Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

A New Image Similarity Measure with Reduced Sensitivity to Interpolation and Generalizability to Multispectral Image Registration

Babak A. Ardekani, *Senior Member, IEEE,* and Alvin Bachman

***Abstract*— Mutual information (MI) has proven to be a useful similarity measure for spatial registration between related pairs of images in various medical imaging applications. Image registration algorithms that utilize the MI assume that the best alignment between a pair of images is reached when their MI is at its maximum. However, this assumption is not always valid because the MI is not only sensitive to dissimilarity between images, but also to the image interpolation operations performed during the optimization process in image registration algorithms. When the images that are being registered are close to their optimum spatial alignment, MI's sensitivity to interpolation may become dominant over its sensitivity to image misalignment, hence limiting the accuracy of the image registration method. In this paper, we present an entropy-based cost function, closely related to MI, that can be made relatively insensitive to interpolation effects, and can be generalized to registration of multispectral images.**

I. INTRODUCTION

The mutual information (MI) measure of similarity between images was introduced independently by Wells et al. [1] and Maes et al. [2] and applied to the problem of multimodality medical image registration. Since its introduction, there has been considerable interest in the medical imaging community in the application of MI to various image registration problems [3].

When MI is used to measure similarity between images, it is sensitive to two factors: (a) the accuracy of spatial registration between the images; and (b) any image interpolation that may have been applied to spatially transform the images before computing the MI. Sensitivity to misalignment between images is a desirable characteristic of MI and is the reason for its utility in image registration applications. However, sensitivity to interpolation operations is a confounding effect that limits the accuracy of MI and should be minimized. When spatial registration between images is at or close to the optimum point, variations in MI due to interpolation effects can dominate those due to registration errors, thus limiting the accuracy of spatial registration.

A number of strategies have been suggested to reduce the interpolation artifact in MI [4], [5]. These often involve resampling, filtering or blurring the images. In some cases, these methods appear *ad hoc* and without sufficient justification for their use. They also require specification of parameters (e.g., resampling ratio or blurring kernel) that are selected subjectively and their optimum values may vary depending on the type of images being registered.

In this paper, we propose a new entropy-based image dissimilarity cost function, which is closely related to MI, but its implementation reduces to variance estimation in spatially contiguous image sub-regions. This allows us to characterize the source of interpolation artifact and devise effective methods to remedy the problem. In addition, the new cost function can be easily generalized to the case of multispectral image registration, where one aims to register two groups of images together, instead of a pair of images.

II. THEORY

The MI between a pair of spatially aligned images A and B can be computed based on any *labeling* of image volume elements (voxels). We will have more to say about this labeling and how it affects the MI between images A and B shortly. The expression for MI can be written in various forms. One that is particularly useful for the development of the ideas presented in this paper is given by:

$$I(A,B) = H(A) - H(A \backslash B), \quad (1)$$

where $H(A)$ denotes the *entropy* of image A, and $H(A \backslash B)$ represents the *conditional entropy* of A given B. The entropy $H(A)$ can be viewed as the expected amount of information to be gained from observation of the label of a random voxel selected from image A. Alternatively, it quantifies the uncertaintly in such an observation [6]. We shall simply refer to $H(A)$ as the uncertaintly in A. The conditional entropy $H(A \backslash B)$ quantifies the uncertainty remaining in A given that B is known. Thus, the MI as given by (1) measures the reduction in uncertainty of image A given the knowledge of image B. It can be shown that $0 \leq I(A,B) \leq I_{max}$, where $I(A,B) = 0$ when voxel labels in A and B are statistically independent, and $I(A,B) = I_{max} = H(A) = H(B)$ when the labels in one image exactly determine the labels in the other. That is, the two images as labeled carry precisely the same information.

We now present a cost function that is closely related to MI, but can be implemented in such as way as to be relatively insensitive to interpolation effects. To this end, consider the expression for the MI given in (1). We first note that the accuracy of spatial registration should not affect the entropy of A, $H(A)$. If this term is affected, it can only be due to interpolation and since we would like to eliminate interpolation effects, we choose to ignore this term and propose to minimize the conditional entropy $H(A \backslash B)$,

B. A. Ardekani is with the Center for Advanced Brain Imaging, The Nathan S. Kline Institute for Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA ardekani@nki.rfmh.org

A. Bachman is Professor Emeritus, The City College of The City University of New York bachman@nki.rfmh.org

1-4244-0033-3/06/$20.00 ©2006 IEEE.

