Overlap Invariance of Cumulative Residual Entropy Measures for Multimodal Image Alignment

Nathan D. Cahill^{a,b}, Julia A. Schnabel^a,
J. Alison Noble^a and David J. Hawkes^c

^aInstitute of Biomedical Engineering, University of Oxford, Oxford OX3 7DQ, UK;

^bResearch and Innovation, Carestream Health, Inc., Rochester, NY 14608, USA;

^cCentre for Medical Image Computing, University College London, London WC1E 6BT, UK

ABSTRACT

Cumulative residual entropy (CRE)^{1,2} has recently been advocated as an alternative to differential entropy for describing the complexity of an image. CRE has been used to construct an alternate form of mutual information (MI)^{3,4} called *symmetric cross cumulative residual entropy* (SCCRE)⁵ or *symmetric cumulative mutual information* (SCMI).⁶ This alternate form of MI has exhibited superior performance to traditional MI in a variety of ways.⁵ However, like traditional MI, SCCRE suffers from sensitivity to the changing size of the overlap between images over the course of registration. Alternative similarity measures based on differential entropy, such as normalized mutual information (NMI),⁷ entropy correlation coefficient (ECC)^{8,9} and modified mutual information (M-MI),¹⁰ have been shown to exhibit superior performance to MI with respect to the overlap sensitivity problem. In this paper, we show how CRE can be used to compute versions of NMI, ECC, and M-MI that we call the *normalized cross cumulative residual entropy* (NCCRE), *cumulative residual entropy correlation coefficient* (CRECC), and *modified symmetric cross cumulative residual entropy* (M-SCCRE). We use publicly available CT, PET, and MR brain images* with known ground truth transformations to evaluate the performance of these CRE-based similarity measures for rigid multimodal registration. Results show that the proposed similarity measures provide a statistically significant improvement in target registration error (TRE) over SCCRE.

Keywords: Registration

1. INTRODUCTION

In multimodal registration applications, there may not exist any functional relationship between data in the reference and floating images. However, a more general relationship can be assumed by analyzing the information content in both images. Intuitively, this general relationship can be stated in the following way: a combined version of two misaligned images should be more complex than the combined version of the same two images when they are aligned.

Viola and Wells³ and Collignon⁴ recognized that this general relationship could be quantified by the mutual information (MI) between the images, and they described how multimodal registration could be performed by maximizing this quantity. In the years since the introduction of MI for registration, many researchers have investigated potential areas for improvement. In this paper, we focus on two such areas: 1) the use of cumulative residual entropy^{1,2} instead of traditional (differential) entropy for constructing MI, and 2) the generalization of MI to be insensitive to changes in the size of the overlap region between the images.^{7,10}

Our first area of focus is on the use of cumulative residual entropy (CRE).^{1,2} CRE uses the survival function of a random variable to describe its information content in a manner that is more general than differential entropy. CRE can be used to construct an alternate form of MI, called *cross cumulative residual entropy* (CCRE)⁵ or cumulative mutual information (CMI).⁶ Wang *et al.*⁵ showed that CCRE is a desirable alternative to MI

Send correspondence to Nathan D. Cahill: cahill@robots.ox.ac.uk

*The images and data were provided as part of the project: "Retrospective Image Registration Evaluation," National Institutes of Health, Project Number 8R01EB002124-03, Principal Investigator, J. Michael Fitzpatrick, Vanderbilt University, Nashville, TN, USA.

for multimodal registration. They performed a variety of experiments to establish that CCRE exhibits better regularity (due to being based on cumulative density functions), faster convergence, robustness to greater noise levels, and better performance for partial overlap problems than MI.

Our second area of focus is on the problem of overlap sensitivity in MI-based registration. Studholme *et al.*⁷ argued that any image similarity measure should be invariant to changes in the size of the overlap region through the course of registration. They used some simple examples to illustrate that MI values vary with changing overlap, and they proposed the *normalized mutual information* (NMI) as an alternative. Cahill *et al.*¹⁰ built off of Studholme's example and showed that even though NMI provides an improvement over MI, it still does not guarantee overlap insensitivity in every situation. They showed that with knowledge of the background statistics of both images, a different type of modification can be made to MI in order to *guarantee* overlap insensitivity under certain assumptions.

