Intensity-based Image Registration

Dr. Ruisheng Su, Assistant Professor

Department of Biomedical Engineering, Medical Image Analysis group

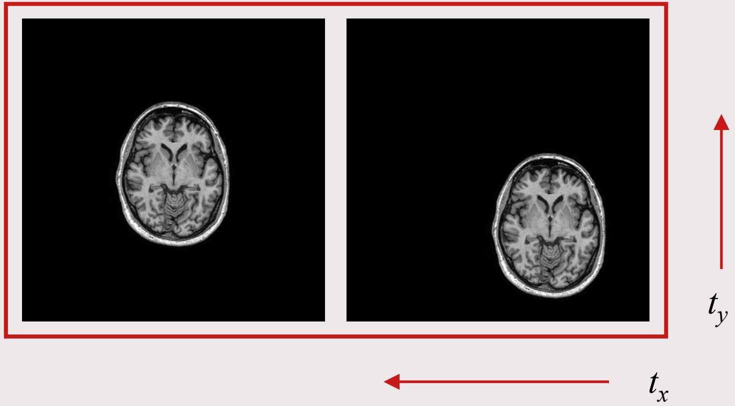
IMAG/e TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Overview & course schedule

Modules	Date	Topic
Registration	April 24 (Thursday)	Course introduction, geometrical transformations
	April 28 (Monday)	Point-based image registration
	May 1 (Thursday)	Intensity-based image registration
Segmentation	May 8 (Thursday)	Introduction to image segmentation
	May 15 (Thursday)	Segmentation in feature space
	May 19 (Monday)	Segmentation using graph-cuts
	May 22 (Thursday)	Statistical shape models
Deep learning for MIA	May 26 (Monday)	Convolutional neural networks
	June 2 (Monday)	Deep learning applications (registration)
	June 5 (Thursday)	Guest lecture by **Danny Ruijters** (principal scientist @ Philips, full professor @ TU/e)
	June 10 (Tuesday)	Deep learning applications (segmentation)
	June 12 (Thursday)	Unsupervised deep learning for medical image analysis

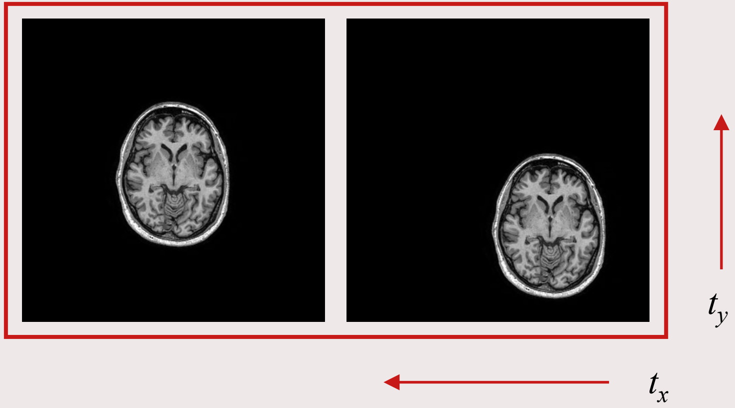
Outline

- Recap of the previous lecture
- Intensity-based similarity metrics
  - Sum of square differences
  - Cross-correlation
  - Mutual information
- Optimization for intensity-based registration
  - Gradient ascent (descent)
- Evaluation metrics



Outline

- Recap of the previous lecture
- Intensity-based similarity metrics
  - Sum of square differences
  - Cross-correlation
  - Mutual information
- Optimization for intensity-based registration
  - Gradient ascent (descent)
- Evaluation metrics



Recap of learning objectives (previous lecture)

The student can:

- explain the role of inverse mapping and interpolation when transforming an image and compute inverse transformation matrices

- design a general algorithm to register two images based on fiducials (i.e., composing an error function and minimizing this function w.r.t. the transformation *T*)
- use optimization to find the minimum of this error function
- give at least four reasons why perfect alignment of multiple fiducials is not possible in practice
- recall the exact solution for *T* (matrix notation) when constrained to affine registration
- describe the algorithm required to find the optimal parameters of *T* when constrained to rigid registration (orthogonal Procrustes problem)
- explain the principle behind the iterative closest point algorithm

- use the target registration error (TRE) to evaluate image registration

Optimization (affine transformation)

Step 2, find the minimum of the error w.r.t. to the parameters.

$$E(\mathbf{T}) = ||\mathbf{T}\mathbf{X}' - \mathbf{X}||_F^2$$

Set of equations:

$$\nabla_{\mathbf{T}} E(\mathbf{T}) = 0$$

Solution for affine transformation:

$$\mathbf{T} = \mathbf{X}'\mathbf{X}^{\intercal}(\mathbf{X}\mathbf{X}^{\intercal})^{-1} \quad \rightarrow \quad \mathbf{T} = \mathbf{X}\mathbf{X}'^{\intercal}(\mathbf{X}'\mathbf{X}'^{\intercal})^{-1}$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_i & \dots & \mathbf{x}_n \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,i} & \dots & x_{1,n} \\ x_{2,1} & x_{2,2} & \dots & x_{2,i} & \dots & x_{2,n} \\ 1 & 1 & \dots & 1 & \dots & 1 \end{bmatrix}$$

Recap

Quiz: When transforming an image, why is inverse mapping preferred over forward mapping?

A. It preserves original pixel intensities exactly
B. It requires no interpolation
C. It avoids holes and overlaps in the target grid
D. It eliminates the need to compute the inverse transformation matrix

Recap

Quiz: Which of the following is **not** a reason that perfect alignment is impossible in point-based registration in practice?

A. Fiducial localization error due to image noise
B. Fiducials moving between acquisitions
C. Numerical round-off error in matrix inversion
D. Too many fiducials are selected

Recap

Quiz: The rigid registration problem constrained to rotation and translation only (no scaling) can be solved by:

A. Computing the pseudo-inverse of point correspondences
B. Performing SVD on the covariance matrix of centered points
C. Direct least-squares fit of an affine matrix
D. Using histogram matching

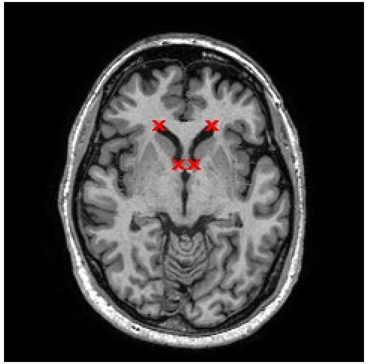
Recap

Quiz: Which factor does not directly affect TRE?

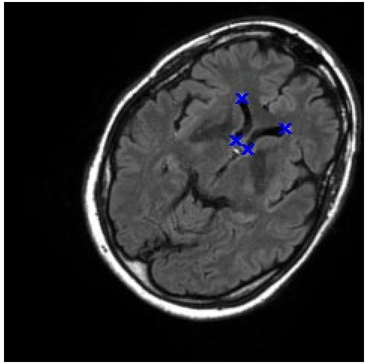
A. Fiducial localization error.
B. Distribution of fiducial points relative to target
C. Measurement noise in target point positions
D. The choice of similarity metric used for image registration

Recap

Point-based registration requires some manual input.
Can it be used with automatic key point selection?



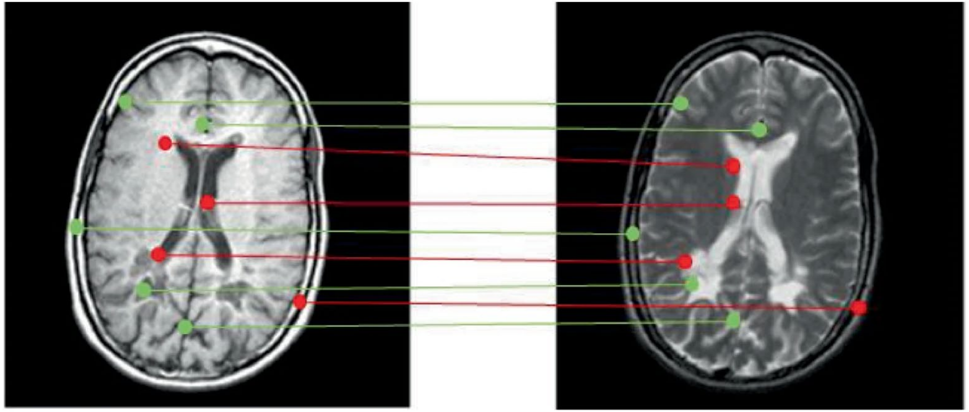
Fixed



Moving

Recap

Point-based registration requires some <u>manual</u> input.
Can it be used with <u>automatic</u> key point selection?



Matching by Scale-invariant feature transform (SIFT)

Common key-point detectors

- SIFT: Scale-invariant feature transform
- SURF: Speeded-Up Robust Features
- FAST: Features from accelerated segment test
- BRIEF: robust independent elementary features
- ORB: Oriented FAST and Rotated BRIEF

Image source: Hossein-Nejad, Z., & Nasri, M. (2024). A new hybrid strategy in medical image registration based on graph transformation matching and mean-based RANSAC algorithms. Multimedia Tools and Applications, 83(35), 82777-82804.

Recap - Classification of image registration

- Image dimensionality: 2D, 3D, 3D + time…
- Registration basis: point sets, intensity…
- Geometrical transformations: rigid, affine, nonlinear…
- Degree of interaction: automatic, manual, semi-automatic
- Optimization procedure: closed-form solution, iterative
- Modalities: multi-modal, intra-modal
- Subject: inter-patient, intra-patient, atlas
- Object: brain, head, vertebra, liver…

Intended learning outcomes of this lecture

The student can:

- **explain** and **implement** three important intensity-based image **similarity metrics**, namely sum of squared differences, normalized cross correlation, and mutual information.
- **select** the correct image similarity metric for an image registration task based on the assumptions of these metrics.

- explain how the **joint probability mass function (p.m.f.)** can be used to measure the similarity between two images if we consider the image intensities as random variables.
- **interpret joint histograms** to judge whether two images are well aligned.

