Deformable MRI-Ultrasound registration using correlation-based attribute matching for brain shift correction: Accuracy and generality in multi-site data

HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2020 November 15.

Published in final edited form as:

Neuroimage. 2019 November 15; 202: 116094. doi:10.1016/j.neuroimage.2019.116094.

Deformable MRI-Ultrasound registration using correlation-based attribute matching for brain shift correction: accuracy and generality in multi-site data

Inês Machado[1,2,*], **Matthew Toews**[3], **Elizabeth George**[1], **Prashin Unadkat**[4], **Walid Essayed**[4], **Jie Luo**[5], **Pedro Teodoro**[2], **Herculano Carvalho**[6], **Jorge Martins**[2], **Polina Golland**[7], **Steve Pieper**[1,8], **Sarah Frisken**[1], **Alexandra Golby**[5], **William Wells III**[1,7], **Yangming Ou**[9,*]

[1]Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

[2]Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

[3]Department of Systems Engineering, École de Technologie Supérieure, Montreal, Canada

[4]Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

[5]Graduate School of Frontier Sciences, University of Tokyo, Tokyo, Japan

[6]Department of Neurosurgery, Hospital de Santa Maria, CHLN, Lisbon, Portugal

[7]Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA

[8]Isomics, Inc., Cambridge, MA, USA

[9]Department of Pediatrics and Radiology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

Abstract

Intraoperative tissue deformation, known as brain shift, decreases the benefit of using preoperative images to guide neurosurgery. Non-rigid registration of preoperative magnetic resonance (MR) to intraoperative ultrasound (US) has been proposed as a means to compensate for brain shift. We present a new method that builds on previous work for addressing the need for accuracy and generality in multi-site clinical datasets. We use high-dimensional texture attributes instead of image intensities for image registration. We propose replacing the standard difference-based attribute matching with correlation-based attribute matching and present a strategy that deals explicitly with the large field-of-view mismatch between MR and US images. We optimize key parameters across independent MR-US brain tumor datasets acquired at three institutions, with a

[*]Corresponding Authors: Inês Machado; ines7.prata.machado@gmail.com; Av. Rovisco Pais, 1,1049-001, Lisboa, Portugal. Yangming Ou; yangming.ou@childrens.harvard.edu; 401 Park Drive, Boston, MA, 02115, USA.

Conflict of interest. The authors declare that there are no conflicts of interest.

Compliance with Ethical Standards. All procedures performed in these studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committees and with the 1964 Declaration of Helsinki and its amendments or comparable ethical standards.

Informed consent. Informed consent was obtained from all participants included in the study.

total of 43 tumor patients and 758 reference landmarks for evaluating accuracy. Despite differences in imaging protocols, patient demographics and landmark distributions, our algorithm was able to reduce landmark errors prior to registration in three data sets (5.37±4.27, 4.18±1.97 and 6.18±3.38 mm, respectively) to a consistently low level (2.28±0.71, 2.08±0.37 and 2.24±0.78 mm, respectively). Our method is competitive with the state-of-the-art on multiple datasets. It was tested with 15 other methods. We show that our method has one of the lowest errors in all datasets (accuracy), and this is achieved while sticking to a fixed set of parameters for multi-site data (generality). In contrast, other algorithms/tools of similar performance need per-dataset parameter tuning (high accuracy but lower generality), and those that stick to fixed parameters have larger errors or inconsistent performance (generality but not the top accuracy). We further characterized landmark errors by brain regions and tumor types, which is so far missing in the literature. We found that landmark errors were higher in high-grade than low-grade glioma patients, and higher in tumor regions than in other brain regions.

Keywords

Brain Shift; Intraoperative Ultrasound; MR-US registration; Multi-site Data

1. Introduction

The proximity of many tumors to critical areas of the brain coupled with the difficulty of differentiating tumor tissue from normal brain parenchyma based on visual inspection renders complete tumor removal a very challenging task (Dimaio et al., 2006). Surgical navigation allows for the use of preoperative images as a map to guide surgery (Golby, 2015). However, when the brain has shifted, MR images may not serve as an accurate guide (Gerard et al., 2017). To compensate for brain shift, non-invasive, high-resolution, but time-consuming MR image acquisitions can be registered to near-real-time intraoperative ultrasound (US) images (Unsgaard et al., 2005)(Rygh et al., 2008)(Miller, Benes, & Sure, 2011)(Coupé, Hellier, Morandi, & Barillot, 2007). In this paper, we focus on the deformable registration of preoperative MR and intraoperative US (prior to the tumor resection).

The primary focus of our work is the accuracy of MR-US image registration. We propose using a new correlation-based similarity measure on a rich set of attributes extracted at each voxel to increase the accuracy in aligning MR-US images. It is different from feature-based methods that attempt to identify and match key features in the images such as landmarks in tumor boundaries, sulci, ventricles or blood vessels (Porter et al., 2001)(Reinertsen, Lindseth, Unsgaard, & Collins, 2007)(Farnia et al., 2018)(Coupé, Hellier, Morandi, & Barillot, 2012). Our method utilizes all voxels in the images because localizing key features itself is difficult and may be subject to errors, also because features existing in one image may not always appear in another image, and features are not always close to the regions of clinical interest (e.g., tumor and surroundings) (Sotiras, Davatzikos, & Paragios, 2013). Among voxel-wise methods, approaches either use single scalar intensities or high-dimensional attributes at voxels to find across-image correspondences. Intensity-based methods use similarity measures such as Mutual Information (MI), Correlation Coefficient (CC) and Correlation Ratio (CR) for MR-US registration (Rivaz, Karimaghaloo, & Collins,

2014)(Hartov, Roberts, & Paulsen, 2008)(Roche et al., 2000)(Letteboer, Viergever, & Niessen, 2003)(Ji, Wu, Hartov, Roberts, & Paulsen, 2008)(Rivaz, Chen, & Collins, 2015) (Rivaz & Collins, 2015)(Myronenko & Song, 2010)(Fuerst, Wein, Müller, & Navab, 2014). Intensities of MR and US images do not always follow a consistent or even linear relationship.

An example is in Figure 1, where we note corresponding structures in MR and US images and number them from 1 to 9. Bright regions in MR can correspond to bright intensities in the US image, e.g., regions 2 and 5; similarly, dark regions corresponding to dark intensities, e.g., regions 7 and 8., However, bright regions can also correspond to dark intensities, e.g., regions 1, 4, 6 and 9, or, grey regions correspond to dark intensities, e.g., region 3. The inconsistent intensity relationship and the measurement of voxel-wise similarity on high-dimensional attributes in patches centered at those voxels have been studied, e.g., (Shen & Davatzikos, 2002)(Ou, Akbari, Bilello, Da, & Davatzikos, 2014)(Toews & Wells, 2013) (Wachinger & Navab, 2013)(Wu, Yap, Kim, & Shen, 2010)(Toews & Arbel, 2003). Self-Similarity Correlation (SSC) (Heinrich, Jenkinson, Papiez, Brady, & Schnabel, 2013) and Linear Correlation of Linear Combination (LC2) methods (Wein et al., 2013) showed increased accuracy in MR-US registration tasks.

We recently developed an attribute-based method, Deformable Registration via Attribute Matching and Mutual-Saliency Weighting (DRAMMS) (Ou, Sotiras, Paragios, & Davatzikos, 2011) that has been validated in multi-site clinical images of brain MRI to MRI registrations across subjects and across time (Diez et al., 2014)(Zhan & Shen, 2006). In contrast to SSC and LC2, which extract attributes from a patch at single scale, DRAMMS uses multi-scale and multi-orientation textures as attributes, arguing that voxels at different brain regions may be distinctive at different scales and orientations. The current study proposes to use the Normalized Correlation Coefficient (aNCC) and Correlation Ratio (aCR) on attribute vectors instead of the original DRAMMS version's Sum of Squared Differences (aSSD) of attribute vectors. aNCC can model linear relationship of attributes and aCR can handle non-linear correlation of attributes. Both are correlation-based similarity measures previously used on image intensities (Roche, Malandain, Pennec, & Ayache, 2006)(Heinrich et al., 2012) but not on attribute vectors. Therefore, we refer to the modified version as correlation-DRAMMS or cDRAMMS.

In addition to using a new similarity measure, we propose a transformation strategy to explicitly handle the field-of-view (FOV) mismatch between the two images. As the dashed yellow circles in Figure 1 show, a large proportion of the content in MR is not included in the US image. We mask the MR image based on the FOV of the US image plus a buffer zone. Recent MR-US registration studies advocate masking the MR image (Rivaz & Collins, 2015)(Heinrich et al., 2013)(Rivaz, Karimaghaloo, Fonov, & Collins, 2014)(Jiang, Shi, Yao, Wang, & Song, 2016), one suggesting a 3 mm radius buffer zone (Drobny, Vercauteren, Ourselin, & Modat, 2018). Our study involves greater detail of the effects of MR masking, the optimal radius of the buffer zone, and the applicability of these approaches when working with multi-site data.

Our second focus is on the generality of the proposed method. Our goal is to achieve consistently high accuracies in multi-site data, without per-dataset parameter tuning. We propose three strategies for this. First, we inherit the use of multi-scale and multi-orientation attributes in the original DRAMMS framework. Single-scale and single-orientation attributes may not always generalize well to other datasets (Wein, 2018)(Heinrich, 2018). Second, we skip ad-hoc pre-processing. Pre-processing is needed in many MR-US registration studies. Some methods require skull stripping (Farnia, Ahmadian, Shabanian, Serej, & Alirezaie, 2014)(Farnia, Makkiabadi, Ahmadian, & Alirezaie, 2016) or tissue segmentation (Hong & Park, 2018)(Morin et al., 2017)(Palombi et al., 2018)(Reinertsen, Lindseth, Askeland, Iversen, & Unsgård, 2014) of MR images, or segmentation and removal of bright strips in skin surfaces (Wein, 2018)(Shams, Boucher, & Kadoury, 2018). These pre-processing steps are non-trivial, error-prone, and often require case-specific human intervention in tumor-bearing MR images (Drobny et al., 2018)(Farnia et al., 2014)(Farnia et al., 2016). Third, we optimize the key parameters, seeking a fixed set of parameter settings that is overall optimal in multi-site data.

A third contribution of our work is thorough evaluation. Most existing methods have been evaluated with only single-site data (Farnia et al., 2014)(Farnia et al., 2016)(Lindseth et al., 2003)(Farnia, Ahmadian, Shabanian, Serej, & Alirezaie, 2015), or up to two datasets but resulted in inconsistent levels of accuracy (Hong et al., 2018)(Shams, Boucher, & Kadoury, 2018)(Zhong et al., 2018), or used dataset-specific parameters (Wein, 2018)(Heinrich, 2018). We used three datasets, so far the most comprehensive multi-site data, to evaluate MR-US registration. The three datasets are: BITE (Brain Images of Tumors for Evaluation, released in 2012) (Mercier et al., 2012), RESECT (REtroSpective Evolution of Cerebral Tumors, released in 2017) (Xiao, Fortin, Unsgärd, Rivaz, & Reinertsen, 2017), and MIBS (Multimodal Imaging of Brain Shift, a proprietary dataset) (Machado et al., 2018). They included a total of 43 patients and 758 reference landmarks for accuracy evaluation. Table 1 displays the multi-site data covering a wide variety of imaging protocols (e.g., scanner vendor, field-of-strength, resolution, contrast), patient demographics (age, gender, tumor type, grade, location, extent, scope) evaluation references (varying landmark locations and distributions), and levels of brain shift (0 to 21 mm). The diversity in data helps characterize the generality and accuracy of totally sixteen methods – the largest number of methods ever being evaluated for registering MR-US images of brain tumor patients.

In addition to evaluating on multi-site data, we provide the first quantitative evaluation of registration error according to brain regions and tumor types. A radiologist physician, with 3.5 years of experience in ultrasound, reviewed the 758 reference landmark pairs one by one, and assigned them to five regions: (i) sulci/gyri, (ii) ventricles, (iii) falx and tentorium, (iv) tumor boundary and (v) others, including vessels, mid brain, pons, bone, white matter and interpeduncular cistern, where landmarks appear less frequently. We found that registration errors increased in this order among the first four regions, and the trend is consistent across datasets. We also found that registration errors are roughly the same in patients with low-grade gliomas and patients with metastatic brain tumors, but larger in patients with high-grade gliomas. Errors by regions and tumor types provide quantitative reference for neurosurgeons.

This study is an extension of our previous workshop version (Machado et al., 2018), with a more detailed algorithm description, additional algorithm optimization, comprehensive validation in multi-site data, and quantification of errors according to brain regions and tumor types.

2. Multi-Site Data

Table 1 shows the three datasets used in this study: BITE (Mercier et al., 2012), RESECT (Xiao et al., 2017), and MIBS (Machado et al., 2018)(Luo et al., 2018)(Tempany et al., 2015). As can be seen in the table, there is a wide variety of (a) imaging sites, (b) patient demographics, (c) MRI and US protocols, and (d) landmark distributions.

From the landmark pairs in these datasets, a radiologist (E.G.) with a total of 4 years of experience and 1 year with intraoperative ultrasound of the brain, visually examined landmarks one by one and assigned them into five regions (sulci/gyri, tumor Boundary, falx and tentorium, and others). The assignment is in the last five rows of Table 1. It allows us to quantify registration accuracy by regions.

3. Methods

Section 3.1 presents a summary of the original DRAMMS framework. Sections 3.1, 3.2 and 3.3 present the three key components in our algorithm: 3.2, the new similarity measure; 3.3, the new transformation strategy; and 3.4, the parameter optimization for generality in multi-site data.

3.1 Original DRAMMS Framework Revisited

The original DRAMMS pipeline is shown in Figure 2. Given two images $I_1{:}\Omega_1 \rightarrow \mathbb{R}$ and $I_2{:}\Omega_2 \rightarrow \mathbb{R}$ in 3D image domains $\Omega_i\,(i = 1,2) \subset \mathbb{R}^3$, DRAMMS seeks a transformation T that maps every voxel $\boldsymbol{u} \in \Omega_1$ to its corresponding point $T(\boldsymbol{u}) \in \Omega_2$, by minimizing a cost function $E(T)$,

$$\min_T E(T) = \int_{\boldsymbol{u} \in \Omega_1} \underbrace{ms(\boldsymbol{u}, T(\boldsymbol{u}))}_{\text{Mutual-Saliency}} \cdot \underbrace{sim\big(A_1^*(\boldsymbol{u}), A_2^*(T(\boldsymbol{u}))\big)d\boldsymbol{u}}_{\text{Attribute Matching}} + \lambda R(T), \tag{1}$$

where $A_i^*(\boldsymbol{u})\ (i = 1, 2)$ is an attribute vector that reflects the geometric texture of voxel $\boldsymbol{u}$ at multiple scales and orientations (four scales: at 3, 5, 9, 17mm neighborhoods (Ou et al., 2011) and six orientations, equally sampling 0 to π in three orthogonal planes at a voxel); $sim\,(\cdot,\cdot)$ measures the similarity of two voxels based on their attribute vectors. The default is the sum of squared differences of attributes (aSSD), which we modify in this paper.

