Nonrigid Registration of Ultrasound and MRI Using Contextual Conditioned Mutual Information

Hassan Rivaz, Zahra Karimaghaloo, Vladimir S. Fonov, and D. Louis Collins

***Abstract*—Mutual Information (MI) quantifies the information that is shared between two random variables and has been widely used as a similarity metric for multi-modal and uni-modal image registration. A drawback of MI is that it only takes into account the intensity values of corresponding *pixels* and not of *neighborhoods*. Therefore, it treats images as "bag of words" and the contextual information is lost. In this work, we present Contextual Conditioned Mutual Information (CoCoMI), which conditions MI estimation on similar structures. Our rationale is that it is more likely for similar structures to undergo similar intensity transformations. The contextual analysis is performed on one of the images offline. Therefore, CoCoMI does not significantly change the registration time. We use CoCoMI as the similarity measure in a regularized cost function with a B-spline deformation field and efficiently optimize the cost function using a stochastic gradient descent method. We show that compared to the state of the art local MI based similarity metrics, CoCoMI does not distort images to enforce erroneous identical intensity transformations for different image structures. We further present the results on non-rigid registration of ultrasound (US) and magnetic resonance (MR) patient data from image-guided neurosurgery trials performed in our institute and publicly available in the BITE dataset. We show that CoCoMI performs significantly better than the state of the art similarity metrics in US to MR registration. It reduces the average mTRE over 13 patients from 4.12 mm to 2.35 mm, and the maximum mTRE from 9.38 mm to 3.22 mm.**

***Index Terms*—Registration, Surgical guidance, Mutual information, Ultrasound, Magnetic resonance**

I. Introduction

Image registration involves transforming different sets of images of the same (possibly deformed) object that are acquired from different coordinate systems into one coordinate system. It has numerous medical applications in diagnosis and in image guided surgery/therapy. Mutual Information (MI) of two random variables quantifies the degree by which one variable can be predicted by knowing the other. When two images are aligned, their MI is usually high [1], [2], [3], and therefore image registration can be performed by maximization of MI. At the heart of the MI, a joint histogram of intensity values of the two images is estimated, from which, a measure of dependence between the intensities is calculated. A major problem of MI is that it does not take contextual information into account and only enforces a compact joint histogram. Therefore, deformable image registration, which has many unknown parameters, is challenging with standard MI. We list three lines of previous work that tackle this problem. Note that these lines have many theoretical and practical similarities and we are not suggesting that they are categorically different.

The first approach to bring local contextual information into consideration is through local features: incorporating gradients and their orientations into MI [4], [5], [6], estimating local entropies and registering the resulting entropy images instead of the intensity images [7] and assigning patch similarities to voxels [8] are three examples. We will further elaborate the latter two techniques in the Background Section. Recently, Wachinger and Navab [9] proposed a probabilistic framework that can model a variety of feature based similarity metrics. An entropic measure related to MI is α-MI, which can be estimated from graphs such as the minimum spanning tree or the k-nearest neighbors instead of the joint histogram [10], [11], [12]. Since the joint histogram need not to be computed, it has been suggested that these techniques are less affected by the "curse of dimensionality". Therefore, multiple local features such as intensities, gradients and higher order derivatives are commonly used with this metric. In a recent work [13], we proposed self-similarity weighted α-MI (SeSaMI) where we incorporated patch similarities into α-MI as graph weights. We showed that SeSaMI is an effective method of incorporating contextual information into MI using experiments on simulation and patient data.

The second approach is to generate joint intensity distribution for higher order image structures, such as patches. The issue here is the "curse of dimensionality": the joint histogram for small 3×3 patches is 18D; too many data samples are required to populate an 18D space. Rueckert *et al.* [14] therefore propose an extension of MI into higher dimensions by considering only one neighbour of every pixel. Russakoff *et al.* [15] calculate the joint entropy for 5×5 patches, but assume normal distributions and decouple the joint entropy estimations into 25 independent distributions, instead of estimating a 50D joint histogram. Yi and Soatto [16] propose performing dimensionality reduction on image patches so that low dimensional labels can be assigned to patches. MI can then be performed on the low dimensional labels.

The third approach is to condition MI on spatial location. Studholme *et al.* [17] and Loeckx *et al.* [18] propose respectively regional MI and conditional MI 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.* [19] propose localized MI (LMI) where samples are

H. Rivaz, V. S. Fonov and D. L. Collins are with the McConnell Brain Imaging Center, Montreal Neurological Institute, McGill University, Montreal, QC, H3A 2B4, Canada. Email: hassan.rivaz@mcgill.ca

Z. Karimaghaloo is with the Center for Intelligent Machines, McGill University, Montreal, QC, H3A 2B4, Canada

Manuscript received August 13, 2013; revised October 30, 2013.

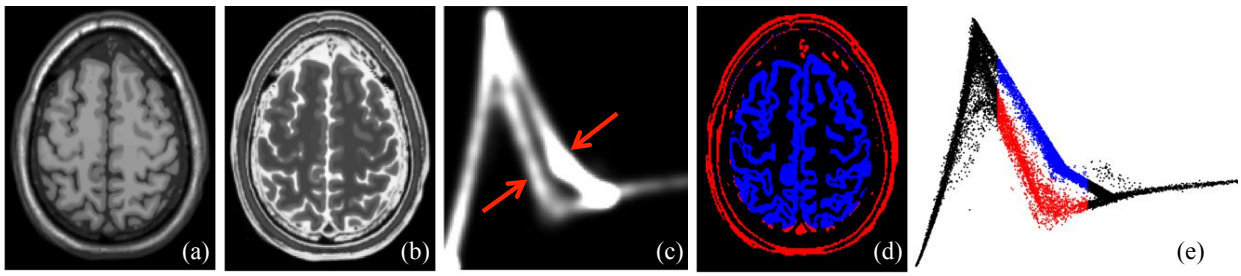


Fig. 1. Different parts of the images can have different intensity relations in multi-modal images. (a) and (b) are aligned T1 and T2 images, and (c) is their joint histogram. (d) and (e) show how the brain anatomy relates to the joint histogram.

randomly selected from local neighborhoods in every iteration and convergence is achieved by using stochastic optimization [20]. Zhuang *et al.* [21] propose 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 significantly outperform global MI in the presence of spatially varying intensity bias.

In this work, we use contextual information to condition MI. Our rationale is that it is more likely for similar structures in one image to follow the same intensity transformations in the second image. For example, an edge pixel and a pixel from a homogeneous region, which both have the same intensity, can have very different intensities in the second image. Figure 1 shows two corresponding slices of T1 and T2 images from BrainWeb [22]. Although the images are aligned, their joint histogram has two relatively close segments, pointed to by red arrows. In (d) and (e), we are showing the anatomical parts that contribute to the two segments. These segments decrease the level of confidence with which the intensity of a pixel in one image can be predicted knowing its intensity in the other image, therefore decreasing MI. We will see that in non-rigid registration, MI will generate erroneous deformations to force these two segments into one. However, if we condition MI on similar contextual regions, these segments are treated separately and are not forced to collapse into one segment. This situation is most common in US-MR registration. Here, a popular line of previous work [23], [24], [25] simulates US images from the MR data by assigning different intensities to different tissue structures. These techniques requires the two steps of first registering the MR volume to a segmented MR atlas, which provides with an MR segmentation, and second, simulating a pseudo-US from the segmented MR using look-up tables. The first step is challenging in clinical images due to the potential variety of possible pathologies that the brain tissue might have, such as different grade gliomas, cysts and cavernomas. The second step is also complex for two main reasons: first, the appearance of these pathologies in US is highly variable [24], [26], and second, changing the time gain compensation (TGC) settings of the US affect the look-up tables. The TGC settings are routinely varied in practice to enhance and reduce the intensity of different parts of the US image. Our proposed algorithm is inherently similar to this line of work, but does not require any of the two challenging steps.

In many surgical planning/guidance and radiation therapy procedures where image registration is required, one of the images (the pre-operative/planning image) is available in advance. We exploit this by performing contextual analysis, which is computationally expensive, on this image only. Our contextual analysis consists of calculating patch similarities in a neighborhood and recording the location of most similar patches. Buades *et al.* [27] first proposed exploiting repetitive regions (or patches) in the form of non local means (NLM) for image denoising. NLM and a similar concept, called self-similarity, have recently been used in image segmentation [28], object detection and image retrieval [29] and image registration [7], [8].

In our previous work [13], we incorporated self-similarities into a graph-based multi-feature similarity metric. The method was however computationally expensive and requires complex implementations for optimization of the similarity metric. In this work, we propose a significantly simpler and faster method to exploit the contextual information of the patches, which we call Contextual Conditioned Mutual Information (CoCoMI). We apply CoCoMI to register pre-operative magnetic resonance (MR) images to intra-operative ultrasound (US) images in the context of image-guided neurosurgery (IGNS).

In IGNS, an optical or electromagnetic device tracks the US probe, which allows registering US images to the pre-operative MRI. Unfortunately, the brain deforms significantly after the craniotomy for biochemical and physical reasons, a phenomenon known as *brain shift*. The errors in tracking and US calibration add to the registration inaccuracies. As a result, current surgical neuronavigation systems can have errors of more than 20 mm, which is not accurate enough to properly locate blood vessels or critical brain structures with the neuronavigation system during the surgical resection of the tumor. Therefore, an accurate registration technique that performs deformable registration of the US and MR images is highly desired in neurosurgery.

Previous work that registers US and MR include the following. Roche *et al.* [30] used the correlation ratio (CR) between US and MR and MR gradient. Arbel *et al.* [23] and Mercier *et al.* [24] registered the MR volume to a probabilistic atlas, segmented the MR using the atlas and assigned different intensity transformations to different MR regions to generate a pseudo-US. They then registered the pseudo-US to US using the ANIMAL uni-modal registration of Collins *et al.* [31]. Similarly, Kuklisova-Murgasova *et al.* [25] generated a US-

like volume from the MR volume, and then registered it with the US volume using a uni-modal block-matching technique. Penney *et al.* [32] generated blood vessels probability maps from from US and MR and registered these maps using cross correlation. Ji *et al.* [33] used normalized MI of US and MR. Zhang *et al.* [34] used MI of local phase information. De Nigris *et al.* [6] used MI of gradient orientations of US and MR as a similarity metric. Recently, Heinrich *et al.* [35] used self-similarity context, and Wein *et al.* [36] used correlation of US intensities and the intensity and gradient of MR. CoCoMI is similar to the previous work that generates pseudo-US from MR in that it treats different structures separately by examining their intensity and gradient content. The advantages are that CoCoMI does not require registering MR to an atlas or simulating pseudo-US from MR, both challenging in clinical applications.

Figure 2 shows an example of the registered US and MR images. Although the volumes are aligned, the joint histogram is very spread. This is mainly due to the high level of spatial bias in the US volume, which is caused by multitude of reasons. First, even though US machines compensate for signal attenuation as it penetrates the tissue, this compensation is only nominal and does not account for different attenuation in different tissue types. Second, wave transmission at tissue boundaries is not compensated, which causes the shadowing artifact behind a strong reflection. Third, US wave varies significantly with depth, adding more to spatial inhomogeneity [37], [38]. And forth, the TGC settings allow the user to vary the intensity gain of a area of interest. Therefore, for registering MR and US we compare CoCoMI against LMI [19], which is robust to intensity bias compared to MI.

In the next section, we first set up image registration as an optimization problem and provide some details for three state-of-the-art similarity metrics that we use for comparing against CoCoMI. These three similarity metrics are sum of square differences (SSD) of entropy representations (eSSD), Modality Independent Neighborhood Descriptor (MIND) and LMI. We then provide a detailed explanation of CoCoMI, and demonstrate using simulated images how CoCoMI outperforms LMI, which conditions MI on spatial locations. For the first time, we show the performance of eSSD and MIND on US to MR registration, and compare the results with LMI and CoCoMI. We end with discussions of the results and conclusions.

II. Background

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

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} C, \quad C = D(I_f(\boldsymbol{x}), I_m(\mathbf{T}(\boldsymbol{x};\boldsymbol{\mu})) + \frac{\omega_R}{2}\|\nabla \boldsymbol{\mu}\|^2 \tag{1}$$

where $I_f(\boldsymbol{x})$ and $I_m(\boldsymbol{x})$ are respectively the fixed and moving images, D is a dissimilarity metric, ω_R is a regularization penalty weight, ∇ is the gradient operator and $\mathbf{T}(\cdot;\boldsymbol{\mu})$ is the transformation parameterized by $\boldsymbol{\mu}$. For similarity metrics such as MI or cross correlation that maximize when the images align, negative of the similarity metric is used as the dissimilarity metric. We choose a free-form transformation parameterized by the location of cubic B-spline nodes, which are shown to be more efficient than other alternatives [39]. Therefore, $\boldsymbol{\mu}$ is a vector of the coordinates of all the nodes. In Sections II-A and II-B, we elaborate two different approaches of defining the dissimilarity metric, which are respectively based on local contextual representations and information theoretic measures.

A. Structural Representation

An early use of example of structural representation is the use of orientation of gradients [4]. Recently, two structural representation techniques have been proposed that transform multi-modal registration into uni-modal registration. These two techniques are described below.

1) Entropy Images: Wachinger and Navab [7] proposed calculating the entropy of local patches around every voxel, and generating a new image where intensity of every voxel is the entropy of its local patch. They showed that although the local intensity appearance of images of different modalities is usually different, the local entropy values are similar. They therefore introduced SSD of the entropy images as a similarity function, which they called eSSD. Using entropy representation, they successfully performed nonrigid registration on multi-modal images with high accuracy. The size of the local patches and number of the histogram bins are two important parameters here. Smaller local patches allow more local entropy estimation, and higher number of bins allow more discrimination power against small intensity changes. These two parameters should be selected together, since local patches have to be large enough to populate the intensity bins. Wachinger and Navab [7] therefore use the Parzen window method to estimate the local intensity distributions as it allows robust histogram estimation with fewer samples.

