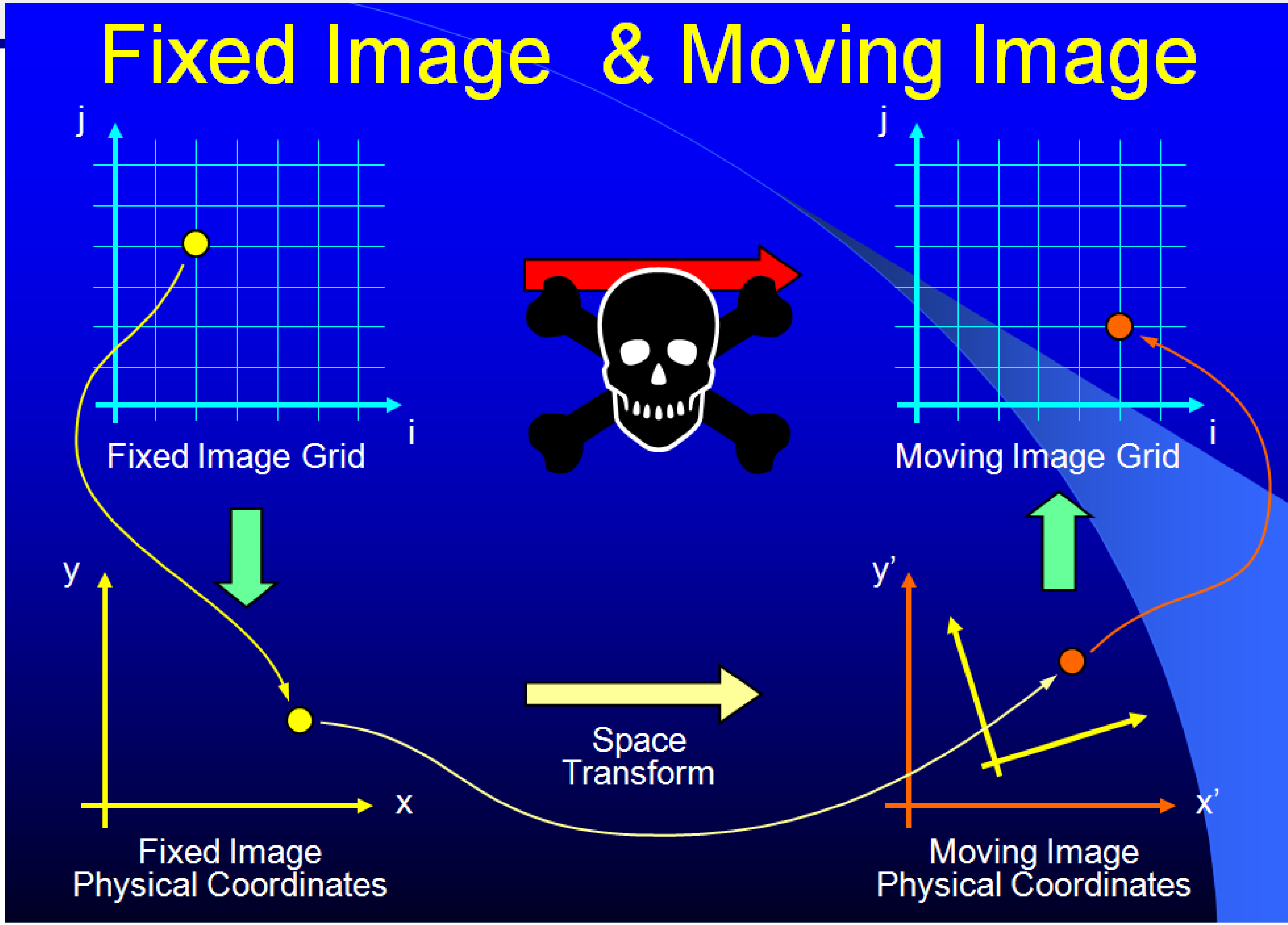
Registration in practice



Registration in practice



Registration in practice

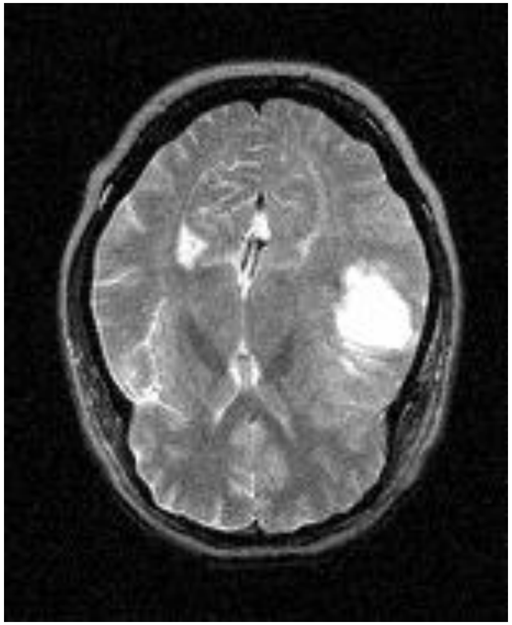


Registration in practice

- In principle the denomination of fixed & moving image is arbitrary
- In practice the moving image is the one that will be resampled into the fixed image coordinate system

