Normalized Measures of Mutual Information with General Definitions of Entropy for Multimodal Image Registration

Nathan D. Cahill

Center for Applied and Computational Mathematics,
School of Mathematical Sciences, Rochester Institute of Technology,
85 Lomb Memorial Drive, Rochester, NY 14618, USA
nathan.cahill@rit.edu

Abstract. Mutual information (MI) was introduced for use in multimodal image registration over a decade ago [1,2,3,4]. The MI between two images is based on their marginal and joint/conditional entropies. The most common versions of entropy used to compute MI are the Shannon and differential entropies; however, many other definitions of entropy have been proposed as competitors. In this article, we show how to construct normalized versions of MI using any of these definitions of entropy. The resulting similarity measures are analogous to normalized mutual information (NMI), entropy correlation coefficient (ECC), and symmetric uncertainty (SU), which have all been shown to be superior to MI in a variety of situations. We use publicly available CT, PET, and MR brain images[1] with known ground truth transformations to evaluate the performance of the normalized measures for rigid multimodal registration. Results show that for a number of different definitions of entropy, the proposed normalized versions of mutual information provide a statistically significant improvement in target registration error (TRE) over the non-normalized versions.

Keywords: Image registration, mutual information, entropy.

1 Introduction

Collignon and Maes [1,2] and Viola and Wells [3,4] introduced the idea that multimodal images could be aligned by maximizing their mutual information (MI), which is an information measure that depends on the marginal and joint entropies of the underlying images. Since that introduction, much research has gone into understanding, applying, and generalizing this idea in a variety of ways. Pluim et al. [5] captured a snapshot of the state of the art in 2003, which showed that much progress had already been attained by that point.

[1] 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.

B. Fischer, B. Dawant, and C. Lorenz (Eds.): WBIR 2010, LNCS 6204, pp. 258–268, 2010.

This paper focuses on two threads of investigation that emerged in the years after mutual information was introduced to the medical imaging community. One thread is the normalization of MI, in the form of Studholme's normalized MI (NMI) [6], Maes' entropy correlation coefficient (ECC) [1], or Melbourne's symmetric uncertainty (SU) [7], which have been empirically shown to improve registration accuracy and robustness to truncation of the images. The other thread is the use of different generalizations of entropy (Rényi and Tsallis entropies [8], cumulative residual entropy [9,10], and generalized survival exponential entropy [11]) to form the MI measure; such generalizations have also been shown to improve registration performance in various situations.

In this paper, we show that normalized versions of MI can be formed when MI is constructed from *any* of the definitions of entropy. Furthermore, we illustrate experimentally that such normalized versions in general outperform their unnormalized counterparts for the task of rigid registration of multimodal brain images. This idea generalizes some results in the literature that extend NMI for use with Rényi and Tsallis entropies [8] and extend NMI and ECC for use with cumulative residual entropy [12,13].

The remainder of this paper is organized as follows: section 2 describes various definitions of entropy, and section 3 provides a general framework for constructing MI from any of these definitions of entropy. Section 4 defines versions of NMI, ECC, and SU from any of the general definitions of entropy, and section 5 illustrates experimentally how these normalized versions of MI outperform their unnormalized counterparts. Finally, section 6 draws some conclusions and presents ideas for future work.

2 Measures of Entropy

In the context of probability theory, entropy describes the amount of uncertainty associated with a random variable. Shannon [14] considered an information source having components $L_1, L_2, \ldots, L_n$ with associated probabilities of occurrence $p_1, p_2, \ldots, p_n$, and he showed that the quantity:

$$H = -\sum_{i=1}^{n} p_i \log p_i \tag{1}$$

is a measure of the uncertainty in the outcome of a particular event. This measure is generalized to the case of a continuous random variable X with density $p(x)$ by the *differential* entropy:

$$H(X) = -\int_{-\infty}^{\infty} p(x) \ln p(x)\, dx \ . \tag{2}$$

A wide variety of competing measures of entropy have emerged in other contexts; a number of these measures that can be considered generalizations of Shannon or differential entropy are shown in Table 1. Rényi [15] presented a family of

entropies that converge to the differential entropy as $\alpha \to 1$. Tsallis [16] proposed a family of entropies that is useful in describing non-additive systems. Exponential and generalized exponential entropies were introduced by Campbell [17] and Koski and Persson [18] and used for data compression. The Tsallis and exponential entropies converge to differential entropy as $\alpha \to 1$; the generalized exponential entropy can be thought of as being related to a generalized version of the Rényi entropy, both of which converge to differential entropy as $(\alpha, \beta) \to (1, 1)$.

