CONFIDENCE REGION ESTIMATION FOR MEDICAL IMAGE REGISTRATION

Michael Allison, Takanori Watanabe, Bhargav Avasarala

1 Introduction

Medical image registration is a pre-processing operation which deforms a floating image so that its anatomical features align with those of a reference image. Its ability to combine physiological and anatomical information has led to its adoption in a variety of clinical settings [1]. However, the registration process is complicated by the fact that these images are often acquired using different modalities (e.g., PET, CT, MR) resulting in a variation in the appearance of the anatomy, Figure 1. Furthermore, the actual anatomy being imaged may change significantly between acquisitions (e.g., beating heart). Thus, there can be significant errors in the registration.

In many clinical applications, it is important to know the level of accuracy of the registration [2]. For example, registered PET and MR images are often used to guide oncological neurosurgery. In this case, the PET images contain information about cancerous activity within the brain, but do not include significant anatomical structure. Conversely, MR images capture anatomical structure, but provide little physiological information. By registering these images, the cancerous anatomical structures can be identified. As it is desirable to remove all of the cancerous tissue, while sparing as much healthy tissue as possible, it is important to know the degree of error in the alignment.

While several methods have been developed to estimate the error distribution of arbitrary points in an image [2], they require fiducial markers which limit their applicability. We therefore propose a novel method, independent of the particular registration technique, which uses machine learning to provide such estimates. Our method provides a confidence region in the reference image based on a given significance level and a pixel location in the transformed floating image. In the following sections, we outline our method and provide an evaluation of its performance using simulated data.

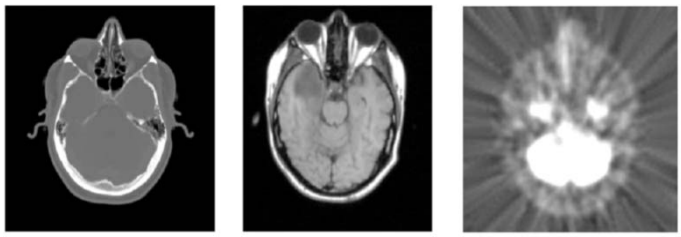

Figure 1: Examples of brain CT, MRI, and PET images. Note that the appearance of the anatomy varies with acquisition method.

2 Background

Medical image registration techniques can be broadly divided into two categories: rigid and nonrigid. Rigid registration is characterized by only rotation and translation, and is particularly applicable when the anatomical structure remains constant (e.g., brain and pelvis). Conversely, nonrigid registration constitutes a complex set of transformations that cannot be generalized by

simple linear motion models. This technique is particularly applicable when the anatomy of interest can deform (e.g., heart and lungs). Similarly, the applications of medical imaging registration can be classified into two categories: unimodal and multimodal. Unimodal registration refers to the alignment of images acquired using the same modality, while multimodal registration operates on images acquired from different modalities. Multimodal registration is typically a more difficult problem, as the appearance of a particular structure may differ between images, Figure 1.

Registration involves maximizing a measure of similarity between two images. One of the earliest medical imaging registration techniques was the point-based method, an approach that requires the placement of fiducial markers around the anatomy of interest [2]. This method involves identifying the optimal transformation that aligns the fiducial markers between the images. In the case of rigid deformation, a closed form solution can be found using the Orthogonal Procrustes method, where optimality is achieved in the least squares sense [2]. Although this method provides satisfactory results when the rigid assumption holds, it quickly reaches its limitation under non-rigid settings. Alternatively, voxel-intensity based registration methods can be adapted for non-rigid deformations. For this reason and because fiducial markers are often not available, these registration methods have quickly gained in popularity.

Currently the most popular similarity measures are based on mutual information [3], a measure of how much knowledge one image provides about the other. These methods use a joint intensity histogram computed between the images and thus implicitly assume spatial independence. Intuitively, this assumption is unrealistic as a significant amount of information about a pixel can be obtained from its neighbors. Hence, there has been recent interest in learning new similarity measures which take into account this additional information. One such method that learns similarity measures uses a set of features extracted from the neighborhoods of both the floating (I_f) and reference image (I_r) pixels [4]. The algorithm begins by selecting corresponding points in pairs of correctly registered images, ($\mathbf{p}_i$,$\mathbf{p}_i'$). A function, **f**, is then applied to a rectangular image patch centered at each pixel of each pair to extract feature vectors, $y_i=\mathbf{f}(I_f,\mathbf{p}_i')$ and $x_i=\mathbf{f}(I_r,\mathbf{p}_i)$. The similarity measure adopted by this method is computed between individual patches, $s(y_i,x_i)$, rather than entire images as is common. A maximum margin structured output learning method (MMSO) is used to learn the optimal similarity measure from the feature vectors. Although the similarity measures developed by this learning method generate promising results for CT to MR registration of brain images, they struggle for the case of MR to PET registration, where few features are shared between the images.