Registration in practice

MRI-T2



256 x 256 pixels

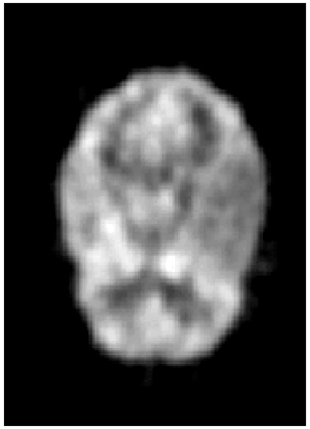
Quiz #1

Images from the same patient

Moving Image ?

Fixed Image ?

PET

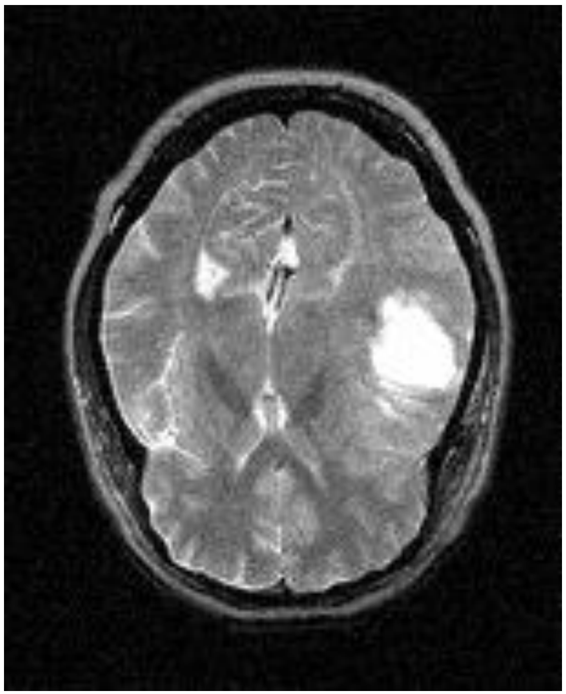


128 x 128 pixels

Reason: Resample
higher resolution image!

Registration in practice

MRI-T2



256 x 256 pixels

Quiz #2

What scale factor ?

