Lecture 20
Registration and Optimization Considerations: Cost Function

MP574: Applications

Sean B. Fain, PhD (sfain@wisc.edu)
Diego Hernando, PhD (dhernando@wisc.edu)
ITK/VTK Applications: Andrew Hahn, PhD (adhahn@wisc.edu)

Housekeeping

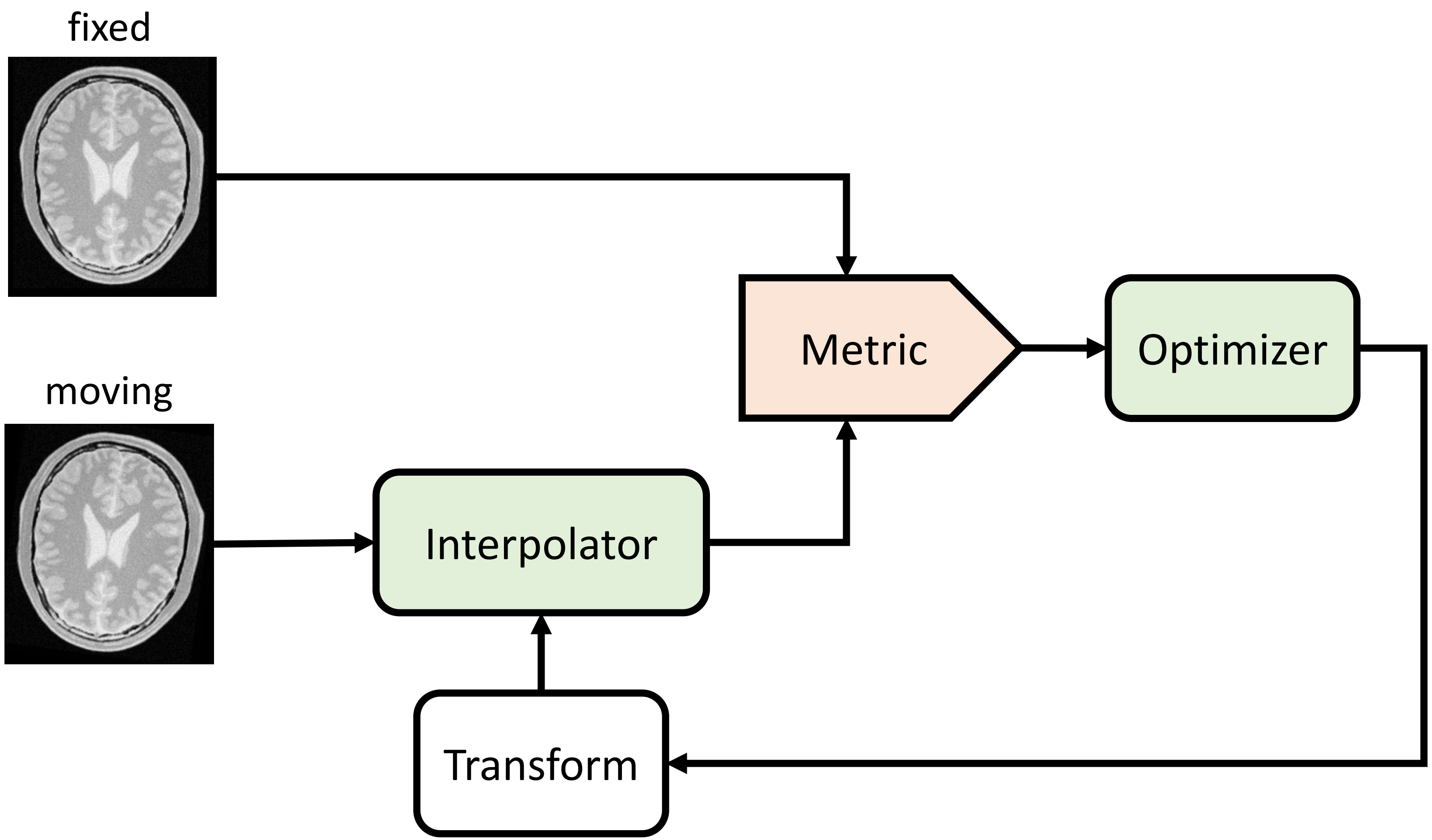
- Group 5 Presentation of project proposal
- Homework 4 questions?
- Install virtual box (https://www.virtualbox.org/wiki/Downloads)
- Download the virtual machine if you haven't already
  - Link: https://uwmadison.box.com/s/saemxuith6lbmurbys9z6xtywp96jdm6

- Lecture 20

Learning Objectives

- Review the optimization problem in image registration
- Introduce major classes of cost or “matching” function formulations for image registration
  - Geometric or Landmark-Based
  - Signal Intensity-Based
  - Information-Based

Registration Workflow



Review/Introduction

- The mapping is highly influenced by the degrees of freedom (DOFs)
- Constraining the mapping and a wide range of DOFs are used in medical image registration depending on the application and desired goals.
  - In general, the preference is to minimize the DOF's required to reach a useful registration solution,
    - …while allowing sufficient DOF to model the actual mapping.
- Choice of DOF and transformation depends on:
  - the organ system,
  - desired accuracy,
  - level of interaction, and
  - mono- vs. multi-modality application.

Level of Interaction

- Interactive or “supervised”
- Semi-automatic
    - User initialized, corrected
    - Supervised segmentation step
- Automatic
    - Usually search-based
    - Iterative with cost function minimization

Choice of Cost Function

- The primary determining factor is the nature of the tissue signal intensity represented in the image.
  - If mono-modal registration problem, then signal intensity-based metrics such as the L1- and L2-norm and cross-correlation are common and useful.
  - If a multi-modal registration method, then signal intensities will vary by tissue type across the fixed and moving images:
    - Surface matching, or
    - Information theory approach (sometimes called “overlap” methods).