2) Modality Independent Neighborhood Descriptor (MIND): Heinrich *et al.* [8], [40] used the similarity of an image patch to its neighbors, in the same image, as features for multi-modal registration. For every voxel, they computed the self-similarity of a small patch around that voxel to the neighboring patches using SSD. They then used an exponential function to transform SSD distances to similarity weight that are in the [0 1] range, which they called Modality Independent Neighborhood Descriptor (MIND). They showed that although image intensities are different in multi-modal images, the similarity weights are similar. For volumetric images, they demonstrated that comparing every voxels to its 6 neighbors (2 in every dimension) provides with 6 features for every voxel that can be used to compute the dissimilarity metric. They successfully performed nonrigid multi-modal registration of CT and MR images with high accuracy.

B. Local Mutual Information (LMI)

Assume we have the intensities of I_f and I_m at M overlapping pixels. Let p_f and p_m be the marginal intensity probabilities of I_f and I_m respectively, and p_{fm} be their joint intensity probability. All of these distributions are estimated from the overlapping M pixels. We estimate p_{fm} in this work using Gaussian Parzen windows. MI is then computed as [1],

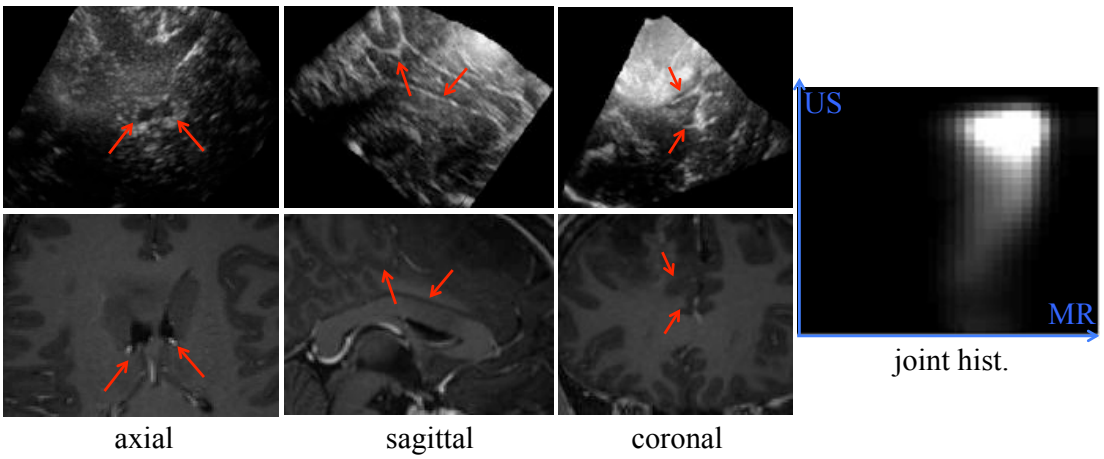


Fig. 2. Corresponding MR and US images of neurosurgery. The ventricles are shown in the axial view, and the sulci, the corpus callosum and the interhemispheric fissure are shown in the sagittal and coronal views. The joint histogram is on the right.

[2]:

$$\mathrm{MI}(I_f, I_m; \mathbf{\Omega}) = \sum_k \sum_i p_{fm}(k,i) log \frac{p_{fm}(k,i)}{p_f(k) p_m(i)}. \quad (2)$$

The probabilities are estimated using samples $\boldsymbol{x}$ from the image domain Ω. A limitation of MI is that it assumes the intensity probabilities do not vary over $\mathbf{\Omega}$, an assumption that can be violated for various reasons such as spatial inhomogeneity. 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; examples are regional MI [17], localized MI (LMI) [19], conditional MI [18] and spatially encoded MI [21]. For comparison, we implement the LMI method [19] where these spatial kernels are box filters. In the other words, LMI is computed by summing MI over multiple local neighborhoods:

$$\mathrm{LMI}(I_f, I_m; \mathbf{\Omega}) = \frac{1}{N} \sum_{j=1}^{N} \mathrm{MI}(I_f, I_m; \mathcal{N}_j) \quad (3)$$

where $\mathcal{N}_j \subset \mathbf{\Omega}$ are spatial neighborhoods and N is the number of these neighborhoods. Each neighborhood should be large enough to contain enough information for MI estimation, and small enough to allow local estimation of MI [19]. Similar to [19], we first randomly select a pixel x_i in $\mathbf{\Omega}$, take a cubic neighborhood $\mathcal{N}_j$ centered at x_i, and randomly select M samples from $\mathcal{N}_j$ (see Figure 3 left). We repeat this for N pixels $i = 1 \cdots N$ and average the results to estimate LMI.

III. Methods

CoCoMI estimates the dissimilarity metric of Eq. 1 in two steps: an offline pre-processing and an online registration steps. The offline step performs contextual analysis and only requires one of the volumes. The online registration step is described in Section III-A, followed by offline contextual analysis step in Section III-B. The optimization of the cost function is detailed in Section III-C.

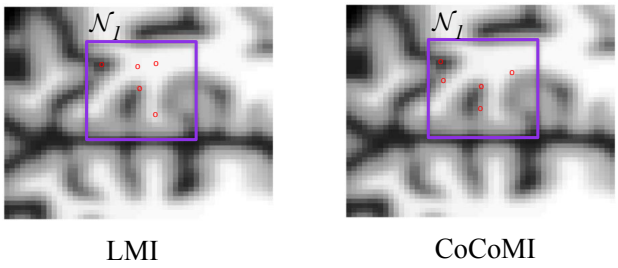


Fig. 3. Computing LMI and CoCoMI using neighborhood $\mathcal{N}_1$ with $M = 5$ pixels. M is usually on the order of 1000. In LMI, M pixels are selected randomly in $\mathcal{N}$. In CoCoMI, M pixels similar to the central pixel, i.e. $\mathcal{S}_1$, are selected.

A. CoCoMI formulation

It is more likely for similar structures to cluster close to each other in the joint histogram when the two images are aligned. An example is two pixels with the same intensity in MR where one is on an edge and the other is in a homogeneous region. These pixels can have very different intensities in US. Therefore, we condition MI to pixels that belong to similar structures. CoCoMI is:

$$\mathrm{CoCoMI}(I_f, I_m; \mathbf{\Omega}) = \frac{1}{N} \sum_{j=1}^{N} \mathrm{MI}(I_f, I_m; \mathcal{S}_j) \quad (4)$$

where $\mathcal{S}_j$ is the set of pixels whose small neighboring patches are similar to the patch around j. For each pixel j, the location of similar pixels is recorded in $\mathcal{S}_j$. One of the volumes is sufficient to find $\mathcal{S}_j$, i.e. $\mathcal{S}_j$ is a self-similarity map. We will elaborate in Section III-B on how we compute it using patch similarities and spatial distance between patches. This equation is very similar to Eq. 3 except that the local neighborhood $\mathcal{N}$ is replaced with the similar pixels $\mathcal{S}$. To compute CoCoMI, a random pixel indexed with j is first selected. The M similar pixels in $\mathcal{S}_j$ are then used to estimate MI. In the other words, MI is only computed from the similar pixels $\mathcal{S}_j$. The process is repeated for N random pixels and the results are averaged to estimate CoCoMI (see Figure 3). To select

N, a tradeoff between running time and performance should be made: Increasing N usually gives better results, with the improvement plateauing for large N. We always set N to 50 for both LMI and CoCoMI, which provides robust and accurate registration in a practical time.

B. Contextual Analysis

The contextual analysis is performed to find $\mathcal{S}_j$, the set of pixels with similar structure to the pixel j. The idea is to consider small patches around pixels and calculate a similarity metric between the patches. We compute these similarities within the same image, and therefore, the similarity metric can simply be SSD. A problem with SSD is that it is not rotationally invariant. Grewenig *et al.* [41] proposed to calculate rotation angles and subsequently rotating patches using interpolation to achieve rotation invariance. This method is however computationally expensive and sensitive due to errors in computation of the rotation angle and interpolations involved. We develop a measure that is invariant to rotation and to small image deformations. We elaborate the measure for 2D images; its extension to 3D is trivial.

We compute two histogram descriptors for each patch, one based on intensities and the other based on gradient orientations. The intensity histogram is similar to the spin image proposed in [42]; we denote it with H^{spin}. It has two axes: the distance from the patch center and the normalized intensity, which is obtained by linearly mapping the intensities of patch pixels to the [0 1] range. Figure 4 (b) illustrates the construction of the spin image. A pixel in the patch with distance to the center of d and normalized intensity of i contributes to bin (d_b, i_b) according to the Gaussian weight:

$$exp\left(-\frac{(i-i_b)^2}{2\sigma_i^2}-\frac{(d-d_b)^2}{2\sigma_d^2}\right). \quad (5)$$

We always set $\sigma_d = \sigma_i = 0.5$, and use 4×4 histograms as illustrated.

The second histogram in Figure 4 (c) is based on the histogram of orientations. For every pixel in the patch, we compute the gradient vector g and compute its radial angle θ using the dot product. This is similar to the rotation invariant feature descriptor (RIFT) of [42] except that here we use the dot product to calculate an angle between 0 to π, while in [42] an angle between 0 to 2π is estimated. Our dot product readily translates to vectors in 3D, while the extension of RIFT in [42] to 3D is not straightforward. Nevertheless, we denote this histogram with H^{RIFT}. Every pixel inside the patch contributes to the histogram proportional to its gradient magnitude and according to a Gaussian weight similar to Eq 5. We use 4×6 histograms as illustrated. The spin image and the RIFT encode complementary kinds of information: the former uses intensity values, while the latter uses the gradient orientations.

We then compare the spin and gradient orientation histograms to obtain two complementary similarity measures. Most histogram comparison metrics are bin-by-bin based, such as sum of L_1 or L_2 norms of differences, Kullback-Leibler (KL) divergence or χ^2 test. A problem with such metrics is that they can give results that are counter intuitive, and also are sensitive to binning and quantization [43]. Therefore, we use the Earth Mover's Distance (EMD) [43], [44] 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. We use EMD-L1 [44], an efficient implementation of EMD where the weight between two bins is their L1 distance. The similarity between two patches i and j is therefore

$$S(i,j) = w_x \cdot \left(\text{EMD}(H_i^{\text{S}}, H_j^{\text{S}}) + w_H \cdot \text{EMD}(H_i^{\text{R}}, H_j^{\text{R}})\right) \quad (6)$$

where w_x and w_H are two weights and H^{S} and H^{R} are respectively spin and RIFT histograms. To set w_H, we have to first calculate the range of the EMD between two histograms. The minimum EMD value is trivially 0. The maximum EMD-L1 value between histograms of size $m \times n$ is $m+n-2$. For the spin and RIFT histograms this translates to respectively 4+4-2=6 and 4+6-2 = 8, and therefore we set $w_H = 6/8 = 0.75$. This weight balances the effect of the two descriptors. The second weigh is $w_x = \exp(x^2/\sigma^2)$, an exponential weight with x as the Euclidean distance between the patch centers. This penalizes pixels that are spatially distant. We always set σ to the width of the $\mathcal{N}$.

We now propose a four step algorithm (see Table I) to find the similar structures. The first two steps are pre-processing, where we mask out part of the image that contain no structure from our computations. These two steps are elaborated in the next two paragraphs. The third step computes the similarities according to Eq. 6. In the forth step, the M pixels with the smallest EMD distance to the central pixel j are located and are marked as $\mathcal{S}_j$. These M pixels will be used to compute MI. Although these steps are offline preprocessing, we downsample the image volume by a factor of 2 to speed the computations. All steps of computing self-similarities for a volume of size 100^3 and neighborhoods of size $\mathcal{N} = 21^3$ take less than 20 minutes on a 3GHz processor.

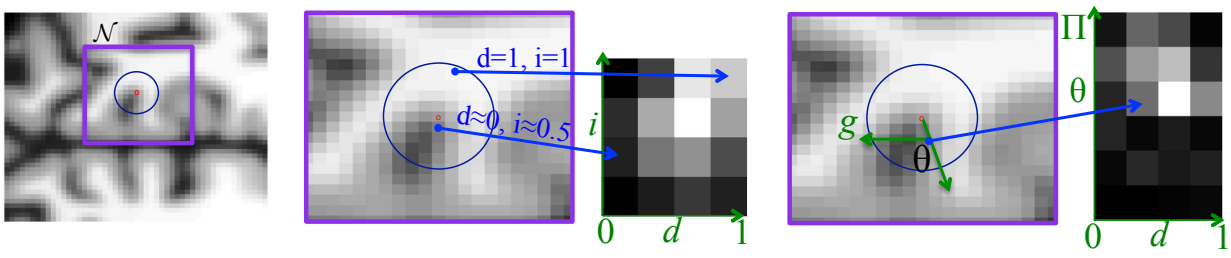
TABLE I
THE OUTLINE OF THE CONTEXTUAL ANALYSIS ALGORITHM.

1) Find $\mathcal{P}_1$ whose surrounding patches contain structure (Eq. 7).
2) Select image pixels $\mathcal{P}_2 \subseteq \mathcal{P}_1$ such that in a neighborhood $\mathcal{N}$ around those pixels there is at least M pixels with structure.
3) For all pixels in $\mathcal{P}_2$, compare the histograms to that of pixels in $\mathcal{P}_1$.
4) For each pixel j in $\mathcal{P}_2$, find the location of M most similar pixels $\mathcal{S}_j$.

Details of step 1: We calculate autocorrelation inside all patches using the Moran's I coefficient [45], [13]. For an image patch with intensities $\{i_j, j = 1 \cdots N\}$ and the mean value $E(i) = \bar{i}$, Moran's I is

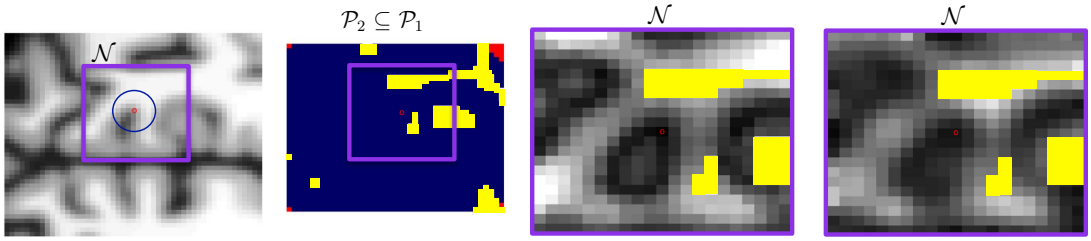
$$I = \frac{N}{\sum_{j,k}^{N} w_{jk}} . \frac{\sum_{j,k}^{N} w_{j,k}(i_j - \bar{i})(i_k - \bar{i})}{\sum_{j}^{N}(i_j - \bar{i})^2} \quad (7)$$

where $W = w_{jk}$ matrix represents the connectivity weights. It can be a binary map or a decaying map based on the distance between j and k. I varies between -1 to 1; values close to 0 translate to random patterns and values close to 1 or -1



(a) $\mathcal{N}$ inside the image (b) Construction of the spin histogram (c) Construction of the RIFT histogram

Fig. 4. Construction of the spin and RIFT histograms H^{spin} and H^{RIFT}. (a) is a slice from BrainWeb. The x axis of H^{spin} in (b) and H^{RIFT} in (c) are distance from the patch center. In (b) and (c), the y axes are respectively intensity and radial gradient angle.