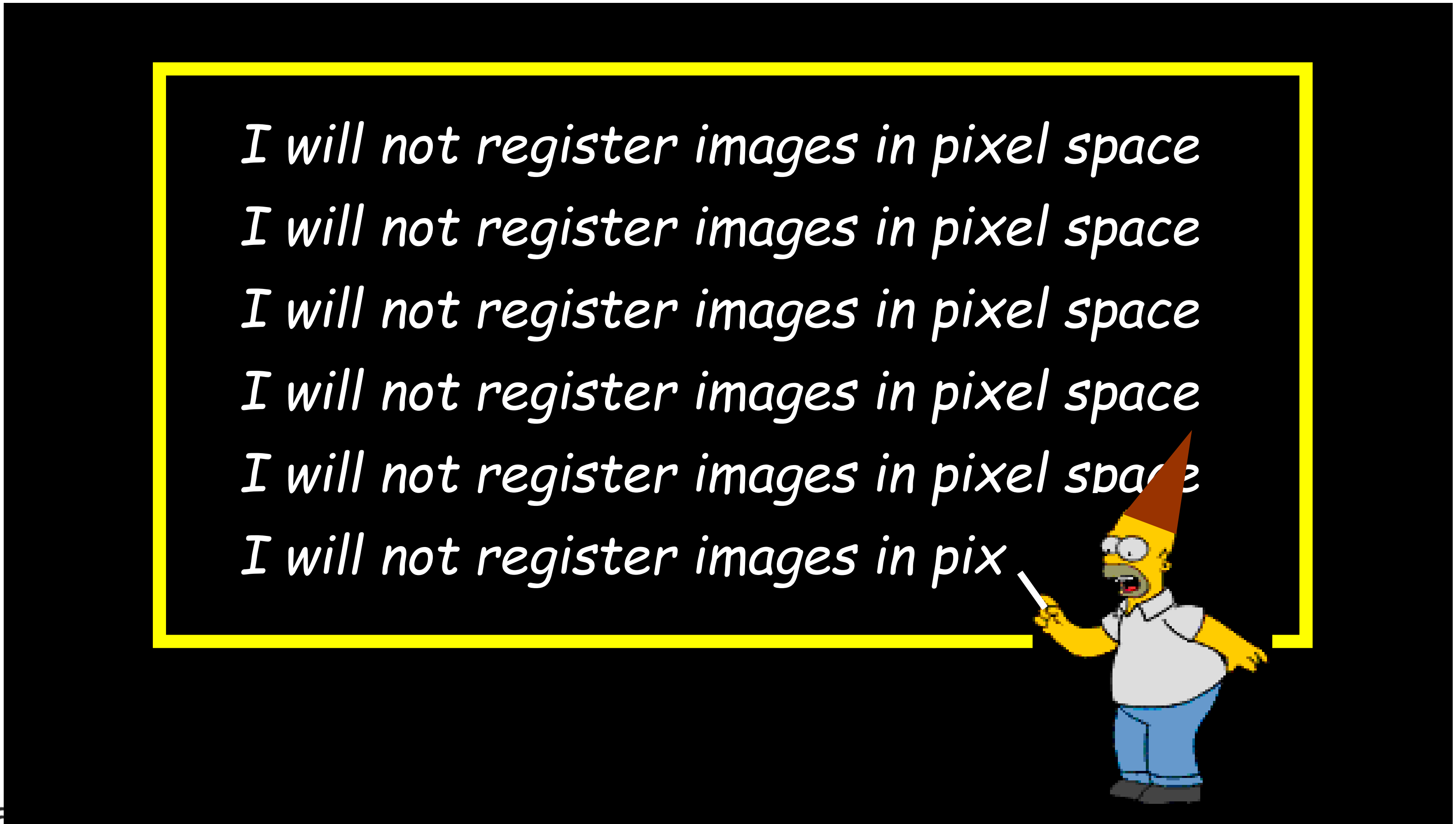
a) 2.0
b) 1.0
c) 0.5

PET

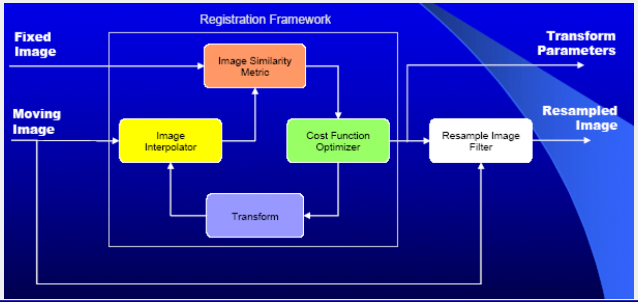
128 x 128 pixels

Images provided as part of the project: "Retrospective Image Registration Evaluation",
NIH, Project No. 8R01EB002124-03, Principal Investigator, J. Michael Fitzpatrick, Vanderbilt University, Nashville, TN.

