Doctoral Thesis

博士論文

Towards a Paradigm Shift from Error to Uncertainty in Neuronavigation

(ニューロナビゲーションにおける誤差から不確実性へのパラダイムシフトに向けて)

羅　捷

Contents

Abstract

Surgical resection is the initial treatment for nearly all brain tumors. The achieved extent-of-resection is strongly correlated with prognosis and is the single greatest modifiable determinant of survival. Since brain tumors are intimately involved in surrounding functioning brain tissue, aggressive resection must be balanced against the risk of causing new neurological deficits.

Due to its potential for minimizing surgical trauma, *neuronavigation* (image-guided neurosurgery system) has been a ubiquitous tool for many neurosurgical procedures. In preoperative (p-) planning, neuronavigation offers surgeons the images, i.e., Magnetic Resonance (MR), necessary to understand patient-specific information and allow them to choose the most appropriate surgical strategy. During the surgery, neuronavigation provides a patient-to-image mapping so that surgeons can point to a specific location on the patient and see the corresponding anatomy in the p-MR image, helping them achieve a complete tumor resection while avoiding damage to surrounding functioning brain tissue. However, intraoperative (i-) deformation of the brain, also known as *brain shift*, invalidates the image-to-patient mapping, thus makes it unreliable to use p-MR for intraoperative surgical guidance.

The most successful way to compensate for the brain shift is using intra-operative images, i.e., i-MR or i-Ultrasound (US). By transforming the brain shift invalidated p-MR image to i-images via *image registration*, surgeons can get an updated view of the pre-surgical planning during surgery. The limitation of this strategy is that, even though the brain is clearly experiencing non-rigid deformation due to tumor resection or retraction, standard neuronavigation in clinical practice only integrates sub-optimal rigid registration, which is insufficient for accurate brain shift compensation. As a result, most surgeons use neuronavigation to approach a surgical target but justifiably do not trust it throughout the entire operation.

Incorporating non-rigid registration has long been a goal for neuronavigation,

yet this goal is hampered because it is harder to predict, validate and understand non-rigid registration error. In practice, if surgeons see a discrepancy between two aligned image features, they may not be able to tell if it is caused by a registration error or an actual tissue deformation. In this case, providing surgeons with a spatial distribution of the expected registration error could help them make more informed decisions, e.g., ignoring the registration where the expected error is high. However, determining this spatial distribution of error is particularly difficult for neurosurgery because: 1) Many existing methods conduct multiple runs of the non-rigid registration algorithm and estimate the error based on the (in)consistency of registration results. These methods are too time-consuming to be practical in the operating room because feedback is required within a few minutes of i-image acquisition; 2) More importantly, most error estimation methods essentially look for discrepancies in aligned image features. They are inappropriate for neurosurgery because tumor resection and retraction significantly alter the vicinity of the surgical field, particularly at tumor margin. Thus inconsistency near the tumor margin, which is often expected, can be mistakenly reported as registration errors.

To address these challenges, we propose to use *registration uncertainty* as a surrogate to indicate registration error in neuronavigation. Registration uncertainty predicts the trustworthiness of registration results and can be helpful in clinical practice. For example, if surgeons observe a large discrepancy at location A and small discrepancy at location B, without knowledge of registration uncertainty, they would most likely assume a large error everywhere and thus ignore the registration. With accurate knowledge of uncertainty, once surgeons know that A lies in an area of high uncertainty while B lies in an area of low uncertainty, they would have greater confidence in the registration at B and other locations of low uncertainty.

In this dissertation, we attempt to establish an important foundation for utilizing registration uncertainty in neuronavigation. The main contributions of this dissertation are as follows:

1) Most registration approaches in neuronavigation have difficulties in registering image pairs with artifacts and missing correspondences, e.g., a part of the tumor volume in the p-image may be missing in the i-image.

Another shortcoming of existing approaches is the lack of an uncertainty measure. Since brain shift is a complex spatio-temporal phenomenon, given the state of registration technology, it is reasonable to expect an indication of the confidence level in the estimated deformation. We developed a fast probabilistic *active image registration* method which provides registration uncertainty and meanwhile is robust against image pairs with missing correspondence.

2) Registration uncertainty is a useful addition to the registration result. However, the majority of research takes registration uncertainty for granted and use it in ad hoc ways. We investigated the applicability of registration uncertainty and categorized it into transformation uncertainty and label uncertainty. We pointed out that using transformation uncertainty to quantify label uncertainty, which is a widely adopted by the registration community, is inappropriate and can be misleading. We also shared a potentially critical finding that making use of the label uncertainty may not always be helpful.

3) A key assumption for using uncertainty to indicate the error is that these two quantities are monotonically related. While this notion is intuitive and believed by some clinicians, it has never been examined in the image registration literature. We systematically investigated the putative monotonic association between Gaussian process registration uncertainty and error based on neurosurgical data and showed empirically that there is a weak-to-moderate positive monotonic correlation between point-wise GP registration uncertainty and error. This work also opens a new vista for the uncertainty/error relationship analysis.

4) During the course of research, we also pursued two related projects. The first project is about using the variogram to screen outliers for vector fields. Since a key step in the proposed *active image registration* is to interpolate a dense deformation field from a set of sparse vectors, adding an outlier screening step can improve the registration accuracy. In the second project, we used the variogram to perform a third-party screening on the annotation of two public datasets. We found that (1) a small number of annotations may have fiducial localization errors; (2) the landmark distribution for some cases is not ideal to offer fair comparisons. If unresolved, both findings could incur bias in registration evaluation.

In summary, this dissertation attempts to establish a foundation for using registration uncertainty to indicate error in the context of neurosurgery. We developed methods that estimate, investigate and understand registration uncertainty. We believe that our contributions can initiate a paradigm shift from error to uncertainty in neuronavigation.

Acknowledgements

Firstly, I would like to express my deepest appreciation to my advisor Professor Masashi Sugiyama for his supervision, encouragement, and continuous support, without which I could not have been where I am today.

Next, I would like to extend my gratitude to Professor William Wells III for his guidance and invaluable conversation (about research and beyond), which helped me tremendously in pursuing my academic goals.

My sincere thanks also go to Professor Sarah Frisken. Her passion and kindness motivated me to become a better researcher and a better person.

Furthermore, I'm very fortunate to have worked with the following people: Professor Alexandra Golby, Professor Tina Kapur, Professor Matthew Toews, Dr. Steve Pieper, Professor Miaomiao Zhang, Dr. Frank Preiswerk, Dr. Alireza Sedghi, Dr. Ines Machado, Professor Polina Golland, and Marianna Jakab.

I also want to thank my thesis committee for their constructive comments that helped me improve the thesis.

Last but not least, I would like to thank my parents for their unwavering and long-lasting support.

Chapter 1

Introduction

This dissertation establishes a foundation for using image *registration uncertainty* to indicate image registration error in the context of *neuronavigation*. In this chapter, we review the background, motivation, and contributions of the dissertation. We also give a brief introduction to brain tumor, *neuronavigation*, *brain shift* and components of medical *image registration*.

1.1 Brain tumor

Brain tumor is a general term that refers to a mass of abnormally growing cells in and around the brain. Like other tumors, a brain tumor can be malignant (cancerous) or benign (non-cancerous). Although it can occur at any age, adults 40 to 70 years old and children 3 to 12 years old are most commonly affected. In 2020, an estimated of at least 23890 adults in the United States will be diagnosed with brain tumors.

This section reviews the classification, diagnosis, and treatment of a brain tumor. The sources of information for this section are primarily from the following articles (Assaf et al., 2017; Baig et al., 2016; cancer.net, 2020; Chilla et al., 2015; Ellingson et al., 2015; health.harvard.edu, 2019; hopkinsmedicine.org, 2019; Krishnatry et al., 2016; Maier et al., 2000, 2010; Schaefer et al., 2000; Villanueva-Meyer et al., 2017; Wikipedia, 2020a,b).

1.1.1 A Taxonomy of Brain Tumor

Brain tumors can be classified by the origin and by grade.

1.1.1.1 Classification by Origin

Based on the origin, there are *primary* and *secondary* brain tumors. Primary brain tumors arise from diseased brain tissues and can be further classified by the tissue in which they begin, for example:

- **Gliomas:** As the most common primary tumors, Gliomas start in the brain's glial cells, a type of supportive cell in the brain, and comprise about 80 percent of all malignant brain tumors.
- **Medulloblastomas:** Originated from early embryonic cells, Medulloblastomas are the most common malignant primary tumor in children.
- **Meningiomas:** Meningiomas is a slow-growing benign tumor that relates to the membranous layers surrounding the brain and spinal cord.

Secondary brain tumors are metastatic and malignant. They occur when cancer from another part of the body spreads, such as lungs and breast, to the central nervous system. Secondary brain tumors are much more common than primary tumors.

1.1.1.2 Classification by Grade

For brain tumors, such as Gliomas, it is common to classify them according to their grades (aggressiveness). A higher grade indicates more aggressive and more likely to overgrow. The types of grades include:

- **Grade 1:** A G1 tumor is biologically benign and have (comparatively) low risk.
- **Grade 2:** A G2 or low-grade brain tumor grows slowly, yet it can spread into nearby tissues and develop into higher grades over time.
- **Grade 3:** G3 or high-grade indicates a cancerous tumor that carries a worse prognosis.
- **Grade 4:** A G4 brain tumor is very aggressive and often fatal. e.g., Glioblastoma is the G4 Glioma that is alarmingly common among adults. It has an estimated 2-year survival rate of around 17 percent.

Regardless of the grade, all brain tumors are serious. A growing tumor is eventually going to compress other structures in the brain and cause implications.

1.1.2 Diagnosis of Brain Tumor

Most brain tumors are not diagnosed until after symptoms appear. The symptoms of a brain tumor often include: headaches, seizures, sensory changes, memory or personality changes, nausea or vomiting, and complex partial.

In the medical examination, doctors usually ask about the patient's symptoms, health habits, and history of illness. Then they perform a neurological test to examine the condition of the patient, e.g., muscle strength, eyesight, reflexes, etc.

In conjunction with the medical examination, doctors may recommend the following medical imaging tests because tumor tissues often show as differently colored masses in medical images. Typical medical image modalities are:

- **Computed tomography (CT) scan.** The CT uses an x-ray camera that rotates around the body to create cross-sectional images of the brain. On a CT image, benign brain tumors look darker than normal brain tissue. Patients are sometimes injected with a dye called contrast agent before the scan to increase tumor visibility.

- **Magnetic resonance imaging (MRI).** The MR test uses magnetic fields to produce detailed images of the brain. An MRI scan captures some brain structures better than a CT scan, and it is the reference standard for brain tumor diagnosis. Magnetic resonance angiogram is an MRI variant that visualizes the blood flow. It can help doctors find aneurysms and better define tumors. Diffusion-weighted imaging is another MR technique that is useful in classifying tumors by grade.

- **Positron emission tomography (PET) scan.** Cancer cells consume more glucose than healthy cells. In the PET test, radioactive glucose is injected into the patient, then areas, where cells are consuming lots of glucose, are highlighted by the PET scan.

Using medical imaging alone can not diagnose the prognosis of a brain tumor, i.e., whether it is a high-grade or low-grade glioma. A sample of the tumor tissues is usually needed to confirm the prognosis. Doctors can obtain the tumor sample via brain biopsy, which is the removal of a small amount of tissue under a needle procedure or during tumor resection. Pathologists then analyze the samples and conclude the prognosis.

1.1.3 Treatment of Brain Tumor

Doctors and other health care professionals often work together to create a patient's treatment plan.

Factors that can be taken into account for making the treatment plan include:

- Tumor size, type, location, and grade
- The patient's age and overall health
- The patient's preferences
- Possible side effects

A treatment plan may combine different types of treatment. There are three primary treatment options:

- **Surgery** is the process of removing the tumor mass and some surrounding healthy tissue.
- **Radiation therapy** uses high-energy x-rays or other particles to kill tumor cells. Since high-dose radiation can damage normal cells, doctors try to precisely target the tumor, limiting the amount of radiation to surrounding parts of the brain.
- **Chemotherapy** is the use of drugs to destroy tumor cells and prevent the tumor from growing. In general, chemotherapy tends to be less effective against gliomas than surgery or radiation therapy.

Surgery is the preferred (initial) treatment for all brain tumors. However, in some cases, performing the surgery is too risky. For example, the tumor may be surrounded by functioning brain tissues. Damage to these tissues during surgery could cause severe neurological deficits.

For a low-grade brain tumor, surgery may be the only treatment needed, especially if all of its volume can be removed.

For higher-grade tumors, the treatment plan typically consists of a combination of surgery, radiation therapy, and chemotherapy. During the surgery, surgeons are likely to remove as much tumor mass as possible. The goal of surgery is to help relieve symptoms and reduce the burden of follow-up radiation therapiey and/or chemotherapy.

1.2 Neuronavigation

Successful neurosurgery is the foundation for brain tumor treatment. *neuronavigation* can provide surgeons with precise localization of surgical targets, thus it has been an indispensable tool for many neurosurgical procedures.

This section gives a brief review of neuronavigation. The sources of information for this section are the following articles (Chartrain et al., 2017; Enchev, 2009; Gerard et al., 2017; Mercier et al., 2010; Muacevic et al., 2000; Orringer et al., 2012; Reinertsen et al., 2007; Seeger and Zentner, 2002; Slavin, 2008; Wikipedia, 2020c; Willems et al., 2006).

1.2.1 Overview

With the advent of modern medical imaging techniques, intracranial anatomy, such as the tumor and surrounding soft tissues, can be presented to the surgeon in detail via, e.g., MRI images.

In essence, neuronavigation is the set of imaging-assisted technologies that allow surgeons to "navigate" within the skull confines during surgery. The benefits of neuronavigation are two folds:

- **Preoperative use:** Based on the information provided by neuronavigation systems, surgeons can preoperatively evaluate the operating risks

and make the most appropriate surgical plan. Moreover, such systems help surgeons find potential surgical corridors through non-critical areas and enable surgeries of previously inoperable cases.

- **Intraoperative use:** During the surgery, neuronavigation systems can help surgeons determine the position of a surgical instrument relative to the surgical field, and in advance, achieve a complete tumor resection while avoiding damage to surrounding functioning brain tissues.

1.2.2 Workflow

The main components of neuronavigation are a computer, a display that shows images and other information, and a tracking device. Optional accessories include navigation probes and reference frames.

A typical workflow of using neuronavigation systems for surgical guidance can be summarized in 4 steps:

1. **Image scanning:** Medical images that are necessary for the surgery are acquired.

2. **Surgical planning:** Clinicians fuse anatomical information from all preoperative images by co-registering them into the same coordinate space and use them to make a plan towards the best surgical outcome.

3. **Patient-to-Image mapping:** Neuronavigation systems create a mapping that relates the patient's intraoperative physical space to preoperative images. This step is often carried out by landmark (or fiducial) matching.

4. **Intraoperative navigation:** Based on the mapping in step 3, surgeons can place a tracked instrument in the patient's physical space, and its corresponding anatomical location can be shown on displayed preoperative images.

1.2.3 Limitations

The major limitation of neuronavigation is the loss of accuracy in the patient-to-image mapping step. If the mapping is inaccurate, a physical location on the patient could be incorrectly related to an anatomy shown on the display, which may cause severe consequences.

There are two sources of inaccuracy:

1. **Technical inaccuracies:** Technical inaccuracies vary between systems and arise when the acquired preoperative images have distortion, or the associated hardware has positioning error.

2. **Brain shift-induced inaccuracies:** Since images are taken before the surgery, for them to remain useful for intraoperative guidance, it is implicitly assumed that structures of the brain stay in the same position throughout the surgery as when they were imaged. However, intraoperative deformation of the brain, also know as *brain shift*, invalidates the patient-to-image mapping, thus makes it less reliable to use preoperative images for intraoperative guidance.

Unlike technical inaccuracies that can be managed with care, there is no obvious way to compensate for the brain shift-induced inaccuracies, which is the most significant source of error in neuronavigation.

1.3 Brain Shift

Brain shift is defined as any factor that creates a discrepancy of anatomy location between the physical and image spaces.

In this section, we review the causes of brain shift and strategies to compensate for it. The sources of information for this section are the following articles (Bayer et al., 2017a; Correa-Arana et al., 2017; Gerard et al., 2017; Luo et al., 2017; Miga, 2016; Miller, 2019; Skrinjar et al., 1998).

1.3.1 Causes of Brain Shift

We classify the causes of brain shift into physical, surgical and biological categories.

Physical factors

- Hardware movement: During the surgery, the patient's head is immobilized by a head clamp. Any changes to the head position or reference frame could cause brain shift.
- Patient positioning: Compared to the position during preoperative image acquisition, if the patient is positioned differently in the operating room, it can lead to a discrepancy between the anatomy location in physical and image spaces.
- Gravity: Gravitational forces are perhaps the most substantial physical factor that contributes to brain shift. The patient has to be well-positioned to minimize the loss of blood and cerebrospinal fluid (CSF), as well as the sagging of brain tissues.

Surgical factors

- Tissue loss: The removal of pathological and healthy tissue is the main contributor to brain shift. Because of the effect of gravity, tumor resection, or retraction can cause the unsupported surrounding tissue to sag. As the surgery progresses, the brain also deforms due to swelling, fluid loss, and other surgical interactions.
- Fluid loss: For many surgical procedures, the CSF is intentionally evacuated to reduce the Intracranial pressure and allow brain relaxation. Since avoiding excessive bleeding is prioritized during surgery, the loss of blood is an insignificant factor for causing brain shift.

Biological factors

- Tumor type: Some research have reported that the type of tumor could affect the underlying movement of the brain. Gliomas tend to cause frequent and unpredictable brain shift during surgeries.

- Drugs: Before surgery, the patient sometimes are given drugs to reduce intracranial pressure. These drugs change the conformation of anatomy, and as a result, induce brain shift.

1.3.2 Brain Shift Compensation

Brain shift compensation is the key to increasing the feasibility of neuronavigation and is an active area of research. There are two main streams in the literature of brain shift compensation. Their difference is whether to take advantage of intraoperative images (images taken during the surgery).

1.3.2.1 Predictive Modeling

Predictive modeling-based methods do not need intraoperative information (images). These methods attempt to directly predict the intraoperative brain deformation to correct “obsolete" preoperative images (Dumpuri et al., 2007; Zhuang et al., 2010). However, any changes caused by the surgery are not taken into account.

Predictive modeling-based methods are useful in scenarios when intraoperative data is absent or insufficient. They can even be regarded as a complementary tool for intraoperative image-based compensation methods.

1.3.2.2 Intraoperative Images

The most successful way to compensate for the brain shift is using intraoperative images, i.e., intraoperative MR (i-MR) or intraoperative Ultrasound (i-US).

Image Registration Image registration-based methods estimate a mathematical transformation that spatially aligns preoperative images to the coordinate space of intraoperative images (Goshtasby, 2012; Maintz and Viergever, 1998; Sotiras et al., 2013). They provide surgeons with updated pre-surgical planning during surgery. Image registration-based methods may add extra

operation time to the surgery. However, they are currently the most accurate brain shift compensation approaches (Gerard et al., 2017).

Biomechanical Modeling Biomechanical modeling methods attempt to describe the mechanics of brain deformation using linear elastic (for small deformation) or hyperelastic (for large deformation) models (Luo et al., 2017, 2019b; Miga, 2016; Miller, 2019). These methods are often used in conjunction with image registration to achieve a faster and more accurate brain shift compensation (Frisken et al., 2020). The drawbacks of biomechanical modeling is that building such a model with unknown boundary conditions is time-consuming and patient-specific, which in general yield poor generality.

In general, image registration is the most promising strategy for brain shift compensation. We explain each step of image registration in detail in the next section.

1.4 Image Registration

Medical *image registration* is the process of bringing two sets of image data into spatial alignment (Goshtasby, 2012; Maintz and Viergever, 1998; Oliveira and Tavares, 2014; Pluim et al., 2003; Sotiras et al., 2013; Viergever et al., 2016). It has become a standard routine for a large number of clinical applications.

In this section, we introduce the basics of image registration to give an overview of topics related to our work for a broader scope of the field.

1.4.1 Introduction

In radiology, *modality* is the term which refers to one form of imaging, e.g., MRI scanning (Robb, 2009). In general, there are two kinds of modalities:

- Anatomical modalities, such as MRI and US, depict the shape morphology of organs and some tissue information.
- Functional modalities, such as functional MRI, capture the information on the metabolism.

Since the information provided by two images acquired during a clinical workflow can be complementary/correlated, proper integration of different images is beneficial and highly sought after by clinicians. Image registration is the first step of this integration process. Based on the modalities of input images, image registration can be categorized into *uni-modal* and *multi-modal* registration (Maintz and Viergever, 1998).

For the two images that are going to be registered, let $S : \Omega \to \mathbb{R}, \Omega \subset \mathbb{R}^d$ and $T : \Omega \to \mathbb{R}, \Omega \subset \mathbb{R}^d$ denote the source (moving) and target (fixed) image respectively, where Ω is the image domain, and d indicates the image dimension ($d = 2$ or 3).

Image registration attempts to spatially map S to the coordinate system of T by estimating a transformation $W : \Omega \to \Omega$ that minimize the cost function:

$$\mathcal{M}(T, S \circ W) + \mathcal{R}(W). \tag{1.1}$$

The cost function has two terms: the first term $\mathcal{M}$ is a similarity metric that measures the disagreement between target image T and transformed source image $S \circ W$. The second term $\mathcal{R}$ is a regularization term which works to encourage specific properties of the transformation W (Maintz and Viergever, 1998; Sotiras et al., 2013).

In summary, the main components of a registration algorithm consist of (1) a transformation model W (including regularization R); (2) a similarity measure $\mathcal{M}$; and (3) an optimization method. W and $\mathcal{M}$ have significant impact on the performance of the registration algorithm.

1.4.2 Deformation Model and Regularization

The choice of transformation model W implies an assumption with the underlying tissue deformation to be compensated. The complexity of W governs the performance of the registration algorithm (Sotiras et al., 2013).

As the degrees of freedom increase, W is expected to have more expressive power. As a result, the number of parameters that have to be estimated also increases. Thus there is a trade-off between the computational efficiency and richness of the description.

Rigid transformation is a simple geometric transformation that preserves the Euclidean distance between every pair of points in the coordinate system. Let $\mathbf{x} \in \Omega$ be the position of a point, and its rigid transformation can be written as:

$$W_{\text{rigid}}(\mathbf{x}) = Q\mathbf{x} + \mathbf{b}. \tag{1.2}$$

Here, Q is a rotation matrix and $\mathbf{b}$ is a translation vector. A rigid transformation has minimal descriptive power with only six parameters. However, due to the simplicity, it has been the most prevalent transformation model and sometimes good enough for skull registration (Goshtasby, 2012; Maintz and Viergever, 1998; Song, 2017).

Non-rigid transformation models offer more degrees of freedom and are more suitable for modeling tissue deformation. The free-form deformation model is a popular non-rigid transformation model that has had great success in many clinical applications (Rueckert et al., 1999).

The most descriptive non-rigid transformation model is the dense displacement field (DDF). For $\mathbf{x} \in \Omega$, its DDF is

$$W_{\text{DDF}}(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x}), \tag{1.3}$$

where $\mathbf{u}$ is a displacement vector associated with position $\mathbf{x}$.

Regularization R has a close relationship with non-rigid transformation models. It is in the cost function of non-rigid image registration for two purposes:

1. In the case that non-rigid transformation model W is parameterized by a large number of variables, the registration is often a under-determined problem, i.e., W has more unknown parameters than equations needed to estimate them. A regularization term can reduce the degree of freedom for W by inducing correlations among parameters. i.e., it is common to minimize the bending energy to correlate displacement vector values in a local neighborhood at each position x (Sotiras et al., 2013).

2. In the case that non-rigid transformation model W only has a small set of parameters, a regularization term can introduce prior knowledge

(task-specific constraints) to improve the registration.

Image registration algorithms can incorporate one or more types of regularization.

1.4.3 Similarity Metrics

Similarity metrics $\mathcal{M}$ measure the disagreement between two images. Most registration approaches use hand-crafted similarity metrics. Designing an appropriate $\mathcal{M}$ should take into account two factors:

1. Convexity: The convexity of the cost function depends on the choice of $\mathcal{M}$ and the images being registered. A convex $\mathcal{M}$ reduces the difficulty of parameter search, and facilitate the inference of registration solutions. However, since being non-convex is the nature of most spatial alignment problems, convexity may also lead to unrealistic deformation (Sotiras et al., 2013).

2. Discrimination: An ideal $\mathcal{M}$ should accurately reflect the quality of image registration: taking low values when two images are well aligned.

Since the same anatomical structures share a similar intensity range on the same image modality, a similarity metric for a uni-modal image registration is straightforward to define. Popular choices include sum-of-squared differences (SSD) (Bajcsy and Kovacic, 1989) and attribute-based methods (Shen and Davatzikos, 2002).

