Quantifying effects of deformable CT-MR registration

FREDRIK ISAKSSON
Quantifying effects of deformable CT-MR registration

FREDRIK ISAKSSON

Master’s Thesis in Mathematical Statistics (30 ECTS credits)
Master Programme in Applied and Computational Mathematics (120 credits)
Royal Institute of Technology year 2016
Supervisors at Elekta: Markus Eriksson
Supervisor at KTH: Timo Koski
Examiner: Timo Koski

TRITA-MAT-E 2016:25
ISRN-KTH/MAT/E--16/25-SE

Royal Institute of Technology
SCI School of Engineering Sciences
KTH SCI
SE-100 44 Stockholm, Sweden
URL: www.kth.se/sci
Abstract

Rigid image registration is an important part of many medical applications. In order to make correct decisions from the registration process the uncertainty of the results should be included. In this thesis a framework for estimating and visualising the spatial uncertainty of rigid image registration without groundtruth measurements is presented. The framework uses a deformable registration algorithm to estimate the errors and a groupwise registration for collocating multiple image sets to generate multiple realisations of the error field. A mean and covariance field are then generated which allows a characterisation of the error field. The framework is used to evaluate errors in CT-MR registration and a statistically significant bias field is detected using random field theory. It is also established that B-spline registration of CT images to themselves exhibit a bias.
Sammanfattning

Acknowledgements

First and foremost, I would like to thank Elekta generally for providing me with the opportunity to work on an interesting project and specifically my Supervisor Markus Eriksson for guidance and support. My gratitude also goes out to my supervisor at KTH, Timo Koski for making this thesis possible.
## Contents

1 Introduction .......................................................... 1  
  1.1 The Leksell Gamma Knife ........................................ 1  
  1.2 MRI and Geometric distortions ................................. 2  
  1.3 Thesis goal & layout ........................................... 2  

2 Background ........................................................... 4  
  2.1 Image registration .............................................. 4  
    2.1.1 The Similarity measure .................................... 5  
    2.1.2 Transformation models .................................... 7  
    2.1.3 Optimization .............................................. 10  
    2.1.4 Interpolation ............................................. 11  
    2.1.5 Related works regarding Uncertainty measure ............ 12  
  2.2 Group-wise registration ......................................... 13  
  2.3 Statistical tests ............................................... 14  
    2.3.1 Multivariate tests ....................................... 16  
    2.3.2 Random Field theory ..................................... 20  

3 Method ................................................................. 24  
  3.1 Deformation field generation ................................... 24  
    3.1.1 CT-MR registration ...................................... 24  
    3.1.2 Baseline deformation ..................................... 27  
    3.1.3 Comparisons .............................................. 27  
  3.2 Implementation details ......................................... 27  
    3.2.1 Code implementation ..................................... 28  
    3.2.2 Visualization ............................................ 28  

4 Results ................................................................. 30  
  4.1 Data-selection .................................................. 30  
  4.2 Auto deformation bias fields ................................... 30  
  4.3 MR-CT Bias deformation field .................................. 31  
  4.4 Prediction visualization ....................................... 33  
  4.5 Model validation ............................................... 34
Chapter 1

Introduction

Image registration, the process of finding a correspondence transform between images, is an important component in many medical applications relating to image processing [1]. It relates to essentially all body parts and image modalities [2], as well as treatment evaluation [3] and population models [4]. Elekta Instrument AB uses image registration between tumour delineated magnetic resonance images (MRI) and cone beam computed tomography (CBCT) images for positioning of patients during radio surgery using the Leksell Gamma Knife® Icon. The focus of this thesis will thus be multimodal image registration, specifically between MR and CT images of the brain where a ground truth measurement is not available. Specifically an assessment of the uncertainties in the registration arising from geometric distortions of the MR image. For this purpose we will make use of a large set of clinical MR and CT images to estimate the errors and uncertainty of rigid registration.

1.1 The Leksell Gamma Knife

The Leksell Gamma Knife is a Gamma-radiation based radio-surgery device built on the principle of focusing radiation from 192 Co$^{60}$-sources on a target. The system uses a mounted CBCT system to locate and align the patient’s position in relation to the radiation iso-center.

Prior to treatment the target area is delineated using the superior soft-tissue properties of an MR image and a treatment plan is created using this information. The MR image is then registered to the CBCT image of the Gamma knife which will position the patient such that the iso-center will target the points given by the treatment plan. This approach allows the utilization of both the superior contrast properties of the MR image and the spatial accuracy of the CBCT image. The registration step in this process is based on a mutual information registration approaches.
1.2 MRI and Geometric distortions

Magnetic Resonance imaging has become a preferred technique in applications requiring soft tissue distinctions such as neurosurgery [5] and tumour delineation [6]. A specific area where MRI is used is dose planning of brain cancer. There are however drawbacks with using MRI for delineation in applications sensitive to spatial errors, one such example being radiation surgery of tumours that is located close to the facial nerve. The drawbacks with MRI primarily stem from the geometric distortions of MRI causing the wrong area to be targeted during treatment. There are three sources of distortions in MRI[7, 8] that can contribute to this miss-targeting, the first is machine related inhomogeneities such as imperfect gradient fields and eddy currents. The second is target related distortions, mainly a variable magnetic susceptibility $\chi_v$ in the subject and chemical-shifts at material edges. The third and final source can be summarised as corrupting objects, this can be everything from implanted metal shunts[8] to mascara, tattoos and hair products [9, 10] that distort the image. Some of these effects can be corrected for, especially from the first group by using phantom models and specific image sequences[11] other effects are harder to correct for and often require time consuming sequencing modalities. The effects of the distortions that vary between subject can be up to the order of millimetres [12, 8, 13], different imagine modalities will also exhibit varying degrees of distortions. These distortions could be considered to be small but in the case of tumours a difference of these sizes can greatly impact the dose received by the facial nerve and accurate results could reduce complications such as facial paralysis. In the context of the Leksell Gamma Knife® that has sub-millimetre precision a shift on the order of a millimetre in the position of the target area could thus be the dominant error in the treatment and thus an interesting target to study in order to be able to present a more accurate description of the errors in a treatment method to clinicians.

1.3 Thesis goal & layout

The aim of this thesis will be to characterise the deformations and the uncertainties induced in the MR images from a clinical treatment setting when registered to a CT image with an aim to provide additional information to the clinical decision making process.

To do this the thesis has been divided into five chapters.

The following chapter, chapter 2 summarises the underlying theory used in the thesis as well as a description of some related areas of uncertainty estimation in image registration. Chapter 3 describes the process of generating deformation fields and how to spatially normalize them, as well as how a baseline deformation is generated. In chapter 4 the results are present and
the implications are discussed in chapter 5 where areas of further work are also described. Finally conclusions are formed and the thesis is summarized in chapter 6.
Chapter 2

Background

This chapter covers the established theory and previous results necessary for understanding the methods and the context of the thesis. It aims to give a beginner’s introduction to image registration, a refresher on basic statistics and how it extends into higher dimensions.

2.1 Image registration

Image registration, sometimes also called image fusion or co-registration, is informally the process of finding a common coordinate system for two or more images. It applies to overlaying two images from different sensors showing different characteristics of the same subject, creating a time sequence of images from the same sensor and finding related structures in groups of different subjects.

To understand image registration we begin by establishing a mathematical model of an image, a image is defined to be a mapping from a spatial domain $\Omega$ to and intensity domain $Y$, $I : \Omega \rightarrow Y$. For a greyscale image the intensity domain could be $Y_{\text{Grey}} = [0, 2^M]$ and for a color image it could be $Y_{\text{RGB}} = [0, 2^M]^3$, the spatial domains $\Omega$ are commonly rectangular subsets of $\mathbb{R}^d$.

Formally the registration process can be considered as finding the transformation $T : \Omega_F \rightarrow \Omega_M$ that maximises some similarity criterion $C(I_F, I_M; T)$. The F subscript stands for the “fixed” image and the M for the “moving” image, this might seem a bit counter-intuitive since the transform $T$, is from the fixed domain into the moving domain, but it refers to the fact that it is the fixed image $I_F$ and the transformed moving image $I_M \circ T$ that are being visualised and not the direction of the underlying mathematical transform. The process for computing the registration can be divided into four primary components:

- The similarity measure
• The transformation model
• The interpolation strategy
• The optimization method

It is also possible to study a fifth component of the registration that is not as commonly included, that is an estimation of the uncertainty or the quality of the registration.

2.1.1 The Similarity measure

The first component, the similarity measure, has the purpose to quantify how similar the appearances of two images are, thus two images are considered similar when the similarity measure is maximized between them. A great deal of different measures have been proposed for registration as described in the review of image registration by Oliveira and Tavares[2]. Many similarity measures are limited in which problems they can handle, for example one of the simpler measures, the Sum of Square Differences (SSD) measure is straightforwardly defined by its name as

\[ C_{SSD}(I_F, I_M; T) = \sum_{x \in S} (I_F(x) - (I_M \circ T)(x))^2, \tag{2.1} \]

where \( S \) is some sampling space i.e. the grid points of a stochastically selected subset. The SSD measure has the problem that it requires that the intensities of the two images are similar. This holds for images of the same modality but when comparing images of different modalities the assumption fails, making the measure highly unsuitable. Note that it is not sufficient that the intensities of the two images are in the same range for the SSD measure to be suitable, since it assumes an identical relationship between the intensities in the two images.

