

Multi-Modal Robust Inverse-Consistent Linear Registration

Christian Wachinger^{a,b*}, Polina Golland^a, Caroline Magnain^b, Bruce Fischl^{a,b}, Martin Reuter^{a,b}

^a Computer Science and Artificial Intelligence Lab, MIT

^b A.A.Martinos Center for Biomedical Imaging, Dept. of Radiology,
Massachusetts General Hospital, Harvard Medical School

Abstract

Registration performance can significantly deteriorate when image regions do not comply with model assumptions. Robust estimation improves registration accuracy by reducing or ignoring the contribution of voxels with large intensity differences, but existing approaches are limited to mono-modal registration. In this work, we propose a robust and inverse-consistent technique for cross-modal, affine image registration. The algorithm is derived from a contextual framework of image registration. The key idea is to use a modality invariant representation of images based on local entropy estimation, and to incorporate a heteroskedastic noise model. This noise model allows us to draw the analogy to iteratively reweighted least squares estimation and to leverage existing weighting functions to account for differences in local information content in multi-modal registration. Further, we employ the non-parametric windows density estimator to reliably calculate entropy of small image patches. Finally, we derive the Gauss-Newton update and show that it is equivalent to the efficient second-order minimization for the fully symmetric registration approach. We illustrate excellent performance of the proposed methods on datasets containing outliers for alignment of brain tumor, full head, and histology images.

Keywords: Computer-assisted image analysis, rigid and affine image alignment, statistical model, magnetic resonance imaging, optical coherence tomography, histology

1. Introduction

Tremendous progress in image registration has led to a vast number of new algorithms over the last decades. One of the important aspects for translating these algorithms from research to clinical practice is their robustness. Clinical images tend to frequently violate model assumptions that were incorporated in the derivation of the algorithm. Typical sources for such inconsistencies are imaging artifacts, noise, anatomical changes, and partial image information. Registration methods based on concepts from robust estimation have been proposed to address such issues by identifying outliers in the images and restricting their influence on the cost function (Reuter et al., 2010). These techniques are targeted at minimizing weighted intensity differences and therefore cannot account for more complex intensity relationships, prevalent in multi-modal images. However, multi-modal registration could largely benefit from robust approaches because inconsistencies arise from the very nature of multi-modal acquisitions: the depiction of complementary information. Structural information that is only available in selected image modalities thus adds to the inconsistencies already observed in mono-

modal registration and therefore highlights the need for *robust multi-modal registration*.

Furthermore, several applications require the calculation of inverse consistent (i.e. unbiased) transformations, that are independent of the order of the two passed images. Usually, an asymmetric registration approach is taken, where only one of the images is transformed during registration and consequently only this image is resampled onto the grid of the fixed image. Even in global (rigid/affine) registration, this asymmetry in the resampling process introduces artifacts to only one of the images and has been shown to bias follow-up analyses (Yushkevich et al., 2010; Thompson and Holland, 2011; Fox et al., 2011; Reuter and Fischl, 2011). *Inverse-consistent* approaches avoid this processing bias and have been a focus of research in registration (Christensen and Johnson, 2001; Smith et al., 2002; Avants et al., 2008; Vercauteren et al., 2009; Reuter et al., 2010). A fully symmetric registration setup where both images are mapped into the mid-space avoids such problems and produces inverse-consistent results. However, existing symmetric approaches describe mono-modal registration settings and focus mainly on non-linear transformation models. The reliable and unbiased inclusion of multi-modal image stacks into modern image processing pipelines therefore requires an *inverse-consistent multi-modal registration*.

In this article, we address these requirements and intro-

*Corresponding Author. Address: 32, Vassar St., D462, Cambridge, MA, 02139. Tel: +1 6176427026. Email: wachinger@csail.mit.edu

duce a *robust, inverse-consistent, multi-modal* registration algorithm. The method is designed to produce a highly accurate affine alignment of images in the presence of potentially severe local outliers. We derive the new registration algorithm via a probabilistic framework that incorporates layers of latent random variables. The layers provide a structural representation of images, transforming the initial multi-modal setup into a mono-modal registration problem via localized entropy estimation (Wachinger and Navab, 2012b). We incorporate a heteroskedastic Gaussian noise model for the similarity term that operates on the latent layers, allowing for *spatially varying* noise. This construction permits employing techniques from robust estimation to model local noise variation via an iteratively re-weighted least squares estimation. Moreover, we derive the Gauss-Newton method with symmetric transformation updates in the Lie group and show that this approach corresponds to an efficient second-order minimization achieving a vastly improved convergence rate and frequency, as it uses gradient information from both images.

1.1. Applications

The **first** target application is the registration of multi-modal, intra-subject, full head images with the objective to accurately align brain structures. As these images are collected very close in time (often within the same scan session) we expect only minimal changes in the brain between these multi-modal image pairs, but potentially large differences in surrounding regions, *e.g.*, in soft tissue, different jaw, tongue, neck and eye placement, or different cropping planes due to subject motion. Applying deformable registration here is not meaningful as brain displacements are only rigid (or at most affine, depending on the acquisition). Non-linear registration bears the risk of propagating unwanted deformations into the brain by regularization constrains. In full head registration, however, standard global alignment can fail because of the non-matching structures and may terminate in local minima. A robust approach helps to overcome this problem by identifying non-matching structures as outliers and by iteratively reducing their influence on the registration.

An alternative approach to this problem, would be to align skull-stripped images and thus remove most of the local differences outside the brain. Unfortunately, brain extraction tools often work only on a single modality. Even when they can be applied to both images independently, they will lead to varying brain masks, as different parts of dura or skull may be included or removed by the algorithm. A robust registration approach is capable of identifying these regions and can produce highly accurate alignments of the brain independent of skull stripping errors. Common cross-modal registration procedures are often severely influenced by strong boundary edges and thus may drive the registration towards a good alignment of the boundary, sacrificing accuracy in the interior. *Consequently, the proposed robust registration method can assist in pre-registering full head images for the simultaneous*

extraction of the brain in both modalities or to accurately align independently extracted brain images.

A **second** motivating application for our method is the alignment of brain tumor images. Tumors have a different appearance depending on the acquisition parameters and the usage of contrast agent (*e.g.*, Gadolinium in MR angiography). The largely varying appearance can deteriorate the performance of common multi-modal registration algorithms. Approaches based on robust statistics are useful in this context; they automatically discount those regions and recover the correct alignment based on the remainder of the image. Additionally, accurate alignment of full head images can also be beneficial for tumor images, since the automatic brain extraction is very challenging in these cases. Common brain extraction tools use prior knowledge about intensity distributions to extract the brain. Since the intensity values vary significantly in tumor regions, such methods fail frequently in the presence of pathology.

A **final** application is the mapping of histology images to optical coherence tomography (OCT) (Huang et al., 1991; Yaqoob et al., 2005) or to high resolution MR images. Accurate registration with histology is necessary to validate the appearance of structures in OCT or MR based on their true appearance on the histological slices. Furthermore, accurate correspondence will be invaluable to guide development and validation of new algorithms for automatic detection of areal and laminar boundaries in the human brain, or of the extent of infiltrating tumors. Histology registration can be extremely challenging due to tears and deformations that occur during slicing, mounting and staining. These artifacts are not present in the previously acquired OCT or MR images and give rise to complex registration problems that can be alleviated when using our proposed robust cross-modal algorithm.

1.2. Related Work

Stewart (1999) presents a general summary of robust parameter estimation in computer vision. Most robust registration techniques focus on *mono-modal* registration. Nestares and Heeger (2000) propose a method based on M-estimators for the robust registration of images. Periaswamy et al. (2006) introduce a method based on a local neighborhood and employ the expectation-maximization algorithm to handle partial data. Reuter et al. (2010) introduce a robust, inverse-consistent registration approach, where both images are transformed into a mid space. Similar to Nestares and Heeger (2000) robust statistics are employed to obtain accurate registration in the presence of temporal differences. This approach was shown to lead to superior results for alignment of mono-modal images acquired in the same session, as well as across time in longitudinal studies (Reuter et al., 2012).

In addition to related methods for mono-modal registration, a few studies have reported results for robust *multi-modal* registration. Itti et al. (1997) extract the brain surface and subsequently perform an iterative anisotropic

chamfer matching. A hybrid registration approach, based on the extraction of the brain surface and intensity information is proposed in (Greve and Fischl, 2009). Wong and Orchard (2009) employ the residual of the local phase coherence representation for registration. This is similar to local phase mutual information (Mellor and Brady, 2005), where mutual information is calculated between local phase estimates from the image intensities for multi-modal registration.

Here, we transform the multi-modal registration problem to a mono-modal one. Several structural representations have been proposed in the literature for this purpose. Andronache et al. (2008) recolor images, depending on the variance in the images. Other algorithms extract edges and ridges from the images (Maintz et al., 1996) or correlate image gradient directions (Haber and Modersitzki, 2007). Wachinger and Navab (2012b) study the theoretical properties of structural representations and propose two approaches: an entropy representation, and Laplacian images with spectral decomposition for dimensionality reduction. Entropy images can easily and quickly be calculated via a dense grid of patches. While they bear similarities to gradient images, they better represent complex configurations such as triple junctions (Wachinger and Navab, 2012b). Shechtman and Irani (2007) employ a concept similar to the Laplacian images, however, construct the self-similarity descriptor based only on local information. Heinrich et al. (2012) also use local self-similarity measures for multi-modal registration, but without dimensionality reduction. With the length of the information vector corresponding to the size of the neighborhood, this approach is limited to small neighborhoods due to memory and computational complexity constraints.