Devising the similarity metric $\mathcal{M}$ for a multi-modal image registration is more challenging because $\mathcal{M}$ should be able to account for the intensity relation between physical principles behind different image acquisition methods. Information theoretical similarity metrics are proven to be effective in some multi-modal registration problems (Maes et al., 1997; Wells et al., 1996).

Some similarity metrics work for both uni-modal and multi-modal registration problems, thus are called modality-independent metrics(Heinrich et al., 2012).

If some anatomical landmarks on S, and their corresponding counterparts on T can both be identified, landmark distance (LD) is another useful similarity metric that works for both uni- and multi-modal registration(Song, 2017).

It's noteworthy that, even though the majority of existing registration methods use hand-crafted similarity metrics, some deep learning-based registration approaches propose to automatically learn a task-specific metric, which may be advantageous (Simonovsky et al., 1998).

1.4.4 Optimization

Some registration methods have an exact solution for their cost functions, e.g., affine registration with corresponding landmarks. More registration methods use iterative optimization to update the transformation parameters so that the cost function is minimized (Sotiras et al., 2013).

Based on the nature of estimated parameters, optimization method can be categorized as:

- Continuous: When the cost function is differentiable, continuous optimization can be used to obtain real-valued transformation parameters.
- Discrete: Discrete optimization, e.g., graph-based methods, are entitled to problems where the variables take discrete values (Glocker et al., 2009).

Traditionally, given a complex transformation model, iterative optimization methods tend to be time-consuming. However, there are exceptions, even though new learning-based registration methods use gradient descent to update the transformation parameters iteratively, these methods transfer the computational burden to the training stage, and can considerably speed up the registration process(Dalca et al., 2018; Hu et al., 2019).

1.4.5 Evaluation

Evaluation for image registration is a challenging and less studied topic. The main criterion for registration evaluation is the registration accuracy, which can be categorized as

1. Qualitative: Qualitative accuracy can usually be assessed using visual inspection, e.g., when registering MRI and US brain images, overlaying the registered region of interest can provide clinicians with a reasonable knowledge of accuracy (Goshtasby, 2012; Maintz and Viergever, 1998).

2. Quantitative: Because of the lack of "ground truth" correspondences, direct estimation of quantitative accuracy is particularly difficult. One strategy is to use the target registration error (TRE). TRE is defined as the sum of Euclidean distance between each pair of registered points. In practice, such points are annotated corresponding landmarks on the source and target images (Fitzpatrick, 2007; Song, 2017; Sonka, 2000; West et al., 1997).

Registration precision is defined as the systematic error of the registration algorithm. It is another useful criterion for registration evaluation, yet less prevalent than registration accuracy (Maintz and Viergever, 1998).

Synthetic, phantom, and clinical data can all be used to evaluate the accuracy and precision of registration algorithms.

1.4.6 Registration Uncertainty

Given the current state of the registration technology and the difficulty of some problems, it is expected that the registration algorithm has errors. An uncertainty measure that highlights locations where the algorithm had difficulty finding a proper alignment can be beneficial. (Risholm et al., 2010; Risholm and Wells, 2011).

Among the approaches that characterize the uncertainty of image registration, the most popular, or perhaps the most successful framework is the probabilistic image registration (PIR) (Luo et al., 2018b; Risholm et al., 2013)

In contrast to traditional "point-estimate" image registration approaches that report a unique set of transformation parameters that best align two images, PIR models transformation parameters as random variables and estimates distributions over them. The mode of the distribution is then chosen as the most likely value of that transformation parameter. PIR has the advantage that the registration uncertainty can be naturally obtained from the distribution of transformation parameters. PIR methods can be broadly categorized into discrete probabilistic registration (DPR) (Cobzas and Sen, 2011; Heinrich et al., 2016; Lotfi et al., 2013; Popuri et al., 2013) and continuous probabilistic registration (CPR) (Dalca et al., 2018; Folgoc et al., 2017; Janoos et al., 2012;

Luo et al., 2018b; Risholm et al., 2013; Sedghi et al., 2018; Simpson et al., 2015, 2012; Wang et al., 2018; Wassermann et al., 2014; Yang and Niethammer, 2015; Zhang et al., 2013)

The transformation distribution estimated by DPR and CPR have different forms. DPR discretizes the transformation space into a set of displacement vectors. It then uses discrete optimization techniques to compute a categorical distribution for every voxel as its transformation. CPR is essentially a probabilistic inference model with the estimated transformation given by a multivariate continuous posterior distribution.

1.5 Motivation

A limitation of the current neuronavigation systems is that, even though the brain is clearly undergoing non-rigid deformation due to tumor resection or retraction, standard neuronavigation in clinical practice only integrates rigid registration, which is insufficient for accurate brain shift compensation. As a result, most surgeons use neuronavigation to approach a surgical target but justifiably do not trust it throughout the entire operation, especially near the end of the procedure when the information is most important (Gerard et al., 2017).

Even though some non-rigid registration approaches have achieved better accuracy in brain shift compensation (Luo et al., 2018b; Machado et al., 2019, 2018; Xiao et al., 2020), in the operating room, rigid-registration is still the clinical routine. Incorporating non-rigid registration has long been a goal for neuronavigation, yet this goal is hampered because it is harder to predict, validate and understand non-rigid registration error.

In practice, if surgeons see a discrepancy between two aligned image features, they may not be able to tell if it is caused by a registration error or an actual tissue deformation. In this case, providing surgeons with a spatial distribution of the expected registration error could help them make more informed decisions, e.g., ignoring the registration where the expected error is high.

However, determining this spatial distribution of error is particularly difficult for neurosurgery because:

1. Many existing methods conduct multiple runs of the non-rigid registration algorithm and estimate the error based on the (in)consistency of registration results. These methods are too time-consuming to be practical in the operating room because feedback is required within a few minutes of intraoperative image acquisition (Datteri and Dawant, 2012; Hub and Karger, 2013; Hub et al., 2009; Kybic, 2010; Shams et al., 2017);

2. More importantly, most error estimation methods essentially look for discrepancies in aligned image features (Saygill, 2018; Saygill et al., 2016; Sokooti et al., 2019). They are inappropriate for neurosurgery because tumor resection and retraction significantly alter the vicinity of the surgical field, particularly at tumor margin. Thus inconsistency near the tumor margin, which is often expected, can be mistakenly reported as registration errors.

To address these challenges, we propose to use *registration uncertainty* as a surrogate to indicate registration error in neuronavigation. Registration uncertainty measures the trustworthiness of registration results and can be helpful in clinical practice (Luo et al., 2020a, 2019a, 2018b). For example, if surgeons observe a large discrepancy at location A and small discrepancy at location B, without knowledge of registration uncertainty, they would most likely assume a large error everywhere and thus ignore the registration. With accurate knowledge of uncertainty, once surgeons know that A lies in an area of high uncertainty while B lies in an area of low uncertainty, they would have greater confidence in the registration at B and other locations of low uncertainty.

1.6 Contributions

In this dissertation, we establish an important foundation for utilizing registration uncertainty in neuronavigation. We develop methods that estimate, investigate and understand registration uncertainty. We believe that contributions in this dissertation could increase the feasibility of non-rigid registration in interventional guidance and advance the state of neuronavigation.

1.7 Dissertation Overview

This dissertation is divided into four main chapters.

Chapter 2: Active Image Registration Most registration approaches in neuronavigation have difficulties in registering image pairs with artifacts and missing correspondences, e.g., a part of the tumor volume in the p-image may be missing in the i-image. Another shortcoming of existing approaches is the lack of an uncertainty measure. Since brain shift is a complex spatio-temporal phenomenon, given the state of registration technology, it is reasonable to expect an indication of the confidence level in the estimated deformation. We develop a fast probabilistic *active image registration* method which provides registration uncertainty and meanwhile is robust against image pairs with missing correspondence (Luo et al., 2018b).

Chapter 3: On the applicability of registration uncertainty Registration uncertainty is a useful addition to the registration result. However, the majority of research takes registration uncertainty for granted and use it in ad hoc ways. We investigate the applicability of registration uncertainty and categorize it into transformation uncertainty and label uncertainty. We point out that using transformation uncertainty to quantify label uncertainty, which is widely adopted by the registration community, is inappropriate and can be misleading. We also share a potentially critical finding that making use of the label uncertainty may not always be helpful (Luo et al., 2019a).

Chapter 4: Are Registration Uncertainty and Error Monotonically Associated? A key assumption for using uncertainty to predict the error is that these two quantities have a monotonic relationship. While this notion is intuitive and believed by many clinicians, it has never been examined in the image registration literature. We systematically investigate the monotonic association between Gaussian process registration uncertainty and error based on neurosurgical data and show empirically that there is a weak-to-moderate positive monotonic correlation between point-wise GP registration uncertainty and

error. This work also opens a new vista for the uncertainty/error relationship analysis (Luo et al., 2020a).

Chapter 5: Related Projects During the course of research, we have also pursued two related projects. The first project is about using the variogram, which is a geostatistical tool for capturing spatial dependence, to screen outliers for vector fields. Since a key step in the proposed *active image registration* is to interpolate a dense deformation field from a set of sparse vectors, adding an outlier screening step can improve the registration accuracy. In the second project, we use the variogram to perform a third-party screening on the annotation of two public datasets. We found that (1) a small number of annotations may have fiducial localization errors; (2) the landmark distribution for some cases is not ideal to offer fair comparisons. If unresolved, both findings could incur bias in registration evaluation (Luo et al., 2018a, 2020b).

Chapter 6: Conclusion and Future Work We conclude our work with some future prospectives of this line of research

Chapter 2

Active Image Registration

A reliable Ultrasound (US)-to-US registration method to compensate for brain shift would substantially improve Image-Guided Neurological Surgery. Developing such a registration method is very challenging, due to factors such as the tumor resection, the complexity of brain pathology and the demand for fast computation. We propose a novel feature-driven active registration framework. Here, landmarks and their displacement are first estimated from a pair of US images using corresponding local image features. Subsequently, a Gaussian Process (GP) model is used to interpolate a dense deformation field from the sparse landmarks. Kernels of the GP are estimated by using variograms and a discrete grid search method. If necessary, the user can actively add new landmarks based on the image context and visualization of the uncertainty measure provided by the GP to further improve the result. We retrospectively demonstrate our registration framework as a robust and accurate brain shift compensation solution on clinical data. This chapter is based on our publication of (Luo et al., 2018b).

2.1 Introduction

Commercial IGNSs assume a rigid registration between preoperative imaging and patient coordinates. However, intraoperative deformation of the brain, also known as brain shift, invalidates this assumption. Since brain shift progresses during surgery, the rigid patient-to-image mapping of IGNS becomes less and less accurate. Consequently, most surgeons only use IGNS to make a surgical

plan but justifiably do not trust it throughout the entire operation (Bayer et al., 2017b; Gerard et al., 2017).

2.1.1 Related Work

As one of the most important error sources in IGNS, intraoperative brain shift must be compensated in order to increase the accuracy of neurosurgeries. Registration between the intraoperative MRI (iMRI) image and preoperative MRI (preMRI) image (preop-to-intraop registration) has been a successful strategy for brain shift compensation (Clatz et al., 2005; Drakopoulos et al., 2014; Hata et al., 1999; Vigneron et al., 2012). However, iMRI acquisition is disruptive, expensive and time consuming, making this technology unavailable for most clinical centers worldwide. More recently, 3D intraoperative Ultrasound (iUS) appears to be a promising replacement for iMRI. Although some progress has been made by previous work on preMRI-to-iUS registration (Arbel et al., 2004; Fuerst et al., 2014; Gobbi et al., 2000; Letteboer et al., 2003; Pennec et al., 2003; Reinertsen et al., 2004; Rivaz and Collins, 2015), yet there are still no clinically accepted solutions and no commercial neuro-navigation systems that provide brain shift compensation. This is due to three reasons: 1) Most non-rigid registration methods (including optical flow and deep learning registration methods) can not handle artifacts and missing structures in iUS; 2) The multi-modality of preMRI-to-iUS registration makes the already difficult problem even more challenging; 3) A few methods (Ou et al., 2011) can achieve a reasonable alignment, yet they take around 50 minutes for an US pair and are too slow to be clinically applicable. Another shortcoming of existing brain shift compensation approaches is the lack of an uncertainty measure. Brain shift is a complex spatio-temporal phenomenon and, given the state of registration technology and the importance of the result, it seems reasonable to expect an indication (e.g. error bars) of the confidence level in the estimated deformation.

2.1.2 Contributions

In this chapter, we propose a novel feature-driven active framework for brain shift compensation. Here, landmarks and their displacement are first estimated

from a pair of US images using corresponding local image features. Subsequently, a Gaussian Process (GP) model (C.E. and Williams, 2006) is used to interpolate a dense deformation field from the sparse landmarks. Kernels of the GP are estimated by using variograms and a discrete grid search method. If necessary, for areas that are difficult to align, the user can actively add new landmarks based on the image context and visualization of the uncertainty measure provided by the GP to further improve the registration accuracy. We retrospectively demonstrate the efficacy of our method on clinical data.

Contributions and novelties of our work can be summarized as follows:

1. The proposed feature-based registration is robust for aligning iUS image pairs with missing correspondence and is fast.
2. We explore applying the GP model and variograms for image registration.
3. Registration uncertainty in transformation parameters can be naturally obtained from the GP model.
4. To the best of our knowledge, the proposed active registration strategy is the first method to actively combine user expertise in brain shift compensation.

2.2 The Role of US-to-US Registration

In order to alleviate the difficulty of preop-to-intraop registration, instead of directly aligning iMRI and iUS images, we choose an iterative compensation approach which is similar to the work in (Riva et al., 2017).

As shown in Figure 2.1, the acquisition processes for pre-duraUS (preUS) and post-resectionUS (postUS) take place before opening the dura and after (partial) tumor resection, respectively. Since most brain-shift occurs after taking the preUS, a standard multi-modal registration may be suffice to achieve a good alignment T_{multi} between preMRI and preUS (Fuerst et al., 2014). Next, we register the preUS to postUS using the proposed feature-driven active framework to acquire a deformable mapping T_{mono}. After propagating T_{multi} and T_{mono} to the preMRI, surgeons may use it as an updated view of anatomy to compensate for brain shift during the surgery.

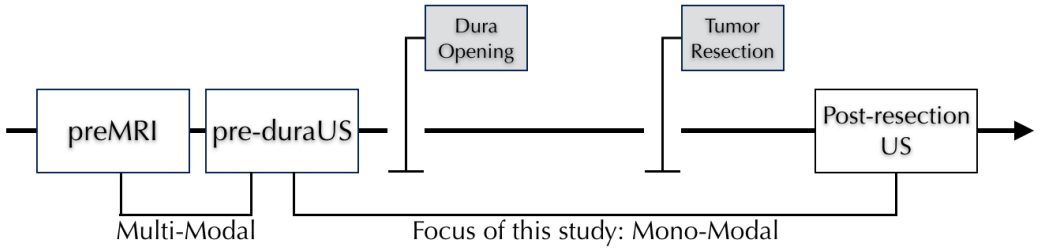


Figure 2.1: Pipeline of the US-based brain shift compensation.

2.3 Feature-based Registration Strategy

Because of tumor resection, compensating for brain shift requires non-rigid registration algorithms capable of aligning structures in one image that have no correspondences in the other image. In this situation, many image registration methods that take into account the intensity pattern of the entire image will become trapped in incorrect local minima.

We therefore pursue a Feature-Based Registration (FBR) strategy due to its robustness in registering images with missing correspondence (Toews and Wells, 2013). FBR mainly consists of 3 steps: feature-extraction, feature-matching and dense deformation field estimation. An optional "active registration" step can be added depending on the quality of FBR.

2.3.1 Feature Extraction and Matching

As illustrated in Figure 2.2(a)(b), distinctive local image features are automatically extracted and identified as key-points on preUS and postUS images. An automatic matching algorithm searches for a corresponding postUS key-point for each key-point on the preUS image (Toews and Wells, 2013).

For a matched key-point pair, let $\mathbf{x}_i$ be the coordinates of the preUS key-point and $\mathbf{x}_i^{\text{post}}$ be the coordinate of its postUS counterpart. We first use all matched PreUS key-points as landmarks, and perform a land-mark based preUS-to-postUS affine registration to obtain a rough alignment. $\mathbf{x}_i^{\text{post}}$ becomes $\mathbf{x}_i^{\text{affine}}$ after the affine registration. The displacement vector, which indicates the movement of landmark $\mathbf{x}_i$ due to the brain shift process, can be calculated as $\mathbf{d}(\mathbf{x}_i) = \mathbf{x}_i^{\text{affine}} - \mathbf{x}_i$. where $\mathbf{d} = [d_x, d_y, d_z]$.

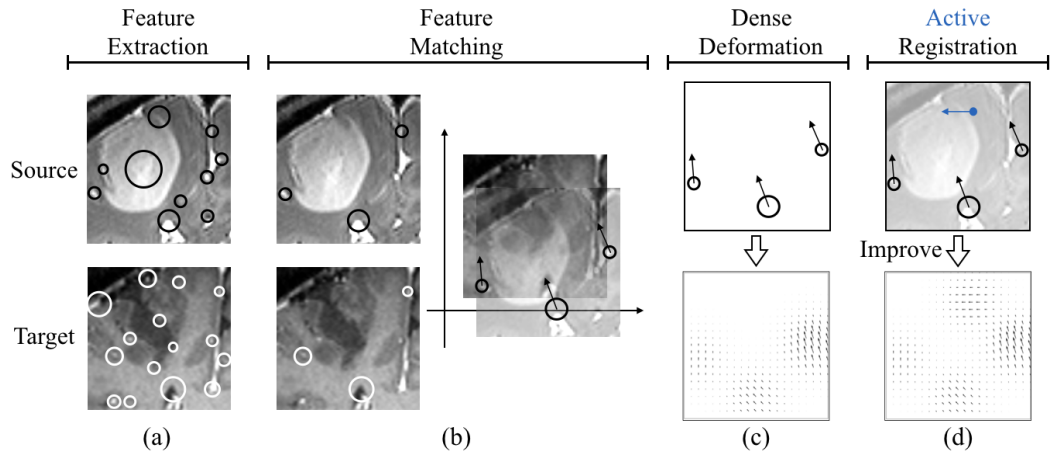


Figure 2.2: Pipeline of the feature-based active preUS-to-postUS registration.

2.3.2 Dense Deformation Field

The goal of this step is to obtain a dense deformation field from a set of N sparse landmark and their displacements $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{d}_i), i = 1 : N\}$, where $\mathbf{d}_i = \mathbf{d}(\mathbf{x}_i)$ is modeled as a observation of displacements.

In the GP model, let $\mathbf{d}(\mathbf{x})$ be the displacement vector for the voxel at location $\mathbf{x}$ and define a prior distribution as $d(\mathbf{x}) \sim \mathrm{GP}(\mathrm{m}(\mathbf{x}), \mathrm{k}(\mathbf{x}, \mathbf{x}'))$, where $\mathrm{m}(\mathbf{x})$ is the mean function, which usually is set to 0, and the GP kernel $\mathbf{k}(\mathbf{x}, \mathbf{x}')$ represents the spatial correlation of displacement vectors.

By the modeling assumption, all displacement vectors follow a joint Gaussian distribution $p(\mathbf{d} \mid \mathbf{X}) = \mathcal{N}(\mathbf{d} \mid \mu, \mathbf{K})$, where $K_{ij} = \mathbf{k}(\mathbf{x}, \mathbf{x}')$ and $\mu = (\mathrm{m}(\mathbf{x}_1), ..., \mathrm{m}(\mathbf{x}_N))$. As a result, the displacement vectors $\mathbf{d}$ for known landmarks and N_* unknown displacement vectors $\mathbf{d}_*$ at location $\mathbf{X}_*$, which we want to predict, have the following relationship:

$$\begin{pmatrix} \mathbf{d} \\ \mathbf{d}_* \end{pmatrix} \sim \mathrm{GP}\left(\begin{pmatrix} \mu \\ \mu_* \end{pmatrix}, \begin{pmatrix} \mathbf{K} & \mathbf{K}_* \\ \mathbf{K}_*^T & \mathbf{K}_{**}. \end{pmatrix} \right). \tag{2.1}$$

In Equation (1), $\mathbf{K} = \mathrm{k}(\mathbf{X}, \mathbf{X})$ is the $N \times N$ matrix, $\mathbf{K}_* = \mathrm{k}(\mathbf{X}, \mathbf{X}_*)$ is a similar $N \times N_*$ matrix, and $\mathbf{K}_{**} = \mathrm{k}(\mathbf{X}_*, \mathbf{X}_*)$ is a $N_* \times N_*$ matrix. The mean $\mu_* = [\mu_{*x}, \mu_{*y}, \mu_{*z}]$ represents values of voxel-wise displacement vectors and can be estimated from the posterior Gaussian distribution $p(\mathbf{d}_* \mid \mathbf{X}_*, \mathbf{X}, \mathbf{d}) = \mathcal{N}(\mathbf{d}_* \mid \mu_*, \Sigma_*)$ as

$$\mu_* = \mu(\mathbf{X}_*) + \mathbf{K}_*^T \mathbf{K}^{-1}(\mathbf{d} - \mu(\mathbf{X})). \tag{2.2}$$

Given $\mu(\mathbf{X}) = \mu(\mathbf{X}_*) = 0$, we can obtain the dense deformation field for the preUS image by assigning μ_{*x},μ_{*y},μ_{*z} to d_x, d_y and d_z, respectively.

2.3.3 GP Kernel Estimation

The performance of GP registration depends exclusively on the suitability of the chosen kernels and its parameters. In this study, we explore two schemes for the kernel estimation: Variograms and discrete grid search.

2.3.3.1 Variograms

The variogram is a powerful geostatistical tool for characterizing the spatial dependence of a stochastic process (Cressie, 1991). While being briefly mentioned in (Ruiz-Alzola et al., 2003), it has not yet received much attention in the medical imaging field.

In the GP registration context, where $\mathbf{d}(\mathbf{x})$ is modelled as a random quantity, variograms can measure the extent of pairwise spatial correlation between displacement vectors with respect to their distance, and give insight into choosing a suitable GP kernel.

In practice, we estimate the empirical variogram of landmarks' displacement vector field using

$$\hat{\gamma}(h \pm \delta) := \frac{1}{2|N(h \pm \delta)|} \sum_{(i,j) \in N(h \pm \delta)} \|\mathbf{d}(\mathbf{x}_i) - \mathbf{d}(\mathbf{x}_j)\|^2. \tag{2.3}$$

For the norm term $\|\mathbf{d}(\mathbf{x}_i) - \mathbf{d}(\mathbf{x}_j)\|$, we separate its 3 components d_x d_y d_z and construct 3 variograms respectively. As shown in Figure 2.3(a), for displacement vectors $\mathbf{d}(x_1)$ and $\mathbf{d}(x_2)$, $\|d_x(\mathbf{x}_2) - d_x(\mathbf{x}_1)\|$ is the vector difference with respect to the $\mathbf{x}$ axis, etc. h represents the distance between two key-points.

To construct an empirical variogram, the first step is to make a variogram cloud by plotting $\|d(\mathbf{x}_2) - d(\mathbf{x}_1)\|^2$ and h_{ij} for all displacement pairs. Next, we divide the variogram cloud into bins with a bin width setting to 2δ. Lastly,

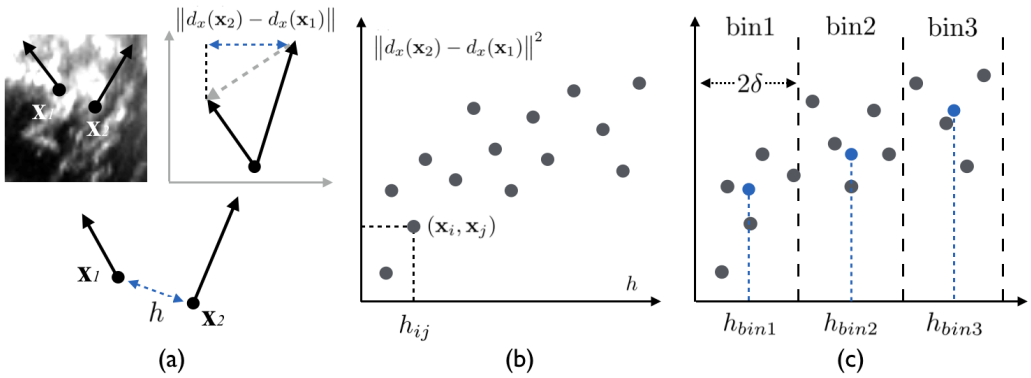


Figure 2.3: (a) $\|d_x(\mathbf{x}_2) - d_x(\mathbf{x}_1)\|$ and h; (b) Empirical variogram cloud; (c) Variogram cloud divided into bins with their means marked as blue.

the mean of each bin is calculated and further plotted with the mean distance of that bin to form an empirical variogram. Figure 2.4(a) shows an empirical variogram of a real US image pair that has 71 corresponding landmarks.

