

A New & Robust Information Theoretic Measure and its Application to Image Alignment ^{*}

F. Wang¹, B. C. Vemuri¹, M. Rao² and Y. Chen²

¹ Department of Computer & Information Sciences & Engr.,

² Department of Mathematics

University of Florida, Gainesville, FL 32611

Abstract. *In this paper we develop a novel measure of information in a random variable based on its cumulative distribution that we dub cumulative residual entropy (CRE). This measure parallels the well known Shannon entropy but has the following advantages: (1) it is more general than the Shannon Entropy as its definition is valid in the discrete and continuous domains, (2) it possess more general mathematical properties and (3) it can be easily computed from sample data and these computations asymptotically converge to the true values. Based on CRE, we define the cross-CRE (CCRE) between two random variables, and apply it to solve the image alignment problem for parameterized (3D rigid and affine) transformations. The key strengths of the CCRE over using the mutual information (based on Shannon's entropy) are that the former has significantly larger tolerance to noise and a much larger convergence range over the field of parameterized transformations. We demonstrate these strengths via experiments on synthesized and real image data.*

1 Introduction

The concept of entropy is central to the field of information theory and was originally introduced by Shannon in his seminal paper [14], in the context of communication theory. Since then, this concept and variants thereof have been extensively utilized in numerous applications of science and engineering. To date, one of the most widely benefiting application has been in data compression and transmission. Shannon's definition of entropy originated from the discrete domain and its continuous counterpart called the *differential entropy* is not a direct consequence of the definition in the discrete case. Note that the Shannon definition of entropy in the discrete case does not converge to the continuous definition [6]. Moreover, the definition in the discrete case, which states that the entropy $H(X)$ in a random variable, X , is $H(X) = -\sum_x p(x)\log(p(x))$ is based on the density of the random variable $p(X)$, which in general may or may not exist e.g., for cases when the cumulative distribution function (*cdf*) is not differentiable. *It would not be possible to define the entropy of a random variable*

^{*} This research was in part funded by the NIH grant NS42075.

for which the density function is undefined. However, it seems plausible that entropy should exist for such cases as well.

In this paper, we will define a new measure of information in a random variable that will overcome the aforementioned drawbacks of the Shannon entropy and has very general properties as a consequence. We will then derive some interesting properties and state some theorems which are proved elsewhere [4]. Following this, we will use this new measure to define cross-CRE (CCRE) and use it in the image alignment problem and compare it to methods that use the Shannon entropy in defining a matching criterion specifically, the mutual information.

1.1 Previous Work

There are several information theoretic measures that have been reported in literature [7] since the inception of Shannon’s entropy in 1948 [14]. Some of these are more general than others but all of them, like Shannon’s entropy, were defined based on the probability density function. This is the point of departure in our approach i.e., our definition is based on cumulative distribution instead and has some very interesting properties some of which are discussed subsequently.

In the context of the image alignment problem, information theoretic measures for comparing image pairs differing by an unknown coordinate transformation has been popular since the seminal works of Viola & Wells [18] and Collignon et.al., [5]. There are numerous methods in literature for solving the image alignment problem. Broadly speaking, these can be categorized as feature-based and direct methods. The former typically compute some distinguishing features and define a cost function whose optimization over the space of a known class of coordinate transforms leads to an optimal coordinate transformation. The latter set of methods involve defining a matching criterion directly on the intensity image pairs. We will briefly review the direct methods and refer the reader to the survey [11] for others.

Sum of squared differences (SSD) has been a popular technique for image alignment [16, 17]. Variants of the original formulation have been able to cope with the deviations from the image brightness constancy assumption [8]. Other matching criteria use of statistical information in the image e.g., correlation ratio [12] and maximum likelihood criteria based on data sets that are pre-registered [10]. Image alignment is achieved by optimizing these criteria over a set of parameterized coordinate transformations. The statistical techniques can cope with image pairs that are not necessarily from the same imaging modality.

