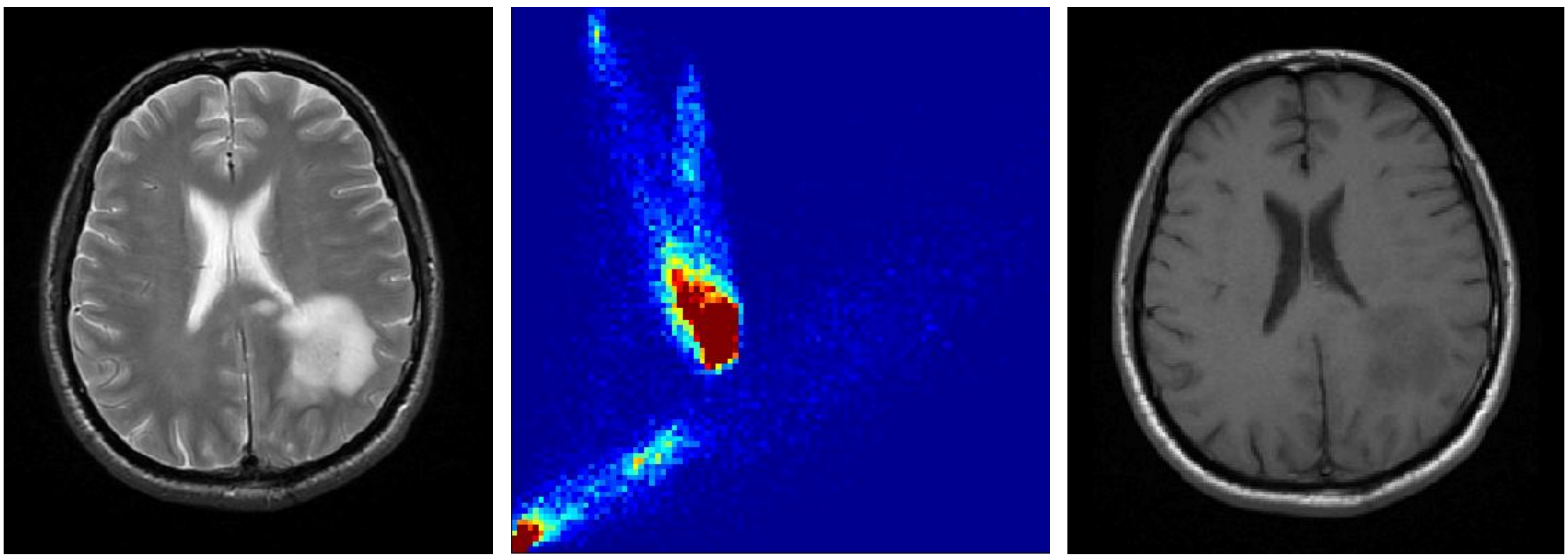
Multi-modality image registration

Mutual information



Now:

Multi-modality image registration

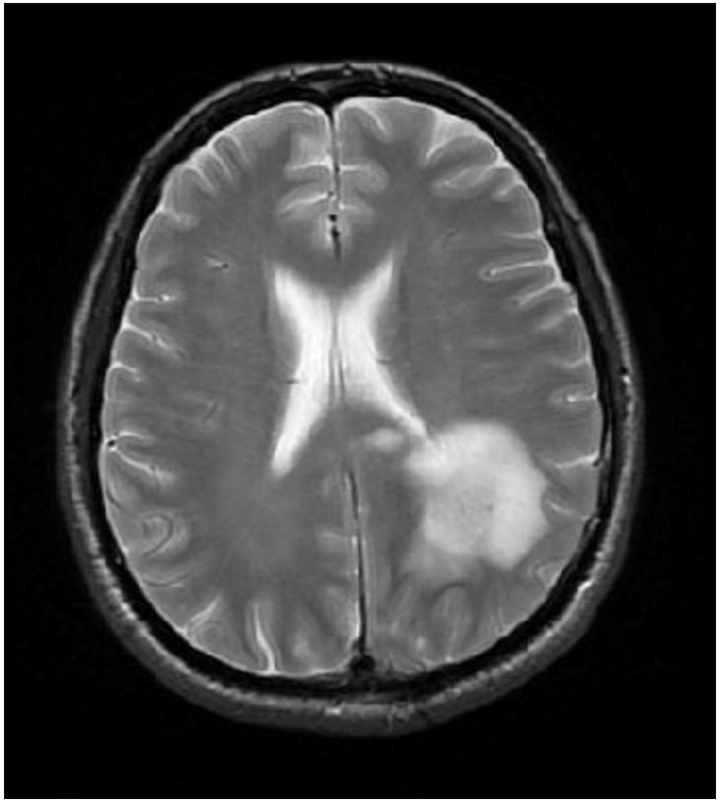
So far: Minimize the difference in pixel intensities

Problem: only works if both images are of the same type

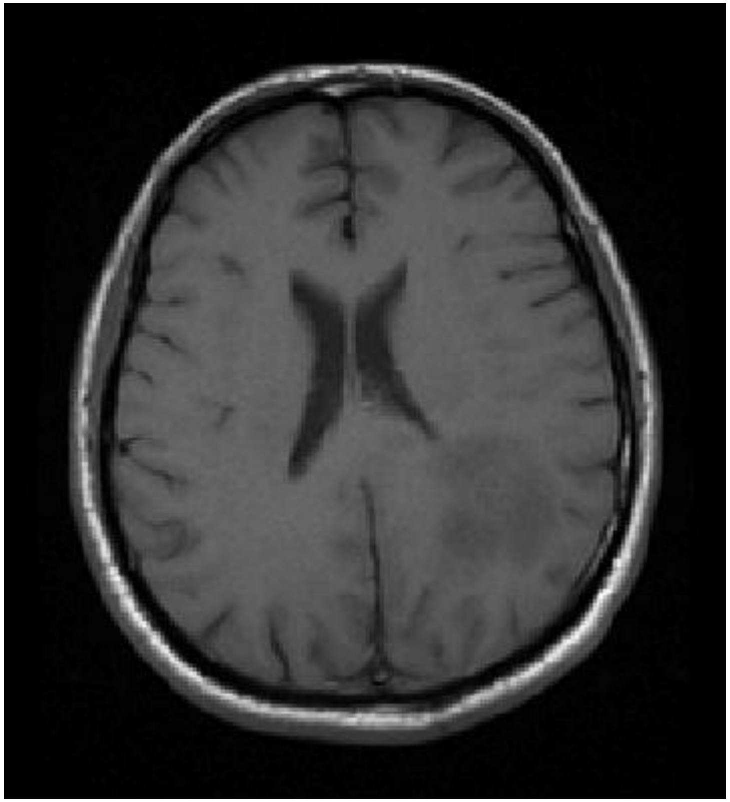
(CT is ideal – Hounsfield number is quantitative / scanner independent)