Registration in practice



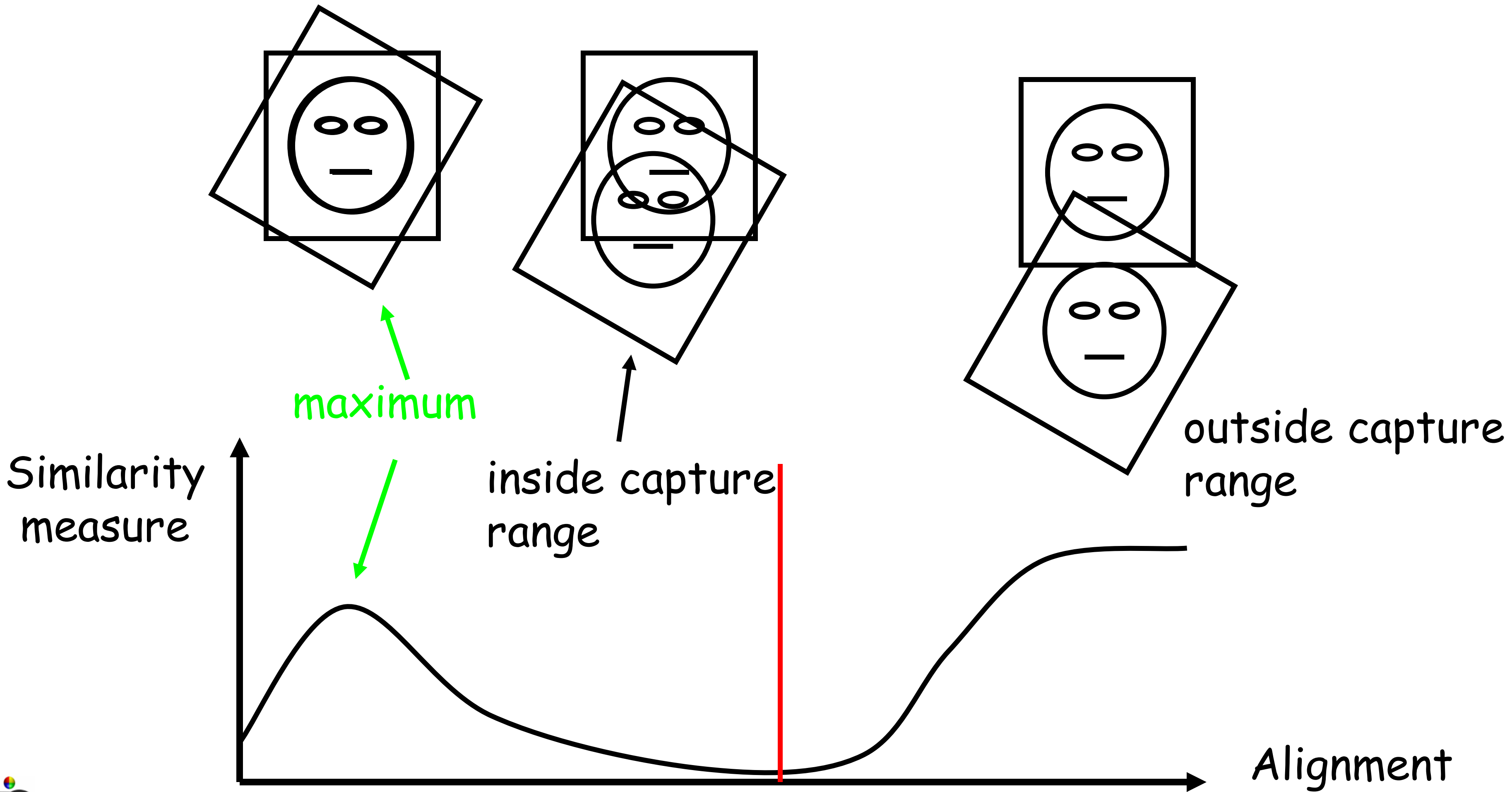
Optimization



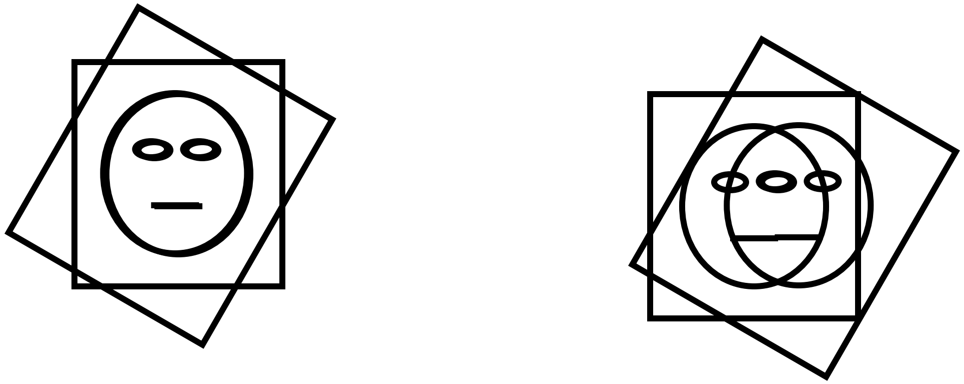
- Optimization of voxel-similarity measures normally requires iterative techniques, i.e.
  - gradient descent
  - Gauss-Newton, Levenberg-Marquardt, BFGS
  - see Numerical Recipes for a description of various optimization schemes
