Image Similarity Metrics in Image Registration

A. Melbourne, G. Ridgway, D. J. Hawkes

Centre for Medical Image Computing, University College London, UK.

ABSTRACT

Measures of image similarity that inspect the intensity probability distribution of the images have proved extremely popular in image registration applications. The joint entropy of the intensity distributions and the marginal entropies of the individual images are combined to produce properties such as resistance to loss of information in one image and invariance to changes in image overlap during registration. However information theoretic cost functions are largely used empirically. This work attempts to describe image similarity measures within a formal mathematical metric framework. Redefining mutual information as a metric is shown to lead naturally to the standardised variant, normalised mutual information.

Description of Purpose

The use of image similarity measures that inspect the intensity distribution of each of the images has proved successful for many image alignment procedures. The combination of the Shannon entropies of the individual and joint image histograms into mutual information was discussed independently by Viola and Wells[1] and Maes[2] . A popular normalised variant was also proposed that the authors claim is more robust to changes in image overlap, although a full theoretical justification is not given[3] . However, neither of these measures is strictly a metric. A later survey of the growth of image registration procedures using mutual information or its variants was produced by Pluim *et al*[4] and later extended[5] . It is normally assumed that pixel intensity values in an image are the result of a stochastic spatially independent process; as a result, mutual information is invariant to an arbitrary re-mapping of pixel intensity values (as demonstrated in Figure 1). This property is often used to allow registration of images produced using different acquisition techniques, assuming that the same structures are visible but at different intensities. Also shown in Figure 2 is an 'image' generated from the the same stochastic spatially independent process as in Figure 1 (this is equivalent to rearranging all pixel positions in Figure 1a); in the terminology to be discussed shortly, these images have the same information content as their counterparts in Figure 1. Separate developments in information theory discuss the notion of distance between two distributions and it is this work that we attempt to combine here in order to formally develop image similarity[6,7] .

Method

Metrics

The fundamental requirements for a measure of image similarity to be considered a metric are listed below for a distance $d(\mathbf{A}, \mathbf{B})$ between two objects (for our purpose these are images) $\mathbf{A}$ and $\mathbf{B}$[6] . These encapsulate a measure of distance between two images. R1 and R2 ensure that the distance between two images $\mathbf{A}$ and $\mathbf{B}$ is always positive and only zero when the two images are identical. R3 ensures that the distance between images $\mathbf{A}$ and $\mathbf{B}$ is always symmetric and therefore the same in both directions. Meeting the triangle inequality in requirement R4 ensures that the distance between the two images $\mathbf{A}$ and $\mathbf{B}$ is always less than or identical to the distance between $\mathbf{A}$ and a third image $\mathbf{C}$ followed by the distance from $\mathbf{C}$ to $\mathbf{B}$.

$$R1) \quad d(\mathbf{A}, \mathbf{B}) > 0 \tag{1}$$

$$R2) \quad d(\mathbf{A}, \mathbf{B}) = 0 \; \textit{iff} \; \mathbf{A} = \mathbf{B} \tag{2}$$

$$R3) \quad d(\mathbf{A}, \mathbf{B}) = d(\mathbf{B}, \mathbf{A}) \tag{3}$$

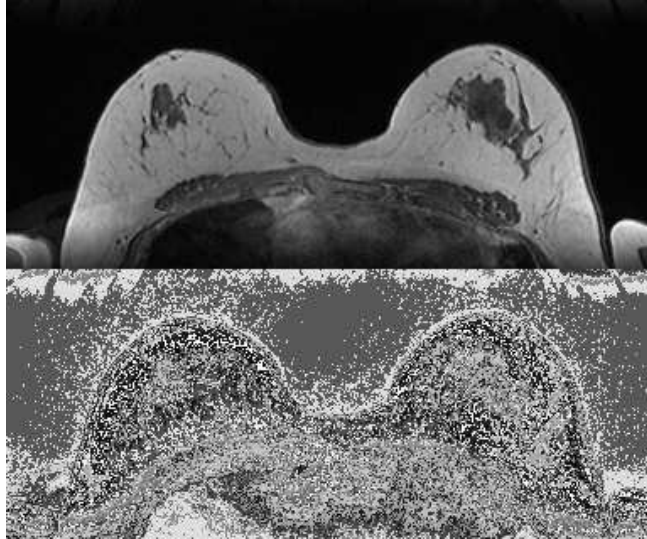$$R4) \quad d(\mathbf{A}, \mathbf{B}) \leq d(\mathbf{A}, \mathbf{C}) + d(\mathbf{C}, \mathbf{B}) \tag{4}$$