A popular class of measures that is applicable to a wide array of problems since it assumes very little about the intensity relationship are information and entropy based measures with Mutual information (MI) [14] probably being the most used. Most of these measures have their origin in information and probability theory and are based on measuring the dependence of probability functions describing the intensities in the images. Inherent to the use of all similarity measures is the assumption that the measure will have its maximum when the images are correctly aligned. This assumption is weakly motivated for MI by Zöllei, et. al. [15] in the fact that image alignment constitutes a local maximum for rigid transformation models but this is not something that holds generally [16] for other measures or transformations.
Mutual information is based on the Kullback-Leibler divergence, $D_{KL}(P||Q)$, and deals with the intensity distribution in the images. Informally MI finds a correspondence between the intensities of the images and aligns the images so the corresponding intensities overlap often. More formally the mutual information is defined in terms of the joint and marginal distribution of the intensity distributions $A$ and $B$ of the two images. That is $p_A(a)$ gives the ratio of how often the intensity $a$ appears in image $A$, and the joint distribution $p_{AB}(a,b)$ describes how common it is for the intensities $a, b$ to occupy the same voxel, the full definition is then defined as

$$MI(A, B) = D_{KL}(P_{AB}||P_A P_B) = \sum_{a,b\in Y_A \times Y_B} p_{AB}(a,b) \log \frac{p_{AB}(a,b)}{p_A(a)p_B(b)}.$$  

(2.2)

In practical implementation such as in the Elastix toolkit [1] or ITK [17] the sum is not taken over all possible intensity values, instead the intensity ranges are divided in $L_A$ and $L_B$ histogram bins and the joint probability is estimated using B-Spline Parzen windows as proposed by Thévenaz and Unser[18], it will also include the transformation model in the evaluation. Thus the mutual information measure definition calculation will follow the definition given by Elastix as:

$$C_{MI}(I_F, I_M; T) = \sum_{m\in L_M} \sum_{f\in L_F} p_{FM}(f, m; T) \log_2 \left( \frac{p_{FM}(f, m; T)}{p_F(f)p_M(m; T)} \right),$$  

(2.3)

with $p_{FM}$, the joint probability distribution, estimated with the B-spline Parzen windows:

$$p_{FM}(f, m; T) = \frac{1}{|\Omega_F|} \sum_{x\in\Omega_F} w_F(f/\sigma_F - I_F(x)/\sigma_F) \times w_M(m/\sigma_M - (I_M \circ T)(x)/\sigma_M),$$  

(2.4)

where $w_F$ and $w_M$ are the Parzen windows and $\sigma$ represents the bandwidths of the underlying histogram bins. The marginal distributions are easily obtained by summing over the joint estimate. The reason that we use these Parzen windows in the estimation is to avoid the undesired properties that comes from normal histograms.

The mutual information can also be related to the Shannon entropy of the images as:

$$MI(A, B) = H(A) + H(B) - H(A, B).$$  

(2.5)

where $H(\cdot)$ and $H(\cdot, \cdot)$ denotes the individual and joint Shannon entropies respectively. The entropy formulation is useful for relating MI to similar measures such as joint entropy and Normalised Mutual information (NMI).
that aims to address problems of non-overlapping areas. A survey of mutual information in registration of medical images can be found in [19]. Mutual information will be the measure used in the rest of the thesis.

A final approach that can be used as a similarity criterion are landmark or feature based measurements in contrast to the intensity based methods described above. Feature based measurements are based on identifying corresponding points or areas in both images and measuring the distance or overlap of these features, either by calculating the mean square error of landmark locations or by using an overlap measure such as the Dice criterion to quantify overlap[20]. These methods are based on being able to accurately locate these features and generally require expert delineation or accurate segmentation algorithms and are thus dependent on how well one can find corresponding points in the two images, this problem is in some sense a sparse image registration problem.

2.1.2 Transformation models

There are two main classes of transformation models, rigid and non-rigid registration. Rigid registration assumes that the transformation consists of only translations and rotations, if scalings are added the transform is called affine. Thus a rigid transformation model assumes that the geometry of the two images are the same, implying that no distortions or biological shifts exists in the given geometry. Non-rigid registration on the other hand does not assume these constraints and allow more general deformations. Allowing completely free deformations on the other hand gives rise to an ill-posed problem with methods creating invalid images by introducing things as folding of the image. Thus non-rigid registration requires regularization [21] to be able to solve the problem. There are many ways to regularize the problem, some based in physics, others based on splines and some on prior-knowledge to name a few [22].

Rigid registration

The rigid estimation problem for 3D is in its essence a 6 (or 9 in the affine case) parametric estimation problem that is solved as an optimization problem over the parameter space with the similarity criterion as objective function. The parametrization is over translations and rotations of the image, in the affine case scaling and possibly shears are added. The problem can be formulated as chained matrix operations describing the four different actions as characterized in the thesis by Ashburner[23]. The definition of the rigid case in spatial coordinates with a parameter set $\mu$ becomes

$$T_\mu(x) = R_\mu(x - c) + t_\mu + c,$$  \hspace{1cm} (2.6)
and for the affine case,

\[ T_\mu(x) = A_\mu R_\mu(x - c) + t_\mu + c, \] (2.7)

where \( A_\mu, R_\mu, t \) and \( c \) are a scaling matrix, rotation matrix, translation vector and center of rotation vector respectively. This transform is often expressed as a four-vector described by Ashburner in order to formulate the entire transform in a single matrix as:

\[ T_\mu = \begin{bmatrix} A_\mu & R_\mu & t_\mu \\ 0_{3\times1} & 1 \end{bmatrix}, \] (2.8)

with the four-vector defined as:

\[ [x \ 1]^T. \] (2.9)

It is also possible to convert spatial locations into the index representation, this requires the images origin location in space and the voxel spacing in all directions. Using this information the transform matrix is formed as

\[ O = \begin{bmatrix} s_x & 0 & 0 & o_x \\ 0 & s_y & 0 & o_y \\ 0 & 0 & s_z & o_z \\ 0 & 0 & 0 & 1 \end{bmatrix}, \] (2.10)

and maps spatial coordinates to voxels. In the matrix \( s_* \) represents the voxel spacing and \( o_* \) the image origin. The upside with this matrix representation is that the inverse transforms are easily calculated by inverting the transform matrices. Finding an inverse in the non-rigid case however is not as simple and sometimes not even possible.

**Non-rigid registration, B-splines approach**

The non-rigid registration case can in its most general case be formulated as a displacement field

\[ T_D(x) = x + D(x), \] (2.11)

where \( D \) is an essentially arbitrary function. This formulation is based on estimating \( d \) unknowns for each grid point in the estimation and thus the image registration will be an under-determined problem. To avoid the ill-posed problem the deformation field or the transformation is constrained in some way. This constriction of the problem is not only motivated by the estimation problem but is also connected to the concept of ”correctly aligned” assumption that was discussed with regard to the similarity measure in section 2.1.1, that is the assumption that the images are correctly aligned when the similarity criterion is maximized. If an unconstrained deformation is allowed then it is unlikely that most of the commonly used intensity similarity
criterias will produce reasonable transformation as demonstrated with the Completely Useless Registration Tool (CURT) [24] that simply orders the intensities of the two images and overlays them. The resulting images perform very well in most metrics and produce images that would be convincing to most experts, however the spatial information from the moving image is completely lost i.e. allowing free deformations will in some sense abandon the notion that the image taking process is spatially accurate in any way. As such it becomes clear that the non-rigid transform requires some constraints.

There are many ways of regularizing the deformable transformation, the most common way is to restrict the transformation to a parametric model. The parametric models usually only estimate small deformations and thus parametric models are often on the form:

$$T = T_{\text{global}} + T_{\text{local}},$$

(2.12)

where the global registration is a rigid or affine transformation. Rigid and affine registration is a degenerate case of this model were the local model is zero. Two common parametric models are B-splines and Thin-plate splines that are quite similar to each other. B-splines where first proposed by Rueckert et. al. [25] and consists of a product of $d$ (cubic) B-spline kernels $\beta^3$ on a regular grid $N_x$ over the fixed image giving the form of the local component as

$$T_{\text{local}}(x) = x + \sum_{x_k \in N_x} p_k \beta^3 \left( \frac{x - x_k}{\sigma} \right),$$

(2.13)

with $\sigma$ denoting the control point spacing. The advantage of the B-splines compared to many other models, including thin-plate splines, is that they are locally supported and are thus computationally efficient, the size of the overlaiden grid can also be tuned to suit most applications.
Another way of restricting the transformation, that can be combined with parametric models, is to introduce a regularizing term \( C_{\text{reg}} \) in the similarity criterion. In this case one aims to maximize

\[
C(I_F, I_M; T) = C_{\text{sim}}(I_F, I_M; T) - C_{\text{reg}}(I_F, I_M; T),
\]

instead of only the first term on the right hand side of the equation. The regularizing penalty term aims to ground the transformation in reality, proposed terms are based on rigidity [26], the determinant of the Jacobian [27], the bending energy of the transformation [28] and many more. One can also introduce hard constraints to the transformation, such as restricting folding of the image structure. Folding can be quantified as having a negative determinant of the Jacobian of the deformation field or as having a folded grid as shown in figure 2.2. The major reason of disallowing folding and thus restricting the determinant of the Jacobian of the deformation field to be strictly positive is to guarantee that the transform will be locally invertible, which is guaranteed by the inverse function theorem given the constraint on the determinant.

