ESTIMATION OF ATLAS-BASED SEGMENTATION OUTCOME: LEVERAGING INFORMATION FROM UNSEGMENTED IMAGES

Orcun Goksel, Tobias Gass, Valery Vishnevsky, Gabor Szekely

Computer Vision Laboratory, ETH Zurich, Switzerland

ABSTRACT

Segmentation via atlas registration is a common technique in medical image analysis. Devising estimates of such segmentation outcome has been of interest in cases with multiple atlases, both for single-atlas selection and for multi-atlas fusion. This paper studies the estimation of expected Dice's similarity metric for registering atlas-target pairs, by employing registration loops with models of such metric (error) accumulation over these loops. In this framework, the use of registration information also from unsegmented images is proposed and is shown to outperform using segmented atlas images alone. We demonstrate a fast, memory-efficient implementation and single-atlas selection results using a CT and an MR dataset.

Index Terms— Quality assessment, registration circuits.

1. INTRODUCTION

Image registration is of major interest to several medical applications. Registration is the act of establishing spatial correspondences between two images, either in the form of sparse corresponding landmarks or as a dense spatial mapping. Automatic registration methods often employ optimization of a voxel similarity metric (VSM). However, VSM itself does not indicate any estimates or bounds on true registration fidelity [1]. This is due to several reasons such as similar appearance anatomy, homogeneous tissue regions and partial volume effect. Nevertheless, fidelity can be assessed through independent measurements, e.g. selecting landmarks or segmenting target images, for evaluating a registration method or a particular case. However, this is impractical to perform on each target image following an automatic method. Registration is often employed precisely to avoid such overhead of labeling target images or in cases when the relevant anatomy is not identifiable in the target image modality. Consequently, a post-registration VSM is typically the only output measure to judge registration fidelity.

Registration uncertainty has been studied in the literature using voxel-statistics methods such as Cramér-Rao bound [2] and bootstrap resampling [3]. Detection of displaced edges between registered and target images using the state-space of multiple Gaussian filters on intensity difference images was studied in [4]. For assessing a group of registrations, voxels-statistics based on active appearance models generated from registered images was used in [5]. Registration fidelity cannot be precisely implied from such voxel-based criteria either, similarly to discussion on VSM above. Another approach to fidelity estimation focuses on the consistency of registrations among themselves. For instance, a pairwise consistency requirement is that the correspondences between two images when registered in either direction shall be unique. This was exploited in devising a symmetric registration method in [6] and as an atlas classification metric in [7]. Such consistency between three serial intra-patient MR images was used to assess the validity of results in [8]. Consistency was measured as the root-mean-square displacement residual of uniformly distributed points transformed through a triangular loop of three registrations (also known as a *registration circuit*). Woods *et al.* evaluated their registration suite for intra-patient registration by considering a norm of such triangular consistency over all possible loops in a given set of images [9]. A similar approach was later used to compare VSMs for serial brain MR registration [10] and to compare nonrigid registration methods [11].

Such registration loops were recently employed by Datteri and Dawant to estimate individual registration fidelity; target registration error in [12] and Dice overlap of resulting atlas-based segmentation in [13]. This technique employs an assumed analytical model of registration error accumulation over a loop. An aggregation of several such loop criteria in a system of equations then leads to a least-squares solution for individual fidelity metrics. In [12, 13], improving multi-atlas to single-target registration has been studied with the goal of atlas-selection and fusion. In this paper, we devise a new solution technique for this method and show that it benefits from estimating several targets *concurrently* by leveraging information from inter-target registrations.

2. METHODS

Let each image in a set of n images be X_i for $i = 1..n$ where ℓ of these images are the so-called atlases with known ground-truth segmentations S_i for $i{=}1..\ell{<}n$. Let T_{ij} be the transformation (deformation field for nonrigid registration) resulting from the registration of X_i to X_j such that $T_{ij}(X_i) \approx X_j$ as seen in Fig. 1(left). Then, based on a known segmentation S_i of image X_i, a segmentation S_j^i can be estimated for X_j using this transformation, i.e. $S_j^i = T_{ij}(S_i)$, namely the *single-atlas segmentation*. Similarly, typical multi-atlas segmentation combines (fuses) multiple such segmentation estimates based on some confidence on each one. Thus, determining such confidence on each individual registration then becomes crucial. Such ranking of atlases is also essential in single-atlas segmentation to select the optimal atlas (image retrieval) for a best segmentation outcome. Below, we show a method for such belief assignment and our contributions to it, which (i.) improve accuracy, (ii.) reduce computation time, and (iii.) reduce memory footprint significantly so that much larger image sets can be processed.