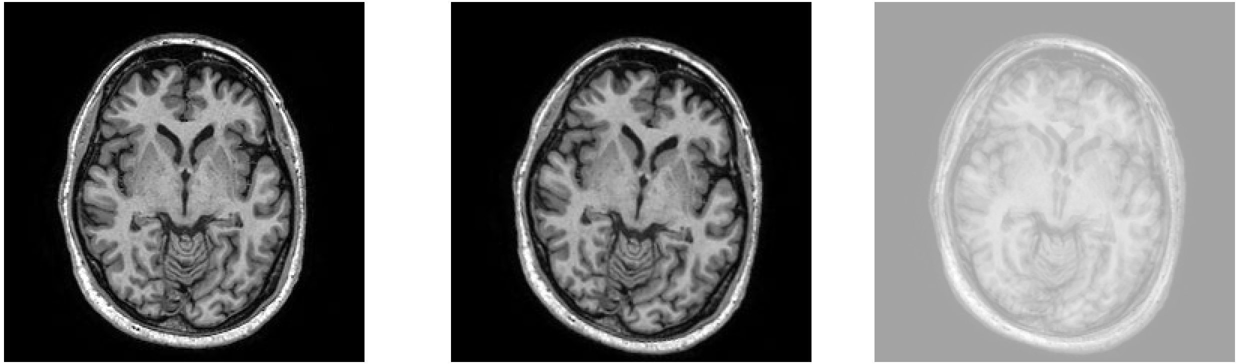
- describe the numerical **procedure** to register two images by maximizing a similarity function (gradient ascent/descent).
- explain the effect of the **learning rate** and the **initialization** of the parameters on the optimalization process when using gradient ascent/descent.

Intensity-based registration

Image intensity is an alternative registration basis to points.

It is the most widely used registration basis.

Compared to point-based registration, requires less user interaction.



Intensity-based registration

Intensity-based image registration works by **iterative optimization** of an

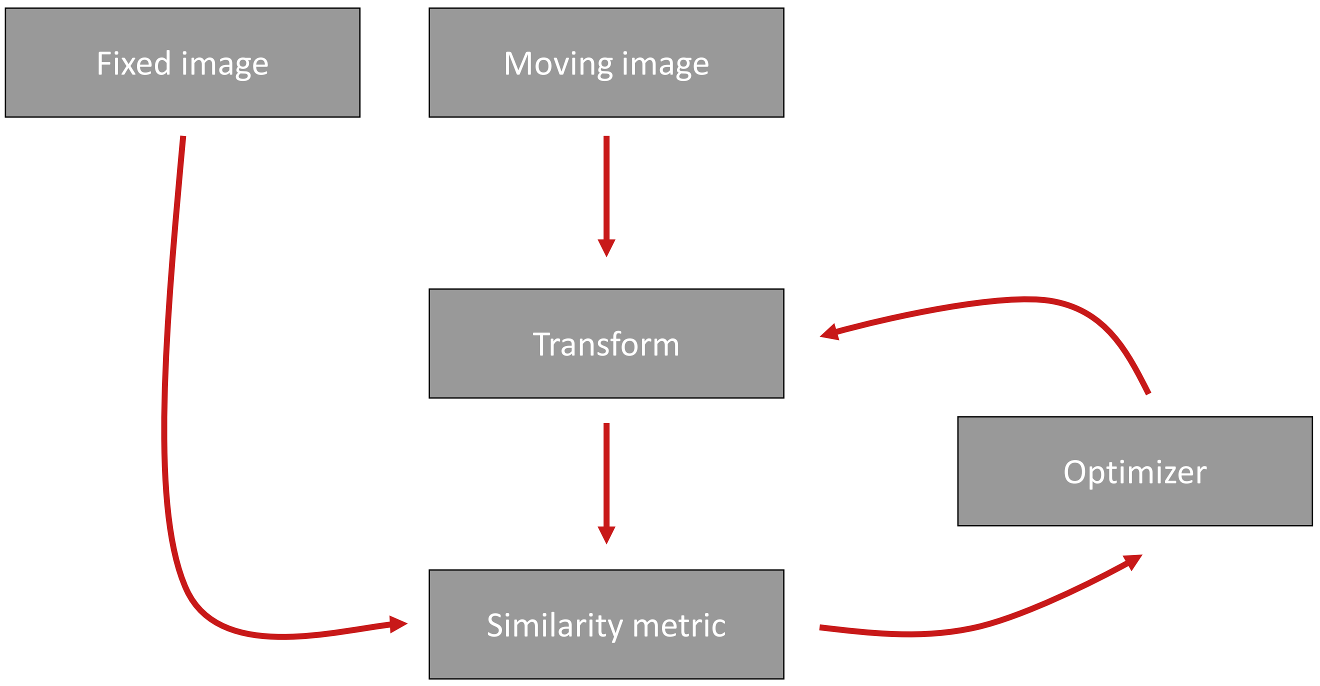
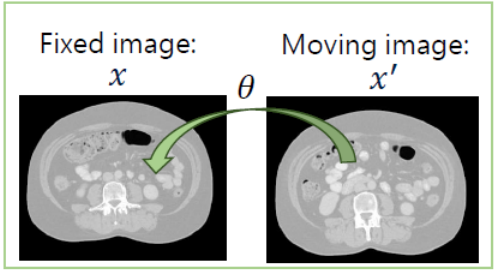


Image registration overview

Medical image registration: general “recipe”



Transformation model:
(e.g., rigid, affine, deformable)

$T(x', \theta)$

Image registration

Similarity measure:
(e.g. SSD, CC, MI, MSE)

$S(x, x')$

Optimization:

$$\hat{\theta} = \max_{\theta} S\big(x, T(x', \theta)\big)$$

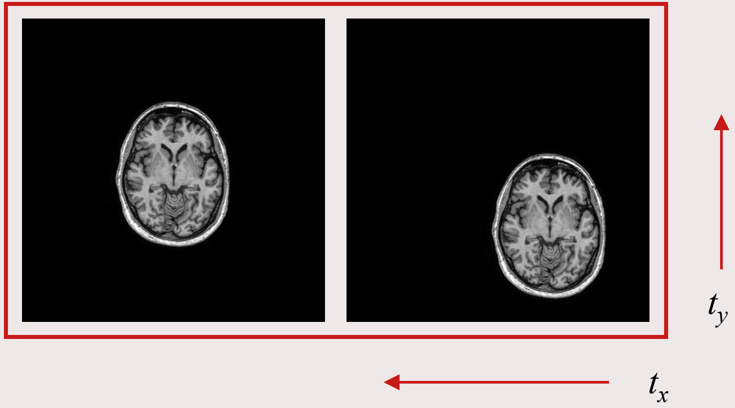
Outline

- Recap of the previous lecture

- Intensity-based similarity metrics
    - Sum of squared differences (SSD)
    - Normalized cross-correlation (NCC)
    - Mutual information (MI)

- Optimization for intensity-based registration
    - Gradient ascent (descent)

- Evaluation metrics



Sum of squared differences (SSD)

Let *I* and *J* be two images and *i* the pixel locations.

A simple and intuitive intensity-based measure of the similarity of *I* and *J* is the sum of squared differences (SSD):

$$\mathrm{SSD}(I,J) = \sum_{i=1}^{n} (I(i) - J(i))^2$$

$$\begin{array}{|c|c|c|}\hline 10 & 12 & 16 \\ \hline 8 & 12 & 18 \\ \hline 4 & 8 & 10 \\ \hline \end{array} \; - \; \begin{array}{|c|c|c|}\hline 8 & 14 & 16 \\ \hline 12 & 12 & 20 \\ \hline 2 & 6 & 12 \\ \hline \end{array} \; \longrightarrow \; \begin{array}{|c|c|c|}\hline (10-8)^2 & (12-14)^2 & (16-16)^2 \\ \hline (8-12)^2 & (12-12)^2 & (18-20)^2 \\ \hline (4-2)^2 & (8-6)^2 & (10-12)^2 \\ \hline \end{array}$$

Sum of squared differences (SSD)

If *I* is the fixed image in a registration problem,
and *J* is the moving image transformed with a transformation **T,**

the similarity measure will be a function of the transformation :

$$\mathrm{SSD}(I, J, \mathbf{T}) = \sum_{i=1}^{n} (I(i) - J_{\mathbf{T}}(i))^2$$

The SSD will be lowest when the images are perfectly aligned and will increase with misalignment.

When lowest == zero?

Assumptions and drawbacks

It can be shown that this measure is optimal when two images differ only by **Gaussian noise**. This is an implicit assumption of this measure.

- This assumption does not hold for inter-modality registration
- This assumption rarely true for intra-modality registration (e.g. MRI noise is not Gaussian, there will be changes between acquisitions etc.)

Nevertheless, SSD can still be used with success in <u>intra-modality registration</u>.

Drawback: It can be very sensitive to a few “**outlier**” intensity differences.

Normalized cross-correlation (NCC)

Another measure that makes slightly less assumptions is normalized cross correlation *C*:

$$C(I,J) = \frac{\sum_{i=1}^{n}(I(i)-\bar{I})(J(i)-\bar{J})}{\sqrt{\sum_{i=1}^{n}(I(i)-\bar{I})^2\sum_{j=1}^{n}(J(i)-\bar{J})^2}}$$

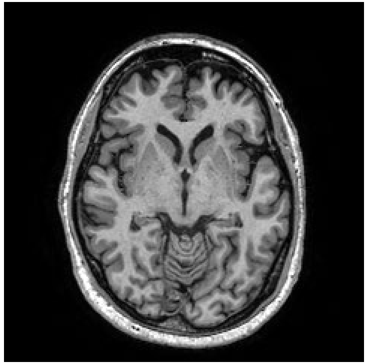
Assumption: the main assumption of normalized cross-correlation is that there is a <u>linear relationship</u> between the pixel intensities in the two images.

