Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2018 March ; 10574: . doi:10.1117/12.2293642.

QUADRATIC: Quality of Dice in Registration Circuits

Shikha Chaganti^a and Bennett A. Landman^{a,b}

^aDepartment of Computer Science, Vanderbilt University

^bDepartment of Electrical Engineering, Vanderbilt University

Abstract

Image registration involves identification of a transformation to fit a target image to a reference image space. The success of the registration process is vital for correct interpretation of the results of many medical image-processing applications, including multi-atlas segmentation. While there are several validation metrics employed in rigid registration to examine the accuracy of the method, non-rigid registrations (NRR) are validated subjectively in most cases, validated in offline cases, or based on image similarity metrics, all of which have been shown to poorly correlate with true registration quality. In this paper, we model the error for each target scan by expanding on the idea of Assessing Quality Using Image Registration Circuits (AQUIRC), which created a model for error “quality” associated with NRR. In this paper, we model the Dice similarity coefficient (DSC) error in the network, for a more interpretable measure. We test four functional models using a leave-one-out strategy to evaluate the relationship between edge DSC and circuit DSC: linear, quadratic, third order, or multiplicative models. We found that the quadratic model most accurately learns the NRR-DSC, with a median correlation coefficient of 0.58 with the true NRR-DSC, we call this the QUADRATIC (QUALity of Dice in RegistrATIOn Circuits) model. The QUADRATIC model is used for multi-atlas segmentation based on majority vote. Choosing the four best atlases predicted from the QUADRATIC model resulted in a 7% increase in the DSC between segmented image and true labels.

Keywords

Non-rigid registration; error modeling; AQUIRC; atlas selection; multi-atlas; majority vote

1. INTRODUCTION

Non-rigid registration (NRR) is used in various applications in the medical image processing community, such as comparing anatomical structures between patients, tracking change in anatomical structures over time within the same patient, and performing segmentation using multi-atlas algorithms [1, 2]. The ability to estimate the accuracy of NRR is vital in the correct interpretation of these results. Specifically, in use cases such as multi-atlas segmentation, the ability to quantify the magnitude and location of error associated with NRR procedures is important as it can propagate to further steps of the segmentation process.

In the past, NRR was evaluated either subjectively by visual inspection [3] or offline evaluation of atlases based on an image similarity metric [4, 5]. The optimization process

used by NRR inherently biases image similarity metrics, and image similarity is often not predictive of structural correspondence. Quantifying the error in correspondence as a result of NRR in a target image is a challenging problem, with some calling automatic registration techniques impossible to validate[5]. Datteri et al have developed the AQUIRC (Assessing Quality Using Image Registration Circuits)[6, 7] model to assess registration quality in target images using registration circuits. The authors proposed to use registration circuits to model the local error associated with NRR and have shown moderate success in applying this to multi-atlas segmentation[8]. In the AQUIRC model, the error “quality” in edges of the circuits is measured. In our model, we use the idea of image registration circuits to estimate the actual error associated with each non-rigid deformation, in this case, the Dice Similarity Coefficient (DSC) of the edges in the network. To model DSC in the network, we test four models to determine the relationship between edge DSC and circuit DSC. The success or failure of the non-rigid registration is predicted using the error model and is used for atlas selection in majority vote multi-atlas segmentation. We find that the quadratic error model is the most successful in modeling DSC error in registration circuits, we call this model QUADRATIC (QUALity of Dice in RegistrATIOn Circuits).

2. METHODS

2.1 Data

The data consists of 16 computed tomography scans of eye orbit collected at the Vanderbilt University Medical Center which were retrieved and de-identified under institutional review board approval and previously described in [9]. The following structures in the eye orbit were labeled by experts to form 16 atlases: the optic nerve, the globe, the rectus muscles, and orbital fat. A leave-one-out approach was used where one of the atlases was treated as a target scan, while the other 15 were used for NRR error modeling and segmentation.

