ECSE-626
Statistical Computer Vision

Image Registration Based on Mutual Information

Slides Courtesy of Dr. Dante DeNigris Moreno

Overview

- Define the problem of **image registration**.
- Describe the **family** of image registration **algorithms**
- Explain **mutual information (MI)** and its relevance.
- Describe a common **implementation** of MI-based image registration.

Image Registration

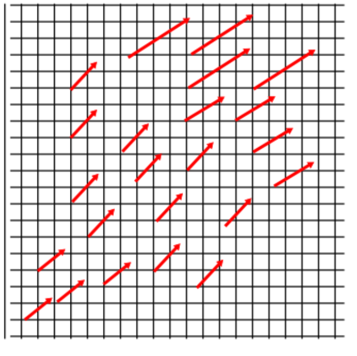
Informal Definition:

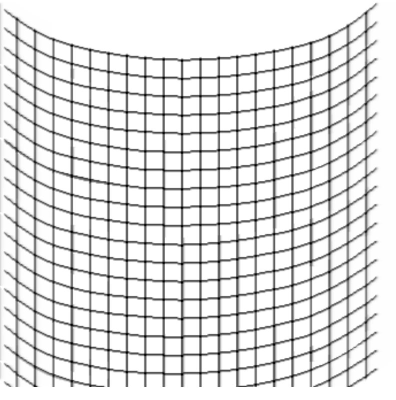
A technique for aligning visual content from two images

Image Registration

Formal Definition:

Find a **mapping** or **transformation**, $T(x)$, that brings a Moving Image, I_m, into alignment with the points of a Fixed Image, I_f.

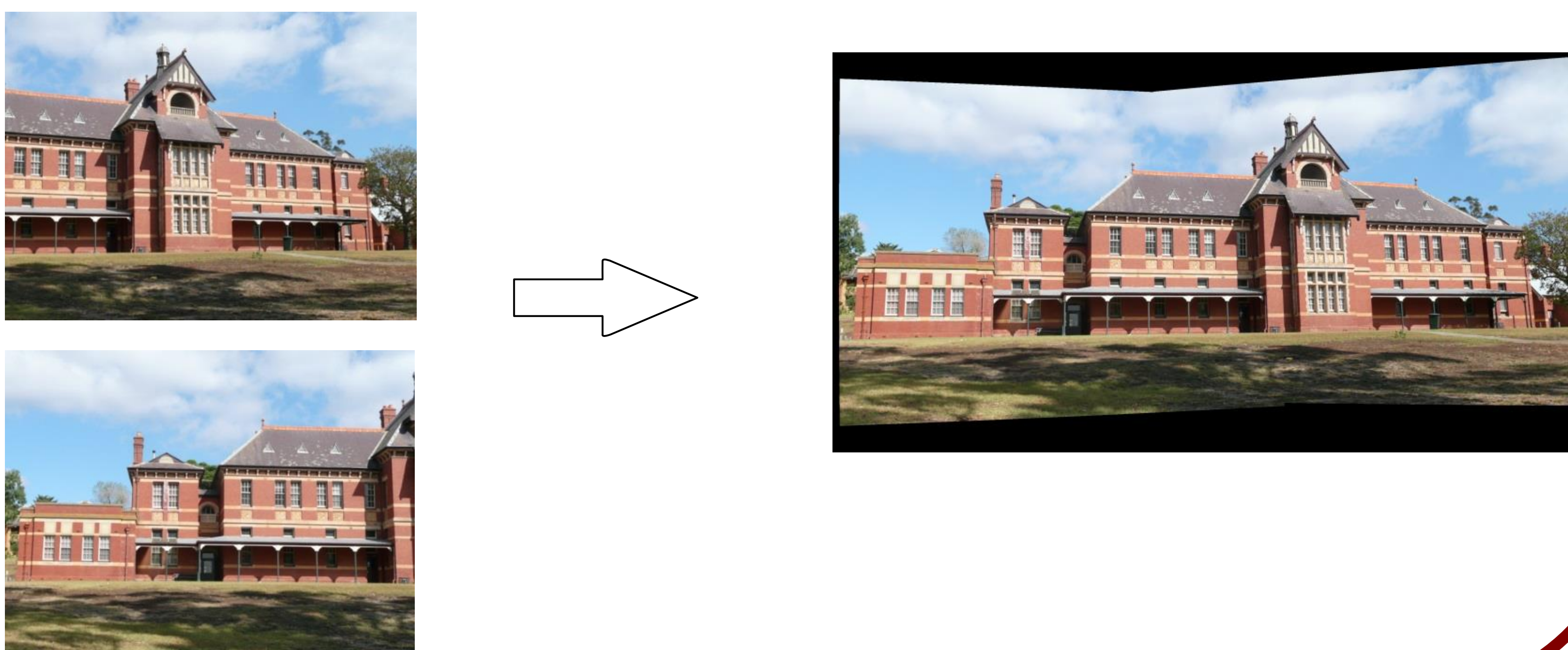




$I_f(x)$ should expose the same physical point as $I_m(T(x))$

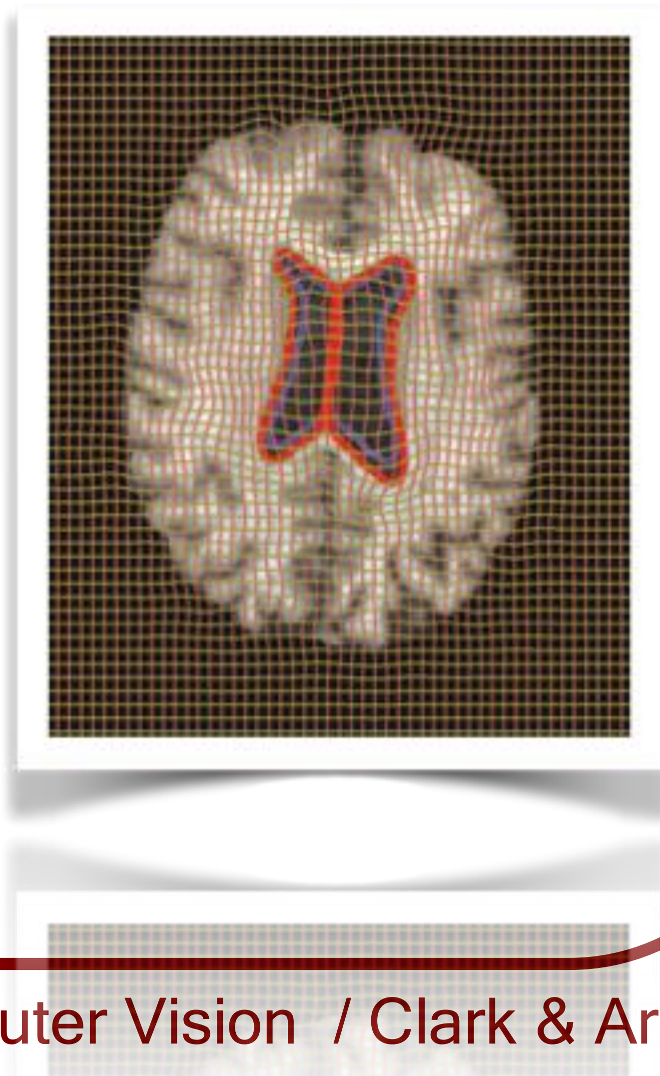
Applications

- **Remote sensing, Consumer Apps**
  - Creating super-resolution images, Cartography updating, Image mosaicing.



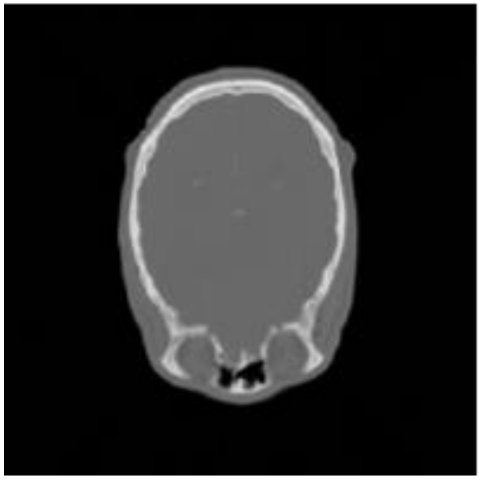
Applications

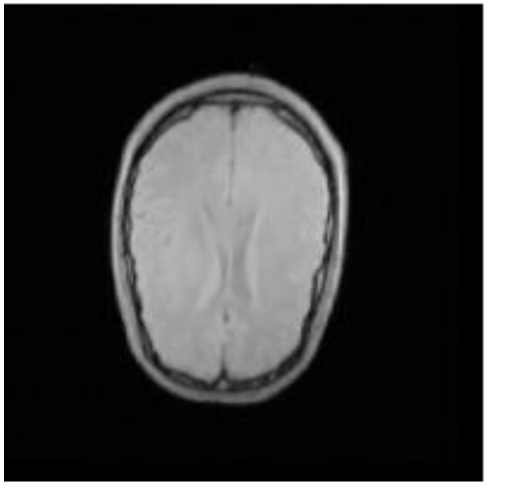
- **Medicine**
  - Registering patient's data to an anatomical atlas



Applications

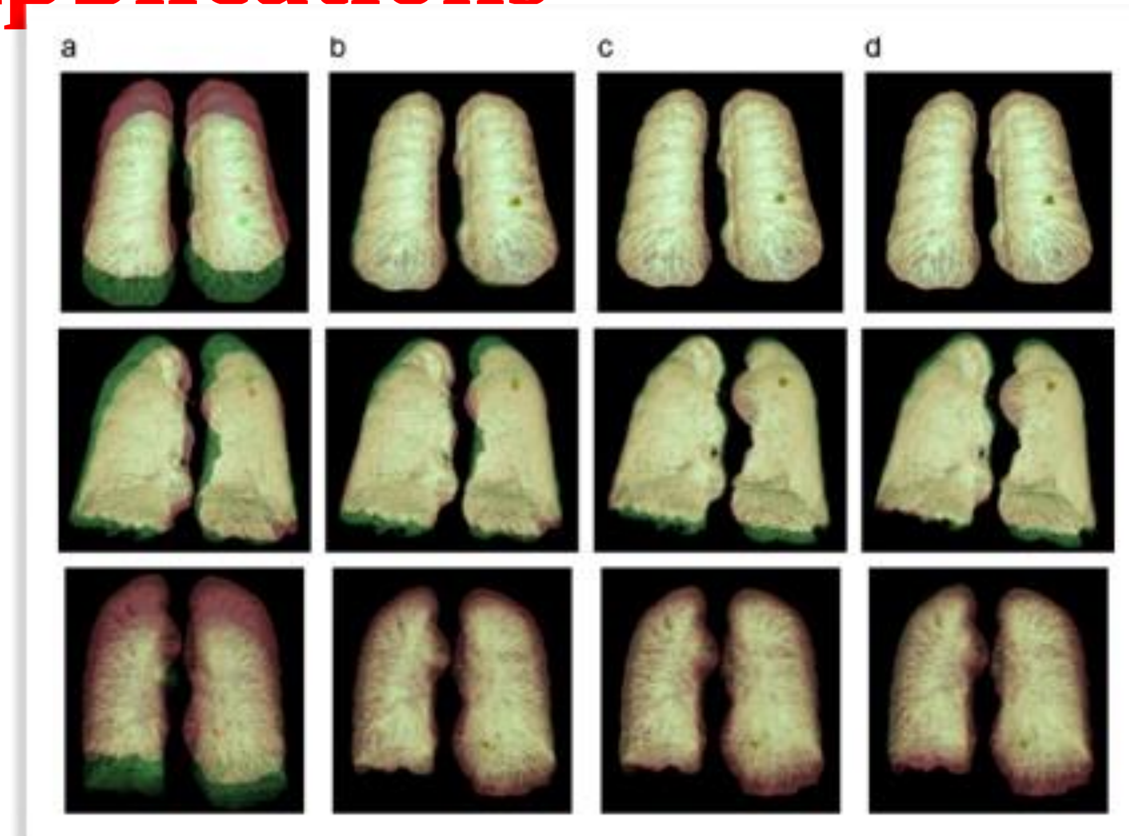
- **Medicine**
  - Registering images from different modalities