In order to obtain the data-driven GP kernel function, we further fit a smooth curve, generated by pre-defined kernel functions, to the empirical variogram. As shown in Figure 2.4(b), a fitted curve is commonly described by the following characteristics:

Nugget The non-zero value at $h = 0$.

Sill The value at which the curve reaches its maximum.

Range The value of distance h where the sill is reached.

Fitting a curve to an empirical variogram is implemented in most geostatistics software. A popular choice is choosing several models that appear to have the right shape and use the one with smallest weighted squared error (**?**). In this study, we only test Gaussian curves

$$\gamma(h) = c_0 + c\{1 - exp(-\frac{h^2}{a})\}. \tag{2.4}$$

Here, c_0 is the nugget, $c = \text{Sill} - c_0$ and a is the model parameter. Once the fitted curve is found, we can obtain a from the equation (5) and use it as the Gaussian kernel scale in the GP interpolation.

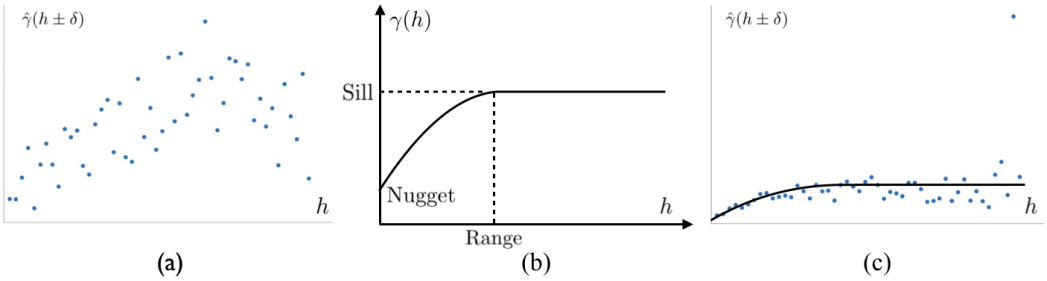


Figure 2.4: (a) X-axis empirical variogram of a US images pair;(b) Sill, range and nugget; (c) Fitting a continuous model to an empirical variogram.

2.3.3.2 Discrete Grid Search

The variogram scheme often requires many landmarks to work well (**?**). For US pairs that have fewer landmarks, we choose predefined Gaussian kernels, and use cross validation to determine the scale parameter in a discrete grid search fashion (C.E. and Williams, 2006).

2.3.4 Active Registration

Automatic approaches may have difficulty in the preop-to-intraop image registration, especially for areas near the tumor resection site. Another advantage of the GP framework is the possibility of incorporating user expertise to further improve the registration result.

From Equation (1), we can also compute the covariance matrix of the posterior Gaussian $p(\mathbf{d}_* \mid \mathbf{X}_*, \mathbf{X}, \mathbf{d})$ as

$$\Sigma_* = \mathbf{K}_{**} - \mathbf{K}_*^T \mathbf{K}^{-1} \mathbf{K}_*. \tag{2.5}$$

Entries on the diagonal of Σ_* are the marginal variances of predicted values. They can be used as an uncertainty measure to indicates the confidence in the estimated transformation parameters.

If users are not satisfied by the FBR alignment result, they could manually, guided by the image context and visualization of registration uncertainty, add new corresponding pairs of key-points to drive the GP towards better results.

2.4 Experiments

The experimental dataset consists of 6 sets 3D preUS and postUS image pairs. The US signals were acquired on a BK Ultrasound 3000 system that is directly connected to the Brainlab VectorVision Sky neuronavigaton system during surgery. Signals were further reconstructed as 3D volume using the PLUS (Lasso et al., 2014) library in 3D Slicer (Fedorov et al., 2012).

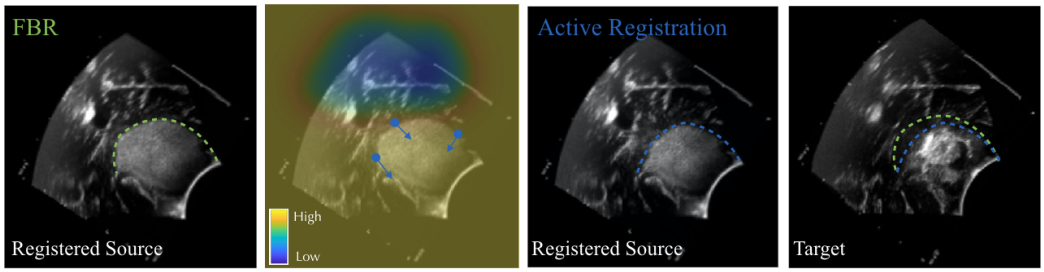
Table 2.1: Registration evaluation results (in mm)

	Before Reg.	Affine	Thin-plate	Variograms	GaussianK
P1	5.56±1.05	2.99±1.21	1.79±0.70	2.11±0.74	1.75±0.68
P2	3.35±1.22	2.08±1.13	2.06±1.18	2.06±1.12	1.97±1.05
P3	2.48±1.56	1.93±1.75	1.25±1.95	n/a	1.23±1.77
P4	4.40±1.79	3.06±2.35	1.45±1.99	n/a	1.42±2.04
P5	2.91±1.33	1.86±1.24	1.29±1.17	n/a	1.33±1.40
P6	3.29±1.09	2.12±1.16	2.02±1.21	2.05±1.40	1.96±1.38

We used the mean Euclidean distance between the predicted and ground truth of key-points' coordinates, measured in mm, for the registration evaluation. We compared affine, thin-plate kernel FBR, variograms FBR, and Gaussian kernel FBR during the assessment. For US pairs with fewer than 50 landmarks, we used leave-one-out cross-validation; otherwise, we used 5-fold cross-validation, i.e., we arbitrarily partitioned all matched key-points into five (almost) equal-sized groups. A single group is retained as the validation data to evaluate the registration accuracy, and the remaining four groups were used to interpolate the deformation field. This process was repeated five times, with each group used exactly once as for the validation. Notice that the cross-validation strategy is not an ideal evaluation. In future works, we can improve the evaluation by using synthetic data, or public datasets (Mercier et al., 2012; Xiao et al., 2017) with known ground truth correspondences.

This could be improved by using manual landmarks in public datasets, such as RESECT (Xiao et al., 2017) and BITE (Mercier et al., 2012).

All of the compared methods were computed in less than 10 minutes.



(a) (b) (c) (d)

Figure 2.5: (a) FBR result of the preUS with a tumor boundary outlined in green; (b) Overlaying the visualization of uncertainty on the preUS image. A characteristic of GP is that voxels near landmarks tend to have smaller uncertainty. In this example, all landmarks happen to be located near the large sulcus, hence the incertitude looks high everywhere else except around the sulcus. (c) Active registration result of the preUS with a tumor boundary outlined in blue; (d) Overlaying the green and blue tumor boundary on the target image.

The pre-defined Gaussian kernel with a discrete grid search generally yields a better result than the variogram scheme. This is reasonable as the machine learning approach stresses the prediction performance, while the geostatistical variogram favors the interpretability of the model. For some cases that only have a few matched key-points, the variogram-based method was unsuccessful.

In addition, we have performed preliminary tests on active registration as shown in Figure 2.5, which illustrate the use of a colour map of registration uncertainty to guide the manual placement of 3 additional landmarks to improve the registration. By visual inspection, we can see the alignment of tumor boundary substantially improved.

2.5 Discussion

One key point of our framework is the “active registration”idea that aims to overcome the limitation of automatic image registration. Human and machines have complementary abilities; we believe that the element of simple user interaction should be added to the pipeline for some challenging medical imaging applications. Although the proposed method is designed for brain shift compensation, it is also applicable to other navigation systems that require

tracking of tissue deformation or even to difficult natural image registration problems. The performance of FBR is highly correlated with the quality of feature matching. In future works, we plan to test different matching algorithms (Jian and Vemuri, 2010), and also perform more validation with public datasets.

Chapter 3

On the Applicability of Registration Uncertainty

Estimating the uncertainty in (probabilistic) image registration enables, e.g., surgeons to assess the operative risk based on the trustworthiness of the registered image data. If surgeons receive inaccurately calculated registration uncertainty and misplace unwarranted confidence in the alignment solutions, severe consequences may result. For probabilistic image registration (PIR), the predominant way to quantify the registration uncertainty is using summary statistics of the distribution of transformation parameters. The majority of existing research focuses on trying out different summary statistics as well as means to exploit them. Distinctively, in this chapter, we study two rarely examined topics: (1) whether those summary statistics of the transformation distribution most informatively represent the registration uncertainty; (2) Does utilizing the registration uncertainty always be beneficial. We show that there are two types of uncertainties: the transformation uncertainty, U_{t}, and label uncertainty U_{l}. The conventional way of using U_{t} to quantify U_{l} is inappropriate and can be misleading. By a real data experiment, we also share a potentially critical finding that making use of the registration uncertainty may not always be an improvement. This chapter is based on the publication of (Luo et al., 2019a).

3.1 Introduction

Non-rigid image registration is the foundation for many image-guided medical tasks (Maintz and Viergever, 1998; Sotiras et al., 2013). However, given the current state of the registration technology and the difficulty of the problem, an uncertainty measure that highlights locations where the algorithm had difficulty finding a proper alignment can be very helpful. Among the approaches that characterize the uncertainty of non-rigid image registration, the most popular, or perhaps the most successful framework is the probabilistic image registration (PIR) (Cobzas and Sen, 2011; Dalca et al., 2018; Folgoc et al., 2017; Heinrich et al., 2016; Janoos et al., 2012; Lotfi et al., 2013; Luo et al., 2018b; Popuri et al., 2013; Risholm et al., 2013; Simpson et al., 2015, 2012; Wang et al., 2018; Wassermann et al., 2014; Yang and Niethammer, 2015; Zhang et al., 2013).

In contrast to traditional “point-estimate" image registration approaches that report a unique set of transformation parameters that best align two images, PIR models transformation parameters as random variables and estimates distributions over them. The mode of the distribution is then chosen as the most likely value of that transformation parameter. PIR has the advantage that the registration uncertainty can be naturally obtained from the distribution of transformation parameters. PIR methods can be broadly categorized into discrete probabilistic registration (DPR) (Cobzas and Sen, 2011; Heinrich et al., 2016; Lotfi et al., 2013; Popuri et al., 2013) and continuous probabilistic registration (CPR) (Dalca et al., 2018; Folgoc et al., 2017; Janoos et al., 2012; Luo et al., 2018b; Risholm et al., 2013; Sedghi et al., 2018; Simpson et al., 2015, 2012; Wang et al., 2018; Wassermann et al., 2014; Yang and Niethammer, 2015; Zhang et al., 2013)

3.1.1 Related Work

Registration uncertainty is a measure of confidence in image alignment solutions. In the PIR literature, the predominant way to quantify the registration uncertainty is using summary statistics of the transformation distribution. Applications of various summary statistics have been proposed in previous research: the Shannon entropy and its variants of the categorical transformation

distribution were used to measure the registration uncertainty of DPR (Lotfi et al., 2013); the variance (Folgoc et al., 2017; Simpson et al., 2012; Yang and Niethammer, 2015), standard deviation (Simpson et al., 2015), inter-quartile range (Risholm et al., 2013, 2010) and the covariance Frobenius norm (Wassermann et al., 2014) of the transformation distribution were used to quantify the registration uncertainty of CPR. In order to visually assess the registration uncertainty, each of these summary statistics was either mapped to a color scheme, or an object overlaid on the registered image. By inspecting the color of voxels or the geometry of that object, end users can infer the registration uncertainty, which suggests the confidence they can place in the registration result. Utilizing the registration uncertainty is presumably an advantage of PIR (Risholm et al., 2010; Risholm and Wells, 2011; Simpson et al., 2013), to date, the majority of existing research focuses on trying out different summary statistics and means to exploit the registration uncertainty.

3.1.2 Clinical Motivation

In image-guided neurosurgery, surgeons need to correctly understand the registration uncertainty so as to make better informed decisions, e.g., If the surgeon observes a large registration error at location A and small error at location B, without knowledge of registration uncertainty, s/he would most likely assume a large error everywhere and thus entirely ignore the registration. With an accurate knowledge of uncertainty, once the surgeon knows that A lies in an area of high uncertainty while B lies in an area of low uncertainty, s/he would have greater confidence in the registration at B and other locations of low uncertainty. If surgeons are influenced by inaccurate amount of registration uncertainty and place unwarranted confidence in the alignment solutions, severe consequences may result (Risholm et al., 2013, 2010; Risholm and Wells, 2011).

3.1.3 Contributions

The majority of research takes the registration uncertainty for granted. In this chapter, we investigate two rarely examined topics: (1) whether summary statistics of the transformation distribution most informatively reflect the

registration uncertainty; (2) Does utilizing the registration uncertainty always be beneficial. In Section 2, we identify and discuss two types of uncertainties: the transformation uncertainty U_{t} and label uncertainty U_{l}. By concrete examples, we show that the conventional way of using U_{t} to quantify U_{l} is inappropriate and can be misleading. In Section 3, by a real data example, we share a potentially critical finding that making use of the registration uncertainty may not always be an improvement. Finally, we summarize in Section 4. It should be noted that registration uncertainty is not equal to registration accuracy. There are excellent works which study standards of registration evaluation (Fitzpatrick, 2009; Min et al., 2020; Rohlfing, 2012). However, here we focus on the relation among different types of registration uncertainty.

3.2 The Ambiguity of Registration Uncertainty

For illustration purpose, we use DPR in all examples.

3.2.1 The DPR Set Up

In the DPR setting, let I_{t} and I_{s} respectively be the target and source images $I_{\mathrm{t}}, I_{\mathrm{s}} : \Omega_I \to \mathbb{R}, \Omega_I \subset \mathbb{R}^d, d = 2 \text{ or } 3$. The algorithm discretizes the transformation space into a set of K displacement vectors, $\mathcal{D} = \{\mathbf{d}_k\}_{k=1}^K, \mathbf{d}_k \in \mathbb{R}^d$. These displacement vectors radiate from voxels on I_{t} and point to their candidate transformation locations on I_{s} (Sotiras et al., 2013). For every voxel v_i, the algorithm computes a unity-sum probabilistic vector $\mathcal{P}(v_i) = \{P_k(v_i)\}_{k=1}^K$ as the transformation distribution. $P_k(v_i)$ is the probability of displacement vector $\mathbf{d}_k$. In a standard DPR, the algorithm takes a displacement vector that has the highest probability in $\mathcal{P}(v_i)$ as the most likely transformation $\mathbf{d}_m$.

Conventionally, the uncertainty of registered v_i is quantified by the Shannon entropy of $\mathcal{P}(v_i)$ (Lotfi et al., 2013). Since the algorithm takes $\mathbf{d}_m$ as its "point-estimate", the entropy provides a measure of the extent of dispersion from $\mathbf{d}_m$ of the rest of displacement vectors in $\mathcal{D}$. If other displacement vectors are all as equally likely to occur as $\mathbf{d}_m$, then the entropy is maximal, which indicates that it is completely uncertain which displacement vector should be chosen as

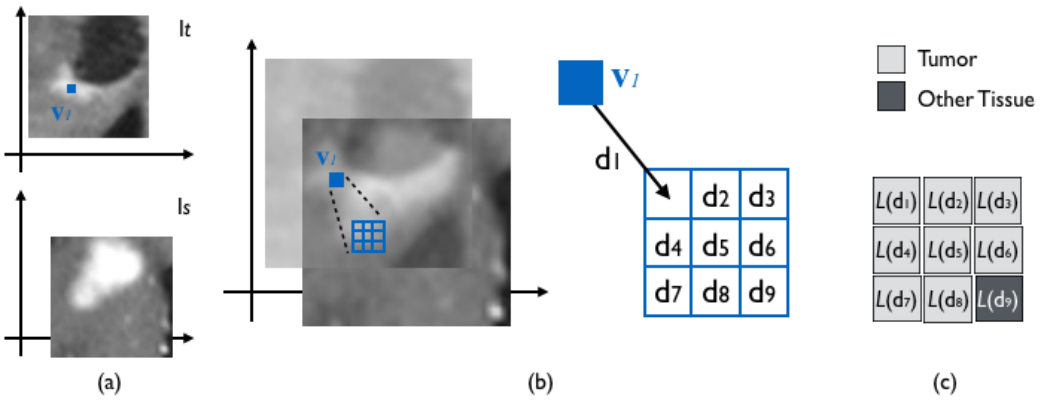


Figure 3.1: The target image I_t and souce image I_s ; (b) The discretized transformation space $\mathcal{D}$; (c) The corresponding tissue label $L(\mathbf{d}_k)$ for $\mathcal{D}$. .

the most likely transformation. When the probability of $\mathbf{d}_m$ is much higher than that of other displacement vectors, the entropy decreases, and there is greater certainty that $\mathbf{d}_m$ is the correct choice.

For example, $\mathcal{P}(v_l) = [0.25, 0.25, 0.25, 0.25]$ and $\mathcal{P}(v_r) = [0.1, 0.7, 0.1, 0.1]$ are two discrete transformation distributions. $\mathcal{P}(v_l)$ is uniformly distributed, and its entropy is $E(\mathcal{P}(v_l)) = 2$. $\mathcal{P}(v_r)$ has an obvious peak, and its entropy is $E(\mathcal{P}(v_r)) \approx 1.36$, which is lower than $E(\mathcal{P}(v_l))$. For a registered voxel, the entropy of its transformation distribution is usually mapped to a color scheme, clinicians can infer the level of confidence of the registration result by the color of the voxel.

3.2.2 Transformation Uncertainty and Label Uncertainty

In the context of neurosurgery, the goal of image registration is frequently to map the pre-operatively labeled tumor, and/or other tissue, onto the intra-operative patient space for resection. Since registration uncertainty is strongly linked to the goal of registration, here it should also reflect the confidence in the registered labels. However, does the conventional uncertainty measure of DPR, which is the entropy of transformation distribution, truly give insight into the trustworthiness of registered labels?

In a hypothetical DIR example, I_t and I_s in Figure 3.1(a) are the intra-operative target and pre-operative source images, respectively. Voxel v_1 on I_t is

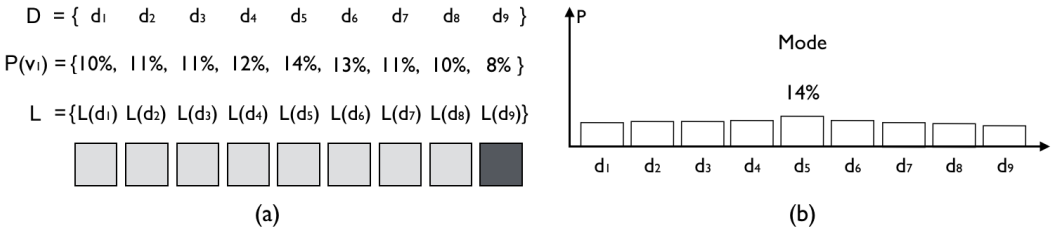


Figure 3.2: (a) $\mathcal{P}(v_1)$ and corresponding labels; (b) The bar chart of $\mathcal{P}(v_1)$.

the voxel we want to register. In Figure 3.1(b), we can see that the discretized transformation space $\mathcal{D} = \{\mathbf{d}_k\}_{k=1}^{9}$ is a set of nine displacement vectors. Each displacement vector is linked to a candidate corresponding voxel of v_1. The labels $L(\mathbf{d}_k)$ are for voxels associated with $\mathbf{d}_k$. In this example, there are labels for the tumor and other tissue, as shown in Figure 3.1(c).

Figure 3.2 shows a transformation distribution $\mathcal{P}(v_1) = \{P_k(v_1)\}_{k=1}^{9}$ and its bar chart. We observe that $P_5(v_1)$ has the highest probability in $\mathcal{P}(v_1)$; therefore, $\mathbf{d}_5$'s corresponding label, $L(\mathbf{d_5})$ = Tumor, will be assigned to the registered v_1.

Although $\mathcal{P}(v_1)$ has its mode at $P_5(v_1)$, the entire distribution is more or less uniformly distributed. The entropy of $\mathcal{P}(v_1)$, $E(\mathcal{P}(v_1)) \approx 3.15$, is close to the maximum. Therefore, the conventional uncertainty measure will suggest that the registration uncertainty of v_1 is very high and highlight it with a bright color. Upon noticing the high degree of uncertainty in registered v_1, surgeons would place less confidence in its tumor label and make surgical plans accordingly.

On the other hand, let us take into account the label $L(\mathbf{d}_k)$ associated with each $\mathbf{d}_k$ and form a label distribution. As shown in Figure 3.3(a), even if $\mathbf{d}_1, \ldots, \mathbf{d}_8$ are different displacement vectors, they correspond to the same label as the most likely displacement vector $\mathbf{d}_5$. If we accumulate the probability for all labels in $\mathcal{L}$, it is clear that “tumor" is the dominant one. Interestingly, despite being suggestive of having high registration uncertainty using the conventional uncertainty measure, the label distribution in Figure 3.3(b) indicates that it is quite trustworthy to assign a tumor label to the registered v_1. In addition, the entropy of the label distribution is as low as 0.4, which also differs from the high entropy value computed from the transformation distribution.

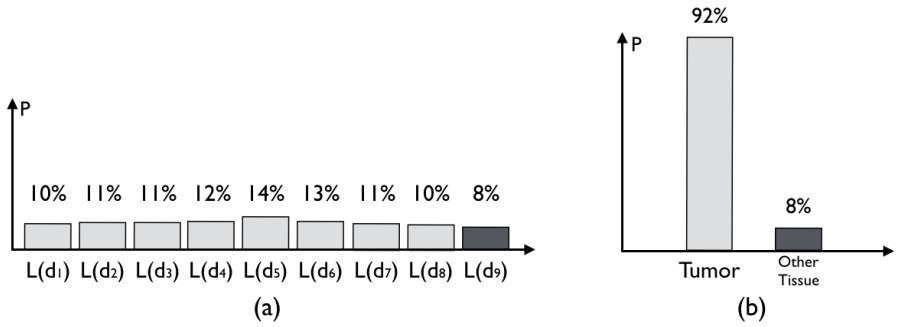


Figure 3.3: (a) Bar chart of the transformation distribution $\mathcal{P}(v_1)$ taking into account $L(\mathbf{d_k})$; (b) The label distribution of the registered v_1.

In the example above, there appear to be two kinds of uncertainty. We name the uncertainty computed from the transformation distribution as the transformation uncertainty U_{t}, and the uncertainty relating to the goal of registration as label uncertainty U_{l}. Examples of U_{l} can be uncertainty in a categorical classification, or uncertainty in the intensity value of registered voxels.

In the PIR literature, the definition of registration uncertainty is ambiguous, because researchers do not differentiate U_{t} from U_{l}, and perhaps subconsciously use U_{t} to quantify U_{l}. The previous counter-intuitive example demonstrates that high U_{t} does not guarantee high U_{l}. In fact, the value of U_{t} can barely guarantee any useful information at all about the U_{l}.

More precisely, for point-estimate image registration, let Ω_T be the set of all estimated transformation, and Ω_L be the set of all possible corresponding labels (categorical labels from semantic segmentation or intensity values). The algorithm assigns a transformation $t \in \Omega_T$ to a voxel. By a non-linear function $f_{point} : \Omega_T \to \Omega_L$, the voxel will have its label $l \in \Omega_L$ as:

$$l = f_{point}(t). \tag{3.1}$$

In this case, the function f_{point} is surjective, and t always has a unique corresponding l. However, in the PIR setting, the voxel transformation becomes a random variable T. The corresponding label L is a function of T:

$$L = f_{prob}(T). \tag{3.2}$$

therefore, it is also a random variable. Even if T and L are intuitively correlated, given different image context, there is no guaranteed analytical way to compute the uncertainty propagation from T to L. Thus it's inappropriate to measure the uncertainty of L, by the summary statistics of T.

In the registration community, researchers routinely distinguish between intensity match and e.g. DICE scores. It also makes sense to distinguish the transformation uncertainty and label uncertainty. In practice, U_{t} and U_{l} around certain areas, i.e., the tumor boundary, is quite dissimilar. Propagating U_{t} to the surgeon, as if it is the U_{l} can mislead them to place unwarranted confidence in the alignment solution and result in severe consequences.

