Published in final edited form as:

IEEE Trans Med Imaging. 2015 January ; 34(1): 86–96. doi:10.1109/TMI.2014.2344911.

Validation of a Non-Rigid Registration Error Detection Algorithm Using Clinical MRI Brain Data

Ryan D. Datteri [Student Member, IEEE], Yuan Liu [Student Member, IEEE], Pierre-François D’Haese, and Benoit M. Dawant [Fellow, IEEE]

Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37235 USA

Abstract

Identification of error in non-rigid registration is a critical problem in the medical image processing community. We recently proposed an algorithm that we call “Assessing Quality Using Image Registration Circuits” (AQUIRC) to identify non-rigid registration errors and have tested its performance using simulated cases. In this article, we extend our previous work to assess AQUIRC’s ability to detect local non-rigid registration errors and validate it quantitatively at specific clinical landmarks, namely the Anterior Commissure (AC) and the Posterior Commissure (PC). To test our approach on a representative range of error we utilize 5 different registration methods and use 100 target images and 9 atlas images. Our results show that AQUIRC’s measure of registration quality correlates with the true target registration error (TRE) at these selected landmarks with an $R^2 = 0.542$. To compare our method to a more conventional approach, we compute Local Normalized Correlation Coefficient (LNCC) and show that AQUIRC performs similarly. However, a multi-linear regression performed with both AQUIRC’s measure and LNCC shows a higher correlation with TRE than correlations obtained with either measure alone, thus showing the complementarity of these quality measures. We conclude the article by showing that the AQUIRC algorithm can be used to reduce registration errors for all five algorithms.

Keywords

Image registration; registration circuits; registration error; non-rigid registration

I. Introduction

Although analytical solutions have been derived to estimate error for the point-based rigid-body registration problem [1], no such solution exists for the non-rigid case or even for the rigid case when the transformations are not estimated with homologous points. Assessing the quality of a particular non-rigid registration between image volumes thus remains a difficult and outstanding problem facing the medical imaging community. Few solutions have been proposed that do not require pre-labeled atlases. These solutions generally fall into two broad categories: Bayesian methods and supervised learning techniques.

Bayesian methods have been proposed for instance to estimate registration uncertainty, and have been used to regularize the deformation field in [2], provide confidence information on volumetric measurements in [3], to estimate uncertainty in intra-subject registration in [4] or to estimate lung elasticity in [5]. These techniques require the estimation of posterior distributions, which is done either with Markov Chain Monte Carlo methods [3-5], a computationally demanding approach, or mean-field variational Bayesian techniques. In [6] a bootstrap resampling technique is proposed to estimate the variability of an estimated transformation. Evaluation of this technique was limited to 2D images and simple transformations. Other methods have been proposed that use supervised learning techniques [7-9] on image features or manually labeled points to identify registration uncertainty. These techniques, which require a training set for each new application and data set, have been used, for instance, to assess registration accuracy in longitudinal CT images of the lungs [7, 8] or to detect misregistered regions in simulated brain tumor images [9].

Because assessing the quality of an individual registration is difficult, a number of techniques have been developed over the years to make the transformation estimation process robust and to produce transformations that are likely to be accurate. These include enforcing desirable transformation properties such as inverse consistency and transitivity consistency [10-12]. A number of algorithms that enforce inverse consistency have been proposed. A representative set of such algorithms can be found in [13-22]. The consistency of three transformations was used by Holden et al. [12], Freeborough [24], Woods et al. [11], and Christensen [10] to compare and evaluate several similarity measures and registration algorithms. This technique involves three images and three transformations, T_{AB} , T_{BC} , T_{CA} , i.e. the transformation from B, from B to C and from C to A, respectively. Composing the three transformations maps the coordinate system of image onto itself through a circuit. If all three transformations are error free, any voxel in image A is mapped exactly onto itself. Errors along the circuit may be detected by computing the distance between a point x and its projection x' , where $x' = T_{CA}(T_{BC}(T_{AB}(x)))$. In past work, registration circuits have also been referred to as triplets and loops. Here we use the term *circuit* that is used in graph theory.

The invertibility and transitivity properties of transformations were also used by Christensen and Johnson [25] to evaluate non-rigid registration algorithms. In more recent work [16], the same group proposed a registration algorithm called TICMR for Transitive Inverse Consistent Manifold Registration (TICMR) that jointly estimates a correspondence between three manifolds while minimizing inverse and transitivity consistency. While, undoubtedly, transitivity consistency is a necessary condition for error-free registration around a registration circuit, it is not sufficient. As discussed by Christensen [25], the identity transformation is a trivial example of a useless transformation that would minimize both inverse consistency and transitivity error. Also, transitivity error is an aggregate measure of error across the circuit. It does not identify on which edge in the registration circuit error occurs, and errors may be masked because an error on one edge can be compensated by an error on another edge in the registration circuit. As a consequence, transitivity error has traditionally been used to compare and evaluate algorithms with the assumption that lower

transitivity error combined with other checks that would rule out unrealistic or useless transformations is synonymous with a better registration algorithm.

In this work, we calculate transitivity error over multiple registration circuits that have images in common. Thus an individual registration is used in multiple registration circuits; using this redundancy allows for the estimation of error for an individual registration. To the best of our knowledge, transitivity error has not been used to assess the quality of a particular registration between two image volumes until recently [26-29]. In this published work, we built on the idea of transitivity error and have shown that it could indeed be used to evaluate the quality of an individual image registration. This algorithm, that we call AQUIRC for Assessing Quality Using Image Registration Circuits, was first proposed in [26] where it was used for global atlas selection. In [27] we have shown that it could be used to detect errors in intensity-based rigid body registration problems. In [28] we presented preliminary and qualitative results on simulated data suggesting that this approach could be used to detect local registration error. In [29] we used the estimation of error at the local error for local atlas selection. Herein, we extend upon the work in [28]. We apply AQUIRC to a data set of 109 medical images with manually identified ground truth, treating 9 images as atlas images and the remaining 100 as target images. We use 5 popular registration algorithms to provide a wide range of error that is representative of what might be expected when using a registration algorithm in practice. Using these data and algorithms, we show that AQUIRC's quality measure correlates well with the ground truth at the Anterior Commissure (AC) and the Posterior Commissure (PC) points. The AC and PC are two reference points used clinically for brain normalization to, for instance, guide the placement of Deep Brain Stimulation (DBS) electrodes. We use the AC and PC points, which can be localized accurately in MR images, as this allows us to correlate our results with a known target registration error locally. Also, we [26, 29] as well as others [30] have used other error measures such a DICE to evaluate our approach. In [26] we applied AQUIRC to global atlas selection and showed that it can be used to improve the DICE value over majority vote as well as over a residual global NMI atlas selection method. We also found in [29] that AQUIRC can be used to improve upon the DICE value over majority vote in the context of local atlas selection for the brainstem and to perform comparably to cutting edge atlas selection techniques on other brain structures.