Another direct approach is based on the concept of maximizing mutual information (MI) – defined using the Shannon entropy – reported in Viola and Wells [18], Collignon et al., [5] and Studholme et al., [15]. MI between the source and the target images that are to be aligned is maximized using a stochastic analog of the gradient descent method in [18] and other optimization methods such as the Powells method in [5] and a multiresolution scheme in [15]. Reported registration experiments in these works are quite impressive for the case of rigid motion. In [15], Studholme et.al., presented a normalized MI (NMI) scheme for

matching multi-modal image pairs misaligned by a rigid motion. Normalized MI was shown to be able to cope with image pairs not having the same field of view (FOV), an important and practical problem; for an alternative competing method, see Liu et.al., [9] wherein local frequency representations of the images are matched in a statistically robust framework. In recent times, most of the effort on the MI-based methods has been focussed on coping with non-rigid deformations between the source and target multi-modal data sets [13, 3].

2 Cumulative Residual Entropy: A new measure of information

In this section we define our new information theoretic measure and present some properties/theorems. We do not delve into the proofs but refer the reader to a more comprehensive mathematical (technical) report [4].

Definition: Let \bar{X} be a random vector in \mathcal{R}^N , we define the cumulative residual entropy (CRE) of \bar{X} , by :

$$\mathcal{E}(\bar{X}) = - \int_{\mathcal{R}_+^N} P(|\bar{X}| > \bar{\lambda}) \log P(|\bar{X}| > \bar{\lambda}) d\bar{\lambda} \quad (1)$$

Where $\bar{X} = (X_1, X_2, \dots, X_N)$, $\bar{\lambda} = (\lambda_1, \dots, \lambda_N)$ and $|\bar{X}| > \bar{\lambda}$ means $|X_i| > \lambda_i$ and $\mathcal{R}_+^N = (\bar{X} \in \mathcal{R}^N; X_i \geq 0)$

Proposition 1 *If X_i are independent, then*

$$\mathcal{E}(\bar{X}) = \sum_i \left(\prod_{i \neq j} E(|X_j|) \right) \mathcal{E}(X_i)$$

Proposition 2 (*Weak Convergence*). *Let the random vectors \bar{X}_k converge in distribution to the random vector \bar{X} ; by this we mean*

$$\lim_{k \rightarrow \infty} E[\varphi(\bar{X}_k)] = E[\varphi(\bar{X})] \quad (2)$$

for all bounded continuous function ϕ on \mathcal{R}^N , if all the \bar{X}_k are bounded in L^p for some $p > N$, then

$$\lim_{k \rightarrow \infty} \mathcal{E}(\bar{X}_k) = \mathcal{E}(\bar{X}) \quad (3)$$

Definition: Given random vectors \bar{X} and $\bar{Y} \in \mathcal{R}^N$, we define the conditional CRE $\mathcal{E}(\bar{X}|\bar{Y})$ by :

$$\mathcal{E}(\bar{X}|\bar{Y}) = - \int_{\mathcal{R}_+^N} P(|\bar{X}| > x|\bar{Y}) \log P(|\bar{X}| > x|\bar{Y}) dx \quad (4)$$

Proposition 3 *For any \bar{X} and \bar{Y}*

$$E[\mathcal{E}(X|\bar{Y})] \leq \mathcal{E}(X) \quad (5)$$

Equality holds iff \bar{X} is independent of \bar{Y} . This is a useful property and is analogous to the Shannon entropy case.

Definition: The continuous version of the Shannon entropy called the differential entropy [6] $\mathcal{H}(X)$ of a random variable X with density f is defined as

$$\mathcal{H}(X) = -E[\log f] = - \int f(x) \log f(x) dx$$

The following proposition describes the relationship between CRE and the differential entropy and we prove that the CRE is exponentially larger than the differential entropy. This in turn will influence the relations between quantities derived from $\mathcal{E}(X)$ and $\mathcal{H}(X)$ such as mutual information which will be used in the alignment process.