3.2.3 Real Data Examples

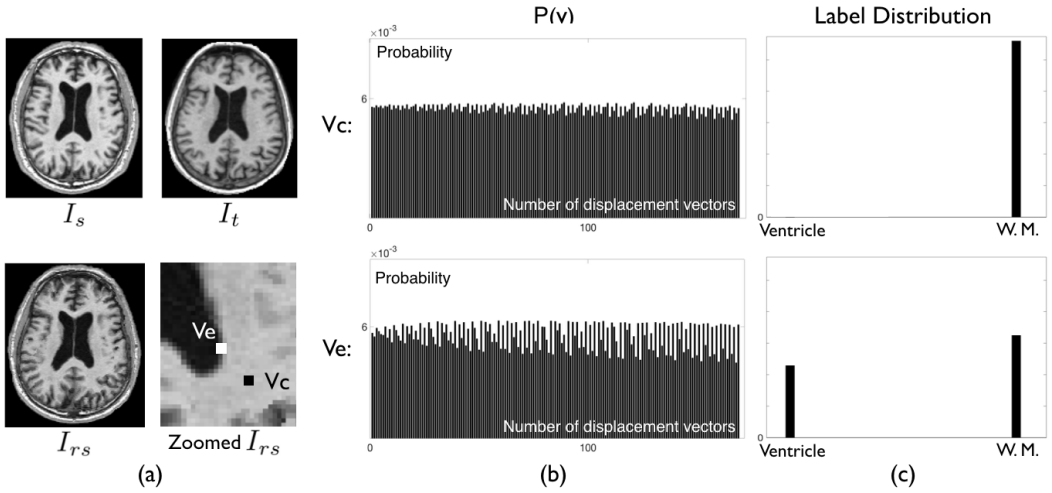


Figure 3.4: (a) Input and result of the CUMC12 data example, v_c and v_e are two voxels of interest on the registered source image; (b) The transformation distribution of v_c and v_e in the DIR; (c) Label distributions of registered v_c and v_e.

As shown in Figure 3.4, I_{t} and I_{s} are two brain MRI images arbitrarily chosen from the CUMC12 dataset. Subsequent to performing a DIR, we obtained the registered source image I_{rs}. The goal of this registration was to determine the categorical label, whether it is a ventricle or a white matter, for registered voxels of interest v_c and v_e. The transformation distribution of v_c is more uniformly distributed than that of v_e. Therefore, conventional entropy-based methods

will report v_c as having higher registration uncertainty than v_e. However, as we form a label distribution in Figure 3.4(c), it is clear that v_e, despite having a lower U_{t}, is assigned a label that is more uncertain. Examples that demonstrate the dissimilarity between U_{t} and U_{l} can be frequently found in the registration of various kind of images.

3.3 Credibility of Label Distribution

Utilizing the registration uncertainty, in particular the full label distribution, to benefit registration-based tasks is presumably an advantage of PIR. Many research of registration uncertainty reported positively over its impact in applications (Risholm et al., 2010; Risholm and Wells, 2011; Simpson et al., 2013). However, to the best of our knowledge, there does not exist any validation study about whether we should use the registration uncertainty. In this section, we design an experiment to explore whether utilizing the registration uncertainty always results in an improvement.

In PIR, the registered voxel has the corresponding label of the most likely transformation $L(\mathbf{d}_m)$, upon which the registration evaluation is also based (Cobzas and Sen, 2011; Folgoc et al., 2017; Heinrich et al., 2016; Janoos et al., 2012; Lotfi et al., 2013; Luo et al., 2018b; Popuri et al., 2013; Risholm et al., 2013; Simpson et al., 2015, 2012; Wang et al., 2018; Wassermann et al., 2014; Yang and Niethammer, 2015; Zhang et al., 2013). Likewise, we can derive the most likely label L_m from the full label distribution. If utilizing the registration uncertainty, like reported in previous research, always be beneficial, then L_m should be always better than $L(\mathbf{d}_m)$.

In the following pilot experiment: an MRI image is arbitrarily chosen from the BRATS dataset (Menze et al., 2015) and synthetically deformed.Then we registered the original data with the deformed data using DIR. By doing so, we know the ground truth intensity for every registered voxel so that we can compare whether it is $L(\mathbf{d}_m)$ or L_m closer to the ground truth.

Here we are interested in the intensity label distributions of four registered voxels $v_{\mathrm{b}}, v_{\mathrm{c}}, v_{\mathrm{d}}$ and v_{e}, shown in Figure 3.5(b), (c), (d), and (e) respectively. In Figure 3.5, the red circle indicates the most likely intensity label L_m given by the

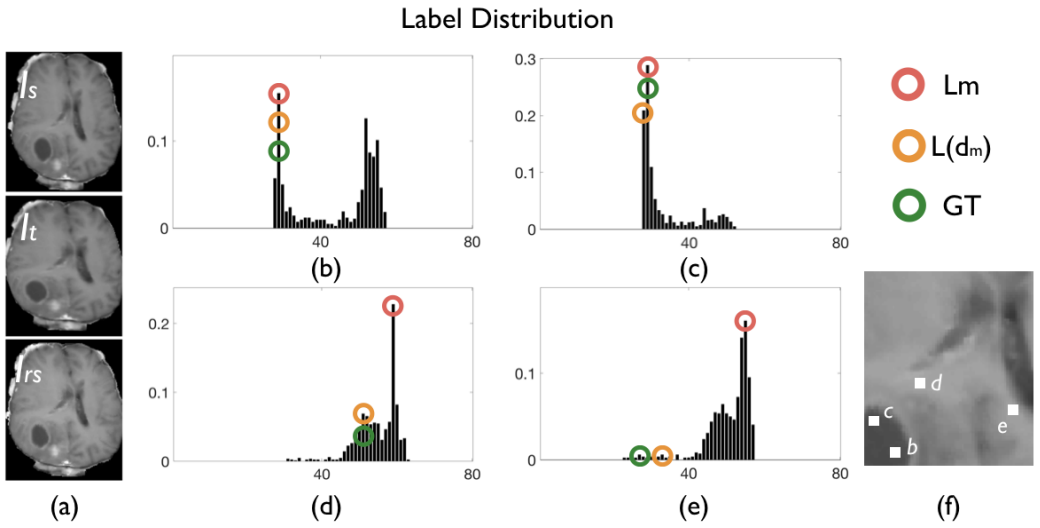


Figure 3.5: (a) Input and result of the registration example; (b,c,d,e) Intensity label distributions of voxels v_b,v_c,v_d and v_e; (f) Approximate locations of tested voxels.

full transformation distribution, the orange circle indicates the corresponding intensity label of the transformation mode $L(\mathbf{d}_m)$, and the green circle is the Ground Truth (GT). We observe that for v_b, L_m and $L(\mathbf{d}_m)$ are both equal to the GT. On the other hand, L_m and $I(\mathbf{d_m})$ for $v_\mathrm{c}, v_\mathrm{d}$ and v_e, are not the same. As seen in Figure 3.5(c), the L_m of the registered v_c is equal to the GT intensity, and is more accurate than $I(\mathbf{d_m})$. Yet, unexpectedly, for v_d and v_e, their $I(\mathbf{d}_m)$ is closer to the GT than their L_m. Voxels such as v_d and v_e were found frequently in our experiments using other real data. This surprising result indicates that utilizing the full transformation distribution can actually give a poorer/less accurate estimation than using the transformation mode alone.

Researchers have attempted to present the visualized full label distribution of functional areas in fMRI to neurosurgeons (Risholm et al., 2010). However, based on the above finding, if L_m can give poorer estimation, the full label distribution might also have questionable credibility. Conveying such false information to surgeons would certainly be detrimental to the outcome of surgery.

It is noteworthy that in PIR, the estimation of T and L is influenced by the choice of hyper parameters, priors, and image context. Other PIR approaches

can yield different findings. Nevertheless, studying the credibility of the label distribution before using it in practice warrants increased investigation.

3.4 Discussion

The majority of research takes the registration uncertainty for granted. We summarize current approaches of quantifying registration uncertainty and point out some fundamental problems which would make researchers rethink, or even re-work approaches for quantifying and applying registration uncertainty.

At this stage, even the uncertainty is a useful addition to the registration result, we recommend treating it with caution: (1) It is advised to distinguish U_{t} and U_{l} in applications. Instead of using the unified term "registration uncertainty", i.e., we can use U_{t} to indicate the confidence for a predicted instrument location in neurosurgery; (2) Since the credibility of label distribution is unclear, we should avoid using U_{l} in clinical settings and put further effort in studying the implication of PIR results. We believe that this chapter will serve as a foundation and draw more attention to this topic.

Chapter 4

Are Registration Uncertainty and Error Monotonically Associated?

In image-guided neurosurgery, current commercial systems usually provide only rigid registration, partly because it is harder to predict, validate and understand non-rigid registration error. For instance, when surgeons see a discrepancy in aligned image features, they may not be able to distinguish between registration error and actual tissue deformation caused by tumor resection. In this case, the spatial distribution of registration error could help them make more informed decisions, e.g., ignoring the registration where the estimated error is high. However, error estimates are difficult to acquire. Probabilistic image registration (PIR) methods provide measures of registration uncertainty, which could be a surrogate for assessing the registration error. It is intuitive and believed by many clinicians that high uncertainty indicates a large error. However, the monotonic association between uncertainty and error has not been examined in image registration literature. In this pilot study, we attempt to address this fundamental problem by looking at one PIR method, the Gaussian process (GP) registration. We systematically investigate the relation between GP uncertainty and error based on clinical data and show empirically that there is a weak-to-moderate positive monotonic correlation between point-wise GP registration uncertainty and non-rigid registration error. This chapter is based on the publication of (Luo et al., 2020a).

4.1 Introduction

In image-guided neurosurgery (IGN), surgical procedures are often planned based on the preoperative (p-) magnetic resonance imaging (MRI). During surgery, clinicians may acquire intraoperative (i-) MRI and/or Ultrasound (US). Image registration can be used (Maintz and Viergever, 1998; Sotiras et al., 2013) to map the p-MRI to the intraoperative coordinate space to help surgeons locate structures or boundaries of interest during surgery (e.g., tumor margins or nearby blood vessels to be avoided) and facilitate more complete tumor resection (Gerard et al., 2017; Luo et al., 2019b; Morin et al., 2017).

Even though the brain clearly undergoes non-linear deformation during surgery, rigid registration is still the standard for clinical practice (Rivaz and Collins, 2015). Although non-rigid registration has long been a goal for IGNs, this goal is hampered because non-rigid registration error is less predictable and harder to validate than rigid registration error. In practice, if surgeons see a discrepancy between two aligned image features, they may not be able to tell if the misalignment is caused by a registration error or an actual tissue deformation caused by tumor resection. In this case, providing surgeons with a spatial distribution of the expected registration error could help them make more informed decisions, e.g., ignoring the registration where the expected error is high. However, determining this spatial distribution is difficult since:

1. Most methods that estimate registration error, such as bootstrapping (Kybic, 2010; Shams et al., 2017), perturbed input (Datteri and Dawant, 2012; Hub and Karger, 2013; Hub et al., 2009), stereo confidence (Saygill et al., 2016) and supervised learning (Saygill, 2018; Sokooti et al., 2016, 2019), require multiple runs of a non-rigid registration algorithm, thus they are too time-consuming to be practical for IGNs where feedback is required within a few minutes of intraoperative image acquisition.

2. More importantly, existing methods estimate the error by detecting misaligned image features (Saygill, 2018; Saygill et al., 2016; Shams et al., 2017; Sokooti et al., 2019). These methods fail in IGNs because tumor resection and retraction significantly alter the brain, particularly at the

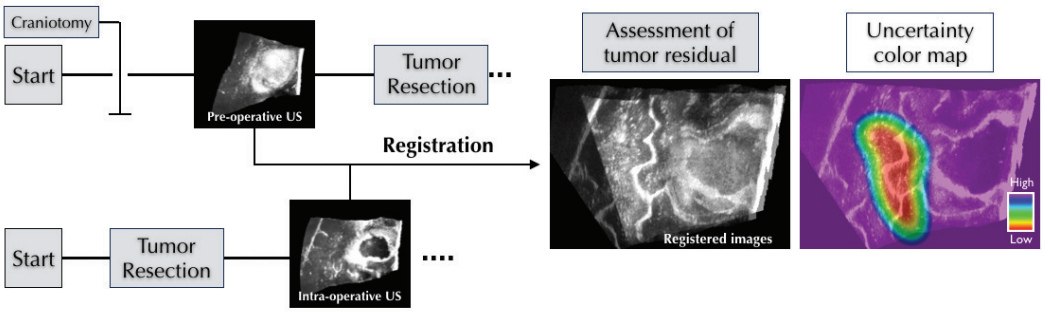


Figure 4.1: An example to illustrate the usefulness of registration uncertainty in IGNs.

tumor margin where precision is most needed. Thus finding consistent image features near the tumor margin may be difficult.

An alternative for directly estimating the registration error is to use registration uncertainty as a surrogate. Registration uncertainty is a measure of confidence in the predicted registration and typically estimated by probabilistic image registration (PIR)(Agn and Van Leeput, 2019; Bayer et al., 2017a; Dalca et al., 2018; Folgoc et al., 2017; Glocker et al., 2009; Heinrich et al., 2016; Lotfi et al., 2013; Luo et al., 2018b; Popuri et al., 2013; Risholm et al., 2013; Sedghi et al., 2018; Simpson et al., 2015; Wang et al., 2018; Wassermann et al., 2014; Yang and Niethammer, 2015). In IGN, utilizing registration uncertainty can be helpful. As shown in Figure 4.1, surgeons can inspect the residual tumor after registering the p-US image to the i-US image and decide whether to continue the resection or end the operation. An uncertainty color map overlaid on top of the registered images, where red indicated regions of low uncertainty, can be used by surgeons to dismiss clear misregistration regions where uncertainty is high and have more confidence in the registration where uncertainty is low (e.g., red regions).

In this example, surgeons might have higher confidence inside red regions because they assume that areas with low uncertainty also tend to have a low error. However, this assumption is only valid if the registration uncertainty and error have a positive monotonic association. While this notion is intuitive and believed by many clinicians, to the best of our knowledge, it has not been examined in the PIR literature.

"Are registration uncertainty and error monotonically associated?" is a crucial question that impacts the applicability of registration uncertainty. In this pilot study, we attempt to address this question by looking at a promising PIR method, Gaussian process (GP) registration (Bayer et al., 2017a; Luo et al., 2018b; Wassermann et al., 2014). We systematically investigate the GP uncertainty and error using point-wise posterior predictive checking and a patch-wise correlation test. We note that the registration uncertainty can be categorized as transformation uncertainty or label uncertainty (Luo et al., 2019a). Since the applicability of label uncertainty is still in question, this chapter will focus solely on the transformation uncertainty when it refers to 'registration uncertainty'.

4.2 Methods

In this section, we briefly review GP registration uncertainty. Then we introduce Spearman's correlation coefficient and provide details about our point-wise and patch-wise experiments.

4.2.1 Review of the GP Registration Uncertainty

The stochastic GP registration approach has shown promising results in IGNs (Bayer et al., 2017a; Luo et al., 2018b; Wassermann et al., 2014). As shown in Figure 4.2, a key step in the GP registration is to estimate N_* unknown displacement vectors $\mathbf{D}_*$ from N known ones $\mathbf{D}$ that were derived from automatic feature extraction and matching.

Let $\mathbf{x}$ be the grid coordinate and $\mathbf{d}(\mathbf{x}) = [d_{\mathrm{x}}, d_{\mathrm{y}}, d_{\mathrm{z}}]$ be the associated displacement vector. For $d(\mathbf{x})$ being one of d_{x}, d_{y}, and d_{z}, it is modeled as a joint Gaussian distribution $d(\mathbf{x}) \sim \mathrm{GP}(m(\mathbf{x}), \mathrm{k}(\mathbf{x}, \mathbf{x}'))$ with mean function $m(\mathbf{x}) = 0$ and covariance function $\mathrm{k}(\mathbf{x}, \mathbf{x}')$. Thus $\mathbf{D}$ and $\mathbf{D}_*$ have the following relationship:

$$\begin{bmatrix} \mathbf{D} \\ \mathbf{D}_* \end{bmatrix} \sim N\left(\mathbf{m}(\mathbf{x}), \begin{bmatrix} \mathbf{K} & \mathbf{K}_* \\ \mathbf{K}_*^T & \mathbf{K}_{**}. \end{bmatrix}\right). \tag{4.1}$$

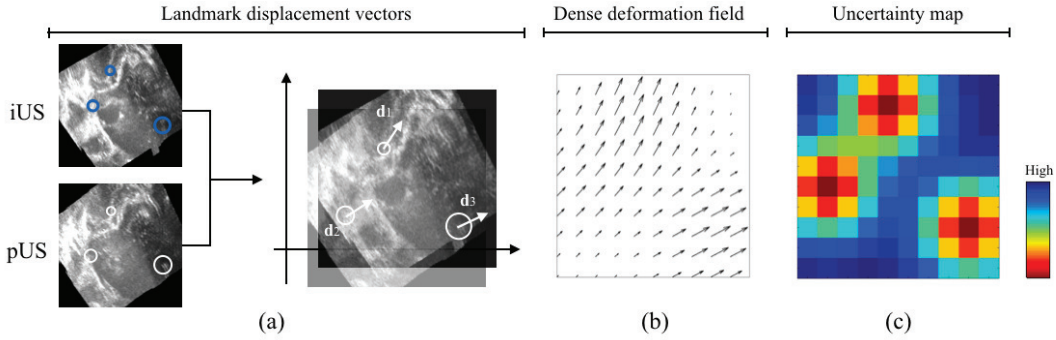


Figure 4.2: (a) 3 displacement vectors; (b) A 10×10 interpolated dense deformation field; (b) A visualization of registration uncertainty.

In Eq.(1), $\mathbf{K} = \mathrm{k}(\mathbf{X}, \mathbf{X}) \in \mathbb{R}^{N \times N}$ and $\mathbf{K}_{**} = \mathrm{k}(\mathbf{X}_*, \mathbf{X}_*) \in \mathbb{R}^{N_* \times N_*}$ are intra-covariance matrices of $\mathbf{d}$ and $\mathbf{d}_*$ respectively. $\mathbf{K}_* = \mathrm{k}(\mathbf{X}, \mathbf{X}_*) \in \mathbb{R}^{N \times N_*}$ is the inter-covariance matrix. The interpolated displacement vector values can be estimated from the mean μ_* of the posterior distribution of $p(\mathbf{D}_* \mid \mathbf{X}_*, \mathbf{X}, \mathbf{D})$:

$$\mu_* = \mathbf{K}_*^T \mathbf{K}^{-1} \mathbf{D}. \tag{4.2}$$

From Eq.(1), the posterior covariance matrix can also be derived as

$$\mathbf{\Sigma}_* = \mathbf{K}_{**} - \mathbf{K}_*^T \mathbf{K}^{-1} \mathbf{K}_*. \tag{4.3}$$

Diagonal entries of $\mathbf{\Sigma}_*$ are the marginal transformation variances, and they can be used as the GP registration uncertainty. In this study, we choose the same kernel $\mathrm{k}(\mathbf{x}, \mathbf{x}') = \exp(-\frac{x^2}{a})$ for all three displacement components.

Figure 4.2(b) shows a 10×10 dense deformation field interpolated from three landmark displacement vectors. Each voxel is associated with an estimated displacement vector and uncertainty value. Figure 4.2(c) is an uncertainty color map for the displacement field.

4.2.2 Spearman's Rank Correlation Coefficient

Spearman's correlation coefficient, often denoted by ρ_s, is a non-parametric measure of statistical dependence between the rankings of two variables. It assesses how well their relationship can be described using a monotonic function (Corder and Foreman, 2014).

In this study we prefer ρ_{s} over Pearson's correlation ρ_{p} for the following reasons:

1. ρ_{p} measures the strength of a linear relationship. To be clinically useful, registration uncertainty does not have to be linearly correlated with the error. In this sense, we prefer ρ_{s} which measures a less "restrictive" monotonic relationship;

2. Since ρ_{s} limits the influence of outliers to the value of its rank, it is less sensitive than ρ_{p} to strong outliers that lie in the tails of the distribution (Corder and Foreman, 2014).

Assume there are M test points, $\mathrm{u}(i)$ and $\epsilon(i)$, which represent the uncertainty and error for point i respectively. Let U and E denote discrete random variables with values $\{\mathrm{u}(1), \mathrm{u}(2), ..., \mathrm{u}(M)\}$ and $\{\epsilon(1), \epsilon(2), ..., \epsilon(M)\}$. To measure ρ_{s}, we have to convert U and E to descending rank vectors rU and rE, i.e., the rank vector for $[0.2, 1.2, 0.9, 0.5, 0.1]$ would be $[2, 5, 4, 3, 1]$. Then ρ_{s} can be estimated as

$$\rho_{\mathrm{s}} = \frac{\mathrm{cov}(rU, rE)}{\sigma_{rU}\sigma_{rE}}, \tag{4.4}$$

where cov is the covariance, σ's are the standard deviations. Noticing that ρ_{s} is by design constrained as $-1 \leq \rho_{\mathrm{s}} \leq 1$, and 1 indicates a perfect positive monotonic relationship.

4.2.3 Point-wise Posterior Predictive Checking

When a surgeon is removing a tumor mass near a critical structure, it is vital that s/he knows how close the predicted instrument location is from the structure and how confident the prediction is. With GP registration, we can predict the instrument location using a displacement vector. Meanwhile, we can also provide the registration uncertainty to indicate how likely the estimated instrument location is accurate. Here, we designed a point-wise experiment to investigate whether the true location is close to the predicted location when the uncertainty is low and vice versa.

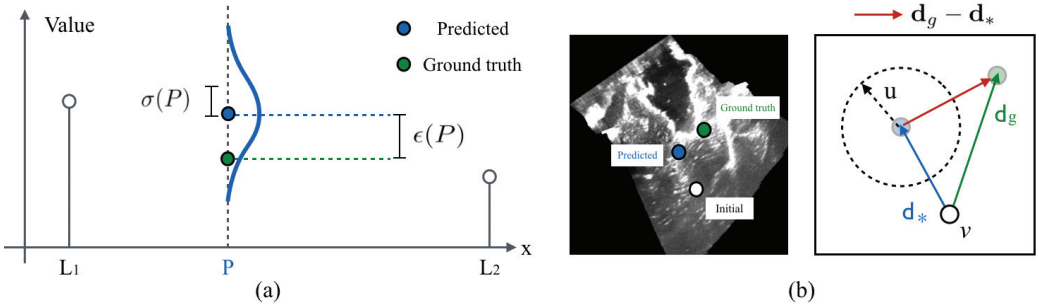


Figure 4.3: (a) An illustrative example for the point-wise posterior predictive checking experiment; (b) An illustration for how to compute ϵ and u in the context of IGNs

The point-wise experiment is inspired by posterior predictive checking (PPC) (Gelman et al., 2004). PPC examines the fitness of a model using the similarity between values generated by the posterior distribution and the observed ones.

In an illustrative 1D example shown in Figure 4.3(a), L_1 and L_2 are two landmarks whose values are indicated by the length of vertical bars. The goal is to interpolate the value at location P. Here the blue bell-curve is the estimated posterior distribution $p(\mathrm{P}|\mathrm{L}_1, \mathrm{L}_2)$ and it has a mean of p_*. Since we know the ground truth value p_g, we can compute the estimation error as $\epsilon(\mathrm{P}) = |\mathrm{p}_\mathrm{g} - \mathrm{p}_*|$. The standard deviation σ of the posterior is often used to represent the uncertainty u of the estimation.

In the context of GP registration shown in Figure 4.3(b), the white circle is the initial location of voxel v on the p-US image. The green circle represents the ground truth location of deformed v on the i-US image and the blue circle is the predicted location. In Figure 4.3(b), $\mathbf{d}_\mathrm{g}$ and $\mathbf{d}_*$ are the ground truth and predicted displacement vectors respectively, and the registration error can be computed as $\epsilon = \|\mathbf{d}_\mathrm{g}(i) - \mathbf{d}_*(i)\|$. As u is the registration uncertainty associated with d_*, it is visualized by a circle where the magnitude of u is the radius of the circle (a larger circle indicates a higher uncertainty).

In the point-wise experiment, we compute u and ϵ for every voxel-of-interest and form two discrete variables U and E. Using ρ_s, we can measure how strong their monotonic relationship is.

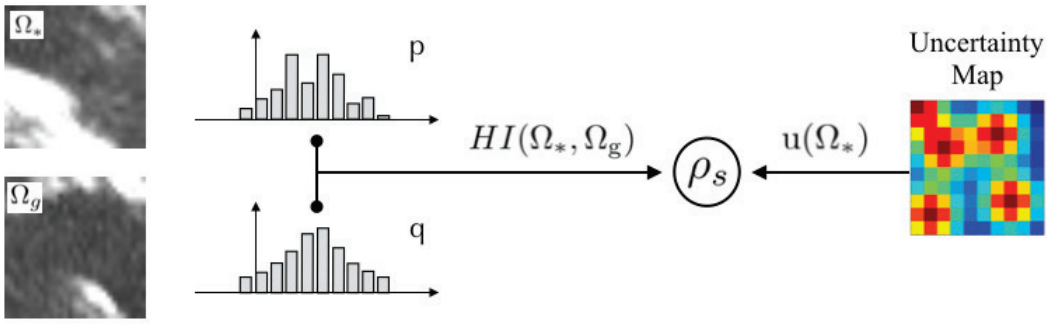


Figure 4.4: An illustration for using the HI metric to compute Spearman's rank correlation coefficient for patches.