In this article we first present the results of AQUIRC qualitatively. Then, we compute the AQUIRC algorithm and correlate the results to the ground truth TRE. To compare our algorithm to a more traditional intensity-based approach [31], we compute the popular LNCC [32] at the AC and PC between the atlases and the target images and we compute the correlation between this quantity and the ground truth TRE. We then show that LNCC and AQUIRC provide complementary information, i.e., that in a multi-linear regression, AQUIRC and LNCC can increase the correlation to TRE. To evaluate the sensitivity of our approach, we vary the smoothness of the transformations computed when registering the images and we vary the quality of the images we use as atlases. Finally, we show one way by which AQUIRC can be used to reduce registration error. The next section describes the methods used in this work. This is followed by the results section. Finally, we discuss the results and potential applications of the AQUIRC algorithm.

II. Methods

The AQUIRC algorithm builds on the idea of registration circuits. The main idea is that, although registration error between two images cannot be estimated with one single circuit, it can be done when multiple circuits are used. Here, a registration circuit involves three images A, B, and C and three transformations T_{AB} , T_{BC} , and T_{CA} . An example registration circuit is shown in Figure 1. In this work, we assume that each registration in the circuit affects the consistency error measure multiplicatively. This allows us to define the error across a registration circuit as $E_C = \varepsilon_A * \varepsilon_B * \varepsilon_C$.

The multiplicative model can be solved linearly by computing the logarithm of the equation, i.e. $\log(E_C) = \log(\varepsilon_A) + \log(\varepsilon_B) + \log(\varepsilon_C)$. We refer to EC as the consistency error and the variables ε_A , ε_B , and ε_C represent the unknown registration quality for each individual registration in the circuit. These are the quantities of interest. We note that an additive model may also be applicable. We have tested it and found that the multiplicative model performed slightly better. As a consequence, we chose not to present results obtained with the additive model to avoid redundancy.

To compute the consistency error E_C we first select a set of points, defined here as X , in an image. We then project these points back to the original image through the transformations relating the images across a circuit, i.e., $X' = T_{CA}(T_{BC}(T_{AB}(X)))$. In the implementation presented in this article, we define E_C as the Euclidian distance between X and X' , i.e., $E_C = \|X - X'\|$.

The variable E_C is affected by error in three registrations: the registrations error between A and B, between B and C, and between C and A. As discussed earlier and also noted by Fitzpatrick [33], using only one registration circuit can lead to an underestimation of registration error because the error made along one edge in the circuit may correct error introduced from a separate edge in the circuit. To address this issue, we expand the idea of registration circuits to multiple circuits. To do this, we start with a set of N images and compute pair-wise registrations between all images in the set, creating a complete graph as shown in Figure 2. The complete graph of registrations is similar to what is done by Christensen [34]. However, in [34], the graph of registrations is used as an overall measure of the quality of a registration algorithm, rather than as a method to determine the quality of individual registrations as proposed in this work. In our paradigm, we consider the network of images to be an undirected graph. Because of this, the error estimation for a registration between two images A and B is the same for the transformation deforming A onto B and for the transformation deforming B onto A. We can do this because either the registration algorithms we utilize are inverse consistent by design or we enforce inverse consistency after the registration is performed (this is fully described in section II C). Early in our work, we tested a directed graph version but this was not pursued because we did not observe an improvement in the results and a directed graph adds computational complexity and requires more memory.

If our data set contains N images, the graph contains $\binom{N}{2}$ edges (i.e. individual registrations) for which we associate an ε value. There are $\binom{N}{3}$ unique registration circuits that can be formed from a complete graph.

The registration quality can be estimated globally, i.e., the overall registration between two images can be evaluated, or locally, i.e., a quality measure can be estimated at each voxel. Here, we focus on the latter, and the consistency error, E_C , is defined for each voxel i , as $E_C^i = \|x_i - x'_i\|$, where x_i are the coordinates of voxel i , and x'_i are the transformed coordinates across a registration circuit of voxel i . $\overline{E}_C = [E_C^1, E_C^2, \dots, E_C^Q]$ then, is a vector of consistency errors across a registration circuit, with Q equal to the number of voxels. We estimate the registration quality between two images at each voxel by defining a vector of registration quality $\overline{\varepsilon} = [\varepsilon^1, \varepsilon^2, \dots, \varepsilon^Q]$, where ε^i represents the registration quality at voxel i and $\overline{\varepsilon}$ represents a rasterized vector of registration quality values, with each entry in that vector corresponding to one particular voxel.

Figure 3 illustrates by means of a graph with 4 nodes the way we set up our system of equations. In this figure, each node contains four voxels placed at the intersection of the gridlines. In practice, an $N = 5$ is required to compute a linear least squares solution. Here, we use $N = 4$ to illustrate the process. In this example, there are 4 possible registration circuits. Thus we can construct a system of equations for voxel 1 as:

$$\begin{aligned} \log(E_{C_1}^1) &= \log(\varepsilon_1^1) + \log(\varepsilon_2^1) + \log(\varepsilon_3^1) \\ \log(E_{C_2}^1) &= \log(\varepsilon_1^1) + \log(\varepsilon_6^1) + \log(\varepsilon_5^1) \\ \log(E_{C_3}^1) &= \log(\varepsilon_3^1) + \log(\varepsilon_4^1) + \log(\varepsilon_5^1) \\ \log(E_{C_4}^1) &= \log(\varepsilon_6^1) + \log(\varepsilon_2^1) + \log(\varepsilon_4^1) \end{aligned}$$

where for $E_{C_p}^i$, p refers to a particular circuit, for ε_l^i , l refers to an edge, i.e. a link between two images and for both terms i again represents a particular voxel; computing E_C requires composing transformations across the circuits, which is done through tri-linear interpolation. This can be rewritten in matrix form as:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \log(\varepsilon_1^1) \\ \log(\varepsilon_2^1) \\ \log(\varepsilon_3^1) \\ \log(\varepsilon_4^1) \\ \log(\varepsilon_5^1) \\ \log(\varepsilon_6^1) \end{bmatrix} = \begin{bmatrix} \log(E_{C_1}^1) \\ \log(E_{C_2}^1) \\ \log(E_{C_3}^1) \\ \log(E_{C_4}^1) \end{bmatrix} \quad (1)$$

The same system of equations can be created for each of the other 3 voxels and, appended on to (1), to produce (2):

