Fast automatic estimation of the optimization step size for nonrigid image registration

Y.Qiao, B.P.F. Lelieveldt, M. Staring

Division of Image Processing, Leiden University Medical Center, Leiden, The Netherlands

ABSTRACT

Image registration is often used in the clinic, for example during radiotherapy and image-guide surgery, but also for general image analysis. Currently, this process is often very slow, yet for intra-operative procedures the speed is crucial. For intensity-based image registration, a nonlinear optimization problem should be solved, usually by (stochastic) gradient descent. This procedure relies on a proper setting of a parameter which controls the optimization step size. This parameter is difficult to choose manually however, since it depends on the input data, optimization metric and transformation model. Previously, the Adaptive Stochastic Gradient Descent (ASGD) method has been proposed that automatically chooses the step size, but it comes at high computational cost. In this paper, we propose a new computationally efficient method to automatically determine the step size, 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 then derived. Experiments have been performed on 3D lung CT data (19 patients) using a nonrigid B-spline transformation model. For all tested dissimilarity metrics (mean squared distance, normalized correlation, mutual information, normalized mutual information), we obtained similar accuracy as ASGD. Compared to ASGD whose estimation time is progressively increasing with the number of parameters, the estimation time of the proposed method is substantially reduced to an almost constant time, from 40 seconds to no more than 1 second when the number of parameters is 10^5.

Keywords: nonrigid image registration, automatic parameter selection, computational performance

1. INTRODUCTION

Image registration is often used in clinical operations such as radiotherapy[1] and image-guide surgery,[2] but also for other general image analysis tasks.[3–7] Currently, this process is often very slow,[8] yet for intra-operative procedures the speed is crucial. The main principle of image registration is to align two images, fixed and moving image, by optimizing the similarity between these two. An often used method for intensity-based image registration is to solve a nonlinear optimization problem, usually by iterative gradient descent. For this iterative scheme,[9] a suitable step size should be chosen to guarantee the convergence and obtain a reasonable convergence speed. This step size can be chosen manually, like with the Robbins-Monro (RM) stochastic gradient descent method.[10] However, this is a difficult task, since it depends on the input data, optimization metric and transformation model. Although Klein's adaptive stochastic gradient descent (ASGD) method[11] selects the step size automatically and with a better performance than manual selection, the estimation time is very long and memory consumption is high, especially when the number of parameters is larger than $\sim 10^4$. Developing a fast and automatic selection method for the step size is still an open research topic.

In this paper, we propose a new estimation method to automatically select the optimization step size by deriving a relation with the observed voxel displacement between iterations. Using the Taylor expansion, we obtain a relationship between the step size and the voxel displacement. From the decay function used in the stochastic gradient descent, we get the maximum step size of each resolution. In Section 2, the method to calculate the parameter a is introduced. Experiments on lung CT data with different metrics are presented in Section 3. Finally, Section 4 concludes the paper.

Further author information: (Send correspondence to Y.Qiao)
E-mail: Y.Qiao@lumc.nl, Telephone: +31 (0)71 526 1391

2. METHODS

Intensity-based image registration tries to align two images following a continuous deformation strategy. We can define the fixed image as $I_F(\boldsymbol{x})$, the moving image as $I_M(\boldsymbol{x})$, and a parameterised coordinate transformation as $\boldsymbol{T}(\boldsymbol{x}, \boldsymbol{\mu})$, in which $\boldsymbol{\mu} \in \mathbb{R}^N$ is a potentially large vector of transformation parameters. When treating the registration problem as a nonlinear optimization problem, we get the following iterative minimization problem, using gradient descent:

$$\begin{aligned} \widehat{\boldsymbol{\mu}} &= \arg\min_{\boldsymbol{\mu}} \mathcal{C}(I_F, I_M \circ \boldsymbol{T}_{\boldsymbol{\mu}}), \\ \boldsymbol{\mu}_{k+1} &= \boldsymbol{\mu}_k - \gamma_k \boldsymbol{g}_k, \\ \gamma_k &= \frac{a}{(A+k)^{\alpha}}, \end{aligned} \tag{1}$$

with k the iteration number, $\mathcal{C}$ the cost function to measure the dissimilarity between the fixed and moving image, $\boldsymbol{g}_k = \partial\mathcal{C}/\partial\boldsymbol{\mu}_k$, γ_k the step size function which is a decaying function with $a > 0$, $A \geq 1$, and $0 < \alpha \leq 1$ usually manually selected before the optimization. In the decaying function γ_k, the parameter α controls the speed of decay, A provides a starting point at the beginning of the optimization and has a relatively small influence for large k, and a determines the overall scale of the step size. From these three, a is most difficult to select, since it is dependent on I_F, I_M, $\mathcal{C}$ and even $\boldsymbol{T}_{\boldsymbol{\mu}}$. The theoretically optimal value for α equals 1, and from experience[11,12] $A = 20$ provides a reasonable value for most situations.