➔ Requires different cost function

Example: T1 / T2 MR images

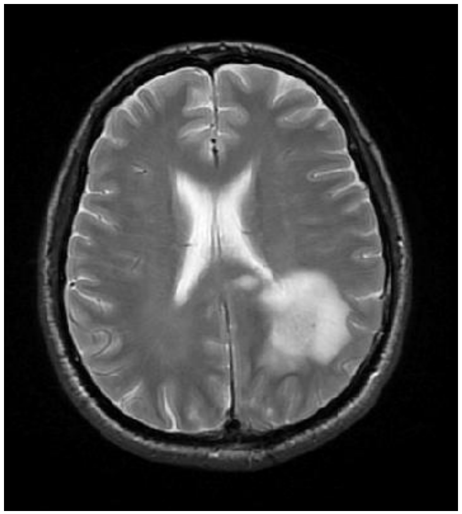


T2 image



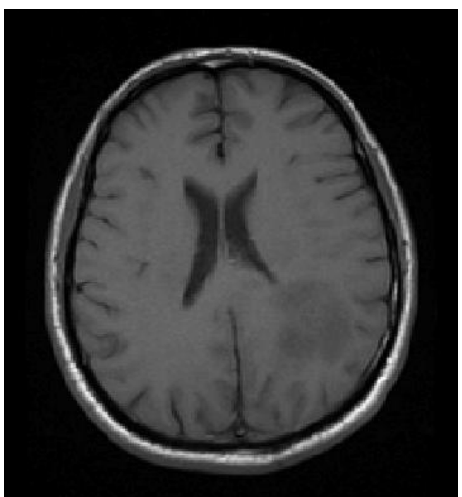
T1 image

Example: T1 / T2 MR images



Intuition:

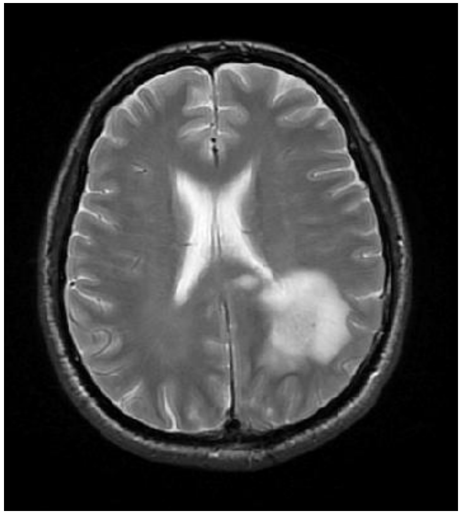
Regions with constant gray value in the image should be mapped to constant gray values in the reference image

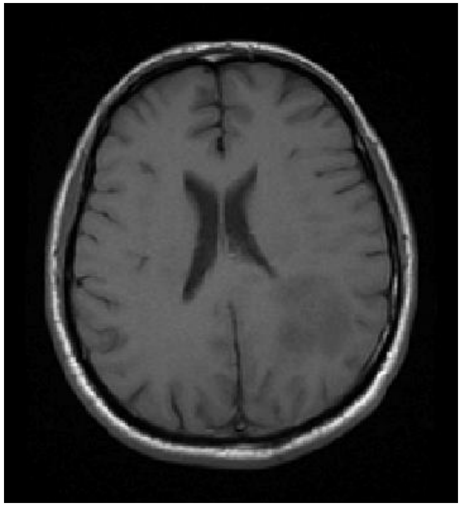


Common approach:

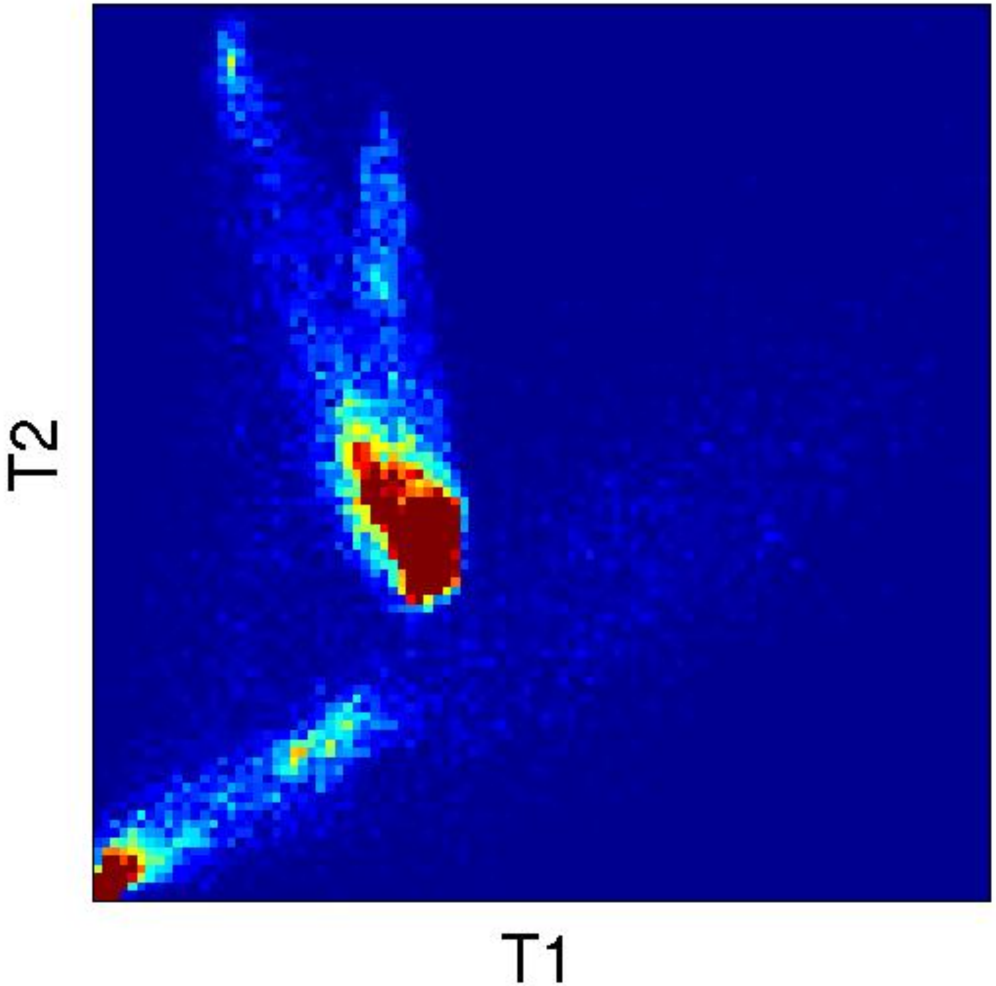
Mutual information cost function

Gray value correlation

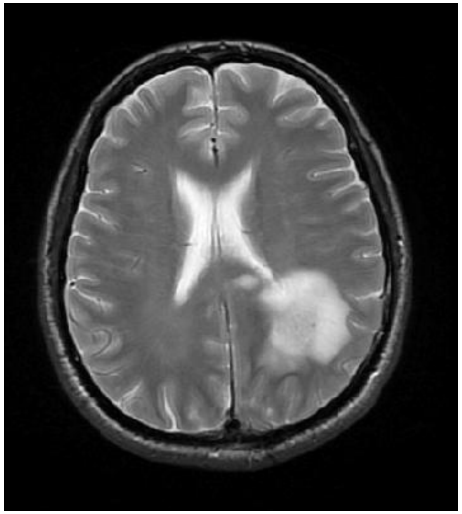


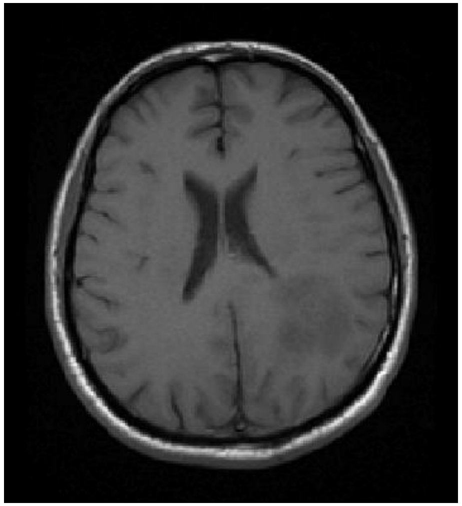


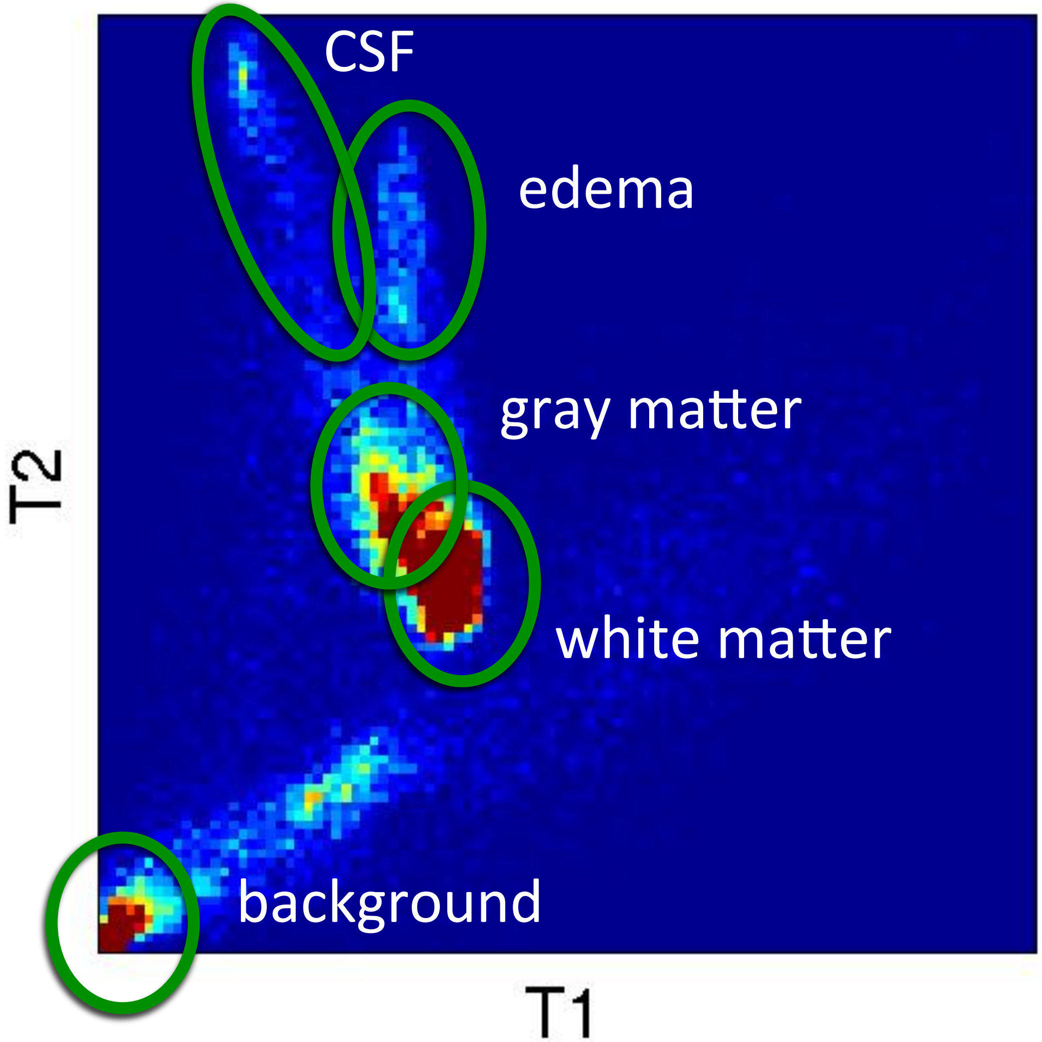
Joint gray value histogram of T1 and T2 image