(a) $\mathcal{N}$ inside the image (b) Moran's I selection (c) EMD of the RIFT (d) EMD of the spin image

Fig. 5. The similarity weights. (a) is a slice from BrainWeb. In (b), the union of red and blue is the pixels with structure, i.e. $\mathcal{P}_1$. Blue is $\mathcal{P}_2$. Yellow regions are masked out by Moran's I. (c) and (d) show the self-similarity of different pixels with respect to the central pixel. Black indicates more similarity.

indicate presence of structure. Let $d(j,k)$ be the Euclidian distance between j and k. We simply set W to

$$w(j,k) = \begin{cases} 1/d(j,k) & j \neq k \\ 0 & j = k \end{cases} \tag{8}$$

We set $\mathcal{P}_1$ to pixels whose patch have a $|I|$ ($|.|$ denotes the absolute value) larger than a threshold, which we set to std($|I|$)/2. In Figure 5 (b), $\mathcal{P}_1$ is the union of the red and blue regions, and the yellow region represents the parts that have low Moran's I. Moran's I eliminates uniform patches and in our experience is robust to noise, compared to simply calculating gradient and discarding parts with low gradient. Some pixels that are in $\mathcal{P}_1$ might not have enough pixels in their neighborhood which are also in $\mathcal{P}_1$. An example is the pixels in the right top corner of Figure 5 (b). Since we only compute the descriptive histograms for pixels in $\mathcal{P}_1$, not enough similar pixels can be found in neighborhood $\mathcal{N}$ of these pixels, and therefore step 2 is necessary.

Details of step 2: $\mathcal{P}_2$ is found as:

$$\mathcal{P}_2 = (\mathcal{P}_1 * B) > M \tag{9}$$

where B is the box kernel, which is zero everywhere except in $\mathcal{N}$, and $*$ denotes the convolution. Pixels in $\mathcal{P}_2$ are guaranteed to have at least M pixels in their $\mathcal{N}$ which are in $\mathcal{P}_1$. In Figure 5 (b), $\mathcal{P}_2$ is shown in blue.

Figure 5 (c) and (d) shows the resulting EMD values. Pixels that are from a similar neighborhood have a smaller EMD distance to the central pixel. These two similarity maps are combined according to Eq. 6, and $\mathcal{S}_j$ is found as the set of most similar pixels to the pixel j. We now show how we optimize the cost function.

C. Optimization

Given the current transformation parameter $\boldsymbol{\mu}$, a typical iterative optimization technique will seek an incremental $\Delta\boldsymbol{\mu}$ so that $\boldsymbol{\mu} + \Delta\boldsymbol{\mu}$, on average, reduces the cost function. Neglecting the regularization term of Eq. 1 to avoid clutter, the dissimilarity

$$D(I_f(\boldsymbol{x}), I_m(\mathbf{T}(\boldsymbol{x}; \boldsymbol{\mu} + \Delta\boldsymbol{\mu})) \tag{10}$$

should decrease. We use a method similar to the stochastic gradient descent optimization method of [20]. Therefore, the gradient of D w.r.t. $\Delta\boldsymbol{\mu}$ must be computed. Using the chain rule,

$$\nabla_{\Delta\boldsymbol{\mu}} D = \frac{\partial D}{\partial \Delta\boldsymbol{\mu}} = \frac{\partial \mathbf{T}}{\partial \Delta\boldsymbol{\mu}} \cdot \frac{\partial I_m}{\partial \mathbf{T}} \cdot \frac{\partial D}{\partial I_m} \tag{11}$$

where $\frac{\partial \mathbf{T}}{\partial \Delta\boldsymbol{\mu}}$ is simply the transformation Jacobian, $\frac{\partial I_m}{\partial \boldsymbol{T}}$ is the image gradient, and $\frac{\partial D}{\partial I_m}$ is the derivative w.r.t. the intensity of all pixels. We see that the derivatives are applied to the moving image I_m. In many multi-modal registration applications, the resolution and signal to noise ratio of the two images can be different. On one hand, it makes more sense to use the higher quality image to perform contextual analysis and find similar patches $\mathcal{S}_j$. Also, in image-guided surgical applications, the pre-operative image usually has a higher quality, making it more desirable to perform contextual analysis on this image to find $\mathcal{S}_j$. One the other hand, it makes more sense to set the image with higher quality as I_m since both $\frac{\partial I_m}{\partial \mathbf{T}}$ and $\frac{\partial D}{\partial I_m}$

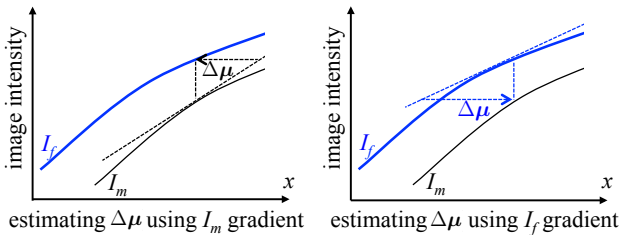


Fig. 6. Two options to find the incremental transformation $\Delta\boldsymbol{\mu}$ for simple 1D images.

will be more accurate. This means that the self-similarity maps need to deform with I_m in every iteration of the registration. Also, the image gradients $\frac{\partial I_m}{\partial \boldsymbol{T}}$ have to be recomputed in every iteration. Therefore, we use an alternative approach to Eq. 10, which solves the aforementioned problems and speeds up the computations as follows. We *virtually deform the fixed image*, and apply *the inverse of the deformations to the moving image* [46], [47], [48], [49], i.e. equation 10 changes to

$$D(I_f(\mathbf{T}(\boldsymbol{x};\Delta\boldsymbol{\mu})), I_m(\mathbf{T}(\boldsymbol{x};\boldsymbol{\mu})). \tag{12}$$

Note that here, we are applying the incremental deformation to I_f. Figure 6 shows the difference between the two approaches. Here, the transformation update $\Delta\boldsymbol{\mu}$ is a simple translation of the moving image I_m. The derivative of D is now

$$\nabla_{\Delta\boldsymbol{\mu}} D = \frac{\partial D}{\partial \Delta\boldsymbol{\mu}} = \frac{\partial \mathbf{T}}{\partial \Delta\boldsymbol{\mu}} \cdot \frac{\partial I_f}{\partial \mathbf{T}} \cdot \frac{\partial D}{\partial I_f}, \tag{13}$$

where the derivatives are applied w.r.t. the fixed image. The updated transformation is

$$\mathbf{T}(\boldsymbol{x};\boldsymbol{\mu}) \circ \mathbf{T}(\boldsymbol{x};\Delta\boldsymbol{\mu})^{-1} = \mathbf{T}(\mathbf{T}(\boldsymbol{x};\Delta\boldsymbol{\mu})^{-1};\boldsymbol{\mu}). \tag{14}$$

Intuitively, the new formulation assumes we want to transform the fixed image, calculates an incremental transformation by minimizing D, and applies the inverse of the incremental transformation to the moving image. Finally, we prove that $\mathbf{T}(\boldsymbol{x};\Delta\boldsymbol{\mu})^{-1} = \mathbf{T}(\boldsymbol{x};-\Delta\boldsymbol{\mu})$ up to the first order approximation. Using the Taylor expansion:

$$\mathbf{T}(\boldsymbol{x};\Delta\boldsymbol{\mu}) = \mathbf{T}(\boldsymbol{x};0) + \frac{\partial \mathbf{T}}{\partial \boldsymbol{\mu}}\Delta\boldsymbol{\mu} = \boldsymbol{x} + \frac{\partial \mathbf{T}}{\partial \boldsymbol{\mu}}\Delta\boldsymbol{\mu}.$$

Therefore,

$$\mathbf{T}(\boldsymbol{x};\Delta\boldsymbol{\mu}) \circ \mathbf{T}(\boldsymbol{x};-\Delta\boldsymbol{\mu}) = \boldsymbol{x} + \frac{\partial \mathbf{T}}{\partial \boldsymbol{\mu}}\Delta\boldsymbol{\mu} - \frac{\partial \mathbf{T}}{\partial \boldsymbol{\mu}}\Delta\boldsymbol{\mu} = \boldsymbol{x} \tag{15}$$

In summary, we use equation 12 to estimate $\Delta\boldsymbol{\mu}$, and update the current transformation parameter $\boldsymbol{\mu}$ to $\boldsymbol{\mu} - \Delta\boldsymbol{\mu}$. Please note that, in general, the combination of two B-spline transformations is not a B-spline transformation and the inverse of a B-spline transformation may not exist. In the other words, B-spline transformations do not form a group. However, for small incremental updates we showed that the inverse can be easily obtained. We can now set the image with higher quality as the fixed image, and do not need to transform the similarity maps $\mathcal{S}_j$.

Our implementation is as follows. In every iteration, we randomly select N pixels $i = 1 \cdots N$. The random selection

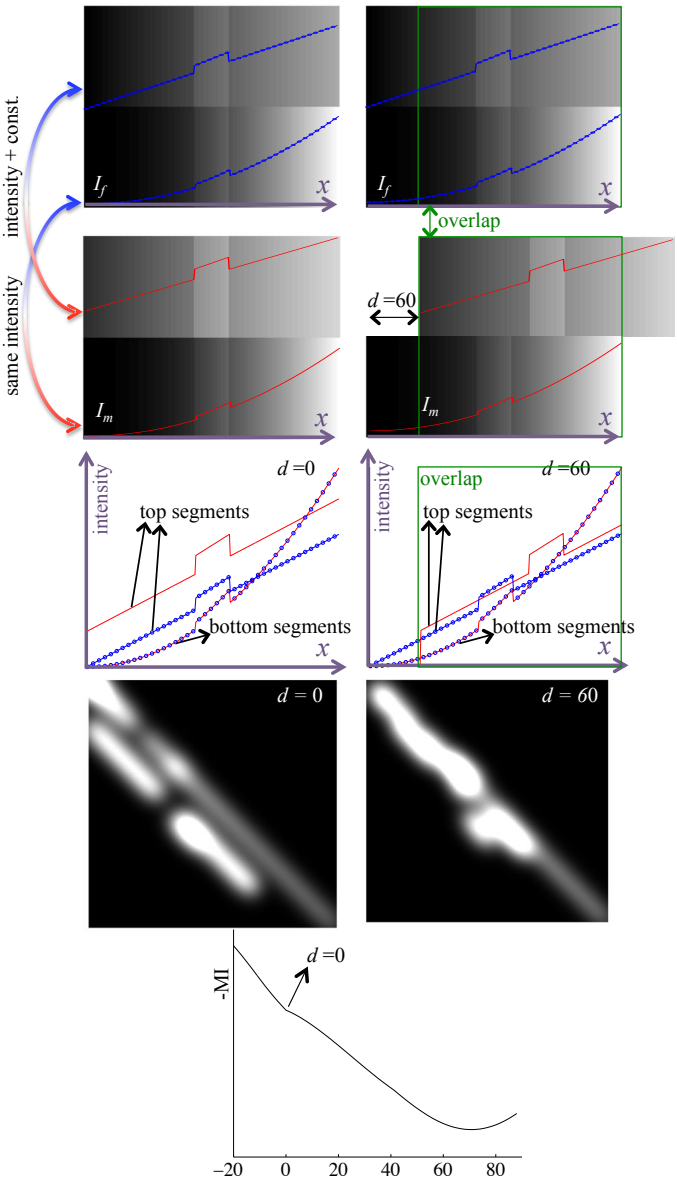


Fig. 7. MI values with deforming I_m by rigidly moving its top segment by d. In the top three rows, the intensity profile of each segment is plotted in blue (with circles overlaid) for I_f and red for I_m: the intensities of the bottom segments are the same, and the intensities of the top segments are drifted by a constant value. We crop the left d columns of I_m and I_f as d increases and compute MI for the overlap between the I_f and I_m. The third row shows the intensity profiles of the images for 2 different values of $d = 0$ and 60 pixels (the width of the images is 300 pixels). The joint histograms are shown in forth row. The last row's plot shows that while the images are registered at $d = 0$, -MI (note the negative sign) reaches its minimum at $d = 72$.

generally picks different pixels in every iteration. For every pixel i, we use the M similar pixels in $\mathcal{S}_j$ to calculate MI, and average the MI results for $i = 1 \cdots N$ to compute CoCoMI and its partial derivatives w.r.t. the location of the B-spline nodes. Letting $\nabla_{\Delta\boldsymbol{\mu}} C$ be the gradient of C in Eq. 1 w.r.t. $\Delta\boldsymbol{\mu}$ (with D as in equation 12), the update equation is $\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t - a_t \nabla_{\Delta\boldsymbol{\mu}} C$. The step size is a decaying function of the iteration number t: $a_t = a/(A+t)^{\tau}$, with $a > 0, A \geq 0$ and $0 < \tau \leq 1$ user-defined constants [20]. The recommended values for these parameters are provided in [20]: A should be

