# Self-Similarity Weighted Mutual Information: A New Nonrigid Image Registration Metric

Hassan Rivaz, Zahra Karimaghaloo, D. Louis Collins

*McConnell Brain Imaging Center (BIC)*
*Montreal Neurological Institute (MNI)*
*McGill University*
*Montreal, QC, Canada*

**Abstract**

Mutual information (MI) has been widely used as a similarity measure for rigid registration of multi-modal and uni-modal medical images. However, robust application of MI to deformable registration is challenging mainly because rich structural information, which are critical cues for successful deformable registration, are not incorporated into MI. We propose a self-similarity weighted graph-based implementation of $\alpha$-mutual information ($\alpha$-MI) for nonrigid image registration. We use a self-similarity measure that uses local structural information and is invariant to rotation and to local affine intensity distortions, and therefore the new Self Similarity $\alpha$-MI (SeSaMI) metric inherits these properties and is robust against signal non-stationarity and intensity distortions. We have used SeSaMI as the similarity measure in a regularized cost function with B-spline deformation field to achieve nonrigid registration. Since the gradient of SeSaMI can be derived analytically, the cost function can be efficiently optimized using stochastic gradient descent methods. We show that SeSaMI produces a robust and smooth cost function and outperforms the state of the art statistical based similarity metrics in simulation and using data from image-guided neurosurgery.

*Keywords:* Nonrigid registration, $\alpha$-Mutual Information, Nonlocal Means, Self-Similarity, Mutual Information

## 1. Intro

Image registration involves finding the transformation that aligns one image to the second, and has numerous medical applications in diagnosis and in image guided surgery/therapy. The joint intensity histogram of two images, be they from different or the same modalities, is spread (i.e. the joint entropy is high) when they are not aligned, and is compact (i.e. the joint entropy is low) when the two images are aligned. Therefore, mutual information (MI) (Wells et al. (1996); Maes et al. (1997); Pluim et al. (2003)) and the overlap invariant normalized MI (NMI) (Studholme et al. (1999)) have been proposed and widely used for rigid registration of multi-modal images.

MI is not robust against spatially varying bias fields since they result in different intensity relations between the two images at different locations. Therefore, Studholme et al. (2006) and Loeckx et al. (2010) proposed respectively regional MI (RMI) and conditional MI (CMI) where spatial information is used as an extra channel for conditioning MI. This essentially leads to summing MI calculated for regions of the images, instead of globally estimating MI. Klein et al. (2008) proposed localized MI (LMI) where samples are randomly selected from regions in every iteration and convergence is achieved by using stochastic optimization Klein et al. (2007, 2009). Zhuang et al. (2011) proposed spatially encoded MI, which instead of giving equal weights to all pixels in a region, hierarchically weights pixel contributions based on their spatial location. These methods have shown to significantly improve the registration results in the presence of bias fields. Recently, Darkner and Sporring (in press) provided a unifying framework for NMI and other common similarity measures and shed more intuition towards local histograms.

A second difficulty rises because MI does not directly take into account local structures. Therefore, nonrigid registration, which has considerably more degrees of freedom, can distort local structures. Utilizing image gradients and their orientations was proposed by Pluim et al. (2000). Recently, De Nigris et al. (2012) proposed a gradient orientation metric that adaptively controls the trade-off between smooth or accurate cost functions. The HAMMER framework of Shen and Davatzikos (2002) sets local geometric moment invariants as attribute vectors of each voxel in the image. These attribute vectors are then used to form a cost function, which is hierarchically optimized to give the transformation parameters. Xue et al. (2004) later used wavelet-based attributes as local morphological signatures for each voxel. Recently, Ou et al. (2011) introduced Gabor attributes which can be used for different imaging modalities and tissue organs, and further utilizes mutual saliency to weight different voxels based on their local appearance. Taking a different approach, Wachinger and Navab (2012) generated entropy images independently from each image by calculating entropy in small patches around every pixel. They show that since different imaging modalities show the same tissue structure, their entropy images are similar and therefore they can be registered using monomodal registration. In addition to the entropy image representation, they show that structural information of patches can be encoded into a scalar value using manifold learning tech-

*Email address:* hassan.rivaz@mcgill.ca (Hassan Rivaz)

niques. Performing the same technique on both images, they again arrive at two representations (one for each image) which can be registered using monomodal techniques.

A third problem with MI lies in the fact that the *infinite dimensional* joint and marginal probability distributions [1] are required to calculate the *scalar* parameter MI. Most MI estimation methods (Wells et al. (1996); Maes et al. (1997); Pluim et al. (2003)) substitute non-parametric density estimators, such as Parzen windows, into the MI formulation, and are called "plug-in" estimation in Beirlant et al. (1997). An inherit problem of these methods is due to the infinite dimension of the unconstrained densities. Strict smoothness constraints or lower dimensional parametrization must be enforced to estimate these densities, which can cause significant bias in the estimate (Hero et al., 2002). Graph-based entropy estimators (Hero et al., 2002; Neemuchwala and Hero, 2005) have been proposed to directly calculate entropy without the need for performing density estimation. Therefore, these methods have faster asymptotic convergence rate especially for non-smooth densities and high dimensional feature spaces (Hero et al., 2002). Two drawbacks of these methods are their computational complexity and the discontinuity of their gradient as the graph topology changes.

Towards developing a bias invariant similarity metric for nonrigid registration that also takes into account structural information, we build on our previous work (Rivaz and Collins, 2012) to incorporate image self-similarity into MI formulation. Self-similarity estimates the similarity of a patch in one of the images to other patches in the *same* image, and attributes the similarity to the pixels in the center of the patches. Based on patches, self-similarity depends on local structures which are ignored by MI. Buades et al. (2005) first proposed exploiting repetitive regions (or patches) in the form of non local means for image denoising. A recent comparative study of these methods is provided in Buades et al. (2010). Self-similarity was later used for object detection and image retrieval (Shechtman and Irani (2007)), and it has since been used successfully in denoising MR (Coupe et al. (2008); Manjon et al. (2012)) and US images (Coupe et al. (2009)), and image segmentation (Coupe et al. (2011)). Compared to our previous work (Rivaz and Collins (2012)), we present significantly more details and in-depth analysis of SeSaMI. We also provide extensive results for validation and more analysis of the results.

Recently, Heinrich et al. (2011, 2012) proposed using self-similarity for multimodal image registration. The similarity of a pixel to its neighbors, calculated using sum of square differences (SSD), are attributed to the pixel as multi-dimensional descriptors. These descriptors are calculated independently for both images. The multi-modal image similarity is then defined as the SSD of the descriptors of the two images.

Since self-similarity is calculated for pairs of points, it is natural to perceive it in a graph representation where image pixels are vertices and self-similarity is the weight of the edges. Graph-based estimators of $\alpha$-mutual information ($\alpha$-MI) similarity metric have recently been proposed for both rigid (Neemuchwala and Hero (2005); Sabuncu and Ramadge (2008); Kybic (2007); Kybic and Vnucko (2012)) and nonrigid (Staring et al. (2009); Oubel et al. (2012)) registration applications. These methods have been shown to outperform the traditional "plug-in" entropy estimators for MI calculation. Therefore, we choose to incorporate self-similarity into this registration framework.

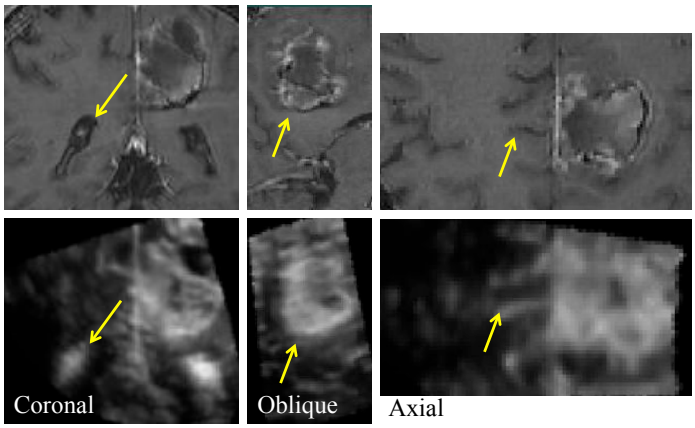



Figure 1: Corresponding pre-operative MR (top row) and intra-operative US (bottom row) images of neurosurgery. A US volume is reconstructed from 2D US slices using tracking data, and is then re-sliced in the shown directions. The lateral ventricles, the boundaries of a tumor and sulci can be seen in both MR and US. While local structures correspond, intensities are not related globally.

We apply the method to register pre-operative magnetic resonance (MR) images to intra-operative ultrasound (US) images in the context of image-guided neurosurgery (IGNS). Previous work that registers US to other modalities is relatively rare: Roche et al. (2001) used the correlation ratio (CR) between US and MR and MR gradient, Arbel et al. (2004) and Mercier et al. (2012b) calculated a lookup table for mapping US and MR intensities and used the monomodal registration of Collins et al. (1999), Kuklisova-Murgasova et al. (2012) segmented the MR volume using a probabilistic atlas, generated a US-like volume from the segmented MR volume, and then registered the US-like volume with the US volume using robust monomodal block-matching techniques, Penney et al. (2004) generated blood vessels probability maps from from US and MR and registered these maps using cross correlation, Ji et al. (2008) used NMI of US and MR, Zhang et al. (2011) used MI of phase information to register US to MR, De Nigris et al. (2012) optimized MI of gradient orientations to register US to MR, Wein et al. (2013) assumed a linear relationship between US intensities and MR intensities and gradient magnitudes, and finally Heinrich et al. (2013) used the self-similarity context along with a discrete optimization approach through block-wise parametric transformation model with belief propagation.

Most of the aforementioned methods simulate US images from the MR data as described. These methods cannot be readily applied to IGNS due to the variety of pathologies that the brain tissue might have, such as different grade gliomas. The

[1] The probability distributions are infinite dimensional if we assume image intensities take real continuous values. However, since intensities of digital images are discrete and finite, the probabilities distributions are finite, but still very high dimensional.

appearance of these pathologies in MR and US are also highly variable (Mercier et al., 2012b,a), adding to the difficulty. We assume no *a priori* relationship between intensities but opt for two non-parametric MI based methods for validating our results.

Figure 1 shows an example of the registered US and MR images. The US images suffer from strong bias field due to signal attenuation, caused by scattering (from smaller than US wavelength inhomogeneities), specular reflection (from tissue boundaries) and absorption (as heat). In addition, US beam width varies significantly with depth, and therefore the same tissue may look different at different depths. A final and important source of spatial inhomogeneities is the time gain compensation (TGC) which is manually adjusted on US machines. Hence, it is critical to exploit local structures.

Our algorithm only needs the self-similarity of one of the images. In most image guided applications, one of the images is pre-operative, and therefore the self-similarity estimation can be performed offline, resulting in a small increase in the online computational complexity. In addition, the pre-operative image is also usually of higher quality, making it a more attractive choice. We use a rotation invariant self-similarity metric that is also robust to bias fields, and utilize it in a graph-based $\alpha$-MI method. We call our method the Self Similarity $\alpha$-MI (SeSaMI) algorithm. We show that SeSaMI outperforms LMI and multi-feature $\alpha$-MI in terms of producing a smooth dissimilarity function and registration accuracy.

This paper is organized as following. We first formulate the problem of image registration as an optimization problem, and provide background information for two popular similarity metrics that we use in this work for comparisons. We then elaborate on how we estimate self-similarity between patches. We explain a graph-based $\alpha$-MI similarity metric, and then formulate SeSaMI by incorporating self-similarity into it. We also show how the derivative of SeSaMI can be efficiently estimated. We finally show the results on simulation and patient data for validation.

## 2. Background

Registration of two images $I_m(\boldsymbol{x})$, $I_f(\boldsymbol{x})$: $\boldsymbol{\Omega} \subset \mathbb{R}^d \rightarrow \mathbb{R}$ can be formulated as

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} C, \quad C = S\left(I_f(\boldsymbol{x}), I_m(\mathbf{T}_{\boldsymbol{\mu}}(\boldsymbol{x}))\right) + \frac{\omega_R}{2}\|\nabla \boldsymbol{\mu}\|^2 \qquad (1)$$

where $I_f(\boldsymbol{x})$ and $I_m(\boldsymbol{x})$ are respectively the fixed and moving images, $S$ is a dissimilarity metric, $\omega_R$ is a regularization penalty weight, $\nabla$ is the gradient operator and $\mathbf{T}_{\boldsymbol{\mu}}$ is the transformation modeled by $\boldsymbol{\mu}$. We choose a free-form transformation parameterized by the location of cubic B-spline nodes. Therefore, $\boldsymbol{\mu}$ is a vector of the coordinates of all the nodes. The dissimilarity metric $S$ is the focus of this work. We now briefly elaborate two similarity metrics that we have used for US to MR registration as benchmarks for comparison.

