ASSESSING REGISTRATION QUALITY VIA REGISTRATION CIRCUITS

By

Ryan Datteri

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Computer Science

December, 2014

Nashville, Tennessee

Approved:

Benoit M. Dawant, Ph.D.

Bennett A. Landman, Ph.D.

Michael J. Fitzpatrick, Ph.D.

Michael I. Miga, Ph.D.

Robert E. Bodenheimer, Ph.D.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to first acknowledge my family for always being there and for their support throughout the trials and tribulations of graduate school. I also want to acknowledge my friends and colleagues at Vanderbilt for their help and support, which kept me mostly sane.

I want to thank Dr. Dawant for accepting me as a research assistant into his lab during my tenure at Vanderbilt University. His direction and guidance have vastly improved the quality of this work. Dr. Dawant deserves enormous credit for the work I have completed at Vanderbilt University and I thank him for his advice, support, and the large amount of time and effort he spent working with me.

I want to also thank Dr. Fitzpatrick for his support throughout my dissertation and specifically for the many hours he spent in one on one lessons which were invaluable to this thesis. I thank him for his support and his encouragement to pursue new ideas.

I want to thank Dr. Landman, who, during an image processing class, brought to my attention the research problem that eventually became the focus of this thesis. I often used Dr. Landman as a sounding board and he provided vital insights.

I want to also thank Dr. Miga and Dr. Bodenheimer, whose input early in the development of this work helped to clarify and orient its direction.

This work would not have been possible without the support provided by NSF grant DMS-0334769 and NIH Grant No. R01EB006193.

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER 1

Introduction

In the field of image registration there is a pressing need for techniques that can assess the quality of a registration. Currently there is no widespread method to determine error in an image registration, a weakness of registration that adds a level of uncertainty to its results. While there do exist analytical solutions for error in fiducial-based registration, they do not allow for the estimation of error in individual cases. This dissertation presents the algorithm Assessing Quality Using Image Registration Circuits (AQUIRC), which estimates the quality of intensity-based and fiducial-based registration.

Image registration is used extensively in the medical imaging community for a variety of applications including, but not limited to, atlas-based segmentation, large scale probabilistic imaging studies, rigid registration of same-subject images acquired at different times, rigid registration of same-subject images acquired with different modalities, non-rigid registration of intra-subject images acquired at different times to track anatomical changes, and inter-subject non-rigid registration used to provide anatomical and functional information acquired from previous subjects. Fiducial-based registration is often used in the context of image-guided surgery (IGS) and is a critical component in determining the accuracy of an IGS procedure.

There are two main contributions to the field of image registration presented in this dissertation. First, it defines an algorithm that is the first to use the idea of redundancy in sets of registration circuits with multiple registrations which is then used to estimate the accuracy of a single registration. Second, the work described in this dissertation has inspired new work and new methods. The new work that has been inspired combined with the work described in this dissertation may expand the knowledge of registration accuracy.

In this dissertation a validation of the AQUIRC algorithm method is also presented. AQUIRC has been tested across multiple applications: fiducial-based registration (Chapter 4), affine and rigid image registration (Chapter 5), global non-rigid registration (Chapter 6), global atlas selection (Chapter 6), local non-rigid registration (Chapter 7, 8) and local atlas selection (Chapter 9). The results are analyzed using simulated and manually defined ground truth, and AQUIRC's applicability to each situation is discussed.

CHAPTER 2

Background

Introduction

In this chapter, a description of intensity-based and fiducial-based registration and their applications is presented. Then, the problem of registration error is discussed. Following this, background information is presented for related work in image registration error estimation and past work that led to the AQUIRC algorithm. Finally, an overview of this dissertation is given.

Intensity-Based and Fiducial-Based Registration

Generally, registration algorithms attempt to find a transformation that optimizes a similarity metric between two sets of data which, for this paper, consists of either sets of intensities or sets of points. Various registration algorithms differ by the similarity measure, the transformation model, the interpolation method, and the optimization method that are used. A review of the variety of techniques used in medical image registration for rigid-body registrations can be found in [1], a review of non-rigid registration transformation models can be found in [2], a general review of image registration can be found in [3] , and a review of medical image registration can be found in [4]. This dissertation considers fiducial-based rigid registration and intensity-based rigid, affine, and non-rigid registration. Here, fiducial-based rigid registration is defined as the registration between two point sets while intensity-based registration is defined as the registration of two images based on their intensities. Rigid registration restricts the transformation between the two images to be rigid, affine registration permits skew and scale as well as a rigid transformation, while non-rigid registration allows for local deformations.

Fiducial-based registration is an important technique in Image Guided Surgery (IGS) as it is often utilized to align image information to the surgical space in an operating room. In this

context, fiducial markers are attached to the patient and an image is acquired. The physical location of the fiducial markers in the operating room is obtained as well as the location of the markers in the image and the two point sets are registered to each other.

Atlas-based segmentation is a use of image registration that is particularly relevant to this dissertation because it is particularly well suited as an application of AQUIRC. In a clinical setting, atlas-based segmentation allows for the accumulation of information that can be costly or dangerous to acquire to be developed for an atlas. This information can then be cheaply and safely projected onto a patient's image. For example, stimulation information from past Deep Brain Stimulation cases can be dangerous to acquire, as the information is collected in surgery with the patient awake on the table. Thus, the ability to project past information on a new patient is beneficial to the surgeon and can be done safely. Another example use of atlas-based segmentation can be found in Intensity Modulated Radiation Therapy (IMRT), which is an important technique to treat head and neck cancer. IMRT requires the segmentation of structures within a Computed Tomography (CT) scan which, when done manually by a physician, is a time consuming and expensive process. Atlas-based segmentation has been proposed to tackle this problem [5, 6]. Here, structures of interest, e.g. the optic nerves and brainstem, are delineated in an atlas volume, the atlas is registered to a new target image, and the structures of interest are projected onto the new volume. Thus, accuracy in atlas-based segmentation is important and is a recent active theme of investigation. One popular method to increase the accuracy of atlas-based segmentation is multi-atlas segmentation. This concept requires the registration of multiple atlases and then the selection or combination of segmentation results provided by each atlas. Multi-atlas segmentation has been shown to increase the accuracy of segmentations, and can reduce the effect of individual catastrophic registration failure on the final segmentation [7].

There have been multiple proposed methods to combine the segmentations that result from multiple registrations. These methods use information from a variety of sources, including intensity-based similarity measures which are calculated between the volumes after registration (e.g., mutual information, correlation coefficient, etc.), properties of the deformation fields that register the images [8, 9, 5, 10, 11, 12, 13, 14], or statistical information gained from the results of multiple atlas-based segmentation [15, 16, 17, 18, 19, 20, 21, 22]. As the quality of a registration affects the results of an atlas-based segmentation, a method to estimate the quality of a registration could be used in a multi-atlas context.

Registration Error and Why it is Important

Error occurs in fiducial-based registration when localizing the fiducials. This error is called Fiducial Localization Error (FLE) [23]. FLE in turn causes error in the registration between the image and surgical space. There exists analytical solutions to estimate this error [23], however, it has been shown that these solutions do not allow for the estimation of error in individual cases [24]. Currently, Fiducial Registration Error (FRE) is often used as a surrogate for Target Registration Error (TRE) which is the quantity of clinical interest [23]. Unfortunately, it has been shown [25] that TRE and FRE are uncorrelated. FRE is thus an unreliable predictor of registration accuracy.

Error occurs in intensity-based registration (rigid, affine, and non-rigid) for multiple reasons. First, the search space is often too large to compute all possible deformations so there is no guarantee of finding a globally optimal correspondence. In fact, non-rigid registration has been shown to be NP-complete [26]. Similarity measures are not guaranteed to result in the correct correspondence. Thus even if the optimization algorithm is able to identify the globally optimal value for a similarity measure this may not result in a perfect correspondence between

images. Also, medical images are discrete representations of a continuous space. Thus discretization error can occur during acquisition and processing, affecting registration. A perfect correspondence may not even be possible because of imaging artifacts, acquisition noise, or anatomical differences, all of which can cause error in registration.

An argument can be made that a technique to estimate registration error is not needed after a registration algorithm has been tested and validated. There are many methods to validate a registration algorithm and, once validated, a registration algorithm could be assumed to produce registrations that fall within the range of quality identified during the validation study, with no verification methods applied when using the registration algorithm in practice. Recent studies have emerged, however, [27, 28] that indicate that many metrics that quantify the quality of a registration algorithm are not accurate. Thus, utilizing standards such as the Dice metric across large structures, inverse consistency, and image similarity measures may not result in a clear understanding of registration error. Even when a particular registration algorithm has been thoroughly tested and a reasonable distribution of the quality of possible registrations is available, catastrophic failure for individual registrations is still possible. Also, the quality of a registration may be affected by the quality of the image acquisition or the anatomical differences between patients, errors that are difficult to quantify when validating a registration algorithm.

Detection of error in non-rigid registration has been reported in various studies as a difficult and important problem [29, 30]. Intensity-based image registration algorithms, which are the most commonly used methods for non-rigid registration tasks, depend on similarity measures to estimate how well two images match. However, similarity measures are unable to determine the quality of the registration results based on image intensity alone. The problem of correspondence error and its potential negative effect on various studies that use intensity-based

non-rigid registration is described in [30]. The paper was published in 2003 and, to the best of our knowledge, no method has yet to be developed in response to the problem they pose, namely that "most widely used methods are essentially dumb in that, for a particular registration task, they report only a measure of image similarity which does not allow a judgment of 'success' or 'failure' to be made. Worse the magnitude and spatial distribution of errors in NRR [Non-Rigid Registration] are unknown".

Related Work in Registration Error Estimation

Although analytical solutions have been derived to estimate error for the point-based rigid-body registration problem [23], no such solution exists for the non-rigid case or even for the rigid case when transformations are not estimated with homologous points. In a well-known comparative study which establishes Mutual Information (MI) [31] and Normalized Mutual Information (NMI) [32] as the most robust and accurate image similarity measures for rigid registration, West et al. [29] recommend to visually check registration results to identify instances when registration fails. This, however, is not feasible in large registration studies where thousands of images are registered together. It is also difficult to identify small rotational or translational error visually for rigid and affine registration [33, 34]. 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 manual intervention or pre-labeled atlases. These solutions generally fall into two categories: Bayesian methods [35, 36, 37, 38, 39] and supervised learning techniques [40, 41, 42].

Bayesian methods require the estimation of posterior distributions, which is done either with Markov Chain Monte Carlo methods [36, 37, 38], a computationally demanding approach, or mean-field variational Bayesian techniques [35]. The method by Simpson et al. [35]

calculates the uncertainty in a non-rigid registration and then regularizes the deformation field by the uncertainty level. The deformation field is regularized to be smoother when there is higher uncertainty, and conversely less smooth when there is lower uncertainty. The method is novel; however, it has not been applied directly to estimate registration error. Iglesias et al. [36] developed an uncertainty measure that they use to add error bars to the volume estimates that result from atlas-based segmentations. This method uses intensity information and information from the deformation field to determine segmentation uncertainty. This method also requires manual label segmentations on the atlas images, requires training data, and is computationally complex. The method shows interesting results for quantifying segmentation uncertainty. However, it was tested on its ability to classify Alzheimer's disease, not directly for the application of registration error detection. Risholm et al. [37, 38] developed a method to provide a level of uncertainty to intra-subject non-rigid registration. They utilize information from an intensity-based similarity measure as well as information from the regularization of the deformation field to create a probability density function. This provides a map of uncertainty that was applied to estimate uncertainty in tumor resection in [37] and estimating lung elasticity in [38]. This method was applied only to intra-patient registrations and is computationally intensive. Kybic [39] proposed a bootstrap resampling technique to estimate the variability of an estimated transformation. Evaluation of this technique was limited to 2D images and simple transformations.

Supervised learning techniques [40, 41, 42] utilize image features or manually labeled points to identify registration uncertainty. Generally, these methods require a training set for each new application and data set to which it is to be applied. Muenzing et al. developed a method to assess quality for intra-patient CT lung registration in [40] and validated their work

on three different registration algorithms in [41]. In [40] they developed a method that identifies image features at distinct landmarks, and use these features in a supervised learning technique to classify registration errors. This algorithm was also applied on intra-patient non-rigid registration and requires a training set for every new application. In [41] Muenzing developed a framework to boost a registration algorithm's performance by focusing on regions that are identified as having registration error. This framework was tested using the error detection method developed in [40], in which Muenzing et al. were able to reduce error by 5-34% using various data sets and registration algorithms in an intra-patient lung CT data set. Lofti et al. [42] also used a supervised learning method, including spatial information from segmented labels into a similarity measure. Lofti et al. [32] use both intensity-based and deformation-based features together to develop their uncertainty measure. They visually show that their uncertainty measure overlaps with regions that are difficult to register due to tumors in the images, and reported an increase in correlation between estimated and known registration errors in a simulated case, with an $R^2 = 0.29$.

One method that cannot be categorized as either a supervised learning technique or Bayesian technique was developed by Glatard et al. [43]. These authors proposed a method that defines a bronze standard ground truth using multiple registration algorithms to calculate the accuracy of each individual registration algorithm in a leave-one-out framework. Requiring multiple registration algorithms is not always feasible and the method was tested without any ground truth, making it difficult to determine the accuracy of the method and its applicability.

Past Work Related to AQUIRC

Verification of individual registrations is difficult and a number of techniques have thus 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 [44, 45, 46]. Assuming two images A and B and two transformations T_{AB} and T_{BA} that map image A onto B and image B onto A, inverse consistency implies that the composition of these transformations, denoted as $T_{\mathrm{BA}} \circ T_{\mathrm{AB}}$, is the identity. When the transformations can be expressed in matrix form, e.g., a 4X4 homogenous matrix, this simply means that $T_{\mathrm{AB}} = T_{\mathrm{BA}}^{-1}$. In 1995, Christensen and colleagues [47, 48] proposed an algorithm that computes the two transformations simultaneously and forces them to be inverses of each other. Inverse consistency was proposed by Thirion as early as 1996 [49]. In that article, and building on the work of Burr [50], inverse consistency was imposed in the Demon's algorithm [51] by computing T_{AB} and T_{BA} separately, composing the transformations, and distributing the residual error at each iteration. Since then, a number of algorithms that enforce inverse consistency have been proposed. A representative set of such algorithms can be found in [42-54]. [52, 53, 54, 55, 56, 57, 58, 59, 60, 61] [62, 63, 64]

The consistency of three transformations was used by Holden et al.. [46], Freeborough et al. [65], Woods et al. [45], and Christensen et al. [44] to compare and evaluate several similarity measures and registration algorithms. This technique, which we refer to as a registration circuit, involves three images A, B, and C and three transformations $T_{\mathrm{AB}}, T_{\mathrm{BC}}, T_{\mathrm{CA}}$. Composing the three transformations maps the coordinate system of image A onto itself. 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 $\mathrm{x}' = T_{\mathrm{CA}}(T_{\mathrm{BC}}(T_{\mathrm{AB}}(\mathrm{x})))$. 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 [44] to evaluate non-rigid registration algorithms. In more recent work [55], the same group proposed a registration algorithm called TICMR for Transitive Inverse Consistent Manifold Registration (TICMR) that jointly estimates a correspondence between three manifolds (curves in 2D, surfaces in 3D) while minimizing inverse consistency and transitivity errors. While, undoubtedly, transitivity consistency is a necessary condition for error-free registration around a registration circuit, it is not sufficient. As discussed by Christensen et al. [44], 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.

Dissertation Overview

This dissertation discusses the work published in [66, 67, 68, 69, 70, 71]. Chapter 3 describes the general AQUIRC algorithm. Chapter 4 applies the AQUIRC algorithm to fiducial-based rigid registration [66], Chapter 5 applies AQUIRC to intensity-based rigid registration [67], Chapter 6 applies AQUIRC to global non-rigid registration and global atlas selection [68], Chapter 7 applies AQUIRC to estimate local non-rigid registration error in a simulated case [69], Chapter 8 applies AQUIRC to estimate local non-rigid regsitration error with manually defined ground truth [70], and Chapter 9 applies AQUIRC to local atlas selection [71]. Chapter 10 concludes

the dissertation by discussing the contributions made in this thesis, and describing methods that have been developed based on the ideas presented herein.

CHAPTER 3

Methods

General AQUIRC Algorithm

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 nodes A, B, and C and three transformations T_{AB}, T_{BC}, and T_{CA}. For this work, each node represents either an image or a set of fiducial markers. An example registration circuit is shown in Figure 1.

In this dissertation we mathematically define a registration circuit using two separate models of error: an additive consistency error measure and a multiplicative consistency error measure. The multiplicative registration circuit model is defined as: $E_C=\varepsilon_A*\varepsilon_B*\varepsilon_C$, while the additive model is defined as: $E_C=\varepsilon_A+\varepsilon_B+\varepsilon_C$. We refer to E_C 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. The multiplicative model can be solved linearly by computing the logarithm of the equation, i.e. $\log(E_C)=\log(\varepsilon_A)+\log(\varepsilon_B)+\log(\varepsilon_C)$. These are the quantities of interest. In the following Chapters, each data set and application has been tested with both the additive and multiplicative error models and when the differences are unsubstantial; only one model is discussed and shown.

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 node through the transformations relating the nodes across a circuit. Referring to Figure 1, the traditional registration circuit (TRC)

is defined as combining transformations in order, starting on node A and applying to X the transformation from A to B, followed by the transformation from B to C, and finally the transformation from C to A. In our notation, we define TRC as $X' = T_{CA}(T_{BC}(T_{AB}(X)))$ [2, 45], i.e. we apply the operators in this expression from right to left. Another method to compute a registration circuit that we refer to as a non-traditional registration circuit (NTRC), would be to combine the transformations out of order, for example applying the transformation from A to B, followed by the transformation from C to A, followed by the transformation from B to C. Thus the NTRC is computed as $X' = T_{BC}(T_{CA}(T_{AB}(X)))$. The reasoning for the use of the non-traditional registration circuit is given in Chapter 4, and the NTRC method is used for the rigid registration applications in Chapters 4 and 5. The E_C value here is defined as the dissimilarity(X, X′). The dissimilarity function can be tailored to the application, and this dissertation presents multiple methods to define the function. Generally we define E_C either globally or locally.

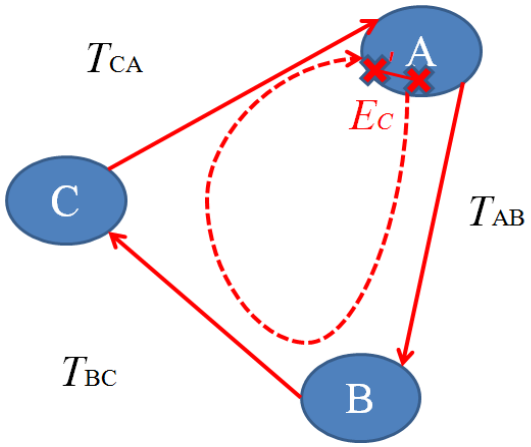


Figure 1: An example of one registration circuit, with the transformations between each node represented by $\boldsymbol{T_{AB}}$, $\boldsymbol{T_{BC}}$ and $\boldsymbol{T_{CA}}$. The distance between the red X and X' represents the registration circuit consistency error.

The variable E_C is affected by error in three registrations: the registration error between A and B, between B and C, and between C and A. As discussed earlier and also noted by Fitzpatrick [72], 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 upon the idea of registration circuits to multiple circuits. To do this, we start with a set of N nodes and compute pair-wise registrations between all nodes 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 [73]. However, in [73], 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. If our data set contains N nodes, 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.

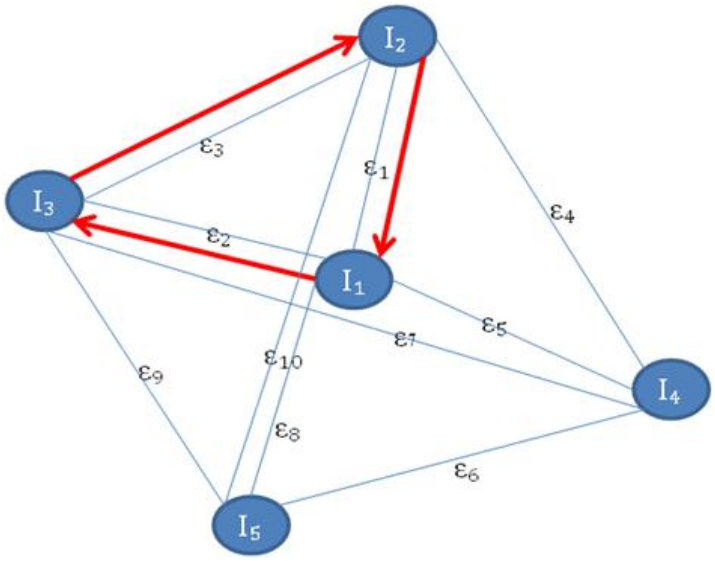


Figure 2: Example of a complete graph with one circuit shown with red arrows

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 over a local region, e.g. at each voxel in an image. In this dissertation, all local estimates are done at the voxel level. To estimate the registration quality globally, a single value is computed for E_C. To

estimate the registration quality locally the consistency error E_C is defined at each local region 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. $\overline{E}_C = \left[E_C^1, E_C^2, \ldots, E_C^Q\right]$ is a rasterized vector of consistency errors across a registration circuit, with Q equal to the number of voxels. Thus, the estimation of registration quality between two nodes is either a global estimation, where a single ε value is defined as the registration quality, or a local estimation (e.g. at each voxel) where we define vector of registration quality $\overline{\varepsilon} = [\varepsilon^1, \varepsilon^2, \ldots, \varepsilon^Q]$, in which ε^i represents the registration quality at voxel i and $\overline{\varepsilon}$ represents a rasterized vector of registration quality values.

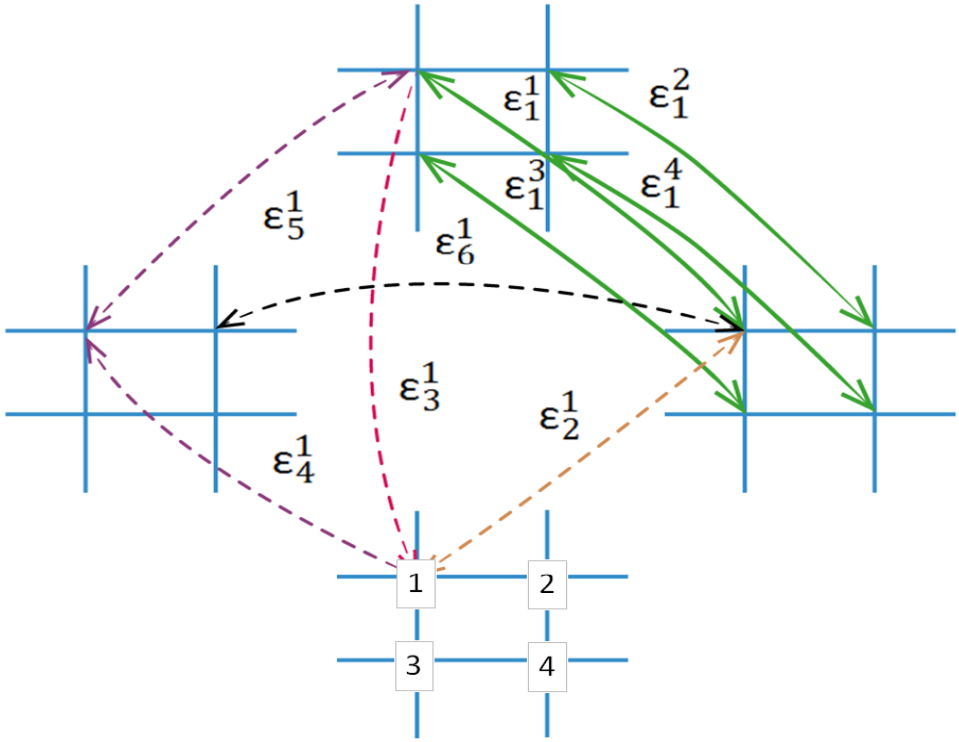


Figure 3: Diagram of a graph with 4 nodes and 4 voxels in each node. The ε values represent the estimation of the quality of registration between each voxel. The 4 voxels are specified on the bottom grid.

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 $N \geq 5$ is required to compute the linear least squares solution used in this work, however, an $N = 4$ is used here to illustrate the process. In this example, there are 4 possible registration circuits. Thus we construct a system of equation using the additive error model as:

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

where for E_{C_p}, p refers to a particular circuit and for ε_l, l, refers to an edge, i.e. a link between two images. This can be rewritten in matrix form as:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix} = \begin{bmatrix} E_{C_1} \\ E_{C_2} \\ E_{C_3} \\ E_{C_4} \end{bmatrix} \quad (2)$$

This formulation estimates a global registration quality. It is also possible to instead estimate a local registration quality for each of the voxels in the nodes. Then the same system of equations can be created for each of the 4 voxels to produce (3), where for $E_{C_p}^i$ and ε_l^i, i represents a particular voxel. This can be written in matrix form as:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} [\varepsilon_1^1, \varepsilon_1^2, \varepsilon_1^3, \varepsilon_1^4] \\ [\varepsilon_2^1, \varepsilon_2^2, \varepsilon_2^3, \varepsilon_2^4] \\ [\varepsilon_3^1, \varepsilon_3^2, \varepsilon_3^3, \varepsilon_3^4] \\ [\varepsilon_4^1, \varepsilon_4^2, \varepsilon_4^3, \varepsilon_4^4] \\ [\varepsilon_5^1, \varepsilon_5^2, \varepsilon_5^3, \varepsilon_5^4] \\ [\varepsilon_6^1, \varepsilon_6^2, \varepsilon_6^3, \varepsilon_6^4] \end{bmatrix} = \begin{bmatrix} [E_{C_1}^1, E_{C_1}^2, E_{C_1}^3, E_{C_1}^4] \\ [E_{C_2}^1, E_{C_2}^2, E_{C_2}^3, E_{C_2}^4] \\ [E_{C_3}^1, E_{C_3}^2, E_{C_3}^3, E_{C_3}^4] \\ [E_{C_4}^1, E_{C_4}^2, E_{C_4}^3, E_{C_4}^4] \end{bmatrix} \quad (3)$$

The global estimation in equation (2) can be re-written in a more compact form as $\overline{\overline{P}}\overline{\varepsilon} = \overline{E}_c$ while the local estimation in equation (3) can be re-written as $\overline{\overline{P}}\overline{\overline{\varepsilon}} = \overline{\overline{E}}_c$. Note the following derivations solve for the local error estimation; however, it is equivalent to finding the global estimation

except $\bar{\bar{\varepsilon}}$ and $\bar{\bar{E}}_c$ are matrices instead of vectors. Here, $\bar{\bar{P}}$ is a matrix that represents the edges used for each circuit, with the value of $\bar{\bar{P}}$ set to 1 if the edge is utilized in the circuit and 0 otherwise. In the matrix $\bar{\bar{P}}$, each row represents a registration circuit while each column represents individual edges in the network of images. We can solve for $\bar{\bar{\varepsilon}}$ using a linear least squares solution as:

$$\bar{\bar{\varepsilon}} = \left(\bar{\bar{P}}^T \bar{\bar{P}}\right)^{-1} \bar{\bar{P}}^T \bar{\bar{E}}_c \qquad (4)$$

If we utilize the multiplicative error model, equation (3) can be re-written as $\bar{\bar{P}} \log(\bar{\bar{\varepsilon}}) = \log(\bar{\bar{E}}_c)$, and because of the multiplicative assumption, $\log(\bar{\bar{\varepsilon}})$ can also be solved for using a linear least squares solution as:

$$\log(\bar{\bar{\varepsilon}}) = \left(\bar{\bar{P}}^T \bar{\bar{P}}\right)^{-1} \bar{\bar{P}}^T \log(\bar{\bar{E}}_c) \qquad (5)$$

and solving for $\bar{\bar{\varepsilon}}$:

$$\bar{\bar{\varepsilon}} = e^{\log(\bar{\bar{\varepsilon}})} \qquad (6)$$

Equation (3) can be defined more generally as:

$$\begin{bmatrix} 1\,1\,1\,0\,..\,0 \\ 1\,0\,1\,1\,..\,0 \\ 1\,1\,0\,1\,..\,0 \\ 0\,1\,1\,1\,..\,0 \\ . \\ . \\ . \end{bmatrix} \begin{bmatrix} \left[\varepsilon_1^1, \varepsilon_1^2, \ldots, \varepsilon_1^Q\right] \\ \left[\varepsilon_2^1, \varepsilon_2^2, \ldots, \varepsilon_2^Q\right] \\ \left[\varepsilon_3^1, \varepsilon_3^2, \ldots, \varepsilon_3^Q\right] \\ . \\ . \\ . \\ \left[\varepsilon_{\binom{N}{2}}^1, \varepsilon_{\binom{N}{2}}^2, \ldots, \varepsilon_{\binom{N}{2}}^Q\right] \end{bmatrix} = \begin{bmatrix} \left[E_{C_1}^1, E_{C_1}^2, \ldots, E_{C_1}^Q\right] \\ \left[E_{C_2}^1, E_{C_2}^3, \ldots, E_{C_2}^Q\right] \\ \left[E_{C_3}^1, E_{C_3}^2, \ldots, E_{C_3}^Q\right] \\ . \\ . \\ . \\ \left[E_{C_{\binom{N}{3}}}^1, E_{C_{\binom{N}{3}}}^2, \ldots, E_{C_{\binom{N}{3}}}^Q\right] \end{bmatrix} \qquad (7)$$

in which E_{Cp}^i is defined as the $dissimilarity(X_i, X_i')$ around circuit p at voxel i and ε_l^i is defined as the quality of the registration l at voxel i. Pseudo-code illustrating the algorithm (with '//' representing a comment describing the algorithm) is as follows:

Define Ind to be the indicator function
Define P to be a matrix of 0's of size $\binom{N}{3}$x$\binom{N}{2}$
RegistrationCircuit = 1;
For each unique combination of three images, call them A B and C
{
// Compute the Registration Circuit, and rasterize the results
$\bar{\bar{E}}_c$(RegistrationCircuit, 1 to Q)= $T_{BC}\left(T_{CA}\left(T_{AB}(X)\right)\right)$
//Compute the *P* Matrix
$\bar{\bar{P}}$ (RegistrationCircuit, ind(A)) = 1;
$\bar{\bar{P}}$ (RegistrationCircuit, ind(B)) = 1;
$\bar{\bar{P}}$ (RegistrationCircuit, ind(C)) = 1;
RegistrationCircuit ++;
}
//Compute the matrix inversion and multiplication for the linear least square
$\bar{\bar{\varepsilon}} = \left(\bar{\bar{P}}^T\bar{\bar{P}}\right)^{-1}\bar{\bar{P}}^T\bar{\bar{E}}_c$

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. One benefit of using the multiplicative model is that the estimation of registration error calculated is always positive. While the additive model can result in negative numbers, it was not observed to occur in any of the Experiments presented in this dissertation. After calculating E_C for the multiplicative model, 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 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 on to A. We can do this because either the registration algorithms we utilize are inverse consistent by design or we force inverse consistency after the registration is performed. Early in this work, a directed graph

version was tested but this was not pursued because an improvement in the results was not observed and a directed graph adds computational complexity and requires more memory.