→ Mostly true for intra-modality registration

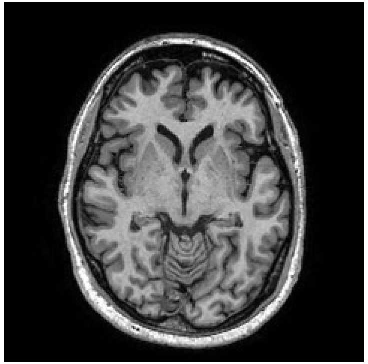
→ Usually not true for inter-modality registration

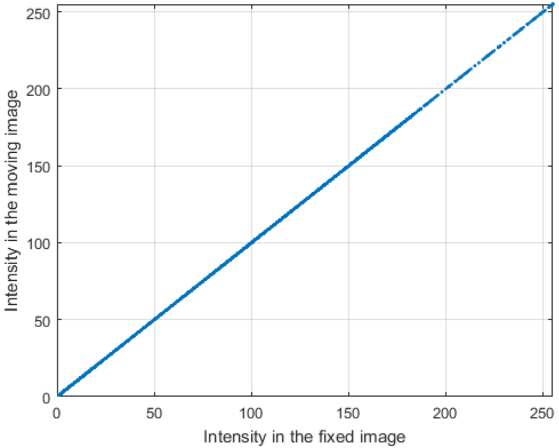
Example

T1



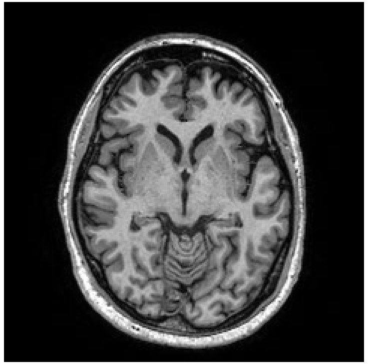
T1



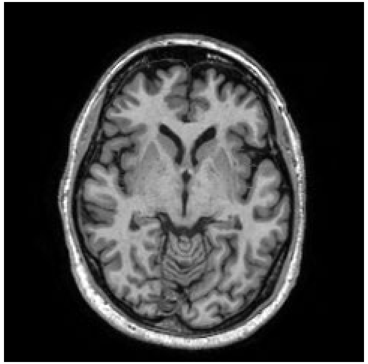


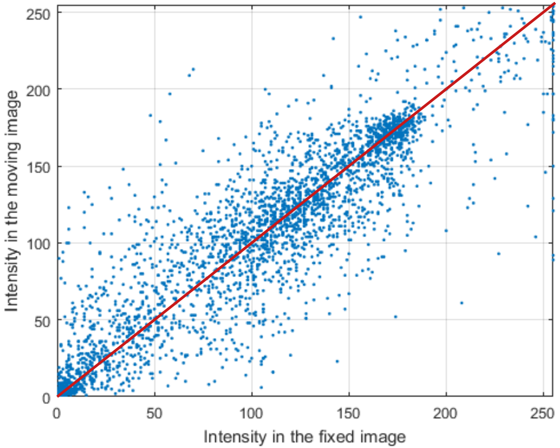
Example

T1



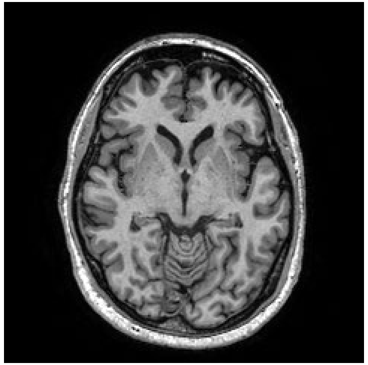
T1



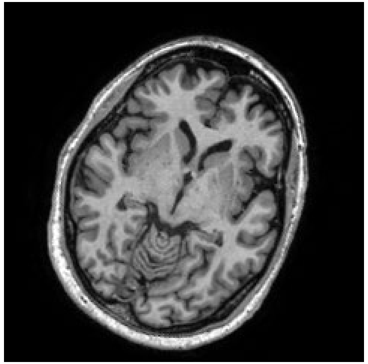


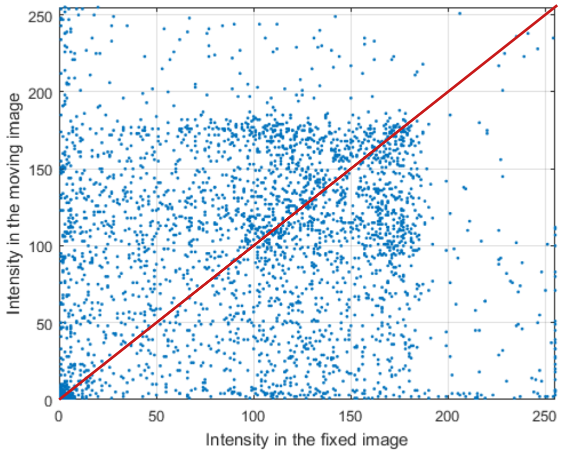
Example

T1



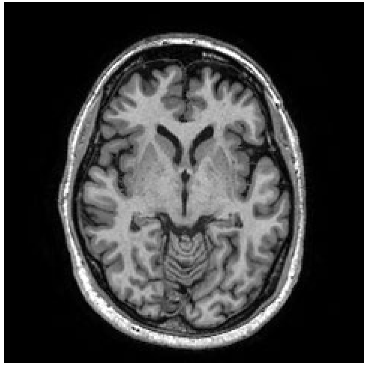
T1



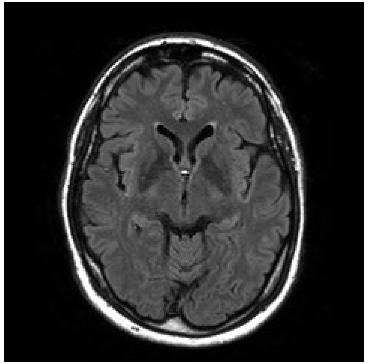


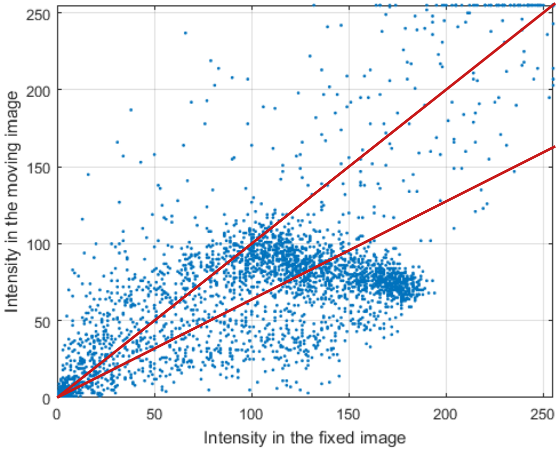
Example

T1



T2

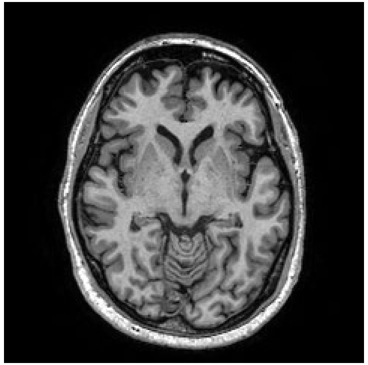




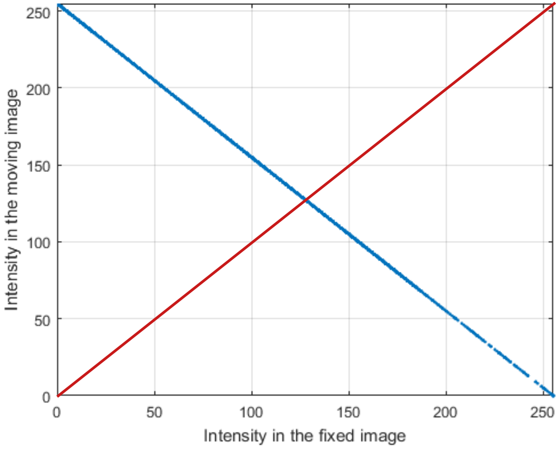
Violates the assumption of linear relationship between the intensities – not optimal

Example

T1



Simulated modality



Completely opposite of the assumption (inverse linear relationship)

Probability theory

Random variables map the outcomes of random phenomena to numbers.

Example random phenomenon: **coin toss**.

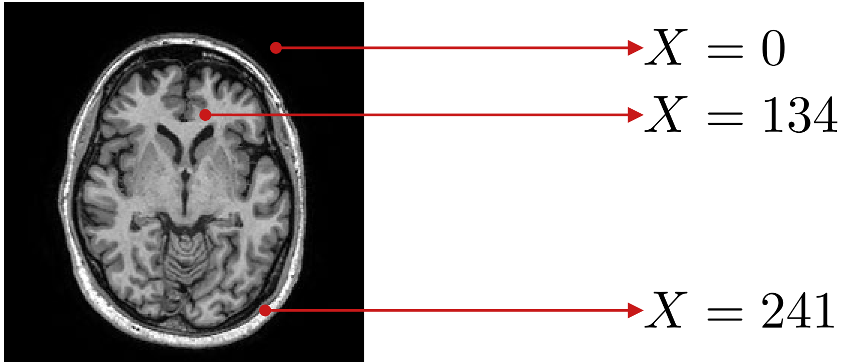
Random variable ***X***: the outcome of the coin toss.
Another random variable ***Y***: the number of heads in a series of 3 tosses.

$$X = \begin{cases} 1, & \text{if heads} \\ 0, & \text{if tails} \end{cases}$$

Probability theory

Image intensities as random variables:

Random phenomenon: pick a random pixel location.
In this case, the pixel intensity can be treated as a random variable.



Probability theory

Each outcome from the random phenomenon we are studying can be associated with a probability.

If a random variable *X* can have a finite set of possible values, we can define a function that maps each possible value to a probability. This function is called probability mass function (p.m.f).

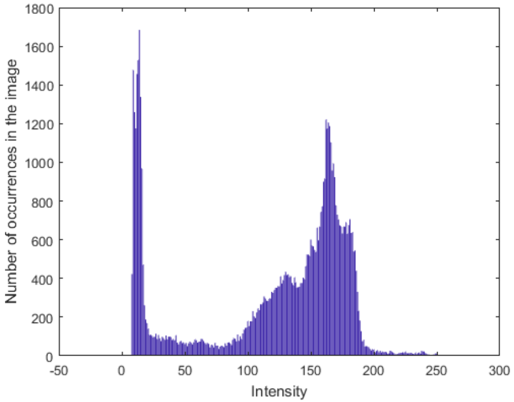
Probability mass function:

$$p_X(x) = P(X = x)$$

Probability theory

How can we define the **probability mass function** for the image intensities?