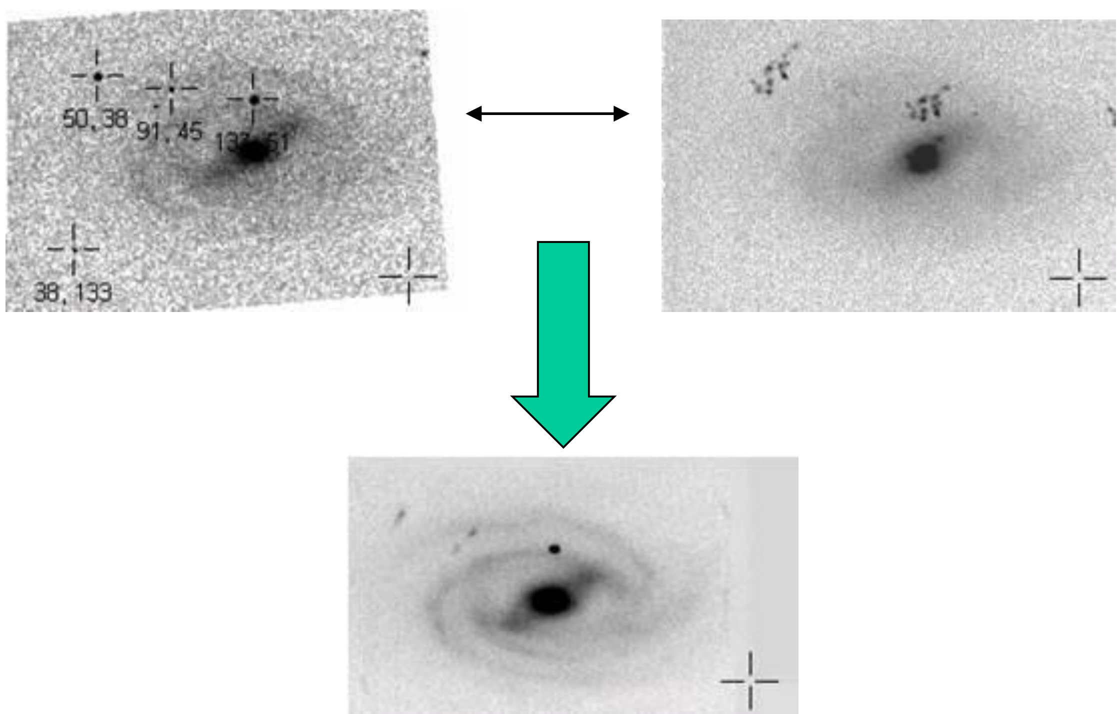
Applications

- **Medicine**
  - Characterizing organ deformation across several timepoints



Applications

- **Astronomy**
  - Aligning images of constellations.



Mathematical Formulation

- Image registration as **maximization** of **image alignment**.

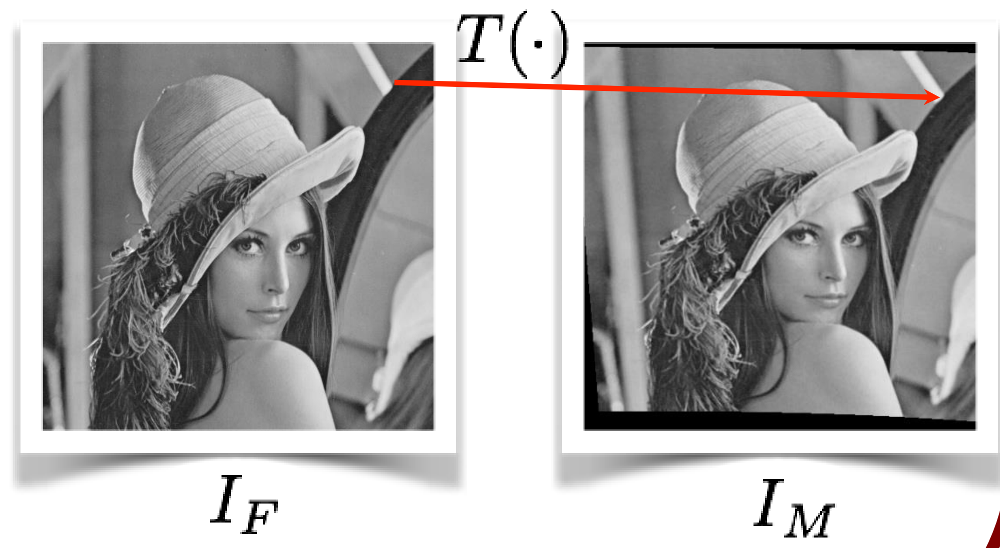
$$\boxed{\mathbf{T}^* = \arg\max_{\mathbf{T}} S\Big(I_F(\mathbf{x}), I_M(\mathbf{T}(\mathbf{x}))\Big)}$$

I_F Fixed (Reference) Image

I_M Moving Image

$S(\cdot)$ Similarity Metric ??

$T(\cdot)$ Transformation Function



Mathematical Formulation

- What is the similarity metric?

$$\mathbf{T}^* = \arg\max_{\mathbf{T}} \boxed{S\Big(I_F(\mathbf{x}), I_M(\mathbf{T}(\mathbf{x}))\Big)}$$

- The ***similarity metric*** is the function that evaluates image alignment.

- How do we define the ***similarity metric***?

  ... varies **largely** with respect to the **type of images** being registered.

Defining Similarity Metric

"Easy" Example

- The case when both images are identical with the true transformation function is applied:

$$I_F(\mathbf{x}) = I_M(\mathbf{T}^*(\mathbf{x}))$$

A reasonable similarity metric would be the **negative sum of squared differences**.

$$S(\cdot) = -\sum_{\mathbf{x}} \left(I_F(\mathbf{x}) - I_M(\mathbf{T}(\mathbf{x}))\right)^2$$

Why?

Defining Transformation

- How do we define the space of plausible transformation functions?

... varies with respect to the **type of context** being registered.

Defining Transformation

- How do we define the space of plausible transformation functions?

... varies with respect to the **type of context** being registered.

“Easy” Examples (in 2D)

-The case when images can be misaligned only by a global horizontal displacement.

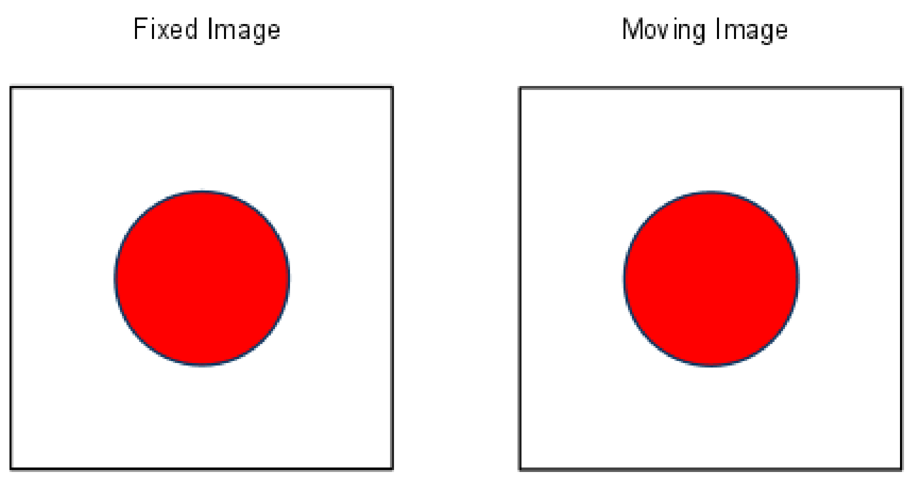
-The case when images can be misaligned only by a global horizontal and vertical displacement.

-The case when images can be misaligned by a global translational and rotational displacement. (i.e. rigid transformation).

Ill-Posedness of Image Registration

- Many image registration contexts lead to ill-posedness: Different ransformation functions yield a maximal similarity metric and an optimal alignment.

Example: *What is the optimal rotation?*



Family of Image Registration Algorithms

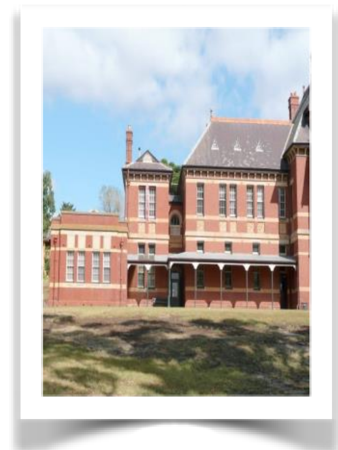
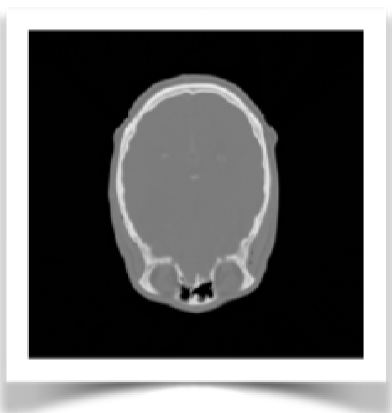
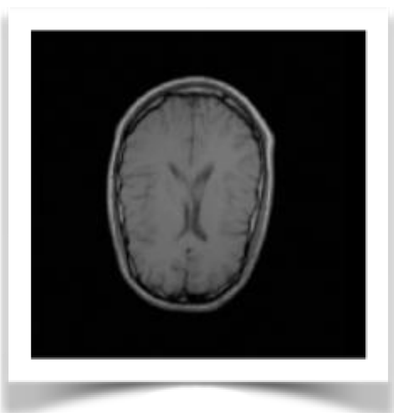
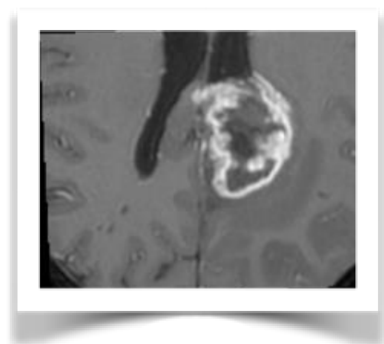
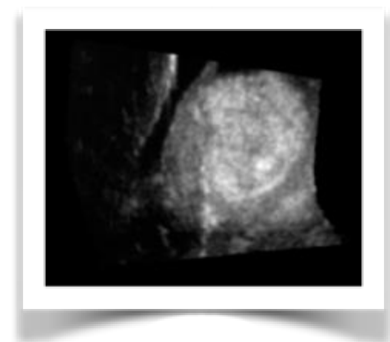
- **Type of Input Images**
  - Image Devices; Number and Dimensions of Images.
- **Transformation Model**
  - Rigid or Non-Rigid Deformation.
- **Similarity Metric.**
  - Based on Image Intensity; Based on Extracted Features; Mono-modal or Multi-modal.
- **Optimization Strategy**
  - Global or Local.