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \log \left(\begin{bmatrix} \varepsilon_1^1, \varepsilon_1^2, \varepsilon_1^3, \varepsilon_1^4 \end{bmatrix} \right) \\ \log \left(\begin{bmatrix} \varepsilon_2^1, \varepsilon_2^2, \varepsilon_2^3, \varepsilon_2^4 \end{bmatrix} \right) \\ \log \left(\begin{bmatrix} \varepsilon_3^1, \varepsilon_3^2, \varepsilon_3^3, \varepsilon_3^4 \end{bmatrix} \right) \\ \log \left(\begin{bmatrix} \varepsilon_4^1, \varepsilon_4^2, \varepsilon_4^3, \varepsilon_4^4 \end{bmatrix} \right) \\ \log \left(\begin{bmatrix} \varepsilon_5^1, \varepsilon_5^2, \varepsilon_5^3, \varepsilon_5^4 \end{bmatrix} \right) \\ \log \left(\begin{bmatrix} \varepsilon_6^1, \varepsilon_6^2, \varepsilon_6^3, \varepsilon_6^4 \end{bmatrix} \right) \end{bmatrix} = \begin{bmatrix} \log \left(\begin{bmatrix} E_{c_1}^1, E_{c_1}^2, E_{c_1}^3, E_{c_1}^4 \end{bmatrix} \right) \\ \log \left(\begin{bmatrix} E_{c_2}^1, E_{c_2}^2, E_{c_2}^3, E_{c_2}^4 \end{bmatrix} \right) \\ \log \left(\begin{bmatrix} E_{c_3}^1, E_{c_3}^2, E_{c_3}^3, E_{c_3}^4 \end{bmatrix} \right) \\ \log \left(\begin{bmatrix} E_{c_4}^1, E_{c_4}^2, E_{c_4}^3, E_{c_4}^4 \end{bmatrix} \right) \end{bmatrix} \quad (2)$$

This expression can be re-written in a more compact form as $\bar{P} \log(\bar{\varepsilon}) = \log(\bar{E}_c)$.

Where $\bar{P}$ is a matrix that represents the edges utilized for each circuit, with the value of $\bar{P}$ set to 1 if the edge is utilized in the circuit and 0 otherwise. In the matrix $\bar{P}$, each row represents a registration circuit while each column represents individual edges in the network of images. As a result of the multiplicative assumption, $\log(\bar{\varepsilon})$ can be solved for using a linear least squares solution:

$$\log(\bar{\varepsilon}) = \left(\bar{P}^T \bar{P} \right)^{-1} \bar{P}^T \log(\bar{E}_c) \quad (3)$$

and solving for $\bar{\varepsilon} = e^{\log(\bar{\varepsilon})}$. The formulation in (2) and (3) can be expanded to an arbitrary number of nodes, i.e., images, and to every voxel in the images.

The multiplicative model assumes that the registration circuit error, E_C , is non-zero and that the error estimate for each registration, ε , is non-zero. After calculating E_C , it is necessary to check for any value equal to zero because we use its log value. In practice, we discard any voxel i in which $E_C^i = 0$. In our experience $E_C^i = 0$ has never occurred inside the foreground region of an image.

In this work we reformat our images to have the same number of voxels and voxel size prior to the creation of the network. It is possible to utilize images with differing voxel sizes and with a different number of voxels, but this is out of the scope of this paper. Experimentally $\bar{P}$ has been observed to be full rank when $N \geq 5$. Finally, while we have used registration circuits with 3 images in this study, the number of images in a circuit can be increased to form additional registration circuits but this has not yet been explored.

A. Data

Our data set consists of 100 target images and 2 sets of 9 atlas images. Atlas set (1) contains five atlas images that are normal clinical images while the remaining four atlas images have features that are known to be challenging for registration algorithms: two are from patients with high grade gliomas, and two contain Deep Brain Stimulation (DBS) electrodes. Atlas set (2) contains 9 atlas images that are normal clinical images, with 5 of those images being

the same 5 clinical images as in Atlas set (1). Figure 4 shows representative slices in an atlas with bilateral implants (left), in a normal atlas (middle), and in an atlas with a high grade glioma (right).

All images are T1-weighted sagittal MR volumes that have $256 \times 256 \times 170$ voxels and 1 mm in each direction. The images were acquired with the SENSE parallel imaging technique from Philips on a 3 Tesla scanner.

B. AC and PC Identification Rules

Two raters manually identified both the AC and PC points in the 109 MRI images. The inter-rater variability between the two raters is 0.573 ± 0.39 mm for the AC and 0.563 ± 0.33 mm for the PC. It is close to half a voxel in our data set. This is lower than published clinical variability [35], which is 1.53 ± 1.44 mm for the AC and is 1.45 ± 1.24 mm for the PC. The average coordinates of the two raters were utilized as the ground truth. Accurate selection of the AC and PC points is important for applications such as Deep Brain Stimulation (DBS) surgical procedures [36]. These require the implantation of electrodes in specific brain nuclei and the AC and PC points are used to define a coordinate system that permits localizing the implantation targets even though these are not easily discernible in the images.

C. Registration Algorithms

Multiple registration algorithms have been proposed in the literature with differing deformation field models, similarity measures, optimization strategies, and constraints. To show that the method we propose is not idiosyncratic to the registration algorithm and to test AQUIRC across a wide range of error, we test our method on five distinct, well-established, registration algorithms. The five algorithms are 1) Adaptive Bases Algorithm (ABA) [13], 2) Diffeomorphic Demons Algorithm (Diff. Demons) [14], 3) Fast Free Form Deformation (F3D) [37], 4) Symmetric Normalization (SyN) [15], and 5) Automatic Registration Tools (ART) [38]. This work is not, however, meant to be a comparison of registration algorithms, so we anonymize the registration algorithms and refer to them as A, B, C, D, and E.

If an algorithm does not calculate the inverse deformation field automatically, we calculate the inverse ourselves to satisfy the assumption we made for the undirected graph. To do this, we rely on the ITK filter `itkInvertDisplacementFieldImageFilter` which iteratively computes the inverse field from the forward field while minimizing the inverse consistency error [39]. As discussed above, because the goal of this work is not to compare registration algorithms performances, each algorithm is parameterized using their default settings and no attempt has been made to tune the parameters values for our data set. Before computing each non-rigid registration algorithm, a pre-processing affine registration is performed, registering every image in our data set to the same atlas image. The affine registration method uses NMI [40] as its similarity metric and the results are visually checked to assure accuracy.