![Figure 2.2: A transformation that has folded the grid.](image)

There also exist non-parametric registration methods such as the DEMOns algorithm [29], however a further description is omitted since the methods are not utilized in this thesis. They could however be used in the framework of the thesis if one desires to.

### 2.1.3 Optimization

As mentioned previously the registration procedure is an optimization problem over the parameters \( \mu \) defined by the transformation with respect to the similarity criterion i.e.

\[
\arg \max_\mu C_{\text{MI}}(I_F, I_M; T_\mu).
\]

This problem is solved numerically and is computationally demanding. The Elastix toolbox can be used to solve the problem and the implementation
details of the optimizer are described in [1, 30]. In particular an iterative adaptive step-size gradient descent approach is used to solve the problem. Two major improvements are added upon the optimization, the first aims to make the optimization more robust with respect to the initial transform and the other aims to decrease the run time.

The first addition can be motivated by the fact that in many cases the initial transform for multi-modal images will result in no overlap at all between the two images. This stems from each image being defined in spatial coordinates in relation to some reference point in its imaging device, thus different imaging devices will produce different offsets based on the local geometry. To be able to correct for this without dramatically increasing the run-time or reducing the accuracy a multi-resolution optimization scheme is used. The multi-resolution scheme is based on solving the problem hierarchically on different scales by smoothing the image, commonly with a Gaussian kernel filter, and using larger step-sizes to solve the coarse problem. The solution of the coarse solution will then be used as the initial alignment of the next, less smoothed, problem. This approach is repeated until the originally problem is solved.

The approach looking to decrease the run time of the solver is based in stochastic-sampling of the problem. Instead of solving the normal gradient descent problem, the stochastic method of Robbins-Monro is used to simplify the estimation of the gradient. This approach can also be combined with the evaluation of the similarity criterion, for example in the case of Mutual information, the sum in equation (2.4) can be taken over a subset of the voxels instead of the entire image. This approach has empirically been show to have similar accuracies as the full search while greatly reducing the run-time by only looking at a small subset of the voxels (∼2000) [30]. This stochastic approach becomes necessary for large images and if the sample points are selected randomly for each iteration no bias will be introduced. It can be noted that in the rigid case the standard deviation in the translation resulting from different seeds to the random generator is on the level of 0.1 mm i.e. less than the accuracy of the images.

2.1.4 Interpolation

In order to perform the steps in the previous sections one practical problem arises relating to the discreteness of a image. The registration rests on evaluating $I_M \circ T(x)$ for the location of voxels in the fixed image. The estimated transform is however a continuous function and the transformed values $T(x)$ will thus rarely coincide with a spatial grid point where a value for $I_M$ has been observed. It is thus necessary to extend our discrete grid of values in the moving image space to the whole region. This is done by interpolation
using B-splines for values inside the moving image grid and by asserting 0 outside of the grid to accommodate for cropped images in the algorithm. Other approaches such as a nearest neighbour method are also possible.

This interpolation step introduces some extra uncertainty to the registration and performing multiple interpolations should be avoided when possible. Multiple interpolations arises when creating a concatenated transform $T_C = T_A \circ T_B$ and an intermediate image $I_M \circ T_A$ is used as the moving image when calculating $T_B$. The correct way of calculating a concatenated transform so that only one interpolation is performed is by using the original fixed and moving images and specifying that the model of $T_B$ is of the form $T_B(T_A(x))$. Thus chained image generation should be avoided when possible.

### 2.1.5 Related works regarding Uncertainty measure

After having performed the above steps to register two images the question of the quality of the registration naturally arises, typically the way to investigate this is to use a ground truth measurement. One form of ground truth typically used for rigid registration is to include fiducial points in both images, this allows accurate determination of the uncertainty in the registration of spatial correct images given at least three non-collinear points [31, 32]. However the problem with ground truth measurements is that they are typically not available in practice, for example the deformations induced by the brain in the MR-image are not easily measurable and of a dense character such that the deformation at the surface of the skull doesn’t necessarily describe the deformation inside the brain. With this in mind the need for uncertainty measures independent of ground truth becomes clear. Many conditions that are necessary, but not sufficient for a correct registration are sometimes used as a surrogate estimate for the quality of the registration, these surrogates usually relates to ensuring the precision of the algorithm and not the accuracy using the reasoning that a good accuracy requires a good precision.

Heuristically it is clear to understand that the certainty of the registration should be more certain where there are clear gradients in the images, i.e. edge matching should be more certain than areas where no clear features exist. A way to use this to estimate the uncertainty stems from stereo vision and is described in [33], the method is based on modelling the uncertainty as the stability of the cost function during small perturbations. In this way perturbations in homogeneous areas will have very small effects on the cost function so the location is uncertain, on the other hand areas with large-gradients will have a very big change, signalling a large certainty. The downside with this approach is that it will only produce qualitatively result
of the certainty, it does not give information of how large or in what direction the actual spatial misregistration might be. A way to get more quantitative estimates of the uncertainties stemming from the registration algorithm itself is to use a bootstrap approach [34]. The bootstrap approach requires a lot of computations to estimate the covariance of the estimated parameters, thus it will become increasingly demanding as the degrees of freedoms increase. A similar approach but in a Bayesian framework is proposed in [35], here the computational load is moved to a MCMC simulation of the posterior distribution. This Bayesian framework is essentially a full registration algorithm where the deformation is estimated as the maximum likelihood of the deformation distribution. One of the major problems with determining the uncertainty in the registration that both of the computational intensive methods try to circumvent is the fact that only one observation is available of the transform, in [36] this is dealt with directly when trying to estimate the covariance of the parameters by imposing a-prior expectations on the form of the covariance matrix, the estimates are then used to construct confidence regions. The general problem formulation described in [36] will be pursued in the thesis with an attempt to create more observations of the deformations by spatial collocating multiple image sets in order to circumvent the single outcome problem.

2.2 Group-wise registration

The conventional use of image registration is the binary problem described above involving two images of the same subject, there are however situations where one has multiple images that one wants to register. One such situation is morphological and atlas studies of different subject’s brains, in such a case one has to find correspondence between the structures in the brain. Such a correspondence can be found via registration into some standard space. The selection or creation of the standard space is however a complicated business. The naive approach to group-wise registration is to select a template image that defines the standard space and then perform pairwise deformable registration of all images to this template. This approach is very straightforward to understand, easy to implement and scales linearly, $O(n)$, with the amount of images. The downside is that the selection of the template image is arbitrary and will bias any results. A way to avoid the bias of the template image is to construct a group mean image and use that as the template. The standard mean image achieved by performing affine registration and then calculating the mean of each pixel is however not a suitable template since the mean image will be fuzzy and have lost certain features. This can be thought of as the brain images laying on a manifold in the full image space, the mean image will however fall outside of this manifold. To avoid this problem several approaches have been proposed including
estimating a sharp-mean image [37] and taking the manifold structure into account [38, 39, 40]. The downside with these approaches are that they are slow and computationally intensive since they take all images into account when creating the mean. In the evaluation of the eHUGS implementation [40] runtimes between 40 and 640 hours where reported for different algorithms for a 80 image set on a 10 core, 2.0 GHz CPU, 256GB/1066MHz RAM machine. These runtimes can be restrictive to the application of the improved estimate algorithms.

It is however possible to combine the manifold techniques with the pairwise registration in order to balance the downsides of both process by using a sub-sampled set of images to create the template image for the pairwise registration method. This will allow a user defined balancing between runtime and accuracy. An illustration of the advantage of the manifold approach can be seen in figure 2.3.

![Figure 2.3: Schematic illustration of different group-wise algorithms. The line represents brain images and the means the results of the registrations.](image)

2.3 Statistical tests

Statistical test for significant univariate effects are well known in science at least for the univariate case. In this section the notions of statistical tests are repeated, some important considerations are presented and how they can be extended to high-dimensional cases is described.