Classes of Cost Functions

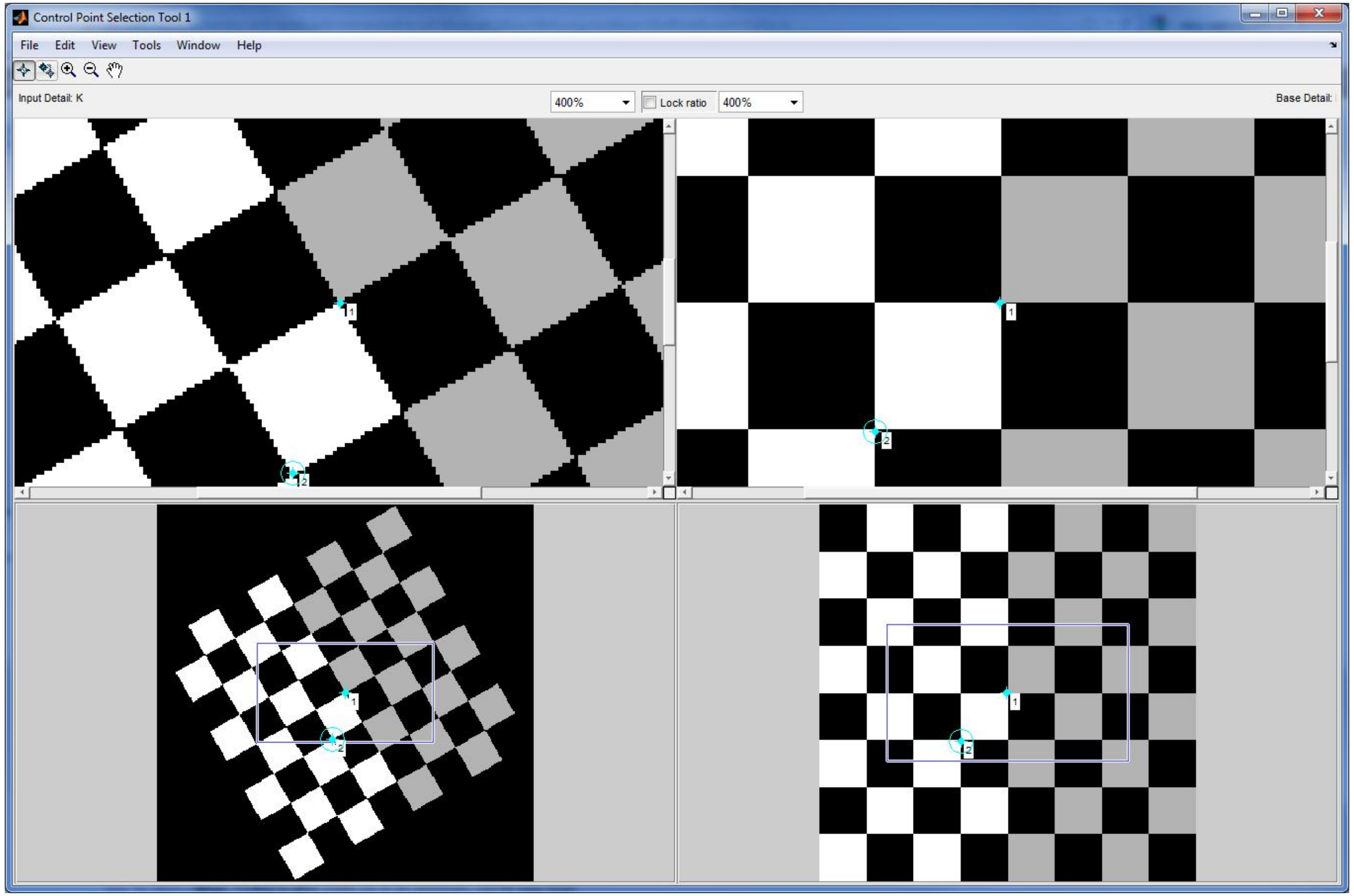
- Geometric/Landmark
  - Advantage = sparse and computational simple
  - Can be noise sensitive and insensitive to local deformation
  - Detection of features in both fixed and moving images can be difficult
- Signal Intensity-Based
  - L1, L2 Norms
  - Cross-correlation
- Information-Based
  - Entropy, Joint Entropy, and Mutual Information (MI)
  - Normalized MI

Geometric/Landmark

- Two-step process:
  - Detecting points of interest
  - Determining correspondence
  - Inferring the transformation
- Examples:
  - Fiducial or Control-point registration
  - Anatomic or Feature-based registration

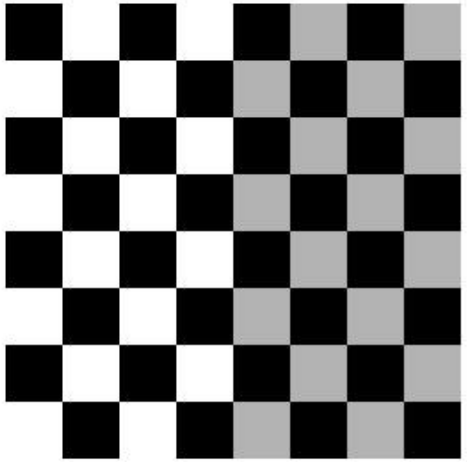
- $K_F = \{\kappa_1, \kappa_2, \kappa_3, \ldots \kappa_n\}$ and $\Lambda_M = \{\lambda_1, \lambda_2, \lambda_3, \ldots \lambda_n\}$,
  - the first set of landmarks contains points corresponding the fixed domain, while the second corresponds to the moving domain.
- These points can be selected manually as in Control-Point registration
- Or in an automated approach to identify “features” corresponding to prominent structures in the image.