In this paper, we bring together these two areas and examine the behavior of CRE-based similarity measures as the overlap region changes through the course of registration. We establish that CCRE suffers from the *same* overlap sensitivity problem as MI. We then show that a normalized version of CCRE can be constructed that is analogous to Studholme's NMI. Furthermore, we use Cahill's technique¹⁰ to develop modified versions of CRE-based similarity measures that have guaranteed overlap insensitivity. Finally, we use publicly available¹¹ CT, PET, and MR brain images along with known ground truth transformations to evaluate the performance of the CRE-based similarity measures in the context of rigid multimodal registration.

2. BACKGROUND

2.1 Mutual Information

In multimodal image registration applications, there is an underlying assumption about the relationship between two images. This assumption can be roughly stated in the following way: a combined version of two misaligned images should be more complex than the combined version of the same two images when they are aligned.

To intuitively describe this assumption, consider the argument made by Studholme *et al.*⁷ When two medical images are mis-aligned, a fused or combined image would contain duplicate versions of much of the underlying anatomy. When the same images are brought into alignment, the fused image should be simpler, containing no duplicate versions of the underlying anatomy. Studholme *et al.* illustrate this concept by showing that a fused image constructed from misaligned cranial images contains four eyes and four ears, and that the fused image becomes simpler when the cranial images are brought into alignment.

If each pixel of an image is considered as a realization of a continuous random variable (c.r.v.), one way to describe the complexity of the image is by the differential entropy of the random variable. For a c.r.v. A , the differential entropy $H(A)$ is given by:

$$H(A) = - \int_{-\infty}^{\infty} p_A(a) \log p_A(a) da, \quad (1)$$

where $p_A(a)$ is the value of the probability density function of A at a . Each pixel of a fused image can be considered a function of a joint realization of two random variables. The complexity of the fused image can therefore be described by the joint entropy of the two underlying random variables. The differential joint entropy of c.r.v.'s A and B is given by:

$$H(A, B) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{A,B}(a, b) \log p_{A,B}(a, b) da db, \quad (2)$$

where $p_{A,B}(a, b)$ is the value of the joint probability density function of (A, B) at (a, b) .

One problem with the use of joint entropy for image registration is that a reduction in joint entropy can be effected by a reduction in marginal entropy. This could yield an "optimal" alignment having the least amount of information rather than the most *corresponding* information. Viola and Wells³ and Collignon *et al.*⁴

independently proposed the use of mutual information to combat this problem. MI is defined as the difference between the sum of the marginal entropies and the joint entropy:

$$MI(A, B) := H(A) + H(B) - H(A, B). \quad (3)$$

Maximizing the mutual information (or minimizing the negative mutual information) will tend to simultaneously reduce joint entropy while maintaining the marginal entropies.

2.2 Cumulative Residual Entropy

Another way in which the complexity of an image can be described is by the *cumulative residual entropy* (CRE) of the underlying random variable. CRE was introduced by Wang *et al.*¹ and Rao *et al.*,² and generalized by Drissi *et al.*⁶ The general form is given by:

$$\varepsilon(A) = - \int_{-\infty}^{\infty} P(A > a) \log P(A > a) da, \quad (4)$$

where $P(A > a)$ is the value of the survival function (or reliability function) of A at a . The analogy of differential joint entropy is the *joint cumulative residual entropy* (JCRE), given by:

$$\varepsilon(A, B) = \varepsilon(A) + E[\varepsilon(B|A)], \quad (5)$$

where

$$\varepsilon(B|A) = - \int_{-\infty}^{\infty} P(B > b | A) \log P(B > b | A) db. \quad (6)$$