2.2 Registration

Each of the 16 atlases was registered to the other 15 atlases using ANTS SyN registration [10], with the cross-correlation similarity metric. The registration process non-rigidly transforms each atlas to the target space such that a voxel-wise correspondence between the atlas and the target is estimated. The deformation field that was obtained by registering the atlases to the target space was used to propagate the expert labels from the atlas space to the target scan using nearest neighbor interpolation.

The quality of the registration was measured based on structural correspondence between the target image and registered atlas. The structural correspondence was measured using the Dice Similarity Coefficient (DSC). It gives a value between 0 and 1 indicating similarity between the target and the registered atlas, by calculating the proportion of voxels that are the same in both images. It is given by,

$$DSC = 1 - \frac{2 |X \cap Y|}{|X| + |Y|}$$

The DSC value is computed for each of the four structures bilaterally as well as for the whole segmentation, we will refer to these DSC values as NRR-DSC (Non-Rigid registration Dice similarity coefficient), in order to differentiate it from DSC values obtained from segmentation results that will be computed later in the paper.

2.3 Estimating the Error Model with Registration Circuit

This paper builds on the idea of AQUIRC to use registration circuits of size 3 to estimate registration error [8]. To model the error associated with NRR for each target scan (A_x), the other 15 scans are used as atlases ($A_1 - A_{15}$), with known true labels, to construct $15p3$ (permutations of size three from 15 atlases) circuits ($n = 2730$), having $15p2$ (permutations of size two from 15 atlases) edges ($m = 210$). Each registration circuit consists of three atlases A_x , A_y , and A_z as shown in Figure 1. The non-rigid transformations from A_x to A_y , A_y to A_z and A_z to A_x , are computed as T_{xy} , T_{yz} and T_{zx} respectively. The deformation fields calculated from the non-rigid transformations are used to propagate the labels associated with A_x through the circuit. A_x is transformed to image space A_y using transformation, $T_{xy}(A_x)$. Next, it is transformed to image space A_z using $T_{yz}(T_{xy}(A_x))$, and then back to its original space to get $A'_x = T_{zx}(T_{yz}(T_{xy}(A_x)))$. The error associated with each of the edges of the circuit is evaluated by computing the NRR-DSC value between the true labels and propagated labels from the non-rigid transformation for each edge to give ϵ_{xy} , ϵ_{yz} and ϵ_{zx} . The NRR-DSC value between true labels of A and the propagated labels of A' is the error associated with the registration circuit given by E_{xyz} . In this paper, we evaluate the non-linear relationship between edge error, ϵ and circuit error E with a Taylor series expansion up to the third order in f_1 (linear), f_2 (quadratic), f_3 (third order), and also a multiplicative relationship in f_4 , which are given by,

$$f_1(\epsilon_{xy}, \epsilon_{yz}, \epsilon_{zx}) = E_{xyz} = p_0 + p_1\epsilon_{xy} + p_2\epsilon_{yz} + p_3\epsilon_{zx} \quad (1)$$

$$f_2(\epsilon_{xy}, \epsilon_{yz}, \epsilon_{zx}) = E_{xyz} = p_0 + p_1\epsilon_{xy} + p_2\epsilon_{yz} + p_3\epsilon_{zx} + p_4\epsilon_{xy}^2 + p_5\epsilon_{yz}^2 + p_6\epsilon_{zx}^2 + p_7\epsilon_{xy}\epsilon_{yz} + p_8\epsilon_{yz}\epsilon_{zx} + p_9\epsilon_{zx}\epsilon_{xy} \quad (2)$$

