



Queensland University of Technology
Brisbane Australia

This may be the author's version of a work that was submitted/accepted for publication in the following source:

[Nazib, Abdullah, Galloway, James, Fookes, Clinton, & Perrin, Dimitri](#)
(2018)

Performance of registration tools on high-resolution 3D brain images.
In Suaning, G & Doessel, O (Eds.) *Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*.

Institute of Electrical and Electronics Engineers Inc., United States of America, pp. 566-569.

This file was downloaded from: <https://eprints.qut.edu.au/129503/>

© Consult author(s) regarding copyright matters

This work is covered by copyright. Unless the document is being made available under a Creative Commons Licence, you must assume that re-use is limited to personal use and that permission from the copyright owner must be obtained for all other uses. If the document is available under a Creative Commons License (or other specified license) then refer to the Licence for details of permitted re-use. It is a condition of access that users recognise and abide by the legal requirements associated with these rights. If you believe that this work infringes copyright please provide details by email to qut.copyright@qut.edu.au

Notice: *Please note that this document may not be the Version of Record (i.e. published version) of the work. Author manuscript versions (as Submitted for peer review or as Accepted for publication after peer review) can be identified by an absence of publisher branding and/or typeset appearance. If there is any doubt, please refer to the published source.*

<https://doi.org/10.1109/EMBC.2018.8512403>

Performance of Registration Tools on High-Resolution 3D Brain Images

Abdullah Nazib¹, James Galloway¹, Clinton Fookes¹, Dimitri Perrin^{1,*}

Abstract—Recent progress in tissue clearing allows the imaging of entire organs at single-cell resolution. A necessary step in analysing these images is registration across samples. Existing methods of registration were developed for lower resolution image modalities (e.g. MRI) and it is unclear whether their performance and accuracy is satisfactory at this larger scale (several gigabytes for a whole mouse brain). In this study, we evaluated five freely available image registration tools. We used several performance metrics to assess accuracy, and completion time as a measure of efficiency. The results of this evaluation suggest that ANTS provides the best registration accuracy, while Elastix has the highest computational efficiency among the methods with an acceptable accuracy. The results also highlight the need to develop new registration methods optimised for these high-resolution 3D images.

I. INTRODUCTION

The conventional way of obtaining high-resolution 3D images of entire brains is by histological sectioning. The brain is cut into thin cross-sections, which are imaged individually, and the complete 3D volume is later reconstructed. However, sectioning the brain damages tissues, and the computational reconstruction is both time-consuming and error-prone.

The main source of light scattering in the fixed brain is the presence of lipids, and a number of tissue-clearing methods have therefore been developed to remove those, thereby making the brain transparent. Imaging an entire organ with cellular-level resolution, without slicing, is now possible [2], [8]. The images obtained with these new methods are several orders of magnitude larger than with standard modalities. A typical MRI image has a voxel resolution of $0.86 \times 0.86 \times 1.5 \text{ mm}^3$, for a total of roughly 8 million voxels. Cleared whole brains imaged with light-sheet fluorescence microscopy (LSFM) have a voxel resolution of $6.45 \times 6.45 \times 10 \text{ } \mu\text{m}^3$, for a total of more than 3 billion voxels [8]. This creates significant challenges for their downstream analysis, and in particular for image registration, (the process transforming different samples into a common coordinate system and enables direct comparison across the samples).

The objective of this paper is to evaluate such image registration tools on CUBIC dataset. We focus on five well-known and freely available tools: Advanced Normalization Tools (ANTS [1]), Insight Registration Toolkit (IRTK [6]), Automated Image Registration (AIR [11]), Elastix [4], and NiftyReg [5]. The performance of the registration tools are evaluated in terms of accuracy (both quantitatively and qualitatively) and computational efficiency.

II. MATERIALS AND METHODS

A. Datasets

In the study, we used the three Arc-dVenus brains available with the CUBIC protocol [9]. In what follows, they are noted B1, B2 and B3. We created downsampled images at 10%, 15%, 20% and 25% resolution for initial analysis. Samples at 100% were also used with the top-performing tools. To complement the assessment performed on inter-subject data (‘inter-brain pipeline’), we also created an artificial dataset by deforming those files, so that each ‘deformed brain’ can then be registered back to its original version to allow for an accurate validation against a known ground truth (‘intra-brain pipeline’). The deformed files were generated using simple Gaussian deformation. A Gaussian distributed deformation vector field was generated from a pixel grid. Pixel intensities were then interpolated over the Gaussian distributed deformation field. Repeatedly applying the Gaussian deformation field over all slices in a given 3D volume generated a Gaussian-deformed volume which was used for evaluation.