3 Previous Work

A method to compute the distribution of the error at each pixel was developed by Fitzpatrick et al. for the case of rigid-body point-based registration using Orthogonal Procrustes [2]. They showed that if the image deformation is truly rigid, the registration error depends only on the error in locating the exact geometrical positions of the fiducial points, the fiducial localization error (FLE), not on the object itself. This error exists because of resolution and calibration limitations of the imaging device. In particular, the expected registration error (TRE) at each spatial position (**r**) is given by:

$$\langle TRE^2(\boldsymbol{r})\rangle = \langle FLE^2\rangle\left(\frac{1}{N}+\frac{1}{3}\sum_{i=1}^{3}\frac{d_i^2}{f_i^2}\right),$$

where N is the number of fiducial points, d_i is the distance to the target from principal axis *i*, f_i^2 is the sum of squared distances of the fiducial points from principle axis *i*, and $\langle FLE^2\rangle$ is the expected squared FLE. They further argued that one can estimate the $\langle FLE^2\rangle$ by performing multiple experimental measurements on a phantom where the true fiducial marker locations are known. Thus, the $\langle TRE(\mathbf{r})\rangle$ can be computed for any spatial location. They further extended this work by deriving a formula for the error at **r** along any direction under the assumption that the error is Gaussian distributed with a diagonal covariance matrix. The resulting error estimations computed using these formulas were found to match closely with numerical simulations. However, the applicability of this method is limited as many registration problems do not involve fiducial markers and the assumption of rigid deformation is not always valid.

4 Methodology

4.1 Overview

We propose a new method for error estimation that computes a confidence region in the reference image based on a given significance and a pixel location in the transformed floating image, Figure 2. This functional relationship is determined using supervised machine learning techniques; in particular, the relationship between the confidence regions and the pixels is learned from a set of training data consisting of registered images and their true deformation models. Thus, unlike the algorithm discussed by Fitzpatrick et al., our method is applicable to any registration technique under any type of image deformation and does not require the presence of fiducial markers.

It is insufficient to simply map the pixel location directly to the confidence region. Rather, we adopt a similar approach to Cahill et al. [4] and use the additional information of the image appearance near the pixel. In particular, we describe the pixel using a similarity measure, **s**, between two feature vectors, **f**, one centered at the pixel location (***p***) in the reference image (I_r) and the other at the pixel location in the transformed floating image ($T(I_f)$):

$$\boldsymbol{s}(\boldsymbol{f}(I_r,\boldsymbol{p}),\boldsymbol{f}(T(I_f),\boldsymbol{p})),$$

where **s** may be a vector and T is the estimated registration transform. The problem therefore becomes finding the mapping between the similarity measures and the confidence regions. We make a further assumption that the error is normally distributed (i.e., elliptical confidence regions) in order to reduce the amount of required training data.

Our overall algorithm, Figure 3, can be divided into two categories: training and testing. The training phase consists of determining the relationship between the similarity measures and the Gaussian parameters (i.e., mean and covariance) from a set of registered images with known registration error. The testing phase computes the similarity measure for a given point and uses this to estimate the Gaussian parameters and subsequently a confidence region for that location. The mapping between similarity measure and confidence region is specific to the particular

registration algorithm and to the application (e.g., CT to MR). Thus, one must retrain the algorithm for every new setting.