$$f_3(\epsilon_{xy}, \epsilon_{yz}, \epsilon_{zx}) = E_{xyz} = p_0 + p_1\epsilon_{xy} + p_2\epsilon_{yz} + p_3\epsilon_{zx} + p_4\epsilon_{xy}^2 + p_5\epsilon_{yz}^2 + p_6\epsilon_{zx}^2 + p_7\epsilon_{xy}\epsilon_{yz} + p_8\epsilon_{yz}\epsilon_{zx} + p_9\epsilon_{zx}\epsilon_{xy} + p_{10}\epsilon_{xy}^3 + p_{11}\epsilon_{yz}^3 + p_{12}\epsilon_{zx}^3 + p_{13}\epsilon_{xy}^2\epsilon_{yz} + p_{14}\epsilon_{yz}^2\epsilon_{zx} + p_{15}\epsilon_{zx}^2\epsilon_{xy} + p_{16}\epsilon_{xy}\epsilon_{yz}^2 + p_{17}\epsilon_{yz}\epsilon_{zx}^2 + p_{18}\epsilon_{zx}^2\epsilon_{xy} + p_{19}\epsilon_{xy}\epsilon_{yz}\epsilon_{zx} \quad (3)$$

$$f_4(\epsilon_{xy}, \epsilon_{yz}, \epsilon_{zx}) = E_{xyz} = p_0 * p_1 \epsilon_{xy} * p_2 \epsilon_{yz} * p_3 \epsilon_{zx} \text{ or } \log E_{xyz} = \log p_0 + \log p_1 \epsilon_{xy} + \log p_2 \epsilon_{yz} + \log p_3 \epsilon_{zx} \quad (4)$$

Note that the edge errors are known for all of the 210 edges of the circuits since the true labels are known. The parameters, p_i in models 1 through 4 are evaluated using the 2730 circuits using a generalized linear model with equations (1)–(4). Here, the response variable is the circuit error E_{xyz} , and the independent variables are edge errors of the form ϵ_{xy} , ϵ_{yz} and ϵ_{zx} . The estimated parameters will be indicated by $\hat{p}_i$.

Next, the estimated parameters and known errors ϵ , are used to learn the error between the target scan A_s , and each of the 15 atlases, A_{1-15} . $15p2(n' = 210)$ circuits are constructed with target scan S , where the circuit starts with an atlas. For example, given atlases A_x , and A_y , the following circuits are constructed: $A_x \rightarrow T_{xy}(A_x) \rightarrow T_{ys}(T_{xy}(A_x)) \rightarrow A'_x$ and $A_x \rightarrow T_{xs}(A_x) \rightarrow T_{sz}(T_{xs}(A_x)) \rightarrow A'_x$. This leads to a system of 210 equations with 30 unknown variables ($\epsilon_{1s}, \epsilon_{2s}, \dots, \epsilon_{15s}, \epsilon_{s1}, \epsilon_{s2}, \dots, \epsilon_{s15}$). For models 1 through 4, the systems of equations are:

1. $\hat{f}_1(\epsilon_{xy}, \epsilon_{ys}, \epsilon_{sx}) = E_{xys} = \hat{p}_0 + \hat{p}_1 \epsilon_{xy} + \hat{p}_2 \epsilon_{ys} + \hat{p}_3 \epsilon_{sx}$
2. $\hat{f}_2(\epsilon_{xy}, \epsilon_{ys}, \epsilon_{sx}) = E_{xys} = \hat{p}_0 + \hat{p}_1 \epsilon_{xy} + \hat{p}_2 \epsilon_{ys} + \hat{p}_3 \epsilon_{sx} + \hat{p}_4 \epsilon_{xy}^2 + \hat{p}_5 \epsilon_{ys}^2 + \hat{p}_6 \epsilon_{sx}^2 + \hat{p}_7 \epsilon_{xy} \epsilon_{ys} + \hat{p}_8 \epsilon_{ys} \epsilon_{sx} + \hat{p}_9 \epsilon_{sx} \epsilon_{xy}$
3. $\hat{f}_3(\epsilon_{xy}, \epsilon_{ys}, \epsilon_{sx}) = E_{xys} = \hat{p}_0 + \hat{p}_1 \epsilon_{xy} + \hat{p}_2 \epsilon_{ys} + \hat{p}_3 \epsilon_{sx} + \hat{p}_4 \epsilon_{xy}^2 + \hat{p}_5 \epsilon_{ys}^2 + \hat{p}_6 \epsilon_{sx}^2 + \hat{p}_7 \epsilon_{xy} \epsilon_{ys} + \hat{p}_8 \epsilon_{ys} \epsilon_{sx} + \hat{p}_9 \epsilon_{sx} \epsilon_{xy} + \hat{p}_{10} \epsilon_{xy}^3 + \hat{p}_{11} \epsilon_{ys}^3 + \hat{p}_{12} \epsilon_{sx}^3 + \hat{p}_{13} \epsilon_{xy}^2 \epsilon_{ys} + \hat{p}_{14} \epsilon_{ys}^2 \epsilon_{sx} + \hat{p}_{15} \epsilon_{sx}^2 \epsilon_{xy} + \hat{p}_{16} \epsilon_{xy} \epsilon_{ys}^2 + \hat{p}_{17} \epsilon_{ys} \epsilon_{sx}^2 + \hat{p}_{18} \epsilon_{sx} \epsilon_{xy}^2 + \hat{p}_{19} \epsilon_{xy} \epsilon_{ys} \epsilon_{sx}$
4. $\log \hat{f}_4(\epsilon_{xy}, \epsilon_{ys}, \epsilon_{sx}) = \log E_{xys} = \log \hat{p}_0 + \log \hat{p}_1 \epsilon_{xy} + \log \hat{p}_2 \epsilon_{ys} + \log \hat{p}_3 \epsilon_{sx}$