D. Experiments

To compute the AQUIRC-derived quality measure, we register, for each set of atlas images, the 9 atlas images to each other in a pair-wise fashion. We also register the two sets of 9

atlas images to the 100 target images. For set (1), this is repeated for each of the 5 registration algorithms, while for set (2) we utilize only algorithm A. For each target image we compute AQUIRC with a network of size $N = 10$, composed of 9 atlas images and 1 target image. The registration quality is estimated at each voxel in the atlas images and this is repeated for each of the five registration algorithms and for each atlas set. This process is illustrated in Figure 5.

A potential use of the AQUIRC algorithm is to identify regions in an atlas that are, on average, difficult to register. This could be used, for instance, to detect abnormal or pathological images in large databases. In our data set there are 100 AQUIRC estimated registration error maps for each atlas and for each registration algorithm (one map per target image for each registration algorithm). Choosing two atlases, we average the 100 AQUIRC-estimated error maps for each atlas, and repeat this for each registration algorithm. We then visualize the resulting averaged maps of error.

To quantitatively evaluate our algorithm, we compare the ε values it computes to the Target Registration Error (TRE) at the AC and PC points. With a target point r (here r is either the AC or the PC), an atlas image A , and a target image B , we define Target Registration Error (TRE) as:

$$\|T_{AB}(r_A) - r_B\| = \text{TRE} \quad (4)$$

where T_{AB} is the transformation between image A and B that results from a registration algorithm. We calculate the TRE for each atlas and each target. This results in 9000 data points (9 atlases * 100 target images * 2 points * 5 algorithms). The local ε value that we compare to TRE is $\varepsilon_{AB}^T(r_A)$. This is the ε value at the position at which the AC or PC point in atlas A has been projected onto target image B . Trilinear interpolation is used to calculate the value at $\varepsilon_{AB}^T(r_A)$ when $T_{AB}(r_A)$ does not fall on a grid point. We choose $\varepsilon_{AB}^T(r_A)$ as the ε value to compare to TRE as this mimics the real life situation in which information is available in an atlas image but not in the target image. An alternative would be to choose the value ε^A to compare to TRE. We tested this location as well, and since the results were very similar to $\varepsilon_{AB}^T(r_A)$, we excluded the results of ε^A to reduce redundancy.

To compare the TRE and AQUIRC's estimated registration error we calculate their correlation and show their scatter plots. This is done for the AC and PC locations separately and together. We also present these results for each registration algorithm and for all five algorithms together.

To compare AQUIRC to a more traditional approach, we compute the LNCC in a region of interest between the target volumes and the deformed atlases. This was done for all algorithms using atlas set (1). The region of interest is centered at the location of $T_{AB}(r_A)$ for both the AC and PC. The region of interest was defined using three volume sizes: $4 \times 4 \times 4$, $8 \times 8 \times 8$, and $16 \times 16 \times 16$ mm. We then combine AQUIRC and LNCC to test whether or not there is a benefit in using intensity-based and transformation-based registration quality measures together.

To investigate how the choice of atlases affects the results of AQUIRC, we compare results obtained with atlas sets (1) and (2) on the 5 image volumes that are included in these two atlas sets, i.e., we compute the correlation between TRE and ε for the 100 target images and the 5 volumes for each atlas set. If the quality of the atlas set affects the AQUIRC algorithm, we would expect to find a lower correlation for atlas set (1) as it contains more images that are difficult to register.

To explore whether or not the stiffness of the deformations affects the results of AQUIRC, we vary the stiffness parameter λ utilized in algorithm A. A randomly chosen registration between two images is shown as an example in Figure 6 to illustrate the effect this parameter has on the deformation field. This figure shows, from left to right, a uniform grid that has been deformed by the transformation field obtained with algorithm A and stiffness values of $\lambda = 0.1, 0.15, 0.2, 0.25$, and 0.3 . As can be seen, the deformation becomes less stiff as the λ value is increased. In this experiment we use atlas set (1) and for each AC and PC

point, we compute the ratio $\frac{TRE}{\varepsilon}$. This experiment is performed to test whether AQUIRC is biased by the stiffness of the deformation field, i.e., whether or not for a fixed TRE, the value of ε changes with properties of the deformation field.

Finally, to show the practical utility of the AQUIRC algorithm we use it to reduce registration error at the AC and PC on atlas set (1) with two separate methods. The first method, termed “Best AQUIRC”, finds, for each target image, the atlas image that AQUIRC estimates as having the best registration quality (i.e. lowest ε value) out of the 9 possible atlases. We repeat this for each of the 100 target images and compute the mean and standard deviation of the TRE. We do this separately for the AC and PC because it is possible that the atlas selected for the AC is not the same as the atlas selected for the PC. The second method, termed “Weighted Average”, uses the local ε value as a weight in a weighted spatial averaging method. That is, for each target, we project the location of the AC and PC points from each of the 9 atlases onto the other 100 images and weigh the coordinates exponentially to estimate the correct AC and PC location on the target image. With $x(k)'$ the coordinates of points projected from atlas image k onto a target image (with $k = 1$ to 9 for each of the 9 atlases), we compute the weighted average of the coordinates, X_{wa} as:

$$w(k) = \frac{\exp\left(\frac{-\varepsilon_k^{x(k)'}}{\alpha}\right)}{\sum_{i=1}^9 \left(\exp\left(\frac{-\varepsilon_i^{x(i)'}}{\alpha}\right)\right)} \quad (5)$$

$$x_{wa} = \sum_{k=1}^9 w(k) * x'(k) \quad (6)$$

Where $\exp$ is the exponential function and the factor α is used to adjust the rate of exponential decay. We then compute the TRE between the weighted average of the coordinates and the actual known coordinates of the AC and PC in the target image. This is repeated for each of the 100 target images, and we compute the mean and standard deviation of the resulting TREs. To determine the α value in equations (5), we split our data set into a

training (10% of the data) and a testing (90% of the data) sets. We vary α from 0 to 1 for the training set and use the α factor that leads to the best results to evaluate our method on the testing set. As a comparison, we compute the mean and standard deviation of the errors obtained with a simple spatial averaging of the coordinates of the AC and PC points from the 9 atlases, which we refer to as the “Un-Weighted Average”. This is computed as in (6)

with each weight $w(k) = \frac{1}{9}$. We also compute the mean and standard deviation obtained when always choosing the best possible atlas, which is a lower bound only achievable if the ground truth is known. We refer to this as “Best Possible”.

III. Results

A. Qualitative Results

Figure 7 illustrates qualitatively the results produced by our algorithm. The left panels show the original images A and B. The next panels show the resulting deformed images obtained after registration, where the top panel shows image A registered to B and the bottom panel shows image B registered to A. The third set of panels show the AQUIRC-derived error map superimposed on the reformatted images, and lastly we show the AQUIRC-derived error map alone. Figure 7 is an example in which AQUIRC identified a large registration error in the ventricle region. The ventricles are close to being correctly registered in the original images but mis-registered after registration. The ventricle in the top image is compressed toward the center of the image and the ventricle in the bottom image is stretched toward the skull. As can be seen, the error map produced by AQUIRC is consistent with registration errors that can be observed visually both in the forward and reverse directions.