Control Point Placement



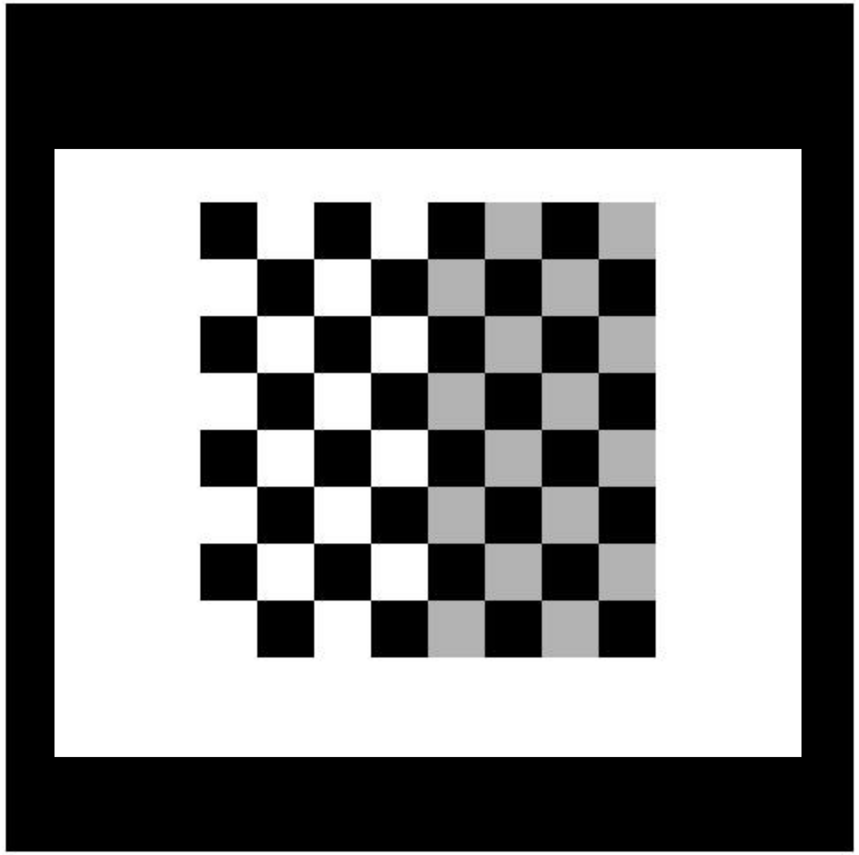
[movingPoints,fixedPoints] = cpselect(moving,fixed)

Registration Result



scale = 0.986

angle = 28.5^{o}



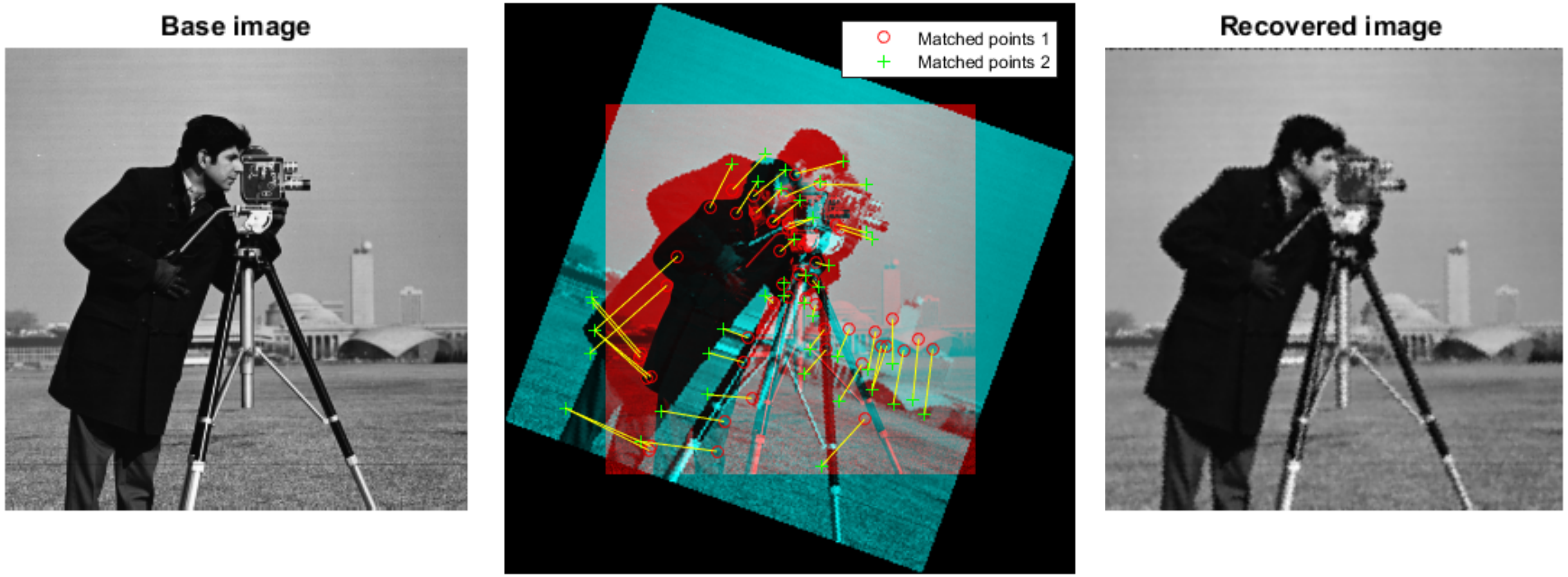
tform = fitgeotrans(movingPoints,fixedPoints,transformationType)

Homework 4: Registration Script

../../../Homeworks/Registration_script.m

Feature-based Registration


```
points1 = detectSURFFeatures(I1);
[f1, vpts1] = extractFeatures(I1, points1);
indexPairs = matchFeatures(f1, f2);
```