Note that unlike differential joint entropy, JCRE is not symmetric; i.e., $\varepsilon(A, B) \neq \varepsilon(B, A)$. A symmetrized version can easily be constructed, however; we define the *symmetric joint cumulative residual entropy* (SJCRE) as:

$$\begin{aligned} \varepsilon_S(A, B) &= \frac{1}{2} \left(\varepsilon(A, B) + \varepsilon(B, A) \right) \\ &= \frac{1}{2} \left(\varepsilon(A) + \varepsilon(B) + E[\varepsilon(A|B)] + E[\varepsilon(B|A)] \right). \end{aligned} \quad (7)$$

Similarity measures for multimodal registration can be constructed using the cumulative residual entropy in place of the differential entropy. Rao *et al.*² use the JCRE itself as a similarity measure. However, similarly to joint entropy, JCRE suffers from the problem that a reduction in JCRE can be effected by a reduction in CRE. A solution to this problem is to use the *cross cumulative residual entropy* (CCRE), which is analogous to mutual information:

$$\begin{aligned} CCRE(A, B) &:= \varepsilon(A) + \varepsilon(B) - \varepsilon(B, A) \\ &= \varepsilon(A) - E[\varepsilon(A|B)]. \end{aligned} \quad (8)$$

CCRE was introduced by Wang *et al.*,⁵ they established through a variety of nonrigid multimodal registration experiments that CCRE outperforms MI, exhibiting better regularity, faster convergence, robustness to greater noise levels, and better performance for partial overlap problems.

Since JCRE is not symmetric, neither is CCRE. However, CCRE can easily be symmetrized; the *symmetric cross cumulative residual entropy* (SCCRE) is given by:

$$\begin{aligned} SCCRE(A, B) &:= \frac{1}{2} \left(CCRE(A, B) + CCRE(B, A) \right) \\ &= \varepsilon(A) + \varepsilon(B) - \varepsilon_S(A, B). \end{aligned} \quad (9)$$

It is this symmetric version that we will use in the experiments in Section 5 of this paper.

(a) A reference disc image is held in position as the floating disc image is overlaid and rotated about its center by an amount θ .

(b) Probability mass values of reference image in overlap region as a function of θ , assuming square images with side length 3 and disc of unit radius.

Figure 1. Overlap sensitivity example presented by Studholme *et al.*⁷

2.3 Overlap Invariance

Studholme *et al.*⁷ provided some simple examples to illustrate that MI is not invariant to changes in the size of the overlap region through the course of registration. One example is of identical images of a disc centered on a dark background, as shown in Figure 1(a). The floating image is rotated about the center of the disc, which has the effect of changing the overlap region between the two images. This in turn affects any quantities that are observed in the overlap region, for example, the probability mass function of the reference image illustrated in Figure 1(b).

Studholme argued that as the floating image in Figure 1(a) is rotated about the center of the reference image, any useful image similarity measure should remain constant. However, this property does not necessarily hold, because image similarity measures are computed from the information contained in the changing overlap region. In fact, MI itself fails this test. Studholme proposed an alternative to MI, called the *normalized mutual information*, which is given by:

$$NMI(A, B) := \frac{H(A) + H(B)}{H(A, B)}. \quad (10)$$

NMI is closely related to Astola's *entropy correlation coefficient* (ECC),⁸ which was explored for multimodal image registration by Maes *et al.*⁹ The ECC is given by:

$$ECC(A, B) := \sqrt{2 - \frac{2H(A, B)}{H(A) + H(B)}}. \quad (11)$$

For the rotating disc example, the NMI and ECC remain constant as a function of θ , satisfying Studholme's notion of overlap invariance.

Cahill *et al.*¹⁰ revisited this problem, examining a slightly more complicated example than the rotating disc. They showed that if the floating image is changed from a disc to a ring whose center has the same intensity as the background, the NMI loses strict overlap invariance. Figure 2(a) shows this rotating ring example, and Figure 2(b) illustrates that the analytically computed NMI is only constant when the inner ring radius is zero (corresponding to the rotating disc example) or one (corresponding to no ring at all).

