

Received November 1, 2013; accepted December 15, 2013. Date of publication 9 January 2014;  
date of current version 3 February 2014.

Digital Object Identifier 10.1109/JTEHM.2014.2299280

# Spatial Mutual Information as Similarity Measure for 3-D Brain Image Registration

QOLAMREZA R. RAZLIGHI<sup>1</sup> (Senior Member, IEEE),  
AND NASSER KEHTARNAVAZ<sup>2</sup> (Fellow, IEEE)

<sup>1</sup>Department of Biomedical Engineering and Neurology, Columbia University, New York, NY 10032, USA

<sup>2</sup>Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080, USA

CORRESPONDING AUTHOR: Q. R. RAZLIGHI (qr2108@columbia.edu)

This work was supported by the U.S. National Institutes of Health – National Institute of Aging  
T32 under Grant AG000261 and K01 under Grant AG044467.

**ABSTRACT** Information theoretic-based similarity measures, in particular mutual information, are widely used for intermodal/intersubject 3-D brain image registration. However, conventional mutual information does not consider spatial dependency between adjacent voxels in images, thus reducing its efficacy as a similarity measure in image registration. This paper first presents a review of the existing attempts to incorporate spatial dependency into the computation of mutual information (MI). Then, a recently introduced spatially dependent similarity measure, named spatial MI, is extended to 3-D brain image registration. This extension also eliminates its artifact for translational misregistration. Finally, the effectiveness of the proposed 3-D spatial MI as a similarity measure is compared with three existing MI measures by applying controlled levels of noise degradation to 3-D simulated brain images.

**INDEX TERMS** Spatial mutual information, spatially dependent similarity measures, brain image registration, spatial entropy.

## I. INTRODUCTION

Registration is a key image processing component in brain image studies. Automatic brain image registration methods predominantly rely on information theoretic-based similarity measures to avoid the time-consuming and subjective process of manual extraction of landmarks, or features and their alignment. Mutual information (MI), based on Shannon's definition of entropy, is a widely utilized similarity measure for intermodal and/or intersubject 3D brain image registration. MI was originally introduced for image registration by Viola [1] and Maes [2]. Despite its widespread use, it has been shown that the use of MI can result in misregistrations and there is room for improvement [3]–[5].

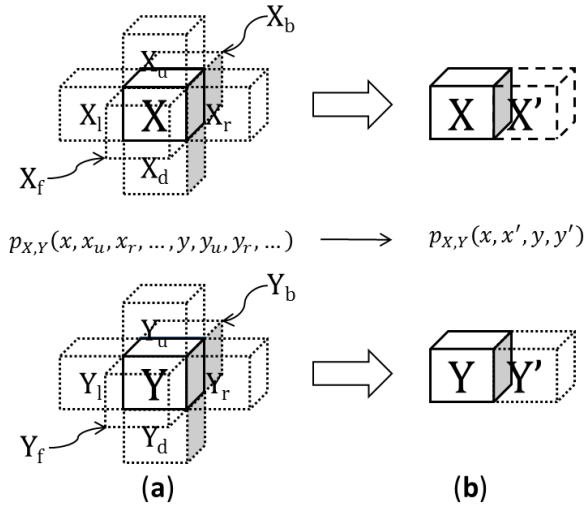
MI computation has been conventionally done based on a global spatial independency assumption over the entire image. This underlying assumption means that there is no statistical relationship among neighboring voxels which is strongly violated in most medical images; this shortcoming was recognized soon after MI was introduced. Studholme *et al.* [3] attempted to enhance the effectiveness of MI by incorporating spatial dependency into its computation,

but the method resulted in limited improvement. Since then, researchers have tried to find new ways by which inter-voxel dependency can be taken into consideration toward computing MI.

Conceptually, the conventional MI provides image similarity based on a single voxel correspondence. However, MI with spatial dependency takes into consideration correspondences of multiple adjacent voxels. Spatially dependent MI is thus more robust to image degradation and consequently provides more accurate image registration. This advantage is the main motivation behind the attempts to incorporate spatial dependency into the computation of MI, but these attempts need to overcome the dimensionality problem when computing similarity across multiple spatially dependent voxels.

Fig. 1(a) illustrates the configuration of spatially dependent voxels in the lowest possible order of the neighboring structure. Even for the lowest order neighboring structure, the high dimensionality of the problem prevents any direct approach from obtaining a tractable solution. The volume of recent works on spatially dependent similarity measures (listed in Section II) is indicative of the keen interest among researchers

in the brain imaging field to incorporate spatial information into useful similarity measures such as *MI*. However, most of the spatially dependent similarity measures, reviewed in Section II, are either an ad-hoc combination of *MI* with various image features to capture the image spatial information or a heuristic use of different definitions of entropy instead of the conventional Shannon entropy.



**FIGURE 1.** Nearest neighbor voxels configuration and joint distribution for (a) anisotropic and (b) isotropic random fields.

The recently introduced spatial mutual information (*SMI*) in [6] provides a method for computing spatially dependent *MI* while addressing the dimensionality problem by applying the Markovianity constraint. A measure that comes closest to *SMI* is the second-order *MI* introduced by Rueckert *et al.* [4].