In this study, we focus on affine and rigid transformation models with low degrees of freedom. Many applications require a highly accurate global registration, e.g. to align different image modalities within subject in close temporal proximity. Non-linear registrations across subjects, or across time, are usually performed within-modality using a high-resolution image or full images stacks. Furthermore, non-linear registrations rely on an accurate global alignment as a pre-processing step. There are a number of freely available registration software packages. The widely used registration tool FLIRT (Jenkinson et al., 2002), part of the FSL package (Smith et al., 2004), implements several intensity based cost functions such as sum of squared differences, correlation ratio and mutual information. It also contains sophisticated optimization schemes to prevent the algorithms from being trapped in local minima. The freely available SPM software package (Ashburner et al., 1999) contains a registration tool based on (Collignon et al., 1995). In our study, we compare against these two programs to evaluate the proposed registration method.

1.3. Outline

Sec. 2 introduces our robust, inverse-consistent, multi-modal registration method by using a probabilistic framework of image registration. In Sec. 3, we derive the steepest descent and Gauss-Newton optimization steps, with the latter one corresponding to the efficient second-order minimization for the proposed symmetric registration setup. The calculation of entropy images with the non-parametric windows estimator is described in Sec 4. Finally, we present experiments on several datasets in Sec 5.

2. Robust Multi-Modal Registration

In this section we detail our robust multi-modal registration approach. First, we describe the parameterization of the transformations for the symmetric setup in Sec. 2.1. Subsequently, we present a probabilistic model for robust multi-modal registration based on entropy images in Sec. 2.2. Sec. 2.3 introduces the spatially varying noise model that allows us to draw the analogy of the derived log-likelihood function to iteratively reweighted least squares estimation in Sec. 2.4. This connection permits to leverage existing robust estimators for the purpose of robust registration in Sec. 2.5.

2.1. Symmetric Transformation Parameterization

For robust registration, we are interested in transforming both images into an unbiased common space. This is achieved by a symmetric registration approach, mapping both images into the mid space (Avants et al., 2008). The result is an inverse-consistent registration (Christensen and Johnson, 2001; Zeng and Chen, 2008). The result is an inverse-consistent registration. This implies that the exact inverse transformation will be obtained by swapping the fixed and moving image. The common setup is to have one moving image I_m and one fixed image I_f together with a transformation T that operates on the moving image, $I_m(T(x)) \approx I_f(x)$, for all spatial locations on the image grid $x \in \Omega$. For the symmetric approach, we want to transform both images half way and therefore need to calculate the half transformation “ $\frac{1}{2}T$ ”. Since the space of most linear and non-linear transformations is not a vector space, we use a parameterization of transformations that has a Lie group structure. This construction ensures that we stay within the transformation space after the parameter update and facilitates calculations, such as computing the inverse transformation, as required for the symmetric approach.

More precisely, consider transformation T that is part of the Lie group \mathcal{G} with the related Lie algebra \mathfrak{g} . Each element in the Lie algebra $\mathfrak{t} \in \mathfrak{g}$ can be expressed as a linear combination $\mathfrak{t} = \sum_{i=1}^n \pi_i \cdot b_i$, with the standard basis of the Lie algebra b_1, \dots, b_n and transformation parameters

$\boldsymbol{\pi} = [\pi_1, \dots, \pi_n]$ (Zefran et al., 1998). The exponential map

$$\exp : \mathfrak{g} \rightarrow \mathcal{G} \quad (1)$$

$$\mathfrak{t} \mapsto \exp(\mathfrak{t}) = \exp\left(\sum_{i=1}^n \pi_i \cdot b_i\right) = T \quad (2)$$

assigns to an element of the Lie algebra \mathfrak{t} that is dependent on the transformation parameters $\boldsymbol{\pi}$ the corresponding transformation T . For notational convenience, we define the transformation operator $T_{\boldsymbol{\pi}}(x)$ that maps the location x with the Lie algebra parameters $\boldsymbol{\pi}$. This is a shorthand for creating a member of the Lie algebra with the basis elements b_i , applying the exponential map to obtain a transformation matrix, and using this matrix to map x .

In the space of the Lie algebra it is straightforward to calculate the mid space transformation with the half parameters $\boldsymbol{\pi}/2$ and the inverse transformation with the negative parameters $-\boldsymbol{\pi}$. The concatenation of two half-way transforms leads to the original transform

$$\begin{aligned} T_{\boldsymbol{\pi}/2} \cdot (T_{-\boldsymbol{\pi}/2})^{-1} &= \exp\left(\sum_{i=1}^n \frac{\pi_i}{2} b_i\right) \cdot \exp\left(\sum_{i=1}^n -\frac{\pi_i}{2} b_i\right)^{-1} \\ &= \exp\left(\sum_{i=1}^n \pi_i b_i\right) = T_{\boldsymbol{\pi}}. \end{aligned} \quad (3)$$

Lie group parameterizations have been applied to rigid, affine, projective, and non-linear registration (Benhimane and Malis, 2004; Vercauteren et al., 2009; Arsigny et al., 2009; Wachinger and Navab, 2009, 2013). In this study we work with rigid and affine transformations. While a standard parameterization for the rigid case exists (Zefran et al., 1998), different parameterizations were proposed for affine transformations (Arsigny et al., 2009; Kaji et al., 2013; Kaji and Ochiai, 2014). For the parameterization in (Arsigny et al., 2009), no closed form of the exponential map exists. The parameterization in (Kaji et al., 2013; Kaji and Ochiai, 2014) describes the transformation matrix as a product of two matrix exponentials, which complicates the computation of the half transform. A viable alternative in practice is to directly update the parameters of the transformation matrix, as gradient descent optimizations are unlikely to produce negative determinants when started from identity. The half transform in this case can be directly computed via the matrix square root. Note, that the matrix square root is equivalent to but $\exp(\log(T)/2)$ computationally more efficient. It can be computed via a Schur decomposition and then applying a recurrence of Parlett for computing the root of the triangular matrix.

2.2. Probabilistic Model for Robust Registration

Probabilistic modeling presents a mathematical framework to derive algorithms for image registration in which

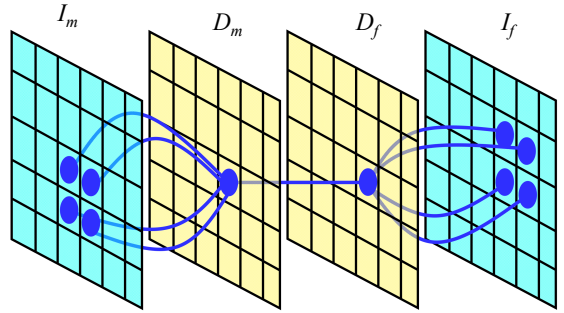


Figure 1: Contextual model for image registration illustrated in 2D. Image layers I_m and I_f consist of observed random variables. Description layers D_m and D_f are latent. Each descriptor variable at location x depends on a local image neighborhood $\mathcal{N}(x)$, which is of size 4 in the illustration. A one-to-one relationship exists between the description layers.

assumptions about the characteristics of the noise and signal are made explicit. In this work, we leverage techniques for robust estimation that require comparable intensity values, as in the mono-modal registration setup. For this purpose, we employ a probabilistic framework (Wachinger and Navab, 2012a) that introduced layers of latent random variables, called description layers. Fig. 1 shows a schematic illustration of the framework in 2D. The description layer D_m depends on I_m and D_f depends on I_f . In our study, we use the description layers to store a structural representation of images.

This framework emphasizes the importance of the local context. It replaces the common assumption of independently distributed coordinate samples with the Markov property. The Markov property allows for splitting up a joint distribution into singleton terms, where each term is conditioned on a neighborhood that contains all the relevant information. In our case, each location x in the description layer is dependent on a local neighborhood or patch $\mathcal{N}(x)$ in the image, as shown in Fig. 1. The joint distribution for the contextual framework factorizes as

$$\begin{aligned} p(I_m, I_f, D_m, D_f, \boldsymbol{\pi}) &= \prod_{x \in \Omega} p(D_m(x) | I_m(\mathcal{N}(x))) \cdot p(D_f(x) | I_f(\mathcal{N}(x))) \\ &\quad \cdot p(D_m(x) | D_f(x), \boldsymbol{\pi}), \end{aligned} \quad (4)$$

with the coupling terms $p(D_m(x) | I_m(\mathcal{N}(x)))$ and $p(D_f(x) | I_f(\mathcal{N}(x)))$ that ensure that the description layers represent the original image information, and the similarity term $p(D_m(x) | D_f(x), \boldsymbol{\pi})$ that compares both description layers. Commonly used description layers result from image filtering, image gradients, or dense

feature descriptors. Since we are interested in a modality-invariant description, we use a structural representation with entropy images, justified by applying the asymptotic equipartition property on the coupling terms (Wachinger and Navab, 2012a). Under the assumption that the information content across modalities is similar, entropy images reduce the multi-modal setup to a mono-modal registration problem. This is essential for the further derivation of the robust registration approach, which assumes a mono-modal registration setup. Sec. 4 details the calculation of entropy images.

Based on the distribution in Eq. (4), we formulate registration as a maximum likelihood estimation (Viola and Wells, 1997; Roche et al., 2000). For the contextual framework, the optimization is performed over the description layers and transformation parameters

$$[\hat{\boldsymbol{\pi}}, \hat{D}_m, \hat{D}_f] = \arg \max_{\boldsymbol{\pi}, D_m, D_f} p(I_m, I_f, D_m, D_f, \boldsymbol{\pi}). \quad (5)$$

In this work, we do not allow for a dynamic adaptation of the description layers during the registration process by specifically choosing to work with entropy images. The coupling terms therefore only motivate the construction of the description layers but do not influence the optimization, yielding the maximization problem

$$\hat{\boldsymbol{\pi}} = \arg \max_{\boldsymbol{\pi}} p(I_m, I_f, D_m, D_f, \boldsymbol{\pi}) \quad (6)$$

$$= \arg \max_{\boldsymbol{\pi}} p(D_m | D_f, \boldsymbol{\pi}). \quad (7)$$

2.3. Heteroskedastic Noise

After introducing the layout of the probabilistic framework, the similarity term $p(D_m | D_f, \boldsymbol{\pi})$ needs to be specified. This requires the introduction of the underlying imaging model.