One of the first statistical test one encounters is the location test of the mean-value, with unknown variance and a normality assumption, i.e.
Student’s $t$-test. It is based on having $n$ i.i.d. observations $x_i$ from a normal distribution with unknown parameters and has the test statistic:

\[ H_0 : \mu = \mu_0, \]
\[ H_1 : \mu \neq \mu_0, \]
\[ t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}, \]

(2.16)

where $\bar{x}$ is the mean and $s$ is the standard deviation. One rejects the null-hypothesis if the observed quantity $t$ is larger than the predefined $\alpha$ quantile of the $t$ distribution. This test can also be viewed in a regression framework as a null regression model, that is

\[ y = X\beta + \epsilon. \]

The test above then checks if $\beta = \mu_0$, when $X = [1 \ldots 1]^T$. Generally the $t$-distribution can be used to test the significance of all predictor variables in an ordinary linear regression model. One is often also interested in visualizing the uncertainty of the estimation by creating a confidence interval, both for the parameter $\mu_0$ as well as a new observation of the regression model i.e. $y_{new}$. An interval for $\mu_0$ is easily given by rearranging the last part of (2.16) under the null hypothesis as:

\[ 1 - \alpha = P \left( t_{\alpha/2} < \frac{\bar{x} - \mu_0}{s/\sqrt{n}} < t_{1-\alpha/2} \right) = P \left( \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}} < \mu_0 < \bar{x} + t_{1-\alpha/2} \frac{s}{\sqrt{n}} \right), \]

and thus the interval is

\[ \left( \bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{1-\alpha/2} \frac{s}{\sqrt{n}} \right). \]

(2.17)

(2.18)

For the prediction of a new observation the uncertainty in both our estimate of $\mu_0$ as well as the inherent noise level of $\epsilon$ has to be taken into account. For the variance of $y$ it holds, given that the assumptions for a linear regression is true,

\[ \text{var}(y) = \text{var}(X\hat{\beta} + \epsilon) = X \text{var}(\hat{\beta}) X^T + \text{var}(\epsilon) = s^2 (X^T X)^{-1} + s^2. \]

(2.19)

From this an interval that will contain $1 - \alpha$ percent of new observation for the null regression where $(X^T X)^{-1} = \frac{1}{n}$,

\[ \left( \bar{x} - t_{\alpha/2} s \sqrt{1 + \frac{1}{n}}, \bar{x} + t_{1-\alpha/2} s \sqrt{1 + \frac{1}{n}} \right), \]

(2.20)

can be created.
**Multiple testing problem**

A problem that arises from the testing framework is the multiple testing problem, stemming from the fact that the tests performed has a $\alpha$ % (usually 5%) risk of rejecting the null hypothesis even though it is correct, a so called Type 1 error. When multiple test are performed the risk that one of the tests fails when it shouldn’t increases, for example if one performs 20 independent tests at a 95% confidence level it is expected that one of the tests will reject the null-hypothesis. In the case of image volumes the multiple testing problem can be a big problem, if the volumes consists of $(512 \times 512 \times 256)$ voxels and a test is performed for each voxel (at a 95% confidence level) more than 3 million false positives are expected.

There are however ways to deal with the multiple testing problem, the simplest is the Bonferroni correction that simply adjusts the significant level of each test to compensate for the multiple testing problem and achieve a global (family-wise) confidence of $\alpha$. The individual tests confidence level after the Bonferroni correction is $\alpha_{\text{new}} = 1 - \left( \frac{1 - \alpha}{k} \right)$ where $k$ is the number of tests. The downside with using the Bonferroni correction is that it is very conservative in that it assumes nothing about the problem. Thus it corrects for the worst case where there is no correlation what so ever between the tests, this is not the case in our image case where there is a clear spatial correlation in the images. A way to account for the correlation in the images is further described in section 2.3.2. In section 2.3.1 it is also described how accounting for correlation between tests allows one to perform infinite many tests without letting the confidence level for each test go to 1.

The multiple testing problem for our image case can also be reduce by downscaling the image to drastically reduce the number of tests or in the case of tumour treatment focusing the tests in a certain area around the tumour.

**2.3.1 Multivariate tests**

When generalizing univariate tests into higher dimensions it is necessary to adapt the tests to include effects that take the entire vector structure into account, i.e. it is not sufficient to just look at each component interdependently since the components can be correlated. Essentially a component-wise result can look reasonable while the combined result becomes extreme, this becomes obvious when looking at figure 2.4.
Hotellings $T^2$ test

The generalization of the Student $t$-test to the multivariate case is based on Hotellings $T^2$ distribution, a generalization of the Student $t$-distribution to the multivariate case. The underlying assumption for a Hotelling test is that the observed outcomes are from a multivariate Gaussian distribution with unknown mean and variance. The one-sample test with $n$ observations then has the following form:

\[
H_0 : \mu = \mu_0, \\
H_1 : \mu \neq \mu_0,
\]

\[
T^2 = n(\bar{x} - \mu_0)^T S^{-1}(\bar{x} - \mu_0),
\]

where $S$ is the estimated covariance matrix and $T^2$ has a $T^2(d, n - 1)$ distribution under the null hypothesis. The Hotelling test is however commonly compared to the $F$-distribution instead of of the $T^2$-distribution using the relation

\[
T^2(d, n - 1; x) = \frac{nd}{n - d + 1} F_{d,n-d+1}(x).
\]

In this case the generalisations of the confidence intervals become confidence ellipses, principally defined by the eigenvectors and values of the covariance matrix, in a similar case to the univariate case. i.e. for the confidence and prediction ellipses it holds:

\[
CI = \left\{ \mu \in \mathbb{R}^d | (\bar{x} - \mu)^T S^{-1}(\bar{x} - \mu) \leq \Phi_{T^2(d,n-1)}(1 - \alpha) \frac{1}{n} \right\}
\]

(2.23)
\[ \mathbf{x}_{\text{new}} \in \mathbb{R}^{d} | (\bar{x} - \mathbf{x}_{\text{new}})^{T} S^{-1} (\bar{x} - \mathbf{x}_{\text{new}}) \leq \Phi_{T_{2}(d,n-1)}(1 - \alpha)(1 + \frac{1}{n}) \] \\
\text{(2.24)}

The way to see that this region is an ellipsoid is by rewriting both forms as:

\[ \mathbf{b}^{T} S^{-1} \mathbf{b} \leq C \]

and remembering that \( S \) is a covariance matrix and is thus non-negative definite. As such it is a standard ellipsoid quadratic form since for the inverse of a matrix,

\[ A \mathbf{v} = \lambda \mathbf{v}, \]

\[ A^{-1} A \mathbf{v} = A^{-1} \lambda \mathbf{v}, \]

\[ \lambda^{-1} \mathbf{v} = A^{-1} \mathbf{v}, \]

i.e. \( S^{-1} \) has the same eigenvectors as \( S \) with the inverted eigenvalues. A change of basis to the eigenbasis thus gives the equation:

\[ \frac{\tilde{b}_{1}^{2}}{\lambda_{1}} + \frac{\tilde{b}_{2}^{2}}{\lambda_{2}} + \frac{\tilde{b}_{3}^{2}}{\lambda_{3}} \leq C \] \\
\text{(2.25)}

where \( \lambda \) is the variance in the direction of each eigenvector. The formulation for the prediction intervals can also be extended in a regression framework in the same way as for the univariate case. Another common test is to test if two sets have the same mean, for this purpose a Two-sample Hotelling test can be performed which has the form

\[ H_{0} : \mu_{1} = \mu_{2}, \]

\[ H_{1} : \mu_{1} \neq \mu_{2}. \] \\
\text{(2.26)}

where \( S_{1,2} \) is the combined variance \( S_{1,2} = ((n_{1}-1)S_{1} + (n_{2}-1)S_{2})/(n_{1} + n_{2}) \).

Here the \( T^{2} \) statistic will have a \( T^{2}(d, n_{1} + n_{2} - 2) \) distribution under the null hypothesis.

**Roy’s Union–Intersection principle**

The Hotelling tests is a simple generalisation of the univariate test to the multivariate case, such a simple generalisation does not always exist. One can however use a heuristic method to generalize univariate tests, known as Roy’s union–intersection principle. The union intersection principle is based on combining multiple (commonly univariate) test into a single test. Formally given, a set of null and alternative hypothesis \( H_{i}, K_{i} \) \( i \in \Gamma \) with \( \Gamma \) as a index set, the combined Hypothesis become:

\[ H_{0} : \bigcup_{i \in \Gamma} H_{i} \]

\[ K_{0} : \bigcap_{i \in \Gamma} K_{i} \] \\
\text{(2.27)}

18
or in words, the combined null is accepted if all of the univariate nulls are accepted and the combined null is rejected if at least one null is rejected. The confidence in each test is then tuned to correspond to the desired global level, something which is not always possible to do analytically. A common way to select the families of Hypotheses is to project the multivariate data on multiple directions \( a_i \) in the higher-dimensional space and then test the projected data with a univariate test. For generalising the t-test the tests would then become:

\[
H_i : \quad a_i \cdot \mu = a_i \cdot \mu_0 \\
K_i : \quad a_i \cdot \mu \neq a_i \cdot \mu_0 \\
t = \frac{a_i \cdot (\bar{x} - \mu_0)}{\sqrt{a_i^T S a_i} / \sqrt{n}}
\]

(2.28)

where one would perform tests in all possible directions, not just in the direction of the coordinates axis. Critically in the case of the t-test it can be shown [41] that constructing a multivariate generalisation using Roy’s Union-intersection principle in all possible directions \( a_i \) is equivalent to performing a Hotelling \( T^2 \) test, the individual t-test are however on a stricter level than \( \alpha \) as illustrated by figure 2.5. The fact that this equivalence exists is something that will arise again in section 2.3.2. It is also possible to see that the set of univariate tests is infinite, but the correlation between the test allows each test to still be on a finite level.