**Normalized Mutual Information (NMI).** NMI is defined in Studholme et al. (1999) as

$$NMI(I_f, I_m) = \frac{H(I_f) + H(I_m)}{H(I_f, I_m)} \qquad (2)$$

where $H(I_f)$ and $H(I_m)$ are the marginal entropies of the fixed and moving images, and $H(I_f, I_m)$ is their joint entropy. This formulation is used by Ji *et al.* Ji et al. (2008) for US and MR registration. A problem with NMI is that it assumes a global statistical relationship between the images, an assumption that can be violated for various reasons such as spatial inhomogeneity.

**Local Mutual Information (LMI).** To take spatial information into account, a popular approach is to consider spatial location as an additional channel and multiply intensities with spatial kernels when calculating the MI (Studholme et al. (2006); Klein et al. (2008); Loeckx et al. (2010); Zhuang et al. (2011)). For comparison, we implement the LMI method (Klein et al. (2008)) where these kernels are box filters. In the other words, LMI is computed by summing MI over multiple local neighborhoods:

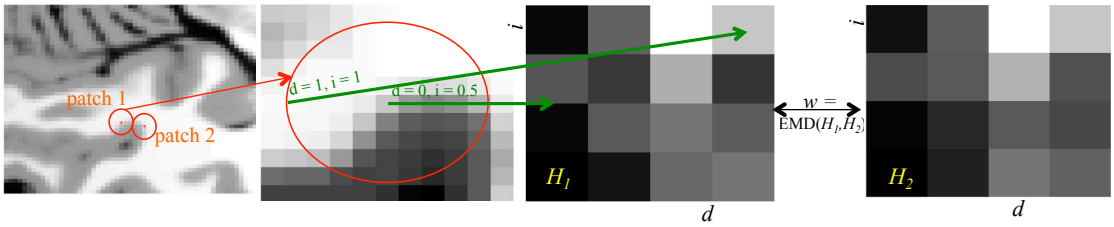
$$LMI(I_f, I_m; \boldsymbol{\Omega}) = \frac{1}{N}\sum_i MI(I_f, I_m; \mathcal{N}_i) \qquad (3)$$

where $\mathcal{N}_i \subset \boldsymbol{\Omega}$ are spatial neighborhoods and $N$ is the number of these neighboorhoods. Each neighborhood should be large enough to contain enough information for MI estimation, and small enough to allow local estimation of MI (Klein et al. (2008)). Similar to (Klein et al. (2008)), we first randomly select a point in $\boldsymbol{\Omega}$ and then select samples from a neighborhood around that point, and repeat this for $N$ points. Also, since US echos are strong at tissue interfaces, they generally show higher correlation with the gradient of MR. We therefore calculate both NMI and LMI between US and the magnitude of the gradient of the MR.

In the next section, we will present a method for calculating self-similarity as a metric that represents contextual information. We will then incorporate this measure into our novel multi-modal similarity metric.

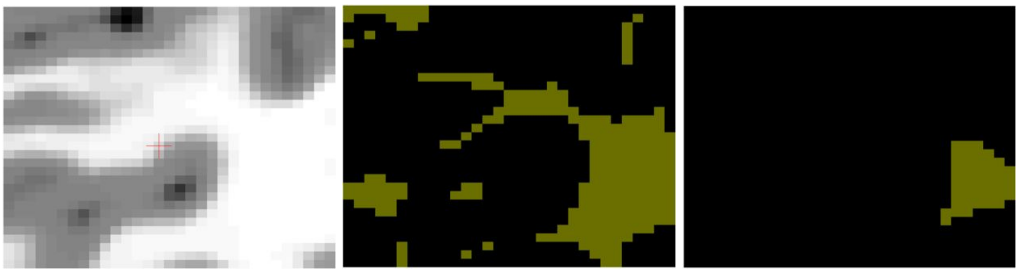
## 3. Rotationally Invariant Self-Similarity Estimation

A small patch centered on a pixel of interest is usually considered for calculating self-similarity. SSD and normalized cross correlation (NCC) of this patch to its neighboring patches are good options for calculating self-similarity, since we are dealing with one imaging modality. NCC is advantageous since it is invariant to affine intensity distortions. However, neither measure is rotation invariant. Grewenig et al. (2011) proposed to calculate rotation angles and subsequently rotating patches using interpolation to achieve rotation invariance. This method is however computationally expensive and does not give good results because of errors due to rotation angle computation and due to interpolation. In this Section, we describe a rotation and bias invariant metric in three steps: construction of histogram descriptors, region selection and histogram comparison. Each step is described below.




(a) T1 image (b) Zoomed patch 1 (c) H (patch 1) (d) H (patch 2)

Figure 2: Construction of the spin image histogram, the output of step 1. (a) is a T1 image from BrainWeb, with two circular patches of radius $r = 4$ pixels around two nominal points. The zoomed image of the patch 1 is shown in (b), and the two descriptor histogram of the patches 1 and 2 in (c) and (d). The $x$ $y$ axes of the histograms are respectively distance $d$ and normalized intensity $i$. Since the two pixels belong to similar structures, their histograms in (c) & (d) are similar.



(a) a 31x31 neighborhood (b) Moran mask, $r = 2$ (c) Moran mask, $r = 4$

Figure 3: Image regions with structure calculated using Moran's $I$, the output of step 2. Two different patch sizes are used as noted. The green parts will be masked out in the next step.

**Step 1: Constructing Histogram Descriptors.** We first estimate a rotationally invariant 2D histogram descriptor for all pixels; such a sample pixel (pointed to by a red arrow) at the center of a circular patch with radius $r = 4$ pixels is shown in Figure 2 left. We show 2D images for clarity; the arguments are trivially extended to 3D images. The rectangle in the left image represents a local neighborhood that the center pixel is compared to, as will be explained in step 3 below. The axes of the histogram are $d$, the normalized Euclidian distance of the pixel from the center point, and $i$, the pixel's normalized intensity. $d = 0$ and $d = 1$ in the histogram respectively correspond to the center pixel and to the pixels on the circle with radius $r = 4$ pixels. Each patch is normalized independently of other patches (i.e. intensities in all patches are mapped to the range [0 1]). Each pixel inside the patch contributes to the 2D histogram: the histogram is constructed using a Gaussian Parzen window of isotropic $\sigma = 0.5$; both distance $d$ and intensity $i$ are normalized and therefore $\sigma$ does not have a unit. In other words, a pixel with distance $d$ to the center and normalized intensity $i$ contributes the bin indexed by $d_b$ and $i_b$ according to

$$exp\left(-\frac{(i-i_b)^2}{2\sigma_i^2}-\frac{(d-d_b)^2}{2\sigma_d^2}\right) \tag{4}$$

where we always set $\sigma_d = \sigma_i = 0.5$. Since $d$ is the distance to the center (i.e. orientation is ignored), the 2D histogram descriptor is rotation invariant. It is also invariant to affine changes in the intensity because of the intensity normalization step. Two histogram descriptors corresponding to the two marked points are shown in Figure 2 right. The histogram descriptor is similar to the spin image used in Lazebnik et al. (2005).

**Step 2: Selecting Regions with Structure.** Parts of the image with little structure do not produce reliable self-similarity measures. Similar to Andronache et al. (2008), we use Moran's $I$ spatial autocorrelation coefficient to limit the self-similarity estimation to parts with structure. The measure is derived from the Pearson's correlation coefficient. For an image patch with intensities $X = \{x_i, i = 1 \cdots N\}$ and the mean value $E(x) = \bar{x}$, Moran's $I$ is

$$I = \frac{N}{\sum_{i,j}^N w_{ij}} \cdot \frac{\sum_{i,j}^N w_{i,j}(x_i-\bar{x})(x_j-\bar{x})}{\sum_i^N (x_i-\bar{x})^2}$$

where $W = w_{ij}$ matrix represents the connectivity weights. It can be a binary map or a decaying map based on the distance between $i$ and $j$. Since $\sum_i^N (x_i-\bar{x})^2 = N\sigma^2$, letting the $z$-value $z_i = (x_i-\bar{x})/\sigma$ allows us to give a more intuitive equation for Moran's $I$:

$$I = \frac{1}{\sum_{i,j}^N w_{ij}} \cdot \sum_i^N z_i \sum_j^N w_{i,j} z_j \tag{5}$$

$I$ varies between -1 to 1; values close to 0 translate to random pattern and values close to 1 or -1 indicate presence of structure. Let $d(i, j)$ be the Euclidian distance between $i$ and $j$. We simply set $W$ to

$$w(i,j) = \begin{cases} 1/d(i,j) & j \neq i \\ 0 & j = i \end{cases} \tag{6}$$

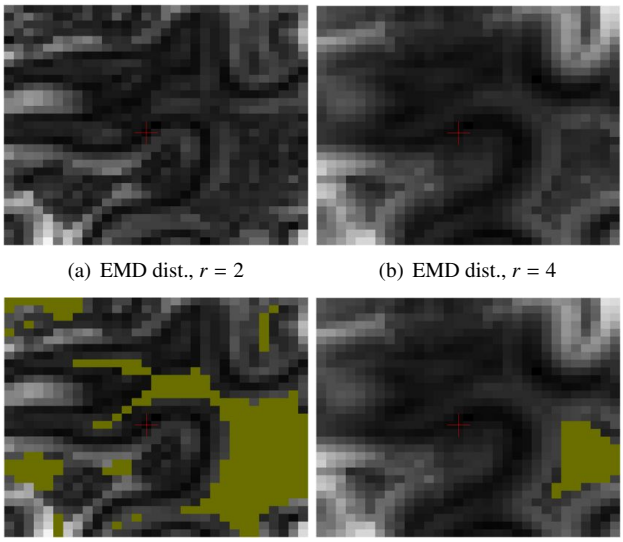
We calculate the $I$ coefficient for patches centered on all pix-

els. We set the patch size to be the same as the one for constructing the histogram descriptors. We then select pixels of the image whose absolute value of Moran's I coefficient $|I|$ is more than the standard deviation of the $|I|$. Figure 3 shows the selected pixels for two different patch sizes. In the next step, we calculate self-similarity between pixels that are selected with Moran's $I$.

**Step 3: Calculating Self-Similarity using Earth Mover's Distance (EMD).** We now calculate self-similarity between two pixels by comparing their 2D histogram descriptor. Most histogram comparison metrics are bin-by-bin similarity metrics where the bins with the same index are compared. Such methods include sum of $L_1$ or $L_2$ norms of differences, Kullback-Leibler (KL) divergence and $\chi^2$ (chi-squared) test. A problem with such metrics is the following. Consider three histograms $H_1 = \{1, 0, 0, 0, 0\}$, $H_2 = \{1/2, 1/2, 0, 0, 0\}$ and $H_3 = \{1/2, 0, 0, 0, 1/2\}$. Such metrics will give the same distance between $H_1$-$H_2$ and $H_1$-$H_3$. However, $H_1$ and $H_2$ are more similar because all of mass of $H_2$ is in its first two bins, compared to $H_1$ which has all of its mass in the first bin, a difference which can be caused by binning. Therefore, we use the Earth Mover's Distance (EMD) (Rubner et al., 2000; Ling and Okada, 2007) which finds the minimum cost required to transform one histogram into another. The cost is the multiplication of the moved mass of the bins and a weight, which depends on the bin indices. EMD is computationally expensive because it needs to optimize a transport cost function; we use EMD-L1 (Ling and Okada, 2007), an efficient implementation of EMD where the weight between two bins is their L1 distance. EMD reduces quantization and binning problems associated with histograms, and has been shown by Rubner et al. (2000) to outperform other histogram comparison techniques.

Figure 4 left shows the EMD distance of all image pixels to the central pixel. The distances are calculated for two different patch radii $r = 2$ and $r = 4$. We empirically found that the histogram sizes of $4 \times 4$ to $7 \times 7$ generate good results; given the computational complexity, we use $4 \times 4$ histogram descriptors throughout this work. Small values of the EMD distance (darker pixels) mean smaller distances and represent more similar regions. We however calculate the EMD distance only to their local neighbors (the rectangle in Figure 2 left). Figure 4 right shows the local self-similarity to pixels that are selected by the Moran's $I$ test. The left two images show that without pixel selection with Moran's $I$, patches that are not similar can get a small EMD distance. Figure 5 shows the EMD distances computed using the 3D volumetric images of the BrainWeb. Here, the local patches are 3D spheres, while the local histogram descriptors are 4×4, the same size as the histograms of 2D images.

We set the EMD distance between the center patch and a structureless patch (with low Moran's $I$ value) to the maximum possible EMD distance. The maximum EMD between two $n \times n$ histograms $H_1$ and $H_2$ happens when $H_1 = \delta[i-1, j-1]$ and $H_2 = \delta[i-n, j-n]$ where $\delta$ is the impulse function: $\delta[0, 0] = 1$ and is zero everywhere else. These "most different" histograms have zero bins everywhere except for the left bottom corner in $H_1$ and the top right corner in $H_2$. It can be shown that the EMD distance between these two $n \times n$ histograms is $\max_{H_1,H_2} \text{EMD}(H_1, H_2) = 2(n-1)$. For our $4 \times 4$ histograms, this number is 6.



