

MICCAI 2010 Tutorial

Intensity-based Deformable Registration

Similarity Measures

Christian Wachinger

Computer Aided Medical Procedures (CAMP), TU München, Germany

Quantify the similarity between images

- Do images contain the same object?

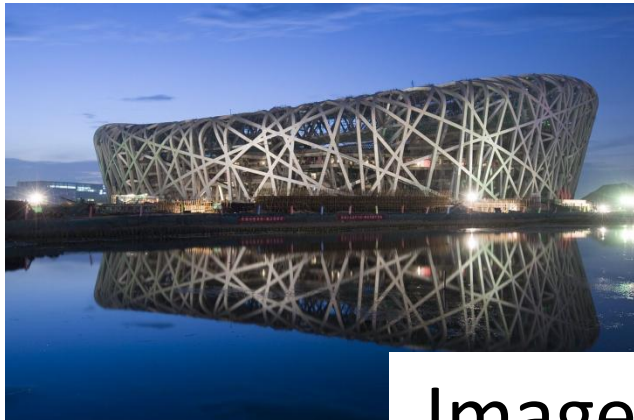


Image retrieval



Quantify the similarity between images

- How are images correctly aligned?

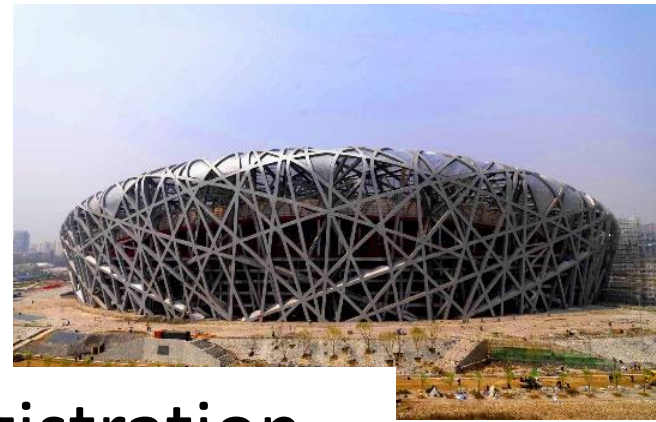
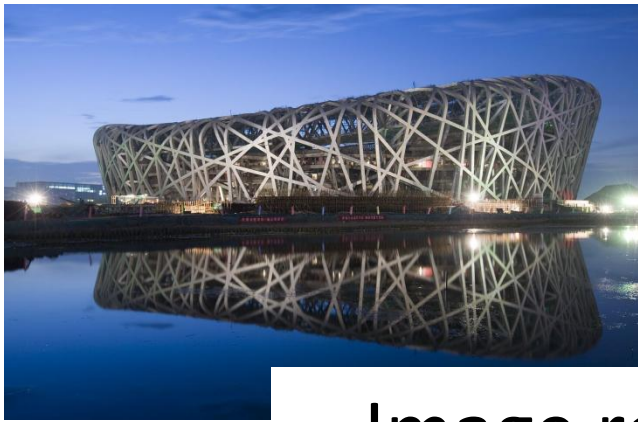
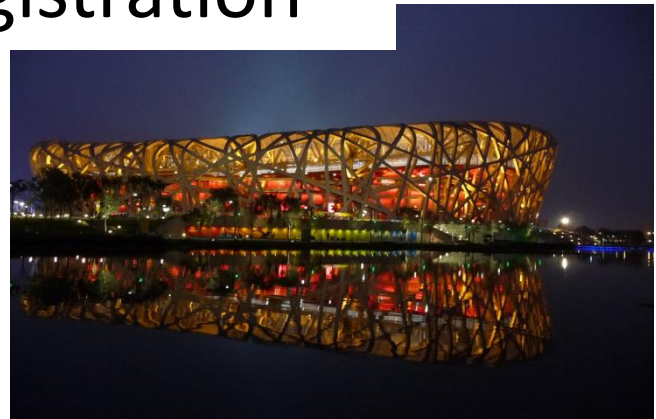
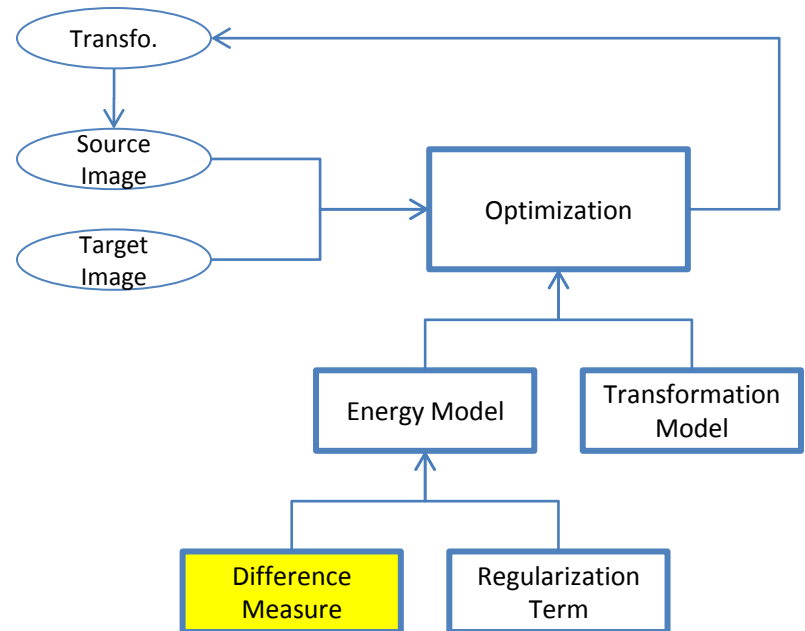


Image registration



Requirements on similarity measure

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



Overview

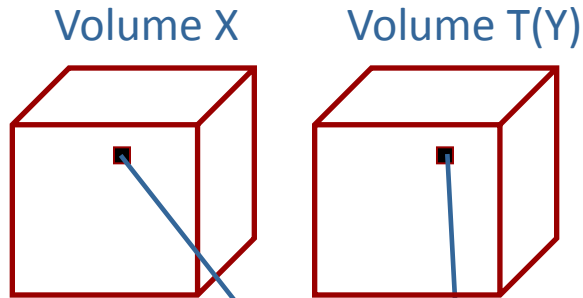
1. Standard similarity measures
2. Probabilistic framework
3. Pre-processing steps
4. Recent approaches
5. Linear vs. non-linear registration

PART I

Standard Similarity Measures

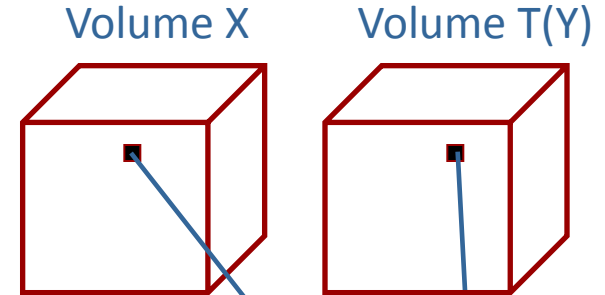
- SSD-SAD
- Cross Correlation
- Mutual information
- Derivatives

Difference Measures



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

Sum of Squared Differences

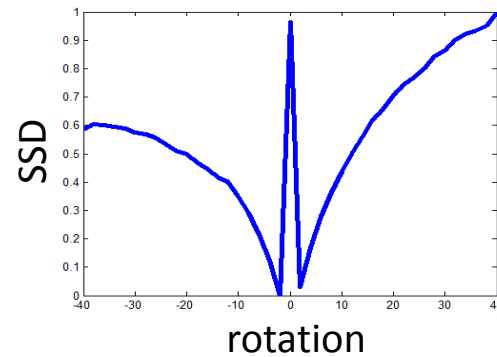


$$SAD = \frac{1}{N} \sum_i |x_i - y_i|$$

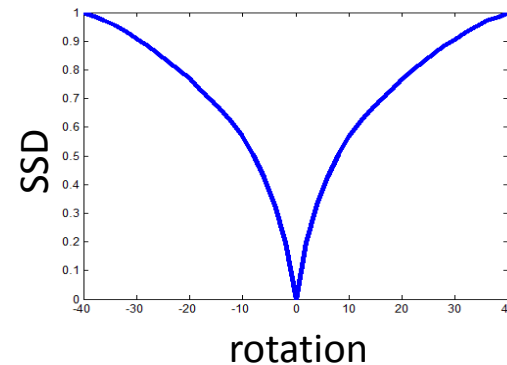
Sum of Absolute Differences:
Less sensitive on large intensity
differences than SSD

Limitations of SSD

- Illumination change affects similarity function



- Idea: normalization of images $\tilde{X} = \frac{X - \mathbb{E}[X]}{\sigma(X)}$

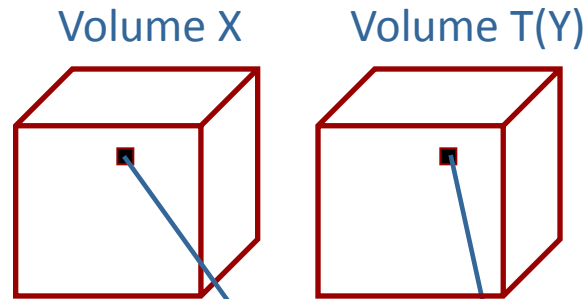


SSD on Normalized Images

$$\mathbb{E}[(\tilde{X} - \tilde{Y})^2]$$

$$\tilde{X} = \frac{X - \mathbb{E}[X]}{\sigma(X)}$$
$$\tilde{Y} = \frac{Y - \mathbb{E}[Y]}{\sigma(Y)}$$

Normalized Cross Correlation (NCC)



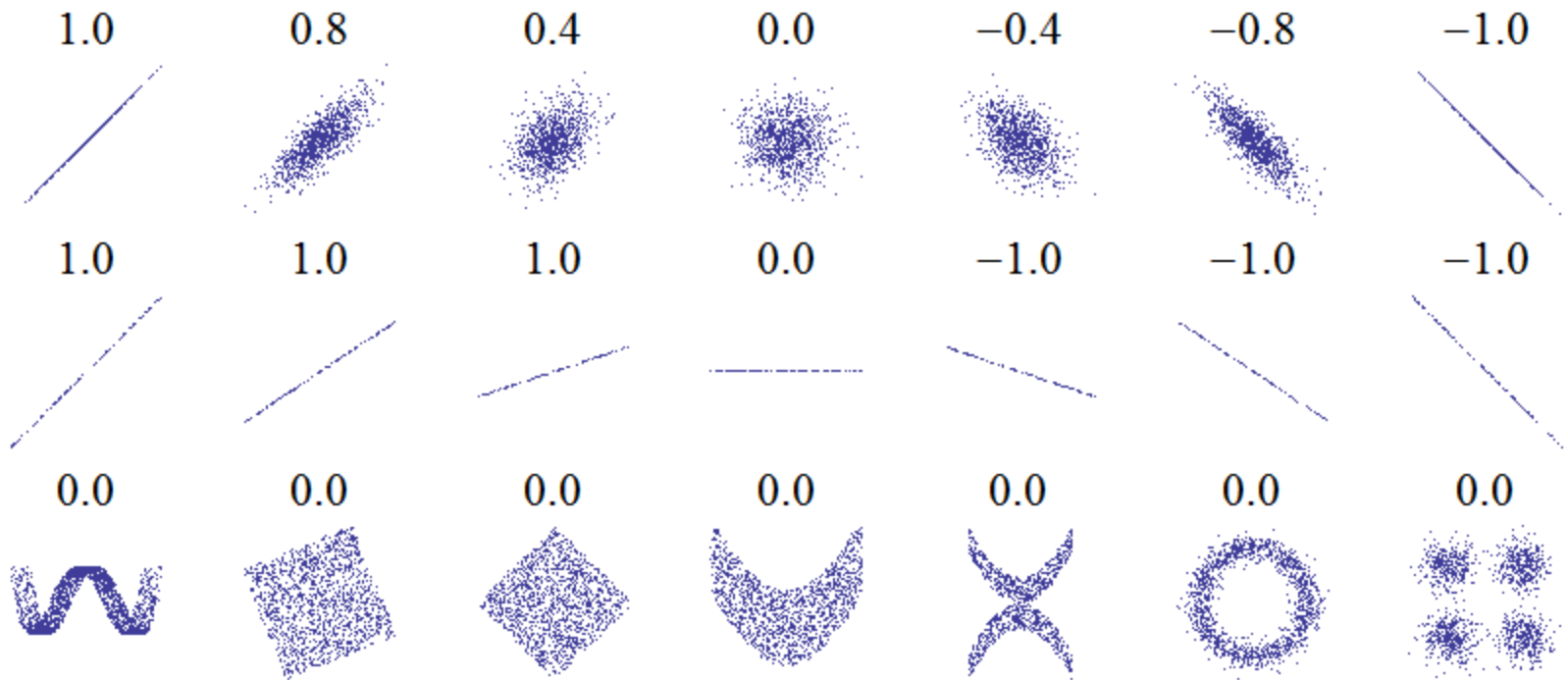
\bar{x} : Mean
 σ_x : Standard deviation
 N : Number of pixels

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

Normalized Cross Correlation:

Expresses the linear relationship between voxel intensities in the two volumes

NCC - example

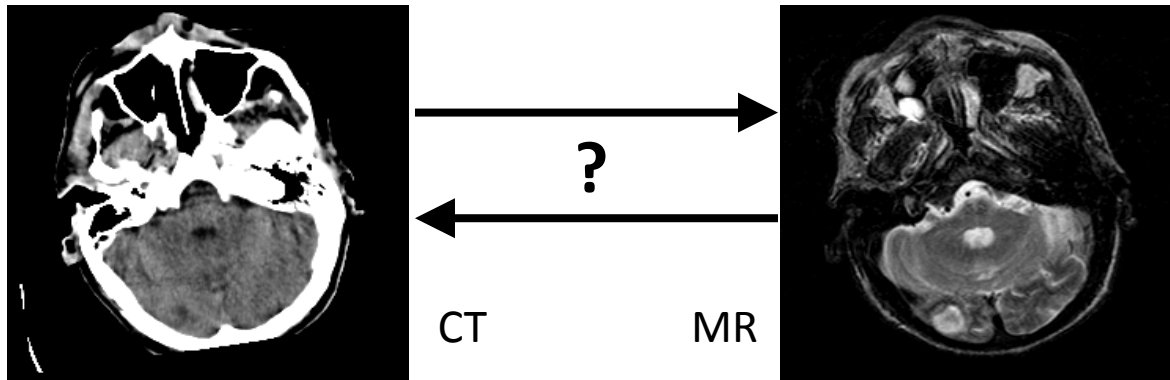


Source: wikipedia

Sets of (x,y) points, with the NCC of x and y for each set.

