A Comparison of Acceleration Techniques for Nonrigid Medical Image Registration

Stefan Klein, Marius Staring, and Josien P.W. Pluim

University Medical Center Utrecht, Image Sciences Institute,
Q0S.459, P.O. Box 85500, 3508 GA Utrecht, The Netherlands
{stefan, marius, josien}@isi.uu.nl.

Abstract. Mutual information based nonrigid registration of medical images is a popular approach. The coordinate mapping that relates the two images is found in an iterative optimisation procedure. In every iteration a computationally expensive evaluation of the mutual information's derivative is required. In this work two acceleration strategies are compared. The first technique aims at reducing the number of iterations, and, consequently, the number of derivative evaluations. The second technique reduces the computational costs *per iteration* by employing stochastic approximations of the derivatives. The performance of both methods is tested on an artificial registration problem, where the ground truth is known, and on a clinical problem involving low-dose CT scans and large deformations. The experiments show that the stochastic approximation approach is superior in terms of speed and robustness. However, more accurate solutions are obtained with the first technique.

1 Introduction

Nonrigid registration is an important technique in medical image processing. A popular class of registration methods is based on maximisation of the mutual information similarity measure, in combination with a deformation field parameterised by cubic B-splines [1,2]. The large computation time of this approach is a big disadvantage for many clinical applications. For practical use, acceleration is required.

Registration is usually stated as a minimisation problem:

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \mathcal{C}\left(\boldsymbol{\mu}; I_F, I_M\right) \quad , \tag{1}$$

where $\mathcal{C}(\boldsymbol{\mu}; I_F, I_M)$ denotes a cost function, and $\boldsymbol{\mu}$ a vector of parameters defining the deformation field that relates the fixed image I_F and the moving image I_M. In this paper the cost function is defined as the negated mutual information similarity metric, and the deformation field is parameterised by cubic B-splines, whose coefficients form the vector $\boldsymbol{\mu}$. To find the solution $\hat{\boldsymbol{\mu}}$ an iterative optimisation strategy is employed. In many optimisation methods the parameter update in each iteration k is based on the derivative of the cost function $\mathcal{C}$ with respect to the parameters $\boldsymbol{\mu}$. The gradient descent algorithm [3] is the most straightforward example:

J.P.W. Pluim, B. Likar, and F.A. Gerritsen (Eds.): WBIR 2006, LNCS 4057, pp. 151–159, 2006.

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k - a_k \boldsymbol{g}(\boldsymbol{\mu}_k), \quad k = 0, 1, 2, \ldots \ , \tag{2}$$

where $\boldsymbol{g}(\boldsymbol{\mu}_k)$ represents the derivative of the cost function, $\partial \mathcal{C}/\partial \boldsymbol{\mu}$, evaluated at $\boldsymbol{\mu}_k$. A scalar gain, a_k, controls the step size. Within certain conditions, the sequence $\{\boldsymbol{\mu}_k\}$ defined by (2) converges to a local minimum of the cost function.

Computation of the derivatives, $\boldsymbol{g}(\boldsymbol{\mu}_k)$, requires a considerable amount of computational effort in nonrigid registration problems. In this work two acceleration strategies are compared. The first technique aims at reducing the number of iterations, and, consequently, the number of required derivative evaluations. Well-known methods with an improved rate of convergence are the quasi-Newton and nonlinear conjugate gradient [3]. The second technique focusses on the computational costs *per iteration* by using approximations of $\boldsymbol{g}(\boldsymbol{\mu}_k)$. Acceleration factors are given with respect to the performance of the standard gradient descent method.

The acceleration strategies are compared in two types of experiments. Firstly, an artificially created problem is considered. An image is registered to itself, after application of a known, randomly generated deformation. Secondly, the registration of a low-dose expiration CT chest scan to a high-dose inspiration scan of the same patient is used as a test problem.