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, yet this would require a reformulation of (3) that could not be expressed as compactly as it is now (e.g. as was done in [74]). Experimentally $\overline{\overline{P}}$ 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 with differing lengths but this has not yet been explored.

Similarity and Validation Measures

Multiple registration algorithms are used in this work. The most common algorithm used in this dissertation is the Adaptive Bases Algorithm (ABA) [52], which uses Normalized Mutual Information (NMI) [32] as its similarity measure. Briefly, ABA computes a deformation field that is modeled as a linear combination of radial basis functions with finite support. This results in a transformation with thousands of degrees of freedom. Two transformations (one from the atlas to the subject and the other from the subject to the atlas) are computed simultaneously and constrained to be inverses of each other. Other registration algorithms are used in Chapter 8; they will discussed in that chapter.

When combining multiple non-rigid transformations, as is done when computing a registration circuit, tri-linear interpolation is used. One common measure of overlap used in multiple chapters of this dissertation is that of DICE [75]. DICE is a measure of overlap of two volumes, defined here as where X and X' are the two volumes in question.

$$\mathrm{DICE} = \frac{|X \cap X'|}{\frac{1}{2}(|X| + |X'|)} \qquad (8)$$

DICE is often used as a surrogate for registration error, where, for example, X is a manually defined segmentation of a structure on a target image and X' is a segmentation that has been projected from an atlas image onto the target image utilizing the results of a registration algorithm. If there is a large overlap between X and X', the registration algorithm is assumed to have been accurate, though small registration errors may not be detected [27, 28].

CHAPTER 4

Estimation of Fiducial-Based Registration Error

Introduction

Fiducial-based registration is often utilized in image guided surgery because of its simplicity and speed. The assessment of target registration error when using this technique, however, is difficult. Although the distribution of the target registration error can be estimated given the fiducial configuration and an estimation of the fiducial localization error, the target registration error for a specific registration is uncorrelated with the fiducial registration error. Fiducial registration error is thus an unreliable predictor of the target registration error for a particular case. In this chapter, a case is made for the Non-Traditional Registration Circuit. Then, AQUIRC is applied to estimate the quality of a fiducial-based registration. The measure resulting from AQUIRC is shown to correlate with the target registration error. Then, this value is used to reduce target registration error.

A Case for the Non-Traditional Registration Circuit

In fiducial-based rigid registration, there exists an analytical expression for target registration error [24]. Fiducial-based registration finds the best possible fit between two sets of points by a rigid rotation and translation and is generally used to register some target point in the image, usually not at a fiducial marker. The error at this target point is called Target Registration Error (TRE).

Error occurs in fiducial-based registration because of error in identifying fiducial points, called Fiducial Localization Error (FLE). In application, it is not possible to measure the error at the target point, yet it is possible to measure the amount of error at the fiducial markers. We can define this error as the root-mean square of the distance between the two sets of points after

registration, called Fiducial Registration Error (FRE). Mathematically the expected value of TRE and FRE (where$\langle . \rangle$ means "expected value of", and boldface denotes vector displacement) have been defined as $\langle |\mathbf{TRE}(x)|^2 \rangle = \frac{1}{N}\left(1 + \frac{1}{3}\sum_{k=1}^{3}\frac{d_k^2}{f_k^2}\right)\langle \mathrm{FLE}^2 \rangle$ and as $\langle \mathrm{FRE}^2 \rangle = \left(1 - \frac{2}{N}\right)\langle \mathrm{FLE}^2 \rangle$, where N is the number of points, x is the target location, d_k is the distance of x from the principle axis k of the fiducial configuration, f_k^2 is the mean of the squared distances of the fiducials from that axis [24]. Because we can measure FRE in application, intuitively it has been mistakenly assumed that FRE and TRE are correlated, and therefore that when the measured FRE^2 is lower than $\langle \mathrm{FRE}^2 \rangle$ that $|\mathbf{TRE}(x)|^2$ will also be lower then $\langle |\mathbf{TRE}(x)|^2 \rangle$. Fitzpatrick however, has shown analytically and through simulation that TRE and FRE are uncorrelated [24].

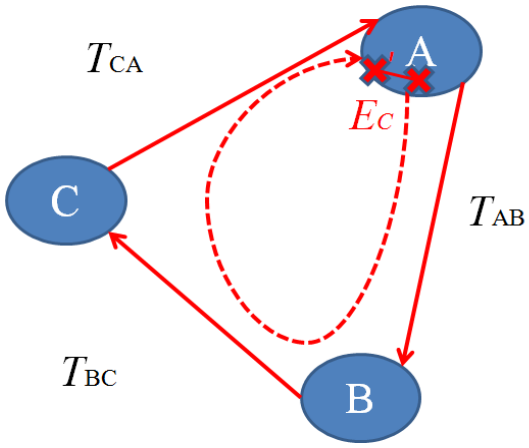


Figure 4: An example of one registration circuit, with the transformations between each node represented by $\boldsymbol{T_{AB}}$, $\boldsymbol{T_{BC}}$ and $\boldsymbol{T_{CA}}$. The distance between the red X and X' represents the $\mathbf{TRE_C}$

Applying the Traditional Registration Circuit (TRC) to fiducial-based registration, Fitzpatrick [76] has recently developed an analytical expression for **TRE** around the TRC as $\mathbf{TRE}_\mathrm{C} = R_\mathrm{p}^{(N_c)} r + t_\mathrm{p}^{(N_c)} - r$. Here, r is a single true target, and $R_\mathrm{p}^{(N_c)}$, $t_\mathrm{p}^{(N_c)}$ are the combined rotation and translation around a TRC. Fitzpatrick generalized $\mathbf{TRE}_\mathrm{C}$ for registration circuits of size $N > 2$, however, in this work we utilize only registration circuits of size $N = 3$. Thus, for

registration circuits of size 3 the $\mathbf{TRE}_C$ defined in [75] is equivalent to $X' - X$, where X is the single true target referred to as r by Fitzpatrick, and $X' = T_{CA}(T_{BC}(T_{AB}(X)))$ (as seen in Figure 4). Fitzpatrick then showed mathematically that $\mathbf{TRE}_C$ for rigid fiducial-based registration is zero to first order in FLE. The proof technique of [75] relies on the fact that the transformations in a traditional registration circuit are ordered. However, if we change the ordering of the transformations in the registration circuit, e.g. $X' = T_{AB}\left(T_{BC}(T_{CA}(X))\right)$, we find that the proof is no longer applicable, suggesting that the first order TRE is no longer zero. Furthermore simulations confirm that $\mathbf{TRE}_C$ has second order form for TRCs and has linear form for NTRCs, supporting the conclusion that, in order to measure a $\mathbf{TRE}_C$ with a strong signal-to-noise ratio (i.e., having a first-order component), it is necessary to utilize a non-traditional ordering. This is our motivation for computing the NTRC in rigid registration. Thus, the NTRC is tested for both fiducial-based registration in this chapter, and intensity-based registration in the following chapter. We note that there are multiple ways to compute an NTRC for three images, however in this work we assume that each NTRC formulation for three specific images are equivalent, although this is not necessarily true.

Applied AQUIRC

In this chapter, we define $E_C = \text{dissimilarity}\left(X, X'\right)$, where X is the target point and X' is the target point transformed around a registration circuit. The error across a circuit, $\text{dissimilarity}(X, X')$, is then defined as the Euclidean distance between X and X'.

In this chapter we utilize a standard fiducial registration method that uses singular-value decomposition as proposed by Golub et al. [76]. Two spaces are defined: the image space and the surgical space. This mimics a typical point-based registration problem where pre-operative images need to be registered to the patient in the operating room. The target point X is selected

in the image space. As discussed below, we introduce a known transformation between the image and surgical space that can be large and we perturb the position of the fiducials in image space. A diagram describing this can be seen in Figure 5. To select a point that is consistent across circuits we fix the position of the target point X in image space and we always start our networks from a configuration in image space.

In this work we also separate the ε values that evaluate the registrations into two classes: image space to image space and image space to surgical space. The ε values calculated to evaluate the registrations from image space to image space are referred to as interior ε values and the ε values calculated to evaluate the registrations from image space to surgical space are referred to as exterior ε values. The blue edges in Figure 5 correspond to the interior ε values and the red edges correspond to the exterior ε values.

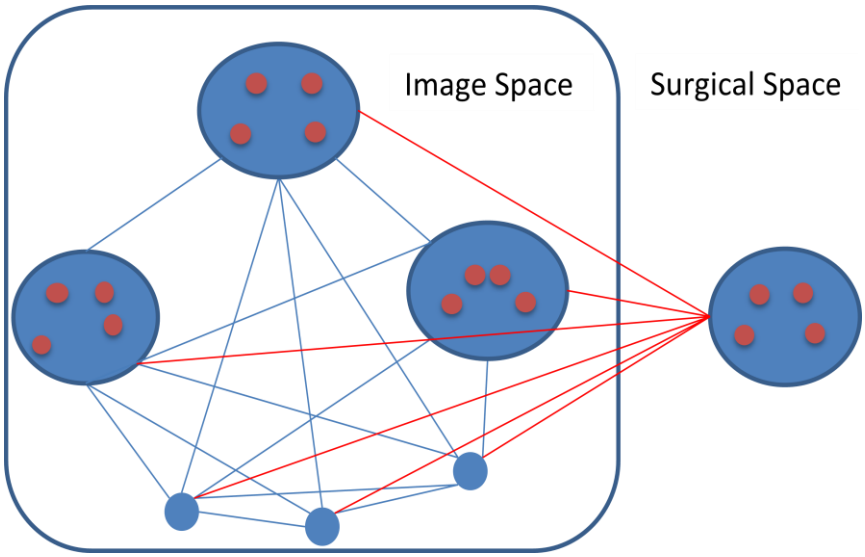


Figure 5: Diagram of the experiment methodology. The image space contains N-1 fiducial configurations that are created by adding FLE to the original fiducial locations. The surgical space contains the rotated and translated fiducial configuration. Each set of fiducial configurations are registered to every other configuration, both in image space and in surgical space. The red links represent the exterior ε values while the blue links represent the interior ε values.

Experiments

To test our method we perform 6 experiments based on the experiment performed by Fitzpatrick which showed analytically and experimentally that FRE and TRE are uncorrelated [25]. Using the same experiment, we show that our quality measure is correlated to the TRE. In [25], Fitzpatrick simulates an actual Deep Brain Stimulation case with 4 fiducials and a target location in the deep brain. The location of the four fiducials x_1, x_2, x_3, and x_4 as well as the target position x_t are as follows in image coordinates:

$$x_1 = \begin{bmatrix}197\\217\\115\end{bmatrix}, \quad x_2 = \begin{bmatrix}109\\225\\121\end{bmatrix}, \quad x_3 = \begin{bmatrix}83\\139\\127\end{bmatrix},$$
$$x_4 = \begin{bmatrix}202\\132\\130\end{bmatrix}, \quad x_t = \begin{bmatrix}144\\155\\57\end{bmatrix}$$

In experiments 1 and 2, we simplify the problem by utilizing just the y and z dimensions, thus point $x_1 = \begin{bmatrix}217\\115\end{bmatrix}$, $x_2 = \begin{bmatrix}225\\121\end{bmatrix}$, $x_3 = \begin{bmatrix}139\\127\end{bmatrix}$, $x_4 = \begin{bmatrix}132\\130\end{bmatrix}$, and $x_t = \begin{bmatrix}155\\57\end{bmatrix}$. For experiments 3-6 we use all three dimensions as was done in [5]. We apply a rotation R and translation t to the location of the fiducials x_1, x_2, x_3, and x_4 as well as to the target position x_t, which results in the corresponding positions y_1, y_2, y_3, y_4, and y_t. We consider the fiducials x_1, x_2, x_3, and x_4 and x_t to be in the image space while the rotated and translated fiducials y_1, y_2, y_3, y_4, and y_t are considered to be in the surgical space, i.e., the physical space in the operating room. In [25], the rotation R is set to be 10, 20, and -30 degrees about the x, y, and z axes and the translation t to be (7, -10, 100) mm (which was chosen as an arbitrary mis-registration in [25]). In experiments 1 and 2, the rotation R is set to 10 degrees about the x axis and the translation t is set to be (7, -10), while in experiments 3-6 R and t are set to the same values as in [25].

For all experiments we perturb the location of the fiducials in image space using a fiducial localization error drawn from a random distribution with zero mean and variance of FLE/3, where FLE is set to 1 mm unless otherwise stated. This simulates the fiducial localization error one would make by localizing the fiducials manually in the images. This is done N-1 times to create a set of N-1 perturbed fiducial configurations (as seen in the left of Figure 5). We demean each fiducial configuration so that the centroid of each perturbed configuration is at the origin. This was done to make the quality of the AQUIRC results consistent as we found that AQUIRC's performance changed depending on the amount of translation between the fiducial configurations. We compute all pair-wise fiducial-based registrations between each of the $\frac{N(N-1)}{2}$ pairs of imagescreating the complete graph of registrations necessary to create the matrices $\overline{\overline{P}}$ and $\overline{E}_C$. We then apply the AQUIRC algorithm and calculate the ε value for each of the registrations.

There are three quantities of importance in this work: TRE, FRE, and the ε value. We compute these for each registration. For experiments 1 and 2 we consider both the interior and exterior ε values for illustration purposes, but in our analysis of experiments 3-6 we consider only the exterior ε values since these evaluate the registrations of interest in our application. This results in N-1 values for each of the three quantities because our simulations involve N fiducial configurations (N-1 in image space and one in surgical space).

In practice, TRE is not known but in our simulations we can compute it because we know the location of the target in both the image space and surgical space as well as the location of the fiducial markers. The ε value is calculated by the AQUIRC algorithm. We repeat this process in a Monte Carlo simulation over thousands of iterations.

For experiments 1 and 2 we use the additive model while for experiments 3-6, we utilize both the additive and multiplicative error-propagation models. All six experiments follow the paradigm used in [25]. In experiment 1 and 2 we use a small network (N = 5) as well as restrict the problem to two dimensions (x and y) for simplicity. The code for this experiment is included in the Appendix of this dissertation. We calculate the TRE and exterior ε values as well as the TRE and the interior ε values across the network 250 times. This results in 1000 points for the exterior ε values and 1500 points for the interior ε values. We then calculate the correlation between the TRE values and both the interior and exterior ε values (separately). For experiment 1 we utilize the TRC method, while in experiment 2 we utilize the NTRC method.

For experiments 3-6 we utilize the NTRC method and include 3-dimensions. The third experiment is performed using an N = 40, i.e., we create a network of 39 configurations in the image space and one in the surgical space, and we repeat the experiment 5000 times. We calculate the correlation between the TRE and FRE values and between the TRE and ε values. In the fourth experiment we test the effect N has on our results by varying N from 5 to 20, using 5000 simulations for each N. Again, we calculate the correlation between TRE and the ε values for each N.

In the fifth experiment, we test the effect FLE has on our results. We vary FLE from 1 to 70 mm using 1000 simulations for each FLE value with an N equal to 20.

Finally, in the sixth experiment, we test the ability of our algorithm to improve the TRE in a fiducial-based registration scenario. To do this we utilize an N = 40 with 5000 simulations and an FLE of 1 mm. For each simulation and for all 39 external registrations we calculate the mean TRE, the min TRE, and the max TRE. We also compute the TRE of the fiducial configuration that the AQUIRC algorithm identifies as being of the highest quality using both the

multiplicative model and with the additive model, and the TRE of the fiducial configuration with the minimum FRE.

Results

The results of experiment 1 are shown in Figure 6 in two scatter plots. The points in red represent a scatter plot between TRE and the exterior ε value and the points in blue represent the scatter plot between the TRE and the interior ε value. In this experiment the TRC method was used to compute the circuit registration error. We did not find a statistically significant correlation.

The results of experiment 2 are shown in Figure 7 as two scatter plots again. The points in red represent a scatter plot between TRE and the exterior ε value and the points in blue represent the scatter plot between the TRE and the interior ε value. These ε values were calculated using the NTRC method. In this case we observe a correlation with TRE between both the exterior ε values (r = 0.4692) and the interior ε values (r = 0.2724) both of which are statistically significant.

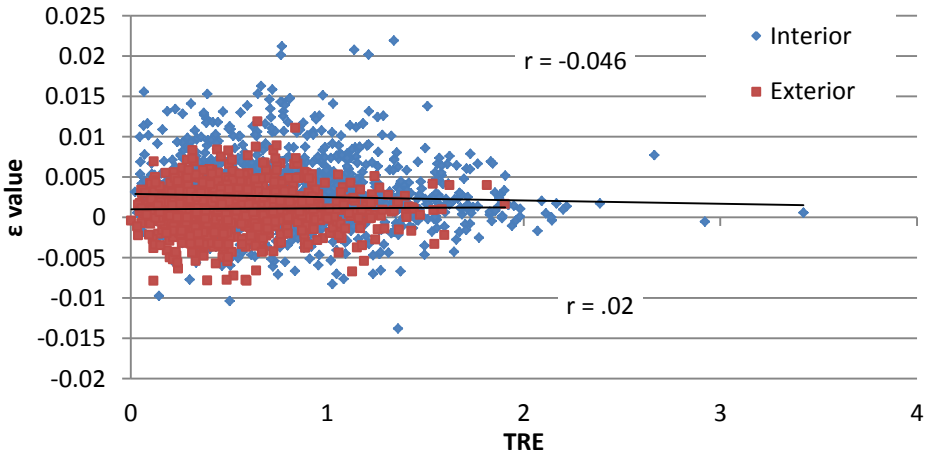


Figure 6: Scatter plot of the TRE and interior ε values in blue, and a scatter plot of the TRE and exterior ε values in red. The ε values were calculated utilizing the TRC method.

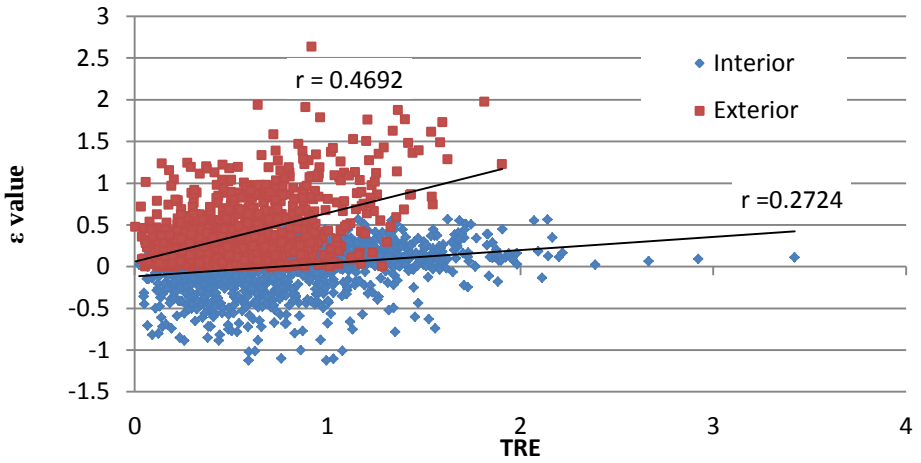


Figure 7: Scatter plot of the TRE and interior ε values in blue, and a scatter plot of the TRE and exterior ε values in red. The ε values were calculated utilizing the NTRC method.

The results of experiment 3 are shown in Figure 8, Figure 9, and Figure 10. In Figure 8 we show a scatter plot of the FRE and TRE values (the points in all three scatter plots were reduced to a random sampling of 1000 data points for better visualization). As can be seen, we produce a correlation that is very similar to the one found in [25], (r = 0.0016), which is not statistically significant (p = 0.4895).

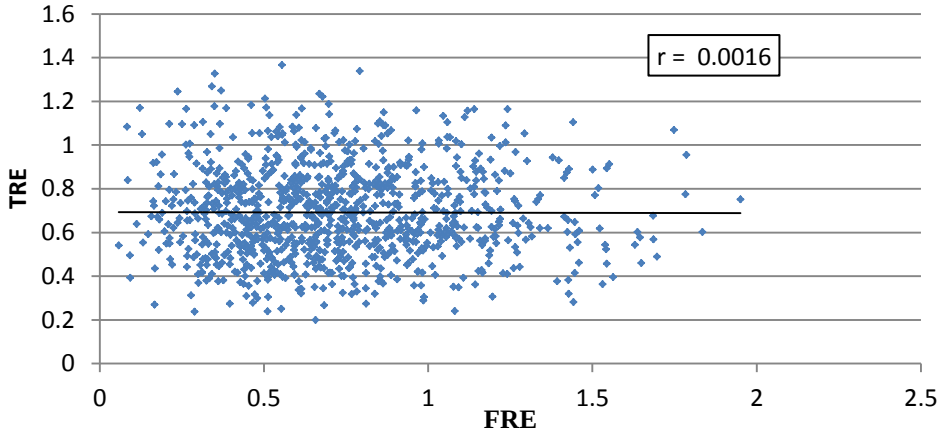


Figure 8: Scatter plot of the TRE and FRE values.

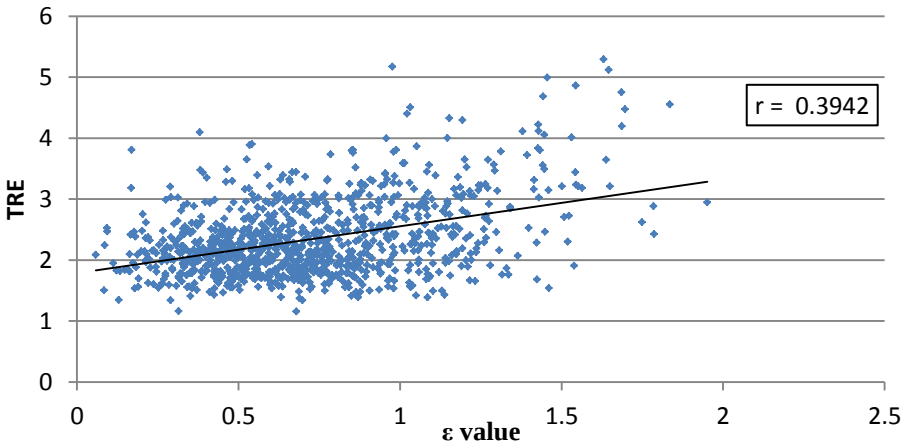


Figure 9: Scatter plot of the TRE and ε values when utilizing the multiplicative model.

In Figure 9 we show the correlation between AQUIRC's ε value and TRE when using the multiplicative model. In this case there is a statistically significant correlation ($r = 0.3916$, $p < 0.001$). In Figure 10 we show the correlation between AQUIRC's ε value and TRE when using the additive model where we find a statistically significant correlation ($r = 0.4294$, $p < 0.001$). The results of experiment 4 are shown in Figure 11. When the number of fiducial configurations that we utilize in the image space is increased, we find a larger correlation between TRE and ε, producing better estimations of the quality of error in fiducial registrations. The increase in correlation is consistent for both the additive and multiplicative model, with a plateau around $N = 20$.

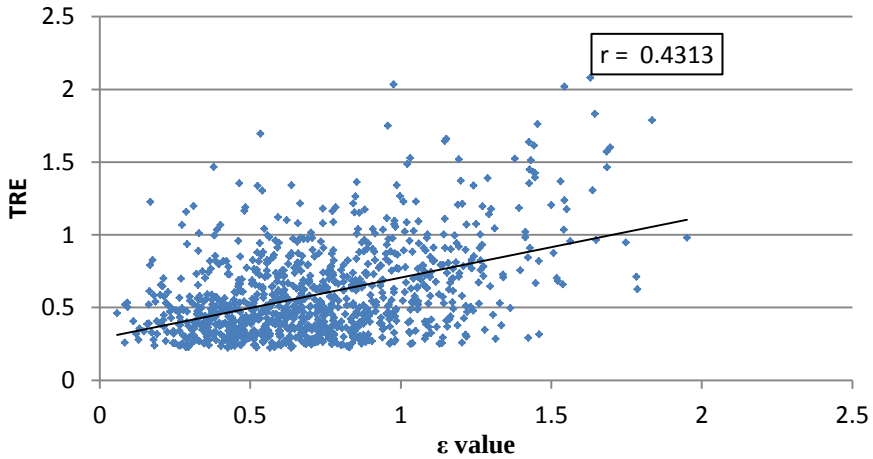


Figure 10: Scatter plot of the TRE and ε values when utilizing the additive model.

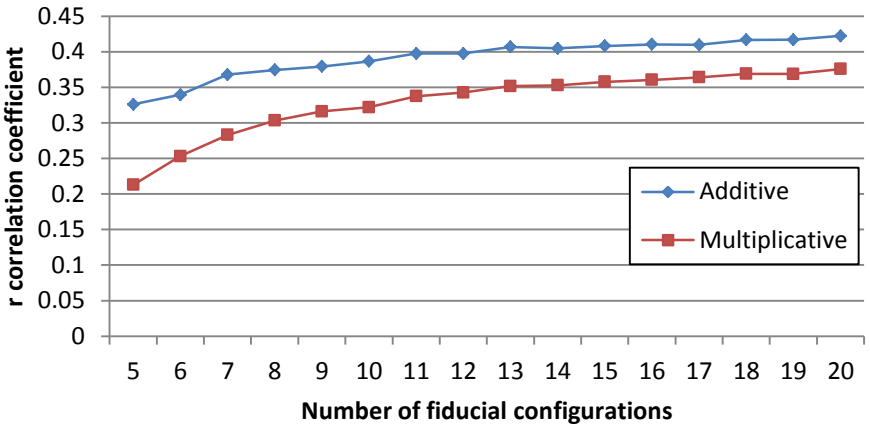


Figure 11: Correlation between TRE and ε as a function of the number of fiducial configurations utilized

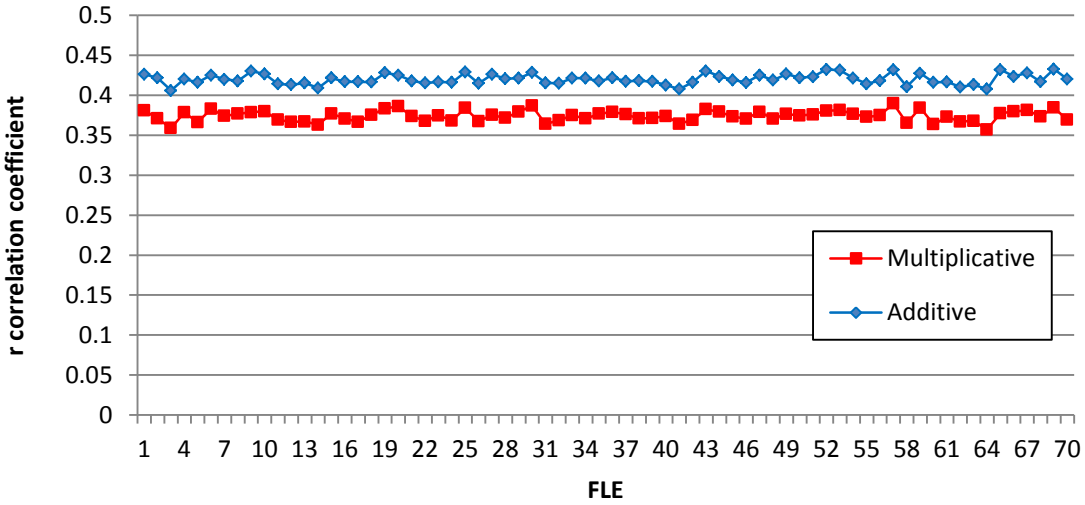


Figure 12: Correlation between TRE and ε as a function of the FLE.

The results of experiment 5 are shown in Figure 12. The graph shows that varying the size of FLE does not change the correlation between the AQUIRC algorithm and the TRE, for both the additive and multiplicative models. The results of experiment 6 are shown in Figure 13. If we utilize the AQUIRC algorithm with the multiplicative model and for every simulation, we choose the configuration with the lowest ε value, the TRE is decreased by a statistically significant 0.1078 mm when compared to the mean TRE value and is decreased by a statistically significant 0.1013 mm when compared to the TRE of the fiducial configuration with the minimum FRE. The additive model results in a statistically significant reduction of 0.1665 mm

compared to the mean TRE and a reduction of 0.16 mm when compared to the TRE of the fiducial configuration with the minimum FRE. We see similar results when looking at the standard deviation; the multiplicative model reduces the standard deviation by 0.0472 mm while the additive model reduces the standard deviation by 0.0801 mm compared to the minimum FRE configuration. This represents a 25.4% reduction in the standard deviation when using the additive model. The figure also shows the mean and standard deviation of the TRE when the configurations with the max and min TRE are selected at every run, respectively. Over the 5000 simulations performed, the worst TRE chosen by AQUIRC using the multiplicative model was 1.7215 mm, while using the additive model the worst TRE was 1.5505 mm both of which improved upon the worst TRE chosen using the minimum FRE, which was 2.1959 mm.