(a) EMD dist., $r = 2$ (b) EMD dist., $r = 4$

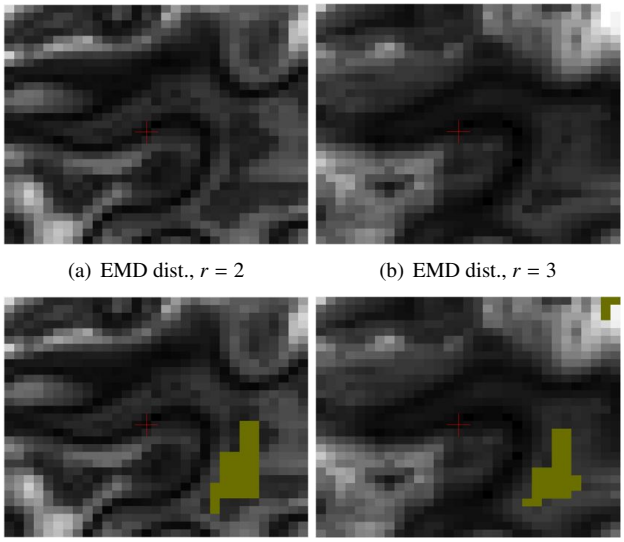
(c) EMD dist., $r = 2$ (d) EMD dist., $r = 4$

Figure 4: The self-similarity distance of the center pixel to other image pixels, computed using the EMD between spin images of Figure 2. Black represents smaller distance, i.e. more similar. The patch radius is $r$. In (a) and (b), the Moran's mask is not used, and therefore some dissimilar pixels have small distance. In (c) and (d), these incorrect small distances are excluded using Moran's $I$.

It can be seen from Figure 4 that the similarity metric is fully rotation invariant. The computational complexity of calculating the EMD distance is not an issue since it can be calculated offline on only the pre-operative image.

The histogram descriptor provides stability against small deformations of structures (due to the binning process), while subdividing the distance to the center ($d$ in the histogram) encodes the spatial information. As a result, it is more robust than filter banks and differential invariants, which are also local descriptors (Lazebnik et al., 2005). Its disadvantage is its computational complexity. Performing self-similarity estimations on a volume of size $100^3$ pixels takes about 5 hours on one core of a 3GHz processor.

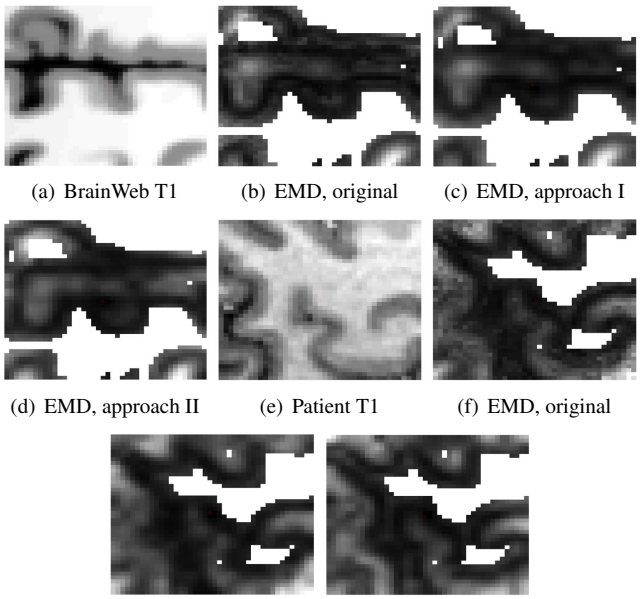
To speed self-similarity estimation, we propose the two following approaches, which are based on the observation that self-similarity maps generated by EMD are smooth and therefore can be sampled at lower resolutions. (I) When computing the self-similarity of a pixel to others, compute only one self-similarity for every $2^3$ cube of pixels, i.e. skip one pixel in all dimensions. (II) Downsample the volume by a factor of 2 in each dimension and perform self-similarity estimation on this volume, with a patch size that is also twice smaller. Using the approaches (I) or (II), the self-similarity at the original scale can be approximated using linear interpolation. The speed gains are approximately factors of $2^3 = 8$ and $2^6 = 64$ respectively. We test these approaches on T1 images of BrainWeb (Collins et al., 1998) and MR images of our IGNS trials. Both volumes have an isotropic pixel size of 1 mm. Figure 6, (b) and (f) show the



(a) EMD dist., $r = 2$ (b) EMD dist., $r = 3$

(c) EMD dist., $r = 2$ (d) EMD dist., $r = 3$

Figure 5: The self-similarity distance of the center pixel to other image pixels, computed in 3D.



(a) BrainWeb T1 (b) EMD, original (c) EMD, approach I

(d) EMD, approach II (e) Patient T1 (f) EMD, original

(g) EMD, approach I (h) EMD, approach II

Figure 6: The effect of downsampling on EMD distance. First and Second row are respectively BrainWeb and patient data. Please see the text for details.

EMD self-similarities to the center pixel at the original resolution, (c) and (g) show the results of the approach (I), and (d) and (h) show the results of the approach (II). The weights in the third and forth columns are very similar to that of the second column ((b)&(f)), which allows computing the weights at a coarse resolution. We use the approach (II) on patient data, and therefore the self-similarities are estimated in less than 10 min on one core of a 3GHz processor for a volume of size $100^3$ pixels.

The rich structural self-similarities between two pixel locations are now encoded in a single number, the EMD distance. We will now use this distance to efficiently incorporate structural information into $\alpha$-MI. In the next section, we first briefly explain $\alpha$-MI and then formulate SeSaMI.

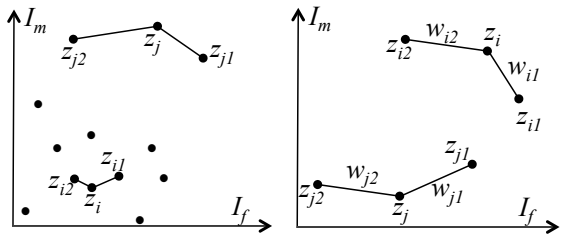



Figure 7: Joint feature space with one feature from each image. The superscript $fm$ is omitted from $z$ for more clarity. In the left image, the density of points is respectively high and low close to $z_i$ and $z_j$, and therefore close (for $z_i$) and far points (for $z_j$) are taken into account for $k$NN estimation of $\alpha$-MI. This can be perceived as having adaptive bin sizes, depending on the data distribution. In the right image, the self-similarity weights $w_{ip}$ are shown for the graph edges.

## 4. Self-Similarity $\alpha$-MI (SeSaMI)

**Multi-feature graph-based $\alpha$-MI.** The MI similarity metric is usually calculated on the intensities only, and therefore the joint histogram is 2D. $\alpha$-MI is usually calculated on multiple features like intensities and their gradients. Adopting the notation of Staring et al. (2009), let $\boldsymbol{z}(\boldsymbol{x}_i) = [z_1(\boldsymbol{x}_i) \cdots z_q(\boldsymbol{x}_i)]^T$ be a $q$-dimensional vector containing all the features at point $\boldsymbol{x}_i$ ($\boldsymbol{z}(\boldsymbol{x}_i)$ is not related to the $z$-score used for estimating Moran's $I$ in Eq. 5). Similar to Staring et al. (2009), we choose image intensities and gradients at two different scales as features, resulting in four total features. Let $\boldsymbol{z}^f(\boldsymbol{x}_i)$ and $\boldsymbol{z}^m(\mathbf{T}_\mu(\boldsymbol{x}_i))$ be respectively the features of the fixed and moving image at $\boldsymbol{x}_i$ and $\mathbf{T}_\mu(\boldsymbol{x}_i)$, and $\boldsymbol{z}^{fm}(\boldsymbol{x}_i, \mathbf{T}_\mu(\boldsymbol{x}_i))$ be their concatenation $[\boldsymbol{z}^f(\boldsymbol{x}_i)^T \;\; \boldsymbol{z}^m(\mathbf{T}_\mu(\boldsymbol{x}_i))^T]^T$. $\boldsymbol{z}^{fm}$ is in the joint feature space. Minimal spanning tree (MST) (Sabuncu and Ramadge, 2008) and $k$-nearest neighbor ($k$NN) (Staring et al., 2009; Oubel et al., 2012) are among different methods for estimating $\alpha$-MI from multi-feature samples. With $N$ samples, the complexities of constructing MST and $k$NN graphs are $O(N^2 \log N)$ and $O(N \log N)$ respectively (Neemuchwala and Hero, 2005). Therefore, we choose $k$NN.

Let $\boldsymbol{z}^f(\boldsymbol{x}_{ip})$, $\boldsymbol{z}^m(\mathbf{T}_\mu(\boldsymbol{x}_{ip}))$ and $\boldsymbol{z}^{fm}(\boldsymbol{x}_{ip}, \mathbf{T}_\mu(\boldsymbol{x}_{ip}))$ be respectively the $p$th nearest neighbors of $\boldsymbol{z}^f(\boldsymbol{x}_i)$, $\boldsymbol{z}^m(\mathbf{T}_\mu(\boldsymbol{x}_i))$ and $\boldsymbol{z}^{fm}(\boldsymbol{x}_i, \mathbf{T}_\mu(\boldsymbol{x}_i))$ (see Figure 7 left). Note that these three nearest neighbors in general do not correspond to the same point. We limit the nearest neighbor search to local neighborhood of $\boldsymbol{x}_i$ and $T_\mu(\boldsymbol{x}_i)$. The size of this local neighborhood is the same as that of the self-similarity analysis. To prevent notation clutter, we show the dependencies on location $\boldsymbol{x}_i$ or $\mathbf{T}_\mu(\boldsymbol{x}_i)$ only through $i$ after this point whenever clear. Let $\boldsymbol{d}^f_{ip} = \boldsymbol{z}^f_i - \boldsymbol{z}^f_{ip}$, $\boldsymbol{d}^m_{ip} = \boldsymbol{z}^m_i - \boldsymbol{z}^m_{ip}$ and $\boldsymbol{d}^{fm}_{ip} = \boldsymbol{z}^{fm}_i - \boldsymbol{z}^{fm}_{ip}$ be the vectors that connect the node $i$ to its $p$th nearest neighbor in respectively the fixed, moving and joint

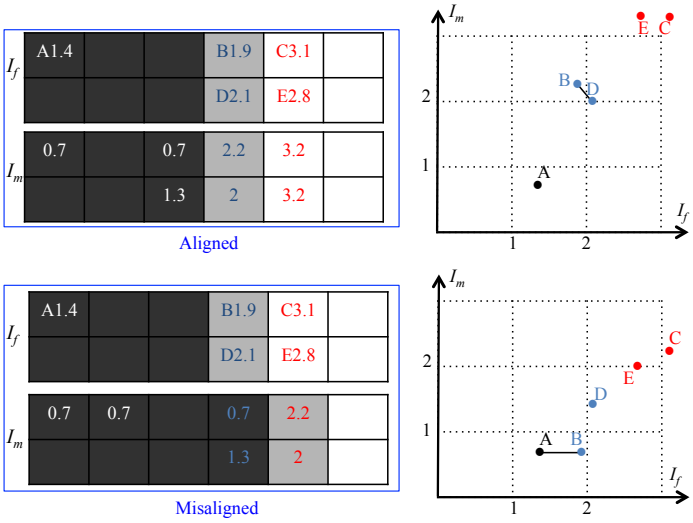



Figure 8: The nearest neighbors in the joint feature space of the aligned images are more likely to be self-similar compared to misaligned images. Please refer to the text for details.

feature spaces. For a node $i$, set

$$\Gamma_i^{space}(\boldsymbol{\mu}) = \sum_{p=1}^{k} \|\boldsymbol{d}_{ip}^{space}\|, \;\; space = f, m, \text{and } fm \qquad (7)$$

where $\|.\|$ is the Euclidian norm. A $k$NN estimator for $\alpha$-MI=$-S$ (the dissimilarity function in Eq. 1) is

$$\widehat{\alpha\text{-MI}}(\boldsymbol{\mu}) = \frac{1}{\alpha - 1} \log \frac{1}{N^\alpha} \sum_{i=1}^{N} \left( \frac{\Gamma_i^{fm}(\boldsymbol{\mu})}{\sqrt{\Gamma_i^f \Gamma_i^m(\boldsymbol{\mu})}} \right)^{2\gamma} \qquad (8)$$

where $\gamma = (1 - \alpha)q$ and $0 < \alpha < 1$ ($q$ is the number of dimensions of $\boldsymbol{z}$ as mentioned at the beginning of this Section); experimental results of rigid registration in (Sabuncu and Ramadge, 2008) suggest that for MST graphs, $\alpha$ close to 1 gives better registration accuracy, while $\alpha$ close to 0.5 yields a wider capture range. To achieve a wide capture range, we perform hierarchical registration. Therefore, to achieve high accuracy we use $\alpha = 0.9$ throughout this work.