B. Registration Process and Settings

The same process was used for all registration tools, both contexts (inter-brain and intra-brain), and all file sizes: rigid registration, followed by affine registration and deformable registration. In the inter-brain context, B3 was the reference and the other two brains were aligned to it. In the intra-subject context, the original samples were used as reference and the deformed ones were aligned back to those. The tests were performed on an HPC cluster. Because the registration

¹ School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, Australia

* Corresponding author, dimitri.perrin@qut.edu.au

tools are not parallelised, we used a single compute core (2.66GHz 64bit Intel Xeon processors, 256GB RAM).

The performance of the tools used in this evaluation can be affected by parameter choice, and we have therefore considered this carefully. For IRTK, we followed the same settings as [13], except B-spline control points. The control point spacing is set to 5mm, which is the highest possible value for this method. For AIR, we applied the default threshold values for rigid and affine transformation. Using affine registered parameters as the starting points for non-linear registration (using 3rd order polynomial spatial transformation model). Elastix includes a number of transformation models, as well as a several optimisation methods. In this evaluation, we used the same parameter settings as [3] for rigid, affine and non-linear (B-spline) registration. The ANTS [1] parameters for affine registration and symmetric normalisation are derived from example scripts. We used cross-correlation as a similarity measure, resolution level 3, and 100 iterations for each sampling level. As for IRTK, the parameter choice for NiftyReg is based on [13]. The number of iterations was set at 1000 for free-form deformation and we used an intensity threshold of 500 for both source and target image.

C. Evaluation Measures

There are multiple methods of measuring similarity between images. In this study, we used cross correlation (CC) and mutual information (MI). To complement the quantitative information obtained from these two metrics, we also qualitatively assessed the results through visual inspection of overlaid registered and reference images. To assess registration efficiency, we also measured the computation time for each method and each input size.

III. RESULTS

A. Quantitative Registration Accuracy

1) *Inter-Brain Registration:* For the inter-brain registration program, brain sample 3 was used as the reference brain, and brain samples 1 and 2 were aligned to this reference.

Table I shows the average cross-correlation and mutual information measures of each registration tool across all downscaled resolutions. ANTS obtained the best cross-correlation average, with Elastix a close second. The average cross-correlation for NiftyReg was under 0.8, while both IRTK and AIR scored under 0.5. Elastix obtained the best mutual information score, followed by ANTS and NiftyReg, and poor scores for IRTK and AIR remained under 1.

We also test the top-performing tools, ANTS and Elastix, on 100%-resolution (running a job with 256Gb and 100-hour wall time). We used 3D images from the dorsal to ventral (D-V) stacks for B1 and B2, and registered them to B3. We then repeated the process for the ventral to dorsal (V-D) images. The Elastix results are shown in Table II, and are similar to those observed at 25%-resolution. ANTS failed to register any brain, even after extending the wall time to 200 hours.

2) *Intra-Brain Registration:* ANTS remained the best performer in terms of cross-correlation, and Elastix continued to outperform ANTS in all but the 10% resolution when considering the mutual information. Finally, Elastix scored highly across both metrics and all resolutions. These results are summarised in Table III.

B. Visual Analysis

Our quantitative results are complemented with a qualitative inspection of the alignments. Figures 1 and 2 are visual representations of the results for ANTS, Elastix, and NiftyReg, where the reference brains (red) are overlaid with their corresponding aligned brains (green).

Figure 1 shows the inter-brain registration results. The first row shows all four resolutions registered by ANTS (Figure 1a–d). There are visible differences in the cerebellum, but this is expected for all methods, and not a problem for subsequent analyses, which typically focus on other regions. There are also smaller differences in regions such as the hippocampal formation, and in particular the dentate gyrus. Elastix gave good results at 10% resolution, but as the resolution increased it underperformed against ANTS. The mismatch of the hippocampal formation is more pronounced, and other regions towards the cerebral cortex are also misaligned. The same issues are present in the NiftyReg results.