Multi-Modal Registration

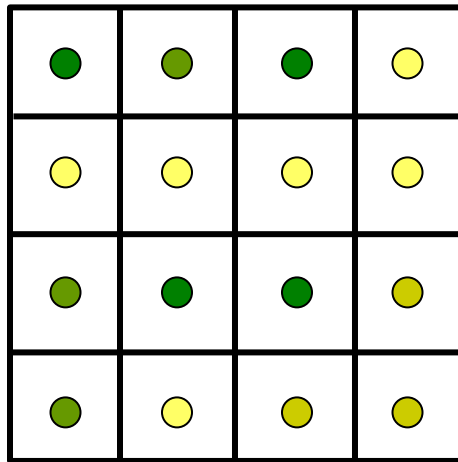
- More complex intensity relationship



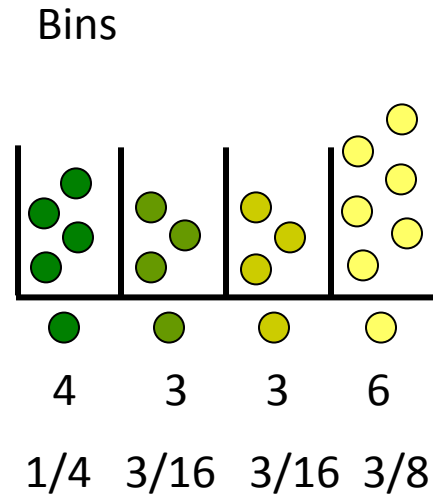
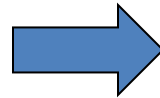
- Approaches:
 - Simulate one modality from the other one
 - Apply sophisticated similarity measures

Information Theoretic Approach

Histogram calculation



Image



counts

probs

Histogram

Joint histogram calculation

Image X

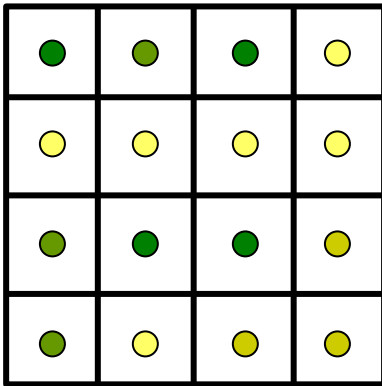
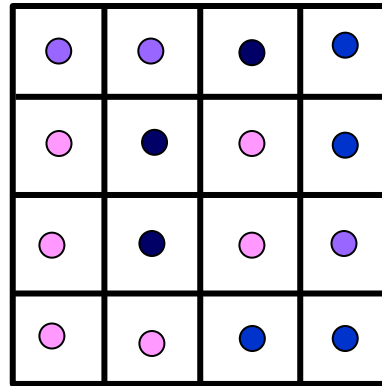
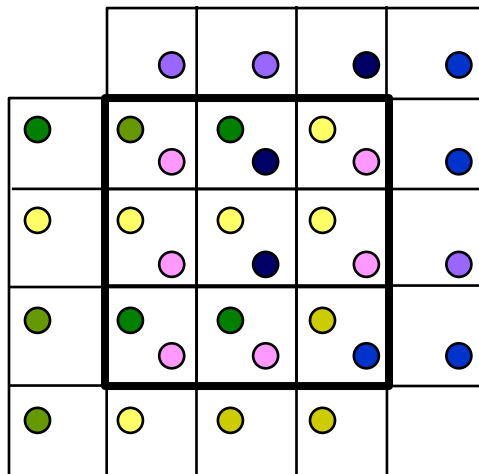


Image Y

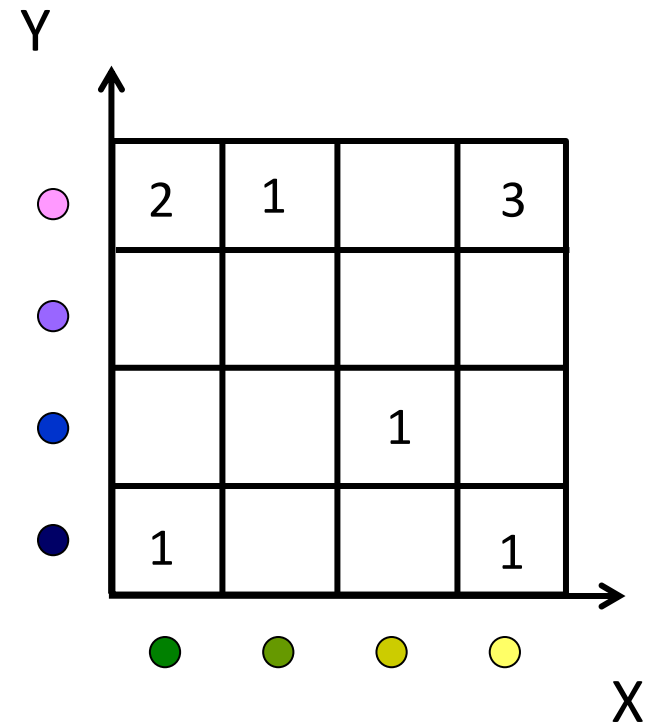


Overlap



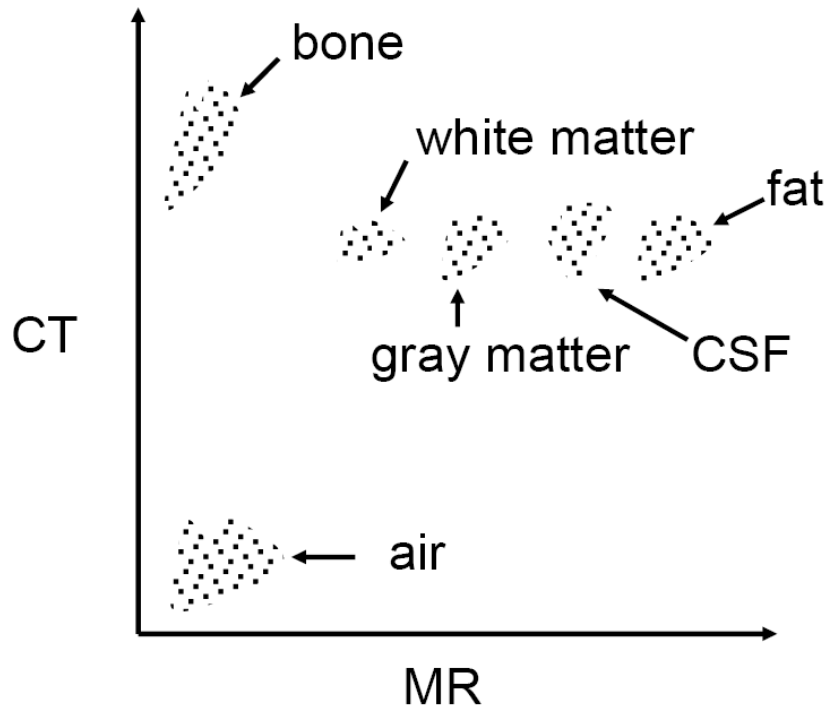
$$\begin{pmatrix} t_x = 1 \\ t_y = 1 \end{pmatrix}$$

Joint Histogram

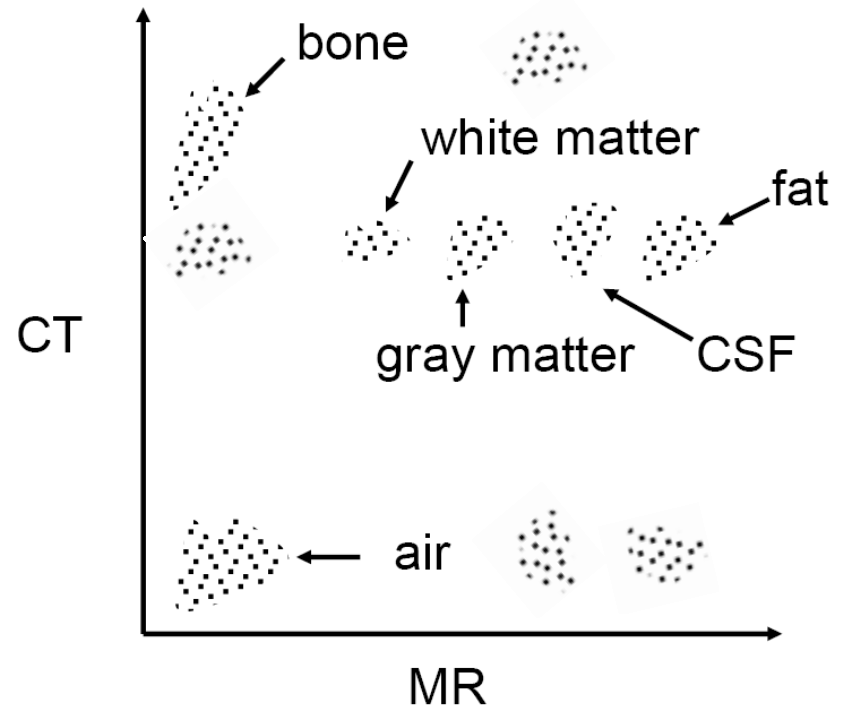


Information Theoretic Approach

Registered



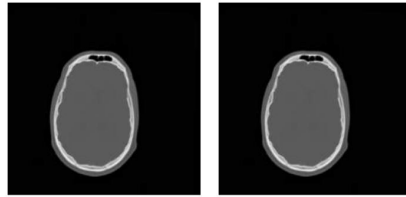
Not Registered



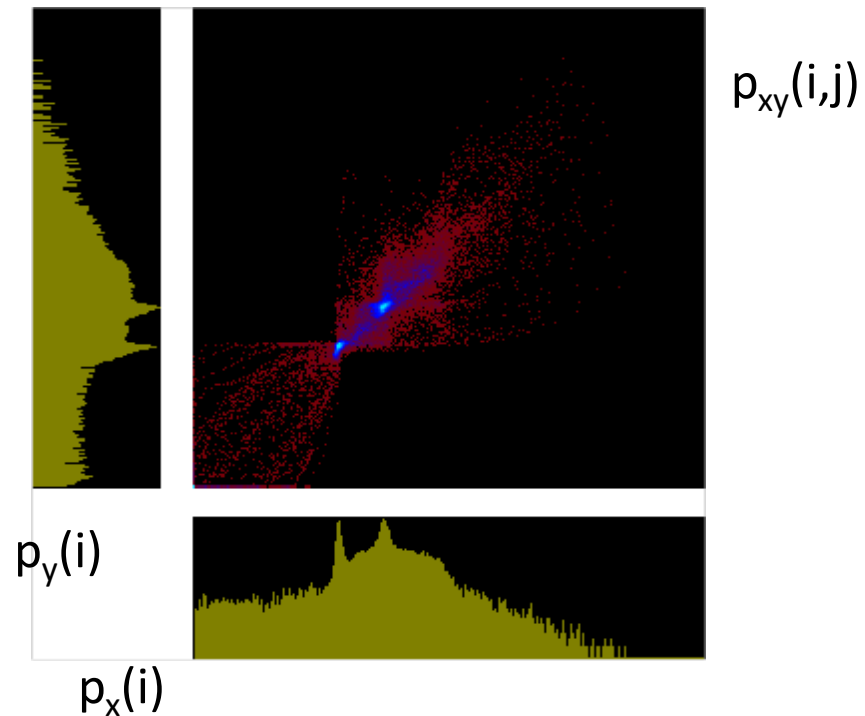
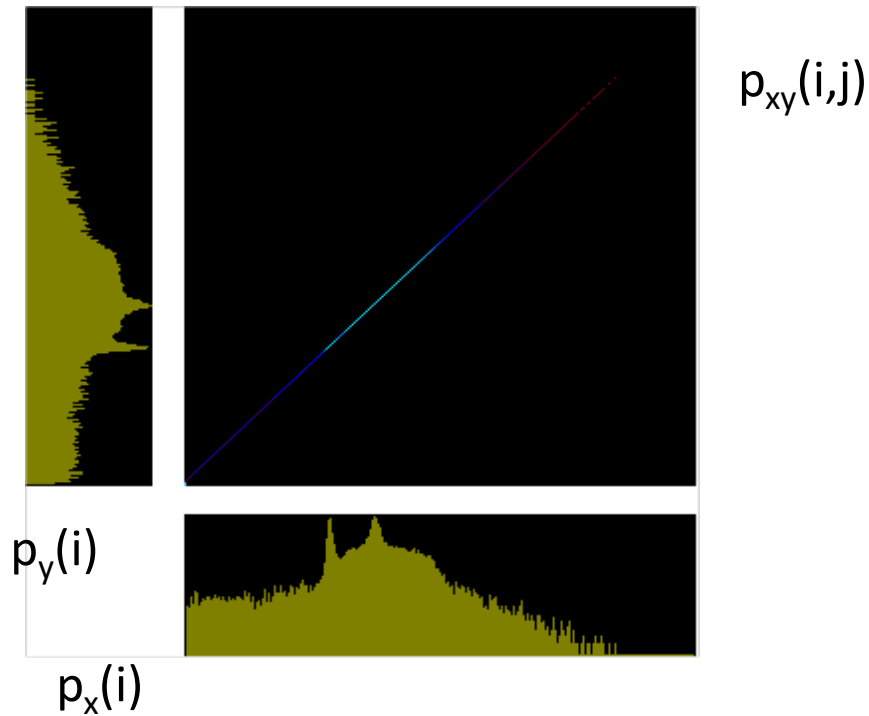
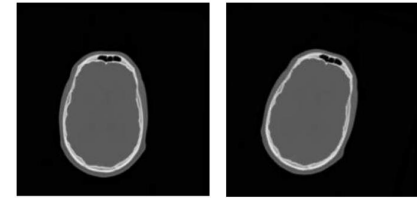
Source: W. Wells, MICCAI 2009

