Fast optimization methods for image registration in adaptive radiation therapy

Qiao, Y.

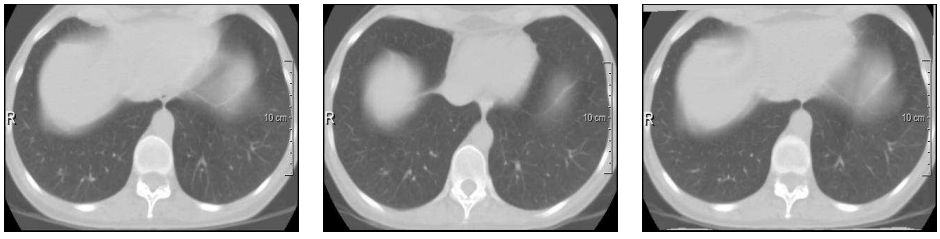
1

Introduction

1.1 Medical image registration

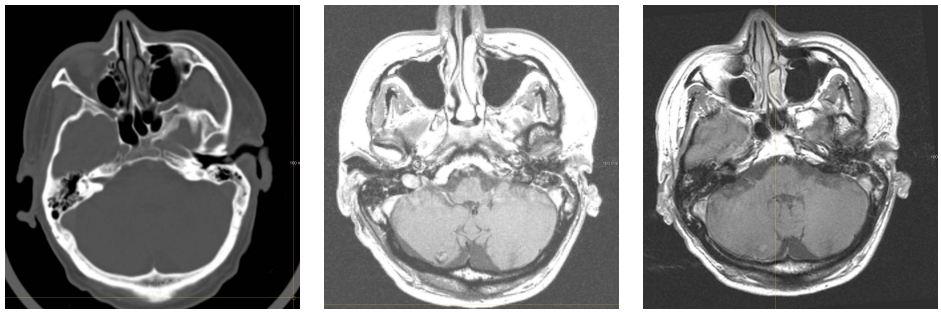
Medical imaging has become an indispensable tool in health care for diagnosis, treatment planning and therapy monitoring. In many cases, medical images are acquired at different stages of the diagnosis and treatment chain. However, medical imaging data is often very heterogenous, in that it can be acquired at different time points (to monitor disease course), or at different imaging devices (providing complementary information). In many cases, the anatomical structures in the images may move or deform due to internal movement (e.g. breathing, bladder filling or cardiac motion) or external function differences between imaging modalities. Also, in studies across multiple subjects, the anatomical structures of different subjects may also differ a lot due to inter-individual differences. The main goal of medical image registration is to find the spatial connection between heterogeneous images or populations.

With the increasing use of medical imaging in routine clinical care, medical image registration is an important driver for the development of innovative image analysis technologies. Application examples are CT screening for lung cancer, atlas-based segmentation and image-guided interventions [1, 2, 3]. For instance, in CT screening for lung cancer, follow-up CT scans of the same subject are compared against a



(a) Fixed image (b) Moving image (c) Registered moving image

Figure 1.1: Example of deformable image registration on lung CT images.



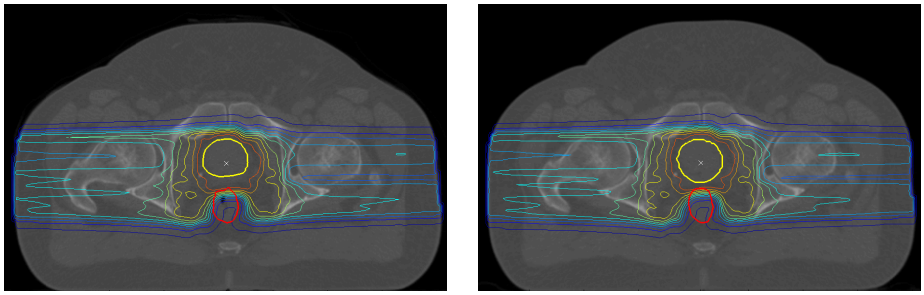
(a) CT image (b) MRI image (c) Registered image

Figure 1.2: Example of CT-MRI registration for target-volume delineation of brain tumors.

baseline CT scan, and a comparison is performed to assess the tumor changes. Even though lung CT scans are acquired at more or less standardized respiration stages, the deformations of the lung can be large. It is essential to register the CT scans to investigate the tumor development in the lung with respect to normal tissues. Figure 1.1 shows an example of deformable image registration to register a follow-up lung CT scan to a baseline CT scan. Besides mono-modal image registration, multi-modal image registration is also used frequently. For example, it can be used to delineate the target volume of brain tumors for the same patient. An example is shown in Figure 1.2. As CT imaging and MR imaging have different resolution and different tissue contrast properties, image registration could integrate these sources of information and provide a better observation of the tumor size change.

In image-guided interventions, for instance image-guided radiation therapy, a planning CT scan is acquired, based on which a treatment plan is generated. The total dose in the treatment plan is usually delivered in daily fractions. The treatment, in particular proton therapy, is sensitive to daily changes in patient setup, the location and shape of the tumor and target volume, and changes in tissue density along the proton beam path. These changes can be captured with the acquisition of a daily CT scan as shown in Figure 1.3. The induced uncertainties by these changes could dramatically distort the dose distribution compared to the planned dose distribution [4, 5, 6, 7, 8]. To achieve highest possible accuracy, the planned dose distribution need therefore be adjusted for the deformations of the tumors over the course of the treatment, which can be computed by image registration.