The term $ms(\boldsymbol{u}, T(\boldsymbol{u}))$ is a continuously-valued mutual-saliency weight quantifying the confidence of two voxels $\boldsymbol{u} \in \Omega_1$ and $T(\boldsymbol{u}) \in \Omega_2$ being a correspondence. Voxels with ambiguous or inexistent correspondence (e.g., outlier regions) are assigned with a smaller mutual-saliency weight and thus down weighted in registration optimization, thereby reducing the impact of missing correspondences. The term $R(T)$ is a smoothness/regularization term using the Laplacian operator (square of Laplacian), also known as the

bending energy, of the deformation field T(Bookstein, 1989). The weight, λ (default at 0.2 (Ou et al., 2011)), is greater if a smoother deformation is needed. We evaluate the sensitivity of this parameter (λ) in Section 5.2.3.

The original DRAMMS framework uses the diffeomorphic Free Form Deformation (FFD) (Rueckert, 1999)(Heckemann et al., 2006) as its transformation model and the discrete optimization (Komodakis, Tziritas, & Paragios, 2008)(Glocker et al., 2008) strategy to optimize the deformation in the FFD model.

We follow the original DRAMMS framework (Ou et al., 2011) for attribute extraction, normalization and selection. Briefly, attributes are extracted as described above under Equation 1. Attributes are then normalized into [0, 256] by linearly mapping from attribute values from the minimum and maximum Gabor responses throughout the image across scales and orientations. Given a similarity metric (regardless of the original or the proposed similarities), attributes are automatically selected by maximizing similarity and reliability of matching (mutual-saliency) using forward-inclusion-and-backward-elimination (FIBE) feature selection algorithm (Fan et al., 2007)(Ou et al., 2009). Overall, attribute extraction, normalization and selection are outside the scope of this paper, but we refer the readers to (Ou et al., 2011) for more details.

3.2 Similarity Measure: Correlation-based Attribute Matching

Figure 3 shows an example of multi-scale and multi-orientation Gabor attributes extracted from (a) preoperative MR and (b) intraoperative US images. Regions 1, 2 and 3 show that attributes follow a relatively more consistent relationship than image intensities. High frequency attributes capture more localized fine-scale edges, whereas low frequency attributes capture more regional and coarse-scale edges. Attributes from $\pi/2$ highlight more vertical edges whereas attributes from π highlight horizontal edges. Overall, attributes at different scales (frequencies) and different orientations extracts a rich set of texture information at each voxel.

The original DRAMMS uses Sum of Squared Differences of attributes (aSSD) as the similarity measure,

$$sim_{aSSD}\left(A_1^\star(u), A_2^\star(T(u))\right) = \frac{1}{1+\frac{1}{D}\left\|A_1^\star(u) - A_2^\star(T(u))\right\|^2} \in [0,1], \tag{2}$$

where D is the dimension of the attribute vector. aSSD prefers identical attributes that may not always be the case in multi-modal image data. Instead, we use Normalized Correlation Coefficient of attributes (aNCC) and Correlation Ratio of attributes (aCR). The two measures have been used for intensity matching but, to the best of our knowledge, not for matching attributes.

$$sim_{aNCC}\left(A_1^\star(\boldsymbol{u}), A_2^\star(T(\boldsymbol{u}))\right) = \frac{d\sum_{i=1}^{d}\left[A_{1i}^\star(\boldsymbol{u})A_{2i}^\star(T(\boldsymbol{u}))\right] - \left(\sum_{i=1}^{d}A_{1i}^\star(\boldsymbol{u})\right)\left(\sum_{i=1}^{d}A_{2i}^\star(T(\boldsymbol{u}))\right)}{\sqrt{d\sum_{i=1}^{d}\left[A_{1i}^\star(\boldsymbol{u})\right]^2 - \left[\sum_{i=1}^{d}A_{1i}^\star(\boldsymbol{u})\right]^2}\sqrt{d\sum_{i=1}^{d}\left[A_{2i}^\star(T(\boldsymbol{u}))\right]^2 - \left[\sum_{i=1}^{d}A_{2i}^\star(T(\boldsymbol{u}))\right]^2}} \quad (3)$$

where $A_{1i}^\star(\boldsymbol{u})$ and $A_{2i}^\star(T(\boldsymbol{u}))$ are d-dimensional attributes selected during the optimization process (within the end-to-end DRAMMS framework), and the index i, next to 1 and 2 in the subscription, denotes the i-th dimension element in the attribute vector.

In a similar notational manner, aCR is defined as

$$sim_{aCR}\left(A_1^\star(\boldsymbol{u}), A_2^\star(T(\boldsymbol{u}))\right) = 1 - \frac{1}{d\sigma^2}\sum_{k=1}^{K} d_k\sigma_k^2 \quad (4)$$

where

$$\sigma^2 = \frac{1}{d}\left(\sum_{i=1}^{d}\left[A_{2i}^\star(T(\boldsymbol{u}))\right]^2\right) - m^2, \quad m = \frac{1}{d}\sum_{i=1}^{d}A_{2i}^\star(T(\boldsymbol{u})) \quad (5)$$

$$\sigma_k^2 = \frac{1}{d_k}\left(\sum_{i\in S_k}\left[A_{2i}^\star(T(\boldsymbol{u}))\right]^2\right) - m_k^2, \quad m_k = \frac{1}{d_k}\sum_{i\in S_k}A_{2i}^\star(T(\boldsymbol{u})) \quad (6)$$

$$S_k = \left\{i \in \left\{1, \ldots d\right\}, A_{1i}^\star(\boldsymbol{u}) \in \max\left(A_1^\star(\boldsymbol{u}) \cdot \left[\frac{k-1}{K}, \frac{k}{K}\right]\right)\right\} \text{ and } d_k = \left|S_k\right| \quad (7)$$

i.e., the attribute elements of a voxel $\boldsymbol{u}$ in image I_1 are divided into totally K disjoint bins $\{S_k\}$ for $k = \{1, \ldots, K\}$ with cardinality d_k, and the aCR measures the functional dependence between the attribute vectors of two voxels. K is set at 32 in our algorithm.

Similar to aSSD (sim_{aSSD}), aNCC (sim_{aNCC}) is symmetric with respect to the two images and has values between zero (no linear dependency between the attribute vectors of two voxels) and one (linearly dependent). Although aCR (sim_{aCR}) is not symmetric, it also takes value of zero (no functional dependency of the target voxel's attribute vector given the source voxel's attribute vector) to one (purely deterministic dependency) as is shown in the following simulated experiment.

Understanding aNCC and aCR in a Simulated Experiment—We simulated variables of linear relationship ($\boldsymbol{y_0} = \boldsymbol{x}$), non-linear but monotonic relationship ($\boldsymbol{y_1} = \boldsymbol{x^2}$) and non-linear but non-monotonic relationship ($\boldsymbol{y_2} = \boldsymbol{\sin(2\pi x)}$), shown in Figure 4 as blue, black, and gray curves, respectively. The similarity values of aSSD, aNCC and aCR are shown in the table at the bottom of Figure 4. Each cell in the table (below Figure 4) corresponds to the similarity value with 0%, 5%, 10% and 20%, respectively, of noise added to the y-axis

(corresponding to the curves in Figure 4, panel a–d). This experiment demonstrates the following:

- **i.** aCR can better capture all three relationships (linear, non-linear but monotonic; non-linear and non-monotonic). All aCR values are close to 1. In contrast, aNCC can capture both linear relationship and non-linear but monotonic relationship (values close to 1); but it does not well capture the non-linear and non-monotonic relationship (values only at 0.75–0.76). aSSD is suitable when two variables follow an identical or linear relationship, but almost completely misses the non-linear relationships (regardless of whether it is monotonic).
- **ii.** aNCC is not as sensitive to non-linear non-monotonic relationships i.e., inconsistent relationships, as aCR. However, aNCC is more stable (0.77➔0.76) than aCR, which quickly drops from 1.0 to 0.94 when noise is present.

Therefore, both aNCC and aCR can likely improve over aSSD for multi-modal MR-US registration. aCR is more sensitive and aNCC is more stable. Comparison of the experiments in real MR-US registration is shown in the Results Section.

3.3. Transformation Strategy to Explicitly Handle MR-US FOV Mismatch

We used an automated transformation strategy to deal with the FOV mismatch between MR and US images. Without explicit skull stripping or other frequently-used semi-automated initialization, we began from rigid registration of raw MR and US images, i.e., with the original FOVs. The robust rigid module in DRAMMS is used (please see Appendix for details, the "−a 4" argument in the software command). It has an improved robustness in the presence of large FOV mismatch by measuring similarity in the overlap instead of union of two images' FOVs (Ou et al., 2018).

In the second step, following the rigid transformation of MR into US space, we dilated the foreground boundary of the US image by a radius of r and used the dilated binary US mask to mask out regions in the rigidly-transformed MR image as shown in Figure 5. Dilation ensures that following masking, the rigidly-transformed MR image has an FOV similar to the US image, leaving the remaining transformation to the third step, the deformable process We call this three-step strategy the "rigid + masking + deformable" strategy and compare it to "rigid only", and rigid followed by deformable without explicit masking, "rigid +deformable".

Similar strategies were used (Drobny et al., 2018)(Hong et al., 2018), in which the dilation radius r was set at 0 and 15 voxels (equivalent to 3mm), respectively for the RESECT dataset. The parameter r was set at 0 for the BITE dataset in other studies (Rivaz & Collins, 2015)(Heinrich et al., 2013)(Rivaz et al., 2014)(Jiang et al., 2016). The following sub-section describes experiments for finding the optimal r and generality in multi-site data.

3.4 Parameter Optimization for Generality and Consistent Accuracy in Multi-Site Data

Our goal is to find a fixed set of parameters that are consistent and highly accurate for multi-site data. The key parameters in our framework are:

i. Similarity measure: we compare aSSD in the original DRAMMS framework and the proposed aNCC and aCR;

ii. Transformation strategy: we compare "rigid only", rigid followed by deformable registration ("rigid+deformable"), and the proposed rigid, dilation, masking, and then deformable registration ("rigid+masking+deformable").

iii. Dilation margin r in the proposed "rigid+masking+deformable" transformation strategy. We vary r from 1mm to 6mm to compensate for any registration errors that may arise from rigid registration, which, as experiments show, are typically less than 6mm (Zimmer, González Ballester, & Piella, 2019).

iv. Smoothness weight λ in Equation (1). We vary λ from 0.6, 1, 1.2, 1.8, 2, 2.3 to 2.5 -- smaller λ, typically 0.1 – 0.8, leads to a more aggressive inter-subject deformation (Ou et al., 2014), whereas larger λ, 1 to 2.5, corresponds to smoother intra-subject registration.

The Similarity measure and Transformation strategy are the most important parameters and they are interleaved. Therefore, we explore an exhaustive combination of these two parameters, test a range of r and λ values, and plot accuracy as a function of those parameters in all datasets.

4. Comprehensive Evaluation

4.1 Accuracy Metric

All three datasets include expert-annotated corresponding landmarks in the MR and US images (see Table 1). Similar to other studies that evaluated MR-US registration accuracy (Heinrich et al., 2013)(Wein et al., 2013)(Masoumi, Xiao, & Rivaz, 2018a), we use the mean target registration error (mTRE), which is the average distance between corresponding landmarks after registration of all landmarks in each patient. Let x_i and $x_i^{'}$ represent expert-annotated corresponding landmark locations in the MR and US images, respectively, indexed by i=1,2,…,N, for a patient, and T the computed deformation by an algorithm. The mTRE of a registration solution is:

$$mTRE = \frac{1}{N}\sum_{i}^{N}\left\|T(x_i) - x_i'\right\|. \quad (8)$$

4.2 Fifteen Other MR-US registration algorithms

Table 2 lists fifteen MR-US registration algorithms. They use a wide variety of image similarity measures, transformation models, and optimization strategies, some of which are also based on machine learning approaches (Zhong et al., 2018)(Sun & Zhang, 2018). The algorithms in Table 2 are noted by an asterisk (*) if they have been validated with self-reported accuracies, or a (#) if they have participated in an international competition in 2018 for third-party independently evaluated accuracies (please see sub-section 4.3).

Two MR-US registration tools, SSC and NiftyReg, are publicly and freely available (Heinrich et al., 2013)(Drobny et al., 2018). SSC uses two different sets of parameters to register MR and US images: from RESECT, set of parameters #1, and BITE, set of parameters #2. We ran every set of parameters in all three datasets. We validated NiftyReg on the RESECT dataset. We ran NiftyReg on both BITE and MIBS. LC2 (Wein et al., 2013) reports a different set of parameters for the two datasets BITE and RESECT, as summarized in Table 3. Unfortunately, LC2 is not freely available.

4.3 Participation in Open Challenges held by Third-Parties

The Correction of Brain shift with Intra-Operative Ultrasound (CuRIOUS) 2018 Challenge (http://curious2018.grand-challenge.org) was held in conjunction with the Medical Image Computing and Computer Assisted Intervention (MICCAI) conference (Xiao et al., 2019). This Challenge included two phases. In Phase I, the training phase, participating teams had access to both MR and US images and reference landmarks for 22 brain tumor patients. They trained their algorithms and reported their own mTRE for each clinical case. In Phase II, the testing phase, participating teams had access to a new set of 10 image pairs (testing dataset), but only the landmark locations in the MR images. They computed the deformed MR landmark locations in the US space, and submitted the locations to the organizers for independent evaluation. The goal of the challenge was to test the accuracy and generality of the algorithm for a cohort of different patients in the RESECT dataset.

4.4 Analysis of Registration Results for Clinical Utility

Tumor type information (low or high-grade gliomas or metastatic brain tumors) is available for each patient in all three datasets (see Table 1). A radiologist (E.G.) with 5 years of experience visually checked the locations of 758 landmark pairs one by one, and assign them into five regions: (i) sulci/gyri, (ii) ventricles, (iii) falx and tentorium, (iv) tumor boundary and (v) others, including vessels, mid brain, pons, bone, white matter and interpeduncular cistern, where landmarks appear less frequently. The information allows us to quantify registration errors by tumor types and by brain regions, which is so far missing in the literature but important for neurosurgeon's references.

5. Results

Section 5.1 reports the computational time. Section 5.2 presents the results for optimizing key parameters. Section 5.3 shows a comparison of sixteen MR-US registration algorithms. Section 5.4 presents an analysis of landmark errors in various brain regions and tumor types.