Another type of entropy, the *cumulative residual* entropy (CRE), was presented by Rao *et al.* [19] and Wang *et al.* [9] in order to provide a way to accommodate random variables that do not have a defined density function. This is done by replacing the density function with the survival function $S(x) = P(|X| > x)$, and then defining the CRE as:

$$\varepsilon(X) = -\int_0^\infty S(x) \ln S(x)\, dx \ . \tag{3}$$

All of the generalizations of differential entropy listed in Table 1 can also be defined in terms of CRE, as shown in Table 2. Zografos and Nadarajah [20] presented and analyzed generalizations of CRE to Rényi, exponential, and generalized exponential entropy, and they named the latter two resulting quantities *survival exponential* and *generalized survival exponential* entropies.

Table 1. Differential entropy and generalizations

	Symbol	Definition
Differential	$H(X)$	$-\int_{-\infty}^{\infty} p(x) \ln p(x)\, dx$
Rényi	$H_\alpha^R(X)$	$\begin{cases} \frac{1}{1-\alpha} \ln \int_{-\infty}^{\infty} p^\alpha(x)\, dx, & \alpha \neq 1 \\ H(X), & \alpha = 1 \end{cases}$
Tsallis	$H_\alpha^T(X)$	$\begin{cases} \frac{1}{1-\alpha} \int_{-\infty}^{\infty} (p^\alpha(x) - p(x))\, dx, & \alpha \neq 1 \\ H(X), & \alpha = 1 \end{cases}$
Exponential	$H_\alpha^E(X)$	$\exp\Big(H_\alpha^R(X)\Big)$
Generalized Rényi	$H_{\alpha,\beta}^{GR}(X)$	$\begin{cases} \frac{1}{\beta-\alpha} \ln \frac{\int_{-\infty}^{\infty} p^\alpha(x)\, dx}{\int_{-\infty}^{\infty} p^\beta(x)\, dx}, & \alpha \neq \beta \\ \frac{-\int_{-\infty}^{\infty} p^\beta(x) \ln p(x)\, dx}{\int_{-\infty}^{\infty} p^\beta(x)\, dx}, & \alpha = \beta \end{cases}$
Generalized Exponential	$H_{\alpha,\beta}^{GE}(X)$	$\exp\Big(H_{\alpha,\beta}^{GR}(X)\Big)$

Table 2. Cumulative residual entropy (CRE) and generalizations

	Symbol	Definition
Cumulative Residual	$\varepsilon(X)$	$-\int_0^\infty S(x)\ln S(x)\,dx$
Rényi	$\varepsilon_\alpha^R(X)$	$\begin{cases} \frac{1}{1-\alpha}\ln\int_0^\infty S^\alpha(x)\,dx, & \alpha\neq 1 \\ \varepsilon(X), & \alpha = 1 \end{cases}$
Tsallis	$\varepsilon_\alpha^T(X)$	$\begin{cases} \frac{1}{1-\alpha}\int_0^\infty (S^\alpha(x)-S(x))\,dx, & \alpha\neq 1 \\ \varepsilon(X), & \alpha = 1 \end{cases}$
Exponential	$\varepsilon_\alpha^E(X)$	$\exp\left(\varepsilon_\alpha^R(X)\right)$
Generalized Rényi	$\varepsilon_{\alpha,\beta}^{GR}(X)$	$\begin{cases} \frac{1}{\beta-\alpha}\ln\frac{\int_0^\infty S^\alpha(x)\,dx}{\int_0^\infty S^\beta(x)\,dx}, & \alpha\neq\beta \\ \frac{-\int_0^\infty S^\beta(x)\ln S(x)\,dx}{\int_0^\infty S^\beta(x)\,dx}, & \alpha=\beta \end{cases}$
Generalized Exponential	$\varepsilon_{\alpha,\beta}^{GE}(X)$	$\exp\left(\varepsilon_{\alpha,\beta}^{GR}(X)\right)$