- Global optimization schemes are not feasible for image registration (exhaustive search, genetic)
- Local optimization schemes are much more efficient but will get trapped in local optima

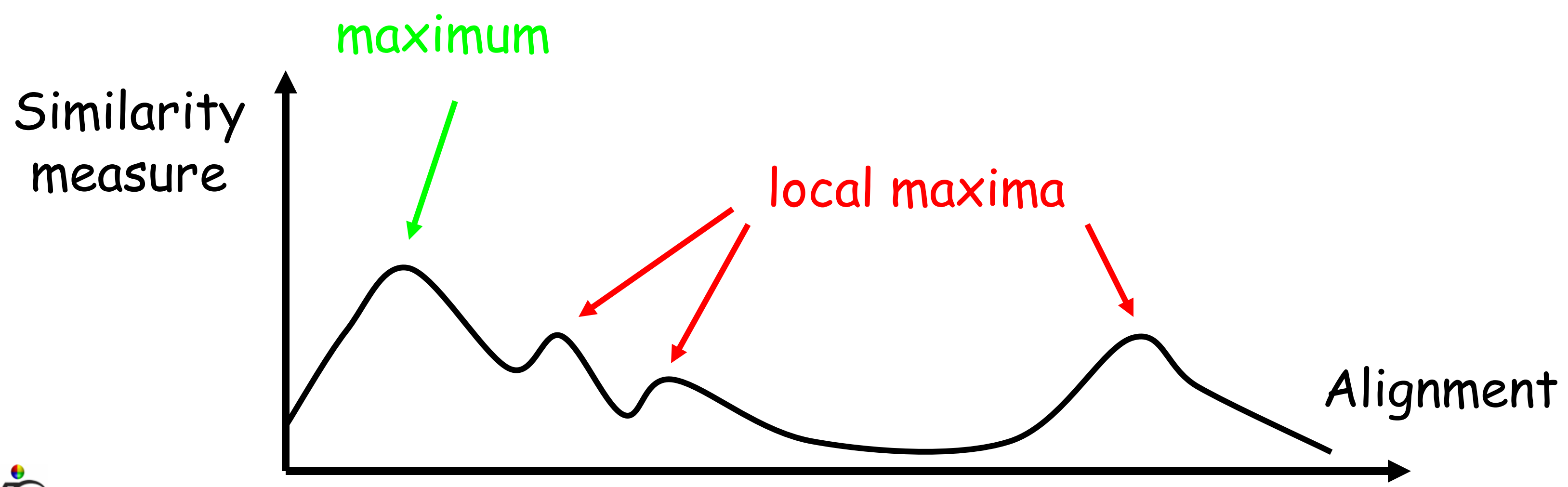
$\Rightarrow$ Registration has a limited capture range

