Fast optimization methods for image registration in adaptive radiation therapy

Qiao, Y.

Author: Qiao, Y.

Title: Fast optimization methods for image registration in adaptive radiation therapy

Issue Date: 2017-11-01

6

Discussion and conclusion

Image registration is important for medical image analysis. However, its clinical application is sometimes limited by the speed of the algorithm. For example, in online adaptive radiation therapy a few seconds is ideal, while it usually takes several minutes, at the least. In this thesis, we consider acceleration techniques for parametric intensity-based image registration problems focussing on the optimization routine, specifically the step size and the search direction. The different proposed methods are thoroughly evaluated on different datasets across modalities, subject, similarity measures and transformation models. Depending on the registration settings, the estimation time of the step size is reduced from 40 seconds to less than 1 second when the number of parameters is 10^5, almost 40 times faster. The total registration time of new acceleration techniques (FASGD) is reduced by a factor of 2.5-7x compared with ASGD for the experiments in this thesis. All methods were implemented using C++ in the open source registration package `elastix`. Based on these acceleration schemes we evaluated `elastix` on the application of automatic contour propagation in online adaptive intensity modulated proton therapy for prostate cancer.

6.1 Summary

A summary of the thesis is given below:

Chapter 2 Step size selection is important for gradient descent optimization. It is difficult to perform manually, because for image registration different fixed or moving images, different similarity measures or transformation models require a different step size. Klein *et al.* proposed a method to automatically estimate the step size, however, for a large number of transformation parameters, i.e. in the order of 10^5 or higher, the runtime is unacceptable and the time used in estimating the step size will dominate the optimization [45]. In this chapter, a new automatic method (FASGD) for estimating the optimization step size parameter a, needed for gradient descent optimization methods, has been presented for image registration. The parameter a is automatically estimated from the magnitude of voxel displacements, randomly sampled from the fixed image. A relation between the step size and the expectation and variance of the observed voxels displacement is derived. The proposed method has a free parameter δ, defining the maximally allowed incremental displacement between iterations. Unlike a, it can be interpreted in terms of the voxel size (mm). In addition, it is mostly independent of the application domain, i.e. setting it equal to the voxel size provided good results for all applications evaluated in this chapter. Compared to the original ASGD method, the time complexity of the FASGD method is reduced from quadratic to linear with respect to the dimension of the transformation parameters P. For the B-spline transformation, due to its compact support, the time complexity is further reduced, making the proposed method independent of P. The FASGD method is publicly available via the open source image registration toolbox `elastix` [10]. The FASGD method was evaluated on a large number of registration scenario's and shows a similar accuracy as the original ASGD method. It however improves the time complexity of the step size estimation from 40 seconds to no more than 1 second, when the number of parameters is $\sim 10^5$: almost 40 times faster. Depending on the registration settings, the total registration time is reduced by a factor of 2.5-7x for the experiments in this chapter.

Chapter 3 This chapter presents a stochastic quasi-Newton optimization method (s-LBFGS) for non-rigid image registration. It uses the classical limited memory BFGS method in combination with noisy estimates of the gradient. Curvature information of the cost function is estimated robustly once every L iterations and then used for the next L iterations in combination with stochastic gradients. A novel restarting procedure, automatically selecting the optimization step size, is shown to be beneficial for accelerated convergence. The new optimization routine is validated on follow-up data of 3D chest CT scans (19 patients). Compared to ASGD the proposed method uses about 5 times fewer iterations to reach the same metric value, resulting in an overall reduction in run time of a factor of two. Compared to deterministic LBFGS, s-LBFGS is almost 500 times faster.

Chapter 4 A generic preconditioner estimation method was proposed in this chapter for the stochastic gradient descent optimizers used in medical image registration. Based on the observed distribution of the voxel displacements, this method automatically constructs a diagonal preconditioner, avoiding the computationally complex calculation of the Hessian matrix. We performed experiments to compare our method with FASGD and also other preconditioning techniques: Jacobian type preconditioned stochastic gradient descent (PSGD-J) [70] and Hessian type preconditioned stochastic gradient descent (PSGD-H) [70]. All tested methods obtained a similar final registration accuracy in all tested datasets. The proposed FPSGD optimizer, however, outperforms FASGD and PSGD-J in terms of convergence rate, while yielding a similar computational overhead. While a previous method (PSGD-H) even further reduces the required number of iterations, it comes at a substantial overhead in computing the preconditioner, especially for high dimensional transformations. Additionally, PSGD-H can only be used in mono-modal problems and requires the implementation of a Hessian matrix computation. We conclude that the proposed method can act as a generic preconditioner for optimization in registration methods, yielding similar accuracy as gradient descent routines while substantially improving the convergence rate with a speedup by a factor of 2-4.

Chapter 5 In this chapter we showed that by integration of our algorithms in the open source registration package `elastix`, repeat CT scans of the prostate can be automatically re-contoured for adaptive online IMPT. The online adaption of IMPT could allow for small margins of the target (the prostate), leading to less complications. The dosimetric performance was evaluated with a margin of 2mm, 3.5mm and 3.5mm for the prostate, seminal vesicles and lymph nodes, respectively. The fastest setting of 13 seconds yielded a promising clinical acceptance rate of 83 in 93 cases (89.2%), 73 in 93 (78.5%), and 91 in 93 (97.9%) in dosimetric coverage for the prostate, seminal vesicles and lymph nodes, respectively. We conclude that the fast setting of open source `elastix` can automatic re-contour the daily scans that meet the dose conformity constraints in 78.5% to 97.9% of cases, and a geometrical criteria of success in 96% of cases. This software may therefore facilitate online adaptive proton therapy of prostate cancer, enabling a reduction in treatment margins.