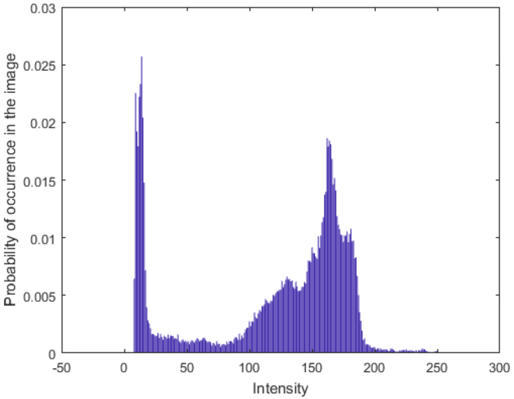
Image histogram – count of the number of occurrences of each intensity value in the image.



Probability theory

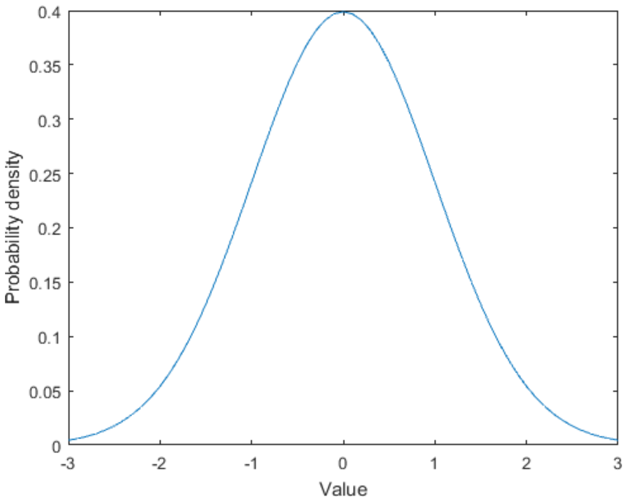
In order to treat the counts of the histogram as probability values, we must **normalize** the histogram in such a way that all values sum to 1.

This is the probability mass function for the pixel intensity as a random variable.



Probability theory

For continuous random variables (can take infinite number of possible values), we can define the probability density function (p.d.f):



Probability theory

What if we have two random variables? For example, the pixel intensity in two images.

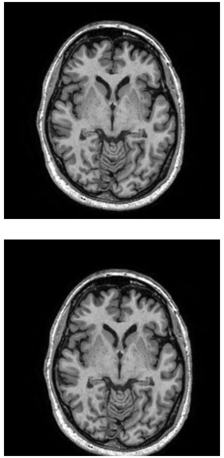
In this case we can define a **joint probability mass function**:

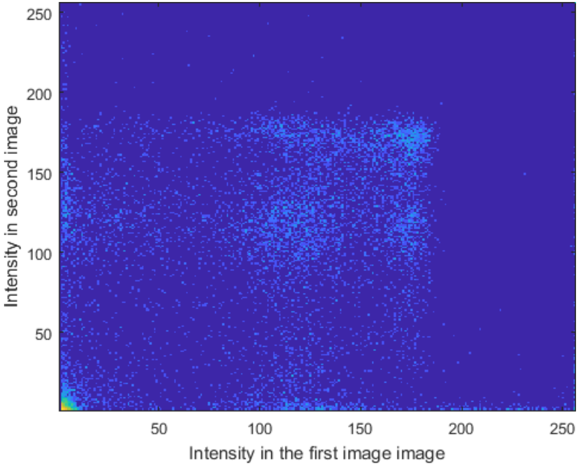
$$p_{X,Y}(x, y) = P(X = x, Y = y)$$

Probability theory

Example: the pixel intensity in two images.

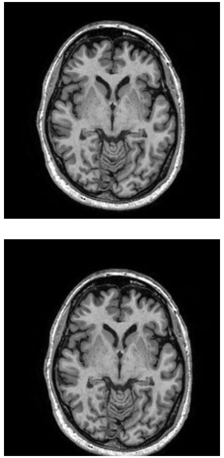
We can estimate this joint p.m.f. from the joint histogram of the two images.

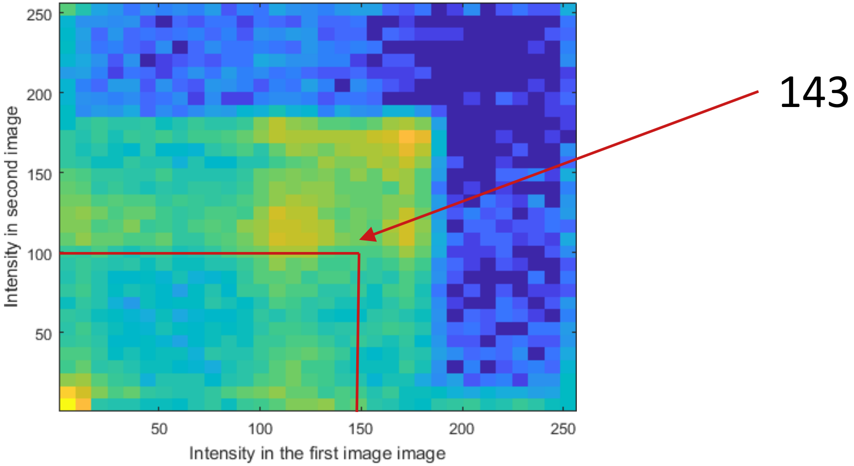




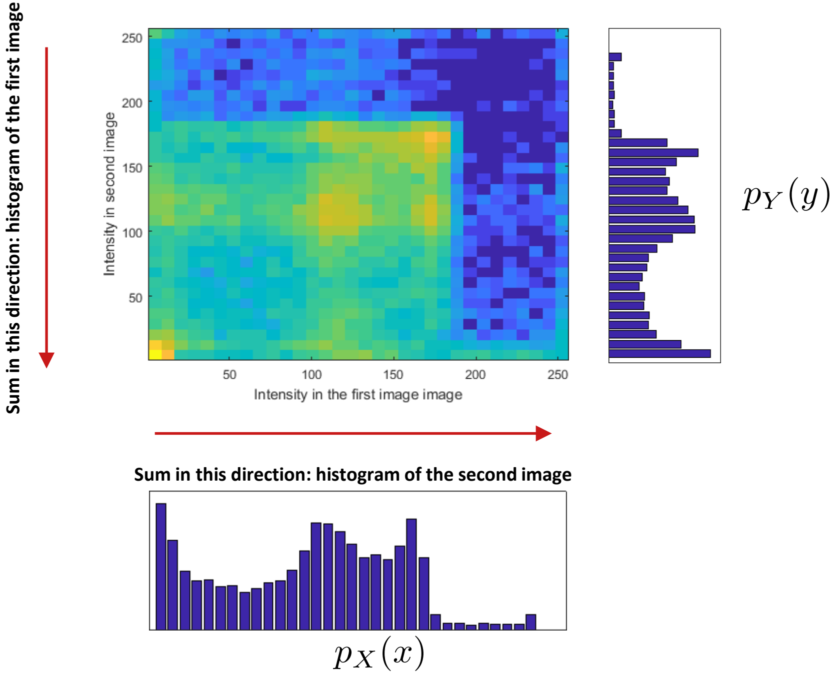
Example: the pixel intensity in two images.

We can estimate this joint p.m.f. from the joint histogram of the two images.





From the joint p.m.f. we can compute the p.m.f.'s of the individual variables (called marginal p.m.f.'s):



Conditional distributions answer the question: what is the probability of distribution over *Y* when we know that *X* takes a certain value?

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$$

$$p_{X,Y}(x,y) = p_{Y|X}(y|x)\, p_X(x)$$

		x				
		0	1	2	3	
y	0	0.84	0.03	0.02	0.01	0.9
	1	0.06	0.01	0.008	0.002	0.08
	2	0.01	0.005	0.004	0.001	0.02
		0.91	0.045	0.032	0.013	1

Example: if we pick a random image location, and the first image has intensity value of 124 at that location, what is the probability distribution for the intensity values in the second image at that location?

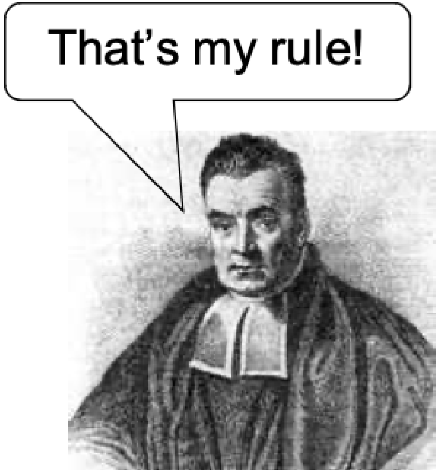
Bayes's rule

$$P(x,y) = P(x|y)P(y) = P(y|x)P(x)$$

$$\Rightarrow \quad p_{Y|X}(y|x) = \frac{p_{Y|X}(x|y)p_Y(y)}{p_X(x)}$$

The random variables *X* and *Y* are independent if:

$$p_{X,Y}(x,y) = p_X(x)p_Y(y)$$



Outline

- Intensity-based similarity metrics
  - Sum of square differences
  - Cross-correlation
  - Mutual information

- Optimization for intensity-based registration
  - Gradient ascent (descent)

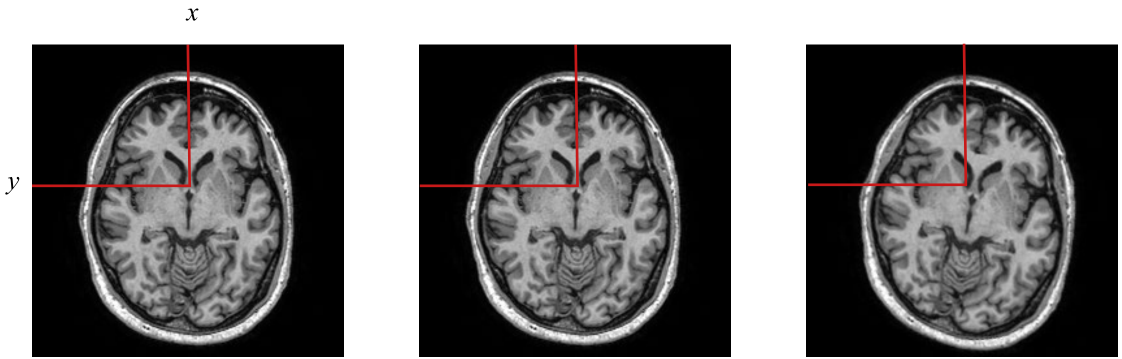
- Evaluation metrics

Mutual information (MI)

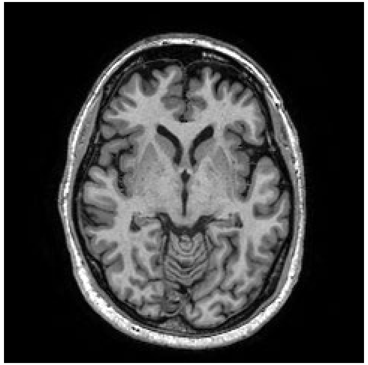
Another measure with even less assumptions: mutual information (MI).