Figure 8 shows the average AQUIRC registration error maps of 2 atlases for each registration algorithm. Figure 8a is of an atlas of a patient with a large lesion while Figure 8b is of an atlas of a patient without a discernible feature that would cause registration error. With AQUIRC, it is possible to identify regions in each atlas which have, on average, a larger registration error. The first four registration algorithms show a much higher registration error in the atlas with the large lesion (Figure 8a) compared to the other example atlas (Figure 8b). Registering volumes with large space-occupying lesions is a notoriously difficult problem and can cause catastrophic registration errors that affect the entire image, which explains why mis-registered regions have been identified throughout the images. We note that the fifth panel (algorithm E) does not show as much error in the region around the lesion compared to the other four registration algorithms. A closer inspection reveals that algorithm E is less affected by the atlases with a lesion, so AQUIRC is correctly estimating that the level of error is lower for algorithm E.

B. Quantitative Results

Figure 9a shows a plot of true TRE vs. the AQUIRC ε values. In this plot, all the data points, i.e., the AC and PC points for all the atlases and all the registration algorithms have been pooled together. Results show that the ε values correlate well with the true registration error ($R^2 = 0.5418$). Figures 9b and 9c separate the results obtained for the AC and PC points. The correlation between ε values and TRE in each of these groups is comparable and reasonable

($R^2 = 0.5034$ for the AC points and $R^2 = 0.5823$ for the PC points). Figures 10a-e show the results for the pooled AC and PC points for each of the five registration algorithms. We use two colors in these plots. The red points correspond to atlases without discernible features. The blue points correspond to the four atlases considered to be challenging for registration algorithms. As expected, TRE values tend to be larger for these volumes than for the standard volumes. The correlation between ϵ and TRE values is good for algorithms A, C, and D ($R^2 = 0.62, 0.52$, and 0.642 , respectively). A weaker correlation is found for algorithms B and E ($R^2 = 0.31$ and 0.31 , respectively). A closer look at the figures shows that algorithms B and E overall have smaller TRE values. This figure suggests that the lower correlation is not a function of the registration algorithm itself but of the magnitude of the registration error produced by the two algorithms at the AC and PC locations. It suggests that either detecting smaller errors with AQUIRC is more difficult than detecting large ones, rater variability is affecting the results at small errors, or a combination of the two.

Figure 11 shows a scatter plot of true TRE versus LNCC value. We show the results for LNCC using a region of interest of $8 \times 8 \times 8$ as that volume resulted in the highest correlation. LNCC correlates to TRE with an $R^2 = 0.6289$ using atlas set (1) and all registration algorithms. This correlation is slightly higher than what was found using the AQUIRC algorithm. We then compute a multi-linear regression using both AQUIRC and LNCC on all 5 registration algorithms and on atlas set (1). We find an increase in correlation with an adjusted $R^2 = 0.776$. We do the same for each individual atlas and report the correlation values for AQUIRC, LNCC, and the multi-linear regression in Table 1. We find that the results of the multi-linear regression using AQUIRC and LNCC consistently improve the correlation with TRE for each registration algorithm.

The correlation using the low quality atlas set compared to the correlation using the high quality atlas set are similar, with the low quality atlas set performing slightly better ($R^2 = 0.206$ vs. $R^2 = 0.1921$) on the five normal images that are part of both atlas sets. The ratio between TRE and ϵ values was found to be insensitive to the stiffness of the deformation field. With $\lambda = 0.1, 0.15, 0.2, 0.25$, and 0.3 , the mean values for the ratios are $1.49, 1.49, 1.51, 1.52, 1.50$ respectively. Computing a paired t-test between the least rigid and most rigid deformation fields ($\lambda = 0.1$ and $\lambda = 0.3$), we find no significant difference between the means of the two sets of ratios ($p = .21$).

The effect of the α value on TRE at the AC and PC for the training set is shown in Figure 12. Based on these results, 0.25 was used as the α value to generate the results on the testing set. Figure 13 shows these results and illustrates the ability of the AQUIRC algorithm to improve registration accuracy when several atlases are available. In each of the panels (one per registration algorithm), we show the “Best Possible”, the “Best AQUIRC” values, the “Un-Weighted Average” values and the “Weighted Average” values where each method was computed on the testing data set.

Using a paired t-test we find that the Weighted Average technique significantly ($p < 0.05$) reduces the TRE for each registration algorithm compared to the Un-Weighted Average technique. While using the Best AQUIRC method, we find that the technique significantly reduces the TRE for all but four cases: two cases show a significant increase in TRE (the PC

for algorithm B and the AC for algorithm E), and two shows no significant difference compared to the Un-Weighted Average method (the PC for algorithm E and the AC for algorithm B). Figure 13a shows the TRE reduction that AQUIRC was able to achieve for each registration algorithm separately while Figure 13b shows the mean TRE reduction for all 5 registration algorithms combined. Using the Best AQUIRC method compared to Un-Weighted Average, the mean TRE is reduced on average by 21% and 19% at the AC and PC respectively, while using the Weighted Average method reduces the mean TRE on average by 32% and 33% at the AC and PC respectively.

IV. Discussion

This article presents AQUIRC, a method that uses redundancy from multiple registration circuits to estimate the local quality of a non-rigid registration and is independent of the registration algorithm used. The results we have obtained show that this idea has merit. It is simple to implement and relatively cheap to execute. The pair-wise registrations between atlases and all registration circuits involving only the atlases can be pre-computed. The

matrix $\bar{P}$ can also be pre-computed as well as its pseudo-inverse. To compute AQUIRC for a new image volume $\bar{E}_c$ needs to be computed for each circuit in which the new image is involved. When this is done solving for the ε values requires a matrix multiplication, i.e.,

$$\left(\begin{matrix} \bar{P}^T & \bar{P}^T \end{matrix} \right)^{-1} \bar{P}^T \log \left(\bar{E}_c \right). \text{ The size of } \bar{P} \text{ is } \left(\frac{N}{3} \right) \times \left(\frac{N}{2} \right) \text{ and is thus determined by the}$$

number of images used in the network while the size of $\bar{E}_c$ is determined by the number of circuits (number of rows) in the network and voxels (number of columns) in the images. The time needed to compute the transformations between a new volume and the other volumes used as atlases depends on the registration algorithm and grows linearly with the number of

images in the network. Computing the circuit error, $\bar{E}_c$, for the circuits in which the new volume is involved requires combining deformation fields across circuits. Timing depends on the size of the memory that is available, i.e., how many deformation fields can be loaded in memory. In our current implementation, without any pre-computation, it requires ~45 minutes to run when programmed in Matlab. For a new image volume, after pre-computing $\bar{P}$, its pseudo-inverse, and the registration circuits that involve only the atlas images, it requires ~20 minutes to compute for an $N = 10$. In this work $N = 10$, thus creating the matrix $\bar{P}$ and computing its pseudo-inverse requires a negligible amount of memory, i.e. less than 1MB. With larger network sizes it may become an issue but the benefit of using larger network sizes is not apparent.