3 Measures of Mutual Information

Mutual information (MI) was introduced as a similarity measure for multimodal image registration by Collignon and Maes [1,2] and Viola and Wells [3,4]. MI can be defined in a number of ways (see Pluim et al. [5] for a comparison of the various definitions); we focus on two such definitions. The first defines the MI between two random variables X and Y in terms of their marginal (Shannon) entropies and their *joint* entropy $H(X,Y)$:

$$\mathrm{MI}(X,Y) = H(X) + H(Y) - H(X,Y) \quad . \tag{4}$$

The second defines MI in terms of the *conditional* entropy:

$$\mathrm{MI}(X,Y) = H(X) - \mathrm{E}[H(X|Y)] \quad . \tag{5}$$

Note that this notation for conditional entropy departs from other references; here, we consider $H(X|Y)$ to be a random variable that is a function of Y, namely,

$$H(X|Y) = -\int_{-\infty}^{\infty} p(x|Y)\ln p(x|Y)\,dx \quad , \tag{6}$$

and $\mathrm{E}[H(X|Y)]$ to be the expected value of (6).

An analogous quantity to MI that uses CRE is the *cross cumulative residual entropy* (CCRE) of Wang *et al.* [10]:

$$\mathrm{CCRE}(X,Y) = \varepsilon(X) - \mathrm{E}[\varepsilon(X|Y)] \quad . \tag{7}$$

CCRE was also investigated in [21] under the alternate name of *cumulative mutual information* (CMI). The conditional CRE $\varepsilon(X|Y)$ is a random variable that is a function of Y, namely,

$$\varepsilon(X|Y) = -\int_0^\infty S(x|Y) \ln S(x|Y)\, dx \ , \tag{8}$$

where $S(x|Y) = P(|X| > x|Y)$.

One key difference between MI and CCRE is that MI is symmetric whereas CCRE is not (i.e., $\mathrm{MI}(X,Y) = \mathrm{MI}(Y,X)$ but $\mathrm{CCRE}(X,Y) \neq \mathrm{CCRE}(Y,X)$). However, CCRE can be easily symmetrized, yielding the *symmetric* CCRE:

$$\begin{aligned} \mathrm{SCCRE}(X,Y) &= \frac{1}{2}\Big(\mathrm{CCRE}(X,Y) + \mathrm{CCRE}(Y,X)\Big) \\ &= \frac{1}{2}\Big(\varepsilon(X) + \varepsilon(Y) - \mathrm{E}[\varepsilon(X|Y)] - \mathrm{E}[\varepsilon(Y|X)]\Big) \ . \end{aligned} \tag{9}$$

The construction of SCCRE can be used to define a general form of mutual information for use with *any* definition of entropy. If we consider $\mathbf{H}$ to denote a placeholder for any of the definitions of entropy listed in Tables 1–2, we can define a general symmetric version of mutual information by:

$$\mathbf{H}\text{-MI}(X,Y) = \frac{1}{2}\Big(\mathbf{H}(X) + \mathbf{H}(Y) - \mathrm{E}[\mathbf{H}(X|Y)] - \mathrm{E}[\mathbf{H}(Y|X)]\Big) \ . \tag{10}$$

Using this general form (10), it is easily seen that H-MI and ε-MI are the specific forms of MI and SCCRE given in (5) and (9), respectively. Other specific forms of (10) have been explored in the medical image registration literature. H_α^R-MI and H_α^T-MI are similar to Rényi and Tsallis entropy based mutual information measures investigated by Wachowiak *et al.* [8], and ε_α^E-MI and $\varepsilon_{\alpha,\beta}^{GE}$-MI are symmetric versions of the SEE-MI and GSEE-MI measures introduced by Liao and Chung [11].

