Gradient Projection Learning for Parametric Nonrigid Registration

Stefan Pszczolkowski[1] ⋆ ⋆⋆, Luis Pizarro[1], Declan P. O'Regan[2], and Daniel Rueckert[1]

[1]Biomedical Image Analysis Group, Imperial College London.
{sp2010,l.pizarro,d.rueckert}@imperial.ac.uk
[2]Robert Steiner MRI Unit, Imperial College London.
{declan.oregan}@imperial.ac.uk

Abstract. A potentially large anatomical variability among subjects in a population makes nonrigid image registration techniques prone to inaccuracies and to high computational costs in their optimisation. In this paper, we propose a new learning-based approach to accelerate the convergence rate of any chosen parametric energy-based image registration method. From a set of training images and their corresponding deformations, our method learns offline a projection from the gradient space of the energy functional to the parameter space of the chosen registration method using partial least squares. Combined with a regularisation term, the learnt projection is subsequently used online to approximate the optimisation of the energy functional for unseen images. We employ the B-spline approach as underlying registration method, but other parametric methods can be used as well. We perform experiments on synthetic image data and MR cardiac sequences to show that our approach significantly accelerates the convergence –in number of iterations and total computational cost– of the chosen registration method, while achieving similar results in terms of accuracy.

1 Introduction

Nonrigid image registration is a very important and widely investigated topic in medical image analysis, since it aids to remove the, potentially very large, natural structural variablity present in pairs or groups of medical images. A popular approach for nonrigid registration is to find a deformation field as the solution of an energy minimisation approach. There exist non-parametric methods [1–5] and parametric methods [6–11]. A typical energy functional E is composed of a data term E_D that measures the degree of alignment of a target (fixed) image and a source (moving) image, and a regularisation term E_R that imposes smoothness on the deformation field that aligns the images:

$$E = E_D + \lambda E_R \tag{1}$$

⋆ This work was partially funded by CONICYT
⋆⋆ We are grateful to Dr. Wenjia Bai for his comments and support

where $\lambda \geq 0$ is a tradeoff parameter between the two terms.

Learning-based image registration techniques have captured the interest of many researchers in the last few years. A popular approach is to capture the statistics of deformation by applying PCA over each band of wavelet coefficients [12] or over the control point values of B-splines that provide a parametric representation of the deformation fields [13–16]. In [17], the parameters of the deformation are estimated by a nearest neighbour search over training images generated according to a special criteria. A low dimensional representation of images, with maximally discriminative power is obtained in [18] by combining generative and discriminative objective functions in a constrained optimisation problem. The work in [19] presents a method where features extracted from regions obtained by adaptively partition brain images and the statistics of deformation field are used to robustly place the control points that parameterise the deformations. Also, a partial least squares approach is employed to relate cardiac deformation due to respiration with surface intensity traces in [20]. Finally, in [21], support vector regression is utilised to estimate the principal modes of a brain deformation model given low dimensionality image features.

One of the main issues of registration thechniques is their high computational cost. Recently, a new type of methods for increasing optimisation convergence have been developed, albeit they are not learning-based. This type of scheme is the so-called *preconditioning schemes*, where the image gradient is scaled differently for different areas of the image. The main contributions on this kind of approaches are [22, 23]. The difference is that in [23], the preconditioner is thought to work specificly for sum of squared differences, while the precontitioning scheme by [22] works for any similarity measure.

In this paper, we propose a learning-based parametric registration method that learns a projection from the gradient space of the energy functional to the parameter space of the chosen registration method using partial least squares. This learnt mapping can be seen as an approximation of the energy gradient, which we utilise to accelerate the convergence of the registration.

The rest of the paper is organised as follows. Section 2 introduces the concepts required by our approach, to then describe the learning and registration procedures, and a multi-resolution extesion for the method. Experiments and results on synthetic and medical data are shown on section 3. Finally, we conclude on section 4.