According to Plakhov's method[13] and the original ASGD method, an adaptive scheme is attained by using the following functions:

$$\begin{aligned} \gamma(t_k) &= \frac{a}{(t_k + A)^{\alpha}}, \\ t_k &= [t_{k-1} + f(-\boldsymbol{g}_{k-1}^T \boldsymbol{g}_{k-2})]^+, \\ f(x) &= f_{min} + \frac{f_{max} - f_{min}}{1 - (f_{max}/f_{min})e^{-x/\omega}}, \end{aligned} \tag{2}$$

in which $f(x)$ is a sigmoid function, f_{max}, f_{min}, ω and a should be (automatically) determined. The iteration number is replaced by an artificial time parameter, which is a function of the inner product of two consecutive gradients. The intuition is that larger steps can be taken when the inner product is large. The sigmoid function f constrains the change in time to the range $[f_{min}, f_{max}]$. In this work, we use the same adaptive scheme and focus on automatically selecting the parameter a in a simple and less time-consuming way.

2.1 Displacement estimation

The displacement of voxel $\boldsymbol{x}_j$ between iteration k and $k+1$ is defined by

$$\boldsymbol{d}_k(\boldsymbol{x}_j) = \boldsymbol{T}(\boldsymbol{x}_j, \boldsymbol{\mu}_{k+1}) - \boldsymbol{T}(\boldsymbol{x}_j, \boldsymbol{\mu}_k). \tag{3}$$

During the optimization process, it is better to constrain the deformation $\boldsymbol{d}_k$ to a reasonable range. This ensures that the transformation is not too big or too small during the iterative optimization.[11]

Klein introduced a user-defined parameter δ as the maximum allowed voxel displacement. We use the same scheme to assume that the maximum voxel displacement for each voxel between two iterations should be not larger than δ: i.e $\|\boldsymbol{d}_k(\boldsymbol{x}_j)\| \leq \delta, \forall \boldsymbol{x}_j \in \Omega_F$. We can use a weakened form for this assumption: $P(\|\boldsymbol{d}_k(\boldsymbol{x}_j)\| > \delta) < \rho$, where ρ is a small probability value often 0.05, which means this situation nearly could not happen. We use the Vysochanskij Petunin[14] inequality to approximate this function, and we get

$$E\|\boldsymbol{d}_k\| + 2\sqrt{Var\|\boldsymbol{d}_k\|} \leq \delta. \tag{4}$$

2.2 Estimation of a

We use the Taylor expansion to make an approximation of $\boldsymbol{d}_k$ around $\boldsymbol{\mu}_k$

$$\boldsymbol{d}_k \approx \frac{\partial \boldsymbol{T}}{\partial \boldsymbol{\mu}} (\boldsymbol{x}_j, \boldsymbol{\mu}_k) \cdot (\boldsymbol{\mu}_{k+1} - \boldsymbol{\mu}_k) = \boldsymbol{J}_j (\boldsymbol{\mu}_{k+1} - \boldsymbol{\mu}_k), \tag{5}$$

in which $\boldsymbol{J}_j = \frac{\partial \boldsymbol{T}}{\partial \boldsymbol{\mu}} (\boldsymbol{x}_j, \boldsymbol{\mu}_k)$. Together with Equation (1), $\boldsymbol{d}_k$ can be rewritten as

$$\boldsymbol{d}_k(x_j) \approx -\gamma_k \boldsymbol{J}(x_j) \boldsymbol{g}_k = -\gamma_k \boldsymbol{M}_k(x_j), \tag{6}$$

where $\boldsymbol{M}_k(\boldsymbol{x}_j) = \boldsymbol{J}(\boldsymbol{x}_j)\boldsymbol{g}_k$. Considering the expectation and variance of all voxels, the Vysochanskij Petunin inequality can be rewritten to

$$\gamma_k \left(E \left\| \boldsymbol{M}_k \right\| + 2\sqrt{Var \left\| \boldsymbol{M}_k \right\|} \right) \leq \delta. \tag{7}$$