Family of Image Registration Algorithms

- **Type of Input Images**
  - Image Devices; Number and Dimensions of Images.
- **Transformation Model**
  - Rigid or Non-Rigid Deformation.
- **Similarity Metric.**
  - Based on Image Intensity; Based on Extracted Features; Mono-modal or Multi-modal.
- **Optimization Strategy**
  - Global or Local.

Types of Input Images



Family of Image Registration Algorithms

- **Type of Input Images**
  - Image Devices; Number and Dimensions of Images.
- **Transformation Model**
  - Rigid or Non-Rigid Deformation.
- **Similarity Metric.**
  - Based on Image Intensity; Based on Extracted Features; Mono-modal or Multi-modal.
- **Optimization Strategy**
  - Global or Local.

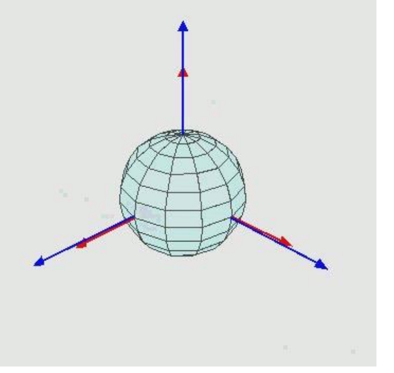
Transformation Model

- Parametric Mapping is characterized by a vector of parameters.
  - Rigid (Translation, Rotational)
  - Affine (Translation, Rotation, Shear, Scale)
  - Free-Form Deformation based on B-Splines, etc..
- Non-Parametric A dense vector field of displacements.
  - Fluid Registration
  - Elastic Registration
  - Optical Flow

Transformation Model - Examples

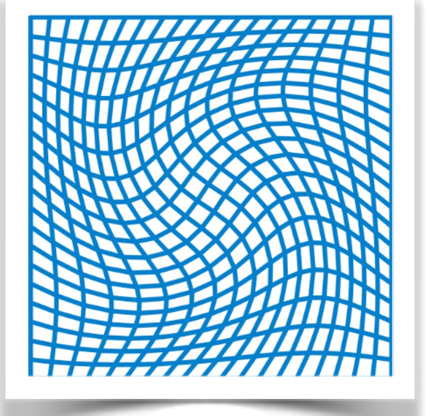
- Rigid

$$\mathbf{T}(\mathbf{x}) = \mathbf{R} \cdot \mathbf{x} + \mathbf{\Delta x}$$



3D Rotation

- Non-Rigid Deformation



Transformation Model Regularization

- <u>(Hard) Regularization via Simple Transformations</u>
  - Choosing a transformation model of **reduced complexity** is a **strong prior**.
  - Hence, we can think of it as a regularization strategy that enforces well-posedness.

Transformation Model Regularization

- <u>(Soft) Regularization via Penalties</u>
  - Regularization can also be embedded in the form of a **penalty** (e.g. bending energy penalty, displacement penalty).

$$\mathbf{T}^* = \arg\max_{\mathbf{T}} S(\cdot) - \lambda \boxed{R(\mathbf{T}(\cdot))}$$

Family of Image Registration Algorithms

- **Type of Input Images**
  - Image Devices; Number and Dimensions of Images.
- **Transformation Model**
  - Rigid or Non-Rigid Deformation.
- **Similarity Metric.**
  - Based on Image Intensity; Based on Extracted Features; Mono-modal or Multi-modal.
- **Optimization Strategy**
  - Global or Local.

Similarity Metric

Pixel-Intensity Based

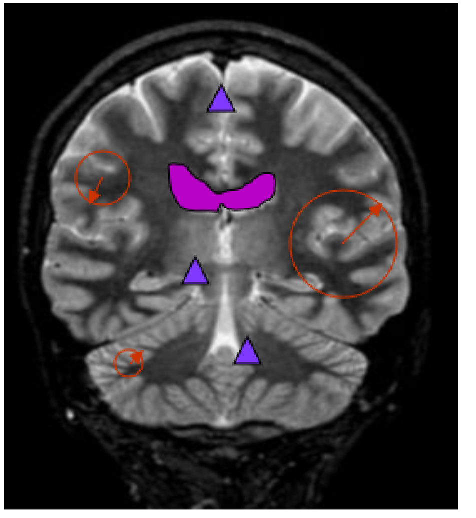
- Sum of Squared Differences $SSD = -\sum_x (I_f(x) - I_m(T(x))^2$
- Normalized Cross-Correlation $NCC = \sum_x (I_f(x) - \overline{I_f})(I_m(T(x)) - \overline{I_m})$
- Mutual Information $MI = H(I_f) + H(I_m) - H(I_f, I_m)$

Similarity Metric

Feature Based

Identify (segment) features of interest first and then match.

- Manual Landmarks
- Segment contours/structures of interest.
- "General" Image Features (e.g. edges, corners, SIFT).



Family of Image Registration Algorithms

- **Type of Input Images**
  - Image Devices; Number and Dimensions of Images.
- **Transformation Model**
  - Rigid or Non-Rigid Deformation.
- **Similarity Metric.**
  - Based on Image Intensity; Based on Extracted Features; Mono-modal or Multi-modal.
- **Optimization Strategy**
  - Global or Local.

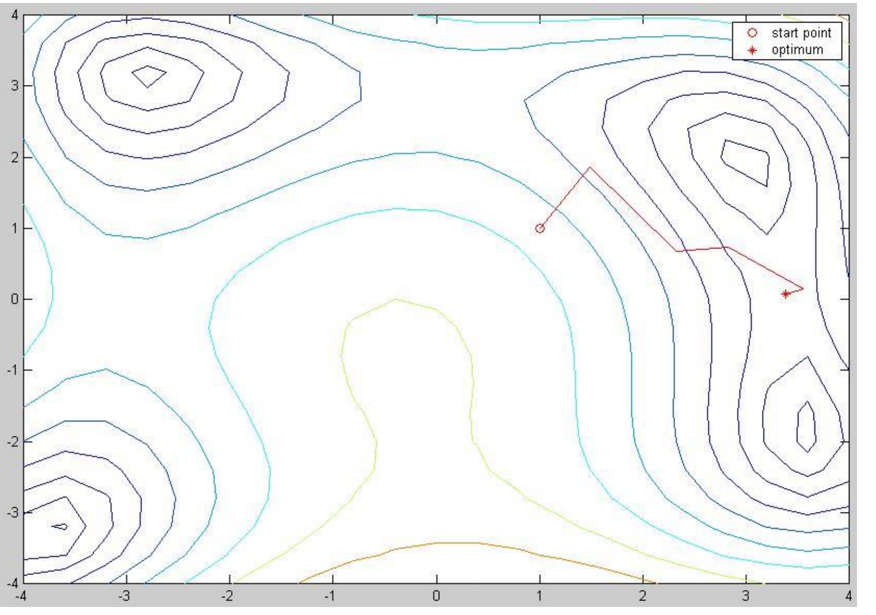
Optimization Method

Global Optimization

- Genetic Algorithm
- Gibbs Sampling
- Simulated Annealing

Local Optimization

- Gradient Ascent
- Quasi-Newton Methods
- Simplex



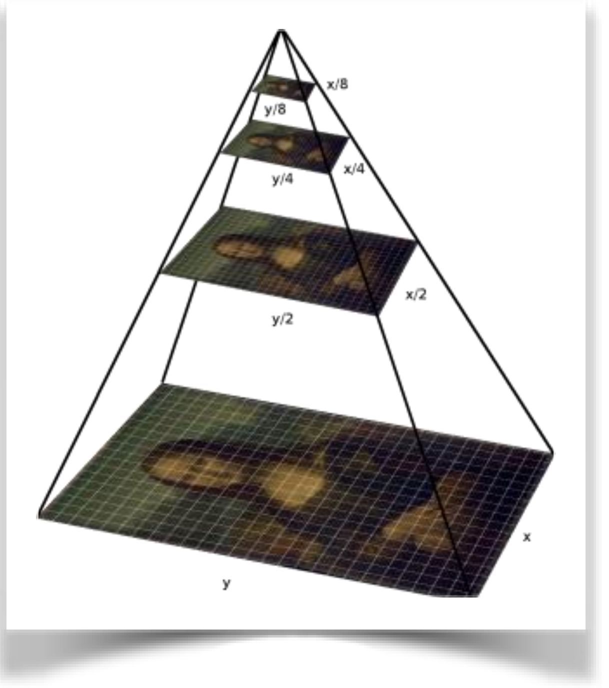
Practical Note

The choice of optimization method typically seeks a balance between <u>accuracy</u> and <u>computational efficiency</u>.