Optimization (Maximization)



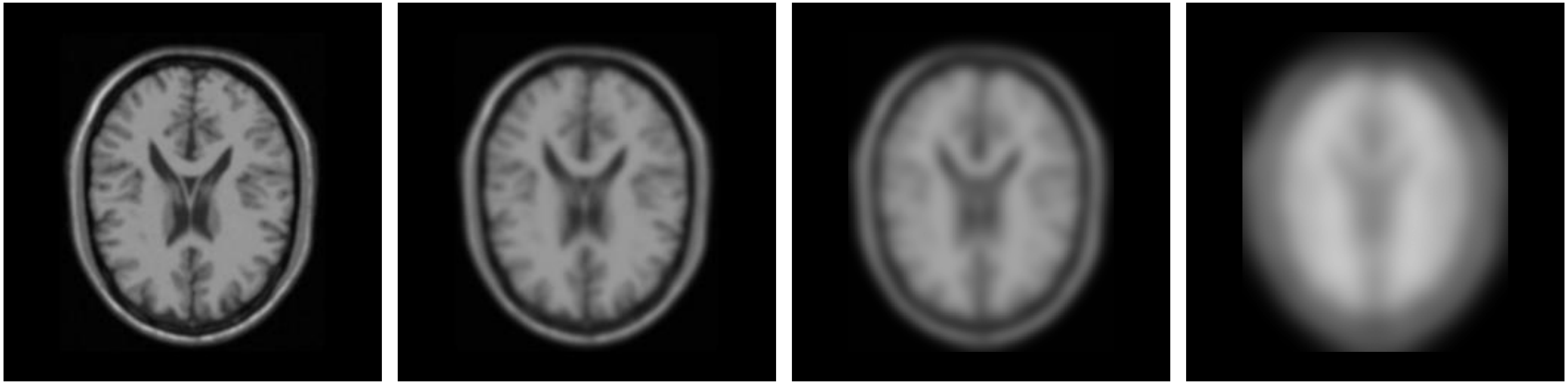
Optimization





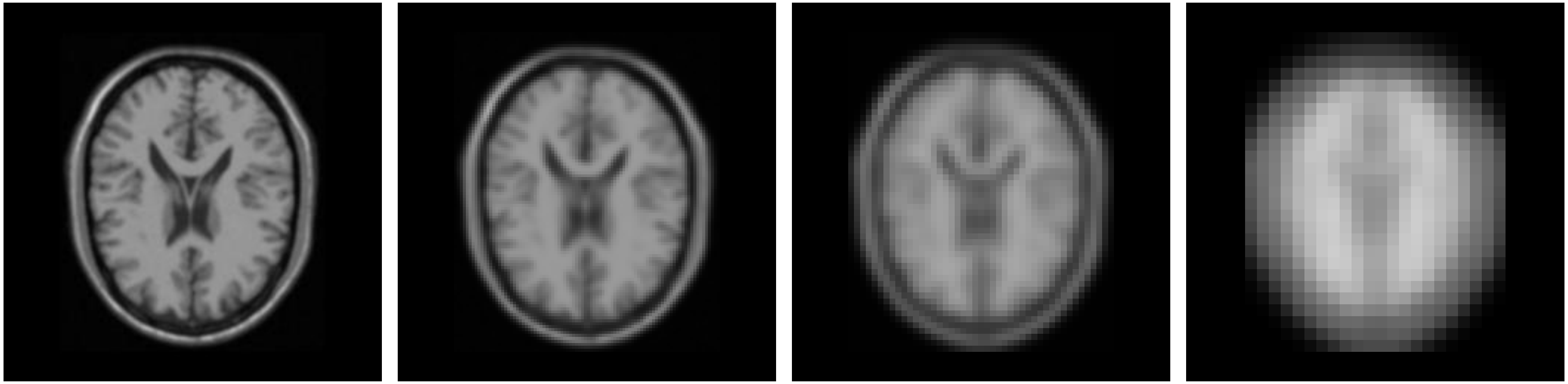
Multi-resolution optimization

- Capture range can be increased by using multi-scale techniques:

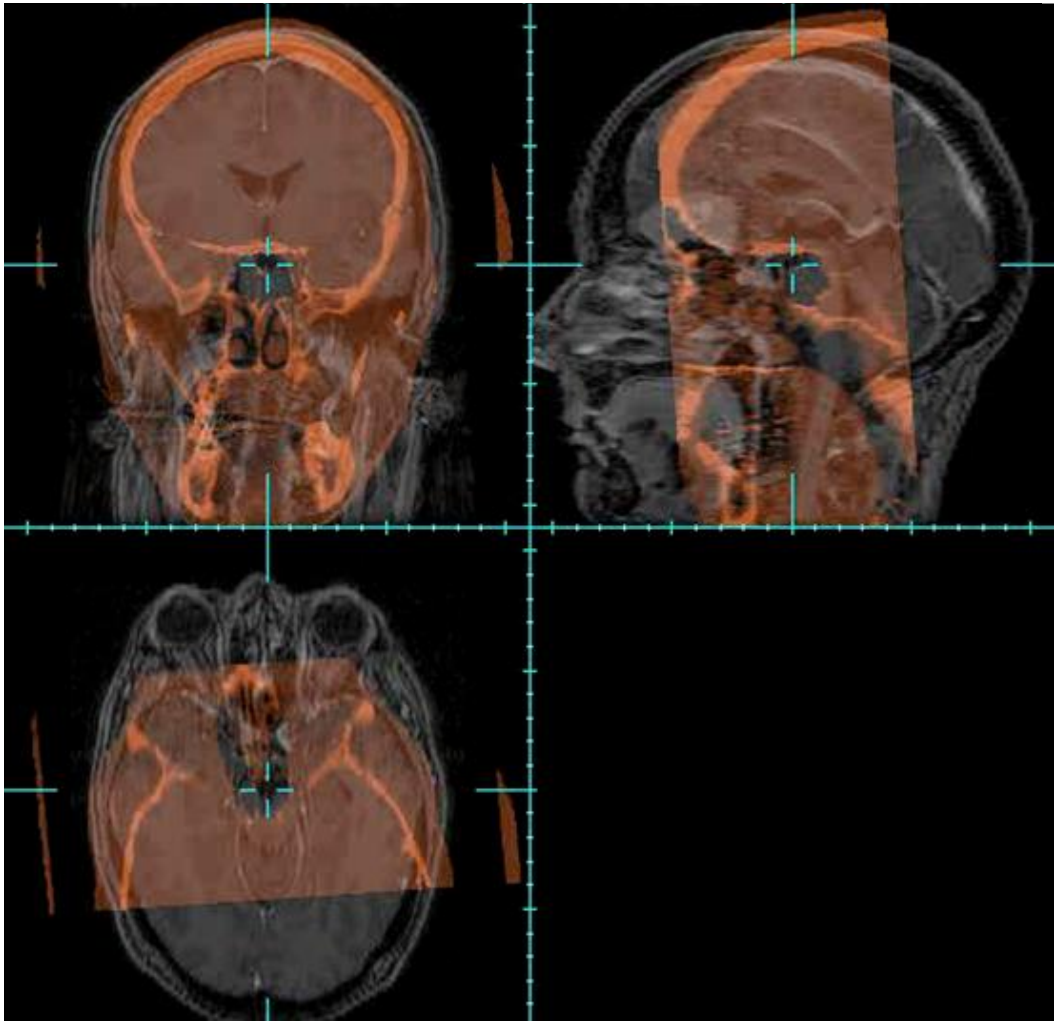


Multi-resolution optimization

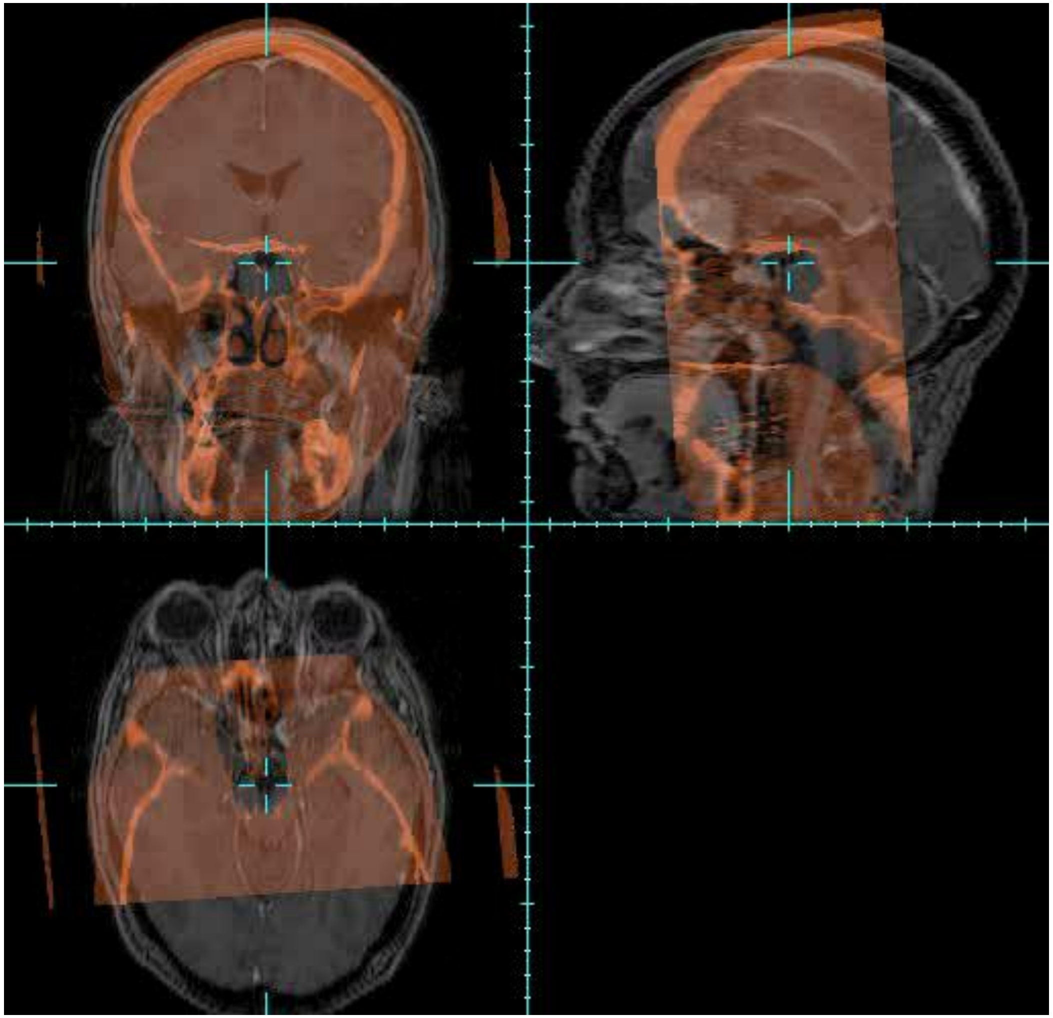
- Registration can be accelerated by using multi-resolution techniques:



Optimization of voxel-similarity measures



Optimization of voxel-similarity measures



Discussion

- Voxel-similarity registration is a powerful technique for automatic image registration
  - Range of voxel similarity measures exist, whose suitability depends on the imaging modality, the image quality, and the anatomy

- Information-theoretic methods are particularly suitable for multi-modal images
  - Joint intensity histograms are used to compute marginal and joint intensity probabilities