First, we map both of the description layers into the mid space to obtain an inverse-consistent transformation. Second, we assume heteroskedastic Gaussian noise to account for outliers and perform robust estimation. Third, instead of intensity mapping, we operate on the modality-invariant description layers. This set of assumptions yields the imaging model

$$D_m(T_{\boldsymbol{\pi}/2}(x)) = D_f(T_{-\boldsymbol{\pi}/2}(x)) + \varepsilon(x) \quad (8)$$

with a spatially varying Gaussian noise $\varepsilon(x) \sim \mathcal{N}(0, \sigma^2(x))$ and variance $\sigma^2(x)$. Note that this model differs from alternative probabilistic approaches (Viola and Wells, 1997; Roche et al., 2000) that consider homoskedastic image noise, $\varepsilon(x) = \varepsilon$, $\forall x \in \Omega$. To facilitate notation we define $D_m^\uparrow(x) = D_m(T_{\boldsymbol{\pi}/2}(x))$ for the movable image mapped forward and $D_f^\downarrow(x) = D_f(T_{-\boldsymbol{\pi}/2}(x))$ for the fixed image mapped backward. We continue referring to images and description layers as fixed and moving to differentiate between them, although both of them are moving.

In the derivation of the similarity measure, we relax the commonly made assumption of independent and

identically distributed samples to allow independent and *not* identically distributed samples. We define the log-likelihood function for heteroskedastic Gaussian noise

$$\log p(D_m | D_f, \boldsymbol{\pi}, \varepsilon) \quad (9)$$

$$= \log \prod_{x \in \Omega} p(D_m(x) | D_f(x), \boldsymbol{\pi}, \varepsilon(x)) \quad (10)$$

$$= \sum_{x \in \Omega} \log \left(\frac{1}{\sqrt{2\pi}\sigma(x)} \exp \left(-\frac{[D_m^\uparrow(x) - D_f^\downarrow(x)]^2}{2\sigma(x)^2} \right) \right) \quad (11)$$

$$= \sum_{x \in \Omega} -\log(\sigma(x)) - \frac{[D_m^\uparrow(x) - D_f^\downarrow(x)]^2}{2\sigma^2(x)} + \text{const}, \quad (12)$$

where const does not depend on the variables of interest. The influence of each sample in the summation is dependent on $\sigma(x)$. In fact, $\sigma(x)^{-2}$ can be seen as a local weight that determines the influence of the local error. Note that, for a constant variance across the image $\sigma(x) = \sigma$, $\forall x \in \Omega$, this reduces to the commonly applied sum of squared differences. The optimization problem changes from estimating just the transformation parameters to additionally estimating the variance at each location

$$[\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\sigma}}] = \arg \max_{\boldsymbol{\pi}, \boldsymbol{\sigma}} \log p(D_m | D_f, \boldsymbol{\pi}, \boldsymbol{\sigma}), \quad (13)$$

with the vector of variances $\boldsymbol{\sigma}^2 = [\sigma^2(x_1) \dots \sigma^2(x_{|\Omega|})]$. The joint estimation of transformation parameters and variance can cause a complex optimization problem. We employ an iterative optimization procedure that alternates between optimizing the transformation parameters and variances, as summarized in Algorithm 1.

Algorithm 1 Alternating Optimization

```

 $\hat{\boldsymbol{\pi}}^{(0)} = \mathbf{0};$ 
for k = 1 to maxIteration do
  (I)  $\hat{\boldsymbol{\sigma}}^{(k)} = \arg \max_{\boldsymbol{\sigma}} \log p(D_m | D_f, \hat{\boldsymbol{\pi}}^{(k-1)}, \boldsymbol{\sigma})$ 
  (II)  $\hat{\boldsymbol{\pi}}^{(k)} = \arg \max_{\boldsymbol{\pi}} \log p(D_m | D_f, \boldsymbol{\pi}, \hat{\boldsymbol{\sigma}}^{(k)})$ 
end for

```

Initially, the transformation parameters are set to zero, corresponding to the identity transformation¹. In step (I), the optimal variances $\hat{\boldsymbol{\sigma}}^{(k)}$ are calculated for the transformation of the last iteration $\hat{\boldsymbol{\pi}}^{(k-1)}$. In step (II), the optimal transformation $\hat{\boldsymbol{\pi}}^{(k)}$ is calculated for fixed variances $\hat{\boldsymbol{\sigma}}^{(k)}$. An important consequence of the decoupled optimization is that the log-likelihood term in Eq. (12) only depends on the second term when maximizing with respect to the transformation parameters in step (II) of Algorithm 1.

In order to solve step (I), we draw the relationship to iteratively reweighted least squares estimation for which

¹An alternative would be to initialize the transformation based on raw image moments, as described in (Reuter et al., 2010).

a large number of robust estimators have been proposed. This allows us to estimate local variances even though we only have a two intensity values at each location x .

2.4. Relationship to Iteratively Reweighted Least Squares

In iteratively reweighted least squares estimation (Holland and Welsch, 1977; Street et al., 1988), the least squares estimator is replaced by the more general class of M-estimators. This yields a cost function

$$E(\boldsymbol{\pi}) = \sum_x \rho(r(x)), \quad (14)$$

where ρ is a symmetric, positive-definite function with a unique minimum at zero and $r(x) = D_m^\dagger(x) - D_f^\dagger(x)$. In the following, we concentrate on ρ being differentiable, in which case the M-estimator is said to be of ψ -type (as $\psi = \rho'$). Calculating the derivative yields

$$\frac{\partial E(\boldsymbol{\pi})}{\partial \pi_i} = \sum_x \rho'(r(x)) \frac{\partial r(x)}{\partial \pi_i}. \quad (15)$$

Defining a weight function $w(z) := \frac{\rho'(z)}{z}$, we obtain

$$\frac{\partial E(\boldsymbol{\pi})}{\partial \pi_i} = \sum_x w(r(x)) r(x) \frac{\partial r(x)}{\partial \pi_i}. \quad (16)$$

To relate this to our optimization problem, we calculate the gradient of the negative log-likelihood function of Eq. (12)

$$-\frac{\partial \log p(D_m | D_f, \boldsymbol{\pi}, \varepsilon)}{\partial \pi_i} = \sum_x \frac{1}{\sigma^2(x)} r(x) \frac{\partial r(x)}{\partial \pi_i}, \quad (17)$$

under the assumption that $\sigma^2(x)$ does not depend on the transformation parameter $\boldsymbol{\pi}$, as it is the case for the alternating optimization. Comparing Eq. (16) and Eq. (17), we see that they can be brought into correspondence by identifying $w(x) = \frac{1}{\sigma^2(x)}$. This shows that the application of gradient-based optimization for the proposed log-likelihood function is equivalent to the gradient-based optimization of the M-estimator function ρ . Since the weights are inversely related to the variances, we have $\sigma^2(x) = \frac{1}{w(x)} = \frac{r(x)}{\rho'(r(x))}$ for different M-estimators ρ .

2.5. Robust Estimators

Having this important connection established, we can now leverage M-estimators ρ , for the robust description of the local variances. Here we specifically model the variances via Tukey's biweight function, as it completely suppresses the effect of outliers:

$$\sigma^2(x) = \begin{cases} \left(1 - \frac{r(x)^2}{c^2}\right)^{-2}, & \text{if } |r(x)| \leq c, \\ \infty, & \text{otherwise,} \end{cases} \quad (18)$$

Fig. 2 illustrates Tukey's biweight function as well as the quadratic function with the corresponding weights. The

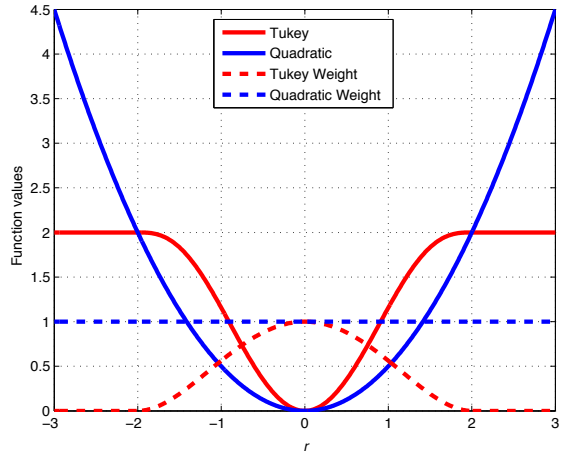


Figure 2: Illustration of Tukey's biweight function for $c = 2$ in red and the quadratic function $\frac{r^2}{2}$ in blue. We also show the corresponding weight function $\frac{\rho'(r)}{r}$ as dashed line. The weights are constant for the quadratic function. For Tukey's biweight the weights are zero for $|r| > c$.

weight function $w(r) = \rho'(r)/r$ for the quadratic function $\rho(r) = r^2/2$ is the constant function $w(r) = 1$ and for Tukey's biweight function is $w(r) = \frac{1}{\sigma^2(r)} = \left(1 - \frac{r^2}{c^2}\right)^2$ for $|r| \leq c$ and zero otherwise. Tukey's biweight (red curve) limits the influence of large residuals ($r > c$), for which the weights (red dashed curve) are zero (and the variance infinity). The regular least squares approach (blue curves) results in a constant weight function and variance, independent of the local residuals, and thus considers the contribution of all voxels equally. Using Tukey's estimator requires the saturation parameter c to be set. A fixed value of $c = 4.685$ is recommended for unit Gaussian noise by Holland and Welsch (1977); Nestares and Heeger (2000). However, it may be more appropriate to specify the value depending on the characteristics of the input images. A method to automatically estimate the saturation parameter for each image pair in a pre-processing step on low resolution images has been recommended in Reuter et al. (2010). The procedure allows for more outliers in the skull, jaw and neck regions, where more variation is to be expected. We employ this method to estimate the saturation parameter for entropy images automatically.