**Weighting $\alpha$-MI by self-similarity.** In an analogy to MI, small $\Gamma_i^{fm}$ for majority of locations $i$ means that data in the joint histogram is clustered and compact, and $\Gamma_i^f$ and $\Gamma_i^m$ are for normalization. Therefore, accurate estimates of $\Gamma_i^{fm}$ are essential. In registered images, generally most of the nearest neighbors in the joint feature space correspond to similar patches in the fixed image. However, due to spatially varying bias, small geometrical distortions, lack of an adequate number of features and misalignment, not all the nearest neighbors are self-similar. Therefore, to penalize points that are close but are not self-similar, we modify $\Gamma_i^{fm}$ by (see Figure 7 right)

$$\Gamma_i^{fm}(\boldsymbol{\mu}) = \sum_{p=1}^{k} w_{ip} \|\boldsymbol{d}_{ip}^{fm}\|, \; w_{ip} = \text{EMD}(H(\boldsymbol{x}_i), H(\boldsymbol{x}_{ip})) \qquad (9)$$

where $\text{EMD}(H(\boldsymbol{x}_i), H(\boldsymbol{x}_{ip}))$ is the EMD between the histogram descriptors. Further intuition for this equation is provided in Figure 8. Images $I_f$ and $I_m$ each have two rows and six columns. The letters in $I_f$ pixel encode location (i.e. indices $i, p$ in Eq. 7) and the numbers represent intensity (i.e. features $\boldsymbol{z}$ in Eq. 7). Pixels are also colored to aid visualization: values close to 1, 2 and 3 are respectively dark gray, gray and white. The intensity of every pixel is the only feature used, and therefore the joint feature space in this figure is simply 2D. When the images are aligned, the nearest neighbor of point B is D, which is from a similar structure (i.e. at the border of low and high intensities). In the misaligned case, the nearest neighbor of B is A, a pixel from a completely different structure. Since this nearest neighbor is not particularly far in the feature space, i.e. $\|\boldsymbol{z}_A^{fm} - \boldsymbol{z}_B^{fm}\|$ is small, this can result in a low $\Gamma^{fm}$ and therefore an incorrect local minima. However, multiplying distances by the self-similarity weight according to Eq. 9 can prevent this since $w_{BA}$ is a large number.

**Optimization of the cost function.** We adopt an iterative stochastic gradient descent optimization method (Klein et al., 2007) for solving Eq. 1, which is fast and is less likely to get trapped in local minima. In this method, the cost function and its derivative are calculated from new randomly selected subset of pixels in each iteration. We select these points from pixels with high values of Moran's $I$ since we do not calculate self-similarity of pixels with low $I$ to other pixels. Some of $k$NN of these pixels, however, can be from pixels with low $I$. Letting $\nabla_{\boldsymbol{\mu}} C$ be the gradient of $C$ (from Eq. 1) w.r.t. $\boldsymbol{\mu}$, the update equation is $\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t + a_t \nabla_{\boldsymbol{\mu}} C$. The step size is a decaying function of the iteration number $t$:

$$a_t = a/(A + t)^\tau \qquad (10)$$

with $a > 0, A \geq 0$ and $0 < \tau \leq 1$ user-defined constants Klein et al. (2007). The recommended values for these parameters are provided in (Klein et al., 2007): $A$ should be around 0.1 of the maximum number of iterations or less and $\tau$ should be more than 0.6. The value of $a$ is critical as it determines the step-size; its value is user-defined. Generally speaking, if $a$ is too small more iterations are required and it is also more likely that the optimization gets trapped in a local minima. On the other hand, the registration can diverge if $a$ is too large. Fortunately, for large enough number of iterations the final registration result varies negligibly if $a$ is varied by as much as 100%. $a$ also depends on the similarity metric; we set it to values between 1 and $10^4$ by multiplying it by 10 each time and evaluated the deformation at each iteration. After we found its order of magnitude, we varied it by smaller steps and finally set it to 500 for NMI and LMI and to 200 for $\alpha$-MI and SeSaMI.

From Eq. 1, we have $\nabla_{\boldsymbol{\mu}} C = -\nabla_{\boldsymbol{\mu}} \widehat{\alpha\text{-MI}} + \omega_R \Delta \boldsymbol{\mu}$ where $\Delta = \nabla.\nabla$ is the Laplacian operator. Numerical methods, such as finite difference and simultaneous perturbation (Spall, 1998)

can be used to estimate the derivative of the cost function. However, analytically estimating the derivative is usually faster and has better convergence properties. Assuming the topology does not change for small changes in $\boldsymbol{\mu}$, the gradient of $\widehat{\alpha\text{-MI}}$ is calculated analytically in Staring et al. (2009); Oubel et al. (2012) using the chain rule; we refer the reader to them for details. The chain rule finally results in computation of the $\nabla_{\boldsymbol{\mu}}\Gamma_i^{fm}(\boldsymbol{\mu})$. From Eq. 9, we have

$$\begin{aligned}\frac{\partial}{\partial\mu_j}\Gamma_i^{fm}(\boldsymbol{\mu}) &= \sum_{p=1}^{k}\left(w_{ip}\frac{\boldsymbol{d}_{ip}^{fm\,T}}{\|\boldsymbol{d}_{ip}^{fm}\|}\cdot\frac{\partial}{\partial\mu_j}\boldsymbol{d}_{ip}^{fm} + \|\boldsymbol{d}_{ip}^{fm}\|\frac{\partial}{\partial\mu_j}w_{ip}\right)\\ &= \sum_{p=1}^{k}\frac{w_{ip}}{\|\boldsymbol{d}_{ip}^{fm}\|}\boldsymbol{d}_{ip}^{m\,T}\cdot\frac{\partial}{\partial\mu_j}\boldsymbol{d}_{ip}^{m}\end{aligned} \tag{11}$$

where $T$ means transpose. We have ignored the term which involves the derivative of $w_{ip}$ because it is calculated for either $I_f$ or $I_m$; in the former case, its derivative w.r.t. $\boldsymbol{\mu}$ is trivially zero, and in the latter case it is zero because the 2D histogram descriptors are invariant to small deformations (Lazebnik et al., 2005). Also, even for large global deformations, the histogram patches can be assumed locally rigid. The equality in the second line is true because $\boldsymbol{z}^{fm}$ is the concatenation of $\boldsymbol{z}^f$ and $\boldsymbol{z}^m$, and $\partial\boldsymbol{z}^f/\partial\boldsymbol{\mu} = 0$. Finally,

$$\boldsymbol{d}_{ip}^{m\,T}\cdot\frac{\partial}{\partial\mu_j}\boldsymbol{d}_{ip}^{m} = \boldsymbol{d}_{ip}^{m\,T}\cdot\left(\frac{\partial}{\partial\mathbf{T}(\boldsymbol{x}_i)}z_i^m\frac{\partial}{\partial\mu_j}\mathbf{T}(\boldsymbol{x}_i) - \frac{\partial}{\partial\mathbf{T}(\boldsymbol{x}_{ip})}z_{ip}^m\frac{\partial}{\partial\mu_j}\mathbf{T}(\boldsymbol{x}_{ip})\right). \tag{12}$$

Note that partial derivative of $\boldsymbol{z}^m$ w.r.t. $\mathbf{T}$ means calculating derivatives in $I_m$'s native space, i.e. w.r.t. its own $x$, $y$ or $z$ coordinate. In our implementation, we pre-compute all the features of the $I_f$ and $I_m$, and the derivatives of $I_m$'s features w.r.t. $x$, $y$ and $z$ directions.

For each $\boldsymbol{d}_{ip}^m$, the partial derivatives of Eq. 12 should be estimated for $2r4^r$ nodes (i.e. $\mu_j$) where $r$ is the image dimension, 4 is the number of the nodes of the cubic B-splines and 2 is because of having two points $i$ and $ip$. For 2D and 3D images, that is 64 and 384 respectively, making gradient estimation computationally expensive. To speed the derivative calculation, we use the chain rule $\frac{\partial}{\partial\mu_j}\widehat{\alpha\text{-MI}} = \frac{\partial}{\partial\mathbf{T}}\widehat{\alpha\text{-MI}}^T\cdot\frac{\partial}{\partial\mu_j}\mathbf{T}$ (again, note that $\mathbf{T}$ and $\boldsymbol{\mu}$ are simply $x$, $y$ or $z$ of respectively all $I_m$ pixels and B-spline nodes); therefore for any $\boldsymbol{d}_{ip}^m$ only partial derivatives w.r.t. $\mathbf{T}(\boldsymbol{x}_i)$ and $\mathbf{T}(\boldsymbol{x}_{ip})$ ($2r$ partial derivatives) are non-zero and should be computed, instead of the $2r4^r$ values of Eq. 12. In our implementation $\frac{\partial}{\partial\mathbf{T}}\widehat{\alpha\text{-MI}}$ is a matrix of size $r\times$`sizeof(`$I_m$`)`, easily multiplied by the transformation Jacobian $\frac{\partial}{\partial\mu_j}\mathbf{T}$ at the last step of the gradient computation.

We then accumulate all the summations into a matrix of dimension $r\times$`sizeof(`$I_m$`)`, where each entry corresponds to the partial derivative w.r.t. $x$, $y$ or $z$ of that pixel. Formally speaking, we factor out the transformation Jacobian $\frac{\partial}{\partial\mu_j}\mathbf{T}(\boldsymbol{x})$ until the very end and multiply it by $\frac{\partial}{\partial\mathbf{T}}\widehat{\alpha\text{-MI}}$ of all pixels according to the chain rule $\frac{\partial}{\partial\mu_j}\widehat{\alpha\text{-MI}} = \frac{\partial}{\partial\mathbf{T}}\widehat{\alpha\text{-MI}}^T\cdot\frac{\partial}{\partial\mu_j}\mathbf{T}(\boldsymbol{x})$

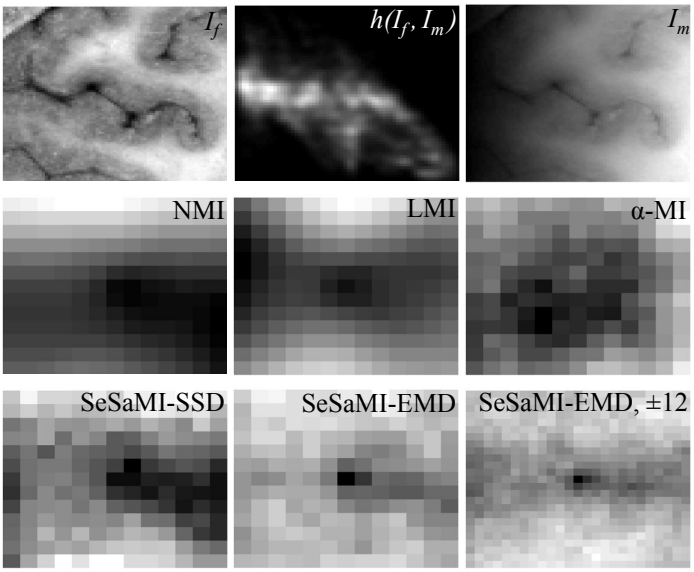



Figure 9: Effect of the bias on the dissimilarity metrics in the human brain images. The similarity metrics are plotted by rigidly moving $I_m$ by a maximum of ±6 pixels, except for the bottom right image where the displacement is ±12 pixels. Black color represents smaller dissimilarity. The self-similarities in the last row are computed from the biased $I_m$ using SSD and EMD as noted. Please refer to the text for details.

## 5. Experiments and Results

Our registration algorithm is implemented in Matlab mex functions. We limit the self-similarity comparisons to local neighborhoods of sizes $40^2$ in 2D and $25^3$ in 3D. To find $k$NN, we use a Matlab wrapper for the approximate nearest neighbor algorithm (Arya et al., 1998). To test the statistical significance of different results, we perform the paired t-test using the MATLAB 2012b software.

### *5.1. Visible human project*

We test the new similarity metric on the red and green channels of brain data of the visible human project, which are intrinsically registered. The data is publicly available at `www.nlm.nih.gov/research/visible/visible_human.html`. We set the red image as $I_m$ and add bias to it (Figure 9 top). The joint histogram for the aligned images is spread due to the bias. We then displace $I_m$ rigidly in the $x$ and $y$ directions, and calculate different similarity metrics at each displacement. The images are aligned at 0 displacement, and therefore the dissimilarity value should be the smallest at the 0 displacement (i.e. the center). The NMI, which is calculated from all $85\times75 = 6375$ image pixels, fails to indicate 0 as the correct alignment due to the strong bias. The LMI is computed by dividing the image in 2 horizontal and 2 vertical pieces, creating a total of 4 neighborhoods each with $21\times19$ pixels. It performs significantly better and predicts 0 as a local minimum, but gives the global minimum at $(x, y) = (-6, 3)$ pixels displacement. $\alpha$-MI and SeSaMI metrics are obtained using 2 features of intensity and gradient. $\alpha$-MI gives an incorrect minimum cost. The self-similarities of the SeSaMI-SSD and SeSaMI-EMD are calculated based on the biased $I_m$, respectively using SSD and the proposed EMD

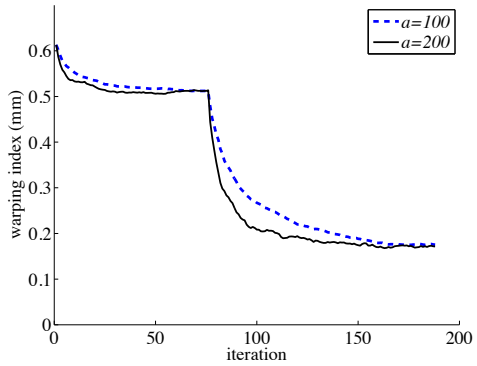



Figure 10: Effect of the optimization parameter $a$, the step size in Eq. 10, on convergence.

method. Comparing SeSaMI-SSD and SeSaMI-EMD, we can see that the EMD self-similarity descriptor outperforms SSD in terms of generating the correct local minimum for the SeSaMI. Note that the results of this figure show that even with a ±6 pixel initialization, only SeSaMI can guide the transformation to the ground truth. To see the extent of the capture range of SeSaMI, we plot its cost function for a larger displacement range of ±12 pixels in the bottom right corner. This figure shows that (0,0) is the global minimum for a relatively large displacement range. This image, however, has incorrect local minima as well, which can trap most common iterative optimization techniques. To solve this problem, a common practice is to perform hierarchical multi-scale optimization, which is what we also adopt.