From the step size function $\gamma(t_k) = a/(t_k + A)^\alpha$, it is easy to find the maximum step size $\gamma_{\max} = \gamma_0 = a/A^\alpha$, from which the maximum value of a, $a_m = \gamma_{\max} A^\alpha$ can be obtained. After computing the value of $E\|\boldsymbol{M}_k\|$ and $\sqrt{Var\|\boldsymbol{M}_k\|}$, we obtain the following:

$$\begin{aligned} a_m &= \frac{\delta A^\alpha}{E\|\boldsymbol{M}_k\| + 2\sqrt{Var\|\boldsymbol{M}_k\|}} \\ &= \delta A^\alpha / \left(\frac{1}{N} \sum_{\boldsymbol{x}_j \in \Omega_F} \|\boldsymbol{M}_k(\boldsymbol{x}_j)\|_F + 2\sqrt{\frac{1}{N-1} \sum_{\boldsymbol{x}_j \in \Omega_F} \left(\|\boldsymbol{M}_k(\boldsymbol{x}_j)\|_F - E\|\boldsymbol{M}_k(\boldsymbol{x}_j)\|_F \right)^2} \right), \end{aligned} \tag{8}$$

in which $\| \cdot \|_F$ is the Frobenius norm. We choose k equal to zero, so that for a given δ, the value of a can be estimated from the initial distribution of $\boldsymbol{M}_0$ at the beginning of each resolution.

Compared to deterministic gradient descent, larger gradients are obtained during the optimization for stochastic gradient descent, due to the estimation error. For larger gradients, a smaller step size is needed, so a compensation is needed when using stochastic gradient descent. We adopt the compensation factor from Klein et al,[11] which is defined as $\eta = E\|\boldsymbol{g}\| / (E\|\boldsymbol{g}\| + E\|\boldsymbol{\epsilon}\|)$, resulting in the compensated step size $a = \eta a_m$.

3. EXPERIMENTS AND RESULTS

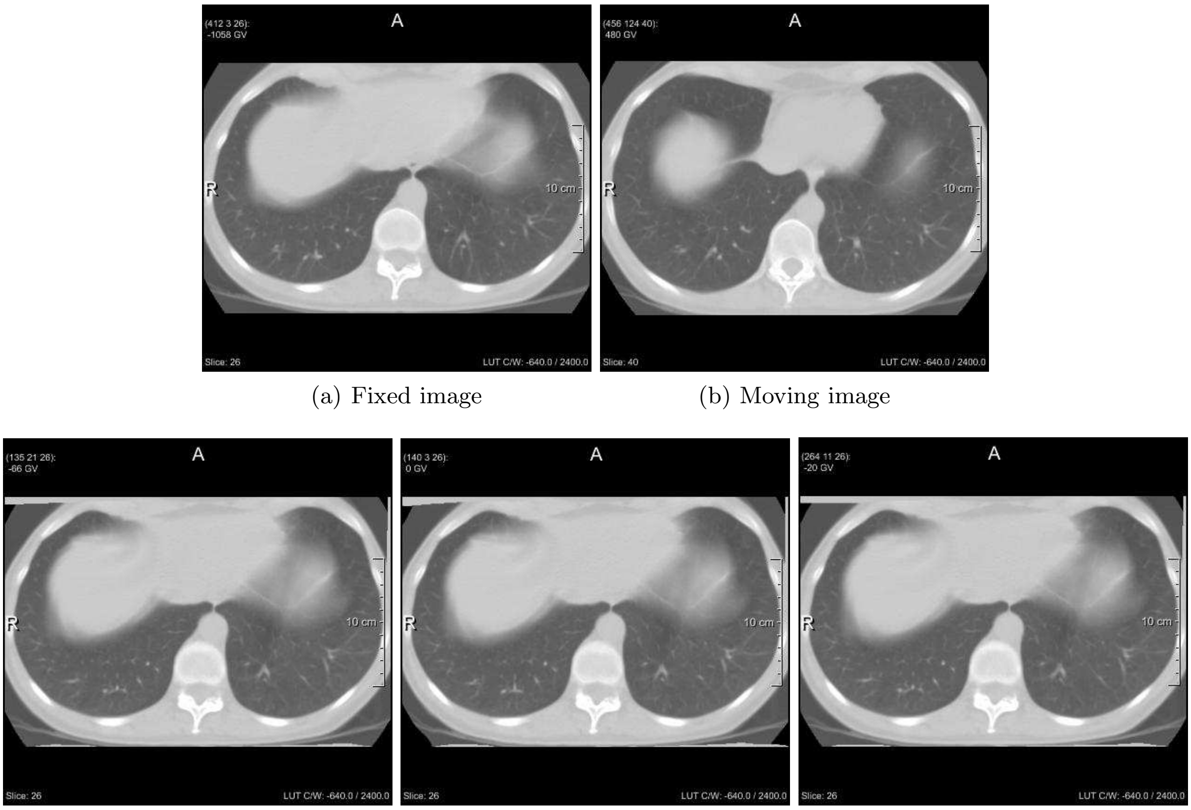
3.1 Data, environment settings and evaluation