Figure 1. Example of image intensity swapping for an axial slice though an MRI breast volume. Pixel intensity values in the original image **a** are binned and arbitrarily rearranged to give image **b**. The mutual information (MI) between these two images is maximal (here, for 64 bins, MI=6) and the images may be regarded as being identical.
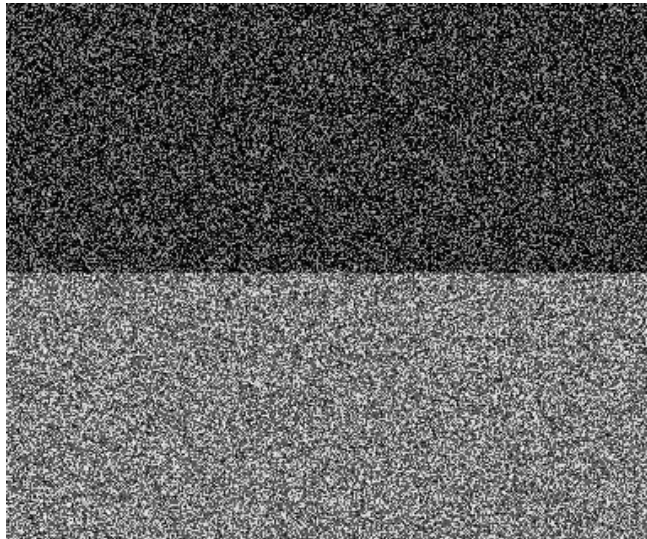

Figure 2. Example of intensity position swapping for an axial slice though an MRI breast volume (1**a**). The pixel positions are arbitrarily rearranged in space. These images have the same marginal entropy as Figure 1**a**. Figure 2**b** has had the intensity values swapped as in Figure 1**b**. Again, the mutual information (MI) between these two images is maximal (here, for 64 bins, MI=6) and the 'images' may be regarded as being identical

Jensen's Inequality

A function $f(x)$ is convex (concave) if its chords lie above (below) the function; Jensen's inequality states that the expected value of the function is above (below) the function of the expected value of x^8 . Stating the concave case more precisely:

$$f(\lambda\,x_1 + (1 - \lambda)x_2) \geq \lambda f(x_1) + (1 - \lambda)f(x_2) \Rightarrow \mathbf{E}f(x) \leq f(\mathbf{E}x). \tag{5}$$

Concepts from information theory

The information theoretic similarity measures discussed in this work make use of the Shannon entropy of a probability distribution (Equation 6) found from the normalised image histogram for image **A**. Defined in this

way the entropy H_A is continuous, symmetric (order independent), additive and greater than or equal to zero.[*]

$$H_A = \mathbf{E}_A \log(1/p_A) = -\mathbf{E}_A \log(p_A) = -\sum_{i=1}^{\text{bins}} p_A(i) \log(p_A(i)) \tag{6}$$

We can find the maximum value of the entropy by noting that log is concave and using Jensen's inequality (5)

$$\mathbf{E}_A \log(1/p_A) \leq \log\left(\mathbf{E}_A \frac{1}{p_A}\right) = \log\left(\sum_{i=1}^{\text{bins}} \frac{p_A(i)}{p_A(i)}\right) = \log(\text{bins}). \tag{7}$$

It is simple to verify that this maximum occurs for uniform distributions, where $p_A(i) = 1/\text{bins}$ for all i.

Since the entropy is defined purely in terms of a probability distribution, it can be trivially extended to joint distributions to express the information present in two images. The joint probability distribution p_{AB} is the two-dimensional distribution obtained when we consider the co-occurences of image intensities between images **A** and **B**. The joint entropy, given in equation 8, can be decomposed into different components by factoring the joint distribution using conditional distributions, where $p_{AB} = p_{A|B}p_B = p_{B|A}p_A$,

$$
\begin{aligned}
H_{AB} &= -\sum_{i=1}^{\text{bins}_A} \sum_{j=1}^{\text{bins}_B} p_{AB}(i,j) \log(p_{AB}(i,j)) \tag{8} \\
&= -\mathbf{E}_{AB} \log p_{AB} = -\mathbf{E}_{AB} \log \frac{p_{AB}^2}{p_{AB}} = -\mathbf{E}_{AB} \log \frac{p_{A|B} p_B \, p_{B|A} p_A}{p_{AB}} \\
&= -\mathbf{E}_{AB} \log p_{A|B} - \mathbf{E}_{AB} \log \frac{p_A p_B}{p_{AB}} - \mathbf{E}_{AB} \log p_{B|A}. \tag{9}
\end{aligned}
$$

The first (or similar third) term can be expanded like:

$$-\mathbf{E}_{AB} \log p_{A|B} = -\sum_j p_B \sum_i p_{A|B}(i,j) \log p_{A|B}(i,j) \tag{10}$$

which is recognised as the expected value (over B) of the entropy of a conditional probability distribution $p_{A|B}$, and is therefore known as the conditional entropy $H_{A|B}$.