### *5.2. BrainWeb*

In this section, we study the performance of SeSaMI using 3D simulated MRI images of the BrainWeb Collins et al. (1998). All the images are with 3% noise. The intensity inhomogeneity fields are from 0% to 40% as noted. The image voxels are 1 mm in all directions. We use 4 features of intensity and gradients at two scales for both $\alpha$-MI and SeSaMI. The registrations are performed in 2 coarse and fine levels to increase the speed and reduce the chance of getting trapped in local minima.

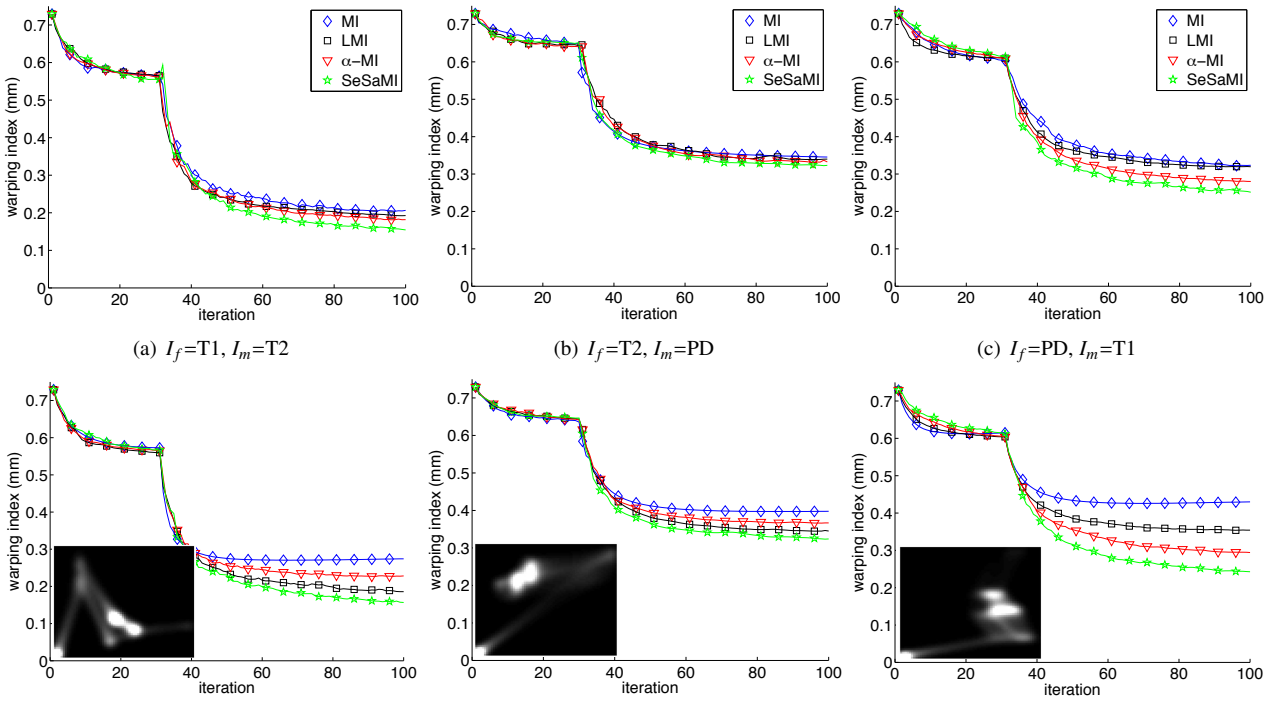
In the first experiment, we test the effect of the only tuneable optimization parameter of SeSaMI, i.e. the parameter $a$ in Eq. 10, in T1-T2 registration. We deform the T1 image with a B-spline deformation field with spacing of 20 mm (or voxels) between nodes in every dimension. We move every node in 3 dimensions by uniformly distributed random numbers between ±9 mm. We then register the T2 image to the T1 image using a B-spline deformation field with 20 mm distance between the nodes. The T1 and T2 images have respectively 0% and 40% intensity inhomogeneities. We measure the *warping index* as an estimate of registration error. The warping index is the root mean square of the displacement error of all image voxels. Figure 10 shows the results. With the larger step size, the method converges faster, but reaches the similar final result error of the small step size. Therefore, an estimate value for this parameter is sufficient to achieve optimal results. We always set it to 200 for both $\alpha$-MI and SeSaMI.

In the second experiment, we perform pairwise registration of T1, T2 and PD volumes. The ground truth deformation field is similar to the previous experiment: the spacing between B-spline nodes is 20 mm and every node is moved randomly by numbers in the range of ±9 mm. We calculate a B-spline deformation field with 20 mm distance between nodes using different similarity metrics of MI, LMI, $\alpha$-MI and SeSaMI. The results are shown in Figure 11. In the top and bottom rows, the bias fields of $I_m$ are respectively 0% and 40%. MI works well in the top row where the intensity inhomogeneity is 0. However, it gives large errors when we set the intensity inhomogeneity to 40%. LMI works well for all the cases, as it significantly reduces the effect of intensity bias. $\alpha$-MI gives very good results for 0% bias, but its results slightly degrade with bias. One reason that the results of LMI and $\alpha$-MI in the second row, in average, are comparable is that the image intensities relate well in the BrainWeb simulation images, which can be seen from the joint histograms. Therefore LMI, which is solely based on image intensities, works well with these cases. SeSaMI gives the best results with or without bias in all the cases.

In the final experiment, we test the effect of the degrees of freedom of the deformation field on PD-T1 registration. The ground truth simulated deformation is the same as the two previous examples: the distance between nodes is 20 mm. The experiments are repeated for 20 random instances of deformation. The deformation that we optimize the cost function for is, however, different: we perform three sets of experiments where the distances between the nodes at the finest level are 10 mm, 20 mm and 40 mm. The results are shown in Figure 12. The numbers 0 and 40 after each method indicate the percentage of bias field in the moving image. The stars indicate that this column's results are significantly better than that of the previous column ($p < 0.05$ using pairwise t-test). Few observations can be made. First, LMI40 is significantly better than MI40 in (a) and (b). Also, $\alpha$-MI40 and LMI40 perform relatively similar. SeSaMI is significantly outperforming all other methods for biased or unbiased cases in (a) and (b). As we constraint the deformation by increasing the spacing between B-spline nodes to 40 in (c), all methods perform similar. Even the bias field does not significantly change the results. This shows that for finer non-rigid registration, the importance of the similarity metric rises. Another interesting observation is that the results of (b) are better than (a) because the ground truth deformation is also generated by B-spline nodes at 20 mm spacings. This shows the importance of choosing a deformation model that is most similar to the unknown underlying deformation.

### *5.3. Phantom Data*

We use US and MR volumes of an anthropomorphic polyvinyl alcohol brain phantom (Chen et al., 2012) to plot different cost functions. The US and MR volumes in this database are registered and there is no deformation between them. Three perpendicular cross sections of the volumes are shown in Figure 13. We move the MR volume in the horizontal and vertical directions by ±6 pixels (each pixel is 1 mm) and calculate different similarity metrics. For SeSaMI, the self-similarity is calculated from the MR image. Figure 13 shows the results. We




(a) $I_f$=T1, $I_m$=T2 (b) $I_f$=T2, $I_m$=PD (c) $I_f$=PD, $I_m$=T1

(d) $I_f$=T1, $I_m$=T2 (e) $I_f$=T2, $I_m$=PD (f) $I_f$=PD, $I_m$=T1

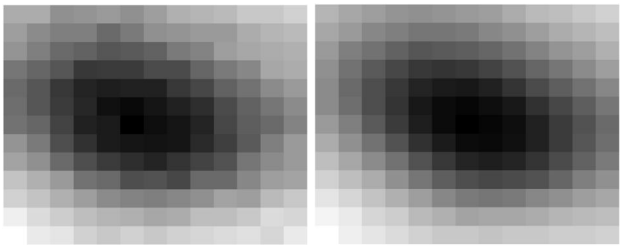
Figure 11: The warping index for different pairwise registrations. $I_f$ has always 0% intensity bias. $I_m$ images in the top and bottom rows have respectively 0% and 40% intensity inhomogeneities. The joint histograms, before deforming any of the images, is also shown for biased images in the second row.

use N=5000 voxels to calculate NMI. For LMI, the volumes are divided into 8 neighborhoods by having two segments in each dimension, and use N=5000 voxels in each neighborhood. We see that the LMI cost function is significantly better than NMI, giving small cost values around the zero displacement. $\alpha$-MI and SeSaMI both predict the alignment around the same location around 0. SeSaMI outperforms $\alpha$-MI in terms of generating a consistently smooth cost function. Please note that we perform this experiment to provide an intuition toward the effect of *different parameters* of the similarity metric, and not to quantitatively compare *different similarity metrics*. To see the improvement of $\alpha$-MI with using more features, we repeat the experiment by using the four features of intensity and gradient at two different scales. Figure 14 shows the results with $k = 10$ and $N = 4000$ using 4 features. It can be seen that the results of $\alpha$-MI significantly improves compared to Figure 13, which uses only 2 features.

### 5.4. US and MR registration of brain images

We apply our registration algorithm to the clinical data from image guided neuro-surgery obtained from 13 patients[2] with gliomas in the Montreal Neurological Institute. The data is available online at `http://www.bic.mni.mcgill.ca/BITE`.

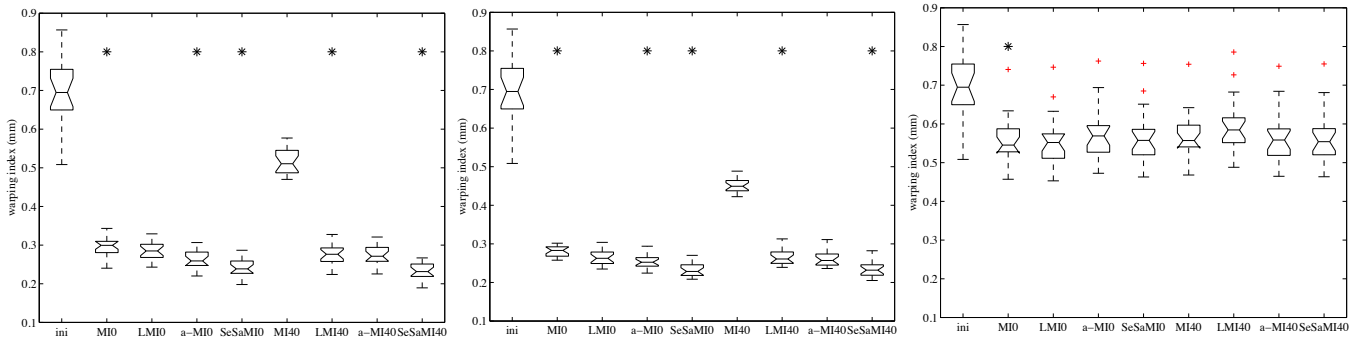
[2]We originally collected data from 14 patients and had the corresponding anatomical landmarks selected by three experts. In one of the datasets, however, some of the landmarks do not match, which we think is due to a problem in a transformation matrix. We therefore exclude this dataset from our analysis.



(a) $\alpha$-MI (b) SeSaMI

Figure 14: The values of $\alpha$-MI and SeSaMI for the experiment of Figure 13, using four features $z_1, \cdots z_4$ of intensities and gradients at two different scales. The $\alpha$-MI cost is significantly smoother compared to using two features as in Figure 13. Both functions are normalized to the 0 to 1 range. Black represents smaller dissimilarity.

The pre-operative MR images are gadolinium-enhanced T1 weighted and are acquired approximately 2 weeks before the surgery. The intra-operative US images are obtained using an HDI 5000 (Philips, Bothell, WA) with a P7-4 MHz phased array transducer. Full description of the data is provided in Mercier et al. (2012a). The ultrasound probe is tracked with a Polaris camera (NDI, Waterloo, Canada), and 3D US volumes are reconstructed using the tracking information. We reconstruct US volumes with a pixel size of 1 mm in the $x$ $y$ $z$ directions. Figure 15 shows the US images before and after 3D reconstruction. This relatively large pixel size means that for every pixel, multiple US measurements from different images and insonification




(a) Spacing 10 (b) Spacing 20 (c) Spacing 40

Figure 12: The warping index for different spacing between the B-spline nodes. The stars indicate statistically significant improvement over the previous column. Please see the text for details.

angles are available. Therefore, the effect of US speckle is minimized, since US speckles look different when imaged from different locations or angles (Rivaz et al., 2007, 2011, 2014). Each volume has a different size because the depths and sweeping areas are different; the typical size is on the order of $100^3$.