In Section III, we first describe the existing artifact for translational misregistration in the recently introduced similarity measure *SMI*. Then, we introduce our method for extending this similarity measure to 3D brain image registration, making *SMI* a viable alternative to *MI*. We also demonstrate that this extension removes the aforementioned artifact for translational misregistration. In section IV, we compare the effectiveness of the proposed 3D *SMI* with *MI*, *SOMI*, and *SMI* similarity measures for 3D brain image registration using simulated T1 and T2 weighted images while applying different levels of image degradation. Finally, we discuss future avenues of this work and conclude the paper in section V.

## II. SPATIALLY DEPENDENT SIMILARITY MEASURES

The aim of this section is to review some of the existing attempts to incorporate spatial dependency into the computation of mutual information, or other information theoretic-based similarity measures. The recent upsurge in the volume of research in spatially dependent similarity measures indicates the importance of this work in the field, and the immediate need for a more robust and effective spatially dependent similarity measure. Fig. 1(a) shows the structure of the nearest neighboring voxels for a pair of 3-dimensional images. Even

though the nearest neighbor structure is the lowest order in the neighboring structure of the Markov random fields, it results in a 14-dimensional joint probability, which is required for the computation of *MI* with spatial dependency. Such a high dimensionality makes any direct approach to compute the joint distribution an intractable problem. The methods reviewed in this section are mainly simplification approaches applied to joint distribution to make the computation of *MI* tractable. We begin with *homogeneity* and *isotropy* as the most common simplifying assumptions.

Second Order Mutual Information (*SOMI*) might be considered the most straightforward extension of *MI* to incorporate spatial dependency under both homogeneity and isotropy assumptions. It involves the use of the co-occurrence, or Aura matrices to estimate the four-dimensional joint probability density function (pdf) of an image pair [4]. This measure is given by

$$SOMI = \sum_{x, x' \in X} \sum_{y, y' \in X} p_{X,Y}(x, x', y, y') \log \frac{p_{X,Y}(x, x', y, y')}{p_X(x, x') p_Y(y, y')} \quad (1)$$

where  $X$  denotes a finite discrete label set,  $p_X(x, x')$  the probability that  $x$  and  $x'$  are adjacent in image  $\mathbf{X}$ ,  $p_Y(y, y')$  the probability that  $y$  and  $y'$  are adjacent in image  $\mathbf{Y}$ ,  $p_{X,Y}(x, x', y, y')$  the joint probability that  $(x, x')$  are adjacent in image  $\mathbf{X}$  and  $(y, y')$  are adjacent in image  $\mathbf{Y}$ , and  $(x, y)$  denotes a corresponding voxel pair.

Unfortunately, in practice the four-dimensional joint histogram for estimating  $p_{X,Y}(x, x', y, y')$  becomes sparse, since a typical brain image contains insufficient data samples to adequately fill its bins. In [4], Rueckert addressed this issue by reducing the size of the discrete label set to 16. However, this reduces the effectiveness of *SOMI* as a similarity measure; this drawback is thoroughly studied by Gao [7] for the classical *MI*, and also in [8] for *SOMI*.

In addition, *SOMI* is an *isotropic* measure, meaning that there is no sense of directionality in the adjacent voxels. In other words, all six voxels in the nearest neighboring structure are treated the same and without direction. As shown in Fig. 1, the 14-dimensional joint distribution for a pair of anisotropic fields is simplified to 4-dimensional joint distribution for a pair of isotropic fields. This simplification is essential for the validity of (1), but it reduces the sensitivity of the similarity measure to changes along different directions.

Gradient Mutual Information (*GMI*) is a spatial similarity measure that is formed by combining *MI* and a gradient measure [9]. *GMI* is formulated as follows:

$$GMI = G(X, Y)I(X, Y) \quad (2)$$

$$G(X, Y) = \sum_{(x, y) \in (X, Y)} w(\alpha_{x, y}(\sigma)) \min(|\nabla x(\sigma)|, |\nabla y(\sigma)|)$$

$$w(\alpha) = \frac{\cos(2\alpha) + 1}{2} \quad \&\alpha_{x, y}(\sigma) = \arccos \frac{\nabla x(\sigma) \cdot \nabla y(\sigma)}{|\nabla x(\sigma)| |\nabla y(\sigma)|}$$

where  $G(X, Y)$  is the gradient part of the similarity measure contributing to spatial information,  $|\nabla x(\sigma)|$  denotes the mag-

nitude of the gradient vector of image  $\mathbf{X}$  at point  $x$  with the scale of  $\sigma$ ,  $|\nabla_x(\sigma)|$  is the magnitude of the gradient vector of image  $\mathbf{Y}$  at point  $y$  with the scale of  $\sigma$ , and  $I(X, Y)$  is the conventional mutual information. The registration outcome when using this similarity measure has shown some improvement for multimodal affine registration in comparison to the conventional  $MI$  [9].

The maximum distance gradient magnitude for capturing image spatial information is another approach discussed in [10]. In this approach,  $MI$  was obtained from a 4-dimensional joint histogram of two images and corresponding maximum distance gradient magnitudes. However, the lack of available data samples to fill the histogram bins still poses a challenge.

Shen *et al.* [11] developed a similarity measure that determines image similarities based on an attribute vector for each voxel including gray matter, white matter, and cerebral spinal fluid interfaces. This similarity measure was specifically devised for inter-modal, inter-subject magnetic resonance brain image registration requiring the segmentation of different tissue types.