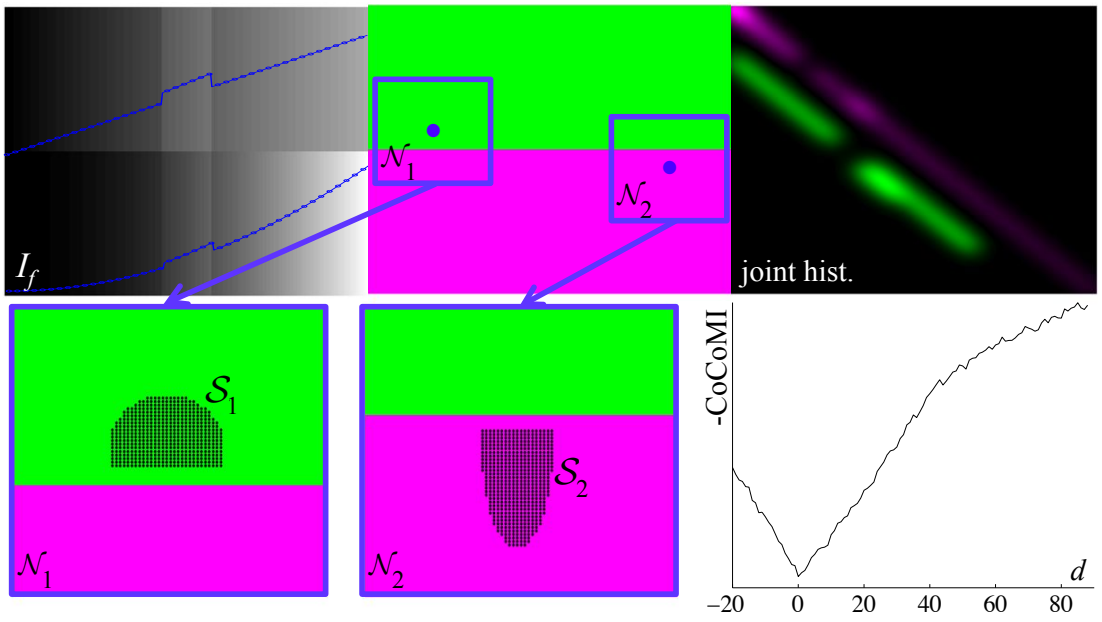


Fig. 8. CoCoMI pixel selection. The green and magenta segments of the joint histogram belong to the top and bottom image segments. In the bottom row, selected pixels for the two zoomed neighborhoods are shown with black dots. Since the intensity profiles of the top and bottom segments of I_f are different (piecewise linear and parabolic as shown), they are not contextually similar, and hence $\mathcal{S}_1$ and $\mathcal{S}_2$ are entirely in the top or bottom segment. The -CoCoMI plot shows the global minimum at $d = 0$.

around 0.1 of the maximum number of iterations or less and τ should be more than 0.6. The value of a is user-defined and is critical as it determines the step-size. 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 100 for both LMI and CoCoMI. We elaborate our implementation for computing similar patches in the next section.

IV. Results

For LMI and CoCoMI, we use our implementation which models deformations with free-form B-splines. We limit the self-similarity comparisons to local neighborhoods of sizes $\mathcal{N} = 81^2$ in 2D and $\mathcal{N} = 21^3$ in 3D. The neighborhood $\mathcal{N}$ of LMI is of the same size, unless otherwise noted.

We use 2D simulated data and 3D US and MR patient data for validation. In the simulation experiments, we compare only LMI and CoCoMI for two reasons. First, LMI and CoCoMI are similar except for the use of contextual information. Detailed comparison of these two methods provides insight into the importance of exploiting contextual information. Second, in the simulation data we introduce known deformations and study different methods by comparing the deformation they recover with the ground truth. On one hand, if the ground truth deformation is a displacement field which is defined for every pixel, entropy images and MIND have an advantage since they model the deformation in a similar manner. On the other hand, if the ground truth deformation is generated with B-splines, LMI and CoCoMI have an advantage since they model the deformation similarly. Therefore, we use B-spline deformations in the simulation experiments and compare LMI and CoCoMI. In the US to MR patient data, however, we compare all four of the similarity metrics.

A. Simulated Images

We first show how MI can force similar intensity relations between different image parts. Figure 7 top shows two images I_f and I_m, each with two segments. The intensity values of I_f and I_m are superimposed on the images with respectively blue and red lines. The bottom segments of I_f and I_m have the same intensities, while the intensities of the top segments are similar up to a constant value. We now deform I_m by simply moving the top segment by d. For $d = 0$ and $d = 60$, we show the intensity profiles of both images in the third row. Note that the red and blue curves that represent the intensity of the bottom segments are identical for both $d = 0$ and $d = 60$. Also, notice that at $d = 60$, the intensity relation of the top segments gets closer to identity (i.e. the constant intensity difference gets close to zero), which means that the top and bottom segments follow the similar intensity transformations (see also Table II). Forth row shows the joint histograms at $d = 0$ and $d = 60$. Note that at $d = 60$, the histogram is more compact, indicating that MI is higher. The bottom plot shows that while the two images are aligned at $d = 0$, the -MI curve is minimized around $d = 72$. This is because as d increases from zero, the constant intensity difference between the top parts becomes smaller. Also note that there is a gradient

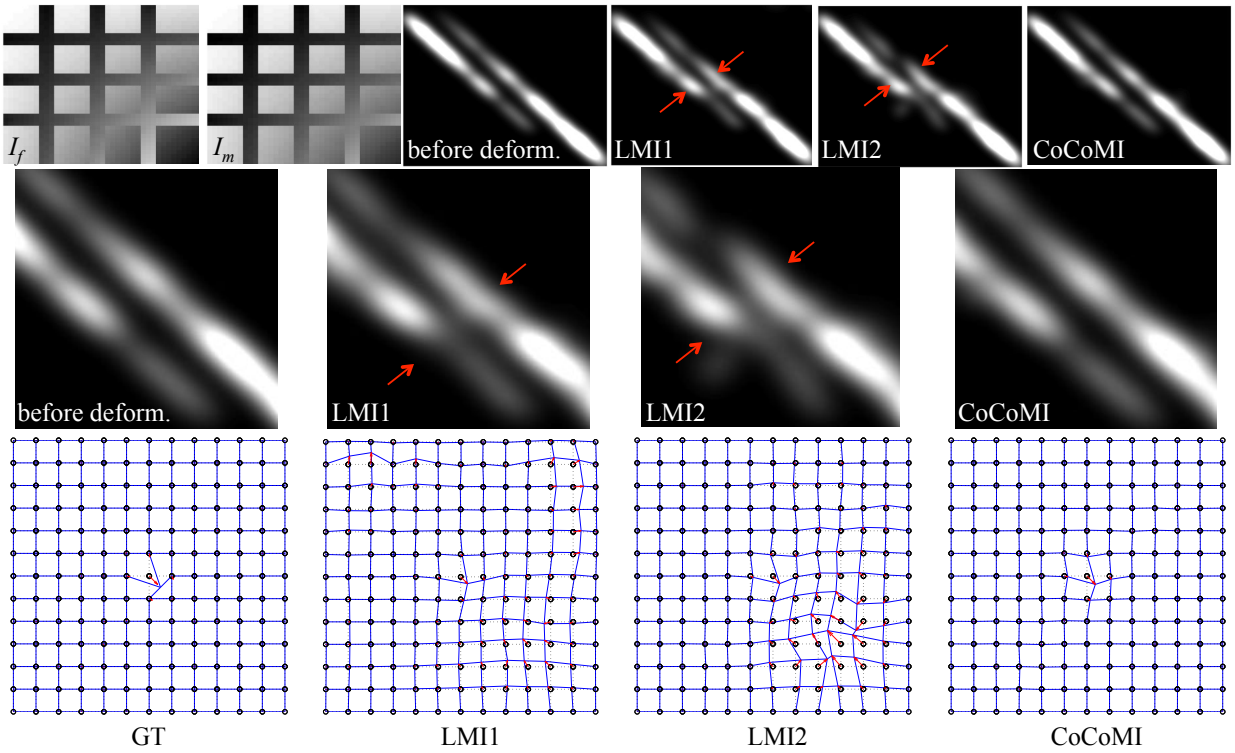


Fig. 9. Deformable registration of two simulated images. The ground truth joint histogram before deforming I_m has two branches as shown. The second row shows a detailed view of the middle of the histograms. The deformable MI and LMI both try to bring these two closer as shown by the arrows. This results in the large erroneous deformation as shown in the third row.

TABLE II
THE INTENSITY RELATIONS OF I_f AND I_m OF FIGURE 7. AT $d = 60$ PIXELS, $C_2 \approx 0 \ll C_1$, AND THEREFORE THE INTENSITY RELATIONSHIPS OF THE TOP AND BOTTOM SEGMENTS ARE SIMILAR.

segment	$d = 0$ pixels	$d = 60$ pixels
top	$I_m = I_f + C_1$	$I_m = I_f + C_2$
bottom	$I_m = I_f$	$I_m = I_f$

discontinuity at $d = 0$, which is caused by the two vertical stripes aligning at $d = 0$. Treating the two segments separately, as in our similarity measure based on contextual information, significantly mitigates this problem, as shown in Figure 8. Since the top and bottom segments of I_f have different intensity profiles (respectively piecewise linear and parabolic), CoCoMI either selects pixels from the top (as in $\mathcal{N}_1$) or bottom (as in $\mathcal{N}_2$) segments. The CoCoMI cost function predicts the correct alignment at $d = 0$. The slight fluctuations in the CoCoMI plot are due to its random neighborhood selection for every d. We use $N = 50$ neighborhoods according to Eq. 4. Increasing N decreases these fluctuations.

As a second example, Figure 9 shows two images I_f and I_m of size 40×40 before deformation, and their joint histogram. We deform I_m using a B-spline mesh with node spacing of 4 pixels, with the middle node moved by 2 pixels in the right and down directions as shown in the ground truth deformation. LMI1 and LMI2 are computed respectively using all the pixels in the neighborhoods $\mathcal{N}$ of size 29^2 and 21^2. Note that the size of $\mathcal{N}$ cannot be very small because it has to contain at least M samples for reliable estimation of MI; 21^2 gives us $M = 441$ samples. N (see Eq. 3) is set to 10; increasing it to 50 did improve the results. For CoCoMI, we similarly set N to 10 and M to 400. LMI produces incorrect deformations which can be observed from both the joint histograms and the deformation fields.

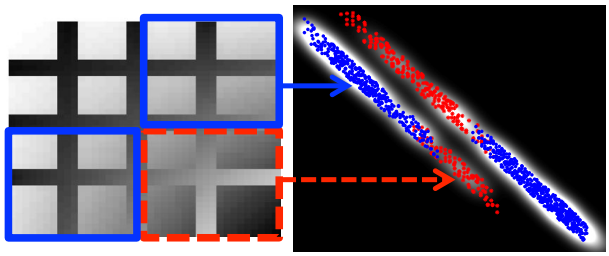


Fig. 11. Spatial locations corresponding to the joint histogram for aligned I_f and I_m. The solid blue and dashed red rectangles respectively correspond to the blue dots and red asterisks as marked.

We then deform the images of Figure 9 by 400 random deformations generated using B-spline nodes that have a 10 pixel spacing between adjacent nodes. In each deformation, every node is displaced by a uniform random number in the ± 1.5 pixels range in both directions. The registration problem is solved using the same B-spline spacing, i.e. 10 pixels. The intensity sum of square differences at every pixel between the un-deformed and deformed I_m is shown in Figure 10. The initial difference is large as expected. The ground truth difference

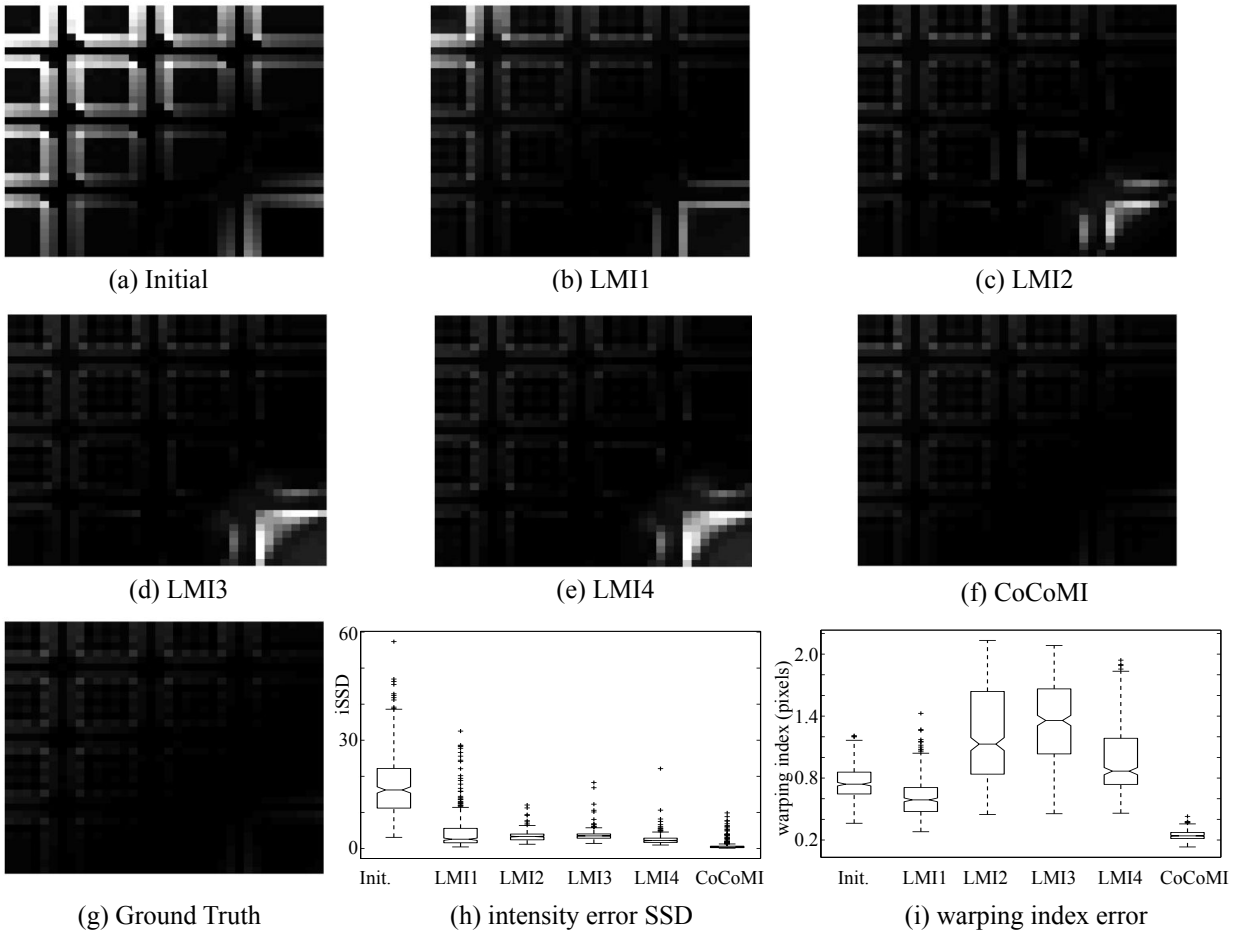


Fig. 10. The I_m in Figure 9 is deformed with 400 random deformations and is registered to I_f. (a) to (g) show the intensity sum of square difference (iSSD) of the deformed I_m and original I_m, averaged over the 400 cases. Deforming I_m back using the ground truth deformation (image in the second row and column) does not give zero iSSD, as shown, because of smoothing induced by two forward and backward interpolations. All images have the same intensity scale.

in (g) is obtained by deforming the I_m back using the ground truth deformation. It is not zero because of the errors incurred in the two interpolations (the random deformation and its inverse). Using the same instance of random deformation, we recover the deformation using LMI with 4 different parameter settings: LMI1 to LMI4 respectively uses neighborhoods $\mathcal{N}$ of sizes 37^2, 29^2, 23^2 and 19^2. The sum of square of differences for CoCoMI in (f) is the most similar to the ground truth in (g) computed over the 400 deformations. The warping index, which is the root mean square (RMS) of the registration error in each pixel, is also significantly smaller in CoCoMI. This example show that LMI with different neighborhood sizes $\mathcal{N}$ cannot achieve the same results of CoCoMI. The results of (b) to (e) show that the error is high at the right bottom corner, and is relatively low at the top right and bottom left corners. To explain this, we plot the spatial correspondence with the joint histogram of *aligned fixed and moving images* in Figure 11. Here, we see that the bottom right corner corresponds to areas with low density in the joint histogram, which are also close to the high-density areas. This causes these points to be pulled towards the high-density areas and results in the large error. In contrast, the corners at top right and bottom left correspond to the high-density areas and therefore have lower deformation error.