The second term of (9), which we denote I_{AB}, is of the general form:

$$-\mathbf{E}_X \log \frac{p_Y}{p_X} = \mathbf{E}_X \log \frac{p_X}{p_Y}, \tag{11}$$

with X denoting the joint distribution, and Y representing a factored approximation to it using the marginal distributions. We can now use Jensen's inequality (5) to prove Gibbs' inequality[8]

$$
\begin{aligned}
\mathbf{E}_X \log \frac{p_Y}{p_X} &\leq \log \mathbf{E}_X \frac{p_Y}{p_X} \\
&= \log \sum p_X \frac{p_Y}{p_X} = \log \sum p_Y = \log(1) = 0 \\
-\mathbf{E}_X \log \frac{p_Y}{p_X} &\geq 0. \tag{12}
\end{aligned}
$$

Hence this is another non-negative entropy-like quantity, sometimes known as the relative entropy. Equality occurs when $p_Y = p_X$, meaning this term can be intuitively considered a distance of p_Y from p_X. It is not a true distance metric, as it is not symmetric (and also does not satisfy the triangle inequality), so it is instead referred to as the information divergence or Kullback-Leibler divergence (KLD).

---

[*]Although probabilities of zero might appear to cause problems with the logarithm, one can use l'Hôpital's rule with $x = 1/p$ to show $\lim_{x \to \infty} \frac{\log x}{x} = \frac{d \log(x)/dx}{dx/dx} = 1/x \to 0$. Hence, certainties of either extreme ($p = 0$ or $p = 1$) contribute nothing to the information, as one might expect.

Returning to the joint entropy decomposition (9), the sum of either the first or last pair of terms gives the marginal entropy:

$$H_{A|B} + I_{AB} = -\mathbf{E}_{AB} \log p_{A|B} - \mathbf{E}_{AB} \log \frac{p_A p_B}{p_{AB}} = -\mathbf{E}_{AB} \log \frac{p_{A|B} p_B \, p_A}{p_{AB}} \tag{13}$$

$$= -\mathbf{E}_{AB} \log \frac{p_{AB} \, p_A}{p_{AB}} = -\mathbf{E}_{AB} \log p_A = -\mathbf{E}_A \log p_A = H_A. \tag{14}$$

The consequent results

$$H_{AB} = H_{A|B} + H_B = H_{B|A} + H_A \tag{15}$$

$$= H_{A|B} + I_{AB} + H_{B|A} \tag{16}$$

$$= (H_{AB} - H_B) + I_{AB} + (H_{AB} - H_A) \tag{17}$$

$$= H_A + H_B - I_{AB}, \tag{18}$$

mean that I_{AB}, which is the KLD from p_{AB} to $p_A p_B$, is also known as the mutual information between A and B, since it is the difference between the information in A and B together and the sum of their separate information contributions (Equation 19).

$$MI(\mathbf{A}, \mathbf{B}) = \sum_{i}^{bins_A} \sum_{j}^{bins_B} p_{AB}(i,j) \log \left(\frac{p_{AB}(i,j)}{p_A(i) p_B(j)} \right) \tag{19}$$

Under independence ($p_{AB} = p_A p_B$) the conditional entropies become the marginal entropies, the mutual information becomes zero, and the joint entropy becomes the sum of the marginal entropies. Perfect dependence, as in $p_A = p_B$, results in the conditional entropies becoming zero (see footnote [*]), and $H_{AB} = H_A = H_B = I_{AB}$. Taken together, all these results motivate a Venn diagram representation of information, illustrated in Figure 3.

Image Similarity Measures

The most popular image (dis)similarity measures are now outlined below; none of these may be considered a metric although all meet the requirement in R3. The joint entropy (Equation 20) is minimised with increasing similarity but identical non-empty images have a non-zero joint entropy. Mutual information (MI) (Equation 21) and the empirical normalised mutual information (NMI) (Equation 22) incorporate the Shannon entropies of the individual normalised image intensity histograms and are maximised with increasing similarity and hence do not obey requirements R2 and R4. Also shown is the symmetric uncertainty coefficient (SUC), an alternative normalisation of mutual information[9] .

$$JE(\mathbf{A}, \mathbf{B}) = H_{AB} \tag{20}$$

$$MI(\mathbf{A}, \mathbf{B}) = H_A + H_B - H_{AB} = I_{AB} \tag{21}$$

$$NMI(\mathbf{A}, \mathbf{B}) = \frac{H_A + H_B}{H_{AB}} \tag{22}$$

$$SUC(\mathbf{A}, \mathbf{B}) = 2 \left(1 - \frac{H_{AB}}{H_A + H_B} \right) \tag{23}$$

Reformulation of Similarity Measures as Metrics

A metric on the shared information content of images $\mathbf{A}$ and $\mathbf{B}$ can be defined as in Equation 24 the *shared information distance*.[6] This distance obeys the definition of a metric in R1-R4 (where $H_{A|B}$ is the conditional entropy of $\mathbf{A}$ given $\mathbf{B}$).

$$d(\mathbf{A}, \mathbf{B}) = H_{A|B} + H_{B|A} \tag{24}$$

$$= H_{AB} - MI \tag{25}$$

$$= 2H_{AB} - H_A - H_B \tag{26}$$

The symmetry requirement (R3) is clear from all the above definitions. Requirements R1 to R3 are met by the definitions of mutual information and d($\mathbf{A}$,$\mathbf{B}$) in Equations 21 and 26 and from the arguments supporting Figure 3, where we see that the distance is the sum of two areas which shrink to zero if the distributions are equal. Using this illustration and the definition $H_{AB} = H_A \cup H_B$: H_{AB} can never have more information than the combined total of H_A and H_B, whilst it must necessarily have at least as many members as its largest component (for instance if H_A is a subset of H_B).