Mutual information of regions was introduced by Russakoff *et al.* [12]. In this approach, a vector of intensity values is created for every voxel in the image. The components of these vectors involve the respective intensity values of the neighboring voxels. These vectors form a matrix in which the rows are assumed to be normally distributed. Thus, the entropy is computed directly from the determinant of the covariance matrix. The entropy of the multivariate normal distribution is given by

$$H(\mathbf{Z}) = \frac{1}{2} \log |\Sigma_{\mathbf{Z}}| + \frac{d}{2} \log(2\pi e) \quad (3)$$

where  $\mathbf{Z}$  is the set of multivariate normal random variables,  $|\Sigma_{\mathbf{Z}}|$  is the determinant of the covariance matrix of  $\mathbf{Z}$ , and  $d$  denotes the dimensionality of  $\mathbf{Z}$ . Another similar attempt was reported in [13] in which the same matrix, with voxel intensity values and average intensity values of the neighboring voxels, was used. Although assuming a normal distribution for each row of the matrix allows estimation of its entropy, the effect of such a substandard assumption has not been studied.

In [14], the image joint histogram was computed along corresponding points on random lines instead of the usual grid pattern. Since the random lines align along any orientation, the spatial information is captured to some degree. However, the outcome depends strictly on the number of points on the lines. Even with a minimal number (2 points), a 4-dimensional histogram is required to be populated which leads to the same scarcity or dimensionality problem as mentioned earlier.

There have also been a number of attempts to compute image spatial mutual information based on multi-feature mutual information [15]–[17]. In these approaches, different features of an image are used to capture the image spatial information instead of incorporating neighboring

voxels information. These approaches still suffer from the dimensionality problem. One solution is to consider the feature probability distribution as normal and obtain the joint entropy directly from the covariance matrix by (3). A reliable estimation of the normal distribution using the covariance matrix requires much fewer samples, yet the error in estimating the feature distribution this way has not been studied. The feature extraction is also another issue when using these methods since this process is often done in an ad hoc manner and there is no systematic way of obtaining the most appropriate features.

Another similarity measure, called  $\alpha$ - $MI$ , was introduced by Hero *et al.* [18], and was later applied to 2D data [19]. The reported results indicated that there was not a significant improvement over the conventional  $MI$  despite the higher computational complexity. Furthermore, the algorithm requires various manual interventions to make the proposed  $\alpha$ - $MI$  a suitable similarity measure in practice for an automatic registration procedure.

The definition of Quantitative-Qualitative  $MI$  ( $QMI$ ) was introduced in [20] and further developed in [21].  $QMI$  is created by adding a utility coefficient into the formulation of the conventional  $MI$  as follows:

$$QMI = \sum_{x \in X} \sum_{y \in Y} U(x, y) p_{X,Y}(x, y) \log \frac{p_{X,Y}(x, y)}{p_X(x) p_Y(y)} \quad (4)$$

where  $p_X(x)$ , and  $p_Y(y)$  are the image intensity distributions obtained from the histograms and  $p_{X,Y}(x, y)$  is the joint distribution of the images  $\mathbf{X}$  and  $\mathbf{Y}$  under voxel independency assumption. The coefficient  $U(x, y)$  is meant to incorporate the spatial information into this measure. In essence,  $U(x, y)$  is an ad-hoc combination of the saliency measure in [22] and the image gradient. Since an optimization process is required for every voxel of an image in order to compute this measure, its computational complexity is quite high. Even though the reported results indicated some improvements over the conventional  $MI$ , it is not practically useable due to its high computational complexity.

There have also been attempts to use different definitions of entropy instead of the classical Shannon entropy [23], as well as other information theoretic measures such as Kullback–Leibler Distance [24], entropy correlation coefficients [25] and normalized mutual information [26]. However, the concept of incorporating spatial information is not taken into consideration. For instance, so-called Jumarie entropy has been used in [27] to define a similarity measure. The expression for the joint entropy in that method resembles the normalized entropy of the absolute difference image. Yet the spatial information is still not taken into consideration, thus they are omitted from this review. Next we describe the spatial mutual information definition and its extension to 3D images.

### III. SPATIAL MUTUAL INFORMATION

#### A. EXISTING DEFINITION FOR 2D IMAGE

$SOMI$  was the first systematic attempt to incorporate the image voxel spatial dependency into the computation

of image mutual information which was based on two simplifying assumptions: homogeneity and isotropy [4]. Despite the inefficiency introduced by the isotropy assumption, the dimensionality of the problem still prevented its full utilization in brain image registration. Markov processes, on the other hand, have facilitated handling high-dimensionality problems under anisotropic conditions. For example, there is a well-established approach using Markov random fields (MRFs) in image modeling [28]. In [6], we made the first attempt to compute image spatial information under the MRF constraint, which is a more relaxed constraint than the independency constraint. In this approach, a causal MRF model, called Quadrilateral Markov Random Field (QMRF), was used to compute image spatial information under the definition of Shannon entropy. The spatial entropy defined in [6] was further simplified for homogenous but anisotropic QMRF in [29] for nearest neighboring structures as per the following equation:

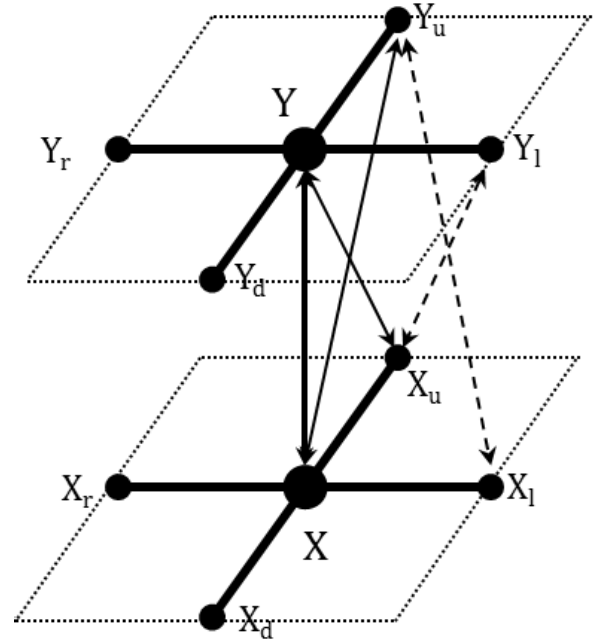
$$H(\mathbf{X}) = mn(H(X, X_u) + H(X, X_l) - H(X)) - \frac{mn}{2}(H(X_u, X_l) + H(X_u, X_r)) \quad (5)$$

where  $m \times n$  denotes the image size,  $H(X, X_u)$  the joint entropy of a voxel with its upper neighbor,  $H(X, X_l)$  the joint entropy of a voxel with its left neighbor,  $H(X_u, X_l)$  the joint entropy of the left and upper neighbors, and  $H(X_u, X_r)$  the joint entropy of the right and upper neighbors; see Fig. 2. Consequently, the computation of spatial mutual information (SMI) was done from the spatial joint entropy as follows [6]:

$$SMI = -mnH(X, Y) + \frac{mn}{2} \left\{ \begin{array}{l} H(X, Y_u) + H(X_u, Y) \\ +H(X, Y_l) + H(X_l, Y) \end{array} \right\} - \frac{mn}{4} \left\{ \begin{array}{l} H(X_u, Y_l) + H(X_l, Y_u) \\ +H(X_u, Y_r) - H(X_r, Y_u) \end{array} \right\} \quad (6)$$

where  $H(X, Y)$  denotes the joint entropy of the voxel  $X$  in image  $\mathbf{X}$  with the corresponding voxel in image  $\mathbf{Y}$ ,  $H(X_u, Y)$  the joint entropy of the voxel  $Y$  in image  $\mathbf{Y}$  with the upper neighbor of its corresponding voxel in image  $\mathbf{X}$  with its counterpart as  $H(X, Y_u)$ ,  $H(X_l, Y)$  denotes the joint entropy of the voxel  $Y$  in image  $\mathbf{Y}$  with the left neighbor of its corresponding voxel  $X$  in image  $\mathbf{X}$  with  $H(X, Y_l)$  as its counterpart,  $H(X_u, Y_l)$  the joint entropy of the left neighbor of voxel  $Y$  in image  $\mathbf{Y}$  with the upper neighbor of the corresponding voxel  $X$  in image  $\mathbf{X}$  with  $H(X_l, Y_u)$  as its counterpart, and finally  $H(X_u, Y_r)$  the joint entropy of the right neighbor of voxel  $Y$  in image  $\mathbf{Y}$  with the upper neighbor of the corresponding voxel  $X$  in image  $\mathbf{X}$  with  $H(X_r, Y_u)$  as its counterpart. Fig. 2 illustrates the configuration of  $H(X, Y)$  and two other joint entropies ( $H(X, Y_u)$  in solid lines and  $H(X_u, Y_l)$  in dashed lines) and their counterpart structures for a pair of 2-dimensional images.

An observation made in (5) and (6) is that they do not contain all the possible joint entropies of the cliques in the first order MRF, as shown in Fig. 2. For instance, the joint entropy  $H(X, X_u)$  is included but not  $H(X, X_d)$ . This is due



**FIGURE 2.** Illustration of sample 2D joint entropies in the definition of spatial mutual information.

to the homogeneity assumption in the computation of SMI in [6]. In other words, from an implementation standpoint, the joint histograms of  $(X, X_u)$  and  $(X, X_d)$  are the same for homogeneous random fields. Therefore, their joint entropies would also be the same in a homogeneous random field, which is the reason they are omitted from (5) and (6).

Next, we state the drawback associated with SMI when it is used as a similarity measure in 3D brain image registration, and present a solution for it.

## B. ARTIFACT OF SMI AS SIMILARITY MEASURE

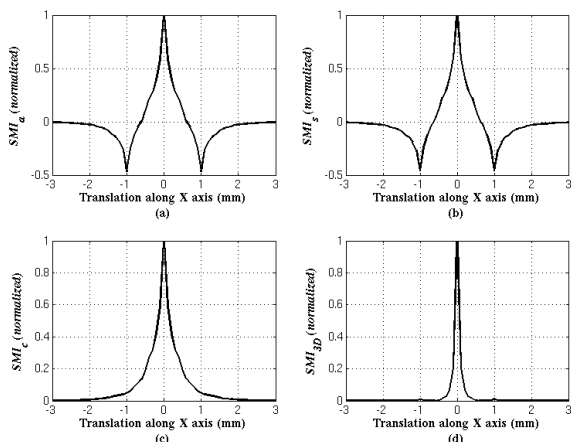
The formulation of SMI in terms of two-dimensional joint entropies is made possible in (6) by the two conditional independency assumptions given in the following equation:

$$Y \perp X_l / (X Y_l) \ \& \ X \perp Y_l / (Y X_l) \quad (7)$$

where  $a \perp b / c$  indicates that  $a$  and  $b$  are independent given  $c$ . These conditional independency assumptions generate an artifact in translational cases since they make the SMI value negative for cases at or around  $\pm 1$  voxel misalignments [see Fig. 3(a)]. The theoretical reason behind this drawback is that when there is such a misalignment between  $\mathbf{X}$  and  $\mathbf{Y}$  images, the assumptions in (7) are no longer valid. For example, if the source image is a translated version of the target image with a  $\pm 1$  voxel misalignment along the  $x$ -axis, then the joint entropies  $H(X, Y_l)$  or  $H(X_l, Y)$  in (6) drop to their minimum value of  $H(X)$ . On the other hand, the conditional independency assumption in (7) implies that  $H(X, Y_l) \geq H(X, Y)$  (data processing inequality mentioned in [30];  $MI(X, Y_l) \leq MI(X, Y)$ ). In fact, all the joint entropies in (6) need to be always greater than  $H(X, Y)$  under the conditional indepen-



density assumptions given in (7). However, this condition is violated on or around  $\pm 1$  voxel misregistration. This is the main reason that it becomes negative around these points.



**FIGURE 3.** *SMI* curves computed between simulated T1 brain image and its translationally misregistered version over (a) axial (b) sagittal, and (c) coronal slices. (d) *SMI*<sub>3D</sub> curve for the same registration.

The above problem is addressed in [6] by using the absolute value of the *SMI*. However, this approach introduces two local optimum points for translational misregistrations causing difficulty in the optimization process of registration. An alternative solution is proposed next by extending the definition of *SMI* to 3D brain image volumes with the added benefit of eliminating its translational artifact. The 3D *SMI*, hereafter called *SMI*<sub>3D</sub>, matches well with magnetic resonance brain images that are captured in 3D.

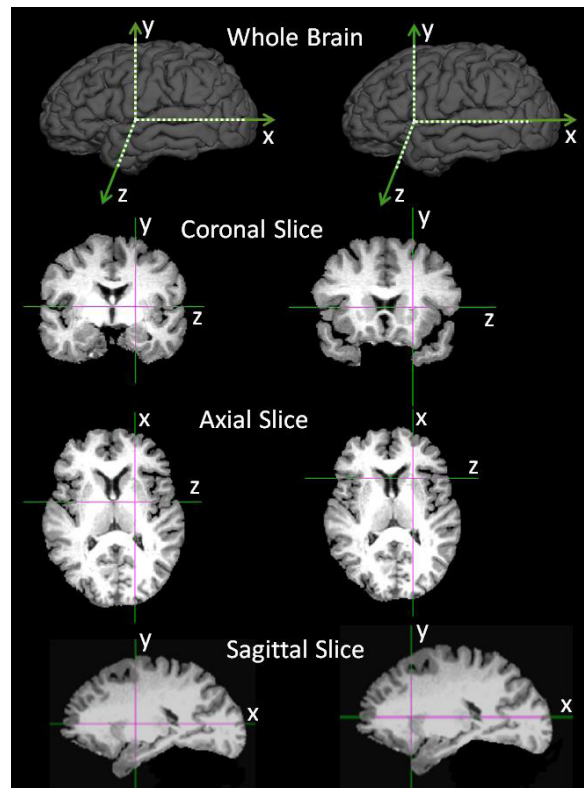
**C. SMI FOR 3D BRAIN IMAGES**

Mathematical computation of *SMI* for pairs of 3D images requires a theoretical expansion of the definition of QMRF to include 3D random fields, an expansion which does not exist at this time. In this work, we have considered a different approach to incorporate the 3D spatial information into the computation of *SMI*. Fig. 4 provides an illustration of the proposed approach. In this figure, a brain volume (top left) is translated along the *x*-axis (top right) and the corresponding cross-sectional slices are shown. As seen in the figure, the effect of 3D translation of the whole volume is different along different cross-sections. While the sagittal and axial slices experience the same translational shift along the *x*-axis, the coronal slices exhibit a total slice change. Therefore, in the case of  $-1$  or  $+1$  voxel translation, the *SMIs* computed for the sagittal and axial slices are negative whereas the *SMI* computed for the coronal slices remains positive. Consequently, one can simply consider a new *SMI* for 3D brain images as the product of the individual cross-sectional *SMIs*, that is

$$SMI_{3D} = SMI_a \times SMI_s \times SMI_c \tag{8}$$

where *SMI*<sub>a</sub> is the *SMI* computed on the axial slices, *SMI*<sub>s</sub> the *SMI* computed on the sagittal slices and *SMI*<sub>c</sub> the *SMI*

computed on the coronal slices. The computed *SMI*<sub>a</sub>, *SMI*<sub>s</sub>, and *SMI*<sub>c</sub> are different due to the fact that they take into account the 2D spatial dependency in different cross-sectional planes. For the case of translational misregistration, only two of the *SMIs* become negative at or around  $\pm 1$  misalignment points, which ensures that the final product is always positive. It should be noted that this drawback occurs only in the translational type of misregistration.



**FIGURE 4.** Illustration of changes in the cross-sectional slices of a typical 3D brain image for a translational misregistration.