The systems of equations are solved using the least squares non-linear (lsqnonlin) function in MATLAB for estimated edge errors: $\hat{\epsilon}_{1s}, \hat{\epsilon}_{2s}, \dots, \hat{\epsilon}_{15s}, \hat{\epsilon}_{s1}, \hat{\epsilon}_{s2}, \dots, \hat{\epsilon}_{s15}$.

2.4 Majority Vote with Atlas Selection

Multi-atlas segmentation based on majority vote is used to segment each target scan, with the other 15 scans used as atlases. The quality of the segmentation is dependent on the quality of the non-rigid registrations. Poor registrations to the target scan can worsen the final result of the segmentation. Predicting the NRR-DSC associated with each registration as described in section 2.3 can help determine the atlases with best registration for atlas selection. The NRR-DSC estimates are sorted in descending order from the atlas with the best predicted-NRR-DSC (A'_1) to the worst predicted-NRR-DSC (A'_{15}): $\hat{\epsilon}'_1, \hat{\epsilon}'_2, \dots, \hat{\epsilon}'_{15}$.

In order to determine the number of atlases to select to achieve the maximum gain in performance, we perform multi-atlas segmentation 15 times, starting with the best atlas and

adding the next best one until all the atlases are used. For the k^{th} segmentation, let $A^{\{k\}} = L^{k \times N}$ be the set of co-registered atlases where N is the number of voxels, and k is the number of registrations. For each target voxel j , the probability that the voxel has a label s , is given by

$$P(X_j = s | D) = \frac{1}{k} \sum_{i=1}^k \delta(s, D_{i,j})$$

where, δ is the Kronecker delta given by,

$$\delta(a, b) = \begin{cases} 1, & \text{if } a = b \\ 0, & \text{otherwise} \end{cases}$$

The label decision at voxel j is given by,

$$\arg_s \max (P(X_j = s | D))$$

For each of the segmentations, the DSC between the true labels and the segmented labels is calculated to assess the quality of the segmentation. We will refer to this as the seg-DSC, in order to differentiate it from the NRR-DSC.