Base image
Matched points 1
Matched points 2
Recovered image

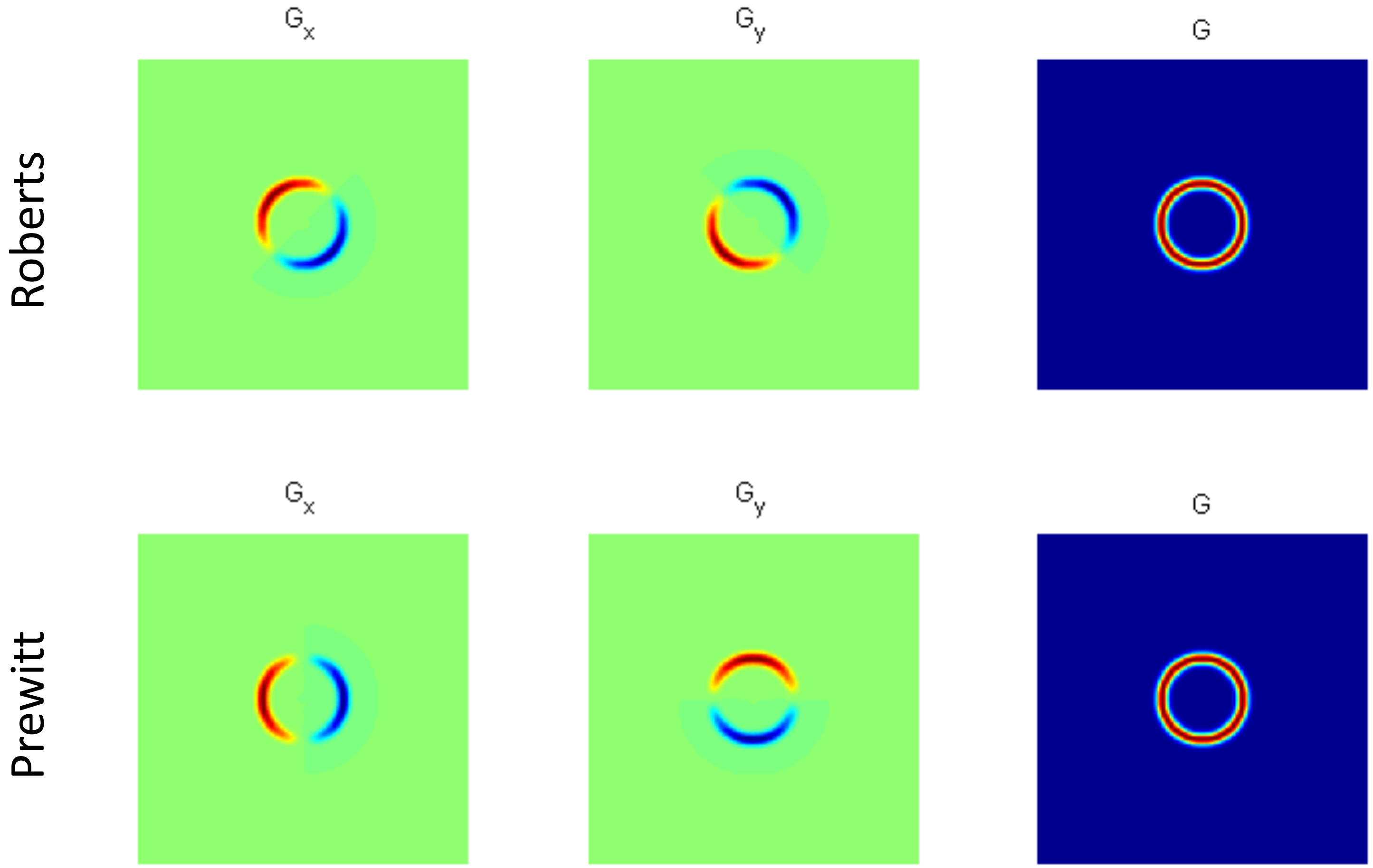
Generalization to 2D

- 2D Laplacian kernels:

Asymmetric: Symmetric:

$$\underbrace{\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}}_{\text{Separable}} \qquad \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

Gradient Magnitude



Noise Insensitive Methods

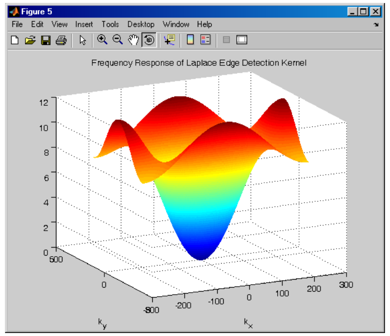
- Threshold based on “strength” of edge relative to maximum gradient

Laplacian of Gaussian Kernel

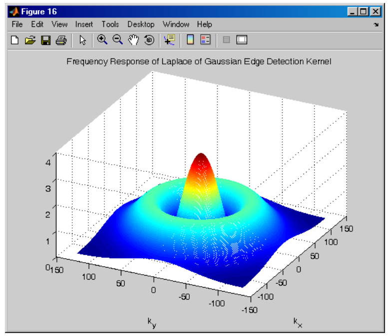
$$-\nabla^2 e^{-\frac{(x^2+y^2)}{2\sigma^2}} = -\left[\frac{(x^2+y^2)-\sigma^2}{\sigma^4}\right] e^{-\frac{(x^2+y^2)}{2\sigma^2}}$$

Reduced noise sensitivity by windowing

Frequency response for symmetric kernel:



Laplacian of Gaussian:

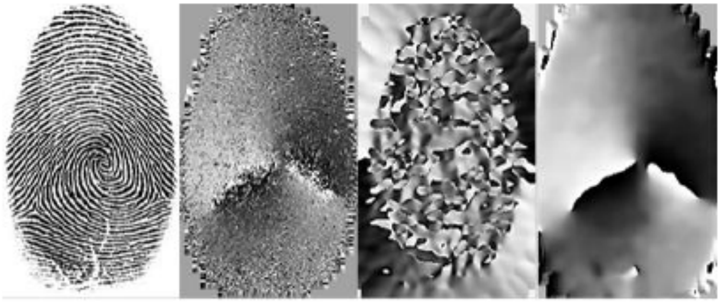


A Structure Tensor for Hyperspectral Images

Maider Marin-McGee and Miguel Velez-Reyes

Laboratory for Applied Remote Sensing and Image Processing, University of Puerto Rico-Mayaguez, PR, USA
E-mail: {maider.marin, miguel.velez-reyes}@upr.edu

$$J_0 = \sum_{i=1}^{m} w_i J_{0i} w_i^T = \begin{bmatrix} \sum_{i=1}^{m} w_{x,i}^2 I_{x,i}^{\,2} & \sum_{i=1}^{m} w_{xy,i} I_{x,i} I_{y,i} \\ \sum_{i=1}^{m} w_{xy,i} I_{x,i} I_{y,i} & \sum_{i=1}^{m} w_{y,i}^2 I_{y,i}^{\,2} \end{bmatrix}$$



(a) (b) (c) (d)

Figure 1. Smoothed Gradient vs. Structure Tensor in a fingerprint image of size 227× 227. (a) Original image. (b) Gradient orientation σ = 0.5. (c) Gradient orientation σ = 2.5. (d) Structure Tensor σ = 0.5, ρ = 4.