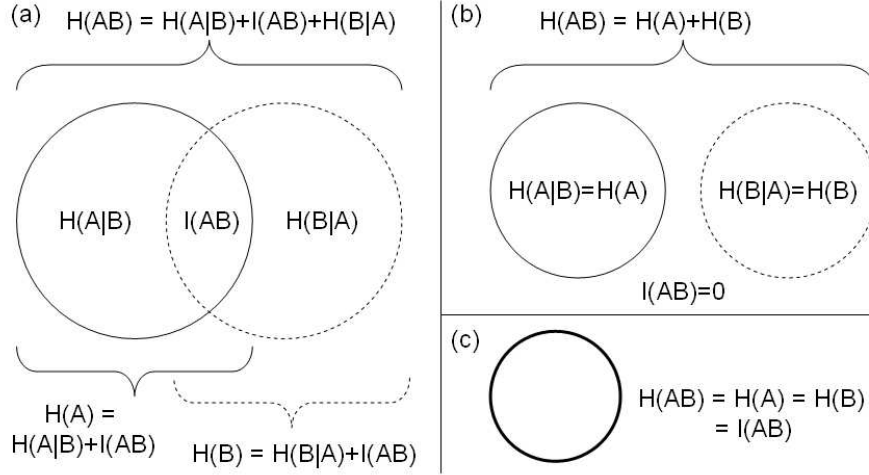$$max(H_A, H_B) \leq H_{AB} \leq (H_A + H_B) \tag{27}$$

Figure 3. Venn diagram representations of the decompositions of joint entropy into constituent parts: (a) the general case; (b) the independent case; and (c) the fully dependent case.

Meeting the triangle inequality is easily shown from Equation 24 and a natural three-set extension of Figure 3, as illustrated in Figure 4. An alternative method is to note that $H_C \cup H_A$ and $H_C \cup H_B$ contain all the members of H_C twice plus a remainder consisting of $H_A \cup H_B$ minus the intersection of the two copies of H_C with $H_A \cup H_B$ (Equation 32). The second part of this remainder necessarily cannot have more members than H_C.

$$
\begin{align}
d(\mathbf{A}, \mathbf{B}) &\leq d(\mathbf{A}, \mathbf{C}) + d(\mathbf{C}, \mathbf{B}) \tag{28}\\
2H_{AB} - H_A - H_B &\leq 2H_{AC} - H_A - H_C + 2H_{CB} - H_C - H_B \tag{29}\\
H_{AB} &\leq H_{AC} + H_{BC} - H_C \tag{30}\\
H_A \cup H_B &\leq H_A \cup H_C + H_B \cup H_C - H_C \tag{31}\\
H_A \cup H_B &\leq 2H_C + H_A \cup H_B - H_C \cap (H_A \cup H_B) - H_C \tag{32}\\
0 &\leq H_C - H_C \cap (H_A \cup H_B) \tag{33}
\end{align}
$$

The shared information distance may also be represented in the form of Equation 19 to produce Equation 34.

$$d(\mathbf{A}, \mathbf{B}) = \sum_i^{bins} \sum_j^{bins} p_{AB}(i,j) \log \left(\frac{p_A(i) p_B(j)}{p_{AB}^2(i,j)} \right) \tag{34}$$