2 Method

2.1 Gradient Projections

Optical flow methods, e.g. [1], estimate a dense (voxel-wise) displacement field $\mathbf{u}$ that aligns the target and source images. This flow $\mathbf{u}$ is obtained as the solution of an energy functional (1). A standard gradient ascent (similarly, descent) scheme updates the solution with a speed-up factor $\eta > 0$,

$$\mathbf{u} \leftarrow \mathbf{u} + \eta \cdot \frac{\partial E}{\partial \mathbf{u}}. \tag{2}$$

where

$$\frac{\partial E}{\partial \mathbf{u}} = \frac{\partial E_D}{\partial \mathbf{u}} + \lambda \frac{\partial E_R}{\partial \mathbf{u}}. \tag{3}$$

The gradients are computed according to the specific choice of the data and regularisation functions. For our later developments, we call the term $\frac{\partial E_D}{\partial \mathbf{u}}$ *similarity gradient image* (SGI).

In parametric registration approaches, such as the B-spline based free-form deformation (FFD) registration algorithm [9], the unknown deformation field is parameterised by $N = n_x \cdot n_y \cdot n_z$ control points. In order optimise the parameters (control point values) Φ_i^ξ, with $i = 1 \dots N$ and $\xi \in \{x, y, z\}$, it is necessary to compute the energy gradient in parametric space rather than voxel space. Thus, the energy gradient is calculated w.r.t. the control point values, by taking the SGI and regularisation terms, and projecting them from voxel space to parameter space

$$\frac{\partial E}{\partial \Phi_i^\xi} = \frac{\partial E_D}{\partial \mathbf{u}} \cdot \frac{\partial \mathbf{u}}{\partial \Phi_i^\xi} + \lambda \frac{\partial E_R}{\partial \mathbf{u}} \cdot \frac{\partial \mathbf{u}}{\partial \Phi_i^\xi}. \tag{4}$$

For the B-spline FFD approach, the projection term $\frac{\partial \mathbf{u}}{\partial \Phi_i^\xi}$ is given by the tensor product of the 1D cubic B-splines

$$\frac{\partial \mathbf{u}}{\partial \Phi_i^\xi} = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_l(u) B_m(v) B_n(w) \tag{5}$$

where u, v, w correspond to relative positions in control point space. Finally, the update step used by the optimisation procedure is

$$\Phi_i^\xi \leftarrow \Phi_i^\xi + \eta \cdot \frac{\partial E}{\partial \Phi_i^\xi}. \tag{6}$$

2.2 Learning the Projection

We introduce an learning-based method (LB-FFD) to project the SGI from the gradient space of the energy functional to parameter space, that yields to faster convergence. In our setting, the projection is not constrained to be computed using the B-Spline tensor product. Instead, it is estimated as follows.

Given M training SGIs based on any similarity metric like normalised mutual information (NMI), sum of squared differences (SSD) or cross correlation (CC), the first step of our method is to extract patches $\mathbf{P}_{i,j} \in \mathbb{R}^{s_x \cdot s_y \cdot s_z \times 1}$ for all control point locations $i = 1 \dots N$ and training SGIs $\left(\frac{\partial E_D}{\partial \mathbf{u}}\right)_j$, $j = 1 \dots M$:

$$\mathbf{P}_{i,j} = \mathbb{P}_i^s \left(\frac{\partial E_D}{\partial \mathbf{u}}\right)_j \tag{7}$$

where $\mathbb{P}_i^s$ is an operator that extracts an intensity patch of size $s = s_x \cdot s_y \cdot s_z$ around the location of the control point i. The motivation of using these patches comes from the fact that the B-spline tensor model has local support, i.e., only

the voxels of the SGI within a neighbourhood of a control point have to be considered to perform the corresponding projections.

Since the patches $\mathbf{P}_{i,j}$ are of high dimensionality, a PCA dimensionality reduction step over the training images is performed, yielding low dimensionality patches $\mathbf{p}_{i,j} \in \mathbb{R}^{C\times 1}$

$$\mathbf{p}_{i,j} = \Gamma_i^\top \cdot (\mathbf{P}_{i,j} - \overline{\mathbf{P}_{i,j}}) \tag{8}$$

where $\overline{\mathbf{P}_{i,j}}$ is the mean patch over the training images and $\Gamma_i \in \mathbb{R}^{s_x \cdot s_y \cdot s_z \times C}$ is the basis matrix containing the first C modes of variation of the patches around control point i.