2D Laplacian vs. Canny

% convert the image to 8 bit

```
>> brain3 = uint8(brain2);
>> brain4 = brain3;
```

% threshold the image at T = 29

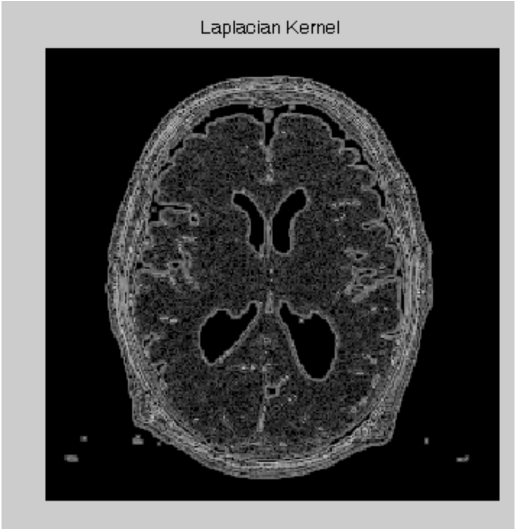
```
>> brain4(find(brain4<=29)) = 0;
```

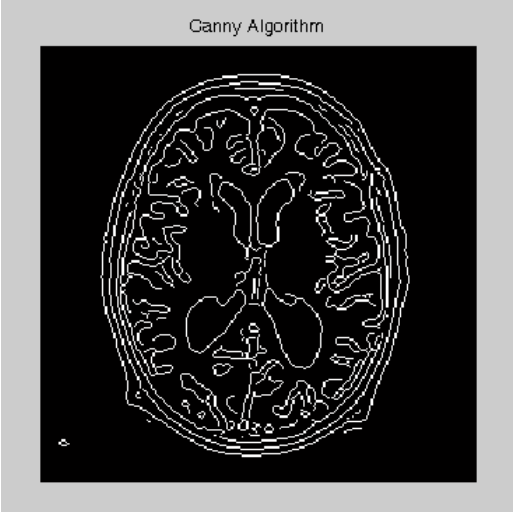
% apply Laplacian edge detection to this image; kernel is symmetric

```
>> brain6 = conv2(double(brain4),double(kernel));
```

% apply the Canny edge detection algorithm to this image

```
>> Brain5 = edge(brain4,'canny');
```





Surface or Chamfer Matching

- Jiang and Robb
  - Rationale – Edge detection to extract surfaces of objects to be registered.
  - Euclidean distance measure from base surface
  - Not all points need be evaluated

Euclidean Distance = Cost Function

$$\frac{1}{3}\left(\frac{1}{n}\sum_{m=1}^{n}\min(t^2, d_{l,i_m}, {}_{j_m}, {}_{k_m})\right)^{1/2}$$