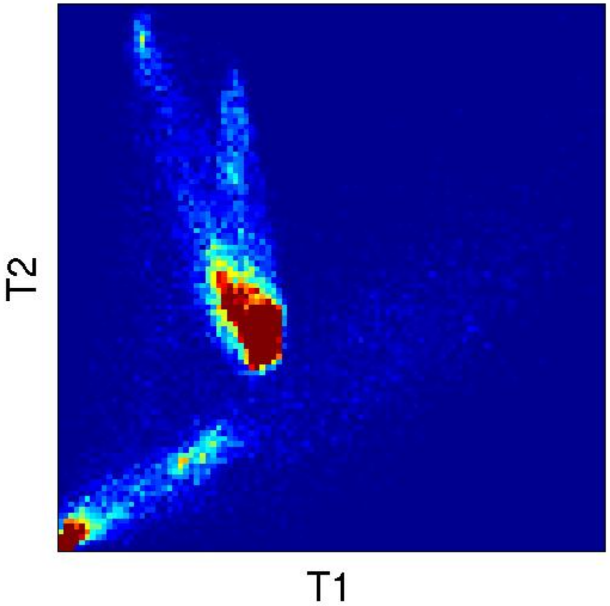
Gray value correlation







Gray value correlation



Idea:

If images are correctly registered, one can predict the gray value in the image by knowing the gray value in the reference image

This is quantified by the

Mutual information

between the two gray value histograms

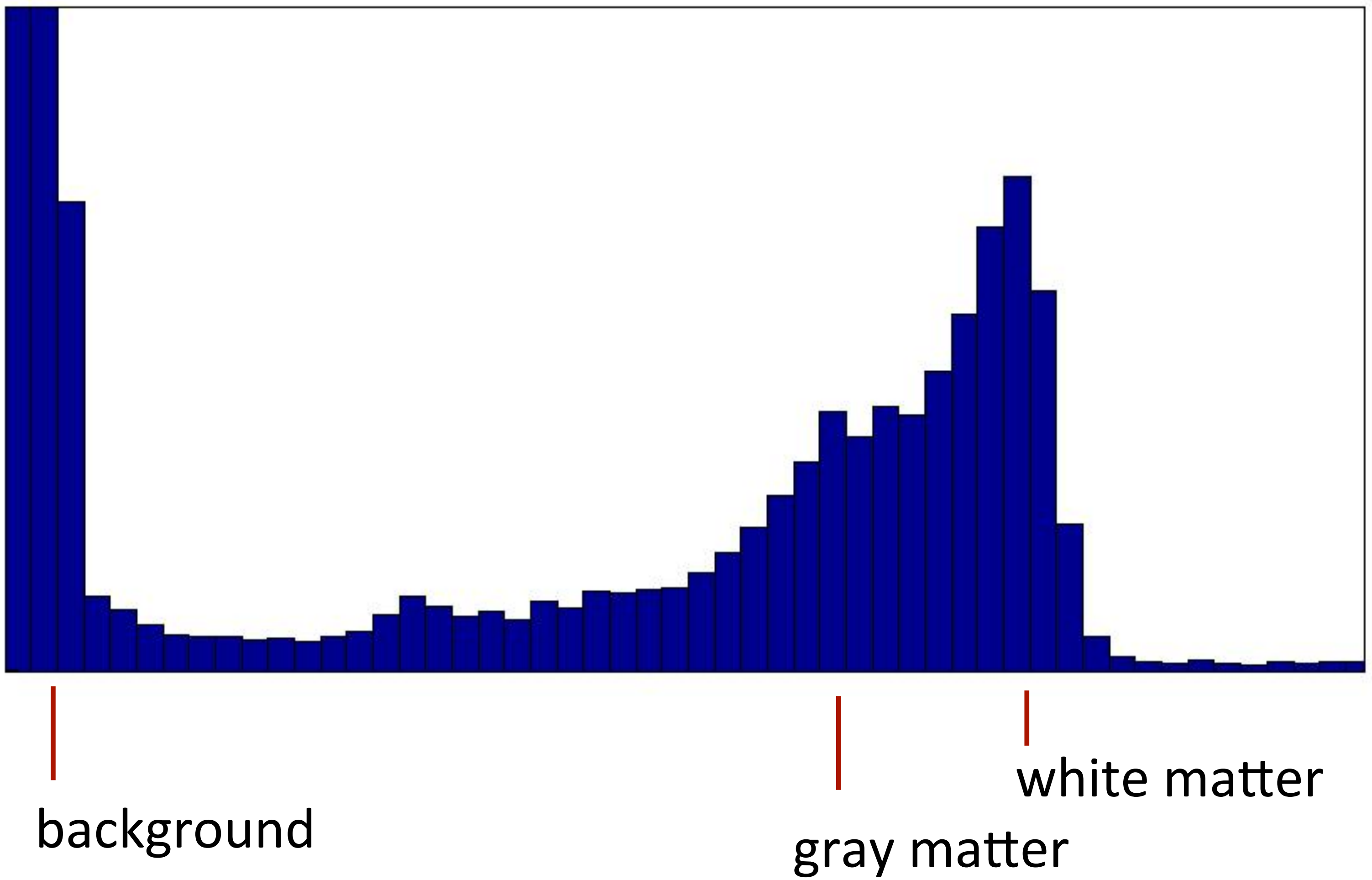
Definition of Entropy (= Shannon Information)