5.1 Computation Time

The cDRAMMS runs in a single thread on a single CPU. It requires 5–15 minutes to register a typical pair of MR and US images. We ran cDRAMMS hundreds of times in parallel using the Partners Enterprise Research Infrastructure & Services (ERIS) Unix cluster to optimize parameter settings across multiple image pairs and multiple patient data sets. It has more than 380 compute nodes, 7000 CPU cores, and a total of 56TB RAM memory. Jobs submitted to the ERIS cluster were handled by the Load Sharing Facility (LSF) job scheduler for high-performance computing in the Unix environment.

5.2 Optimization of Key Parameters

5.2.1 Similarity Measures and Transformation Strategies—Figure 6 shows the landmark errors with various combinations of similarity measures and transformation strategies in the three datasets and with all data combined.

The observations below are generally applicable to all three datasets, suggesting the consistent performance of cDRAMMS across the multi-site data. Comparing aSSD, aNCC and aCR, we found:

- **i.** Regardless of transformation strategies, correlation-based attribute matching aNCC and aCR have significantly smaller mTREs than the sum-of-square-based attribute matching aSSD.
- **ii.** Between the two different correlation-based attribute matching: aCR leads to significantly lower errors than aNCC in the BITE and RESECT dataset, and lower, but statistically-equivalent errors, in the MIBS dataset.
- **iii.** The observations above echo results in the simulated data (Figure 4). Relaxing the requirement on the equality relationship (aSSD) to a non-linear relationship (aNCC and aCR) better models the MR and US image characteristics. aNCC can model non-linear consistent relationships of attributes, which is further relaxed to aCR that can model a non-linear inconsistent relationships among attributes, leading to further a decrease of registration errors.
- **iv.** Comparing transformation strategies, we found that landmark errors decrease when moving from “rigid” to “rigid+deformable”, and then to “rigid+masking +deformable”.

5.2.2. Dilation Radius *r*—Other studies set r=0 mm (Rivaz & Collins, 2015)(Heinrich et al., 2013)(Rivaz et al., 2014)(Jiang et al., 2016) or r=3 mm (Drobny et al., 2018). Figure 7 shows:

- **i.** Landmark errors are sensitive to the dilation radius r. Therefore, minor FOV differences between MR and US images will affect the registration accuracy.
- **ii.** It appears that the observed optimal value appearing at 5mm is related to our dilating the US mask mainly to compensate for errors in rigid MR-US registration that, on average, are approximately 3mm but can be as large as 4.5–5.5mm (Nigris et al., 2013).

The parameter r reaches the same optimal value in all datasets, suggesting generality of our overall algorithm with multi-site data.

5.2.3 Smoothness Weight λ—The default λ is 0.2 that has been shown to work well for aggressive across-patient registration (Ou et al., 2014). We expect λ=1.5~2.5 for smoother deformations as MR and US are from the same patient. Figure 8 shows λ=2 is observed as optimal. The optimum is the same for all datasets, suggesting consistency of our algorithm with multi-site data.

1.1 5.3 Comparison with Fifteen Additional MR-US Registration Algorithms: Table 4 summarizes the mTREs pre- and post-registration for sixteen algorithms. Cells in the table are empty if the accuracies are not reported and the software is not publicly available.

Accuracy: SSC, LC2 and cDRAMMS present the lowest mTRE among known results in three datasets. Although the initial mTRE ranges from 0–21mm (an average of 4.18–6.41mm for the three datasets), the top ranking methods can reduce the errors to approximately 2mm, and in some datasets, to even as low as 1–5mm. LC2 decreases the mTREs to 1.57–1.75mm in RESECT and 2.52 mm in BITE (with a different set of parameters). SSC reduces the mTREs to 1.67–1.87mm in RESECT, 2.34 in BITE (with a different set of parameters) and 2.91–3.61mm in MIBS. cDRAMMS lowers the mTREs to 2.08–2.28mm in all three datasets.

Generality: cDRAMMS ranks first in the two datasets, BITE and MIBS, and third in the RESECT dataset. This is based on using a fixed set of parameters. Using different parameters, SSC ranks second in both RESECT and MIBS and third in BITE. LC2 ranks second in RESECT's training subset, first in RESECT's testing subset, and fourth in the BITE dataset. MLP, a deep learning algorithm, ranks first in RESECT's training subset, and fifth among those results submitted for the unseen testing subset of the RESECT dataset. The lowest mTRE averages among the three datasets are 1.57mm for LC2, 1.67mm for SSC, and 2.08mm for cDRAMMS. The highest mTRE averages among the three datasets are 2.52mm for LC2, 3.61mm for SSC, and 2.28mm for cDRAMMS. MLP, a deep learning based algorithm, shows a far lower mTRE (1.21±0.55mm) in the RESECT training subset but the errors in the RESECT testing subset are 5.72±2.82mm.

5.4 Additional Analysis of Results for Clinical Utility

5.4.1. Qualitative Assessment by Physicians—Figure 9 shows MR-US registration results for three randomly selected patients from each of the three datasets. Blue and white arrows point out tumor boundaries, yellow arrows the falx, and red arrows the sulci. These regions are aligned following cDRAMMS registration (bottom row) compared to those before registration (third row).

A neurosurgeon in training (P.U) and an experienced neurosurgeon (W.E.) with two and nine years' of clinical practice, respectively, and each with four years' intraoperative ultrasound experience, visually assessed the results of the cDRAMMS registration in the tumor boundary, sulci, vessels, choroid plexus, falx, and ventricles. Table 5 shows their grading of no registration as "bad" (grossly visible misregistration) as 49% and 44% of registrations, respectively, as "good" (minor visible misalignments), and 51% and 56% of the registrations, respectively, as "great" (negligible or nearly undetectable misregistration).

5.4.2 Corrected Brain Shift—Figure 10 compares mTRE for all 43 patients before and after cDRAMMS registration. Regardless of datasets, patient cases, and the scale of initial brain shift (*x*-axis, ranging from 0 to 21mm), the results show consistently low mTRE after registration at 0 to 4mm, and most patients have an mTRE around 2mm after registration

5.4.3 Distribution of Registration Errors by Brain Regions—Figure 11 presents color code registration errors in various brain regions in nine randomly selected patients, three from each of the three datasets. In the selected patients, the errors are smaller in sulci and gyri (0.5–1.5mm) than in tumor boundaries (2–3.5mm). This is echoed in Figure 12, which shows the statistics for landmark errors before and after registration in all datasets. In tumor boundary, the average landmark errors are approximately 2.7mm; in ventricles, approximately 2.7mm; in falx and tentorium, approximately 2mm; and in sulci and gyri, approximately 1.4mm.

5.4.4 Distribution of Errors by Type of Tumor—High-grade gliomas have higher landmark errors, averaging 3.1–3.3mm, than low-grade gliomas averaging approximately 1.25mm. Figure 13 shows and highlights the challenge in registering images in patients with HGG. The registration errors for metastatic brain tumor, averaging approximately 2.1mm, are between the errors for HGG and LGG.

6. Discussion

Accuracy and generality are two major issues in registering preoperative MR and intraoperative US images for neurosurgical guidance. This study demonstrates that using correlation-based attribute matching, aNCC and aCR, to replace difference-based attribute matching, improves the accuracy for the non-rigid MR-US registration tasks. aNCC and aCR are now publicly available in the DRAMMS software (https://www.nitrc.org/projects/dramms). Explicitly normalizing the FOVs further improves registration accuracy. The recent availability of multi-site data allows us to optimize key parameters and test generality of the proposed algorithm. We show that a fixed set of thoroughly optimized parameters can have relatively consistent performance with multi-site data. We found that landmark errors are larger in tumor boundary or tentorium regions than in sulci and gyri regions, larger in patients with high-grade gliomas than in those with metastatic tumors and low-grade gliomas. This suggests the need for future work to further improve accuracy as discussed below.

Our results with both simulated and real data show that Correlation Ratio of attributes (aCR) can match regions with inconsistent attribute values. Normalized Correlation Coefficient of attributes (aNCC) can match attributes of both linear and non-linear relationships so long as the relationship is consistent or monotonic.

Attribute-matching based algorithms, including LC2, SSC and cDRAMMS, are among the top accurate and general methods. This highlights the merit of using the high-dimensional texture attributes other than one-dimensional intensity of a voxel for matching. Texture attributes render each voxel more distinctive than the intensity information. The framework is open to other attributes. Other example attributes include, but are not limited to, moments (Shen & Davatzikos, 2002), local histogram analysis (Shen, 2007), local entropy and Laplacian attributes (Wachinger & Navab, 2012), rotationally invariant difference-of-Gaussian filter responses (Toews & Arbel, 2009), uniform spherical region descriptor (Liao & Chung, 2012), rotation-invariant attributes based on alpha stable filter banks (Liao & Chung, 2010), and so on.

In order to normalize the MR image FOV to that of the US image, we dilated the binary mask of US images to compensate for errors in the initial rigid alignment of MR and US images. Errors in MR-US rigid registration are approximately 3mm, but can be as much as 4.5–5.5mm or even higher (Nigris et al., 2013). We believe this explains our observation that the optimal dilation of the FOV of the US image should be approximately 5mm in the datasets studied here.

There is ongoing debate as to whether MR-US (mostly pre-section US) registration should be rigid (Coupé et al., 2012)(Letteboer et al., 2003)(Nigris et al., 2013), affine (Masoumi et al., 2018a)(Masoumi et al., 2018b), or deformable (Rivaz et al., 2015)(Wein, 2018) (Heinrich, 2018). Our results show that predurotomy brain shift is a largely rigid plus highly regularized (smooth) non-rigid process.

Generality remains a major issue in multi-site data. As Table 4 shows, LC2 is highly accurate for the RESECT dataset – the average mTREs across RESECT training and testing patients are 1.57±0.62 and 1.75±0.55mm, respectively. It uses a second set of parameters to maintain a high level of accuracy in the BITE dataset, 2.52±0.87mm. SSC achieves a mTRE of 2.18±0.37mm (the third lowest errors) on the BITE dataset, and 1.67±0.54mm or 1.87±0.51mm (the second lowest errors) on the RESECT training and testing datasets. However, the same parameters lead to a larger mTRE, at 3.61±0.82mm for the MIBS dataset. Changing to another set of parameters allows SSC to achieve a lower mTRE (2.91±0.88mm), which ranks the second lowest errors in the MIBS dataset. In contrast, cDRAMMS scores relatively consistently in all three datasets while using a fixed set of parameters. The average mTRE equals 2.08mm in BITE (lowest errors among methods), 2.19–2.28mm in RESECT (training and testing, respectively, third lowest errors among methods), and 2.24mm in MIBS (lowest errors among methods). The standard deviation is 0.37mm, 0.71–0.87mm, and 0.78mm, respectively in those datasets, which shows the stability across all patients. It was further demonstrated in Figure 10 that the patient mTRE following cDRAMMS registration is stable within a narrow range (1–3.5mm) across patients and datasets, irrespective of landmark distribution, tumor grade, and the initial mTRE (sometimes as great as approximately 20mm).

Besides efforts in algorithm development aiming for accuracy and generality, the evaluation of sixteen algorithms in three datasets is the most comprehensive so far in the MR-US registration tasks. CuRIOUS in 2018 marks so-far the first independent evaluation of MR-US registration algorithms in a well-organized open platform that consists of training and testing data (Xiao et al., 2019). Our work uses results from the CuRIOUS 2018 Challenge (Table 4, results for the RESECT dataset). The extensions are: *(a)* we used two additional datasets to make it multi-site evaluation (details in Table 1); therefore, not only the accuracy but also the generality of sixteen algorithms (Table 2, compared to six algorithms included in (Xiao et al., 2019), some with multiple sets of parameter configurations (Table 3), can be further evaluated in multi-site datasets (Table 4 for accuracy and generality); and *(b)* we further annotated the regions of the landmark pairs, which led to the first report of the registration accuracy by the extent of brain shift that can be corrected (Figure 10), and the registration accuracy by brain regions (Figures 11 and 12) and by tumor grades (Figures 13).

Our detailed statistics (Section 5.4) found that landmark errors are larger in tumor boundary and tentorium regions than in sulci/gyri (Figure 12), and larger in patient with high-grade glioma than in patient with low-grade gliomas or metastatic tumor (Figure 13) suggesting the need for future work to further improve accuracy. Toward a higher accuracy, the deep learning based MLP algorithm shows great promise in the RESECT training subset (N=22). The MLP algorithm reduced landmark errors to an average of 1.21mm per patient (Zhong et al., 2018), much lower than the − 2mm, and even up to 5.6mm, for other algorithms. One may suspect that this is overfitting because, in the RESECT testing subset (N=10), MLP's errors increased to 5.72mm in unseen, and probably more challenging, cases. It leaves the question open as to whether with refined network designs and the greater availability of training data, the accuracy of deep learning-based algorithms will stabilize for this task as has been observed in other across subject healthy brain image registration tasks (Yang, Kwitt, Styner, & Niethammer, 2017)(Wang, Kim, Wu, & Shen, 2017)(De Vos, Berendsen, Viergever, Staring, & Išgum, 2017). This future direction requires not only algorithm development, but also multi-site collaboration for releasing more MR-US data. When large-scale multi-site data is available, it becomes difficult for experts to annotate thousands of landmarks. The proposed cDRAMMS algorithm can offer voxel-wise correspondences at a consistent level of accuracy for hundreds of thousands of voxels. Therefore, future deep learning algorithms can first learn the large number of voxel-wise correspondences found by highly accurate algorithm, and then fine tune the deep learning neural network on expert-annotated, a smaller number of landmark pairs.

Speed is an additional factor in considering the clinical utility of an algorithm. cDRAMMS takes tens of minutes (single-thread single-CPU desktop) whereas several other algorithms may require only two to five minutes (Masoumi et al., 2018a), one to two minutes (Shams, R., Boucher, M. A. and Kadoury, 2018), or even tens of seconds (Heinrich, 2018)(Zhong et al., 2018). cDRAMMS uses cubic B-spline-based free form deformation and discrete optimization. Recent studies have shown that the same components can be optimized such that the overall computation time is reduced to only a few seconds (Jiang et al., 2016) (Heinrich, 2018). This is an additional area for future research.

A further limitation and subject for future work is how to best evaluate registration accuracy. This study, similar to most MR-US brain tumor registration studies (Table 2 and Table 4), use manually annotated landmark correspondences. However, expert landmarks are subject to human error, are non-uniformly distributed in the image, and may, in extreme cases, not be representative of errors in a clinical target such as the tumor edges (Wein, 2018) (Fitzpatrick, 2009). A more rigorous validation would be the exploration of the Dice overlap on the actual segmentation of the tumor and surrounding key regions (Wein, 2018), and the properties of the deformation field (Fitzpatrick, 2009)(Rohlfing, 2012). The three datasets in our study, as of now, do not contain regional segmentations in both MR and US image space. This is another future direction toward thoroughly evaluating clinical utility.