3. RESULTS

3.1 Error Model

We had a set of 16 atlases of the eye orbit with expertly marked labels for bilateral orbital structures including the globe, the optic nerve, the extraocular muscles, and periorbital fat. We used a leave-one-out approach to find the error model of the non-rigid registration by considering one of the atlases to be a target scan whose labels were unknown and using the other 15 scans as atlases. The NRR-DSC of the non-rigid registration from each of the 15 atlases to the target space is calculated as described in section 2.3 for each of the structures and the overall segmentation. The predicted NRR-DSC values using each model are compared to the true NRR-DSC values for each target atlas. The mean of the correlation values for the 16 target scans is shown in Figure 2. The quadratic error model has the highest correlation (p-value<0.05) with the true NRR-DSC values for all of the structures except periorbital fat. The mean correlation between true and predicted NRR-DSC values for overall segmentation using the quadratic model was 0.58 (p-value<0.001). The R^2 values between true and predicted NRR-DSC values are shown in Figure 3. The median R^2 value for the mean registration was highest for the quadratic model, at 0.34. We call this quadratic error model, QUADRATIC (QUALity of Dice in RegistrATIOn Circuits).

3.2 Majority vote results

The mean NRR-DSC predicted using QUADRATIC is used in atlas selection. Figure 4A shows the concordance at the top between the true and estimated mean NRR-DSC values using the quadratic model. Concordance at the top for k atlases calculates the overlap between top k atlases based on true NRR-DSC and top k atlases based on estimated NRR-DSC. As described in section 2.4, 15 segmentations are performed for each target scan. For each of the segmentations, the quality of the segmentation is measured by computing the seg-DSC between the true labels and the segmented labels. The ratio of seg-DSC between the top k atlases and all 15-atlases is computed to evaluate the gain in performance. Figure 4B shows the gain in performance by selecting the best atlases using the true NRR-DSC (red dotted line), and predicted NRR-DSC (blue line). The X-axis shows number of atlases selected, and the Y-axis shows the ratio of k :all seg-DSC. Using the best 4 predicted atlases gives an improvement of 7% on an average. Figure 5 shows the comparison between using the best 4 atlases to perform the majority vote and using all of the atlases. It can be seen that a gain in performance is observed in segmentation of all structures except the optic nerve using the atlas selection procedure. An example of the segmentation result can be seen in Figure 6.

4. CONCLUSION

Assessing the structural correspondence error in non-rigid registrations is one of the most challenging problems in medical image-processing. In this paper, we expand on the previous work in AQUIRC to model error using registration circuits. In AQUIRC a multiplicative relationship was assumed between the error quality and circuit error. However, this relationship is not predictive of DSC error. We used a Taylor series expansion of up to third order to test other non-linear relationships between edge and circuit DSC. We found that the QUADRATIC error model showed the best prediction of NRR-DSC.

In the past, atlas-selection was performed by selecting atlases based on image similarity metrics[11], manifolds computed on image similarity[12, 13], or using meta-data such as age[14]. However, neither method is a predictor of true local structural correspondence, which affects the quality of multi-atlas segmentation. In order to show that prediction of NRR-DSC using QUADRATIC can improve multi-atlas segmentation, we performed the simplest case of a majority vote based on the best overall predicted DSC. On an average, we observe a 7% improvement in the segmentation DSC using the atlases predicted as best from the quadratic error model. From Figure 3C, it can be seen that when there is a perfect prediction of the NRR-DSC, the improvement in segmentation-DSC can be up to 16%. This improvement can be achieved in future work through better optimization procedures to solve the non-linear system of equations to prevent local minima. Additionally, the NRR-DSC value is predicted for each structure of the atlas, which could lead to innovative multi-atlas segmentation techniques where atlas selection can be applied to each structure to improve overall segmentation. The ability to have a quantitative measure of non-rigid registration can lead to several improvements in multi-atlas segmentation algorithms.

Acknowledgments

This research was supported by NSF CAREER 1452485 and NIH grants 5R21EY024036. This research was conducted with the support from Intramural Research Program, National Institute on Aging, NIH. This study was in part using the resources of the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University, Nashville, TN. This project was supported in part by ViSE/VICTR VR3029 and the National Center for Research Resources, Grant UL1 RR024975-01, and is now at the National Center for Advancing Translational Sciences, Grant 2 UL1 TR000445-06. This work was also supported by the National Institutes of Health in part by the National Institute of Biomedical Imaging and Bioengineering training grant T32-EB021937.