Figure 2.5: Example of component-wise t-tests prediction intervals at three different levels, as well as a Hotelling T2 prediction ellipse. The levels are for naive levels, Bonferroni correction and the Hotelling equivalent level.
2.3.2 Random Field theory

Random Field theory (RFT) has its origin in functional MRI where the goal is to detect which area of the brain is activated when certain tasks are performed. In this case one wants to predict which areas in a scalar field are significantly bigger than a resting level. The main goal of RFT is to correct for the multiple testing problem while keeping the inherent spatial dependencies in mind. To achieve this the same regression model is fitted at each pixel, a $t$-test is then performed for the predictors of each fit. This would, given that the null hypothesis holds, create a spatially correlated $t$-field. The question is then at which level should this field be thresholded in order to find significant activations, i.e. for a given region $S$ which level $M$ is it expected will give $P(\max_{s \in S} t(s) \geq M) = \alpha$?

In the case of image registration, random field theory relates to detecting if a bias field is statistically significant and altogether allows statistically valid statements about the form of the whole deformation field. This allows us to understand the systematic effects of a registration algorithm in correct manner, quantitative comparisons between different algorithms on a global level also become possible.

The deformation field will however give rise to a Hotelling $T^2$-field instead of a $t$-field like in the fMRI case, fortunately this does not present to much difficulty since it it was showed by Taylor and Worsley, using Roy’s Union–intersection principle, that this case could be reduced to the already studied $t$-field [42]. The following presentation summarises the relevant practical details from [42] that will be used to establish a significant threshold for the deformation field as well as the major strokes of the underlying theory.

**Euler characteristic approximation**

We begin with approximating the probability that the maximum of a stochastic field exceeds the threshold $M$ by using the expected Euler characteristic (EC) of the excursion set, that is the set of the field exceeding the threshold $M$. The Euler characteristic is a topological property of a set and can be summarized as $\varphi = \#$ blobs $- \#$ handles or tunnels $+ \#$ interior hollows, this approximation is quite unintuitive, however this approximation aims to be valid for large threshold where the excursion set essentially reduces to a single blob above the threshold. The validity and accuracy of this approximation is described in [43] and it is particularly useful since it has a valid expression for all threshold levels, that being

$$\alpha = P(\max_{s \in S} t(s) \geq M) \approx \mathbb{E}(\varphi\{s \in S : t(s) > M\}) = \sum_{i=0}^{N} L_i(S) \rho_i(M), \tag{2.29}$$
where $\mathcal{L}$ is the Lipschitz-killing curvature (LKC) and $\rho$ is the Euler characteristic density of the field above $M$. The Lipschitz-killing curvature relates to the smoothness of the field and the size of the search region and can be estimated in various ways, described below. The need to use a complex construct such as the LKC is to deal with non-isotropic fields, if the field is isotropic the LKC can be replaced with the intrinsic volumes of the search area instead. The Euler characteristic density on the other hand depends on the distribution of the underlying field. For the Hotelling field with $\nu$ degrees of freedom that through the Union-intersection principle can be written as:

$$T(s) = \max_{||u||=1} \tilde{T}(s,u),$$

where $\tilde{T}(s,u) = \frac{(Z(s)^T u)^2}{u^T W(s) u / \nu}$,

with $Z, W$ and $u$ being a standard $d$ Gaussian field, a Wishart random field with $\nu$ degrees of freedom and $u \in \mathbb{R}^d$ respectively. This being that the Hotelling field can be written as the maximum of a square of a $t$-field restricted to the $d$-dimensional unit sphere. The EC densities can then be related to the $t$-field as described in [42] and given as:

$$\rho^H_t(M) = \sum_{j=0}^d \mu_j(U_d) \rho_{t+j}^I(\sqrt{M}),$$

(2.30)

where $\mu_j$ is the intrinsic volumes of sphere $U_d$, i.e. constants and $\rho_{t+j}^I$ is the EC density of $t$-field given in [44]. These quantities can be readily

Figure 2.6: Example of a thresholded spatially correlated Gaussian random field. The Euler characteristic of excursion set, the set above the threshold, is $4$. 
calculated without any observed data, and is for example implemented in the open source NiPy python package [45].

The result in (2.30) follows from relating the Hotelling $T^2$ to a $t$-field and utilizing invariances between the two fields and using symmetries over the sphere.

**Estimating Lipschitz-killing curvature**

To estimate the LKC the simplest method is the use of Resels (resolution elements) first proposed in [46] and given a introduction to in [47], which is based on applying Gaussian kernels to the field in order to guarantee a minimal level of smoothness, the full width at half maximum (FWHM) for the kernel is then used to substitute the LKC as

$$\hat{L}_N(S) = \frac{\text{vol}(S)}{(4 \log 2)^{3/2} (\text{FWHM}_x \text{FWHM}_y \text{FWHM}_z)},$$  \hspace{1cm} (2.31)

where FWHM$_\bullet$ is the FWHM of the applied Gaussian kernel in the direction $\bullet$. An approach based on estimating the LKC is also to estimate the variogram of the field and through this get an estimate for the largest dimensional LKC, this is usually the term that contributes the most to the LKC.

$$\mathcal{L}_N(S) = \int_S \det(\text{Var}(\hat{Z}_1(s)))^{1/2} ds.$$  \hspace{1cm} (2.32)

Although the lower dimensional LKC can be estimate as well by filling the search area with a mesh this is usually not done, instead the lower dimensional LKC are assumed to be the values corresponding to a sphere with the same volume as estimated through equation (2.32). The estimate of the highest LKC is performed by normalizing the columns of the residuals in the regression model. The normalized residuals are then used to create derivative images, which are then combined to estimate $\text{Var}(\hat{Z})$. Formally for each column $j$ of the Residuals:

$$Q_j(s) = R_j(s)/||R_j(s)||.$$  \hspace{1cm} (2.33)

$$D_j(s) = (Q_j(s + e_1) - Q_j(s), ..., Q_j(s + e_n) - Q_j(s),$$  \hspace{1cm} (2.34)

$$\text{Var}(\hat{Z}(s))\Delta = D_j(s)^T D_j(s),$$  \hspace{1cm} (2.35)

$$\hat{L}_n(S) = \sum_{j=1}^d \sum_{s \in S} \det(D_j(s)^T D_j(s))^{1/2} / d.$$  \hspace{1cm} (2.36)

This estimator has been shown to be unbiased as the grid-spacing $\Delta$ goes to zero.
Additional details about RFT

It should be noted that the RFT threshold can give rise to a higher threshold than that of a Bonferoni correction. This occurs since the RFT corrects for a continuous field and the Bonferoni correction works on points, basically RFT corrects for all points in the search area while Bonferoni only corrects the sampled points, and it is unlikely that the sampled points where the field has been observed is the location of the true maximum of the field. That the Bonferoni correction is smaller than the RFT however implies that the spatial correlation of the field is small.
Chapter 3

Method

This chapter describes the process of generating spatially normalized deformation fields, how to detect and compensate for inherent model biases and the implementation details for performing the calculations.

3.1 Deformation field generation

The generation of the deformation field aims to estimate the misslocation between the CT and MR from the rigid registration, the goal of the inter-subject registration is to supply multiple realisations of this estimate in order to facilitate a comparison of the fields. Essentially two things are of interest in relation to the deformation fields, firstly an estimate of any systematic deformations in the skull following the deformable registration, and secondly an estimate of how certain these deformations are in order to get a measure of how accurate the registration is. It is also of interest to study what parts of the algorithm affects the data.

3.1.1 CT-MR registration

To get comparable deformations it is required to separate the local and global registration parts of the registration. In order to do this a rigid registration is performed between the MR and CT image, to generate a transformation model $T_{i,R}$. This model is then used as an initial transformation, as described in section 2.1.4, to a deformable registration between the MR and CT image to generate a deformation model $T_{i,D}$. The deformable registration in this case is viewed as an estimator of the error in the rigid process, in a regression framework it can be formulate as $d_i = \mu + \epsilon_i$ where the error distribution incorporates both the error from the rigid registration as well as the error in the estimation of the deformation field. This process is visualized in figure 3.1.
These transforms are generated for each pair of images in the data set. Independent of this registration a group-wise normalisation step is performed. This normalisation step is a group-wise registration as described in section 2.2, specifically a pair-wise registration to a frame striped template image is used. There is an additional consideration in this group-wise registration in the form of what kind of deformations we are investigating. The normalising step have two plausible directions, one aiming to normalise the locations with respect to the machine and the other with respect to the anatomy of the skull. These are related to the classes of deformations that arise, if we want to characterise the machine specific deformation we have to normalise the image to the machine and if we want to study the biological effect we have to normalize the anatomy. In our image set a machine specific normalisation can be performed by relating the stereotactic frames. Regardless of which normalising process is selected a transformation $T_{i,N}$ is generated from the CT-images, it is also plausible to base the normalisation on the MR-images but since the CT-images generally have better spatial accuracy they are used. However one reason one might want to use the MR-images for the normalisation arise in the case where the CT-images target area are reduced in size in order to decrease the dose to the patient. In this case the registration process becomes less stable for the CT-images since there are less corresponding structures to align. This is not a problem in our case where the CT images cover the whole skull. Thus an anatomical normalization w.r.t. the CT images are performed.