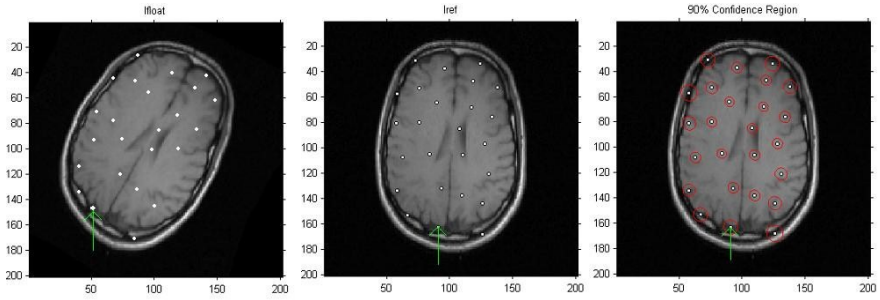

Figure 2: Example demonstrating the confidence region estimation algorithm functionality. The left image presents a floating image with multiple pixels of interest (white dots). The center image shows the same pixels transformed onto the reference image. The right image presents examples of 90% confidence intervals for the pixels of interest (red circles).

4.2 Training Phase

The inputs to the training phase are a set of m registered images $\{T(I_f^k), I_r^k\}_{k=1}^m$ and their associated registration errors $\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$ at known pixel locations $\{\mathbf{p}_1,\ldots,\mathbf{p}_n\}$. The first step in the training algorithm is to extract similarity measures for each known pixel location, $\{\mathbf{s}_1,\ldots,\mathbf{s}_n\}$. The optimal similarity measure, **s**, and feature vector extractor, **f**, used to generate $\mathbf{s}_i$ may depend on the application. For instance, a feature extractor that vectorizes the pixel intensities of a local patch centered at the pixel location, in conjunction with a Euclidean distance similarity measure, was found to work well in MR to MR brain registration. However, a feature extractor that computes the distance between the pixel of interest and the centroid of the image, and a similarity measure that selects this value for the floating image, was found to work well in CT to MR brain registration.

The similarity measures are then normalized to have values on the range [0, 1]. This facilitates the use of multiple similarity measures for regression and was found to improve the regression analysis. One possible normalization method uses the following equation:

$$k(s_i) = e^{(-\frac{s_i' s_i}{2\,BW})},$$

where BW is the estimated variance of the similarity measures and k($\mathbf{s}_i$) is the normalized value of $\mathbf{s}_i$. This normalization is equivalent to the Gaussian kernel when the similarity measure is the square root of the Euclidean distance.

Next, the maximum likelihood estimation (MLE) of the mean ($\hat{\boldsymbol{\mu}}$) and covariance ($\hat{\Sigma}$) of the registration error is computed for all pixels with the same similarity score. There may however be situations when few training pixels share the same similarity score. Thus, in order to obtain reasonable estimates of the Gaussian parameters, the MLE is computed for all pixels with "close" similarity scores. In this case, similarity scores are considered "close" if the norm of their difference is less than some predetermined threshold, ε (i.e., $\|\boldsymbol{s}_i - \boldsymbol{s}_j\| < \varepsilon$). One then calculates a corresponding similarity score for the MLE by taking the mean of the "close" similarity scores. This process results in pairs of vectors ($\mathbf{s}_v$, $\hat{\boldsymbol{\mu}}_v$) and ($\mathbf{s}_v$, $\hat{\Sigma}_v$) for v = 1,…,w where w is the number of

sets of "close" pixels. Furthermore, if a specific structure for the covariance matrix (e.g., $\hat{\Sigma} = \hat{\sigma}^2 I$) is assumed, the estimation process remains the same but the MLE changes accordingly.

The final step in the training phase is to estimate the parameters of a least squares linear regression model for each Gaussian parameter, $\widehat{\boldsymbol{\beta}}_i$, from the training data ($\mathbf{s}_v$, $\widehat{\boldsymbol{\mu}}_v$) and ($\mathbf{s}_v$, $\hat{\Sigma}_v$) for v = 1,…,*w*. This is easily done for both $\hat{\mu}_x$ and $\hat{\mu}_y$ as they are unimodal and assumed to be independent. However, basic regression algorithms cannot be used to estimate arbitrary $\hat{\Sigma}$ as it is multidimensional with mutually dependant components. Rather, we compute the principle eigenvectors and eigenvalues of $\hat{\Sigma}$. We can then perform regression analysis on one of the eigenvectors and both of the eigenvalues as they are independent variables. Since the eigenvectors are orthanormal, one can generate an estimate of arbitrary $\hat{\Sigma}$ from the estimates of these parameters. Depending on the application, one might apply kernel ridge regression or assume a polynomial regression model instead of the linear regression model. However, it was found that kernel ridge regression often overfit the simulation data and did not obtain the same level of performance as the linear regression model.