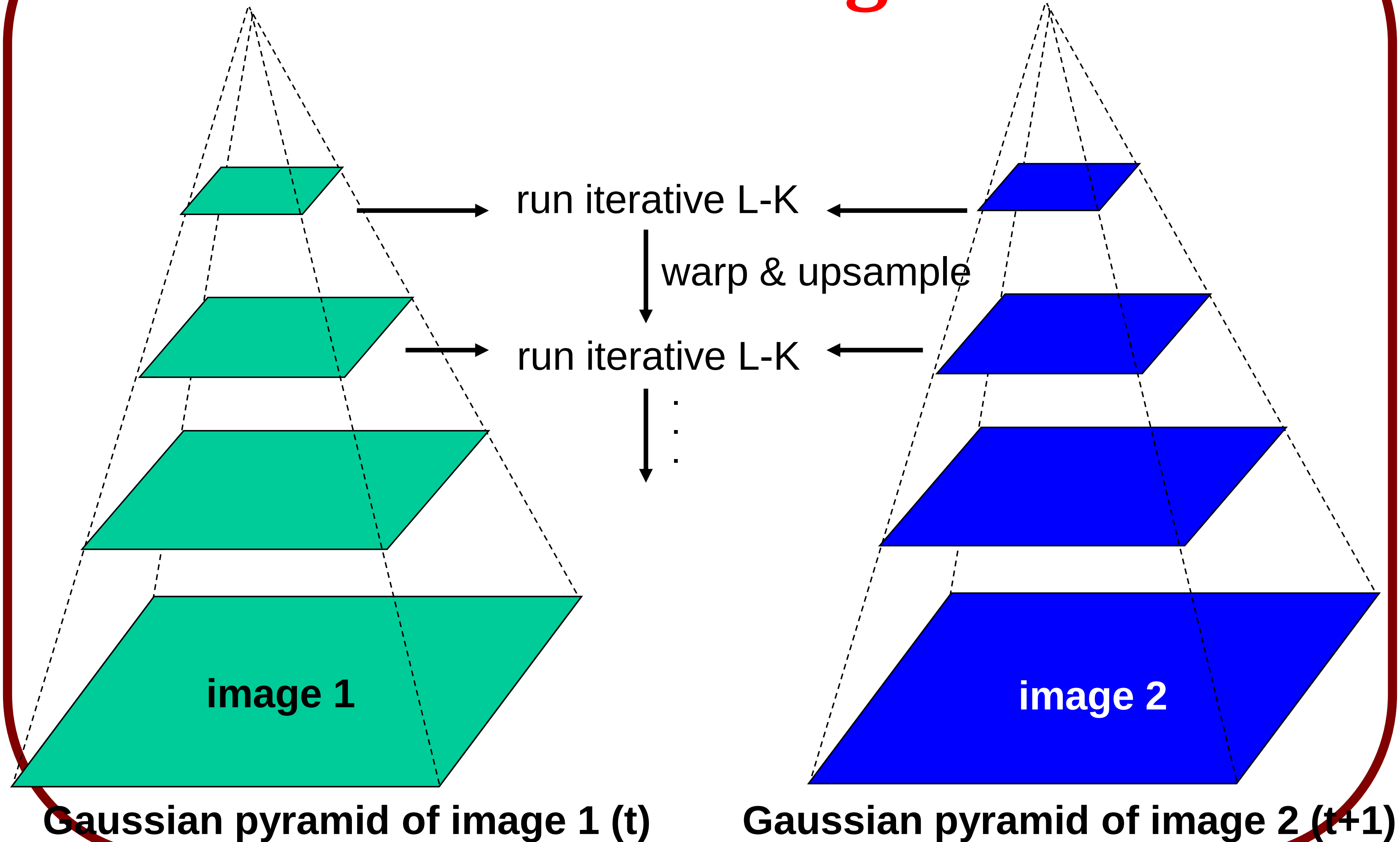
Optimization Method

Practical Note:

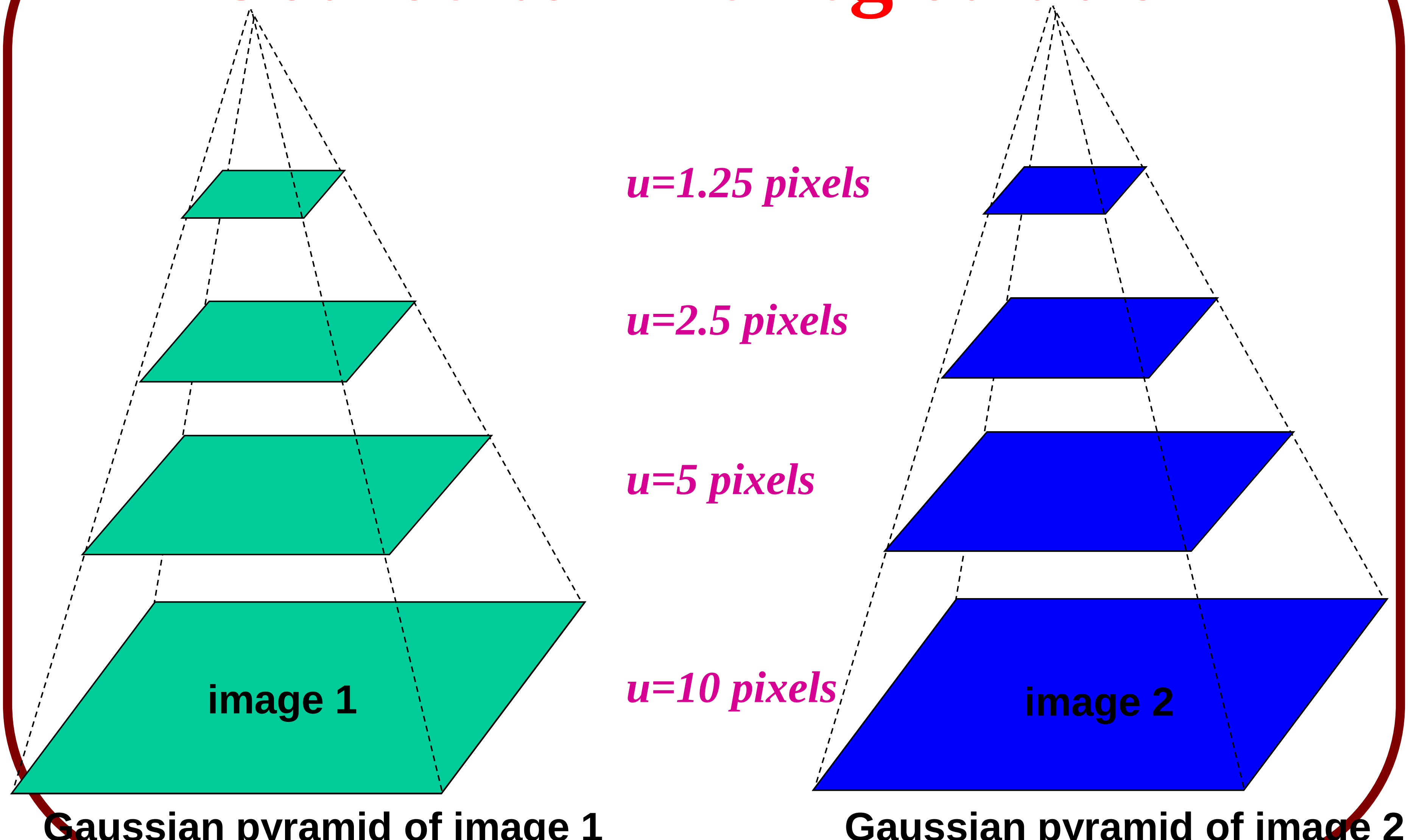
- Many image registration optimization strategies rely on <u>multi-scale image pyramids</u> for improved performance



Coarse-to-fine registration

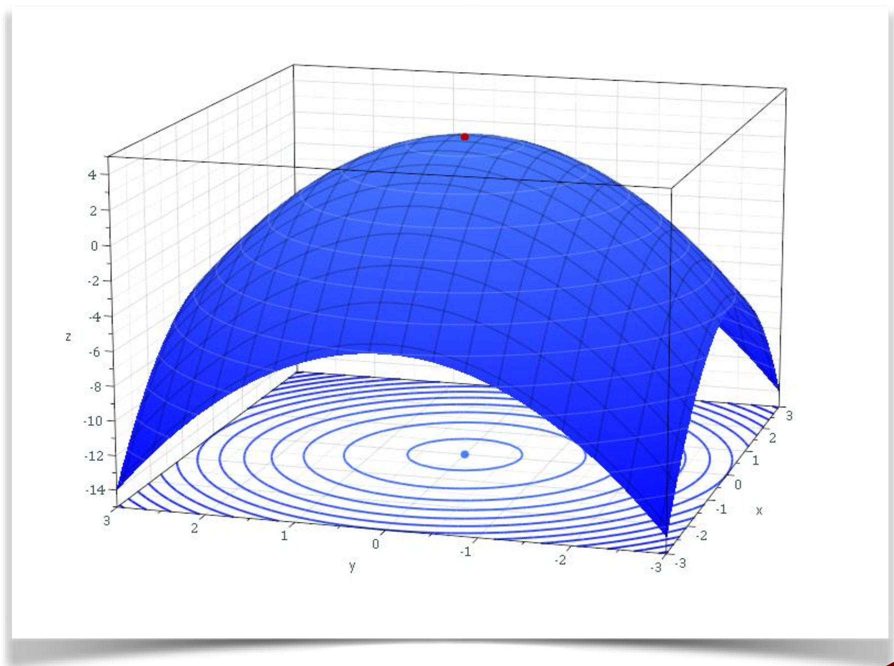


Coarse-to-fine registration



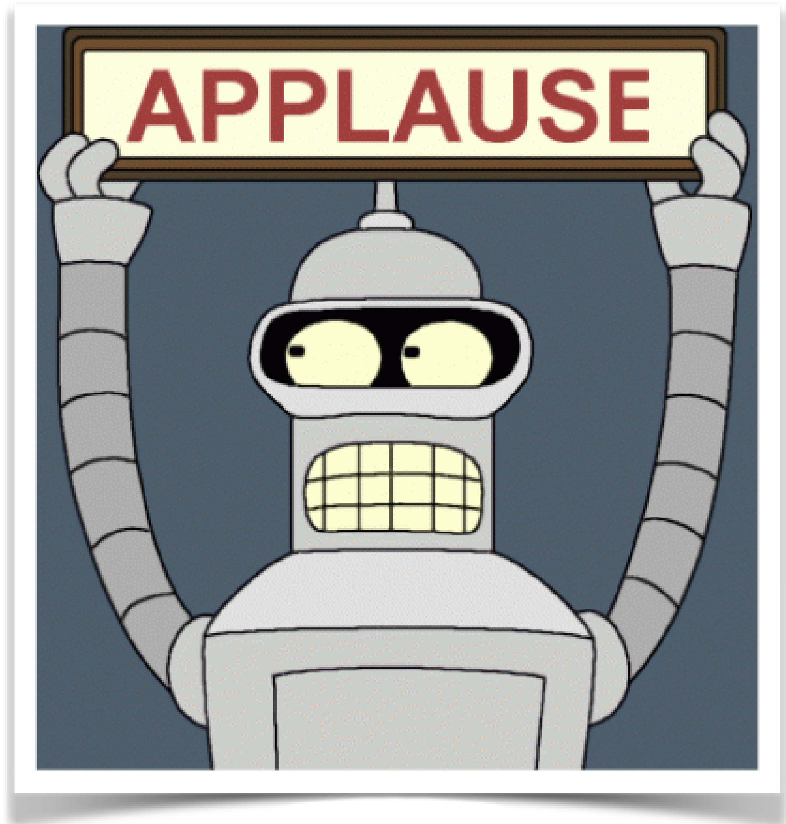
Optimization Method

Optimization is a BIG research field by itself and heavily used in Computer Vision problems.