2 Nonrigid Registration Framework

In this section the various components of the nonrigid registration framework are described. The design of our algorithm is largely based on the papers by Mattes et al. [1], Rueckert et al. [2], and Thévenaz and Unser [4].

For computation of the mutual information the approach described in [4] is used. The joint histogram is constructed using B-spline Parzen windows, which makes it possible to formulate the mutual information as a continuous, differentiable function of the parameters describing the deformation field. In all experiments described in this paper, the joint histogram size is set to 32×32. The deformation field is parameterised by B-splines.

The minimisation problem (1) is solved with a multiresolution strategy. For the image data, we use a Gaussian image pyramid. The complexity of the deformation model is defined by the B-spline control point resolution. In our tests we let it follow the image resolution: when the image resolution is doubled, the control point resolution is doubled as well. The number of resolution levels and the final B-spline control point spacing depend on the specific problem. At each resolution a minimisation is performed, using one of the tested optimisation methods. When necessary, images are rigidly registered before the nonrigid registration.

3 Acceleration Strategies

A standard gradient descent algorithm, see (2), is used as a benchmark, to which further results are compared. The method is implemented using a slowly decaying gain sequence:

$$a_k = a/(A + k + 1)^\alpha \ , \tag{3}$$

where $a > 0$, $A \geq 0$, and $0 \leq \alpha \leq 1$ are user-defined constants.

3.1 Acceleration by Faster Convergence

In the literature many optimisation methods can be found with (theoretically) a better rate of convergence than the gradient descent. Perhaps the most well-known ones are the quasi-Newton and nonlinear conjugate gradient methods.

Quasi-Newton methods [3] use the following iterative scheme:

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k - a_k L_k \boldsymbol{g}(\boldsymbol{\mu}_k) \ . \tag{4}$$

In this equation, L_k is a positive definite matrix that serves as an approximation to the inverse Hessian of the cost function. For computation of the matrix L_k second order derivatives of the cost function are *not* needed; only the already computed first order derivatives are used. The scalar a_k is again a gain factor that controls the step size.

Several ways to construct the series $\{L_k\}$ are proposed in the literature. The method used in this work is a popular variant of the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm: the Limited memory BFGS (LBFGS) [5], which eliminates the need for storing the matrix L_k in memory.

Quasi-Newton methods are usually implemented in combination with an inexact line search routine, which determines a gain factor a_k that ensures sufficient progress towards the solution. In this work we use the line search routine described by Moré and Thuente [6]. If no gain factor can be found that gives sufficient progress, the optimisation is assumed to have converged.

Nonlinear conjugate gradient methods [3, 7] are based on the following iterative scheme:

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k + a_k \boldsymbol{d}_k \ , \tag{5}$$

where the search direction $\boldsymbol{d}_k$ is defined as a linear combination of the current derivative $\boldsymbol{g}(\boldsymbol{\mu}_k)$ and the previous search direction $\boldsymbol{d}_{k-1}$:

$$\boldsymbol{d}_k = -\boldsymbol{g}(\boldsymbol{\mu}_k) + \beta_k \boldsymbol{d}_{k-1} \ . \tag{6}$$

Many expressions for the scalar β_k can be found in the literature [7]. In this study we use a so-called 'hybrid' version, proposed in [8] and shown to be very efficient compared to other methods. The gain factor a_k is determined by the same inexact line search routine as used with the quasi-Newton method.

3.2 Acceleration by Stochastic Approximation

By using approximated derivatives instead of the exact ones the computation time per iteration can be reduced significantly.

The computation time of the derivative of mutual information is linearly dependent on the number of voxels $|I_F|$ in the fixed image, and on the number of B-spline coefficients N (the length of the parameter vector $\boldsymbol{\mu}$):

$$t_{\boldsymbol{g}(\boldsymbol{\mu}_k)} \sim p|I_F| + qN \ , \tag{7}$$

where p and q are constants. For most medical nonrigid registration problems $p|I_F|$ tends to be much larger than qN. It is clear that we can lower the computation time of a derivative evaluation by not using all the voxels, but only a small subset of voxels.