4 Normalized Measures of Mutual Information

Studholme *et al.* [6] argued that any image similarity measure should be invariant to changes in the overlap region through the course of registration. They showed that traditional MI does in fact vary with changing overlap, and they proposed the *normalized mutual information* (NMI) as an alternative:

$$\mathrm{NMI}(X,Y) := \frac{H(X) + H(Y)}{H(X,Y)} \ . \tag{11}$$

Studholme *et al.* validated the invariance of NMI to changing overlap for some simple examples and illustrated how NMI exhibits better behavior than MI on the rigid registration of MR to CT and MR to PET volumes.

NMI is closely related to Astola's *entropy correlation coefficient* (ECC) [22], which is given by:

$$\mathrm{ECC}(X,Y) := \sqrt{2 - \frac{2H(X,Y)}{H(R) + H(Y)}} \ . \tag{12}$$

Maes *et al.* [1] and Collignon [2] use the square of Astola's ECC for multimodal registration, which is equivalent to the *symmetric uncertainty* (SU) [7,23]:

$$\mathrm{SU}(X,Y) := \mathrm{ECC}(X,Y)^2 = 2 - \frac{2H(X,Y)}{H(R) + H(Y)} \ . \tag{13}$$

In order to generalize these normalized versions of mutual information, we first recognize that $H(X,Y)$ and can be rewritten as:

$$H(X,Y) = \frac{1}{2}\Big(H(X) + H(Y) + \mathrm{E}[H(X|Y)] + \mathrm{E}[H(Y|X)]\Big) \ . \tag{14}$$

Now, general forms of NMI, ECC, and SU can be defined analogously to the general mutual information (10). The resulting forms are shown in Table 3. It is straightforward to show that the use of differential entropy for $\mathbf{H}$ causes H-NMI, H-ECC, and H-SU to reduce to (11), (12), and (13), respectively.

A few specific examples of normalized MI measures that do *not* use differential entropy have been developed for use in medical image registration. Wachowiak *et al.* [8] investigated Rényi and Tsallis entropy based normalizations which are equivalent to H_α^R-NMI and H_α^T-NMI. Cahill *et al.* [12,13] showed that SCCRE exhibits the same overlap sensitivity problem as Studholme established with MI. Furthermore, they defined normalized versions of CRE-based MI called *normalized cross cumulative residual entropy* (NCCRE) and *cumulative residual entropy correlation coefficient* (CRECC), which are equivalent to ε-NMI and ε-ECC, respectively, and they established that NCCRE and CRECC exhibit the same type of improvement over SCCRE that NMI and ECC exhibit over MI.

Table 3. Normalized measures of MI with general definitions of entropy

Similarity Measure	Definition
$\mathbf{H}$-NMI(X,Y)	$\dfrac{2\mathbf{H}(X) + 2\mathbf{H}(Y)}{\mathbf{H}(X) + \mathbf{H}(X) + \mathrm{E}\left[\mathbf{H}(X\|Y)\right] + \mathrm{E}\left[\mathbf{H}(Y\|X)\right]}$
$\mathbf{H}$-ECC(X,Y)	$\sqrt{1 - \dfrac{\mathrm{E}\left[\mathbf{H}(X\|Y)\right] + \mathrm{E}\left[\mathbf{H}(Y\|X)\right]}{\mathbf{H}(X) + \mathbf{H}(Y)}}$
$\mathbf{H}$-SU(X,Y)	$1 - \dfrac{\mathrm{E}\left[\mathbf{H}(X\|Y)\right] + \mathrm{E}\left[\mathbf{H}(Y\|X)\right]}{\mathbf{H}(X) + \mathbf{H}(Y)}$

5 Multimodal Rigid Registration Experiment

In order to illustrate the behavior of the various 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.*[24].

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 $0.65 \times 0.65 \times 4.0$ mm^3, the MR images have approximate resolution $1.25 \times 1.25 \times 4.0$ mm^3, and the PET images has resolution $2.59 \times 2.59 \times 8.0$ mm^3. For ease of computation, we resampled each image to $3.0 \times 3.0 \times 3.0$ mm^3 isotropic resolution.