And that's Image Registration 101

Now, let's go deeper into this Mutual Information stuff.



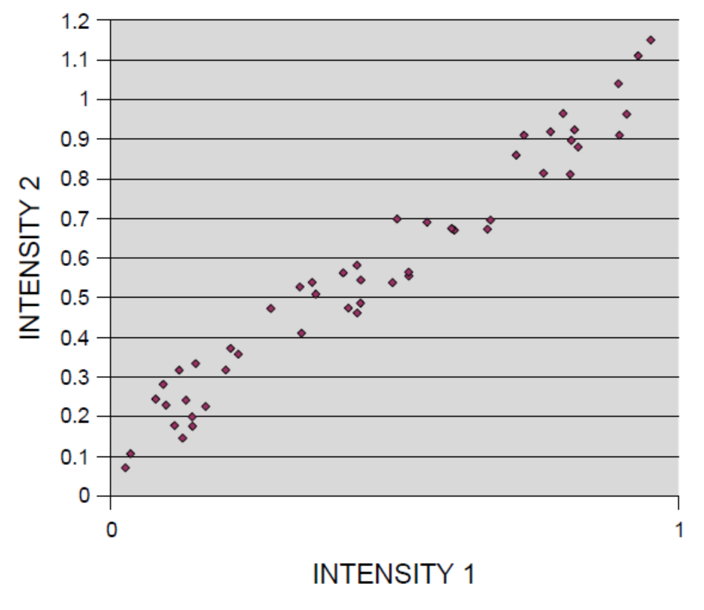
Intensity-based Similarity Metrics

- **Negative Sum of Squared Differences (SSD)**

$$S(\cdot) = -\sum_{\mathbf{x}} \Big(I_F(\mathbf{x}) - I_M(\mathbf{T}(\mathbf{x})) \Big)^2$$

  - Assumes a <u>direct</u> pixel-intensity correspondence
  - When aligned:

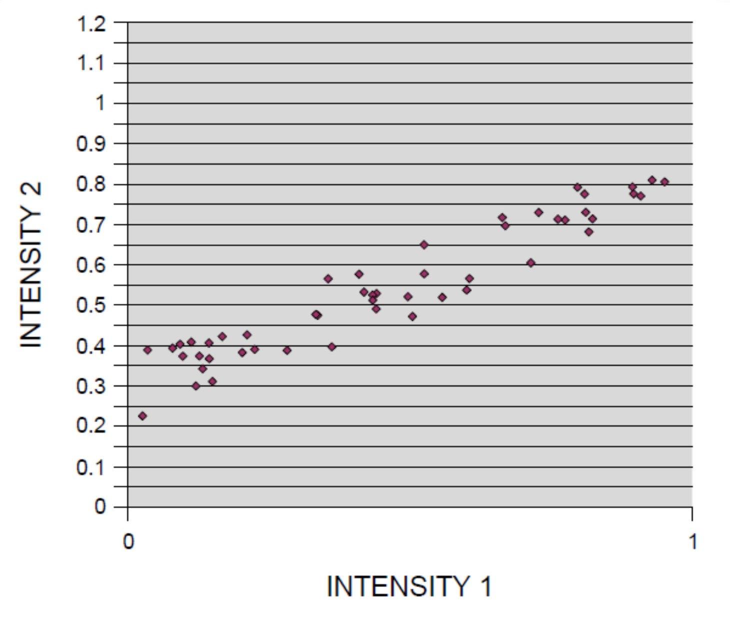
$$I_F(\mathbf{x}) = I_M(\mathbf{T}^*(\mathbf{x}))$$



Intensity-based Similarity Metrics

- **Normalized Cross-Correlation**
  - Assumes a linear relationship between pixel-intensities.
  - When aligned:

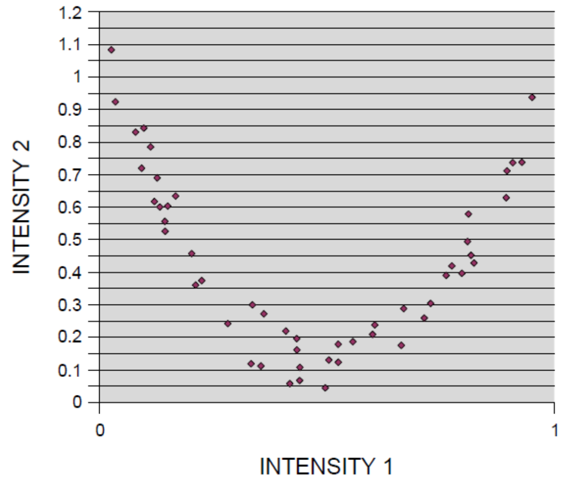
$$I_F(\mathbf{x}) = \alpha \cdot I_M(\mathbf{T}^*(\mathbf{x})) + \beta$$



Intensity-based Similarity Metrics

- **Mutual Information**
  - No <u>hard</u> assumptions! Simply evaluates statistical image-intensity correspondence.
  - Assumes that statistical image intensity correspondence is maximal when images are aligned.
  - Defined as: $MI(\cdot) = H(I_F) + H(I_M) - H(I_F, I_M)$

H( ) : Entropy Function

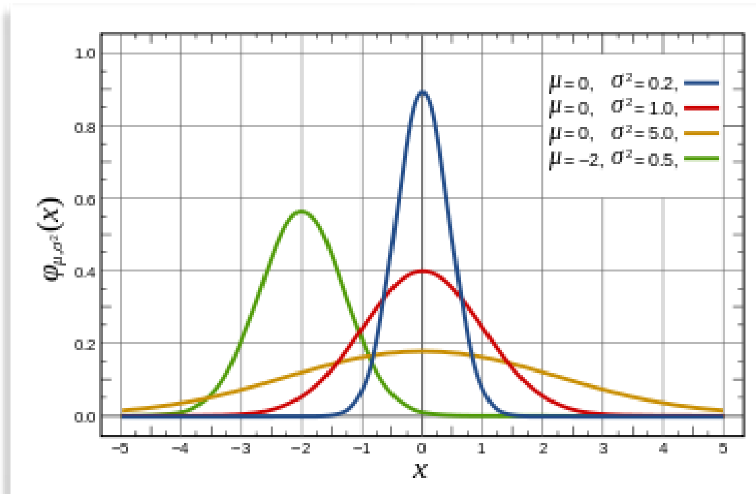


Entropy - recall

- Entropy is a **measure** of the **uncertainty** in a **random variable**.

$$H(X) = -\sum_i p(x_i)\log(p(x_i))$$

$$H(X) = -\int_x p(x)\log(p(x))$$

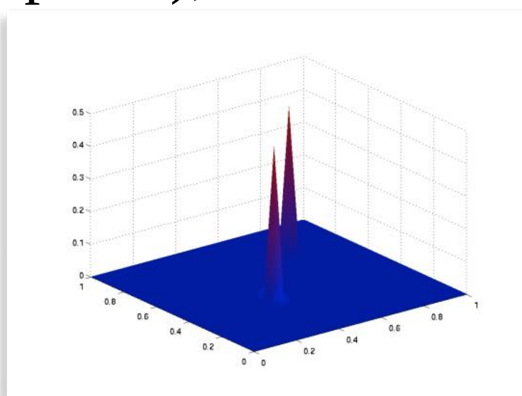


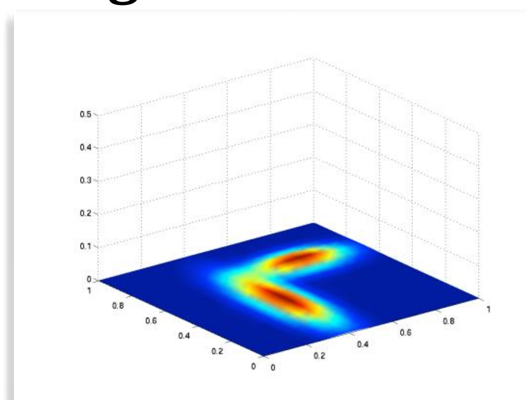
Which curve has highest entropy? Lowest?

Mutual Information

$$MI(\cdot) = \underbrace{\boxed{H(I_F) + H(I_M)}}_{\text{marginal entropies}} - \underbrace{\boxed{H(I_F, I_M)}}_{\text{joint entropy}}$$

- **Joint Entropy**
  - Measures the “spread” of joint pixel intensity correspondence.
  - If the joint density, $p(I_F, I_M)$, has **low entropy** (e.g. composed of a few peaks), it reflects a **strong statistical correspondence**.

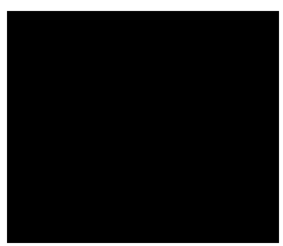




Mutual Information

$$MI(\cdot) = \underbrace{\boxed{H(I_F) + H(I_M)}}_{\text{marginal entropies}} - \underbrace{\boxed{H(I_F, I_M)}}_{\text{joint entropy}}$$

- **Marginal Entropies**
  - Measure the "complexity" of each image.
  - Hence, it penalizes transformations that lead to a reduction in the complexity of any image.



No complexity
H(I) = 0

Maximal complexity

Mutual Information

$$MI(\cdot) = \underbrace{H(I_F) + H(I_M)}_{\text{marginal entropies}} - \underbrace{H(I_F, I_M)}_{\text{joint entropy}}$$

- Mutual Information evaluates the statistical correspondence of joint pixel intensities plus the complexity of each image.
- Why do we need both these terms? Why not just minimize joint entropy?

Mutual Information

$$MI(\cdot) = \underbrace{H(I_F) + H(I_M)}_{\text{marginal entropies}} - \underbrace{H(I_F, I_M)}_{\text{joint entropy}}$$

- By finding a balance between relatively large entropy for the two images separately and small joint entropy, mutual information ensures that the region of overlap (for which the measure is computed) contains (most of) the information in the images.
- Otherwise, a transformation could be favoured that forces the images so far apart that only background is contained in the region of overlap, since this gives minimum joint entropy.

