MP574 Lecture 20: Cost Functions

Learning Objectives

- Review the optimization problem in image registration
- Introduce cost or “matching” function formulations for image registration

References:
Sotiras et al., IEEE Transactions on Medical Imaging, Vol. 32 (7): 1153-1190 (2013).
Marin-McGee et al., A Structure Tensor for Hyperspectral Images, Conference Paper (2011): https://www.researchgate.net/publication/220004170_A_structure_tensor_for_hyperspectral_images.
Studholme et al, An overlap invariant entropy measure of 3D medical image alignment, Pattern Recognition 32(1): 71-86. (1999).

I. Introduction
   a. We have discussed transformations in Lectures 18 and 19. Transformations map coordinates from one space to another to facilitate linear or deformable image registration.
   b. 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.
      i. 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.
      ii. Choice of DOF and transformation depends on the organ system, desired accuracy, supervision/automation, and mono- vs. multi-modality application.
   c. A similar calculus extends to the choice of cost (or “matching”) function and optimization algorithms.
      i. The primary determining factor for the choice of cost function and optimization algorithm is the nature of the tissue signal intensity represented in the image.
      ii. For example, if a mono-modal registration problem, then signal intensity-based metrics such as the L1- and L2-norm and cross-correlation are common and useful.
      iii. But if a multi-modal registration method then signal intensities will vary by tissue type across the fixed and moving images, suggesting that a surface matching or information theory approach (sometimes called “overlap” methods) will be more robust.

MP574 Lecture 20: Cost Functions

II. Cost functions for image registration
   a. Geometric/Landmark-based methods
      i. Two-step process:
         1. Detecting corresponding points of interest
         2. Inferring the transformation from these points
      ii. Known variables = points of interest.
      iii. $K_F = \{\kappa_1, \kappa_2, \kappa_3, \ldots \kappa_n\}$ 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.
      iv. These points can be selected manually in a user supervised approach, as for the Control-Point registration in Homework 4, or in an automated approach to identify "features" corresponding to prominent structures in the image.
      v. Several methods for automated selection of features are built on fundamental approaches to segmentation.
         1. Two common methods are the use of noise-insensitive edge detectors framed on the Laplacian,
            a. including the "Laplacian of Gaussian" filter designed to detect prominent edges in the image (Fig. 1), and
            b. The structure tensor which identifies strong boundaries and their direction using Laplacian-based edge detection.
         2. The structure tensor consists of a matrix, A, taken over some neighborhood of points within a window, w, centered in 3D space at p = (x, y, z):

$$A = \begin{bmatrix} I_x(p)^2 & I_xI_y & I_xI_z \\ I_yI_x & I_y(p)^2 & I_yI_z \\ I_zI_x & I_zI_y & I_z(p)^2 \end{bmatrix}, \qquad [1]$$

Where $I_\zeta(p) \equiv$ is the partial derivative of signal intensity, I, with respect to direction x, y, or z centered at around a neighborhood of pixels at p.

The structure tensor can be calculated for both fixed and moving images with additional criteria to rank potential corresponding points based on strength of the edge information in feature space. One example of feature space would be the eigenvalue of the structure tensor A, given by the characteristic function:

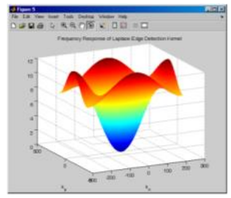
$$\lambda^2 - tr(A)\lambda + \det(A) = 0, \text{ such that}$$

$$|\det(A) - \alpha\, tr(A)| \propto \lambda^2 \qquad [2]$$

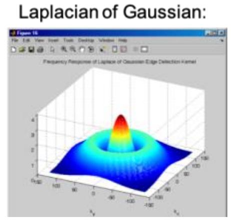
$$-\nabla^2 e^{-\frac{(x^2+y^2)}{2\sigma^2}} = -\left[\frac{(x^2+y^2)-\sigma^2}{\sigma^4}\right] e^{-\frac{(x^2+y^2)}{2\sigma^2}}$$

Reduced noise sensitivity by windowing

Frequency response for symmetric kernel:



Laplacian of Gaussian:

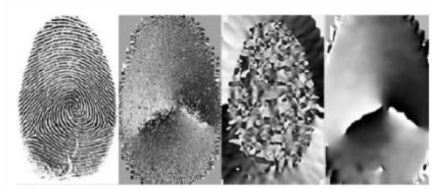


A Structure Tensor for Hyperspectral Images

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

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

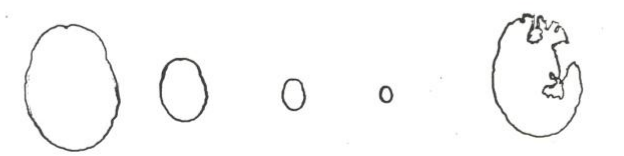


(a) (b) (c) (d)

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.

Figure 1: The Gaussian windowed Laplace filter for noise-insensitive edge detection and the structure tensor matrix are commonly used to automatically and semi-automatically detect corresponding landmarks for feature-based registration.

The second expression above is proportion to the eigenvalues of the structural tensor matrix and represents the principal components of signal intensity in the image, i.e. corners and edges that can serve as reproducible landmarks in both fixed and moving images (Fig. 1). In this



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.

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

where,

t = threshold

l = resolution level

d = Euclidian distance vector

Figure 2: Summary of a chamfer matching approach with upper left showing the extracted surfaces for different image resolutions and the corresponding feature space for the distance cost function at different threshold/resolution combinations. The expression for the Euclidean distance cost function at right.

manner, the most likely corresponding edges between fixed and moving images are then matched as "corresponding points" sorted according to the sets, K_F and Λ_M above, and the transform to align them is then determined. An example from Homework 4 is the Feature-Based registration method (Fig. 4, Lecture 18).