Examples of some of the RIRE images are shown in Figures 1 and 2. Figure 1 shows axial views of the CT, MR-T1, and PET images from patient 5. Figure 2 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. Initial estimates of the solution were selected by translating the images to match their computed centroids. To constrain the parameters, we employed bounds of $\pm\pi/6$ radians on each Euler angle, and $\pm 1/5$ of the width of the corresponding reference image dimension on each translation component. (We visually verified for each case that after the prealignment step is performed, the true rigid transformation parameters fall within these bounds.) For each dissimilarity measure, we carried out a bound-constrained optimization

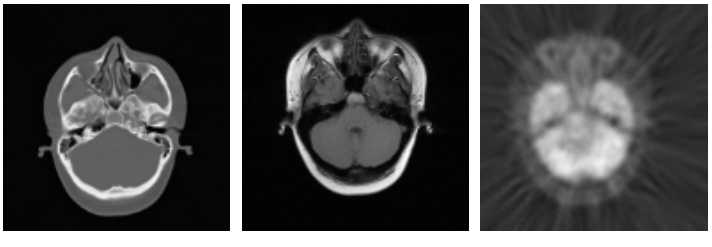
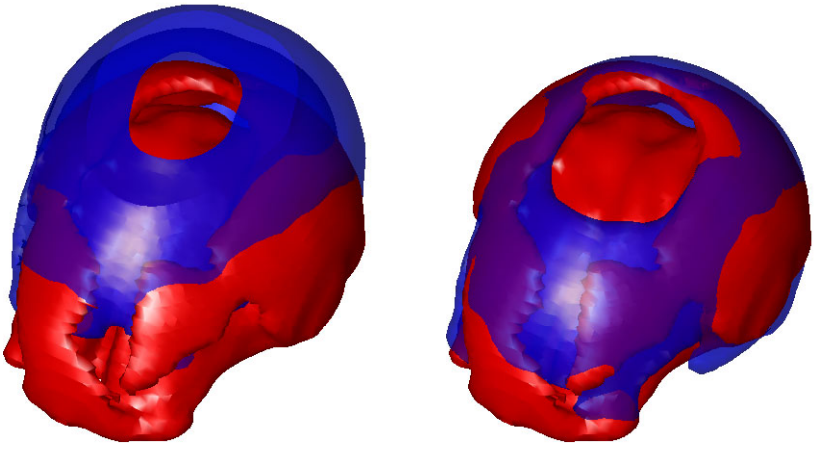


Fig. 1. Axial slices of CT, MR (T1-weighted) and PET images from the patient 5 dataset



(a) Before registration (b) After registration

Fig. 2. Isosurfaces of CT (blue) and MR-T1 (red) images of patient 4, before and after rigid registration

procedure using the active set algorithm employed by the `fmincon` function of MATLAB's Optimization Toolbox. All gradient vectors and Hessian matrices were estimated numerically via finite differences. The optimization was terminated when the maximum change in magnitude in any parameter was less than 10^{-4} or after 500 iterations, whichever occured first.

All probability densities (and joint densities) were estimated via histograms (and joint histograms) that were constructed with 32 (or 32×32) equally spaced bins. Linear (or bilinear) interpolation was used to accumulate partial weights in neighboring bins.

Registration is performed using the general versions of MI and ECC based on the original, Renyi, and Tsallis versions of differential entropy and CRE defined in Tables 1–2. Values of α are drawn from the set $\{0.5, 1, 2\}$.

5.1 Results

The performance of similarity measures on the various registration tasks is measured via Target Registration Error (TRE). For the RIRE project [24], 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 4 reports the mean, median, and standard deviation of the TRE values measured across the VOI's in every patient for various similarity measures.