Similar results can be observed for the intra-brain context (Figure 2). The deformation is larger than the typical difference between two mouse brains, but ANTS still performs

TABLE I: Mean CC and MI for inter-brain pipeline

Methods	B1 (CC)	B1 (MI)	B2 (CC)	B2 (MI)
ANTS	0.9303	1.4317	0.9546	1.5189
AIR	0.0124	0.8879	0.0154	0.9172
Elastix	0.9163	2.4115	0.9388	2.2887
NiftyReg	0.7487	1.4625	0.7645	1.3856
IRTK	0.3268	0.4563	0.3484	0.3456

TABLE II: Elastix – 100%-resolution – inter-brain pipeline

Samples	CC	MI
Brain 1 D-V	0.8412	0.8018
Brain 2 D-V	0.8676	0.8385
Brain 1 V-D	0.4403	0.4781
Brain 2 V-D	0.4278	0.4852

TABLE III: Mean CC and MI for intra-brain pipeline

Methods	B1 (CC)	B1 (MI)	B2 (CC)	B2 (MI)	B3 (CC)	B3 (MI)
ANTS	0.9805	1.4136	0.9822	1.9872	0.9823	1.8591
AIR	0.1657	1.2865	0.1651	1.2340	0.0097	1.2451
Elastix	0.9719	2.2385	0.9692	2.3453	0.9713	2.7193
NiftyReg	0.9411	1.4952	0.9494	1.8642	0.9394	2.4223
IRTK	0.5156	0.8534	0.5160	0.5156	0.4881	0.7286

relatively well, and visibly better than Elastix and NiftyReg.

Overall, taking together the quantitative and qualitative results, ANTS is the best performer in this evaluation.

C. Computational Efficiency

Computational efficiency is shown in Tables IV and V. We performed experiments using HPC jobs with 32GB memory for 10%- and 15%-resolution files, and 64GB memory for 20%- and 25%-resolution files. We did not observe any relationship between accuracy and efficiency. AIR is very efficient, but inaccurate on this type of data. IRTK had similar accuracy issues, but also was inefficient, with a 25% resolution dataset taking more than a day to process.

Amongst the top three performers in terms of accuracy, ANTS is the least efficient, with completion times of over 12 hours for intra-brain registration and over 8 hours for inter-brain registration at 25% resolution. Elastix proved to be the most efficient of the three, with all resolution sizes being processed in less than one hour (in an HPC environment).

The most promising two tools, ANTS and Elastix, were also tested on 100%-resolution samples, using 256GB memory, 100-hour wall time HPC jobs. ANTS failed to register any brain, even after extended the wall time to 200 hours. Elastix successfully completed all registrations, and the results were in line with those obtained at lower resolutions (as discussed above). However, registration of the D-V samples took 27 to 27.5 hours, while registration of the V-D samples took 32.5 hours. While it was possible to align the brains, 30-hour registrations are not practical for realistic scenarios,

TABLE IV: Inter-brain - Computation time (hh:mm:ss)

Methods	10%	15%	20%	25%
ANTS	00:42:45	01:38:31	4:56:48	08:40:11
AIR	00:00:27	00:06:52	00:05:04	00:37:15
Elastix	00:02:17	00:05:15	00:08:36	00:16:21
NiftyReg	00:04:35	00:19:40	01:39:20	01:45:50
IRTK	00:03:52	00:26:57	02:28:10	25:42:26

TABLE V: Intra-brain - Computation time (hh:mm:ss)

Methods	10%	15%	20%	25%
ANTS	01:23:44	04:13:18	08:24:49	12:24:28
AIR	00:00:31	00:25:51	00:10:18	00:50:14
Elastix	00:05:03	00:06:30	00:17:42	00:24:30
NiftyReg	00:16:32	00:44:55	02:12:56	04:06:51
IRTK	00:02:42	03:09:26	06:05:20	30:22:47

where studies would routinely involve dozens of brains. When tested on a desktop workstation (Core i7 3.4 GHz, 16 GB RAM, 64-bit Ubuntu 16.04 OS), neither method could perform the registration.