With these three deformations, $T_{i,R}$, $T_{i,D}$ and $T_{i,N}$ a deformation field can be generated as

$$DF_i : \Omega_N \rightarrow \mathbb{R}^3,$$

$$DF_i(x) = (T_{i,D} - \text{id}) \circ T_{i,R} \circ T_{i,N}(x). \quad (3.1)$$

This field is generated for each of the $n$ patients. In this formulation a comparison of the difference between the deformable registration and the
rigid registration is performed. It is however easy to see that this framework allows comparisons of different deformable algorithms by replacing the identity transformation in equation (3.1), it is also possible to evaluate the convergence of the rigid registration by letting the deformable registration be a rigid registration. The generation process is visualised in figure 3.2, and can be thought of as a function, parametrized by the configurations of the registration algorithms, taking two-images as input and giving the deformation field as output.

\[
\begin{align*}
\text{Rigid} & : \Omega_{CT} \to \Omega_{MR} \\
T_R : \Omega_{CT} & \to \Omega_{MR} \\
M_R \circ T_R & \to \Omega_{MR}
\end{align*}
\]

\[
\begin{align*}
\text{Deformable} & : \Omega_{MR} \to \Omega_{MR} \\
T_D & : \Omega_{MR} \to \Omega_{MR} \\
M_R \circ T_D \circ T_R & \to \Omega_{MR}
\end{align*}
\]

\[
\begin{align*}
\text{Normalization} & : \Omega_T \to \Omega_i \\
T_N & : \Omega_T \to \Omega_i \\
M_R \circ T_N & \to \Omega_i
\end{align*}
\]

\[
DF_i = \{(T_D - \text{id}) \circ T_R \circ T_N(x) | x \in \Omega_N\}
\]

Figure 3.2: The process for creating the deformation fields.

The deformable registration we perform are based on [48], that evaluated a B-splines deformable registrations effect on landmark locations and showed that the norm of the displacement vectors decrease. No directional analysis was however performed, so the extent of any bias in the method is unknown. The parameters of the deformable registration were tuned by leaving one out cross-validation in [48] to minimize the MSE of landmark localisation and should thus be a good estimator of the error deformation.
3.1.2 Baseline deformation

In order to study the effects of the MR-image some baseline properties of the process are studied. This allows for an accurate null level for the deformations. The baseline deformation is established utilising the deformations and uncertainties introduced by the approximations in the registration process itself. This process uncertainty has as a consequence that an auto-registration, a registration from an image $I_F$ onto itself, will not be a perfect identity transform. Thus performing the process of calculating (3.1) with the MR-image $MR_i$ substituted with another copy of $CT_i$ will create a process bias field and the inherent variance in the process itself, this is essentially the minimum bound on the uncertainty in the process and any MR deformations should be significantly different from this. This will also quantify if any biases in the algorithm exist.

3.1.3 Comparisons

After the construction of the deformation fields a summary of the results are generated, primarily in the form of a mean field and a covariance "9D-image" since these can be used to create most other results, such as an effect size image or the $T^2$ field. Primarily two mean images and variance images are created, one set for the MR-CT field and one for the baseline field. A two sample test is then performed to estimate the effect size and $T^2$ fields to assess the bias field. This field is thresholded using RFT at a 95% significance level where we estimate the LKC of the fields by estimating the largest dimensional LKC and estimating the lower dimensions using a sphere assumption. The threshold levels are then calculated using the NiPy[45] package Hotellings function which calculates the $T^2$ p-value for a given LKC and threshold.

The mean and variance image for the MR-CT set are also used to create 95% univariate prediction ellipses at target points of tumours in the brain for image sets not included in the data used in the mean and covariance estimation. These ellipses can then be used to estimate if the deformations of the image are in line with the rest of the set and visualize the uncertainty that the process implies from the large data set.

3.2 Implementation details

When dealing with 3D data at this scale, the computation time and space requirements become significant hurdles to overcome. Thus sequential implementations that do not load the entire set in memory are needed, calculations that can be done on a GPU are also of interest. Another aspect that can be quite time consuming and hard to improve on is the IO-speed of
the disk where the deformation fields are stored, the best that can be done
in this regard is to fully utilize the disk speed. Below the approach used
to deal with some of these problems are described, mostly in the form of
utilizing existing frameworks. Some aspects of how the visualization of the
deformation and covariance fields can be done are also described.

3.2.1 Code implementation

The code implementations where primarily done in C++, using the Insight
Segmentation and Registration Toolkit (ITK) [17] version 4.8 for general
image handling and statistical calculations. The registration process makes
use of the Elastix (4.8) software [1] for all registration steps. Elastix facil-
ities utilization of the GPU. In order to handle the large data sets Python
is used for preprocessing and scripting the calls to both ITK and elastix as
well as extracting relevant data from the images. Finally some data analysis
is done in R.

The implementations of the statistical calculations are done in two ways,
firstly a "NaryFunctorImageFilter" framework is used. This filter will take
N images of the same size and apply a supplied functor, a parametrized
function object, to a vector of pixels at the corresponding location, thus
new algorithms such as a new regression model will only require a new func-
tor to be implemented. The filter is parallelized and handles the memory
constrains by only loading small sections of the image into memory at one
time in a streaming fashion. This filter is used to calculate the mean de-
formation and covariance fields. Secondly filters for updating the mean and
variance images in a recursive doubling manner are implemented as well as
filters for calculating the $T^2$-field from the mean and variance field. The
filter facilitating a recursive doubling is primarily useful from a space saving
viewpoint, where the algorithm can be run for a subset of images, summa-
rized in a mean and variance image and then the intermediate fields can be
removed to create space. This can performed multiple times and finally all
of these intermediate images can be used to generate the final images.

The registrations performed with elastix is one rigid, one affine and one
B-spline, the parameter files can be found in the appendix.

3.2.2 Visualization

3D Image visualisation was performed using the vv software [49]. A big
question is how to visualize the covariance in a scalar way. A simple way
to visualize this is to plot the root of the trace of the variance, one might
think this overestimates the variance due to correlation effects but since the
covariance is diagonalizable we have

$$\text{tr}(S) = \text{tr}(V^{-1}DV) = \text{tr}(DVV^{-1}) = \text{tr}(D),$$  \quad (3.2)

where we have used that the trace is invariant under cyclic permutations. We also visualize the effect size, which represents the size of the bias with respect to the underlying variance, in the multivariate case:

$$(\mu_1 - \mu_2)^T S^{-1} (\mu_1 - \mu_2).$$  \quad (3.3)
Chapter 4

Results

4.1 Data-selection

To examine the effects of the deformable registration we use \( n = 494 \) image pairs of CT and MR images provided by Elekta. This data has been anonymized w.r.t. the names of the patients in order to preserve the integrity of the patients. The MR modality is the T1 weighted post contrast agent sequence f3d1 of the SIEMENS MAGNETOM Avanto with a 1 mm voxel spacing except for patients with AVM that has a \([0.66, 0.66, 1]\) mm spacing. The MR images are corrected for intensity non-uniformity [50] using the N4 algorithm in ITK[51, 52]. The CT scans are of two forms, either with a \([0.5, 0.5, 1]\) mm spacing or a \([0.5, 0.5, 2]\) mm spacing. In order to avoid subject bias, patients with multiple image series are discarded. The final images are all rescaled to 1 mm spacing in all directions, this downscaling is done for two reasons firstly even the downscaled deformation fields are quite large (118 MB) and this downscaling reduces the field size by a factor 4 and thus speeding up the read times from the disk and secondly the B-spline spacing is quite large (60 mm) so a sub millimetre deformation field is not necessary to capture the relevant effects.

4.2 Auto deformation bias fields

The auto deformation process is applied to the CT images in order to get a minimum uncertainty of the deformation field. The LKC of this field is low, 5.18. As thus essentially all deformations are statistically significant due to the high number of observations. This does not mean that they are necessarily practically significant. The effect size is plotted in 4.1 and it is apparent that the effects are centred to the brain region. The largest norm of the mean deformations is 0.18 mm compared to the smallest voxel spacing of 0.5 mm, and as such not something that impacts a single image. The fact that the effect size of the deformations is quite high, 1.81 as a
max, does however indicate that the bias is systematic in a way that is not dominated by the underlying noise.