Joint Histogram

X and Y identical



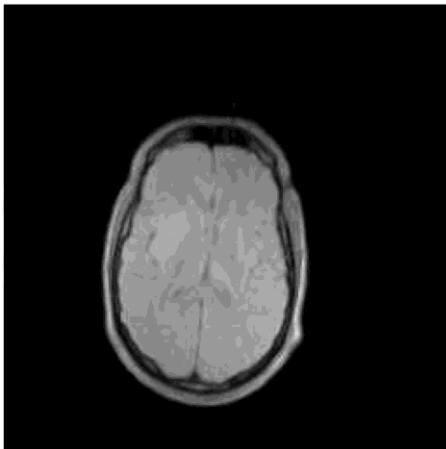
X and Y misaligned



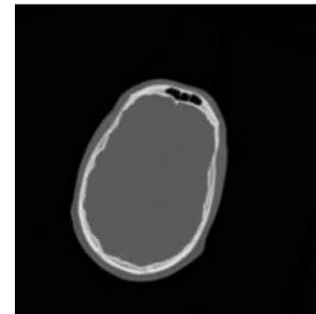
Joint Histogram

- Histogram for images from different Modalities

Source Image

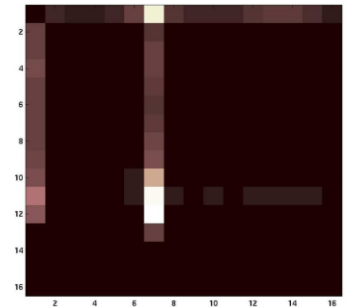


Target Image

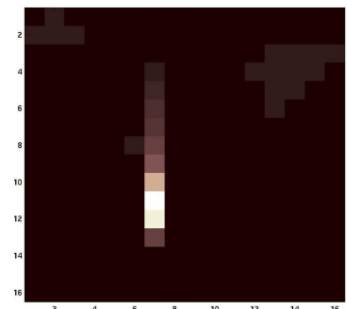
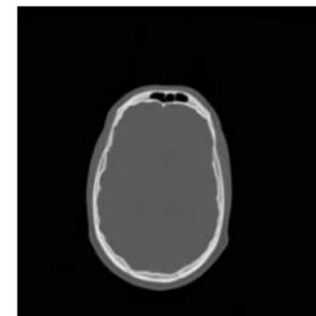


Not Aligned

Joint Histogram

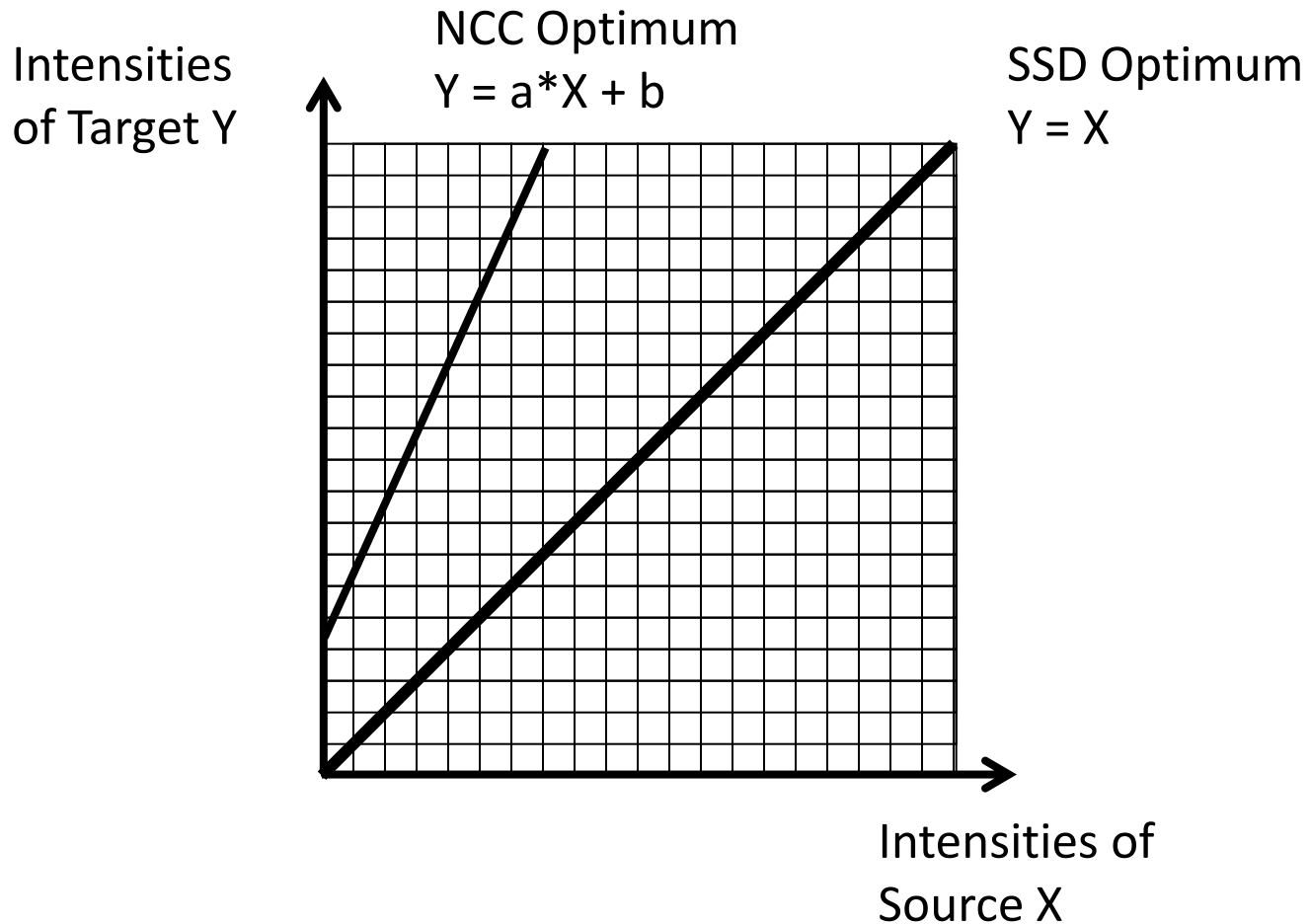


Aligned

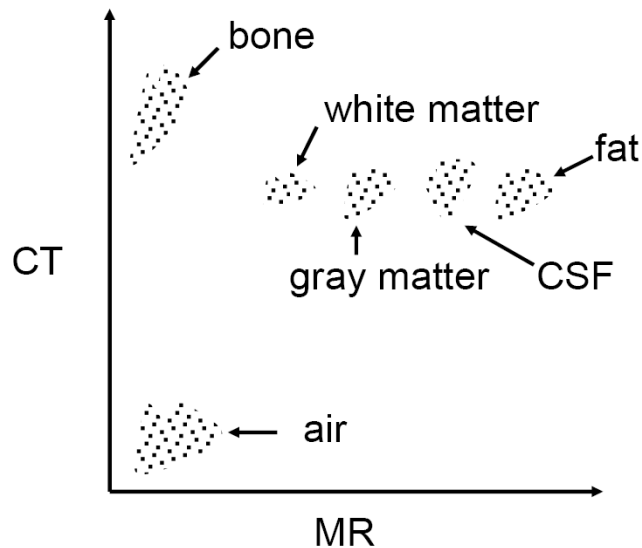


Source: PhD Thesis, L. Zöllei

Joint Histogram



Information Theoretic Approach



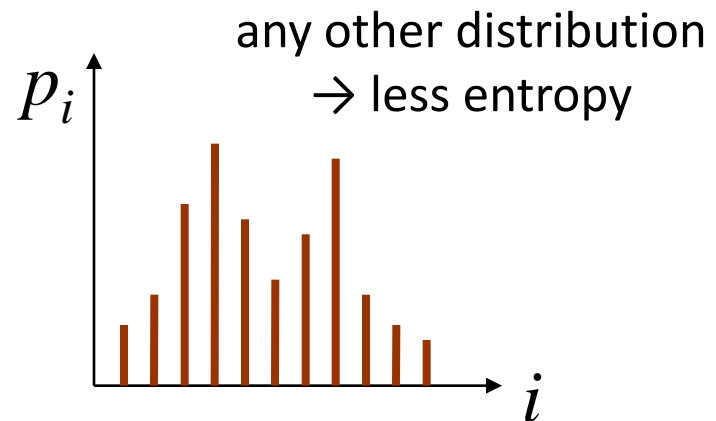
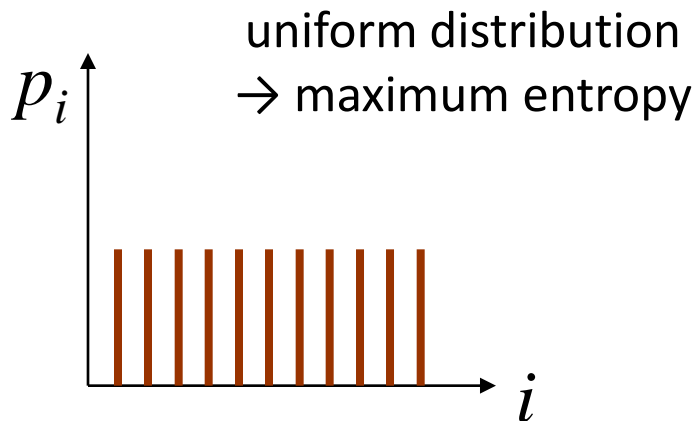
Source: W. Wells, MICCAI 2009

- How to quantify the quality of alignment between MR and CT?
- ✓ Measure the *structure* of the joint distribution
- How to measure the structure?
- ✓ Shannon Entropy

Entropy

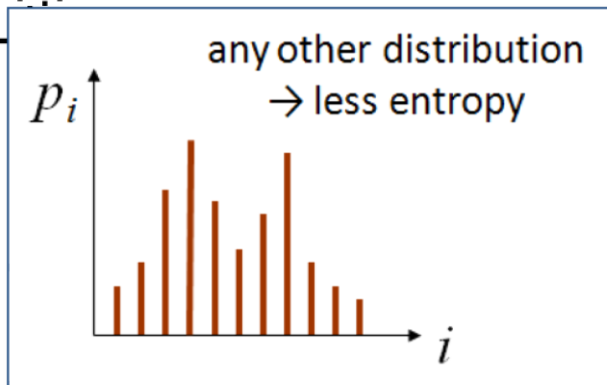
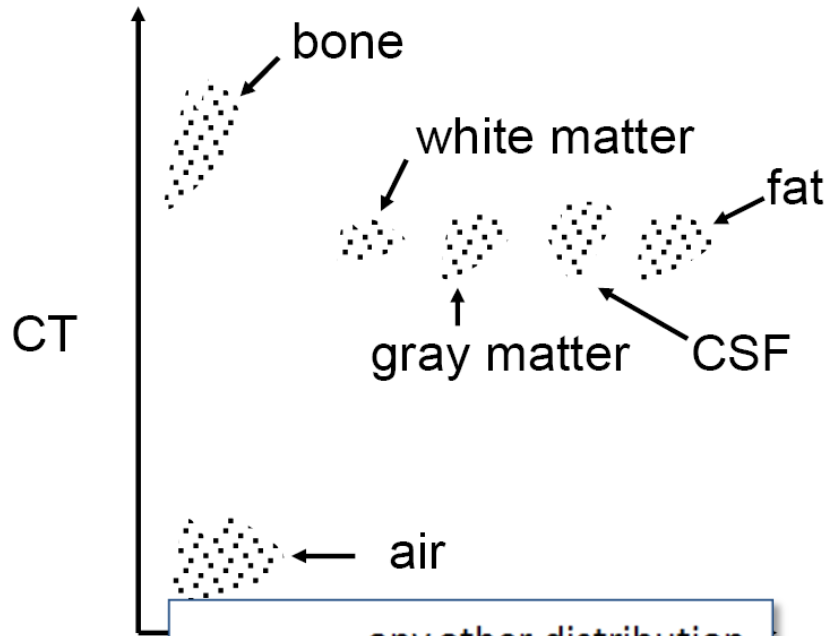
Shannon Entropy, developed in the 1940s
(communication theory)

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

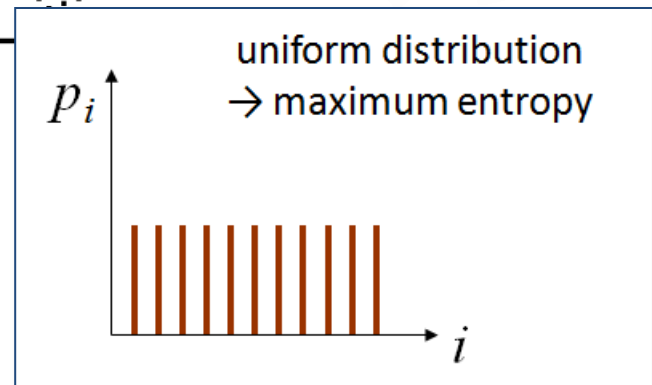
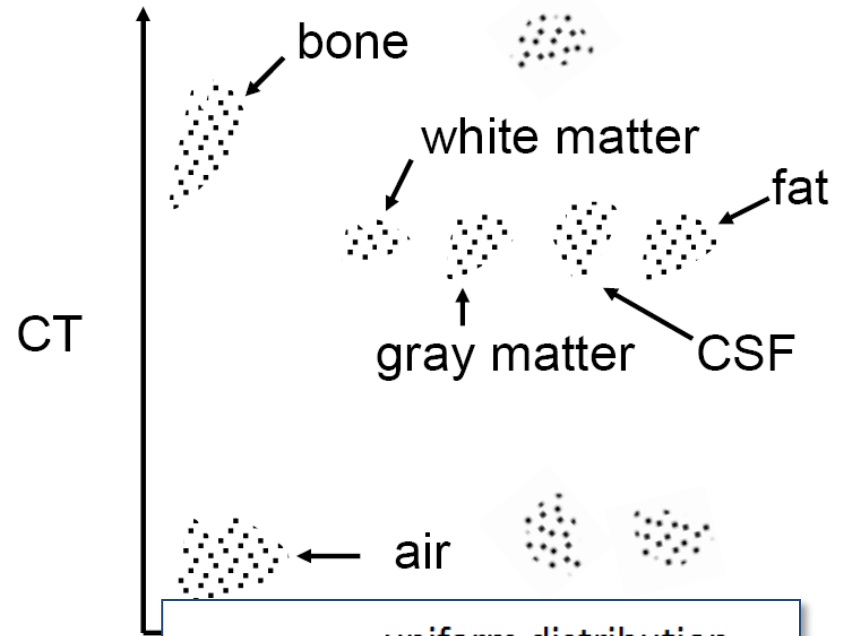


Maximize or minimize entropy?