To further Studholme's goal of devising truly overlap invariant similarity measures, Cahill *et al.*¹⁰ proposed modifying similarity measures with terms based on the background statistics of each image. Under the assumption that the only content moving in and out of the overlap region is background, they showed that these modified similarity measures are guaranteed to possess the desired invariance.

To describe the similarity measure modifications, suppose the original overlap region between images A and B is denoted Ω_0 , and the overlap region at time t is denoted Ω_t , where "time" measures the progress of registration.

(a) A reference disc image is held in position as the floating ring image is overlaid and rotated about its center by an amount θ .

(b) NMI values for rings with various inner radii, expressed as functions of θ .

Figure 2. Overlap sensitivity example presented by Cahill *et al.*¹⁰

Cahill *et al.*¹⁰ use conditioning arguments to show that the distributions of A and B are related by:

$$p_A^{\Omega_0}(a) = (1 + \omega_t) p_A^{\Omega_t}(a) - \omega_t \tilde{p}_A(a), \quad (12)$$

$$p_B^{\Omega_0}(b) = (1 + \omega_t) p_B^{\Omega_t}(b) - \omega_t \tilde{p}_B(b), \quad (13)$$

$$p_{A,B}^{\Omega_0}(a, b) = (1 + \omega_t) p_{A,B}^{\Omega_t}(a, b) - \omega_t \tilde{p}_{A,B}(a, b), \quad (14)$$

where $\omega_t = (|\Omega_t| - |\Omega_0|) / |\Omega_0|$, and where $\tilde{p}_A$ and $\tilde{p}_B$ are the distributions of the background content in images A and B , respectively. These background distributions are assumed to be determined prior to the registration process. The backgrounds are also assumed to be independent, so the joint background distribution $\tilde{p}_{A,B}$ can simply be defined by the product of the marginals; i.e., $\tilde{p}_{A,B}(a, b) := \tilde{p}_A(a) \tilde{p}_B(b)$.

As t increases and the overlap region changes, the right hand sides of (12)–(14) remain constant. This property is exploited to form overlap invariant versions of entropy and joint entropy, which are defined by:

$$MH(A) := - \int_{-\infty}^{\infty} q_A(a) \log q_A(a) da, \quad (15)$$

$$MH(A, B) := - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} q_{A,B}(a, b) \log q_{A,B}(a, b) da db, \quad (16)$$

where

$$q_A(a) = (1 + \omega) p_A(a) - \omega \tilde{p}_A(a), \quad (17)$$

$$q_{A,B}(a, b) = (1 + \omega) p_{A,B}(a, b) - \omega \tilde{p}_{A,B}(a, b). \quad (18)$$

A modified version of MI, called *modified mutual information* (M-MI), is then constructed from the modified entropy and modified joint entropy; i.e.,

$$M-MI(A, B) := MH(A) + MH(B) - MH(A, B). \quad (19)$$

Modified versions of NMI and ECC are also constructed in the same manner, by substituting modified versions of entropy and joint entropy into (10) and (11). Cahill *et al.*¹⁰ illustrate that these modified similarity measures, when applied to the rotating ring example of Figure 2(a), are constant with respect to θ independent of the inner ring radius.

3. CRE-BASED VERSIONS OF NMI AND ECC

It is fairly easy to establish that SCCRE suffers from the same overlap sensitivity problem as MI. If we analyze the rotating disc example in Figure 1(a), we can see from Figure 3(a) that the survival function of the reference image is sensitive to changes in θ . This causes the CRE and JCRE values to vary with θ , which in turn makes the SCCRE sensitive to θ , as shown in Figure 3(b).

Figure 3. Example of overlap sensitivity of SCCRE for the rotating disc example of Figure 1(a).