References

1. Asman AJ, Landman BA. Non-local statistical label fusion for multi-atlas segmentation. *Medical image analysis*. 2013; 17(2):194–208. [PubMed: 23265798]
2. Sabuncu MR, Yeo BT, Van Leemput K, et al. A generative model for image segmentation based on label fusion. *IEEE transactions on medical imaging*. 2010; 29(10):1714–1729. [PubMed: 20562040]
3. Good CD, Scathill RI, Fox NC, et al. Automatic differentiation of anatomical patterns in the human brain: validation with studies of degenerative dementias. *Neuroimage*. 2002; 17(1):29–46. [PubMed: 12482066]
4. Sdika M. Combining atlas based segmentation and intensity classification with nearest neighbor transform and accuracy weighted vote. *Medical Image Analysis*. 2010; 14(2):219–226. [PubMed: 20056473]
5. Sjöberg C, Ahnesjö A. Multi-atlas based segmentation using probabilistic label fusion with adaptive weighting of image similarity measures. *Computer methods and programs in biomedicine*. 2013; 110(3):308–319. [PubMed: 23339900]
6. Datteri R, Asman A, Landman B, et al. Estimation of registration accuracy applied to multiatlas segmentation. 2011:78–87.
7. Datteri RD, Liu Y, D’Haese PF, et al. Validation of a nonrigid registration error detection algorithm using clinical MRI brain data. *IEEE transactions on medical imaging*. 2015; 34(1):86–96. [PubMed: 25095252]
8. Datteri RD, Asman AJ, Landman BA, et al. Applying the algorithm” assessing quality using image registration circuits”(AQUIRC) to multi-atlas segmentation. 90341F.
9. Chaganti S, Nelson K, Mundy K, et al. Structural functional associations of the orbit in thyroid eye disease: Kalman filters to track extraocular rectal muscles. :9784.
10. Avants BB, Epstein CL, Grossman M, et al. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*. 2008; 12(1):26–41. [PubMed: 17659998]
11. Aljabar P, Heckemann RA, Hammers A, et al. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage*. 2009; 46(3):726–738. [PubMed: 19245840]
12. Cao Y, Yuan Y, Li X, et al. Segmenting images by combining selected atlases on manifold. :272–279.
13. Duc AKH, Modat M, Leung KK, et al. Using manifold learning for atlas selection in multi-atlas segmentation. *PloS one*. 2013; 8(8):e70059. [PubMed: 23936376]
14. Aljabar P, Heckemann R, Hammers A, et al. Classifier selection strategies for label fusion using large atlas databases. :523–531.

Figure 1.

Example of an image registration circuit. The three atlases in the circuit, A_x , A_y , A_z and the transformations T_{xy} , T_{yz} and T_{zx} . A_x is transformed to $A'_x = T_{zx}(T_{yz}(T_{xy}(A_x)))$

Figure 2.

The correlation between true and predicted NRR-DSC for each structure and the overall segmentation are shown for the four models. Red indicates the linear model, blue indicates the quadratic model, cyan indicates the third order model, and green indicates the multiplicative model.

Figure 3.

The R-squared between true and predicted NRR-DSC for each structure and the overall segmentation are shown for the four models. Red indicates the linear model, blue indicates the quadratic model, cyan indicates the third order model, and green indicates the multiplicative model.

Figure 4.

A. Concordance at the top between true and estimated NRR-DSC, i.e., the overlap between top k atlases and the predicted top k atlases. B. Majority vote results for the 15 segmentations, where k^{th} segmentation has the top k predicted atlases (blue line), and true top k atlases (red dotted line).

Figure 5. Comparison between seg-DSC for each structure using atlas selection with $k=4$ (shown in red) and no atlas selection (shown in blue). The seg-DSC values are calculated for the 16 scans using a leave-one-out strategy.

Figure 6.

A. Shows the true labels of a scan. B. Shows **labels** segmented using the atlas selection ($k=4$) based on the QUADRATIC error model. C. Shows the labels segmented using all 15 atlases, without atlas selection.