Recent progress has been made toward clinical utility. A representative direction is the filling of the gap between the laboratory-based algorithm development and actual clinical bedside navigation system. CustusX (Askeland et al., 2016), published in 2016, represents a step in this direction. In a 2018 study (Iversen et al., 2018), the authors of CustusX teamed

up with authors of the LC2 algorithm, and tested the performance of a commercial MR-US registration software, ImFusion GmbH (Munich, Germany), on the CustusX platform. They reported a mean landmark error mTRE=4.47mm (median at 2.72mm) (Iversen et al., 2018). Future work includes disseminating laboratory-developed software into clinical platforms for further evaluations and improvement in clinical settings.

7. Conclusions

We present a correlation-based attribute matching algorithm for registration of MR-US images in neurosurgery and show its generality and accuracy in multi-site data. Our work suggests that by using multi-scale, multi-orientation attributes, coupled with correlation-based robust similarity measures on attributes and explicit handling of field of view differences, and by thoroughly optimizing key parameters, it is possible that an algorithm with a fixed set of parameters can be applied to multi-site data and achieve consistent accuracies. This brings us a step closer to the clinical use of registration for neurosurgery assistance. It motivates the continuing development of GPU-powered, faster, more sophisticated learning-based algorithms for use with larger-scale multi-site data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This study was supported by the National Institute of Health Grants P41-EB015898, P41-EB015902 and R01-NS049251, and a Natural Sciences and Engineering Research Council of Canada Discovery grant. The authors would like to acknowledge the financial support from the Portuguese Foundation for Science and Technology under the references PD/BD/105869/2014 and IDMEC/LAETA UID/EMS/50022/2013. The authors also acknowledge the financial support to YO from the Faculty Development Award of Harvard Medical School and Boston Children's Hospital, and the Career Development Award from Thrasher Research Fund (THF13411). We thank the organizers of the CuRIOUS 2018 Challenge, Drs. Ingerid Reinertsen, Hassan Rivaz, Yiming Xiao and Matthieu Chabanas, for organizing the first open challenge for MR-US registration, and for their insightful discussions. We thank Enterprise Research Infrastructure & Services (ERIS) at Partners Healthcare for their in-depth support and for the provision of the ERIS-ONE high performance cluster computing environment.

References

Askeland C, Solberg OV, Bakeng JBL, Reinertsen I, Tangen GA, Hofstad EF, … Lindseth F (2016). CustusX: an open-source research platform for image-guided therapy. International Journal of Computer Assisted Radiology and Surgery. 10.1007/s11548-015-1292-0

Bookstein FL (1989). Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence. 10.1109/34.24792

Coupé P, Hellier P, Morandi X, & Barillot C (2007). A probabilistic objective function for 3D rigid registration of intraoperative us and preoperative MR brain images. In 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro - Proceedings 10.1109/ISBI.2007.357103

Coupé P, Hellier P, Morandi X, & Barillot C (2012). 3D rigid registration of intraoperative ultrasound and preoperative MR brain images based on hyperechogenic structures. International Journal of Biomedical Imaging. 10.1155/2012/531319

de Vos BD, Berendsen FF, Viergever MA, Staring M, & Išgum I (2017). End-to-end unsupervised deformable image registration with a convolutional neural network. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 10.1007/978-3-319-67558-9_24

Diez Y, Oliver A, Cabezas M, Valverde S, Martí R, Vilanova JC, … Lladó X (2014). Intensity based methods for brain MRI longitudinal registration. A study on multiple sclerosis patients. Neuroinformatics. 10.1007/s12021-013-9216-z

Dimaio SP, Archip N, Hata N, Talos I, Warfield SK, Majumdar A, … Kikinis R (2006). Image-guided neurosurgery at Brigham and Women's Hospital. IEEE Engineering in Medicine and Biology Magazine. 10.1109/MEMB.2006.1705749

Drobny D, Vercauteren T, Ourselin S, & Modat M (2018). Registration of MRI and iUS data to compensate brain shift using a symmetric block-matching based approach. In Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation (pp. 172–178). 10.1007/978-3-030-01045-4_21

Fan Y, Shen D, Gur RC, Gur RE, & Davatzikos C (2007). COMPARE: Classification of morphological patterns using adaptive regional elements. IEEE Transactions on Medical Imaging. 10.1109/TMI.2006.886812

Farnia P, Ahmadian A, Shabanian T, Serej ND, & Alirezaie J (2014). A hybrid method for non-rigid registration of intra-operative ultrasound images with pre-operative MR images. Engineering in Medicine and Biology Society (EMBC), 5562–5565. 10.1109/EMBC.2014.6944887

Farnia P, Ahmadian A, Shabanian T, Serej ND, & Alirezaie J (2015). Brain-shift compensation by non-rigid registration of intra-operative ultrasound images with preoperative MR images based on residual complexity. International Journal of Computer Assisted Radiology and Surgery, 10(5). 10.1007/s11548-014-1098-5

Farnia P, Makkiabadi B, Ahmadian A, & Alirezaie J (2016). Curvelet based residual complexity objective function for non-rigid registration of pre-operative MRI with intraoperative ultrasound images. In Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS) (Vol. 2016-Octob, pp. 1167–1170). 10.1109/EMBC.2016.7590912

Farnia P, Najafzadeh E, Ahmadian A, Makkiabadi B, Alimohamadi M, & Alirezaie J (2018). Co-Sparse Analysis Model Based Image Registration to Compensate Brain Shift by Using Intra-Operative Ultrasound Imaging. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 1–4).

Fitzpatrick JM (2009). Fiducial registration error and target registration error are uncorrelated. In Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling. 10.1117/12.813601

Fuerst B, Wein W, Müller M, & Navab N (2014). Automatic ultrasound-MRI registration for neurosurgery using the 2D and 3D LC2 Metric. Medical Image Analysis. 10.1016/j.media.2014.04.008

Gerard IJ, Kersten-Oertel M, Petrecca K, Sirhan D, Hall JA, & Collins DL (2017). Brain shift in neuronavigation of brain tumors: A review. Medical Image Analysis. 10.1016/j.media.2016.08.007

Glocker B, Komodakis N, Tziritas G, Navab N, & Paragios N (2008). Dense image registration through MRFs and efficient linear programming. Medical Image Analysis. 10.1016/j.media.2008.03.006

Golby A (2015). Image-guided Neurosurgery. Academic Press.

Hartov A, Roberts DW, & Paulsen KD (2008). A comparative analysis of coregistered ultrasound and magnetic resonance imaging in neurosurgery. Neurosurgery, 62(3 SUPPL. 1). 10.1227/01.neu.0000317377.15196.45

Heckemann RA, Rueckert D, Aljabar P, Hammers A, & Hajnal JV (2006). Diffeomorphic Registration Using B-Splines. 10.1007/11866763_86

Heinrich MP (2018). Intra-operative ultrasound to MRI fusion with a public multimodal discrete registration tool. In Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation (pp. 159–164).

Heinrich Mattias P., Jenkinson M, Bhushan M, Matin T, Gleeson FV, Brady SM, & Schnabel JA (2012). MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. Medical Image Analysis. 10.1016/j.media.2012.05.008

Heinrich, Paul Mattias, Jenkinson M, Papiez BW, Brady SM, & Schnabel JA (2013). Towards realtime multimodal fusion for image-guided interventions using self-similarities. In Lecture Notes in

Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 10.1007/978-3-642-40811-3_24

Hladuvka J (2003). Establishing point correspondence on training set boundaries. Technical Report, 40(2003), 1–15. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.6.3459&rep=rep1&type=pdf

Hong J, Park H, & Hong J and Park H (2018). Non-linear approach for MRI to intra-operative us registration using structural skeleton. In Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation (pp. 138–145). 10.1007/978-3-030-01045-4_16

Iversen DH, Wein W, Lindseth F, Unsgård G, & Reinertsen I (2018). Automatic Intraoperative Correction of Brain Shift for Accurate Neuronavigation. World Neurosurgery. 10.1016/j.wneu.2018.09.012

Ji S, Wu Z, Hartov A, Roberts DW, & Paulsen KD (2008). Mutual-information-based image to patient re-registration using intraoperative ultrasound in image-guided neurosurgery. Medical Physics, 35(10), 4612–4624. 10.1118/1.2977728 [PubMed: 18975707]

Jiang D, Shi Y, Yao D, Wang M, & Song Z (2016). miLBP: a robust and fast modality-independent 3D LBP for multimodal deformable registration. International Journal of Computer Assisted Radiology and Surgery. 10.1007/s11548-016-1407-2

Komodakis N, Tziritas G, & Paragios N (2008). Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal-dual strategies. Computer Vision and Image Understanding. 10.1016/j.cviu.2008.06.007

Letteboer MMJ, Viergever MA, & Niessen WJ (2003). Rigid registration of 3D ultrasound data of brain tumours. International Congress Series. 10.1016/S0531-5131(03)00305-4

Liao S, & Chung ACS (2010). Feature based nonrigid brain MR image registration with symmetric alpha stable filters. IEEE Transactions on Medical Imaging. 10.1109/TMI.2009.2028078

Liao S, & Chung ACS (2012). Nonrigid brain MR image registration using uniform spherical region descriptor. IEEE Transactions on Image Processing. 10.1109/TIP.2011.2159615

Lindseth F, Kaspersen JH, Ommedal S, Langø T, Bang J, Hokland J, … Nagelhus Hemes TA (2003). Multimodal image fusion in ultrasound-based neuronavigation: improving overview and interpretation by integrating preoperative MRI with intraoperative 3D ultrasound. Computer Aided Surgery, 8(2), 49–69. 10.3109/10929080309146040 [PubMed: 15015719]

Luo J, Frisken S, Machado I, Zhang M, Pieper S, Golland P, … Wells WM (2018). Using the variogram for vector outlier screening: application to feature-based image registration. International Journal of Computer Assisted Radiology and Surgery. 10.1007/s11548-018-1840-5

Machado I, Toews M, Luo J, Unadkat P, Essayed W, George E, … Ou Y (2018). Deformable MRI-ultrasound registration via attribute matching and mutual-saliency weighting for image-guided neurosurgery. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 10.1007/978-3-030-01045-4_20

Machado I, Toews M, Luo J, Unadkat P, Essayed W, George E, … 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. 10.1007/s11548-018-1786-7

Masoumi N, Xiao Y, & Rivaz H (2018a). ARENA: Inter-modality affine registration using evolutionary strategy. International Journal of Computer Assisted Radiology and Surgery. 10.1007/s11548-018-1897-1

Masoumi N, Xiao Y, & Rivaz H (2018b). MARCEL (Inter-Modality Affine Registration with CorrELation Ratio): An application for brain shift correction in ultrasound-guided brain tumor resection. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 10.1007/978-3-319-75238-9_5

Mercier L, Del Maestro RF, Petrecca K, Araujo D, Haegelen C, & Collins DL (2012). Online database of clinical MR and ultrasound images of brain tumors. Medical Physics. 10.1118/1.4709600

Mercier L, Del Maestro RF, Petrecca K, Kochanowska A, Drouin S, Yan CXB, … Collins DL (2011). New prototype neuronavigation system based on preoperative imaging and intraoperative freehand ultrasound: System description and validation. International Journal of Computer Assisted Radiology and Surgery. 10.1007/s11548-010-0535-3

Miller D, Benes L, & Sure U (2011). Stand-alone 3D-ultrasound navigation after failure of conventional image guidance for deep-seated lesions. Neurosurgical Review. 10.1007/s10143-011-0314-9

Morin F, Courtecuisse H, Reinertsen I, Le Lann F, Palombi O, Payan Y, & Chabanas M (2017). Brain-shift compensation using intraoperative ultrasound and constraint-based biomechanical simulation. Medical Image Analysis, 40, 133–153. 10.1016/j.media.2017.06.003 [PubMed: 28651099]

Munkvold BKR, Bo HK, Jakola AS, Reinertsen I, Berntsen EM, Unsgard G, … Solheim O (2017). Tumor Volume Assessment in Low-Grade Gliomas: A Comparison of Preoperative Magnetic Resonance Imaging to Coregistered Intraoperative 3-Dimensional Ultrasound Recordings. Neurosurgery.

Myronenko A, & Song X (2010). Intensity-based image registration by minimizing residual complexity. IEEE Transactions on Medical Imaging. 10.1109/TMI.2010.2053043

Nigris D. De, Collins DL, & Arbel T (2013). Fast rigid registration of pre-operative magnetic resonance images to intra-operative ultrasound for neurosurgery based on high confidence gradient orientations. International Journal of Computer Assisted Radiology and Surgery. 10.1007/s11548-013-0826-6

Ou Y, Akbari H, Bilello M, Da X, & Davatzikos C (2014). Comparative evaluation of registration algorithms in different brain databases with varying difficulty: Results and insights. IEEE Transactions on Medical Imaging. 10.1109/TMI.2014.2330355

Ou Y, Shen D, Zeng J, Sun L, Moul J, & Davatzikos C (2009). Sampling the spatial patterns of cancer: Optimized biopsy procedures for estimating prostate cancer volume and Gleason Score. Medical Image Analysis. 10.1016/j.media.2009.05.002

Ou Y, Sotiras A, Paragios N, & Davatzikos C (2011). DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting. Medical Image Analysis. 10.1016/j.media.2010.07.002

Ou Y, Zöllei L, Da X, Retzepi K, Murphy SN, Gerstner ER, … Gollub RL (2018). Field of View Normalization in Multi-Site Brain MRI. Neuroinformatics 10.1007/s12021-018-9359-z

Palombi O, Morin F, Payan Y, Chabanas M, Le Lann F, Courtecuisse H, & Reinertsen I (2018). Resection-induced brain-shift compensation using vessel-based methods. 10.1117/12.2293640

Porter BC, Rubens DJ, Strang JG, Smith J, Totterman S, & Parker KJ (2001). Three-dimensional registration and fusion of ultrasound and MRI using major vessels as fiducial markers. IEEE Transactions on Medical Imaging. 10.1109/42.921484

Reinertsen I, Lindseth F, Unsgaard G, & Collins DL (2007). Clinical validation of vessel-based registration for correction of brain-shift. Medical Image Analysis, 11(6), 673–684. 10.1016/j.media.2007.06.008 [PubMed: 17681484]

Reinertsen Ingerid, Lindseth F, Askeland C, Iversen DH, & Unsgård G (2014). Intra-operative correction of brain-shift. Acta Neurochirurgica. 10.1007/s00701-014-2052-6

Rivaz H, Chen SJS, & Collins DL (2015). Automatic deformable MR-ultrasound registration for image-guided neurosurgery. IEEE Transactions on Medical Imaging, 34(2). 10.1109/TMI.2014.2354352

Rivaz H, & Collins DL (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. 10.1007/s11548-014-1099-4 [PubMed: 25373447]

Rivaz H, Karimaghaloo Z, & Collins DL (2014). Self-similarity weighted mutual information: a new nonrigid image registration metric. Medical Image Analysis. 10.1016/j.media.2013.12.003

Rivaz H, Karimaghaloo Z, Fonov VS, & Collins DL (2014). Nonrigid registration of ultrasound and MRI using contextual conditioned mutual information. IEEE Transactions on Medical Imaging. 10.1109/TMI.2013.2294630

Rizzini DL, & Caselli S (2007). Improved Mixture Representation in Real-Time Particle Filters for Robot Localization. In EMCR.