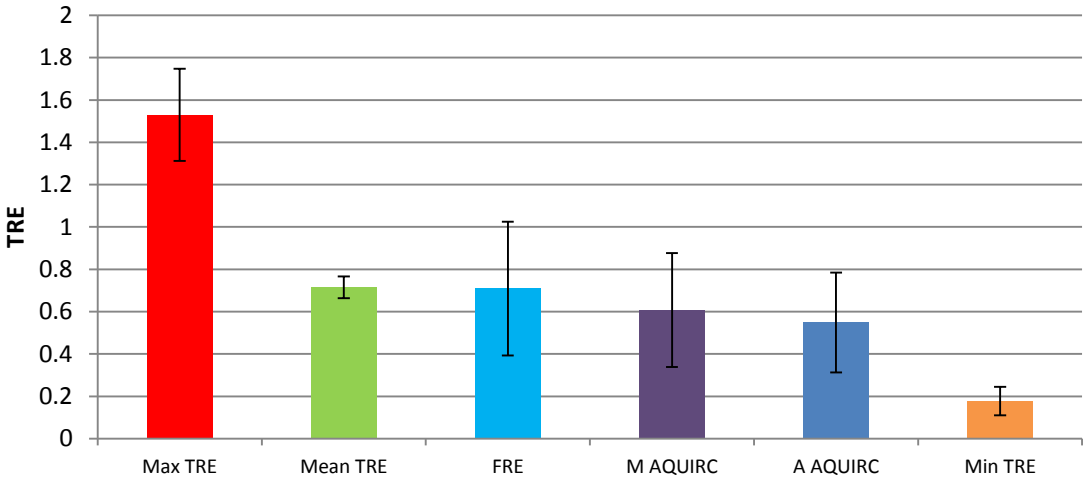


Figure 13: Bar graph of the mean value and standard deviation of the TRE value across 5000 simulations. For each simulation we calculate the mean TRE of all 39 fiducial configurations, the min TRE of all 39 fiducial configurations, the max TRE of all 39 fiducial configurations, the TRE of the fiducial configuration that AQUIRC identifies as being of the highest quality using the additive model, the TRE of the fiducial configuration that AQUIRC identifies as being of the highest quality using the multiplicative model and TRE of the fiducial configuration with the minimum FRE.

Conclusion

This chapter showed that the measure of registration quality resulting from AQUIRC is correlated to the target registration error. We are unaware of any other published work that describes a technique that is able to do so. As shown in experiment 1 and 2, computing the TRC method compared to the NTRC has a large effect on the results of the algorithm. In fact we do not observe any correlation with the traditional method.

We have shown that the number of fiducial configurations that are utilized in the complete graph of registrations affects the quality of AQUIRC's estimation and that by increasing the number of configurations the correlation between the ε value and TRE increases, though with diminishing returns. For fiducial-based registration, the additive model performs better than the multiplicative model. AQUIRC's performance does not appear to be affected by the size of the FLE.

We also show that by choosing the configuration that our algorithm identifies as producing the best registration, the average TRE can be reduced. Using the additive model and an actual IGS case, the AQUIRC algorithm on average reduces the TRE by 22.67%, compared to utilizing the minimum FRE. We also show that the worst case scenario for the AQUIRC algorithm is 29.39% smaller than the worst case scenario using the minimum FRE. The percentage decreased is important, since FLE can vary widely, and TRE can be large in some instances.

While performing our experiments, we have observed a spatial dependence in the correlation between TRE and the ε value. This spatial dependence has not been elucidated but this is a possible area of future research. Also other types of models that could better represent the combination of error that occurs when combining multiple transformations since, though

mathematically simple, the multiplicative and additive models may not be the best methods to describe how error is combined across multiple fiducial registrations.

We note that, in application, we have found little practical use for AQUIRC in regards to fiducial-based rigid registration. This is because to apply AQUIRC in practice would require someone to identify the location of the fiducial markers multiple times to create the multiple configurations necessary for AQUIRC. While using the AQUIRC algorithm would allow someone to better identify which configuration has lower TRE on average, a much simpler method would be to take the mean of the multiple fiducial markers. Taking the average marker location eliminates much of the FLE that occurs, more so as the amount of configurations identified increases, which makes using AQUIRC superfluous in this case (taking the average marker coordinate would not reduce a biased error, yet neither would the AQUIRC algorithm). AQUIRC, however, is useful for intensity-based registration as we can utilize multiple medical images, and there is no related method to averaging fiducial markers for intensity-based registration. It is also possible, that with a better understanding of the spatial dependency and the error model involved when computing the NTRC, that AQUIRC could become useful in the context of image guided surgery.

CHAPTER 5

Estimation of Rigid-Body Registration Quality

Introduction

Most rigid and affine registration methods rely on optimizing an intensity-based similarity criterion between images. Once registered, however, it is difficult to assess the quality of a rigid and affine registration based solely on the value of the similarity measure. This chapter presents AQUIRC applied to estimate error in rigid and affine registrations. AQUIRC is applied to estimate the registration quality of 10 MR and CT pairs and 10 MR and MR-contrast pairs. In this chapter the Non-Traditional Registration Circuit (NTRC) method is again utilized and compared to the Traditional Registration Circuit (TRC) because, although not equivalent, intensity-based rigid-registration is related to fiducial-based registration, and the assumptions made for the NTRC in the previous chapter are made again. We test and compare the NTRC and TRC method to test the efficacy of both methods. This chapter shows that the technique is capable of detecting most cases with registration error larger than 2° around one axis. AQUIRC is also shown to be better able to identify error than popular similarity measures both quantitatively, in the case of intra-patient rigid registration, and visually, for inter-patient affine registration

Applied AQUIRC

In this chapter, we define $E_C = \text{dissimilarity}\left(X, X'\right)$, where X is a grid evenly distributed with a point every 16 voxels and X' is X transformed around a registration circuit. The error across a

circuit, dissimilarity(X, X′), is then defined as the mean Euclidean distance between X and X′. An example registration circuit with a rigid transformation is shown in Figure 14.

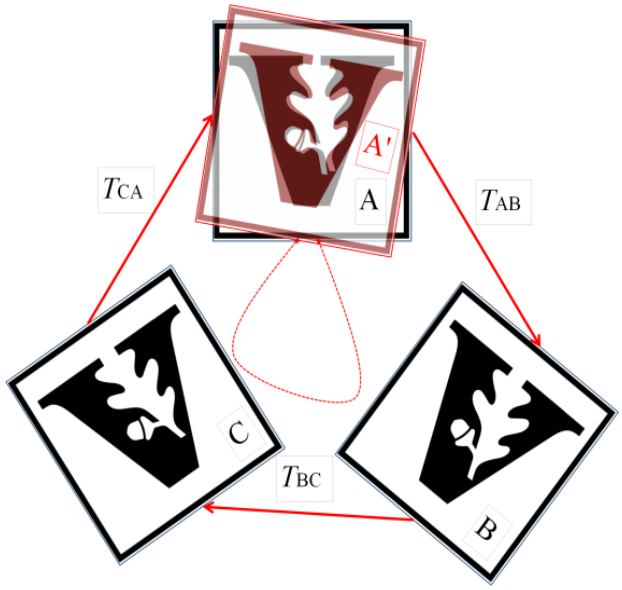


Figure 14: An example of one registration circuit, with the transformations between each image represented by $\boldsymbol{T_{AB}}$, $\boldsymbol{T_{BC}}$ and $\boldsymbol{T_{CA}}$. The red Vanderbilt logo has been transformed around the circuit, and the distance between the red Vanderbilt logo (A') and the original Vanderbilt logo (A) represents the registration circuit consistency error.

In this chapter, we define two sets of images: the reference and the test set. The reference set consists of 30 image volumes, 10 CT images and 20 MR images with each image acquired from a different patient, with no patient being used more than once. The images in the reference set are pair-wise registered together, and the transformations that register these volumes remain constant throughout all experiments. Visually, they form the network inside the square in Figure 15. The test set consists of 20 additional pairs of images. Ten of these test set pairs are MR and MR-contrast images while the other ten pairs are MR and CT images. Each pair is acquired from a different patient. The test sets are used to study our method's ability to detect registration errors between images with one pair shown schematically in Figure 15. AQUIRC requires affinely registering both images in the test pair to all of the images in the reference set. Once the required transformations are computed, the AQUIRC algorithm is used to compute ε_t, which we define as

the estimated registration quality between a test pair. We also compute the ε values between the images in a testing pair and all the volumes in the reference set.

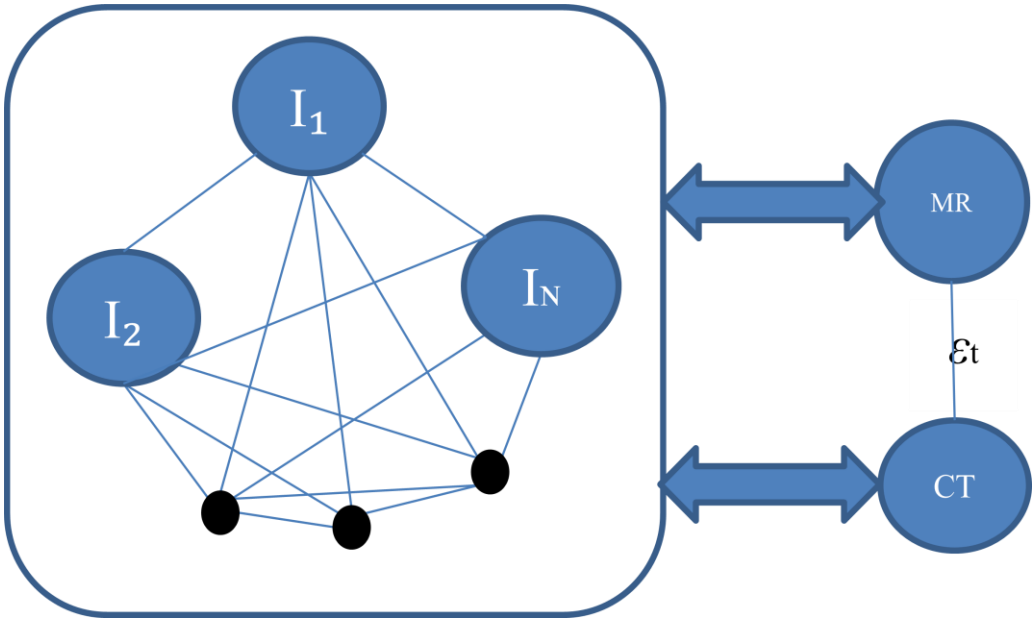


Figure 15: Reference set registered to two test images.

NMI is used as the similarity measure for registration. When calculating NMI and MI a histogram size of 64 bins is utilized (for both the registration and for a later comparison to AQUIRC). Affine transformations are used when registering volumes in the test set to reference volumes because these pertain to different patients. This is also the case for the pair-wise registration of the volumes in the reference set.

Image Information

The CT images we use were typically acquired at kVp = 120 V, exposure 350 mAs and 512 x 512 pixels. In-plane resolution and slice thickness were 0.5 mm and 0.75 mm respectively. The MR images were T1-weighted sagittal volumes with 256x256x170 voxels 1 mm in each direction acquired with the parameters TE = 2.4 ms, TR = 12.2 ms. The contrast images were Gadolinium enhanced and were acquired with the same parameters and in the same scanner

space. The images were acquired with the SENSE parallel imaging technique (T1 W/3D/TFE) from Philips on a 3 Tesla scanner. All images are resampled to 256x256x256 voxels and 1 mm in each direction.

Experiments

We perform two experiments to test AQUIRC in this chapter. We test AQUIRC using the TRC and the NTRC methods and compare AQUIRC to the NMI and MI similarity measures. For the test set of images in experiment 1, we utilize a set of MR images without contrast and MR images with contrast acquired from 10 patients. These volumes have been acquired sequentially with the patient sedated and we can reasonably assume that they are registered. This was verified visually and image pairs in which there was no apparent mis-registration were selected. The MR and MR-contrast test images are affinely registered to each of the images in the reference set of images. The AQUIRC algorithm is applied using both the TRC and NTRC, and a baseline ε_t value is calculated for the link between the test images as well as for each of the affine registrations in the network of images.

To simulate a case in which there is a registration error, each MR image in the test set is rotated along the x-axis by positive and negative 0.5°, 1°, 2°, and 4° angles. We have also repeated this test by rotating the test set along the axis that is the normalized sum of the y and z-axis and have obtained comparable results. This was done to ensure that there is no bias in the results for using the x-axis; for simplicity we only show the results only for x-axis simulated registration error. The reference set is affinely registered to each of the rotated MR test images independently while the transformation between the MR to MR-contrast test images remains the identity transformation. Thus the transformation between the MR and MR-contrast images is

incorrect by the amount of simulated rotational error applied to the MR image. The ε_t value is re-calculated for each rotation.

In experiment 2 we assess the capability of our algorithm to detect errors in intra-patient rigid registrations between MR and CT images of the same patient. In this experiment, 10 MR and CT test images are utilized (an MR and CT image were acquired for each patient). For each pair, the MR and CT test images are rigidly registered to each other. The rigid registration between the MR and CT images is visually checked to verify its quality. After visual assessment, we assume that this registration is the ground truth, although in practice small errors are difficult to detect visually and residual errors may be present. The MR and CT images are then affinely registered to the images in the reference set. The MR image is rotated along the x-axis in the center of the image space by positive and negative 0.5°, 1°, 2°, and 4°. The reference set is again affinely registered to each of the rotated MR images while the transformation between the MR and CT test images remains unchanged; the ε_t value is re-calculated for each rotation. Thus the transformation between the MR and CT images is incorrect by the amount of simulated rotational error applied to the MR image. An example MR image is shown in Figure 16, where the top MR image has been rotated by 2° and the bottom image is the original orientation of the image.

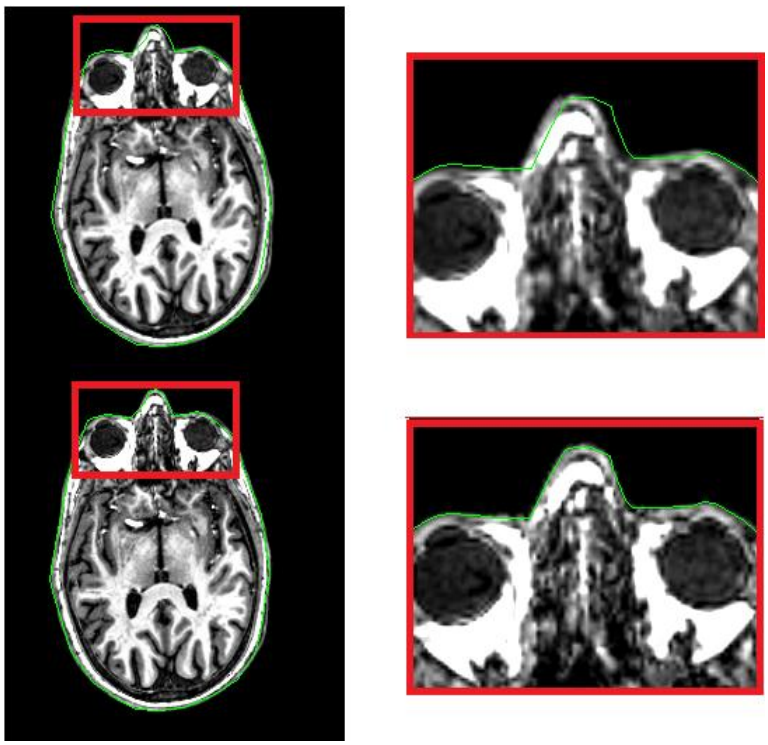


Figure 16: Example of a 2 degree rotation along the x-axis

As a comparison, we also test the ability of the similarity measures Normalized Mutual Information (NMI) and Mutual Information (MI) to assess registration quality. This is similar to what was done by Crum in [77] where they used simulated 2d images to identify error using a pixel-based similarity measure. Here we apply NMI and MI to identify registration error in clinical 3d images. The values of NMI and MI can be affected by the number of time the images that are being measured have been interpolated. Therefore, at each degree of simulated rotation error that is added (0°, 0.5°, 1°, 2°, and 4° angles), we ensure that each image is interpolated only once for the calculation of NMI and MI.

Results

AQUIRC MR to MR-contrast registration

In this section we show the results obtained for Experiment 1 and 2 in Figure 17, Figure 18, Figure 19, and Figure 20. The figures on the left show the mean and standard deviation of the ε

value computing using the TRC and NTRC as well as the MI and NMI values for the baseline 0° error case and for each rotational error that is added. The figures on the right show each of the individual ε values when using the TRC and NTRC to compute AQUIRC as well the MI and NMI values for each registration. This is done for both the MR to MR-contrast data set (top) and the MR to CT data set (bottom). The results of AQUIRC when using the NTRC are shown in Figure 17, the results of AQUIRC when using the TRC are shown in Figure 18, the results of MI are shown in Figure 19, and the results of NMI are shown in Figure 20.

We define an arbitrary error detection threshold to be the value which is one standard deviation above the mean value at 0° (i.e. no error applied). This value is calculated for AQUIRC using the TRC, AQUIRC using the NTRC, NMI and MI, and is shown as a black line in each of the figures. In Table 1, we list the mean value, minimum value, maximum value, standard deviation, and error detection threshold for all values computed using both the MR to MR-contrast data set and the MR to CT data set. An interesting result from Table 1 is that AQUIRC using TRC resulted in a higher ε value compared to AQUIRC using NTRC. This is counter-intuitive to the conclusion of [76], which showed that, for fiducial-based registration, AQUIRC using TRC results in a lower value compared to NTRC. A potential reason for this is that here, AQUIRC is being applied to transformations that include affine transformations, which may cause the results to vary compared to using purely rigid transformations and may be the reason AQUIRC using TRC results in higher ε values compared to NTRC. Generally, the goal of the error detection threshold is to correctly identify registration cases in which there is no error, as well as to detect cases in which there is error. Here we define a method to be able to detect error when the error detection threshold is below (for AQUIRC) or above (for NMI and MI) one standard deviation at a specific degree of rotational error. For instance, we can see in Figure 17

that AQUIRC using the NTRC method is able to detect rotational error at 2 degrees using our definition. Comparing Figure 17 and Figure 18, AQUIRC using the NTRC method is able to identify registration error better then when using the TRC method as AQUIRC using NTRC can identify 2 degrees of error in both cases. While for AQUIRC using the TRC method identifies error at 4 degrees for the MR to MR-contrast case, and does not detect 4 degrees of error in the MR to CT case. We can see from Figure 19 and Figure 20 that NMI and MI both perform comparably to AQUIRC when using the NTRC method for the MR to MR-contrast case, yet neither similarity method is able to detect even 4 degrees of error for the MR to CT case.

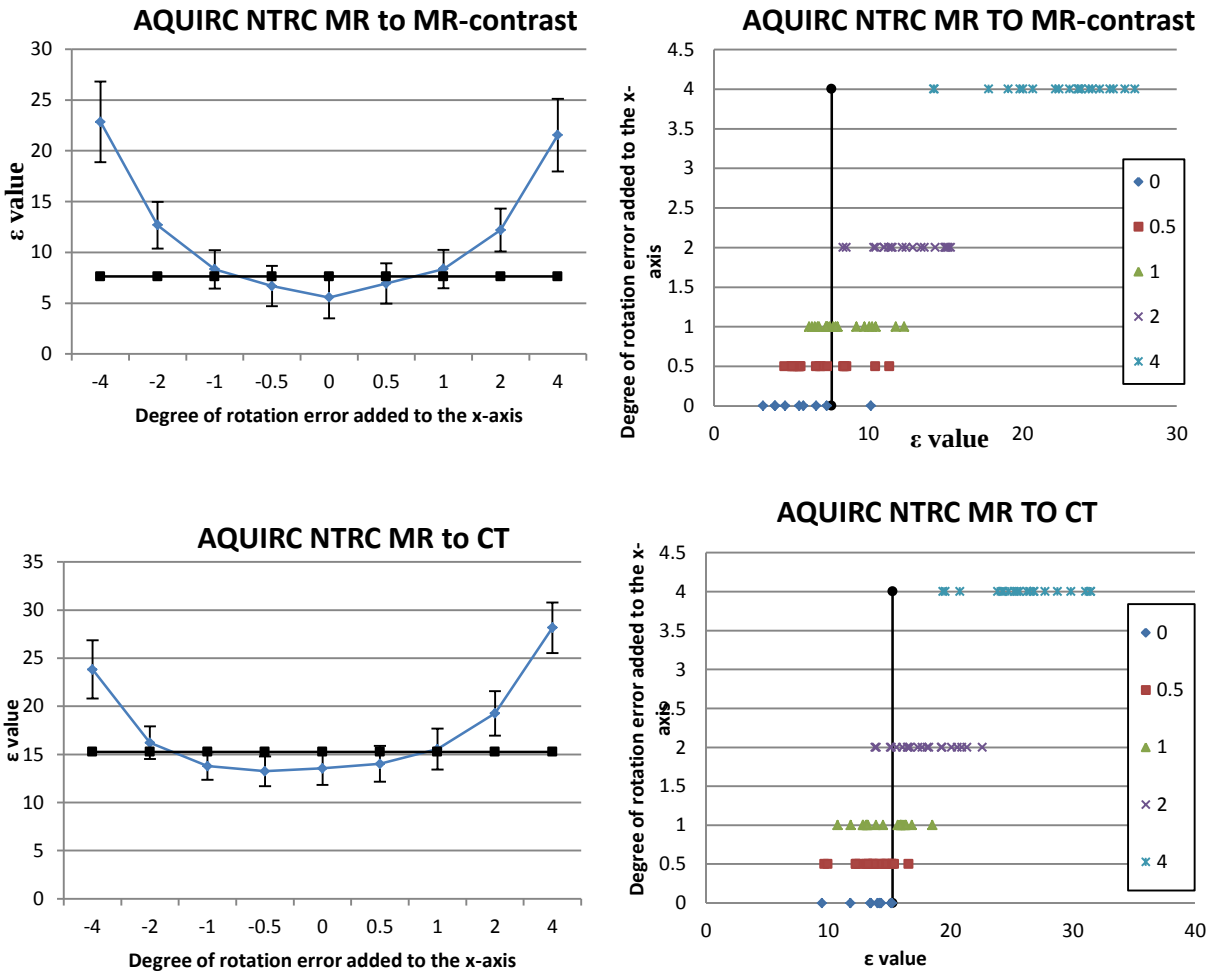


Figure 17: AQUIRC using the NTRC method, Left: the mean and standard deviation of 10 intra-patient rigid rotations, with error being added to the rigid rotation and recalculating ε value each time error is added. Right: Distribution of the ε values.

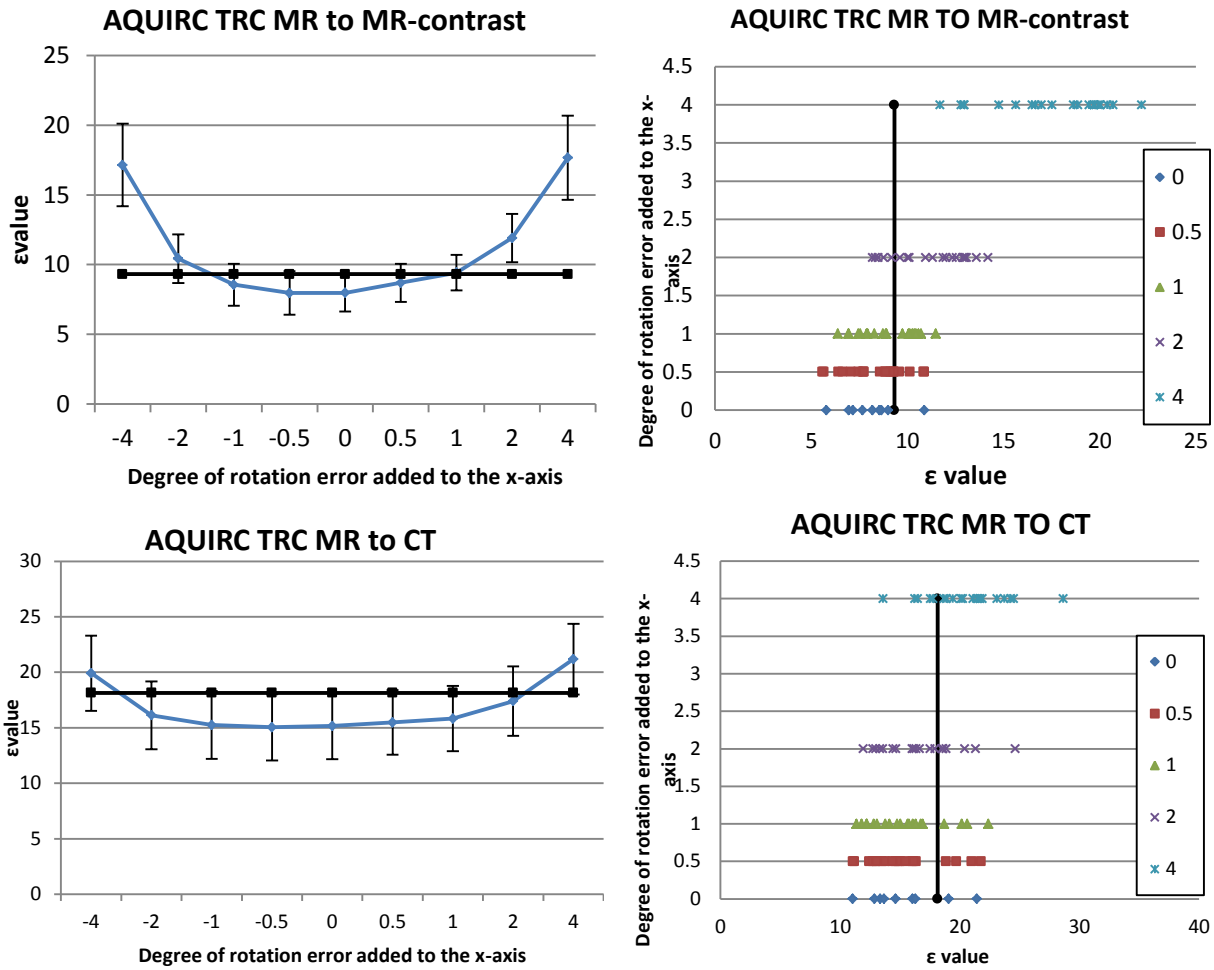
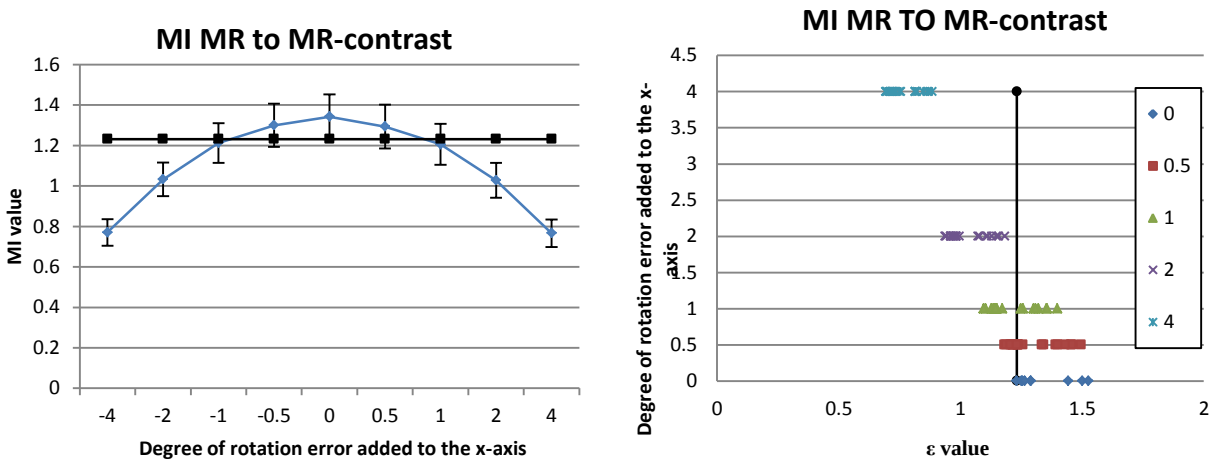


Figure 18: AQUIRC using the TRC method, Left: the mean and standard deviation of 10 intra-patient rigid rotations, with error being added to the rigid rotation and recalculating the ε value each time error is added. Right: Distribution of the ε values.