which can be written as follows:

$$H(A\backslash B) = \sum_{j=1}^{m} p(b_j) H(A\backslash b_j), \qquad (2)$$

where $\{b_j\}$ represents a set of m labels assigned to the voxels of image B. When using MI or other entropy-based criteria for image registration, it is common practice to assign these labels solely based on the voxel intensities, by defining m intensity bins and assigning labels b_j to all voxels whose intensities fall within the jth bin. However, voxel labeling based on histograms is not the only way for classification of voxels. Entropies can be computed based on *any* labeling of image voxels. We take advantage of this fact in the current work and label voxels not only based on their intensities, but also based on their spatial contiguity. That is, two voxels are given the same label if and only if they fall within the same histogram bin *and* are spatially contiguous. So two voxels that fall within a given histogram bin are not necessarily assigned the same label. They are only assigned to the same class if they belong to the same *connected component*. Such classification of voxels in image B for the purpose of computing entropies has the advantage that the class conditional densities of the overlapping voxels in image A may be fairly accurately approximated by Gaussian distributions when the two images are close to spatial registration. That is:

$$p(a\backslash b_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp \frac{-(a-\mu_j)^2}{2\sigma_j^2}, \qquad (3)$$

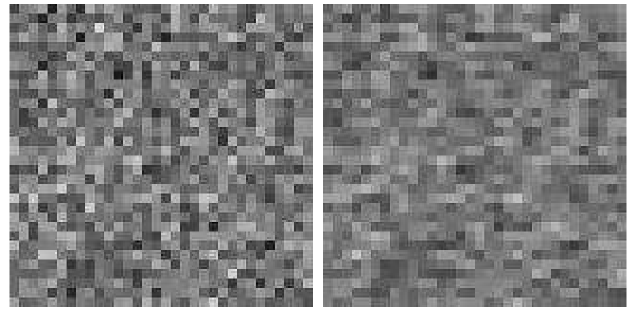
where μ_j and σ_j^2 represent the population mean and variance of the subset of voxels in image A that coincide with the volume occupied by the voxels in image B with label b_j when the two images are in spatial registration. Note that it is convenient to treat the class conditional densities of A as continuous distributions. Given (3) and the expression for entropy of continuous distributions, it is easy to show that the terms $H(A\backslash b_j)$ in (2) can be written as [6]:

$$H(A\backslash b_j) = 0.5 + \ln\sqrt{2\pi} + \ln\sigma_j. \qquad (4)$$

By substituting (4) in (2), it reduces to:

$$H(A\backslash B) = \sum_{j=1}^{m} p(b_j) \ln\sigma_j + constant. \qquad (5)$$

Thus, in order to compute our registration measure, it suffices to estimate (for all j) the within class population variance σ_j^2 of voxels in image A that spatially coincide with the voxels with label b_j in image B. Hence, the computation of our registration measure reduces to one of estimating variances. There are two advantages to this fact: firstly, variance can be estimated fairly accurately even if a class consists of relatively few voxels. Traditional implementations of the MI based on non-parametric probability distributions have had to ensure that a sufficiently large number of voxels occupy each histogram bin in order to obtain a reasonably accurate representation of the marginal and joint probability densities. More importantly, it is easier to deal with the interpolation effects when estimating the variances, reducing the sensitivity of our registration measure to interpolation effects.



(a) (b)

Fig. 1. (a) Simulated Gaussian white noise image. (b) The image in part (a) translated 0.5 pixels horizontally. Image in part (b) is slightly blurred due to the linear interpolation that is applied.