Registered



Not Registered



Mutual Information (MI)

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

- 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*. CVRME 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.

Normalization of MI

- Problem: changing overlap affects MI
- Entropy Correlation Coefficient (ECC)

$$\text{ECC}(X, Y) = \sqrt{2 - \frac{2 \cdot H(X, Y)}{H(X) + H(Y)}}$$

Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. *Multimodality image registration by maximization of mutual information*, TMI, 1997

- Normalized MI (NMI)

$$\text{NMI}(X, Y) = \frac{H(X) + H(Y)}{H(X, Y)}$$

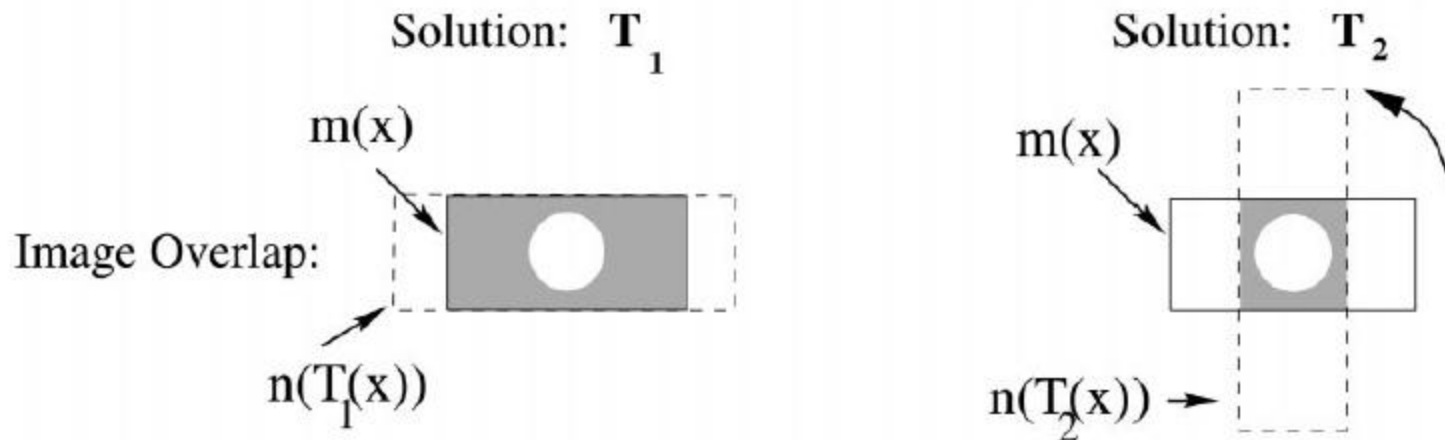
C. Studholme, D.L.G.Hill, D.J. Hawkes, *An Overlap Invariant Entropy Measure of 3D Medical Image Alignment*, Pattern Recognition, Vol. 32(1), Jan 1999, pp 71-86.

- Revisiting overlap invariance

Nathan D. Cahill, Julia A. Schnabel, J. Alison Noble, David J. Hawkes, *"Revisiting overlap invariance in medical image alignment,"* MMBIA, 2008

NMI – change of overlap

- Two images



- T_2 has higher information content
- Overlap changes for deformable registration?

Improvements to MI

- Density estimation
 - Parzen window
 - Partial volume distribution
 - Uniform volume histogram
 - NP windows
- Spatial information
- Alternative entropy measures

- 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>

Derivatives of Similarity Measures

- General form of derivative of similarity metrics

$$\frac{\partial \text{SM}(X, Y(T_p))}{\partial p} = \boxed{\frac{\partial \text{SM}(X, Y)}{\partial Y}} \frac{\partial Y}{\partial T_p} \frac{\partial T_p}{\partial p}$$

– SSD:

$$\nabla \text{SSD}(X, Y) = -2 \cdot (X - Y)$$

– MI

$$\nabla \text{MI}(X, Y) = G_{\Psi} * \frac{1}{|\Omega|} \left(\frac{\partial_2 p(X, Y)}{p(X, Y)} - \frac{p'(Y)}{p(Y)} \right)$$

Hermosillo, G., Chéfd'Hotel, C., Faugeras, O.: *Variational Methods for Multimodal Image Matching*, International Journal of Computer Vision 50(3) (2002)

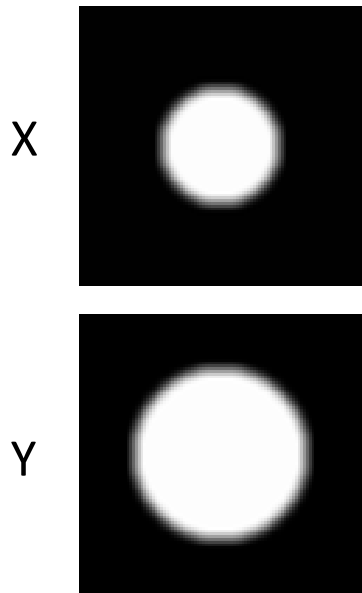
Derivatives of Similarity Measures

- General form of derivative of similarity metrics

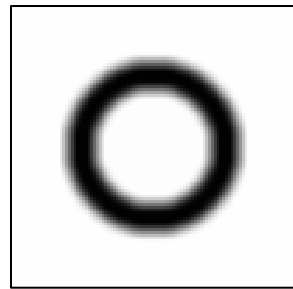
$$\frac{\partial \text{SM}(X, Y(T_p))}{\partial p} = \frac{\partial \text{SM}(X, Y)}{\partial Y} \frac{\partial Y}{\partial T_p} \quad \cancel{\frac{\partial T_p}{\partial p}}$$

Derivatives of Similarity Measures

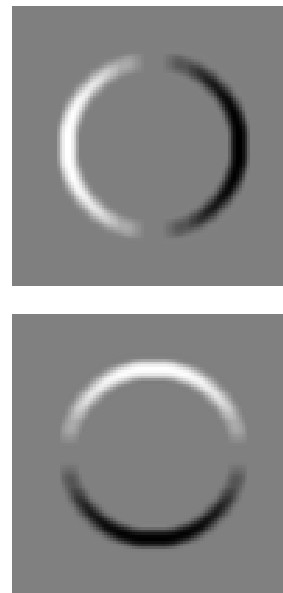
Images



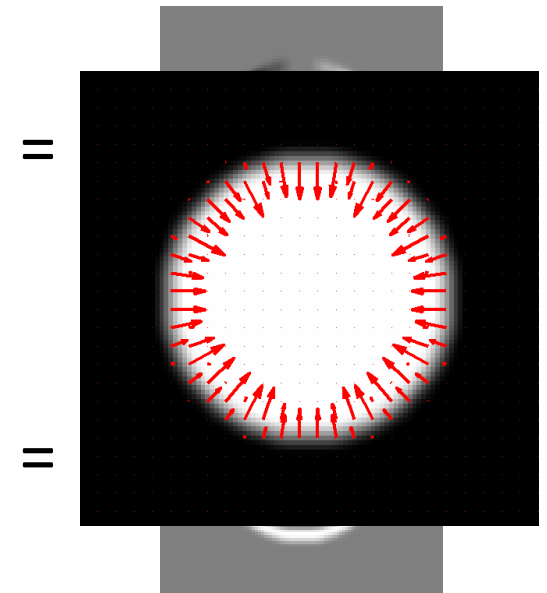
∇SSD



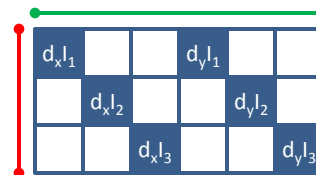
∇Y



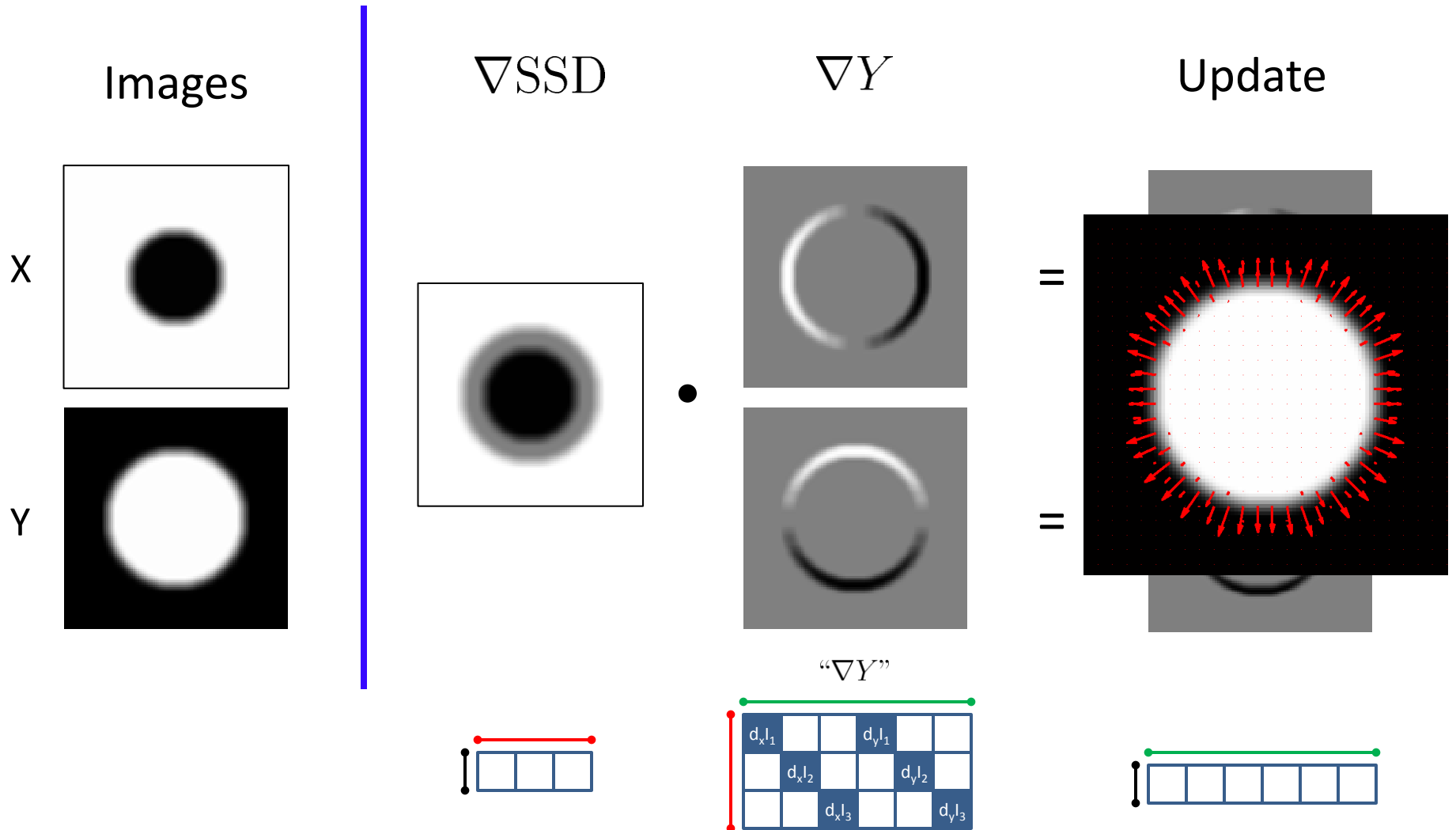
Update



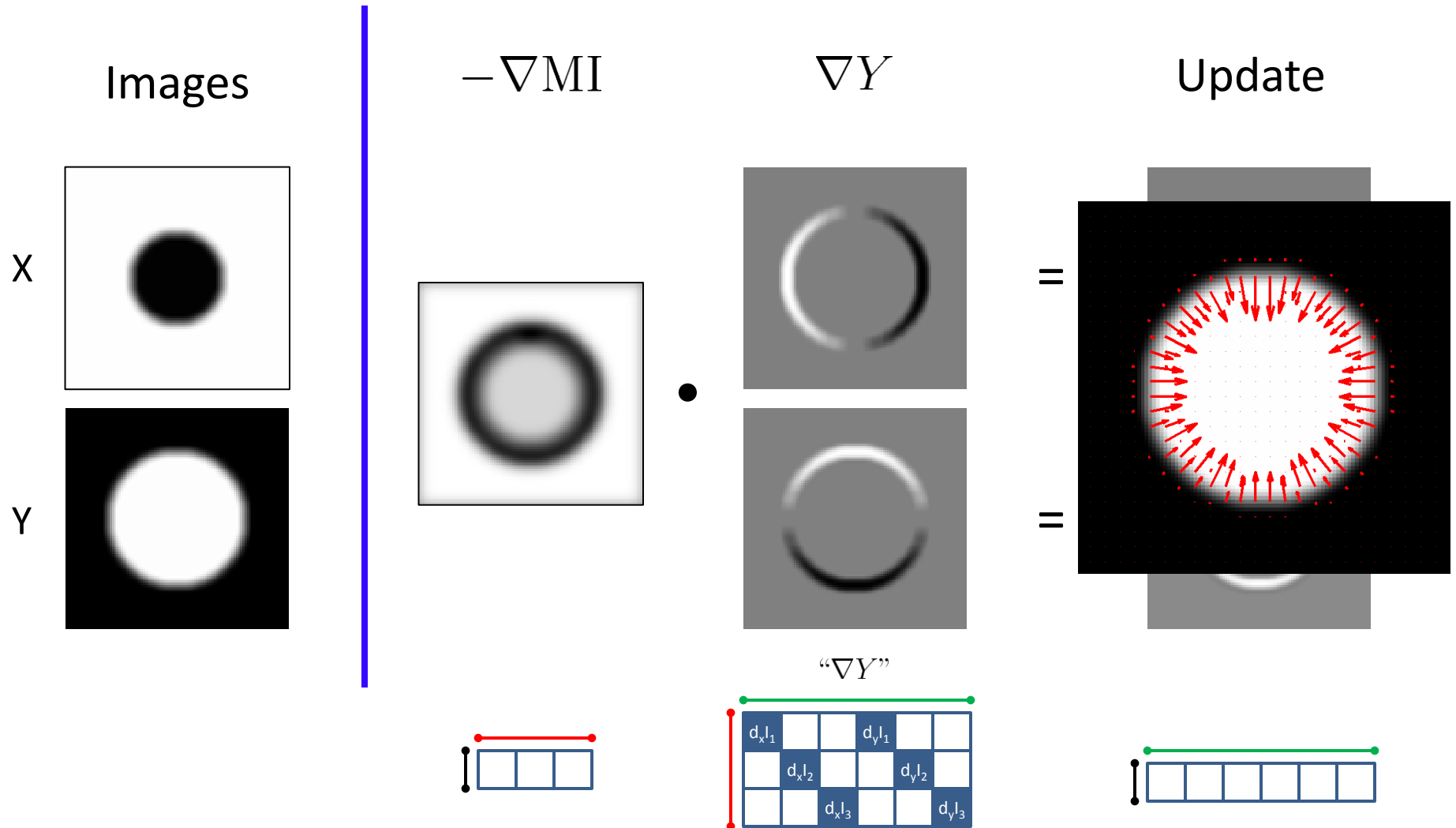
“ ∇Y ”