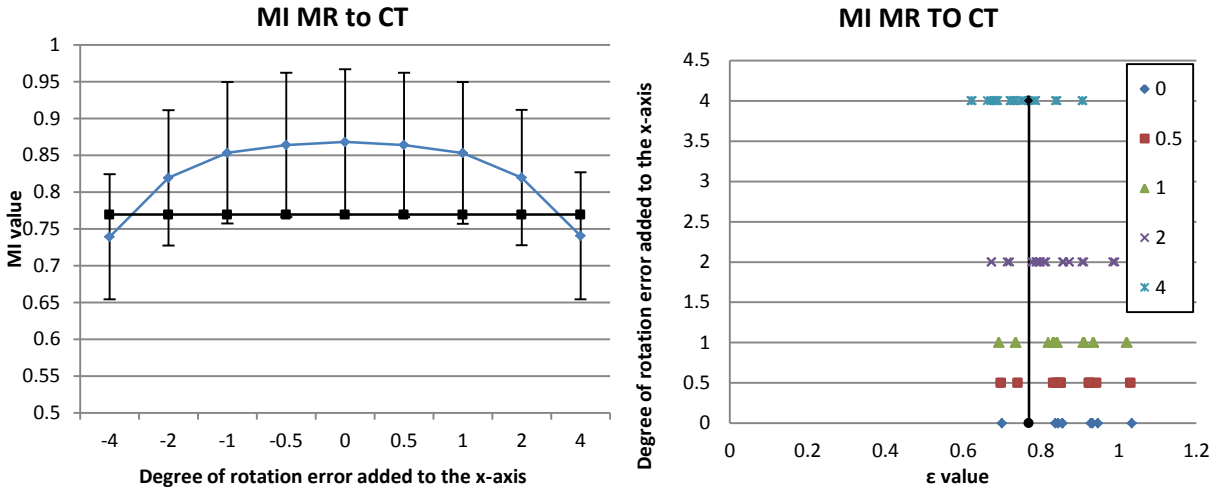


Figure 19: MI, Left: the mean and standard deviation of 10 intra-patient rigid rotations, with error being added to the rigid rotation and recalculating the MI value each time error is added. Right: Distribution of the MI values.

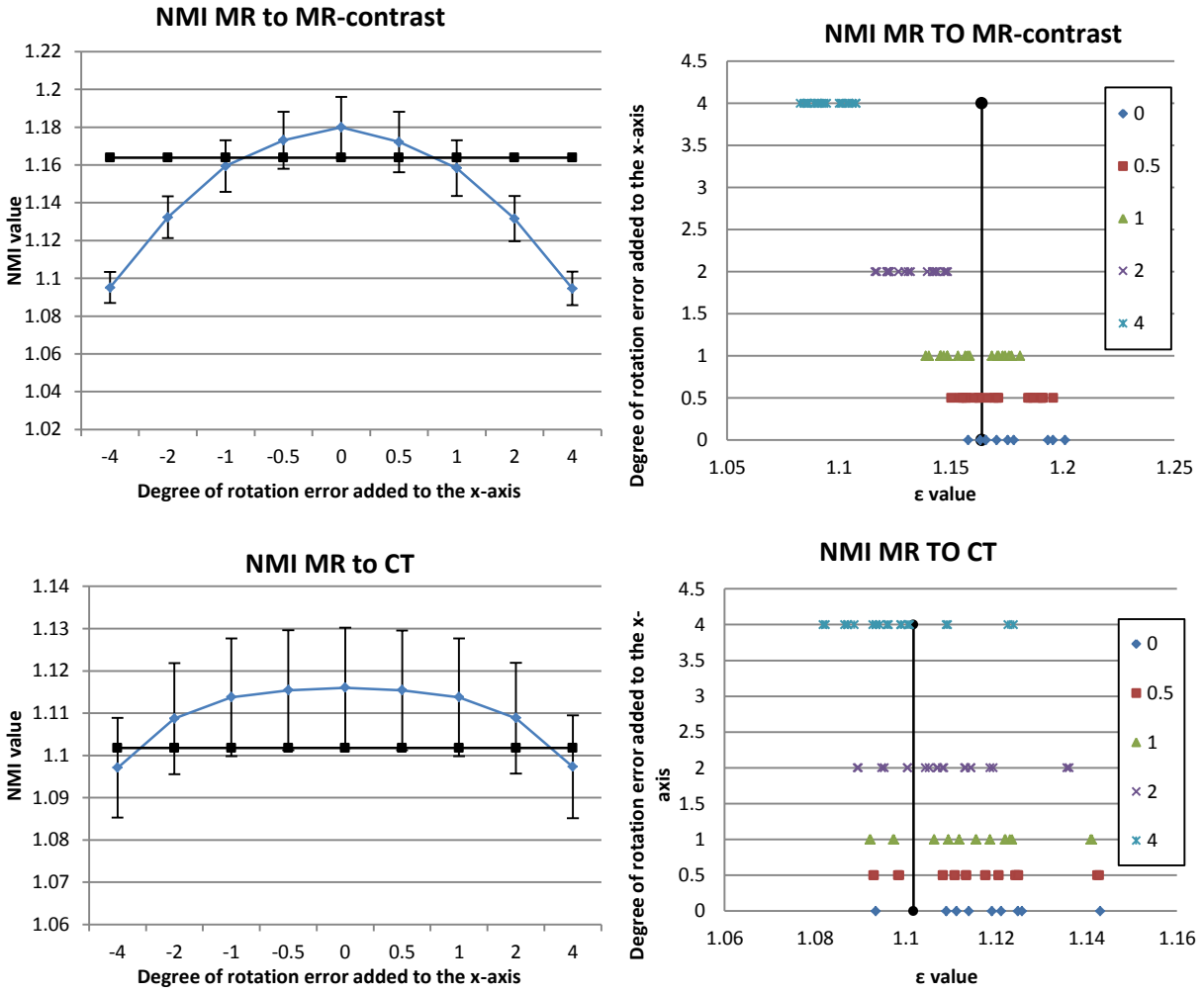


Figure 20: NMI, Left: the mean and standard deviation of 10 intra-patient rigid rotations, with error being added to the rigid rotation and recalculating the NMI value each time error is added. Right: Distribution of the NMI values.

Table 1: Mean value, minimum value, maximum value, standard deviation, and the error detection threshold for AQUIRC NTRC, AQUIRC TRC, MI and NMI are shown for the case in which no error is added.

Detection Method	**Data Set**	**Mean Value**	**Minimum Value**	**Maximum Value**	**Standard Deviation**	**Detection Threshold**
AQUIRC NTRC	MR-CT	13.55	9.51	15.23	1.72	15.27
AQUIRC NTRC	MR-MR contrast	5.57	3.19	10.18	2.06	7.63
AQUIRC TRC	MR-CT	15.1619	11.0433	21.4461	3.1456	18.146
AQUIRC TRC	MR-MR contrast	7.9783	5.7879	10.8804	1.4139	9.3197
MI	MR-CT	0.868	0.6998	1.0347	0.0987	0.9667
MI	MR-MR contrast	1.3429	1.2309	1.5249	0.1107	1.4536
NMI	MR-CT	1.116	1.0934	1.143	0.0142	1.1302
NMI	MR-MR contrast	1.18	1.1577	1.2011	0.016	1.196

Comparison

To compare the results of NMI, MI and the AQUIRC algorithm using the TRC method and the NTRC method and to compute their specificity and sensitivity, we calculate an ROC curve with the results shown in Figure 21. To do this, we use all the values computed as measures to identify when error occurs in a registration. We fix the degree of rotational error and vary the value of each of the measures, calculating the true positives and false positives at each variation, plotting the values to produce Figure 21. Here, a true positive is defined as a mis-registration that is identified correctly while a false positive is defined as a registration without error identified as being a mis-registration. We calculate the ROC curve with the negative and positive registration error cases combined.

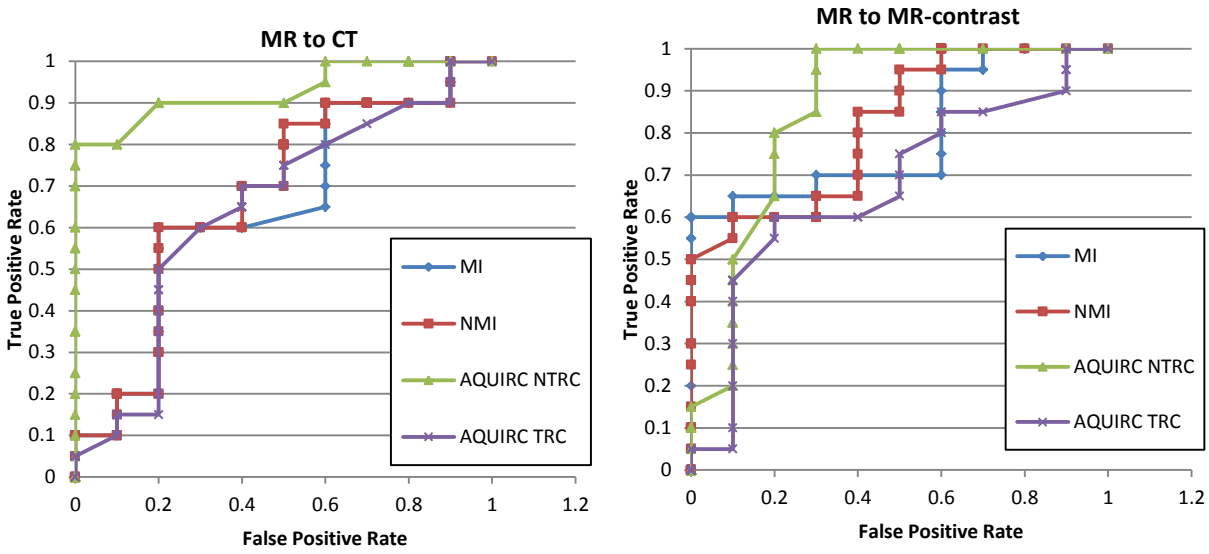


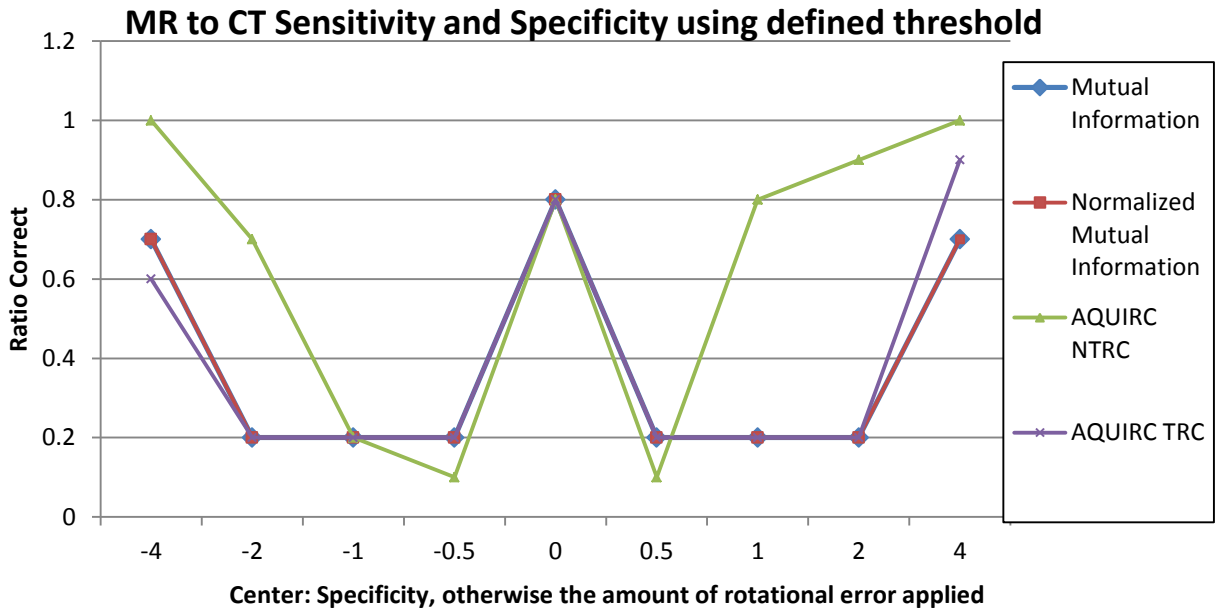
Figure 21: ROC curve utilizing the MI, NMI, NTRC ε, and TRC ε values as classifiers to identify of rotational error. Left: MR to CT test set at **2**°error, Right: MR to MR-contrast test set at **1**°error.

The MR to CT data set shown in Figure 21 uses the 2° rotation error case (i.e. both 2° and -2° combined), while the MR to MR-contrast data set uses the 1° rotation error case. We do this because all 4 methods perform better on the MR to MR-contrast data set compared to the MR to CT data set; thus, differences between the four algorithms are more apparent using the 1° rotation for the MR to MR-contrast and more apparent using the 2° rotation error cases for the MR to CT data set. The left of Figure 21 shows the results of the ROC curve for the MR to CT data set, while the right of Figure 21 shows the results of the ROC curve for the MR to MR-contrast test set. Compared to NMI and MI, AQUIRC using the NTRC method performs better for the MR to CT test set and, once a 0.2 false positive rate is acceptable, AQUIRC using the NTRC method performs slightly better for the MR to MR-contrast test set as well. While AQUIRC using the TRC method performs similarly to NMI and MI for the MR to CT test set, and slightly worse than NMI and MI for the MR to MR-contrast test set.

Leave One Out

We compare the results of NMI, MI and the AQUIRC algorithm using the NTRC and TRC methods in a leave one out framework. To do this we utilize the error detection threshold defined earlier except here the threshold is calculated using 9 out of the 10 registrations for each data set. We then test whether the threshold correctly identifies cases in which no error is applied (defined as 0 degrees in Figure 22) and whether the threshold correctly identifies when error is applied. This is repeated for all 10 cases and the percentage of cases in which the threshold correctly identifies error or no error is shown in a line graph in Figure 22.

The defined threshold for AQUIRC using the NTRC method correctly identifies the cases in which no error is applied in 8 out of 10 cases in the MR to CT case and is able to identify error at around 2° rotational error, correctly identifying 8 out of 10 cases. AQUIRC using the TRC method, NMI and MI correctly identify when no error is applied in 8 out of 10 cases as well, but none of the three methods are able to identify error well until 4° of rotational error, and even then they correctly identify only 7 out of the 10 cases. For the MR to MR contrast data set, MI, AQUIRC using the NTRC method, and AQUIRC using the TRC method identified 9 out of 10 cases in which no error is added, and NMI identified 8 out of 10 cases, with the NMI, MI and AQUIRC using the NTRC method performing similarly, and AQUIRC using the TRC method performing worse. NMI and MI perform slightly better at identifying cases of 1° rotational error compared to AQUIRC using the NTRC method, yet all three methods are able to correctly identify all cases of 2° rotational error while AQUIRC using the TRC method was able to identify all cases of 4° of error.



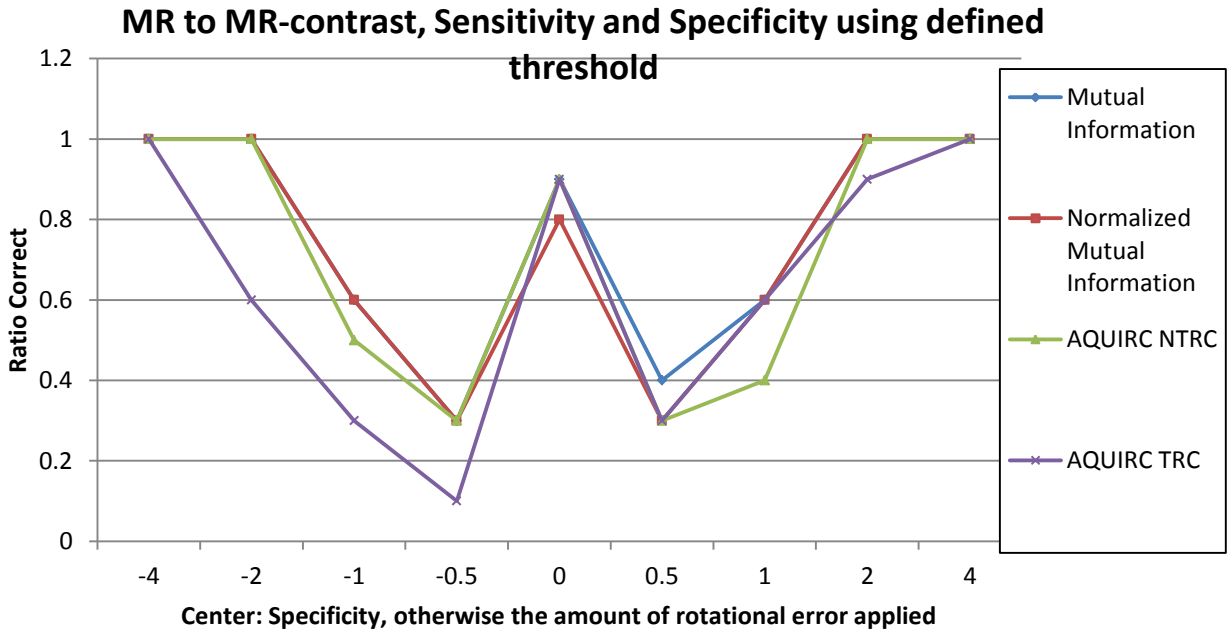


Figure 22: A comparison between NMI, MI, and AQUIRC using the NTRC and TRC methods in a leave one out framework rotating along the x-axis.

Visualizing Registration Quality

In this work, we have also computed the NMI values and ε values of the affine, inter-patient registrations between the images in our reference set. Thus we can visually assess the quality of

the affine inter-patient registrations that AQUIRC and NMI identify as being high or low quality. In order to simplify comparisons, MI and AQUIRC using the TRC method were not used here.

A random CT image from the reference set was chosen, and the ε value and NMI value were calculated between it and the other 29 images in the reference set. Figure 23, Figure 24, Figure 25, and Figure 26 show the best and worst inter-patient registrations between the randomly chosen CT image and the other images in the reference set. The reference set has both CT images and MR images. Thus we compute the best and worst registrations for both image types, with the 10 CT images registered to the target in Figure 23 and Figure 24, and the 20 MR images registered to the target in Figure 25 and Figure 26. The CT image on the top row is the target image for the affine registration, with a manually drawn contour in green; the bottom row is the image registered to the target image. For each registration shown we have chosen a representative slice of the target image and the same slice of the registered source image. The manual contour is projected onto the bottom row of registered images. In order from left to right appear the three registrations identified as the worst quality followed by the three registrations identified as the best quality (with the best image on the far right). The ε value and NMI value calculated for each specific registration are shown below the images.

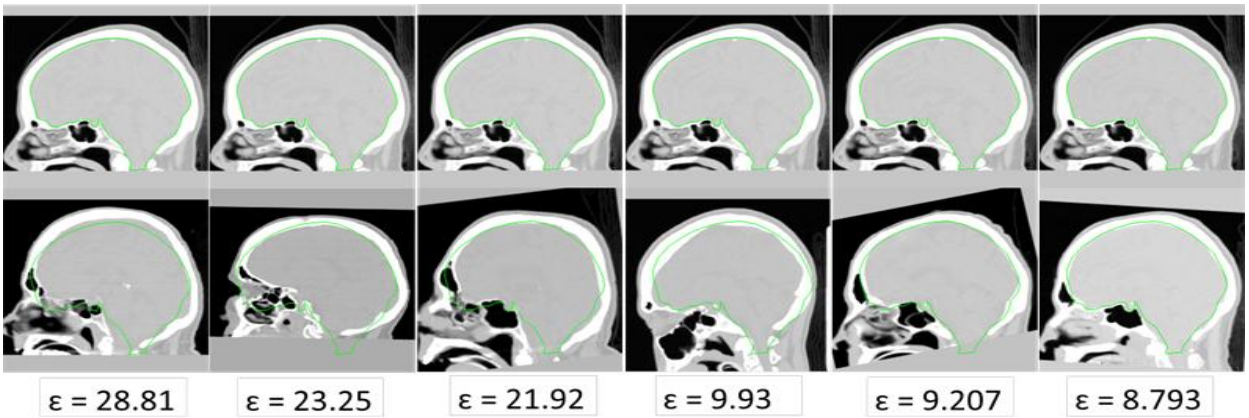


Figure 23: Top row: target image and manually drawn contour in green. Bottom row: CT image registered to the target image with an overlay of the contour. In order from left to right appear the three

registrations that AQUIRC identifies as the worst quality followed by the three registrations that AQUIRC identifies as the best quality. The ε value is shown below each registration.

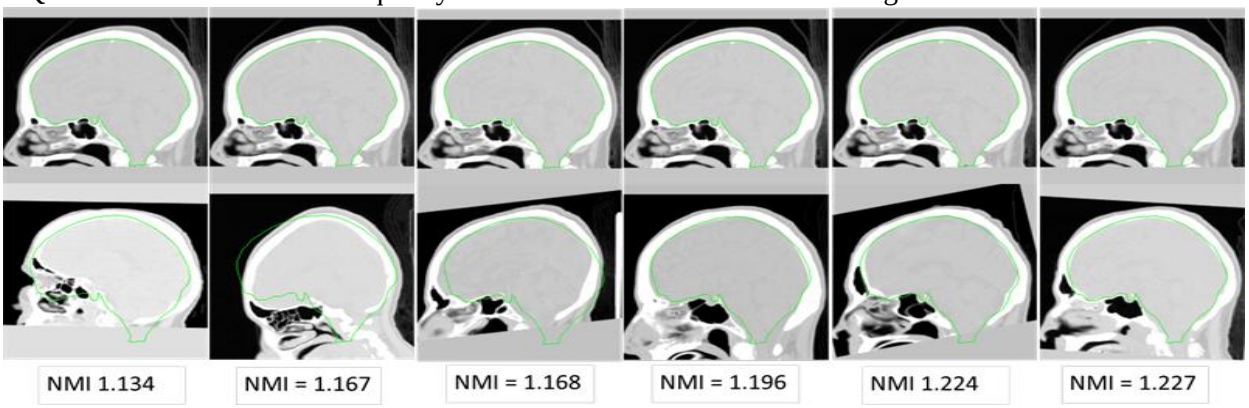


Figure 24: Top row: target image and manually drawn contour in green. Bottom row: CT image registered to the target image with an overlay of the contour. In order from left to right appear the three registrations that NMI value identifies as the worst quality followed by the three registrations that NMI value identifies as the best quality. The NMI value is shown below each registration.

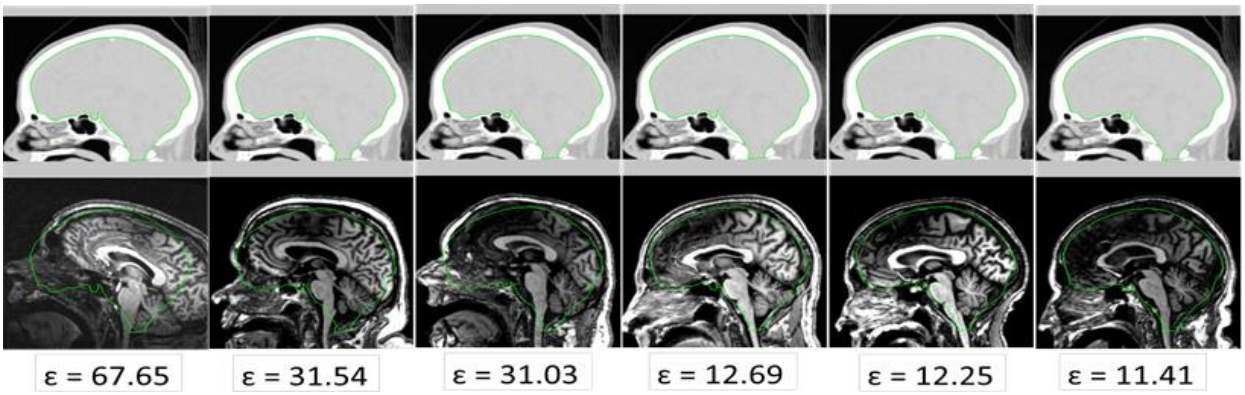


Figure 25: Top row: target image and manually drawn contour in green. Bottom row: MR image registered to the target image with an overlay of the contour. In order from left to right appear the three registrations that AQUIRC identifies as the worst quality followed by the three registrations that AQUIRC identifies as the best quality. The ε value is shown below each registration.

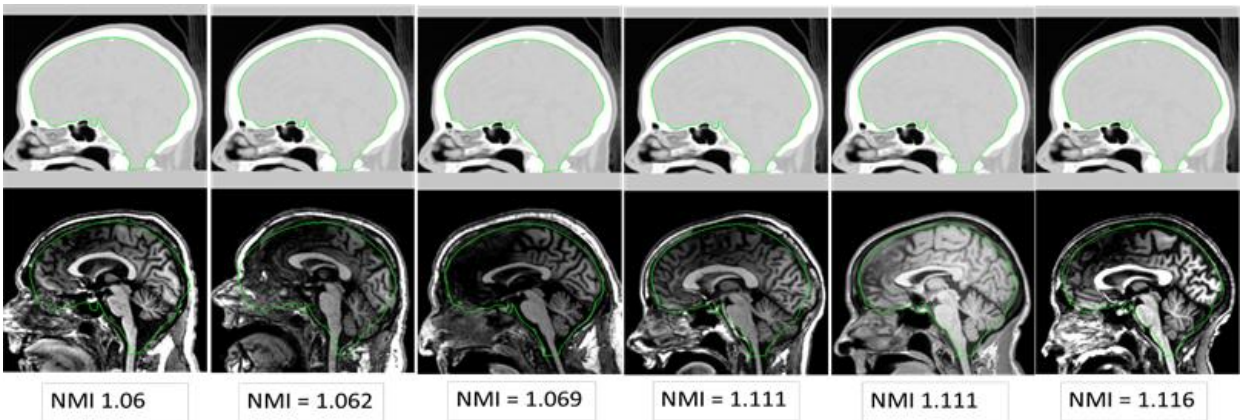


Figure 26: Top row: target image and manually drawn contour in green. Bottom row: MR image registered to the target image with an overlay of the contour. In order from left to right appear the three

registrations that NMI value identifies as the worst quality followed by the three registrations that NMI value identifies as the best quality. The NMI value is shown below each registration.

Comparing Figure 23 and Figure 24, we can see that NMI and AQUIRC both identify the same two CT images as being the best registrations to the target image, and the worst registration identified by NMI is the second worst registration identified by AQUIRC. The other registrations that AQUIRC and NMI identify as being of low quality in Figure 23 and Figure 24 all appear to be visually not well registered. It does appear, however, that the third best registration identified by NMI is visually better registered than the third best registration identified by AQUIRC. Overall the rankings of the quality of registrations do not differ widely between NMI and AQUIRC when considering the CT to CT inter-patient registrations.

Comparing Figure 25 and Figure 26 and the overall registration rankings of the MR to CT inter-patient registrations we see a larger difference between AQUIRC and NMI. The registration that AQUIRC identifies as being the worst registration in Figure 25 can visually be identified as low quality, and easily the worst registration that we visually identified in the MR to CT data set. NMI, however, identifies this registration as the 5th best registration. AQUIRC and NMI do not completely differ, as the second worst registration identified by NMI is the third worst registration identified by AQUIRC. However, the registration that NMI identifies as the third worst registration visually appears to be a good quality registration, while the AQUIRC algorithm identifies this registration as a moderate-quality registration, ranking it as the 12th best registration. AQUIRC and NMI generally agree when considering the high quality registrations, except for the registration that NMI identifies as the second best registration - AQUIRC ranks this registration as the 11th best registration.

Conclusion

In this work we tested the ability of the AQUIRC algorithm to detect error in rigid and affine registration. We utilized two test sets of images, MR to MR-contrast and MR to CT images and have tested two methods to compute a registration circuit (NTRC, TRC) and have evaluated their effect on the results of AQUIRC. We have shown that the NTRC method outperforms the TRC method for rigid-body image registrations. The results for the AQUIRC algorithm when using the NTRC method are similar for both data sets and in a practical application we have calibrated our network to identify most instances in which 2° of rotational error has been added.

For a comparison, we utilized two popular similarity measures: NMI and MI. Both methods performed worse at identifying registration error when compared to the AQUIRC using the NTRC method for the MR to CT data set while NMI and MI performed similar to AQUIRC using the NTRC method in the MR to MR contrast data set. Intensity-based similarity measures are dependent on the image quality, acquisition sequence, modality type, and noise level; thus they are not often used for error detection. On the other hand, the AQUIRC method relies on the quality of the correspondence (i.e. registration) between the reference set and the images in question and is un-affected by the image modality and acquisition sequence, and less affected by noise level and image quality.

Our results also showed that the AQUIRC algorithm is able to identify large inter-patient affine registration errors. Figure 23, Figure 24, Figure 25, and Figure 26 show that the ε value visually correlates to registration quality. While NMI was able to identify large inter-patient affine registration errors in the CT to CT inter-patient registration case, it was not able to identify a visually apparent mis-registration, a case that was identified by the AQUIRC algorithm. The

quality of the inter-patient and inter-modality registrations better corresponds visually to AQUIRC compared to NMI.

AQUIRC can be utilized as a method to determine registration quality without manual intervention. AQUIRC requires having multiple atlases that can be pre-registered together into a network of registrations, into what we refer to as the reference set. We can pre-compute the registration circuits that are unique to the registrations of these atlases. Then, to compute AQUIRC on a new image, the atlas images are registered to the new image, the registration circuits involving the new image are computed, and then AQUIRC is run to solve for the unknown registration quality. With a reference set of 10 and the reference set registration circuits pre-computed, the AQUIRC algorithm for rigid registration can be computed in 0.83 seconds. Therefore, it is feasible that AQUIRC can be used in an automatic registration framework, without adding an unreasonable amount of computation time to the process.

CHAPTER 6

Estimation of Global Non-Rigid Registration Error and its Application to Atlas Selection

Introduction

Multi-atlas registration-based segmentation has recently become a popular technique in medical imaging. This technique requires multiple registrations as well as a method to combine the multiple segmentations. Thus, the quality of the registrations has a large impact on the final segmentation. In this chapter, AQUIRC is applied to global registration quality estimation and it is shown that the measure that results from AQUIRC correlates strongly (R^2 = 0.79) with the ground truth DICE similarity index. When applied to global atlas fusion, AQUIRC performs statistically better than both an un-weighted majority vote technique and a technique using residual Normalized Mutual Information.