4.3 Testing Phase

The inputs to the testing phase are the pixel location, **p**, on $T(I_f)$ and the significance level. The first step of this phase is to compute and normalize the similarity score for pixel **p** in the same manner as the training phase. We then input this similarity score into the regression models to obtain estimates of the Gaussian parameters. These Gaussian parameters are then used to compute the appropriate confidence region for the given significance level.

In the case of $\widehat{\boldsymbol{\mu}}$=0 and $\hat{\Sigma} = \hat{\sigma}^2 I$, the confidence region is given by the following formula:

$$\{\forall \boldsymbol{x} \epsilon \mathbb{R}^2 : (\boldsymbol{x} - \boldsymbol{p})'(\boldsymbol{x} - \boldsymbol{p}) < 2s^2 F(2, v, 1 - \alpha)\},$$

where **p** is the pixel location, s^2 is the unbiased estimate of the variance, *v* is the total number of observations minus 2, $(1 - \alpha)$ is the significance level, and F is the inverse cumulative distribution function of the F-distribution [5]. Since the estimated mean was generally found to be close to zero, this confidence region calculation is used whenever one assumes that $\hat{\Sigma} = \hat{\sigma}^2 I$.

In the case of a more complicated $\hat{\Sigma}$, we make the assumption that the $\hat{\Sigma}$ and $\widehat{\boldsymbol{\mu}}$ are correct. This allows for the computation of the radius of the ellipse, *r*, using:

$$P\left((\boldsymbol{x} - \widehat{\boldsymbol{\mu}})'\hat{\Sigma}^{-1}(\boldsymbol{x} - \widehat{\boldsymbol{\mu}}) < r^2\right) = 1 - \alpha,$$

where $(1 - \alpha)$ is the significance level. Thus, the confidence region is given by the following formula:

$$\left\{\forall \boldsymbol{x} \epsilon \mathbb{R}^2 : (\boldsymbol{x} - \boldsymbol{p} - \widehat{\boldsymbol{\mu}})'\hat{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{p} - \widehat{\boldsymbol{\mu}}) < r^2\right\},$$

where **p** is the pixel location. Note that the confidence regions are expressed parametrically as they are assumed to be ellipses.

Goal: Compute a confidence region for a point in the reference image corresponding to a selected point in the floating image

Training Phase

Input : *Registered images* $\left\{I_r, T(I_f)\right\}_{k=1}^{m}$ *and error values* $\{\mathbf{e}_1, \ldots \mathbf{e}_n\}$ *for pixel locations* $\{\mathbf{p}_1, \ldots \mathbf{p}_n\}$

1 Extract similarity features, $\mathbf{s}_i$, for each pixel location $\mathbf{p}_i$

2 Normalize similarity features

3 For $\forall i, j$ s.t. $\|\mathbf{s}_i - \mathbf{s}_j\| < \varepsilon,\ i \neq j, \varepsilon :=$ threshold

Compute the MLE of the mean, $\hat{\boldsymbol{\mu}}$, and covariance, $\hat{\Sigma}$ of the error

Average similarity features: $\bar{\mathbf{s}}$

End

4 Estimate the parameters of a regression model for each parameter β_i

Output : *Model parameters :* $\boldsymbol{\beta}$

Testing Phase

Input : *Pixel location* $\mathbf{p}_i$ *on* $T(I_f)$ *and the significance level*

1 Extract similarity features, $\mathbf{s}_i$, and normalize

2 Apply regression model, f to compute $\left(\hat{\boldsymbol{\mu}}_i, \hat{\Sigma}_i\right)$

3 Use estimated parameters to define a confidence region

Output : *Confidence region*

Figure 3: Outline of the confidence region estimation algorithm. Note the division between the training and testing phases.

5 Experimentation

5.1 Data Generation

We performed several simulations using a simplified version of our algorithm in order to assess its feasibility. In particular, we addressed the rigid registration of unimodal and multimodal brain images. Since the existing registration datasets had limited numbers of points with known error, we had to create a set of training data from previously aligned pairs of images. For each application, a set of floating images (*F*) was created by deforming one of the aligned images (*A*) by a set of known but random rigid transformations. To ensure the image remained in the field of view, the random translation was restricted to [-10, 10] pixels in both the x and y axes and the random rotation was restricted to [-50°, 50°]. The other aligned image is considered the reference image (*R*). Slight nonrigid transformations were then applied to the floating images (*F*) to ensure that the rigid registration algorithm could not perfectly register the images, thereby forcing error.