Mutual Information - Implementation

$$MI(\cdot) = H(I_F) + H(I_M) - \boxed{H(I_F, I_M)}$$

joint entropy

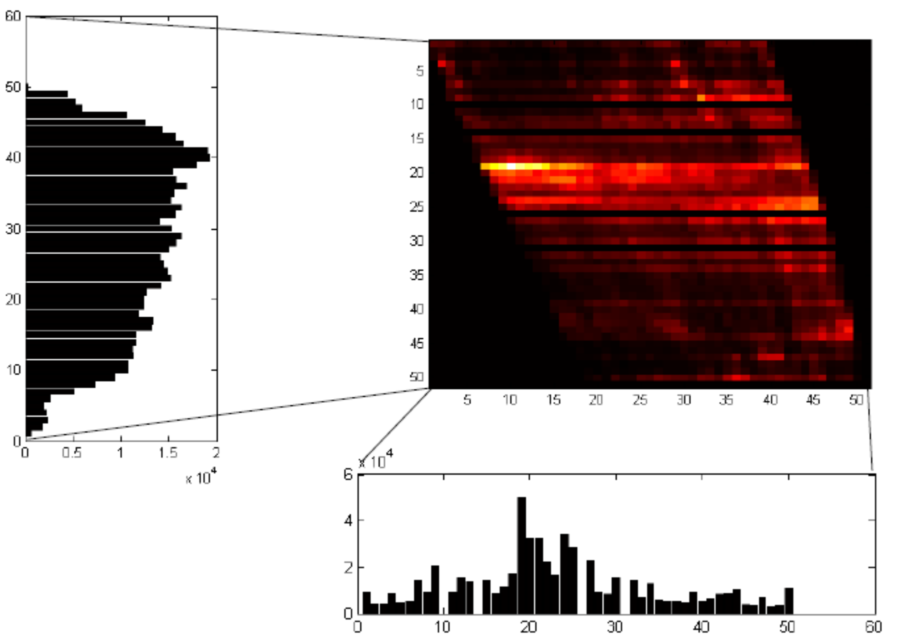
- **Estimating the Joint Density** $p(I_F, I_M)$
  - Estimating the joint pixel intensity density is the <u>main implementation challenge</u> for MI
  - Many different approaches have been proposed.

Mutual Information - Implementation

- **Estimating the Joint Density** $p(I_F, I_M)$
  - Normalized Joint Histogram
    - Very simple,
    - Sensitive to bin selection,
    - Limitations in accuracy and robustness
  - Parzed-Window Density Estimation
    - Very popular,
    - Increased computational cost,
    - Leads to a continuous joint density and derivative.
  - And many many others.....

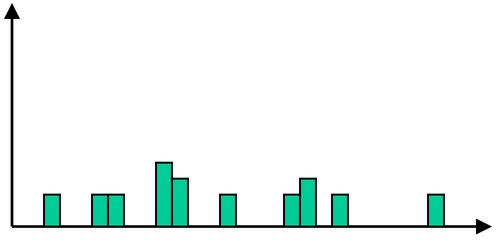
Joint Density Estimation

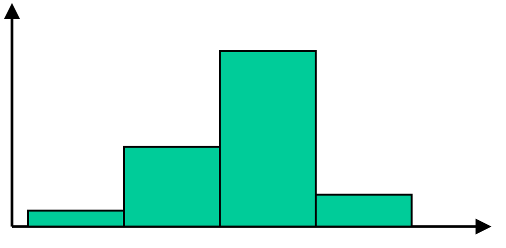
- Normalized Joint Histogram
  - Divide the joint image intensity range in bins and count the number of occurrences.

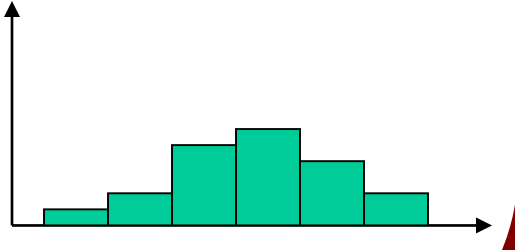


Joint Density Estimation

- Normalized Joint Histogram
  - Divide the joint image intensity range in bins and count the number of occurrences.
    - *How do we select the number and width of bins*?
      - Using "many" bins increases the resolution of the histogram, but will reduce its estimation accuracy since there are less samples per bin.







Joint Density Estimation

- Normalized Joint Histogram
  - Divide the joint image intensity range in bins and count the number of occurrences.
    - *How do we select the number and width of bins*?
      - Using "many" bins increases the resolution of the histogram, but will reduce its estimation accuracy since there are less samples per bin.
    - *How can we construct "smooth" histograms?*
      - Blurring the histogram, creating more samples by spatially-interpolating new samples, ...

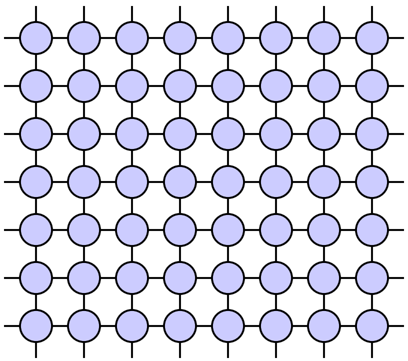
Joint Density Estimation

- Normalized Joint Histogram
  - Divide the joint image intensity range in bins and count the number of occurrences.
    - *How do we select the number and width of bins*?
      - Using "many" bins increases the resolution of the histogram, but will reduce its estimation accuracy since there are less samples per bin.
    - *How can we construct "smooth" histograms?*
      - Blurring the histogram, creating more samples by spatially-interpolating new samples, ...
    - *How to deal with non-aligned pixel-grids? Interpolation?*

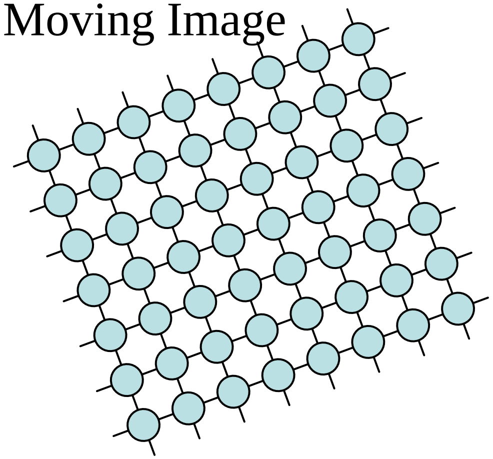
Joint Density Estimation

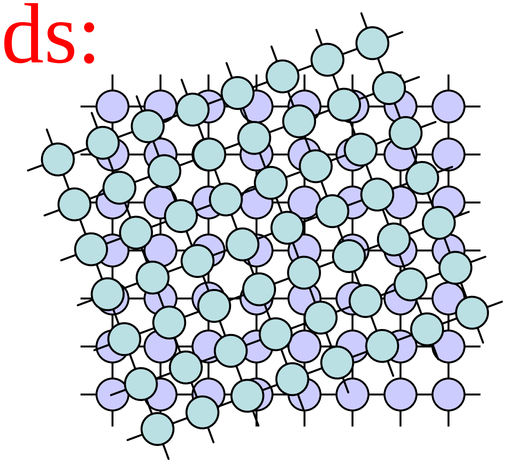
Dealing with non-aligned pixel grids:

Fixed image

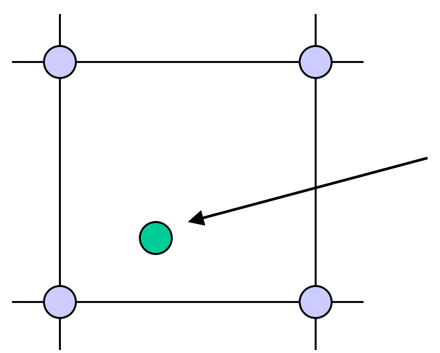


Moving Image





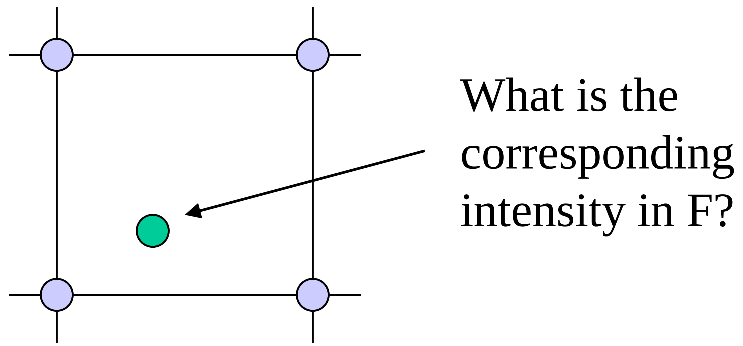
Joint histogram?



What is the
corresponding
intensity in F?

Joint Density Estimation

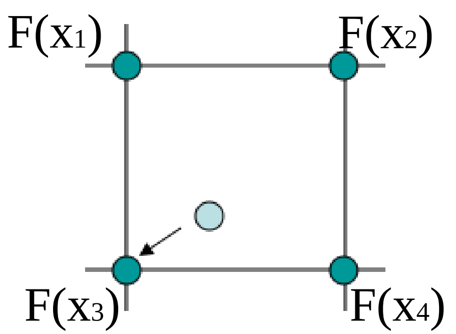
Dealing with non-aligned pixel grids:



- Easy way out is to simply estimate the intensity value by **interpolating** with neighbouring pixels.
- More advanced strategies **populate** the histogram with **weights** based on distance from location of interest.

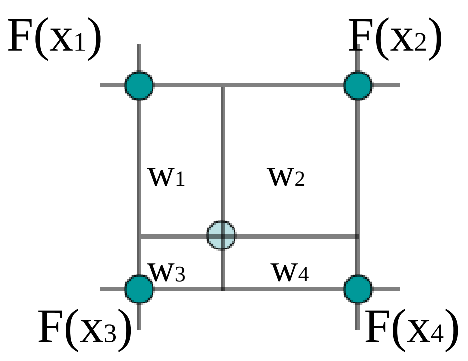
Joint Density Estimation

Dealing with non-aligned pixel grids:



Nearest neighbor interpolation:

Fill the histogram at $\left[F(x_3);M(y)\right]$



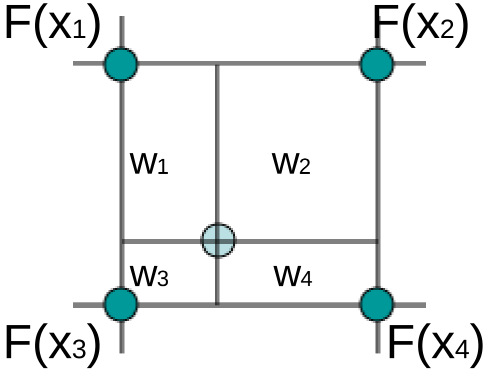
Trilinear interpolation:

Fill the histogram at

$$\left[w_1F(x_1) + w_2F(x_2) + w_3F(x_3) + w_4F(x_4);\ M(y)\right]$$

Joint Density Estimation

Dealing with non-aligned pixel grids:



Partial Volume Interpolation

Fill the histogram at: with:

$\left[F(x_1); M(y)\right] = w_1$

$\left[F(x_2); M(y)\right] = w_2$

$\left[F(x_3); M(y)\right] = w_3$

$\left[F(x_4); M(y)\right] = w_4$

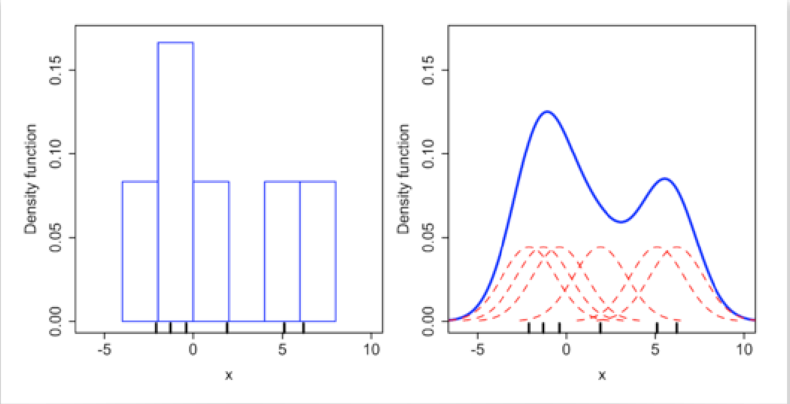
Mutual Information - Viola/Wells

- Same idea proposed independently by two groups.
- We will briefly review the implementation proposed by Viola and Wells.
  - Alignment by Maximization of Mutual Information. Paul Viola and William M. Wells III
  - Multi-modal volume registration by maximization of mutual information. William M. Wells III, Paul Viola, Hideki Atsumid, Shin Nakajimae and Ron Kikinise
  - Multi-modal image registration by maximization of mutual information. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.