Roche A, Pennec X, Rudolph M, Auer DP, Malandain G, Ourselin S, … Ayache N (2000). Generalized Correlation Ratio for Rigid Registration of 3D Ultrasound with MR Images. Miccai. 10.1007/978-3-540-40899-4_58

Roche Alexis, Malandain G, Pennec X, & Ayache N (2006). The correlation ratio as a new similarity measure for multimodal image registration. 10.1007/bfb0056301

Rohlfing T (2012). Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable. IEEE Transactions on Medical Imaging. 10.1109/TMI.2011.2163944

Rueckert D (1999). Nonrigid registration using free-form deformations: Application to breast mr images. IEEE Transactions on Medical Imaging. 10.1109/42.796284

Rygh OM, Selbekk T, Torp SH, Lydersen S, Hernes TAN, & Unsgaard G (2008). Comparison of navigated 3D ultrasound findings with histopathology in subsequent phases of glioblastoma resection. Acta Neurochirurgica 10.1007/s00701-008-0017-3

Shams R, Boucher MA and Kadoury S (2018). Intra-operative brain shift correction with weighted locally linear correlations of 3D US and MRI. In Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation (pp. 179–184).

Shen D (2007). Image registration by local histogram matching. Pattern Recognition. 10.1016/j.patcog.2006.08.012

Shen D, & Davatzikos C (2002). HAMMER: Hierarchical attribute matching mechanism for elastic registration. IEEE Transactions on Medical Imaging. 10.1109/TMI.2002.803111

Sotiras A, Davatzikos C, & Paragios N (2013). Deformable medical image registration: A survey. IEEE Transactions on Medical Imaging. 10.1109/TMI.2013.2265603

Sun L and Zhang S (2018). Deformable MRI-ultrasound registration using 3D convolutional neural network In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (pp. 129–137). Springer.

Tempany CMCC, Jayender J, Kapur T, Bueno R, Golby A, Agar N, & Jolesz FA (2015). Multimodal imaging for improved diagnosis and treatment of cancers. Cancer, 121(6), 817–827. 10.1002/cncr.29012 [PubMed: 25204551]

Toews, & Arbel. (2003). Entropy-of-likelihood feature selection for image correspondence. In Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV) (pp. 1041–1047). 10.1109/iccv.2003.1238464

Toews M, & Arbel T (2009). Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion. IEEE Transactions on Pattern Analysis and Machine Intelligence. 10.1109/TPAMI.2008.233

Toews M, & Wells WM (2013). Efficient and robust model-to-image alignment using 3D scale-invariant features. Medical Image Analysis. 10.1016/j.media.2012.11.002

Unsgaard G, Selbekk T, Brostrup Müller T, Ommedal S, Torp SH, Myhr G, … Nagelhus Hernes TA (2005). Ability of navigated 3D ultrasound to delineate gliomas and metastases - Comparison of image interpretations with histopathology. Acta Neurochirurgica 10.1007/s00701-005-0624-1

Vercauteren T, Pennec X, Perchant A, & Ayache N (2007). Non-parametric Diffeomorphic Image Registration with the Demons Algorithm. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007. 10.1007/978-3-540-75759-7_39

Wachinger C, & Navab N (2012). Entropy and Laplacian images: Structural representations for multi-modal registration. Medical Image Analysis. 10.1016/j.media.2011.03.001

Wachinger C, & Navab N (2013). Simultaneous registration of multiple images: Similarity metrics and efficient optimization. IEEE Transactions on Pattern Analysis and Machine Intelligence. 10.1109/TPAMI.2012.196

Wang S, Kim M, Wu G, & Shen D (2017). Scalable High Performance Image Registration Framework by Unsupervised Deep Feature Representations Learning. In Deep Learning for Medical Image Analysis. 10.1016/B978-0-12-810408-8.00015-8

Wein W (2018). Brain-shift correction with image-based registration and landmark accuracy evaluation In Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation (pp. 146–151). Springer.

Wein W, Ladikos A, Fuerst B, Shah A, Sharma K, & Navab N (2013). Global registration of ultrasound to MRI using the LC2metric for enabling neurosurgical guidance. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 10.1007/978-3-642-40811-3_5

Wu G, Yap PT, Kim M, & Shen D (2010). TPS-HAMMER: Improving HAMMER registration algorithm by soft correspondence matching and thin-plate splines based deformation interpolation. NeuroImage 10.1016/j.neuroimage.2009.10.065

Xiao Y, Fortin M, Unsgärd G, Rivaz H, & 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: A. Medical Physics. 10.1002/mp.12268

Xiao Y, Rivaz H, Chabanas M, Fortin M, Machado I, Ou Y, … Reinertsen I (2019). Evaluation of MRI to ultrasound registration methods for brain shift correction: The CuRIOUS2018 Challenge. ArXiv E-Prints, arXiv:1904.10535.

Xiao Y, Rivaz H, Chabanas M, Fortin M, Machado I, Ou Y, Heinrich M, Schnabel JA, Zhong X, Maier A, Wein W, Shams R, Kadoury S, Droby D, Modat M, Reinertsen I (n.d.). Evaluation of MRI to Ultrasound registration methods for bran shift correction: the CuRIOUS Challenge 2018 Submitted to IEEE TMI.

Yang X, Kwitt R, Styner M, & Niethammer M (2017). Quicksilver: Fast predictive image registration – A deep learning approach. NeuroImage. 10.1016/j.neuroimage.2017.07.008

Zhan Y, & Shen D (2006). Deformable segmentation of 3-D ultrasound prostate images using statistical texture matching method. IEEE Transactions on Medical Imaging. 10.1109/TMI.2005.862744

Zhong X, Bayer S, Ravikumar N, Strobel N, Birkhold A, Kowarschik M, … Maier A (2018). Resolve intraoperative brain shift as imitation game In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Vol. 11042 LNCS, pp. 129–137). Springer 10.1007/978-3-030-01045-4_15

Zimmer VA, González Ballester MÁ, & Piella G (2019). Multimodal image registration using Laplacian commutators. Information Fusion. 10.1016/j.inffus.2018.09.009

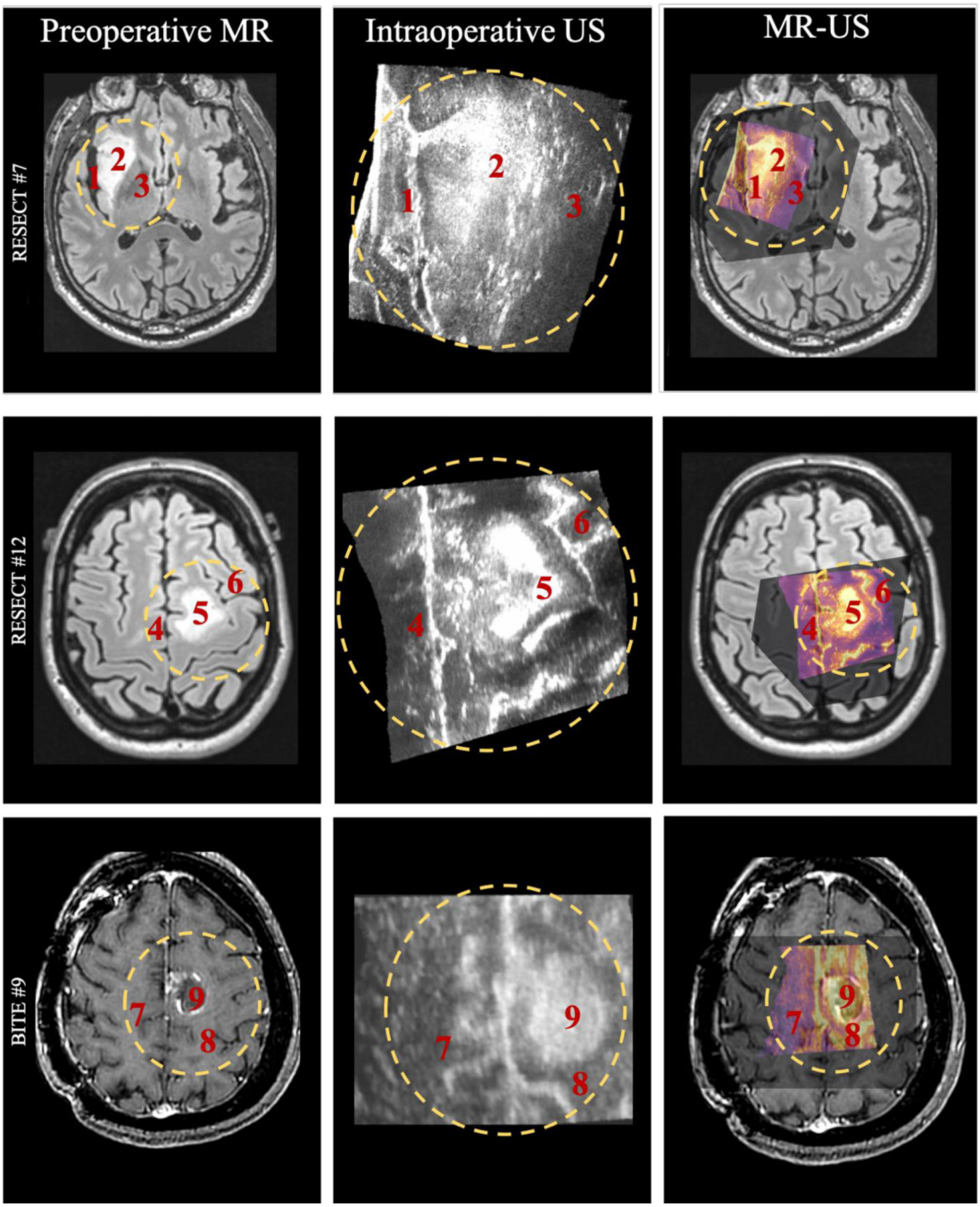


Figure 1:

MR-US registration challenges. Each row corresponds to a different clinical case (a patient). The left, middle and right columns are the preoperative MR, intraoperative US and the superposition of the two images, respectively. Dashed blue circles show the different fields-of-view of the two modalities. MR and US images of corresponding structures are numbered from 1 to 9. Bright regions in MR can correspond to bright intensities in US, e.g., regions 2 and 5; similarly, dark regions of MR can correspond to dark intensities of US, e.g., regions 7 and 8. However bright regions of MR can also correspond to dark intensities of US, e.g.,

regions 1, 4, 6 and 9, or grey regions of MR can correspond to dark intensities of US, e.g., region 3.

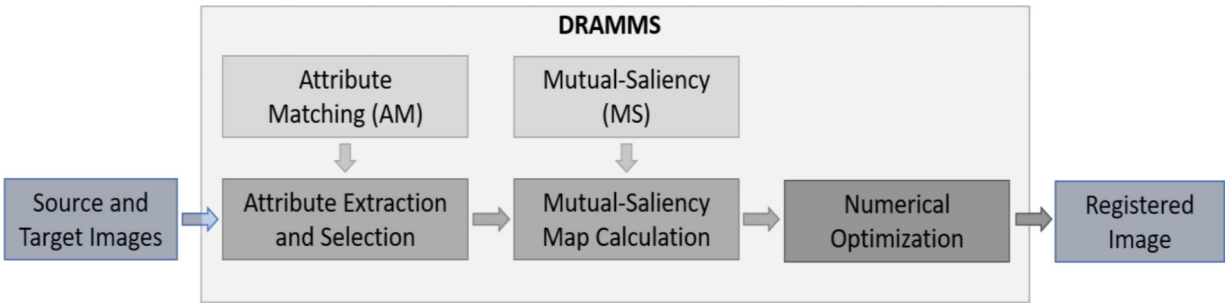


Figure 2:
The original DRAMMS framework (sketch of the framework in (Ou et al., 2011)).