An intuitive interpretation of the MI between two images:

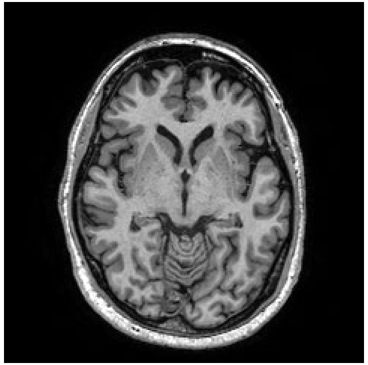
- If we know the pixel intensity value at some location in the fixed image, how much information do we have about the pixel intensity at the same location in the other image?



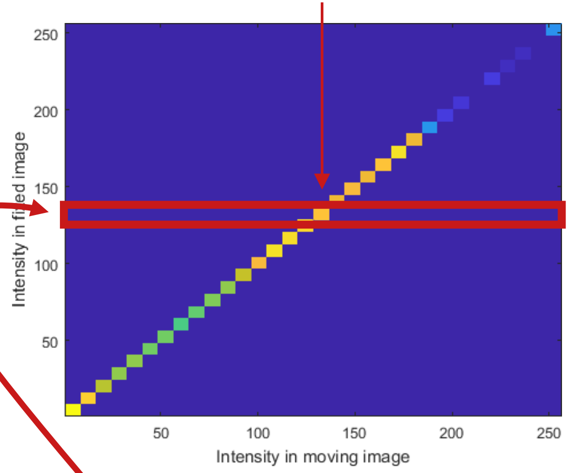
T1



T1

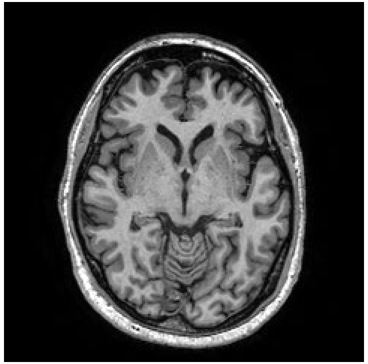


There is only one option for the other value (all other values have zero probability).

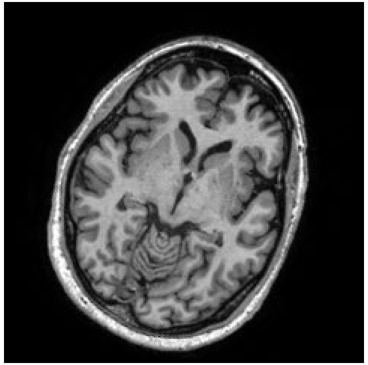


One value is “fixed”.

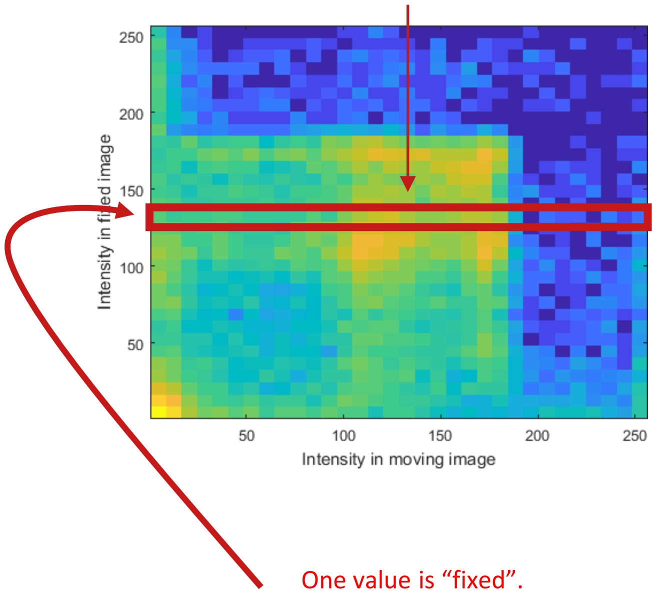
T1



T1

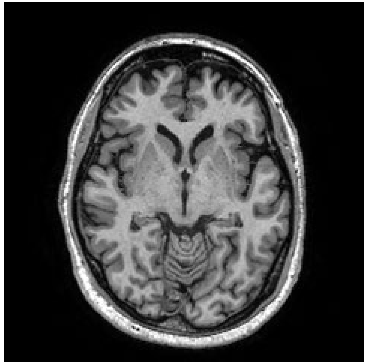


There are many probable values.

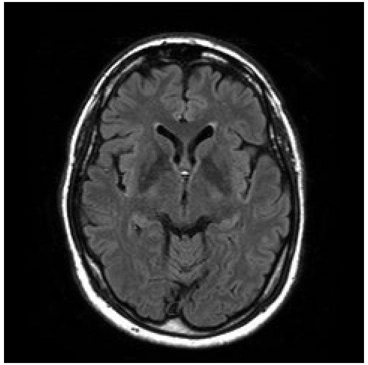


One value is “fixed”.

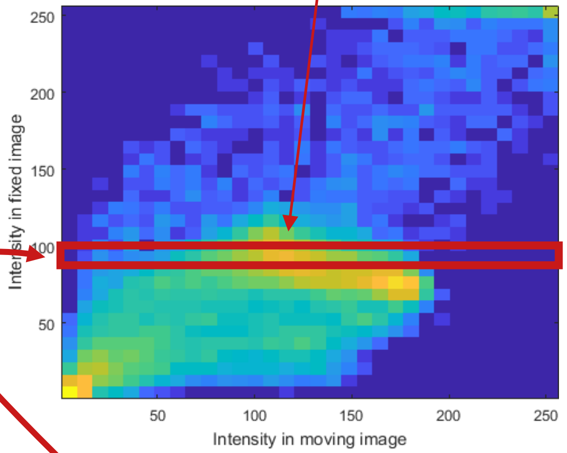
T1



T2

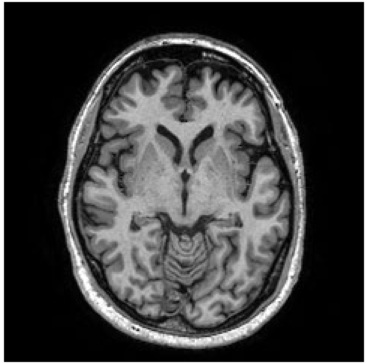


There are a few values with very high probability.

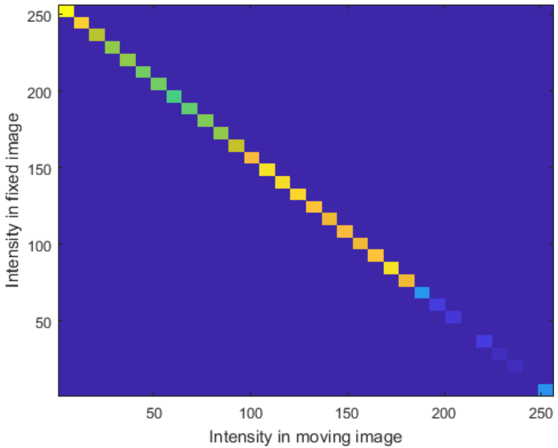


One value is “fixed”.

