Parameter Scale Estimation

BY BAOHUA WU

Mar. 15, 2011

1. Search space regularization

Different parameters may have different impacts on the Jacobian of a transform, depending on the scale of parameters. Due to that, the derivative of a metric function over different parameters may have different ranges. As a result, in gradient descent method, some parameters may always take bigger steps than others. We will illustrate this phenomenon below.

Let us consider the case of the mean square metric M of a fixed image I_f and a moving image I_m:

$$M(I_f, I_m) = \frac{1}{N} \sum_{i=1}^{N} (I_{m_i} - I_{f_i})^2$$

where N is the number of voxels and $I_m = I_m(\phi(x, \theta, t_1, t_2))$ in which ϕ is a rigid transformation with a rotation θ and a translation (t_1, t_2), and x is the coordinates.

The partial derivatives of M with respect to transformation parameters $p = (\theta, t_1, t_2)$ is

$$\nabla_p M = \frac{2}{N} \sum_i (I_{m_i} - I_{f_i}) \cdot \nabla_\phi I_{m_i} \cdot J_p^i \phi \qquad \text{where } \nabla \text{ is the gradient operator and } J \text{ is the Jacobian matrix.}$$

In a real example, we have a gradient as below in a gradient deepest descent optimization:

$$\nabla_p M = (-13.016, 0.00476, -0.0359)$$

It is not optimal to update the parameters with $\frac{\nabla_p M}{\|\nabla_p M\|_2}$ since the update of rotation θ is about thousand times larger than that of translation t_1 and t_2. This scale difference between rotation and translation makes the search space look like a long narrow valley which may entail more steps of optimization.

Our scale estimation will make the search space look more like a round basin. That is, the partial derivative $\frac{\partial M}{\partial p_j}$ should be roughly on the same magnitude for different p_j. A rescaled parameter $q_j = s_j * p_j$ will change the partial derivative to

$$\frac{\partial M}{\partial q_j} = \frac{\partial M}{\partial p_j} / s_j$$

2. Step size limitation

The parameter gradient $\nabla_p M$ won't approximate well over a large step size since the image gradient $\nabla_\phi I_{m_i}$ may change big when the transformation ϕ shifts over one voxel, especially near segment boundaries. Therefore, it is good to have a step small enough such that the maximum shift of voxel positions is less than one.

For affine transformations and convex image domains, it is enough to check the corners of the image domain for the maximum voxel shift. Here comes the proof. For a change $\triangle A$ and $\triangle t$ in an affine transformation $A\,x + t$, the voxel shift is $\triangle A x + \triangle t$, which is still affine. For any positive k_1 and k_2 such that $k_1 + k_2 = 1$, and two vertices x_1 and x_2, we have

$$\triangle A(k_1 x_1 + k_2 x_2) + \triangle t = k_1(\triangle A x_1 + t) + k_2(\triangle A x_2 + t)$$

This means any voxel on the line segment between x_1 and x_2 will be mapped onto the mapped segment. It implies that a convex domain will be mapped as a convex and the corner vertices are kept. So we just need to check the corners for the maximum shift.

For each unit change in only one parameter p_j, the maximum voxel shift may be used as its scale s_j since a change of $1/s_j$ will leads to a unit voxel shift. With all parameters considered, we have a change $(1/s_1, 1/s_2, ..., 1/s_P)$ and get another maximum voxel shift s. And with a change of $1/s * (1/s_1, 1/s_2, ..., 1/s_P)$, we will have a unit voxel shift. Therefore, the overall scales may be $s * s_j$.

[1] M. Jenkinson and S.M. Smith, "A global optimisation method for robust affine registration of brain images," Medical Image Analysis, vol. 5(2), pp. 143-156, June 2001.