Finally, it should be added that in the case of 3D images all the existing cliques in the neighboring structure of the first order QMRF need to be included. Even though these additional cliques would not change the final outcome for the spatial mutual information of a homogeneous random field (as described in [29]) for *SMI*<sub>3D</sub>, it is necessary to include them to ensure the positivity constraint of this similarity measure. This way, the new *SMI* is given by the following equation,

$$SMI = -mnH(X, Y) + \frac{mn}{4} \left\{ \begin{aligned} &H(X, Y_u) + H(X_u, Y) + H(X, Y_d) \\ &+ H(X_d, Y) + H(X, Y_l) + H(X_l, Y) \\ &+ H(X, Y_r) + H(X_r, Y) \end{aligned} \right\} - \frac{mn}{8} \left\{ \begin{aligned} &H(X_u, Y_l) + H(X_l, Y_u) + H(X_d, Y_l) + H(X_l, Y_d) \\ &+ H(X_u, Y_r) + H(X_r, Y_u) + H(X_d, Y_r) + H(X_r, Y_d) \end{aligned} \right\} \tag{9}$$

where the 2-dimensional joint entropies are the same or counterparts of the ones stated in (6), see Fig. 2. This 3D extension

not only solves the artifact of  $SMI$  in translational misregistration, it also produces a more effective similarity measure due to the fact that it captures 3D spatial dependency. Next the effectiveness of the new measure  $SMI_{3D}$  is examined using simulated T1 and T2 weighted brain images. In addition, its performance is compared with the classical  $MI$ ,  $SOMI$  and the 2-dimensional  $SMI$ .

#### IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we used simulated 3D brain MRI scans to examine the effectiveness of  $SMI_{3D}$  as a similarity measure for 3D brain image registration.

##### A. DATA

In order to evaluate and compare the performance of  $SMI_{3D}$  for 3D brain image registration, we used digital brain phantom images of the *BrainWeb* database with two simulated structural MR images: T1-weighted (T1), and T2-weighted (T2). The *BrainWeb* images have been used extensively to study the performance of anatomical brain mapping techniques such as nonlinear co-registration, cortical surface extraction, and tissue classification [31]. The main advantages of using this database are: (i) the answer is known prior to experimentation, and (ii) imaging parameters can be controlled independently. Since the source for simulation of all the images is the same digital phantom, one has a systematic means of establishing a gold standard for registration and control over the level of image degradation for all the modalities. We obtained T1 and T2 brain images with 1mm isometric voxel resolution directly from the *BrainWeb* database. These images were next intensity normalized to the range of (0 ~255) by scaling the range of the original image histogram which contains 99% of the total image energy. Different levels of image degradation (noise) were applied to those images by adding a random Gaussian noise to the images with the variances that gave the desired percentages of noise energy.

##### B. EVALUATION

We first examined the elimination of the translation misregistration artifact in  $SMI$  when using the new measure  $SMI_{3D}$ . Next we compared the effectiveness of  $SMI_{3D}$  to the classical  $MI$ ,  $SOMI$ , and  $SMI$ .

Using bilinear interpolation, we generated translated versions of the T1 scan with step size of 0.1 mm and along the  $x$ -axis. The translated version of the image simulated the misregistered image for our experiments. We then computed  $SMI$  over all the three slices and finally  $SMI_{3D}$  in every step of the simulation. Fig. 3 shows the outcome of this experiment; Fig. 3(a) shows the curve for  $SMI_a$ , Fig. 3(b) for  $SMI_s$ , and Fig. 3(c) for  $SMI_c$  which were computed for the axial, sagittal and coronal slices, respectively. As shown in this figure, the translation along the  $x$ -axis caused both  $SMI_a$  and  $SMI_s$  to become negative at or around  $\pm 1$ , whereas  $SMI_c$  remained positive during the entire translational misregistration process. It is important to note that even though Fig. 3(a) and (b)

look similar, they are different since they show the  $SMI$  value computed on different slices.

Fig. 3(d) shows the  $SMI_{3D}$  curve for the translational misregistration along the  $x$ -axis. As shown in this figure, all the negative drops in the  $SMI$  curves got removed in  $SMI_{3D}$ . In addition, the  $SMI_{3D}$  curve appeared monotonically and smoothly decreasing which is a suitable characteristic of similarity measures for optimization purposes. It is worth pointing out that the two tiny spikes visible at  $\pm 1$  are due to interpolation.

Fig. 3 shows that the translational artifact was removed by using  $SMI_{3D}$  for the translation misregistration along the  $x$ -axis; however, it is also straightforward to show that for the translation along the  $y/z$  axes, this artifact will also be removed by  $SMI_{3D}$ . The only difference would be that for the translation along the  $y/z$  axis,  $SMI_s/SMI_c$  and  $SMI_c/SMI_a$  will be negative at or around  $\pm 1$  misregistration, and  $SMI_a/SMI_s$  will remain positive at all times.

Next we examined the effectiveness of  $SMI_{3D}$  in registering intermodal and noisy images from the *BrainWeb* dataset. We started with the noiseless case and then increased the noise level step by step till it reached 20% of the image energy. The first column in Table 1 lists all the eight registration cases in this experiment. We manually generated translational (30 mm) and rotational (5 degree) misregistrations on one of the images, which is indicated by italic font in the first column. The registration start point was the same for all the eight cases and no optimization or regularization was applied in this experiment. This ensured that only similarity measures got evaluated and the other aspects of the registration were kept constant or got eliminated. We considered a registration to fail when the registration result deviated by more than 0.5 mm / 0.1 degree from a perfect alignment for translational/rotational misregistration.