Parzen-Windows Density Estimation

$$p(I_F = a, I_M = b) = \frac{1}{Z}\sum_i K(I_F(\mathbf{x}_i) - a, I_M(\mathbf{x}_i) - b)$$

- The joint density is estimated as a sum of displaced kernel functions (e.g. Gaussian function).
- Each kernel, **K**, corresponds to a sample and its centre is displaced with the value of the sample.



Note: Z is a normalization constant.

Mutual Information - Viola/Wells Paper

Algorithmic Outline

- For every iteration
  - 1. Randomly pick N_A pixel-pair samples to characterize the joint density.
  - 2. Randomly pick N_B pixel-pair samples to evaluate entropies and corresponding gradients (with respect to transformation parameters).
  - 3. Update transformation parameters based on gradient ascent strategy.

Mutual Information - Viola/Wells Paper

Algorithmic Outline

- For every iteration
  - 1. Randomly pick N_A pixel-pair samples to characterize the joint density.

$$p(I_F = a, I_M = b) = \frac{1}{Z} \sum_{i}^{N_A} K(I_F(\mathbf{x}_i) - a, I_M(\mathbf{T}(\mathbf{x}_i)) - b)$$

where K() represents the Parzen-Window Kernel (e.g. a Gaussian kernel or a cubic B-Spline kernel).

Mutual Information - Viola/Wells Paper

Algorithmic Outline

- For every iteration
  - 2. Randomly pick N_B pixel-pair samples to evaluate entropies and corresponding gradients (with respect to transformation parameters).

$$H(I_F, I_M) = -\frac{1}{N_B}\sum_j N_B \log\left(p(I_F(\mathbf{x_j}), I_M(\mathbf{T}(\mathbf{x_j})))\right)$$

$$\frac{\partial MI}{\partial \mathbf{T}} = \frac{\partial H(I_M)}{\partial \mathbf{T}} - \frac{\partial H(I_F, I_M)}{\partial \mathbf{T}}$$

Mutual Information - Viola/Wells Paper

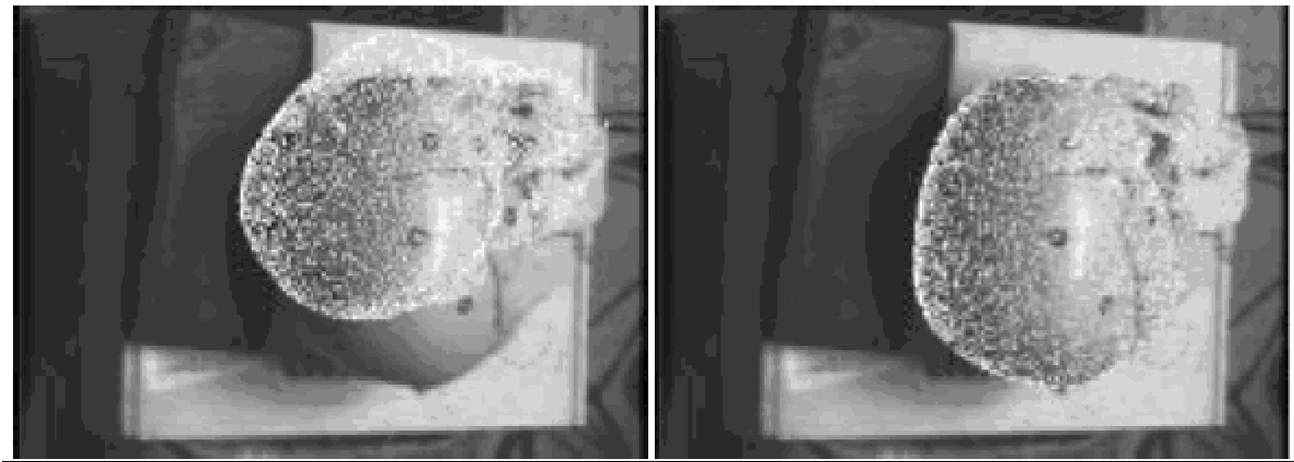
Algorithmic Outline

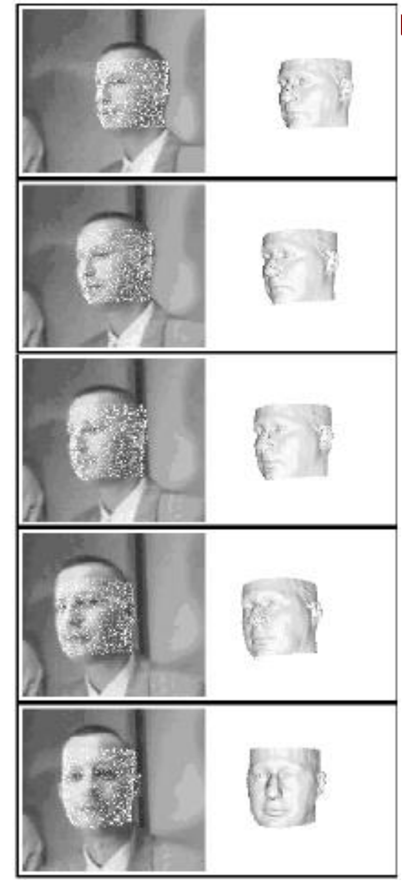
- For every iteration
  - 3. Update transformation parameters based on gradient ascent strategy.

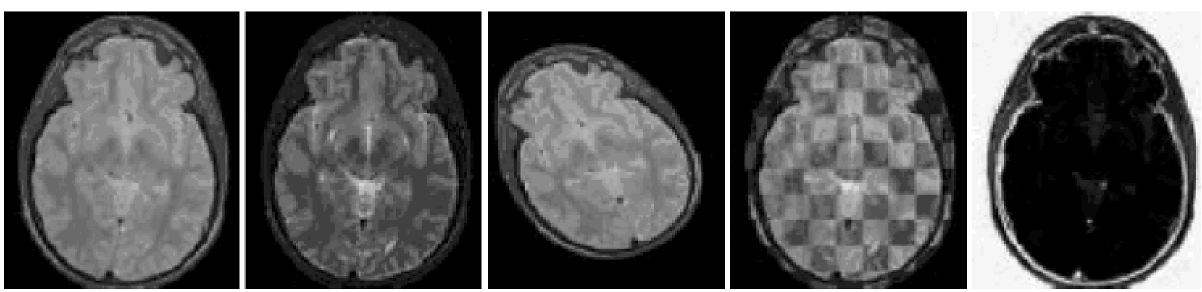
$$\mathbf{T} \leftarrow \mathbf{T} + \lambda \frac{\partial MI}{\partial \mathbf{T}}$$

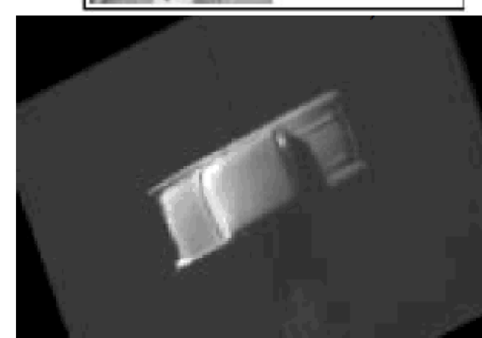
where λ is known as the **learning rate** and is a tuning parameter of the optimization strategy.

Viola and Wells Results



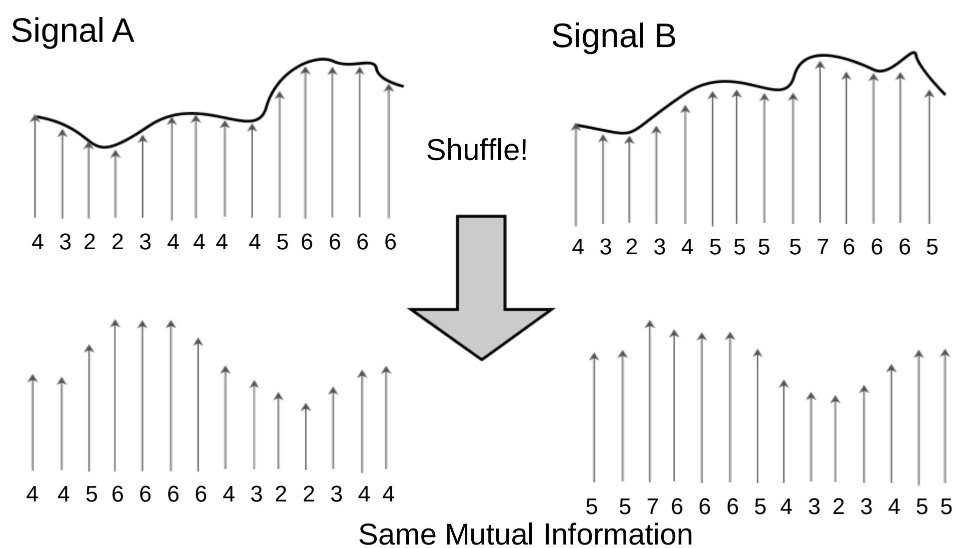






Mutual Information - Limitations

- Mutual Information is a **global** metric.
- In other words, we see the image-pair as a *bag of pixel-pair features*.
- Do you see any limitations at this might present? If you do, how would you try and fix them?