The MRI volume has a pixel size of 0.5 mm in the imaging plane and 1 mm slice thickness. We resample the volume to the isotropic 1 mm pixel size, same as the reconstructed US pixel size. We then use the tracking information to perform the initial rigid registration of MR to US (a sample of this initial registration is shown in Figure 1) and crop the MR volume to the same size as the US volume. The very top of the US image is the acoustic coupling gel that is placed to prevent pressure to the brain tissue. The very bottom is also very noisy due to acoustic wave attenuation. We crop these parts during an automatic preprocessing step for all registration methods. Therefore, the cropped MR volume is larger than the useful US volume and always covers it during the registration.

To validate the results, three experts have selected corresponding anatomical landmarks in US and MR images (Mercier et al., 2012a). For this task, US and MR images are sampled at 0.3 mm voxel size in all dimensions. The landmarks are selected independently by all the three experts throughout the volumes as shown in Figure 15. They are used to calculate mean target registration error (mTRE). The mTRE of $n$ corresponding marks at locations $\boldsymbol{x}$ and $\boldsymbol{x}'$ in the two images is calculated according to

$$mTRE = \frac{1}{n}\sum_{i}^{n} \|\boldsymbol{T}(\boldsymbol{x}_i) - \boldsymbol{x}'_i\| \tag{13}$$

where $\boldsymbol{T}$ is the transformation (see Eq. 1) and $\|\boldsymbol{v}\|$ is the Euclidian length of the vector $\boldsymbol{v}$.

**Parameter Selection and Qualitative Analysis.**

We perform hierarchical registration to minimize the number of false local minima and increase the capture range, with two spacing of 40 and 20 mm between the B-spline nodes. For LMI, we found that three hierarchical levels with spacings of 80, 40 and 20 mm gives the best results. With a data of size $100^3$ pixels, we have $9^3$ nodes at the fine 20-pixel level and therefore $\boldsymbol{\mu}$ has $3 \times 9^3 = 2187$ elements. We empirically found that $N =$ 4000 random points give good results for $\alpha$-MI and SeSaMI.

Out of the 13 patients, we use 6 (patients 1, 2, 3, 8, 12 and 13, which have the highest initial mTRE) to set different parameters of the NMI, LMI, $\alpha$-MI and SeSaMI. In the supplementary material, we perform four-fold cross validation for setting parameters, instead of using only 6 patients. These results show that for different training sets, same parameters are selected for SeSaMI, meaning that the same parameters should work well for new sets of experiments. We found that the performance of the global NMI is poor in the presence of the high level of bias in the US volumes (see Figures 13 and 18 for example). Therefore, we do not include it in our IGNS registration study. For LMI, 32 bins gave the best results. For LMI, we tested three different neighborhood sizes of $13^3$, $25^3$ and $51^3$ mm, and found that the size of $25^3$ mm gives the best results (for the size $13^3$ we selected all the $13^3 \approx 2200$ pixels in the neighborhood for MI computation). We also found that the number of neighborhoods $N = 50$ in Eq. 3 provides a good compromise between the running time and the performance. Figure 16 shows the mTRE values for different $k$ values of $\alpha$-MI and SeSaMI ranging from 1 to 20. The reduction of the mTRE at each $k$ compared to its previous value is statistically significant in both $\alpha$-MI and SeSaMI with $p$-values of less than 0.05. However, we choose $k = 10$ over $k = 20$ as a tradeoff between computational time and small mTRE.

We then vary the number of features from 1 to 6: the intensity and gradient values at three different scales of $\sigma$ =0.5, 1.5 and 3 pixels. Figure 17 shows that increasing the number of features from 1 to 4 decreases the mTRE by statistically significant values ($p < 0.05$). We also point out the difference in performance of $\alpha$-MI and SeSaMI in this figure: At 2 and 4 features, the improvement of SeSaMI over $\alpha$-MI has $p$-values of 0.001 and 0.04 respectively, a significantly higher improvement with fewer features. Similarly, the average reductions in the mTRE in these 6 patients are 0.8 mm with 2 features (mTRE of $\alpha$-MI and SeSaMI are respectively 3.7 mm and 2.9 mm) and 0.2 mm with 4 features (mTRE of $\alpha$-MI and SeSaMI are respectively 2.5 mm and 2.3 mm). This shows that SeSaMI significantly outperforms $\alpha$-MI when few features are used. Another interesting observation is that there is no statistically significant difference

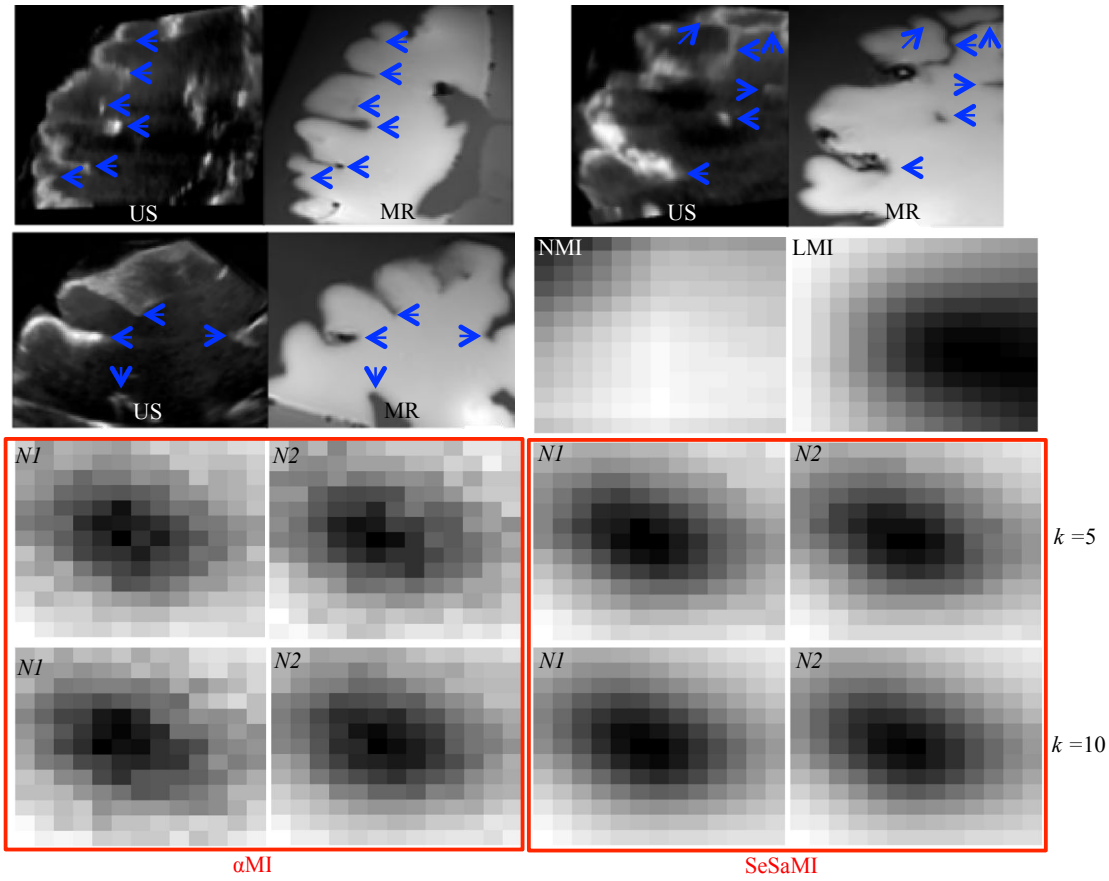



Figure 13: Qualitative comparison of different similarity metrics by rigidly translating the MR volume in the $x$ $y$ directions by ±6 pixels. The arrows in the mutually perpendicular US and MR slices point to corresponding structures. The minimum cost should be at the center. Black represents smaller dissimilarity. Two features $z_1$, $z_2$ are used for both $\alpha$-MI and SeSaMI. The number of nearest neighbors $k$ are 5 and 10 as marked, and $N1 = 2000$ and $N2 = 4000$.

between the results of $\alpha$-MI with 4 features and SeSaMI with 2 features. This is due to the encoding of the structural information as self-similarity weights in SeSaMI. Increasing the number of features from 4 to 6 has no statistically significant impact on mTRE in either method. We therefore use 4 features for the patient data. With these settings, the running times on one core of a 3GHz processor are: 1 hour for LMI and 2 hours for $\alpha$MI and SeSaMI (all implemented in Matlab mex functions).

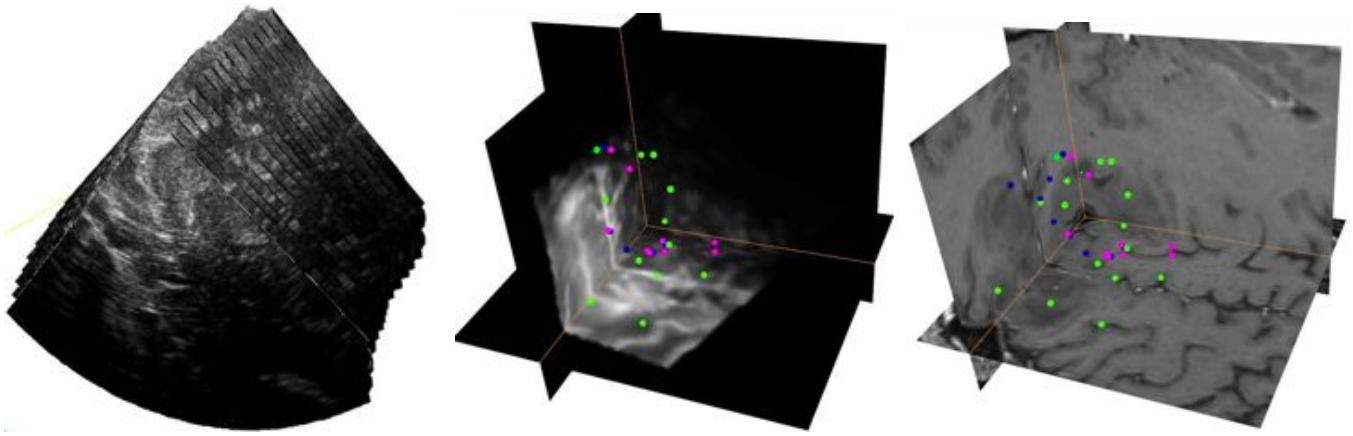
Figure 18 shows the US and MR images before and after nonrigid registration. The initial alignment does not provide enough guidance for avoiding blood vessels and critical brain tissue. For example, three misaligned anatomical and pathological structures of tumor boundary, choroid plexus and the lateral ventricles are marked. Visual inspection of corresponding structures shows that $\alpha$-MI and SeSaMI significantly improve the alignment.

**Quantitative Results.** Table 1 shows that multi-feature $\alpha$-MI and SeSaMI significantly reduce the mTRE by nonlinear registration of MR to US in all the 13 cases. The $\alpha$-MI and SeSaMI mTRE results are significantly smaller than the initial mTRE values with $p$-values of respectively 0.007 and 0.005. Also, LMI gives good results in all patients except for P2; it improves the initial mTRE by a $p$-value of 0.02 if we exclude the outlier case of P2. The improvement of SeSaMI results over $\alpha$-MI is also statistically significant with a $p$-value of 0.009. The most accurate results generated by SeSaMI are due to its robust self-similarity measure incorporated into the powerful multi-feature $\alpha$-MI similarity metric. In Ou et al. (2012), different registration methods are compared by studying the Jacobian of their deformations. We tested the standard deviation of the deformation Jacobians in different patients and found them to be similar (except for patient P2 where LMI gives very large mTRE). This is due to the fact that we use the same regularization and the same deformation model for all methods.

## 6. Discussion and Conclusions

We introduced SeSaMI, a similarity metric that incorporates contextual self-similarity measures into graph-based $\alpha$-MI. SeSaMI exploits self-similarity in a $k$NN $\alpha$-MI registration framework by penalizing clusters (i.e. the nearest neighbors) that are not self-similar. The self-similarity measure that we use is rotation invariant and is robust to small deformations and to intensity bias.

We have also, for the first time, shown that multi-feature $\alpha$-MI and SeSaMI significantly increase the registration accuracy



(a) Individual US images (b) Reconstructed 3D US (c) MRI

Figure 15: The original 2D US slices and the reconstructed 3D US MR images. In (b) and (c), the landmarks selected by 3 experts are shown. Each of the 3 colors (blue, green and red) corresponds to an expert. The images are from patient 1 in Table 1.