We tested the effectiveness of the four different similarity measures of  $MI$ ,  $SOMI$ ,  $SMI$ , and  $SMI_{3D}$  in our final experiment. The second and third columns in Table 1 give the registration results for the translational and rotational misregistration for the classical  $MI$ . The fourth and fifth columns give the results for  $SOMI$ , the sixth and seventh columns for  $SMI$ , and the eighth and ninth columns for  $SMI_{3D}$ . As can be seen from this table,  $SMI_{3D}$  outperformed all the other similarity measures by successfully registering all the registration cases except the one with the highest noise level.  $SMI$  failed in three cases,  $SOMI$  failed in 6 cases, and the classical  $MI$  failed in 10 cases.

While the experimentations here have been limited to simulated images due to our ability to control the noise level and access to gold standard, one can clearly see the superiority of  $SMI_{3D}$  to the other similarity measures. Our experimentations have shown a linear stepwise improvement in this order:  $MI$ ,  $SOMI$ ,  $SMI$ , and  $SMI_{3D}$ . However, full evaluation of the proposed similarity measure under different registration problems, similar to the one introduced in [32], is still required to make the final conclusion about the superiority of the  $SMI_{3D}$  in comparison to the existing similarity measures.

**TABLE 1. Registration test results (pass/failed = X) for registering T1 target image to degraded and spatially transformed T2 source image using MI, SOMI, SMI, and SMI<sub>3D</sub>.**

Registration Cases	MI		SOMI		SMI		SMI <sub>3D</sub>	
	Translation	Rotation	Translation	Rotation	Translation	Rotation	Translation	Rotation
Transformation								
T1 & T2	pass	pass	pass	pass	pass	pass	pass	pass
10% Noisy T1 & T2	pass	pass	pass	pass	pass	pass	pass	pass
T1 & 10% Noisy T2	pass	pass	pass	pass	pass	pass	pass	pass
20% Noisy T1 & T2	X	X	pass	pass	pass	pass	pass	pass
T1 & 20% Noisy T2	X	X	pass	pass	pass	pass	pass	pass
10% Noisy T1 & 20% Noisy T2	X	X	X	X	pass	pass	pass	pass
20% Noisy T1 & 10% Noisy T2	X	X	X	X	pass	X	pass	pass
20% Noisy T1 & 20% Noisy T2	X	X	X	X	X	X	pass	X

Finally, it is important to note that the product of three different 2D spatial mutual information mathematically does not give another mutual information. The proposed measure is merely a similarity measure which combines the three existing spatial information for three different cross-sections of a 3D image. The fully characterized 3D mutual information requires the extension of QMRF to 3D random fields, and the derivation of SMI from such fields.

## V. CONCLUSION

This paper has described the importance of incorporating spatial information into the computation of image mutual information and reviewed previous attempts at computing mutual information with spatial dependency. It was shown that the recently defined spatial mutual information has an artifact in the case of translational misregistration, which was remedied by considering the proposed spatial mutual information for 3D brain images. The proposed similarity measure not only addresses the shortcoming associated with translational misregistration, but it also captures 3D, instead of 2D, brain image spatial dependency. The effectiveness of SMI<sub>3D</sub> as a similarity measure was assessed by applying controlled noise levels to simulated brain images and compared to some of the existing similarity measures. Even though the utilization of simulated T1/T2 images has facilitated the evaluation of the introduced similarity measure in this work, an evaluation of this measure via real images seems to be the next natural step of this work.