We can attempt to remedy this problem by using the CRE to form analogous versions of NMI and ECC, which we call the *normalized cross cumulative residual entropy* (NCCRE) and the *cumulative residual entropy correlation coefficient* (CRECC). These similarity measures are constructed in the same manner as NMI and ECC, but with differential entropy and differential joint entropy replaced by CRE and symmetric JCRE:

$$\begin{aligned} NCCRE(A, B) &:= \frac{\varepsilon(A) + \varepsilon(B)}{\varepsilon_S(A, B)} \\ &= \frac{2\varepsilon(A) + 2\varepsilon(B)}{\varepsilon(A) + \varepsilon(B) + E[\varepsilon(A|B)] + E[\varepsilon(B|A)]}, \end{aligned} \quad (20)$$

$$CRECC(A, B) := \sqrt{1 - \frac{E[\varepsilon(A|B)] + E[\varepsilon(B|A)]}{\varepsilon(A) + \varepsilon(B)}}. \quad (21)$$

If we analytically compute NCCRE and CRECC values for the rotating disc example, we find that they exhibit Studholme's desired behavior. Figure 4 illustrates the constancy of NCCRE and CRECC with respect to θ .

Since NCCRE and CRECC are based on the same building blocks as CCRE and SCCRE (namely, CRE and JCRE), implementation of registration algorithms using NCCRE and CRECC is straightforward. Any derivatives of NCCRE and CRECC with respect to transformation parameters can be constructed by applying various calculus rules to the already established derivatives^{1,5} of CRE and JCRE. In Section 5, we show some results from applying NCCRE and CRECC to the rigid multimodal brain image registration problem.

Figure 4. NCCRE and CRECC values for the reference and floating images of Figure 1(a), computed over the overlap region and plotted as functions of θ .

4. MODIFIED VERSIONS OF CRE-BASED SIMILARITY MEASURES

Even though NCCRE and CRECC do tend to diminish the overlap sensitivity problem, they do not guarantee overlap invariance. If we compute their values for the rotating ring example of Figure 2(a), we see from Figure 5 that they behave in a manner consistent with NMI. If the inner radius of the ring takes on the values of $r = 0$ or $r = 1$, the NCCRE and CRECC do not vary with θ ; however, they *do* vary with θ for any other inner radius.

To address this problem, we can modify SCCRE (and NCCRE and CRECC) in a manner similar to the modification of MI in Section 2.3. We break this process down into two parts, separately considering how to modify the CRE and the JCRE.

To modify the CRE, we begin by integrating both sides of (12) to observe the behavior of the survival function as the overlap region changes from Ω_0 to Ω_t :

$$P^{\Omega_0}(A > a) = (1 + \omega_t) P^{\Omega_t}(A > a) - \omega_t \tilde{P}(A > a), \quad (22)$$

where $\tilde{P}(A > a)$ is the survival function of the background content of images A evaluated at a . We exploit the fact that the right hand side of (22) remains constant as t increases in order to form a modified version of CRE:

$$M\varepsilon(A) := - \int_{-\infty}^{\infty} Q_A(a) \log Q_A(a) da, \quad (23)$$

Figure 5. NCCRE and CRECC as functions of rotation angle θ and inner ring radius r for the example in Figure 2(a).

where

$$Q_A(a) = (1 + \omega) P(A > a) - \omega \tilde{P}(A > a). \quad (24)$$

To modify the JCRE, we first concentrate on the term $E[\varepsilon(B|A)]$, which is defined by:

$$E[\varepsilon(B|A)] = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_A(a) P(B > b|A = a) \log P(B > b|A = a) db da. \quad (25)$$

If we apply Bayes' rule to (14), integrate the conditional p.d.f.'s, and assume that the backgrounds of A and B are independent, we have:

$$p_A^{\Omega_0}(a) P^{\Omega_0}(B > b|A = a) = (1 + \omega_t) p_A^{\Omega_t}(a) P^{\Omega_t}(B > b|A = a) - \omega_t \tilde{p}_A(a) \tilde{P}(B > b). \quad (26)$$