The new method was implemented in the C++ language in the open source platform `elastix`.[12] The experiments were performed on a workstation with 24 GB memory, and 8 cores running at 2.4 GHz. We used 3D lung CT images acquired during the SPREAD study[15] (around $446 \times 315 \times 129$ voxels) of 19 patients scanned without contrast media. The patient group, aging from 49 to 78 with 36%-87% predicted FEV_1 had moderate to severe COPD at GOLD stage II and III, without $\alpha 1$ antitrypsin deficiency. One hundred anatomical corresponding points from each lung CT image were chosen semi-automatically using Murphy's method[16] to obtain a ground truth.

In this paper two aspects of the new method were considered: registration accuracy and estimation speed. To evaluate the registration accuracy, the anatomical points of the fixed image were transformed using the obtained transformation and then compared with the corresponding points in the moving image. If one of the observers marked a point as unsure it was considered an unreliable point, and those points were excluded from validation to obtain the ground truth as reliable as possible. The Euclidean distance between the corresponding points was used as a measure of registration accuracy:

$$dist_i = \|\boldsymbol{T}(p_F^i) - p_M^i\|. \tag{9}$$

To evaluate the method, four commonly used similarity metrics were tested using a B-spline transformation model: mean squared difference (MSD), normalized correlation (NC), mutual information (MI) and normalized mutual information (NMI). The registration scheme includes a three level multi-resolution framework and



(a) Fixed image (b) Moving image

(c) ASGD (d) a_m (e) ηa_m

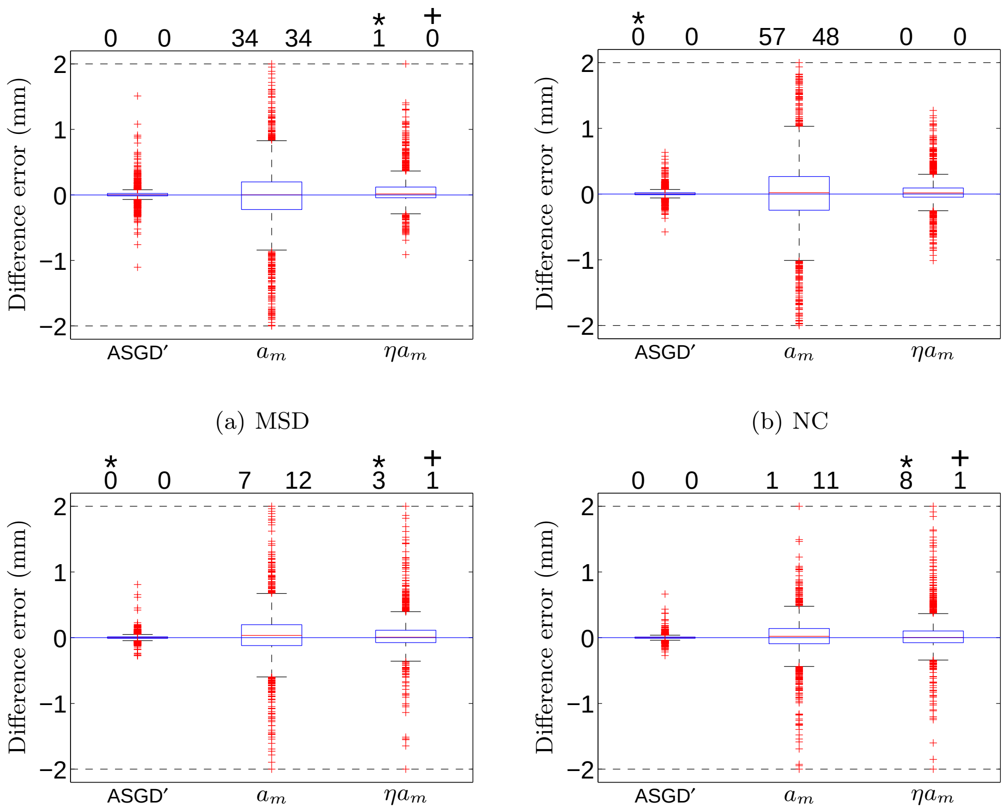
Figure 1. Lung CT images. One example slice of the lung CT data from (a)fixed image, (b)moving image, (c) original ASGD result image, (d)proposed method without noise compensation, (e)proposed method with noise compensation.