Applied AQUIRC

In this chapter, the dissimilarity function, $\mathrm{E_C} = \mathrm{dissimilarity}(\mathrm{X}, \mathrm{X}')$, is defined as the DICE [74] metric. To keep the language of this dissertation consistent, we continue to refer to the dissimilarity function, however, DICE is a similarity metric, which AQUIRC can handle in the same manner as a dissimilarity function. Thus, X is a set of manual segmentations and X' is the projection of the manual segmentations across the circuit. DICE is calculated as in (8). In the context of atlas selection, we do not have manual segmentations on the target image. Therefore we utilize all possible circuits except those circuits that begin on the target image. For example, in Figure 27, if image I_1 is the target image, we would not utilize any circuit that begins on image I_1. Thus, a circuit such as $T_{\mathrm{I}_1\mathrm{I}_2}(T_{\mathrm{I}_2\mathrm{I}_3}(T_{\mathrm{I}_3\mathrm{I}_1}(\mathrm{X})))$ would not be utilized because there is no manual segmentation X, to transform. We do, however, consider all registration circuits of size 3.This mimics the situation where there is a set of atlases that have been manually delineated and

one target image to segment. Empirically P has been observed to be full rank in this case when $N \geq 10$ (i.e. 9 atlases and 1 target image). N is required to be larger here than in the general case because there are circuits that are not utilized.

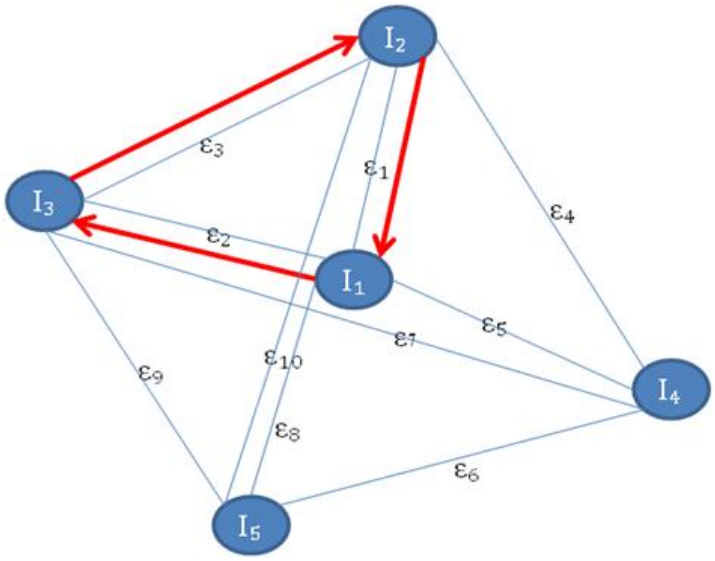


Figure 27: Example of a complete graph with one circuit shown with red arrows

Image Information and Registration Method

In this experiment, we use 10 T1-weighted head MR images with dimensions of 256x256x256 and voxels of 1 mm in each direction from the OASIS data set [78], each containing manual delineations (referred to here as labels) of 40 structures, with each label representing a different structure. The images are registered utilizing the Adaptive Bases Algorithm (ABA) [52], which uses Normalized Mutual Information (NMI) [32] as its similarity measure. Once the deformation field is computed between two MR images, the manual labels are transferred using nearest neighbor interpolation. When transferring labels across a circuit, $T_{\mathrm{AB}}(T_{\mathrm{BC}}(T_{\mathrm{CA}}(\mathrm{X})))$, the deformation fields are first composed using tri-linear interpolation. The composed deformation field is then applied to the labels which are transferred using nearest neighbor interpolation. DICE is calculated for each individual label and averaged over the 40 regions of interest.

Label Fusion

To combine the labels for multi-altlas segmentation, we use a weighted voting technique described as

$$\psi_i = \arg\max_s \sum_j \mathrm{I}(\mathrm{D}_{ij} = s) * w_j \quad (13)$$

where ψ_i is the estimated label for voxel *i*, *s* is the list of 40 labels, D_{ij} is the decision of rater (in this case, the segmentation of one atlas) *j*, I is the indicator function, and w_j is the assigned weight of rater *j*. When we utilize an un-weighted voting scheme, w_j is set equal to one for every rater.

Experiments and Results

Two experiments have been performed to illustrate the performance of the AQUIRC algorithm. The first experiment correlates our measure of registration quality with the ground truth DICE metric; the second experiment applies the method to the problem of label fusion, in which we compare the results of AQUIRC to the residual NMI of the registered images, a method that is often used for atlas selection and atlas fusion.

Comparison to Ground Truth

To validate the AQUIRC algorithm we compare the results of AQUIRC to the DICE value of each atlas-based segmentation since DICE is often utilized to verify the quality of a registration. We first compute a pair-wise registration for each pair of the 10 images in the dataset. This results in 45 unique registrations. The DICE similarity value is then computed between the projected structures and the original structures and then correllated to the ε values. The AQUIRC algorithm is then applied in a leave-one-out framework. Each of the 10 images is treated as a target image to be segmented, using the other 9 images as atlases. All registration circuits are used except for those that begin with the target image to be segmented. This is done

to simulate an applicable situation, in which there is no manual segmentation on a new target image.

Repeating this method for each image in the data set results in 90 estimations of the quality of a registration, because we estimate each registration's ε value twice. For example, in Figure 27, ε_1, is calculated when I_1 is the target image and again when I_2 is the target image. This dual estimation evaluates the consistency of AQUIRC across different circuit configurations, which we found to be the case. When comparing the two ε values calculated for the same registration, we found that there is no statistical difference between the two sets of 45 ε values ($p > 0.05$). The mean difference between the ε values calculated for the same registration, is -0.0012 with a standard deviation of 0.0075.

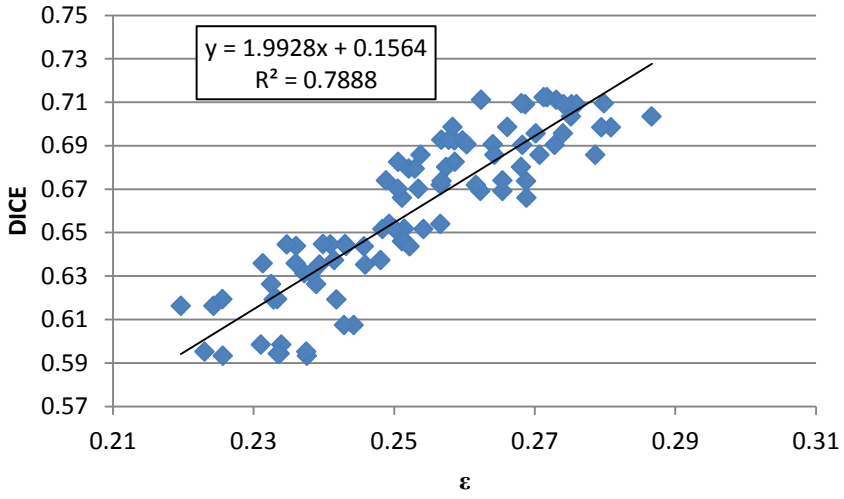


Figure 28: Scatter plot of the ε quality of registration value compared to the ground truth DICE of 90 pair-wise registrations

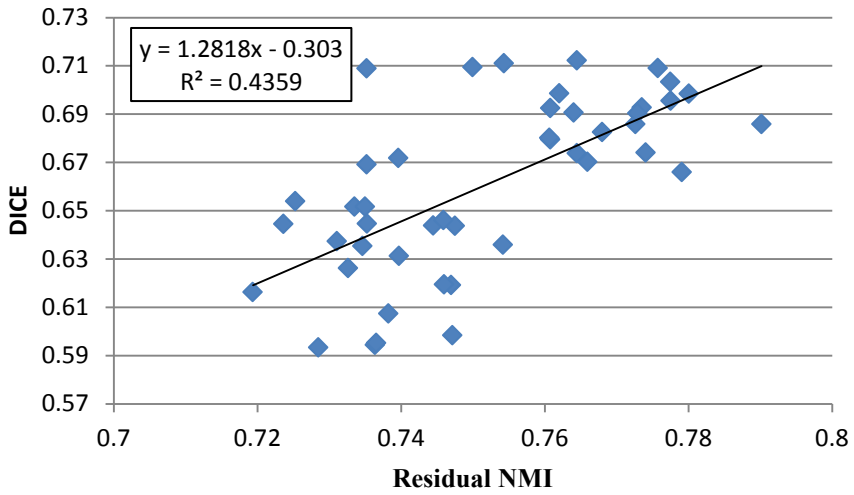


Figure 29: Scatter plot of the residual NMI compared to the ground truth DICE of the 45 pair-wise registrations

As a comparison, we use a popular [7, 79] method often used in multi-atlas based segmentation to rank atlases: residual NMI. This method is calculated by taking the transformed atlas after registration, and calculating the residual NMI between the atlas and target MR images. To calculate NMI, we utilized a bin size of 16 for the histogram, and we calculated this for each of the 45 unique registrations. Residual NMI is calculated over the region included in the segmentation on the atlas image. The registration algorithm used, ABA, is bijective, therefore we did not calculate the residual NMI in both directions (i.e. atlas to target and target to atlas). Figure 29 shows the DICE and ε values in a scatterplot. There is a strong correlation between the ground truth error metric (DICE) and the ε values, with an $R^2 = 0.79$. The correlation for residual NMI on the other hand, is moderate, with an $R^2 = 0.44$.

Label Fusion Results

The quantitative measure of registration quality that results from AQUIRC lends itself to a label fusion technique. Therefore, we utilized an implementation of a weighted majority vote technique to fuse the labels of the multiple atlas segmentations. This was done by weighting the labels of each registered atlas using the ε values calculated in the AQUIRC algorithm. Utilizing

Equation (13), we assign $w_j = \varepsilon_i$ where ε_i is the estimated quality of the registration between atlas j and the target image. This process is then repeated for each target atlas. For example, in Figure 27, if I_1 is the target image, I_2 through I_5 are each registered to I_1 and their labels are projected to I_1 resulting in four observed segmentations. Each segmentation is then combined using Equation (13), weighting their labels as $w_2 = \varepsilon_1, w_3 = \varepsilon_2, w_4 = \varepsilon_5, w_5 = \varepsilon_8$.

As a comparison to the AQUIRC method, we utilized residual NMI with the same implementation of a weighted voting scheme, weighting the observed segmentations by the residual NMI between the registered atlas and the target image. That is, $w_j = NMI(\text{registered Atlas}, \text{target image})$, with j being the atlas that is used. Going back to Figure 27, if I_1 is again the target image, with I_2 through I_5 registered to I_1, the weighting of each of the segmentations using Equation (13) is $w_2 = NMI\left(T_{I_2 I_1}(I_2), I_1\right)$, $w_3 = NMI\left(T_{I_3 I_1}(I_3), I_1\right)$, $w_4 = NMI\left(T_{I_4 I_1}(I_4), I_1\right)$, $w_5 = NMI(T_{I_5 I_1}(I_5), I_1)$.

As a base line, we use a majority vote technique with equal weights. Each of the 10 images in the dataset is treated as a target image and the number of fused atlas-based segmentations is varied from 3 to 9. The order in which the atlases are fused is chosen randomly for each target image, but kept the same among each of the three methods. Figure 30 shows the average DICE score of all three methods, fusing 3 to 9 atlases, averaged over each of the 10 target images.

AQUIRC performs consistently better than both the un-weighted voting method and residual NMI weighted voting method. To show this, the DICE value for each individual fused atlas using both the residual NMI and un-weighted techniques are subtracted from the DICE value of the fused atlas of the AQUIRC method. Figure 31 contains the box plot of the DICE of the AQUIRC method subtracted by the DICE of the NMI weighted voting method and the pure

weighted voting method. The increase in DICE score is statistically significant compared to both the majority vote method as well as compared to the residual NMI weighted vote method, with $p < 0.001$.

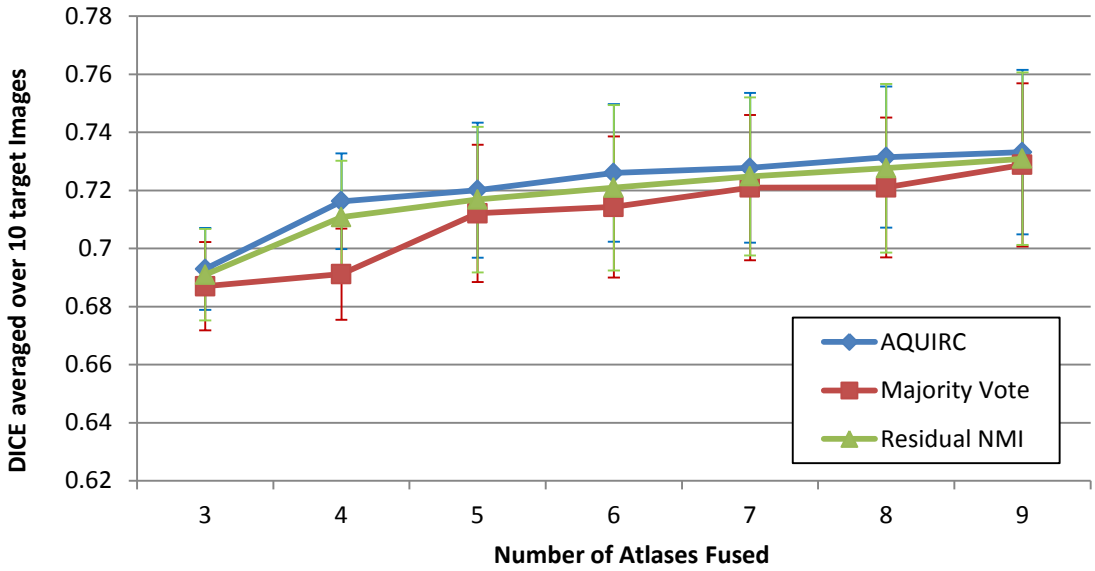


Figure 30: Average and standard deviation of the DICE values of the AQUIRC, Majority Vote and Residual NMI atlas fusion methods, starting with 3 atlases and increasing the number of atlases fused to 9.

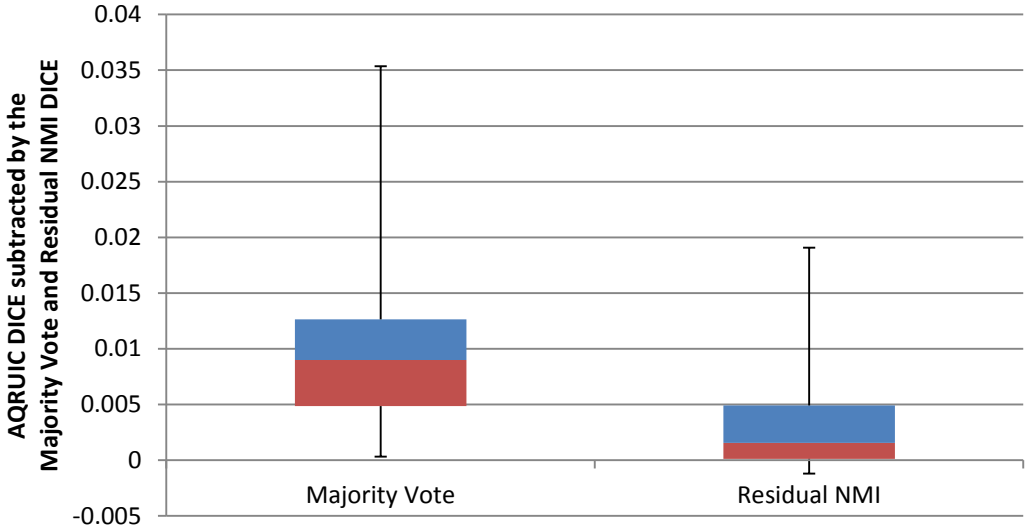


Figure 31: Box plot of the DICE value of the atlas-fused segmentation of AQUIRC subtracted by the DICE value of the atlas-fused segmentation of the Majority Vote and Residual NMI weighted techniques. The whiskers represent the minimum and maximum values. The red segment is the 2nd quartile and the blue segment is the 3rd quartile. The median is the line between the red and blue segments.

Conclusion

We have presented a new algorithm, AQUIRC, to measure global non-rigid registration quality. AQUIRC correlates highly with the DICE ground truth, more so than the popular technique of residual NMI. This suggests that AQUIRC is able to estimate the quality of a registration and that the algorithm can be applied to global atlas selection.

Used as an atlas fusion method, the AQUIRC algorithm outperformed residual NMI as well as an un-weighted majority vote method. Although the AQUIRC algorithm resulted in a small improvement, this improvement was statistically significant and consistent over both the majority vote method and residual NMI method - two popular methods for multi-atlas segmentation.

CHAPTER 7

Local Simulated Non-Rigid Registration Error

Introduction

In this chapter, AQUIRC is applied to estimate the magnitude and spatial location of local error in non-rigid registration. AQUIRC is shown to be capable of estimating the distribution and magnitude of registration error in a simulated case. Using clinical data, the algorithm also estimates registration error that is consistent with error that can be seen visually.

Applied AQUIRC

In this chapter, we calculate AQUIRC at the local level as in equation (5). We define $dissimilarity(\mathrm{X_i}, \mathrm{X_i}')$ to be the Euclidean distance between $\mathrm{X_i}$ and $\mathrm{X_i}'$, where $\mathrm{X_i}$ is the coordinate location of voxel i and $\mathrm{X_i}'$ is $\mathrm{X_i}$ transformed around a registration circuit. This is calculated at each voxel.

Image Information and Registration Method

In the experiments discussed in this chapter, we use 5 MR images that are T1-weighted sagittal volumes with 256x256x170 voxels that are 1 mm in each direction acquired with the parameters TE = 2.4 ms, TR = 12.2 ms. The images were acquired with the SENSE parallel imaging technique (T1 W/3D/TFE) from Philips on a 3 Tesla scanner. The images are registered utilizing the Adaptive Bases Algorithm (ABA) [52], which uses Normalized Mutual Information (NMI) [32] as its similarity measure. When calculating a circuit, $T_{\mathrm{AB}}(T_{\mathrm{BC}}(T_{\mathrm{CA}}(\mathbf{X})))$, the deformation fields are interpolated using tri-linear interpolation.

Experiments

In this chapter we perform three experiments. In the first experiment we simulate error and apply AQUIRC to estimate this error. We compare the estimated error and the known simulated error

visually and statistically. In the second experiment, we add simulated error to a non-rigid registration between two randomly chosen images. We then apply AQUIRC to estimate the error. Finally we estimate the error in a non-rigid registration between two images without adding any simulated error. For all experiments, five separate images are pair-wise non-rigidly registered together. We visually show AQUIRC 's overall magnitude estimation of error for each experiment.

Experiment 1

In the first experiment, we create a simulated image which is a deformed copy of an image from our data set. We call the original image I_1 and the deformed image I_{1d}. The simulated image is deformed using a radially symmetric growth model. There are three sizes of deformations that are utilized in the growth model in three different locations in the image. The largest deformation in the center of the image has a radius of 30 mm with a maximum displacement of 8 mm, the moderately sized deformation has a radius of 15 mm with a maximum displacement of 5 mm, and the smallest deformation has a radius of 10 mm with a maximum displacement of 3 mm. The simulated image and the growth model can be seen in Figure 32. The transformation between the simulated image and the image it was copied from is defined to be the identity transformation. This simulates a case in which the two images are mis-registered by the growth added to the deformed image. The other images in our data set are then non-rigidly registered to the simulated image, creating a network of six images as illustrated in Figure 33. AQUIRC is then applied, calculating a voxel-level error estimation of the simulated error. We calculate the correlation between the growth model and the error estimation and show a scatter plot of the values.

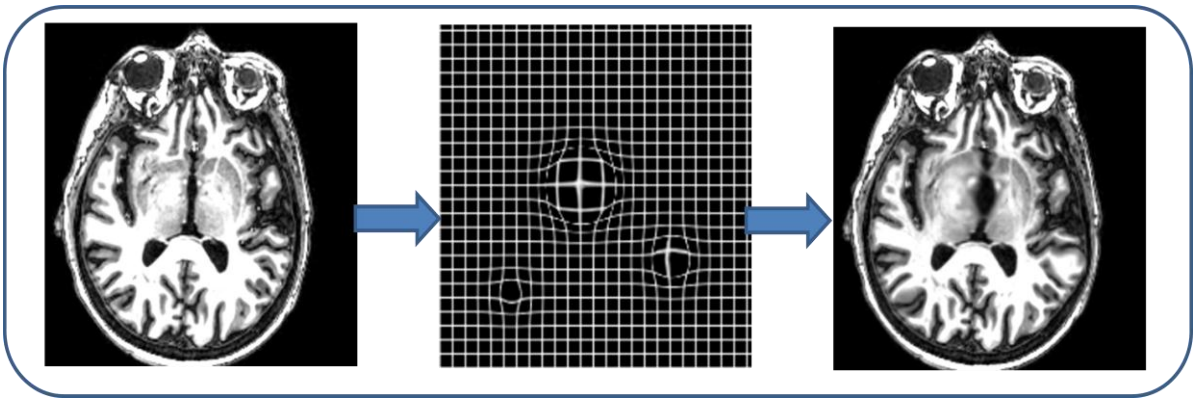


Figure 32: Left: Original image. Middle: Grid deformed by the simulated growth model. Right: Simulated image

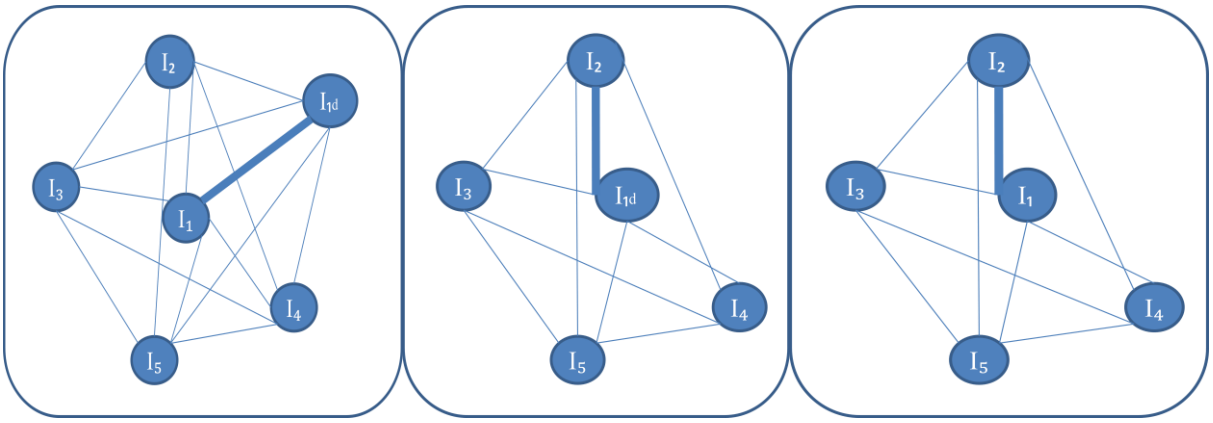


Figure 33: Diagram illustrating the 3 experiments performed. Left: Experiment 1, Middle: Experiment 2, Right: Experiment 3. The larger link represents the registration that each experiment is testing.

Experiment 2

In the second experiment, two images are chosen, which we call images I_1 and I_2. The images I_1 and I_2 are non-rigidly registered together. Then I_1 is deformed utilizing the same growth model that was used in Experiment 1, resulting in I_{1d}. The non-rigid registration between I_1 and I_2 is used for the transformation between I_{1d} and I_2, but the other images within the data set are re-registrered to I_{1d} as can be seen in Figure 33. Similar to Experiment 1, this simulates a case in which the growth model adds a known mis-registration, however in this experiment there is an unknown base mis-registration. AQUIRC is computed and we compare the estimated voxel-wise

error that AQUIRC detects with the simulated error. Because we do not know the base mis-registration, we present qualitative results for this experiment.

Experiment 3

In the third experiment, we run AQUIRC across the five reference set images. We consider images I_1 and I_2, and the error in the registration between them is estimated. This is illustrated in the right panel of Figure 33. Through this experiment we qualitatively investigate the ability of AQUIRC to localize mis-registration errors in practical situations.

Results

An example slice of the results obtained in Experiment 1 is shown in Figure 34. The left panel of the figure is the magnitude of the simulated error, the middle panel is magnitude of AQUIRC 's voxel-wise error estimation and the right panel is a grid that has been deformed by the simulated growth model overlayed on top of the error estimation. Figure 35 shows a magnified area where the moderate sized simulated error was added, alongside the original image and the deformed image.

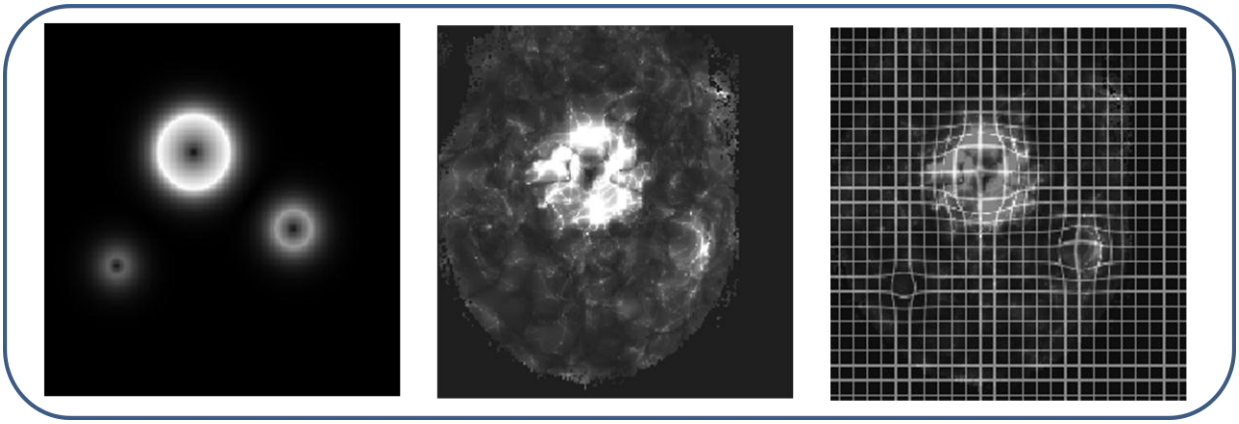


Figure 34: Left: magnitude of the growth model. Middle: Error estimation. Right: overlay of the error estimation onto a grid that has been deformed by the growth model.

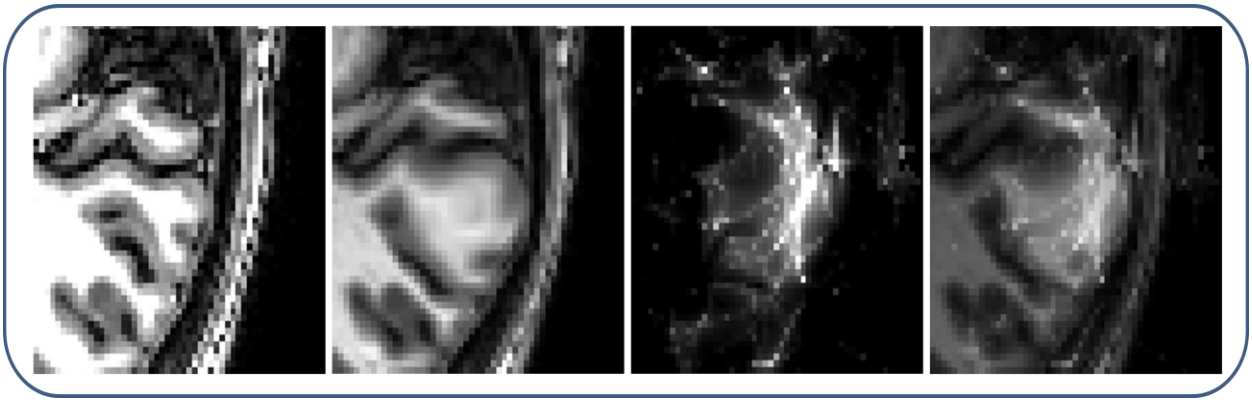


Figure 35: Left: the original image. Middle Left: region of moderate deformation. Middle Right: Error estimation. Right: overlay of the error estimation onto the deformed image.

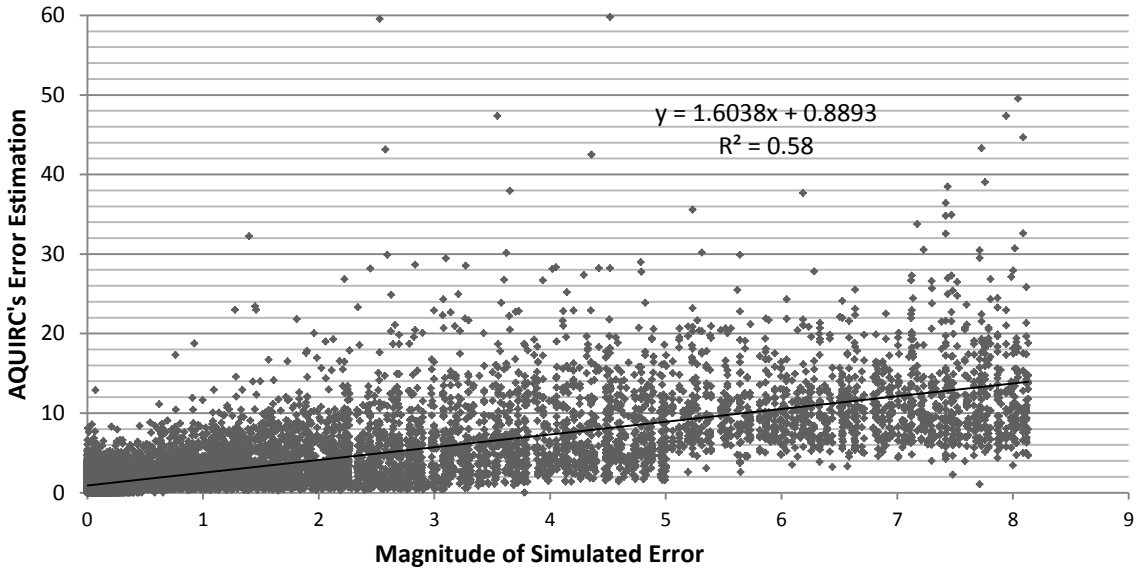


Figure 36: Scatter plot comparing the estimation of error to the magnitude of the simulated deformation field. Voxels within the head and in one transverse slice were included.

