Available online at www.sciencedirect.com
SCIENCE DIRECT®
Medical Image Analysis 8 (2004) 465–473
MEDICAL IMAGE ANALYSIS
www.elsevier.com/locate/media

Adaptive registration using local information measures

Hyunjin Park [a], Peyton H. Bland [a], Kristy K. Brock [b], Charles R. Meyer [a,*]

[a] *Department of Radiology, University of Michigan Medical School, Ann Arbor, MI 48109-0533, USA*
[b] *Radiation Medicine Program,Princess Margaret Hospital, Toronto, Ont., Canada*

Received 12 December 2002; received in revised form 22 September 2003; accepted 4 March 2004
Available online 23 April 2004

Abstract

Rapidly advancing registration methods increasingly employ warping transforms. High degrees of freedom (DOF) warpings can be specified by manually placing control points or instantiating a regular, dense grid of control points everywhere. The former approach is laborious and prone to operator bias, whereas the latter is computationally expensive. We propose to improve upon the latter approach by adaptively placing control points where they are needed. Local estimates of mutual information (MI) and entropy are used to identify local regions requiring additional DOF.

Keywords: Adaptive registration; Thin plate splines; Warping registration; Grid refinement

1. Introduction

Associated with rapid developments of medical imaging there is an increasing need for nonlinear registration. Registration literature has shifted its focus from affine/rigid registrations to high degrees of freedom (DOF) warping registrations (Hill et al., 2001; Johnson and Christensen, 2001; Meyer et al., 1998; Meyer and Boes, 1998; Reuckert et al., 1999). Warping registration algorithms also employ different similarity measures (i.e., measures of alignment) and geometric transforms to suit their purposes. A popular choice of similarity measure has been mutual information (MI) (Wells et al., 1996; Collignon et al., 1995). Additionally there are many geometric transforms to choose from where B-splines and thin plate splines (TPS) are notable (Lee et al., 1996; Bookstein, 1991). Recently, a warping registration algorithm with normalized MI as the similarity measure and B-splines as the geometric transform has gained much support (Reuckert et al., 1999). Herein we employ MI as the similarity measure and TPS as the geometric transform as in our previous work (Kim et al., 1999; Meyer et al., 1997, 1999). Although TPS may be less efficient to compute than B-splines due to the local support properties of B-splines, TPS is supported by a rich literature in shape statistics and Morphometrics (Bookstein, 1991, 1997; Dryden and Mardia, 1998).

A warping registration starts with an initial set of control points in both the reference and homologous dataset and then optimizes the loci of the control points typically in the homologous dataset to maximize MI while control points on the reference side remain fixed. The initial set of control points may be realized either by *manually specifying all control points, or by instantiating a regular grid of control points*. The first method may be impractical for high DOF since manually specifying all control points is *laborious and prone to operator bias*. Given the presence of local minima in numerical optimization of the MI, different initial locations of control points may lead to different final optimized control point locations. Thus, removing operator bias is important. The second, i.e., the regular grid method, suffers from increased computational expense instead. Typical DOF may be in hundreds or even in thousands. In this paper, we propose to *improve the second method by adaptively placing control points only where they are needed* rather than placing control points regularly everywhere to

* Corresponding author. Tel.: +1-734-763-5881; fax: +1-734-764-8541.
E-mail address: cmeyer@umich.edu (C.R. Meyer).
URL: http://www.dipl.med.umich.edu.

1361-8415/$ - see front matter
doi:10.1016/j.media.2004.03.001

improve overall registration. Our adaptive registration results in an irregularly spaced grid of control points with fewer DOF than a regular grid of control points and lesser computational expense.

There are other adaptive registration methods (Rohde et al., 2001, 2003; Rohlfing and Maurer, 2001; Schnabel et al., 2001). They typically share a common approach, i.e., first they identify a region where registration can be improved, and then increase DOF in that region. The geometric transforms and the methods to identify the region requiring additional DOF may be different. Rohde et al. use Wu's radial basis function as the geometric transform and the gradient of global MI to identify the region to increase warping DOF (Rohde et al., 2001, 2003). Others use B-splines as the geometric transform and measures based on entropy to identify the region to increase DOF (Rohlfing and Maurer, 2001; Schnabel et al., 2001). Our method uses TPS as the geometric transform and a novel information measure based on a *local* MI and entropy to identify the region to increase DOF. We refer to this local information measure as a mismatch measure. Note that in our paper the global MI used to optimize the control point locations is calculated over all of both datasets but our mismatch measure is calculated only over sub-regions of the datasets. In summary, we propose to *improve a regularly spaced TPS warping registration method by instantiating an irregular grid of control points derived from a novel local information measure* to determine where to increase DOF.