	ASGD	ASGD′	a_m	ηa_m
MSD	1.09	1.10	1.09	1.12
NC	1.50	1.51	1.56	1.55
MI	1.65	1.65	1.69	1.66
NMI	1.66	1.65	1.67	1.68

Table 1. The median Euclidean distance error (mm)

Gaussian smoothing filter with a fixed standard deviation of 2, 1 and 0.5 mm, for each resolution. The grid size is halved in B-spline control point grid in order to increase the transformation accuracy. In the experiments, we chose the same value of α and A as Klein: $\alpha = 1$ from the theoretically optimal setting and $A = 20$ from experience. The given value of δ is equal to the image voxel size. The sigmoid function parameters $f_{max} = 1.0, f_{min} = -0.8, \omega = 10^{-6}$ were chosen for three more parameter selection methods — ASGD only estimating a (ASGD′), the new method without (a_m) and with compensation (ηa_m). The original method (ASGD) automatically selects all these parameters.

Then the difference of the Euclidean distance errors was calculated between the proposed methods and original method. To assess the registration performance, a Wilcoxon signed rank test ($P = 0.05$) for the registration results was performed. For 19 patients, we first obtained the mean distance error of 100 points for each patient then performed Wilcoxon signed rank test to these mean errors. To evaluate the estimation speed, the parameter estimation time and pure registration time of each resolution were measured. The scopes of estimated values of a of the first resolution for each patient are shown to present the difference between the proposed and original method.



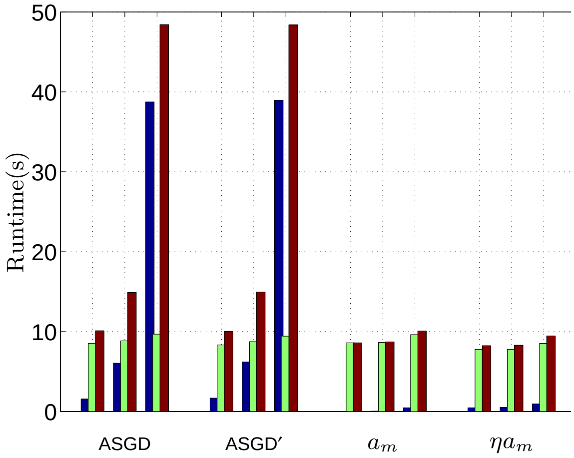
(a) MSD (b) NC

(c) MI (d) NMI

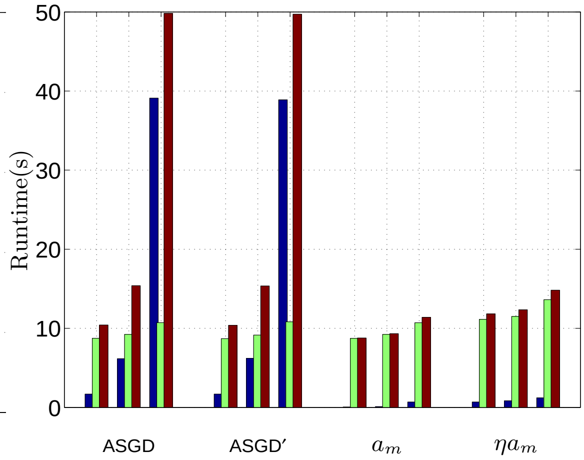
Figure 2. The difference of Euclidean distance error in mm of the registration results compared to ASGD. The median Euclidean distance error of ASGD is 1.09, 1.50, 1.65 and 1.66 for MSD, NC, MI and NMI, respectively. The notion of * indicates a statistically significant difference with ASGD, and + with ASGD′. The two numbers on the top of each box denote the number of the outliers larger (worse) and smaller (better) than 2 and -2 mm, respectively.