4.2.4 Patch-wise Correlation Test

We also investigate the correlation between u and ϵ over image-patches. Because in IGNs, surgeons may be more interested in registration errors over region of interest. We plan to present the uncertainty to surgeons via color overlays so that they can get a higher level understanding of registration error.

Given a voxel v located at $\mathbf{x}_v \in \mathbb{R}^3$, we define an image patch $\Omega \subset \mathbb{R}^3$ as a sub-volume centered at $\mathbf{x}_v$, let Ω have size N. Assuming $u_{\mathbf{x}}$ is the voxel-wise uncertainty at location $\mathbf{x}$, we can compute the patch-wise uncertainty as the mean voxel uncertainty over Ω as $u(\Omega) = \frac{1}{N} \sum_{\mathbf{x} \in \Omega} u_{\mathbf{x}}$. The estimation of ϵ over a patch is not straightforward. An ideal way for measuring the patch-wise registration error would be to use the residual Euclidean distance over densely-labeled and well-distributed landmarks placed on both patches. However, to our knowledge, none of the existing neurosurgical datasets has such landmarks.

In this study, since all experiments are based on uni-modal registration, we use intensity-based dissimilarity metrics to measure the error between ground truth patches Ω_g's and predicted patches Ω_*'s. In a previous study that attempted to use patch-wise dissimilarity measures to indicate registration quality, the Histogram Intersection (HI) metric achieved the best result (Schlachter et al., 2016). Therefore, we use HI as a dissimilarity metric together with the commonly known Sum of Squared Differences (SSD).

For Ω_{g} and Ω_{*}, let $p(t)$ and $q(t)$ be the intensity probability mass functions. K is the number of intensity bins in the histogram. HI can be estimated as

$$HI(\Omega_{*}, \Omega_{\mathrm{g}}) = 1 - \sum_{i=1}^{K} \min(p(t_i), q(t_i)). \tag{4.5}$$

Figure 4.4 illustrates using HI in the patch-wise correlation test. For SSD and HI, their scalar outputs are used as $\epsilon(\Omega)$ for estimating ρ_{s}. Noticing that the size of patches may influence the test result, thus we conduct multiple patch-wise experiments using different patch sizes.

4.3 Experiments

We conducted the experiments on two clinical datasets for neurosurgical registration, RESECT (Xiao et al., 2017) and MIBS. RESECT is a public benchmark dataset for IGN (Xiao et al., 2020), while MIBS is a proprietary dataset from a local hospital. Both datasets in total contain 23 sets of p-US and i-US scans that were acquired from patients with brain tumors. US data were provided as a reconstructed 3D volume. In the p-US to i-US GP registration context, we tested manually annotated landmarks in the RESECT dataset and automatically detected landmarks in MIBS [35,36], which does not have manual annotations. Noticing that all tested points were not used for GP interpolation.

4.3.1 Point-wise Experiment

In the point-wise experiment, for each landmark on the i-US image, GP registration estimated $\mathbf{d}_{*}$ and σ. Since $\mathbf{d}_{\mathrm{g}}$ is known, we can calculate ϵ and then combine all points to compute ρ_{s} for a pair of images.

The estimated point-wise ρ_{s}'s are summarized inFigure 4.5. For manual landmarks in RESECT, the mean value of ρ_{s} is 0.2899, which indicates a weak-to-moderate positive monotonic correlation. Automatically extracted landmarks achieved an average ρ_{s} of 0.4014, which can be categorized as a moderate-to-strong correlation. However, both scores are significantly lower

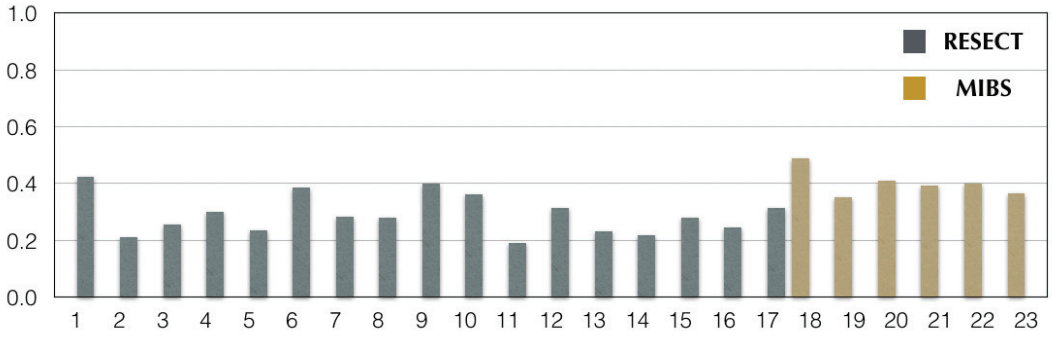


Figure 4.5: The estimated ρ_s for the point-wise experiment. We can see a moderate positive monotonic relationship between u and ϵ.

than what is required for a perfect positive monotonic relationship. At this stage, it's still too early to conclude definitively whether it is safe to use GP registration uncertainty to assess the accuracy of predicted, e.g., instrument location.

We suspect that the ρ_s discrepancy between these two groups of landmarks is due to the nature of GP uncertainty and the distribution of landmarks: In GP registration, the uncertainty of a voxel depends on its distance to neighboring interpolating points, e.g., the closer to interpolating points, the lower uncertainty it has. If an annotated landmark is far away from all interpolating landmarks, it is likely to have high uncertainty. In case it happens to be located in a region with less severe deformation, that highly uncertain landmark would have a low registration error, thus lower the overall score for ρ_s.

4.3.2 Patch-wise Experiment

In the patch-wise experiment, we padded $\pm k$ surrounding voxels to each landmark. For example, ± 2 padding generates a patch of the size of $5 \times 5 \times 5$. Tested values of k include 3 and 5. We calculated ϵ using SSD/HI for all patches and computed $\rho_s(\Omega)$ afterward.

The estimated patch-wise ρ_s's are shown in Figure 4.6. It can be seen that values of $\rho_s(\Omega)$'s are consistently low for both datasets. We deduce the reasons for low $\rho_s(\Omega)$ values are: (1) In the presence of large deformation, e.g., tumor resection, a pair of well-matched patches may look drastically different. In this

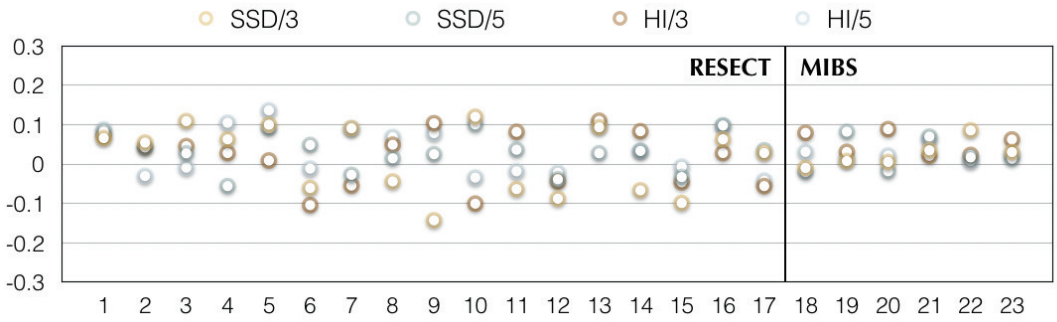


Figure 4.6: The estimated $\rho_s(\Omega)$ for the patch-wise experiment. Values of $\rho_s(\Omega)$'s are consistently low for both datasets.

case, instead of the residual Euclidean distance over densely-labeled and well-distributed landmarks, other appearance-based dissimilarity measures become sub-optimal for estimating the registration error; (2) σ used in the calculation is transformation uncertainty, while intensities over patches is label uncertainty (Luo et al., 2019a). These two quantities may be inherently uncorrelated in GP registration. (3) Features that surgeons are interested in, e.g., tumor margins or nearby blood vessels, may be limited to a small region. It may make more sense to estimate the regional $\rho_s(\Omega)$ instead of using the whole image.

4.4 Conclusion

"Are registration uncertainty and error monotonically associated?' is a fundamental question that has been overlooked by researchers in the medical imaging community. There has been significant progress in the development of fast and accurate methods for performing non-rigid registration. Since all of these methods are subject to some error and rarely used in the operating room, an answer to this question, which enables the use of registration uncertainty as a surrogate for assessing registration error, can increase the feasibility of non-rigid registration in interventional guidance and advance the state of image-guided therapy.

In this pilot study, we systematically investigate the monotonic association between Gaussian process registration uncertainty and error in the context of Image-guided neurosurgery. At the current stage, the low-to-moderate correlation between GP uncertainty and error indicates that it may not be

feasible to apply it in practice. Nevertheless, this work opens a research area for uncertainty/error relationship analysis and may inspire more research on this topic to verify and enhance the applicability of registration uncertainty.

Chapter 5

Related Projects

During the course of research, we also pursued two related projects which are intertwined with the proposed *active image registration* (AIR) method. In the first project, we added a *variogram*-based outlier removal step to improve the feature extraction and matching in AIR. In the second project, we used the variogram to perform a third-party screening on annotations of two public datasets. Since both projects are based on the *variogram*, we summarize them together in this chapter. In Section 5.1 and Section 5.2, we explain in detail how to use the variogram for vector outlier screening and annotation quality control. This chapter is based on the publications (Luo et al., 2018a) and (Luo et al., 2020b).

5.1 Vector-outlier Screening for Feature-based Image Registration

Matching points that are derived from features or landmarks in image data is a key step in some medical imaging applications.For tasks such as feature-based registration in image-guided neurosurgery, even a few mismatches, in the form of invalid displacement vectors, could cause serious consequences. As a result, having an effective tool by which operators can manually screen all matches for outliers could substantially benefit the outcome of those applications. We introduce a novel variogram-based outlier screening method for vectors. The variogram is a powerful geostatistical tool for characterizing the spatial dependence of stochastic processes. Since the spatial correlation of invalid displacement vectors, which are considered as vector outliers, tends to behave

differently than normal displacement vectors, they can be efficiently identified on the variogram. We validate the proposed method on 9 sets of clinically acquired ultrasound data.

5.1.1 Introduction

Matching points that are derived from features or landmarks in image data is a key step in many medical imaging applications, such as 3D reconstruction and image registration for surgical navigation. Since most robust point matching algorithms claim to be able to deal with outliers (Chui and Rangarajan, 2013; Fischler and Bolles, 1981; Fitzpatrick, 2009; Jian and Vemuri, 2010; Myronenko and Song, 2010), users may place high confidence in the matching result and use it without further examination. However, for tasks where precision is of paramount importance, even a few mismatches can cause serious consequences. As a result, having an effective tool by which operators can manually screen all matches for outliers could substantially benefit the outcome of those applications.

Image-guided neurosurgery is a task that requires precise localization of tumor boundaries. Conventionally, surgeons use commercial Image Guided Neuro-navigation Systems (IGNSs) to map the preoperative image data to an intraoperative patient coordinate system so as to get an updated view of the brain. Unfortunately, intraoperative brain deformation, also known as brain shift, invalidates this mapping and limits the trustworthiness of using preoperative images in intraoperative surgical navigation (Bayer et al., 2017b; Gerard et al., 2017). In order to compensate the brain shift and increase the accuracy of neurosurgery, non-rigid image registration (Hata et al., 1999; Letteboer et al., 2003; Reinertsen et al., 2007; Riva et al., 2017) is sometimes adopted to spatially align the preoperative and intraoperative images.

Because of tumor resection, the same structure may not exist on both preoprative and intraoperative images, hence Feature-Based Registration (FBR) becomes a promising strategy due to its robustness in registering images with missing correspondence (Toews and Wells, 2013). FBR consists of three steps. In the feature extraction step, distinctive local image features are automatically extracted and identified as key-points on preoperative and intraoperative

images. In the feature matching step, a matcher searches for a corresponding intraoperative key-point for each key-point on the preoperative image. From every matched key-point pair, a displacement vector can be obtained to indicate the movement of the preoperative key-point due to the brain shift process. In the last step of FBR, the algorithm generates a dense deformation field for the entire preoperative image from the displacement vectors and uses it to map the preoperative image to the intraoperative space. If invalid displacement vectors, which falsely reflect the tissue movement under brain shift, are obtained from mismatched key-point pairs, it will negatively affect the registration result.

In this section, we introduce a simple yet informative variogram-based outlier screening method for vectors and demonstrate its usefulness in the context of feature-based registration. While variograms are extensively used in geostatistics to capture the spatial dependence of stochastic processes (Cressie, 1991), they have not yet received much attention in the medical imaging field. Since the spatial correlation of invalid displacement vectors, which are considered vector outliers, tend to behave differently than valid displacement vectors, they can be identified on the variogram cloud and flagged for further examinations.

5.1.2 Summarizing Variogram in the Context of FBR

In the feature matching step of FBR, we obtain a displacement vector from every matched key-point pair. For example, as shown in Figure 5.1(a)(b), assume $\mathbf{k}_p$ is the coordinate of an extracted preoperative key-point, and $\mathbf{k}_i$ is the coordinate of its corresponding intraoperative counterpart. After overlaying both images in Figure 5.1(c), the displacement vector $\mathbf{d}$ from the preoperative key-point pointing to the intraoperative key-point can be calculated as $\mathbf{k}_i - \mathbf{k}_p$.

All displacement vectors form a vector field, which is supposed to indicate the intraoperative brain shift. We attempt to model the spatial dependence of the vector field and use the dependency as a criterion for identifying displacement vector outliers.

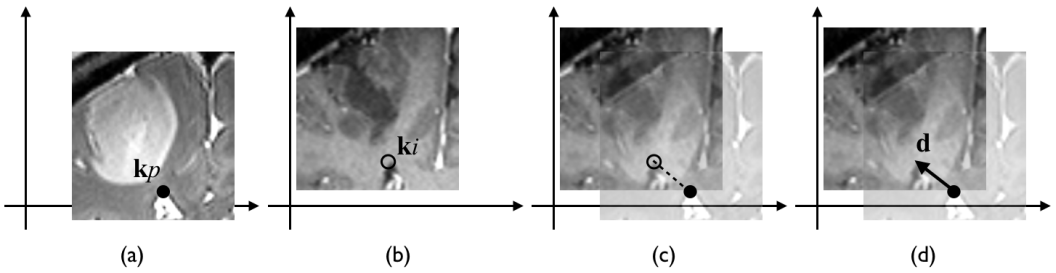


Figure 5.1: (a) The extracted key-point on the preoperative image; (b) The extracted key-point on the intraoperative image; (c) Overlaying two images; (d) The displacement vector **d** for the preoperative key-point.

5.1.3 Random Process Model

Brain shift is caused by the interaction of physical, surgical and biological factors. Due to the complexity and incomplete understanding of the phenomenon, we model the deformation of the brain as a stochastic process. In our model, let $\mathbf{x} \in \mathbb{R}^d$ be a generic data location in the d-dimensional Euclidean space. Suppose $\mathbf{Z}(\mathbf{x})$ is a random quantity at location $\mathbf{x}$, its realization $\mathbf{z}(\mathbf{x})$ is a displacement vector like the one in Figure 5.1(d). Given an index set $D \subset \mathbb{R}^d$, we can model the vector field as

$$\{\mathbf{Z}(\mathbf{x}) \colon \mathbf{x} \in D\}. \tag{5.1}$$

For brain shift, the deformation at a particular location is likely to have influence in all directions. Capturing the spatial correlation of deformation is essential for understanding the phenomenon. The definition of the variogram comes naturally as we quantify the spatial dependence of $\mathbf{Z}(\mathbf{x})$.

5.1.4 The Variogram and Empirical Variogram

The pairwise dependence between displacement vectors in a random field is the variance of the difference between the values at two locations across the field (Cressie, 1991):

$$2\gamma(\mathbf{x}_1, \mathbf{x}_2) := \mathrm{var}(\mathbf{Z}(\mathbf{x}_1) - \mathbf{Z}(\mathbf{x}_2)). \tag{5.2}$$

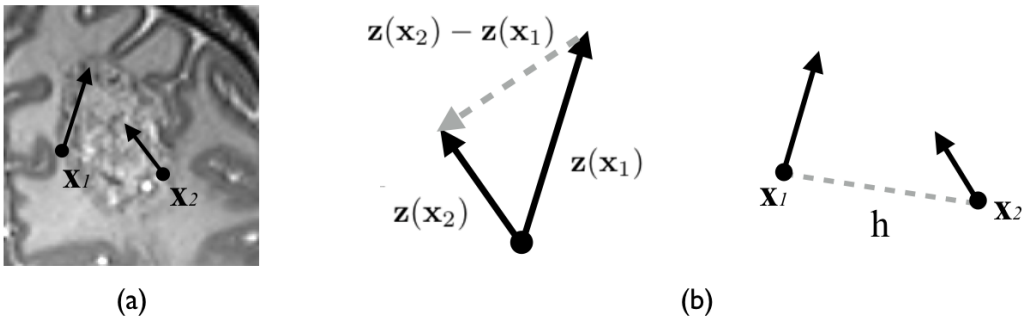


Figure 5.2: (a) Two key-points $\mathbf{x}_1, \mathbf{x}_2$; (b) The displacement differences between $\mathbf{z}(\mathbf{x}_1)$ and $\mathbf{z}(\mathbf{x}_2)$, and the lag h.

We make the mild assumption that the random field has a constant mean and is intrinsically stationary. Intrinsic stationarity states that the variance of difference is the same between any two points that are separated by the same distance. Adopting a terminology from time series analysis, we represent the separation between two spatial locations by a lag vector $\mathbf{h} = \mathbf{x}_2 - \mathbf{x}_1$ and rewrite (2) as

$$\begin{aligned} 2\gamma(\mathbf{x}_1, \mathbf{x}_2) &= E[((\mathbf{Z}(\mathbf{x}_1) - \mathbf{Z}(\mathbf{x}_1 + \mathbf{h}))^2] \\ &= 2\gamma(h). \end{aligned} \tag{5.3}$$

$\gamma(h)$, also known as the variogram (sometimes is called the semivariogram) of a random field $\mathbf{Z}(\mathbf{x})$, is defined as one-half the average squared difference between the value at locations separated by distance h Cressie (1991). The variogram function $\gamma(h)$ depends only on h and it measures the correlation as a function of distance.

Notice that $\gamma(h)$ is a theoretical function. In practice, we do not have the displacement vector for every location to estimate $\gamma(h)$, so we use its empirical alternative (Matheron, 1963). This definition will be clarified in Section 2.3.

$$\hat{\gamma}(h \pm \delta) := \frac{1}{2|N(h \pm \delta)|} \sum_{(i,j) \in N(h \pm \delta)} \|\mathbf{z}(\mathbf{x}_i) - \mathbf{z}(\mathbf{x}_j)\|^2. \tag{5.4}$$

5.1.5 Estimating the Empirical Variogram in FBR

In FBR, there are several ways to represent $\|\mathbf{z}(\mathbf{x}_i) - \mathbf{z}(\mathbf{x}_j)\|$ and h. In this paper, we use the following representations. In Figure 5.2(a), on a preoperative

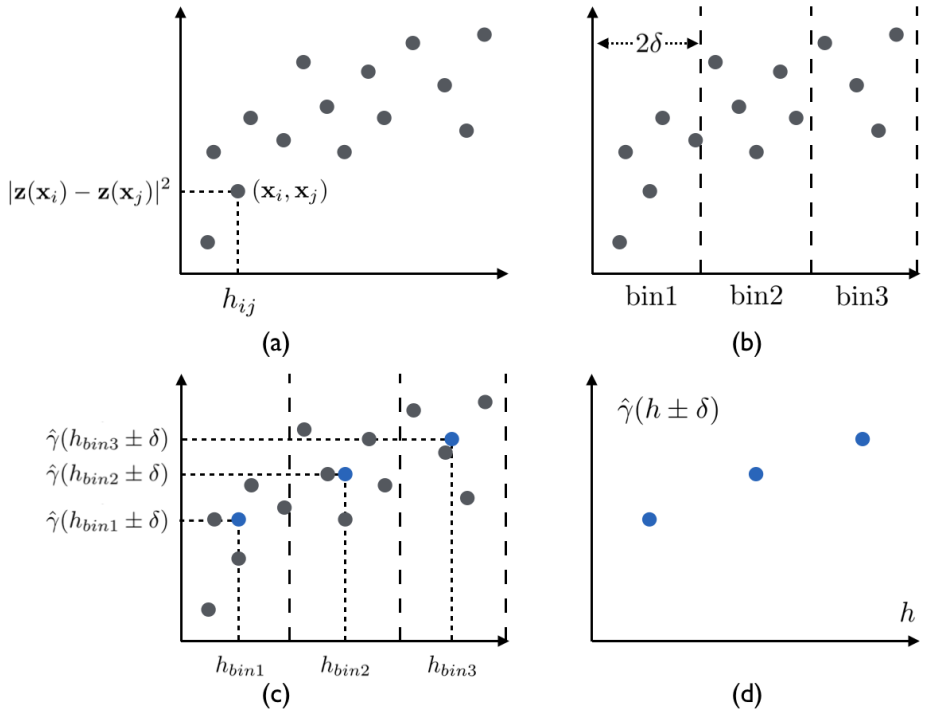


Figure 5.3: (a) The variogram cloud; (b) Binning the variogram cloud with the bin width 2δ; (c) Computing the mean for every bin; (d)Plotting $\hat{\gamma}(h_{\text{bin}k} \pm \delta)$ and $h_{\text{bin}k}$.

Magnetic Resonance Image (MRI), there are two key-points whose coordinate are $\mathbf{x}_1$ and $\mathbf{x}_2$. $\mathbf{z}(\mathbf{x}_1)$ and $\mathbf{z}(\mathbf{x}_2)$ are their associated displacement vectors. As can be seen in Figure 5.2(b), $\|\mathbf{z}(\mathbf{x}_1) - \mathbf{z}(\mathbf{x}_2)\|$ is equal to the norm of the vector $\mathbf{z}(\mathbf{x}_2) - \mathbf{z}(\mathbf{x}_1)$, and lag h is the Euclidean distance between the two key-points $\|\mathbf{x}_2 - \mathbf{x}_1\|$.

Estimating an empirical variogram can be analogous to building a histogram based on samples from a continuous distribution. The strategy is to group pairs of points that have similar h into the same bin, and approximate the variogram by combining information from all of the bins. Here we illustrate the 4 steps of estimating an empirical variogram using Figure 5.3.

(a) Construct the variogram cloud by plotting $\|\mathbf{z}(\mathbf{x}_i) - \mathbf{z}(\mathbf{x}_j)\|^2$ for all key-point pairs with their lag h_{ij}.

(b) Introduce a variable δ as the tolerance range for lag h, and divide the variogram cloud into bins whose width is set to 2δ.

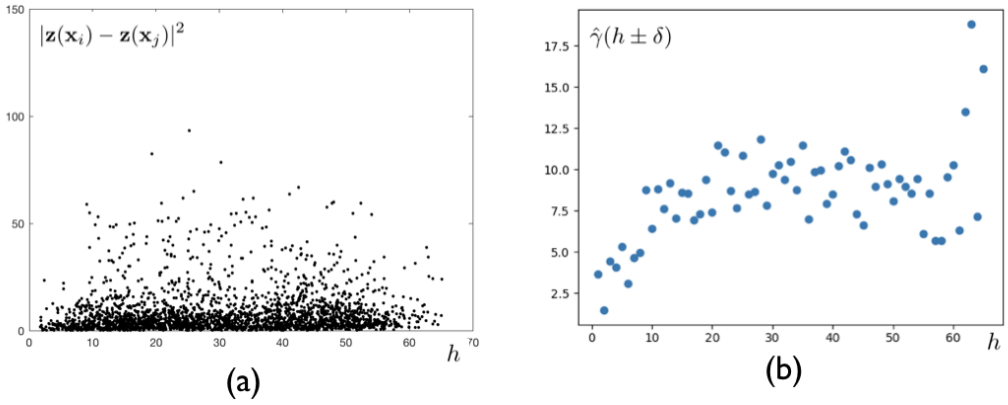


Figure 5.4: (a)The variogram cloud for a set of US data; (b)The empirical variogram for the same US data.

(c) Calculate the mean $\hat{\gamma}(h_{\text{bin}k} \pm \delta)$, which are marked as blue in Figure 5.3(c), for every bin using Equation (4). $|N(h_{\text{bin}k} \pm \delta)|$ represents the number of key-point pairs of bin k.

(d) Plotting all $\hat{\gamma}(h_{\text{bin}k} \pm \delta)$ with their $h_{\text{bin}i}$ to obtain the empirical variogram.