Metric Normalisation

The metric defined in Equation 24 may be normalised. Normalisation by H_{AB} produces a metric formulated in terms of NMI and bounded between 0 and 1. We give this quantity the capitalisation $D(\mathbf{A}, \mathbf{B})$: a value of 0
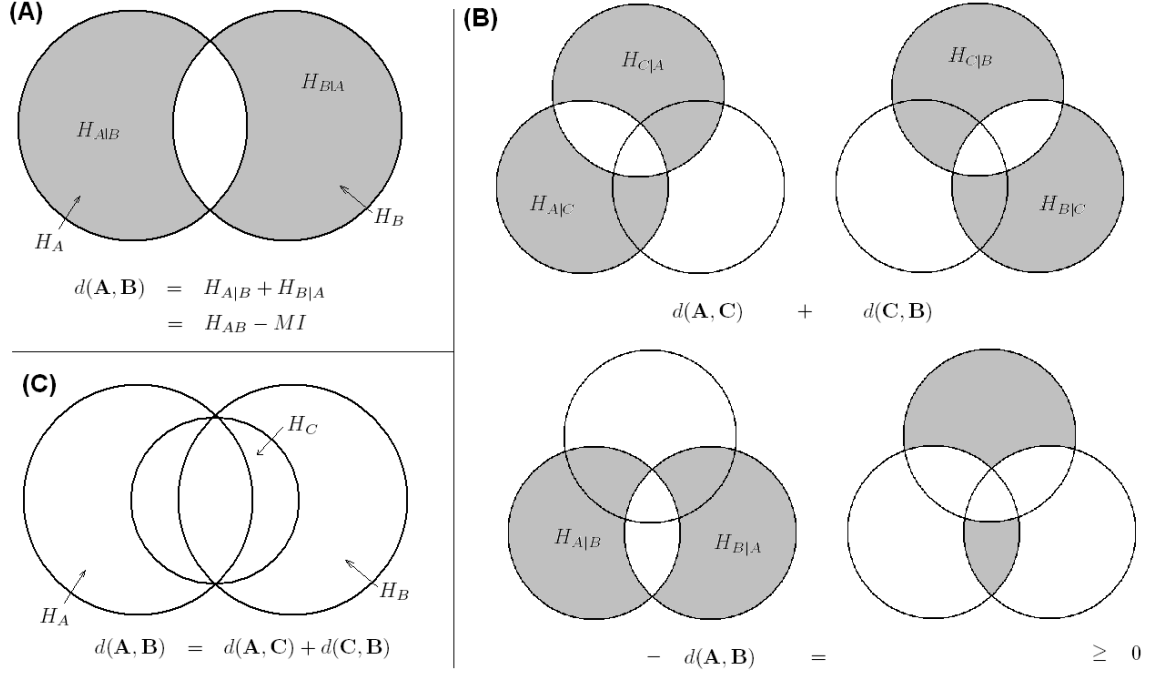
Figure 4. Visualisation of (a) the shared information distance in terms of set areas; (b) a graphical proof of the triangle inequality; and (c) the special case of equality where C is intuitively contained on the informational 'path' from A to B.

corresponding to identical images and a value of 1 corresponding to maximum possible dissimilarity (Equation 36). A key result is that the empirical NMI used in image registration is seen to arise from a natural normalisation of an underlying distance metric; it can be shown to obey the density requirement in Inequality 37, demonstrating that there are finite values of $D(\mathbf{A}, \mathbf{B})$ for an image $\mathbf{B}$ at a given distance from image $\mathbf{A}$. Under certain conditions Equation 36 may also be considered a *universal metric*[6]

$$D_1(\mathbf{A}, \mathbf{B}) = \frac{d(\mathbf{A}, \mathbf{B})}{H_{AB}} \tag{35}$$

$$= 2 - \mathrm{NMI}(\mathbf{A}, \mathbf{B}) \tag{36}$$

$$\sum_{B:B \neq A} 2^{-D(A,B)} \leq 1 \tag{37}$$

The shared information distance metric in Equation 24 may also be normalised by $H_A + H_B$. In this instance a modification of the SUC is seen to arise. We persist with the analysis for the SUC to demonstrate a result in the calculation of the gradients of the image similarity measures.

$$D_2(\mathbf{A}, \mathbf{B}) = \frac{d(\mathbf{A}, \mathbf{B})}{H_A + H_B} \tag{38}$$

$$= 1 - \mathrm{SUC}(\mathbf{A}, \mathbf{B}) \tag{39}$$

Meeting the triangle inequality is shown by the condition in Inequality 40 for $D_1(\mathbf{A}, \mathbf{B})$ and typographical inversion for $D_2(\mathbf{A}, \mathbf{B})$ (Inequality 41), which, in the case $H_A = H_B \neq H_C$ can be seen by inspection to hold. A proof of the triangle inequality for D_1 may be found in Kraskov et al[10] . Proving the inequality for D_2 remains as further work.

$$\frac{H_A + H_C}{H_{AC}} + \frac{H_C + H_B}{H_{CB}} - \frac{H_A + H_B}{H_{AB}} \leq 2 \tag{40}$$

$$\frac{H_{AC}}{H_A + H_C} + \frac{H_{CB}}{H_C + H_B} - \frac{H_{AB}}{H_A + H_B} \leq \frac{1}{2} \tag{41}$$

Image Registration Gradients

Image registration driving forces can be calculated by finding the local change to the image similarity measures discussed above over some spatial parameter, x. Mutual information may be maximised by considering its constituent parts; we aim to minimise the joint entropy, whilst maximising (or more likely, maintaining) the marginal entropies (Equation 42). Hence we have a correcting cost-function that prevents minimisation of the joint entropy by the simplification of the intensity distributions of the registered images.

$$\frac{d\mathrm{MI}(\mathbf{A},\mathbf{B})}{dx} \quad = \quad \frac{dH_A}{dx} + \frac{dH_B}{dx} - \frac{dH_{AB}}{dx} \tag{42}$$