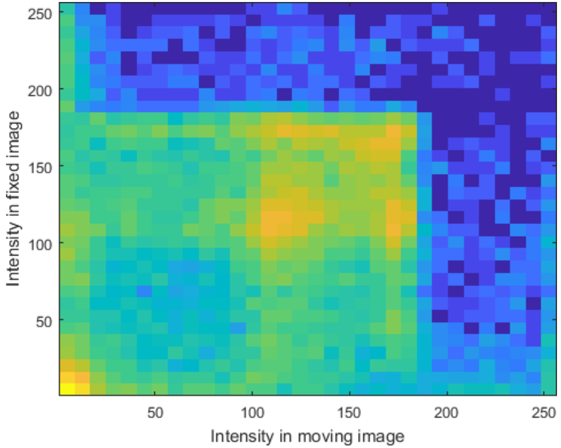
T1

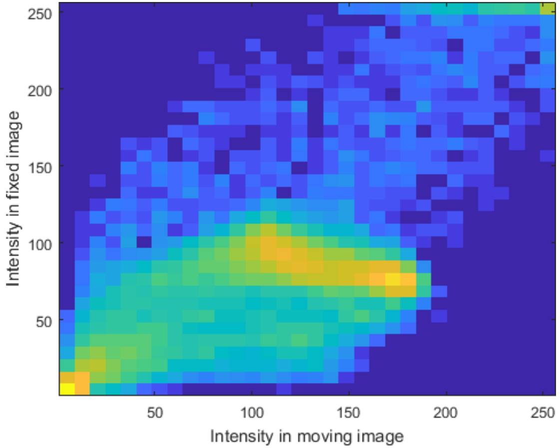


Simulated modality

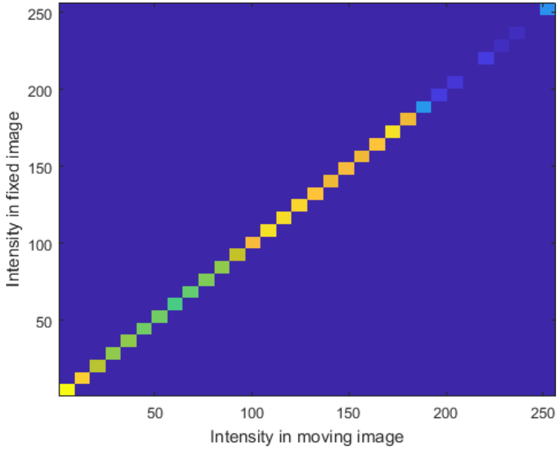


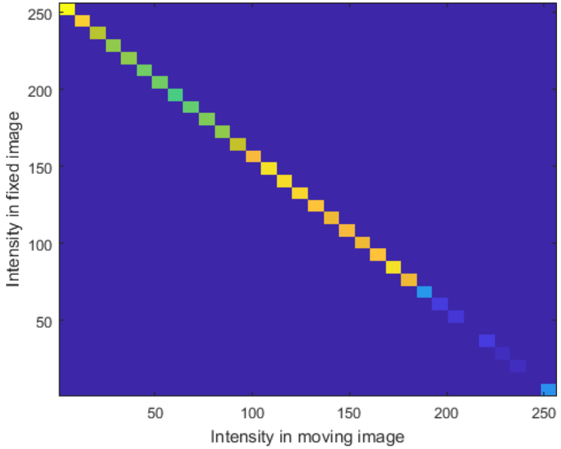
Which pair of images is better aligned according to the joint histogram?
Which histogram is more "compact"?





Which pair of images is better aligned according to the joint histogram?
Which histogram is more "compact"?





Given the joint p.m.f. of two images and the two marginal p.m.f.'s,
the mutual information between the two images can be computed with the following formula:

$$MI(I,J) = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{I,J}(i,j) \log \frac{p_{I,J}(i,j)}{p_I(i)p_J(j)}$$

The unit of MI depends on the particular *log* function:

- when using the natural logarithm the unit is *nats*,
- when using base 2 logarithm the unit is *bits*.

MI in essence is a measure of the “compactness” of the joint p.m.f. of two images.

When the two images are well registered the joint p.m.f. is compact.

When the two images are not well aligned the joint p.m.f. is “spread out”.

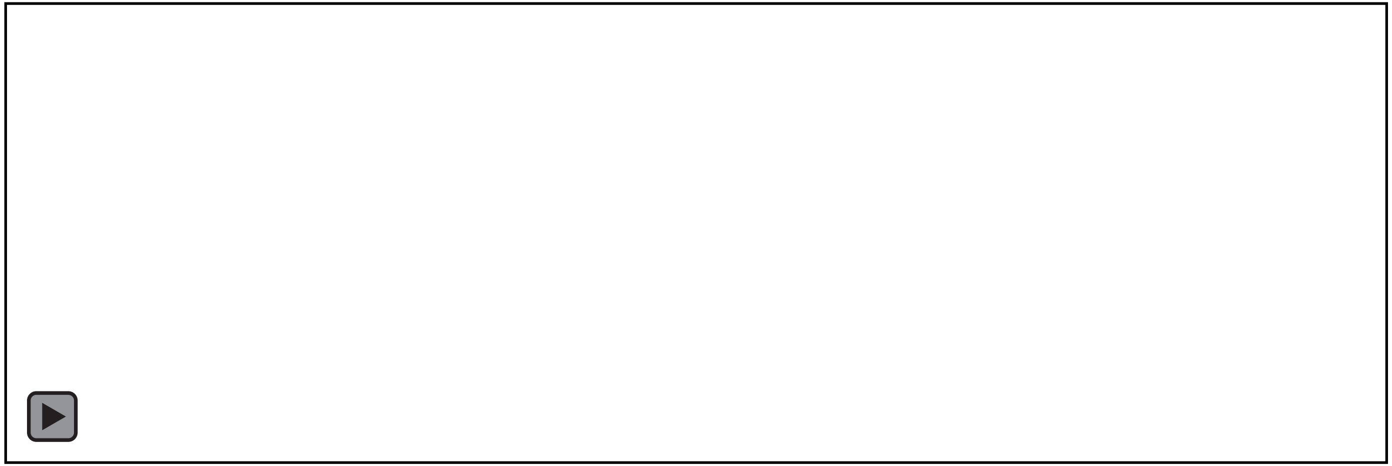
We have now defined several intensity-based similarity measures.

When one of the images is being transformed, the similarity measures are a function of the image transformation.

This is “step 1” in our general approach to registering two images.

“Step 2” is finding the parameters that find the transformation that maximizes the similarity between two images.

Similarity as a function of transformation (T1 to T1):



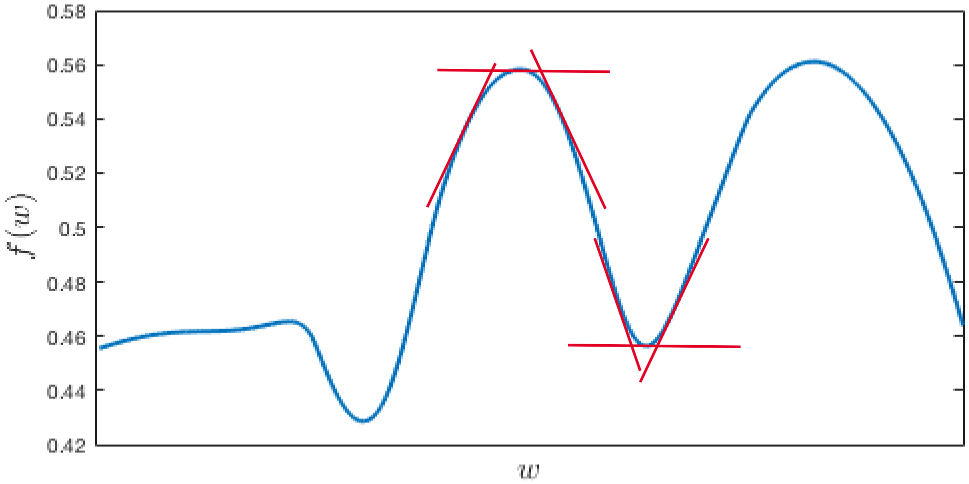
Similarity as a function of transformation (T1 to sim. modality):

Similarity as a function of transformation (T1 to T2):

Outline

- Intensity-based similarity metrics
  - Sum of square differences
  - Cross-correlation
  - Mutual information
- Optimization for intensity-based registration
  - Gradient ascent (descent)
- Evaluation metrics

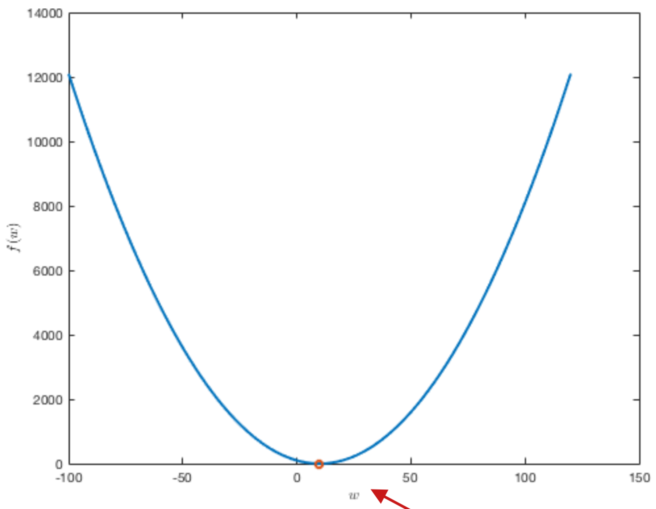
How to find the min. and max. of this function **analytically**?

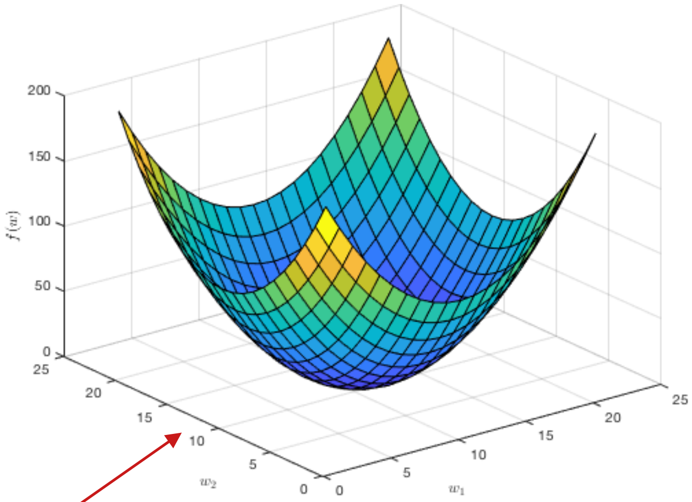


Compute the derivative and set it to zero.

If the function has more than one variable, set the partial derivatives (or gradient vector) to zero.