3. Optimization

In this section, we present more details on the optimization with respect to the transformation parameters (step (II) in Algorithm 1) and derive the Gauss-Newton method. Moreover, we show that Gauss-Newton corresponds to the efficient second-order minimization for the special case of a fully symmetric setup. For the alternating optimization, only the second term of the log-likelihood function in Eq. (12) is a function of the transformation. We rewrite

the resulting cost function in vector notation

$$C(\boldsymbol{\pi}) = \frac{1}{2} \mathbf{r}(\boldsymbol{\pi})^\top W^\top W \mathbf{r}(\boldsymbol{\pi}) = \frac{1}{2} \|W \mathbf{r}(\boldsymbol{\pi})\|^2, \quad (19)$$

with the diagonal weight matrix $W_{xx} = \frac{1}{\sigma(x)}$ being of size $\mathbb{R}^{|\Omega| \times |\Omega|}$. As before, the weights do not depend on the transformation parameters $\boldsymbol{\pi}$.

Steepest descent: For steepest descent, we obtain the update rule

$$\mathbf{h} = \alpha \cdot \mathbf{J}_C \quad \boldsymbol{\pi} \leftarrow \boldsymbol{\pi} + \mathbf{h} \quad (20)$$

where α is the learning rate. Note that the additive update in Lie algebra is equivalent to a compositional update in the Lie group $T_{\boldsymbol{\pi}} \circ \exp(\sum h_i \cdot b_i)$. The Jacobian $\mathbf{J}_C(\boldsymbol{\pi})$ contains the partial derivatives of all n parameters, i.e., $\mathbf{J}_C(\boldsymbol{\pi}) = \left[\frac{\partial C(\boldsymbol{\pi})}{\partial \pi_1}, \dots, \frac{\partial C(\boldsymbol{\pi})}{\partial \pi_n} \right]$ with

$$\frac{\partial C(\boldsymbol{\pi})}{\partial \pi_i} = W^2 \mathbf{r}(\boldsymbol{\pi}) \frac{\partial \mathbf{r}(\boldsymbol{\pi})}{\partial \pi_i}. \quad (21)$$

Gauss-Newton and Efficient Second-Order Minimization: For least squares optimization problems, as in Eq. (19), Gauss-Newton methods can be applied. From the Taylor series, we obtain the following approximation of the residual for small updates $\|\mathbf{h}\|$

$$\mathbf{r}(\boldsymbol{\pi} + \mathbf{h}) \approx \mathbf{r}(\boldsymbol{\pi}) + \mathbf{J}_r(\boldsymbol{\pi}) \cdot \mathbf{h} \quad (22)$$

with the Jacobian $\mathbf{J}_r(\boldsymbol{\pi}) = \left[\frac{\partial \mathbf{r}(\boldsymbol{\pi})}{\partial \pi_1}, \dots, \frac{\partial \mathbf{r}(\boldsymbol{\pi})}{\partial \pi_n} \right]$. The linear approximation yields

$$\begin{aligned} C(\boldsymbol{\pi} + \mathbf{h}) &= \frac{1}{2} \|W \mathbf{r}(\boldsymbol{\pi} + \mathbf{h})\|^2 = \frac{1}{2} \mathbf{r}(\boldsymbol{\pi} + \mathbf{h})^\top W^\top W \mathbf{r}(\boldsymbol{\pi} + \mathbf{h}) \\ &\approx \frac{1}{2} \mathbf{r}(\boldsymbol{\pi})^\top W^2 \mathbf{r}(\boldsymbol{\pi}) + \mathbf{h}^\top \mathbf{J}_r(\boldsymbol{\pi})^\top W^2 \mathbf{r}(\boldsymbol{\pi}) \\ &\quad + \frac{1}{2} \mathbf{h}^\top \mathbf{J}_r(\boldsymbol{\pi})^\top W^2 \mathbf{J}_r(\boldsymbol{\pi}) \mathbf{h} \\ &= C(\boldsymbol{\pi}) + \mathbf{h}^\top \mathbf{J}_r(\boldsymbol{\pi})^\top W^2 \mathbf{r}(\boldsymbol{\pi}) \\ &\quad + \frac{1}{2} \mathbf{h}^\top \mathbf{J}_r(\boldsymbol{\pi})^\top W^2 \mathbf{J}_r(\boldsymbol{\pi}) \mathbf{h} \end{aligned} \quad (23)$$

We obtain the optimal update by differentiating with respect to \mathbf{h} and setting the gradient to zero, leading to the linear system of equations

$$(\mathbf{J}_r(\boldsymbol{\pi})^\top W^2 \mathbf{J}_r(\boldsymbol{\pi})) \mathbf{h} = -\mathbf{J}_r(\boldsymbol{\pi})^\top W^2 \mathbf{r}(\boldsymbol{\pi}). \quad (24)$$

The update is as before $\boldsymbol{\pi} \leftarrow \boldsymbol{\pi} + \mathbf{h}$. We take a closer look at the Jacobian to see an interesting consequence of the fully symmetric setup

$$\begin{aligned} \mathbf{J}_r(\boldsymbol{\pi}) &= \frac{\partial (D_m^\dagger - D_f^\dagger)}{\partial \boldsymbol{\pi}} = \frac{\partial (D_m(T_{\boldsymbol{\pi}/2}) - D_f(T_{-\boldsymbol{\pi}/2}))}{\partial \boldsymbol{\pi}} \\ &= \frac{1}{2} ((\nabla D_m)^\dagger + (\nabla D_f)^\dagger) \frac{\partial T_{\boldsymbol{\pi}}}{\partial \boldsymbol{\pi}}. \end{aligned} \quad (25)$$

Since the first image is mapped exactly in the inverse direction of the second, the derivatives of the transformations only differ by a minus sign. Thus the transformation derivative $\frac{\partial T_{\boldsymbol{\pi}}}{\partial \boldsymbol{\pi}}$ can be factored out. The resulting formula for the Jacobian is equivalent to the Jacobian of the efficient second-order minimization (ESM) (Benhimane and Malis, 2004; Vercauteren et al., 2007; Wachinger and Navab, 2009, 2013). For the fully symmetric registration setup, Gauss-Newton and ESM are therefore equivalent. Independently, the same relationship was discovered by Lorenzi et al. (2013) in an analysis of the log-Demons algorithm.

In a general setup, ESM was shown to require significantly fewer steps in the optimization than Gauss-Newton (Benhimane and Malis, 2004; Wachinger and Navab, 2013) and to have a cubic convergence rate (Malis, 2008). The update procedure of ESM is closely related to Gauss-Newton: it uses the addition of the image gradients of both images, instead of just one of the images. We use the Gauss-Newton optimization in our experiments and are therefore benefiting from the attractive convergence properties of ESM. To simplify implementation we explicitly state the derivative at a point $p = (x, y, z, 1)$ for the rigid case (Wachinger and Navab, 2013)

$$\left[\frac{\partial T_{\boldsymbol{\pi}}}{\partial \boldsymbol{\pi}} \right]_p = \begin{pmatrix} 0 & z & -y & 1 & 0 & 0 \\ -z & 0 & x & 0 & 1 & 0 \\ y & -x & 0 & 0 & 0 & 1 \end{pmatrix} \quad (26)$$

and for the affine case (Ashburner, 2000; Reuter et al., 2010)

$$\left[\frac{\partial T_{\boldsymbol{\pi}}}{\partial \boldsymbol{\pi}} \right]_p = \begin{pmatrix} x & y & z & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & x & y & z & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x & y & z & 1 \end{pmatrix}. \quad (27)$$

3.1. Additional Algorithmic Details

We employ a multiresolution approach to estimate large displacements between the images (Roche et al., 1999). The Gaussian pyramid is constructed by halving each dimension on each level until the image size is approximately 16^3 . This procedure results in 5 resolution levels for typical image sizes (256^3). The iterations at each level are terminated if the transformation update is below a specified threshold (0.01mm) or a maximum number of iterations (5) is reached. The registration of the subsequent resolution level is initialized with the result of the previous one.