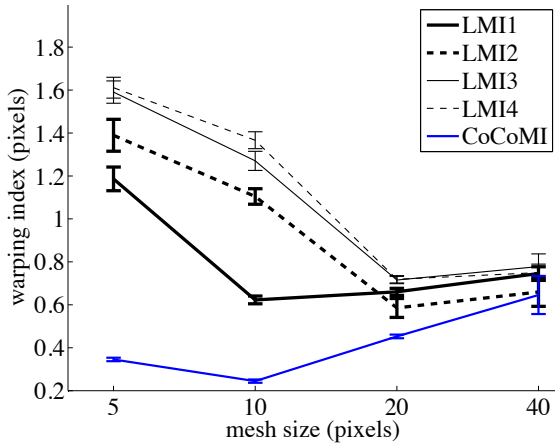


Fig. 12. The effect of mesh size (the distance between adjacent nodes) on different methods using 400 random deformation of images of Figure 9. The error-bars are divided by 5 to aid visualization.

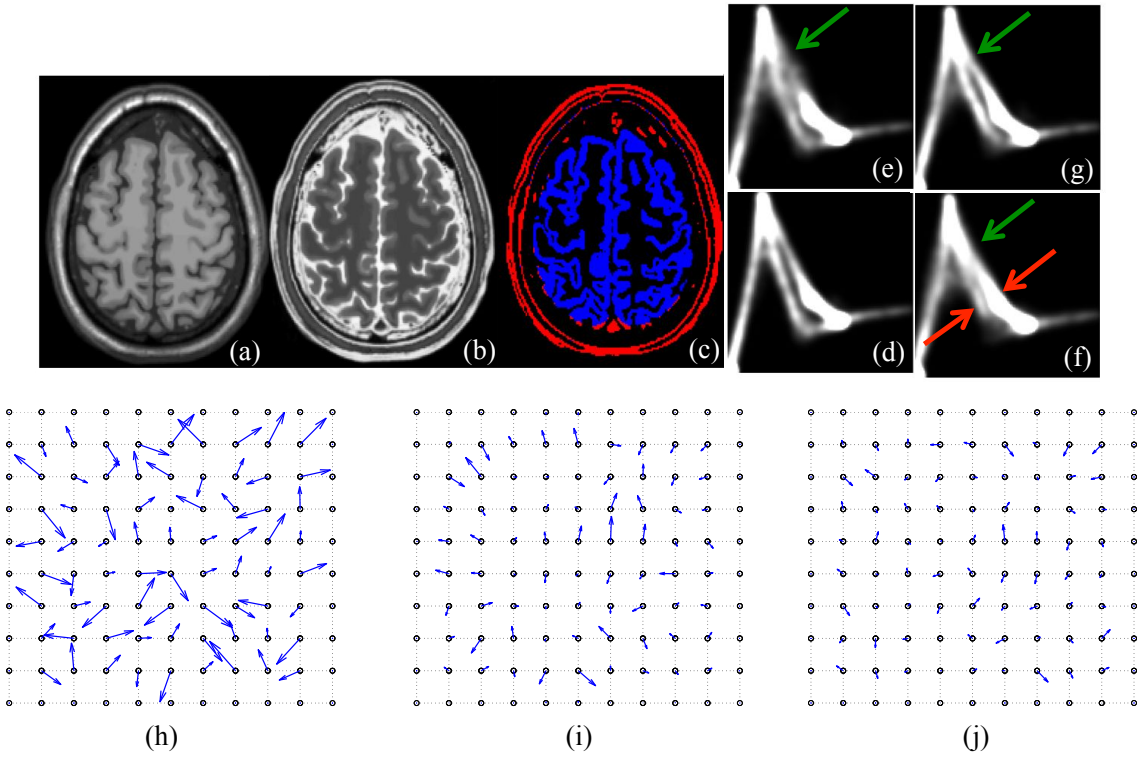


Fig. 13. Non-rigid registration of the BrainWeb images. (a) and (b) are respectively T1 and T2. (c) shows the two red and blue regions that construct the two branches in the joint histogram. (d) and (e) are respectively the joint histograms before and after deforming T2. In (f) and (g) the joint histograms are shown using the deformations found with respectively LMI and CoCoMI. Both methods correct the spread in the joint histogram marked by the green arrow. LMI also brings the two branches closer as marked with the red arrows. (h) is the ground truth random deformation. (i) and (j) are the errors in the deformations recovered by LMI and CoCoMI respectively. The deformation vectors are scaled, and the nodes are downsampled by a factor of 2 to ease visualization.

To show the effect of the mesh size, i.e. $\mathbf{T}_{\boldsymbol{\mu}}(\boldsymbol{x})$ in Eq. 1, we vary the spacing between B-spline nodes from 5 to 40 pixels. The 400 random deformations are the same as the previous experiment: node spacing is 10 pixels and the range for each node is ± 2 pixels in each dimension. Figure 12 shows the results. LMI1 to LMI4 are using neighborhoods of sizes 37^2 to 19^2 similar to the previous experiment. Here, again, CoCoMI significantly outperforms LMI. As the mesh size increases to 40 pixels, all methods perform similarly. The difference is greatest at smaller mesh sizes where deformations can be more arbitrary. An interesting observation is that CoCoMI and LMI1 give the smallest warping index at the 10 pixel mesh spacing because the random deformation field is also generated with the 10 pixel mesh spacing (i.e. the ground truth deformation and the deformation model $\mathbf{T}$ of Eq. 1 are identical). This, however, is not the case for LMI2 to LMI4 where increasing the mesh spacing to 20 pixels, and therefore limiting spurious deformations, generates the best results. A final observation from this figure is that CoCoMI is more robust to allowing a mesh size that is actually finer than the actual deformation (i.e. node distance of 5 pixels for a ground truth deformation with node distance of 10 pixels). In practice, obviously the scale of the actual deformation is unknown, and CoCoMI thus probably will not stray too much from the correct deformation if the mesh size is chosen too small.

In the next experiment, we use T1 and T2 images from BrainWeb [22] at 1% noise and 20% non-uniformity. The images are of size 160×180 and the spacing between the nodes is 10 pixels. Figure 13 (a) and (b) show the images, and (c) shows the blue and red anatomical regions that constitute the two branches in the joint histogram. (d) shows the joint histogram. We then deform the T2 image by randomly moving all nodes in the range of ± 1 pixel in both dimensions, as shown in (h). Please note that to keep the size of the figure small, we are showing every other node (i.e. at 20 pixel spacing) and are also scaling up the deformation vectors. The joint histogram of the images before registration is shown in (e) and after registration with LMI and CoCoMI in (f) and (g). Focusing on the green arrows, we see that both LMI and CoCoMI reduce

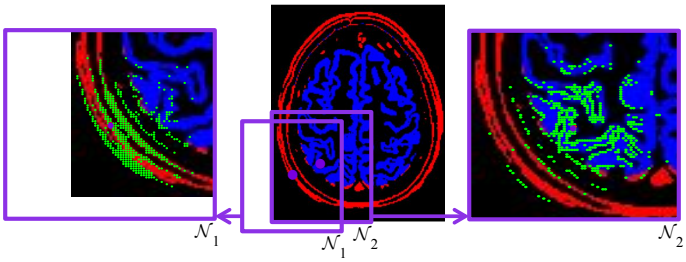


Fig. 14. Pixel selection in CoCoMI. In $\mathcal{N}_1$ and $\mathcal{N}_2$, the center pixel (the violet circle) is respectively on the red and blue anatomy (see also Figure 1.). The selected pixels are marked with green dots in $\mathcal{N}_1$ and $\mathcal{N}_2$. Note than in $\mathcal{N}_1$ and $\mathcal{N}_2$, very few pixels from respectively blue and red are selected.

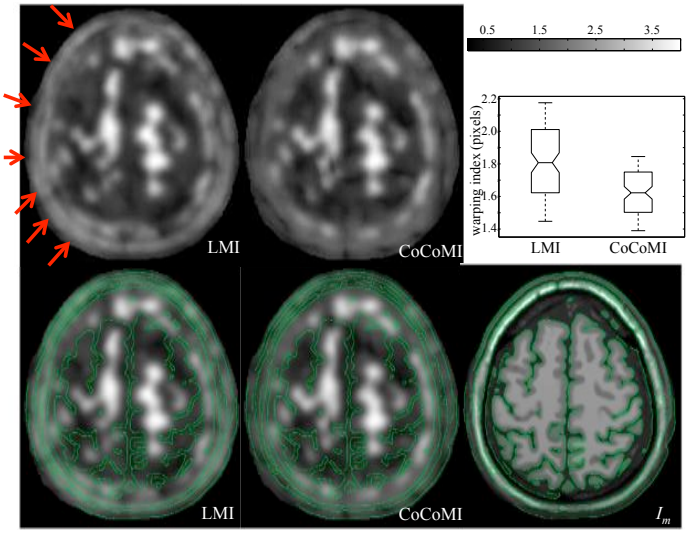


Fig. 15. The average registration error of LMI and CoCoMI. In the bottom, the images are shown with superimposed contours (automatically computed from I_m). The skin, pointed to by the red arrows, has significantly more error in LMI. The warping index boxplot also shows lower error for CoCoMI.

the spread of the joint histogram, compared to (e). However, the parts marked by the red arrow have been mistakenly made compact. We computed the global MI between T1 and T2 after registration with LMI and CoCoMI, and, interestingly, found out that LMI produces higher global MI. However, as this example shows, higher global MI and compact joint histogram does not always translate to higher deformation fidelity, as the error vectors in (i) and (j) show.

To better see why CoCoMI keeps the two histogram branches separate and produces smaller errors, we show the results of CoCoMI pixel selection in Figure 14. Neighborhoods $\mathcal{N}_1$ and $\mathcal{N}_2$ are respectively centered on a red and blue pixels. Selected pixels are shown in green. CoCoMI selects pixels similar to the center, and as a result, we see that very few blue pixels are selected in $\mathcal{N}_1$. Similarly, very few red pixels are selected in $\mathcal{N}_2$. When the blue and red pixels are not mixed for MI computations, they are not forced to move closer together in the joint histogram in the optimization process, which is visible from joint histogram (g) in Figure 13.

Finally, we deform the T2 image by moving each node of the B-spline mesh by a uniformly distributed random number in the range of ± 4 pixel in both xy directions. We generate 100 of these deformation instances and recover the motion using LMI and CoCoMI. Figure 15 shows the results of the average deformation error. A common pattern with both LMI and CoCoMI is the high error in the white matter area, where there is not enough texture to track the motion. Comparing the results around the skin, we see that LMI produces much larger errors because it treats the skin region similar to the gray matter, and tries to push the two branches of the two histogram into one as we showed before. This example illustrates how conditioning the MI estimation on similar structures can improve the registration results for images of modalities other than US.

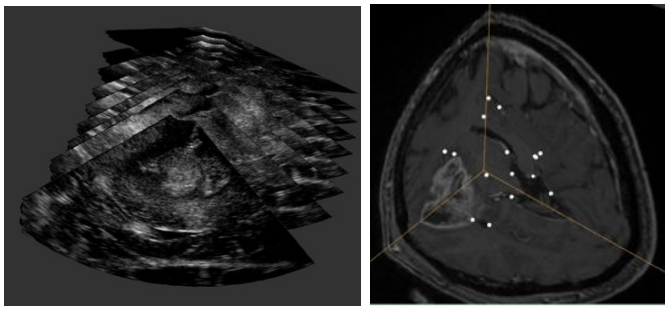


Fig. 16. US and MR images in the BITE database. Left shows the US slices before 3D reconstruction. Right shows the MR volume as well as the landmarks, which are used for assessing the registration accuracy.

B. US-MR Images of Image-Guided Neurosurgery

The clinical data from 14 patients are acquired at the Montreal Neurological Institute, and are part of the BITE database available online at http://www.bic.mni.mcgill.ca/BITE. 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. The pixel size of 2D ultrasound images is 0.3 mm. The ultrasound probe is tracked with a Polaris camera (NDI, Waterloo, Canada), which provides the three locations and angles (6 DOF) of each image. Figure 16 shows the US slices which are placed in 3D using the tracking information. To perform image registration, we reconstruct US volumes with a voxel size of 1 mm in the x y z directions from 2D US slices [50]. This relatively large voxel size means that for every voxel, multiple US measurements from different images and insonification angles are available. Therefore, the effect of US speckle noise is minimized, since US speckles look different when imaged from different locations or angles [51], [52], [53]. Each volume has a different size because the depths and sweeping areas are different; the typical size is on the order of 100^3 voxels of size 1^3 mm.