Convax functions:





Single minimum

For point based affine registration we had:

$$E(\mathbf{T}) = ||\mathbf{T}\mathbf{X}' - \mathbf{X}||_F^2$$

$$\nabla_{\mathbf{T}} E(\mathbf{T}) = 0$$

Find the expression of the gradient and set it to zero. This will result in a system of equations.

$$\mathbf{T} = \mathbf{X}\mathbf{X}'^{\mathsf{T}}(\mathbf{X}'\mathbf{X}'^{\mathsf{T}})^{-1}$$

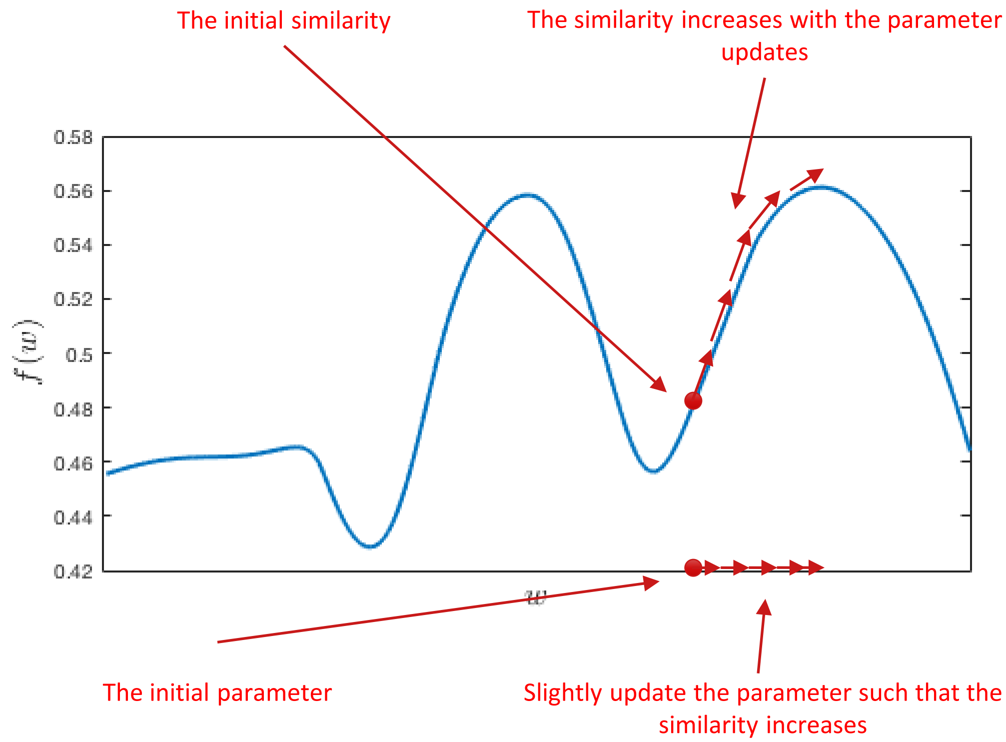
The solution of this system is the optimal value of $\mathbf{T}$.

However, it might be the case that the system of equations produced by setting the gradient to zero is **not solvable**.

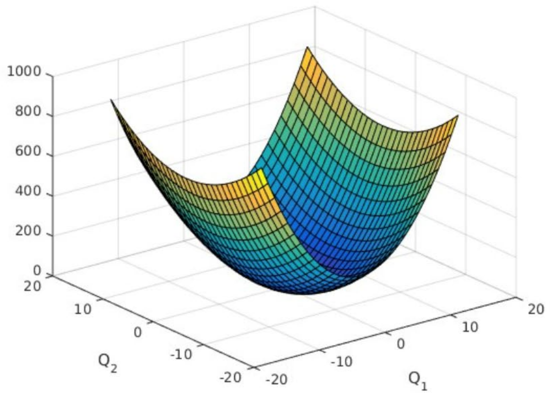
In that case we have to result to numerical methods for finding the minimum of the error (or the maximum of the similarity).

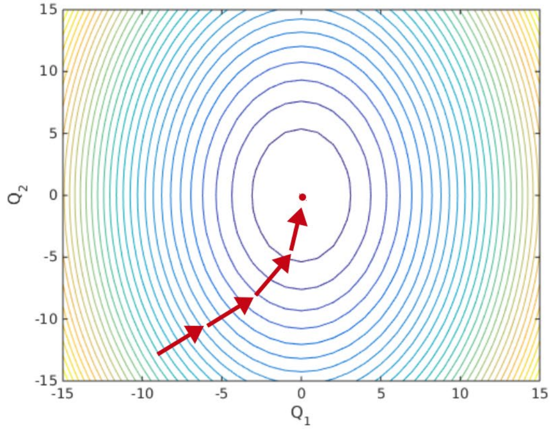
General procedure (for maximization of a similarity function):

1. Start with some initial values for the parameters (in this case the transformation **T**).
2. Slightly update the parameters in such a way that the similarity will slightly increase.
3. Repeat until the similarity stops increasing.

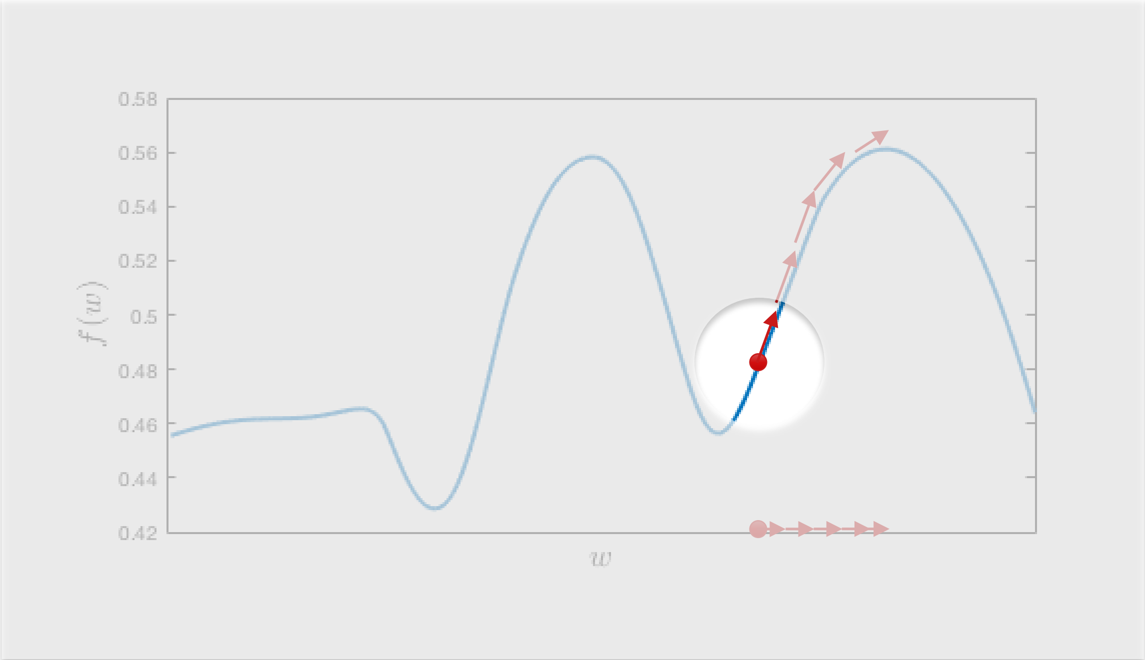

The initial similarity
The similarity increases with the parameter updates
0.58
0.56
0.54
0.52
0.5
0.48
0.46
0.44
0.42
$f(w)$
w
The initial parameter
Slightly update the parameter such that the similarity increases

Minimizing a function with two parameters:



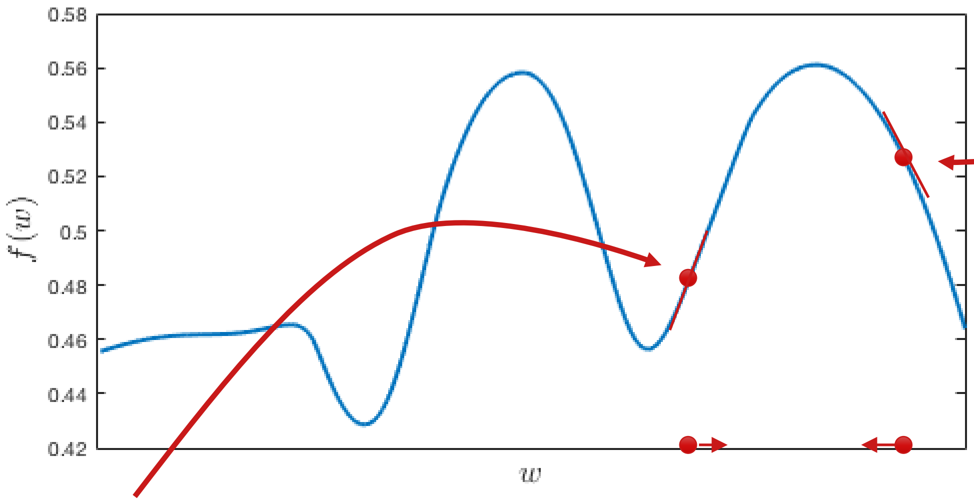


How to find out in which direction to do the parameter update?



How to find out in which direction to do the parameter update?

How to find out in which direction to do the parameter update?



If we are here → positive gradient → increase the parameter (move to the right)

If we are here → negative gradient → decrease the parameter (move to the left)

Gradient ascent algorithm for maximizing a function $f(\mathbf{w})$:

1. Choose some initial values of the parameters **w**
2. Calculate the value for the gradient of $f(\mathbf{w})$ for the current parameters
3. Update the parameters in the direction of the gradient:

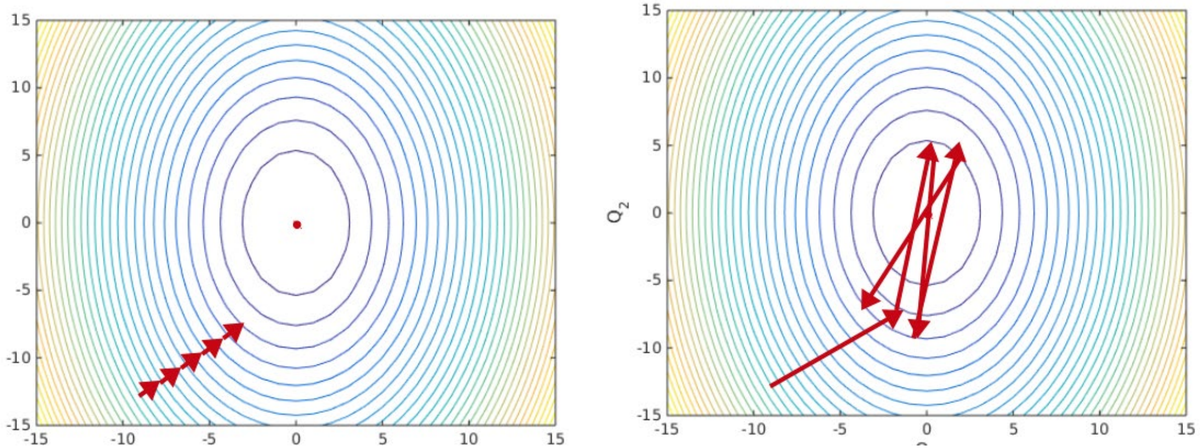
$$\mathbf{w} \leftarrow \mathbf{w} + \mu \nabla_{\mathbf{w}} f(\mathbf{w})$$

If we want to minimize the function, we move in the direction opposite of the gradient (gradient descent):

$$\mathbf{w} \leftarrow \mathbf{w} - \mu \nabla_{\mathbf{w}} f(\mathbf{w})$$

The parameter μ is called **learning rate**. It controls how fast we move towards the maximum (minimum).