Some key tradeoffs of geometric landmark-based registration include one important advantage, which is the sparsity of the transform. Since only a few landmark points are registered, the problem is sparse and computationally straight-forward. However, major disadvantages are noise sensitivity leading to circumstances where complex features are not easily identified as corresponding, leading to a need of user-supervision in many cases.

Another example of feature-based registration is surface or "chamfer" matching in which the Euclidean distance from segmented surfaces is used as a cost function to drive iterative, automated registration. An example is shown in Fig. 2 in which a multi-scale approach is used to speed and constrain convergence.

Automated methods depend on cost functions that compare every voxel in the image or use a hybrid approach to weight high confidence landmarks while including other pixels in the image in the calculation of the cost-function. These include signal intensity-based and information-based cost functions.

B. Signal Intensity-based

1. Simple concept is that similar tissue types have similar signal intensities in the fixed and moving images such that the criterion takes on low (or high depending on the cost function) values when points belonging to the same tissue class are examined and high (or low) values when points from different tissue classes are compared.
2. The signal-intensity based cost functions include the sum of absolute signal difference (SAD), L1-norm, sum of squares (SSD), L2-norm, and cross-correlation functions:

$$SAD = \frac{1}{N}\sum_i^N |A(i) - B\,(i)| \quad [3a]$$

$$SSD = \frac{1}{N}\sum_i^N (A(i) - B\,(i))^2 \quad [3b]$$

$$CC = \frac{\sum_i (A(i)-\bar{A})(B(i)-\bar{B})}{(\sum_i (A(i)-\bar{A})^2 \sum_j (B(i)-\bar{B})^2)^{1/2}} \quad [3c]$$

These cost functions are most robust when signal intensities in the fixed and moving images are expected to have similar signal intensity values, and thus are most useful for mono-modal registration problems.

The second and most widely used set of cost functions for automated registration problems are information-based, including the widely used mutual information and normalized mutual information cost functions.

C. Information-based

1. Statistical overlap methods were the precursor to mutual information methods. They were introduced deliberately to address the challenge of multi-modality registration where tissue signals are expected to differ markedly between the fixed and moving images because of the different underlying physics of the image contrast (e.g. PET and MRI).
2. The Woods algorithm assumes that the joint statistics of similar tissue types will behave similarly if not the actual signal intensities. The idea is that the coefficient of variation for alignments between fixed and moving images will be minimized when similar tissue types are overlapping.
3. This concept of joint statistical properties extends naturally to the more general concept of joint probability and joint entropy that are the basis of the mutual information cost function.
4.
5. Consider from information theory the concept that the information content of an event is given by the log of the inverse probability of that event. The information content is weighted by the probability that the event occurs, and summed over all the vents to give the entropy. For an image, A, with probability of pixel value, a, of $p_A(a)$, then:

$$H(A) = \sum_a p_A(a)\, log\left(\frac{1}{p_A(a)}\right) = -\sum_a p_A(a)\, log\big(p_A(a)\big), \quad [4]$$

where H(A) is the entropy of image A.

Recall that the probability events here are the occurrences of particular pixel values. As an intuition check, consider an image of the same uniform value in every pixel. Such an image would have zero entropy since log (1) = 0. For the mutual information cost function, we are interested in the joint entropy of the fixed image, A, and the moving image, B. The joint probability distribution function (PDF) is given by:

$$PDF(a,b) = \frac{HIST(a,b)}{\sum_{a,b} HIST(a,b)}, \quad [5]$$

and the joint entropy by:

$$H(A,B) = -\sum_{a,b} PDF(a,b)\ln\big(PDF(a,b)\big).[6]$$

Note also that it stands to reason from set theory (Fig. 3) that the joint entropy will always be less the sum of the individual image entropies, H(A) and H(B):

$$H(A,B) \le H(A) + H(B). \quad [7]$$

Equations 5-7 lead directly to the idea of mutual information (MI). When the joint entropy, that is the joint probabilities are reduced to a sparser histogram, similar tissue classes are overlapping such that H(A,B) is minimized and the MI as defined below is maximized:

$$MI(A,B) = H(A) + H(B) - H(A,B). \qquad [8]$$

The MI defined this way is not always convex depending on signal intensities in the respective images and degree of overlap. The MI value tends to increase when the overlapped region is getting smaller for partial overlapping conditions. A more robust and more often, but not always, symmetric and convex cost function is the normalized mutual information (NMI):

$$NMI(A,B) = \frac{H(A)+H(B)}{H(A,B)} \qquad [9]$$

The inclusion of the joint entropy in the denominator creates a more robust local maximum under circumstances of partial overlap, but numerical analyses have shown that NMI is not always convex either depending on specific conditions of the registration, so modifying the NMI cost function with additional constraints and workflows remains an active area of research. For most practical medical imaging applications NMI has become the most widely used cost function for automated image registration problems and works well when the initial condition begins with highly overlapped fixed and moving images.

III. Summary/Conclusions

In this chapter we have reviewed the optimization problem in image registration from the perspective of the cost or “matching” function formulations used. The basic cost functions are dependent on the approach. For user-supervised or sparse image registration problems, control-point or feature-based registration can be a robust a computationally simple solution. For more automated registration problems, a more dense cost function that takes into account every pixel in the image is typically used. If the registration problem is mono-modal, similarity metrics based on image signal intensity are typically used. If the registration problem is multi-modal, then information-based methods based on joint statistical properties between the fixed and moving images are used, most commonly the normalized mutual information (NMI).