The lag tolerance δ should be neither too large nor too small, since it may veil the short-distance correlations or produce empty bins. Typically, δ is determined in Ad hoc fashion. When $\delta = 0$, the empirical variogram $\hat{\gamma}(h \pm \delta)$ is an unbiased estimator of the theoretical variogram $\gamma(h)$ (Cressie, 1991). Notice that the proposed variogram based outlier screening method primarily uses the variogram cloud, hence the choice of δ has no impact on its performance.

In Figure 5.4, we display a variogram cloud and the empirical variogram estimated from a set of preoperative and intraoperative ulstrasound (US) data. For this US data, we were able to find 71 pairs of matched key-points, and the δ used to estimate the empirical variogram is 0.5mm. Here the x-axis represent lag h, and the y-axis is the value difference among pairs. The empirical variogram does indicate that nearby key-points tend to have more similar displacement vectors than those that are far apart. In the rest of this article, we use the variogram and empirical variogram interchangeably.

5.1.6 Variogram and Vector-outlier Removal

In FBR, the feature matching algorithm, e.g., (Toews and Wells, 2013), may identify feature correspondences between preoperative and intraoperative images solely based on the local image appearance similarity. While the algorithm is claimed to be robust against outliers, similar image patches and artefacts at non-corresponding locations of the brain will sometimes be extracted and matched. These mismatched key-point pairs produce invalid displacement vectors, as outliers, that fail to reflect the actual brain deformation. Since all displacement vectors will be used to generate the dense deformation field, we want a simple yet informative tool so that operators can efficiently screen all displacement vectors for outliers.

In Section 2, we mentioned that the variogram cloud can capture the spatial dependence of displacement vectors. Since the spatial correlation of vector outliers tend to behave differently than valid displacement vectors, they can be spotted on the variogram. There are two key features for the variogram-based outlier screening: 1) It is effective and easy to use. Operators can quickly spot and flag potential outliers on the variogram cloud. 2) The variogram is also very informative. Operators also can grasp the quality/trend of the vector field by observing variogram patterns.

5.1.7 Global and Local Outliers in FBR

Outliers are extreme values that deviate from other observations in the data. In FBR, there are generally two types of outliers:

1. A global outlier is a sample point that has a very high or a very low value compared to all the values in a dataset. For example, due to the symmetry of cerebral hemispheres, similar intensity patterns exist on the opposite half of the brain for MRI images. If these patterns are mismatched as corresponding key-point pairs, they produce global outliers. These global outliers are clearly incorrect, and should be removed.

2. A local outlier is a sample key-point that has a value within the normal range for the entire dataset, but is unusually high or low compared to

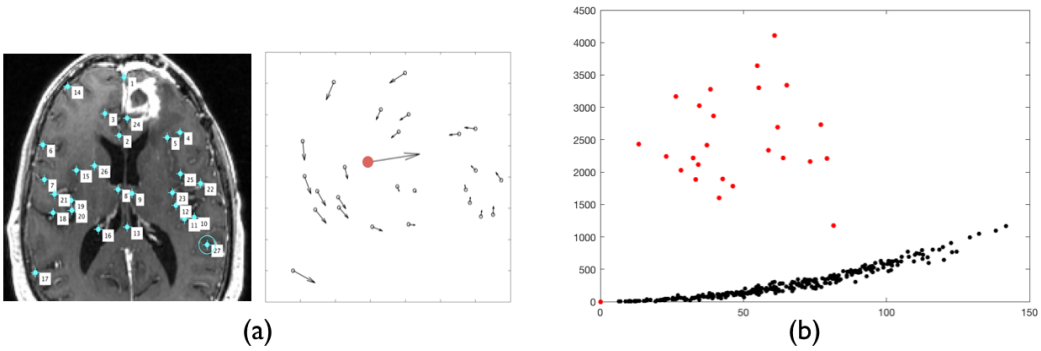


Figure 5.5: (a) Displacement vectors and a red global outlier; (b) The global outlier identified in the variogram cloud.

its neighboring points. In FBR, some displacement vectors may seem to be local outliers at first, yet they capture the actual local deformations that are induced by the brain shift phenomenon. These key-points are vital for understanding and correcting the brain shift, and may be the most significant points in the registration. Therefore, it is unnecessary to remove all local outliers, but suppress them to a certain degree or perhaps to flag them for manual intervention.

5.1.8 Outliers in the Variogram Cloud

Here we explain how to look for global and local outliers in a variogram cloud by the following two examples.

In a synthetic vector field shown in Figure 5.5(a), we have 25 valid displacement vectors and a manually added global outlier. Since the displacement vector of that global outlier is drastically different from other key-points, all pairings of points involving that global outlier, regardless of the lag distance h, have high $\|\mathbf{z}(\mathbf{x}_i) - \mathbf{z}(\mathbf{x}_j)\|^2$ values. From the variogram cloud in Figure 5.5(b), we can see two main strata of points. All the higher values, which are marked as red, come from pairings with the manually added global outlier. On the other hand, the lower strata is composed of pairings among the rest of valid displacement vectors.

In another synthetic vector field shown in Figure 5.6(a), we have 25 valid displacement vectors and a manually added local outlier. The local outlier does not have high values across all distance in the variogram cloud. However,

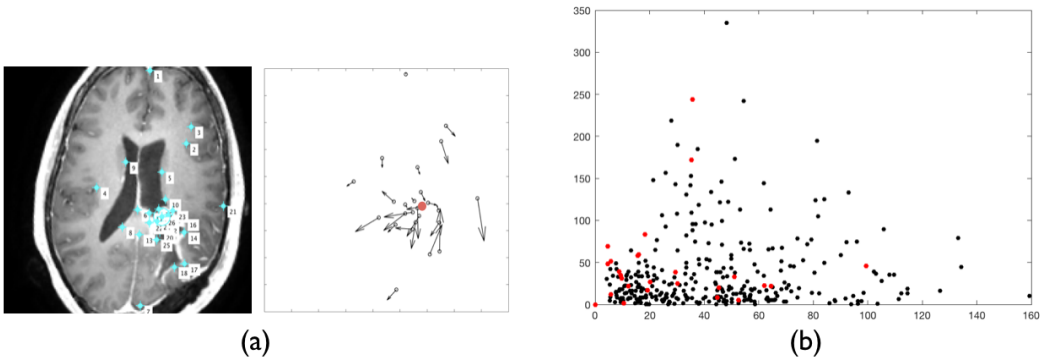


Figure 5.6: (a) Displacement vectors and a red local outlier; (b) The local outlier identified in the variogram cloud.

the variogram values of its close neighbors are relatively high. In Figure 5.6(b), all values from pairings with the local outlier are marked as red. On the bottom-left of the variogram cloud, we can notice a small cluster of abnormally high values as expected. Sometimes, even the local outlier is hard to notice in the displacement vector field, it is easy to spot in the variogram cloud.

Given the characteristics of global and local outliers, screening using the variogram is very straightforward. Since the variogram provides a distinct visualization for outliers, users can get an overview of potential outliers by simply looking at the patterns on the top and bottom-left corner of the variogram cloud. If abnormal variogram values are found, it is easy to trace back to the key-point matches for further examinations.

5.1.9 Experiments

In the experiments we tested the proposed variogram-based outlier screening method on a clinically acquired US data set. This 9-patient data set contains US images taken from 3 different stages of neurosurgery: before opening the dura (predura), after opening the dura (postdura) and during the tumor resection (intraoperative). We registered the intraoperative image to the predura image. In case the intraoperative image is not available, we register the postdura image instead.

After the feature matching, an operator quickly looked for abnormal patterns on the top and bottom-left corner of the variogram cloud and flagged the

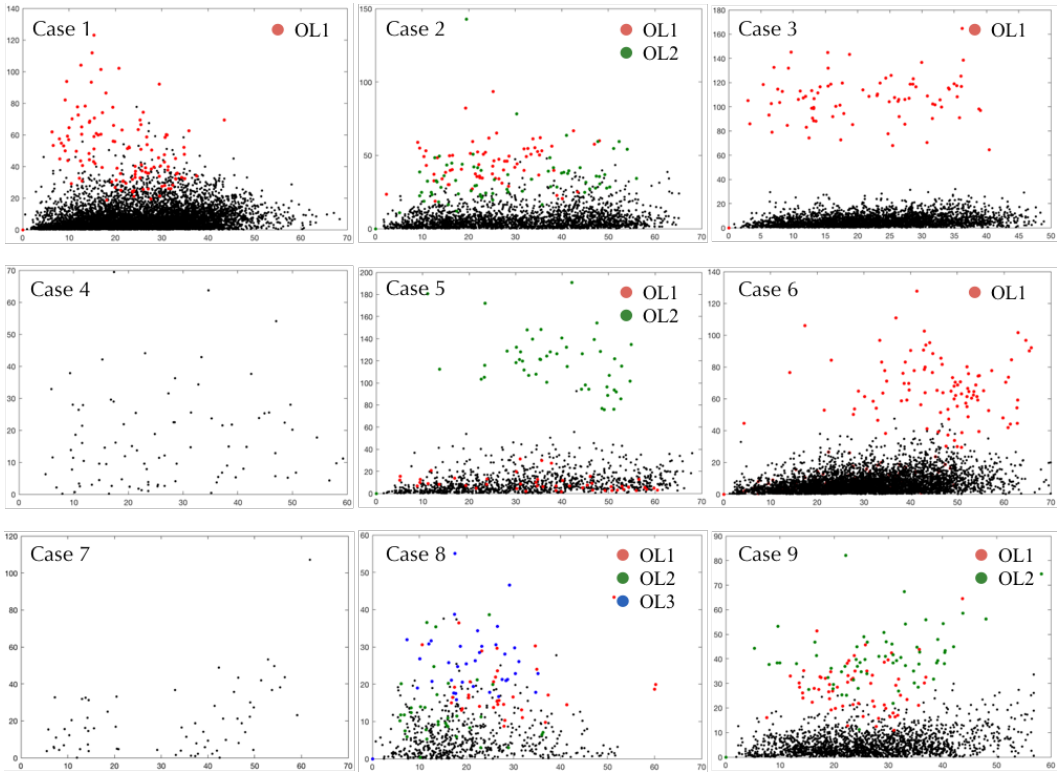


Figure 5.7: The variogram for all potential outliers that were flagged by the operator in a total of 9 cases.

matches that are in correspondence with those patterns. All flagged matches, mixed with some valid matches, are further sent to 8 experienced medical imaging researchers for rating.

In the rating process, without knowing how many potential outliers are in the group, raters assign a score from 1 (bad) to 5 (good) to every match. After a zero-mean normalization for each rater, we calculated the score difference of every potential outlier to the mean score of valid matches and recorded it as Ourlier Score (OLS). The number of matches, number of potential outliers (#OL) suggested by the variogram, and the OLS of each potential outlier are presented in table 1. Here OLS1 means the OLS for potential outlier 1, etc.

The feature matching algorithm we use is claimed to be robust (Toews and Wells, 2013), yet despite the two cases which have less than 15 displacement vectors, we were able to flag at least one potential outlier using the variogram. Given the fact that all flagged potential outliers for cases 1–3 and cases 5–6 and 9 have substantially lower scores than normal matches, it is reasonable to

Table 5.1: Experimental results of the proposed outlier screening method

Case	1	2	3	4	5	6	7	8	9
Data	intra	intra	postd	postd	intra	intra	intra	postd	intra
Matches	123	71	84	14	49	98	12	37	64
#OL	1	2	1	0	2	1	0	3	2
OLS1	-1.3	-1.4	-2.3	n/a	-0.9	-2.3	n/a	-0.8	-2.0
OLS2	n/a	-1.9	n/a	n/a	-2.2	n/a	n/a	-0.4	-2.2
OLS3	n/a	n/a	n/a	n/a	n/a	n/a	n/a	-1.8	n/a

conclude that adding an outlier screening step after the feature matching is beneficial.

Figure 5.7 shows the variogram for all potential outliers that were flagged by the operator. In Figure 5.8, we show the axial, sagittal and coronal views of some found outlier matches together.

5.1.10 Conclusion

Matching points that are derived from features or landmarks in image data is a key step in some medical application. For application that demand a high precision, such as feature-based preop-to-intraop registration, adding an outlier screening step could make it less prone to mismatches and improve the overall result. In this paper, we proposed to use the variogram to estimate the spatial dependence of displacement vectors. Since the spatial correlation of vector outliers behave differently than valid displacement vectors, they can be efficiently identified on the variogram and flagged for further examinations.

5.2 Do Public Datasets Assure Unbiased Comparisons for Registration Evaluation?

With the increasing availability of new image registration approaches, an unbiased evaluation is becoming more needed so that clinicians can choose the most suitable approaches for their applications. Current evaluations typically use landmarks in manually annotated datasets. As a result, the quality of annotations is crucial for unbiased comparisons. Even though most data

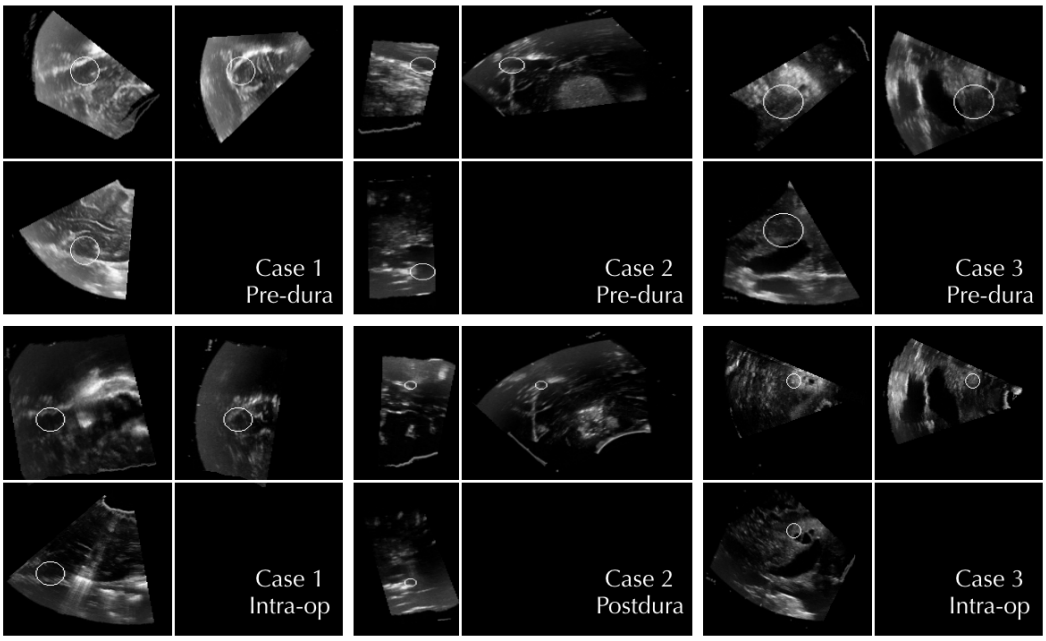


Figure 5.8: Screen shots of axial (top left), sagittal (bottom left) and coronal (top right) views of 3 outliers found by the variogram. Note the matching quality of the axial slice of case 1, the sagittal slice of case 2 and the cases 3 are considerably poor.

providers claim to have quality control over their datasets, an *objective third-party screening* can be reassuring for intended users. In this study, we use the variogram to screen the manually annotated landmarks in two datasets used to benchmark registration in image-guided neurosurgeries. The variogram provides an intuitive 2D representation of the spatial characteristics of annotated landmarks. Using variograms, we identified potentially problematic cases and had them examined by experienced radiologists. We found that (1) a small number of annotations may have fiducial localization errors; (2) the landmark distribution for some cases is not ideal to offer fair comparisons. If unresolved, both findings could incur bias in registration evaluation.

5.2.1 Introduction

The evaluation of non-rigid registration is challenging for two reasons: firstly, quantitative validation between aligned images is sometimes difficult due to lack of ground truth. Secondly, the location where accurate registration is needed

may vary by surgical procedure, e.g., in brain atlas building, good alignment of the ventricle region is sought after, while in image-guided neurosurgery, surgeons are interested in accurate registration of the preoperative (p-) tumor boundary to the intraoperative (i-) coordinate space (Gerard et al., 2017; Luo et al., 2018b, 2019b; Morin et al., 2017). Because of these issues, it is difficult to set up a unified standard to characterize registration error (Maintz and Viergever, 1998; Sotiras et al., 2013).

In early work, image similarity measures, such as the sum of squared differences or mutual information, were used as evaluation criteria for registration (Maintz and Viergever, 1998; Song, 2017; Sotiras et al., 2013). The "Retrospective Image Registration Evaluation" (RIRE) project (Fitzpatrick, 2007; Sonka, 2000; West et al., 1997) introduced Target Registration Error (TRE) and Fiducial Registration Error (FRE) to evaluate registration. TRE is the true error of registered target points in physical space, while FRE represents the error of registered annotated fiducial markers in image space. Even though annotated fiducial markers may not be truly accurate due to operator error, FRE is often used as a surrogate of TRE for convenience (Danilchenko and J.M., 2011; Datteri and Dawant, 2012; Min et al., 2020). The Vista (Hellier et al., 2003) and NIREP (Christensen et al., 2006) projects included additional registration error measures, e.g., the region of interest (ROI) overlap, the average volume difference and the Jacobian of the deformation field. Recently, multiple registration approaches were compared based on Computed Tomography of the abdomen (Kabus et al., 2009; Murphy et al., 2011; Xu et al., 2016) and Magnetic Resonance (MR) images of the brain (Klein et al., 2009, 2010; Ou et al., 2011; Yassa and C.E., 2009).

Due to its simplicity and the reliability issues of other criteria (Rohlfing, 2012), FRE has become the most widely used registration error measure. However, using FRE has certain limitations:

1. Because fiducial markers (landmarks) are annotated by localization algorithms (manual, automatic or semi-automatic methods), they may contain Fiducial Localization Error (FLE) (Sonka, 2000), which measures the discrepancy between an annotated landmark and its true location. FLEs cause false registration errors and should be avoided.

2. The FRE only estimates the error at specific landmark locations, thus a dense population of landmarks is preferred. If landmarks are sparse or are not distributed evenly across the entire ROI, a bias that favors regions with landmarks in the registration evaluation may be introduced.

Recently, more annotated data sets are becoming publicly available and these sets are being used to compare existing algorithms and evaluate new methods. Newly proposed registration algorithms are then characterized only by demonstrating superior performance on these datasets (Machado et al., 2018; Xiao et al., 2020).

To provide an unbiased evaluation of registration, the quality of the annotations is crucial. However, to the best of our knowledge, objective quality control over annotation in public datasets has been overlooked by the image registration community. Even though many providers claim to have mitigated FLEs and other problems by having multiple observers localize the landmarks (and averaging their results) (Song, 2017), an *objective third-party screening* can be reassuring for intended users.

In this study, we perform a third-party screening of the annotations of two benchmark datasets for image-guided neurosurgery, RESECT (Xiao et al., 2017) and BITE (Mercier et al., 2012). Both datasets have corresponding landmarks on p-MR and i-Ultrasound (US) images. Minimizing the FLE for these landmarks is the standard evaluation method in related registration challenges (Xiao et al., 2020). The tool we choose for the screening is called the variogram, which has been extensively used to describe the spatial dependence of minerals in geostatistics. The variogram has been brought to the medical imaging field as a means to identify vector outliers for landmark-based image registration (Luo et al., 2018a). In this screening, we want to (1) detect any obvious FLEs; and (2) examine the distribution of annotated landmarks. We also provide constructive discussion about the impact of our findings.

5.2.2 Method

For each pair of annotated images, we compute displacements between pairs of corresponding landmarks to generate a 3D vector field $\mathcal{D}$. By analyzing $\mathcal{D}$,

we can assess the quality of the annotations. The variogram characterizes the spatial dependence of $\mathcal{D}$ and provides an intuitive 2D representation for visual inspection (Luo et al., 2018a).

In this section, we explain how to use the variogram to flag potential FLEs and problematic landmark distributions.

5.2.3 Constructing the Variogram

In an image registration example, let $\Omega \subset \mathbb{R}^3$, $I_\mathrm{f} : \Omega \to \mathbb{R}$ and $I_\mathrm{m} : \Omega \to \mathbb{R}$ be the fixed and moving images. $(\mathbf{x}, \mathbf{x}') \in \Omega$ represents a pair of manually labeled corresponding landmarks on I_f and I_m. $\mathbf{d}(\mathbf{x}) = \mathbf{x}' - \mathbf{x}$ is the displacement vector from $\mathbf{x}$ to $\mathbf{x}'$. For K pairs of landmarks, we have a set of displacement vectors $\mathcal{D} = \{\mathbf{d}(\mathbf{x}_k)\}_{k=1}^{K}$.

Given a landmark location $\mathbf{s}$, let h represent the distance to $\mathbf{s}$. The theoretical variogram $\gamma(h)$ is the expected value of the differences between $\mathbf{d}(\mathbf{s})$ and other $\mathbf{d}$'s whose starting points are h away from $\mathbf{s}$:

$$\gamma(h) = \frac{1}{2} E[(\mathbf{d}(\mathbf{s}) - \mathbf{d}(\mathbf{s} + h))^2], \tag{5.5}$$

here $\gamma(h)$ describes the spatial dependency of displacement vectors as a function of the distance.

In this study, we are interested in the pairwise spatial dependence of all displacement vectors. Therefore we use the empirical variogram cloud $\hat{\gamma}(h)$ instead of $\gamma(h)$. Given a vector field $\mathcal{D}$, $\hat{\gamma}(h)$ can be constructed as follows:

1. For each pair of vectors $(\mathbf{d}(\mathbf{x}_i), \mathbf{d}(\mathbf{x}_j))$, compute $\epsilon_{ij} = \|\mathbf{d}(\mathbf{x}_i) - \mathbf{d}(\mathbf{x}_j)\|^2$;
2. Compute $h_\mathrm{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$;
3. Plot all (h_{ij}, ϵ_{ij}) to obtain $\hat{\gamma}(h)$.

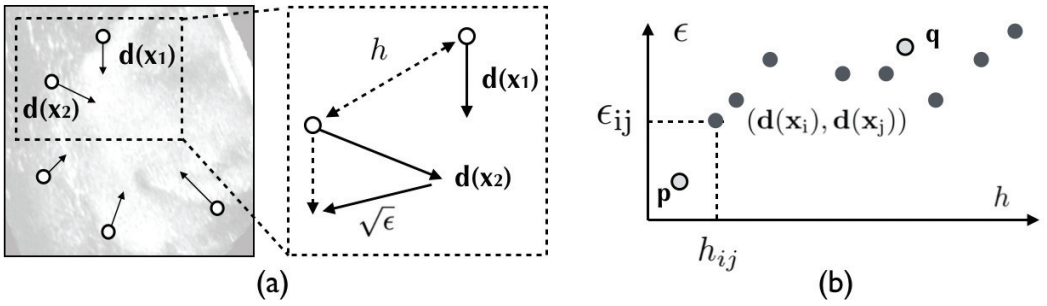


Figure 5.9: (a) An Illustration of how to compute ϵ and h. Here $\mathbf{d}(\mathbf{x}_1)$ and $\mathbf{d}(\mathbf{x}_2)$ are two displacement vectors for landmark $\mathbf{x}_1$ and $\mathbf{x}_2$ respectively. $h = \|\mathbf{x}_1 - \mathbf{x}_2\|$ is the distance between $\mathbf{x}_1$ and $\mathbf{x}_2$. While $\epsilon = \|\mathbf{d}(\mathbf{x}_1) - \mathbf{d}(\mathbf{x}_2)\|^2$ measures the displacement difference; (b) A hypothetical $\hat{\gamma}(h)$ generated from the vector field in Fig.1(a). Since the vector field has 5 displacement vectors, $\hat{\gamma}(h)$ has $\frac{5(5-1)}{2} = 10$ value points. p and q are two value points that demonstrate the typical increasing trend of $\hat{\gamma}(h)$. Here $\mathrm{h_p} < \mathrm{h_q}$, thus $\epsilon_\mathrm{p} < \epsilon_\mathrm{q}$.

Figure 5.9(a) illustrates computing ϵ and h for two vectors. Figure 5.9(b) shows a hypothetical variogram cloud generated from the vector field in Figure 5.9(a). The horizontal and vertical axes represent h and ϵ respectively. Given K displacement vectors, each $\mathbf{d}(\mathbf{x})$ has $K-1$ pairs of corresponding (h_{ij}, ϵ_{ij}), thus $\hat{\gamma}(h)$ contains $\frac{K(K-1)}{2}$ value points.

A common dependency assumption is that displacement vectors which are close to each other tend to be more similar than those far apart (Cressie, 1991). In other words, for a point in $\hat{\gamma}(h)$, a smaller h usually corresponds to a smaller ϵ. As a result, a typical $\hat{\gamma}(h)$ tends to exhibit an increasing trend. For conciseness, In the rest of this article, we call $\hat{\gamma}(h)$ variogram.