Finally, for each training SGI, a nonrigid FFD registration is performed between the target and source images that define it, yielding the optimal FFD control point values $\Phi_{i,j}^\xi$. After that, we estimate the projection term for each of these control points by fitting a partial least squares regression model on each direction ξ that relates the control point value with the low dimensionality patch centered on it. Thus, the regressed coefficients β_i^ξ satisfy

$$\Phi_{i,j}^\xi \approx \left(\beta_i^\xi\right)_0 + \sum_{k=1}^{C} \left(\beta_i^\xi\right)_k \cdot \left(\mathbf{p}_{i,j}\right)_k \tag{9}$$

At this point we regard $\Phi_{i,j}^\xi$ as a good approximation of the term $\frac{\partial E}{\partial \Phi_i^\xi}$ in (6), because it approximates a solution to the registration problem. As a consequence, if we use the regressed coefficients and the low dimentional patches to compute this approximation for the optimisation, the speed of convergence should increase.

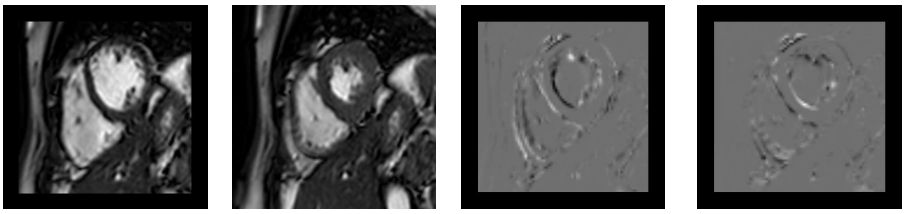
2.3 Registration of Unseen Images

Once the learning procedure described in the previous section is performed, it is possible to register an unseen image to any of the target images used to produce the training SGIs, like the ones depicted in figures 1(c) and 1(d). For this purpose, we devise an optimisation scheme similar to (6) using our approximation of the energy gradient, and an additional regularisation term to account for possible non-smoothness coming from errors in the regression model

$$\Phi_i^\xi \leftarrow \Phi_i^\xi + \eta \cdot \mathbf{G}_{\text{LB}} \tag{10}$$

$$\mathbf{G}_{\text{LB}} := \left(\left(\beta_i^\xi\right)_0 + \sum_{k=1}^{C} \left(\beta_i^\xi\right)_k \cdot \left(\mathbf{p}_{i,j}^*\right)_k \right) + \lambda_2 \frac{\partial E_R}{\partial \mathbf{u}} \cdot \frac{\partial \mathbf{u}}{\partial \Phi_i^\xi} \tag{11}$$

where $\lambda_2 \geq 0$ weights the regularisation term, and the unseen low dimensional patch $\mathbf{p}_i^*$ is computed by projecting the corresponding high dimensionality patch, taken from the SGI between the target and the current transformed source, using the PCA projection rule (8).



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

Fig. 1: One example of the cardiac data used and its similarity gradient in both x and y-direction. **a)** Target ROI. **b)** Source. **c)** SGI in x-direction. **d)** SGI in y-direction.

2.4 Multi-resolution Framework

The procedure described in section 2.2 is only valid for a single-resolution registration framework. In order to extend it to be able to perform in a multi-resolution framework with L levels starting from the coarsest level L down to the finest level 1, some considerations have to be made. For level L, the method needs no change, but for all other levels l with $1 \leq l \leq L-1$, the source images used to generate the training SGIs are previously deformed according to the FFD deformation field obtained down to level $l+1$ (while the target remains the same), and the FFD control point values used for regression are the difference between the FFD control point values down to level l and the FFD control point values down to level $l+1$.

3 Results

To test our method we perform two experiments, where we use the SSD between the target and source images as a data term, and the bending energy of the deformation field [24] as a regularisation term.

First, we generate 130 synthetic images by randomly scaling a target circle image in both x- and y-direction. 100 of these images are used in the training phase, by computing the SGI and performing FFD registrations between the target and source images that define them. The remaining 30 images are used for testing. The number of principal components retained is set to $C = 5$ and $\lambda = 0.001$. The value for λ_2 is set to 0.01 due to the unsmooth nature of some of the synthetic deformations.