Proposition 4 *Let $X \geq 0$ have density f , then,*

$$\mathcal{E}(X) \geq C \exp(\mathcal{H}(X)), \quad C = \exp\left(\int_0^1 \log(x|\log x|) dx\right) \quad (6)$$

Proof: Let $G(x) = P[X > x] = \int_x^\infty f(u) du$ using the Log-Sum inequality [6] we have,

$$\int_0^\infty f(x) \log \frac{f(x)}{G(x)|\log G(x)|} dx \geq \log \frac{1}{\int_0^\infty G(x)|\log G(x)| dx} = \log \frac{1}{\mathcal{E}(X)}$$

The left hand side in (7) equals

$$-\mathcal{H}(X) - \int_0^\infty f(x) \log(G(x)|\log G(x)|) dx$$

so that,

$$\mathcal{H}(X) + \int_0^\infty f(x) \log(G(x)|\log G(x)|) dx \leq \log \mathcal{E}(X)$$

Finally a change of variable gives:

$$\int_0^\infty f(x) \log(G(x)|\log G(x)|) dx = \int_0^1 \log(x|\log x|) dx$$

Using the above and exponentiating both sides of (7), we get (6) □

Definition: The mutual information $I(X, Y)$ of two random variable X and Y using Shannon entropy is defined as :

$$I(X, Y) = \mathcal{H}(X) - E[\mathcal{H}(X/Y)] \quad (7)$$

We define a quantity called the cross-CRE (CCRE) given by

$$\mathcal{C}(X, Y) = \mathcal{E}(X) - E[\mathcal{E}(Y/X)] \quad (8)$$

Note that $I(X, Y)$ is symmetric but $\mathcal{C}(X, Y)$ need not be. We can define a symmetrized version of CCRE by adding $\mathcal{E}(Y) - E[\mathcal{E}(X/Y)]$ to $\mathcal{C}(X, Y)$ and pre-multiplying it by a factor of $\frac{1}{2}$. From Proposition 3, we know that symmetrized

CCRE is non-negative. We will however use the non-symmetric CCRE in all our image alignment experiments as it was sufficient to yield the desired results. We empirically show its superior performance under low SNR and also depict its larger capture range with regards to the convergence to the optimal parameterized transformation. The image alignment problem that we solve in this paper can now be stated as finding the coordinate transformation T – in our case, represented as a parameterized rigid or affine transformation in 3D – that would maximize the CCRE measure $\mathcal{C}(I_1(T(\mathbf{x})), I_2(\mathbf{x}))$, $\mathbf{x} = (x, y)^T$, over all the appropriate (rigid, affine) class of transformations T .

2.1 Estimating Empirical CRE

In order to compute CRE of an image, we use the histogram of an image to estimate the $P(X > \lambda)$ where X corresponds to the image intensity which is considered as a random variable. Note that as a consequence of property 2, empirical CRE computation based on the samples will converge in the limit to the true value. *This is not the case for the Shannon entropy computed using histograms to estimate the probability density functions, as is usually done in current literature.* In the case of CRE, we have,

$$\begin{aligned} \mathcal{E}(X) &= - \int_0^\infty P(X > \lambda) \log P(X > \lambda) d\lambda \\ &= - \sum_\lambda P(X > \lambda) \log P(X > \lambda) \end{aligned} \quad (9)$$

Hence, using a histogram to compute the CRE is well defined and justified theoretically.

Note that estimating $\mathcal{E}(X/Y)$ is done using the joint histogram and then marginalizing it with respect to the conditioned variable.

3 Experiment Results

In this section we present *four sets of experiments*, two of them are on simulated data sets and two on real data sets. The *first* set consists of an example of several 3D registrations under varying noise conditions used to depict the robustness property of CCRE over MI and NMI in the presence of noise. The *second set* consists of 3D MR T1 and T2 weighted data sets obtained from the Montreal Neurological Institute (MNI) database. The data sets were artificially misaligned by known rigid transformations and our algorithm as well as the MI and NMI schemes were used to estimate the transformation and then compared. The *third set* consists of a real data experiments involving rigid motion. The data which we use in this set are CT-MR volume data sets obtained locally from the University Hospital. The fourth sets consists of MR-T1 and DWIs of a mouse brain where images differ by an affine motion in 3D (scaling between the images is due to the voxel resolution differences).