D. Parameter exploration

To confirm that the choice of parameters did not affect our results, we also performed a parameter exploration for the two promising tools (ANTS, Elastix), at 25% resolution. All tests were performed on 32GB memory, 100-hour wall time HPC jobs, with larger resources used if initial tests failed.

ANTS performed well under all configurations, with parameter sets having a small impact on the registration quality (Table VI). Increasing the number of optimisation levels, and using mutual information as the similarity measure, improved the quality of the results. This was confirmed by visual inspection of the registered images. Mutual information was also associated with lower computation time. Inspection of the logs revealed this is due to convergence in fewer iterations. Cross-correlation takes longer, and there is one configuration where ANTS was not able to finish the registration within the 200-hour wall time.

Parameter choice also had some impact on the quality of the Elastix results (Table VII). However, for Elastix, we did

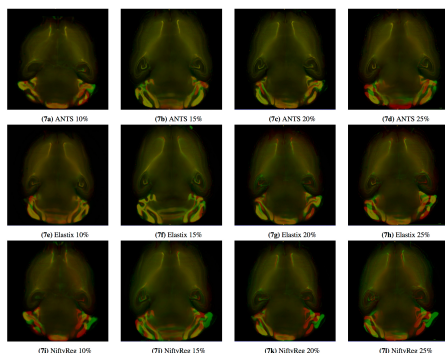


Fig. 1: Qualitative results for inter-brain registration.

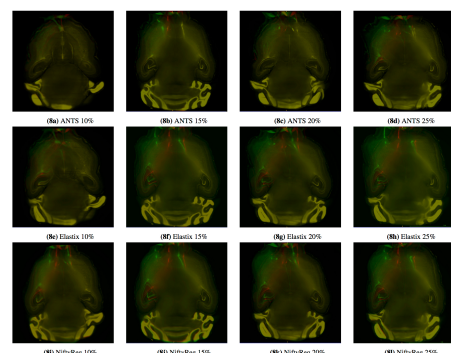


Fig. 2: Qualitative results for intra-brain registration.

TABLE VI: ANTS – Parameter optimization

Parameter Set	Similarity	Interp	Levels	CC	MI	Time
1	CC,1.2	Linear	1,1,3	0.7070	0.6702	03:59:53
2	ML,1.32	Bspline	3,3,3	0.7868	0.6208	01:04:58
3	ML,1.64	Bspline	5,5,5	0.7359	0.8623	01:16:06
4	CC,1.5	Bspline	5,5,5	Unable to complete within 200h.		
5	CC,1.2	Bspline	3,3,5	0.6868	0.6557	38:59:59
6	ML,1.32	Bspline	3,3,5	0.7429	0.8622	01:02:36
7	ML,1.32	Linear	3,3,3	0.7825	0.9257	01:00:25

TABLE VII: Elastix – Parameter optimization

Parameter Set	Similarity	Iterations	Levels	CC	MI	Time
1	CC	1000	3	0.9301	0.8854	01:54:04
2	ML,32	1000	3	0.8752	1.0234	02:03:46
3	ML,64	1000	5	0.8695	1.1337	02:41:27
4	CC	1000	5	0.9538	0.9933	02:41:08
5	CC	2500	6	Unable to complete. Program stops.		
6	ML,32	2500	6	Unable to complete. Program stops.		

not observe a faster convergence using mutual information. The computation time remained largely similar, and the quantitative results were better with cross-correlation as the similarity measure. As one would expect, quantitative results were improved by increasing the number of optimisation levels, up to the point where Elastix could not complete the registration. However, visual inspection of the results, reveal that deformations induced by these altered parameters were too drastic, and created anatomically incorrect features. The initial parameter set was the best configuration.

IV. CONCLUSION

Imaging whole organs at the single-cell resolution, taken together with other recent advances such as gene editing [7], has the potential to lead to important breakthroughs in systems biology and neurosciences [10], but only if the very large datasets that are produced can be analysed accurately and efficiently. Image registration is particularly challenging, because most tools pre-date tissue clearing methods and were developed for low-resolution images such as MRI.