## REFERENCES

- [1] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, 1997.
- [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, Apr. 1997.
- [3] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "Incorporating connected region labelling into automated image registration using mutual information," in *Proc. Workshop Math. Methods Biomed. Image Anal.*, 1996, pp. 23–31.
- [4] D. Rueckert, M. J. Clarkson, D. L. G. Hill, and D. J. Hawkes, "Non-rigid registration using higher-order mutual information," *Proc. SPIE Conf. Med. Imaging*, vol. 3979, pp. 438–447, Feb. 2000.
- [5] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.
- [6] Q. R. Razlighi, N. Kehtarnavaz, and A. Nosratinia, "Computation of image spatial entropy using quadrilateral Markov random field," *IEEE Trans. Image Process.*, vol. 18, no. 12, pp. 2629–2639, Dec. 2009.
- [7] J. Lin, Z. Gao, B. Xu, Y. Cao, and Z. Yingjian, "The affection of grey levels on mutual information based medical image registration," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 3, Jan. 2004, pp. 1747–1750.
- [8] C. Bei-Jing, L. Jun-Li, and C. Gang, "Study of medical image registration based on second-order mutual information," in *Proc. 1st Int. Conf. Bioinforma. Biomed. Eng.*, Jul. 2007, pp. 956–959.
- [9] J. P. W. Pluim, J. B. Maintz, and M. A. Viergever, "Image registration by maximization of combined mutual information and gradient information," *IEEE Trans. Med. Imag.*, vol. 19, no. 8, pp. 809–814, Aug. 2000.
- [10] R. Gan and A. C. S. Chung, "Multi-dimensional mutual information based robust image registration using maximum distance-gradient-magnitude," *Inf. Process. Med. Imag.*, vol. 19, pp. 210–221, Jan. 2005.
- [11] D. Shen and C. Davatzikos, "HAMMER: Hierarchical attribute matching mechanism for elastic registration," *IEEE Trans. Med. Imag.*, vol. 21, no. 11, pp. 1421–1439, Nov. 2002.
- [12] D. Russakoff, C. Tomasi, and T. Rohlfing, "Image similarity using mutual information of regions," in *Proc. 8th ECCV*, 2004, pp. 596–607.
- [13] C. Yang, T. Jiang, J. Wang, and L. Zheng, "A neighborhood incorporated method in image registration," *Med. Imag. Augment. Real.*, vol. 15, no. 1, pp. 244–251, 2006.
- [14] A. Bardera, M. Feixas, I. Boada, and M. Sbert, "High-dimensional normalized mutual information for image registration using random lines," in *Proc. Biomed. Image Registration*, 2006, pp. 264–271.
- [15] D. Tomazevic, B. Likar, and F. Pernus, "Multi-feature mutual information image registration," *Image Anal. Stereol.*, vol. 31, pp. 43–53, 2012.
- [16] M. Holden, L. Griffin, and N. Saeed, "Multi-channel mutual information using scale space," in *Proc. MICCAI*, 2004, pp. 797–804.
- [17] T. Butz and J. P. Thiran, "Affine registration with feature space mutual information," in *Proc. MICCAI*, 2010, pp. 549–556.
- [18] A. O. I. Hero, B. M. B. Ma, O. J. J. Michel, and J. Gorman, "Applications of entropic spanning graphs," *IEEE Signal Process. Mag.*, vol. 19, no. 5, pp. 85–95, Sep. 2002.
- [19] H. Neemuchwala, A. Hero, and P. Carson, "Image matching using alpha-entropy measures and entropic graphs," *Signal Process.*, vol. 85, no. 2, pp. 277–296, 2005.
- [20] H. Luan and F. Qi, "Multi-modal image registration by quantitative-qualitative measure of mutual information (q-mi)," in *Proc. Comput. Vis. Biomed. Image*, 2005, pp. 378–387.
- [21] H. Luan, F. Qi, Z. Xue, L. Chen, and D. Shen, "Multimodality image registration by maximization of quantitative-qualitative measure of mutual information," *Pattern Recognit.*, vol. 41, no. 1, pp. 285–298, Jan. 2008.
- [22] T. Kadir and M. Brady, "Saliency, scale and image description," *Int. J. Comput. Vis.*, vol. 45, no. 2, pp. 83–105, 2001.
- [23] C. E. Shannon, "The mathematical theory of communication. 1963.," *MD Comput. Med. Pract.*, vol. 14, no. 4, pp. 306–317, 1948.
- [24] H.-M. Chan, A. C. S. Chung, S. C. H. Yu, E. Norbashi, and W. M. W. Iii, "Multi-modal image registration by minimizing Kullback-Leibler distance," in *Proc. Int. Conf. Med. Image Comput. ComputerAssist. Intervent*, 2002, pp. 525–532.

- [25] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multi-modality image registration based on information theory," *Proc. Inf. Process. Med. Imag.*, vol. 3, no. 6, pp. 263–274, 1995.
- [26] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognit.*, vol. 32, no. 1, pp. 71–86, 1999.
- [27] C. E. Rodriguez-Carranza and M. H. Loew, "Weighted and deterministic entropy measure for image registration using mutual information," *Proc. SPIE Conf. Med. Imaging*, vol. 3338, pp. 155–166, 1998.
- [28] Z. L. Stan, *Markov Random Field Modeling in Image Analysis*, 3rd ed. New York, NY, USA: Springer-Verlag, 2009, p. 357.
- [29] Q. R. Razlighi, M. T. Rahman, and N. Kehtarnavaz, "Fast computation methods for estimation of image spatial entropy," *J. Real-Time Image Process.*, vol. 6, no. 2, pp. 137–142, Jan. 2011.
- [30] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.
- [31] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, *et al.*, "Design and construction of a realistic digital brain phantom," *IEEE Trans. Med. Imag.*, vol. 17, no. 3, pp. 463–468, Jul. 1998.
- [32] Q. R. Razlighi, N. Kehtarnavaz, and S. Yousefi, "Evaluating similarity measures for brain image registration," *J. Vis. Commun. Image Represent.*, vol. 24, no. 7, pp. 977–987, Oct. 2013.



**NASSER KEHTARNAVAZ** (S'82–M'86–SM'92–F'12) received the Ph.D. degree from Rice University, Houston, TX, USA, in 1987.

He is a Professor of electrical engineering and the Director of the Signal and Image Processing Laboratory, University of Texas at Dallas.

His research interests include digital signal and image processing, real-time signal and image processing, pattern recognition, and biomedical image analysis. He has authored or co-authored eight books and more than 250 papers. Dr. Kehtarnavaz is a fellow of SPIE. He is an Editor-in-Chief of *Journal of Real-Time Image Processing*.



**QOLAMREZA R. RAZLIGHI** (M'09–SM'12) received Ph.D. degree in electrical engineering from the University of Texas at Dallas, TX, USA, in 2009.

He is currently an Assistant Professor with the Department of Neurology and Adjunct Assistance Professor with the Department of Biomedical Engineering, Columbia University. He was a Senior System Engineer with Alcatel Lucent, Plano TX, USA, before joining Columbia University.

His research interests lie in the broad area of signal and image processing with application in neuroimaging. He is a recipient of the National Institutes of Health K01 Career Award. Dr. Razlighi has been a member of the Human Brain Mapping Organization since 2012.

His research interests lie in the broad area of signal and image processing with application in neuroimaging. He is a recipient of the National Institutes of Health K01 Career Award. Dr. Razlighi has been a member of the Human Brain Mapping Organization since 2012.