- If μ is **too small**, the maximum (or minimum) might not be reached in reasonable time.
- If μ is **too large**, the maximum (minimum) might be missed.



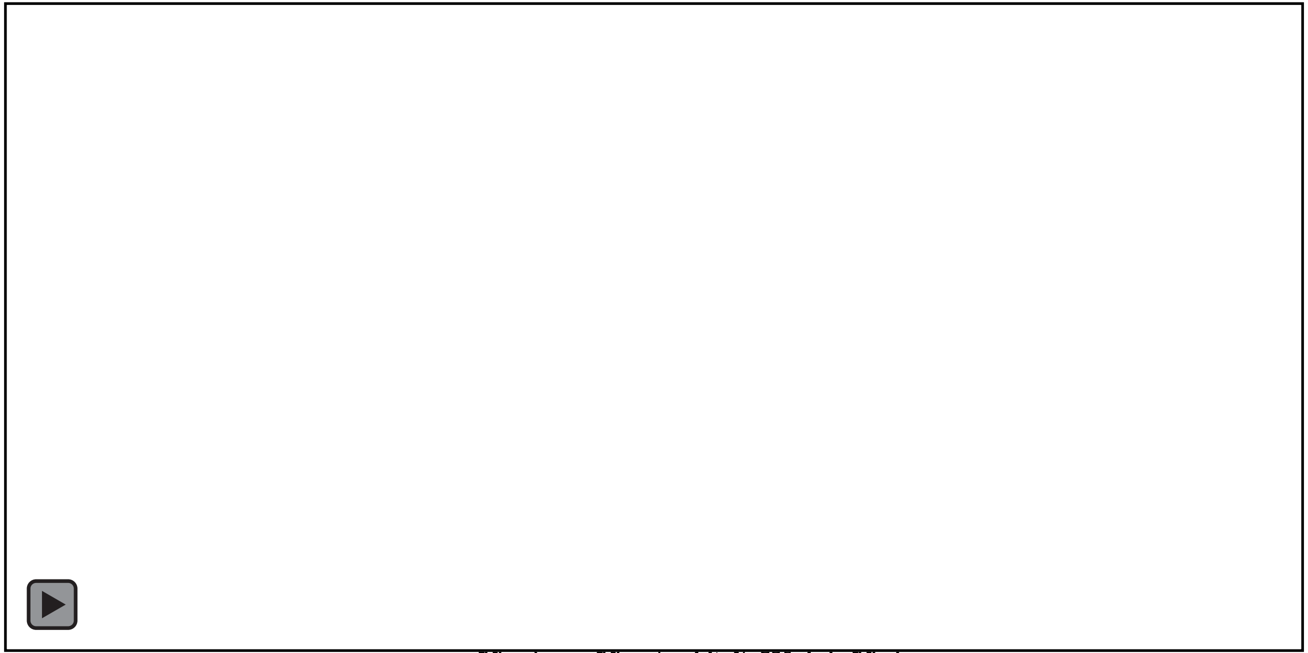
Example – intensity-based image registration

parameters of affine transformation

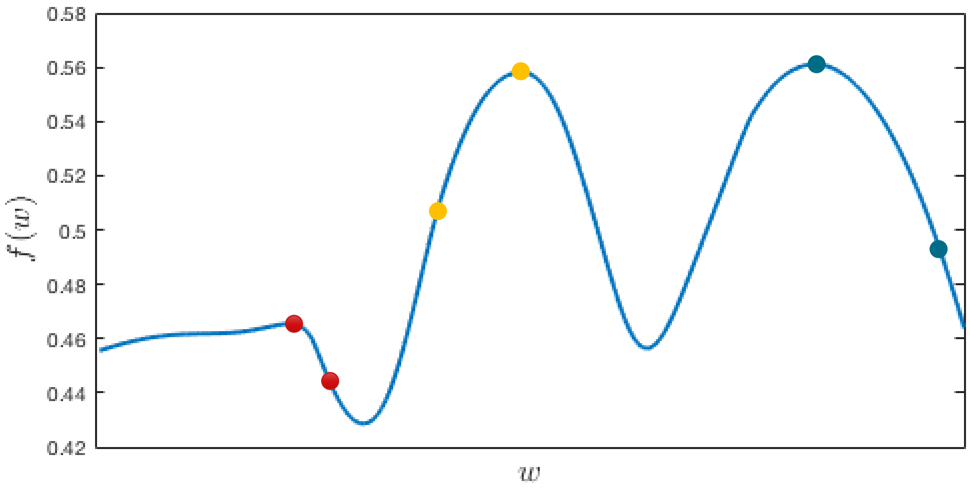
milarity:
SSD, CC or MI

Example – intensity-based image registration

Example – intensity-based image registration



Initialization is important.
Different starting points will result in different found maxima (and not always global).



Outline

- Intensity-based similarity metrics
  - Sum of square differences
  - Cross-correlation
  - Mutual information

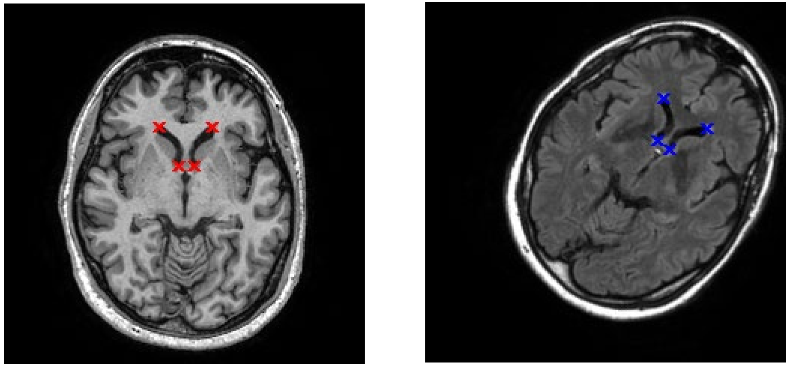
- Optimization for intensity-based registration
  - Gradient ascent (descent)

- Evaluation metrics

Evaluation of registration accuracy

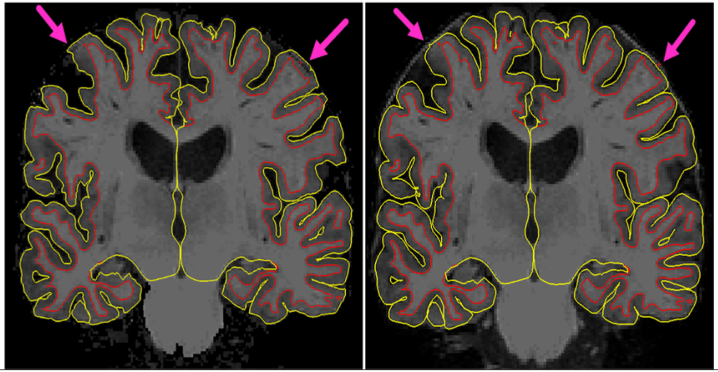
Quantitative evaluation metrics

- Landmark-based metrics
- Segmentation-based metrics
  - Dice similarity coefficient (DSC)
  - Jaccard index (Intersection over union)
  - Hausdorff distance and mean surface distance



Fixed Moving

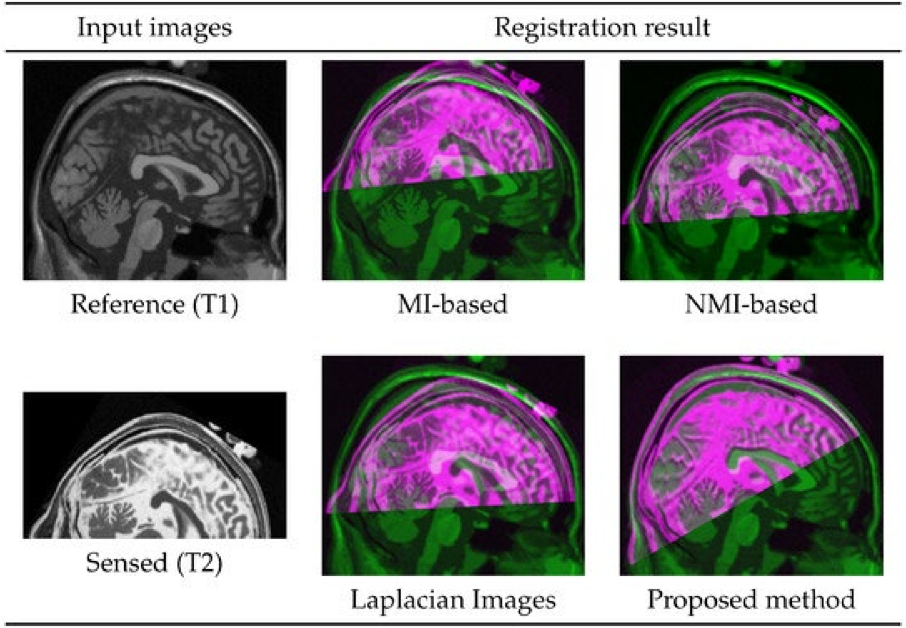
A. T1-Only Segmentation B. T1+T2-FLAIR Segmentation



Evaluation of registration accuracy

Qualitative evaluation metrics

- Overlay or fusion visualization
- Difference imaging
- Clinical evaluation/survey by experts, e.g., Likert scale.



Thank you

r.su@tue.nl

Next: Introduction to image segmentation