In our experiences with entropy images, different smoothness or signal-to-noise ratio can lead to entropy images that vary in overall scale. To account for this variation, we add a global intensity scaling of entropy images. Since we are interested in a fully symmetric registration setup, we apply the intensity scaling to both images. The equation for the imaging model (8), changes to

$$e^{\frac{\varepsilon}{2}} \cdot D_m(T_{\boldsymbol{\pi}/2}(x)) = e^{-\frac{\varepsilon}{2}} \cdot D_f(T_{-\boldsymbol{\pi}/2}(x)) + \varepsilon(x), \quad (28)$$

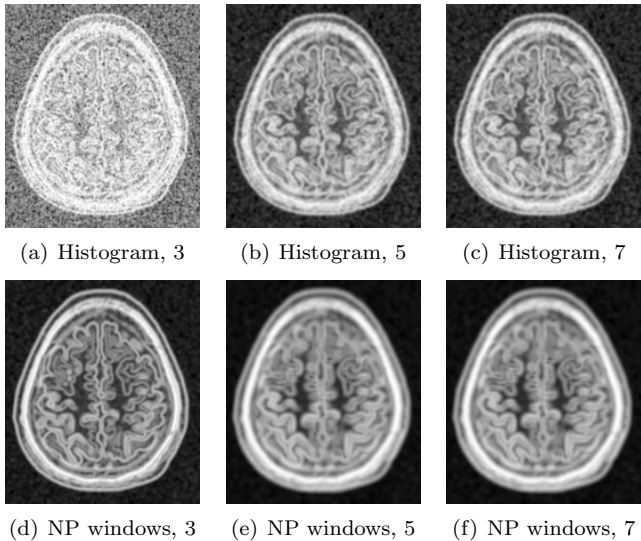


Figure 3: Comparison of entropy images for patch sizes $3 \times 3 \times 3$ to $7 \times 7 \times 7$ created with density estimators based on on histogram (a-c) and non-parametric (NP) windows (d-f). Appearance of entropy images is smoother for larger patch sizes. Note the striking difference between the two estimation approaches for patches of size 3 (a vs. d).

where s is the scaling parameter². We choose an exponential intensity scaling, in contrast to square root multiplication of s in (Reuter et al., 2010), because it is symmetric (with respect to the parameter update) and improves the stability of the model.

4. Entropy Images with non-parametric Windows

We employ entropy images with the non-parametric windows estimator to obtain a structural representation of the images. Considering an image I and a local neighborhood $\mathcal{N}(x)$, the corresponding description layer is calculated as

$$D(x) = H(I|_{\mathcal{N}(x)}), \quad (29)$$

with the Shannon entropy $H(Y) = -\sum_{y \in Y} p(y) \log p(y)$ and $I|_{\mathcal{N}(x)}$ the restriction of the image to the neighborhood $\mathcal{N}(x)$. The calculation of the local entropy for all locations in the image leads to entropy images. Compared to mutual information, entropy images have advantages for the registration of images affected by bias fields (Wachinger and Navab, 2012b). An essential step for the entropy calculation is the estimation of the intensity distribution $p_{I|_{\mathcal{N}(x)}}$ within patches $I|_{\mathcal{N}(x)}$. On the one hand, the patch size has to be large enough to provide enough samples for a

²We consider separate noise for fixed and moving image, which are subject to the intensity scaling. Since linear combination of mutually independent normal random variables is again normal distributed, we summarize it into a global noise term.

reliable density estimation. On the other hand, a large patch size leads to smoothing in the images. Since we are interested in a highly accurate multi-modal registration, we want to reduce the patch size as much as possible to allow for accurate localization and sharp structures.

One approach is to upsample the original image and to utilize the interpolated values for the density estimation. A recently proposed method of non-parametric (NP) windows (Dowson et al., 2008) considers the asymptotic case by letting the number of interpolated samples go to infinity. The method involves constructing a continuous space representation of the discrete space signal using a suitable interpolation method. NP windows requires only a small number of observed signal samples to estimate the density function and is completely data driven. Originally, Dowson et al. (2008) only presented closed form solutions for certain 1D and 2D density estimation cases. Recently, Joshi et al. (2011) introduced a simplified version of NP windows estimation. Importantly they also include closed form solutions for 3D images. The density is estimated per cube in the volume, where each cube is divided in five tetrahedra. Finding the density estimate over the cube then reduces to finding densities for each of the tetrahedra. We employ the NP windows density estimation for entropy images, leading to highly localized entropy images. Fig. 3 shows entropy images for density estimation based on histograms and NP windows for different patch sizes. The entropy estimation with NP windows leads to a clear improvement, especially for small patch sizes.

5. Experiments

We evaluate the robust, multi-modal registration in the following experiments. For quantifying the registration results, we calculate the root mean squared (RMS) error between two transformations Jenkinson (1999). This error describes the average voxel displacements in millimeter inside a ball with radius r located at the center of the image. Given two affine transformations in 3D, (A_1, t_1) and (A_2, t_2) , with the 3×3 transformation matrices A_i and the 3×1 translation vectors t_i , the RMS error is

$$E_{RMS} = \sqrt{\frac{1}{5}r^2 \text{tr}[(A_2 - A_1)^\top (A_2 - A_1)] - (t_2 - t_1)^\top (t_2 - t_1)}. \quad (30)$$

We calculate the RMS error on transformations in RAS (right, anterior, superior) coordinates, where the origin lies approximately in the center of the image. We set $r = 100\text{mm}$, to define a sphere that is large enough to include the full brain. For example, an E_{RMS} error of 0.1 mm corresponds to 1/10 of a voxel displacement (for 1mm isotropic images), which can easily be picked up in visual inspection.

5.1. Brain Registration with Ground Truth

In the first experiment, we work with simulated T1 and T2 images from the BrainWeb dataset (Cocosco et al.,

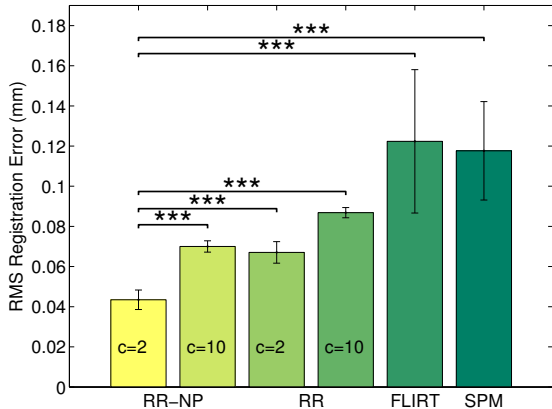


Figure 4: Statistical analysis of RMS errors for the skull-strip T1-T2 registration study over 100 repetitions. Bars indicate mean error, error bars correspond to two standard deviations. *** indicates significance level at 0.001. Robust registration with non-parametric windows (RR-NP) yields a significant reduction in registration error with respect to robust registration with histograms (RR), FLIRT, and SPM. Further, setting the parameter in Tukeys biweight function to $c = 2$, which allows for more outliers than $c = 10$, yields to significantly better results.

1997) which are perfectly aligned. Always knowing the ground truth transformations allows us to establish results on the accuracy of different registration procedures. The image resolution is $256 \times 256 \times 181$ with 1mm isotropic spacing. We automatically skull-strip the T1 image with the brain extraction tool (BET) (Smith, 2002), modify the brain mask (asymmetric and localized dilation), and apply it to the T2 image to induce variations in brain extraction between both modalities. We apply random transformations to the skull-stripped images with translations of 30mm and rotations around an arbitrary axis of 25° . Different registration approaches are compared to determine how well they recover the correct registration: the statistical analysis of the RMS registration error for 100 repetitions is shown in Fig. 4.

For the robust registration with NP windows (RR-NP) the patch size is $3 \times 3 \times 3$ and for the robust registration with histograms (RR) it is $5 \times 5 \times 5$. We manually set the parameter in Tukey’s biweight function to simulate a robust approach with many outliers ($c = 2$) and a less robust approach with few outliers ($c = 10$). We compare to popular and freely available registration software using mutual information: FLIRT in FSL and the coregistration tool in SPM. Our results show the significant reduction in registration error of our approach in contrast to these reference methods. Moreover, we observe a significant improvement for the density estimation with NP windows in comparison to histograms, highlighting the importance of the better localization with NP windows and confirming the qualitative results from Fig. 3. Finally, the significant

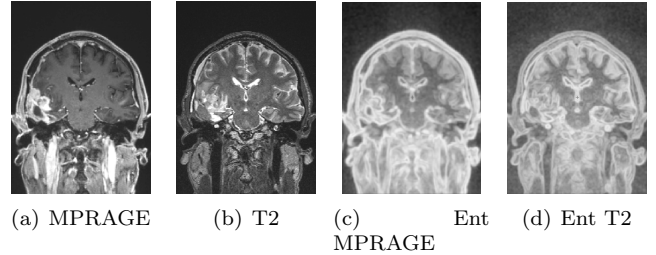


Figure 5: Coronal view of full head MR MPRAGE (a) and T2 (b) tumor images with corresponding entropy images (c-d).

improvement for $c = 2$ over $c = 10$ demonstrates the necessity for a robust approach to limit the influence of outliers on the registration. The creation of the entropy images took about 9s with histograms and about 320s with NP windows. The remaining registration took about 39s.

5.2. Tumor Registration

In the second experiment, we evaluate registration accuracy based on a pair of real brain tumor MR T1 (magnetization-prepared rapid gradient-echo, MPRAGE, Mugler and Brookeman (1990)) and MR T2 images, both acquired within the same session. The image resolution is $0.94 \times 0.94 \times 0.9\text{mm}^3$ for T1 and isotropic 1mm for T2. Fig. 5 shows the pair of images that we use for the experiments. Both the automatic skull stripping tool in FreeSurfer and the brain extraction tool (BET) (Smith, 2002) frequently fail for tumor images because of the different appearance of enhancing tumor compared to regular brain tissue, violating intensity priors used by the methods. In order to produce a useful result, we manually refined the brain extraction obtained from FreeSurfer on the T1 image and propagate the brain mask to the T2 image by registering the full head T1 and T2 images with FLIRT.

When performing robust multi-modal registration on the skull-stripped brain images we expect the tumor regions to be marked as outliers because of the different appearance in both modalities. Fig. 6 depicts the variance $\hat{\sigma}^2$ calculated with Tukey’s biweight function (Sec. 2.5) at each spatial location after successful registration. Larger variances are shown in red and yellow. We observe that tumor regions are marked as outliers (yellow), as expected. Also note the increased variance at the interface between the brain and the skull, which is likely caused by the inconsistent brain extraction (mask propagation) and differential distortion between the two modalities. These non-matching structures in tumor regions and artifacts from the brain extraction, that can cause problems for non-robust techniques, are correctly identified automatically by our method and their influence is limited during the registration.