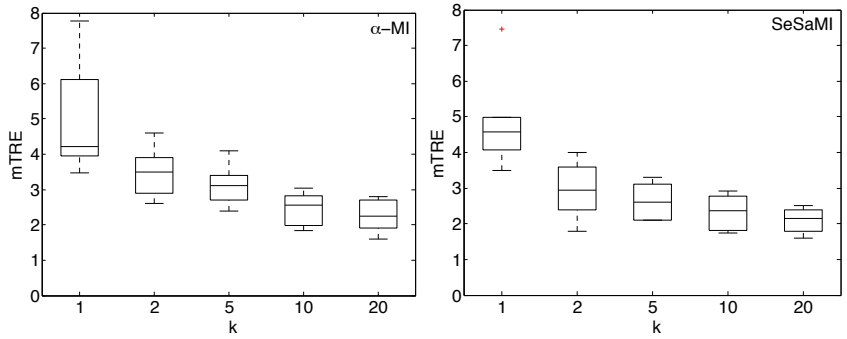



Figure 16: The effect of the number of nearest neighbors $k$.

of MR to US registration in our on-going IGNS project. The large initial misalignment between US and MR images are due to errors in the tracking system, US calibration and brain shift. The misalignment is large enough to make the rigidly registered MR image unreliable.

The entropy images of Wachinger and Navab (2012) also generate rotationally invariant features using local histograms. Similar to our work, these features are obtained in two steps: first by constructing a histogram, and second, by inferring the features from the histogram. Our approach is different in both steps. Firstly, the entropy images only contain intensities[3], while our 2D histogram descriptor contains both intensity and spatial information. Secondly and more importantly, the entropy images directly map the histograms into scalar entropy values, where different histograms can generate the same entropy. Our approach however utilizes all the histogram bins and performs pair-wise comparisons using EMD to map differences between patches into a scalar value. Encoding the rich structural information into the pair-wise distances are performed offline in SeSaMI. Since SeSaMI is graph-based, these pair-wise differences readily relate to the edge weights.

SeSaMI can reduce the number of incorrect local minima as Figures 9 and 13 appear to support. An intuitive explanation for this was given in Figure 8: SeSaMI may penalize close neighbors in the joint feature space that are not self-similar. Therefore, accidental clusters in the joint feature space are less likely to generate incorrect local minima.

The results of Figure 17 revealed that two is the minimum number of features required; with one feature even the SeSaMI results degrade considerably. This is due to the fact that the $\Gamma^f$ and $\Gamma^m$ terms in the denominator of Eq. 8 can become close to zero, or even exactly zero depending on the precision of the data, when the nearest neighbors are found in just one dimension. In fact, Kybic (2007) proposes a number of remedies for "too close" nearest neighbors. In our experience, we found that two or more features were enough to prevent having $k$ very close nearest neighbors.

We trained the parameters of $\alpha$-MI and SeSaMI on 6 patients to achieve the best results for either method, and presented these best results in Table 1. Our goal was to show that

[3]Wachinger and Navab (2012) proposed a weighting map to differentiate different patches that have identical histograms. However, in order to completely discriminate different patches, the weight map should contain numbers in a very large dynamic range. This leads to locations that become negligible in the entropy calculation as discussed by Wachinger and Navab (2012).

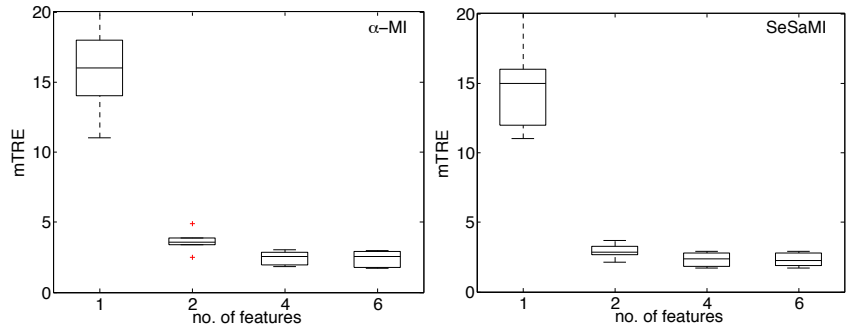



Figure 17: The effect of the number of features.

Table 1: The mTRE (mm) in initial alignment (obtained by rigidly registering US and MR using tracking information) and after registration using 3 different nonlinear registration methods. $n$ is the number of landmarks in Eq. 13. The numbers inside brackets are minimum and maximum TRE values. Note: the highest initial mTRE patients (1, 2, 3, 8, 12 & 13) were used to determine the registration parameters. See the supplementary material for a cross-validation analysis of the results. The minimum value of each row is in bold font.

| Patient | $n$ | Initial | LMI | $\alpha$-MI | SeSaMI |
|---|---|---|---|---|---|
| P1 | 35 | 6.30 (1.9-9.1) | 4.05 (2.0-9.1) | 2.32 (1.7-6.9) | **1.82 (1.5-6.0)** |
| P2 | 40 | 9.38 (6.3-14.6) | 3.74 (1.5-10.1) | 3.14 (1.4-7.2) | **2.54 (1.4-5.5)** |
| P3 | 32 | 3.93 (1.0-6.1) | 2.49 (0.8-5.8) | **1.83(0.7**-5.3) | 1.96 (0.8-**4.8**) |
| P4 | 31 | 2.62 (0.5-6.9) | 2.72 (0.6-5.3) | 2.62 (**0.6**-3.7) | **2.59** (0.7-**3.4**) |
| P5 | 37 | 2.30 (0.2-4.4) | 2.23 (**0.3**-5.0) | 1.97 (0.6-4.3) | **1.73** (0.5-**3.8**) |
| P6 | 19 | 3.04 (0.3-6.3) | 2.51 (0.9-5.2) | 2.28 (1.0-5.0) | **1.94 (0.8-4.4)** |
| P7 | 23 | 3.75 (0.0-8.5) | **2.56 (0.8**-6.2) | 3.05 (1.1-6.6) | 2.91 (1.0-**5.1**) |
| P8 | 21 | 5.09 (2.5-7.6) | 2.49 (**1.6**-7.0) | **2.44** (1.6-**5.4**) | 2.52 (1.7-5.4) |
| P9 | 25 | 3.00 (0.8-5.3) | 2.84 (1.1-5.3) | 2.83 (1.2-**5.2**) | **2.74 (1.0**-5.4) |
| P10 | 25 | 1.52 (0.6-3.5) | 3.94 (0.4-5.1) | 1.44 (**0.2**-4.4) | **1.35** (0.7-**3.4**) |
| P11 | 21 | 3.70 (0.9-7.0) | **2.29 (0.5-3.2**) | 2.81 (1.4-6.3)) | 2.78 (1.1-5.5) |
| P12 | 23 | 5.15 (1.5-10.4) | **2.67 (0.8**-6.7) | 3.37 (2.0-5.9) | 2.91 (1.2-**5.5**) |
| P13 | 23 | 3.78 (1.2-5.7) | 2.90 (2.0-6.0) | 2.45 (1.8-5.5) | **2.16 (1.6-4.9)** |
| mean | 27 | 4.12 (1.4-7.3) | 2.88 (1.1-6.2) | 2.50 (1.2-5.5) | **2.29 (1.1-4.8)** |
| std | 6.9 | 2.03 (2.7-2.9) | 0.62 (0.6-1.8) | 0.55 (0.5-1.0) | **0.52 (0.4-0.9)** |

multi-feature $\alpha$-MI can be applied to US to MR registration. While this table shows statistically significant improvement of SeSaMI over $\alpha$-MI, greater improvements can be seen when using non-optimized parameters. For example, Figure 13 shows that with 2 features and fewer samples, SeSaMI generates significantly smoother cost functions. Comparing these results to Figure 14, we see that the improvement becomes less obvious when using 4 features and more samples. A similar trend can be seen in Figure 17 by comparing the results of $\alpha$-MI and SeSaMI at 2 and 4 features: the average reductions in mTRE are 0.8 mm and 0.2 mm with respectively 2 and 4 features.

Graph-based $\alpha$-MI estimators do not require the joint probability distribution unlike the plug-in MI estimators; they rather estimate $\alpha$-MI directly from the data. Therefore they have a faster convergence rate especially in higher dimensional data (Hero et al., 2002). However, even the graph based methods can suffer from the so called "curse of dimensionality". To see this, consider $N$ data points uniformly distributed in a $2d_f$ dimensional unit ball. In the image registration paradigm, $d_f$ is the number of features used from each of the fixed and moving images. It can be easily shown that the median distance from the origin to the first nearest neighbor is $(1-0.5^{1/N})^{1/2d_f}$ (Hastie and Tibshirani, 2009). For $d_f = 15$, even for a large set of $N = 10,000$ sample points we have $(1-0.5^{1e-4})^{1/30} = 0.73$, about 3/4 away from the center and in the 1/4 boundary layer. Hence, in high dimensions most data points are closer to the boundary of the sample space where estimations are less accurate.

In this work, we used self-similarity to weight $\alpha$-MI. These weights can also be incorporated into the quantitative-qualitative MI approach of Luan et al. (2008) to weight the MI formulation, which can be an area for future work.

For LMI results of Table 1, we used 3 hierarchical levels, while for $\alpha$-MI and SeSaMI we used 2 levels. The coarsest optimization level computes a better initial alignment for the second level, and therefore the results improve. In the supplementary material, we also provide the LMI results for 2 hierarchical levels, a setting that is consistent with $\alpha$-MI and SeSaMI.

In the future, we will compare the histogram and EMD based self-similarity estimation against NCC and SSD that are not rotationally invariant. The use of the SeSaMI with the MST entropy estimator is also a subject of future work. Since calcu-

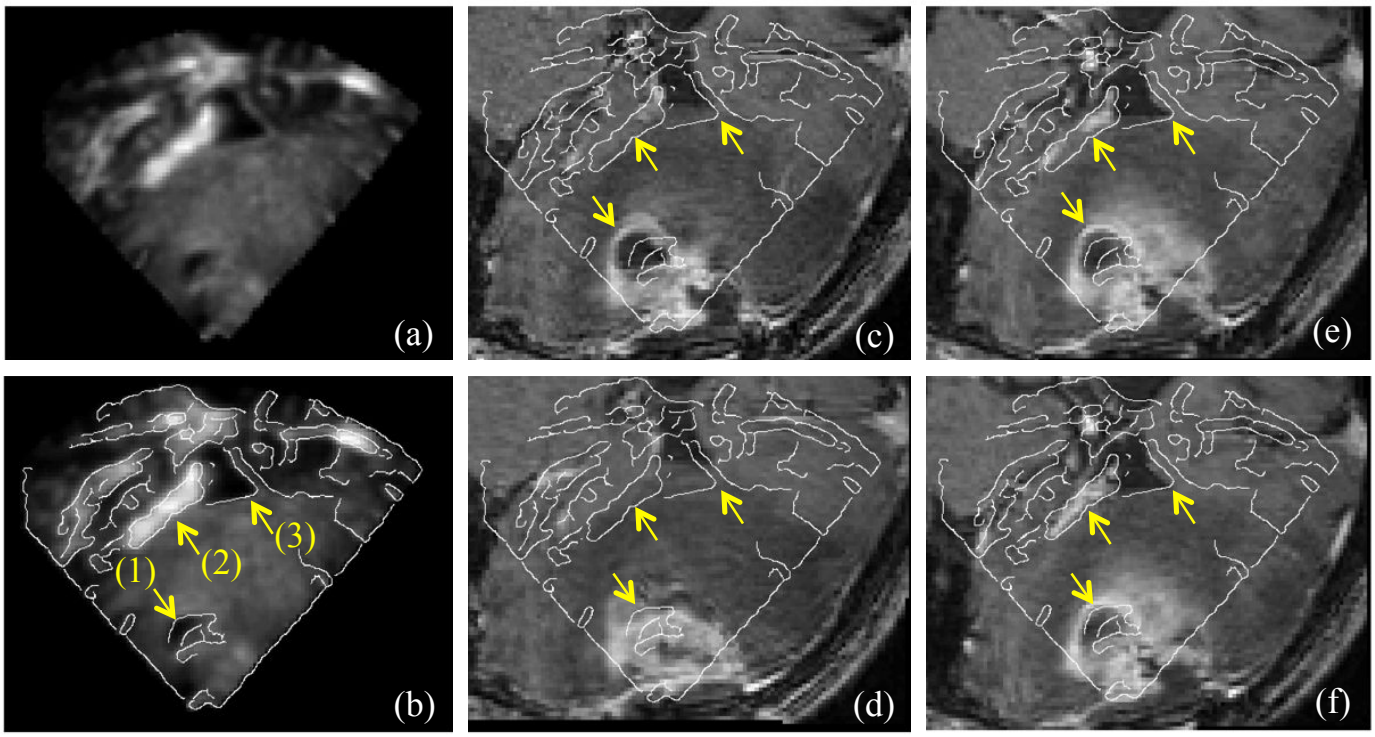



Figure 18: Corresponding intra-operative US and pre-operative MR images. The structures (1) to (3) are respectively the tumor boundary, choroid plexus and the lateral ventricle. (a) and (b) are the same US image, without and with contours overlaid. (c) to (f) are the corresponding MR images, with the contour overlaid to aid comparison. (c) initial alignment. (d) to (f): nonrigid registrations with respectively NMI, $\alpha$-MI and SeSaMI. Note that all registrations are done in 3D.

lating the derivative of the MST graph is computationally more efficient compared to the $k$NN graph, this can lead to shorter running times. A problem here is that as the graph topology changes, the gradient of the cost function can undergo very large discontinuities. We also plan to generate elasticity images of the tumor using ultrasound elastography (Rivaz et al., 2008) to allow better resection and to help registration. Finally, we will investigate GPU implementations of SeSaMI to achieve our goal of near-real time intra-operative US-MRI registration.