For three images $\{X_i, X_j, X_k\}$ seen in Fig. 1(center), let T_{ijk} denote the composition (ordered combination) of shown three pairwise transformations in a loop, i.e.:

$$T_{ijk} = T_{ki}(T_{jk}(T_{ij}(\cdot))) = T_{ki} \circ T_{jk} \circ T_{ij} \qquad (1)$$

where $\{i, j, k\}{\leq}n$ and $i{\neq}j{\neq}k{\neq}i$. Holden *et al.* [10] suggested that when all these three transformations are ideal (i.e., "correct"

This work is supported by the NCCR Co-Me of the Swiss National Science Foundation.

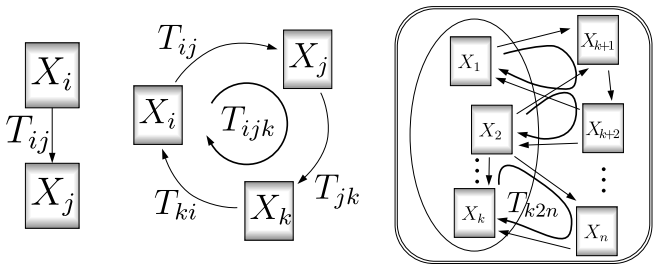


Fig. 1. Registration of two images (left), a loop of three registrations (center), and examples loops in an image set (right).

anatomical correspondences are identified between each pair) then T_{ijk} should map the image X_i back onto itself. Conversely, deviation of T_{ijk} from the identity transform was used to assess and compare different registration algorithms. Such deviations can also be computed only at selected target points in X_i, leading to a target registration error along this loop. Extending on this idea of registration loops or so-called circuits, one shall also expect that a segmentation propagated over such a loop should be similar (and in fact identical) to itself, i.e. $T_{ijk}(S_i) \simeq S_i$. Thus, any deviation from such similarity, which can be measured by known segmentation quality metrics such as Dice's similarity metric (DSM), can be used as an evaluation criteria for the registration method or its parameters [9, 10]. This criteria is used in this paper as a measure of how well particular registrations in a loop (or, precisely, the combination thereof) do perform. Accordingly, let d be the quality metric, such that for DSM the error in a loop $\tilde{d}_{ijk}$=DSM$(T_{ijk}(S_i), S_i)$. Then, our goal is to estimate all pair-wise metrics d_{ij}=DSM$(T_{ij}(S_i, S_j)$ using such computed loop metrics $\tilde{d}_{ijk}$. For this, a model describing how each pair-wise registration error might contribute to an entire loop error has to be assumed. Datteri and Dawant [13] suggested an additive error model so that a closed-form least-squares solution is identified considering multiple loops in a set of images. Assuming that error accumulates additively along a registration loop:

$$\tilde{d}_{ijk} \simeq d_{ij} + d_{jk} + d_{ki} \; . \tag{2}$$

where $i \leq \ell$ because DSM can only be computed for loops starting from atlas images, which have known segmentations.

Using a symmetric registration method as in [13], $T_{ij} = T_{ji}^{-1}$ and hence one can assume $d_{ij} = d_{ji}$ and $\tilde{d}_{ijk} = \tilde{d}_{ikj}$. However, symmetry is not a general condition for typical registration methods and, therefore, the registrations in both direction and the involved errors therewith may be different. Collocating the linear combinations (2) from all unique registration loops as in Fig. 1(right) in a system of equations yields:

$$\begin{bmatrix} \boldsymbol{\delta}_{123} \\ \boldsymbol{\delta}_{124} \\ \vdots \\ \boldsymbol{\delta}_{ijk} \\ \vdots \end{bmatrix} \begin{bmatrix} d_{12} \\ d_{13} \\ \vdots \\ d_{n(n-1)} \end{bmatrix} = \begin{bmatrix} \tilde{d}_{123} \\ \tilde{d}_{124} \\ \vdots \\ \tilde{d}_{ijk} \\ \vdots \end{bmatrix} \tag{3}$$

$$\boldsymbol{A}\,\boldsymbol{d} \;=\; \tilde{\boldsymbol{d}} \tag{4}$$

where $\boldsymbol{\delta}_{ijk}$ are Kronecker delta row-vectors of length $n(n-1)$ with all zeros but three ones as follows:

$$\boldsymbol{\delta}_{ijk} = \begin{cases} \delta_{ijk}[t] = 1 & \text{, if} \quad t \in \{i, j, k\} \\ \delta_{ijk}[t] = 0 & \text{, otherwise} \end{cases} \tag{5}$$

Recall that $i \in [1, \ell]$ and $\{j, k\} \in [1, n]$ where $i \neq j \neq k \neq i$.