Derivatives of Similarity Measures



Derivatives of Similarity Measures



Overview

1. Standard similarity measures
2. Probabilistic framework
3. Pre-processing steps
4. Recent approaches
5. Linear vs. non-linear registration

PART II

Probabilistic Framework for Image Registration

- Viola, PhD thesis, 1995
- Roche et al., 2000
- Zöllei, PhD thesis, 2006

Probabilistic Framework for registration

- Maximum Likelihood Estimation (MLE)

- Probability for the model m having the observations a

$$\arg \max_m P(a = \{a_1, \dots, a_n\} | m) = \arg \max_m \prod_{i=1}^n P(a_i | m)$$

- Log-likelihood function

$$\log \mathcal{L}(m | a) = \sum_{i=1}^n \log P(a_i | m)$$

- Formulate registration as likelihood maximization

MLE framework for registration

- Model

$$X = f(Y(T)) + \varepsilon$$

X, Y : Images

ε : Noise

f : Intensity map

- Probability function

$$P(X|Y, T, \varepsilon, f)$$

- Log-likelihood function

$$\begin{aligned} \log \mathcal{L}(T, \varepsilon, f) &= \log P(X|Y, T, \varepsilon, f) \\ &= \frac{1}{N} \sum_{v \in \Omega} \log P(X(v)|Y(T(v)), \varepsilon, f) \end{aligned}$$

$$\log \mathcal{L}(T, \varepsilon, f) = \frac{1}{N} \sum_{v \in \Omega} \log P(X(v) | Y(T(v)), \varepsilon, f)$$

- Stationary white Gaussian noise

$$P(X|Y, T, f, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \cdot e^{-\frac{1}{2} \frac{[X - f(Y(T))]}{\sigma^2}}$$

- Log-likelihood with Gaussian noise

$$\log \mathcal{L}(T, \sigma, f) = -N \log \sqrt{2\pi\sigma} - \frac{1}{2\sigma^2} \frac{1}{N} \sum_{v \in \Omega} (X(v) - f(Y(T(v))))^2$$

Assuming f to be the identity

$$\log \mathcal{L}(T, \varepsilon, f) = \frac{1}{N} \sum_{v \in \Omega} \log P(X(v) | Y(T(v)), \varepsilon, f)$$

- Stationary white Gaussian noise

$$P(X|Y, T, f, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \cdot e^{-\frac{1}{2} \frac{[X - f(Y(T))]}{\sigma^2}}$$

- Log-likelihood with Gaussian noise

$$\log \mathcal{L}(T, \sigma) = -N \log \sqrt{2\pi\sigma} - \frac{1}{2\sigma^2} \frac{1}{N} \sum_{v \in \Omega} (X(v) - Y(T(v)))^2$$

SSD

Intensity relationships

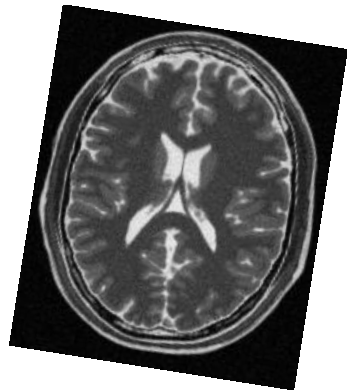
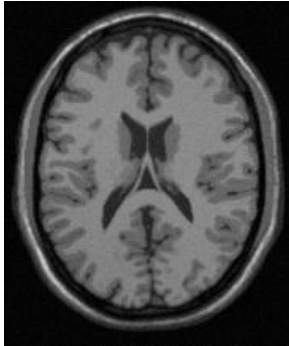
- Identity : SSD
- Affine : NCC
- Functional : Correlation Ratio
- Statistical : Mutual Information

Viola, P. *Alignment by maximization of Mutual Information*. MIT PhD Thesis, June 1995.
Roche et al., *Unifying maximum likelihood approaches in medical image registration*, 2000

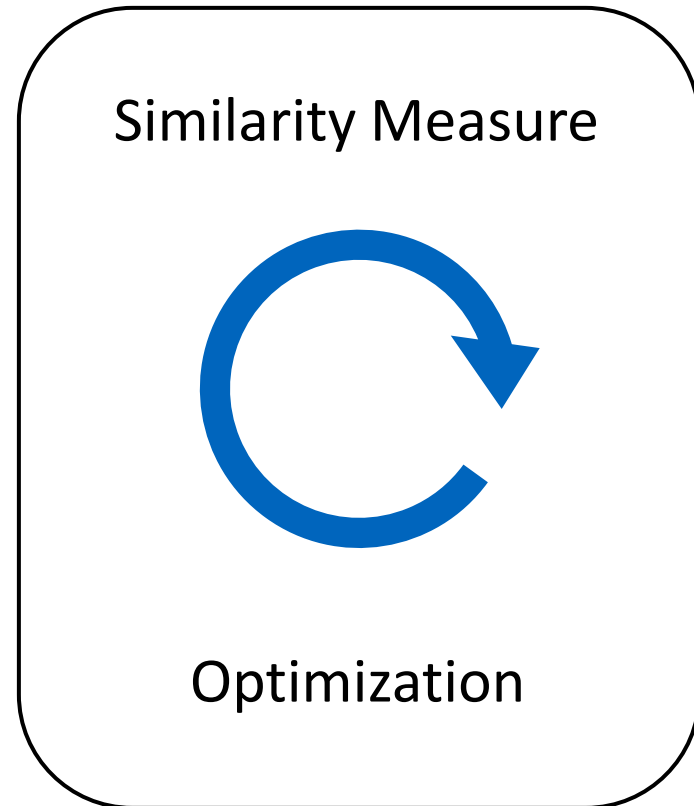
PART III

Pre-Processing Steps

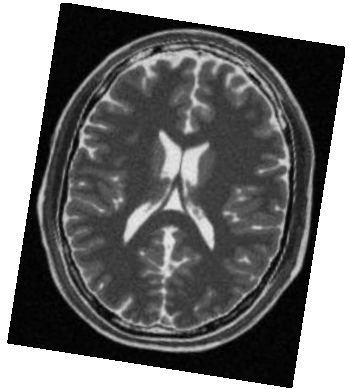
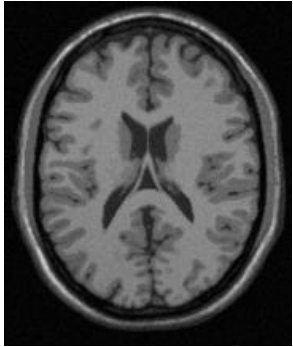
Images



Registration Framework



Images

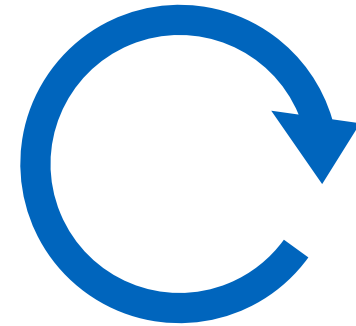


Pre-processing

1. Image gradients
2. Entropy images
3. Phase
4. Multi-resolution
5. Attribute vectors

Registration Framework

Similarity Measure



Optimization

Image Gradients

- Normalized gradient fields

$$n(X) := \begin{cases} \frac{\nabla X}{\|\nabla X\|}, & \text{if } \nabla X \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

“two images are considered similar, if intensity changes occur at the same locations”

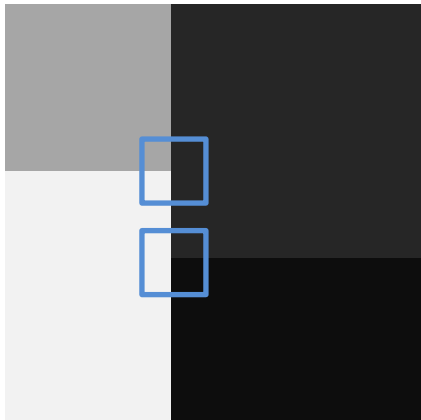
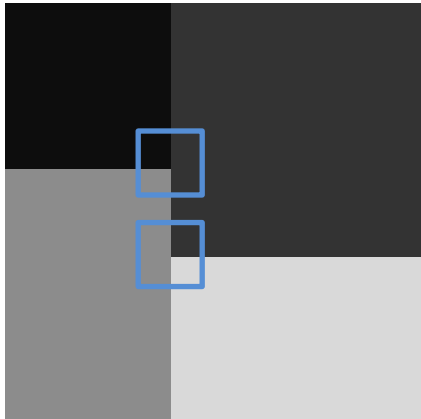
- Maximize square of cosine

$$d(X, Y) = \langle n(X), n(Y) \rangle^2$$

Haber, E. and Modersitzki, J., *Intensity gradient based registration and fusion of multi-modal images*, MICCAI 2006

Image Gradients

Synthetic images



Gradient images

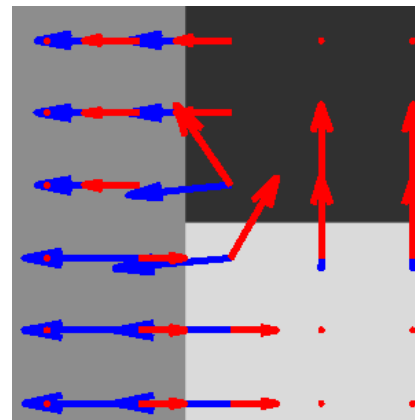
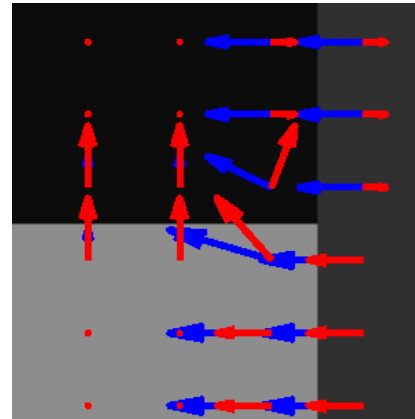
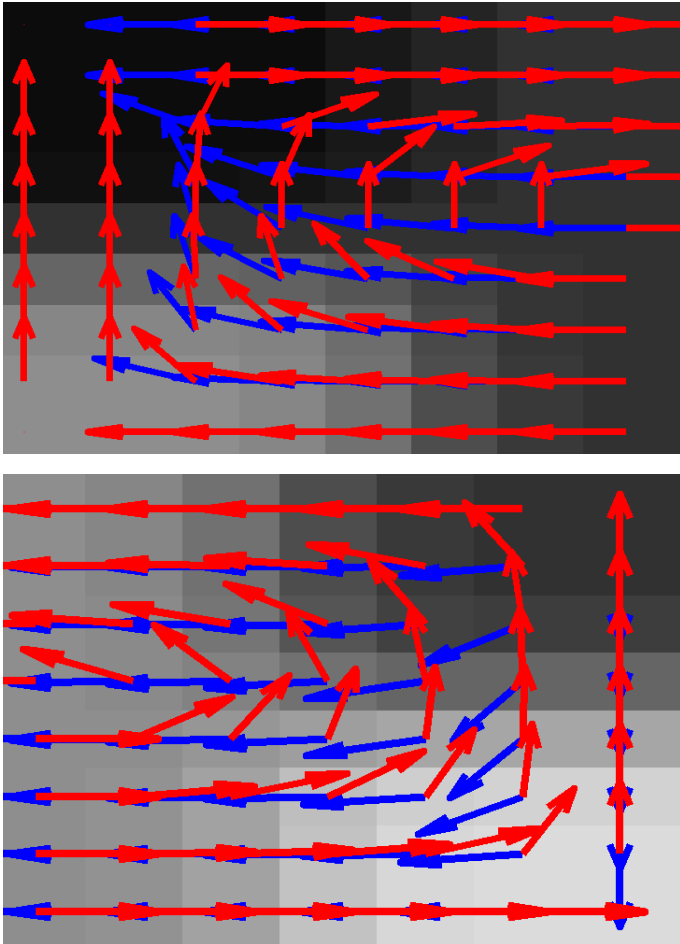
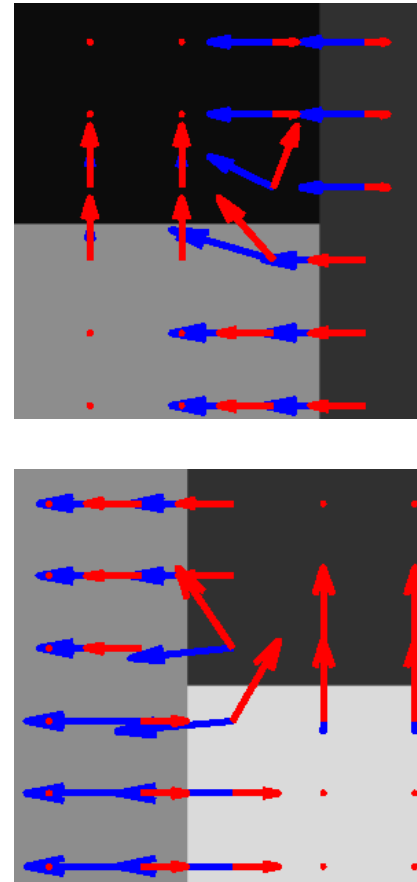