The exact same operations were also performed on a modified version of the original aligned image (*A*) in which a grid of artificial fiducial markers was introduced by maximizing the intensity of randomly chosen patches. These fiducial markers were also introduced into the reference image (*R*)**.** The original floating (*F*) and reference (*R*) images were then registered using mutual information under the assumption of rigid deformation. Next, the resulting estimated transformation matrix was applied to the floating images containing the artificial fiducial markers. The misalignment of the fiducial markers between the registered floating images and the reference image (*R*) was used to calculate the registration error. This procedure was repeated 100 times for each application to generate 100 pairs of registered images. To validate the accuracy of our learned regression models, a second set of 100 pairs of registered images was generated like

before. However, in this case, the location of the fiducial markers were selected randomly within the brain region in order avoid systematically selecting patches that were used in the training phase.

5.2 Training and Testing

For each application, we applied the previously outlined training algorithm to the set of 100 training image pairs. The particular similarity measures used in the training varied depending on the application as specified in the results section. We made a further simplifying assumption that $\Sigma = \sigma^2 I$. This facilitates an intuitive understanding about the behavior of the algorithm.

We then used the previously described testing algorithm on the set of 100 testing image pairs. Specifically, we apply the algorithm, with a given significance level, on the pixels for which the errors are known. This generates estimates of the desired confidence regions for each of the fiducial locations. Since the errors are known, we can determine if they are in the estimated confidence regions. We then compute the percentage of times the error lies within the estimated region over the entire dataset and compared this to the pre-specified significance level. The results of this investigation are presented in the following section.

5.3 Results

Table 1 summarizes the result of the validation of the confidence region estimation. The similarity measure used was a combination of the following two measures: the Euclidean distance between the pixel intensity values of two local patches and the distance between the pixel of interest and the centroid of the transformed floating image. This similarity measure was adopted because it was found to yield the best confidence region estimates during algorithm development. Although the degree of accuracy of the confidence region varies heavily among different applications, the results obtained were reasonable.

Table 1: Percentage of pixels with error lying within the estimated confidence region.

Floating/Reference Image Modalities	**Significance Levels**				
	80%	**85%**	**90%**	**95%**	**99%**
MR/ MR	80.42	83.25	86.15	90.38	94.46
CT / CT	77.92	79.88	81.79	84.22	86.87
PET / PET	75.16	76.66	78.03	79.14	79.87
CT / MR	79.25	83.40	88.31	94.16	99.17
PET / MR	84.88	87.40	90.25	92.84	95.78
PET / CT	83.61	86.11	88.32	90.65	94.14

The validation table above shows that among the unimodal registration investigated, the MR to MR registration produced the best result, whereas for the multimodal case, the CT to MR produced some exceptionally accurate result. Thus we performed further analysis of these two applications.

MR to MR and CT to MR Test Results

It is interesting to contrast the performance of individual similarity measures for these two applications. Table 2 presents the results of the same validation algorithm as was used for Table 1, but with individual similarity measures.

Table 2: Comparison of percentage of pixels with error lying within the estimated confidence region for individual similarity measures.

	Similarity Measure	Significance Levels				
		80%	**85%**	**90%**	**95%**	**99%**
MR/MR	**Euclidean Norm**	83.01	85.2	88.34	91.72	95.14
	Distance to Centroid	80.36	83.19	86.14	90.33	94.42
	Harris Corner Metric	81.04	83.99	87.05	90.36	94.48
CT/MR	**Euclidean Norm**	77.53	81.31	85.97	92.14	98.56
	Distance to Centroid	77.90	82.54	87.38	93.70	98.97
	Harris Corner Metric	78.51	82.32	86.54	92.37	98.32

We found that the best similarity measure for MR to MR, the Euclidean norm, is not the best for CT to MR, the distance to the centroid. Furthermore, in some situations, the performance from the individual similarity measures exceeds that of the combined similarity measure. However, it was found that the combined similarity measure generally outperformed that of the individual measures. Finally, it is interesting to note that the Harris corner measure, which is the Euclidean distance of the two patches filtered using the Harris corner metric, had similar performance to the centroid.