Casting (4) as the optimization problem $(\min\|\boldsymbol{A}\boldsymbol{d}-\tilde{\boldsymbol{d}}\|_2)$, a least squares solution can be found using pseudo-inverse as:

$$\boldsymbol{d} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\,\tilde{\boldsymbol{d}} \tag{6}$$

A multiplicative error model can be cast as in (2) after variable substitutions $\boldsymbol{d'}$=log$(\boldsymbol{d})$ and $\tilde{\boldsymbol{d}}'$=log$(\tilde{\boldsymbol{d}})$ as follows:

$$\tilde{d}_{ijk} \;\simeq\; d_{ij}\, d_{jk}\, d_{ki} \tag{7}$$

$$\log(\tilde{d}_{ijk}) \;=\; \log(d_{ij}) + \log(d_{jk}) + \log(d_{ki}) \tag{8}$$

$$\tilde{d}'_{ijk} \;=\; d'_{ij} + d'_{jk} + d'_{ki} \tag{9}$$

and thus solved similarly to (4).

The solution of (6) cannot guarantee any bounds on the estimated values $\boldsymbol{d}$. However, in (2) $\boldsymbol{d}$ was assumed to have a unit of DSM. Therefore, $\boldsymbol{d} \in [0, 1]$ should be satisfied. In this paper, we formulate (4) as a constrained optimization problem:

$$\min \|\boldsymbol{A}\boldsymbol{d} - \tilde{\boldsymbol{d}}\|_2 \qquad \text{s.t.} \quad 0 \leq \boldsymbol{d} \leq 1 \tag{10}$$

This is then solved using the reflective Newton trust-region method, an iterative optimization method chosen due to its computational and memory efficiency [14]. This proposed solution technique using the multiplicative model (7) is called OPT* in this paper. The original analytical solution method of [13] with the additive assumption is called ALG$^+$.

In a set of n images with given ℓ atlases, consider the goal of finding the best possible single-atlas segmentations for the rest of the $n-\ell$ images (targets). In ALG$^+$, it was proposed to use one single target image together with the available atlas images in order to generate estimates from each atlas to this target. Then, for each target in $[\ell+1, n]$ a separate system of (4) with $(\ell+1)\ell$ unknowns and $\ell^2(\ell-1)$ equations are to be solved. Having accordingly estimated $d_{ij} \forall_{i\in[1,\ell]\,, j\in[\ell+1,n]}$, one would then choose (retrieve) the best atlas for each target:

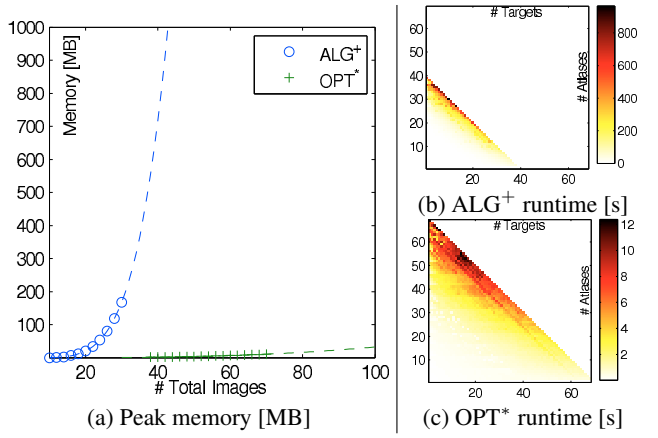
$$\forall_{j\in[\ell+1,n]} \quad i_j^* = \arg\max_{i\in[1,\ell]} d_{ij} \tag{11}$$

where i_j^* is the number of optimal atlas choice for target j. In this paper, we instead propose by OPT* to solve for all available targets within the same equation system concurrently. This yields a system of $n(n-1)$ unknown estimations with $\ell(n-1)(n-2)$ known equations. Having solved all d_{ij} estimations concurrently, the best atlases are picked as in (11).