2. Methods

In this paper the following notations are assumed. $A(x)$ is the reference dataset and $B(x)$ is the homologous, or floating, dataset. $T(x)$ is the geometric transform between two datasets, where x is the coordinates in 2D or 3D. The homologous dataset is mapped onto the reference dataset before calculating the similarity measure. Once T is found, all the coordinates are assumed to in the reference coordinate frame since the homologous coordinate frame can always be found by applying the transform T,

$$\hat{T} = \arg\max_{T \in F} \mathrm{MI}(A(\bullet), BT(\bullet)), \tag{1}$$

where $\hat{T}$ is the estimate of the transform and F is the family of feasible transforms.

2.1. Thin plate splines

TPS is used to specify the geometric transform between two datasets of interest. *Control points are needed to formulate TPS*, which are placed in ordered pairs on the corresponding loci of both datasets. The loci can be anatomically or mathematically identified (Bookstein, 1991; Dryden and Mardia, 1998). Assuming that x is the reference coordinates, x' is the homologous coordinates, and that there are N control point pairs $(x_1, \ldots, x_N)$, and $(x'_1, \ldots, x'_N)$, the formulation of TPS is the following, where $x' = f(x)$ is the geometric transform, $U(r)$ is the basis function, r_i is the Euclidean distance between x_i and x (i.e., $|x - x_i|$):

$$T(x) = a_0 + a_1 x + \sum_{i=1}^{N} w_i U(r_i), \tag{2}$$

where a_0, a_1 are affine parameters and w_i is the warp coefficient,

$$U(r) = \begin{cases} r^2 \log(r^2) & \text{in 2D,} \\ |r| & \text{in 3D.} \end{cases}$$

B-splines are constructed to have a local support property, but with TPS the local support property is not strictly true because the basis function is globally defined everywhere. Thus, it is more expensive to compute TPS than B-splines. Still the effect of TPS is fairly local if the DOF are high enough.

2.2. Mutual information

Mutual Information is used as the similarity measure under the TPS geometric transform. The MI used here is the classical Shannon mutual information. *Many papers have shown the effectiveness of the MI similarity measure in registration problems* (Wells et al., 1996; Collignon et al., 1995; West et al., 1997; Meyer et al., 1997; Hill et al., 2001). Basically, two co-registered datasets yield a joint probability density function (PDF) with tight clusters, whereas un-registered datasets yield a joint PDF with disperse clusters. Tighter clusters (i.e., more correlation) translate into a higher MI value than disperse ones (i.e., less correlation).

$$\mathrm{MI}(A, B) = \sum \sum p(a, b) \log_2(p(a, b)/p(a)p(b)), \tag{3}$$

where $p(a)$, $p(b)$ are marginal densities of A and B, and $p(a, b)$ is the joint density of A and B.

While other papers used Parzen windowing for the estimation of PDFs (Wells et al., 1996), in our implementation all PDFs (both marginal and joint PDFs) are estimated using histograms with fixed bin width. The bin width of the histogram is calculated from Izenman (Izenman, 1991), where the optimal bin width h is chosen to minimize integrated mean squared error over all squared-integrable PDFs.

$$h = 2(IQR)n^{-1/3}, \tag{4}$$

where IQR is the inter quartile range and n is the number of samples.

Under a hypothetic geometric transform the homologous dataset is interpolated onto the reference coordinate frame using a simple linear, grayscale, interpolation, and then the joint histogram is formed to calculate an MI value.