The gradient of the NMI (or the negative of the D_1 metric) can be seen to be a weighted variant of Equation 42. In this case the gradient contributions of the marginal entropies and the joint entropy are balanced by the complementary term (Equation 44). In the case of high joint entropy we give a higher weighting to maintaining (or increasing) the marginal entropies. In the case of high marginal entropies we allow more weight to be given to the minimisation of the joint entropy.

$$\frac{d\mathrm{NMI}(\mathbf{A},\mathbf{B})}{dx} \quad = \quad \frac{d}{dx}\left(\frac{H_A + H_B}{H_{AB}}\right) \tag{43}$$

$$= \quad \frac{1}{H_{AB}^2}\left[H_{AB}\left(\frac{dH_A}{dx} + \frac{dH_B}{dx}\right) - (H_A + H_B)\frac{dH_{AB}}{dx}\right] \tag{44}$$

The gradient of the SUC (or the negative of the D_2 metric) is identical to that for NMI, with the exception of the scaling denominator. An interesting property of this denominator in the case of maximising NMI is that the denominator H_{AB}^2 gets smaller as we approach the solution (although it will not vanish for non-empty images, Inequality 27). In the case of minimisation of the SUC, the denominator $(H_A + H_B)^2$ will maximise (or at least stay constant) so that the gradient will become smaller as we approach the solution.

$$\frac{d\mathrm{SUC}(\mathbf{A},\mathbf{B})}{dx} \quad = \quad -\frac{d}{dx}\left(\frac{2H_{AB}}{H_A + H_B}\right) \tag{45}$$

$$= \quad \frac{2}{(H_A + H_B)^2}\left[H_{AB}\left(\frac{dH_A}{dx} + \frac{dH_B}{dx}\right) - (H_A + H_B)\frac{dH_{AB}}{dx}\right] \tag{46}$$

Data

The metrics defined above are used to monitor changes in image similarity when a known deformation is applied to one of the images. The metric is calculated under a range of simulated deformations generated by the application of a finite element model of breast deformation to MRI data using the method and four breast volumes from Guy's Hospital, UK found in Tanner *et al*[11] .

Deformations are modelled over a range of force types and magnitudes to simulate differences due to a range of common breast imaging deformations. Six force types are modelled[11] , a push over a local area of one breast; simulations of both one and two plate-like compressions; a point puncture simulation (representing deformation due to biopsy) and simulation of pectoral muscle movement from tension to relaxation and vice versa. Biomechanical deformations were applied at two different displacement magnitudes and here we linearly interpolate between these to generate a range of ten displacements for each force type so that it is possible to inspect variation in the distance metric with deformation extent.

Results

Image Similarity

Similarity between different groups of images can be compared using the notion of the un-normalised shared information distance in Equation 26. For the range of displacements and force types discussed in the methods, we find the distance between the original and deformed volume. The distance as measured by Equation 26 shows good correlation with the extent of the applied deformation. It is also clear that there is considerable

overlap between the type of deformation and the distance between the final volumes; it is not necessarily possible to distinguish the type of deformation from the calculated distance. The only exception is the result for the simulated point puncture where, since the deformation is over only a small volume, the distance between the volumes before and after the point puncture is also small. One possible method of demarcation may be to monitor the rate of change of the similarity distance with the applied physical deformation.
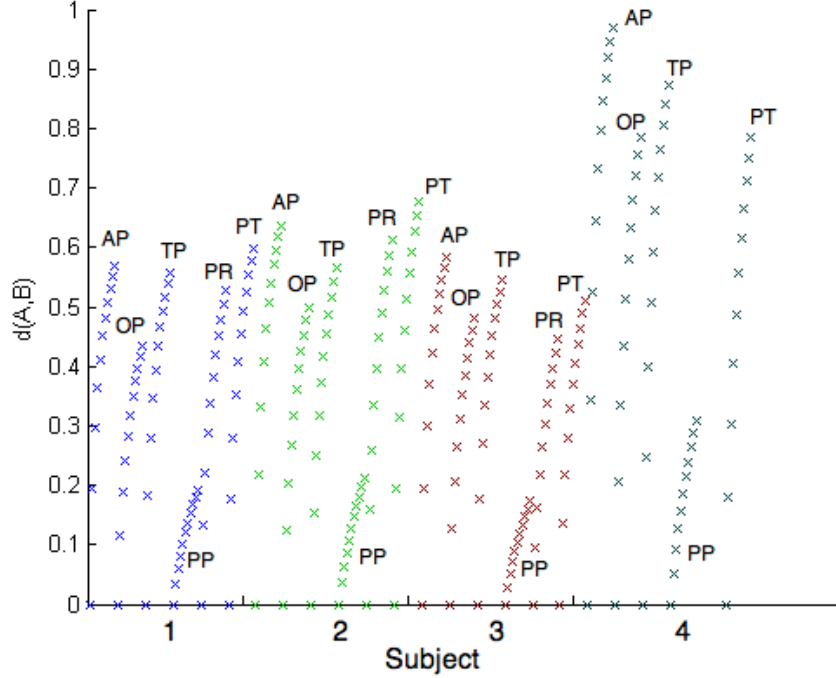