3. RESULTS AND DISCUSSION

We compared the proposed OPT* with the state of the art method ALG$^+$ on two datasets: 70 midsagittal MR slices of brain with segmented corpus collasum (CC) and 15 3D head CT scans with segmented (mandibular) jaw bone (JB). These datasets are, respectively, referred as CC MR and JB CT. All images of each dataset were first registered to each other using a Markov random field based deformable registration technique [15], which led to the transformations T_{ij} that we employed in the methods described above.

Although $\boldsymbol{A}$ in (6) is sparse, $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}$ is a dense matrix with $\mathcal{O}(n^4\ell^2)$ elements. In contrast, the iterative trust-region method in OPT* does not build the Hessian, hence yielding a small memory footprint. In Fig. 2(a), peak memory requirements for both methods are plotted when processing datasets of n images (with ℓ=n−1



(a) Peak memory [MB] (b) ALG$^+$ runtime [s] (c) OPT* runtime [s]

Fig. 2. Comparisons of (a) peak memory requirements, and (b-c) computation time.

atlases). ALG$^+$ requirement is seen to increase exponentially. We fitted polynomials to experimental points in the figure to indicate this trend. Indeed, with $n{>}45$ images our setup cannot perform ALG$^+$ anymore, whereas all our images can be processed concurrently using OPT*. For the CC MR dataset of $n{=}70$, the polynomial fit estimates that ALG$^+$ would require 10 GB of memory, whereas OPT* in practice necessitated a mere 11 MB.

The large matrix inversion in (6) also affects the algorithmic speed. Fig. 2(b-c) shows the time required by each method for different numbers of atlas (ℓ) and target ($n{-}\ell$) combinations. For each combination, an average from 20 experiments using randomized atlas-target selections from the CC MR datasets was plotted. Due to memory limitation, ALG$^+$ was only computed for $n{\leq}40$. For $n{=}40$ it already takes nearly 15 min, whereas for all $n{\leq}70$ images OPT* performs under 12 s on average. The worst case OPT* runtime was not more than 18 s for any experiment.

For comparing accuracy similarly to [13], we have observed the correlation of the estimated atlas-to-target d_{ij} values to their ideal DSM $\tilde{d}_{ij}$, which are computed from the registrations using their ground-truth segmentations. We use two metrics: Pearson's product-moment correlation coefficient $r(d_{ij}, \tilde{d}_{ij})$, which seeks for an underlying linear relation between the values; and Spearman's rank correlation coefficient $\rho(d_{ij}, \tilde{d}_{ij})$ which quantifies the rank (ordering) relation between those [16]. 20 randomized experiments for each different atlas-target numbers (with $n{\leq}40$) was conducted on each dataset. Fig. 3 shows the improvement in both linear and rank relationship from using our proposed method. The correlation difference images show that in no case OPT* performs worse r than ALG$^+$. The average values of such correlations per method is given in Table 1. For comparisons, we also included the methods ALG* for algebraic solution (6) with multiplicative assumption (7) and OPT$^+$ for constrained trust-region method (10) with additive assumption (2). The proposed OPT* outperforms all other three combinations.

Although for Fig. 3 we performed ALG$^+$ for more than one targets as well, the original method in fact proposes its use in only single target case (the left-most column of the correlation images). Nevertheless, we noted that for the same atlas number, in particular for small number of atlases, the correlation indeed increases as the target number is increased (observe a single row of the correlation images). We hypothesize that this is due to the leverage of registration information between target images in the system solution (4). Note that in the original ALG$^+$ method, no inter-target registrations