The MRI volume has a voxel size of 0.5 mm in the imaging plane and 1 mm slice thickness. We resample the volume to the isotropic 1 mm voxel size, same as the reconstructed US voxel size. To find the transformation between the patient and the MR, landmarks are identified on the MR and the patient's skin. This transformation, along with the US tracking information and US calibration matrix are all provided in BITE, which we use to perform the initial rigid registration of MR to US and crop the MR volume to the same size as the US volume. The intensity of the US images is not reliable at the top and bottom of the images. We exclude the boundaries of the US volume from our computations to take this into account. Therefore, the cropped MR volume is larger than the useful US volume and always covers it.

To validate the results, three experts have selected corresponding anatomical landmarks in the US and MR images, which is available in BITE. Figure 16 shows these landmarks on the MR data. These landmarks are used to calculate mean target registration errors (mTRE). The mTRE of n

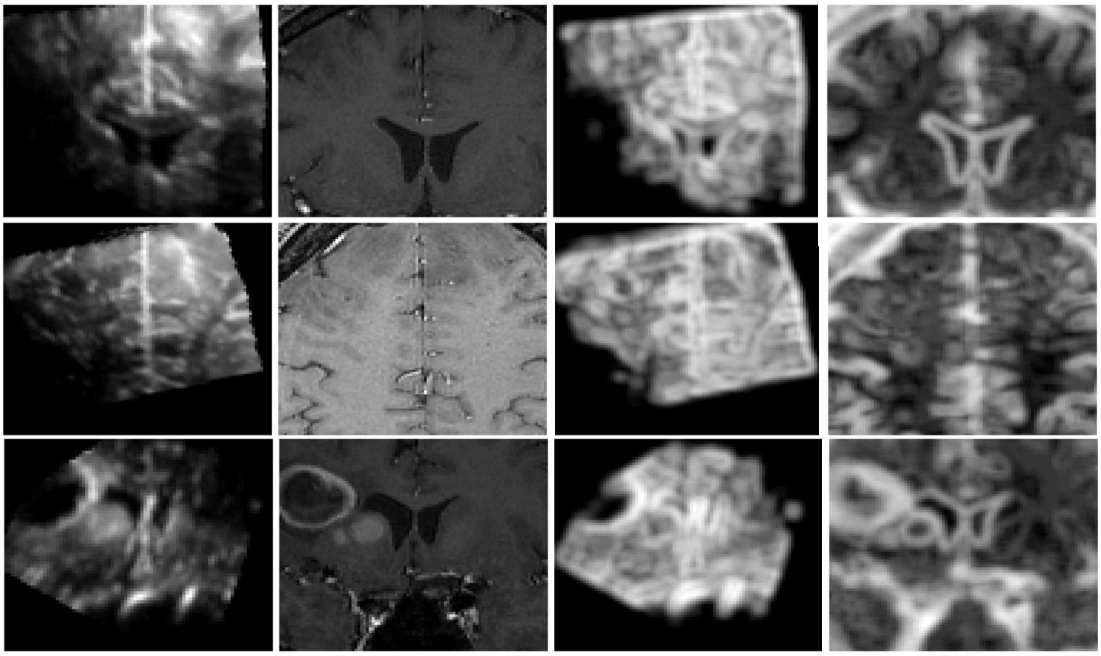


Fig. 17. Three different corresponding slices of US, MR, and their entropy images. The entropy images are calculated using patches of size 5^3 with 32 bins. At tissue boundaries, both US and MR have high values in the entropy images. However, the low intensity areas of the MR entropy images, mostly in the white matter, do not have corresponding low intensities in the US entropy images.

corresponding marks at locations $\boldsymbol{x}$ and $\boldsymbol{x}'$ in the two images is calculated according to

$$\mathrm{mTRE} = \frac{1}{n}\sum_{i}^{n} \|\boldsymbol{T}(\boldsymbol{x}_i) - \boldsymbol{x}'_i\| \tag{16}$$

where $\boldsymbol{T}$ is the transformation (see Eq. 1) and $\|\boldsymbol{v}\|$ is the Euclidian length of the vector $\boldsymbol{v}$. BITE contains data from 14 patients. We have carefully inspected all landmarks in all patients and noticed that the landmarks in one of the patients do not match in the US and MR volumes. We are including a picture of this case as a supplementary material. We have also asked a neuro-radiologist to check the accuracy of the landmarks in two randomly selected patients, who confirmed their accuracy. Therefore we analyze 13 patients in this study. The initial mTRE values are shown in Table III. Patient P10 in this table has a small initial mTRE of 1.52 mm. We use the MR and US images of this patient as a bronze standard registered data to plot different similarity metrics as a function of the rigid translation of the MR.

C. Implementation Details

SSD of entropy images, eSSD: For computing entropy images, the patch size and number of intensity bins are two important parameters. Smaller patches give a more local estimate of the entropy, while larger patches generate smoother entropy images. More bins give the entropy image more discrimination power against intensity changes, while they need larger patches for populating the bins. Similar to Wachinger and Navab [7], we varied the number of bins from 8 to 128 and the patch size from 5^3 to 13^3 to analyze and optimize the results. Figure 17 shows three different corresponding slices of MR, US, entropy of MR and entropy of US. Note that the US entropy images are invariant to the large spatial bias in the US images. We see that while the US and MR entropy images share substantial similarities, they have different intensities in some regions especially in the white matter. Therefore, a robust similarity metric is necessary to limit the effect of these outliers. Using P10 in Table III as a bronze standard registered data, we plot the eSSD cost function by rigidly moving the MR entropy image in the x y directions by ± 5 pixels in Figure 18 . The minimum cost is expected to be at zero displacement. The results of Figures 17 and 18 show that the eSSD is not an optimal similarity metric for US to MR registration. We exploited US speckle denoising algorithms and also reconstructed US volumes with larger voxel size (up to 2 mm) to minimize possible adverse effect of US speckle on entropy images, but the problem persisted. We also used two mono-modal registration methods for deformable registration of the entropy images: the block matching algorithm of ANIMAL+INSECT [31], and the cross correlation similarity metric with symmetric diffeomorphic deformation of [54]. The registration did not improve the alignments or reduce the mTRE values, and therefore we do not present them here.

MIND: For the MIND descriptors, the number of neighbors for patch similarity can be tuned from 6 (2 neighboring voxels in every dimension) to $d^3 - 1$ for considering all voxels in a cube of size d^3. The number of neighbors is equal to the number of similarity features that are assigned to every voxel. More neighbors provide every voxel with more features, but also makes the MIND descriptor sensitive to deformations. We

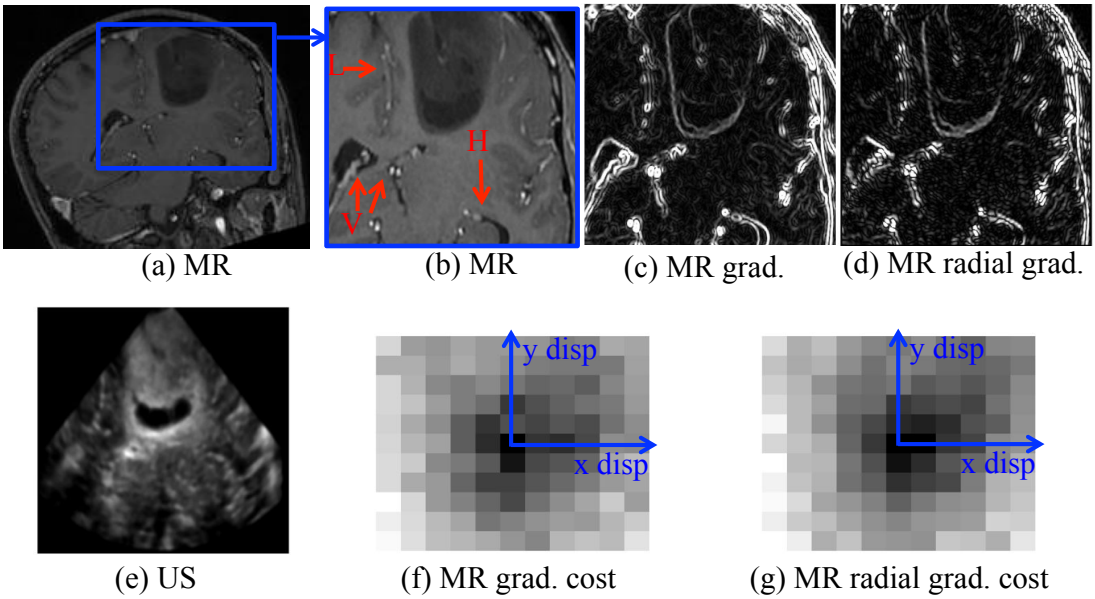


Fig. 19. Gradient magnitude and radial gradient of MR versus US. (a) is a slice of MRI, with a zoomed view in (b). Longitudinal fissure, ventricles and hippocampus are respectively marked by L, V and H. (c) is the magnitude of the MR gradient. (d) is the MR radial gradient. (e) is the corresponding US. (f) and (g) are the LMI values obtained by rigidly moving the MR gradient of respectively (c) and (d) in the x y directions by ± 5 pixels. The minimum cost is expected to be at the zero displacement. Black represents smaller dissimilarity.

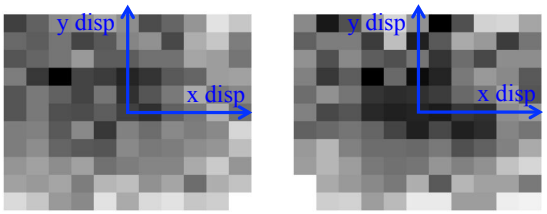


Fig. 18. Qualitative eSSD similarity metric versus rigid translation of the MR in the x y directions by ± 5 pixels. The minimum cost is expected to be at the zero displacement. Black represents smaller dissimilarity.

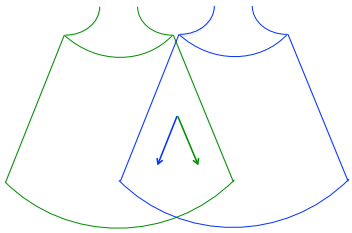


Fig. 20. US images acquired from different locations or different probe orientations have different US insonification angles, i.e. the radial directions are not aligned.

found that 6 neighbors give substantially the best results. We use the implementation provided in [8] for deformable registration, which computes symmetric diffeomorphic deformations. We tested different pixel spacings and numbers of iterations and found that the following values give the optimal results for our data: 3 hierarchical levels at spacings of 4, 2 and 1 pixels, with the number of iterations at each level respectively set to 8, 128 and 4. At these settings, the running time is approximately 5 min.

LMI and CoCoMI For LMI and CoCoMI, we set N (number of neighborhoods) to 50 and M (number of voxels) to 1000. More neighborhoods increases the accuracy and robustness of both methods, but also increases the computation time. We use the MRI for self-similarity estimation since it has a higher quality and is also available in advance. Hence, the running time of LMI and CoCoMI is similar. Our registration has 3 hierarchical levels at 80, 40 and 20 pixel spacing between the B-spline nodes. Note that since our deformation model is different from that of MIND, our pixel spacings are different. The running time is approximately 10 min for both LMI and CoCoMI.

Since US echoes are strong at tissue interfaces, they generally show high correlation with the gradient of MR. As a result, previous work [23], [24] have registered the magnitude of the MR gradient to US. In this work, we present results of US registration to both MR intensity and MR gradient. When using MR gradients, we compute the descriptive histograms from the original MR intensity and not the MR gradient. Regarding SeSaMI, we do not provide the results for both MR intensity and gradients, because SeSaMI utilizes intensities and gradient features for registration. Both qualitative and quantitative results are presented.

D. Qualitative Analysis

Radial gradient versus magnitude of the gradient of the MR. The US probe used in most IGNS systems, including the one used in the BITE database, is a phased-array probe. A phased-array probe has a small head and generates US waves in the radial direction. Therefore, its advantage compared to linear probes is that it can produce a fan-shaped image (see

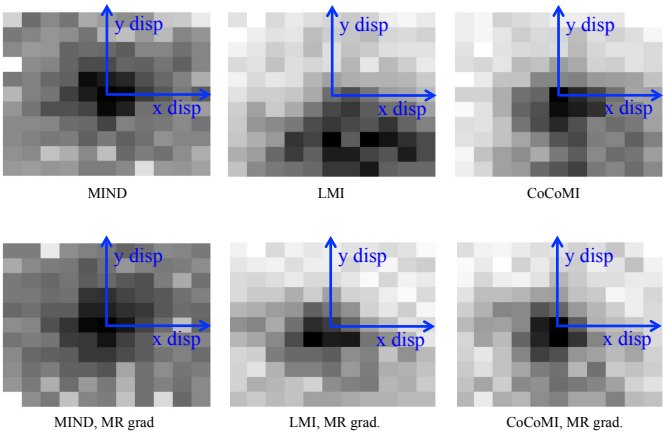


Fig. 21. Qualitative comparison of different similarity metrics by rigidly translating the MR image in the horizontal and vertical directions by ± 5 pixels. In the top and bottom rows, MR intensity and MR gradient magnitude are respectively utilized. The minimum cost is expected to be at the zero displacement. Black represents smaller dissimilarity. All the figures are normalized to [0 1] range to enhance visualization.

Figure 19 (e) and 20) from a small opening. It seems natural to calculate the MR gradient also in the same radial direction. Calculating the radial gradient of the MR is more difficult compared to simply calculating the magnitude of the gradient because it depends on the alignment of the US and MR. To compare the correlations of the radial gradient and gradient magnitude with US, we perform the following experiment.

Using P10 in Table III as a bronze standard registered data, we plot LMI as a function of the rigid translation of the MR in the x y directions by ± 5 pixels (see Figure 19). Here, the radial gradient of the MR can be simply computed since MR and US are registered: we perform a dot product between the gradient vector and the vector that connects that point to the US probe. The results in (f) and (g) show that both the magnitude of the gradient, and the radial component of the gradient can be used to perform image registration. In fact, the gradient magnitude seems to generate a sharper minimum at 0 displacement. The reason, in our opinion, lies in the fact that the 3D US volume is generated from many 2D US images. Both unparalleled imaging planes, and in-plane translation or rotation of the probe cause the radial directions to be different; see Figure 20 for an example of two US images with in-plane probe translation. In fact, this type of averaging reduces the dependence of image on the US insonification angle and is used to improve the quality of US image [52]. We therefore use the magnitude of the MR gradient, which does not need to be recomputed after every iteration.