To further investigate the differences in the performance between these two applications, we look at the scatter plots between the variance and their best similarity measures, the Euclidean norm and the distance to the centroid, Figure 4 (b-c) (e-f). Note that all four plots demonstrate a downward trend with increasing similarity measure value, although at different rates.

Using the same learned regression model as before, we computed the Gaussian parameter estimates at each pixel location within the brain of one test image. It was found that the value of the estimated means $(\hat{\mu}_x, \hat{\mu}_y)$ were roughly zero everywhere within the brain for both MR to MR and CT to MR cases. This result implies that the registration error is close to unbiased for both applications. Alternatively, a more structured pattern was obtained for the estimates of the variance of the error for the pixels within the brain. Figure 4 (a) illustrates the intensity map images representing the estimated variance $(\hat{\sigma}^2)$ at each pixel location within the brain for the MR to MR registration. It is clear that the variance in the error is estimated to be small at the center region of the brain, while it increases radially as the pixel location moves away from the center. A similar pattern, but with higher variance values at all locations, was found for the case of CT to MR registration, Figure 4 (d).

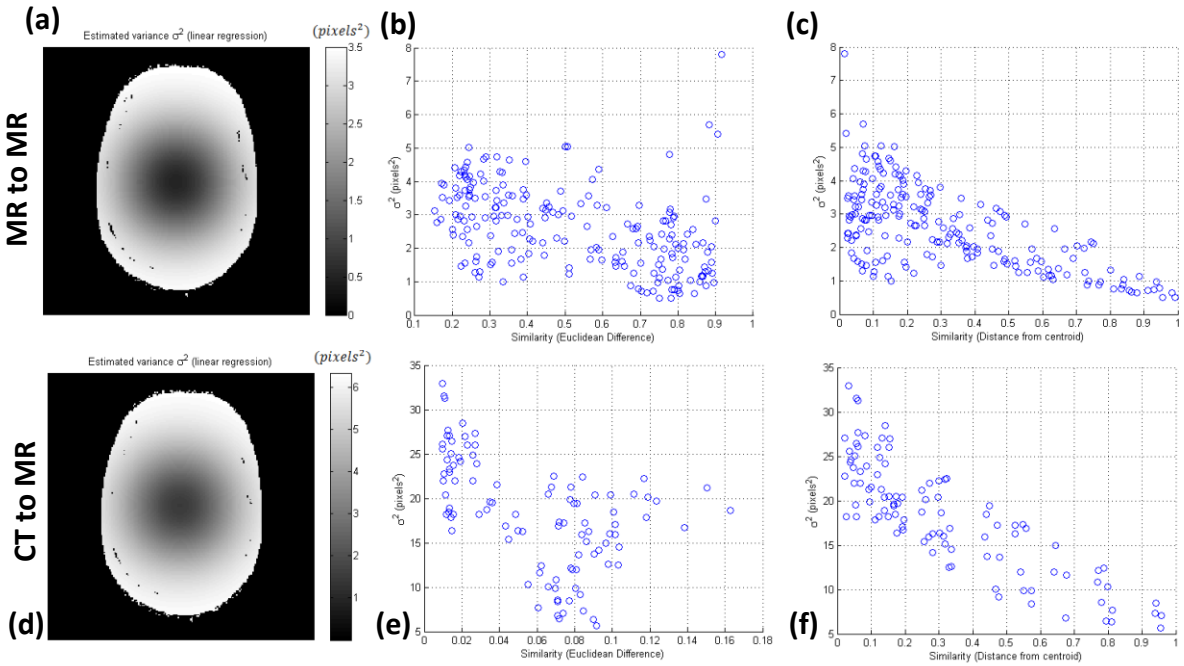

Figure 4: (a) and (d) are the image intensity maps of the estimated variance for MR- MR and CT-MR registration, respectively. (b) and (e) are the scatter plots of estimated variance versus the Euclidean distance similarity measure for MR-MR and CT-MR, respectively. (c) and (f) are the scatter plots of estimated variance versus the distance to centroid similarity measure for MR-MR and CT-MR, respectively.