Mutual Information - Limitations

- **No Spatial Context.**
  - Bag of Pixel-Pairs Features.
  - No spatial context ... .(as in what is the voxel value in relation to the neighbourhood voxels)
- **Mutual information is a global measure.**
  - Sensitive to non-homogeneous image intensity response (e.g. intensity attenuation in part of the image).
  - Sensitive to un-corresponded features or occlusion (e.g. blob appears in one modality but not in another).

Mutual Information - Limitations

- Fixing some of these problems:
- Embed spatial information!
    - *Regional Mutual Information*
    - *Local Mutual Information*
    - *Conditional Mutual Information*

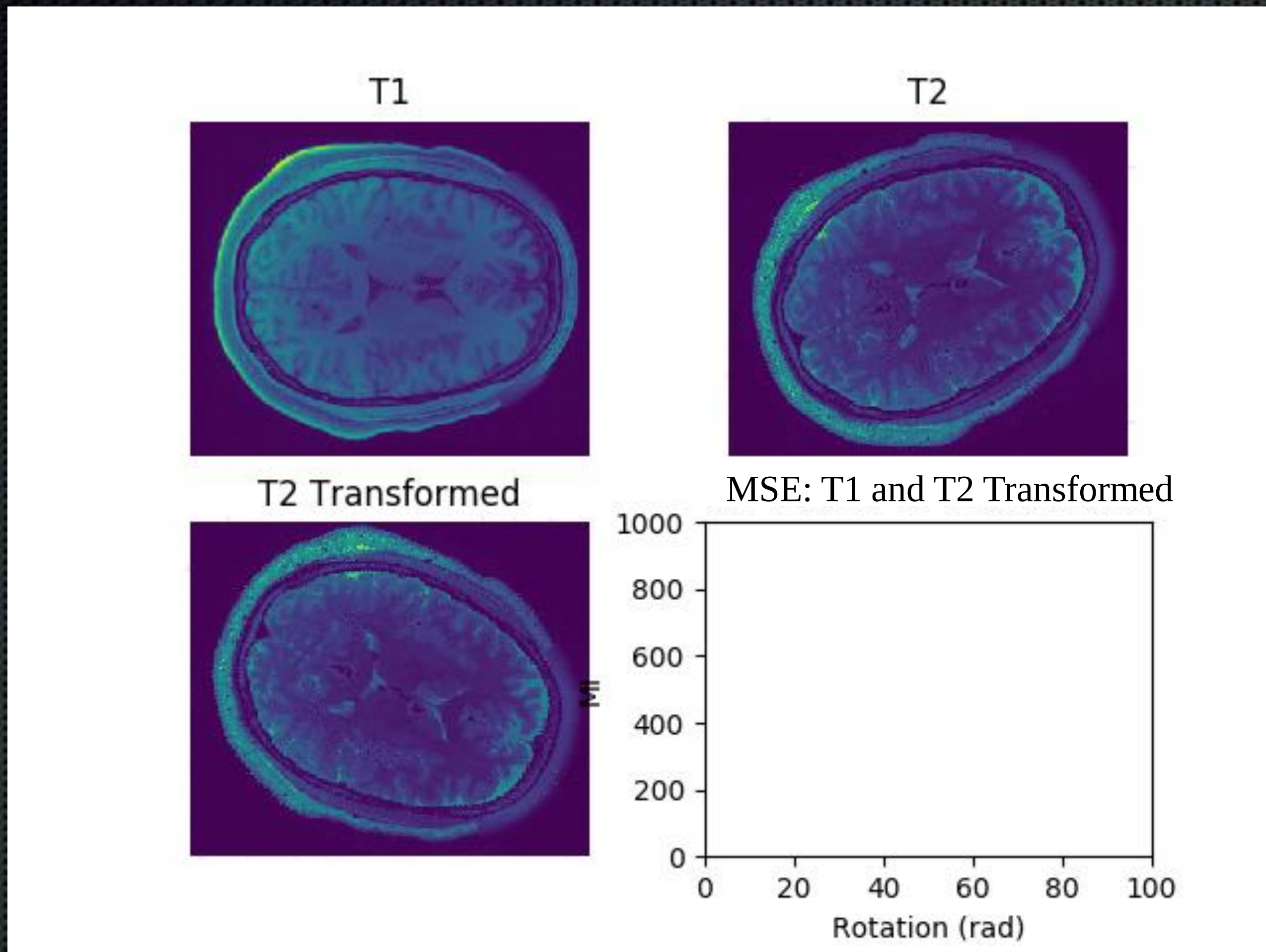
Recap

- Reviewed the general definition of the Image Registration Problem.
- Described the main variants that characterize any particular image registration algorithm:
  - Input Images, Transformation Model, Similarity Metric, Optimization Strategy.
- Reviewed underlying assumptions behind common similarity metric.
- Described the intuition behind using Mutual Information as a similarity metric.
- Explained technical challenges and outlined a popular algorithm.

References

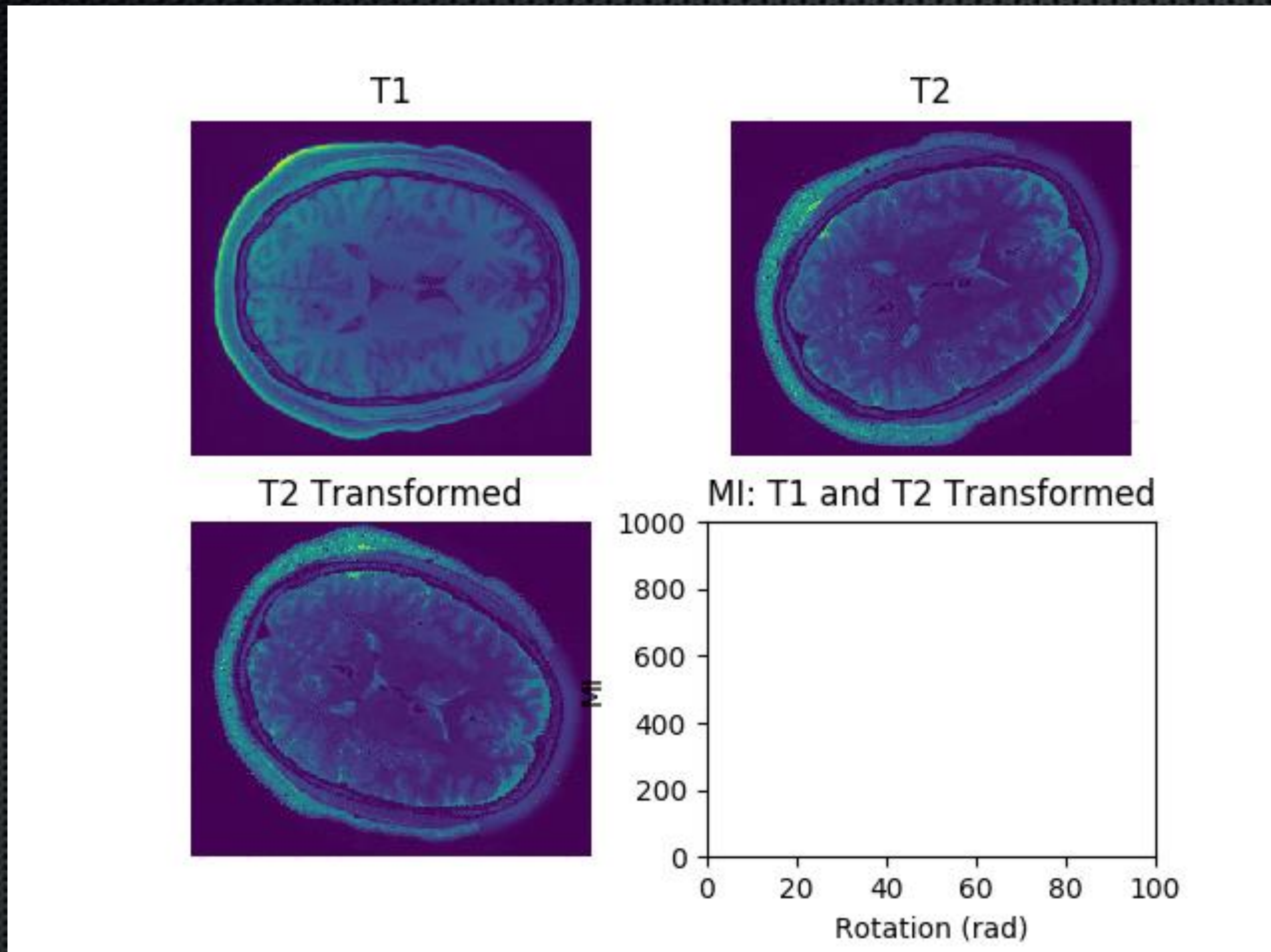
- A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, Automated multimodality image registration using information theory, Proc. Information Processing in Medical Imaging (Ile de Berder, France) (Y. Bizais, C. Barillot, and R. Di Paola, eds.), 1995, pp. 263-274. 259-302.
- Lisa Gottesfeld Brown, A survey of image registration techniques, ACM Computing Surveys 24 (1992), no. 4, 325- 376.
- P. Cachier, E. Bardinet, D. Dormont, X. Pennec, and N. Ayache, Iconic Feature Based Nonrigid Registration: The PASHA Algorithm, CVIU -- Special Issue on Nonrigid Registration 89 (2003), no. 2-3, 272-298.
- Joseph V. Hajnal, Derek L. G. Hill, and David J. Hawkes (eds.), Medical image registration, CRC Press, 2001.
- Jan Modersitzki, Numerical methods for image registration, Numerical Mathematics and Scientific Computation, Oxford University Press, 2004.
- Josien P. W. Pluim, J. B. Antoine Maintz, and Max A. Viergever, Mutual-information based registration of medical images: A survey, IEEE Trans. Medical Imaging 22 (2003), no. 8, 986-1004.
- Alexis Roche, Grégoire Malandain, Nicholas Ayache, and Sylvain Prima, Towards a better comprehension of similarity measures used in medical image registration, Proc. of MICCAI 1999, 1999.
- Philippe Thévenaz and Michael Unser, Optimization of mutual information for multi-resolution image registration, IEEE Trans. Image Processing 9 (2000), no. 12, 2083-2099.
- W. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, Multi-modal volume registration by maximization of mutual information, Medical Image Analysis 1 (1996), no. 1, 35-51.

Demo!



- Data from HCP:
- David C. Van Essen, Stephen M. Smith, Deanna M. Barch, Timothy E.J. Behrens, Essa Yacoub, Kamil Ugurbil, for the WU-Minn HCP Consortium. (2013). The WU-Minn Human Connectome Project: An overview. NeuroImage 80(2013):62-79.

Demo!



- Data from HCP:
- David C. Van Essen, Stephen M. Smith, Deanna M. Barch, Timothy E.J. Behrens, Essa Yacoub, Kamil Ugurbil, for the WU-Minn HCP Consortium. (2013). The WU-Minn Human Connectome Project: An overview. NeuroImage 80(2013):62-79.

MRI/US ANIMATION