This suggests that the term $p_A(a) P(B > b|A = a)$ in (25) can be replaced with:

$$Q_{A,B}(a, b) = (1 + \omega) p_A(a) P(B > b|A = a) - \omega \tilde{p}_A(a) \tilde{P}(B > b). \quad (27)$$

To determine how to modify the $\log P(B > b|A = a)$ term in (25), we divide both sides of (26) by $p_A^{\Omega_0}(a)$ and plug in (12) to yield:

$$P^{\Omega_0}(B > b|A = a) = \frac{(1 + \omega_t) p_A^{\Omega_t}(a) P^{\Omega_t}(B > b|A = a) - \omega_t \tilde{p}_A(a) \tilde{P}(B > b)}{(1 + \omega_t) p_A^{\Omega_t}(a) - \omega_t \tilde{p}_A(a)}. \quad (28)$$

This suggests that the term $\log P(B > b|A = a)$ in (25) can be replaced with $\log (Q_{A,B}(a, b) / q_A(a))$, where $q_A(a)$ is given by (17). We can form a modified version of $E[\varepsilon(B|A)]$ by combining these results:

$$ME[\varepsilon(B|A)] := - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Q_{A,B}(a, b) \log \left(\frac{Q_{A,B}(a, b)}{q_A(a)} \right) db da. \quad (29)$$

This allows us to define modified versions of the JCRE and symmetric JCRE by combining (23) and (29):

$$M\varepsilon(A, B) := M\varepsilon(A) + ME[\varepsilon(B|A)], \quad (30)$$

$$M\varepsilon_S(A, B) := \frac{1}{2} (M\varepsilon(A, B) + M\varepsilon(B, A)). \quad (31)$$

Finally, we can use (23), (30) and (31) to define modified versions of CCRE and SCCRE:

$$M\text{-}CCRE(A, B) := M\varepsilon(A) + M\varepsilon(B) - M\varepsilon(B, A), \quad (32)$$

$$M\text{-}SCCRE(A, B) := M\varepsilon(A) + M\varepsilon(B) - M\varepsilon_S(A, B). \quad (33)$$

The NCCRE and CRECC can also be modified using the same components; i.e.,

$$M\text{-}NCCRE(A, B) := \frac{M\varepsilon(A) + M\varepsilon(B)}{M\varepsilon_S(A, B)}, \quad (34)$$

$$M\text{-}CRECC(A, B) := \sqrt{1 - \frac{ME[\varepsilon(A|B)] + ME[\varepsilon(B|A)]}{M\varepsilon(A) + M\varepsilon(B)}}. \quad (35)$$

If we now go back to the ring example of Figure 2(a) and analytically compute the modified versions of SCCRE, NCCRE, and CRECC for various values of r and θ , we see from Figure 6 that these modifications yield similarity measures that are constant with respect to θ for *any* inner ring radius r .

Figure 6. Values of modified SCCRE, NCCRE and CRECC as functions of rotation angle θ and inner ring radius r for the example in Figure 2(a).

5. REGISTRATION EXPERIMENT

In order to illustrate the behavior of the modified similarity measures on real-world data, we focus on the rigid registration case, and we use images from the Retrospective Image Registration Evaluation project. The RIRE project database contains CT, MR, and PET images for a variety of patients, and has a sequestered set of ground truth rigid body transformations that were computed from fiducial markers implanted in the skull. (The fiducial markers were removed from the images prior to retrospectively evaluating registration algorithms.) Results of the original RIRE study are provided by West *et al.*¹¹

In this paper, we used the images from nine patient datasets. Each patient dataset contains MR images from some or all of the following protocols: T1-weighted, T2-weighted, PD-weighted, and rectified versions of the T1, T2, and PD-weighted images. Five of the nine datasets contain both CT and PET images in addition to the MR images. Two of the datasets contain CT but not PET images, and the remaining two datasets contain PET but not CT images. The CT images have resolution $0.65 \times 0.65 \times 4.0 \text{ mm}^3$, the MR images have approximate resolution $1.25 \times 1.25 \times 4.0 \text{ mm}^3$, and the PET images has resolution $2.59 \times 2.59 \times 8.0 \text{ mm}^3$. For ease of computation, we resampled each image to $6.0 \times 6.0 \times 6.0 \text{ mm}^3$ isotropic resolution.