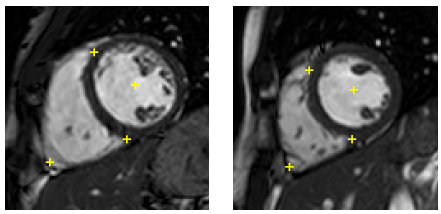


Fig. 2: Examples of the 4 landmarks used by the point-based affine registration.

We compare the mean number of iterations, mean error and total execution time over the 30 testing images of both the FFD approach and our learning-based method, in a single resolution setting. The results are summarised in Table 1.

Table 1: Mean iterations, mean error and total time for the FFD and LB-FFD methods using synthetic data

Algorithm	Iterations	Error (normalised SSD)	Total time (sec)
FFD	29.53	46.32	47.75
LB-FFD	2.70	48.44	3.88

It is possible to see that a reduction in the number of iterations for the learning-based approach is achieved and that, as a consequence, our algorithm performs faster, but mantaining good enough accuracy.

We also evaluate our method on 1.5T Philips Achieva MR cardiac sequences from 10 subjects covering one complete cardiac cicle over 30 frames. Fig. 1 depicts an example of the cardiac data used and its similarity gradient. From the obtained images, we extract one mid-ventricular short-axis slice to produce 2D sequences and then, since the images are on different spatial coordinates for different subjects, a point-based affine registration using 4 landmarks was used to align them, taking the first frame of one of the subjects as a reference. Fig. 2 shows two examples of the landmarks used.

For our experiments on cardiac data, we use $C = 15$ as the number of principal components to retain for our method, since we empirically found that almost no variation in the number of iterations needed to converge is appreciated for $C \in [5, 10, 15, 20, 25, 30]$. The values of λ and λ_2 are both set to 0.001.

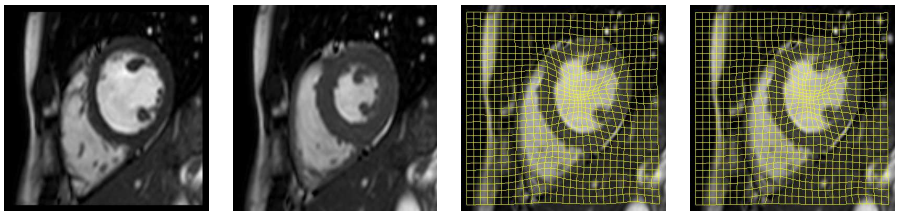
We perform a cross-validation procedure where a region of interest from the first frame of each subject is fixed as target image. Then, the second frame of each subject is registered to its corresponding target using both FFD registration with 3 resolution levels and our learning-based method trained using all other frames, also with 3 resolution levels. We repeat this process for the third frame, fourth frame, and so on. Table 2 summarizes the results over the 290 registrations performed for both methods.

Table 2: Mean iterations and total time for the FFD and LB-FFD methods using cardiac data

Algorithm	Iterations Lev. 3	Iterations Lev. 2	Iterations Lev. 1	Total time (sec)
FFD	4.57	7.73	7.93	94.52
LB-FFD	2.02	2.38	2.87	43.08

A clear reduction in the number of iterations can be seen on each resolution level, which in turn makes our algorithm achieve a speedup of over 50%.

Finally, to demonstrate the registration quality of our method, Fig. 3 depicts the registration result between the region of interest on the first frame of a subject and frame number 10, where the end of systole occurs, for both the FFD



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

Fig. 3: Registraton results for both FFD and learning-based methods with superimposed deformation grid. **a)** Target ROI. **b)** Source. **c)** FFD result with linear interpolation. **d)** LB-FFD result with linear interpolation.

and learning-based methods, with superimposed deformation grid. It can be seen that our learning-based method performs comparably to the FFD approach.