3.1 Robust property of the CCRE measure

In this section, we demonstrate the robustness property of CCRE and hence justifying the use of the CCRE measure in the registration problem. This is achieved by showing the noise immunity for CCRE over the MI and NMI algorithms. Noise, as one of the most typical degradation effects in the medical image, alters the intensity distribution of the image which may affect the MI registration criterion. The data we use for this experiment is an MR T1 and T2 image pair from the brainweb site at the Montreal Neurological Institute [2]. They are originally aligned with each other. A single corresponding slices are shown in Fig1. The two volumes are defined on a 1mm isotropic voxel grid in Talairach space, with dimension $(181 \times 217 \times 181)$. The effect of noise in

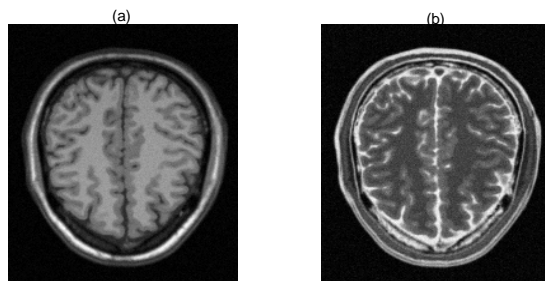


Fig. 1. Display of a single slice of aligned a) T1 weighted MR and b) T2 weighted MR images used in the computation of \mathcal{C} and \mathcal{I} over the range of rotations.

the input image pairs was evaluated by comparing CCRE, MI and NMI traces obtained for the degraded (via addition of noise) T1 and T2 image pair. The traces are computed over the rotation angle that was applied to the T2 image to achieve the misalignment between the T1 and T2 pair. In each plot of the Fig 2 the X-axis shows the rotation angle, while the Y-axis shows the values of CCRE, MI and NMI computed between the misaligned (by a rotation) image pairs. The original MR T1 and T2 data intensities range from 0-255 with the mean 55.6 and 60.6 respectively. Zero mean Gaussian noise of varying standard deviation 0, 20, 37 was added to the image pair and for each level of variance, CCRE, MI and NMI were computed between the two noisy pair of images. *Note that the SNR in these experiments was at a level where the Riccian noise in the MRI magnitude data can be well approximated by the Gaussian.* Fig.(2) shows (X-axis: rotation angle, Y-axis: CCRE, MI, NMI) that increasing the level of noise results in a decrease in magnitude of CCRE, MI and NMI values respectively. The range of the traces of all the three match measures also decreases. For each noise level, CCRE shows a significantly larger range of values compared to MI and NMI. This has a significant influence in shaping the robustness property of CCRE with respect to noise in the data as well as the capture range in finding the optimal alignment.

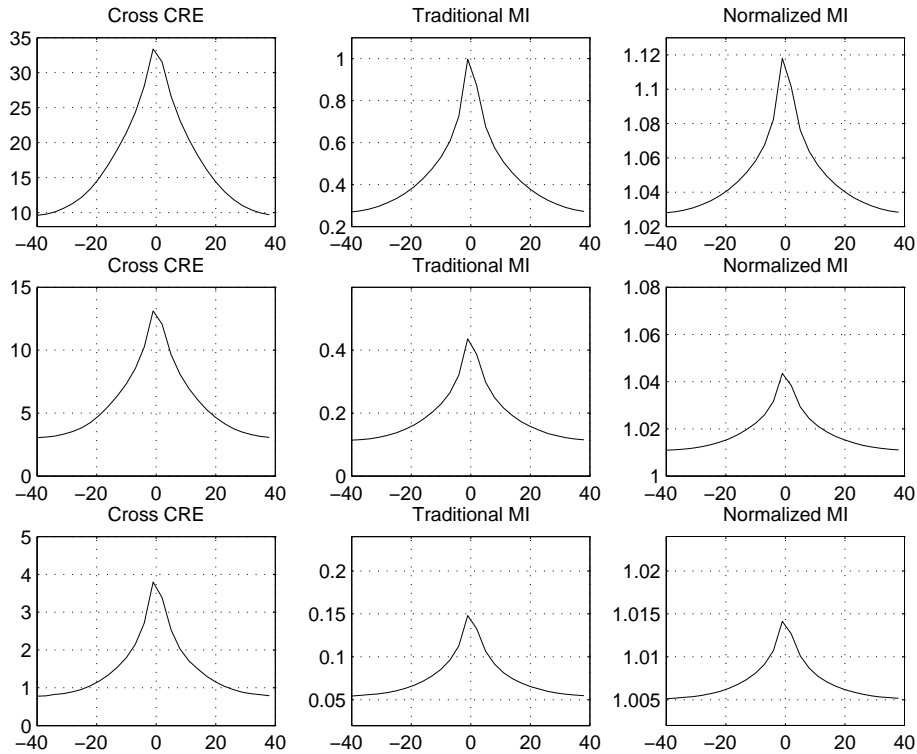


Fig. 2. Effects of the presence of the noise on CCRE, MI and NMI. CCRE, MI and NMI traces plotted for the misaligned T1 & T2 image pair where misalignment is generated by a rotation of the MR T2 weighted image over the range -40° to 40° . First row: no noise; Second row: $\sigma = 20$; Third row: $\sigma = 37$.