Image Gradients

Gradient of smoothed images



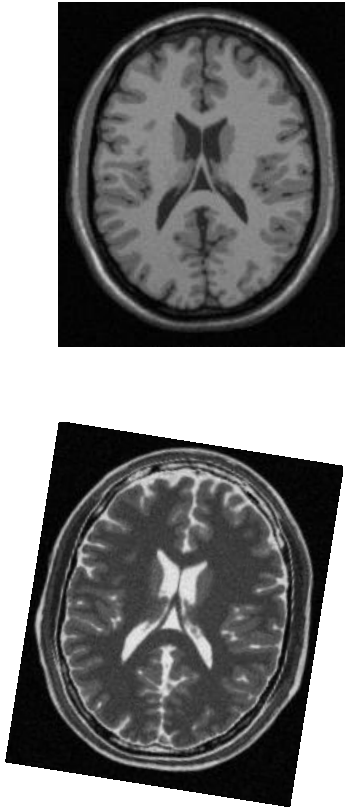
Gradient images



Images

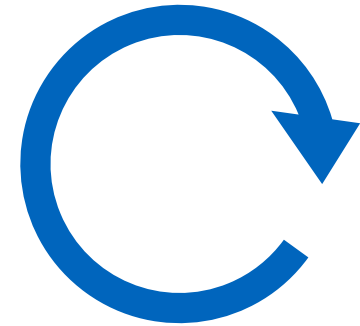
Pre-processing

Registration Framework



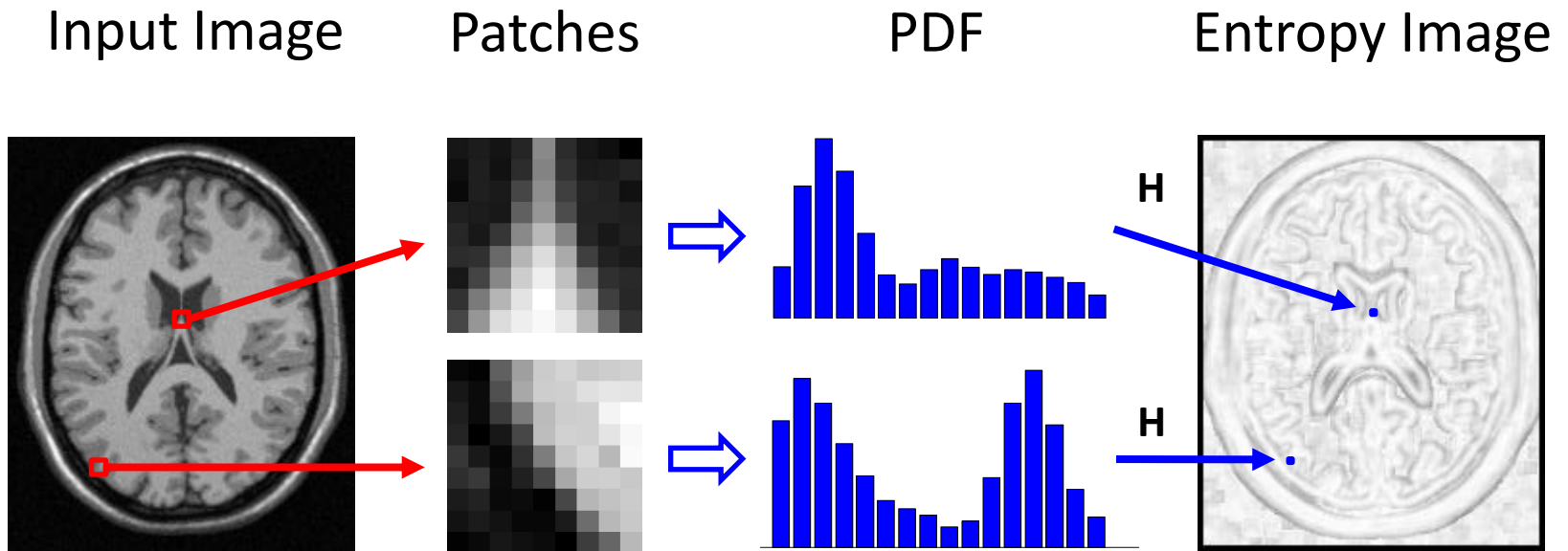
1. Image gradients
2. Entropy images
3. Phase
4. Multi-resolution
5. Attribute vectors

Similarity Measure



Optimization

Entropy Images

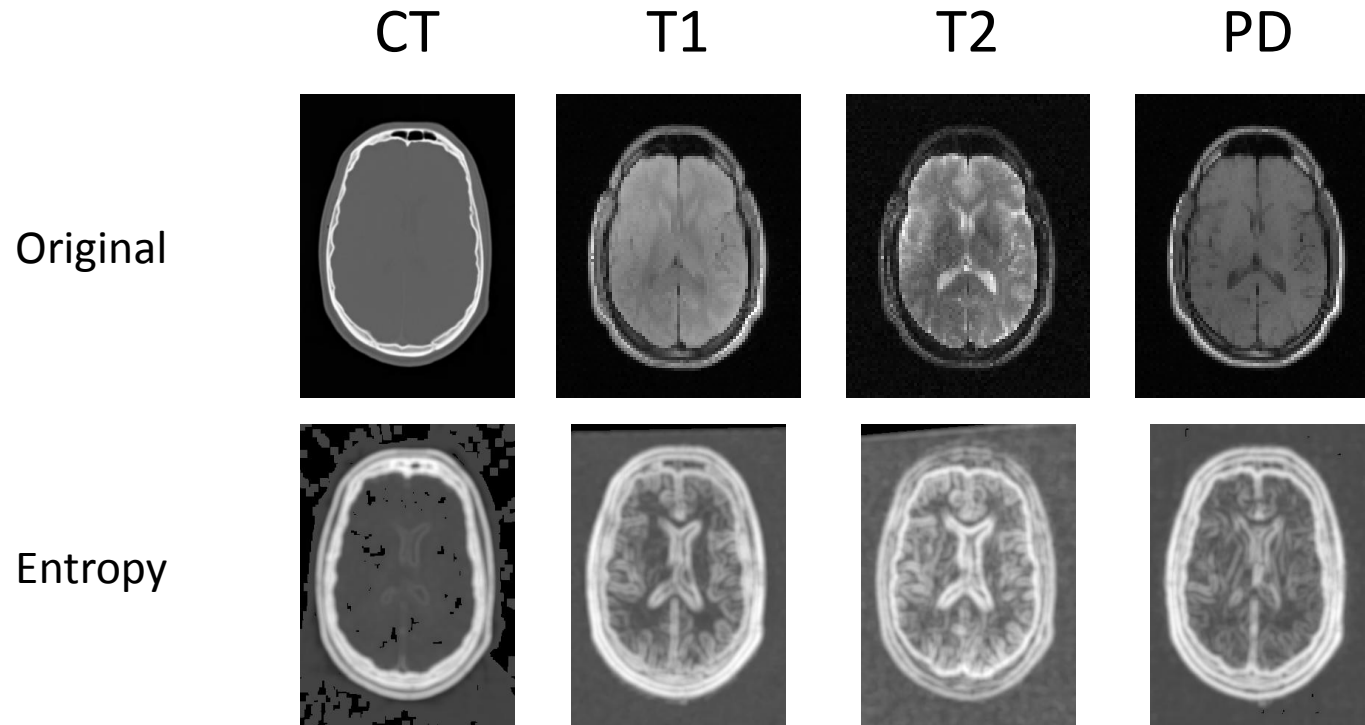


Wachinger, Navab, *Structural Images for Image Registration*, MMBIA, 2010.

Entropy Images

- Examples from the RIRE dataset

<http://www.insight-journal.org/rire/>



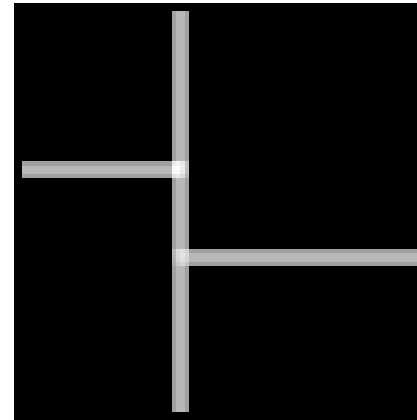
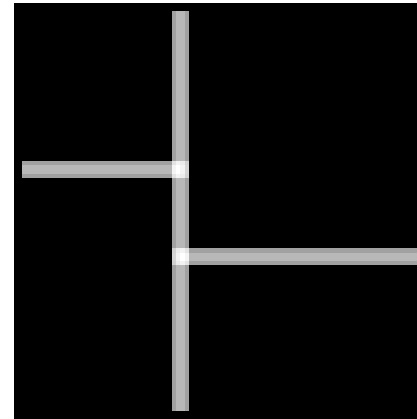
Wachinger, Navab, *Structural Images for Image Registration*, MMBIA, 2010.

Entropy Images

Synthetic images



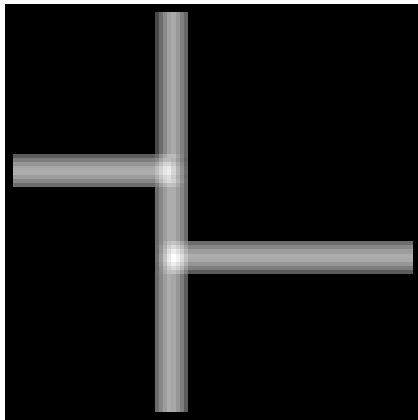
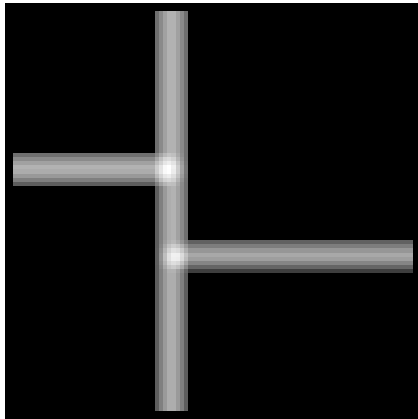
Entropy images



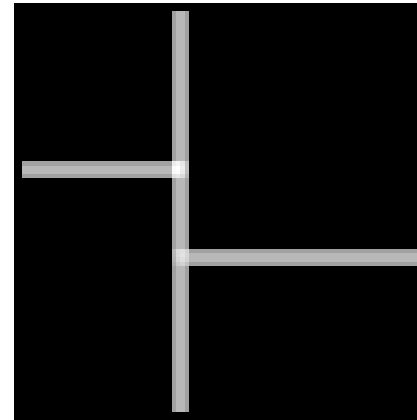
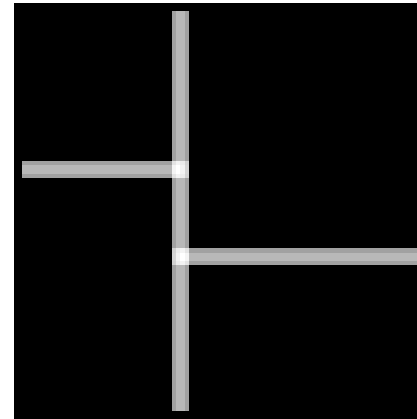
Wachinger, Navab, *Structural Images for Image Registration*, MMBIA, 2010.

Entropy Images

Entropy on smoothed images

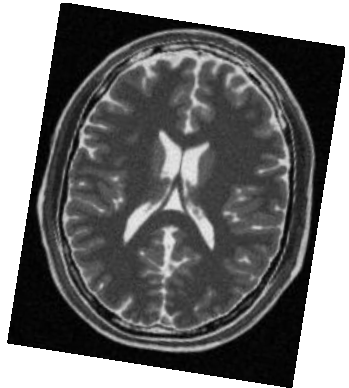
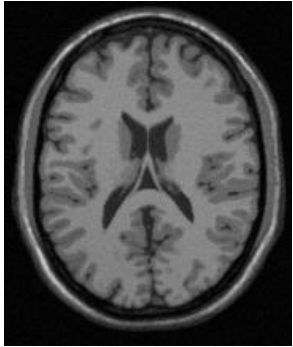


Entropy images



Wachinger, Navab, *Structural Images for Image Registration*, MMBIA, 2010.