4.3 MR-CT Bias deformation field

For the full MR-CT field we first begin by verifying that the rigid deformation has converged. This is done by looking at the mean deformation field generated by running a new rigid registration, after discarding a single outlier from the set of deformations we get norm deformations on the $10^{-5}$ mm scale with standard deviations of at most 0.03 mm, thus a clear convergent result. Secondly the full deformable MR-CT field is compared, using a Hotelling $T^2$ two-sample test, to the auto deformation field from the result in the previous section above. The LKC of the deformation field is estimated from a subset of 100 pairs randomly selected from the 494 pairs and is estimated to be 50.67 and does not vary significantly with respect to the number of samples. This volume corresponds to a radius of $r = 2.30$, which gives us the three other LKCs, thus the sum becomes:
This implies a threshold level of the $T^2$ field of 20.8, this can also be seen as a Bonferroni correction of $\sim$500 tests. Which can be related to the about $\sim$1000 parameters estimated in the deformable registration. This corresponds to an effect size of 0.29, which is an threshold that is exceeded almost everywhere. These biases are plotted in 4.2. This gives an indica-

\begin{table}
\centering
\begin{tabular}{|l|c|c|c|}
\hline
Part & $L_0$ & $L_1$ & $L_2$ & $L_3$ \\
\hline
Auto Field & 1 & 4.29 & 7.23 & 5.18 \\
MR-CT Field & 1 & 9.18 & 33.10 & 50.67 \\
Total LKC & 2 & 13.47 & 40.33 & 55.85 \\
\hline
\end{tabular}
\end{table}

Figure 4.2: The effect size of the final deformation field overlayed a sagittal slice of the CT-mean image. The deformations field magnitude is scaled by 5.

We are also interested in how the variance affects the displacements, i.e. the actual displacements can be large while the bias is zero. The mean square error (MSE), which in the multivariate case can be expressed as

$$E[||\hat{D} - D||^2 = \text{trace}(\text{var}(\hat{D})) + ||\text{bias}(\hat{D})||^2, \quad (4.1)$$

where $||\cdot||$ represents the euclidean norm of a vector. If the mean deformation is taken as the bias of the rigid registration then it is possible to create an mean square error image representing both aspects of the registration. This is not a true MSE image since the bias is unknown, instead it is an overestimate of the error implied by the deformation algorithm, since it
includes the variance in the deformable estimation as well. The square root of such an image can be seen in 4.3, the root is taken in order to preserve the unit. The root of the variance trace image can also be seen in figure 4.4.

Figure 4.3: The mean square error field overlayed a sagittal slice of the mean MR image. The largest root mean square error inside the skull is 1.4 mm.

4.4 Prediction visualization

The quantification in the previous section is good at displaying how the certainty in the areas compare, it does however not given an indication in which directions the uncertainty is in. Thus a more spatial description displaying how individual points are affected is desired. This is solved using 95% prediction ellipses in the normalized images to display an area that the expected actual location will contain. The ellipse could also be presented in the original subject space by inverting the normalizing transform. The transform is created along the lines described in (2.24), and relates primarily to visualizing the variance in different points. A clear example can be seen in figure 4.5.

Some of these ellipses become rather small in relation to the voxel size such that the variance directions becomes hard to distinguish thus it might be better to display the eigenvectors instead. Using the magnitude of the largest eigenvector can also be used to describe the variance since it corresponds to the largest magnitude of a vector contained in the prediction.
ellipse. These ellipse can also be used to detect suspicious deformations.

4.5 Model validation

The detection of suspicious deformations is related to the detections of outliers that relate to verifying the underlying assumptions that the residuals are multivariate normally distributed. This assumption is not straightforward to verify since there are a great deal of tests to verify. Some representative points are thus examined, such as the the center of the brain, a bone edge and a point outside the brain. A major question in this verification is the question of outliers, if and how they should be identified and excluded, for example the question if image pairs where the initial rigid registration failed should be included is such a central question.

If all the deformations at a single voxel are extracted it is possible to look for outliers with a simple QQ-plot between the chi-square distribution and the empirical distribution of the Mahalanobis distance for the observations. Such a plot for a central point is shown in figure 4.7 and clearly shows five outliers, these correspond to an initial rigid transform that is rotationally misaligned, an MR image with severe metal distortions, an image of patient having undergone surgery removing a great deal of the skull and a two cropped CT-images. Even after these outliers are removed it is possible to see in figure 4.8 that the residuals are heavy tailed. Potential causes of this
Figure 4.5: Two examples of confidence regions being displayed. The second case have been scaled to make the illustration clearer.

Figure 4.6: Example of two outliers.

and the effects this will cause is discussed in the following chapter. It can however be noted that the removal of these outliers have a negligible effect on the estimated covariance matrix.
Figure 4.7: $\chi^2$ plot of the deformations at central voxel, clearly showing at least five outliers.

Figure 4.8: Three univariate normality plots of the central point distribution as well as a $\chi^2$ QQ plot with the clear outliers removed. It can also be seen that the non-normality is primarily in one direction.
Chapter 5

Discussion

In this chapter some aspects of the validity of the proposed method and the underlying assumptions will be discussed, as well as how the naive estimation approaches might influence the outcome. These considerations also give rise to some suggestions for further work.

5.1 Inter subject correspondence

A major assumption that the proposed method depends on is the ability of the group-wise registration algorithm, in our case the pairwise affine registration, to create an accurate correspondence between parts of different subjects. That such a correspondence can be found is not clear since unlike the intra subject registration no true underlying transform is obvious. Essentially this is a question of how well the results from other patients can be generalized to new patients. Currently all patients are treated as a single group, this is probably optimistic and a clear example of this is the patients with AVM, whose CT-images suffer a greater deal of artefacts than other patients and also have another voxel spacing. The actual head size also comes up here since we will be comparing scaled versions of the heads, thus a larger head might be subjected to larger deformations due to, for example, a thicker skull. Currently the deformation fields are only collocated so the scalings of the original deformations are preserved but moved to the ”correct” locations. This scaling effect of the head size is plausibly something that could be corrected in the regression formulation by including the head size as a predictor. The largest scaling are however at most 15% and should thus have only minor effects on the results.

Concretely the question of assessing the quality and validity of this inter subject registration arises and secondly the question of how this affects the validity of the deformation results. I perform no quantitative assessment of the group-wise registration, this could however be performed through a seg-
mentation overlap analysis as used to compare different groupwise methods as done in [40]. A qualitative assessment of the group-wise registration is however harder to construct. Fortunately the underlying deformation fields are in our case rather smooth so a voxel perfect registration should not be necessarily required to get representative results.

A simple way to motivate that the group-wise registration does produce reasonable results can be achieved by looking at the mean and standard deviation images of the normalized heads. These can be seen in figure 5.1 and qualitatively motivates that the algorithm succeeds at a reasonable degree for the whole set. In the standard deviation image a single miss-registration in the upper part of the brain can be seen.

If the registration fails for a subject this is essentially an introduction of an outlier to the data set, and as such the validity of the final result is related to the robustness of the algorithm w.r.t. outliers. In this sense a soft detection can be achieved in the form that a failed pairwise registration could be detected from abnormal deformations, this however depends on that the pairwise registration usually works, i.e. a failed registration actually has to be an outlier. The intra subject registration is primarily a problem in the generation of the mean and covariance fields since we are dealing with such a large data set that manual verification of the quality is infeasible.

5.2 The effect of outliers & and Robust estimates

In section 4.5 it was established that some outliers are generated in the process, either through abnormal anatomies or through failed rigid registrations and the errors are longer tailed than one might desire for the underlying normality assumption. The longer tails in particular will affect the validity of the generated prediction ellipses. Ideally one would want to like to use robust methods in the mean and covariance estimation for example using a minimum covariance determinant (MCD) scheme to estimate the covariance matrix, this is however computationally heavy and just switching to

Figure 5.1: The collocated MR images.
MCD estimation at this sample size would increase the estimation time by a factor 150, additionally we would lose the possibility of simply updating the covariance estimate when new data is introduced. It is thus important to understand what this might imply for the clinical decision making process.

There are two aspects to this, the visualization of the prediction region and the ability to detect outliers. These two aspects are in some sense counterparts to each other. Firstly the effect of outliers will generally have a larger effect on the covariance estimation than on the mean estimates. This implies that the covariance will be overestimated, which is consistent with the robust covariance estimates observed for the data set. This means that the prediction region will be larger than it should be without the outlier. Essentially making our ability to detect outliers lower by being more uncertain about the location. The effect of the longer tail is harder to predict since it departs from the normality assumption that the prediction intervals depends on, ideally the effect would be evaluated using cross-validation or similar techniques.

### 5.3 Validity of the error estimation

Another part crucial to the usefulness of any statements made by the process described in chapter 3 is the choice of the deformable registration model. In this thesis a B-spline model was used to estimate errors, it is however clear that the framework allows for other models to be used, one could select the identity transform as a model which would produce results of no interest. The model used subjected smaller MSE than rigid registration\[48\] w.r.t. ground truth measurements and thus indicates that the estimators are correlated to the actual errors. However the bias of the algorithm is unknown, but remains an interesting property since it is probably non-zero since the transform was selected in a MSE minimizing fashion. It should thus be stated that the suggested method is not a means for selecting the most accurate deformable algorithm, this requires ground truth measurement but instead the method will describe the effects and variations in a selected method.