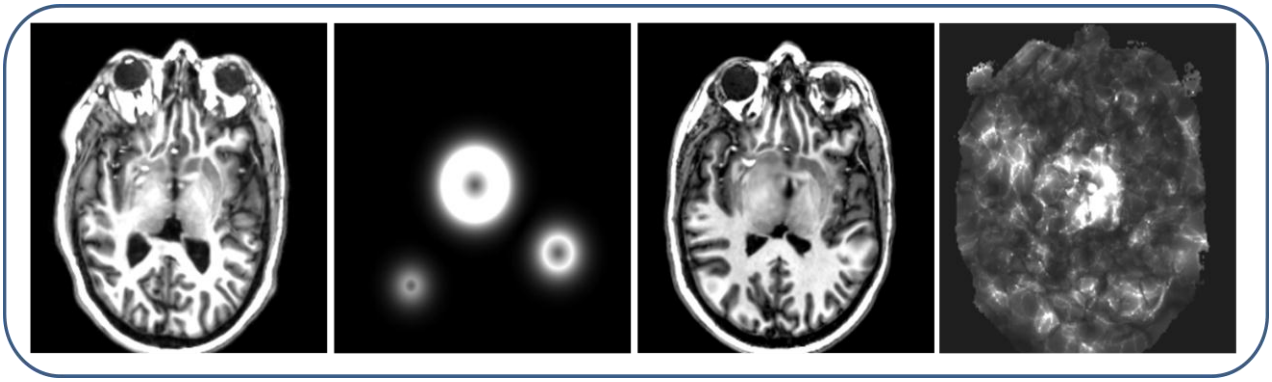


Figure 37: Left: $\mathbf{I_2}$ registered to $\mathbf{I_1}$. Middle Left: magnitude of the growth model. Middle Right: $\mathbf{I_{1d}}$. Right: Error estimation of the registration between $\mathbf{I_{1d}}$ and $\mathbf{I_2}$.

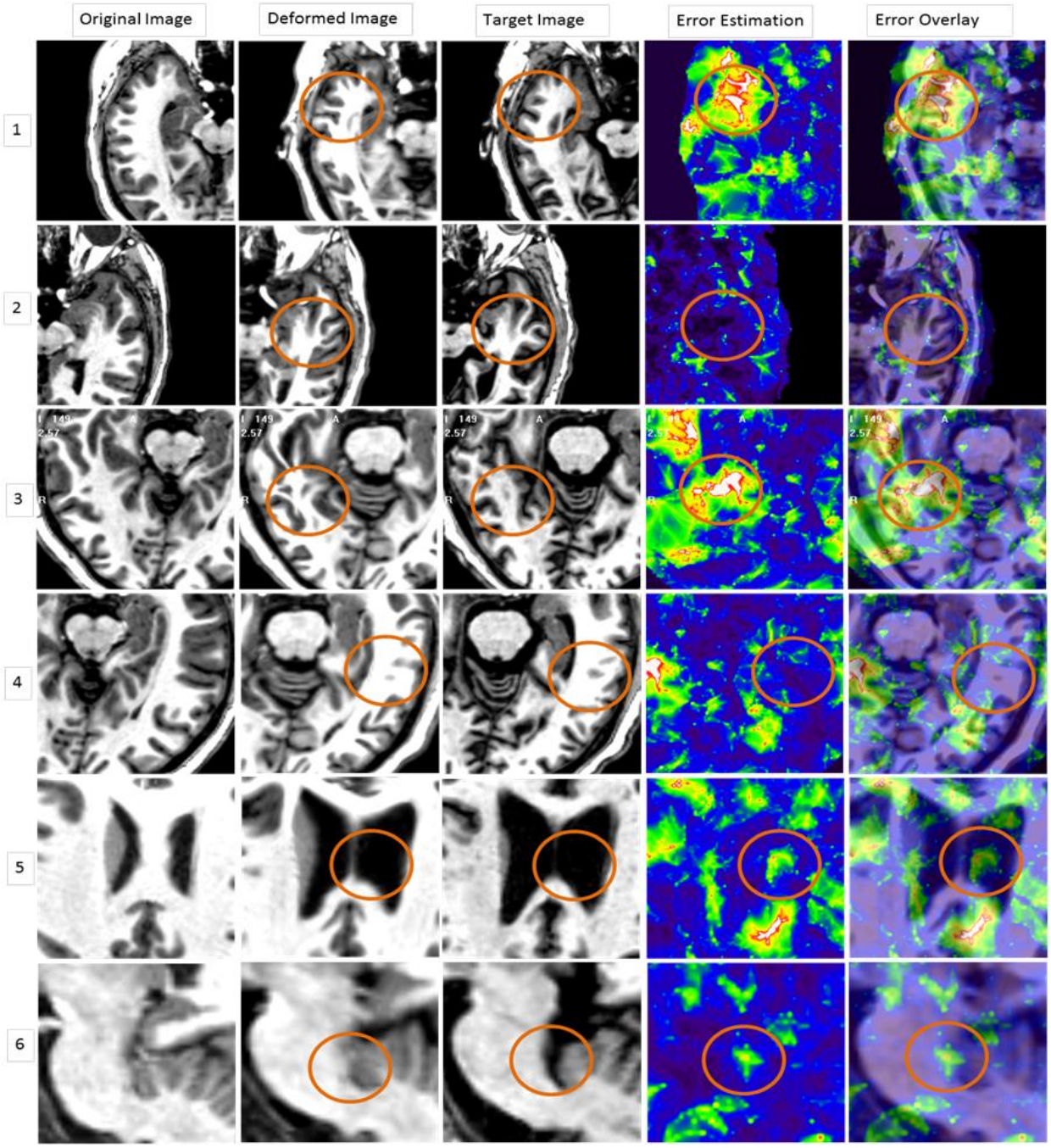


Figure 38: Row 1-5 shows the transverse plane and row 6 shows the sagittal plane. From left to right the column rows are the original image, the original image registered to the target, the target image, estimation of the error, and error overlayed on top of the original image registered to the target image. Areas of interest are highlighed with an orange circle.

We show a scatter plot of the magnitude of the simulated growth error and the AQUIRC's error estimation in Figure 36. The scatter plot utilizes data from one slice (the slice shown in Figure 34) since otherwise there are too many data points to produce a useful visualization of the results. We consider only voxels within the brain for both the scatter plot and correlation. AQUIRC was able to identify the simulated error well in the region of the large deformation and was able to identify some of the simulated error in the region of the moderate deformation; however, AQUIRC was unable to identify the simulated error in the region of the small deformation. The results from Experiment 2 are shown in Figure 37. The image on the left is I_2 registered to I_1, the left middle image is the magnitude of the growth model, the image in the middle right is the image I_{1d}, and on the right is the final error estimation of the registration between I_{1d} and I_2. As can be seen in Figure 37, the added simulated error is apparent in the center of AQUIRC's error estimation figure, although it is difficult to identify the medium and small growth deformations from the background registration error estimation. We also calculated the correlation between the AQUIRC's error estimation and the magnitude of the simulated growth that was added to the image, and find a reasonable correlation with an $r = 0.7008$ that is statistically significant with a $p < 0.001$.

The results from Experiment 3 are shown in Figure 38. In the very left column of Figure 38 we show the original image, followed by the original image registered to the target, followed by the target image, followed by AQUIRC's estimation of the error and finally the error overlayed on top of the original image registered to the target image. In rows 1-5 we show the results in the transverse plane and row 6 shows the results in the sagittal plane. We found that most of the highest registration error levels were found in the sulci of the brain, which is logical since there is a high level of inter-patient variability in sulci regions. In rows 1 and 2 we show

the left and right side of the head in the same image slice. AQUIRC estimates a high amount of registration error for the left side of the head, and a small amount of error for the right side. Rows 3 and 4 also show the left and right side of the same image slice, and we again find a high estimated error on the left side compared to a small amount of estimated error on the right side. Qualitatively, it appears that regions with low estimated error are better registered than the regions of high estimated error. In row 5 we show a region in the CSF where the registration appears correct, but that AQUIRC identifies as having some registration error. A potential cause of this is that in the CSF region there is no contrast to direct the registration, and without a consistent correspondence, AQUIRC may mis-identify error, which is expected. In row 6 of Figure 38 we see an example in the brainstem where visually we can identify a clear registration error that AQUIRC is also able to identify as being a registration error.

Conclusion

We have presented an algorithm to estimate voxel-wise registration error. In Experiment 1 AQUIRC is able to identify simulated errors, with a high correlation (r = 0.7008) between the estimated and actual error. In Experiment 2, AQUIRC is able to identify the large registration error that has been added to the center of an inter-patient registration. However, it is not obvious whether or not the moderate and small simulated deformations are identified by AQUIRC in Experiment 2. In Experiment 3, it appears that AQUIRC has correctly identified areas that are visually mis-registered, however, the ground truth is unknown in this instance.

Our results suggest that AQUIRC is able to identify regions of simulated error better when the growth model affects regions with contrast, which is expected. This can be seen in Figure 35 in the region of moderate growth where the simulated error affects a region with contrast as well as a homogeneous region. AQUIRC is not able to identify the simulated growth

in the homogeneous region but is able to identify the simulated growth in the area with contrast. When the growth region is larger and contains homogeneous regions as well as regions with contrast, AQUIRC is able to identify error in both types of regions.

CHAPTER 8

Validation of AQUIRC for Local Non-Rigid Registration Using Clinical Data

Introduction

In this chapter, AQUIRC's ability to detect local non-rigid registration errors using clinical data is tested and validated quantitatively at specific clinical landmarks, namely the Anterior Commissure (AC) and the Posterior Commissure (PC). To test AQUIRC on a representative range of error, 5 different registration methods are utilized, as well as 100 target images and 9 atlas images. The 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 the method to a more conventional approach, Local Normalized Correlation Coefficient (LNCC) is computed. AQUIRC is shown to perform similarly to LNCC. Importantly, it is shown that AQUIRC and LNCC provide complementary information and can be combined to improve the correlation with TRE. This chapter is concluded by showing that the AQUIRC algorithm can be used to reduce registration errors for all five algorithms.

Applied AQUIRC

In this chapter, we calculate AQUIRC at the local level as in equation (5). We define $dissimilarity(X_i,\ X_i')$ to be equal to the Euclidean distance between X_i and X_i', where X_i is the coordinate location of voxel i and X_i' is X_i transformed around a registration circuit. This is repeated for each voxel in the image.

Image Information and Registration Method

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 Set (1). Figure 39 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).

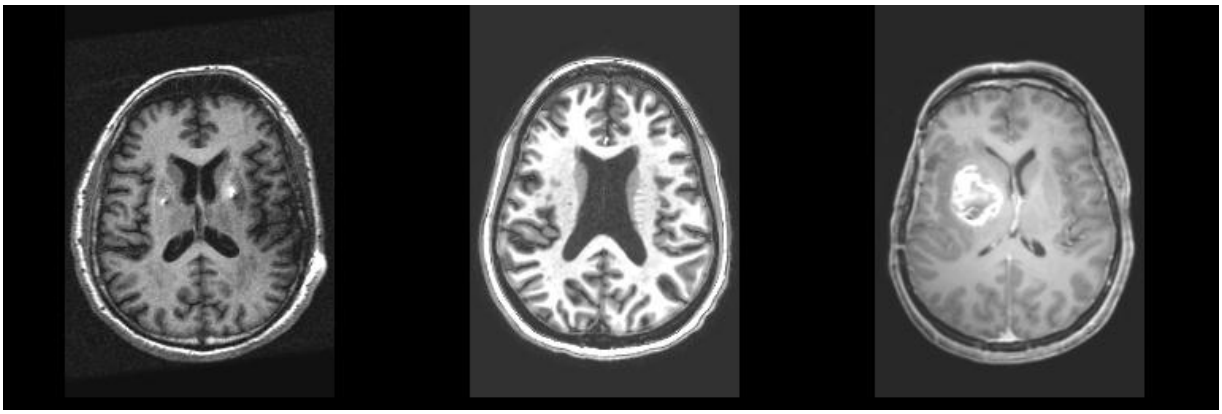


Figure 39: An example image of a patient with bilateral electrode implant (left), an example image without a discernible feature that would cause registration error (middle), and an example image of a patient with a large lesion (right).

All images are T1-weighted sagittal MR volumes that have 256x256x170 voxels and 1 mm in each direction. The images were acquired with the SENSE parallel imaging technique from Philips on a 3 Tesla scanner.

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) [52], 2) Diffeomorphic Demons Algorithm (Diff. Demons) [53], 3) Fast Free Form Deformation (F3D) [80], 4) Symmetric Normalization (SyN) [54], and 5) Automatic Registration Tools (ART) [81]. 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 [82]. As discussed above, because the goal of this work is not to compare registration algorithm's 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 [32] as its similarity metric and the results are visually checked to assure accuracy. If there was an inaccurate affine registration, the algorithm was re-initialized with an added rotation to an input image until the registration was determined to be accurate.

AC and PC Identification Rules

Two raters manually identified both the AC and PC points in the 113 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 [83], 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.

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 40.

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 at random, 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:

$$\|\boldsymbol{T_{AB}}(\boldsymbol{r_A}) - \boldsymbol{r_B}\| = \mathbf{TRE} \qquad (9)$$

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 $\varepsilon^{T_{AB}(r_A)}$. 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 $\varepsilon^{T_{AB}(r_A)}$ when $T_{AB}(r_A)$ does not fall on a grid point. We choose $\varepsilon^{T_{AB}(r_A)}$ 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^{r_A} to compare to TRE. We tested this location as well, and since the results were very similar to $\varepsilon^{T_{AB}(r_A)}$, we excluded the results of ε^{r_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: 4x4x4, 8x8x8, and 16x16x16 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 normal image volumes that are common to both of these two atlas sets, i.e., we compute the correlation between the TRE and ε values of 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) because it contains more images that are difficult to register.

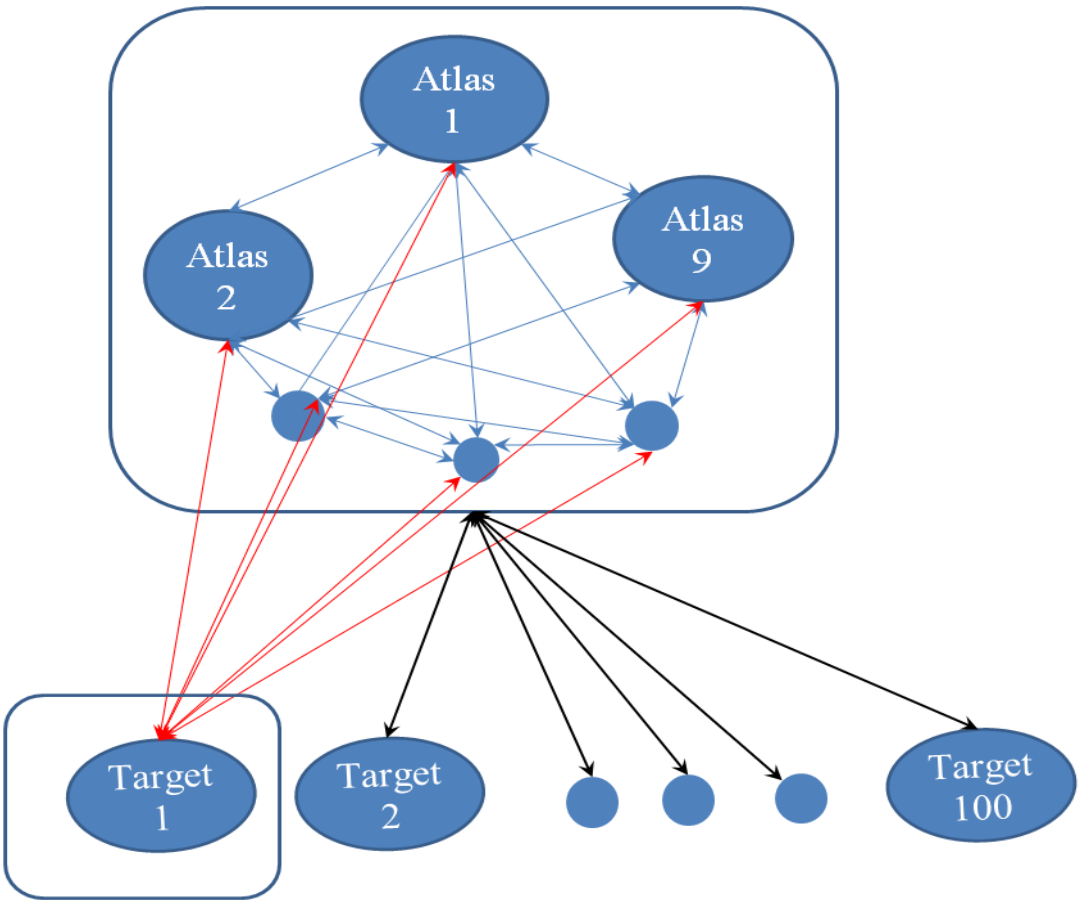


Figure 40: Illustration of the process used to calculate the AQUIRC-derived ε values. Nine atlases are registered pair-wise together and to each of the target images. AQUIRC is run for each target image, with a network of 10 images. This is repeated 5 times for each target image, once for each registration algorithm.

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 41 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 λ = 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 $\frac{\text{TRE}}{\varepsilon}$. 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.

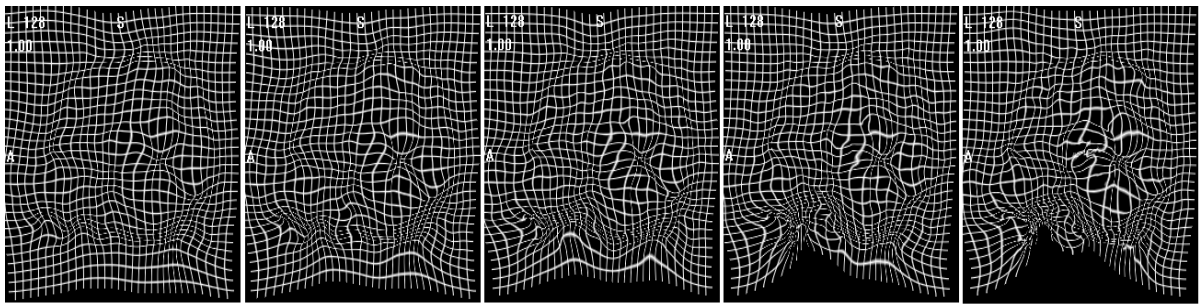


Figure 41: Example deformations that result from changing the stiffness parameter of algorithm A. From left to right: λ = 0.1, 0.15, 0.2, 0.25, and 0.3

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:

$$w(k) = \frac{e^{\frac{-\varepsilon_k}{\alpha}^{x(k)'}}}{\Sigma_{i=1}^{9}\left(e^{\frac{-\varepsilon_i}{\alpha}^{x(i)'}}\right)} \qquad (10)$$

$$x_{wa} = \sum_{k=1}^{9} w(k) * x'(k) \qquad (11)$$

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 value 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".

Results

Qualitative Results

Figure 42 illustrates qualitatively the results produced by our algorithm. The left panels show two 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 42 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 (i.e. by the identity transformation) 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 43 shows the average AQUIRC registration error maps of 2 atlases for each registration algorithm. Figure 43a is of an atlas of a patient with a large lesion while Figure 43b 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 43a) compared to the other example atlas (Figure 43b). Registering volumes with large space-occupying lesions is a notoriously difficult problem and can cause catastrophic registrations 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.

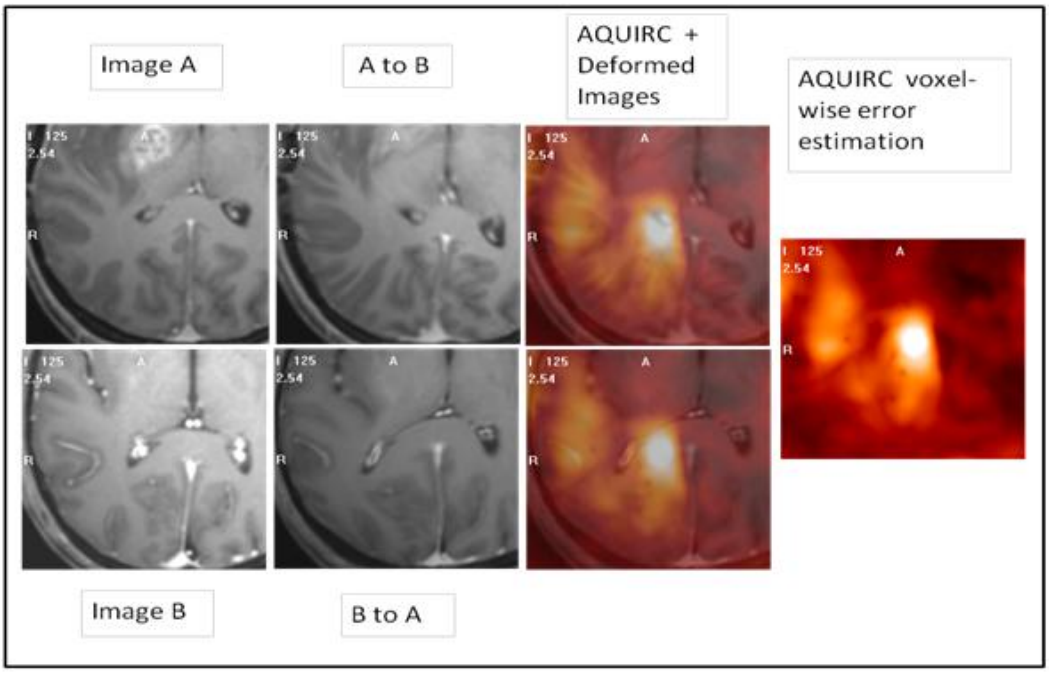


Figure 42: An example of AQUIRC's registration quality map, with error mapped on both the forward and reverse transformations.

Quantitative Results

Figure 44a 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). Figure 44b and Figure 44c 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). Figure 45a-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 E (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 detecting smaller errors with AQUIRC is more difficult than detecting large ones. Figure 46 shows a scatter plot of true TRE versus LNCC value. We show the results for LNCC using a region of interest of 8x8x8 as that volume resulted in the highest

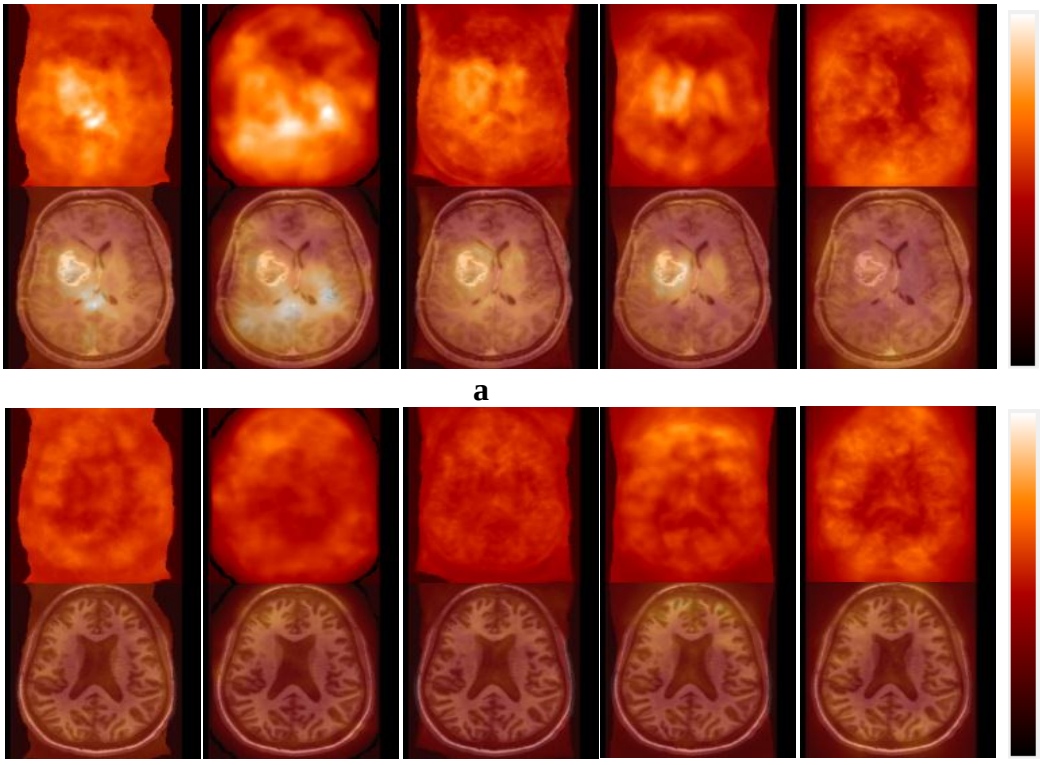


Figure 43: Average AQUIRC error maps from the 100 images in our dataset, using an atlas with a large lesion (a), and an atlas without any discernible features that might cause a registration problem (b). The error maps are averaged for each registration algorithm, which are, from left to right, algorithms 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.

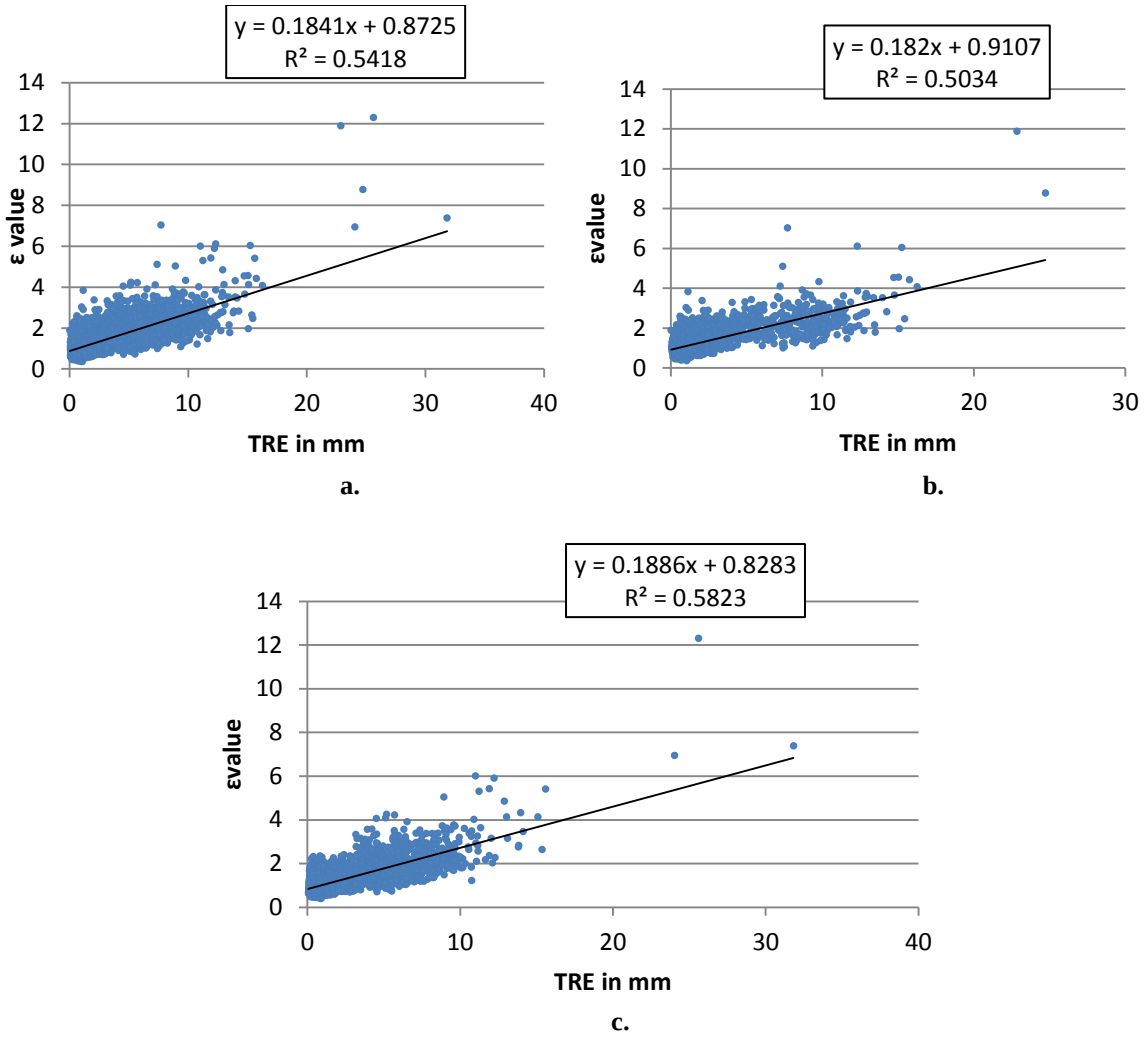
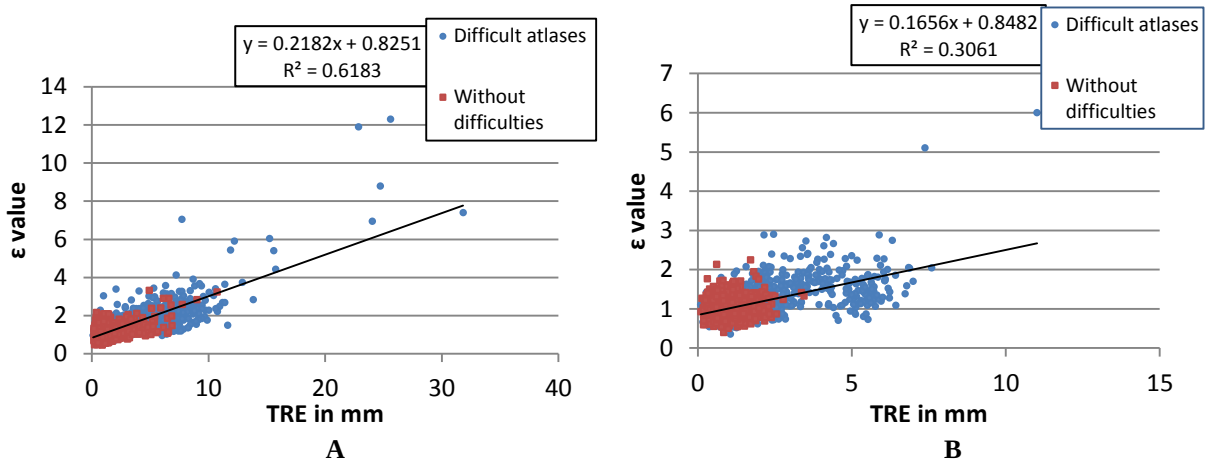


Figure 44: Scatter plot of the 9000 points using atlas set (1), using all 9 atlases and all five registration algorithms for (a) both the AC and PC, (b) AC only and (c) PC only.