Figure 5. Variation of shared information distance, d($\mathbf{A}$,$\mathbf{B}$), between volumes $\mathbf{A}$ and $\mathbf{B}$ for four subjects under a varying biomechanical test deformation on $\mathbf{B}$ compared to undeformed volume $\mathbf{A}$ *see text*. AP - push over a local area; OP, TP - one and two plate-like compressions; PP - a point puncture simulation and PR, PT - simulation of pectoral muscle movement from tension to relaxation and vice versa.

Image Registration

Image registration can be carried out by the minimisation of the D_1 or D_2 metrics. Here, registrations use a 2D fluid registration incorporating a strategy to minimise information theoretic cost-functions[12,13] . Non-rigid image registration gradients use the expressions in Equation 44 and 46. A central slice from each of the volumes in Figure 5 is registered using both the D_1 and D_2 metrics to determine any differences in registration performance (a total of 528 registrations).

The number of registration iterations required for registration using either the D_1 and D_2 metric is shown in Figure 6 up to a maximum of 150 iterations. The median number of iterations reached before registration termination is 39 for D_1 and 44 for D_2: this may be some indication of the effect of the smaller gradient step size of Equation 46, implying that the registration takes slightly longer to converge.

Registration of the deformed volume back to the original undeformed state can also be investigated. Figure 7a assesses the final values of the Shared Information Distance metric after registration using either D_1 or D_2. The results show no difference between metrics, although a substantial increase in image similarity is seen using both methods. Results having a value of $d(\mathbf{A}, \mathbf{B}) = 0$ arise from images for which no deformation has been applied. Figure 7b assesses the distribution of the change in image-similarity from the unregistered value of d($\mathbf{A}$,$\mathbf{B}$) by each method. Again, there is no obvious difference between registration by either D_1 or D_2. Decreases in image similarity ($d(\mathbf{A}, \mathbf{B}) > 0$) are likely to be the result of the image interpolation method used.
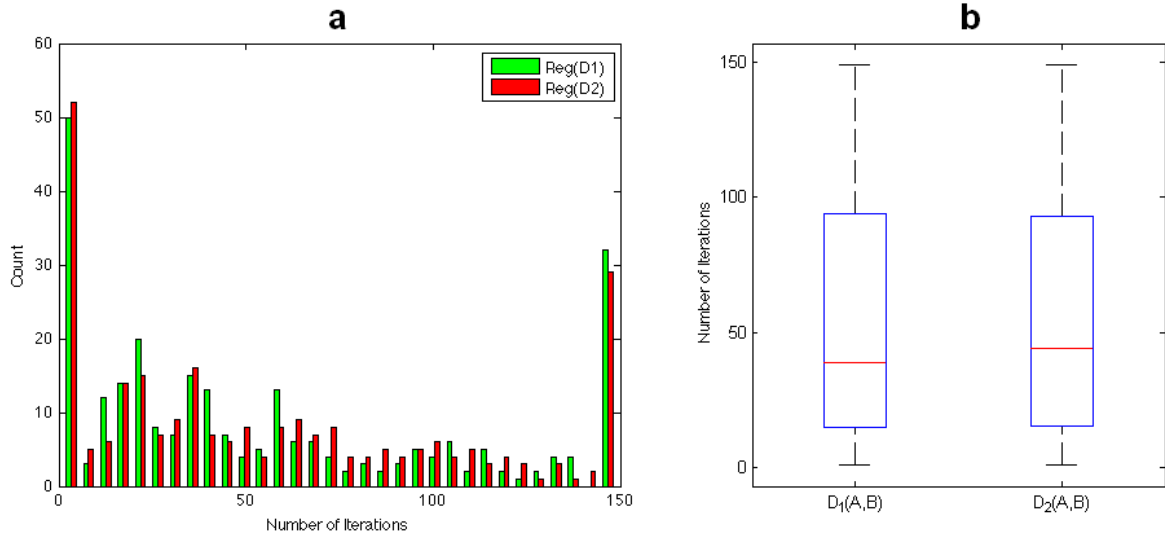
Figure 6. a) Histogram of iteration number for all registrations with either D_1 or D_2 metric. b) Box and whisker plot of the distribution of iteration number for registrations with either D_1 or D_2 metric.
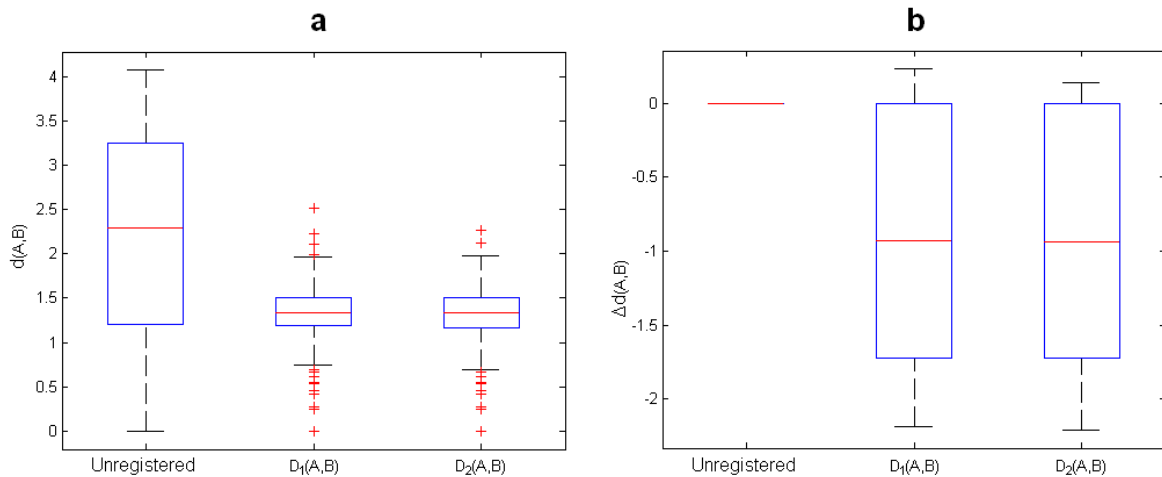