Cost functions of MIND, LMI and CoCoMI. Using P10 in Table III, we now compare LMI and CoCoMI by looking at their values as a function of rigid translation of the MR in the x y directions. The minimum dissimilarity is expected to be at 0 displacement. The results are shown in Figure 21. We see that MIND performs better using MR gradients, where it gives the smallest dissimilarity at 0. LMI works significantly better with the MR gradient. CoCoMI predicts 0 as the correct alignment with both MR intensity and MR gradient.

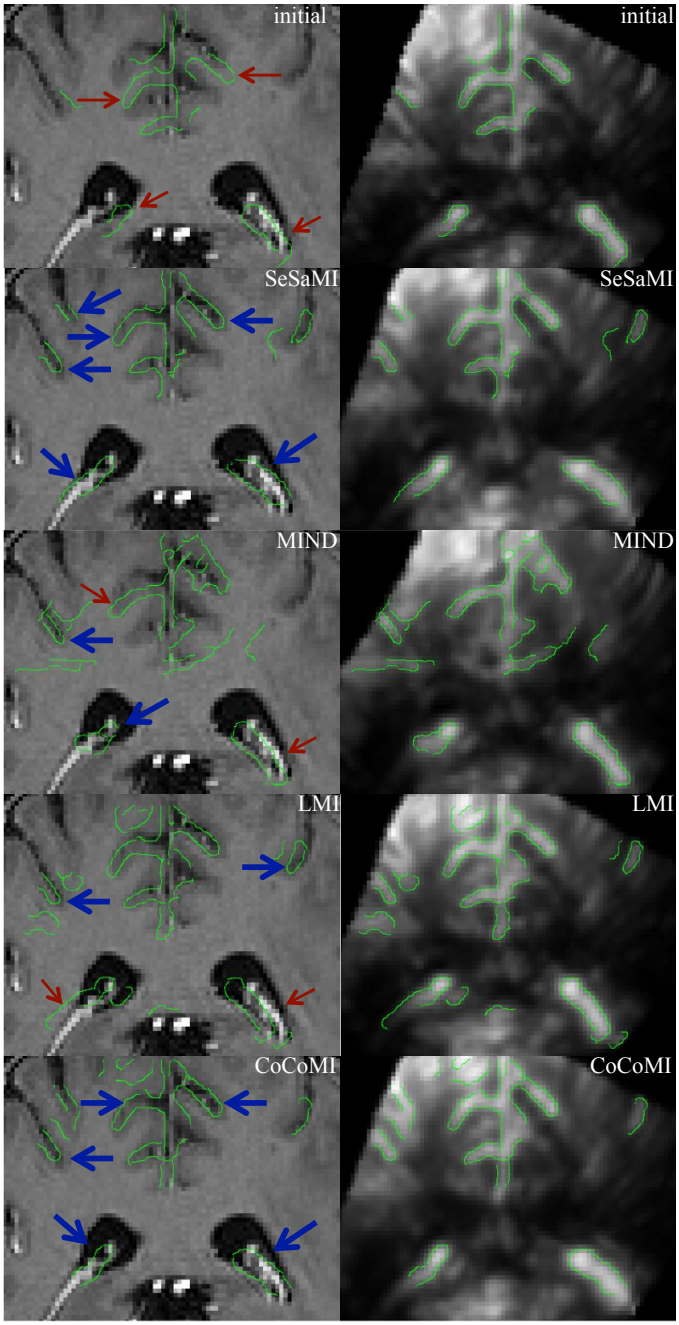


Fig. 22. The alignment quality before and after registration. The underlying MR image is always the same. The US slice is reformatted according the the recovered 3D transformation by the named technique. The edges (in green) are from US. The think blue and thin red arrows respectively point to correct and incorrect alignments.

Visual assessment of the registration results. The next experiment provides visual comparison of the registration results (see Figure 22). The left column is the same MR slice, and the right column is the corresponding US slice after registration. Note that the registration is done in 3D. The thick blue and thin red arrows respectively point to correct and incorrect alignments. The green edge contour is found automatically from the US images, and is overlaid on the MR images to aid assessing the alignment quality. Here, we see

the alignment accuracy improves significantly from initial to LMI, and further from LMI to CoCoMI.

E. Quantitative Analysis

Finally, the mTRE results for all the 13 patients are shown in Table III. The SeSaMI results are from [13]. They show that either MR intensity or MR gradient can be used to drive the non-rigid registration. The last two rows show that both the mean and standard deviation of the mTRE are decreased with non-rigid registration. The average mTRE of MIND with MR intensity and MR gradient (the last two rows of the table) are respectively 3.77 mm and 3.68 mm, showing that MIND significantly reduces the initial mTRE. The p-value of statistical significance of the improvement of CoCoMI over LMI using a paired student t-test are: $p = 0.0004$ for intensity of MR and $p = 0.002$ for gradient of MR. This shows that the CoCoMI is consistently outperforming LMI in terms of mTRE results. While the average mTRE of CoCoMI is similar to that of SeSaMI (2.36 mm versus 2.30 mm) and the difference in the results is not statistically significant, CoCoMI runs 10 times faster. The 10^{th} and 90^{th} percentile values also show that the reduction in the high TRE values is the greatest in CoCoMI: in average, it reduces the 90% value from the initial value of 6.0 mm to 3.9 mm and 3.8 mm using respectively MR intensity and MR gradient. In addition, the results of the last row show that the maximum 90% value for LMI and CoCoMI are 7.1 mm and 4.9 mm respectively, while the initial value is 11.8 mm. This significant reduction indicates the robustness of CoCoMI compared to LMI.

V. Discussions

Many important anatomical references that are well visualized in the pre-operative MR are not visible in the intra-operative US, and therefore fusing MR and US images is of significant clinical value. CoCoMI performs automatic registration of MR and US, while accounting for brain shift. As a result, it reduces the maximum initial mTRE value of 9.38 mm in Table III to 3.22 mm. Without such intra-operative re-registration, surgeons cannot rely on the images presented by standard neuronavigation systems when brain shift is present. CoCoMI significantly improves the alignment between US and MR images, which results in improved confidence in the neuronavigation system, and can potentially reduce surgical time and complications.

We used two complementary histogram descriptors for contextual analysis: the spin image, which is based on intensities, and RIFT, which is based on gradient orientations. We then used the contextual information to constrain the shape of the joint histogram. Richa *et al.* [55] also proposed a similar strategy for tracking surgical tools, where they impose weights on the contribution of different pixels to prevent optimality of undesirable shapes for the joint histogram.

The offline pre-processing step allows us to use local histograms to encode the rich structural self-similarities between two voxel locations to a single number, the EMD distance. Wachinger and Navab [7] generate entropy images also 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. In the first step, our histogram descriptors explicitly contain both intensity and spatial location, while the entropy images, in their standard form, contain only intensities. Wachinger and Navab [7] proposed a spatial 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, which leads to locations that become negligible in the entropy calculation as discussed in [7]. In the second step, 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. Our asymmetric algorithm, which only requires these contextual analysis on one of the images, allows us to perform them offline on the pre-operative image.

Wachinger and Navab [7] showed that the entropy images can be used to transform multi-modal registration of T1, T2, PD, computed tomography (CT) and positron emission tomography (PET) to uni-modal registration. In this work, we showed that entropy images of US and MR have many similarities, but SSD is not enough for registering them. Our experiments showed that US entropy images have many desirable features, such as eliminating the large spatial bias in the US image. In the future, we will explore incorporating speckle statistics in estimation of US entropy images. We will also use robust similarity metrics for registering entropy images of US to entropy images of other modalities.

We successfully used the MIND technique for the first time for registration of US and MR images, and showed that the cost function generated by MIND is generally smooth (Figure 21) even with the high level of noise and attenuation bias in the US volume. The results of Table III show that the initial mTRE values are also significantly reduced with deformable registration using MIND. Comparing the MIND and CoCoMI results, one should also keep in mind that they use different transformation models: MIND uses symmetric diffeomorphic while CoCoMI uses free-form B-splines.

Contextual analysis allowed us to group similar pixels together for MI estimation. This is, to some extent, similar to the popular line of work [23], [24], [25] which segments the MR to simulate a pseudo-US from MR. In another work, Zhuang *et al.* [21] showed that performing Parzen window weighting on the spatial distances is superior to using a box kernel, which is 1 or 0 for respectively spatially close or far pixels. Therefore, an avenue for future work is to further utilize contextual information as follows. We will add self-similarities to image intensities as a third channel of the joint histogram and consider Parzen window weighting in the third channel as well as the first two channels (i.e. the image intensities). This is in contrast with this work where we used box kernel (i.e. value 1 for the most similar, and 0 for others) for the third channel (i.e. the self-similarities) and Parzen windows for the first two channels. Incorporation of tracked US elastography [56] into IGNS is also an area of future work.

TABLE III
TRE AND MTRE RESULTS OF MR/US REGISTRATION. EACH ENTRY IS MEAN ± STD, FOLLOWED BY 10^{th} AND 90^{th} PERCENTILES IN THE BRACKET. ALL NUMBERS IN MM. THE SMALLEST VALUES USING MR INTENSITY OR GRADIENT ARE IN BOLD.

				MR intensity			MR gradient		
case	tags	initial	SeSaMI	MIND	LMI	CoCoMI	MIND	LMI	CoCoMI
P1	35	6.30±1.5 (4.2,8.2)	1.82±1.3 (1.1,5.8)	4.59±1.8 (**2.0**,6.5)	4.05±1.5 (2.9,7.1)	**3.22±1.0** (2.6,**4.6**)	4.42±1.3 (2.7,5.9)	4.61±1.5 (**1.5**,6.0)	**3.47±1.1** (2.7,**4.4**)
P2	40	9.38±1.9 (6.7,11.8)	2.54±1.8 (2.0,5.4)	4.55±2.3 (**1.9**,7.5)	3.74±2.1 (2.1,7.0)	**3.03±1.6** (2.2,**4.9**)	4.65±2.4 (2.2,7.4)	3.86±2.0 (**1.9**,7.0)	**3.27±1.6** (2.1,**4.8**)
P3	32	3.93±1.1 (2.8,5.3)	1.96±1.0 (1.0,3.5)	2.48±1.2 (1.1,4.2)	2.49±1.1 (**1.0**,4.1)	**2.17±1.0** (1.2,**3.9**)	2.51±1.3 (0.9,4.3)	2.94±**1.1** (0.9,**4.0**)	**2.81**±1.2 (**0.8**,4.1)
P4	31	2.62±1.4 (1.0,4.0)	2.59±1.1 (1.0,2.9)	2.34±1.1 (1.1,3.6)	2.72±1.3 (**0.9**,4.0)	**2.18±1.1** (0.9,**2.9**)	2.31±1.2 (1.4,3.3)	2.58±**1.0** (**1.4**,**3.0**)	**2.47**±1.0 (1.4,3.1)
P5	37	2.30±1.0 (0.8,3.4)	1.73±0.9 (0.8,3.1)	4.17±1.6 (2.6,7.0)	2.23 ±1.1 (**0.7**,**3.2**)	**2.20±0.9** (0.9,3.2)	4.85±1.8 (2.7,7.6)	2.44±1.0 (0.8,3.3)	**2.11±0.9** (**0.7**.**2.8**)
P6	19	3.04±1.6 (1.2,5.3)	1.94±1.2 (1.2,3.9)	2.34±1.1 (1.2,4.3)	2.51±1.2 (**1.1**,4.4)	**2.13±1.0** (1.1,**3.6**)	2.65±1.2 (1.1,4.2)	2.34 ±1.0 (**1.2**,**3.8**)	**1.70±1.0** (1.3,3.8)
P7	23	3.75±2 (1.2,5.9)	2.91±1.3 (1.3,4.5)	5.35±2.2 (2.6,8.4)	2.56±1.7 (**1.4**,4.9)	**2.00 ±1.0** (1.6,**3.6**)	5.46±2.1 (2.8,8.8)	**2.11**±1.4 (**1.2**,4.0)	2.24±**1.3** (1.2,**3.7**)
P8	21	5.09±1.4 (3.1,6.9)	2.52±1.1 (1.6,4.9)	5.55±1.9 (3.1,8.1)	2.49 ±1.1 (**2.1**,5.0)	**2.18 ±1.1** (2.2,**4.1**)	4.86±1.9 (2.4,7.6)	2.41±**1.0** (2.9,4.9)	**2.05**±1.0 (**1.5**,**4.4**)
P9	25	3.00±1.4 (1.4,5.0)	2.74±1.4 (1.9,5.4)	3.56±2.1 (1.6,6.4)	2.84±**1.4** (**1.3**,**4.7**)	**2.04**±1.6 (1.3,4.9)	2.96±2.2 (0.7,6.5)	2.91± 1.9 (**1.0**,5.0)	**2.41±1.6** (1.1,**4.6**)
P10	25	1.52±0.7 (0.8,2.9)	1.35±0.8 (0.9,3.3)	2.45±1.1 (1.1,4.2)	3.94±1.4 (1.0,3.9)	**2.48±1.0** (**1.0**,**2.9**)	2.48±0.9 (1.1,3.6)	2.19± **0.9** (**0.9**,**2.8**)	**2.12**±0.9 (1.0,2.8)
P11	21	3.70±1.9 (1.1,6.3)	2.78±1.6 (1.1,5.0)	2.73±1.6 (1.2,4.3)	2.29±**1.1** (1.2,**2.9**)	**2.16**±1.3 (**0.9**,3.0)	2.69±1.4 (1.1,4.3)	2.45±1.2 (1.2,4.2)	**1.90±1.1** (**1.1**,**2.9**)
P12	23	5.15±2.8 (2.0,8.6)	2.91±1.9 (1.9,4.8)	4.41±2.0 (2.3,6.8)	2.67±**1.7** (**2.1**,**5.0**)	**2.64** ±1.7 (2.2,5.0)	4.21±1.9 (2.3,6.2)	2.31±1.6 (2.0,5.0)	**2.10±1.0** (**2.0**,**3.9**)
P13	23	3.78±1.2 (2.0,5.0)	2.16±1.1 (2.1,4.1)	4.53±1.4 (3.1,6.6)	2.90±1.3 (2.9,5.1)	**2.07 ±1.0** (**2.1**,**4.2**)	3.85±1.3 (2.6,5.7)	2.25±1.4 (2.1,5.1)	**2.03±0.9** (**2.0**,**3.9**)
mean	27	4.12±1.5 (2.2,6.0)	2.30±1.3 (1.4,4.4)	3.77±1.6 (1.9,6.0)	2.88±1.4 (1.6,4.7)	**2.35±1.2** (**1.5**,**3.9**)	3.68±1.8 (1.8,5.8)	2.72±1.3 (1.5,4.5)	**2.36±1.1** (**1.5**,**3.8**)
std	6.9	2.03±0.5 (1.7,2.4)	0.52±0.3 (0.4,1.0)	1.18±0.4 (0.7,1.6)	0.62±0.3 (0.7,1.2)	**0.39±0.3** (**0.6**,**0.9**)	1.12±0.5 (0.8,1.5)	0.73±0.4 (0.6,1.2)	**0.53±0.3** (**0.5**,**0.9**)
max	40	9.38±2.8 (6.7,11.8)	2.91±1.8 (2.5,5.4)	5.55±2.3 (3.1,8.4)	4.05±2.1 (2.9,7.1)	**3.22±1.7** (**2.6**,**4.9**)	5.46±2.2 (2.7,8.8)	4.61±2.0 (2.9,7.0)	**3.47±1.6** (**2.7**,**4.7**)