$$H(A) = -\sum_{n} P(a_n)\log P(a_n)$$

random variable A with realizations a_n

Here: "A" corresponds to gray value of the image

Histogram of T1 image



Entropy

Definition of Entropy

$$H(A) = -\sum_n P(a_n) \log P(a_n)$$

random variable A with realizations a_n

Here: "A" corresponds to gray value of in image voxel

Entropy = average information contained in one gray value (in "bit" if $\log_2$ is used)

- Uniform image (peaked histogram) contains no information $H(A) = 0$
- Uniform histogram has maximum entropy

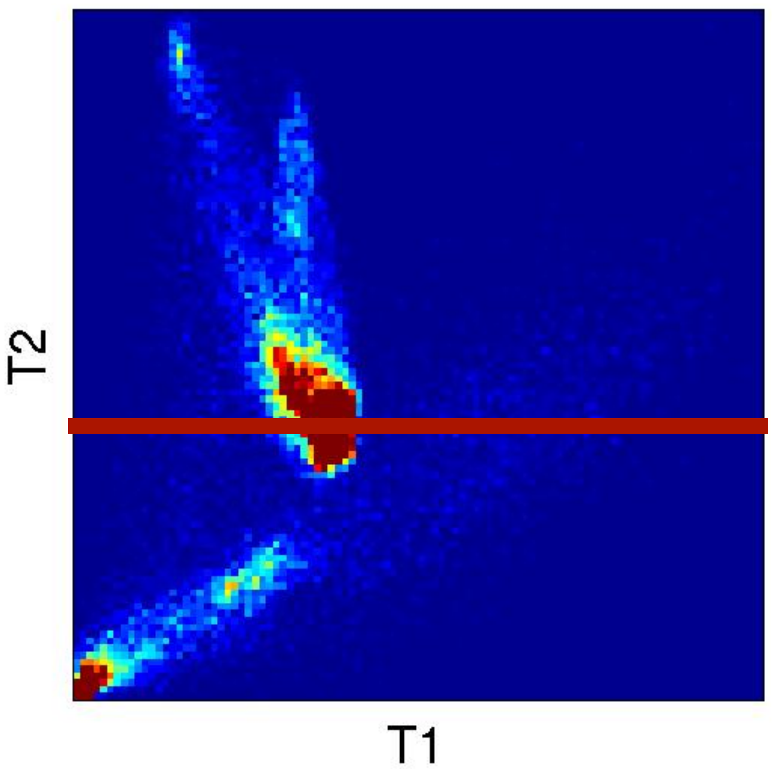
Conditional Entropy

conditional probability

$$H(A \mid B) = \sum_m P(b_m) \underbrace{\left[-\sum_n P(a_n \mid b_m) \log P(a_n \mid b_m) \right]}_{H(A \mid B = b_m)}$$

average information contained in a gray value in image A given that the gray value in image B is b_m

Gray value correlation



Conditional Entropy:

$$H(A \mid B) \leq H(A)$$

knowing the gray value in image B reduces the average information contained in image A

(histogram becomes more peaked)

$$H(A \mid B) = H(A)$$

if and only if A and B are independent random variables

Mutual information

$$MI(A,B) = H(A) - H(A \mid B)$$

Mutual information of both histograms

"how much information does the reference image contain over the image"

Conditional entropy of image given the reference image

"how undetermined is the gray value of the image after knowing the reference image"

Entropy of image histogram A

"how undetermined is the gray value of an image pixel"

Mutual information

$$MI(A,B) = H(A) - H(A \mid B)$$

Mutual information of both histograms

Conditional entropy of image given the reference image

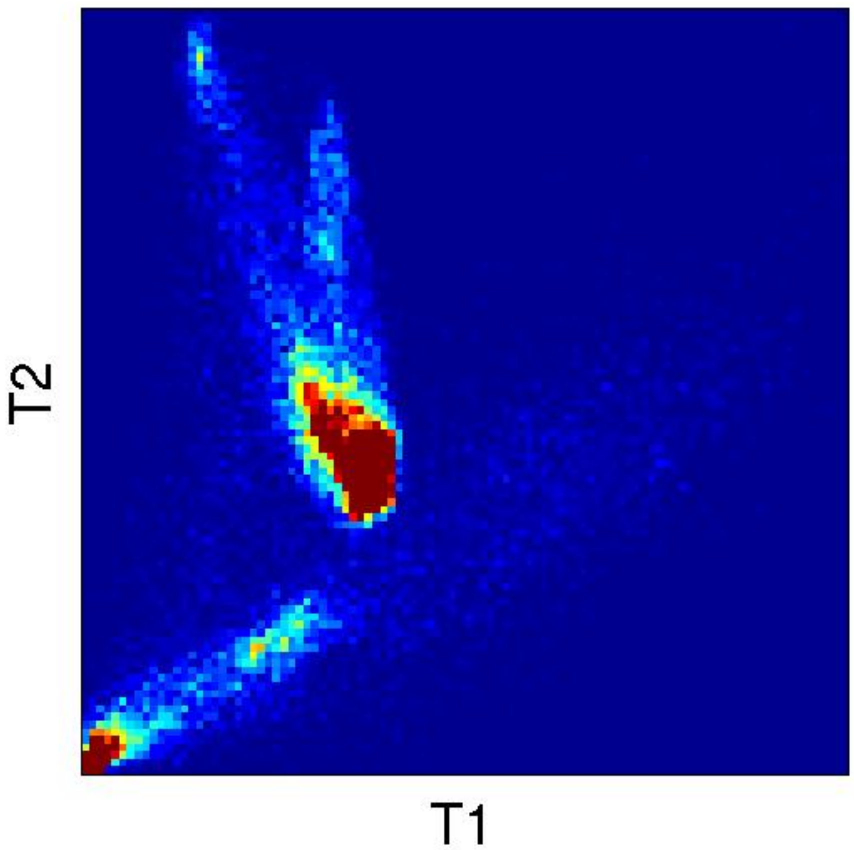
if histogram A|B is peaked, then H(A|B) ≈ 0