## Acknowledgment

We used the Intraoperative Brain Imaging System (IBIS) program, developed by S. Drouin and A. Kochanowska for generating the images of Figure 15. We acknowledge their technical support. The authors also would like to thank Dr. Fonov for valuable discussions, and Mr. De Nigris, Dr. Oreshkin, Dr. Arbel, Dr. Mercier, Ms. Bailey and Mr. Chen for their input. H. Rivaz is supported by the NSERC PDF.

## References

Andronache, A., Siebenthal, M. V., Szekely, G., Cattin, P., 2008. Non-rigid registration of multi-modal images using both mutual information and cross-correlation. Medical Imag Analysis 12 (1), 3–15.

Arbel, T., Morandi, X., Comeau, R., Collins, D. L., 2004. Automatic non-linear mri-ultrasound registration for the correction of intra-operative brain deformations. Compt Aided Surg 9 (4), 123–136.

Arya, S., Mount, D., Netanyahu, N., Silverman, R., Wu, A., December 1998. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. J. of ACM 45 (6), 891–923.

Beirlant, J., Dudewicz, E., Gyorfi, L., van der Meulen, E., 1997. Nonparametric entropy estimation: An overview. Int. J. Math. Stat. Sci. 6 (1), 17–39.

Buades, A., Coll, B., Morel, J., 2005. A non-local algorithm for image denoising. Computer Vision Pattern Recognition (CVPR), 149–157.

Buades, A., Coll, B., Morel, J., 2010. Image denoising methods. a new nonlocal principle. SIAM Review 52 (1), 113–147.

Chen, S. J.-S., Hellier, P., Marchal, M., Gauvrit, J.-Y., Carpentier, R., Morandi, X., Collins, D. L., 2012. An anthropomorphic polyvinyl alcohol brain phantom based on colin27 for use in multimodal imaging. Medical Physics 39, 554.

Collins, D. L., Zijdenbos, A., Baare, A., Evans, A., 1999. ANIMAL+INSECT: improved cortical structure segmentation. Inform Proc Med Imag 1613, 210–223.

Collins, D. L., Zijdenbos, A., Kollokian, V., Sled, J., Kabani, N., Holmes, C., Evans, A., 1998. Design and construction of a realistic digital brain phantom. IEEE Trans. Medical Imag 17 (3), 463–468.

Coupe, P., Eskildsen, S. F., J. V. Manjon, V. F., Collins, D. L., 2011. Simultaneous segmentation and grading of hippocampus for patient classification with alzheimer's disease. Medical Image Computing Computer Assisted Intervention (MICCAI), 149–157.

Coupe, P., Hellier, P., Kervrann, C., Barillot, C., 2009. Nonlocal means-based speckle filtering for ultrasound images. IEEE Trans Imag. Proc. 18 (10), 2221–2229.

Coupe, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C., 2008. An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images. IEEE Trans Med Imag 27 (4), 425–441.

Darkner, S., Sporring, J., in press. Locally orderless registration. IEEE Trans Pattern Anal Machine Int.

De Nigris, D., Collins, D. L., Arbel, T., 2012. Multi-modal image registration based on gradient orientations of minimal uncertainty. IEEE Trans. Medical Imag 31 (12), 2343–2354.

Grewenig, S., Zimmer, S., Weickert, J., 2011. Rotationally invariant similarity

measures for nonlocal image denoising. J. Vis. Commun. Imag. R. (JVCI) 22, 117–130.

Hastie, T., Tibshirani, R., 2009. The elements of statistical learning: Data mining, inference, and prediction, second edition. Springer Series in Statistics.

Heinrich, M., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F., Brady, M., Schanbel, J., 2011. Non-local shape descriptor: a new similarity metric for deformable multi-modal registration. Medical Image Computing Computer Assisted Intervention (MICCAI), 541–548.

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

Heinrich, M. P., Jenkinson, M., Papież, B. W., Brady, M., Schnabel, J. A., 2013. Towards realtime multimodal fusion for image-guided interventions using self-similarities. In: MICCAI. pp. 187–194.

Hero, A., Ma, B., Michel, O., Gorman, J., 2002. Applications of entropic spanning graphs. IEEE Signal Proc Mag, 85–95.

Ji, S., Wu, Z., Hartov, A., Roberts, D., Paulsen, K., 2008. Mutual-information-based image to patient re-registration using intraoperative ultrasound in image-guided neurosurgery. Med. Phys. 35 (10), 4612–4624.

Klein, S., Pluim, J., Staring, M., Viergever, M., 2009. Adaptive stochastic gradient descent optimisation for image registration. Int J Comput Vis 81, 227–239.

Klein, S., Staring, M., Pluim, J., December 2007. Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines. IEEE Trans. Imag Proc 16 (12), 2879–2890.

Klein, S., van der Heide, U., Lips, I., van Vulpen, M., Staring, M., Pluim, J., 2008. Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. Med. Phys. 35 (4), 1407–1417.

Kuklisova-Murgasova, M., Cifor, A., Napolitano, N., Papageorghiou, A., Quaghebeur, G., Noble, A., Schanbel, J., 2012. Registration of 3d fetal brain us and mri. Medical Image Computing Computer Assisted Intervention (MICCAI), 667–674.

Kybic, J., 2007. High-dimensional entropy estimation for finite accuracy data: R-NN entropy estimator. Information Processing Medical Imag. (IPMI), 569–580.

Kybic, J., Vnucko, I., 2012. Approximate all nearest neighbor search for high dimensional entropy estimation for image registration. Signal Processing 92 (5), 1302–1316.

Lazebnik, S., Schmid, C., Ponce, J., 2005. A sparse texture representation using local affine regions. IEEE Trans Pattern Anal Machine Int 27 (8), 1265–1278.

Ling, H., Okada, K., 2007. An efficient earth mover's distance algorithm for robust histogram comparison. IEEE Trans Pattern Anal Machine Int. 29 (5), 840–853.

Loeckx, D., Slagmolen, P., Maes, F., Vandermeulen, D., Suetens, P., 2010. Nonrigid image registration using conditional mutual information. IEEE Trans. Medical Imag 29 (1), 19–29.

Luan, H., Qi, F., Xue, Z., Chen, L., Shen, D., 2008. Multimodality image registration by maximization of quantitative-qualitative measure of mutual information. Pattern Recognition 41, 285–298.

Maes, F., Collignon, A., Vandermeulen, D., G, M., Suetens, P., 1997. Multimodality image registration by maximization of mutual information. IEEE Trans. Medical Imag 16 (2), 187–198.

Manjon, J., Coupe, P., Buades, A., Collins, L., Robles, M., 2012. New methods for mri denoising based on sparseness and self-similarity. Medical Image Analysis 16 (1), 18–27.

Mercier, L., Del Maestro, R. F., Petrecca, K., Araujo, D., Haegelen, C., Collins, D. L., 2012a. Online database of clinical mr and ultrasound images of brain tumors. Medical Physics 39, 3253.

Mercier, L., Fonov, V., Haegelen, C., Maesstro, R., Petrecca, K., Collins, D. L., 2012b. Comparing two approaches to rigid registration of three-dimensional ultrasound and magnetic resonance images for neurosurgery. Compt Aided Surg 7 (1), 125–136.

Neemuchwala, H., Hero, A., 2005. Entropic graphs for registration. In R. Blum and Z. Liu, editors, Multi-sensor image fusion and its applications, CRC Press.

Ou, Y., Sotiras, A., Paragios, N., Davatzikos, C., August 2011. DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting. Medical Imag. Anal. 15 (4), 622–639.

Ou, Y., Ye, D., Pohl, K., Davatzikos, C., 2012. Validation of dramms among 12 popular methods in cross-subject cardiac mri registration. Biomedical Image Registration, LNCS 7359, 209–219.

Oubel, E., Craene, M., Hero, A., Frangi, A., 2012. Cardiac motion estimation by joint alignment of tagged mri sequences. Med. Imag. Anal. 16, 339–350.

Penney, G., Blackall, J., Hamady, M., Sabharwal, T., Adam, A., Hawks, D., 2004. Registration of freehand 3d ultrasound and magnetic resonance liver images. Med Imag Anal 8 (1), 81–91.

Pluim, J., Maintz, J., Viergever, M., 2000. Image registration by maximization of combined mutual information and gradient information. IEEE Trans. Medical Imag. 19 (8), 809–814.

Pluim, J., Maintz, J., Viergever, M., 2003. Mutual-information-based registration of medical images: a survey. IEEE Trans. Medical Imag. 22 (8), 986–1004.

Rivaz, H., Boctor, E., Foroughi, P., Zellars, R., Fichtinger, G., Hager, G., 2008. Ultrasound elastography: a dynamic programming approach. Medical Imaging, IEEE Transactions on 27 (10), 1373–1377.

Rivaz, H., Boctor, E. M., Choti, M. A., Hager, G. D., Apr 2011. Real-time regularized ultrasound elastography. IEEE Trans Med Imaging 30 (4), 928–945.

Rivaz, H., Boctor, E. M., Choti, M. A., Hager, G. D., 2014. Ultrasound elastography using multiple images. Medical Image Analysis (0), –.
URL http://www.sciencedirect.com/science/article/pii/S1361841513001709

Rivaz, H., Collins, D. L., 2012. Self-similarity weighted mutual information: A new nonrigid image registration metric. Medical Image Computing Computer Assisted Intervention (MICCAI), 91–98.

Rivaz, H., Zellars, R., Hager, G., Fichtinger, G., Boctor, E., Oct. 2007. Beam steering approach for speckle characterization and out-of-plane motion estimation in real tissue. IEEE Int. Ultrasonics Symp., 781–784.

Roche, A., Pennec, X., Malandain, G., Ayache, N., 2001. Rigid registration of 3-d ultrasound with mr images: a new approach combining intensity and gradient information. IEEE Trans Med Imag 20 (72), 291–237.

Rubner, Y., Tomasi, C., Guibas, L., 2000. The earth mover's distance as a metric for image retrieval. IEEE Trans Pattern Anal Machine Int. 40 (2), 99–121.

Sabuncu, M., Ramadge, P., 2008. Using spanning graphs for efficient image registration. IEEE Trans. Imag. Proc. 17 (5), 788–797.

Shechtman, E., Irani, M., 2007. Matching local self-similarities across images and videos. Computer Vision and Pattern Recognition (CVPR), 1–8.

Shen, D., Davatzikos, C., November 2002. HAMMER: hierarchical attribute matching mechanism for elastic registration. IEEE Trans. Med. Imag. 21 (11), 1421–1439.

Spall, J. C., 1998. Implementation of the simultaneous perturbation algorithm for stochastic optimization. Aerospace and Electronic Systems, IEEE Transactions on 34 (3), 817–823.

Staring, M., Heide, U., Klein, S., Pluim, J., 2009. Registration of cervical mri using multifeature mutual information. IEEE Trans. Medical Imag. 28 (9), 1412–1421.

Studholme, C., Drapaca, C., Iordanova, B., Cardenas, V., 2006. Deformation-based mapping of volume change from serial brain mri in the presence of local tissue contrast change. IEEE Trans. Medical Imag 25 (5), 626–639.

Studholme, C., Hill, D., Hawkes, D., 1999. An overlap invariant entropy measure of 3d medical image alignment. Pattern Recognition 32, 71–86.

Wachinger, C., Navab, N., 2012. Entropy and laplacian images: Structural representations for multi-modal registration. Medical Imag Analysis 16 (1), 1–17.

Wein, W., Ladikos, A., Fuerst, B., Shah, A., Sharma, K., Navab, N., 2013. Global registration of ultrasound to mri using the LC2 metric for enabling neurosurgical guidance. In: MICCAI. pp. 34–41.

Wells, W., Viola, P., Atsumid, H., Nakajimae, S., Kikinis, R., 1996. Multimodal volume registration maximization of mutual information. Med Imag Anal 1 (1), 35–51.

Xue, Z., Shen, D., Davatzikos, C., December 2004. Determining correspondence in 3-D MR brain images using attribute vectors as morphological signatures of voxels. IEEE Trans. Med. Imag. 23 (10), 1276–1291.

Zhang, W., Brady, M., Becher, H., Noble, A., 2011. Spatio-temporal (2d+t) non-rigid registration of real-time 3d echocardiography and cardiovascular mr image sequences. Physics Med Biol 56, 1341–1360.

Zhuang, S., Arridge, D., Hawkes, D., Ourselin, S., 2011. A nonrigid registration framework using spatially encoded mutual information and free-form deformations. IEEE Trans. Medical Imag 30 (10), 1819–1828.