3.2 Results

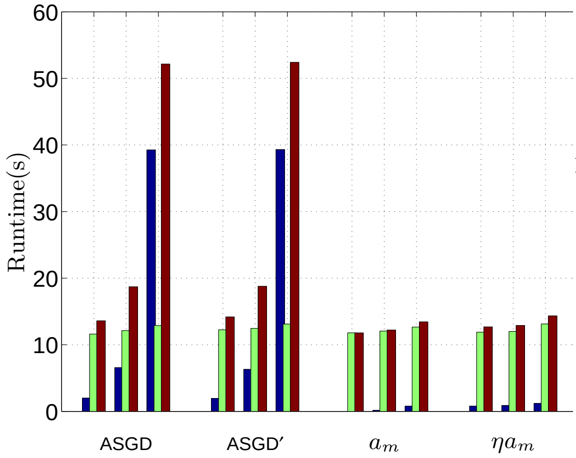
One slice of the lung CT data is shown in Figure 1 for the fixed image, moving image, original ASGD result image and proposed method result image, respectively. Table 1 shows the median Euclidean distance error of all corresponding points (after excluding 18 unsure points) for four different metrics and methods. The results in Figure 2 show the difference with ASGD of the registration errors for each of the four metrics using 500 iterations and 3 resolutions. The number of parameters N is 4×10^3, 2×10^4 and 9×10^4, for each resolution respectively. It can be seen that the median accuracy of the new method and the original ones is almost the same. Compared with ASGD and ASGD′, the method ηa_m had a significant difference for metrics MSD, MI and NMI, but this difference is smaller than 0.05 mm. The improvement in runtime is shown in Figure 3, in which the original estimation time takes a large part of the total runtime per resolution, while the new method consumes only a small fraction of the total runtime. The estimation time of ASGD and ASGD′ are progressively increasing from 3 seconds to 40 seconds with the number of parameters increasing from 4×10^3 to 9×10^4, while the new method maintains a constant estimation time no more than 1 second. When $N = 10^6$, the estimation time is even more than 1000 seconds for ASGD, see Figure 4. From Figure 5 we can see that with the compensation factor, the value of a is more similar to ASGD and ASGD′, while without compensation the value is quite large. Figure 6 presents the step size γ_k from which we can see that the step size of ASGD and ASGD′ are usually located between ηa_m and a_m, except for NC. Considering Figure 5 and Figure 6, we can find that with the noise compensation the step size more closely resembles the original result.



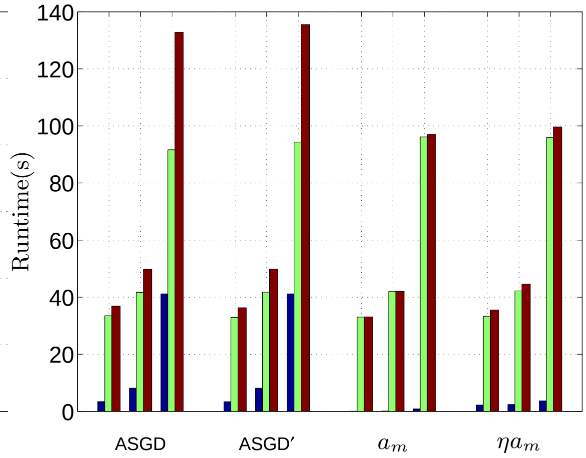
(a) MSD



(b) NC



(c) MI



(d) NMI

Figure 3. Runtime in seconds. The blue, green and red bar indicate estimation time, pure registration time and total time elapsed in each resolution, respectively.

4. CONCLUSION AND DISCUSSION

A new automatic estimation method for selecting the optimization parameter a of the step size for nonrigid image registration has been presented. Our main contribution is to select parameter a automatically from the observed distribution of voxels displacements between iterations. A relation between the step size and the expectation and variance of the observed distribution is then derived. The new method has been verified on four similarity metrics with 3D lung CT image data using a nonrigid B-spline transformation model.

From Figure 3, we can see that the estimation time of the original ASGD increases sharply from a few seconds to almost 40 seconds when the number of parameters changes from 10^3 to 10^4, and even to 1000s for 10^6 (Figure 4), but the pure registration time is almost constant for each resolution. The reason is that for the original ASGD algorithm, the estimation method mostly depends on a huge covariance matrix computation of the Jacobian of the transformation, where its size is determined by the number of transformation parameters. With the new method, the computation time is sharply reduced. However, in this paper we only consider an almost constant image size around $446 \times 315 \times 129$, and the image size also influences the estimation time of a, which should be further studied.