When using our method to reduce registration errors, we show a larger improvement with the Weighted Average method than with the Best AQUIRC method. In the field of atlas selection, a similar phenomenon has been observed, i.e., that a weighted combination of atlases generally outperforms choosing a single ‘Best’ atlas [41].

As shown in Figure 10, the correlation between the ε values and true registration error is weaker for algorithms B and E than it is for the other algorithms. Algorithms B and E have smaller registration errors. For algorithm B the average TRE error for the set of AC and PC points is 1.56mm; it is 0.84mm for algorithm E. These errors are thus on the order of one voxel and AQUIRC was only able to reduce error by 15% for algorithm B and 18% for algorithm E, lower than the average of 33% reduction for all algorithms combined. This indicates that it is difficult for AQUIRC to detect errors of that magnitude. As discussed earlier, no attempt was made to optimize the parameters used by the registration algorithm and they were only evaluated quantitatively at two points that were chosen because of the accuracy with which they can be manually localized. It is possible that AQUIRC would perform differently at other locations. We will be testing AQUIRC in different regions and in different conditions in future work.

To compare our approach with a more traditional intensity-based approach, we utilized the similarity measure LNCC. The correlation between LNCC and TRE is slightly higher than the correlation between ε and TRE. We note, however, that the AC and PC points at which we evaluate our algorithm are well contrasted. The atlas and test images are also both MR images. This, as opposed to a situation in which the target would be chosen in a uniform region, is arguably one the scenarios for which an intensity-based method would provide the strongest results. Because AQUIRC only relies on the deformation fields, it is less susceptible to intensity-related image properties. We thus suspect that AQUIRC would be at an advantage in regions of low contrast but testing this quantitatively is difficult on real images. This is because selecting points in low contrasted regions is difficult to do with high accuracy. One important finding, however, is that intensity-based quality measures and AQUIRC provide complementary information. Indeed, we have shown that combining these measures leads to stronger correlations than using either of them independently.

We found that the quality of the images in the atlas set did not affect the correlation in the test we have performed. The correlation that was achieved for the five normal atlas volumes common to the two sets were $R^2 = 0.206$ vs. $R^2 = 0.1921$ for the set with the difficult-to-register atlases and without, respectively. These results indicate that adding “bad” atlases in the atlas set does not affect the results, at least with the number of atlases we have used. It is possible that with fewer atlases than 9, the number of “bad” atlases would have affected the results. The low correlation values reflect the fact that the normal atlases register well to the target volumes, because, in this case, we compute the correlation on only normal atlases the range of TRE values is smaller. As discussed earlier, lower TRE value ranges also result in a lower correlation between AQUIRC and TRE.

Our results also did not show that AQUIRC was biased by the stiffness of the deformation fields. One potential use of the AQUIRC algorithm is to optimize the parameters of a registration algorithm without using a ground truth. This could be done by computing a registration algorithm, estimating the local registration error, and varying the parameters of the registration algorithm in an attempt to lower the AQUIRC registration error value. This would not be possible, however, if the value of ε was a function of the deformation field, i.e., if for a fixed TRE, ε was changing with the property of the fields. Future testing will determine the full applicability of applying AQUIRC to optimizing registration algorithm

parameters. As noted earlier, AQUIRC is not able to identify the registration error of trivial correspondences (i.e. the identity or transformation obtained with very high stiffness values). In this work, we assume that there are no trivial correspondences. In practice, the inclusion of a similarity measure to check for trivial correspondences may be necessary.

We note that methods inspired by AQUIRC have already been used by others with success. Donoghue et al. [42] proposed a graph-based method to register knee images in a large data base. They use a method similar to AQUIRC to estimate the quality of the registration between nodes in the graph and show that it leads to lower registration error. Goksel et al [30] developed a variant of the AQUIRC algorithm proposed in [26] for global atlas selection that slightly increased the correlation and reduced the memory and speed needed for computation compared to AQUIRC in the context of a global non-rigid registration estimation when labels are present. This variant utilized more registration circuits than we had proposed originally and used a reflective Newton trust-region method to solve the linear least squares solution compared to the pseudo inverse that is used in this work. Goksel showed that the results of AQUIRC and their variant of AQUIRC correlate highly with DICE. They tested two data sets and show AQUIRC correlates to DICE with an $R^2 = 0.7903$ and $R^2 = 0.7744$, while their variant correlates to DICE with an $R^2 = 0.865$ and $R^2 = 0.812$.

The circuit multiplicative error model we currently use is relatively simple and, although we have experimented with an additive model, other, more complex models may lead to further improvement. A method that uses multiple loops in a network of images as we have done and a more complex error model was recently introduced by Gass in [43]. These authors utilize information about the deformations to update the error model in their equivalent to $\bar{P}$ in our algorithm. Although this is a more accurate model of registration error, it requires orders of magnitude more memory to compute. As discussed earlier, the $\bar{P}$ matrix in our work only depends on the network of images, i.e., the number of nodes and the number of circuits; their equivalent to our $\bar{P}$ matrix in [43] is a sparse matrix of size

$\sim \left(\frac{N}{3}\right) * Q \times \left(\frac{N}{2}\right) * Q$ with Q the number of voxels in the image. Thus the more in realistic error model in [43] is difficult to scale up and to apply at the voxel level to 3D images. One instance in which AQUIRC would not be able to identify registration error is if the error is biased, for example if the error is the same across all registrations into one image. Every registration circuit would not be affected by the biased error as the error is added, then subtracted from the combination of the transformations. We do not believe error bias is likely to appear in practice, but its impact is under continuing evaluation.

Potential applications for AQUIRC include comparing registration algorithms, determining error in regions where determining ground truth is difficult such as regions of low contrast, providing a level of confidence to applications in which accuracy is of critical importance, or tuning parameters in registration algorithms. Muenzing in [8] proposed a framework to reduce registration error in which any method that estimates registration quality can be applied. It would be of interest to investigate the usefulness of the local registration quality

measure AQUIRC provides for this framework either in isolation or when combined with others.