5.2.3.1 Potential FLEs

A pair of annotated landmarks $(\mathbf{x}, \mathbf{x}')$ forms a displacement vector $\mathbf{d}(\mathbf{x})$, which should indicate the deformation of $\mathbf{x}$. If $\mathbf{d}(\mathbf{x}) \in \mathcal{D}$ exhibits a spatial dependency that differs from other vectors, it could indicate FLE for $(\mathbf{x}, \mathbf{x}')$. We call these abnormally behaved vectors *outliers*. In general, there are global outliers λ_G and local outliers λ_L:

λ_G: have large differences with the majority of displacement vectors in $\mathcal{D}$.

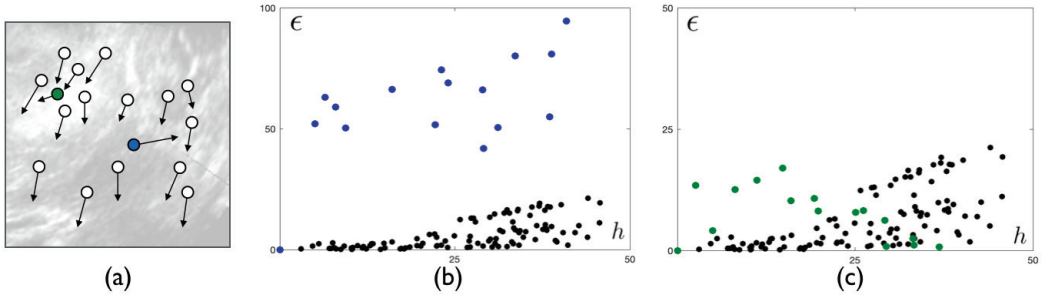


Figure 5.10: (a) Manually added atypically behaved displacement vectors. The blue point is λ_{G}, the green point is λ_{L} (b,c) Colorized global and local outliers identified in $\hat{\gamma}(h)$. Each point represents the difference between a pair of displacement vectors, e.g., blue points are vector pairs that involve the global outlier.

λ_{L}: do not have extreme values, but are considerably different from their neighbors.

Vector outliers tend to have different spatial dependence from other landmarks, which can be captured by the values of (h, ϵ), hence we can use $\hat{\gamma}(h)$ to distinguish λ_{G} and λ_{L}. In the example of Figure 5.10(a), we deliberately added two problematic landmarks, one with global error λ_{G} (blue) and one with local error λ_{L} (green), to a vector field. In Figure 5.10(b), all blue points in $\hat{\gamma}(h)$ are from adding λ_{G}, which can be easily identified because all of its corresponding points have distinctively higher values of ϵ. In Figure 5.10(c), all green points in $\hat{\gamma}(h)$ are from adding λ_{L}. We can also distinguish λ_{L} at the bottom-left corner of $\hat{\gamma}(h)$, because some of its points yield small h while having unusually large ϵ.

It is noteworthy that some critical displacement vectors, which indicate large tissue deformation, may share the same features as outliers. Therefore, $\hat{\gamma}(h)$ is used to flag suspicious λ_{G} and λ_{L}, which can be further examined by experienced radiologists.

5.2.3.2 Atypical Variogram Patterns

Since h represents the distance between a pair of vectors, the distribution of annotated landmarks can also be reflected in the pattern of $\hat{\gamma}(h)$. Compared to observing a 3D visualization of $\mathcal{D}$, where atypical patterns may be hidden

because of the viewpoint,, the variogram's 2D representation provides a clearer representation of landmark distribution.Figure 5.11(a) shows the typical smooth and steadily increasing pattern of an evenly distributed vector field $\mathcal{D}$. Other variogram patterns may indicate undesirable distributions of the vector field. Two undesirable patters are clustered landmarks and isolated landmarks:

1. If landmarks are clustered into two (or more) distinct groups, the clustering is evident in $\hat{\gamma}(h)$ as illustrated in Figure 5.11(b).

2. If a landmark is isolated from other landmarks, its points in $\hat{\gamma}(h)$ only exist in areas where h is large. Figure 5.11(c) shows an isolated landmark and its values in $\hat{\gamma}(h)$.

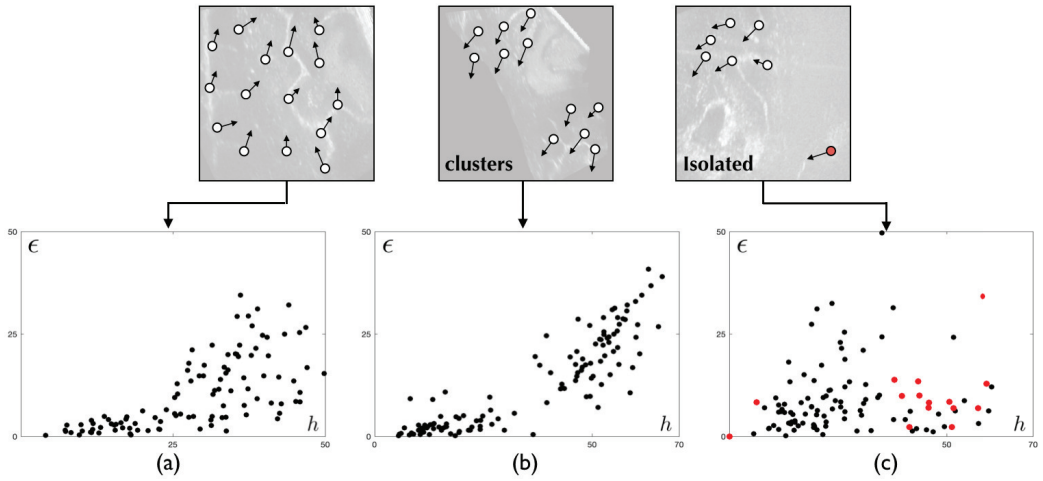


Figure 5.11: (a) $\hat{\gamma}(h)$ of an evenly distributed vector field; (b) $\hat{\gamma}(h)$ of a vector field that has clusters; (c) $\hat{\gamma}(h)$ of a vector field that has an isolated point.

We construct $\hat{\gamma}(h)$ for all data and manually flag cases with the above atypical patterns for further examinations.

5.2.4 Experiments

RESECT (Xiao et al., 2017) and BITE (Mercier et al., 2012) are two high-quality clinical datasets that contain p-MR and i-US images of patients with brain tumors. They are widely used to evaluate registration algorithms for image-guided neurosurgery. RESECT includes 23 patients with each having

p-MR (pMR), before-resection US (bUS), during-resection US (dUS) and after-resection US (aUS) images. Four image pairs, i.e., pMR-aUS, pMR-bUS, bUS-aUS and bUS-dUS, have been annotated with corresponding landmarks. For BITE, pre- and post-operative MR, and i-US images have been acquired from 14 patients. These images were further annotated and put into three groups (1) Group 1: bUS and aUS; (2) Group 2: pMR-bUS; (3) Group 4: pMR and post-MR.

In order to provide an objective, thrid-party screening of the annotations in these two datasets, we generated $\hat{\gamma}(h)$ for all 700+ landmark pairs and flagged those landmark pairs with potential FLE issues. Two operators visually inspected the flagged landmark pairs and together categorized them into three categories: (1) They are certain that the landmark pair is problematic; (2) $\hat{\gamma}(h)$ looks atypical, but they are unsure whether the landmark pair is problematic; (3) $\hat{\gamma}(h)$ looks normal. In addition, they also flagged cases with clusters or isolated landmark $\hat{\gamma}(h)$ problems.

5.2.4.1 Findings

After the *objective screening*, we found that the vast majority of landmarks have normal-looking $\hat{\gamma}(h)$, which indicates that both datasets have high-quality annotations. In total, there are 29 pairs of landmarks that potentially have FLEs. In addition, we also identified 4 cluster cases and 11 isolated landmarks. All flagged data are summarized in Table 1. Figure 5.12 and Figure 5.13 show the $\hat{\gamma}(h)$ of some landmark pairs that were flagged as potentially having FLE's. Figure 5.14 gives two examples of $\hat{\gamma}(h)$'s of a flagged *cluster* and isolated landmarks.

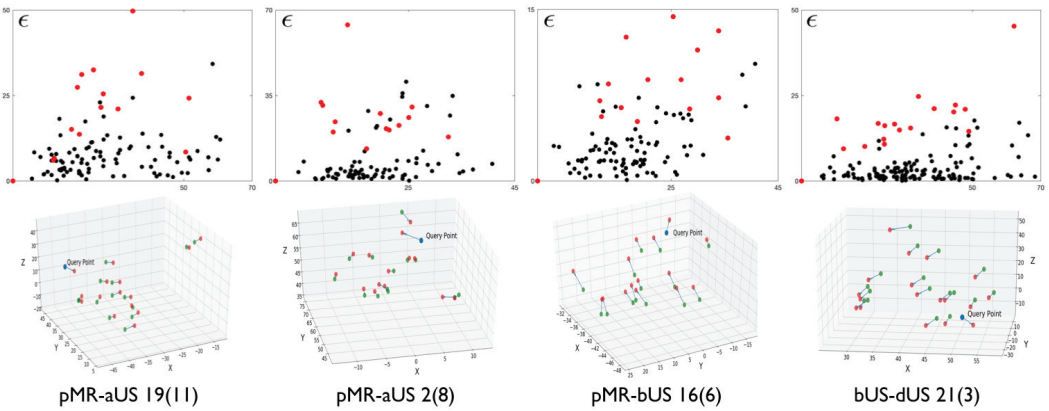


Figure 5.12: Examples of category one atypical landmark pairs (red). The first row shows their $\hat{\gamma}(h)$'s while the second row displays their 3D displacement vectors.

		Certain	Unsure	Cluster	Isolated
R	pMR-aUS	1(9), 2(10) 19(11), 2(8)	3(5), 4(3), 7(1, 4), 14(1),	n/a	18(14)
	pMR-bUS	1(9), 16(6)	1(13), 2(14), 3(1) 15(13), 25(15)	19	2(14), 19(1)
	bUS-aUS	1(7), 7(8) 12(11) , 24(14)	15(3)	25	1(11), 18(12) 19(1)
	bUS-dUS	21(3), 27(11)	6(10), 7(22)	19	n/a
B	G1	3(4), 10(1)	n/a	12	2(3), 4(4) 9(6)
	G2	9(5)	12(1)	n/a	3(21)
	G4	n/a	1(6)	n/a	3(16)

Table 5.2: Indexes of problematic landmark pairs, e.g., 1(9) means patient 1 and landmark pair No.9.

Since the brain may undergo deformation during surgery, atypical behaviors of $\mathbf{d}$'s may indicate actual deformation of the brain. In order to investigate whether "problematic" landmarks contain localization errors, we sent the findings (mixed with good landmarks) to 3 experienced neuro-radiologists for validation and rating. They carefully examined the landmark coordinates in physical space using Slicer (Fedorov et al., 2012) and assigned a score from [1(poor), 2(questionable), 3 (acceptable) , 4 (good)]. Landmarks in Category

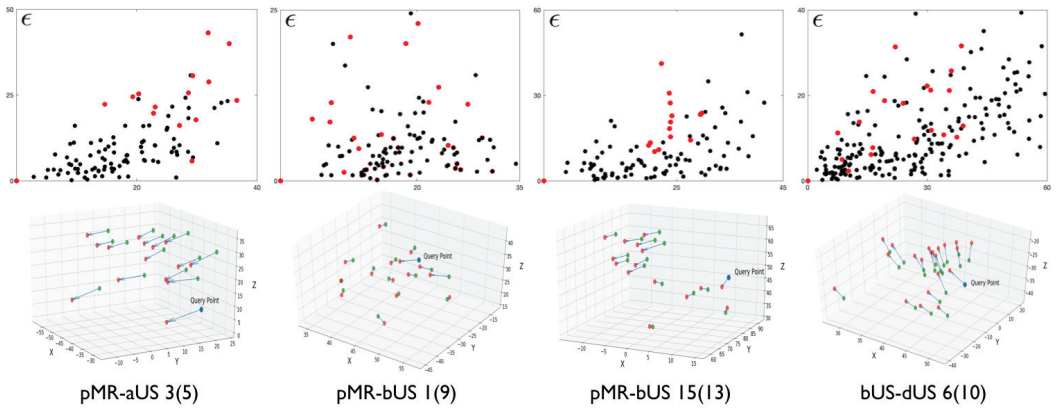


Figure 5.13: Examples of category two landmark pairs. These landmarks have atypical $\hat{\gamma}(h)$'s, but their displacement vectors could be reasonable.

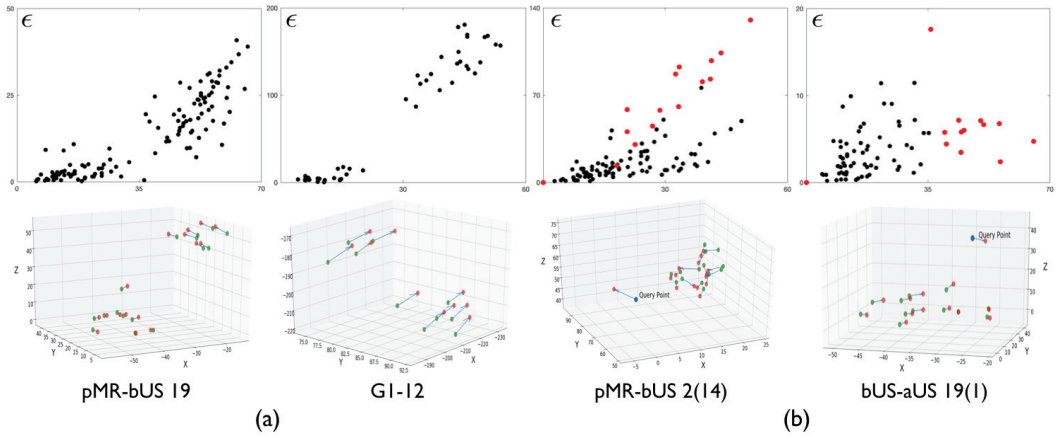


Figure 5.14: Examples of (a) two flagged *clusters* and (b) two isolated landmarks.

one and category two received an average score of 1.5 and 2.4 respectively. Figure 5.15(a) shows the user interface for the validation procedure.

5.2.5 Potential Evaluation Bias

FLEs and non-evenly distributed landmarks can incur bias in the registration evaluation:

1. Since most FRE metrics takes into account all landmarks equally (not weighted), landmarks with FLEs produce false registration error and can drive the algorithm towards aligning those inaccurately located markers.

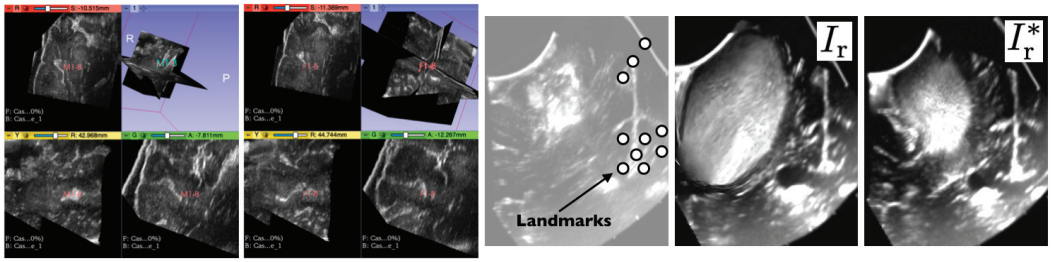


Figure 5.15: (a) The user interface for the landmark validation; (b) Evaluation bias caused by un-evenly distributed landmarks. On the left is an i-US image. I_r and I_r^* are registered p-US's using two different registration methods.

2. *cluster* and isolated landmarks both are not (densely) evenly distributed, thus they incur evaluation bias that prioritizes regions that have landmarks. As in the p-US to i-US registration example shown in Figure 5.15(b), landmarks only exist in the sulcus region (as two clusters). Here, we have registered p-US images I_r and I_r^*, which are registered by two different registration methods. In the landmark-based evaluation, I_r has a better FRE score than I_r^* because it perfectly aligns the sulcus region (while ignoring the rest of image). However, in surgeons' eyes, it is I_r^* which is more reasonable (useful) since it provides accurate tumor boundary alignment.

5.2.6 Discussion

Manual landmark annotation is (mostly) a subjective task, thus public datasets may inevitably have FLEs. To mitigate the evaluation bias caused by FLEs, one strategy is to apply a landmark weighting (selection) scheme (Danilchenko and J.M., 2011; Shamir R.R. and Y., 2012; Thompson et al., 2013). However, at the current stage, these methods are still not thoroughly validated and should be approached with caution. RESECT and BITE are both purposefully created for image-guided neurosurgery. We understood that anatomical features in the brain, mainly corner points and small holes, will most likely not appear uniformly in image space. In order to achieve a non-biased comparison between registration algorithms, dataset providers can add notes to describe

the distribution (limitations) of landmarks so that users, if necessary, can incorporate more criteria, such as surfaces (dos Santos et al., 2014), in the evaluation.

Whether or not public datasets assure unbiased comparisons for registration evaluation is a crucial question that deserves more attention from the image registration community. From this *objective third-party screening*, we conclude that RESECT and BITE are both reliable datasets with a small number of problematic landmarks and landmark distributions that may bias the non-weighted FRE evaluation. Besides the fore-mentioned using advanced evaluation criteria, some data providers offer to update their repository based on users' feedback, e.g., adding or correcting annotations, which is another solution to assure unbiased registration evaluation.

As a tool for *objective screening*, variogram may have limitations. Nevertheless, we believe that this paper will serve as a foundation and draw more attenton to this topic.

Chapter 6

Conclusions and Future Work

This chapter concludes the thesis and discusses potential future research directions.

6.1 Conclusions

This dissertation is devoted to establishing a foundation for utilizing registration uncertainty in the context of neurosurgery. Since the spatial distribution of registration error is particularly difficult to acquire, we propose to use *registration uncertainty* as a surrogate to indicate error in neuronavigation. We develop methods to estimate, investigate and understand registration uncertainty, and validate them retrospectively using clinical data.

The dissertation is based on our publications (Luo et al., 2018a, 2020a,b, 2019a, 2018b). We summarize the contributions below.

- In Chapter 2, we proposed a novel landmark-driven active framework for brain shift compensation. The proposed method is robust for aligning preoperative and intraoperative Ultrasound images with missing correspondences, and is fast enough for clinical applications. Besides, we can obtain registration uncertainty in transformation parameters from the probabilistic registration model. For areas that are difficult to align, users can actively add new landmarks based on the image context and visualization of the uncertainty measure provided by our method to improve the registration accuracy further. To the best of our knowledge, the proposed framework is the first method to actively combine user expertise in brain shift compensation.

- Misinterpreting the registration uncertainty and placing unwarranted confidence in the alignment solutions may result in severe consequences. In Chapter 3, we investigated the applicability of registration uncertainty and categorized it into transformation uncertainty and label uncertainty. We pointed out that using transformation uncertainty to quantify label uncertainty is inappropriate and can be misleading. We also shared a potentially critical finding that making use of the label uncertainty may not always be helpful. How to apply the registration uncertainty is an overlooked problem. We believe that this work will serve as a foundation and draw more attention to this topic.

- A key assumption for using uncertainty to indicate the error is that these two quantities have to be in a monotonic relationship. "Are registration uncertainty and error monotonically associated?" is a fundamental question that has been overlooked by researchers in the medical imaging community. In Chapter 4, we systematically investigate the monotonic association between Gaussian process registration uncertainty and error in the context of Image-guided neurosurgery. We show empirically that there is a weak-to-moderate positive monotonic correlation between point-wise GP uncertainty and non-rigid registration error. At the current stage, our findings indicate that it may not be ideal to apply the GP uncertainty in practice. Nevertheless, this work opens a new avenue for uncertainty/error relationship analysis and may inspire more research on this topic.

- The variogram is a useful tool in geostatistics for capturing the spatial dependence of stochastic processes. However, it has not received much attention in the medical imaging field. In Chapter 5, we introduce two variogram-based related projects. In the first project, we use a variogram in the feature-matching step to remove vector outliers for the proposed active image registration. In the second project, we share our concern, according to findings from variogram-based experiments, about the quality control of some public datasets that are used to evaluate registration methods. By way of this chapter, we also expect researchers

to find the variogram useful in other medical applications that involve motion vectors analyses.

We believe that we have established a foundation for initiating a paradigm shift from error to uncertainty in neuronavigation.

6.2 Future Work

We discuss further extensions, investigations, and other future work related to the dissertation.

Component correlation in the deformation field interpolation The proposed active image registration method uses the Gaussian process (GP) to interpolate a dense deformation field from a sparse set of landmark displacement vectors. The current version interpolates each component (i.e., elements in the x and y directions of a 2D vector) with an independent GP and disregard any component correlation. However, given certain physical properties of tissues, e.g., incompressibility, the independence assumption is not ideal. In future works, it is possible to leverage co-Kriging or multivariate GP interpolation to take into account the interaction among components and improve the registration.

-

User study for the interaction step One key point of the proposed brain shift compensation framework is the "active registration" idea that aims to overcome the limitation of automatic image registration. If users are not satisfied by the initial registration result, they could manually add new corresponding landmarks to drive the algorithm towards better results, guided by the image context and visualization of registration uncertainty. Human and machines have complementary abilities; we believe that an element of simple user interaction should be added to the pipeline for some challenging medical imaging applications. However, in the current version of the work, this step has not been extensively tested. In the future, we will design more

experiments (with clinical and synthetic data), together with user studies, to further validate the "active registration step".

Patch-level analysis of registration uncertainty Our analysis on the applicability of registration uncertainty almost entirely focuses on the voxel level discrepancies. In practice, surgeons typically may not look at these uncertainty maps on such a fine scale and would be much more interested in the pattern across a broad region. In future work, we will extend the work to investigate correlation between the transformation uncertainty and label uncertainty at the level of patches. In addition, we should further study the implication of probabilistic registration.

In depth investigation of the association between registration uncertainty and error The low-to-moderate correlation between point-wise GP uncertainty and error from our findings indicates that it may not be ideal for practical applications. Since the choice of the inference model, hyperparameters, priors, and image context all influence the relationship between uncertainty and error, findings obtained by other probabilistic image registration approaches may yield different results. As a result, searching for a more appropriate PIR framework that provides uncertainty highly correlated with error may be another meaningful future direction,

Approaches to estimate uncertainty for non-probabilistic registration Only probabilistic image registration approaches can estimate registration uncertainty. However, non-probabilistic registration approaches are the choice for the majority of clinical applications. In future work, we plan to develop a universal strategy that estimates uncertainty for non-probabilistic registration, increasing the feasibility of non-rigid registration in interventional guidance, and advancing the state of image-guided therapy.

Beyond neuronavigation Given the trend to develop navigation system relying on unobtrusive imaging modalities such as the intraoperative Ultrasound, methods developed in the dissertation can be not only useful for brain

shift compensation, but also many other clinical applications that require the tracking of tissue deformations.

Bibliography

Agn, M. and Van Leeput, K. (2019). Fast nonparametric mutual-information-based registration and uncertainty estimation. In *Proceedings of the International Workshop on Uncertainty for Safe Utilization of Machine Learning in Medical Imaging*, pages 42–51. Springer.

Arbel, T., Morandi, X., Comeau, R., and Collins, D. (2004). Automatic non-linear mri-ultrasound registration for the correction of intra-operative brain deformations. *Computer Aided Surgery*, 9(4).

Assaf, Y., Johansen-Berg, H., and de Schotten, M. (2017). The role of diffusion mri in neuroscience. *NMR in Biomedicine*, 32(4):e3762.

Baig, M., Klein, J., and Mechtler, L. (2016). Imaging of brain tumors. *Neuroimaging*, 5:1529–1552.

Bajcsy, R. and Kovacic, S. (1989). Multiresolution elastic matching. *Computer Vision, Graphics, and Image Processing*, 46:1–21.

Bayer, S., Maier, A., Ostermeier, M., and Fahrig, R. (2017a). Intraoperative imaging modalities and compensation for brain shift in tumor resection surgery. *International Journal of Biomedical Imaging*, 2017:ID6028645.

Bayer, S., Maier, A., Ostermeier, M., and Fahrig, R. (2017b). Intraoperative imaging modalities and compensation for brain shift in tumor resection surgery. *International Journal of Biomedical Imaging*, 2017:Article ID 6028645.

cancer.net (2020). *Brain Tumor Statistics*. Cancer.net.

C.E., R. and Williams, C. (2006). *Gaussian Processes for Machine Learning*. MIT press.

Chartrain, A., Kellner, C., Fargen, K., Spiotta, A., Chesler, D., Fiorella, D., and Mocco, J. (2017). A review and comparison of three neuronavigation systems for minimally invasive intracerebral hemorrhage evacuation. *Neuroimaging*, 10(1):66–74.

Chilla, G., Tan, C., Xu, C., and Poh, C. (2015). Diffusion weighted magnetic resonance imaging and its recent trend—a survey. *Quantitative Imaging in Medicine and Surgery*, 5(3):2223–4292.

