Lecture 21: Optimization Algorithms for Image Registration

Learning Objectives

- Introduce the different optimization algorithms used for iterative image registration and their tradeoffs
- Review and synthesize the components for optimized image registration.
- Present Feature-Based image registration as a specific example from Homework 4 utilizing all components of the iterative optimization approach: Transform, Interpolator, Cost Function, Optimization Algorithm.

References:

Klein et al., IEEE Transactions in Image Processing, Vol. 16 (12): 2879-2890.

I. Mathematical Formulation of Optimized Iterative Image Registration

Recall the mathematical shorthand for the registration optimization problem from Lecture 19, Eq. 1.

$$M_{i+1} = C(F, T \circ M_i) + R(T),$$

where

M ≡ “Moving” image
F ≡ “Fixed” image
i ≡ current iteration index
T ≡ transform operator
C ≡ cost function
R(T) ≡ regularization of the transform operation (usually a smoothing operation to minimize noise and interpolation error)

Recall that regularization is closely coupled with the transform. In many cases regularization is a smoothing operation that removes discontinuities and noise in the transform deformation field. In inherently smooth transforms, such as affine or B-spline, regularization may also serve to introduce prior knowledge regarding the solution such as placing constraints on the scale and polarity of the transform. In the case that we seek the displacement of every image element (i.e., nonparametric deformation model) regularization may incorporate physical constraints, e.g. only certain values of the Jacobian (J>0 or J = 1) allowed, thus imposing task-specific constraints on the transformation.

In this lecture we discuss the implementation of the optimization algorithm. At each iteration there is a set of transformation vectors, $\vec{r}_{i+1}(p) = \vec{r}_i(p) + \alpha_i \vec{g}_i\Big(C\big(\vec{r}_i(p)\big) + R(\vec{r}_i(p))\Big)$, with the last term indexed in a simplified syntax as: $\vec{g}_i\big(\vec{r}_i(p)\big)$. In words, $\vec{g}_i(\cdot)$ is the optimization algorithm that defines the search direction for each step of the iteration and operates on a feature space with dimensionality equal to the cost function as constrained by the regularized transform.

Lecture 21: Optimization Algorithms for Image Registration

The cost function and the regularization are essential to enable the system to be solved. Specifically, both the cost function and regularization reduce the dimensionality of the system when applied over the number of degrees of freedom (e.g. from a handful of control-points using a rigid transform up to a fully non-parametric deformational model operating on all voxels in the image).

II. Optimization Algorithms

Several common mathematical formulations and methods are adapted as optimization algorithms for iterative image registration. The choice of optimization algorithm depends on the registration problem and degrees of freedom. In general, gradient descent and related methods perform well in the general case and are preferred in multi-modality registration problems. Slower rates of convergence is the main tradeoff for gradient descent methods. Stochastic gradient descent is sometimes used to speed convergence, but typically only works well in registration problems where the degrees of freedom are limited (i.e. B-spline deformable registration or lower). The respective mathematical forms are as listed in Table 1.

Algorithm	Formulation	Application
Gradient descent	$\vec{g}_i\left(\vec{r}_i(p)\right) = -\nabla_{\vec{r}_i(p)}(\vec{r}_i(p))$	Multi-modality, non-parametric
Conjugate gradient	$\vec{g}_i(\vec{r}_i) = -\nabla_{\vec{r}_i}(\vec{r}_i) + \beta_i\vec{g}_{i-1}$	Multi-modality, non-parametric, β_i provides a scalar weighting of the previous gradient to accelerate convergence.
Stochastic gradient descent	$\vec{g}_i(\vec{r}_i) = \vec{r}_i(p) + \alpha_i\hat{\vec{g}}_i(\vec{r}_i(p))$	Lower degree of freedom deformable registration transforms such as piecewise affine or B-spline. Several versions exist, here $\alpha_i$is used to indicate variable step-size with iteration, typically reducing with greater iterations to refine the search; $\hat{\vec{g}}_i$ approximates the gradient along randomized directions and for a subsample of uniformly sampled voxels.
Quasi-Newton	$\vec{g}_i(\vec{r}_i) = -(J^T(\vec{r}_i)J(\vec{r}_i))^{-1}\nabla_{\vec{r}_i}(\vec{r}_i)$	Mono-modality where $(J^T(\vec{r}_i)J(\vec{r}_i))$ approximates the change in local magnification and similarity of signal intensity is assumed across fixed and moving images. Thus, the gradient is weighted by the inverse of the local change in magnification (i.e. intuitively towards more trusted values).

Levenberg-Marquardt	$\vec{g}_i(\vec{r}_i) = -\left(H^{-1}(\vec{r}_i) + \xi\vec{I}\right)\nabla_{\vec{r}_i}(\vec{r}_i)$	The $H^{-1}(\vec{r}_i)$ is the inverse of the Hessian matix, which in the case of Gaussian distributed signal intensities with similar variances, is equivalent to the covariance matrix. As we know from last semester, the diagonal elements of this matrix are the principal components of variance and represent the maximum curvature of the objective function in the corresponding directions. Note that $\vec{I}$ is the identity matrix, so that when the scalar ξ is zero, the algorithm becomes the Gauss-Newton method. ξ is used to adjust the rate of convergence.