4 Conclusion

We developed a new general learning-based nonrigid registration approach to accelerate the convergence rate of any chosen parametric energy-based image registration method. A projection from the gradient space of the energy functional to the parameter space is learnt offline and subsequently used online to approximate the optimisation of the energy functional for unseen images. Our preliminary results from experiments on synthetic image data and MR cardiac sequences show that our approach significantly accelerates the convergence, both in number of iterations and total computational cost of the chosen registration method, while achieving similar results in terms of accuracy.

References

1. B.K.P. Horn and B.G. Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3):185–203, 1981.
2. G.E. Christensen, R.D. Rabbitt, and M.I. Miller. Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 5(10): 1435–1447, 1996.
3. J.P. Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical image analysis*, 2(3):243–260, 1998.
4. M.F. Beg, M.I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61(2):139–157, 2005.
5. J. Ashburner. A fast diffeomorphic image registration algorithm. *NeuroImage*, 38 (1):95–113, 2007.
6. J.A. Little, D.L.G. Hill, and D.J. Hawkes. Deformations incorporating rigid structures. In *Computer Vision and Image Understanding 66*, pages 223–232, 1997.
7. H. Yoshida. Removal of normal anatomic structures in radiographs using wavelet-based nonlinear variational method for image matching. In *Proceedings of SPIE*, volume 3458, page 174, 1998.

8. J. Ashburner and K.J. Friston. Nonlinear spatial normalization using basis functions. *Human brain mapping*, 7(4):254–266, 1999.
9. D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.
10. Y.T. Wu, T. Kanade, C.C. Li, and J. Cohn. Image registration using wavelet-based motion model. *International Journal of Computer Vision*, 38(2):129–152, 2000.
11. M. Fornefett, K. Rohr, and H.S. Stiehl. Radial basis functions with compact support for elastic registration of medical images. *Image and Vision Computing*, 19(1):87–96, 2001.
12. Z. Xue, D. Shen, and C. Davatzikos. Statistical representation of high-dimensional deformation fields with application to statistically constrained 3D warping. *Medical Image Analysis*, 10(5):740–751, 2006.
13. D. Loeckx, F. Maes, D. Vandermeulen, and P. Suetens. Temporal subtraction of thorax CR images using a statistical deformation model. *IEEE Transactions on Medical Imaging*, 22(11):1490–1504, 2003.
14. D. Rueckert, A.F. Frangi, and J.A. Schnabel. Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging*, 22(8):1014–1025, 2003.
15. S. Tang, Y. Fan, G. Wu, M. Kim, and D. Shen. RABBIT: rapid alignment of brains by building intermediate templates. *NeuroImage*, 47(4):1277–1287, 2009.
16. S. Pszczolkowski, L. Pizarro, R. Guerrero, and D. Rueckert. Nonrigid free-form registration using landmark-based statistical deformation models. In *Proceedings of SPIE*, volume 8314, page 831418, 2012.
17. Y. Tian and S.G. Narasimhan. A globally optimal data-driven approach for image distortion estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010*, pages 1277–1284. IEEE, 2010.
18. N. Batmanghelich, B. Taskar, and C. Davatzikos. Generative-Discriminative basis learning for medical imaging. *IEEE Transactions on Medical Imaging*, 2011.
19. G. Wu, F. Qi, and D. Shen. Learning best features and deformation statistics for hierarchical registration of MR brain images. In *Information Processing in Medical Imaging*, pages 160–171. Springer, 2007.
20. N.A. Ablitt, J. Gao, J. Keegan, L. Stegger, D.N. Firmin, and G.Z. Yang. Predictive cardiac motion modeling and correction with partial least squares regression. *IEEE Transactions on Medical Imaging*, 23(10):1315–1324, 2004.
21. M. Kim, G. Wu, P. Yap, and D. Shen. A General Fast Registration framework by learning deformation-appearance correlation. *IEEE Transactions on Image Processing*, (99):1823–1833, 2011.
22. D. Zikic, A. Kamen, and N. Navab. Natural gradients for deformable registration. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010*, pages 2847–2854. IEEE, 2010.
23. S. Klein, M. Staring, P. Andersson, and J. Pluim. Preconditioned stochastic gradient descent optimisation for monomodal image registration. *Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2011*, pages 549–556, 2011.
24. G. Wahba. *Spline models for observational data*, volume 59. Society for Industrial Mathematics, 1990.