6 Discussion

The results from Table 1 show a trend that more accurate results are obtained for multimodal registration than for unimodal registration. This agrees with our intuition, because in the unimodal case, the similarity between patches coming from uniform intensity pattern will not be heavily affected by misregistration. For instance, the Euclidean norm will remain near zero for misregistered uniform regions. However, in other regions where there are high variations in the intensity pattern, a perfect registration also has a Euclidean norm near zero. Thus, the algorithm cannot distinguish between these two situations. This is less prevalent in the case of multimodal registration, as patches between the images do not have similar pixel intensities. Hence it makes sense that the CT to MR case produced an exceptionally accurate result, since both CT and MR contain significant structural information that complements one another. Finally, it should be noted that the PET to PET case generated erroneous results because this is an application where the image varies smoothly resulting in many patches that look similar.

The scatter plot shown in Figure 4 (b-c) shows the error variance plotted against the Euclidean distance and centroid distance respectively for the case of MR to MR registration. Here, although only a weak correlation can be seen for the Euclidean distance measure, a relatively strong negative correlation can be observed for the distance to centroid measure. This negative relationship between the variance and the similarities is shown to be much stronger in the case for CT to MR registration, as depicted in Figure 4 (e-f). This matches our intuition that a similarity measure that has high correlation with the parameter of interest will significantly help improve the confidence region estimation. Lastly, it should be noted that since the important similarity measures seem to be negatively correlated, applying nonlinear regression methods such as kernel ridge regression is not appropriate here.

The intensity map image representing the estimated variance ($\hat{\sigma}^2$) at each pixel location within the brain for the MR to MR registration shows that the variance value increases as one moves away from the center of the image. This behavior was expected, since the shift caused by rotation grows larger as one move away from the center of rotation. Hence, a slight misregistration of the rotation parameter can cause significant deviation in the outer edge of the brain. Figure 4 (d) reaffirms this trend for the case of CT to MR, but with higher level of estimated variance throughout the brain.

Finally, a few notable computer vision descriptors, such as the Harris corner metric and SIFT were also used as similarity measures for the patches. However, they generally did not provide meaningful improvement to the confidence region result. In fact, they often degraded the result, particularly in the case of SIFT. In general, these descriptors tend to provide meaningful information only at specific points in an image (e.g., corners or edges). Thus, they usually do not provide any discrimination benefit when applied to arbitrary point.

7 Conclusions

Based on the results, our proposed algorithm shows promise. For instance, the computed error estimates for the case of multi-modal registration are close to the true values. However, there is still significant room for improvement; particularly, in unimodal registration and applications that involve PET. As we have previously discussed, the choice of the similarity measure significantly affects the performance of the algorithm. Therefore, a reasonable starting point to improving our algorithm is to test similarity measures that are specialized for each method.

The main advantage of our current framework is that it is applicable to any registration technique under any type of image deformation and does not require the presence of fiducial markers. Although we may use different similarity measures for different applications the general framework of the algorithm remains the same. However the method is not without its limitations. Specifically, the efficacy of our algorithm relies on a fairly large training dataset. Such a dataset would require a significant amount of time to require.

The next steps in the development of our algorithm include the exploration of a larger variety of similarity measures across several registration algorithms in order to improve our error estimates. Furthermore, we must apply our algorithm to a real world dataset in order to evaluate its performance under real world uncertainty. Finally, we should extend our investigation of our approach to non-rigid registration applications.

References

[1] J. V. Hajnal, D. l. Hill, and D. J. Hawkes. Medical Image Registration. CRC Press: Boca Raton, 2001.

[2] J. Michael Fitzpatrick, J. B. West, and C.R. Maurer, Jr. Predicting Error in Rigid-Body Point-Based Registration. *IEEE Transactions on Medical Imaging*, Vol. 17, (5): 1998

[3] J. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions of Medical Imaging*, (8):986-1004, 2003.

[4] D. Lee, M. Hofmann, F. Steinke, Y. Altun, N. D. Cahill, and B. Schölkpof. Learning similarity measure for multi-modal 3D image registration. In *CVPR*, 2009.

[5] N.R. Draper and H. Smith. Applied Regression Analysis. John Wiley & Sons, Inc.: New York, 1998.

CONTRIBUTIONS

Michael Allison evaluated similarity measures and set up the confidence region validation. Bhargav Avasarala explored the SIFT similarity measure and implementations of the Cahill algorithm. Takanori Watanabe generated training and testing data and evaluated similarity measures. Everyone contributed equally to the final report