### 5.4 Further work

In regards to further developments there are some specific areas to consider.

**Additional predictors:** As described above the current framework assumes that all subjects belong to the same group. There are some obvious factors that could have effects on the problem, such as gender, age, size, voxel spacing and affliction to name a few. These effects could reasonable
be included into the regression models as logistic or continuous variables to possibly allow a more accurate prediction of the outcome.

**Groupwise methods:** The current groupwise registration is a simple pairwise affine registration, as described in section 2.2 this will bias the results. Although the deformation fields are very smooth it could still be of interest to use a more deformable groupwise model in order to generate a more accurate correspondence, especially if the deformation model is replaced.

**Sparse correspondence:** Currently the subject normalization tries to find a correspondence on a global level between the subjects brains. Such a correspondence is not guaranteed to exist. An alternative could thus be to pursue a sparse correspondence relying on automatically identifying certain landmarks for each subject, such as lenses in the eyes, tips of the ventricles and other easily detected points. The displacement could then be performed at each of these regions and interpolated out to other areas of the brain. The detection of the landmarks would then avoid any subject bias and make an evaluation of the inter subject registration simple, it would simply be the ability of detecting each landmark compared to a ground truth measurement.

**Modality comparisons:** The bias field generation gives an idea of the systematic deformations, an area to pursue could thus be to qualitatively compare different registration methods as well as comparing different imagining modalities to each other in order to be able to relate differing modalities better.

**Bias estimation:** The process could be repeated on an image set with manually marked landmarks so as to generate a bias vector of the deformable registration process which would indicate the true bias. The deformable registration algorithm could then be tuned in a true mean square error minimization fashion while still giving information regarding the error distribution. It would also be possible to study how applying the mean field deformation to a new image would impact the error to reduce computation times. This would also allow selection of a suitable predictor model.
Chapter 6

Conclusions

The aim of this thesis was to provide a spatial uncertainty and error estimate for the rigid registration process in medical image registration in a situation where no ground truth is available. An spatial error estimate was created using a deformable registration and a normalizing scheme to get multiple realisations of the error distribution at each voxel, allowing for a statistical treatment of the estimated errors. This established the quantitative effects of the deformable registration process. The process produced a statistically significant bias field using random field theory correction indicating that the errors are not only the result of noise. Spatial prediction regions were successfully created in order to visualize the uncertainty in the error estimation.

Nevertheless the method is very dependent on the deformable registration algorithm correctly estimating the spatial errors and will only approximate the true errors. This aspect needs further examination as it was noted that the non-rigid registration has a bias in the trivial registration case. A ground truth bias estimation and MSE evaluation of the algorithm is thus of interest as it could assists in creating a variance-bias trade off for the error estimating algorithm. The method for visualization also depends on the new subject’s similarity to the reference group, this similarity dependence could probably be reduced by using additional predictors.

As a summary, a method for using a-priori knowledge to visualize the spatial uncertainty has been presented and shows promise, however further verification of the the choice of estimator is needed and more developed distinctions between subjects should be investigated.
Bibliography


Appendices
Appendix A

Registration parameters

A.1 Rigid

// ImageTypes
(FixedInternalImagePixelType "float")
(FixedImageDimension 3)
(MovingInternalImagePixelType "float")
(MovingImageDimension 3)

// Components
(Registration "MultiResolutionRegistration")
(FixedImagePyramid "FixedSmoothingImagePyramid")
(MovingImagePyramid "MovingSmoothingImagePyramid")
(Interpolator "BSplineInterpolator")
(Metric "AdvancedMattesMutualInformation")
(Optimizer "StandardGradientDescent")
(ResampleInterpolator "FinalBSplineInterpolator")
(Resampler "DefaultResampler")
(Transform "EulerTransform")

// ********** Pyramid

// Total number of resolutions
(NumberOfResolutions 6)

// ********** Transform

//(Scaler 600000.0)
//(CenterOfRotation 128 128 ??) center by default
(AutomaticTransformInitialization "true")
(AutomaticTransformInitializationMethod "Origins")
(HowToCombineTransforms "Compose")

// ********** Optimizer

// Maximum number of iterations in each resolution level:
(MaximumNumberOfIterations 1200 1200 1200 2000 2000 2000)

//SP: Param_a in each resolution level. a_k = a/(A+k+1)^alpha
(SP_a 3000.0 2000.0 1000.0 1000.0 500.0 500.0)

//SP: Param_alpha in each resolution level. a_k = a/(A+k+1)^alpha
(SP_alpha 0.602)

//SP: Param_A in each resolution level. a_k = a/(A+k+1)^alpha
(SP_A 50.0)

// ********** Metric

// Number of grey level bins in each resolution level:
(NumberOfHistogramBins 32)
(FixedLimitRangeRatio 0.0)
(MovingLimitRangeRatio 0.0)
(FixedKernelBSplineOrder 1)
(MovingKernelBSplineOrder 3)

48
A.2 Affine

// Description: affine, MI, ASGD

// ImageTypes
(FixedInternalImagePixelType "float")
(FixedImageDimension 3)
(MovingInternalImagePixelType "float")
(MovingImageDimension 3)

// Components
(Registration "MultiResolutionRegistration")
(FixedImagePyramid "FixedSmoothingImagePyramid")
(MovingImagePyramid "MovingSmoothingImagePyramid")
(Interpolator "BSplineInterpolator")
(Metric "AdvancedMattesMutualInformation")
(Optimizer "AdaptiveStochasticGradientDescent")
(ResultImagePixelType "short")
(ResultImageFormat "mha")
(WriteTransformParametersEachIteration "false")
(WriteResultImage "true")
(Compression "false")
(ShowExactMetricValue "false")
(NumberOfResolutions 4)
(HowToCombineTransforms "Compose")
(AutomaticTransformInitialization "true")
(AutomaticTransformInitializationMethod "CenterOfGravity")

// Maximum number of iterations in each resolution level:
(MaximumNumberOfIterations 2000)

// Number of grey level bins in each resolution level:
(NumberOfHistogramBins 32)
(FixedLimitRangeRatio 0.0)
(MovingLimitRangeRatio 0.0)
(FixedKernelBSplineOrder 3)
(MovingKernelBSplineOrder 3)

// Number of spatial samples used to compute the mutual information in each resolution level:
(ImageSampler "Random")
NumberOfSpatialSamples 2000
NewSamplesEveryIteration "true"
CheckNumberOfSamples "true"
MaximumNumberOfSamplingAttempts 10

// Order of B-Spline interpolation used in each resolution level:
BSplineInterpolationOrder 1

// Order of B-Spline interpolation used for applying the final deformation:
FinalBSplineInterpolationOrder 3

// Default pixel value for pixels that come from outside the picture:
DefaultPixelValue 0

(MaximumStepLength 4.0)

(ResultImagePixelType "short")
(ResultImageFormat "mha")

A.3 B-spline

// ********** Image Types
(FixedInternalImagePixelType "float")
(FixedImageDimension 3)
(MovingInternalImagePixelType "float")
(MovingImageDimension 3)
(UseDirectionCosines "true")

// ********** Components
(Registration "MultiResolutionRegistration")
(FixedImagePyramid "FixedSmoothingImagePyramid")
(MovingImagePyramid "MovingSmoothingImagePyramid")
(Interpolator "BSplineInterpolator")
(Metric "AdvancedMattesMutualInformation")
(Optimizer "AdaptiveStochasticGradientDescent")
(ResampleInterpolator "FinalBSplineInterpolator")
(Resampler "OpenCLResampler")
(OpenCLResamplerUsesOpenCL "true")
(Transform "BSplineTransform")

// ********** Pyramid
// Total number of resolutions
NumberOfResolutions 2
(ImagePyramidSchedule 2 2 1 1 1 1)

// ********** Transform
(FinalGridSpacingInPhysicalUnits 60)
(AutomaticScalesEstimation "true")
(AutomaticTransformInitialization "false")
(HowToCombineTransforms "Compose")

// ********** Optimizer
// Maximum number of iterations in each resolution level:
(MaximumNumberOfIterations 1000)

// ********** Metric
// Number of grey level bins in each resolution level:
(NumberOfHistogramBins 32)
(ErodeMask "false")

// ********** Several
(WriteTransformParametersEachIteration "false")
(WriteTransformParametersEachResolution "true")
(WriteResultImageAfterEachResolution "false")
(WriteResultImage "false")
(ShowExactMetricValue "false")

// ********** ImageSampler
// Number of spatial samples used to compute the mutual information in each resolution level:
(ImageSampler "RandomCoordinate")
(NumberOfSpatialSamples 4000)
(NewSamplesEveryIteration "true")
(CheckNumberOfSamples "true")
// ********** Interpolator and Resampler
// Order of B-Spline interpolation used in each resolution level:
(BSplineInterpolationOrder 1)
// Order of B-Spline interpolation used for applying the final deformation:
(FinalBSplineInterpolationOrder 3)
// Default pixel value for pixels that come from outside the picture:
(DefaultPixelValue 0)
(ResultImagePixelType "short")
(ResultImageFormat "mha")
(WriteDiffusionFiles "true")