$$where,$$

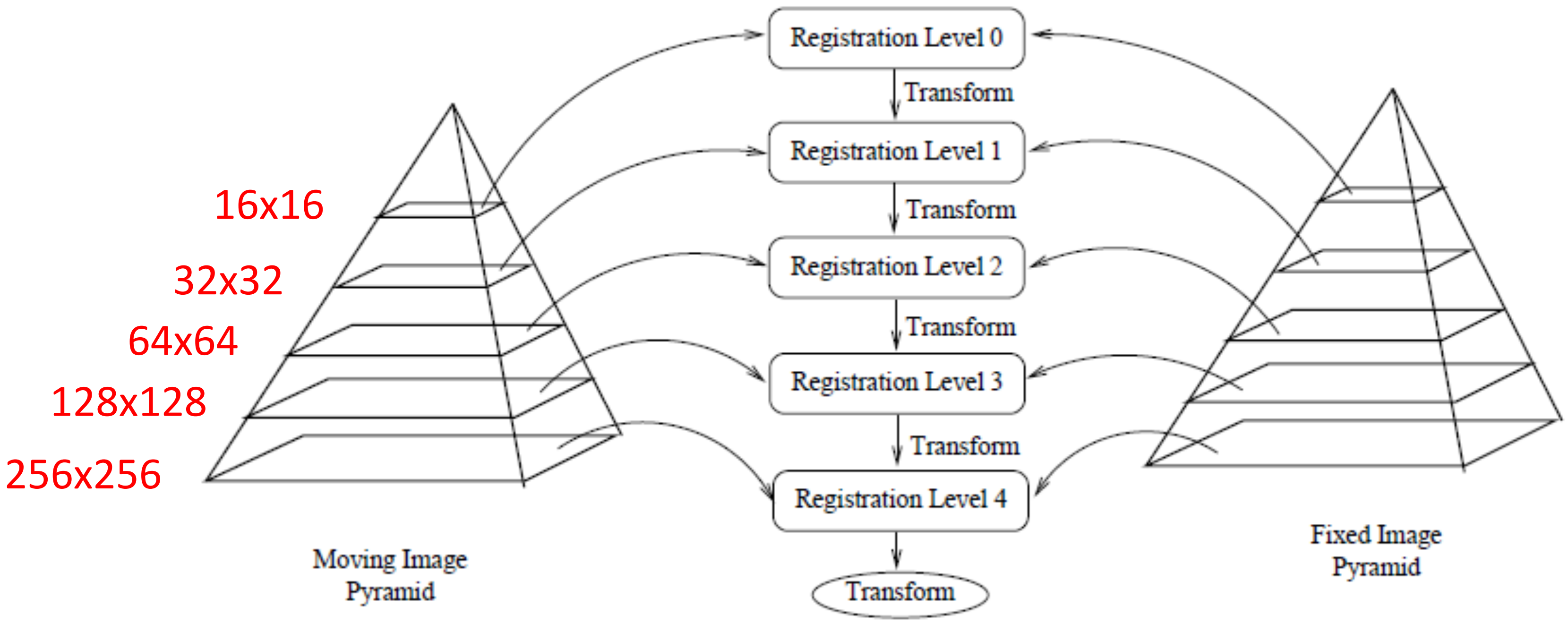
$$\mathrm{t} = \mathrm{threshold}$$

$$\mathrm{l} = \mathrm{resolution\ level}$$

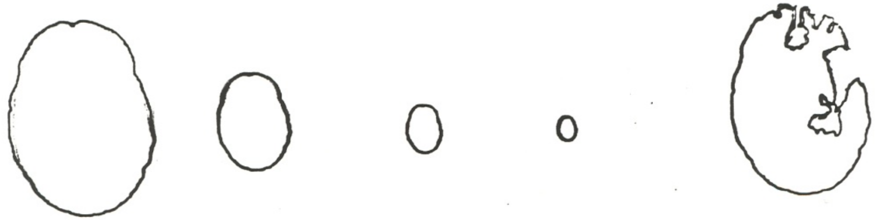
$$\mathrm{d} = \mathrm{Euclidian\ distance\ vector}$$

- Algorithm:
  - Segment surfaces
  - Create “distance image” for base surface
  - Select random points from other surface (Monte Carlo)
  - Calculate cost function: sum of values where points land on the distance image
  - Optimize
    - Coarse resolution - emphasizes global minima
    - Threshold to eliminate outliers.

Multi-resolution



Robustness to localized surface changes



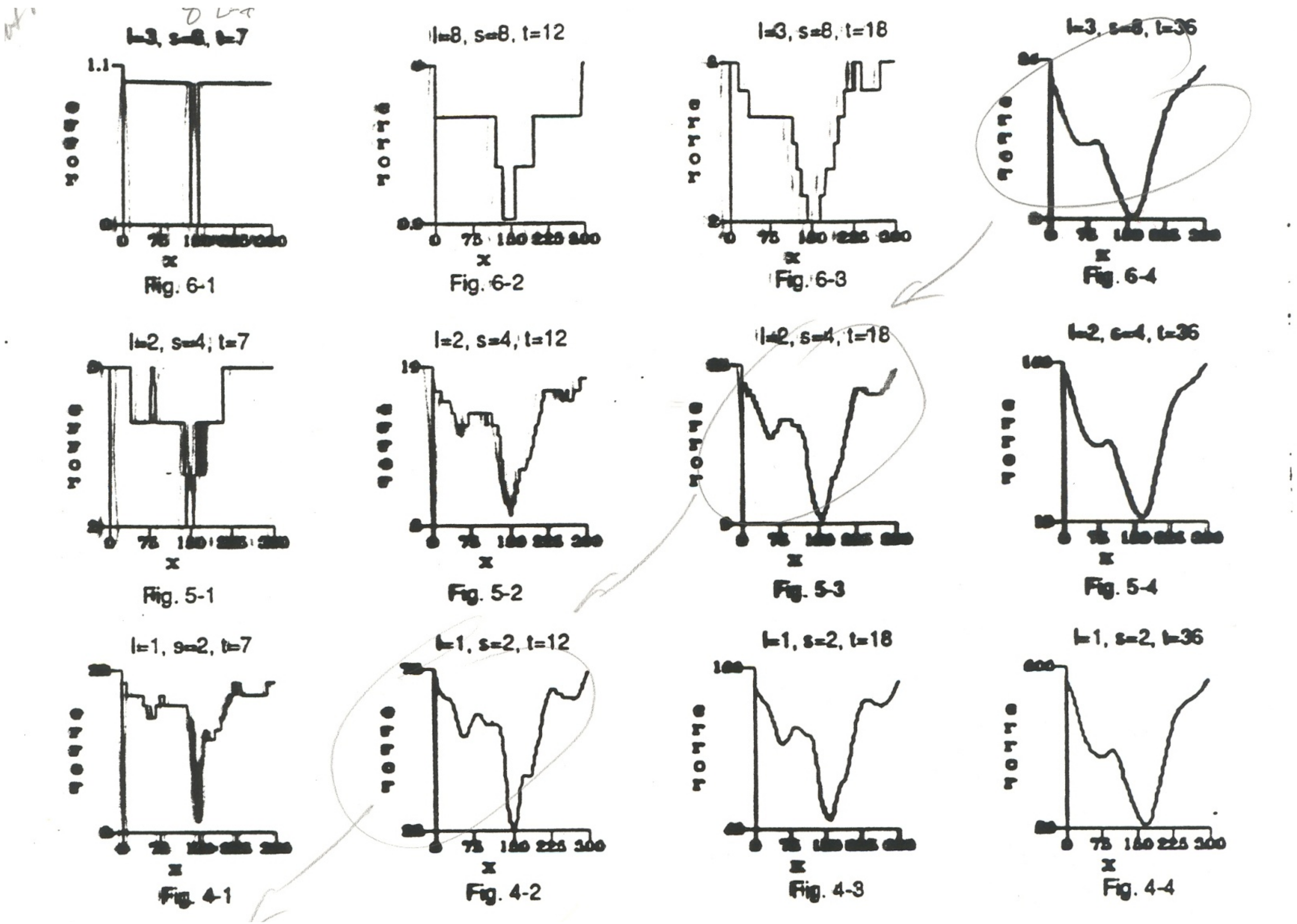
Scale=1 Fig. 1-1 | Scale=2 Fig. 1-2 | Scale=4 Fig. 1-3 | Scale=8 Fig. 1-4 | Scale=1 Fig. 1-5

Fig. 1-1 to Fig. 1-4 are base contours at different scales. Fig. 1-5 is the match contour. At scale=1, the sizes of contours are 144x186.

t=7 Fig. 2-1 | t=12 Fig. 2-2 | t=18 Fig. 2-3 | t=36 Fig. 2-4 | t=∞ Fig. 2-5

Fig. 2-1 to Fig. 2-5 are distance transformed images of the base contours with different threshold values t (unit in pixels). Scale=1 for all images.