3.2 Synthetic Motion Experiments

In this section, we demonstrate the robustness property of CCRE and will make a case for its use over MI and NMI in the alignment problem. The case will be made via experiments depicting superior performance in matching the misaligned image pairs under noisy inputs and in addition via the depiction of a larger capture range in the estimation of the motion parameters.

Rigid Motion In this section, we show the algorithm performance for inter-modality rigid registrations. All the examples contain synthesized misalignments applied to same MR data sets as in the first experiment. With the MR T1 weighted image as the source, the target image is obtained by applying a known 3D rigid transformation to the MR T2 image.

Next, we applied CCRE, MI and NMI algorithms to estimate motion parameters in **30** cases of misaligned MR T1 and T2 pairs. The 30 cases were generated using 30 randomly generated rigid transformations to misalign the T1 and T2 image pair. These 30 transformations are normally distributed around the values

of $(10^\circ, 5mm)$, with standard deviations of $(3^\circ, 3mm)$ for rotation and translation respectively. Table 3.2 shows the statistics of errors resulting from the three different methods (CCRE, MI and NMI). Six parameters are displayed in each cell. The first three are rotation angles(in degrees), while the next three values show the translations(in mm). Both the rotation and translation parameters are in (x, y, z) order. For most of the cases, the average error of rotation angle is less than 0.5° and the translation is less than $0.5mm$. Out of the 30 trials, the traditional MI and NMI failed 3 times and 4 times respectively, while our CCRE does not fail for any of the 30 trials. (*“failed” here and subsequent usages means that the numerical algorithm for optimizing the cost function did not converge within 500 iterations of the optimizer*). If we only count the cases which resulted in acceptable results, as shown in the first (for CCRE), second (for MI) and third (for NMI) rows, CCRE, MI and NMI have a comparable performance, all being quite accurate.

	mean	standard deviation
CCRE	0.1255° 0.0905° 0.0654° 0.0764 0.0617 0.0536	0.1254° 0.1266° 0.0492° 0.0541 0.0540 0.0484
Traditional MI	0.1497° 0.1049° 0.1004° 0.0728 0.0499 0.0413	0.1792° 0.0857° 0.0952° 0.0616 0.0410 0.0374
Normalized MI (NMI)	0.2167° 0.1224° 0.1663° 0.0629 0.0881 0.0335	0.1411° 0.1003° 0.0869° 0.0574 0.1261 0.0372

Table 1. Statistics (computed from 30 cases) comparison for rigid motion estimation errors between CCRE, traditional MI and NMI.

In the second experiment, we compare the robustness of the three methods (CCRE, MI and NMI) in the presence of noise keeping the misalignment fixed. Still choosing the MR T1 image pair from the previous expt. as our source image, we generate the target image by applying a known rigid motion to the T2 weighted MR image. We conduct this experiment by varying the amount of Gaussian noise added and then for each instance of the added noise, we register the two images using the three techniques. We expect all the schemes to fail at some level of noise. By comparing the noise variance of the failure point, we can comment on the degree to which these methods are tolerant to noise. We choose the fixed motion to be 10° rotation, and 5 pixel translation in X and Y directions respectively. The numerical schemes we used to implement these registrations are all based on the sequential quadratic programming (SQP) algorithm [1]. Table 2 shows the registration results for the three schemes. From the table, we observe that the traditional MI fails when the σ of the noise is increased to 13. It is slightly better for NMI, which fails at 16, while CCRE is tolerant until $\sigma = 40$, a significant difference when compared to the traditional MI and the NMI methods. This experiment conclusively depicts that CCRE has more noise immunity than both traditional MI and the NMI. Next, we fix the variance of noise and vary the magnitude of the synthetic motion until all of the methods fail. With this experiment, we can compare the convergence range for each registration scheme. In all the methods, we use the sequential quadratic programming technique