Christensen, G., Geng, X., Kuhl, J., Bruss, J., Grabowski, T., Pirwani, I., Vannier, M., Allen, J., and Damasio, H. (2006). Introduction to the non-rigid image registration evaluation project (nirep). In *Proceedings of the International Workshop on Biomedical Image Registration*, pages 128–135. Springer.

Chui, H. and Rangarajan, A. (2013). A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(2-3):114–141.

Clatz, O., Delingette, H., Talos, I., Golby, A., Kikinis, R., Jolesz, F., Ayache, N., and Warfield, S. (2005). Robust nonrigid registration to capture brain shift from intraoperative mri. *IEEE Transactions on Medical Imaging*, 24(11):1417–1427.

Cobzas, D. and Sen, A. (2011). Random walks for deformable registration. In *Proceedings of the 14th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 557–565. Springer.

Corder, G. and Foreman, D. (2014). *Nonparametric Statistics: A Step-by-Step Approach*. Wiley.

Correa-Arana, K., Vivas-Alban, O., and Sabater-Navarro, J. (2017). Neurosurgery and brain shift: review of the state of the art and main contributions of robotics. *TecnoLogicas*, 20(40):125–138.

Cressie, N. (1991). *Statistics for spatial data*. Wiley.

Dalca, A., Balakrishnan, G., Guttag, J., and Sabuncu, M. (2018). Unsupervised learning for fast probabilistic diffeomorphic registration. In *Proceedings of the 21th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 729–738. Springer.

Danilchenko, A. and J.M., F. (2011). General approach to first-order error prediction in rigid point registration. *IEEE Transactions on Medical Imaging*, 30(3):679–693.

Datteri, R. and Dawant, B. (2012). Automatic detection of the magnitude and spatial location of error in non-rigid registration. In *Proceedings of the International Workshop on Biomedical Image Registration*, pages 21–30. Springer.

dos Santos, T., Seitel, A., Kilgus, T., Suwelack, S., Wekerle, A., Kenngott, H., Speidel, S., Schlemmer, H., Meinzer, H., Heimann, T., and Maier-Hein, L. (2014). Pose-independent surface matching for intra-operative soft-tissue marker-less registration. *Medical Image Analysis*, 18(7):1101–1114.

Drakopoulos, F., Foteinos, P., Liu, Y., and Chrisochoides, N. (2014). Toward a real time multi-tissue adaptive physics-based non-rigid registration framework for brain tumor resection. *Frontiers in Neuroinformatics*, 8(11).

Dumpuri, P., Thompson, R., Dawant, B., Cao, A., and Miga, M. (2007). An atlas-based method to compensate for brain shift: Preliminary results. *Medical Image Analysis*, 11(2):128–145.

Ellingson, B., Bendszus, M., and Boxerman, J. (2015). Consensus recommendations for a standardized brain tumor imaging protocol in clinical trials. *Neuro-Oncology*, 17(9):1188–1198.

Enchev, Y. (2009). Neuronavigation: geneology, reality, and prospects. *Journal of Neurosurgery*, 27(3):Online.

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J., Pieper, S., and R., K. (2012). 3d slicer as an

image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*, 30(9):1323–1341.

Fischler, M. and Bolles, R. (1981). Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Communication of the ACM*, 24(6):381–395.

Fitzpatrick, J. (2007). *The retrospective image registration evaluation project*. Insight-Journal.

Fitzpatrick, J. (2009). Fiducial registration error and target registration error are uncorrelated. In *Proceedings of SPIE Medical Imaging*, page 726102G. SPIE.

Folgoc, L., Delingette, H., Criminisi, A., and Ayache, N. (2017). Quantifying registration uncertainty with sparse bayesian modelling. *IEEE Transactions on Medical Imaging*, 36(2):607–617.

Frisken, S., Luo, M., Juvekar, P., Bunevicius, A., Machado, I., Unadkat, P., Bertotti, M., Toews, M., Wells, W., Miga, M., and Golby, A. (2020). A comparison of thin-plate spline deformation and finite element modeling to compensate for brain shift during tumor resection. *International journal of computer assisted radiology and surgery*, 15(1):75–85.

Fuerst, B., Wein, W., Muller, M., and Navab, N. (2014). Automatic ultrasound-mri registration for neurosurgery using the 2d and 3d lc(2) metric. *Medical Image Analysis*, 18(8):1312–1319.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2004). *Bayesian data analysis*. CRC.

Gerard, I., Kersten-Oertel, M., Petrecca, K., Sirhan, D., Hall, J., and Collins, D. (2017). Brain shift in neuronavigation of brain tumors: A review. *Medical Image Analysis*, 35:403–420.

Glocker, B., Paragios, N., Komodakis, N., Tziritas, G., and Navab, N. (2009). Optical flow estimation with uncertainties through dynamic mrfs. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1063–6919. IEEE.

Gobbi, D., Comeau, R., and Peters, T. (2000). Ultrasound/mri overlay with image warping for neurosurgery. In *Proceedings of the 3rd International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 106–114. Springer.

Goshtasby, A. (2012). *Image Registration: Principles, Tools and Methods*. Springer.

Hata, N., Nabavi, A., Warfield, S., Wells, W., Kikinis, R., and Jolesz, F. (1999). A volumetric optical flow method for measurement of brain deformation from intraoperative magnetic resonance images. In *Proceedings of the 2nd International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 928–935. Springer.

health.harvard.edu (2019). *Brain Tumor Overview*. Harvard Health Publishing.

Heinrich, M., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F., Brady, M., and Schnabel, J. (2012). Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical Image Analysis*, 16(7):1423–1435.

Heinrich, M., Simpson, I., Papiez, B., Brady, M., and Schnabel, J. (2016). Deformable image registration by combining uncertainty estimates from supervoxel belief propagation. *Medical Image Analysis*, 27:57–71.

Hellier, P., Barillot, C., Corouge, L., Gibaud, B., Le Goualher, G., Collins, D., Evans, A., Malandain, G., Ayache, N., Christensen, G., and Johnson, H. (2003). Retrospective evaluation of inter-subject brain registration. *IEEE Transactions on Medical Imaging*, 22(9):1120–1130.

hopkinsmedicine.org (2019). *Brain Tumor: Introduction*. hopkinsmedicine.org.

Hu, Y., Modat, M., Gibson, E., Li, W., and Vercauteren, T. (2019). Weakly-supervised convolutional neural networks for multimodal image registration. *Medical Image Analysis*, 49:1–13.

Hub, M. and Karger, C. (2013). Estimation of the uncertainty of elastic image registration with the demons algorithm. *Physics in medicine and biology*, 58(9):3023–3036.

Hub, M., Kessler, M., and Karger, C. (2009). A stochastic approach to estimate the uncertainty involved in b-spline image registration. *IEEE Transactions on Medical Imaging*, 28(11):1708–1716.

Janoos, F., Risholm, P., and Wells, W. (2012). Bayesian characterization of uncertainty in multi-modal image registration. In *Proceedings of the International Workshop on Machine Learning in Medical Imaging*, pages 50–59. Springer.

Jian, B. and Vemuri, B. (2010). Robust point set registration using gaussian mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1633–1645.

Kabus, S., Klinder, T., Murphy, K., van Ginneken, B., Lorenz, C., and Pluim, J. (2009). Evaluation of 4d-ct lung registration. In *Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 747–754. Springer.

Klein, A., Andersson, J., Ardekani, B., Ashburner, J., Avants, B., Chiang, M., Christensen, G., Collins, D., Gee, J., Hellier, P., Song, J., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R., Mann, J., and R.V., P. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *Neuroimage*, 46(3):786–802.

Klein, A., Ghosh, S., Avants, B., Yeo, B., Fischl, B., Ardekani, B., Gee, J., Mann, J., and R.V., P. (2010). Evaluation of volume-based and surface-based brain image registration methods. *Neuroimage*, 51(1):214–220.

Krishnatry, R., Zhukova, N., de Guerreiro-Stucklin, A., and Tabori, U. (2016). Clinical and treatment factors determining long-term outcomes for adult

survivors of childhood low- grade glioma: a population-based study. *Cancer*, 122(8):1261–1269.

Kybic, J. (2010). Bootstrap resampling for image registration uncertainty estimation without ground truth. *IEEE Transactions on Image Processing*, 19(1):64–73.

Lasso, A., Heffter, T., Rankin, A., Pinter, C., Ungi, T., and Fichtinger, G. (2014). Plus: open-source toolkit for ultrasound-guided intervention systems. *IEEE Transactions on Biomedical Engineering*, 61(10):2527–2537.

Letteboer, M., Willems, P., Viergever, M., and Niessen, W. (2003). Non-rigid registration of 3d ultrasound images of brain tumours acquired during neurosurgery. In *Proceedings of the 6th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 408–415. Springer.

Lotfi, T., Tang, L., Andrews, S., and Hamarneh, G. (2013). Improving probabilistic image registration via reinforcement learning and uncertainty evaluation. In *Proceedings of the International Workshop on Machine Learning in Medical Imaging*, pages 187–194. Springer.

Luo, J., Frisken, S., Machado, I., Zhang, M., Pieper, S., Golland, P., Toews, M., Unadkat, P., Sedghi, A., Zhou, H., Mehrtash, A., Preiswerk, F., Cheng, C., Golby, A., Sugiyama, M., and Wells, W. (2018a). Using the variogram for vector outlier screening: application to feature-based image registration. *International Journal of Computer Assisted Radiology and Surgery*, 13(12):1871–1880.

Luo, J., Frisken, S., Wang, D., Golby, A., Sugiyama, M., and Wells, W. (2020a). Are registration uncertainty and error monotonically associated? In *Proceedings of the 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention*, page tbd. Springer.

Luo, J., Ma, G., Frisken, S., Juvekar, P., Haouchine, N., Xu, Z., Xiao, Y., Golby, A., Codd, P., Sugiyama, M., and Wells, W. (2020b). Do public datasets assure unbiased comparisons for registration evaluation? In *arXiv*, page 2003.09483. arXiv.

Luo, J., Sedghi, A., Popuri, K., Cobzas, D., Zhang, M., Preiswerk, F., Toews, M., Golby, A., Sugiyama, M., Wells, W., and Frisken, S. (2019a). On the applicability of registration uncertainty. In *Proceedings of the 22th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 107–115. Springer.

Luo, J., Toews, M., Machado, I., Frisken, S., Zhang, M., Preiswerk, F., Sedghi, A., Ding, H., Pieper, S., Golland, P., Golby, A., Sugiyama, M., and Wells, W. (2018b). A feature-driven active framework for ultrasound-based brain shift compensation. In *Proceedings of the 21th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 30–38. Springer.

Luo, M., Frisken, S., Weis, J., Clements, L., Unadkat, P., Thompson, R., Golby, A., and Miga, M. (2017). Retrospective study comparing model-based deformation correction to intraoperative magnetic resonance imaging for image-guided neurosurgery. *Journal of Medical Imaging*, 4(3):035003.

Luo, M., Larson, P., Martin, A., Konrad, P., and Miga, M. (2019b). An integrated multi-physics finite element modeling framework for deep brain stimulation: Preliminary study on impact of brain shift on neuronal pathways. In *Proceedings of the 22th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 682–690. Springer.

Machado, I., Toews, M., George, E., Unadkat, P., Essayed, W., Luo, J., Teodoro, P., Carvalho, H., Martins, J., Golland, P., Pieper, S., Frisken, S., Golby, A., Wells, W., and Ou, Y. (2019). Deformable mri-ultrasound registration using correlation-based attribute matching for brain shift correction: Accuracy and generality in multi-site data. *Neuroimage*, 202(15):116094.

Machado, I., Toews, M., Luo, J., Unadkat, P., Essayed, W., George, E., Teodoro, P., Carvalho, H., Martins, J., Golland, P., Pieper, S., Frisken, S., Golby, A., and Wells, W. (2018). Non-rigid registration of 3d ultrasound for neurosurgery using automatic feature detection and matching. *International Journal of Computer Assisted Radiology and Surgery*, 13(10):1525–1538.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997). Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198.

Maier, S., Bogner, P., Bajzik, G., Mamata, H., Mamata, Y., Repa, I., Jolesz, F., and Mulkern, R. (2000). Normal brain and brain tumor: Multicomponent apparent diffusion coefficient line scan imaging. *Radiology*, 219:842–849.

Maier, S., Sun, Y., and Mulkern, R. (2010). Diffusion imaging of brain tumors. *NMR in Biomedicine*, 23(7):849–864.

Maintz, J. and Viergever, M. (1998). A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58:1246–1266.

Menze, B., Jakab, A., Bauer, S., and Geremia, F. (2015). The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024.

Mercier, L., Del Maestro, R., Petrecca, K., Araujo, D., Haegelen, C., and Collins, D. (2012). Online database of clinical mr and ultrasound images of brain tumors. *Medical Physics*, 39(6):3253–3261.

Mercier, L., Del Maestro, R., Petrecca, K., Kochanowska, A., Drouin, S., Yan, C., Janke, A., Chen, S., and Collins, D. (2010). New prototype neuronavigation system based on preoperative imaging and intraoperative freehand ultrasound: system description and validation. *International Journal of Computer Assisted Radiology and Surgery*, 6:507–522.

Miga, M. (2016). Computational modeling for enhancing soft tissue image guided surgery: An application in neurosurgery. *Annals of Biomedical Engineering*, 44(1):128–138.

Miller, K. (2019). *Biomechanics of the brain*. Springer.

Min, Z., Ren, H., and Meng, M.-H. (2020). Statistical model of total target registration error in image-guided surgery. *IEEE Transactions on Automation Science and Engineering*, 17(1):151–165.

Morin, F., Courtecuisse, H., Reinertsen, I., Lann, F., Palombi, O., Payan, Y., and Chabanas, M. (2017). Brain-shift compensation using intraoperative ultrasound and constraint-based biomechanical simulation. *Medical Image Analysis*, 40:133–153.

Muacevic, A., Uhl, E., Steiger, H., and H.J., R. (2000). Accuracy and clinical applicability of a passive marker based frameless neuronavigation system. *Journal of Clinical Neuroscience*, 7(5):414–418.

Murphy, K., Van Ginneken, B., Reinhardt, J., Kabus, S., Ding, K., Deng, X., Cao, K., Du, K., Christensen, G., and Garcia, V. (2011). Evaluation of registration methods on thoracic ct: the empire10 challenge. *IEEE Transactions on Medical Imaging*, 30(11):1901–1920.

Myronenko, A. and Song, X. (2010). Point set registration: Coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2262–2275.

Oliveira, F. and Tavares, J. (2014). Medical image registration: A review. *Computer Methods in Biomechanics and Biomedical Engineering*, 17(2):73–93.

Orringer, D., Golby, A., and Jolesz, F. (2012). Neuronavigation in the surgical management of brain tumors: current and future trends. *Expert Rev Med Devices*, 9(5):491–500.

Ou, Y., Sotiras, A., Paragios, N., and Davatzikos, C. (2011). Dramms: Deformable registration via attribute matching and mutual-saliency weighting. *Medical Image Analysis*, 15(4):622–639.

Pennec, X., Cachier, P., and Ayache, N. (2003). Tracking brain deformations in time sequences of 3d us images. *Pattern Recognition Letters*, 24(4):801–813.

Pluim, J., Maintz, J., and Viergever, M. (2003). Mutual-information-based registration of medical images: A survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004.

Popuri, K., Cobzas, D., and Jagersand, M. (2013). A variational formulation for discrete registration. In *Proceedings of the 16th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 187–194. Springer.

Reinertsen, I., Descoteaux, M., Drouin, S., Siddiqi, K., and Collins, D. (2004). Vessel driven correction of brain shift. In *Proceedings of the 7th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 208–216. Springer.

Reinertsen, I., Lindseth, F., Unsgaard, G., and Collins, D. (2007). Clinical validation of vessel-based registration for correction of brain-shift. *Medical Image Analysis*, 11(6):673–684.

Risholm, P., Janoos, F., Norton, I., Golby, A., and W.M., W. (2013). Bayesian characterization of uncertainty in intra-subject non-rigid registration. *Medical Image Analysis*, 17(5):538–555.

Risholm, P., Pieper, S., Samset, E., and Wells, W. (2010). Summarizing and visualizing uncertainty in non-rigid registration. In *Proceedings of the 13th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 554–561. Springer.

Risholm, P. and Wells, W. (2011). Estimation of delivered dose in radiotherapy: the influence of registration uncertainty. In *Proceedings of the 14th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 548–557. Springer.

Riva, M., Hennersperger, C., Milletari, F., Katouzian, A., Pessina, F., Gutierrez-Becker, B., Castellano, A., Navab, N., and Bello, L. (2017). 3d intra-operative ultrasound and mr image guidance: pursuing an ultrasound-based management of brainshift to enhance neuronavigation. *International Journal of Computer Assisted Radiology and Surgery*, 12(10):1711–1725.

Rivaz, H. and Collins, D. (2015). Deformable registration of preoperative mr, pre-resection ultrasound, and post-resection ultrasound images of neurosurgery. *International Journal of Computer Assisted Radiology and Surgery*, 10(7):1017–1028.

Robb, R. (2009). *Handbook of Medical Image Processing and Analysis*. Elsevier.

Rohlfing, T. (2012). Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable. *IEEE Transactions on Medical Imaging*, 31(2):153–163.

Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., and Hawkes, D. (1999). Non-rigid registration using free-form deformation: application to breast mr images. *IEEE Transactions on Medical Imaging*, 18(8):712–721.

Ruiz-Alzola, J., Suarez, E., Alberola-Lopez, C., Warfield, S., and Westin, C.-F. (2003). Geostatistical medical image registration. In *Proceedings of the 6th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 894–901. Springer.

Saygill, G. (2018). Local-search based prediction of medical image registration error. In *Proceedings of SPIE medical imaging*, page 105771F. SPIE.

Saygill, G., Staring, M., and Hendriks, E. (2016). Confidence estimation for medical image registration based on stereo confidences. *IEEE Transactions on Medical Imaging*, 35(2):539–549.

Schaefer, P., Grant, P., and Gonzalez, R. (2000). Diffusion-weighted mr imaging of the brain. *Radiology*, 217(2):331–345.

Schlachter, M., Fechter, T., Jurisic, M., Schimek-Jasch, T., Oehlke, O., Adebahr, S., Birkfellner, W., Nestle, U., and Buhler, K. (2016). Visualization of deformable image registration quality using local image dissimilarity. *IEEE Transactions on Medical Imaging*, 35(10):2319–2328.

Sedghi, A., Luo, J., Mehrtash, A., Pieper, S., Tempany, C., Kapur, T., Mousavi, P., and Wells, W. (2018). Semi-supervised image registration using deep learning. In *Proceedings of SPIE Medical Imaging*, page 109511G. SPIE.

Seeger, W. and Zentner, J. (2002). *Neuronavigation and Neuroanatomy*. Springer.

Shamir R.R., Joskowicz, L. and Y., S. (2012). Fiducial optimization for minimal target registration error in image-guided neurosurgery. *IEEE Transactions on Medical Imaging*, 31(3):725–737.

Shams, R., Xiao, Y., Hebert, F., Abramowitz, M., Brooks, R., and Rivaz, H. (2017). Assessment of rigid registration quality measures in ultrasound-guided radiotherapy. *IEEE Transactions on Medical Imaging*, 37(2):428–437.

Shen, D. and Davatzikos, C. (2002). Hammer: Hierarchical attribute matching mechanism for elastic registration. *IEEE Transactions on Medical Imaging*, 21(11):1421–1439.

Simonovsky, M., Gutierrez-Becker, B., Mateus, D., Navab, N., and Komodakis, N. (1998). A deep metric for multimodal registration. In *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 641–649. Springer.

Simpson, I., Cardoso, M., Modat, M., Cash, D., Woolrich, M., Andersson, J., Schnabel, J., and Ourselin, S. (2015). Probabilistic non-linear registration with spatially adaptive regularisation. *Medical Image Analysis*, 26(1):203–216.

Simpson, I., Schnabel, J., Groves, A., Andersson, J., and Woolrich, M. (2012). Probabilistic inference of regularisation in non-rigid registration. *NeuroImage*, 59(3):2438–2451.

Simpson, I., Woolrich, M., Andersson, J., Groves, A., and Schnabel, J. (2013). Ensemble learning incorporating uncertain registration. *IEEE Transactions on Medical Imaging*, 32(4):748–756.

Skrinjar, O., Spencer, D., and Duncan, J. (1998). Brain shift modeling for use in neurosurgery. In *Proceedings of the 1st International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 641–649. Springer.

Slavin, K. (2008). Neuronavigation in neurosurgery: current state of affairs. *Expert Rev Med Devices*, 5(1):1–3.

Sokooti, H., Saygili, G., Glocker, B., Lelieveldt, B., and Staring, M. (2016). Accuracy estimation for medical image registration using regression forests. In *Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 107–115. Springer.

Sokooti, H., Saygili, G., Glocker, B., Lelieveldt, B., and Staring, M. (2019). Quantitative error prediction of medical image registration using regression forests. *Medical Image Analysis*, 56:110–121.

Song, J. (2017). *Methods for evaluating image registration*. UIOWA.

Sonka, M. (2000). *Handbook of Medical Imaging: Medical image processing and analysis*. SPIE.

Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190.

Thompson, S., Penney, G., Dasgupta, P., and Hawkes, D. (2013). Improved modelling of tool tracking errors by modelling dependent marker errors. *IEEE Transactions on Medical Imaging*, 32(2):165–177.

Toews, M. and Wells, W. (2013). Efficient and robust model-to-image alignment using 3d scale-invariant features. *Medical Image Analysis*, 17(3):271–282.

Viergever, M., Maintz, J., Klein, S., Murphy, K., Staring, M., and Pluim, J. (2016). A survey of medical image registration. *Medical Image Analysis*, 33:140–144.

Vigneron, L., Noels, L., Warfield, S., Verly, J., and Robe, P. (2012). Serial fem/xfem-based update of preoperative brain images using intraoperative mri. *International Journal of Biomedical Imaging*, 2012:Article ID 872783.

Villanueva-Meyer, J., Mabray, M., and Cha, S. (2017). Current clinical brain tumor imaging. *Neurosurgery*, 81(3):397–415.

Wang, J., Wells, W., Golland, P., and Zhang, M. (2018). Efficient laplace approximation for bayesian registration uncertainty quantification. In *Proceedings of the 21th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 880–888. Springer.

Wassermann, D., Toews, M., Niethammer, M., and Wells, W. (2014). Probabilistic diffeomorphic registration: Representing uncertainty. In *Proceedings of the International Workshop on Biomedical Image Registration*, pages 72–82. Springer.

Wells, W. M., Viola, P., Atsumi, H., Nakajima, S., and Kikinis, R. (1996). Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51.

West, J., Fitzpatrick, J., Wang, M., and Woods, R. (1997). Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Computer Assisted Tomography*, 21(4):554–566.

Wikipedia (2020a). *Brain Tumor*. Wikipedia.

Wikipedia (2020b). *Glioma*. Wikipedia.

Wikipedia (2020c). *Neuronavigation*. Wikipedia.

Willems, P., van der Sprenkel, J., Tulleken, C., Viergever, M., and Taphoorn, M. (2006). Neuronavigation and surgery of intracerebral tumours. *Journal of Neurology*, 253:1123–1136.

Xiao, Y., Fortin, M., Unsgard, G., Rivaz, H., and Reinertsen, I. (2017). Retrospective evaluation of cerebral tumors (resect): A clinical database of pre-operative mri and intra-operative ultrasound in low-grade glioma surgeries. *Medical Physics*, 44(7):3875–3882.

Xiao, Y., Rivaz, H., and Chabanas, M. (2020). Evaluation of mri to ultrasound registration methods for brain shift correction: The curious2018 challenge. *IEEE Transactions on Medical Imaging*, 39(3):777–786.

Xu, Z., Lee, C., Heinrich, M., Modat, M., Rueckert, D., Ourselin, S., Abrasmson, R., and Landman, B. (2016). Evaluation of six registration methods for the human abdomen on clinically acquired ct. *IEEE Transactions on Biomedical Engineering*, 63(8):1563–1572.

Yang, X. and Niethammer, M. (2015). Uncertainty quantification for lddmm using a low-rank hessian approximation. In *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 289–296. Springer.

Yassa, M. and C.E., S. (2009). A quantitative evaluation of cross-participant registration techniques for mri studies of the medial temporal lobe. *Neuroimage*, 44(2):319–327.

Zhang, M., Singh, N., and Fletcher, T. (2013). Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In *Proceedings of International Conference on Information Processing in Medical Imaging*, pages 37–48. Springer.

Zhuang, D., Liu, Y., Wu, J., Yao, C., Mao, Y., Zhang, C., Wang, M., Wang, W., and Zhou, L. (2010). A sparse intraoperative data-driven biomechanical model to compensate for brain shift during neuronavigation. *American Journal of Neuroradiology*, 32(2):395–402.

List of Figures

List of Tables