In this study, we analysed five registration tools on high-resolution whole-brain images obtained with the CUBIC method. Three of these (ANTS, Elastix and NiftyReg) gave good quantitative results. IRTK was unstable, sometimes producing completely black or over-saturated images. AIR proved unsuitable for the discrete single-cell signal, as it registers images by minimizing voxel ratio.

Taking together all the evaluation measures, ANTS (with MI) is the best choice for the registration of 3D samples obtained with tissue clearing and LFSM imaging when the overall analysis is not degraded by registering at 25% resolution, as in [10]. In datasets where registration at 100% resolution is required, it may be substituted with Elastix. However, even Elastix struggles with these files. It took more

than a day per registration, which limits its practical use. This study therefore highlights the need to develop dedicated tools for this new type of data. New approaches such as deep learning are being considered in an image registration context [12], and may provide suitable alternatives.

REFERENCES

- [1] B. Avants, C. Epstein, M. Grossman, and J. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, 2008.
- [2] K. Chung and K. Deisseroth. CLARITY for mapping the nervous system. *Nature Methods*, 10(6):508–513, 2013.
- [3] L. Hammelrath, S. Škokić, A. Khmelinskii, A. Hess, N. van der Knaap, M. Staring, B. P. Lelieveldt, D. Wiedermann, and M. Hoehn. Morphological maturation of the mouse brain: An in vivo MRI and histology investigation. *NeuroImage*, 125:144–152, 2016.
- [4] S. Klein, M. Staring, K. Murphy, M. Viergever, and J. Pluim. elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205, 2010.
- [5] M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin. Fast free-form deformation using graphics processing units. *Computer Methods and Programs in Biomedicine*, 98(3):278–284, 2010.
- [6] 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.
- [7] G. A. Sunagawa, K. Sumiyama, M. Ukai-Tadenuma, D. Perrin, H. Fujishima, H. Ukai, O. Nishimura, S. Shi, R. Ohno, R. Narumi, Y. Shimizu, D. Tone, K. L. Ode, S. Kuraku, and H. R. Ueda. Mammalian Reverse Genetics without Crossing Reveals Nr3a as a Short-Sleeper Gene. *Cell Reports*, 14(3):662–677, 2016.
- [8] E. A. Susaki, K. Tainaka, D. Perrin, F. Kishino, T. Tawara, T. M. Watanabe, C. Yokoyama, H. Onoe, M. Eguchi, S. Yamaguchi, T. Abe, H. Kiyonari, Y. Shimizu, A. Miyawaki, H. Yokota, and H. R. Ueda. Whole-Brain Imaging with Single-Cell Resolution Using Chemical Cocktails and Computational Analysis. *Cell*, 157(3):726–739, 2014.
- [9] E. A. Susaki, K. Tainaka, D. Perrin, H. Yukinaga, A. Kuno, and H. R. Ueda. Advanced CUBIC protocols for whole-brain and whole-body clearing and imaging. *Nature Protocols*, 10(11):1709–1727, 2015.
- [10] F. Tatsuki, G. A. Sunagawa, S. Shi, E. A. Susaki, H. Yukinaga, D. Perrin, K. Sumiyama, M. Ukai-Tadenuma, H. Fujishima, R. Ohno, D. Tone, K. L. Ode, K. Matsumoto, and H. R. Ueda. Involvement of Ca²⁺-Dependent Hyperpolarization in Sleep Duration in Mammals. *Neuron*, 90(1):70–85, 2016.
- [11] R. P. Woods, S. R. Cherry, and J. C. Mazziotta. Rapid automated algorithm for aligning and reslicing PET images. *Journal of Computer Assisted Tomography*, 16(4):620–33, 1998.
- [12] G. Wu, M. Kim, Q. Wang, B. C. Munsell, and D. Shen. Scalable High-Performance Image Registration Framework by Unsupervised Deep Feature Representations Learning. *IEEE Transactions on Biomedical Engineering*, 63(7):1505–16, 2016.
- [13] Z. Xu, C. P. Lee, M. P. Heinrich, M. Modat, D. Rueckert, S. Ourselin, R. G. Abramson, and B. A. Landman. Evaluation of Six Registration Methods for the Human Abdomen on Clinically Acquired CT. *IEEE Transactions on Biomedical Engineering*, 63(8):1563–1572, 2016.