We employed hypothesis testing to gauge the statistical significance of these results. Given similarity measures **H**-MI and **H**-ECC, the null hypothesis states that the TRE's from **H**-MI and **H**-ECC arise from the same distribution. The alternative hypothesis is that the TRE's from **H**-MI are "worse" than the TRE's

Table 4. Statistics of TRE (in mm) for CT/MR and PET/MR registration, aggregated across all patients

Measure	CT/MR Registration			PET/MR Registration		
	Mean TRE	Median TRE	Std. Dev. TRE	Mean TRE	Median TRE	Std. Dev. TRE
H-MI	2.0	1.8	1.0	6.6	3.1	8.3
H-ECC	2.3	2.0	1.3	3.1	2.8	1.6
$H^R_{1/2}$-MI	2.1	1.8	1.3	3.3	3.2	1.3
$H^R_{1/2}$-ECC	2.2	1.8	1.1	2.7	2.9	0.9
H^R_2-MI	24.0	27.5	14.9	6.4	3.6	7.6
H^R_2-ECC	13.6	2.9	15.0	4.0	2.8	4.7
$H^T_{1/2}$-MI	1.8	1.6	0.9	2.7	2.7	0.9
$H^T_{1/2}$-ECC	1.9	1.6	1.0	2.6	2.5	1.0
H^T_2-MI	5.4	3.4	6.0	6.2	2.9	8.1
H^T_2-ECC	6.2	4.6	6.2	4.4	2.6	5.9
ε-MI	3.2	3.0	1.7	3.6	3.4	2.2
ε-ECC	3.4	3.0	1.9	3.6	3.0	2.5
$\varepsilon^R_{1/2}$-MI	4.2	3.7	3.1	4.0	3.7	2.5
$\varepsilon^R_{1/2}$-ECC	4.4	4.4	2.1	3.8	3.3	2.3
ε^R_2-MI	13.2	11.3	14.9	5.2	3.3	5.0
ε^R_2-ECC	7.9	7.3	4.5	5.5	3.4	5.3
$\varepsilon^T_{1/2}$-MI	5.9	4.4	4.3	4.1	4.0	1.7
$\varepsilon^T_{1/2}$-ECC	3.1	3.0	1.8	3.2	3.3	1.3
ε^T_2-MI	14.6	14.4	10.2	14.3	11.6	13.9
ε^T_2-ECC	3.6	3.2	2.4	2.5	2.6	0.9

from **H**-ECC, 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:

- CT/MR Registration: H^R_2, ε^R_2, $\varepsilon^T_{1/2}$, ε^T_2
- PET/MR Registration: H^R_2, $\varepsilon^T_{1/2}$, ε^T_2

When the same analysis was done with the roles of **H**-MI and **H**-ECC interchanged, there were no versions of **H** for which **H**-MI exhibited a statistically significant improvement over **H**-ECC.

This analysis indicates that in both the CT/MR and PET/MR registration cases, statistically significant improvements can be made by choosing the ECC versions of MI over MI itself when H^R_2, $\varepsilon^T_{1/2}$, or ε^T_2 are selected for entropy. When ε^R_2 is selected, a statistically significant improvement is made in CT/MR registration when ECC is used over MI.

6 Conclusion

In this article, we showed how to construct normalized versions of MI using a variety of definitions of entropy, including Rényi, Tsallis, exponential, generalized exponential, and cumulative residual entropy. The normalized similarity measures are analogous to NMI, ECC, and SU, which have previously been established to outperform MI in a variety of situations. To test the proposed normalized similarity measures, we used publicly available multimodal brain imaging data [24] that allowed us to perform CT/MR and PET/MR rigid registration and compare the results to known ground truth transformations. Results indicate statistically significant improvements in target registration error for a variety of the proposed similarity measures.