σ	Cross CRE	traditional MI	normalized MI
10	9.9809° 9.9924° 10.0497° 6.9833 6.8536 7.0635	9.9523° 10.0038° 10.0497° 7.0067 6.9599 7.0213	10.1356° 9.7861° 10.5424° 7.0156 6.9311 7.0976
12	9.9351° 10.0325° 10.0554° 7.0134 6.9272 7.0341	10.1471° 10.1643° 10.2158° 6.9985 6.8495 6.9777	10.1757° 10.4617° 10.4603° 6.9515 6.7816 6.9114
15	9.9363° 10.0327° 10.0550° 7.0130 6.9271 7.0343	FAILED	10.0268° 9.9580° 9.9064° 6.8633 6.7924 7.1617
16	9.9679° 9.9924° 10.0140° 6.9693 7.0948 7.0126		FAILED
40	9.6314° 9.8377° 10.2273° 6.9840 7.0960 7.0080		
42	FAILED		

Table 2. Comparison of the registration results between CCRE and other MI algorithms for a fixed synthetic motion and varying noise. The image intensity range before adding noise is 0-255. The true motion here is $(10^\circ, 10^\circ, 10^\circ, 7, 7, 7)$

to estimate the optimal motion. From Table 3, we find that the convergence range of traditional MI and NMI is estimated at $(12^\circ, 12^\circ, 14^\circ, 12, 12, 12)$ and $(15^\circ, 15^\circ, 15^\circ, 15, 15, 15)$ respectively, while our new algorithm has a much larger capture range at $(32^\circ, 32^\circ, 25^\circ, 31, 31, 31)$. It is evident from this experiment that the capture range for reaching the optimum is significantly larger for CCRE when compared with MI and NMI in the presence of noise.

GroundTruth	CCRE	traditional MI	normalized MI
7° 7° 7° 5 5 5	6.970° 6.912° 7.014° 4.952 5.001 5.073	6.973° 7.003° 7.057° 5.003 4.942 5.069	6.955° 7.186° 6.954° 5.015 4.931 5.097
12° 12° 15° 12 12 12	11.961° 12.033° 15.055° 12.013 11.921 12.034	FAILED	12.025° 12.980° 11.990° 11.843 12.094 11.916
15° 15° 15° 15 15 15	15.031° 15.008° 15.012° 14.963 14.955 15.020		FAILED
32° 32° 25° 31 31 31	32.086° 32.034° 25.092° 32.037 31.932 31.967		
33° 33° 33° 33 33 33	FAILED		

Table 3. Comparison of the convergence range of the rigid registration between CCRE and other MI schemes for fixed noise standard deviation 7.

3.3 Real data 3D rigid motion

In this section, we present the performance of our method on data containing real rigid misalignments. The results are compared to ground-truth registration obtained semi-automatically by an "expert", whose registrations are currently used in clinical practice at University hospital. For the purpose of comparison, we also apply traditional MI implemented as was presented in Collignon et al. [5], to these data. We tested our algorithm and the MI based technique on MR-CT data from eight different subjects. The MR-CT pairs were misaligned due to the motion of the subject. The CT image is of size $(512, 512, 120)$ while the MR image size is $(512, 512, 142)$, and the voxel dimensions are $(0.46, 0.46, 1.5)$ and $(0.68, 0.68, 1.05)$ for CT and MR respectively.

Table 4 summarizes the results of the comparison for the *eight* data sets. The row labeled "True" in the table shows the ground truth rotation and translation parameters (as assessed semi-automatically by the local expert) of each data set. Both rotation and translation parameters are in (x,y,z) order. The row labeled, "CCRE Err." depicts the absolute error between the ground truth and the estimated parameters using our method, while "MI Err." indicates the corresponding errors for the MI method. As evident, our (CCRE-based) algorithm has achieved higher accuracy in the registration of these eight dataset. For most of the cases, the average error of rotation angle is less than 0.5° . and the translation error is less than 0.5mm for our algorithm. The maximum error of six parameters are $(0.412^\circ, 0.366^\circ, 1.0188^\circ, 0.2519\text{mm}, 0.5110\text{mm}, 0.5471\text{mm})$, while for traditional MI, the maximum errors in the parameters are $(0.795^\circ, 1.177^\circ, 2.681^\circ, 1.060\text{mm}, 1.525\text{mm}, 0.584\text{mm})$, which shows that our algorithm is more reliable than the MI based alignment algorithm. In the large motion cases, we observed that the CCRE converges very fast as compared to the traditional MI scheme where we need to resort to a multi-resolution implementation to find the solution in a coarse-to-fine framework and this takes extra computational effort to get the two images aligned.