➔ *MI*(A,B) ≈ H(A) is maximized

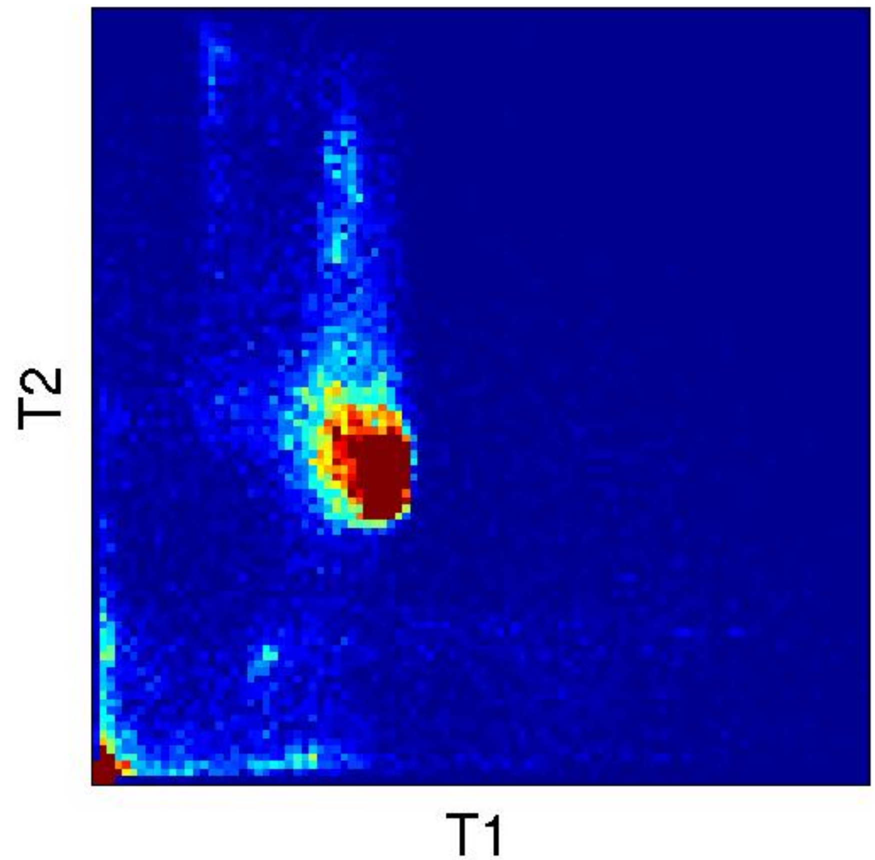
➔ Image B contains as much information as possible about A

Image misalignment

Gray value correlation is reduced if images are not registered



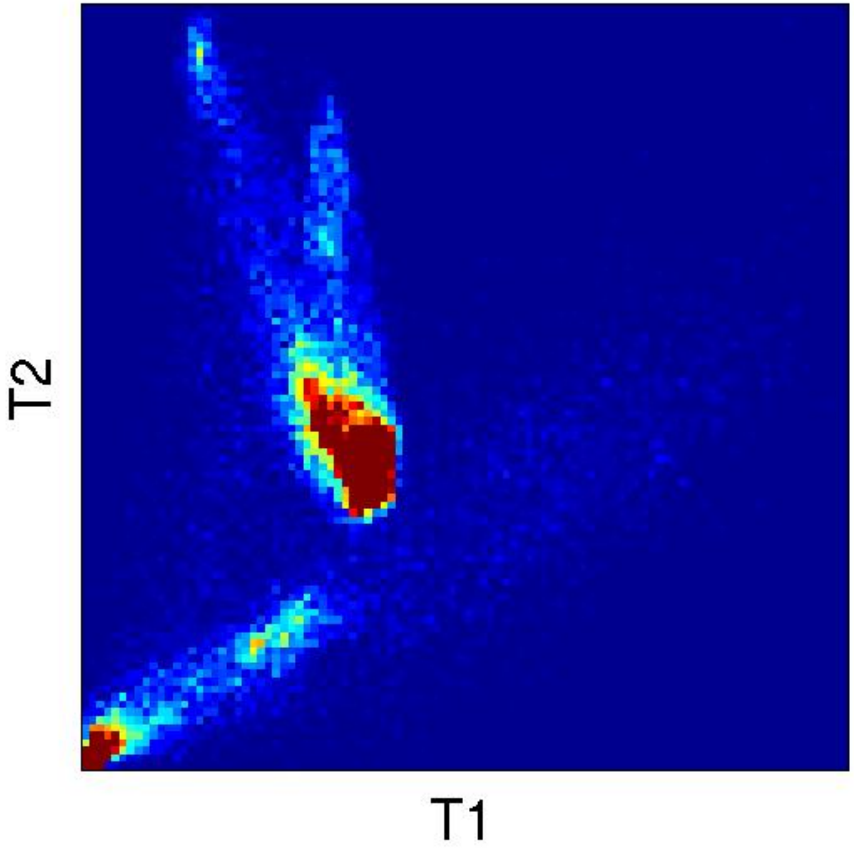
(registered)