6.2 Discussion

The aim of this thesis was to accelerate the procedure of image registration for clinical applications, such as online adaptive radiation therapy. The image registration procedure includes sampling, transformation, optimization, interpolation and similarity measure selection [10], where the core part and most time-consuming component is optimization. From the evaluation of different optimization strategies performed by Klein *et al.* [25], we realized that the speed of image registration could be improved by subsampling. Later he proposed adaptive stochastic gradient descent (ASGD), which is powerful and achieved a good registration accuracy and fast convergence speed in terms of runtime. However, this method is still not fast enough for large scale problems, for example for 10^6 transformation parameters and 3D volumetric image registration. If the runtime of ASGD could be further reduced, the speed of image registration will be accelerated.

We first focus on the core part of the optimization algorithm, specifically in finding a suitable initial step size and determining a search direction yielding a faster convergence rate. These two parts are not only important for stochastic gradient type methods but also for deterministic gradient type methods. In Chapter 2 we found that the good initialization of the step size is important for the optimization and especially for stochastic gradient type methods. Therefore, different schemes to choose a suitable step size are still an actively pursued topic in the optimization field. Considering search direction schemes, first order gradient methods have the inherit shortcoming of a linear or sublinear convergence rate. A new scheme using second order gradient information with stochastic gradient was thereby proposed and first used in medical image registration problems in Chapter 3. This chapter provides an insight to take use of second order gradient information with an averaging and restarting scheme, both of which are useful to improve the optimization speed. Besides directly using the second order gradient information, we found that we could scale or transform multi-variate problems from an ill-conditioned status to a well-conditioned one at the very beginning of the optimization. A more generic preconditioning scheme was then proposed for (stochastic) gradient descent methods in Chapter 4. The experimental results evaluated on different clinical data had shown that the proposed methods work well for different image registration problems parameterized by different transformation models and different similarity measures. In summary, the step size and the search direction are both important, and essential parts for optimization of image registration.

The successful speedup of the proposed methods on the evaluated datasets encourages further research in this direction. The possibilities are either on the adaptive step size selection or on the efficient search direction scheme. There are many other adaptive step size selection methods such as Adagrad [138], Adadelta [139] and Adaptive Moment Estimation (Adam) [140], which use the properties of the current gradient together with the past gradient. For the search direction, the conjugate stochastic gradient descent [141], variance reduction stochastic gradient descent [142], stochastic gradient descent with momentum [143, 144] and others provide some avenues for future research. Reducing the variance of the stochastic gradient estimation may improve the convergence rate and averaging the stochastic gradient can reduce the noise. Finally, combining these two techniques may yield of further performance improvement. For iterative optimization schemes, these three proposed methods, fast initial step size estimation, stochastic second order gradient method and fast preconditioning scheme, are not independent. Our fast preconditioning scheme was based on the work of fast initial step size estimation, so this scheme could also be used to accelerate second order gradient methods.

In this thesis, we applied these proposed methods to online adaptive IMPT for the prostate cancer and achieved a registration time of 13 seconds for automatic propagation of the contours for most cases. We found that this speed is fast enough for the current procedure of online adaptive IMPT. There are still some aspects that should be considered for improvement. The registration time assessed in this thesis is determined by the number of iterations, which is not case-specific [137, 100]. For cases that are geometrically close, image registration may finish the task with less than the average required iterations, while for difficult cases the number of iterations may be much larger. An adaptive stopping condition for stochastic gradient descent, such as considering a moving average of the noisy cost function values (or gradients), may remedy this. To apply this in clinical practice, the robustness of image registration is also critical. Robustness may be further improved by taking into account automatic estimates of targets for instance the bladder in the prostate cancer [135] in the registration by optimizing a joint function.

Besides the speedup improvements in the optimization, there are several approaches to further accelerate the image registration procedure. The first is the importance-driven sampling strategy used in image registration [44], which could reduce the runtime and improve the registration accuracy. Second, more efficient calculation techniques in the transformation models have become available, such as using a non-uniform cubic B-spline transformation model [145], using a fast recursive implementation [122], and using a random perturbation to smooth the B-spline control grid [146, 103]. Thirdly, fast implementation of for example the interpolator is also useful for acceleration, for example, Shamonin *et al.* [22] proposed to use the GPU for the acceleration. Lastly, other strategies with learning could be applied, such as fast image registration using prior knowledge [147].

6.3 Conclusion

In this thesis we developed several stochastic optimization methods for fast image registration, leading to a 5-10 fold speedup over previous approaches. All proposed methods are implemented using C++ and integrated in the open source registration package `elastix`. We also exploited the usage of high performance computation resources – the life science grid (`lsgrid`) to perform over 10^6 registrations, which significantly reduced computation time for large scale computational tasks. As we have evaluated the proposed method in the application of online adaptive IMPT for prostate cancer, we expect that these methods can achieve the desirable performance for use in clinical practice.