In [9] it is demonstrated that, when using a new, randomly selected, subset of voxels in every iteration of the optimisation process, the parameter sequence $\{\boldsymbol{\mu}_k\}$ still converges to the correct solution. Selecting a new subset of voxels in every iteration ensures that the approximation errors will, on average, cancel each other out. The approximation errors can be considered a source of noise $\boldsymbol{\varepsilon}_k$ entering the optimisation process:

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k - a_k \left(\boldsymbol{g}(\boldsymbol{\mu}_k) + \boldsymbol{\varepsilon}_k \right) \quad . \tag{8}$$

This scheme is often referred to as a *stochastic gradient descent algorithm* or a Robbins-Monro procedure [10, 11]. It can be proven that the sequence $\{\boldsymbol{\mu}_k\}$ defined by (8) still converges to the solution $\hat{\boldsymbol{\mu}}$, provided that the *bias* of the approximation error goes to zero.

The experiments in [9] indicate that for the registration of large 3D images as few as 2048 voxels are required in each iteration, which is adopted in our tests. The gain sequence $\{a_k\}$ is defined as in the gradient descent method, see (3).

3.3 Combining the Acceleration Strategies

Naturally, the question rises whether it is possible to combine the two acceleration strategies. Unfortunately, the quasi-Newton and nonlinear conjugate gradient optimisation methods are not designed to work with stochastic approximations of the derivatives. They expect noise-free derivatives to be available.

A possible strategy for these methods is to select *a single* subset of voxels in the fixed image and use these samples throughout the registration process [12,1]. A disadvantage of this method is that convergence to the correct solution cannot be guaranteed, because the approximation error bias does not go to zero.

In our tests with quasi-Newton and nonlinear conjugate gradient the samples are selected on a regular grid using identical downsampling factors for each image dimension.

4 Experiments and Results

In two types of nonrigid registration problems, the following methods are compared:

- Gradient Descent (GD), see Sec. 3,
- Quasi-Newton (QN-*df*), see Sec. 3.1,
- Nonlinear Conjugate Gradient (NCG-*df*), see Sec. 3.1,
- Stochastic Gradient Descent (SGD), see Sec. 3.2.

The extension *df* denotes the downsampling factor. Downsampling factors of 1 (full image), 2, 4, 8, and 16 are tested. In case of GD, the full image is always

used. As explained in Sec. 3.2, the SGD method uses only 2048 voxels to calculate the derivatives.

For all optimisation methods, the computation time per iteration is assumed to be dominated by the time required for computing the derivative of the mutual information. Moreover, the derivative's computation time is assumed to be mainly related to the number of voxels used. These simplifications allow us to define the *normalised computation time (NCT)* up to iteration k of the optimisation process:

$$\mathrm{NCT} = (k+1)V/|I_F| \ , \tag{9}$$

with V the number of voxels used to compute the derivative.

4.1 Artificial Deformation Fields

In the first evaluation procedure an image I is registered with a deformed version of itself. To avoid interpolation errors, the deformation field $\tilde{\boldsymbol{v}}$ is added to the B-spline deformation field $\boldsymbol{v}_{\boldsymbol{\mu}}$ that is updated during optimisation. Since the image I is registered with itself, the desired solution is a total deformation field that is zero everywhere. The ground truth is known, so an error measure, the average displacement error e, can be defined:

$$e(\boldsymbol{\mu}) = \frac{1}{|I|} \sum_{\boldsymbol{x}_i \in I} \| \tilde{\boldsymbol{v}}(\boldsymbol{x}_i) + \boldsymbol{v}_{\boldsymbol{\mu}}(\boldsymbol{x}_i) \| \ , \tag{10}$$

where $\boldsymbol{x}_i$ is the position of voxel i in the image volume I, and $|I|$ the total number of voxels in I. The speed of convergence of a method is visualised by plotting this error measure against the normalised computation time NCT.