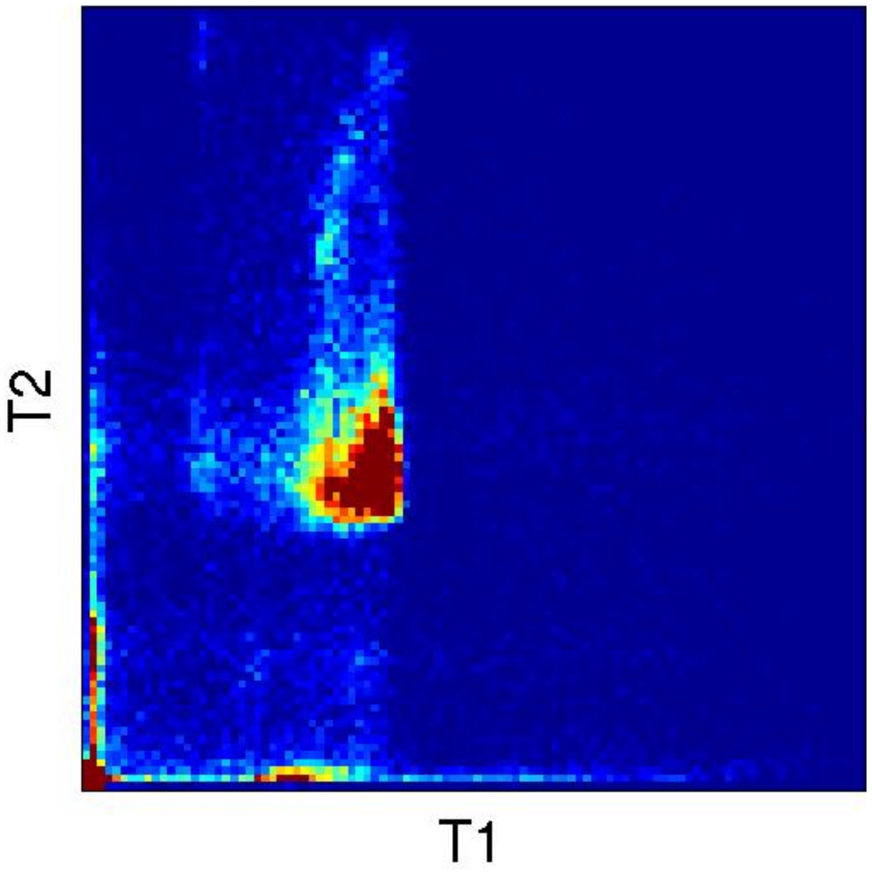
(shifted by 3 pixel (1.5 mm))

Image misalignment

Gray value correlation is reduced if images are not registered



(registered)



(shifted by 10 pixel (5 mm))

Registration using MI - interpretation

Registration: Maximize mutual information

$$MI(A,B) = H(A) - H(A \mid B)$$

We are minimizing the entropy of the joint histogram

We are minimizing the conditional entropy of the transformed image given the reference image