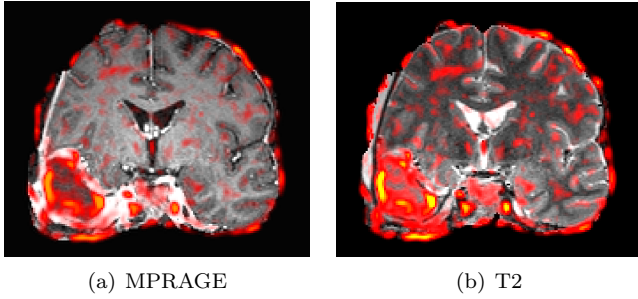


Figure 6: Coronal view of MR MPRAGE (a) and T2 (b) skull-stripped brain tumor images after registration. Estimated weights are overlaid as heat map.

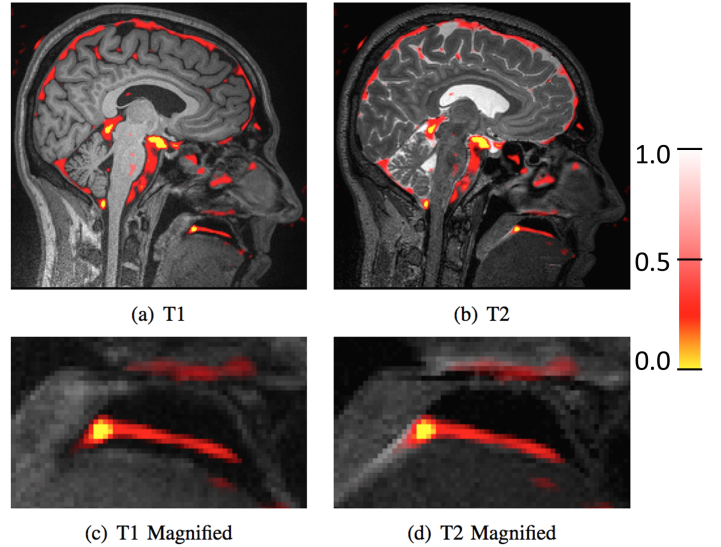


Figure 8: Sagittal view of MR T1 (a) and T2 (b) full head images with an overlay of the weights as a result of the robust registration. Magnifications of the tongue area (c-d) that is susceptible to motion. Areas with motion differences are assigned low weights to limit their influence on the registration. Also regions in the periphery of the brain are assigned low weights because of the different appearance of CSF and dura mater and resulting dissimilarities in the entropy images.

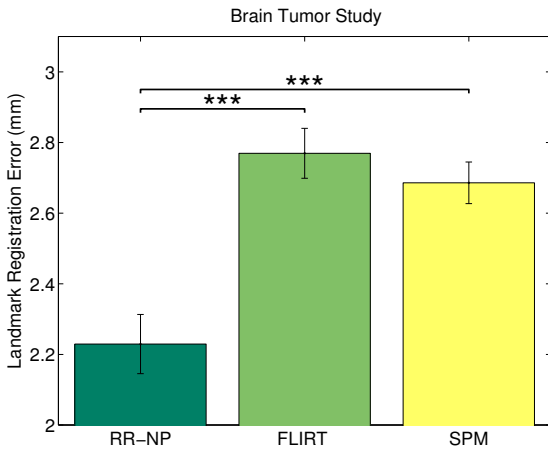


Figure 7: Statistical analysis of RMS errors for tumor registration study over 100 repetitions. Bars indicate mean error, error bars correspond to two standard deviations. *** indicates significance level at 0.001. Robust registration with NP windows (RR-NP) yields a significant reduction in registration error. The range of the y-axis is adjusted to highlight the differences of the methods.

In the case of clinical tumor images, we do not know the ground truth alignment. To quantify the registration error, landmarks were manually selected in both modalities, permitting the computation of landmark distances after the registration. Fig. 7 shows the statistical analysis for 100 repetitions with the random displacement of the images similar to the previous experiment. We compare RR-NP with patch size 3 to FLIRT and SPM. The results indicate a significantly lower registration error for the proposed robust registration approach. The manual refinement of the brain extraction and the manual selection of landmarks make the extension of this study to many subjects complicated, but results on a large cohort are presented in the next section.

5.3. Full Head Registration

In this experiment we register T1 (Multi-Echo MPRAGE van der Kouwe et al. (2008)) and T2 full head images of 106 healthy subjects. These images have a 1mm isotropic resolution ($256 \times 256 \times 176$). Although each image pair was acquired within the same session, we expect local differences caused by jaw, tongue, and small head movements. Such local displacements can deteriorate the registration quality of full head scans. Fig. 8 shows a registered pair of T1 and T2 images with an overlay of the estimated variances $\hat{\sigma}^2$, together with a magnified view of the area around the tongue. Due to the movement of the tongue between the scans, the robust registration algorithm detects outliers in that region. Note that also regions in the periphery of the brain show low weights, which are caused by the different appearance of dura mater and CSF in T1- and T2-weighted images. The local information content is different in these regions, yielding differences between the entropy images. The benefit of the robust approach is to identify these contradictory regions as outliers and reduce their influence on the registration.

Again no ground truth alignment is available, but the number of scans in this study is too large for the manual identification of landmarks. Instead, we assume that the registration of skull-stripped images is more accurate than the registration of the full head images because most structures that are susceptible to patient motion are removed.

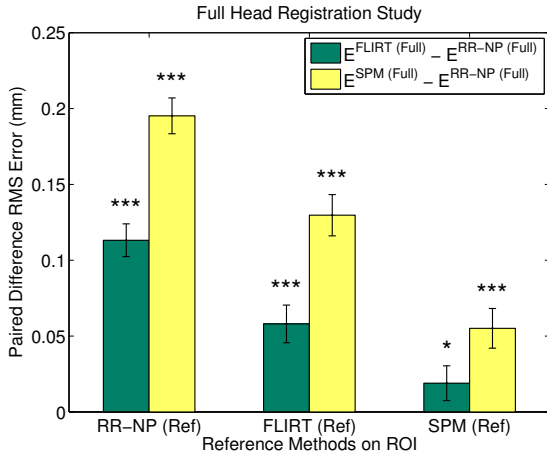


Figure 9: Analysis of difference of RMS errors for the large control study on 106 subjects with 10 repeats. The RMS error is computed between full head registrations and the reference transformation, yielding $E^{Type} = E_{RMS}(T^{Ref}, T_{Type}^{Full})$, where we use RR-NP, FLIRT and SPM as registration *Type* on the full head scans. Differences of RMS errors $E^{FLIRT} - E^{RR-NP}$ and $E^{SPM} - E^{RR-NP}$ are plotted to compare the methods, where positive values show better performance of RR-NP. Bars indicate mean difference error and correspond to standard error of difference. *, **, and *** indicate significance levels from paired t-test at 0.05, 0.01, and 0.001, respectively. To avoid biasing the results, we use each method (RR-NP, FLIRT, and SPM) to establish the reference transformation on the skull-stripped images (x-axis).

Since both scans were collected within the same session, we do not expect any changes in the brain. Local differences are mainly expected in the scalp, jaw, neck and eye regions. It is therefore reasonable to assume a higher registration accuracy when using the brain as a stable region of interest (ROI). Given a reference transformation T^{Ref} , we compute the RMS error $E^{Type} = E_{RMS}(T^{Ref}, T_{Type}^{Full})$, where T_{Type}^{Full} is the transformation computed on the full head scans with registration method *Type* that we want to evaluate. Fig. 9 shows the difference in RMS error between the usage of FLIRT and SPM compared to RR-NP, so $E^{FLIRT} - E^{RR-NP}$ and $E^{SPM} - E^{RR-NP}$, where positive differences indicate that a better approximation of the reference transformation with RR-NP. To avoid biasing results by selecting only one registration to establish the reference transformation, we report results for using all three registration methods to compute T^{Ref} , shown on the x-axis of Fig. 9. For each subject, we randomly transform the full head images (up to 15° and 20mm) ten times and perform the registration with each method, yielding 3180 full head registrations. The results show that RR-NP is significantly better in recovering the reference transformation established on the brain ROI independent of which reference method was used.

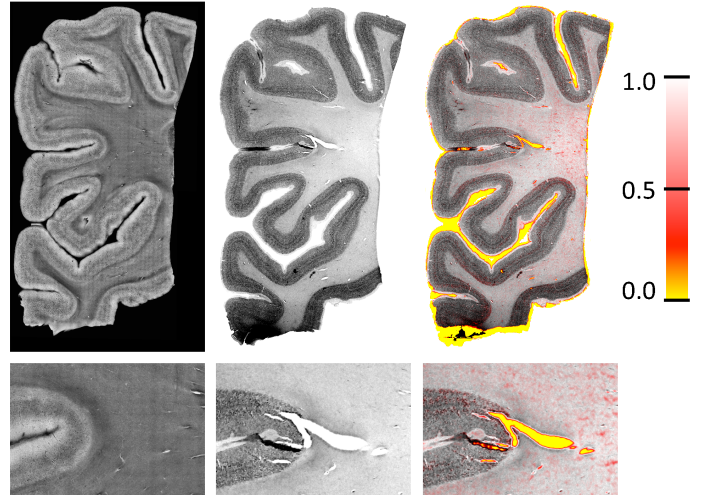


Figure 10: OCT (first column) and histology (second column) slices together with an overlay of weights on the histology slices (third column). The top row shows the entire slices, while the bottom row shows a magnifications around the large crack in the center of the images. Tears in the histology image and alignment errors at the boundary (caused by non-linear deformations) are marked as outliers.