The experiments are performed on four 3D CT images of the heart. The images originate from chest scans. These were cropped to the area of the heart and downsampled by a factor of two, resulting in images of $97 \times 97 \times 97$ voxels with an isotropic size of 1.4 mm. For each image a deformation field $\tilde{\boldsymbol{v}}$ is generated, composed of randomly placed Gaussian blobs with a standard deviation of 14 mm. A $10 \times 10 \times 10$ grid of B-spline control points defines the deformation field $\boldsymbol{v}_{\boldsymbol{\mu}}$, yielding 3000 parameters to be optimised. No multiresolution schemes are used in this experiment, which makes comparison of the results more straightforward. The maximum number of iterations is limited to 2048. Three constants must be set for the gain sequence in (3): $a = 3200$, $A = 50$, and $\alpha = 0.602$.

In the following we present the test results for one of the four images. The outcome for the other images is similar. Figure 1 shows the average displacement error as a function of the normalised computation time for the tested methods. A logarithmic scale is used for the time axis. It is clear that both acceleration strategies realise considerable speed improvements compared to the standard GD procedure. The QN and NCG method do not give acceptable results anymore with a downsampling factor of 8, which results in 2197 selected voxels. The methods QN-16 and NCG-16, which are omitted from the figure, perform even worse. Without downsampling QN and NCG achieve a slightly higher accuracy

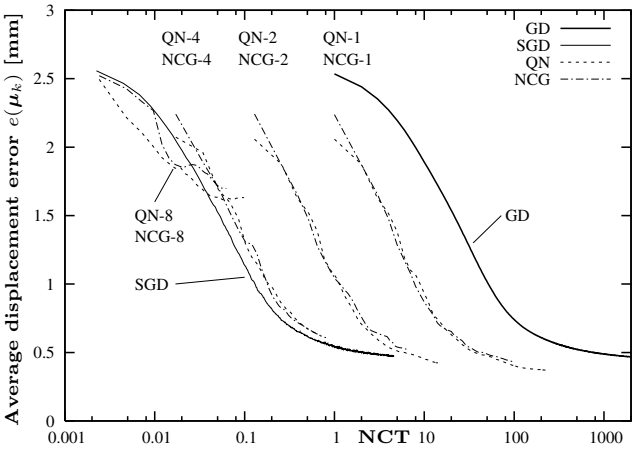

Fig. 1. Results for one of the heart images. The graph shows the average displacement error as a function of the normalised computation time for the tested methods.

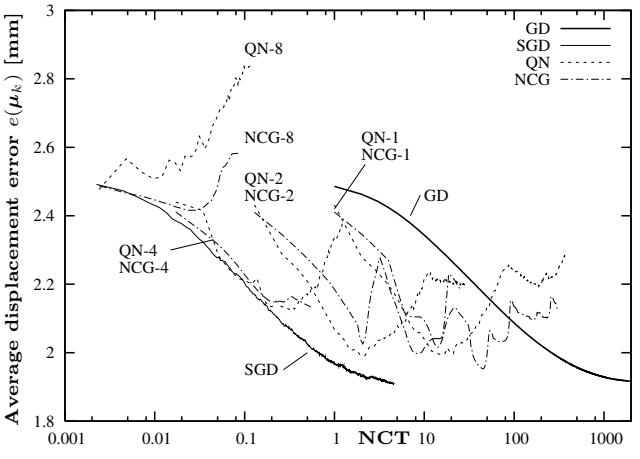

Fig. 2. The imposed deformation field is composed of Gaussian blobs with a standard deviation of 7 mm. The graph shows the average displacement error as a function of the normalised computation time for the tested methods.

than the GD and SGD method. The SGD method, which works well with only 2048 voxels, is clearly the fastest.