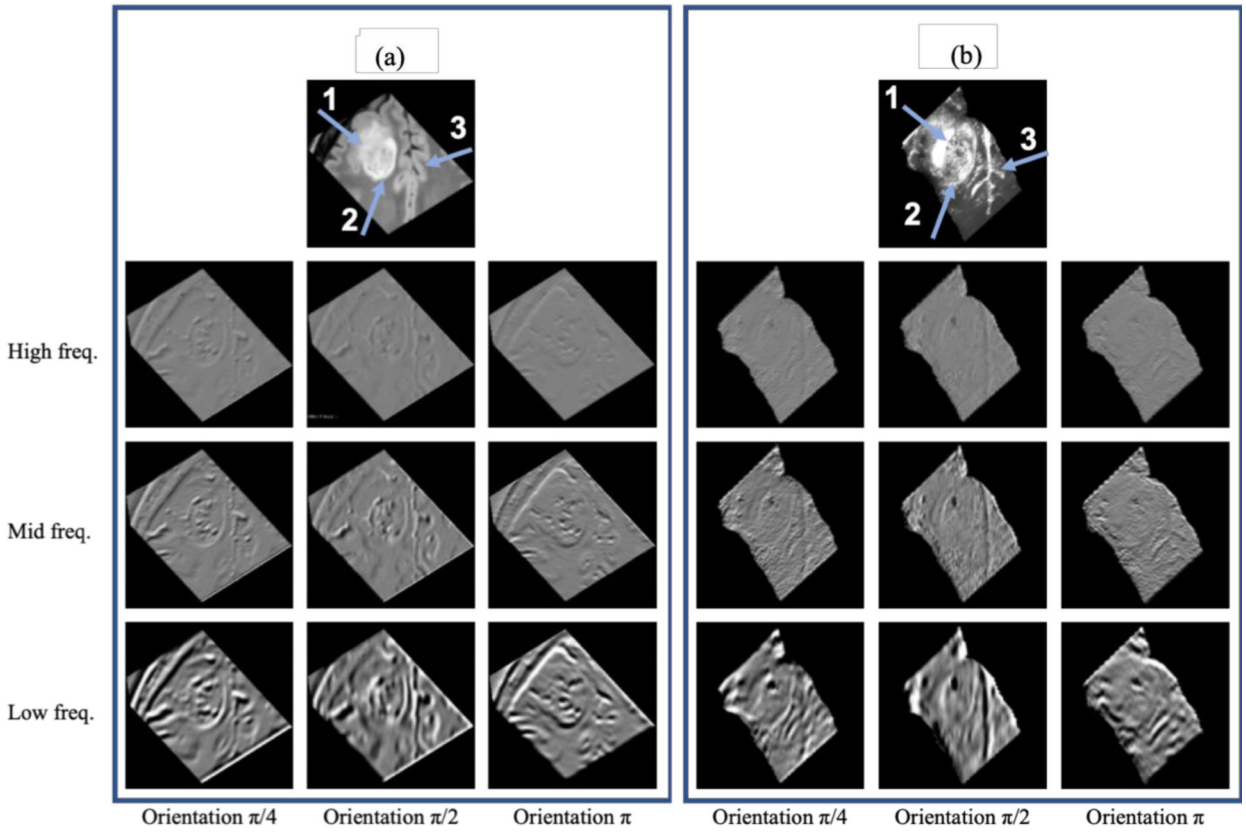
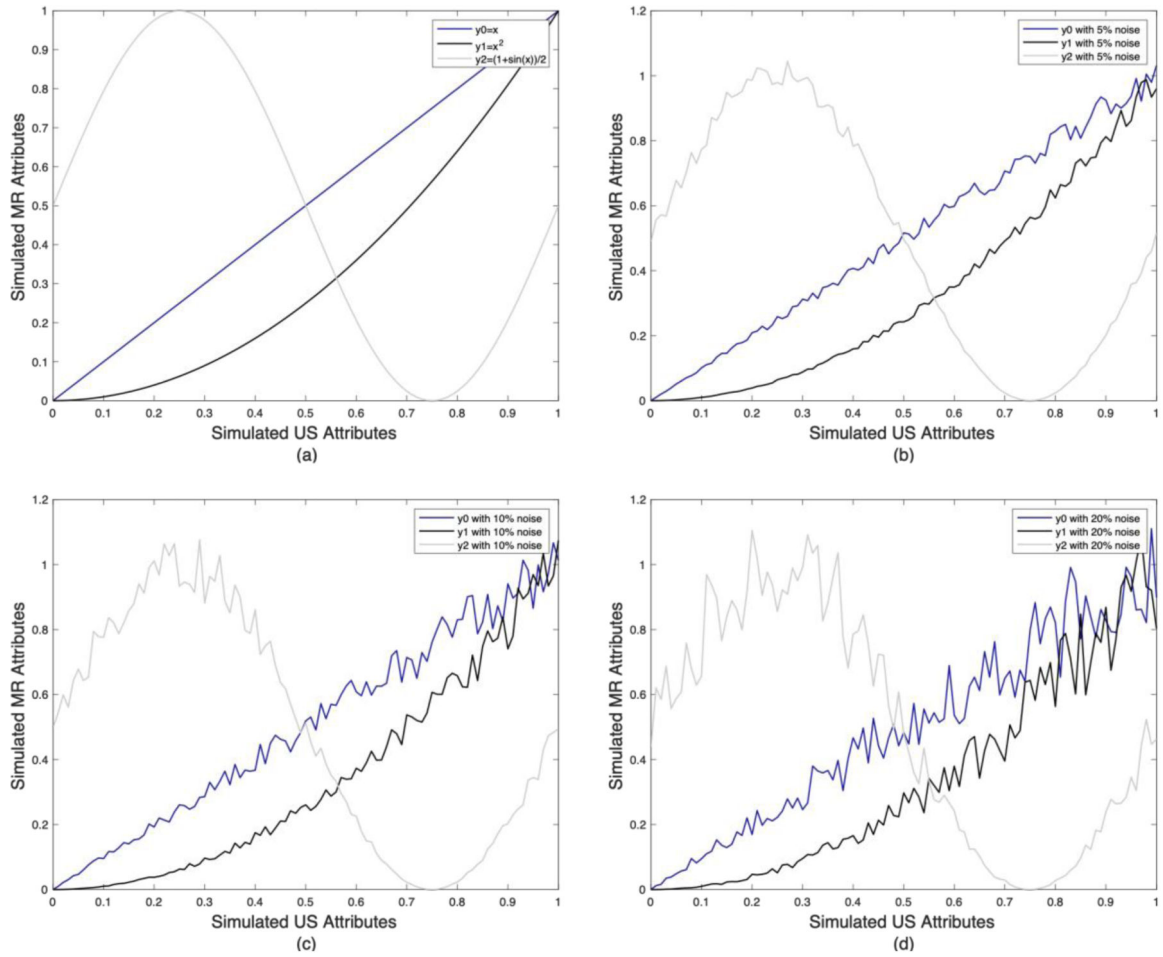


Figure 3:

Multi-scale and multi-orientation Gabor attributes in a patient's MR image (left panel) and US image (right panel) visualizing the benefit of using attributes over intensities for image matching. Regions 1–3 show the inconsistent relationship of MR and US image intensities disqualifying the use of NCC or CR directly on image intensities. Multi-scale and multi-orientation Gabor attributes are extracted from (a) preoperative MR and (b) intraoperative US to better handle the registration challenges.



Similarity Measure	*Linear relationship* ($x \leftrightarrow y_0$)	*Non-linear monotonic relationship* ($x \leftrightarrow y_1$)	*Non-linear non-monotonic relationship* ($x \leftrightarrow y_2$)
aSSD (Eqn. (2))	1.0 → 0.98 → 0.90 → 0.69	0.23 → 0.23 → 0.22 → 0.20	0.026 → 0.026 → 0.027 → 0.026
aNCC (Eqn. (3))	1.0 → 1.0 → 0.99 → 0.97	0.97 → 0.97 → 0.96 → 0.96	0.77 → 0.77 → 0.76 → 0.76
aCR (Eqn. (4))	1.0 → 1.0 → 1.0 → 1.0	1.0 → 1.0 → 1.0 → 1.0	1.0 → 1.0 → 0.98 → 0.94

Figure 4:

Comparison of aSSD, aNCC, and aCR in simulated data. In each panel, blue curves are the simulated linear relationship ($y_0 = x$), black curves are the simulated non-linear but monotonic relationship ($y_0 = x^2$), and the gray curves are the simulated non-linear but non-monotonic relationship ($y_2 = \sin(2\pi x)$. Upper left: no noise; Upper right: 5% noise added on the y-axis. Lower left: 10% noise added on the y-axis; Lower right: 20% noise added on the y-axis. In the table under the four panels, each cell shows how the similarity values change when 0%, 5%, 10%, and 20% noises, respectively, are added to y_0, y_1 and y_2, corresponding to figure panels *a-d*.

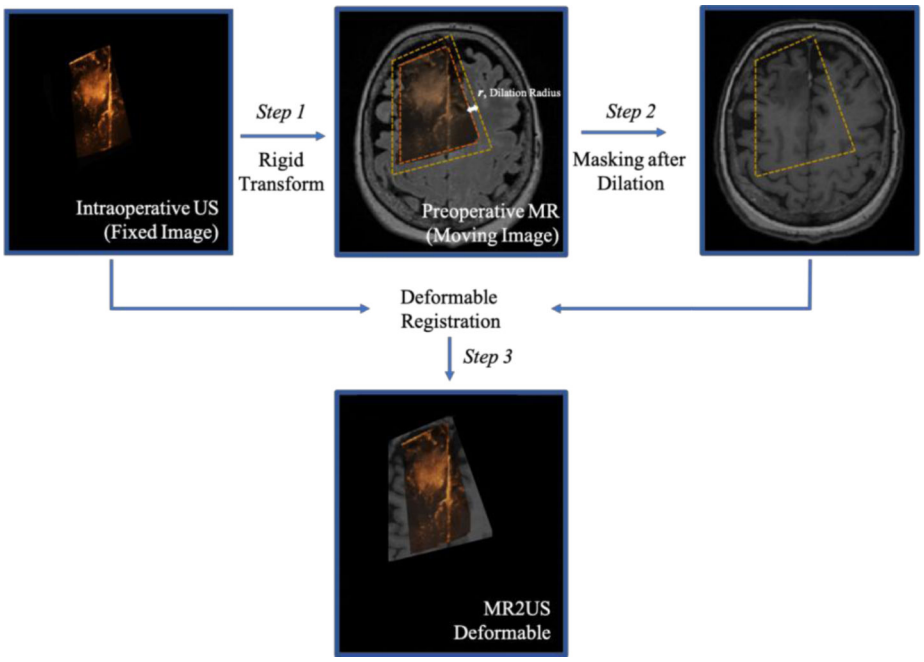


Figure 5:

Sketch of the “rigid+masking+deformable” transformation strategy. US is the fixed image and MR is the moving image throughout the process. The MR image is rigidly aligned to the US (Step 1), masked by the foreground binary mask of the US image with a dilation margin r (Step 2), and the masked rigidly-transformed MR image patch is then deformed onto the US image (Step 3).

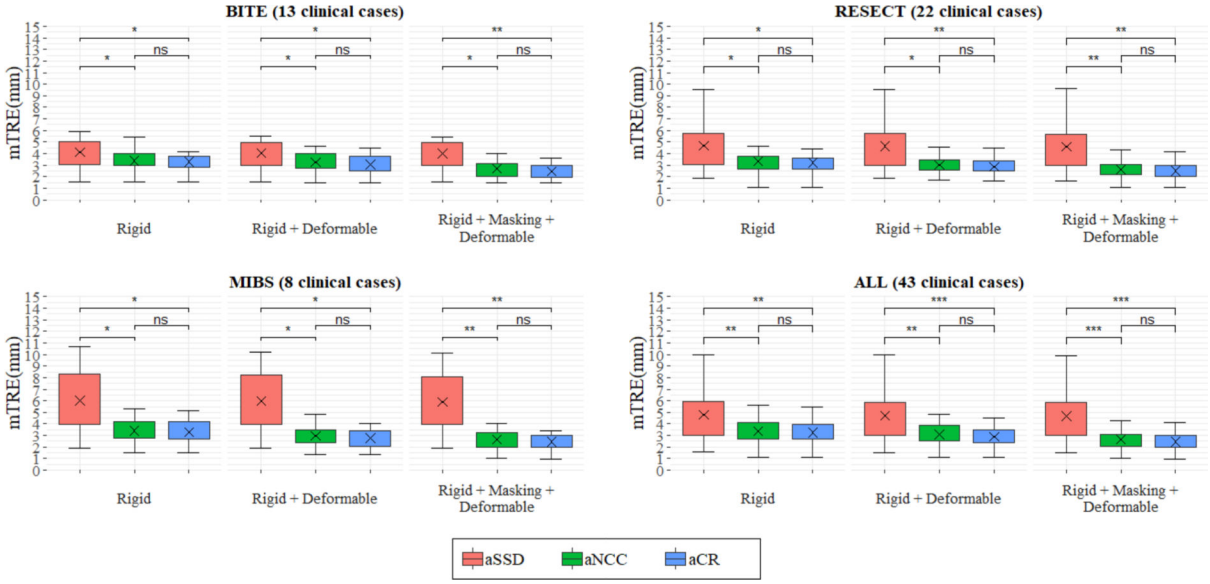


Figure 6.

Distributions of mTRE for various similarity measures and transformation strategies. Pair-wise T-test is used to determine statistically significant difference: * : $p < 0.05$; **: $p < 0.01$; *** : $p < 0.001$, ****:$p < 0.0001$ and ns: $p > 0.05$(not significant).

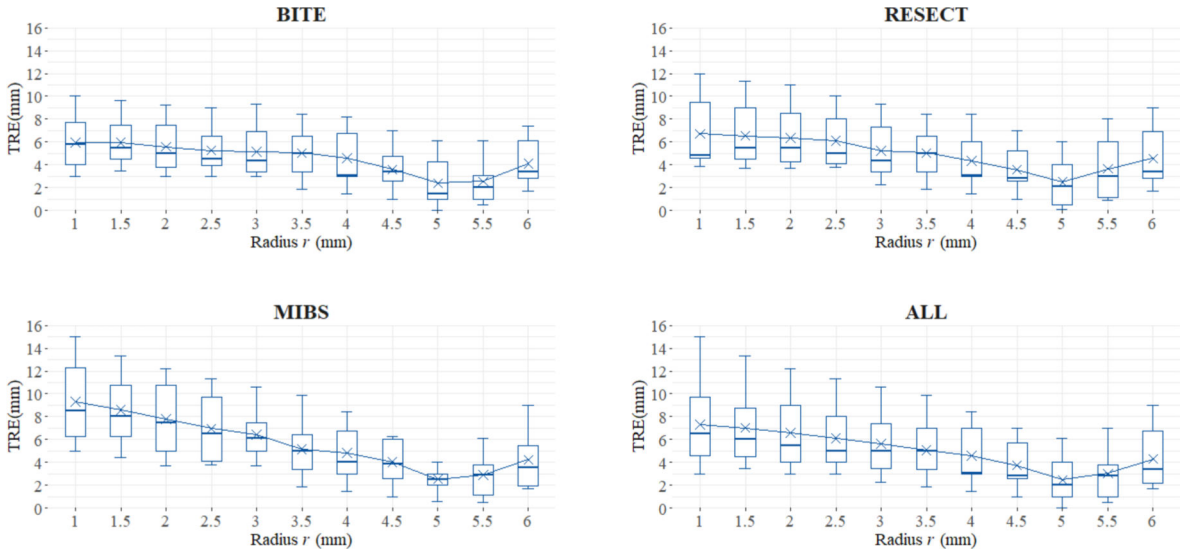


Figure 7.

Registration error (mTRE) as a function of dilation radius *r*. There are four panels: one for each of three datasets and one for all the data combined. We used aCR as the similarity measure and "rigid+masking+deformable" as the transformation strategy. Pair-wise T-test is used to determine statistically significant difference* : $p < 0.05$; **: $p < 0.01$; *** : $p < 0.001$, **** : $p < 0.0001$ and ns: $p > 0.05$(not significant).

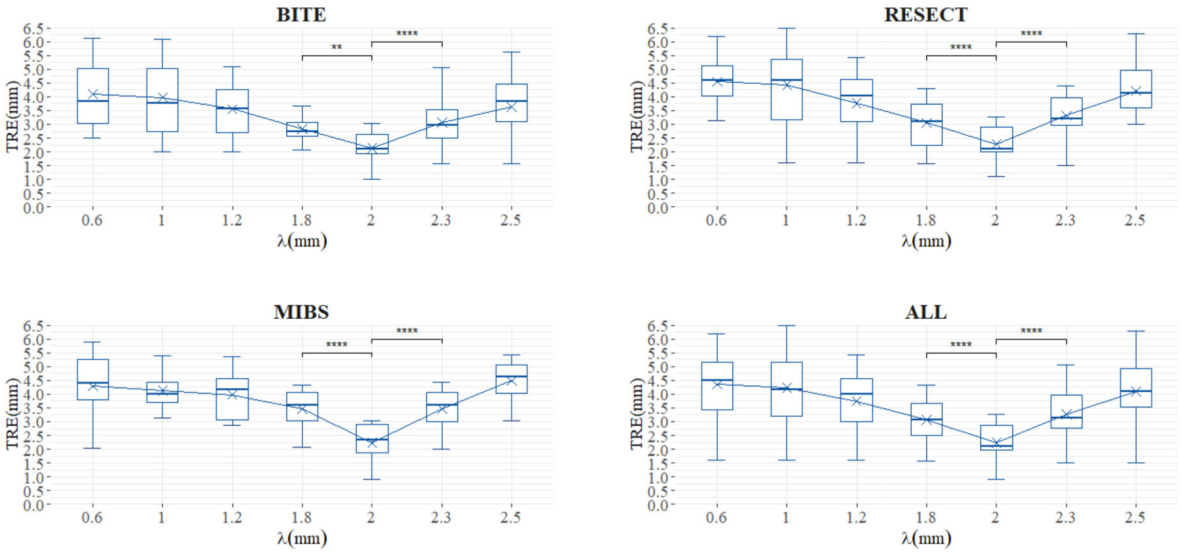


Figure 8.
Registration errors (mTRE) at various deformation smoothness levels.