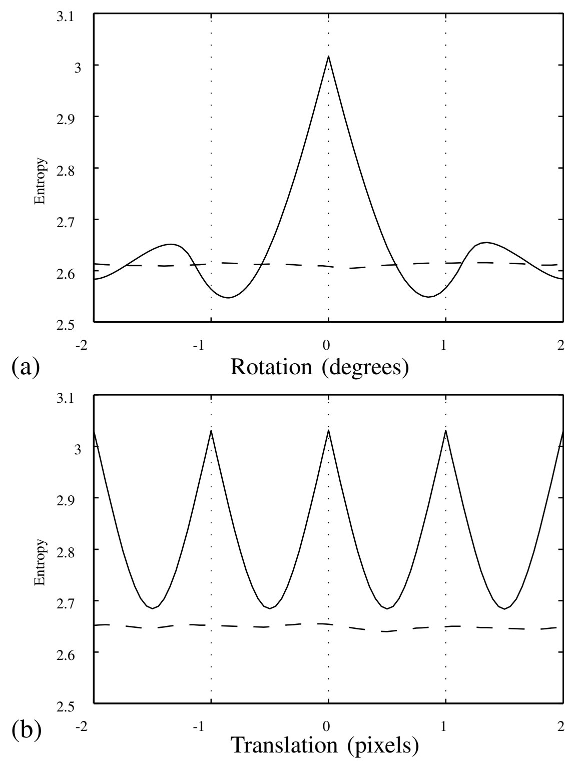


Fig. 2. The entropy of the Gaussian white noise image in Fig. 1a as it is rotated from $-2°$ to $2°$ (a) and from -2 pixels to 2 horizontally (b).

Having reduced the computation of our cost function to one of estimating variances in regions of image A that spatially overlap with specific connected components in image B, we can now characterize the adverse effect of interpolation on variance estimation and devise a scheme to remedy the situation. To this end, consider the the image shown in Fig. 1a. This image consists entirely of simulated Gaussian white noise with with $\mu = 100$ and $\sigma^2 = 25.0$. The entropy of the image is computed to be approximately 3.03. If this noise image is translated, say horizontally, by one pixel, then the entropy estimate does not change (ignoring the edge effects). However, if the image is translated by only half of a pixel, then an interpolation operation will be required, which slightly blurs the translated image and hence reduces the entropy. Fig. 1b shows the original image shifted by half of a voxel in the horizontal direction using linear interpolation. Note the small amount of smoothing introduced by the interpolation operation. Due to this smoothing, the entropy decreases to approximately 2.68.

The solid lines in Fig. 2 illustrate the variations in the entropy estimate as a function of rotations (a) and translations (b) of the original image. It can be seen that when the image is rotated using interpolation, a spatially varying blurring is introduced such that the entropy decreases (Fig. 2a). Similarly, when the translation is not a multiple of whole pixels, the entropy decrease due to the interpolation (Fig. 2b). These variations are precisely those that cause the so called interpolation artifact in MI. Since computation of the conditional entropy cost function in (5) is basically variance estimation, and the effect of interpolation on image intensity variance is easy to characterize, methods may be devised to overcome the interpolation artifact when implementing (5). One approach is described below.

An image voxel in 3D can be considered as representing the volume of a rectangular parallelepiped centered at coordinates $\mathbf{r}$. Variance estimation of a subset of voxels in image A that correspond to a connected component $\mathcal{C}_j$ in image B proceeds as follows. For each of N_j voxels $\mathbf{r} \in \mathcal{C}_j$, we find their transformed position $\mathbf{r}' = f(\mathbf{r})$ under the current transformation f. Then estimate the variance as follows:

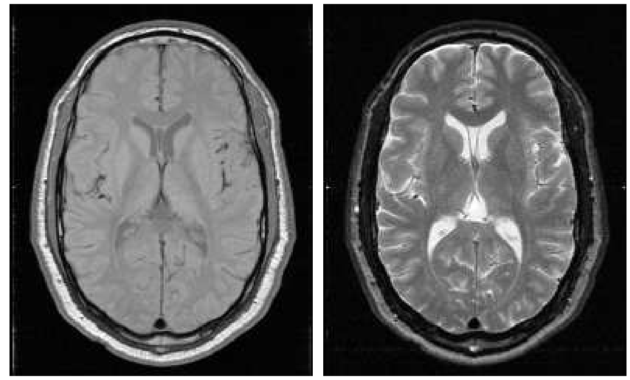
$$\hat{\sigma}_j^2 = \frac{1}{N_j - 1} \sum_{\mathbf{r} \in \mathcal{C}_j} [I_A(\mathbf{r}') - \hat{m}_j]^2 \tag{6}$$

where $I_A(\mathbf{r}')$ denotes the interpolated intensity of image A at point $\mathbf{r}' = f(\mathbf{r})$ and $\hat{m}_j$ is the average of the voxels in A that correspond to voxels $\mathcal{C}_j$ in B:

$$\hat{m}_j = \frac{1}{N_j} \sum_{\mathbf{r} \in \mathcal{C}_j} I_A(\mathbf{r}'). \tag{7}$$

One scheme to make this variance estimation independent of the amount of blurring introduced by image interpolation is to consider the voxel coordinates $\mathbf{r}$, not to be the center of the rectangular parallelepiped, but a random point selected within that region. In our implementation, the random point is drawn from a uniform distribution. In the simulations shown in Fig. 1 and Fig. 2, when the variance, or equivalently the entropy, was estimated using the above scheme, we obtained the dashed lines in Fig. 2. It can be seen that the sensitivity of the entropy measure to the amount of smoothing introduced by interpolation is greatly reduced.