The tests are repeated for a more difficult registration problem. The imposed deformation field $\tilde{\boldsymbol{v}}$ is composed of Gaussian blobs with a standard deviation of 7 mm. This smaller standard deviation results in a deformation field that is very

hard to recover, since the B-spline control point grid used during registration is not dense enough. The test results for the same image as before are shown in Fig. 2. It is interesting to see that the QN and NCG methods can not handle this very ill-posed registration problem. The GD and SGD procedures remain stable. As expected, none of the optimisation methods is able to achieve a very large reduction of the initial average displacement error, since the B-spline control point grid is not fine enough.

Note that the QN and NCG methods *do* still find a set of parameters that decrease the mutual information. The Moré-Thuente line search, employed in both QN and NCG to set the gain factor a_k, guarantees that the cost function decreases in every iteration, $\mathcal{C}(\boldsymbol{\mu}_k) < \mathcal{C}(\boldsymbol{\mu}_{k-1})$. However, in an ill-posed problem a decreasing cost function does not imply that actual progress is made towards the correct solution.

4.2 Clinical Data

In this section a number of experiments with 3D CT chest scans are described. The patients were scanned after inspiration and after expiration. The inspiration scans were recorded with a high radiation dose; the expiration scans with a low dose. The large deformations in combination with the noisy nature of the low-dose scans make this a challenging registration problem.

The images were acquired with a Philips Mx8000IDT 16-slice CT scanner. We use data of seven patients. The original images, with in-plane dimensions of 512×512 and a number of slices ranging from 400 to 800, were downsampled by a factor of 2 in each dimension to be able to register the images on a standard PC with one gigabyte of memory. The resulting voxel size is approximately 1.4 mm in all directions, and the images consist of about 10^7 voxels.

A four-level multiresolution approach is applied. At each resolution the number of iterations is fixed to 256. At the highest resolution the B-spline control point spacing is set to 22 mm, yielding a grid of about 19^3 control points (approximately 20000 parameters). The following settings are used for the gain sequence in GD and SGD: $a = 60000$, $A = 50$, and $\alpha = 0.602$.

As is common in clinical applications of nonrigid registration, the ground truth is not known. We assess the registration results by computing the overlap of the lungs, L_1 and L_2:

$$\text{Overlap} = \frac{2 \cdot |L_1 \cap L_2|}{|L_1| + |L_2|} \quad . \tag{11}$$

Segmentations of the lungs were made by means of a method based on the work of Hu et al. [13]. In the segmentations large pulmonary vessels are not considered part of the lungs. For reference: a translation of one voxel in each dimension results in a lung overlap of about 0.95.

Table 1 shows the overlap measures after rigid registration and nonrigid registration with the studied methods. Each column displays the results for a single

Table 1. The results of CT chest scan registration for patients P1-P7. The overlap of segmented lungs is shown after rigid registration (first row), and after nonrigid registration with different methods. Each column contains the results for a single patient.

	P1	P2	P3	P4	P5	P6	P7
rigid	0.736	0.453	0.729	0.883	0.654	0.811	0.671
GD	0.925	0.802	0.951	0.972	0.951	0.975	0.914
SGD	0.922	0.799	0.945	0.966	0.945	0.972	0.909
QN-1	0.932	0.855	0.956	0.974	0.962	0.979	0.930
QN-2	0.931	0.836	0.955	0.973	0.960	0.979	0.926
QN-4	0.926	0.815	0.951	0.971	0.955	0.975	0.916
QN-8	0.905	0.797	0.934	0.960	0.936	0.965	0.898
QN-16	0.862	0.751	0.906	0.938	0.897	0.946	0.858
NCG-1	0.931	0.795	0.941	0.966	0.942	0.969	0.929
NCG-2	0.912	0.796	0.955	0.973	0.961	0.969	0.904
NCG-4	0.911	0.792	0.940	0.966	0.943	0.969	0.903
NCG-8	0.895	0.794	0.935	0.959	0.934	0.965	0.894
NCG-16	0.854	0.752	0.903	0.938	0.900	0.946	0.858

patient. All methods result in a considerable improvement on the rigid registration. The overlap measures confirm the results found in Sec. 4.1. The final accuracy of the nonrigid registration is, compared to GD, very little affected by the random subsampling strategy employed by SGD. The methods QN-1, QN-2, NCG-1, and NCG-2 result in a somewhat better accuracy than GD and SGD. With higher downsampling factors the accuracy decreases.