Multi-scale improvement of convergence



Signal Intensity-Based: Similarity Metrics

- Sum of absolute differences (L1-norm):

$$SAD = \frac{1}{N}\sum_{i}^{N}|A(i) - B\ (i)|$$

- Sum of squares of differences (L2-norm):

$$SSD = \frac{1}{N}\sum_{i}^{N}(A(i) - B\ (i))^2$$

- Correlation coefficient:

$$CC = \frac{\sum_i (A(i) - \bar{A})(B(i) - \bar{B})}{(\sum_i (A(i) - \bar{A})^2 \sum_j (B(i) - \bar{B})^2)^{1/2}}$$

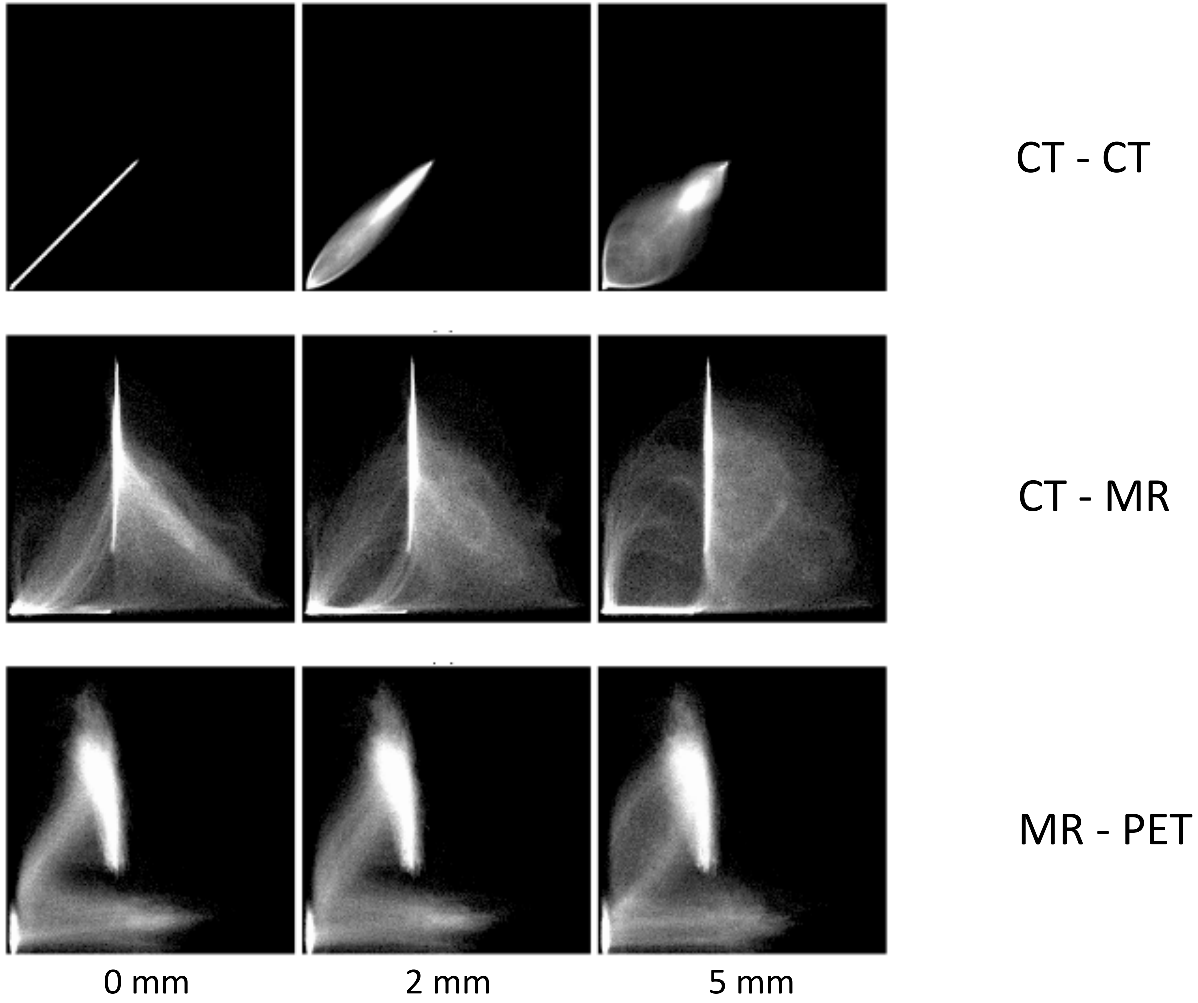
- Ratio-image uniformity:

$$RIU = \frac{\sigma_R}{\mu_R} \qquad R(i) = B\ (i)/A(i)$$

Information Theory: Overlap Methods

- Wood's Algorithm (multi-modality)
  - Statistical method
  - Rationale – gray level in one image (e.g. MRI) represent a certain tissue type in another modality image (e.g. PET).
  - Define a uniformity measure based on the correlation of voxels representing the same tissue types.
  - Define cost function:
    - Consider voxel *i* with value *j* in the MRI (base image)
      - $\alpha_{i,j}$ is the value of the corresponding PET voxel at the present overlap
      - σ_j is the standard deviation of $\alpha_{i,j}$ for all voxels *i* with an MRI voxel value *j*
      - *Let* α'_j be the mean of $\alpha_{i,j}$ for all voxels *i* with an MRI voxel value of *j*
      - *Define* $\sigma'_j = \sigma_j / \alpha'_j$, i.e. a coefficient of variation
      - n_j is the total number of voxels within the brain with MRI value of *j* and:
    - Cost Function = $\sigma'' = \sum_j \sigma'_j \frac{n_j}{N}$, where $N = \sum_j n_j$

Joint Histograms



Mutual Information (MI)

- Recall:

$p_A(a)$ = (No. of pixels with value a)/ (Total pixels),

$p_{AB}(a,b)$ is the joint probability of pixel values a and b occurring at the same coordinates in images A and B.