References

1. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality Image Registration by Maximization of Mutual Information. IEEE Trans. Med. Imag. 16(2), 187–198 (1997)
2. Collignon, A.: Multi-modality Medical Image Registration by Maximization of Mutual Information. Ph.D. Thesis, Catholic University of Leuven, Leuven, Belgium (1998)
3. Viola, P.A.: Alignment by Maximization of Mutual Information. Ph.D. Thesis, Artificial Intelligence Laboratory, Massachusetts Institute of Technology (1995)
4. Viola, P.A., Wells III, W.M.: Alignment by Maximization of Mutual Information. Int. J. Comp. Vis. 24(2), 137–154 (1997)
5. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: Mutual-Information-Based Registration of Medical Images: A Survey. IEEE Trans. Med. Imag. 22(8), 986–1004 (2003)
6. Studholme, C., Hill, D.L.G., Hawkes, D.J.: An Overlap Invariant Entropy Measure of 3D Medical Image Alignment. Patt. Rec. 32, 71–86 (1999)
7. Melbourne, A., Hawkes, D., Atkinson, D.: Image Registration Using Uncertainty Coefficients. In: Proc. ISBI 2009, pp. 951–954 (2009)
8. Wachowiak, M.P., Smoliková, R., Tourassi, G.D., Elmaghraby, A.S.: Similarity Metrics Based on Nonadditive Entropies for 2D–3D Multimodal Biomedical Image Registration. In: Sonka, M., Fitzpatrick, J.M. (eds.) Proc. SPIE of Medical Imaging 2003: Image Processing, vol. 5032, pp. 1090–1100 (2003)
9. Wang, F., Vemuri, B.C., Rao, M., Chen, Y.: A New & Robust Information Theoretic Measure and Its Application to Image Alignment. In: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 388–400. Springer, Heidelberg (2003)
10. Wang, F., Vemuri, B.C.: Non-rigid Multi-modal Image Registration Using Cross-cumulative Residual Entropy. Int. J. Comp. Vis. 74(2), 201–205 (2007)
11. Liao, S., Chung, A.C.S.: Multi-modal Image Registration Using the Generalized Survival Exponential Entropy. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 964–971. Springer, Heidelberg (2006)
12. Cahill, N.D.: Constructing and Solving Variational Image Registration Problems. D.Phil. Thesis, Department of Engineering Science, University of Oxford (2009)
13. Cahill, N.D., Schnabel, J.A., Noble, J.A., Hawkes, D.J.: Overlap Invariance of Cumulative Residual Entropy Measures for Multimodal Image Alignment. In: Pluim, J.P.W., Dawant, B.M. (eds.) Proc. SPIE of Medical Imaging 2009: Image Processing, vol. 7259 (2009)

14. Shannon, C.E.: A Mathematical Theory of Communication. Bell Sys. Tech. J. 27, 379–423, 623–656 (1948)
15. Rényi, A.: On measures of entropy and information. In: Proc. 4th Berkeley Symp. Mathematical Statistics and Probability, Berkeley, CA, vol. I, pp. 547–561 (1961)
16. Tsallis, C.: Possible Generalizations of Boltzmann-Gibbs Statistics. J. Stat. Phys. 52, 479–487 (1988)
17. Campbell, L.L.: Exponential Entropy as a Measure of Extent of a Distribution. Zeitschr. für Wahrsch. und verw. Geb. 5, 217–255 (1966)
18. Koski, T., Persson, L.-E.: Some Properties of Generalized Exponential Entropies with Applications to Data Compression. Inf. Sci. 62(1-2), 103–132 (1992)
19. Rao, M., Chen, Y., Vemuri, B.C., Wang, F.: Cumulative Residual Entropy: A New Measure of Information. IEEE Trans. Inf. Th. 50(6), 1220–1228 (2004)
20. Zografos, K., Nadarajah, S.: Survival Exponential Entropies. IEEE Trans. Inf. Th. 51(3), 1239–1246 (2005)
21. Drissi, N., Chonavel, T., Boucher, J.M.: Generalized Cumulative Residual Entropy for Distributions with Unrestricted Supports. Res. Let. Sig. Proc. Article ID 790607, 2008 (2008)
22. Astola, J., Virtanen, I.: A Measure of Overall Statistical Dependence Based on the Entropy Concept. In: Proc. Univ. Vaasa. No. 91 (1983)
23. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, Amsterdam (2005)
24. West, J., et al.: Comparison and Evaluation of Retrospective Intermodality Brain Image Registration Techniques. J. Comp. Assist. Tom. 21(4), 554–566 (1997)