Acknowledgments

This work partially supported by NSF grant DMS-0334769 and NIH Grant No. R01EB006136. An early version of this work was presented at WBIR 2012 in Nashville TN, USA.

References

- [1]. Fitzpatrick JM, West JB, Maurer CR Jr. Predicting error in rigid-body point-based registration. *IEEE Transactions on Medical Imaging*. 1998; 17(5):694–702. [PubMed: 9874293]
- [2]. Simpson IJ, Schnabel JA, Groves AR, Andersson JL, Woolrich MW. Probabilistic inference of regularisation in non-rigid registration.”. *NeuroImage*. 2012; 59(3):2438–2451. [PubMed: 21939772]
- [3]. Iglesias JE, Sabuncu MR, van Leemput K. Improved Inference in Bayesian Segmentation Using Monte Carlo Sampling: Application to Hippocampal Subfield Volumetry. *Medical Image Analysis*. 2013; 17(7):766–778. [PubMed: 23773521]
- [4]. Risholm P, Janoos F, Norton I, Golby AJ, Wells WM III. Bayesian characterization of uncertainty in intra-subject non-rigid registration. *Medical image analysis*. 2013; 17(5):538–555. [PubMed: 23602919]
- [5]. Risholm, P.; Ross, J.; Washko, GR.; Wells, WM. *Information Processing in Medical Imaging*. Springer Berlin Heidelberg; 2011. Probabilistic elastography: estimating lung elasticity; p. 699-710.
- [6]. Kybic J. Bootstrap resampling for image registration uncertainty estimation without ground truth. *IEEE Transactions on Image Processing*. 2010; 19(1):64–73. [PubMed: 19709978]
- [7]. Muenzing SE, van Ginneken B, Murphy K, Pluim JP. Supervised quality assessment of medical image registration: Application to intra-patient CT lung registration. *Medical Image Analysis*. 2012; 16(8):1521–1531. [PubMed: 22981428]
- [8]. Muenzing SE, van Ginneken B, Viergever AM, Pluim JP. DIRBoost-An algorithm for boosting deformable image registration: Application to lung CT intra-subject registration. *Medical Image Analysis*. 2014; 18(3):449–459. [PubMed: 24556079]
- [9]. Lotfi T, Tang L, Andrews S, Hamarneh G. Improving Probabilistic Image Registration via Reinforcement Learning and Uncertainty Evaluation. *Machine Learning in Medical Imaging*. 2013:187–194.
- [10]. Christensen GE, Johnson HJ. Consistent image registration. *IEEE Transactions on Medical Imaging*. 2001; 20(7):568–582. [PubMed: 11465464]
- [11]. Woods RP, Grafton ST, Holmes CJ, Cherry SR, Mazziotta JC. Automated image registration: I. General methods and intrasubject, intramodality validation. *J. Comput. Assist. Tomogr*. 1998; 22:139–152. [PubMed: 9448779]
- [12]. Holden M, Hill DLG, Denton ERE, Jarosz JM, Cox TCS, Rohlfing T, Goodey J, Hawkes DJ. Voxel similarity measures for 3D serial MR brain image registration. *IEEE Transactions on Medical Imaging*. 2000; 19(2):94–102. [PubMed: 10784281]
- [13]. Rohde GK, Aldroubi A, Dawant BM. The adaptive bases algorithm for intensity-based nonrigid image registration. *IEEE Transactions on Medical Imaging*. 2003; 22(11):1470–1479. [PubMed: 14606680]
- [14]. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: efficient non-parametric image registration. *NeuroImage*. 2009; 45:S61–S72. [PubMed: 19041946]
- [15]. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*. 2008; 12(1):26–41. [PubMed: 17659998]
- [16]. Geng, X.; Kumar, D.; Christensen, GE. Transitive inverse-consistent manifold registration. In: Christensen, GE.; Sonka, M., editors. *Information Processing in Medical Imaging*. Vol. 19. Springer-Verlag; Berlin: 2005. p. 468-479.

- [17]. Rao A, Chandrashekara R, Sanchez-Ortiz GI, Mohiaddin R, Aljabar P, Hajnal JV, Puri BK, Rueckert D. Spatial transformation of motion and deformation fields using nonrigid registration. *IEEE Transactions on Medical Imaging*. 2004; 23(9):1065–1076. [PubMed: 15377115]
- [18]. Beg MF, Miller MI, Trounevé A, Younes L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int J Comput Vis*. 2005; 61(2):139–157.
- [19]. Ashburner J. A fast diffeomorphic image registration algorithm. *NeuroImage*. 2007; 38(1):95–113. [PubMed: 17761438]
- [20]. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: efficient non-parametric image registration. *NeuroImage*. 2009; 45(1):61–72.
- [21]. Yeo BTT, Sabuncu MR, Vercauteren T, Ayache N, Fischl B, Golland P. Spherical demons: fast diffeomorphic landmark-free surface registration. *IEEE Transactions on Medical Imaging*. 2010; 29(3):650–668. [PubMed: 19709963]
- [22]. Auzias G, Colliot O, Glaune's JA, Perrot M, Mangin J-F, Trounevé A, Baillet S. Diffeomorphic brain registration under exhaustive sulcal constraints. *IEEE Transactions on Medical Imaging*. 2011; 30(6):1214–1227. [PubMed: 21278014]
- [23]. Leung KK, Ridgway GR, Ourselin S, Fox NC. Consistent multi-time-point brain atrophy estimation from the boundary shift integral. *NeuroImage*. 2012; 59(4):3995–4005. [PubMed: 22056457]
- [24]. Freeborough PA, Woods RP, Fox NC. Accurate registration of serial 3D MR brain images and its application to visualizing change in neurodegenerative disorders. *J. Comput. Assist. Tomogr*. 1996; 20:1012–1022. [PubMed: 8933812]
- [25]. Christensen GE, Johnson HJ. Invertibility and transitivity analysis for nonrigid image registration. *Journal of Electronic Imaging*. 2003; 12(1):106–117.
- [26]. Datteri, RD.; Dawant, BM. Estimation of Rigid-Body Registration Quality Using Registration Networks. In: Dawant, BM.; Haynor, DR., editors. *SPIE Medical Imaging 2011: Image Processing*. 2011.
- [27]. Datteri, RD.; Asman, AJ.; Landman, BA.; Dawant, BM. In: Landman, BA.; Warfield, S., editors. *Estimation of Registration Quality Applied to Multi-Atlas Segmentation; MICCAI Workshop on Multi-Atlas Labeling and Statistical Fusion*; 2011; p. 78-87.
- [28]. Datteri, RD.; Dawant, BM. *Biomedical Image Registration*. Vol. 7359. LNCS; 2012. Automatic Detection of the Magnitude and Spatial Location of Error in Non-rigid Registration; p. 21-30.
- [29]. Datteri, RD.; Asman, AJ.; Landman, BA.; Dawant, BM. Applying the Algorithm “Assessing Quality Using Image Registration Circuits” (AQUIRC) to Multi-Atlas Segmentation. In: Ourselin, S.; Styner, A., editors. *SPIE Medical Imaging 2014: Image Processing*. 2014.
- [30]. Goksel, O.; Gass, T.; Vishnevsky, V.; Székely, G.. Estimation of Atlas-Based Segmentation Outcome: Leveraging Information From Unsegmented Images; *Biomedical Imaging (ISBI), IEEE 10th International Symposium on*; 2013; p. 1215-1218.
- [31]. Asman AJ, Landman BA. Non-Local Statistical Label Fusion for Multi-Atlas Segmentation. *Medical Image Analysis*. 2013; 17(2):194–208. [PubMed: 23265798]
- [32]. Brown LG. A survey of image registration techniques. *ACM Computing Surveys (CSUR)*. 1992; 24(4):325–376.
- [33]. Fitzpatrick, JM. Detecting failure, assessing success. In: Hajnal, JV.; Hill, DLG.; Hawkes, DJ., editors. *Medical Image Registration*. CRC Press; Baton Rouge, Florida: 2001. p. 117-139.
- [34]. Christensen, GE. Transitivity Error [Online]. Non-Rigid Image Registration Evaluation Project (NIREP); 2006. Available: Web site: <http://www.nirep.org/te>
- [35]. Pallavaram S, Yu H, Spooner J, D'Haese P-F, Bodenheimer B, Konrad PE, Dawant BM. Inter-surgeon variability in the selection of anterior and posterior commissures and its potential effects on target localization. *Stereotact Funct Neurosurg*. 2008; 86(2):113–119. [PubMed: 18270482]
- [36]. Benabid AL. Deep brain stimulation for Parkinson's disease. *Curr Opin Neurobiol*. 2003; 13:696–706. [PubMed: 14662371]
- [37]. Modat M, Ridgway GR, Taylor ZA, Lehmann M, Barnes J, Fox NC, Hawkes DJ, Ourselin S. Fast free-form deformation using graphics processing units. *Comput. Methods Programs Biomed*. 2009; 98(3):278–284. [PubMed: 19818524]