VI. CONCLUSION

We presented CoCoMI, a similarity metric that conditions MI estimation on contextual information. We utilized ideas from pattern recognition to develop a robust contextual analysis technique, and used it as an offline step so that the on-line registration time changes negligibly. Using simulation and patient data, we showed that CoCoMI significantly outperforms state of the art non-rigid registration methods. CoCoMI is easily parallelizable; we will investigate its GPU implementation in the future. We will also add the contextual information as a third channel to joint intensity histogram and use Parzen windowing in all three dimensions to estimate the joint probabilities. Finally, we plan to apply CoCoMI to other problems such as MR to CT and CT to US image registration.

ACKNOWLEDGEMENTS

The authors would like to thank anonymous reviewers for their constructive feedback, Dr. David Araujo for checking the accuracy of the landmark selection of the BITE database, Dr. Pierrick Coupé and Dr. Tal Arbel for valuable discussions, and Dr. Christian Wachinger for kindly providing his implementation of entropy estimation for 3D volumes. We used the Intraoperative Brain Imaging System (IBIS) software developed by Simon Drouin and Anka Kochanowska to generate Figure 16. This work was financed by the Fonds Québécois de la recherche sur la nature et les technologies, the Canadian Institute of Health Research (MOP-97820), and the Natural Science and Engineering Research Council of Canada. H. Rivaz is supported by a Natural Science and Engineering Research Council of Canada post-doctoral fellowship.

REFERENCES

[1] W. Wells, P. Viola, H. Atsumid, S. Nakajimae, and R. Kikinis, "Multi-modal volume registration maximization of mutual information," *Med Imag Anal*, vol. 1, no. 1, pp. 35–51, 1996.

[2] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med Imag*, vol. 16, no. 2, pp. 187–198, 1997.

[3] J. Pluim, J. Maintz, and M. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Trans. Medical Imag.*, vol. 22, no. 8, pp. 986–1004, 2003.

[4] ——, "Image registration by maximization of combined mutual information and gradient information," *IEEE Trans Med Imag*, vol. 19, no. 8, pp. 809–814, 2000.

[5] B. Karaçali, "Information theoretic deformable registration using local image information," *Int. J. Comp. Vis.*, vol. 72, no. 3, pp. 219–237, 2007.

[6] D. D. Nigris, D. L. Collins, and T. Arbel, "Multi-modal image registration based on gradient orientations of minimal uncertainty," *IEEE Trans. Medical Imag*, vol. 30, pp. 2343–2354, 2012.

[7] C. Wachinger and N. Navab, "Entropy and Laplacian images: Structural representations for multi-modal registration," *Medical Imag Analysis*, vol. 16, no. 1, pp. 1–17, 2012.

[8] M. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. Gleeson, M. Brady, and J. Schnabel, "MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration," *Medical Image Analysis*, vol. 16, no. 7, pp. 1423–1435, 2012.

[9] C. Wachinger and N. Navab, "A contextual maximum likelihood framework for modeling image registration," *Computer Vision Pattern Recognition (CVPR)*, pp. 1995–2002, 2012.

[10] M. Sabuncu and P. Ramadge, "Using spanning graphs for efficient image registration," *IEEE Trans. Imag. Proc.*, vol. 17, no. 5, pp. 788–797, 2008.

[11] M. Staring, U. A. van der Heide, S. Klein, M. A. Viergever, and J. Pluim, "Registration of cervical mri using multifeature mutual information," *IEEE Trans. Medical Imag.*, vol. 28, no. 9, pp. 1412–1421, 2009.

[12] E. Oubel, M. Craene, A. Hero, and A. Frangi, "Cardiac motion estimation by joint alignment of tagged MRI sequences," *Med. Imag. Anal.*, vol. 16, pp. 339–350, 2012.

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

[14] D. Rueckert, M. Clarkson, M. Hill, and D. Hawkes, "Non-rigid registration using higher-order mutual information," *SPIE*, pp. 438–447, 2000.

[15] D. Russakoff, C. Tomasi, T. Rohlfing, and C. Maurer, "Image similarity using mutual information of regions," *ECCV, LNCS*, pp. 596–607, 2004.

[16] Z. Yi and S. Soatto, "Multimodal registration via spatial-context mutual information," *IPMI*, pp. 569–580, 2007.

[17] C. Studholme, C. Drapaca, B. Iordanova, and V. Cardenas, "Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change," *IEEE Trans. Medical Imag*, vol. 25, no. 5, pp. 626–639, 2006.

[18] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens, "Nonrigid image registration using conditional mutual information," *IEEE Trans. Medical Imag*, vol. 29, no. 1, pp. 19–29, 2010.

[19] S. Klein, U. A. van der Heide, I. Lips, M. van Vulpen, M. Staring, and J. Pluim, "Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information," *Med. Phys.*, vol. 35, no. 4, pp. 1407–1417, 2008.

[20] S. Klein, M. Staring, and J. Pluim, "Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines," *IEEE Imag Proc*, vol. 16, no. 12, pp. 2879–2890, 2007.

[21] S. Zhuang, D. Arridge, D. Hawkes, and S. Ourselin, "A nonrigid registration framework using spatially encoded mutual information and free-form deformations," *IEEE Trans. Medical Imag*, vol. 30, no. 10, pp. 1819–1828, 2011.

[22] D. L. Collins, A. Zijdenbos, V. Kollokian, J. Sled, N. Kabani, C. Holmes, and A. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Trans. Medical Imag*, vol. 17, no. 3, pp. 463–468, 1998.

[23] T. Arbel, X. Morandi, R. Comeau, and D. L. Collins, "Automatic non-linear MRI-ultrasound registration for the correction of intra-operative brain deformations," *Compt Aided Surg*, vol. 9, no. 4, pp. 123–136, 2004.

[24] L. Mercier, V. Fonov, C. Haegelen, R. Maesstro, K. Petrecca, and D. L. Collins, "Comparing two approaches to rigid registration of three-dimensional ultrasound and magnetic resonance images for neurosurgery," *Compt Aided Surg*, vol. 7, no. 1, pp. 125–136, 2012.

[25] M. Kuklisova-Murgasova, A. Cifor, N. Napolitano, A. Papageorghiou, G. Quaghebeur, A. Noble, and J. Schnabel, "Registration of 3D fetal brain us and MRI," *Medical Image Computing Computer Assisted Intervention (MICCAI)*, pp. 667–674, 2012.

[26] L. Mercier, R. Maestro, K. Petrecca, D. Araujo, C. Haegelen, and D. L. Collins, "Online database of clinical MR and ultrasound images of brain tumors," *Medical Physics*, 2012.

[27] A. Buades, B. Coll, and J. Morel, "A non-local algorithm for image denoising," *CVPR*, pp. 149–157, 2005.

[28] P. Coupe, S. Eskildsen, J. Manjon, V. Fonov, and D. L. Collins, "Simultaneous segmentation and grading of hippocampus for patient classification with Alzheimer's disease," *MICCAI*, pp. 149–157, 2011.

[29] E. Shechtman and M. Irani, "Matching local self-similarities across images and videos," *CVPR*, pp. 1–8, 2007.

[30] A. Roche, X. Pennec, G. Malandain, and N. Ayache, "Rigid registration of 3-d ultrasound with MR images: a new approach combining intensity and gradient information," *IEEE Trans Med Imag*, vol. 20, no. 10, pp. 1038–1049, 2001.

[31] D. L. Collins, A. Zijdenbos, A. Baare, and A. Evans, "ANIMAL+INSECT: improved cortical structure segmentation," *Inform Proc Med Imag*, vol. 1613, pp. 210–223, 1999.

[32] G. Penney, J. Blackall, M. Hamady, T. Sabharwal, A. Adam, and D. Hawkes, "Registration of freehand 3D ultrasound and magnetic resonance liver images," *Med Imag Anal*, vol. 8, no. 1, pp. 81–91, 2004.

[33] S. Ji, Z. Wu, A. Hartov, D. Roberts, and K. Paulsen, "Mutual-information-based image to patient re-registration using intraoperative ultrasound in image-guided neurosurgery," *Med. Phys.*, vol. 35, no. 10, pp. 4612–4624, 2008.

[34] W. Zhang, M. Brady, H. Becher, and A. Noble, "Spatio-temporal (2D+T) non-rigid registration of real-time 3D echocardiography and cardiovascular MR image sequences," *Physics Med Biol*, vol. 56, pp. 1341–1360, 2011.

[35] M. P. Heinrich, M. Jenkinson, B. W. Papież, M. Brady, and J. A. Schnabel, "Towards realtime multimodal fusion for image-guided interventions using self-similarities," in *MICCAI*, 2013, pp. 187–194.

[36] W. Wein, A. Ladikos, B. Fuerst, A. Shah, K. Sharma, and N. Navab, "Global registration of ultrasound to mri using the LC2 metric for enabling neurosurgical guidance," in *MICCAI*, 2013, pp. 34–41.

[37] H. Rivaz, E. Boctor, M. Choti, and G. Hager, "Real-time regularized ultrasound elastography," *IEEE Trans Med Imaging*, vol. 30, no. 4, pp. 928–945, Apr 2011.

[38] H. Rivaz, E. Boctor, P. Foroughi, G. Fichtinger, and G. Hager, "Ultrasound elastography: a dynamic programming approach," *IEEE Trans Med Imaging*, vol. 27, no. 10, pp. 1373–1377, Oct. 2008.

[39] J. Kybic and M. Unser, "Fast parametric elastic image registration," *IEEE Trans. Med. Imag.*, vol. 12, pp. 1427–1442, 2003.

[40] M. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. Gleeson, M. Brady, and J. Schnabel, "Non-local shape descriptor: a new similarity metric for deformable multi-modal registration," *MICCAI*, pp. 541–548, 2011.

[41] S. Grewenig, S. Zimmer, and J. Weickert, "Rotationally invariant similarity measures for nonlocal image denoising," *J. Vis. Commun. Imag. R. (JVCI)*, vol. 22, pp. 117–130, 2011.

[42] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Trans Pattern Anal Machine Int*, vol. 27, no. 8, pp. 1265–1278, 2005.

[43] Y. Rubner, C. Tomasi, and L. Guibas, "The earth mover's distance as a metric for image retrieval," *IEEE Trans Pattern Anal Machine Int.*, vol. 40, no. 2, pp. 99–121, 2000.

[44] H. Ling and K. Okada, "An efficient earth mover's distance algorithm for robust histogram comparison," *IEEE Trans Pattern Anal Machine Int.*, vol. 29, no. 5, pp. 840–853, 2007.

[45] A. Andronache, M. V. Siebenthal, G. Szekely, and P. Cattin, "Non-rigid registration of multi-modal images using both mutual information and cross-correlation," *Medical Imag Analysis*, vol. 12, no. 1, pp. 3–15, 2008.

[46] G. Hager and P. Belhumeur, "Efficient region tracking with parametric models of geometry and illumination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 10, pp. 1025–1039, Oct. 1998.

[47] S. Baker and I. Matthews, "Lucas-Kanade 20 years on: A unifying framework," *Inter. J. Comput. Vis.*, vol. 56, no. 3, pp. 221–255, 2004.

[48] C. Wachinger and N. Navab, "Simultaneous registration of multiple images: Similarity metrics and efficient optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 5, pp. 1221–1233, May 2013.

[49] H. Rivaz, E. M. Boctor, M. A. Choti, and G. D. Hager, "Ultrasound elastography using multiple images," *Medical Image Analysis*, no. 0, pp. –, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361841513001709

[50] H. Rivaz, E. Boctor, and G. Fichtinger, "A robust meshing and calibration approach for sensorless freehand 3d ultrasound," in *SPIE Medical Imaging*, 2007, pp. 651 318–651 326.

[51] H. Rivaz, H. J. Kang, P. J. Stolka, R. Zellars, F. Wacker, G. Hager, and E. Boctor, "Novel reconstruction and feature exploitation techniques for sensorless freehand 3d ultrasound," in *SPIE Medical Imaging*, 2010, pp. 76 291D–76 291D.

[52] R. Rohling, A. Gee, and L. Berman, "Three-dimensional spatial compounding of ultrasound images," *Medical Image Analysis*, vol. 1, no. 3, pp. 177–93, 1997.

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

[54] B. Avants, C. Epstein, M. Grossman, and J. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, 2008.

[55] R. Richa, M. Balicki, E. Meisner, R. Sznitman, R. Taylor, and G. Hager, "Visual tracking of surgical tools for proximity detection in retinal surgery," in *Information Processing in Computer-Assisted Interventions*. Springer, 2011, pp. 55–66.

[56] H. Rivaz, P. Foroughi, I. Fleming, R. Zellars, E. Boctor, and G. Hager, "Tracked regularized ultrasound elastography for targeting breast radiotherapy," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*. Springer, 2009, pp. 507–515.