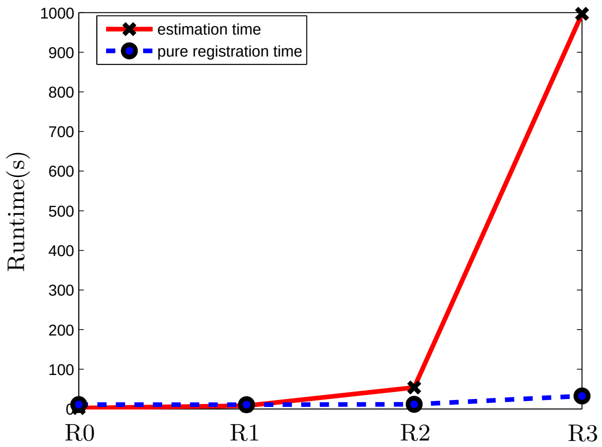
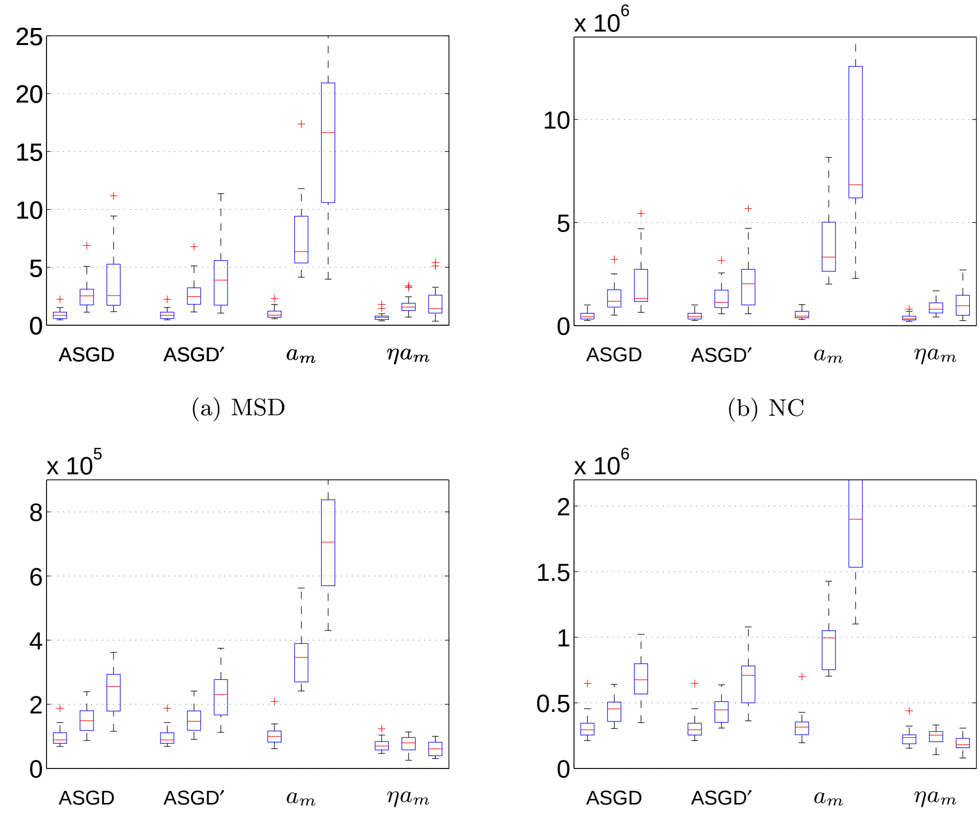


Figure 4. One example of runtime in seconds of original ASGD method for Mutual Information metric. The blue line is the pure registration time and red one the estimation time. The number of parameters for four resolution is 4680, 22440, 141081 and 975966, respectively.



(a) MSD (b) NC (c) MI (d) NMI

Figure 5. The value of a for MSD, NC, MI, NMI for the three resolutions.

With respect to registration accuracy, we can see that the median Euclidean distance error of these method is almost the same. Without the noise compensation factor, the Euclidean distance error compared with the original method is within 0.3 mm, while with the noise compensation factor it is within 0.2 mm. Comparing the number of outliers ($> 2mm$ or $< -2mm$) without the noise compensation factor the positive and negative numbers are