Images

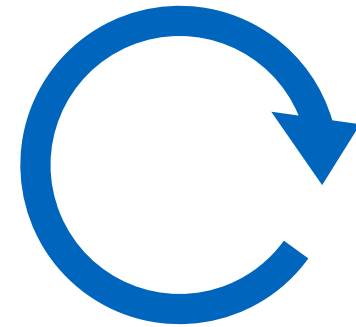


Pre-processing

1. Image gradients
2. Entropy images
3. Phase
4. Multi-resolution
5. Attribute vectors

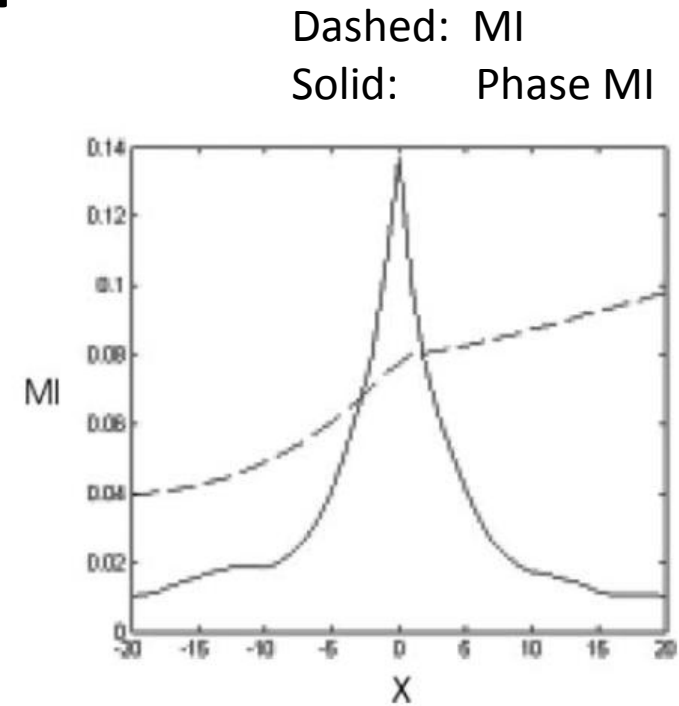
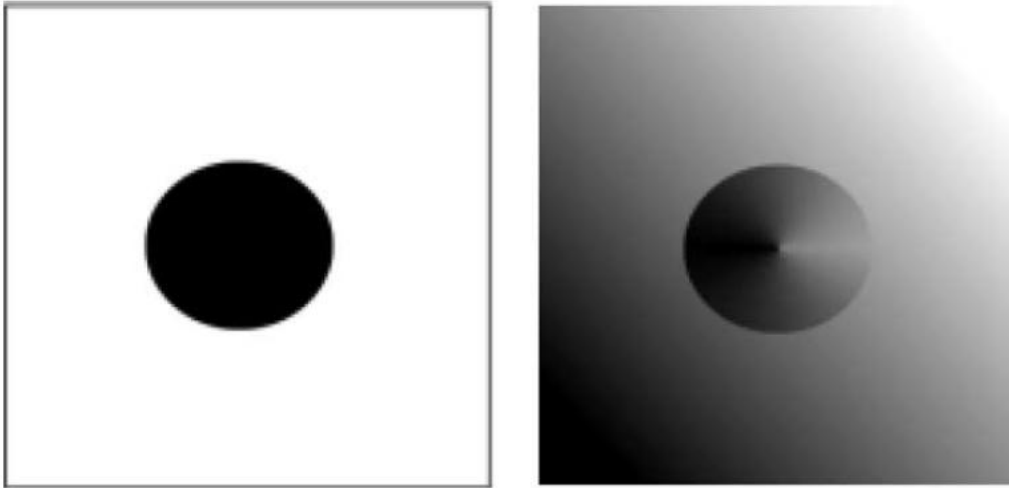
Registration Framework

Similarity Measure



Optimization

Phase-base registration



- Fourier representation

$$f(x) = \int_0^{\infty} A(\omega) \cdot \sin(\omega x + \phi(\omega)) d\omega$$

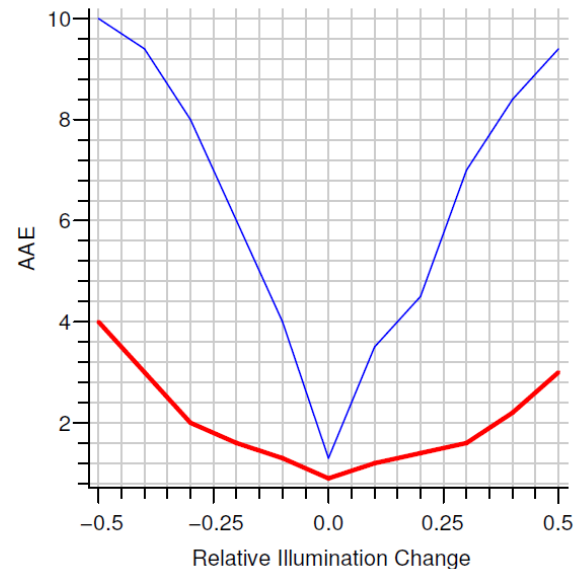
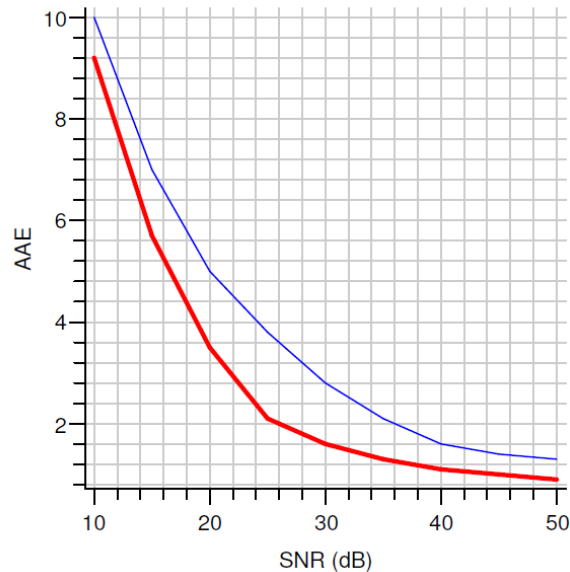
$A(\omega)$: Amplitude

$\phi(\omega)$: Phase

M. Mellor and M. Brady. *Phase mutual information as a similarity measure for registration*. Medical Image Analysis, 2005.

Optical Flow with Phase

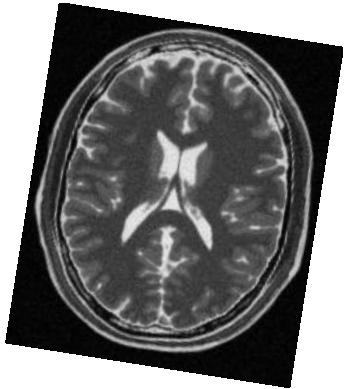
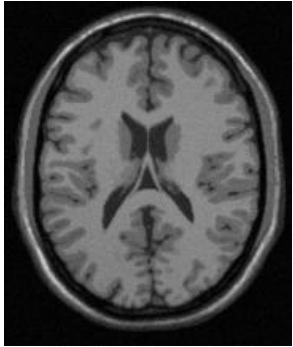
- Idea: replace the assumption of brightness constancy with phase constancy
- Error for changing noise and illumination



blue: gradient
red: phase

L. Wietzke, G. Sommer, O. Fleischmann, *The Geometry of 2D Image Signals*, CVPR, 2009

Images

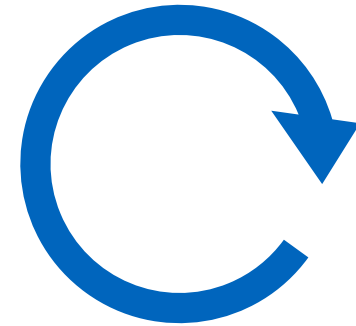


Pre-processing

1. Image gradients
2. Entropy images
3. Phase
4. Multi-resolution
5. Attribute vectors

Registration Framework

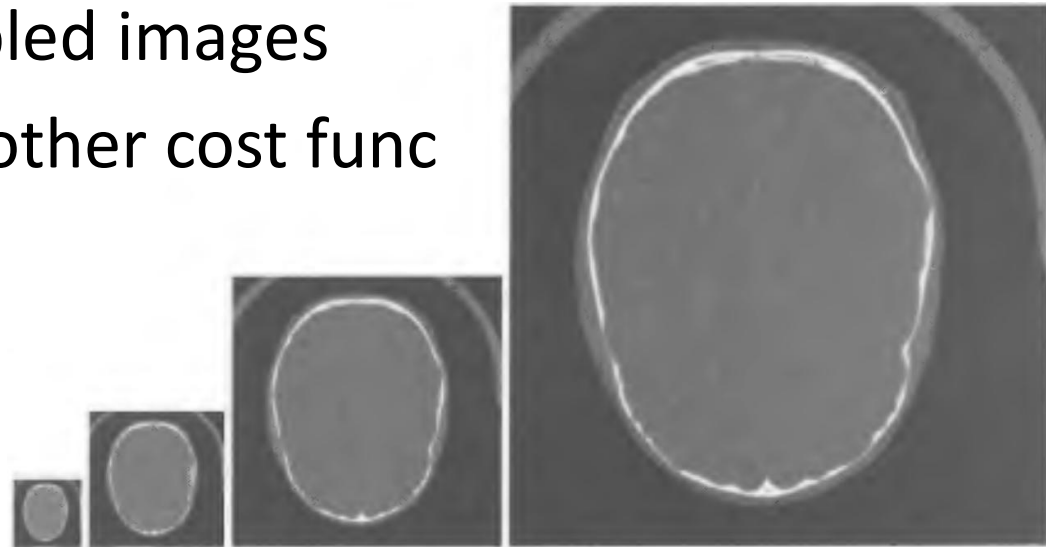
Similarity Measure



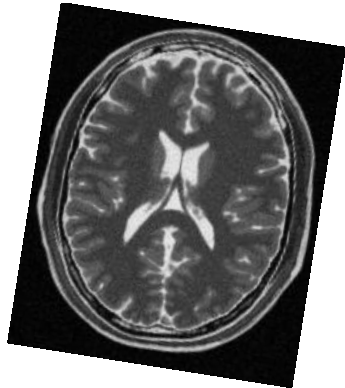
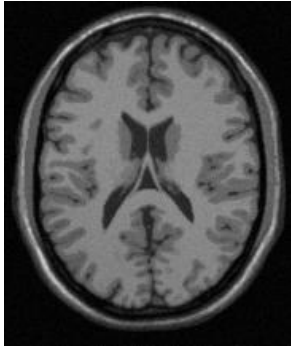
Optimization

Multi-Resolution Registration

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



Images

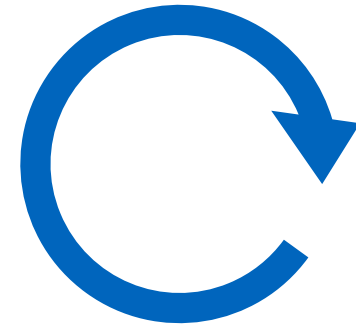


Pre-processing

1. Image gradients
2. Entropy images
3. Phase
4. Multi-resolution
5. Attribute vectors

Registration Framework

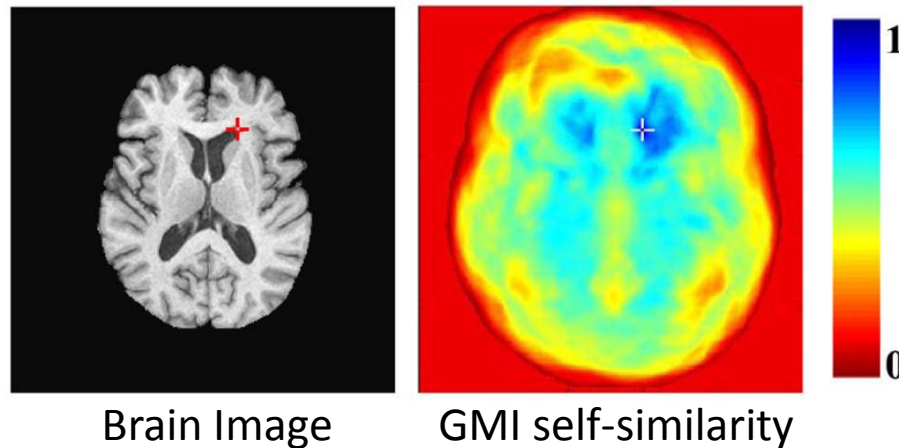
Similarity Measure



Optimization

Attribute Vectors

- Registration with attribute vectors
 $a = [\text{gradient}, \text{intensity}, \text{GMI}]$
- GMI = geometric moment invariants



Shen, D. and Davatzikos, C., *HAMMER: hierarchical attribute matching mechanism for elastic registration*, IEEE Transactions on Medical Imaging, 2002

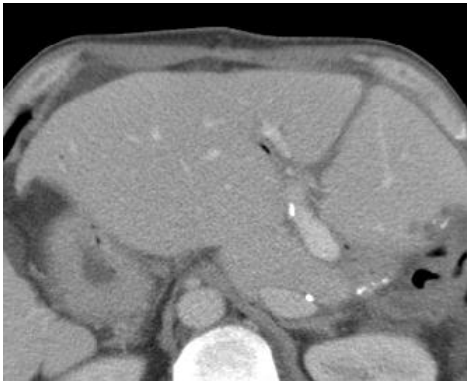
M. Wacker and F. Deinzer. *Automatic robust medical image registration using a new democratic vector optimization approach with multiple Measures*, MICCAI, 2009

PART IV

Recent Approaches

Recent examples...