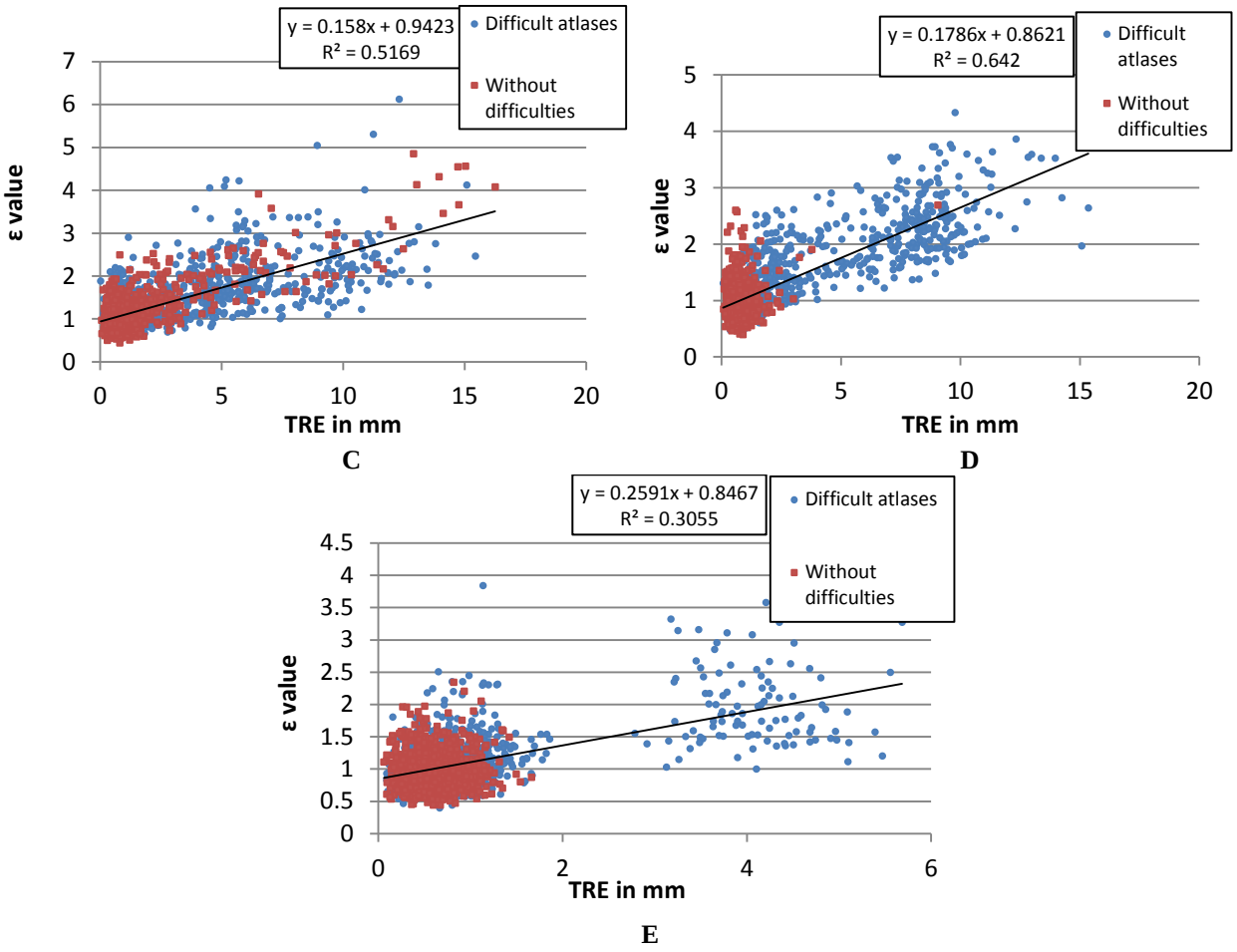


Figure 45: Scatter plot of the AC and PC points using algorithm A, B**,** C, D, and E on atlas set (1). The red square points are the atlases without any discernible feature that would cause registration error and the blue circles are the difficult to register atlases.

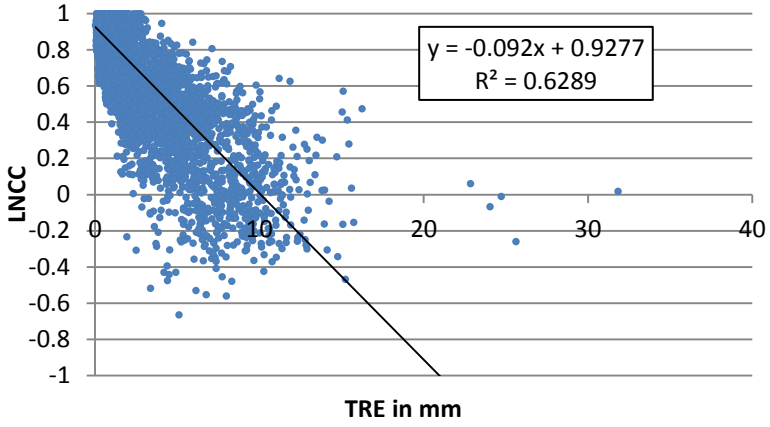


Figure 46: Scatter plot of LNCC compared to TRE for both the AC and PC.

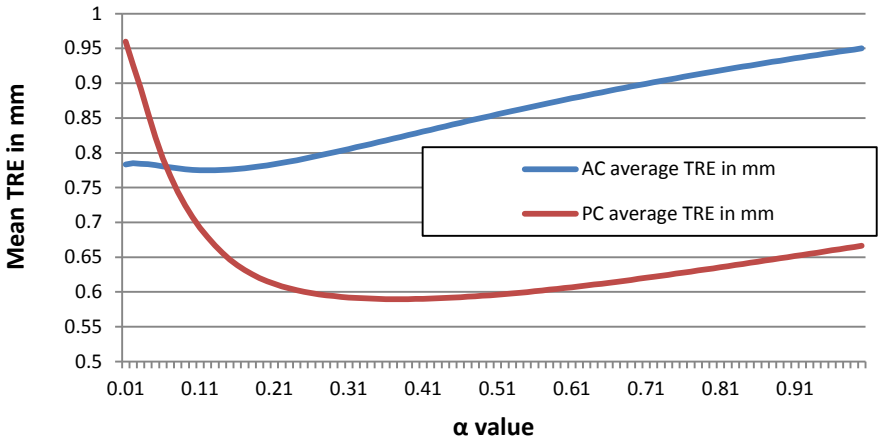


Figure 47: Mean TRE at the AC and PC as the weighting factor α is varied from 0.01 to 1.

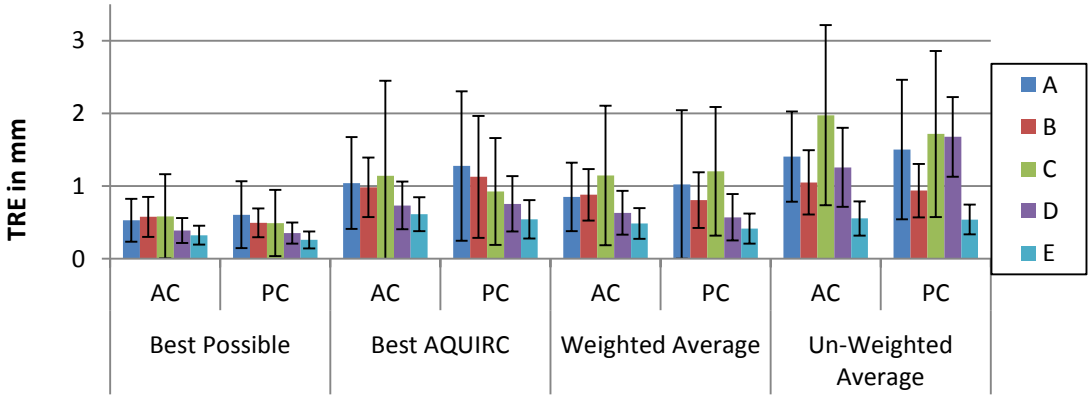


Figure 48: Reduction in registration error when using AQUIRC for each registration algorithm A, B, C, D, and E.

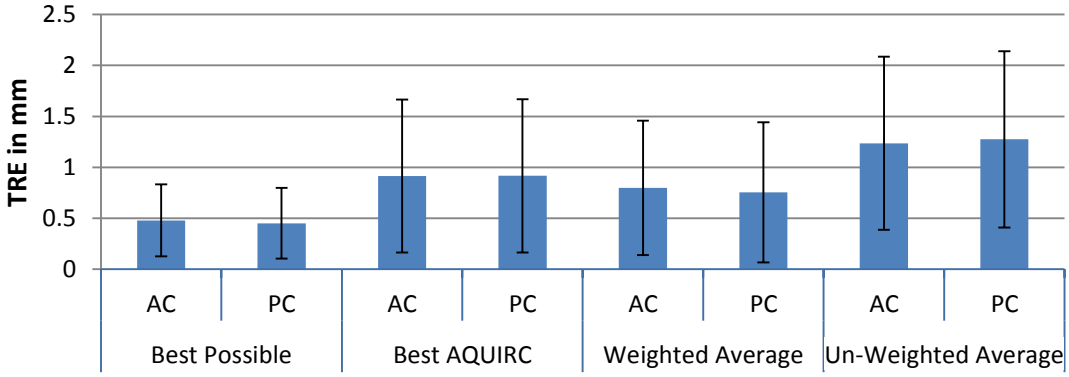


Figure 49: Reduction of registration error when using AQUIRC for all registration algorithms combined. From left to right, this plot shows the Best Possible, the Best AQUIRC, the Weighted Average, and the Un-Weighted Average results.

Table 2: R^2 values for the correlation between TRE and the results of AQUIRC, LNNC and both methods combined 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
Combined	0.79	0.57	0.75	0.82	0.52

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 with both combined in Table 2. We find that combining the results of AQUIRC with LNCC consistently improves the correlation with TRE for each registration algorithm.

The correlation using the low quality atlas set compared to using 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 47. Based on these results, 0.25 was used as the α value to generate the results on the testing set. Figure 48 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 48a shows the TRE reduction that AQUIRC was able to achieve for each registration algorithm separately while Figure 48b 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.

Conclusion

This chapter applies AQUIRC to estimate the local quality of a non-rigid registration using clinical data. The results we have obtained show that AQUIRC has merit. It is simple to implement and relatively cheap to execute. For instance, the pair-wise registrations between atlases and all registration circuits involving only the atlases can be pre-computed. The matrix $\overline{\overline{P}}$ can also be pre-computed as well as its pseudo-inverse. To compute AQUIRC for a new image volume $\overline{\overline{E}}_c$ 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., $\left(\overline{\overline{P}}^T\overline{\overline{P}}\right)^{-1}\overline{\overline{P}}^T\log(\overline{\overline{E}}_c)$. The

size of $\overline{\overline{P}}$ is $\left(\frac{N}{3}\right) x \left(\frac{N}{2}\right)$ and is thus determined by the number of images used in the network while the size of $\overline{\overline{E}}_c$ 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, $\overline{\overline{E}}_c$, 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 $\overline{\overline{\mathrm{P}}}$, 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 $\overline{\overline{P}}$ 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 [79].

As shown in Figure 45, 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 1.56 mm; it is 0.84 mm 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 of the scenarios for which an intensity-based method would provide the strongest results. Because AQUIRC relies only on the deformation fields, it is less susceptible to intensity-related image properties. We thus suspect that AQUIRC would be at an advantage relative to LNCC 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 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.

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.

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 [84]. These authors utilize information about the deformations to update the error model in their equivalent to $\overline{\overline{P}}$ 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 $\overline{\overline{P}}$ matrix in our work depends only on the network of images, i.e., the number of nodes and the number of circuits; their equivalent to our $\overline{\overline{P}}$ matrix in [84] is a sparse matrix of size $\sim\left(\frac{N}{3}\right) * Q$ x $\left(\frac{N}{2}\right) * Q$ with Q the number of voxels in the image. Thus the more realistic error model in [84] is difficult to scale up and to apply at the voxel level to 3D images.

Muenzing in [41] 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.

CHAPTER 9

Applying AQUIRC to Multi-Atlas Segmentation

Introduction

Multi-atlas registration-based segmentation is a popular technique in the medical imaging community, used to transform anatomical and functional information from a set of atlases onto a new patient that lacks this information. The accuracy of the projected information on the target image is dependent on the quality of the registrations between the atlas images and the target image. In this chapter, AQUIRC is applied to atlas selection at the local level across multiple structures in cases in which non-rigid registration is difficult. AQUIRC is applied to 6 structures, the brainstem, optic chiasm, left and right optic nerves, and the left and right eyes. The results of AQUIRC are compared to that of popular techniques, including Majority Vote, STAPLE, Non-Local STAPLE, and Locally-Weighted Vote. AQUIRC is shown to be useful as a method to combine multiple segmentations and increase the accuracy of the projected information on a target image, and is comparable to cutting edge methods in the multi-atlas segmentation field.

Applied AQUIRC

In this chapter, we calculate AQUIRC at the local level as in equation (5). We define $dissimilarity(X_i, X_i')$ to be equal to the Euclidean distance between X_i and X_i', where X_i is the coordinate location of voxel i and X_i' is X_i transformed around a registration circuit. This is repeated for each voxel in the image.

Image Information and Registration Method

Our data set consists of 20 clinical images with high grade gliomas. Figure 50 shows a representative slice of an image with a high grade glioma. All images in our data set are T1-weighted sagittal MR image volumes with approximately 256x256x170 voxels and 1 mm in each direction acquired with the parameters TR = 7.92 ms, TE = 3.65 ms. The images were acquired with the SENSE parallel imaging technique (T1 W/3D/TFE) from Philips on a 3 Tesla scanner. The 20 images were segmented by 8 expert raters. The raters were asked to delineate the brainstem, optic chiasm, eyes and optic nerves for 20 cases utilizing fused CT/MR imaging within a clinical system. A discussion of the data-set and the ground truth segmentation used in this paper can be found in [6].

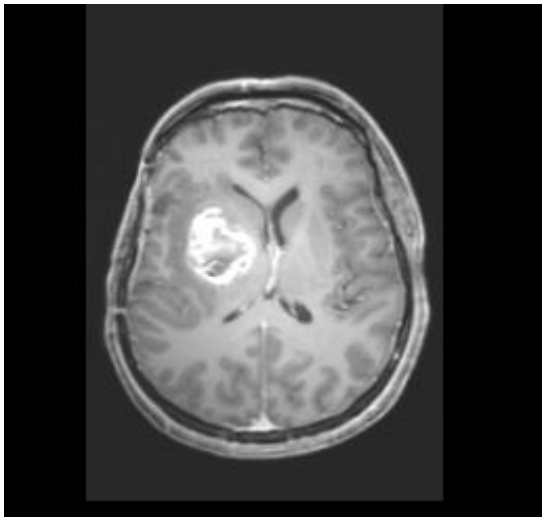


Figure 50: Example MR with large lesion

To create the ground truth, we performed a variant of majority vote on the segmentations of the 8 expert raters. This method, termed p-maps [85] and applied to this data-set in [6] was used since, in some cases, there was a large disagreement among the expert raters. Using majority vote in these cases would lead the ground truth to be under-segmented which can greatly affect the results in the cases of the small structures. The p-maps were created by

summing the binary masks of each rater for a particular structure, and then thresholding the p-map at the level of the mean value of all non-zero voxels. That is, each voxel with a value greater than or equal to the mean was included in the ground truth binary segmentation. The non-rigid registration algorithm used in this work is the Adaptive Basis Algorithm (ABA) [52], which uses Normalized Mutual Information (NMI) [32] as its similarity measure.

Atlas Selection Techniques

In this work we test 6 methods: 1) Majority Vote, 2) STAPLE [15], 3) Non-Local STAPLE [20], 4) Locally-Weighted Vote [7], 5) Un-Weighted AQUIRC and 6) Weighted AQUIRC. For methods 1), 5) and 6) we fuse the segmentations as:

$$\psi_i = \arg\max_s \sum_j I(D_{ij} = s) * w_{ij} \qquad (14)$$

where ψ_i is the estimated label for voxel i, s is the list of labels, D_{ij} is the decision of rater (in this case, atlas segmentation) j, I is the indicator function, and w_{ij} is the assigned weight of rater j at voxel i. Here, equation (14) is similar to equation (13) except that it allows for local weighting. For Majority Vote w_{ij} was set equal to one for every rater. When using the Un-Weighted AQUIRC method we include only the 10 raters (i.e. atlas segmentations) with the lowest ε^i value and set $w_{ij} = 1$. This provides an un-weighted majority vote segmentation of the 10 atlases that AQUIRC identifies as having the best registration quality. The AQUIRC-Weighted voting scheme also utilizes the top ten atlases with the lowest ε^i value for each voxel but with $w_{ij} = e^{-\varepsilon_j^i}$.

For STAPLE [15], label probabilities were initialized using a global prior. For Non-Local STAPLE [20], the label probabilities were initialized using a global prior, the cost

function used was locally normalized correlation coefficient (LNCC), the cost function's standard deviation was 0.35 with a patch radius of 4 mm isotropic and a search radius of 1 mm isotropic. For Locally-Weighted Vote [7] the cost function used was LNCC, with a standard deviation of 0.35 and a radius of 4 mm isotropic.

Experiments

For each image we perform a multi-atlas segmentation treating 19 images as atlases, in a leave-one-out scheme. We combine the atlases using the six atlas selection methods described above. To calculate the quality of a segmentaiton, we compute a DICE score between the manual segmentation on the target, and the multi-atlas segmentation that has been transformed onto the target. If we define X as the manual segmentation on a target image, and X' as the multi-atlas segmentation, we can compute DICE as in equation (8)

Results

The results from the experiment are shown in a box plot in Figure 51. The box plots from left to right are Majority Vote (blue), STAPLE (red), Locally Weighted Vote (green), Non-Local STAPLE (purple), Un-weighted AQUIRC (light blue), and Weighted AQUIRC (black). AQUIRC performs statistically the same as the cutting edge techniques for every case, except the Left Nerve, where Non-Local STAPLE is statistically better. AQUIRC is significantly better compared to Majority Vote and STAPLE for the brainstem. AQUIRC's performance relative to the other methods is strongest in the Brainstem and Chiasm. A closer look at the results of AQUIRC's estimation of registration error shows that the registrations for the Brainstem and Chiasm have a higher average ε value compared to the other structures (1.43 and 1.42 versus 1.29, 1.26, 1.25, and 1.23).

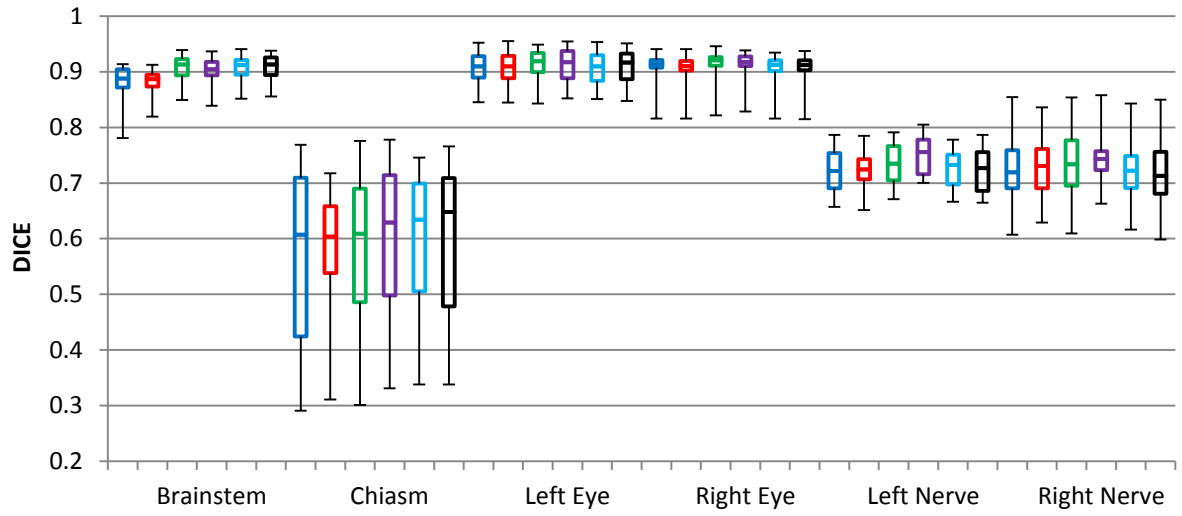


Figure 51: A box plot of the results. From left to right are Majority Vote (blue), STAPLE (red), Locally Weighted Vote (green), Non-Local STAPLE (purple), Un-weighted AQUIRC (light blue), and Weighted AQUIRC (black).

Figure 52 illustrates qualitatively the results produced by our algorithm. The top panels show the original images. The next panels show the deformation fields computed by one of the registration algorithms to register the images to each other by means of a deformed grid. The third panels show the images reformatted with these deformation fields. The fourth panels show the AQUIRC-derived error map superimposed on the reformatted images shown on the third panels. The fifth panels show the AQUIRC-derived error map superimposed on the deformed grid shown on the second panels. Figure 52 shows an example in which the brainstem is overly stretched and conversely compressed in the transformation. AQUIRC correctly identifies the location of the error.

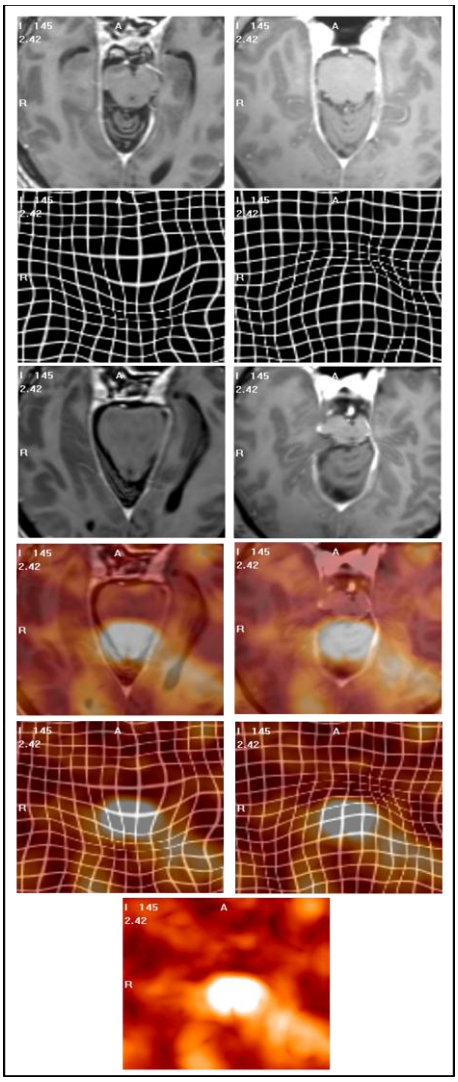


Figure 52: An example of AQUIRC's registration quality map, with error mapped onto both the forward and reverse transformations.

Conclusions

In this chapter, we have shown that the AQUIRC algorithm can be useful in the context of multi-atlas selection. The AQUIRC algorithm performed statistically better than both Majority Vote and STAPLE when segmenting the brainstem. AQUIRC also performed comparably to popular techniques such as Non-Local STAPLE and Locally-Weighted Vote, with only Non-Local STAPLE performing better than AQUIRC for the Left Optic Nerve. In the structures in which AQUIRC performs the best, the average ε value is higher. If indeed AQUIRC is estimating

registration error, this suggests that the algorithm performs better in cases in which there is a high registration error. It is logical that the algorithm would perform best in instances in which there is a higher registration error. This is because there are other reasons that the multi-atlas segmentation can be degraded (e.g. poor manual segmentations), and this algorithm is considering only registration error, so when registration error is the main source of error, AQUIRC should perform better in comparison to other algorithms that use other sources of information.

The AQUIRC algorithm does not have any parameters that need to be optimized unlike STAPLE, Non-Local STAPLE and Locally-Weighted Vote. AQUIRC is also unique as it estimates error only from the deformation fields that result from an image registration, and does not explicitly utilize any information from the segmentations of the structures and the intensity of the images. Non-Local STAPLE and STAPLE conversely use the information provided from the segmentation of the structures and Non-Local STAPLE and Locally-Weighted Vote use intensity information. AQUIRC thus uses separate and complementary information compared to Non-Local STAPLE and Locally-Weighted Vote. In the future, we aim to create a framework that utilizes all sources of information, as we believe this will lead to an increase in accuracy for multi-atlas based segmentation.

CHAPTER 10

Conclusions and Contributions

Identifying registration error is of critical importance in the field of medical imaging. This dissertation presents AQUIRC, a method that uses redundancy from multiple registration circuits to estimate the error of intensity-based and fiducial-based registration. This dissertation also presents an assessment of the capabilities of the AQUIRC algorithm. AQUIRC has been applied to estimate error for fiducial-based registration, rigid intensity-based registration, global intensity-based registration, local simulated non-rigid registration, local non-rigid registration at the AC and PC, and local non-rigid registration applied to atlas selection. AQUIRC was shown to better estimate error for intensity-based rigid registration compared to MI and NMI, to outperform NMI for global atlas selection, to correlate well with simulated non-rigid registration error, to correlate and reduce error at the AC and PC, and to perform comparably to cutting edge local atlas selection methods. While AQUIRC was shown to correlate to TRE for fiducial-based registration, a practical use for this finding has not been determined.

One instance in which AQUIRC is not 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. In that case, no registration circuit would not identify the biased error because the error is added to, and then subtracted from, the combination of the transformations. The possibility of biased error needs to be considered for any application in which accuracy is important.

There are still open questions concerning AQUIRC. An area of future interest concerns the number and length of registration circuits necessary to compute the AQUIRC algorithm.

While a small amount of work has been done concerning the length of registration circuits and how that affects the accuracy of the error estimation [66], there is much left to be done. An interesting question is whether using registration circuits greater than length three provides extra information and better results compared to = using registration circuits exclusively of size three. It is also possible that using more registration circuits of greater size allows for a smaller *N* with equivalent accuracy. A related question is how many registration circuits are sufficient to estimate registration error. In this work, we utilize all unique registration circuits $\binom{N}{3}$. However, for large *N*, the resulting calculations become infeasible to compute. Therefore, an open question is whether there is a subset of registration circuits that would be sufficient to estimate registration error which would alleviate the computation overhead. In [86] it was found that some subset of all the unique registration circuits still provided useful informaiton to their registration algorithm, yet for most instances this is still an open question. It would also be interesting to consider the information gained from using multiple registration algorithms and how that extra information could be applied to AQUIRC. With multiple registration algorithms, circuit of size two would be possible, as well as many other possible combinations would be available.

From Chapter 8 we found that AQUIRC is applicable across five separate registration algorithms and applicable across varying amounts of deformation stiffness. We do know, however, that if all transformations are so constrained to be the Identity, the AQUIRC algorithm would fail. Thus, another open question is what level of stiffness can be allowed that still provides enough information for AQUIRC to produce accurate results. AQUIRC needs a level of correspondence to perform well, and identifying the edge cases of that boundary is an interesting open question. We also found that the results from AQUIRC were consistent across the five registration algorithms, however, it is possible that the uncertainty measure may need to be

calibrated for other registration algorithms. While this work suggests that such calibration is not needed, evaluation on other algorithms is important future work. Identifying a method that would allow for this calibration for each registration algorithm is also of interest.

It is notable that methods inspired by AQUIRC have already been used by others with success. Donoghue et al. [86] 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 multiple registrations and show that it leads to lower registration error. Goksel et al. [87] developed a variant of the AQUIRC algorithm proposed in [68] (discussed in Chapter 6), for global atlas selection. Goksel's variaent slightly increased the correlation and speed and reduced the memory needed for computation compared to AQUIRC. Goksel et al. [87] showed that AQUIRC and their variant of AQUIRC correlates highly with DICE: 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. A more complex error model using the AQUIRC framework was recently introduced by Gass et al. in [84] where they utilize information about the deformations for a more accurate model of registration error, but this requires orders of magnitude more memory to compute. This work contributes to the field as it presents, to the best of our knowledge, the first algorithm to utilize the redundancy of registration circuits to estimate the error for an individual registration.Thus, the work presented in this dissertation has contributed to the field of image registration in two ways. First, it has provided the first algorithm that utilizes sets of registration circuits to exploit redundancy among multiple registrations in order to estimate the error of a single registration. Second, it has inspired new work and new methods that expand and improve on our initial algorithm with the future promise of even greater knowledge of the limits of registration accuracy.