The most remarkable results are found in patient P2. The QN-1 and QN-2 methods seems to outperform all other methods. However, visual inspection of the results taught us that the good overlap results come at the price of some very unrealistic deformations. In patients P5 and P7 the same problem was observed for the QN and NCG methods. The GD and SGD procedures only have this problem in patient P5. This is in line with the results of Sec. 4.1, where GD and SGD also seem to be more robust than QN and NCG.

5 Conclusion

We have compared acceleration techniques for both artificially deformed and real clinical data. The experiments indicate that SGD achieves the largest acceleration, and seems to be more robust for badly defined problems than the QN and NCG algorithms. Without downsampling QN and NCG yield a slightly smaller error than GD or SGD. Downsampling increases the error, and does not result in the same acceleration as obtained by SGD.

In summary, we can conclude that the acceleration technique focussing on reduction of the computational costs per iteration is the preferred approach.

Acknowledgements

Funding of this research has been provided by the Netherlands Organisation for Scientific Research (NWO). Additionally, this work benefited from the use of the Insight Segmentation and Registration Toolkit (ITK), an open source software developed as an initiative of the U.S. National Library of Medicine and available at `http://www.itk.org`.

References

1. Mattes, D., Haynor, D.R., Vesselle, H., Lewellen, T.K., Eubank, W.: PET-CT image registration in the chest using free-form deformations. IEEE Transactions on Medical Imaging **22** (2003) 120–128
2. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: Application to breast MR images. IEEE Transactions on Medical Imaging **18** (1999) 712–721
3. Nocedal, J., Wright, S.J.: Numerical optimization. Springer-Verlag, New York (1999)
4. Thévenaz, P., Unser, M.: Optimization of mutual information for multiresolution image registration. IEEE Transactions on Image Processing **9** (2000) 2083–2099
5. Nocedal, J.: Updating quasi-Newton matrices with limited storage. Mathematics of Computation **35** (1980) 773–782
6. Moré, J.J., Thuente, D.J.: Line search algorithms with guaranteed sufficient decrease. ACM Transactions on Mathematical Software **20** (1994) 286–307
7. Dai, Y.H.: A family of hybrid conjugate gradient methods for unconstrained optimization. Mathematics of Computation **72** (2003) 1317–1328
8. Dai, Y.H.: An efficient hybrid conjugate gradient method for unconstrained optimization. Annals of Operations Research **103** (2001) 33–47
9. Klein, S., Staring, M., Pluim, J.P.W.: Comparison of gradient approximation techniques for optimisation of mutual information in nonrigid registration. In Fitzpatrick, J., Reinhardt, J., eds.: SPIE Medical Imaging: Image Processing. Volume 5747 of Proceedings of SPIE., SPIE Press, Bellingham, WA (2005) 192–203
10. Robbins, H., Monro, S.: A stochastic approximation method. The Annals of Mathematical Statistics **22** (1951) 400–407
11. Kushner, H.J., Clark, D.S.: Stochastic approximation methods for constrained and unconstrained systems. Springer-Verlag, New York (1978)
12. Kybic, J., Unser, M.: Fast parametric elastic image registration. IEEE Transactions on Image Processing **12** (2003) 1427–1442
13. Hu, S., Hoffman, E.A., Reinhardt, J.M.: Automatic lung segmentation for accurate quantitation of volumetric X-Ray CT images. IEEE Transactions on Medical Imaging **20** (2001) 490–498