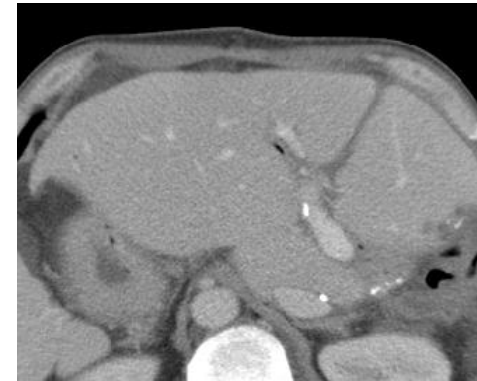
- mainly compare to MI as state of the art
- address the following problems of MI
 - intensity does not represent tissue (e.g. US-CT)



CT



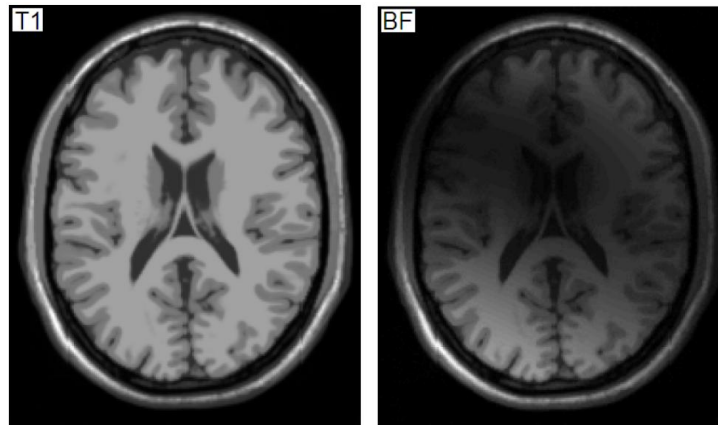
US



US-CT

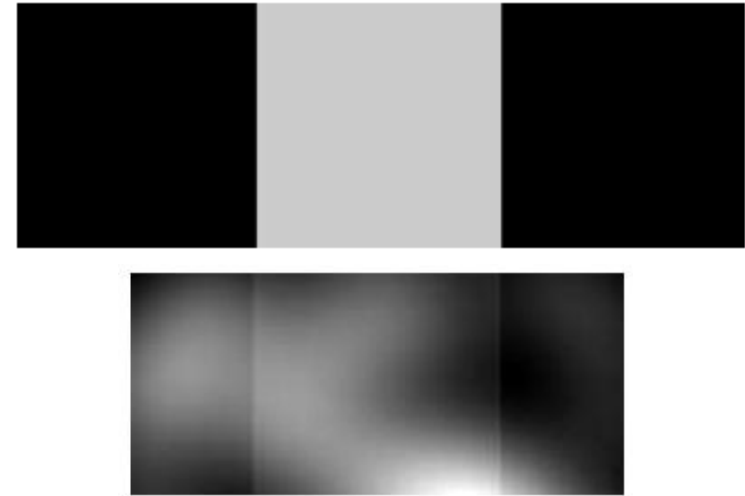
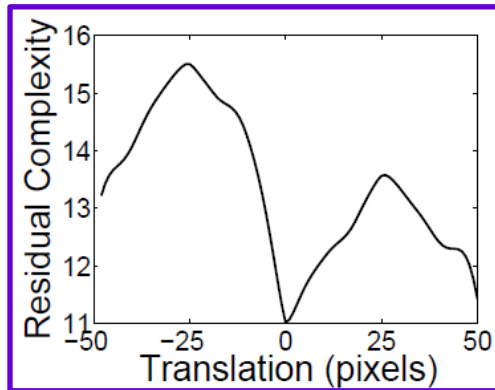
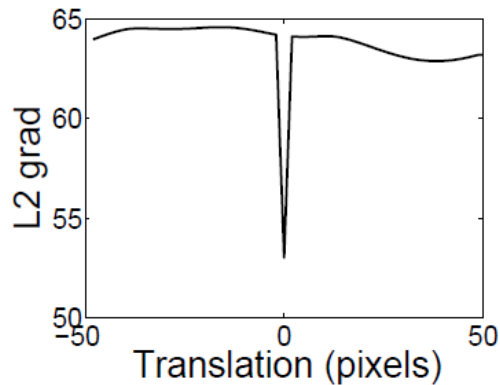
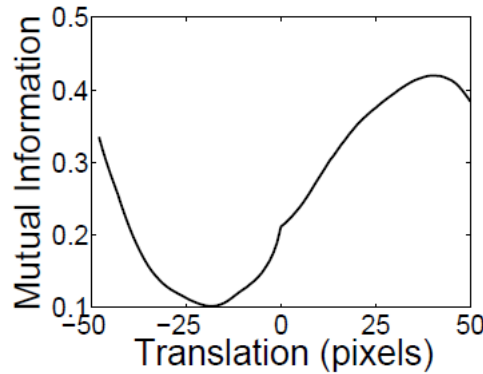
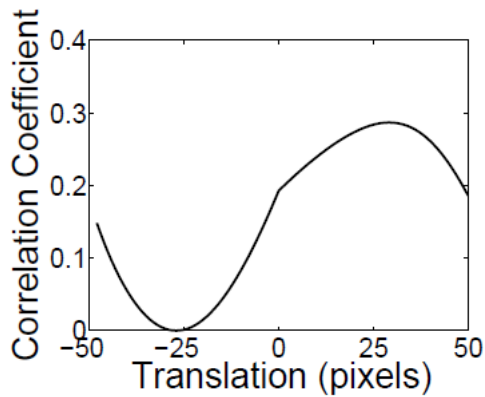
Recent examples...

- mainly compare to MI as state of the art
- address the following problems of MI
 - intensity does not represent tissue (e.g. US-CT)
 - intensity non-uniformity (bias)



Minimizing Residual Complexity

- Motivation: MI fails for these images



Model

$$X = Y(T) + S + \varepsilon$$

Myronenko A., Song X.: "Intensity-based Image Registration by Minimizing Residual Complexity", IEEE Trans. on Medical Imaging, 2010

Minimizing Residual Complexity

- Compress the difference image

$$\sum_{i=1}^N \mathbf{q}_i \cdot (X - Y) \quad \mathbf{q}_i : \text{i-th DCT basis function}$$

- Similarity measure

$$RC(X, Y) = \sum_{i=1}^N \log \left(\frac{(\mathbf{q}_i(X - Y))^2}{\alpha} + 1 \right)$$

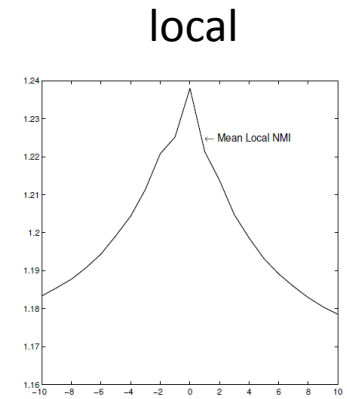
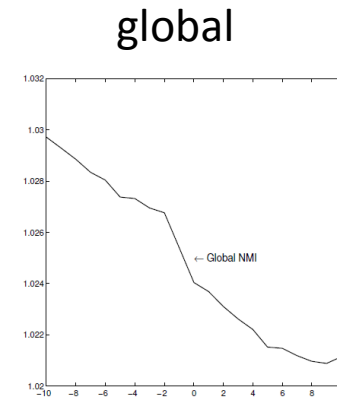
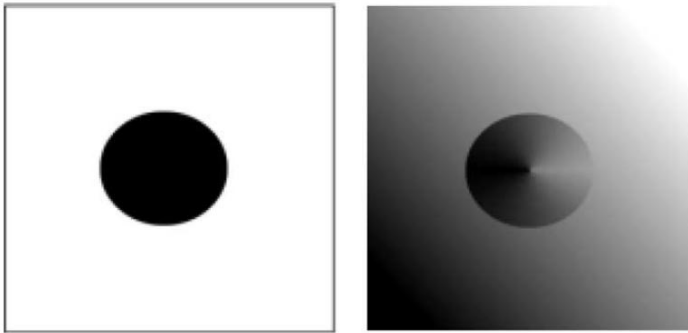
- Related to entropy of difference image

$$H(X - Y)$$

Buzug et al.: *Image registration: convex weighting functions*, CVRMed, 1997

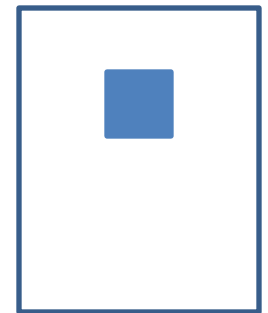
Myronenko A., Song X.: *"Intensity-based Image Registration by Minimizing Residual Complexity"*, IEEE Trans. on Medical Imaging, 2010

Local and Global Statistics



NMI_g

$NMI_l(x)$



$$NMI_{l+g} = \frac{1}{|\Omega|} \sum_{x \in \Omega} (1 - \alpha_x) \cdot NMI_g + \alpha_x \cdot NMI_l(x)$$

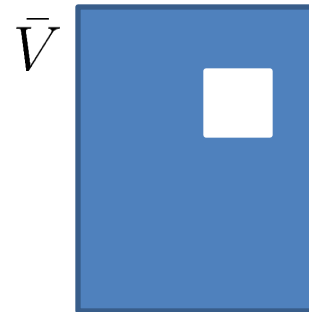
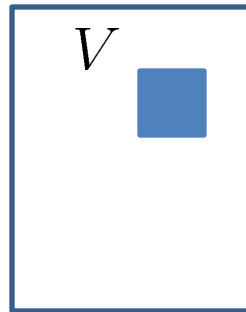
$$\alpha_x = NMI_l(x) - 1$$

Z. Yi and S. Soatto. *Nonrigid Registration Combining Global and Local Statistics*, CVPR, 2009.

Local and Global Statistics

- Combining local and global densities

$$P(X, Y) = \alpha \cdot P_V(X, Y) + (1 - \alpha) \cdot P_{\bar{V}}(X, Y)$$



- α : manually, emphasize on local or global influence

Zhuang, X. and Hawkes, D. and Ourselin, *Unifying Encoding of Spatial Information in Mutual Information for Nonrigid Registration*, IPMI, 2009

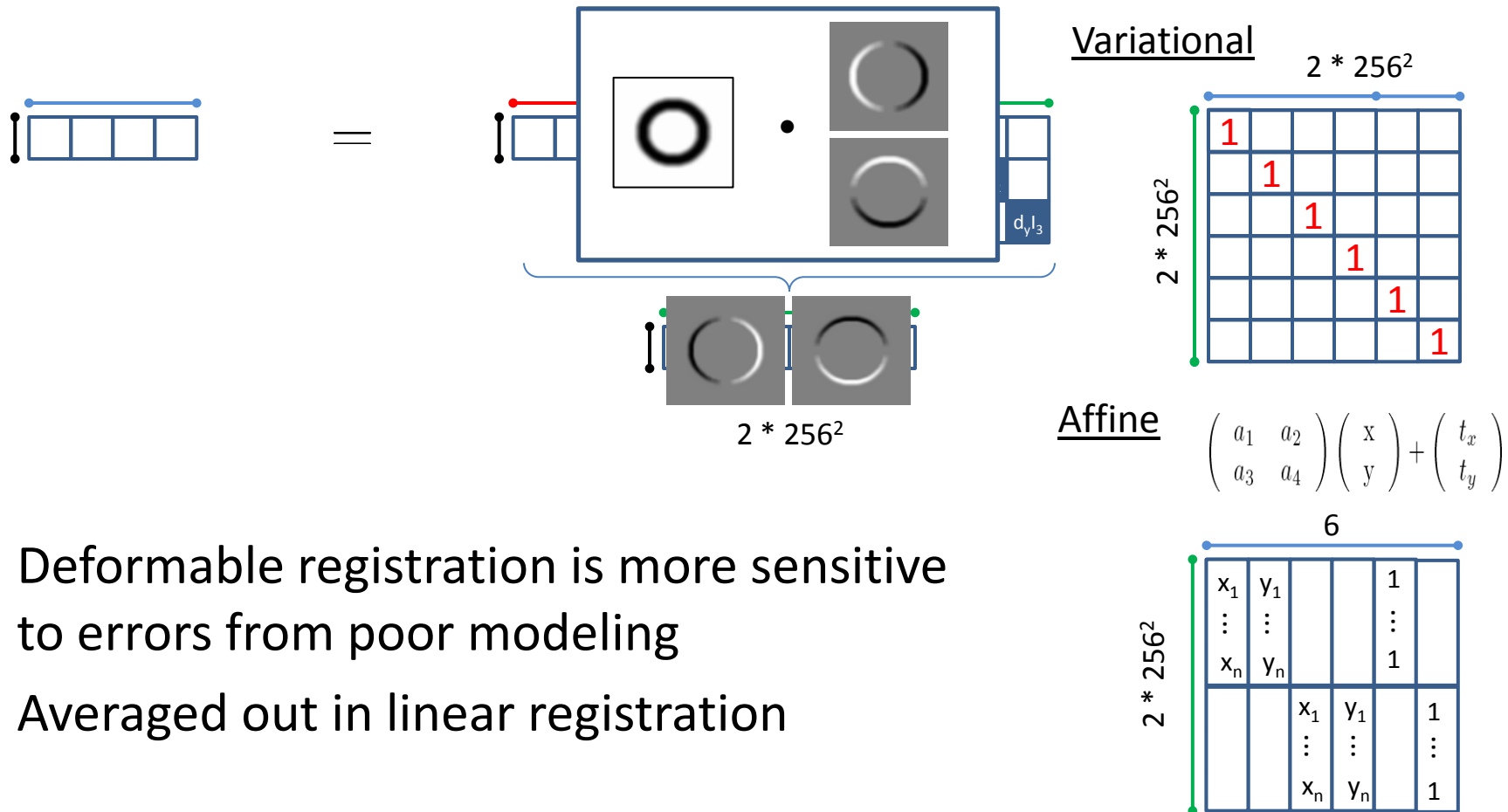
Loeckx, Slagmolen, Maes, Vandermeulen, Suetens, *Nonrigid image registration using conditional mutual information*, IPMI, 2007

PART V

Linear vs. Non-Linear Registration

Linear vs. Non-Linear Registration

$$\frac{\partial \text{SM}(X, Y(T_p))}{\partial p} = \frac{\partial \text{SM}(X, Y)}{\partial Y} \frac{\partial Y}{\partial T_p} \frac{\partial T_p}{\partial p}$$



- Deformable registration is more sensitive to errors from poor modeling
- Averaged out in linear registration

Linear vs. Non-Linear Registration

- Choice of similarity measure has influence on the optimization strategy

EXAMPLE: Gauß-Newton Optimization with SSD

$$\left(\begin{array}{ccc|ccc}
 \nabla I_S^x & \nabla I_S^x & \nabla I_S^x & \Delta & & \\
 \nabla I_S^y & \nabla I_S^x & \nabla I_S^y & & \Delta & \\
 \nabla I_S^z & \nabla I_S^x & \nabla I_S^z & & & \Delta \\
 \hline
 \nabla I_S^x & \nabla I_S^y & \nabla I_S^z & & & \\
 \nabla I_S^y & \nabla I_S^x & \nabla I_S^y & & & \\
 \nabla I_S^z & \nabla I_S^x & \nabla I_S^z & & &
 \end{array} \right) - \lambda \begin{array}{ccc}
 \Delta & & \\
 & \Delta & \\
 & & \Delta
 \end{array} \begin{array}{c} h_x \\ h_y \\ h_z \end{array} = \begin{array}{c} f_x \\ f_y \\ f_z \end{array}$$

$$J_e^\top J_e \quad h = f$$

Linear vs. Non-Linear Registration

- Overlap invariance only for linear ?
 - MI vs. NMI
 - Joint Entropy vs. MI
- Transition between linear and non-linear
 - Parameterization
 - Regularization

Summary

- A large variety of concepts for measuring the similarity is available in the literature
- Choose similarity measure that best fits the application
- Appropriate choice is more important for deformable registration

END