5.4. Histology-OCT Registration

In this multi-modal experiment we register images from optical coherence tomography (OCT) with histology slices. OCT is an optical microscopy technique, analogue to ultrasound imaging, which provides a resolution close to 1 micron. This technique requires neither staining nor dyes, as it relies on the intrinsic optical properties of the tissue. The block sample can be imaged prior to any cutting, which greatly reduces the distortions contrary to the histological protocol. Details on the technique and the acquisition of those images can be found in Magnain et al. (2014). Histology-OCT alignment is of clinical importance to validate the appearance of structures in OCT based on the true appearance on the sliced tissue. The registration is challenging as histology images show artifacts due to local deformations and tearing of the tissues during slicing, which are not present in the previously acquired OCT images. Fig. 10 shows the result of the registration between an OCT and histology slice, as well as the estimated variance $\hat{\sigma}^2$. The magnified view illustrates the differences between both images and shows that cracks in the histology slices are correctly marked as outliers. Fig. 11 shows another set of histology and OCT images with the corresponding entropy images. The gyrus was cropped from larger OCT and histology images independently, which introduces small differences at the top boundary edge due to slightly different cropping angles. The variance map shows that these differences together with internal cracks are marked as outliers by the robust registration.

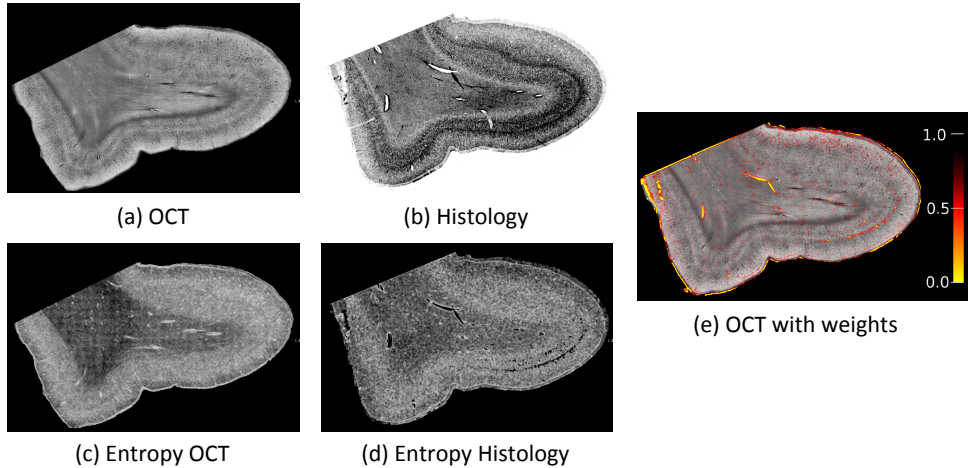


Figure 11: OCT (a) and histology (b) slices with corresponding entropy images (c-d). Estimated weights are shown on the OCT slice (e). Outliers are found in cracks of the histology slices and around the boundary.

To confirm the promising results from the qualitative analysis, an expert identified 26 landmark correspondences in four histology and corresponding OCT images. We use the transformation from image registration to map the landmarks from the OCT to the histology domain. The average Euclidean distance between mapped OCT and histology landmark pairs measures the registration error. Fig. 12 shows the mean and standard error for the robust registration (RR), FSL, and an alternative 2D affine registration method using mutual information (MI) and a Powell minimizer (based on the SPM algorithm). The original SPM registration method is implemented only for the full 3D case and does not model 2D affine transformations. Because of the high resolution of the images in our experiment (1400×2500), we select a larger patch size and work with histogram-based density estimation in the robust registration method. The results from the quantitative analysis confirm the improved registration accuracy of the robust multi-modal registration.

5.5. Bias Field Robustness

Intensity bias fields in images can severely affect cost functions in cross-modal registration, such as mutual information (Myronenko and Song, 2010). On the contrary, the presented approach is very robust with respect to bias fields, as it is based on local entropy estimation. To demonstrate robustness with respect to changes in the bias field, we employ 5 different cases from the BrainWeb dataset (Cocosco et al., 1997) and increase the size of this test set by simulating brain tumors Prastawa et al. (2009) (including contrast enhancement, local distortion of healthy tissue, infiltrating edema adjacent to tumors, destruction and deformation of fiber tracts). For each of the 5 original cases tumor seed were placed at two different locations and two different parameter settings were selected to vary the characteristics of the tumor (conservative vs. aggressive growth), yielding 20 tumor cases. Each simulation

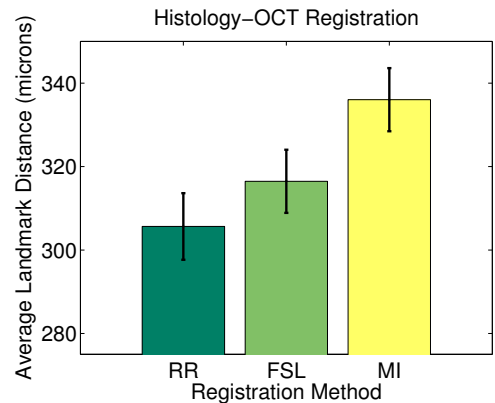


Figure 12: Registration error for histology-OCT alignment measured as landmark distance in microns. Results are shown for different registration methods. Bars indicate mean error; error bars correspond to standard error. The range of the y-axis is selected to highlight the differences of the methods.

produced aligned T1, T2, and T1 Gadolinium enhanced (T1Gad) tumor images (1mm isotropic, $256 \times 256 \times 181$) where we added noise and bias fields.

We apply random rigid transformations with translations of 30mm and rotations of 25° around an arbitrary axis with the center of rotation corresponding to the image center, and measure how accurately each test method can estimate the true transformation. We perform the robust registration for patch sizes between $3 \times 3 \times 3$ and $7 \times 7 \times 7$ and again compare to FLIRT and SPM. Fig. 13 shows the RMS registration errors for the different cross-modal registration pairs. It can be seen that the robust registration approach is not affected by the bias field. SPM performs better than FLIRT, but in both methods regis-

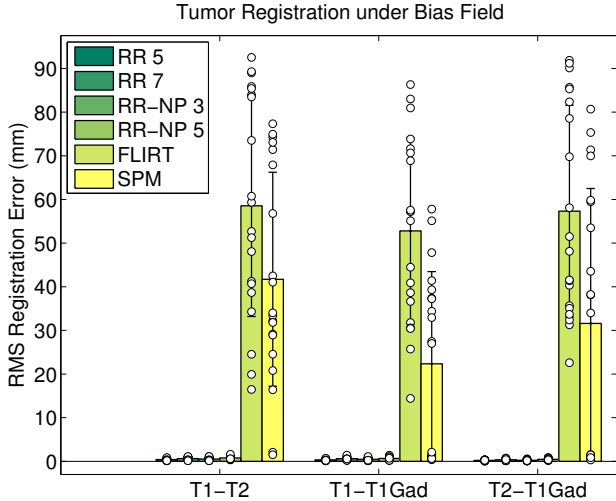


Figure 13: RMS errors of the different registration methods for 20 synthetic multi-modal brain tumor images (T1, T2, T1Gad) with bias field. Bars indicate mean error, error bars correspond to two standard deviations. White discs correspond to the individual data points. The proposed methods (RR, RR-NP) are robust to bias field due to the local entropy estimation, while the bias field has a strong influence on FLIRT and SPM.

tration fails completely in a large number of cases. Without adding a bias field none of the methods produce any of these severe registration failures (not shown). Bias field correction can therefore be expected to remove most of the intensity bias and resulting registration problems, but this additional processing step is not required for our entropy based cross-modal registration approach.

6. Conclusion

We presented a novel registration approach for inverse-consistent, multi-modal registration, that is robust with respect to local differences and with respect to bias fields. To achieve outlier robustness, we incorporated a heteroskedastic noise model and established the relationship to iteratively reweighted least squares estimation. We derived the Gauss-Newton optimization, which we showed to be equivalent to the efficient second-order minimization in case of our robust and inverse-consistent registration setup. To allow for better localization of structures when constructing entropy images we employed a non-parametric density estimator and demonstrated its advantages. We evaluated our method on different multi-modal datasets and demonstrated increased accuracy and robustness of the proposed method. This work focuses on global registration and it remains to investigate the performance and feasibility of the proposed robust multi-modal approach for non-linear registration. One concern is that locations with large differences may be marked as outliers

and therefore produce no force on the deformation field, although they could potentially be correctly aligned in subsequent steps. On the other hand, deformation fields are never estimated only locally. Regularizers and parametric models combine the forces from several locations. This combination may still push the deformation field in the correct direction, in spite of a reduced weight at certain locations. Adjustments to the robust estimation approach may be required, for instance, the regularization of weights via local smoothing.

The presented registration method will be made freely available within the FreeSurfer software package.

Acknowledgements

We are very thankful to Niranjan Joshi and Timor Kadir for providing Matlab code of the NP windows estimator. We are very grateful to the unknown reviewers for their insightful comments.