2.3. General regular grid warping registration

The goal of a regular grid warping registration is to find a warping transform that registers two datasets according to maximizing a chosen similarity measure. Before attempting any high order registration, a simple low order affine registration removes any large linear global effects. This ensures that the subsequent high order transform deals with only relatively small local transforms. After the affine registration, a warping registration is performed in an iterative multi-scale fashion to save computational time. For lower resolution data (i.e., sub-sampled data), a sparse grid of control points is used. As the resolution increases, the grid becomes denser. Also within a specific resolution, different sized grids from large to small scale are used to speed convergence of optimization process. Any time grid density is increased, a new denser grid is initialized using the previously optimized sparse grid. This multi-scale grid optimization continues for a fixed number of resolutions and a fixed number of scales within a resolution. Typically, at every grid refinement in 3D, the existing grid is halved in all dimensions resulting in an 8-fold DOF increase.

2.4. Adaptive grid warping registration

The regular grid warping registration works well, but suffers major computational complexity at high DOF. For a typical CT dataset of $512 \times 512 \times 60$ with voxel dimension $1 \times 1 \times 5$ mm^3, control points placed every 5 mm results in 1,880,000 DOF, which leads to significant computational and convergence related problems. Additionally, with *control points placed regularly everywhere, there are many control points placed in background areas* where there is little information to judge the registration and little interest in local registration result. To remedy these computational and convergence related issues, *adaptive registration algorithms* have been developed (Rohde et al., 2001, 2003; Rohlfing and Maurer, 2001; Schnabel et al., 2001). *They place control points or B-splines only in areas where they are needed* to improve the overall registration. Thus, the registration is computed using an irregularly spaced grid having fewer DOF. Moreover, for a given DOF adaptive algorithms can allocate a dense grid of control points (or B-splines) to areas of interest without wasting them in backgrounds, and thus can achieve better registration accuracies.

2.5. Existing adaptive grid warping registration

Rohde et al. use the gradient of global MI to adaptively refine a grid (Rohde et al., 2001, 2003). They use Wu's radial basis function for the geometric transform, which has a finite local support property as does B-splines, and MI as the similarity measure. The authors use the following algorithm to accomplish grid refinement. At a given scale, they put basis functions at regular intervals and then change one coefficient of a basis function at a time and observe a change in global MI values. Rohde et al. argue that if the gradient of global MI is large, then the global MI is not maximum with respect to that specific location of the basis function and the registration can be improved further at that location. Once a new location of the basis function to be refined is determined, 8 smaller (i.e., octree) scale basis functions occupy the area where there previously was one larger basis function.

Rohlfing et al. use B-splines and a similarity measure where one term is global normalized MI and the other term is smoothness of the deformation (Rohlfing and Maurer, 2001). They use local entropies to estimate locally mis-registered areas and use active (i.e., allowed to move) and inactive (i.e., not allowed to move) B-splines to reduce DOF and effectively implement an irregular B-spline grid. Schanbel et al. also use B-splines and a two-term similarity measure, but refine their grid using local entropy, local SD, or gradient of the global cost function (Schnabel et al., 2001).

2.6. Local mismatch measure

Similarly, our goal is to increase DOF selectively in regions where they are needed rather than increase them globally everywhere in the data set. In context of TPS, we are trying to increase the density of control points in areas where additional control points are needed to empower the algorithm with the ability to correct local deformations. Noting that the effects of control points are fairly local, a good candidate region is the region with the largest local mismatch. Mismatched local areas are poorly correlated by definition, and thus have a low local MI. Additionally, a local MI between two corresponding areas can be low if the entropy of either of the corresponding areas is low, because MI is bounded by the entropies of the individual datasets (i.e., $\mathrm{MI}(a,b) \leqslant \min\{H(a), H(b)\}$). Because areas of low entropies are basically featureless, they are not good candidate loci for a grid refinement. Therefore, local *areas of interest* for the grid refinement are confined to those with *a high local entropy and a low local MI* These notions are formalized in the normalized measure of mismatch M introduced below:

$$M = 1 - \frac{\mathrm{MI}(a,b)}{\min(H(a), H(b))}. \tag{5}$$

Note that all values (i.e., $\mathrm{MI}(a,b)$, $H(a)$ and $H(b)$) are local measures (i.e., they are computed over a finite subblock, not over the whole data set). The mismatch