The mutual information between images A and B can be rewritten as:

$$I(A,B) = \sum_a \sum_b p_{A,B}(a,b) \log_2\left(p_{A,B}(a,b) / p_A(a) p_B(b)\right)$$

Joint Entropy

Joint probability distribution function (PDF):

$$PDF(a,b) = \frac{HIST(a,b)}{\sum_{a,b} HIST(a,b)}$$

Joint entropy:

$$H(A,B) = -\sum_{a,b} PDF(a,b)\ln(PDF(a,b))$$

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

Mutual Information (MI)

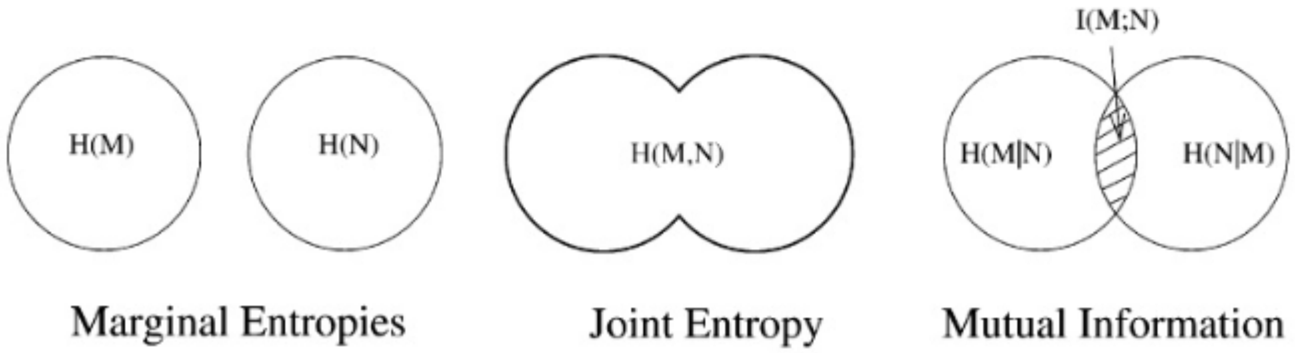
Statistical image information concepts tied to entropy

$$H(A) = \sum_a \left(\sum_b PDF(a,b) \ln \sum_b PDF(a,b) \right)$$

$$H(B) = \sum_b \left(\sum_a PDF(a,b) \ln \sum_a PDF(a,b) \right)$$

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

We seek to maximize the MI by minimizing $H(A,B)$.



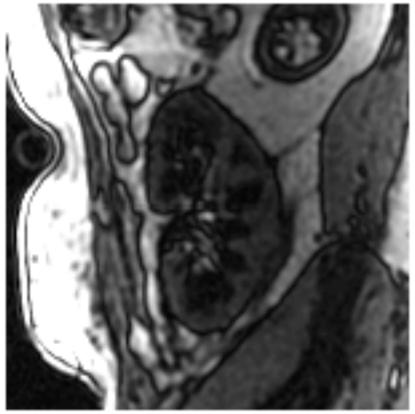
Normalized MI (NMI)

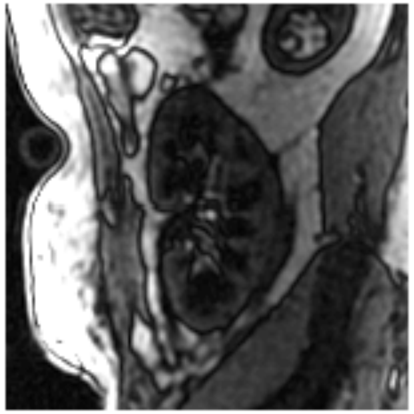
The normalized mutual information between images A and B can be written as (Studholme *et al*):

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

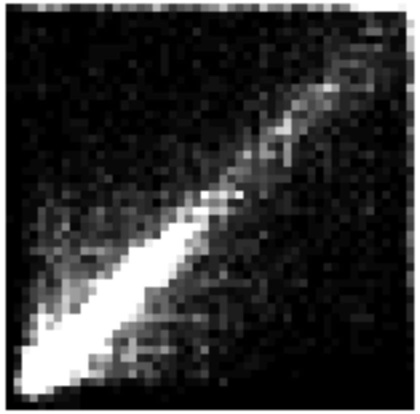
We also seek to maximize the NMI by minimizing $H(A, B)$.
Alternative, and most common, cost function.

Matlab





HIST



```
pdf = HIST/sum(HIST(:));
pa = sum(pdf,1);
pb = sum(pdf,2);
pab = sum(pdf(:));

HA = sum(pb.*log2(pb));
HB = sum(pa.*log2(pa));
HAB = sum(pab*log2(pab));
```

Optimizers

- Optimizers are used to optimize the metric criterion with respect to the transform parameters.
- The basic input to an optimizer is a cost function object.
- Some optimizers also allow rescaling of their individual parameters.

Summary

- Image registration is one of the most complex post-processing operations
  - Plethora of methods classified broadly by:
    - Rigid or deformable
    - Supervised or automatic
    - Multi-modality or mono-modality
  - Method chosen will depend on application and desired accuracy
- Affine transform with mutual information cost function
  - Most universally applied algorithm routinely used in radiation therapy treatment planning
  - Locally constrained deformable registration after affine transformation when needed