- [38]. Ardekani BA, Guckemus S, Bachman A, Hoptman MJ, Wojtaszek M, Nierenberg J. Quantitative comparison of algorithms for inter-subject registration of 3D volumetric brain MRI scans. *Journal of Neuro. Methods.* 2005; 142:67–76.
- [39]. Insight Segmentation and Registration Toolkit. <http://www.itk.org/http://www.itk.org/>
- [40]. Studholme C, Hill DLG, Hawkes DJ. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition.* 1999; 32(1):71–86.
- [41]. Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *Neuroimage.* 2009; 46(3):726–738. [PubMed: 19245840]
- [42]. Donoghue, CR.; Rao, A.; Pizarro, L.; Bull, AMJ.; Rueckert, D. Fast and accurate global geodesic registrations using knee MRI from the Osteoarthritis Initiative; *IEEE Computer Vision and Pattern Recognition Workshops (CVPRW)*; 2012; p. 16-21.
- [43]. Gass, T.; Szekely, G.; Goksel, O. Detection and Correction of Inconsistency-based Errors in Non-Rigid Registration. In: Ourselin, S.; Styner, A., editors. *SPIE Medical Imaging 2014: Image Processing.* 2014.

Figure 1.

An example of one registration circuit, with the transformations between each image represented by T_{AB} , T_{BC} and T_{CA} . The distance between the red X and X' represents the registration circuit consistency error.

Figure 2.
Example of a complete graph with one circuit shown with red arrows

Figure 3.

Diagram of a graph with 4 nodes and 4 voxels in each node. The ε values represent the estimation of the quality of registration between each voxel. The 4 voxels are specified on the bottom grid. As discussed in the text, in this work the values are estimated on a regular grid but computing the consistency error E_C across a circuit involves composing transformations, which is done with tri-linear interpolation.

Figure 4.

An example image of a patient with bilateral electrode implant (left), an example image without a discernible feature that would cause registration error (middle), and an example image of a patient with a large lesion (right).

Figure 5.

Illustration of the process used to calculate the AQUIRC-derived ε values. Nine atlases are registered pair-wise together and to each of the target images. AQUIRC is run for each target image, with a network of 10 images. This is repeated 5 times for each target image, once for each registration algorithm.

Figure 6.

Example deformations that result from changing the stiffness parameter of algorithm A.
From left to right: $\lambda = 0.1, 0.15, 0.2, 0.25$, and 0.3

Figure 7.

An example of AQUIRC's registration quality map, with error mapped on both the forward and reverse transformations.

Figure 8.

Average AQUIRC error maps from the 100 images in our dataset, using an atlas with a large lesion (a), and an atlas without any discernible features that might cause a registration problem (b). The error maps are averaged for each registration algorithm, which are, from left to right, algorithms A, B, C, D, and E. The error map is shown on top and the atlas image with an overlay of the color map is shown on the bottom. The ε values in the AQUIRC estimation error map vary from 3.5 to 0.

Figure 9.

Scatter plot of the 9000 points using atlas set (1), using all 9 atlases and all five registration algorithms for (a) both the AC and PC, (b) AC only and (c) PC only.

Figure 10.

Scatter plot of the AC and PC points using algorithm A, B, C, D, and E on atlas set (1). The red square points are the atlases without any discernible feature that would cause registration error and the blue circles are the difficult to register atlases.

Figure 11.
Scatter plot of LNCC compared to TRE for both the AC and PC.

Figure 12.

Mean TRE at the AC and PC as the weighting term is varied from 0.01 to 1.

Figure 13.

The left graph shows the reduction in registration error when using AQUIRC for each registration algorithm A, B, C, D, and E. The right graph shows the reduction of registration error when using AQUIRC for all registration algorithms combined. From left to right, this plot shows the Best Possible, the Best AQUIRC, the Weighted Average, and the Un-Weighted Average results.

Table 1

R^2 values for the correlation between TRE and the results of AQUIRC, LNCC and a multi-linear regression using AQUIRC and LNCC for each registration algorithm.

	A	B	C	D	E
AQUIRC	0.62	0.31	0.52	0.64	0.31
LNCC	0.58	0.49	0.67	0.72	0.43
Multi-Linear Regression	0.79	0.57	0.75	0.82	0.52