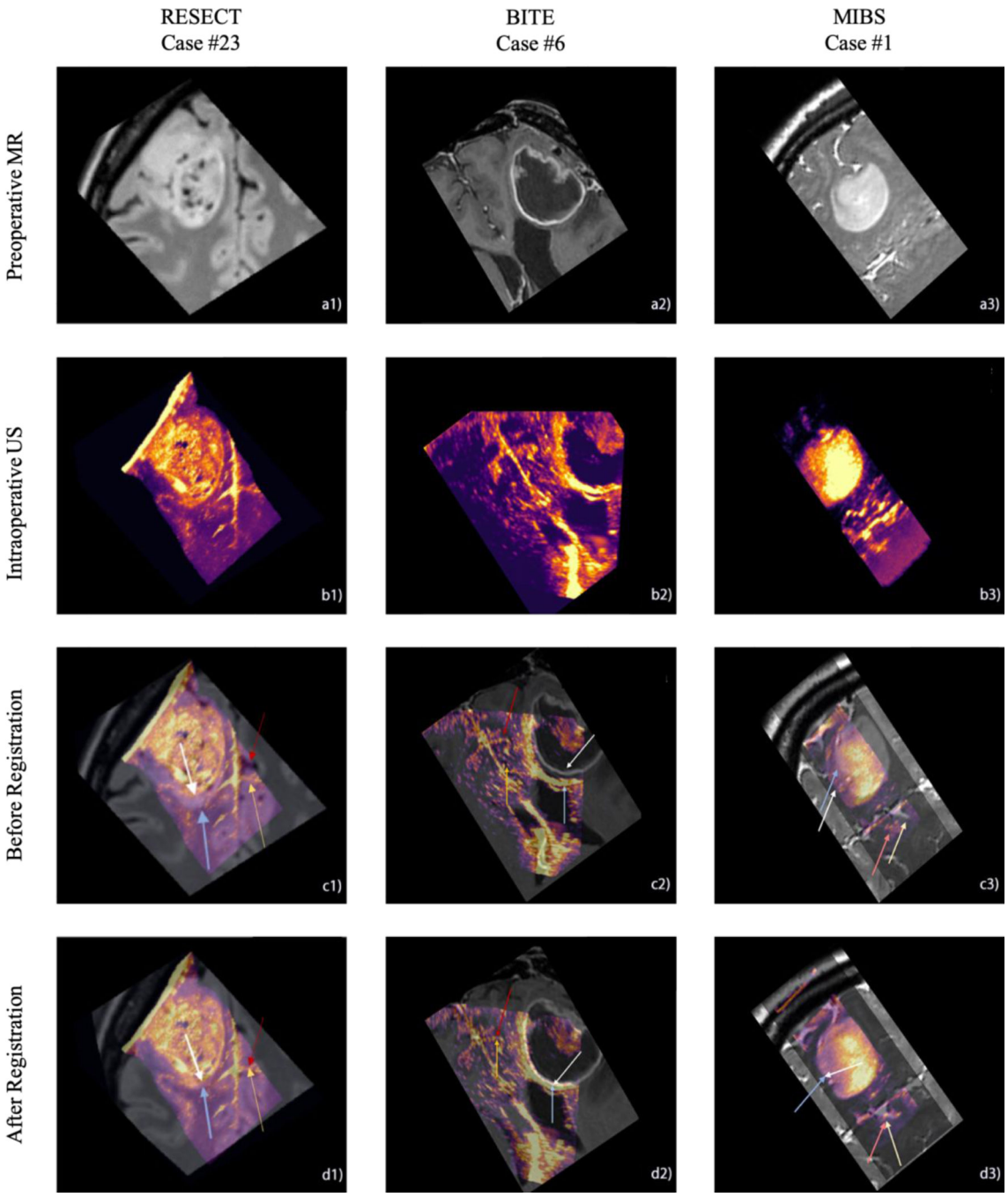


Figure 9.

Preoperative MR to intraoperative US registration in three randomly chosen patients: RESECT Case #23, BITE Case #6 and MIBS Case #1. Row designations: a) preoperative T2-FLAIR MR; b) intraoperative US; c) MR superimposed on US before registration; d) MR superimposed on US after registration. Arrows indicate tumor boundaries (blue and white) and sulci/falx (red and yellow).

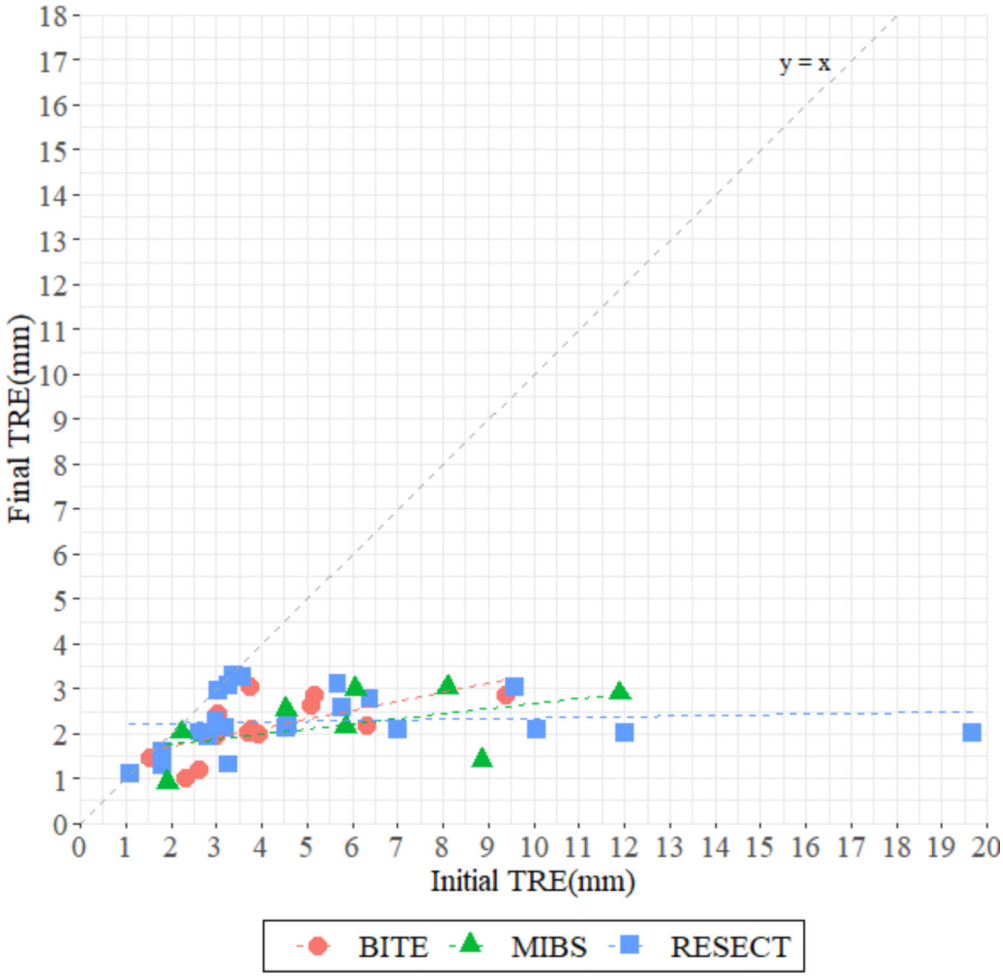


Figure 10.

Brain Shift represented by mTREs before and after cDRAMMS registration. Each dot represents one patient. Dots in red circles, green triangles, and blue squares represent patients in the BITE, MIBS, and RESECT datasets, respectively.

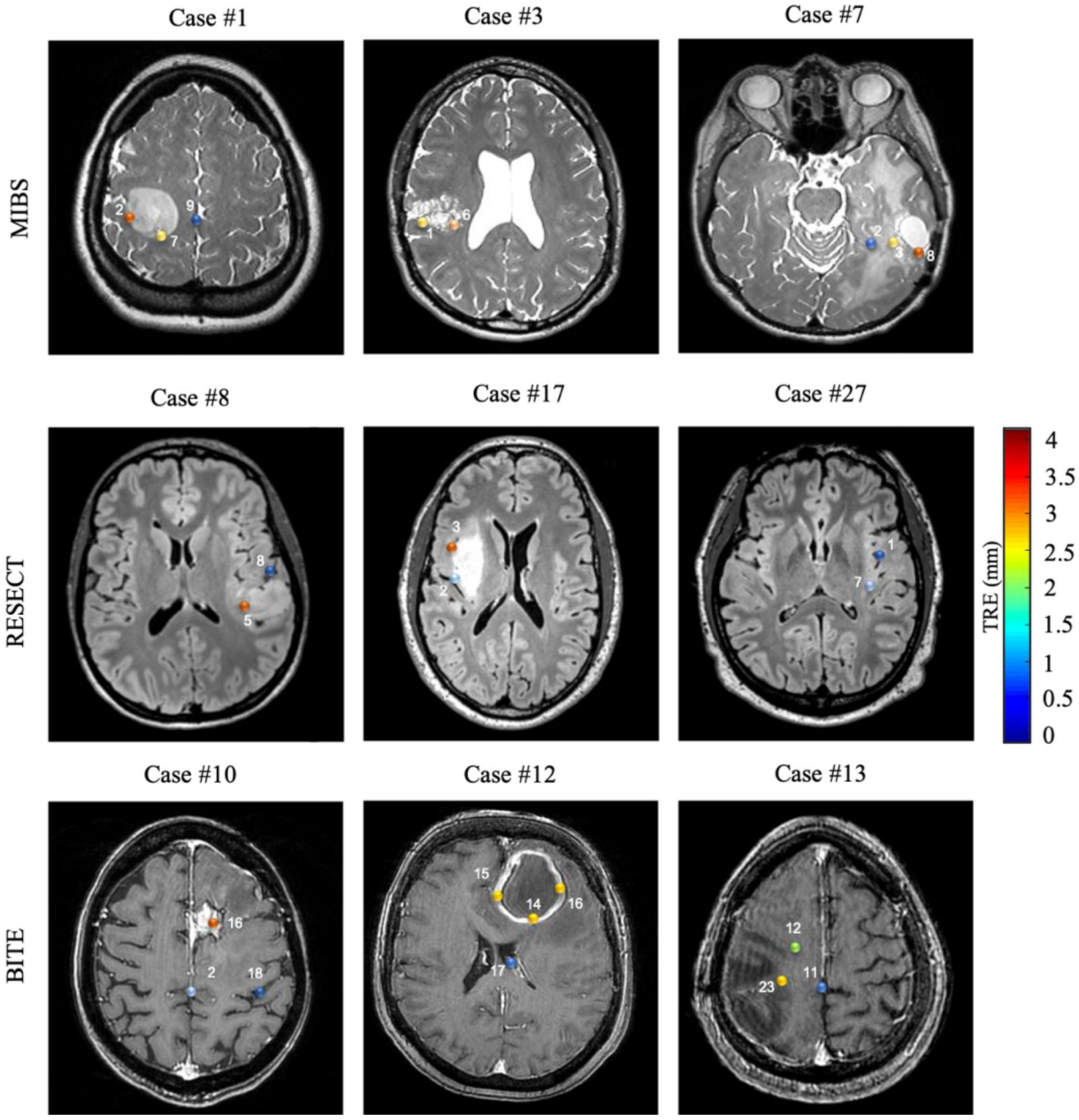


Figure 11.

Landmark errors in representative axial MR slices for nine randomly selected patients from three datasets. Color encodes the magnitude of errors: blue for small (0−1 mm) and red for large errors (3mm or above).

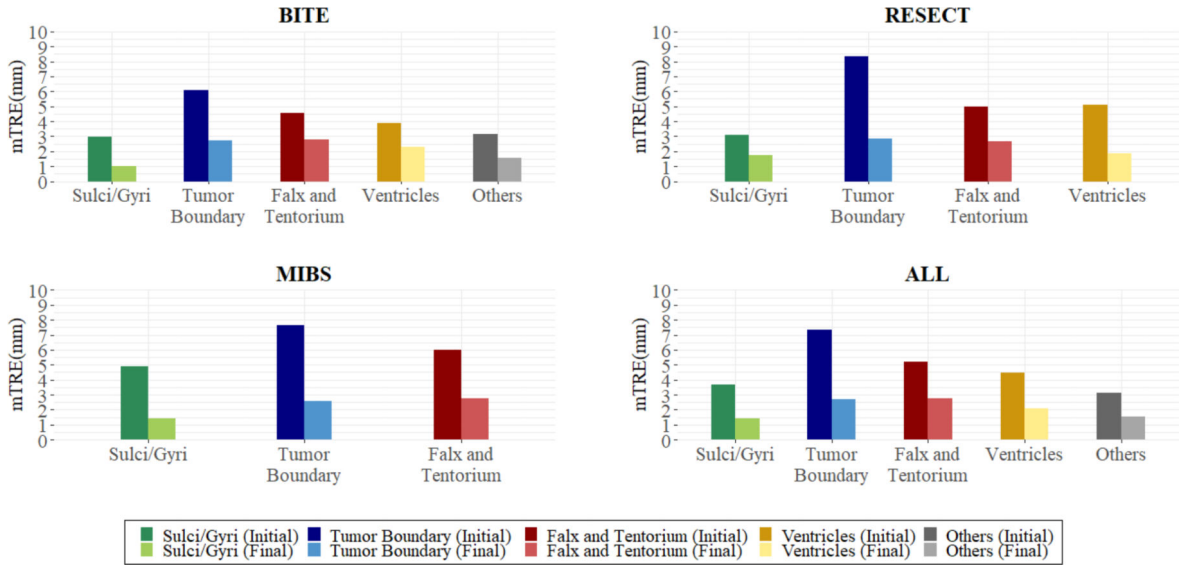


Figure 12:

Landmark errors in various brain regions. Different colors indicate different brain regions, with dark and light color for the same type of before and after registrations.