The procedure of online adaptive image-guided radiation therapy requires a fast, online image registration to automatically and efficiently re-contour the target and organs-at-risk (OARs) of repeat CT scans by establishing the spatial correspondence with the planning CT scan. Image registration then enables the use of small margins and high robustness without losing dose coverage. It is of high practical importance that image registration can be performed on the fly, so that treatment adaptation can be applied before new intra-fraction motions occur in the patient [9]. Nowadays, the registration computing time is quite long (usually several minutes), and it is difficult for image registration to obtain the optimal solution within a few seconds due to the



(a) Planning CT image (b) Daily CT image

Figure 1.3: Example of organ motion in the planned dose distribution for IMPT of prostate cancer. The prostate and rectum are delineated and represented as a yellow and red solid line, respectively. The shape change of the prostate can be observed.

complicated cost functions, transformation models and optimization methods.

It would therefore be highly desirable to accelerate the procedure of image registration, to enable its use in real-time interventions.

1.2 The image registration framework and acceleration approaches

Many approaches can be applied for image registration, such as feature-based image registration and intensity-based image registration. As intensity-based image registration is widely used and most algorithms are developed based on it, we focus on this type of problem in this thesis.

The procedure of intensity-based image registration can be formulated as a parametric optimization problem to minimize the dissimilarity between a d-dimensional fixed image I_F and moving image I_M:

$$\widehat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \mathscr{C}(I_F, I_M \circ \boldsymbol{T}_{\boldsymbol{\mu}}), \tag{1.1}$$

in which $\boldsymbol{T}_{\boldsymbol{\mu}}(\boldsymbol{x})$ is a coordinate transformation parameterized by $\boldsymbol{\mu}$. Often used dissimilarity measures $\mathscr{C}$ for intensity-based image registration include mutual information (MI), normalized correlation (NC) and the mean squared intensity difference (MSD) [1, 2, 3]. To account for rotations, translation, global scaling, shrinking and local deformations that occur in medical images, different transformation models are adopted including the translation transform, affine transform and B-spline transform. In particular, complex local deformation models require more degrees of freedom for the transformation models, and are thus more computationally expensive. Multi-resolution strategies on both the image data and the transformation model, allow for a fast and robust image registration [10].

To solve this registration optimization problem, the following iterative scheme is commonly used:

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k - \gamma_k \boldsymbol{d}_k, \tag{1.2}$$

where k is the iteration number, γ_k is the step size at iteration k, and $\boldsymbol{d}_k$ is the search direction in the parameter space. For fast registration methods the search direction $\boldsymbol{d}_k$ as well as the estimation of the step size γ_k need to be performed with high efficiency.

Gradient descent directions are widely used for the search direction $\boldsymbol{d}_k$. Gradient-type search directions include steep gradient descent, conjugate gradient descent, Newton gradient descent and their ***stochastic*** variations. Because of the exponential growth of data and parameter spaces in the past twenty years, the computational burden eventually became a bottleneck to find the optimal solution. Stochastic variations of these methods are therefore commonly used with its commendable properties: efficient implementation, little computational burden per iteration and overall less computation cost. These type of methods approximate the deterministic gradient by subsampling the fixed image. However, the inherent drawbacks of stochastic gradient methods are its slow convergence rate and unstable oscillations even if sufficient iterations are provided.

To improve the convergence rate, there are two common approaches. One can use the second order gradient to capture the curvature information of the cost function. A different option is the use of a preconditioning scheme to transform an ill-conditioned cost function to a well-conditioned one at the very beginning of the optimization. Both are well-known for deterministic gradient type methods. For ***stochastic*** type gradient methods, the noise in the curvature calculation or the preconditioner estimation may amplify the errors, resulting in a slow convergence rate or a failed registration. Besides this, the calculation of the Hessian or the preconditioner should also be fast, otherwise the gain in the convergence will be lost. New schemes of fast calculation of the Hessian or the preconditioner for ***stochastic*** type gradient methods are therefore needed.

Besides the acceleration schemes in the calculation of the search direction $\boldsymbol{d}_k$, the selection of the step size γ_k is also important. There are two classes of methods to determine the step size γ_k: exact and inexact methods. An exact way could be the conjugate gradient method to determine the step size. An example of an inexact approximation uses for instance a line search method to find the step size that satisfies the Wolfe conditions [11]. However, both schemes are developed for deterministic optimization methods and could not guarantee the convergence of stochastic type methods. For stochastic methods such as stochastic gradient descent, the step size is also an important condition to ensure the convergence, which should meet the following constraints [12, 13, 14, 15],

$$\sum_{k=1}^{\infty} \gamma_k = \infty, \qquad \sum_{k=1}^{\infty} \gamma_k^2 \quad < \infty. \tag{1.3}$$

A common choice for the step size that satisfies this constraint is a monotonically non-increasing sequence. Consider the following decay function for the stochastic methods:

$$\gamma_k = \frac{a}{(A+k)^{\alpha}}, \tag{1.4}$$

with $a > 0$, $A \geq 1$, $0 < \alpha \leq 1$, where $\alpha = 1$ gives a theoretically optimal rate of convergence [16].