The similarity measure introduced in this paper may be generalized to multispectral image registration. In this case, instead of a pair of images, we have two sets of images $\mathcal{A} = \{A_1, A_2, \ldots A_N\}$ and $\mathcal{B} = \{B_1, B_2, \ldots B_M\}$. Within each set, the images are assumed to be registered. The problem is to find a transformation that matches the images in one set to the images in the other set. For the case of multispectral image registration, the cost function proposed in this paper directly generalizes to minimization of the joint conditional entropy $H(A_1, A_2, \ldots A_N \backslash B_1, B_2, \ldots B_N)$. Since the images in set $\mathcal{B}$ are assumed to be in spatial registration, the corresponding voxels in these images must be given the same label. Thus, the joint conditional entropy may be written as:



(a) (b)

Fig. 3. Double-echo proton density (a) and T_2-weighted (b) images.

$$H(\mathcal{A} \backslash \mathcal{B}) = \sum_{j=1}^{m} p(b_j) H(A_1, A_2, \ldots A_N \backslash b_j), \tag{8}$$

where $\{b_j\}$ represents a set of m labels assigned to the voxels of images in set $\mathcal{B}$. As before, this assignment of labels is based on both the multispectral data in image set $\mathcal{B}$ *and* spatial contiguity of voxels. Then the class conditional densities of voxels in image set $\mathcal{A}$ may be approximated as N-dimensional Gaussian distributions. In that case, $H(\mathcal{A} \backslash b_j)$ reduces to [6]:

$$H(\mathcal{A} \backslash b_j) = \frac{N}{2}(1 + \ln 2\pi) + \ln |\mathbf{\Sigma}_j|^{1/2}, \tag{9}$$

where $\mathbf{\Sigma}_j$ is the $N \times N$ covariance matrix of the subset of voxels in image set $\mathcal{A}$ that overlap with the region labeled as b_j in image set $\mathcal{B}$. Therefore, (8) reduces to:

$$H(\mathcal{A} \backslash \mathcal{B}) = \sum_{j=1}^{m} p(b_j) \ln |\mathbf{\Sigma}_j|^{1/2} + constant. \tag{10}$$

Thus, the computation of the conditional entropy cost function given the image sets $\mathcal{A}$ and $\mathcal{B}$ reduces to (a) assigning labels b_j $j = 1, 2, \ldots m$ to the corresponding voxels in images in set $\mathcal{B}$ based on an M-dimensional clustering algorithm (e.g, K-means) followed by a connected component analysis; and (b) for each j, estimating the population covariance matrix $\mathbf{\Sigma}_j$ which corresponds to the voxels in image set $\mathcal{A}$ that are spatially aligned to the voxels in image set $\mathcal{B}$ that have been assigned label j. This estimation should be made in such as way so that it does not depend heavily on the amount of blurring introduced by image interpolation.

III. EXPERIMENTS AND RESULTS

To illustrate the interpolation artifact and how it can be remedied by using the cost function presented in this paper, we obtained MRI images scanned on a $1.5\,$T Siemens Vision system (Erlangen, Germany). The images were acquired in an axial orientation using a double-echo turbo spin-echo

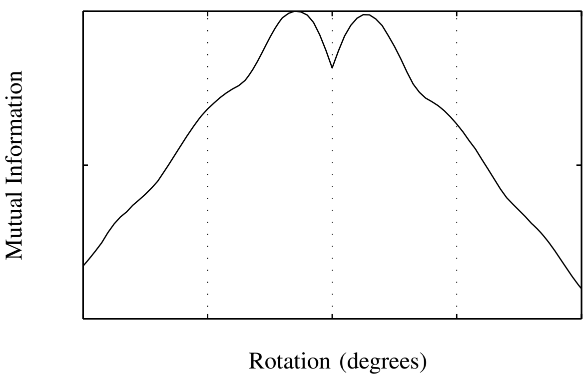


Fig. 4. The mutual information measure as a function of rotational misregistration between the proton density and T_2-weighted images.