Examples of some of the RIRE images are shown in Figures 7 and 8. Figure 7 shows axial views of the CT, MR-T1, and PET images from patient 5. Figure 8 illustrates overlayed isosurfaces of the CT (blue) and MR-T1 images from patient 4, both before (left) and after (right) rigid registration.

For each dataset, we rigidly registered the CT and/or PET image to all of the MR images. Rigid transformations were parameterized by three Euler angles and three translation parameters. Each of the similarity

Figure 7. Axial slices of CT, MR (T1-weighted) and PET images from the patient 5 dataset.

Figure 8. Isosurfaces of CT (blue) and MR-T1 (red) images of patient 4, before and after rigid registration.

measures and their modified versions were optimized over the six-dimensional parameter space using the sequential quadratic programming (SQP) method implemented in MATLAB's Optimization Toolbox. Initial estimates of the solution were selected by translating the images to match their computed centroids.

All probability densities (and joint densities) were estimated via histograms (and joint histograms) that were constructed with 64 (or 64×64) equally spaced bins. Linear (or bilinear) interpolation was used to accumulate partial weights in neighboring bins. For the modified similarity measures, background distributions for each volume were estimated via histograms that were constructed from the data in manually selected background regions.

6. RESULTS

The performance of similarity measures on the various registration tasks is measured via Target Registration Error (TRE). For the RIRE project,¹¹ a number of anatomically meaningful volumes of interest (VOI) were annotated. TRE is computed as the average Euclidean distance (in mm) between VOI centroids in the reference image and their predicted positions after registration.

Table 1 reports the mean, median, and standard deviation of the TRE values measured across the VOI's in every patient for each similarity measure. Figure 9 shows box and whiskers plots that illustrate the median, quartiles, interquartile range, and outliers of the TRE values. These results suggest that all of the proposed measures outperform SCCRE for both CT/MR and PET/MR registration. Improvements are seen not only in mean/median TRE, but also in the extent of outliers in the PET/MR case.

We employed hypothesis testing to gauge the statistical significance of these results. Given similarity measures S_1 and S_2 , the null hypothesis states that the TRE's from S_1 and S_2 arise from the same distribution. The

Modality	Statistic	SCCRE	NCCRE	CRECC	Modified SCCRE	Modified NCCRE	Modified CRECC
CT/MR	Mean TRE (mm)	6.6	4.5	4.1	4.6	4.5	4.3
	Median TRE (mm)	5.7	3.3	3.1	3.5	3.2	3.3
	Std. dev. TRE (mm)	4.0	3.4	2.9	3.3	3.3	3.1
PET/MR	Mean TRE (mm)	7.6	4.6	4.3	4.3	4.3	4.0
	Median TRE (mm)	5.1	3.9	3.6	3.4	3.2	3.3
	Std. dev. TRE (mm)	7.3	2.7	2.6	2.6	2.8	2.4

Table 1. Statistics of VOI TRE aggregated across all patients for each of the CRE-based similarity measures.

Figure 9. Box and whisker plots of TRE (in mm) after rigid registration of RIRE data using each similarity measure. "M-" prefixes denote modified similarity measures.

alternative hypothesis is that the TRE's from S_1 are "worse" than the TRE's from S_2 , in the sense that the c.d.f. values are smaller everywhere. Using the two-sample Kolmogorov-Smirnov test at a level $\alpha = 0.05$, we found that the null hypothesis can be rejected in favor of the alternative hypothesis for the following cases:

1. CT/MR Registration:
 - (a) $S_1 = \text{SCCRE}$; $S_2 = \text{NCCRE}, \text{CRECC}, \text{M-SCCRE}, \text{M-NCCRE}, \text{and M-CRECC}.$
 - (b) $S_1 = \text{M-SCCRE}$; $S_2 = \text{CRECC}.$
2. PET/MR Registration:
 - (a) $S_1 = \text{SCCRE}$; $S_2 = \text{NCCRE}, \text{CRECC}, \text{M-SCCRE}, \text{M-NCCRE}, \text{and M-CRECC}.$
 - (b) $S_1 = \text{NCCRE}$; $S_2 = \text{CRECC}, \text{M-SCCRE}, \text{M-NCCRE}, \text{and M-CRECC}.$
 - (c) $S_1 = \text{CRECC}$; $S_2 = \text{M-NCCRE}.$

This shows that in both the CT/MR and PET/MR registration cases, statistically significant improvements can be made by choosing *any* of the similarity measures proposed in this paper over SCCRE. The differences between NCCRE/CRECC versus modified similarity measures are not significant in most cases, except with respect to the inferior performance of NCCRE for PET/MR registration.

7. CONCLUSIONS AND FUTURE WORK

In this paper, we have shown how CRE-based similarity measures can be constructed to diminish and avoid the overlap sensitivity problem inherent in MI-based and SCMI-based multimodal image registration. These similarity measures are defined in ways that are analogous to Studholme's NMI,⁷ Astola's ECC⁸ and Cahill's modified similarity measures.¹⁰ Using CT, PET, and MR brain data with known ground truth transformations, we showed that statistically significant improvements can be achieved by using the proposed similarity measures for rigid multimodal registration.

In the future, we plan to analyze how the proposed similarity measures compare to SCCRE with respect to noise robustness, sensitivity to initial alignment estimate, and severe partial overlap. Furthermore, we plan to explore the behavior of the similarity measures for nonrigid registration.

ACKNOWLEDGMENTS

The first author would like to thank Fei Wang and Baba Vemuri for helpful discussions.

REFERENCES

1. F. Wang, B. C. Vemuri, M. Rao, and Y. Chen, "A new & robust information theoretic measure and its application to image alignment," in *Proceedings of the International Conference of Information Processing in Medical Imaging (IPMI 2003)*, pp. 388–400, July 2003.
2. M. Rao, Y. Chen, B. C. Vemuri, and F. Wang, "Cumulative residual entropy: A new measure of information," *IEEE Transactions on Information Theory* **50**, pp. 1220–1228, June 2004.
3. P. Viola and W. M. Wells, III., "Alignment by maximization of mutual information," in *Proceedings of the Fifth International Conference on Computer Vision*, pp. 16–23, June 1995.
4. A. Collignon, D. Vandermeulen, P. Suetens, and G. Marchal, "3D multi-modality medical image registration using feature space clustering," in *Proceedings of the First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine, Lecture Notes in Computer Science* **905**, pp. 193–204, Springer Berlin, 1995.
5. F. Wang and B. C. Vemuri, "Non-rigid multi-modal image registration using cross-cumulative residual entropy," *International Journal of Computer Vision* **74**, pp. 201–205, August 2007.
6. N. Drissi, T. Chonavel, and J. M. Boucher, "Generalized cumulative residual entropy for distributions with unrestricted supports," *Research Letters in Signal Processing* **2008**(790607), 2008.
7. C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognition* **32**, pp. 71–86, 1999.
8. J. Astola and I. Virtanen, "A measure of overall statistical dependence based on the entropy concept," in *Proceedings of the University of Vaasa*, (91), 1983.
9. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging* **16**, pp. 187–198, April 1997.
10. N. D. Cahill, J. A. Schnabel, J. A. Noble, and D. J. Hawkes, "Revisiting overlap invariance in medical image alignment," in *Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pp. 1–8, June 2008.
11. J. West *et al.*, "Comparison and evaluation of retrospective intermodality brain image registration techniques," *Journal of Computer Assisted Tomography* **21**, pp. 554–566, Jul–Aug 1997.