As we can see, in Equation (1.4) the selection of a is important. In medical image registration, this selection is case-specific for different transformation models

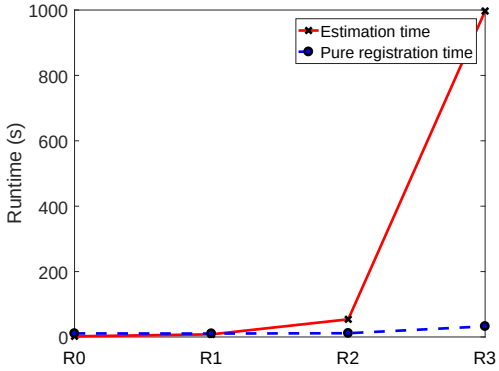


Figure 1.4: An example of runtime in seconds of ASGD for the mutual information measure and a B-spline transformation model. The blue line is the pure registration time and the red line the estimation time of the step size. R0 until R3 are the four resolution levels. The number of transformation parameters for these four resolutions is around 10^3, 10^4, 10^5 and 10^6, respectively. It can be seen that the estimation time for the step size becomes even larger than the registration time.

and different similarity measures. To select a good step size, the magnitude of a should be not too large, otherwise the estimated optimal value of cost function will be "bouncing", and not too small, otherwise the convergence will be slow [17, 18]. Choosing a suitable step size therefore is difficult to perform manually. Klein *et al.* [15] proposed a method to automatically estimate the step size for adaptive stochastic gradient descent (ASGD) by considering the distribution of transformations. This method works for few parameters within a reasonable time, but 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 procedure. An example to illustrate this limitation is given in Figure 1.4. This limitation disqualifies ASGD for real-time image registration tasks. A fast alternative is therefore needed for real-time registration problems.

1.3 Outline of the thesis

The aim of this thesis is to develop novel optimization strategies for fast image registration. In particular, we address the following specific aims: 1) to investigate strategies to determine the step-size and search direction to accelerate image registration; 2) to develop new stochastic schemes for second order gradient optimization methods; 3) to investigate a new time-efficient preconditioner for preconditioned gradient descent optimization; 4) to validate these novel fast image registration techniques in the context of online adaptive image-guided radiation therapy. The thesis is further structured as follows:

Chapter 2 The Adaptive Stochastic Gradient Descent (ASGD) method has been proposed to automatically choose the optimization step size, but it comes at a high computational cost, depending on the number of transformation parameters. In Chapter 2, we propose a new computationally efficient method (fast

ASGD) to automatically determine the step size for gradient descent methods, by considering the observed distribution of the voxel displacements between iterations. A relation between the step size and the expectation and variance of the observed distribution is derived. While ASGD has quadratic complexity with respect to the transformation parameters, the fast ASGD method only has linear complexity. Extensive validation has been performed on different datasets with different modalities, inter/intra subjects, different similarity measures and transformation models. To perform a large scale experiment on 3D MR brain data, we have developed efficient and reusable tools to exploit an international high performance computing facility. This method is already integrated in an open source deformable image registration package `elastix`.

Chapter 3 ASGD not only outperforms deterministic gradient descent methods but also quasi-Newton method in terms of runtime. ASGD, however, only exploits first-order information of the cost function. In this chapter, we explore a stochastic quasi-Newton 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 once every L iterations and then used for the next L iterations in combination with a stochastic gradient. The method is validated on follow-up data of 3D chest CT scans (19 patients), using a B-spline transformation model and a mutual information metric.

Chapter 4 In case of ill-conditioned problems, ASGD only exhibits sublinear convergence properties. In Chapter 4, we propose an efficient preconditioner estimation method to improve the convergence rate of ASGD. Based on the observed distribution of voxel displacements in the registration, we estimate the diagonal entries of a preconditioning matrix, thus rescaling the optimization cost function. This makes the preconditioner suitable for stochastic as well as for deterministic optimization. It is efficient to compute and can be used for mono-modal as well as multi-modal cost functions, in combination with different transformation models like the rigid, affine and B-spline models.

Chapter 5 In Chapter 5, we have investigated the performance of the method developed in Chapter 2, for fast and robust contour propagation in the context of online-adaptive IMPT for prostate cancer. The planning CT scan and 7-10 repeat CT scans of 18 prostate cancer patients were used in this study. Automatic contour propagation of repeat CT scans was performed and compared with manual delineations in terms of geometric accuracy and runtime. Dosimetric accuracy was quantified by generating IMPT plans using the propagated contours expanded with a 2-mm (prostate) and 3.5-mm margin (seminal vesicles and lymph nodes) and calculating coverage based on the manual delineation. A coverage of $V_{95\%} \geq 98\%$ was considered clinically acceptable.

Chapter 6 In Chapter 6, the overall achievements of this thesis are summarized and discussed.