3.4 Real data affine motion

In this section, we demonstrate the experimental results for the affine motion estimation. The data we used in our experiments is a pair of MR images of mouse brain. The source image is $(46.875 \times 46.875 \times 46.875)$ micron resolution with the field of view $(2.4 \times 1.2 \times 1.2\text{cm})$, while the target is 3D diffusion-weighted image with $(52.734 \times 52.734 \times 52.734)$ micron resolution with the field of view $(2.7 \times 1.35 \times 1.35\text{cm})$. Both the images have the same acquisition matrix $(256 \times 512 \times 256)$.

Figure 3.4 shows the registration results for the datasets. As is visually evident, the misalignment has been fully compensated for after the application of the estimated affine deformation.



Fig. 3. Affine registration of an MR-T1 & MR-DWI mouse brain scan. Left to Right: an arbitrary slice from the source image, a slice of the transformed source overlaid with the corresponding slice of the edge map of the target image and the target image slice.

Set	Item	Rotation (degree)	Translation (mm)
1	True	(5.455 - 1.146 - 14.003)	(3.822 9.254 3.1094)
	CCRE Err.	(0.257 0.247 0.606)	(0.076 0.364 0.124)
	MI Err.	(0.714 0.755 0.004)	(0.013 0.223 0.441)
2	True	(1.765 - 2.023 - 12.284)	(6.661 2.340 6.280)
	CCRE Err.	(0.297 0.316 1.018)	(0.252 0.108 0.053)
	MI Err.	(0.453 0.025 0.109)	(0.401 0.002 0.130)
3	True	(-5.099 7.357 - 18.472)	(9.790 - 0.901 - 0.228)
	CCRE Err.	0.040 0.272 0.419)	(0.072 0.168 0.212)
	MI Err.	(0.137 0.533 0.218)	(0.127 0.297 0.012)
4	True	(-5.581 - 2.865 - 21.675)	(10.561 - 4.306 18.874)
	CCRE Err.	(0.317 0.092 0.359)	(0.165 0.283 0.145)
	MI Err.	(0.539 0.254 1.025)	(0.121 0.253 0.415)
5	True	(-2.498 7.540 - 23.737)	(3.249 2.425 3.734)
	CCRE Err.	(0.298 0.238 0.414)	(0.001 0.0338 0.485)
	MI Err.	(0.795 1.313 1.479)	(0.334 0.115 0.584)
6	True	(-1.381 - 3.386 - 30.028)	(0.6022 7.4366 - 7.125)
	CCRE Err.	(0.155 0.060 0.111)	(0.133 0.148 0.415)
	MI Err.	(0.031 0.498 0.195)	(0.173 0.191 0.313)
7	True	(13.922 - 3.965 - 18.719)	(8.700 7.328 - 22.421)
	CCRE Err.	(0.412 0.1145 0.472)	(0.251 0.395 0.371)
	MI Err.	(0.487 1.461 1.833)	(1.060 0.818 0.179)
8	True	(14.024 3.569 - 21.778)	(0.120 12.970 - 9.870)
	CCRE Err.	(0.063 0.3662 0.515)	(0.082 0.5110 0.547)
	MI Err.	(0.115 1.177 2.681)	(0.500 1.525 0.770)
<i>Mean</i>	CCRE Err.	(0.230 0.213 0.489)	(0.129 0.251 0.294)
	MI Err.	(0.443 0.654 0.886)	(0.322 0.492 0.331)
<i>STD</i>	CCRE Err.	(0.131 0.111 0.258)	(0.089 0.164 0.184)
	MI Err.	(0.269 0.503 0.880)	(0.296 0.485 0.241)

Table 4. 3D rigid motion estimates for eight MR-CT data sets.