REFERENCES

[1] J. Fitzpatrick, D. L. Hill and C. R. Maurer Jr., Image registration, Handbook of Medical Imaging, vol. 2, Bellingham, WA: SPIE Press, 2000.

[2] M. Holden, "A review of geometric transformations for nonrigid body registration," *IEEE Trans. Med. Imaging,* vol. 27, pp. 111-128 , 2008.

[3] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image Vis. Comput,* vol. 21, no. 11, pp. 977-1000, 2003.

[4] D. L. G. Hill, P. G. Batchelor, M. Holden and D. J. Hawke, "Medical image registration," *Physics in Medicine and Biology,* vol. 46, no. 3, 2001.

[5] A. Chen, K. Niermann and M. a. D. B. Deeley, "Evaluation of multi atlas-based approaches for the segmentation of the thyroid gland in IMRT head-and-neck CT images," *Phys. Med. Biol.,* vol. 57, pp. 11-93, 2011.

[6] M. e. a. Deeley, "Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: a multi-expert study," *Phys. Med. Biol.* , vol. 56, pp. 4557-4577, 2011.

[7] M. Sabuncu, B. T. T. Yeo, K. Van Leemput, B. Fischl and P. Golland, "A generative model for image segmentation based on label fusion," *IEEE Transactions on Medical Imaging* , vol. 29, pp. 1714-1729, 2010.

[8] O. Commowick, S. Warfield and G. Malandain, "Using Frankenstein's creature paradigm to build a patient specific atlas," *MICCAI,* pp. 993-1000, 2009.

[9] T. Rohlfing, R. Brandt, R. Menzel and C. Maurer, "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains," *NeuroImage* , vol. 21, no. 4, pp. 1428-1442, 2004.

[10] X. Artaechevarria, A. Muñoz-Barrutia and C. Ortiz-de-Solorzano, "Combination strategies in multi-atlas image segmentation: application to brain MR data," *IEEE Transactions on Medical Imaging* , vol. 28, p. 1266–1277, 2009.

[11] I. Isgum, M. Staring, A. Rutten, M. Prokop, M. Viergever and B. van Ginneken, "Multi-atlas-based segmentation with local decision fusion—application to cardiac and aortic segmentation in CT scans," *IEEE Transactions on Medical Imaging* , vol. 28, pp. 1000-1010, 2009.

[12] M. Sabuncu, B. Yeo, K. Van Leemput, B. Fischl and P. Golland, " A generative model for image segmentation based on label fusion," *IEEE Transactions on Medical Imaging ,* vol. 29, pp. 1714-

1729, 2010.

[13] H. Wang, J. Suh, S. Das, J. Pluta, C. Craige and P. Yushkevich, "Multi-Atlas Segmentation with Joint Label Fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2012.

[14] P. Coupé, J. Manjķn, V. Fonov, J. Pruessner, M. Robles and D. Collins, "Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation," *Neuroimage,* vol. 54, pp. 940-954, 2011.

[15] S. Warfield, K. Zou and W. Wells, "Simultaneous truth and performance level estimation (STAPLE): and algorithm for the validation of image segmentation," *IEEE TMI ,* vol. 23, no. 7, pp. 903-921, 2004.

[16] A. Asman and B. Landman, "Non-local statistical label fusion for multi-atlas segmentation," *Medical Image Analysis,* vol. 17, no. 2, pp. 194-208, 2013.

[17] M. Cardoso, K. Leung, M. Modat, J. Barnes and S. Ourselin, "Locally Ranked STAPLE for Template based Segmentation Propagation," *MICCAI Workshop on Multi-Atlas Labeling and Statistical Fusion,* vol. 25, p. 25–26, 2011.

[18] A. Asman and B. Landman, "Characterizing spatially varying performance to improve multi-atlas multi-label segmentation," in *Information Processing in Medical Imaging (IPMI)*, 2011.

[19] A. Asman and B. Landman, " Robust statistical label fusion through consensus level, labeler accuracy and truth estimation (COLLATE)," *IEEE Transactions on Medical Imaging ,* vol. 30, pp. 1779-1794, 2011.

[20] A. Asman and B. Landman, "Non-local STAPLE: an intensity-driven multi- atlas rater model," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2012.

[21] T. Rohlfing, D. Russakoff and C. Maurer, "Performance-based classifier combination in atlas-based image segmentation using expectation–maximization parameter estimation," *IEEE Transactions on Medical Imaging ,* vol. 23, pp. 983-994, 2004.

[22] O. Commowick and S. Warfield, " Incorporating priors on expert performance parameters or segmentation validation and label fusion: a maximum a posteriori STAPLE," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI ,* 2010.

[23] J. Fitzpatrick, J. West and C. Maurer Jr., "Predicting error in rigid-body point-based registration," *IEEE Transactions on Medical Imaging ,* vol. 17, pp. 694-702, 1998.

[24] J. Fitzpatrick and J. West, "The distribution of target registration error in rigid-body point-based registration," *IEEE Transactions on Medical Imaging ,* vol. 20, pp. 917-927, 2001.

[25] J. M. Fitzpatrick, "Fiducial registration error and target registration error are uncorrelated," in *SPIE*

Medical Imaging , Lake Buena Vista, FL, 2009.

[26] D. Keysers and W. Unger, "Elastic Image Matching is NP-complete," *Pattern Recognition Letters,* vol. 24, pp. 445-453, 2003.

[27] T. E. Bender and W. A. Tome, "The utilization of consistency metrics for error analysis in deformable image registration," *Phys. Med. Biol,* vol. 54, pp. 5561-5577, 2009.

[28] T. Rohlfing, " Image Similarity and Tissue Overlaps as Surrogates for Image Registration Accuracy: Widely Used but Unreliable," *IEEE Transactions on Medical Imaging,* vol. 99, 2011.

[29] J. West et al., "Comparison and evaluation of retrospective intermodality brain image registration techniques," *Journal of Computer Assisted Tomography,* vol. 21, pp. 554-566, 1997.

[30] W. L. Crum, D. Hill and D. Hawkes, " Zen and the art of medical image registration: correspondence, homology, and quality," *Neuroimage,* vol. 20, no. 3, pp. 1425-1437, 2003.

[31] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging,* vol. 16, no. 2, pp. 187-198, 1997.

[32] C. Studholme, D. Hill and D. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognition,* vol. 32, no. 1, pp. 71-86, 1999.

[33] T. Veninga, H. Huisman, R. van der Maazen and H. H., "Clinical validation of the normalized mutual information method for registration of CT and MR images in radiotherapy of brain tumors," *J. Appl. Clin. Med. Phys,* vol. 5, pp. 66-79, 2004.

[34] J. M. Fitzpatrick, D. Hill, Y. Shyr, J. West, C. Studholme and C. Maure Jr, "Visual assessment of the accuracy of retrospective registration of MR and CT images of the brain," *IEEE Trans. Med. Imaging* , vol. 17, pp. 571-585, 1998.

[35] I. J. Simpson, J. A. Schnabel, A. R. Groves, J. L. Andersson and M. W. Woolrich, "Probabilistic inference of regularisation in non-rigid registration," *NeuroImage* , vol. 17, no. 5, pp. 2438-2451, 2012.

[36] J. Iglesias, M. Sabuncu and K. van Leemput, " Improved Inference in Bayesian Segmentation Using Monte Carlo Sampling: Application to Hippocampal Subfield Volumetry," *Medical Image Analysis* , vol. 17, no. 7, pp. 766-778, 2013.

[37] P. Risholm, F. Janoos, I. Norton, A. J. Golby and W. M. Wells III, "Bayesian characterization of uncertainty in intra-subject non-rigid registration," *Medical Image Analysis* , vol. 17, no. 5, pp. 538-555, 2013.

[38] P. Risholm, J. Ross, G. Washko and W. Wells, "Probabilistic elastography: estimating lung

elasticity," *Information Processing in Medical Imaging,* no. 2011.

[39] J. Kybic, "Bootstrap resampling for image registration uncertainty estimation without ground truth," *IEEE Trans. on Im. Proc,* vol. 19, no. 1, pp. 64-73, 2010.

[40] S. E. Muenzing, B. van Ginneken, K. Murphy and J. P. Pluim, "Supervised quality assessment of medical image registration: Application to intra-patient CT lung registration," *Medical Image Analysis* , vol. 16, no. 8, pp. 1521-1531, 2012.

[41] S. Muenzing, B. van Ginneken, A. Viergever and J. Pluim, "DIRBoost-An algorithm for boosting deformable image registration: Application to lung CT intra-subject registration," *Medical Image Analysis* , vol. 18, no. 3, pp. 449-459, 2014.

[42] T. Lotfi, L. Tang, S. Andrews and G. Hamarneh, "Improving Probabilistic Image Registration via Reinforcement Learning and Uncertainty Evaluation," *Machine Learning in Medical Imaging,* pp. 187-194, 2013.

[43] T. Glatard, X. Pennec and J. Montagnat, "Performance evaluation of grid-enabled registration algorithms using bronze-standards," in *MICCAI*, 2006.

[44] G. E. Christensen and H. J. Johnson, "Consistent image registration," *IEEE Trans. Med. Imaging* , vol. 20, no. 7, pp. 568-582, 2001.

[45] R. Woods, S. Grafton, C. Holmes, S. Cherry and J. Mazziotta, "Automated image registration: I. General methods and intrasubject, intramodality validation," *J. Comput. Assist. Tomogr.*, vol. 22, pp. 139-152, 1998.

[46] M. Holden, D. Hill, E. Denton, J. Jarosz, T. Cox, T. Rohlfing, J. Goodey and D. Hawkes, "Voxel similarity measures for 3D serial MR brain image registration," *IEEE Transactions on Medical Imaging* , vol. 19, pp. 94-102, 2000.

[47] G. E. Christensen, R. D. Rabbitt, M. I. Miller, S. C. Joshi, U. Grenander, T. A. Coogan and D. C. Van Essen, "Topological properties of smooth anatomic maps," *Information Processing in Medical Imaging,* vol. 3, pp. 101-112, 1995.

[48] G. E. Christensen, S. C. Joshi and M. I. Miller, "Volumetric transformation of brain anatomy," *IEEE Transactions on Medical Imaging,* vol. 16, pp. 864-877, 1997.

[49] J. P. Thirion, "Non-rigid matching using demons," in *Computer Vision and Pattern Recognition*, San Francisco, CA, 1996.

[50] D. J. Burr, "A dynamic model for image registration," *Computer Graphics Image Processing,* vol. 15, pp. 102-112, 1981.

[51] J. P. Thirion, " Image matching as a diffusion process: an analogy with maxwell's demons," *Medical*

Image Analysis, vol. 2, pp. 243-260, 1998.

[52] G. Rohde, A. Aldroubi and B. Dawant, "The adaptive bases algorithm for intensity-based nonrigid image registration," *IEEE Transactions on Medical Imaging ,* vol. 22, pp. 1470-1479, 2003.

[53] T. Vercauteren, X. Pennec, A. Perchant and N. Ayache, "Diffeomorphic demons: efficient non-parametric image registration," *NeuroImage ,* vol. 45, pp. 61-72, 2009.

[54] B. B. Avants, C. L. Epstein, M. Grossman and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Medical Image Analysis,* vol. 12, no. 1, pp. 26-41, 2008.

[55] X. Geng, D. Kumar and C. G.E., "Transitive inverse-consistent manifold registration," *Information Processing in Medical Imaging,* 2005.

[56] A. Rao, R. Chandrashekara, G. Sanchez-Ortiz, R. Mohiaddin, P. Aljabar, J. Hajnal, B. Puri and D. Rueckert, " Spatial transformation of motion and deformation fields using nonrigid registration," *IEEE Transaction on Medical Imaging,* vol. 23, no. 9, pp. 1065-1076, 2004.

[57] M. Beg, M. Miller, A. Trouve and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," *Int J Comput Vis,* vol. 61, no. 2, pp. 139-157, 2005.

[58] J. Ashburner, "A fast diffeomorphic image registration algorithm," *NeuroImage,* vol. 38, pp. 95-113, 2007.

[59] T. Vercauteren, X. Pennec, A. Perchant and N. Ayache, " Diffeomorphic demons: efficient non-parametric image registration," *NeuroImage,* vol. 45, no. 1, pp. 61-72, 2009.

[60] B. Yeo, M. Sabuncu, T. Vercauteren, N. Ayache, B. Fischl and P. Golland, "Spherical demons: fast diffeomorphic landmark-free surface registration," *IEEE Transactions on Medical Imaging,* vol. 29, no. 3, pp. 650-668, 2010.

[61] X. Geng, T. Ross, H. Gu, W. Shin, W. Zhan, Y.-P. Chao, L. Ching-Po, N. Schuff and Y. Yang, "Diffeomorphic image registration of diffusion MRI using spherical harmonics," *IEEE Transactions on Medical Imaging,* vol. 30, no. 3, pp. 747-758, 2011.

[62] G. Auzias, O. Colliot, J. Glaunes, M. Perrot, J.-F. Mangin, A. Trouve and S. Baillet, "Diffeomorphic brain registration under exhaustive sulcal constraints," *IEEE Transactions on Medical Imaging,* vol. 30, no. 6, pp. 1214-1227, 2011.

[63] K. Leung, G. R. Ridgway, S. Ourselin and N. Fox, "Consistent multi-time-point brain atrophy estimation from the boundary shift integral," *NeuroImage,* vol. 29, pp. 3995-4005, 2012.

[64] M. Reuter, H. Rosas and B. Fischl, "Highly accurate inverse consistent registration: a robust approach," *Neuroimage,* vol. 23, pp. 1181-1196, 2010.

[65] P. Freeborough, R. Woods and N. Fox, "Accurate registration of serial 3D MR brain images and its application to visualizing change in neurodegenerative disorders," *J. Comput. Assist.Tomogr.*, vol. 20, pp. 1012-1022, 1996.

[66] R. Datteri and B. Dawant, "Estimation and Reduction of Target Registration Error," in *Medical Image Computing and Computer-Assisted Intervention*, Nice, France, 2012.

[67] R. Datteri and B. Dawant, "Estimation of Rigid-Body Registration Quality Using Registration Networks," in *SPIE Medical Imaging*, San Diego, California, 2012.

[68] R. Datteri, A. Asman, B. Landman and B. Dawant, "Estimation of Registration Quality Applied to Multi-Atlas Segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, Toronto, 2011.

[69] R. Datteri and B. Dawant, "Automatic Detection of the Magnitude and Spatial Location of Error in Non-rigid Registration," in *WBIR*, Nashville, TN, 2012.

[70] R. Datteri and B. Dawant, "Validation of a Non-Rigid Registration Error Detection Algorithm using Clinical Data," *IEEE Transactions on Medical Imagin*, In Review.

[71] R. Datteri, A. Asman, B. Landman and B. Dawant, "Applying the Algorithm "Assessing Quality Using Image Registration Circuits" (AQUIRC) to Multi-Atlas Segmentation," in *SPIE Medical Imaging*, San Diego, California, 2014.

[72] J. FitzPatrick, Detecting failure, assessing success in Medical Image Regsitration, Baton Rouge: CRC Press, 2001.

[73] G. Christensen, "Transitivity Error: Non-Rigid image Registration Evaluation Project," 2006. [Online]. Available: http://www.nirep.org/te. [Accessed 2011].

[74] T. Gass, G. Szekely and O. Goksel, "Detection and Correction of Inconsistency-based Errors in Non-Rigid Registration," in *SPIE Medical Imaging* , San Diego, 2014.

[75] L. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, pp. 297--302 , 1945.

[76] J. Fitzpatrick, "Rigid point registration circuits," in *SPIE Medical Imaging*, San Diego, CA, 2014.

[77] G. Golub and C. van Loan, Matrix Computations, Baltimore, MD: Johns Hopkins University Press, 1983.

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

[79] D. Marcus et al., "Open Access Series of Imaging Studies (OASIS): Cross-Sectional MRI Data in

Young, Middle Aged, Nondemented, and Demented Older Adults," *Journal of Cognitive Neuroscience*, vol. 19, pp. 1498-1507, 2007.

[80] P. Aljabar, R. A. Heckemann, A. Hammers, J. Hajnal and D. Rueckert, "Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy," *Neuroimage*, vol. 46, no. 3, pp. 726-736, 2009.

[81] M. Modat, G. R. Ridgway, Z. A. Taylor, M. B. J. Lehmann, N. Fox, D. Hawkes and S. Ourselin, "Fast free-form deformation using graphics processing units," *Comput. Methods Programs Biomed*, 2009.

[82] B. Ardekani et al., "Quantitative comparison of algorithms for inter-subject registration of 3D volumetric brain MRI scans," *Journal of Neuroscience Methods* , vol. 142, pp. 67-76, 2005.

[83] "Insight Segmentation and Registration Toolkit," [Online]. Available: http://www.itk.org/.

[84] S. Pallavaram, H. Yu, J. Spooner, P.-F. D'Haese, B. Bodenheimer, P. E. Konrad and B. M. Dawant, "Inter-surgeon variability in the selection of anterior and posterior commissures and its potential effects on target localization," *Stereotact Funct Neurosurg*, vol. 86, no. 2, pp. 113-119, 2008.

[85] C. Meyer, T. Johnson, D. McLennan, D. Aberle, E. Kazerooni, H. M. B. MacMahon, D. Yankelevitz, v. B. J., S. Armato III, M. McNitt-Gray, A. Reeves and D. e. a. Gur, "Evaluation of lung MDCT nodule annotation across radiologists and methods," *Acad Radiol*, vol. 13, pp. 1254-1265, 2006.

[86] C. Donoghue, A. Rao, L. Pizarro, A. Bull and D. Rueckert, " Fast and accurate global geodesic registrations using knee MRI from the Osteoarthritis Initiative," in *IEEE Computer Vision and Pattern Recognition Workshops*, 2012.

[87] O. Goksel, T. Gass, V. Vishnevsky and G. Székely, "Estimation of Atlas-Based Segmentation Outcome: Leveraging Information From Unsegmented Images," in *IEEE Int Symp on Biomedical Imaging* , 2013.

[88] P.-F. D'Haese, S. Pallavaram, R. Li, M. S. Remple, C. Kao, J. S. Neimat, P. E. Konrad and B. M. Dawant, "CranialVault and its CRAVE tools: a clinical computer assistance system for Deep Brain Stimulation (DBS) therapy," *Med Image Anal.*, vol. 16, no. 3, pp. 744-753, 2012.

[89] P. Freeborough, R. Woods and N. Fox, "Accurate registration of serial 3D MR brain images and its application to visualizing change in neurodegenerative disorders," *J. Comput. Assist.Tomogr.*, vol. 20, pp. 1012-1022, 1996.

[90] G. E. Christensen and H. J. Johnson, "Invertibility and transitivity analysis for nonrigid image registration," *Journal of Electronic Imaging*, vol. 12, no. 1, pp. 106-117, 2003.

[91] G. E. Christensen and H. J. Johnson, "Invertibility and transitivity analysis for nonrigid image registration," *Journal of Electronic Imaging*, vol. 12, pp. 106-117, 2003.

APPENDIX

```
function  [tre, lsNIRC_Eps, lsIRC_Eps] =corrRegCircuitAndTRE()
%Rotate by 20 degrees and translate by -10 and 100
   R_fitz1 = createRotation(20,-10,100);
%The location of the four fiducials and target point
   coordI = [217,225,139,132;115,121,127,130;1,1,1,1];
   targetI= [155;57;1];
%Place the target point in the 4 image space configurations
   targetB = targetI; targetC = targetI;
   targetD = targetI; targetE = targetI;
%Amount of FLE applied
   N_fle = sqrt(1/3);
%Add FLE to create the configurations in  image space
   coordB = coordI + N_fle.*randn(size(coordI)).*[1,1,1,1;1,1,1,1;0,0,0,0];
   coordC = coordI + N_fle.*randn(size(coordI)).*[1,1,1,1;1,1,1,1;0,0,0,0];
   coordD = coordI + N_fle.*randn(size(coordI)).*[1,1,1,1;1,1,1,1;0,0,0,0];
   coordE = coordI + N_fle.*randn(size(coordI)).*[1,1,1,1;1,1,1,1;0,0,0,0];
%Demean each image space configuration and their target point
   meanB = mean(coordB,2);  meanB(length(meanB)) = 0;
   targetB = targetB - meanB;
   coordB = coordB - meanB*ones(1,length(coordB));
   meanC = mean(coordC,2);  meanC(length(meanC)) = 0;
   targetC = targetC - meanC;
   coordC = coordC - meanC*ones(1,length(coordC));
   meanD = mean(coordD,2);  meanD(length(meanD)) = 0;
   targetD = targetD - meanD;
   coordD = coordD - meanD*ones(1,length(coordD));
   meanE = mean(coordE,2); meanE(length(meanE)) = 0;
   targetE = targetE- meanE;
   coordE = coordE - meanE*ones(1,length(coordE));
%Demean the original configuration and then rotate/translate the
%coordinates to create the surgical space fiducial configuration
   meanI = mean(coordI,2); meanI(3) = 0;
   coordI = coordI - meanI*ones(1,length(coordI));
   targetI = targetI - meanI;
   coordA =  R_fitz1*(coordI);
   targetA = R_fitz1*(targetI);
%Compute the point-wise registrations between each of the configurations
%and their inverses
   [Ra_b] = point_register(coordA,coordB);
   [Ra_c] = point_register(coordA,coordC);
   [Ra_d] = point_register(coordA,coordD);
   [Ra_e] = point_register(coordA,coordE);
   [Rb_c] = point_register(coordB,coordC);
   [Rb_d] = point_register(coordB,coordD);
   [Rb_e] = point_register(coordB,coordE);
   [Rc_d] = point_register(coordC,coordD);
   [Rc_e] = point_register(coordC,coordE);
```

```
    [Rd_e] = point_register(coordD,coordE);
    [Rb_a] = inv(Ra_b); [Rc_a] = inv(Ra_c); [Rd_a] = inv(Ra_d);
    [Re_a] = inv(Ra_e); [Rc_b] = inv(Rb_c); [Rd_b] = inv(Rb_d);
    [Re_b] = inv(Rb_e); [Rd_c] = inv(Rc_d); [Re_c] = inv(Rc_e);
    [Re_d] =inv(Rd_e);
%Compute the true TRE between each point configuration
    TREa_b= sqrt(sum((targetB - Ra_b*targetA).^2));
    TREa_c = sqrt(sum((targetC - Ra_c*targetA).^2));
    TREa_d = sqrt(sum((targetD - Ra_d*targetA).^2));
    TREa_e = sqrt(sum((targetE - Ra_e*targetA).^2));
    TREb_c = sqrt(sum((targetC - Rb_c*targetB).^2));
    TREb_d = sqrt(sum((targetD - Rb_d*targetB).^2));
    TREb_e = sqrt(sum((targetE - Rb_e*targetB).^2));
    TREc_d = sqrt(sum((targetD - Rc_d*targetC).^2));
    TREc_e = sqrt(sum((targetE - Rc_e*targetC).^2));
    TREd_e = sqrt(sum((targetE - Rd_e*targetD).^2));
%Compute the traditional network error (TNEM).  Always start in a node
%in the image space.
    i_error = zeros(10,1);
    combinedBCA = Ra_b*Rc_a*Rb_c;
    i_error(1) = sqrt(sum((targetB - combinedBCA*targetB).^2));
    combinedBDA = Ra_b*Rd_a*Rb_d;
    i_error(2) = sqrt(sum((targetB - combinedBDA*targetB).^2));
    combinedBEA = Ra_b*Re_a*Rb_e;
    i_error(3) = sqrt(sum((targetB - combinedBEA*targetB).^2));
    combinedCDA =Ra_d*Rc_a*Rd_c;
    i_error(4) = sqrt(sum((targetD - combinedCDA*targetD).^2));
    combinedCEA = Ra_e*Rc_a*Re_c;
    i_error(5) = sqrt(sum((targetE - combinedCEA*targetE).^2));
    combinedDEA = Ra_e*Rd_a*Re_d;
    i_error(6) = sqrt(sum((targetE - combinedDEA*targetE).^2));
    combinedBCD = Rc_b*Rd_c*Rb_d;
    i_error(7) = sqrt(sum((targetB - combinedBCD*targetB).^2));
    combinedBCE = Rc_b*Re_c*Rb_e;
    i_error(8) = sqrt(sum((targetB - combinedBCE*targetB).^2));
    combinedBDE = Rd_b*Re_d*Rb_e;
    i_error(9) = sqrt(sum((targetB - combinedBDE*targetB).^2));
    combinedCDE = Rd_c*Re_d*Rc_e;
    i_error(10)  = sqrt(sum((targetC - combinedCDE*targetC).^2));
    tre = [  TREa_b;TREa_c;TREa_d;TREa_e;TREb_c;
    TREb_d;TREb_e;TREc_d;TREc_e;TREd_e ];
%Compute the non traditional network error measure (NTNEM). Always starts in a
%node in the image space.
    n_error = zeros(10,1);
    combinedBCA = Ra_b*Rb_c*Rc_a;
    n_error(1) = sqrt(sum((targetC - combinedBCA*targetC).^2));
    combinedBDA = Ra_b*Rb_d*Rd_a;
    n_error(2) = sqrt(sum((targetD - combinedBDA*targetD).^2));
    combinedBEA = Ra_b*Rb_e*Re_a;
    n_error(3) = sqrt(sum((targetE - combinedBEA*targetE).^2));
    combinedCDA = Ra_d*Rd_c*Rc_a;
```

```
    n_error(4) = sqrt(sum((targetC - combinedCDA*targetC).^2));
    combinedCEA = Ra_e*Re_c*Rc_a;
    n_error(5) = sqrt(sum((targetC - combinedCEA*targetC).^2));
    combinedDEA = Ra_e*Re_d*Rd_a;
    n_error(6) = sqrt(sum((targetD - combinedDEA*targetD).^2));
    combinedBCD = Rd_c*Rc_b*Rb_d;
    n_error(7) = sqrt(sum((targetB - combinedBCD*targetB).^2));
    combinedBCE = Re_c*Rc_b*Rb_e;
    n_error(8) = sqrt(sum((targetB - combinedBCE*targetB).^2));
    combinedBDE = Re_d*Rd_b*Rb_e;
    n_error(9) = sqrt(sum((targetB - combinedBDE*targetB).^2));
    combinedCDE = Re_d*Rd_c*Rc_e;
    n_error(10) = sqrt(sum((targetC - combinedCDE*targetC).^2));
%Calculate the connectivity graph:
    numberOfNodes = 5;
    connectivityMap = zeros(numberOfNodes,numberOfNodes, 'single');
    count1 = 1;
%Assign a single number in the path matrix to each connection i <-> k.
%For instance the edge between node 1 and node 2 is assigned the number 1.
    for i = 1:numberOfNodes
        for k = 1:numberOfNodes
            if (i < k)
                connectivityMap(i,k) = count1;
                count1 = count1 +1;
            end
        end
    end
    connectivityMap = connectivityMap + connectivityMap';
    count = 0;
%Create the path matrix using the connectivity map.
    for i = 1:numberOfNodes
        for j = i+1:numberOfNodes
            for k = j+1:numberOfNodes
                if ( i ~= j && i ~= k && j ~= k)
                    count = count + 1;
                    Path(count,connectivityMap(i,j)) = 1;
                    Path(count,connectivityMap(j,k)) = 1;
                    Path(count,connectivityMap(k,i)) = 1;
                end
            end
        end
    end
%Compute the least squares solution on the connectivity graph
    A= inv(Path'*Path)*Path';
%Calculate the NTNEM epsilon value
    lsNIRC_Eps = A*n_error;
%Calculate the TNEM epsilon value
    lsIRC_Eps = A*i_error;
%Visualize the TNEM epsilons from the surgical space to the image space
    figure(2);hold on;scatter(tre(1:4), lsIRC_Eps(1:4), 'r')
%Visualize the TNEM epsilons from the image space to the image space
```

```
   hold on;scatter(tre(5:10), lsIRC_Eps(5:10), 'b')
%Visualize the NTNEM epsilons from the surgical space to the image space
   figure(1);hold on;scatter(tre(1:4), lsNIRC_Eps(1:4), 'r')
%Visualize the NTNEM epsilons from the image space to the image space
   hold on;scatter(tre(5:10), lsNIRC_Eps(5:10), 'B')
end
%Register the point set X to Y by minimizing FRE using SVD. Returns a
%transformation that projects X to Y.
function [R] = point_register(X,Y);
   [xr, xc]=size(X);
   [yr, yc]=size(Y);
   N=xc;
%Compute the centroid of the fiducial config in each space
   Xbar=mean(X,2);
   Ybar=mean(Y,2);
%Compute the displacement form the centroid  to each fiducial point in
%each space
   Xnew=X-Xbar*ones(1,N);
   Ynew=Y-Ybar*ones(1,N);
%Compute the fiducial covariance matrix
   H=Xnew*Ynew';
%Perform svd of H
   [U, Lambda, V] = svd(H);
%Compute the rotation matrix R
   r=V*diag([1,1, det(V*U)])*U';
%Compute the translation vector t
   t=Ybar-r*Xbar;
   R = includeTranslation(r,t);
end
function [R3d] = includeTranslation(R,T)
   R3d = zeros(3,3);
   R3d(1:3, 1:3) =R;
   R3d(1:3, 3) =T;
   R3d(3, 3) =1;
end
%creates a rotation matrix with a rotation of 'degree' around the x-axis
%and translates in the X and Y direction by 'transX' and 'transY'.
function R = createRotation(degree, transX, transY)
   angle = degree*pi/180;
   R = [cos(angle), -sin(angle), 0; sin(angle), cos(angle), 0;0,0,1];
   R(1,3) = transX;
   R(2,3) = transY;
end
```