This work was partly supported by the Humboldt foundation. Further support for this research was provided in part by the National Cancer Institute (K25-CA181632-01A1), the National Center for Research Resources (P41-RR13218, P41-RR14075, U24 RR021382), the National Institute for Biomedical Imaging and Bioengineering (R01EB006758, NAMED U54-EB005149), the National Institute on Aging (AG022381, 5R01AG008122-22), the National Center for Alternative Medicine (RC1 AT005728-01), the National Institute for Neurological Disorders and Stroke (R01 NS052585-01, 1R21NS072652-01, 1R01NS070963), and was made possible by the resources provided by Shared Instrumentation Grants 1S10RR023401, 1S10RR019307, and 1S10RR023043. Additional support was provided by The Autism & Dyslexia Project funded by the Ellison Medical Foundation, and by the NIH Blueprint for Neuroscience Research (5U01-MH093765), part of the multi-institutional Human Connectome Project. In addition, BF has a financial interest in CorticoMetrics, a company whose medical pursuits focus on brain imaging and measurement technologies. BF's interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies.

References

- Andronache, A., von Siebenthal, M., Székely, G., Cattin, P., 2008. Non-rigid registration of multi-modal images using both mutual information and cross-correlation. *Medical Image Analysis* 12, 3–15.
- Arsigny, V., Commowick, O., Ayache, N., Pennec, X., 2009. A fast and log-euclidean polyaffine framework for locally linear registration. *Journal of Mathematical Imaging and Vision* 33, 222–238.
- Ashburner, J., 2000. *Computational Neuroanatomy*. Ph.D. thesis. University College London.
- Ashburner, J., Friston, K., et al., 1999. Nonlinear spatial normalization using basis functions. *Human brain mapping* 7, 254–266.

- Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis* 12, 26–41.
- Benhimane, S., Malis, E., 2004. Real-time image-based tracking of planes using efficient second-order minimization, in: *IEEE/RSJ*, pp. 943–948.
- Christensen, G., Johnson, H., 2001. Consistent image registration. *Medical Imaging, IEEE Transactions on* 20, 568–582.
- Cocosco, C.A., Kollokian, V., Kwan, R.S., Evans, A.C., 1997. Brainweb: Online interface to a 3d mri simulated brain database. *NeuroImage* 5.
- Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G., 1995. Automated multi-modality image registration based on information theory, in: *Information processing in medical imaging*, pp. 264–274.
- Dowson, N., Kadir, T., Bowden, R., 2008. Estimating the joint statistics of images using nonparametric windows with application to registration using mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 1841–1857.
- Fox, N.C., Ridgway, G.R., Schott, J.M., 2011. Algorithms, atrophy and alzheimer’s disease: cautionary tales for clinical trials. *Neuroimage* 57, 15–18.
- Greve, D., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* 48, 63.
- Haber, E., Modersitzki, J., 2007. Intensity gradient based registration and fusion of multi-modal images. *Methods of information in medicine* 46, 292–299.
- Heinrich, M., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F., Brady, S., Schnabel, J., 2012. Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical Image Analysis* 16, 1423–1435.
- Holland, P., Welsch, R., 1977. Robust regression using iteratively reweighted least-squares. *Communications in Statistics-Theory and Methods* 6, 813–827.
- Huang, D., Swanson, E.A., Lin, C.P., Schuman, J.S., Stinson, W.G., Chang, W., Hee, M.R., Flotte, T., Gregory, K., Puliafito, C.A., et al., 1991. Optical coherence tomography. *Science* 254, 1178–1181.
- Itti, L., Chang, L., Mangin, J., Darcourt, J., Ernst, T., 1997. Robust multimodality registration for brain mapping. *Human brain mapping* 5, 3–17.
- Jenkinson, M., 1999. Measuring transformation error by RMS deviation. Tech. Rep. TR99MJ1, Oxford Center for Functional Magnetic Resonance Imaging of the Brain (FMRIB).
- Jenkinson, M., Bannister, P., Brady, M., Smith, S., et al., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841.
- Joshi, N., Kadir, T., Brady, M., 2011. Simplified computation for nonparametric windows method of probability density function estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 1673–1680.
- Kaji, S., Hirose, S., Ochiai, H., Anjyo, K., 2013. A lie theoretic parameterization of affine transformation. *Proceedings of MEIS2013 (Mathematical Progress in Expressive Image, Synthesis 2013)*, 134–140.
- Kaji, S., Ochiai, H., 2014. A concise parametrization of affine transformation. submitted for publication 52, 53.
- van der Kouwe, A., Benner, T., Salat, D., Fischl, B., 2008. Brain morphometry with multiecho mprage. *Neuroimage* 40, 559–569.
- Lorenzi, M., Ayache, N., Frisoni, G.B., Pennec, X., 2013. Ledemons: a robust and accurate diffeomorphic registration algorithm. *NeuroImage*.
- Magnain, C., Augustinack, J.C., Reuter, M., Wachinger, C., Frosch, M.P., Ragan, T., Akkin, T., Wedeen, V.J., Boas, D.A., Fischl, B., 2014. Blockface histology with optical coherence tomography: A comparison with nissl staining. *NeuroImage* 84, 524–533.
- Maintz, J., van den Elsen, P., Viergever, M., 1996. Comparison of edge-based and ridge-based registration of CT and MR brain images. *Medical image analysis* 1, 151–161.
- Malis, E., 2008. Méthodologies d’estimation et de commande à partir d’un système de vision. Habilitation. Nice-Sophia Antipolis.
- Mellor, M., Brady, M., 2005. Phase mutual information as a similarity measure for registration. *Medical Image Analysis* 9, 330–343.
- Mugler, J.P., Brookeman, J.R., 1990. Three-dimensional magnetization-prepared rapid gradient-echo imaging (3d mp rage). *Magnetic Resonance in Medicine* 15, 152–157.
- Myronenko, A., Song, X., 2010. Intensity-based image registration by minimizing residual complexity. *Medical Imaging, IEEE Transactions on* 29, 1882–1891.
- Nestares, O., Heeger, D., 2000. Robust multiresolution alignment of mri brain volumes. *Magnetic Resonance in Medicine* 43, 705–715.
- Periaswamy, S., Farid, H., et al., 2006. Medical image registration with partial data. *Medical Image Analysis* 10, 452–464.
- Prastawa, M., Bullitt, E., Gerig, G., 2009. Simulation of brain tumors in MR images for evaluation of segmentation efficacy. *Medical image analysis* 13, 297–311.
- Reuter, M., Fischl, B., 2011. Avoiding asymmetry-induced bias in longitudinal image processing. *NeuroImage* 57, 19–21.
- Reuter, M., Rosas, H.D., Fischl, B., 2010. Highly accurate inverse consistent registration: A robust approach. *NeuroImage* 53, 1181–1196.
- Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage* 61, 1402–1418.
- Roche, A., Malandain, G., Ayache, N., 2000. Unifying maximum likelihood approaches in medical image registration. *International Journal of Imaging Systems and Technology: Special Issue on 3D Imaging* 11, 71–80.
- Roche, A., Malandain, G., Ayache, N., Prima, S., 1999. Towards a better comprehension of similarity measures used in medical image registration, in: *Medical Image Computing and Computer Assisted Interventions*, Springer Verlag. pp. 555–566.
- Shechtman, E., Irani, M., 2007. Matching Local Self-Similarities across Images and Videos. *IEEE Conference on Computer Vision and Pattern Recognition*.
- Smith, S., 2002. Fast robust automated brain extraction. *Human brain mapping* 17, 143–155.
- Smith, S., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P., Federico, A., De Stefano, N., et al., 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage* 17, 479–489.
- Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-berg, H., Bannister, P.R., Luca, M.D., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., Stefano, N.D., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural mr image analysis and implementation as fsl. *NeuroImage* 23, 208–219.
- Stewart, C.V., 1999. Robust parameter estimation in computer vision. *SIAM review* 41, 513–537.
- Street, J., Carroll, R., Ruppert, D., 1988. A note on computing robust regression estimates via iteratively reweighted least squares. *The American Statistician* 42, 152–154.
- Thompson, W.K., Holland, D., 2011. Bias in tensor based morphometry stat-roi measures may result in unrealistic power estimates. *NeuroImage* 57, 1–4.
- Vercauteren, T., Pennec, X., Malis, E., Perchant, A., Ayache, N., 2007. Insight into efficient image registration techniques and the demons algorithm. *Information Processing in Medical Imaging (IPMI)*.
- Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2009. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage* 45, 61–72.
- Viola, P., Wells, W., 1997. Alignment by Maximization of Mutual Information. *International Journal of Computer Vision* 24, 137–154.
- Wachinger, C., Navab, N., 2009. Similarity Metrics and Efficient Optimization for Simultaneous Registration, in: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Wachinger, C., Navab, N., 2012a. A Contextual Maximum Likelihood Framework for Modeling Image Registration, in: *IEEE*

- Computer Society Conference on Computer Vision and Pattern Recognition (CVPR).
- Wachinger, C., Navab, N., 2012b. Entropy and laplacian images: Structural representations for multi-modal registration. *Medical Image Analysis* 16, 1 – 17.
- Wachinger, C., Navab, N., 2013. Simultaneous registration of multiple images: Similarity metrics and efficient optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 1221–1233.
- Wong, A., Orchard, J., 2009. Robust multimodal registration using local phase-coherence representations. *Journal of Signal Processing Systems* 54, 89–100.
- Yaqoob, Z., Wu, J., Yang, C., 2005. Spectral domain optical coherence tomography: a better oct imaging strategy. *BioTechniques* 39.
- Yushkevich, P., Avants, B., Das, S., Pluta, J., Altinay, M., Craige, C., 2010. Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: an illustration in adni 3 t mri data. *Neuroimage* 50, 434–445.
- Zefran, M., Kumar, V., Croke, C., 1998. On the generation of smooth three-dimensional rigid body motions. *Robotics and Automation, IEEE Transactions on* 14, 576–589.
- Zeng, Q., Chen, Y., 2008. Accurate inverse consistent non-rigid image registration and its application on automatic re-contouring. *Bioinformatics Research and applications* , 293–304.