4 Summary

In this paper, we presented a novel measure of information that we dub cumulative residual entropy (CRE). This measure has several advantages over the traditional Shannon entropy whose definition is based on probability density functions which need not necessarily exist for some random variables. In addition, CRE can be easily computed from the sample data and these computations asymptotically converge to the true value. Unlike the Shannon entropy, the same CRE definition is valid for both discrete and continuous domains.

We defined a new measure of match called cross-CRE (CCRE) and applied it to estimate the parameterized misalignments between 3D image pairs and tested it on synthetic as well as real data sets from multi-modality (MR T1 and T2 weighted, MR & CT, MR T1 and DWIs) imaging sources. Comparisons were made between CCRE and traditional MI and normalized MI both of which were defined using the Shannon entropy. Experiments depicted significantly better performance of CCRE over the MI and NMI-based methods currently used in literature. Our future work will focus on extending the class of transformations to non-rigid motions.

Acknowledgements

Authors would like to thank, Dr. S.J. Blackband, Dr. S. C. Grant both of the Neuroscience Dept., UFL and Dr. H. Benveniste of SUNY, BNL for providing us with the mouse DWI data set, Dr. F. Bova of the Neurosurgery Dept. of UFL for providing us with the MR-CT pairs. SJB and SCG were supported by NIH-NCRR P41 RR16105 and HB was supported by NIH-RO1 EB00233-04.

References

1. D.P. Bertsekas, Nonlinear programming, Athena Scientific.
2. Simulated brain database [Online]. <http://www.bic.mni.mcgill.ca/brainweb/>
3. C. Chéfd'Hotel, G. Hermosillo and O. Faugeras, "A variational approach to multi-modal image matching," in IEEE Workshop on VLSM, pp. 21-28, 2001, Vancouver, BC, Canada.
4. Y. Chen, M. Rao, B. C. Vemuri and F. Wang, "Cumulative residual entropy, a new measure of information," Institute of Fundamental Theory, Technical Report IFT-Math-01-02, University of Florida, Gainesville, Florida.
5. A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, and P. S. ang G. Marchal, "Automated multimodality image registration using information theory," *Proc. IPMI*, Y.J.C.Bizais, Ed., pp. 263-274, 1995.
6. Thomas M. Cover, Joy A. Thomas, *Elements of Information Theory*, John Wiley and Sons, 1991.
7. J. N. Kapur, Measure of Information and their Applications, John Wiley & Sons Inc.
8. S. H. Lai and M. Fang, "Robust and efficient image alignment with spatially varying illumination models," in IEEE CVPR 1999, pp. 167-172.
9. J. Liu, B. C. Vemuri and J. L. Marroquin, "Local frequency representations for robust multi-modal image registration," *IEEE Trans. on Medical Imaging*, Vol. 21, No. 5, 2002, pp. 462-469.
10. M. Leventon and W. E. L. Grimson, "Multi-modal volume registration using joint intensity distributions," in MICCAI 1999.
11. J.B. Maintz and M. A. Viergever, "A Survey of Medical Image Registration," *MedIA* Vol. 2, pp. 1-36, 1998.
12. A. Roche, G. Mandalain, X. Pennec and N. Ayache, "The correlation ratio as new similarity metric for multi-modal image registration, in MICCAI'98.
13. D. Ruckert, C. Hayes, C. Studholme, M. leacha nd D. Hawkes, " Non-rigid registration of breast MRI using MI," in MICCAI'98.
14. C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379-423 and 623-656, July and October, 1948.
15. C. Studholme, D. L. G. Hill and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognition*, Vol. 32, pp. 71-86, 1999
16. R. Szeliski, J. Coughlan, "Spline-based image registration," *IJCV*, v.22 n.3, p.199-218, March/April 1997
17. B. C. Vemuri, S. Huang, S. Sahni, C. M. Leonard, C. Mohr, R. Gilmore and J. Fitzsimmons, "An efficient motion estimator with application to medical image registration", *Medical Image Analysis*, Oxford University Press, Vol.2, No. 1, pp. 79-98, 1998 .
18. P. A. Viola and W. M. Wells, "Alignment by maximization of mutual information," in *Fifth ICCV*, MIT, Cambridge, MA, pp. 16-23, 1995