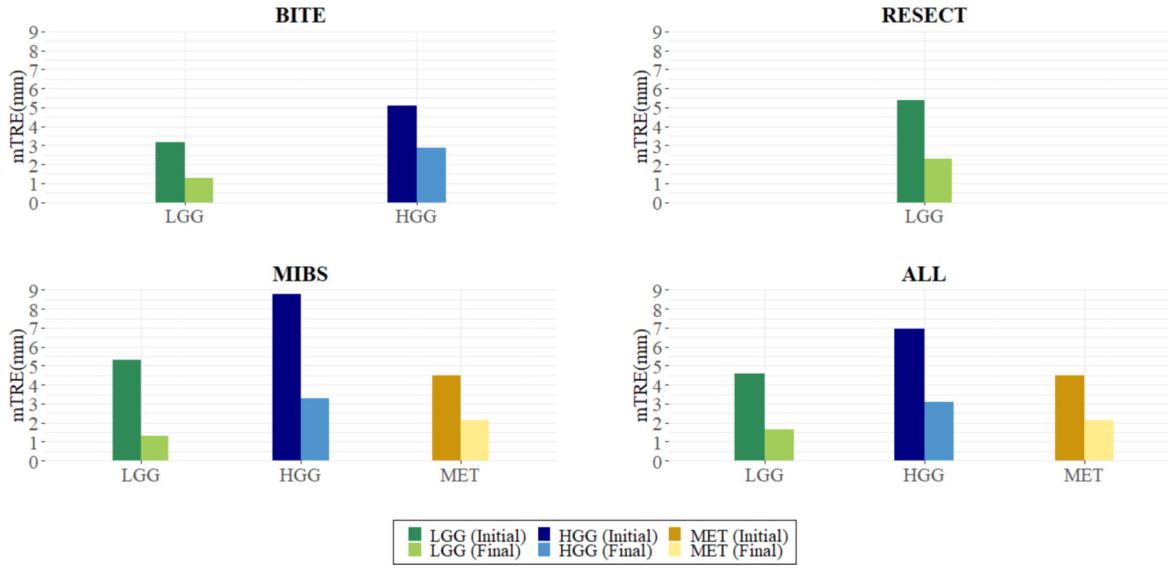


Figure 13.
Landmark errors in different tumor types before and after registration.

Table 1.

Detailed information for three datasets used in this study. They cover a wide variety of patient demographics, MRI and US protocols, and landmark distributions.

			RESECT	BITE	MIBS
Imaging Site (a)			**St. Olavs University Hospital, Trondheim, Norway**	**Montreal Neurological Institute, Montreal, Canada**	**Brigham and Women's Hospital, Boston, USA**
Patient and Tumor Demographics (b)	**No. of Patients (No. Male)**		22[1]	13 (9 males)	8 (5 males)
	Ages (years): Mean±SD (ranges)		>18[1]	52.77±18.16 (23–76)	46.37±15.08 (27–69)
	Tumor volume[2] (cm^3): Mean±SD (range)		40.0±44.09 (1.4–165.9) (Munkvold et al.,2017)	34.65±23.62 (0.2–79.2)	21.93±17.90 (3.1–57)
	Tumor Diagnosis: Astrocytoma (32.6%) Oligodendroglioma (34.9%) Oligoastrocytoma (4.7%) Glioblastoma (25.6%) Metastatic Brain Tumor (2.3%)		Astrocytoma (10) Oligodendroglioma (10) Oligoastrocytoma (2)	Astrocytoma (3) Oligodendroglioma (3) Glioblastoma (7)	Astrocytoma (1) Metastasis (1) Oligodendroglioma (2) Glioblastoma (4)
	Tumor Type[3] LGG (67.4%) HGG (30.2%) MET (2.3%)		LGG (22)	LGG (4), HGG (9)	LGG (3), HGG (4), MET (1)
	Tumor Location: Left Hemisphere (51.2%) Right Hemisphere (48.8%) Frontal lobe (51.2%) Temporal lobe (20.9%) Insular lobe (9.3%) Parietal lobe (7%) Parietoccipital lobe (4.7%) Frontoparietal lobe (4.7%) Temporoparietal lobe (2.3%)		Right brain: - Frontal (5), - Temporal (3), - Insular (2), - Frontoparietal (1) Left brain: - Temporal (3) - Frontoparietal (1) - Insular (2) - Frontal (5)	Right brain: - Frontal (2), - Temporal (1), - Parietal (2) Left brain: - Parietoccipital (1) - Frontal (6), - Temporal (1)	Right brain: - Frontal (4), - Temporal (1) Left brain: - Parietal (1), - Parietoccipital (1), - Temporoparietal (1)
MRI Protocol (c)	**MRI Machine**		1.5T Siemens Magnetom Avanto	1.5T GE Signa	3T Siemens Magnetom Verio
	MRI Sequences		T2-FLAIR	T2 and T1w gadolinium enhanced MR	T1 and T2-FLAIR
	Date of Acquisition		1 day before the surgery	avg. 17 days pre-surgery (range = 1–72 days)	avg. 18 days pre-surgery (range = 1–77 days)
	Time of Echo and Repetition		388ms and 5000ms	8ms and 23ms	232ms and 2000ms
	Flip Angle		120°	20°	120°
	Voxel Size		1.0 × 1.0 × 1.0 mm^3	1.0 × 1.0 × 1.0 mm^3	1.0 × 1.0 × 1.0 mm^3
US Protocol (c)	**US Frequency Range**		5 – 15 MHz	4 – 7 MHz	3.8 – 10 MHz
	No. of Frames		200 to 400	200 to 600	100 to 300
	Resolution of US Reconstruction		0.14 × 0.14 × 0.14 mm^3 to 0.24 × 0.24 × 0.24 mm^3	0.3 × 0.3 × 0.3 mm^3(Mercier et al., 2011)	0.5 × 0.5 × 0.5 mm^3
Landmark Distribution (d)	**Landmark pairs/patient**		15–16	19–40	5–10
	No. Landmark pairs		338	355	65
	Dispersion of landmarks[4]		336.25±155.38	828.28±237.56	529.43±194.96
	Landmark Locations	Sulci/Gyri	206	198	25

			RESECT	BITE	MIBS
Imaging Site (a)			**St. Olavs University Hospital, Trondheim, Norway**	**Montreal Neurological Institute, Montreal, Canada**	**Brigham and Women's Hospital, Boston, USA**
		Tumor Boundary	123	35	26
		Falx and Tentorium	8	47	14
		Ventricles	1	62	–
		Others[5]	–	13	–

[1]RESECT does not provide detailed age and gender information (Xiao et al., 2017).

[2]Tumor volume is based on manually drawn boundaries in MR images.

[3]Tumor type is categorized in low-grade glioma (LGG), high-grade glioma (HGG) or metastasis (MET)

[4]The spatial dispersion of the 3D point cloud is measured by the trace of the covariance matrix of the coordinates of points where higher traces correspond to higher degrees of dispersion (Hladuvka, 2003)(Rizzini & Caselli, 2007).

[5]"Others" means vessels, mid brain, pons, bone, white matter and interpeduncular cistern, where landmarks appear less frequently.

Table 2:

Major components of sixteen MR-US registration algorithms we evaluated in this paper. An asterisk (*) denotes algorithms with self-reported accuracies and # if authors have participated in the CuRIOUS 2018 challenge for the third-party independently reported accuracies. "_" under the "Publicly Released" column means the algorithm is not yet publicly released. Abbreviations: CMA-ES – Covariance matrix adaptation evolutionary strategy; FFD – Free Form Deformation based on cubic B-splines; SGD – Stochastic Gradient Descent; DO – Discrete Optimization; BOBYQA – Bound Optimization by Quadratic Approximation; CCOD – Closest commuting operator distance.

Author, Publication Yr.	Algorithm	Similarity Measure	Deformation Model	Optimization Strategy	Publicly Released
(Nigris, Collins, & Arbel, 2013)	Gradient orientation (*)	Local gradient orientation	Rigid	CMA-ES[1]	–
(Rivaz, Karimaghaloo, & Collins, 2014)	SeSaMI (*)	α-MI	FFD	SGD	_
(Rivaz, Karimaghaloo, Fonov, et al., 2014)	CoCoMi (*)	Contextual Conditioned Mutual Information	FFD	SGD[1]	–
(Rivaz et al., 2015)	RaPTOR (*)	CR	FFD	SGD	_
(Jiang et al., 2016)	miLBP (*)	miLBP	FFD	DO	_
(Masoumi, Xiao, & Rivaz, 2018b)	MARCEL (*)	CR	Affine	SGD	_
(Hong et al., 2018)	SS+Demons (#)	Demons (Vercauteren, Pennec, Perchant, & Ayache, 2007)			_
(Drobny et al., 2018)	Symmetric Block-Matching Based Approach (NiftyReg) (#)				Yes
(Heinrich, 2018)	SSC (#)	Self-similarity context metric	FFD	DO	Yes
(Wein, 2018)	LC2 (#)	LC2	Demons	BOBYQA	Partially[2]
(Shams, Boucher, and Kadoury, 2018)	LC2+P (#)	LC2+P[3]	FFD	BOBYQA	_
(Masoumi et al., 2018a)	ARENA (*)	CR	Affine	CMA-ES[1]	–
(Zimmer et al., 2019)	Laplacian Commutators (*)	CCOD	FFD	SGD	_
(Zhong et al., 2018)	MLP (#)	Learning-based Approaches			_
(Sun & Zhang, 2018)	CNN+STN (#)				_
cDRAMMS (#)		aNCC and aCR	Rigid+Masking +FFD	DO	Yes

[1]The authors used a similar version of stochastic gradient descent optimization (please see (Rivaz et al., 2014) for additional details).

[2]The similarity calculator is freely released, whereas the whole registration is implemented in the ImFusion SDK platform that is author proprietary material.

[3]The registration includes a rigid step with LC2 and a non-rigid step with LC2+P where P is a pixel weighting term.

Table 3:

Publicly-available tools for MR-US registration usually use different parameter settings for different datasets. SSC and NiftyReg are fully released, whereas LC2 only has the similarity metric tool released.

Registration Tool (Reference)	**Variable**	***for* BITE Dataset**	***for* RESECT Dataset**
SSC (used for BITE in (Heinrich et al., 2013) and for RESECT in (Heinrich, 2018))	Number of coarse-to-fine pyramid levels	3	5
	Control point spacing	{6, 5, 4}mm in 3 levels	{4, 3.5, 3, 2.5, 2}mm in 5 levels
	Search range	{12, 5, 2}mm in 3 levels	{6.4, 6.125, 6, 4.125, 8}mm in 5 levels
	Discretization of search space	[6, 5, 4] intervals in 3 levels	[5, 4, 3, 2, 1] intervals in 5 levels
	Patch size	3×3×3 voxels = 1.5×1.5×1.5mm	3×3×3 voxels = 1.5×1.5×1.5mm
	Weight for regularizer	0.5	1.6
NiftyReg (used for RESECT in (Drobny et al., 2018))	Coarse-to-fine pyramid levels	--	2
	Iterations per level	--	10 (coarse level), 5 (fine level)
	Block size	--	4×4×4 voxels
	% of blocks	--	25% (blocks with highest intensity variations)
LC2 (used for BITE in (Wein et al., 2013); for RESECT in (Wein, 2018))	Patch size	7×7×7 voxels	7×7×7 voxels
	Transformation	Free-form Cubic B-Spline	Rigid

Table 4:

Pre- and post-registration errors (mTRE) for sixteen algorithms in three datasets. The top three ranked algorithms in each subset are noted in bold text, underlined, and in italics, respectively. Results are mostly from publications from the authors of these algorithms, except for SSC (parameter #1) in BITE and MIBS, SSC (parameter #2) in RESECT training patients and MIBS, and NiftyReg in BITE and MIBS.

	BITE	RESECT		MIBS
		Training	Testing	
Initial mTRE (mm)	4.18±1.97 (Mercier et al., 2012)	5.37±4.27 (Xiao et al.,2017)	6.41±4.46 (Xiao et al., 2019)	6.18±3.38
ARENA	2.82±0.71 (Masoumi et al., 2018a)	2.77±1.13 (Masoumi et al., 2018a)	-	-
CoCoMi	3.22 ±1.70 (Rivaz et al., 2014)	-	-	-
CNN+STN	-	3.91±0.53 (Sun & Zhang, 2018)	-	-
Gradient Orientation	2.57±0.74 (Nigris et al., 2013)	-	-	-
Laplacian Commutators[1]	3.01±1.22 (Zimmer et al., 2019)	-	-	-
LC2 (Parameter #1)	-	*1.75±0.62* (Wein, 2018)	**1.57±0.55**	-
LC2 (Parameter #2)	2.52±0.87 (Wein et al., 2013)	-	-	-
LC2+P	-	4.6±3.4 (Shams et al., 2018)	6.62±2.80	-
MARCEL	-	2.32±0.68[2] (Masoumi et al., 2018b)	-	-
RaPTOR	2.65±0.98[3] (Rivaz et al., 2015)	-	-	-
miLBP	2.15±1.10 (Jiang et al., 2016)	-	-	-
MLP	-	**1.21±0.55** (Zhong et al., 2018)	5.72±2.82	-
SeSaMI	2.44±0.40 (Rivaz et al., 2014)	-	-	-
SSC (Parameter #1)	*2.18±0.37*	1.67±0.54 (Heinrich, 2018)	1.87±0.51	3.61±0.82
SSC (Parameter #2)	2.34±0.52 (Heinrich et al., 2013)	2.52±0.60	-	2.91±0.88
SS+Demons	-	5.60±3.94 (Hong et al., 2018)	6.55±4.55	-
NiftyReg	2.67±0.58	2.90[4] (Drobny et al., 2018)	3.24±3.63	*3.18±0.62*
Ours (cDRAMMS)	**2.08±0.37**	2.28±0.71 (Machado et al., 2018)[5]	*2.19±0.87*	**2.24±0.78**

[1] This group uses nine of 13 BITE clinical cases (#2,3,5,6,10–14).

[2] MARCEL was validated in five of 22 RESECT cases. They reported an initial mTRE equal to 5.13±2.78 mm (N=5).

[3] RaPTOR was validated in a previous version of the BITE dataset in which the mTRE pre-registration was 5.92±3.15 mm and the number of landmarks per case ranged from nine to 21.

[4] This group reported only the mean mTRE for all patients in the RESECT training subset.

[5] An earlier submission was used in (Machado et al., 2018).

Table 5.

Visual grading of registration accuracy by two neurosurgery fellows of 43 patients from 3 datasets.

Grader	Grading		
Neurosurgeon	**Bad**	**Good**	**Great**
P.U.	0	21 (49%)	22 (51%)
W.E.	0	19 (44%)	24 (56%)