In mono-modality registration problems, Gauss-Newton and Levenberg-Marquardt, are more robust and converge more rapidly. “Quasi-Newton” methods use the Jacobian to adjust the rate of convergence such that the transpose-square of the Jacobian approximates the second derivative of the local deformation.

III. Specific Example of Automated Image Registration

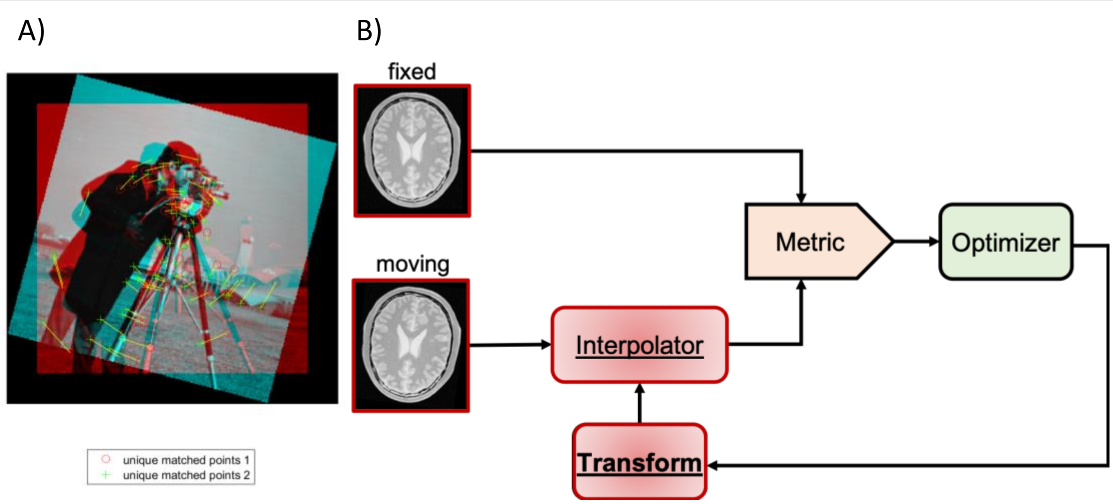

Figure 1: A) Feature-based image registration depends on a sparse set of corresponding points found typically using an automated or semi-automated algorithm (edge or corner detection), whereas in generalized automated image registration, B), the process employs a cost function (“metric”) with an optimization algorithm in an iterative manner to find the global minimum (maximum) of the metric with interpolation performed at each transform step. Additional regularization constraints, typically smoothness, are applied iteratively also.

Lecture 21: Optimization Algorithms for Image Registration

Feature-Based registration as performed in Problem 5 of Homework 4 with the function call to matchFeatures() identifies corner points that correspond between the two images using a distance threshold (Figure 1). In this particular example of image registration the cost function is used to identify the set of points to register, and the transform is then typically applied in a one step process based on the inverse mapping of the features to the fixed coordinate space.

More typically in automated registration, the cost function is used to optimize the alignment of the image in an iterative process of incremental mappings as presented in Lecture 19, Eq. 1 and represented schematically in Figure 2. An example cost function commonly used for evaluating alignment is the normalized mutual information (NMI). This cost function is then iteratively calculated along with regularization constraints and associated interpolation is performed and an optimization algorithm is used to determine likely global maxima of the NMI. Once the NMI is found to be unchanged for a search space around several candidate transformations, and assuming any other regularization constraints are satisfied, such as topological constraints as described in Lecture 19, then the iteration stops and presented as a registration solution.

Certain generalized approaches, such as the NMI cost function with conjugate gradient and bilinear or cubic spline interpolation, are commonly used because they are robust to multi-modality, converge in most situations, and mitigate interpolation errors. However, this is at a cost of additional computation time and may be untenable for large scale registration problems. By knowing the strengths and limitations of the possible transformation degrees of freedom, cost functions for circumstances of mono, or multi-modality problems, and the convergence of the different optimization algorithms, you are now in a position to adapt your choices to the complexity of the research problem to improve the efficiency while maintaining robustness of your solution. Moreover, the optimization problem as presented here is commonly used to solve multiple image processing and reconstruction problems in slightly different forms and thus the student can adapt what they have learned in the image registration application widely to other problems the will encounter in the research setting.

IV. Summary

General tradeoffs between gradient descent (multi-modality) and quasi-Gauss-Newton methods (mono-modality) are computational cost, and rate of convergence depending on the registration problem.

Certain generalized approaches, such as the NMI cost function with conjugate gradient and bilinear or cubic spline interpolation, are commonly used because they are robust to multi-modality, converge in most situations, and mitigate interpolation errors.

Feature-based and control-point registration are distinct from the more generalized approaches discussed here and addressed in the itk/vtk in-class exercises. In many instances the registration problem can be treated as a modular problem in which transform, cost function, and optimization algorithm can be chosen to be tailored to the problem of interest and with tradeoffs defined by the investigator.