let's assume that's independent of the image transformation

Mutual information

Express MI through joint histogram

$$MI(A,B) = H(A) - H(A \mid B)$$

$$= H(B) - H(B \mid A)$$

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

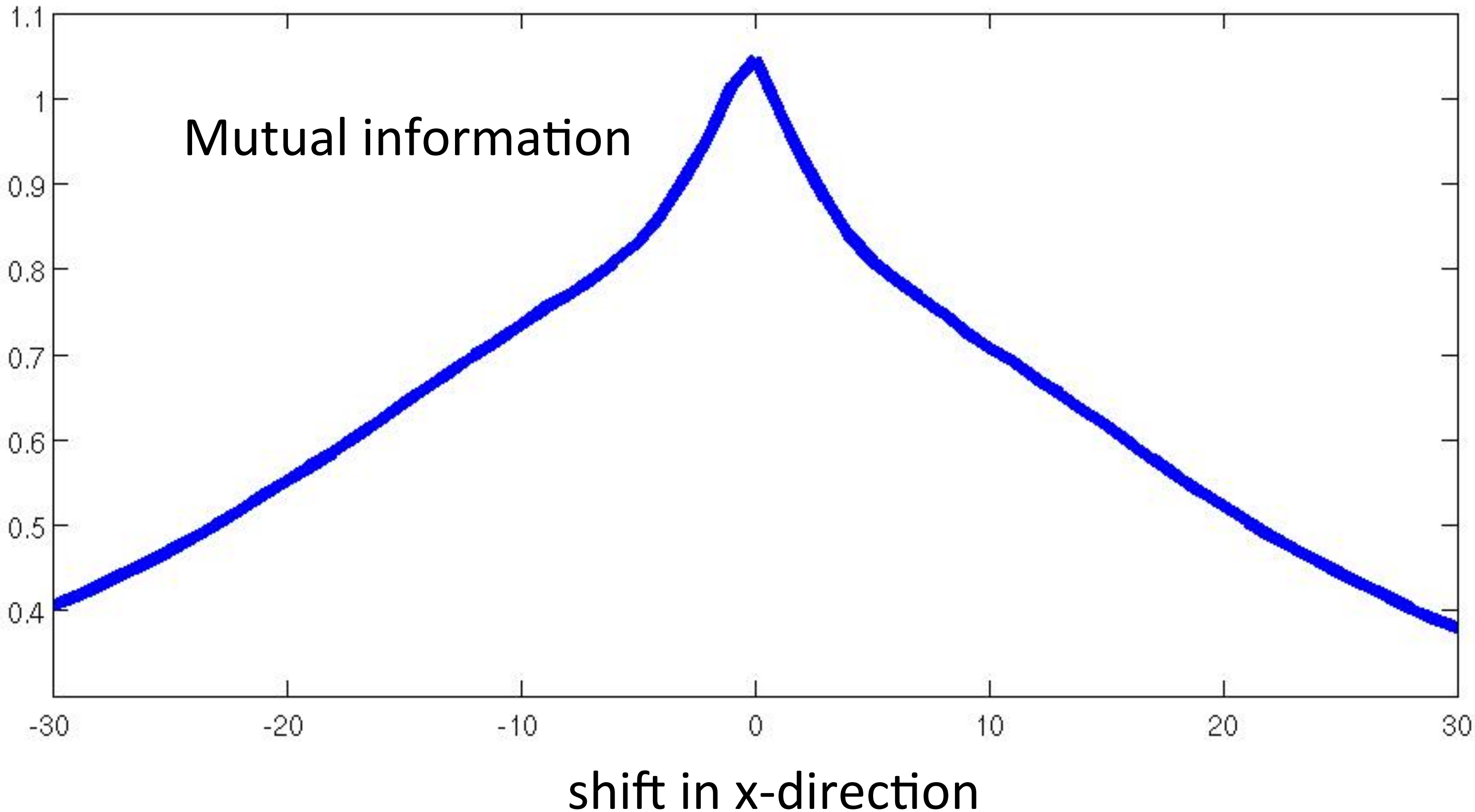
joint gray value histogram

$$MI(A,B) = \sum_{n,m} P(a_n, b_m) \log\left(\frac{P(a_n, b_m)}{P(a_n)P(b_m)}\right)$$

individual histograms

Calculating gradients

Mutual information (T1-T2) as a function of misalignment:



Maximizing the Mutual information cost function

$$MI\left(p_{mn}\left(I_R, I_T(\Theta)\right)\right) = \sum_{m,n} p_{mn} \log\left(\frac{p_{mn}}{p_m p_n}\right)$$

Mutual information depends on registration parameters through the transformed image and the joint histogram

Calculating gradients

Two options for gradient calculation:

1. **approximate gradients through finite differences**

2. **calculate derivatives via the chain rule**

Optimizing Mutual Information

Gradient calculation:

$$\frac{\partial MI}{\partial \Theta_k} = \sum_{n,m} \frac{\partial MI}{\partial p_{nm}} \frac{\partial p_{nm}}{\partial \Theta_k}$$

Derivative of MI with respect to histogram bin (n,m)
(can be calculated analytically)

How does the histogram depend on registration parameters

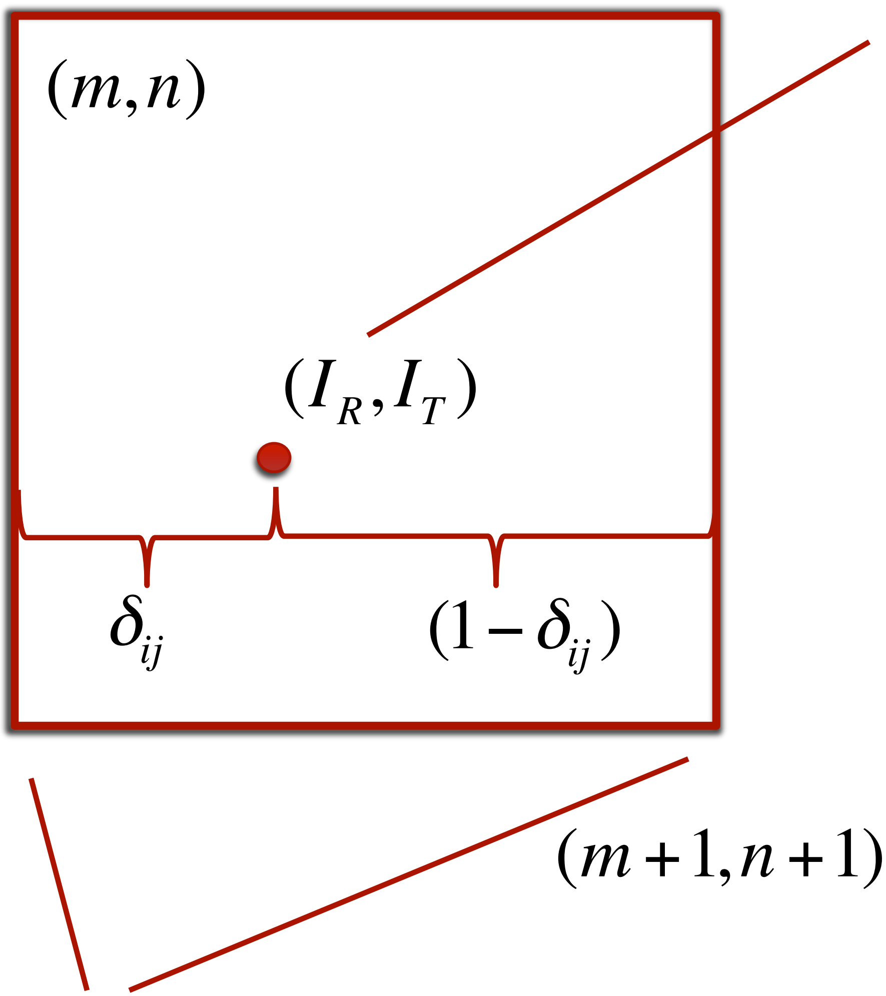
Optimizing Mutual Information

$$\frac{\partial p_{nm}}{\partial \Theta_k} = \sum_{i,j} \Bigg[\underbrace{\frac{\partial p_{nm}}{\partial I}\bigg|_{\left(x(i,j;\Theta),y(i,j;\Theta)\right)}}_{} \; \underbrace{\frac{\partial I\left(x(i,j;\Theta),y(i,j;\Theta)\right)}{\partial \Theta_k}}_{\text{see first lecture}} \Bigg]$$

change of the joint histogram when changing the gray value in the transformed image (at the point that corresponds to voxel (i,j))

sum over the contributions of all voxels to the joint histogram

Gradient calculation



gray value pair of voxel (i,j)

4 adjacent histogram bins

- Voxel (i,j) contributes to 4 histogram bins

  (weights determined by bilinear interpolation)

- changing the gray value changes the weightings of the 4 bins

Summary

Multi-modality registration requires different cost function

Maximizing mutual information is popular choice

Advantage: no prior assumptions on gray value correlations needed

Assumption:

When images are registered, the gray value in the reference image predicts the gray value in the image best

Gradient calculation:

a bit messy