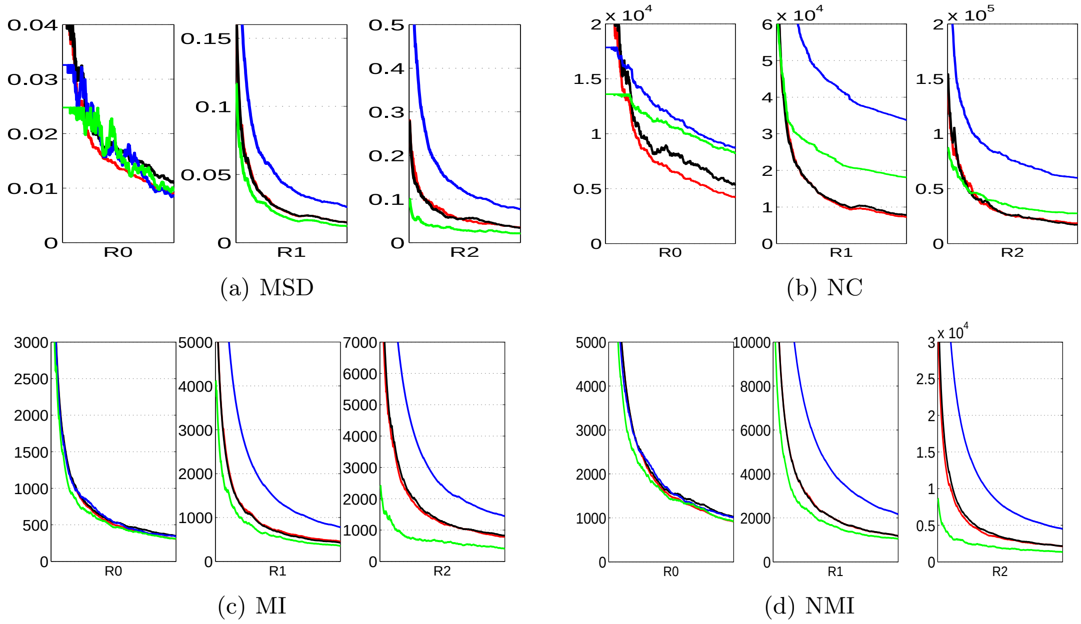


Figure 6. An example of the step size using 500 iterations in each resolution for each method. The red line is original ASGD, the black line ASGD′, blue line is a_m (without compensation) and green line is ηa_m (with compensation).

larger, but this only means that it is different from the original method. With the noise compensation factor, the number of outliers is very small but with a significant difference. So, the influence of the noise compensation factor should be further considered.

In comparison with the original ASGD method, the results of the new method show that there is a sharp cut of the estimation time, i.e. the estimation time is reduced from minutes to seconds when the number of parameters is larger than 10^4, while maintaining the registration accuracy as tested for four metrics. The methodology will be released as open source via the next release of `elastix`. Future work will focus on experimental validation of the proposed method in different transformation models, different clinical applications and different image sizes.

REFERENCES

1. J. Cai, J. C. Chu, D. Recine, M. Sharma, C. Nguyen, R. Rodebaugh, V. A. Saxena, and A. Ali, "Ct and pet lung image registration and fusion in radiotherapy treatment planning using the chamfer-matching method," *International Journal of Radiation Oncology* Biology* Physics* **43**(4), pp. 883–891, 1999.
2. D. Tomazevic, B. Likar, T. Slivnik, and F. Pernus, "3-d/2-d registration of ct and mr to x-ray images," *Medical Imaging, IEEE Transactions on* **22**(11), pp. 1407–1416, 2003.
3. J. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis* **2**(1), pp. 1–36, 1998.
4. B. Zitov and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing* **21**(11), pp. 977 – 1000, 2003.
5. L. G. Brown, "A survey of image registration techniques," *ACM computing surveys (CSUR)* **24**(4), pp. 325–376, 1992.
6. B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and vision computing* **21**(11), pp. 977–1000, 2003.
7. A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *Medical Imaging, IEEE Transactions on* **32**(7), pp. 1153–1190, 2013.
8. A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M.-C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier, *et al.*, "Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration," *Neuroimage* **46**(3), pp. 786–802, 2009.

9. H. J. Kushner and G. Yin, *Stochastic approximation and recursive algorithms and applications*, vol. 35, Springer, 2003.
10. J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *The Annals of Mathematical Statistics* **23**(3), pp. 462–466, 1952.
11. S. Klein, J. Pluim, M. Staring, and M. Viergever, "Adaptive stochastic gradient descent optimisation for image registration," *International Journal of Computer Vision* **81**(3), pp. 227–239, 2009.
12. S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim, "Elastix: a toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging* **29**(1), pp. 196–205, 2010.
13. A. Plakhov and P. Cruz, "A stochastic approximation algorithm with step-size adaptation," *Journal of Mathematical Sciences* **120**(1), pp. 964–973, 2004.
14. D. Vysochanskij and Y. I. Petunin, "Justification of the 3σ rule for unimodal distributions," *Theory of Probability and Mathematical Statistics* **21**, pp. 25–36, 1980.
15. J. Stolk, H. Putter, E. M. Bakker, S. B. Shaker, D. G. Parr, E. Piitulainen, E. W. Russi, E. Grebski, A. Dirksen, R. A. Stockley, *et al.*, "Progression parameters for emphysema: a clinical investigation," *Respiratory medicine* **101**(9), pp. 1924–1930, 2007.
16. K. Murphy, B. van Ginneken, S. Klein, M. Staring, B. de Hoop, M. A. Viergever, and J. P. Pluim, "Semi-automatic construction of reference standards for evaluation of image registration," *Medical Image Analysis* **15**(1), pp. 71–84, 2011.