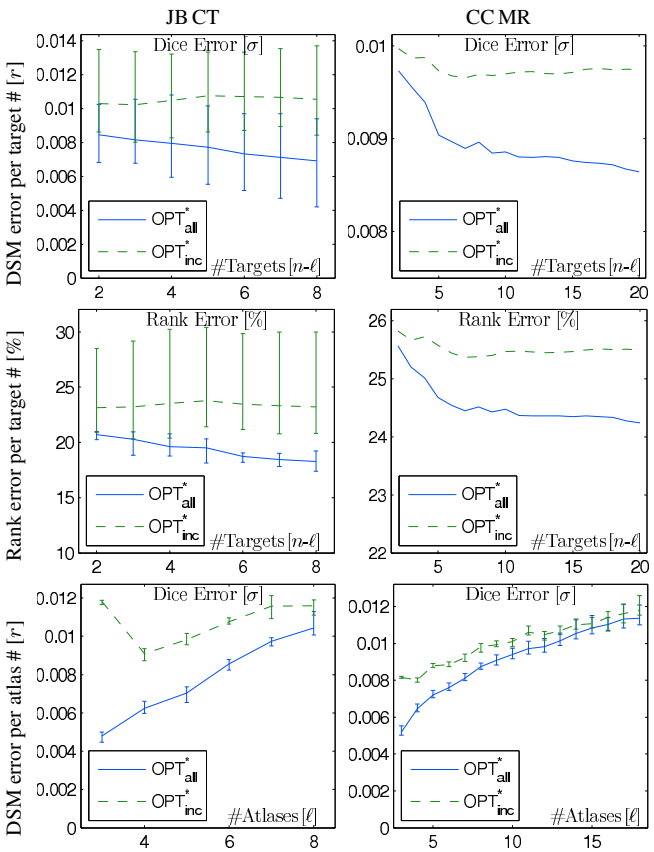


Fig. 4. For selecting an atlas with each method, average DSM (top) and rank (middle) errors per number of targets, and DSM error (bottom) per number of atlases. For JB CT, the extent of errors for all experiments are also shown.

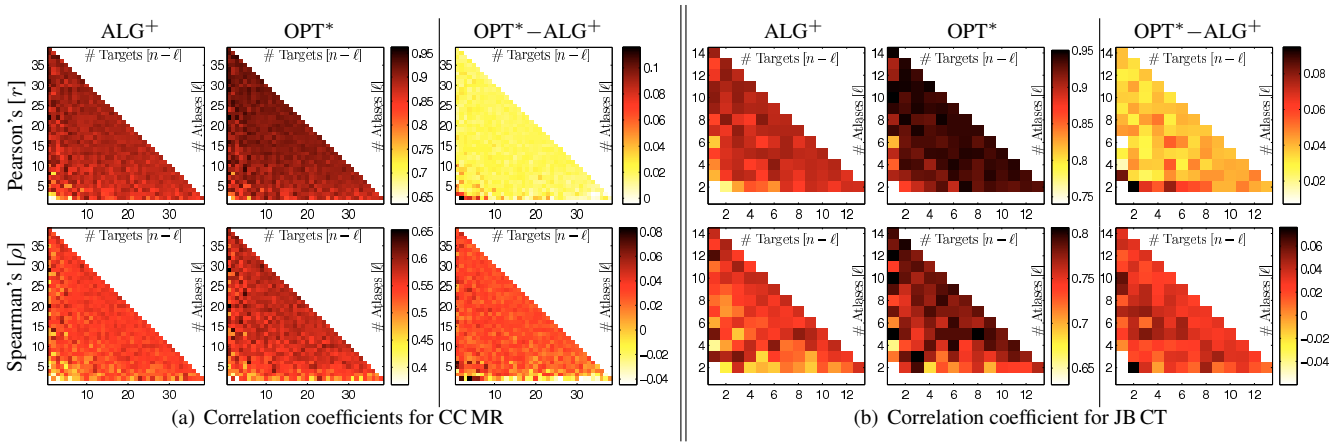
are employed, whereas our proposed method includes those. To test this hypothesis, we measured the success of our image retrieval task in (11). For each target j of each randomized experiment, we compared the best theoretically achievable DSM $\max_{i\in[1,\ell]} \tilde{d}_{ij}$ to the DSM achieved in practice by selecting the atlas i_j^* with the best estimation using (11), i.e. $\tilde{d}_{i_j^* j}$. We accordingly define an error:

$$\sigma_j \overset{\text{def}}{=} -\tilde{d}_{i_j^* j} + \max_{i\in[1,\ell]} \tilde{d}_{ij} \quad , \tag{12}$$

which is zero for predicting the *correct* atlas. Fig. 4(top) shows the error from randomized experiments averaged over different number of atlases. We also computed a rank error $p \in [1, \ell]$ as the rank of each retrieval i_j^* within the atlases sorted by $\tilde{d}_{ij}$. This rank is normalized to a percentage, i.e. by $\frac{p-1}{\ell-1}$, for each experiment and it is shown as averaged in Fig. 4(middle).

When atlas supervision is weak using few atlases, inter-target registrations can provide valuable additional information. Error σ is shown in Fig. 4(bottom) for using different small numbers of atlases. The improvement with our method is seen, in particular for using three atlases with JB CT images where the error is more than halved.

$\boldsymbol{A}$ in (4) is not necessarily full-rank, when *nonsymmetric* registrations, which are more readily available, are used. As both the algebraic and the iterative methods seek for a minimum-norm solution, a unique minimizer is found nonetheless. Such solutions are



(a) Correlation coefficients for CC MR

(b) Correlation coefficient for JB CT

Fig. 3. Pearson's linear and Spearman's rank correlation coefficients are shown using ALG$^+$ and OPT* methods on both datasets. The improvement from using the proposed method is highlighted by showing the difference between them.