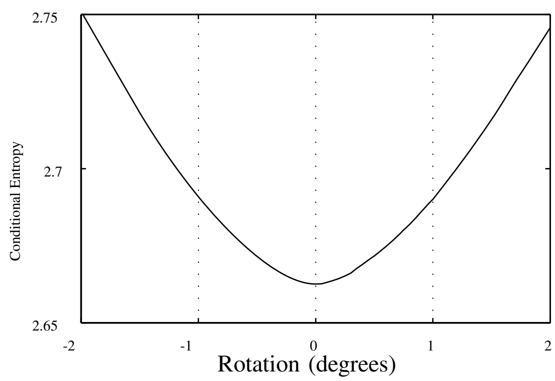


Fig. 5. The conditional entropy measure (5) as a function of rotational misregistration between the proton density and T_2-weighted images.

(TSE) sequence with (TR=5 s, TE=22 and 90 ms, matrix size = 256×256, FOV=240 mm, 26 slices, 5 mm slice thickness, no gap). The short TE of 22 ms produces a proton density weighted image (PD) and the longer TE of 90 ms produces a T_2-weighted image (T_2). Since the PD and T_2 images are acquired simultaneously, they are inherently registered. Fig. 3 shows a pair of image slices obtained by the TSE sequence.

To demonstrate the interpolation artifact present in the MI registration measure, we computed the MI between the PD and T_2 images as we rotated the PD image from -2 to 2 degrees in steps of $0.05°$. The MI was computed using the conventional method based on intensity histograms with 40 bins of equal width. The results are shown in Fig. 4, scaled by the maximum MI value. It can be seen that when the mismatch between the two images becomes less than roughly $0.3°$, the interpolation artifact dominates the MI measure creating the local minimum at $0°$. Thus, maximization of the MI without accounting for the interpolation artifact would cause the registration algorithm to find the maximum at $-0.3°$ instead of the true point of optimum registration, which in this case we know to be at $0°$.

The same experiment was performed with the new cost function given in (5) and the variance estimation scheme proposed in the previous section. The labeling was performed on the T_2 image using a K-means clustering algorithm with 8 classes, followed by a connected component analysis. Details of the labeling method can be found in [7]. Thus, voxels were given the same label if their intensities were classified to be in the same class by the K-means algorithm *and* they were in the same 4-connected component. The resulting curve is shown in Fig. 5. As it can be seen, the interpolation artifact at $0°$ that is apparent in Fig. 4 disappears in Fig. 5.

IV. Discussion and Conclusions

In this paper we proposed to use the conditional entropy as a cost function for image registration. We assume that a labeling of the target image can be done in such a way that the class conditional probability density functions in the subject image can be well approximated as Gaussian distributions. In such a case, the conditional entropy function reduces to a form that only dependends on population variance estimates within voxels in the subject image that overlap with particular class labels in the target image. When estimating those variances, effects of interpolation may be characterized and minimized, hence reducing the interpolation artifacts that are present in entropy-based similarity measures, such as the mutual information. One such scheme for variance estimation with reduced sensitivity to blurring introduced by interpolation was proposed in this paper. Other methods may also be devised but were not considered in this paper. In general, one should aim to create algorithms that minimize sensitivity to interpolation effects and maximize sensitivity to image misalignment. Finally, the implementation of the the conditional entropy cost function presented in this paper can be readily generalized to the case of multispectral image registration.

References

[1] W.M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Med. Image Anal.*, vol. 1, 1996, pp. 35-51.

[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, 1997, pp. 187-198.

[3] J.P. Pluim, J.B. Maintz, M.A. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Trans. Med. Imag.*, vol. 22, 2003, pp. 986-1004.

[4] J. Tsao, "Interpolation artifacts in multimodality image registration based on maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 22, 2003, pp. 854-864.

[5] J.P.W. Pluim, J.B.A. Maintz, M.A. Viergever, "Interpolation artefacts in mutual information-based image registration," *Comp. Vis. Image Understanding*, vol. 77, 2000, pp. 211-232.

[6] C.E. Shannon, "A mathematical theory of communication," *The Bell Sys. Tech. J.*, vol. 27, 1948, pp. 379-423, 623-656.

[7] B.A. Ardekani, M. Braun, B.F. Hutton, I. Kanno, H. Iida, "A fully automatic multimodality image registration algorithm," *J. Comp. Assis. Tomog.*, vol. 19, 1995, pp. 615-623.