Table 1. Average correlation values for each method.

		ALG$^+$	ALG*	OPT$^+$	OPT*
LJ CT	Pearson's [r]	0.889	0.861	0.880	0.930
	Spearman's [ρ]	0.739	0.704	0.703	0.769
CC MR	Pearson's [r]	0.880	0.872	0.865	0.901
	Spearman's [ρ]	0.538	0.521	0.521	0.564

seen to perform well, as demonstrated in both the correlation and atlas-selection results; with our method outperforming the others.

The reason for the observed DSM improvements being relatively small is that all original DSMs from our MRF-based registration is relatively high, e.g. $\geq$90% of DSM are larger than 0.9 for CC MR. Then, even retrieving a wrong atlas can give a large DSM.

4. CONCLUSIONS

We present an improved technique for registration estimation, enabling rapid computation in large image datasets. We show in two different datasets that information from registrations between target images can indeed be leveraged to improve such estimation accuracy. The achieved fidelity estimation improvement in single-atlas selection is expected to translate similarly to multi-atlas fusion.

5. REFERENCES

[1] WR Crum, LD Griffin, DLG Hill, and DJ Hawkes, "Zen and the art of medical image registration: correspondence, homology, and quality," *Neuroimage*, vol. 20, pp. 1425–1437, 2003.

[2] D Robinson and P Milanfar, "Fundamental performance limits in image registration," *IEEE T Img Process*, vol. 13, no. 9, pp. 1185 –1199, 2004.

[3] J Kybic, "Bootstrap resampling for image registration uncertainty estimation without ground truth," *IEEE T Img Proc*, vol. 19, no. 1, pp. 64–73, Jan 2010.

[4] W Crum, L Griffin, and D Hawkes, "Automatic estimation of error in voxel-based registration," in *Proc MICCAI*, 2004.

[5] R Schestowitz, C Twining, and V Petrovic, "Non-rigid registration assessment without ground truth," in *Proc Med Img Unders Anal (MIUA)*, 2006.

[6] GE Christensen and HJ Johnson, "Consistent image registration," *IEEE T Med Imag*, vol. 20, no. 7, pp. 568–582, 2001.

[7] T Gass, G Szekely, and O Goksel, "Semi-supervised segmentation using multiple segmentation hypotheses from a single atlas," in *MICCAI W Med Comp Vision*, Oct 2012.

[8] L Lemieux, UC Wieshmann, NF Moran, DR Fish, and SD Shorvon, "The detection and significance of subtle changes in mixed-signal brain lesions by serial MRI scan matching and spatial normalization," *Med Img Anal*, pp. 227–242, 1998.

[9] RP Woods, ST Grafton, CJ Holmes, SR Cherry, and JC Mazziotta, "Automated image registration: I. general methods and intrasubject, intramodality validation," *J Comp Assisted Tomog*, vol. 22, pp. 139–52, 1998.

[10] M Holden, DL Hill, ER Denton, JM Jarosz, TC Cox, T Rohlfing, J Goodey, and DJ Hawkes, "Voxel similarity measures for 3-D serial MR brain image registration," *IEEE T Med Imag*, vol. 19, pp. 94–102, 2000.

[11] GE Christensen and HJ Johnson, "Invertibility and transitivity analysis for nonrigid image registration," *J Electronic Imaging*, vol. 12, no. 1, pp. 106–117, 2003.

[12] RD Datteri and BM Dawant, "Estimation and reduction of target registration error," in *MICCAI*, 2012, pp. 139–146.

[13] RD Datteri, AJ Asman, BA Landman, and BM Dawant, "Estimation of registration accuracy applied to multi-atlas segmentation," in *MICCAI W Multi-Atlas Label Stat Fusion*, 2011.

[14] TF Coleman and Y Li, "A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables," *SIAM J Optimization*, vol. 6, pp. 1040–1058, 1996.

[15] Ben Glocker, Nikos Komodakis, Georgios Tziritas, Nassir Navab, and Nikos Paragios, "Dense image registration through mrfs and efficient linear programming," *Med Img Anal*, vol. 12, no. 6, pp. 731–741, 2008.

[16] DJ Best and DE Roberts, "Algorithm as 89: The upper tail probabilities of spearman's rho," *Applied Statistics*, vol. 24, pp. 377–379, 1975.