Figure 7. a) Values of the distribution of image similarity (as measured by $d(\mathbf{A},\mathbf{B})$) before and after image registration with either D_1 or D_2 metric. b) box and whisker plot for the distribution of the decrease (or otherwise) of $d(\mathbf{A},\mathbf{B})$ after registration.

Conclusion

It is important with the growing application of mathematically well-defined deformations to medical image registration that the definition of image similarity also has sufficient mathematical foundations. This work provides an early step toward the necessary mathematics of image similarity; we have attempted to develop some formal development of information-based image similarity measures. Information theory has been applied far more extensively in other areas than in medical imaging and this early work develops a sound mathematical framework for medical image similarity. The choice of mutual information is particularly natural for representation as a metric and normalised mutual information can be presented as a justifiable normalised variant of this. The use of a metric obeying mathematical definitions of distance should allow an improved definition of image similarity, allowing intra-subject separation of deformation type, extent and inter-subject identification. Unfortunately, although the distance between two intensity distributions is naturally represented as a metric, the interpretation of those intensity distributions is difficult since they do not explicitly contain any notion of the spatial informa-

tion that defines an image. This might require a definition of the Shannon entropy that incorporates the spatial information[14][15] . As a result, metricated mutual information (the shared information distance) is useful but unlikely to be sufficient to define image similarity. Further work will quantify the limits of applicability of mutual information and investigate extensions to the conventional assumption of a stochastic spatially independent pixel intensity distribution in the formation of an image.

Many registration objective criteria include a sum of dis(similarity) and a norm on some aspect of the deformation, which may be viewed as a distance of the transformation from the identity. In the case of sum-of-squared differences for mono-modal registration, the dissimilarity term is also a distance metric; this paper extends information theoretic measures to allow the mathematical neatness of having a sum of distances (between the images and the transformations) . This aesthetic result could gain practical importance if one attempts to provide a single overall measure of distance between two images based on their combined geometric and photometric differences, as in Miller *et al*[16] . The notion of image similarity being a measure of distance, in combination with a measure of distance for the regularisation term, provides an elegant registration framework.

REFERENCES

[1] Viola and Wells, "Alignment by maximization of mutual information," *International Journal of Computer Vision* **24**, p. 137154, 1997.

[2] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information.," *IEEE Trans Med Imaging* **16**, pp. 187–198, Apr 1997.

[3] C. Studholme, D. Hill, and D. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognition* **32(1)**, pp. 71–86, 1999.

[4] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a survey.," *IEEE Trans Med Imaging* **22**, pp. 986–1004, Aug 2003.

[5] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "F-information measures in medical image registration.," *IEEE Trans Med Imaging* **23**, pp. 1508–1516, Dec 2004.

[6] C. H. Bennett, P. Gacs, M. Li, M. B. Vitanyi, and W. H. Zurek, "Information distance," *IEEE Transactions on Information Theory* **44**, pp. 1407–1423, July 1998.

[7] M. Li, "Information distance," in *Proc. IEEE International Conference on Granular Computing*, pp. 1–1, 10–12 May 2006.

[8] D. Mackay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.

[9] A. Melbourne, D. Hawkes, and D. Atkinson, "Image registration using uncertainty coefficients," in *ISBI*, (1059), 2009.

[10] A. Kraskov, H. Stgbauer, R. G. Andrzejak, and P. Grassberger, "Hierarchical clustering using mutual information," *Europhys. Lett.* **70**, pp. 278–284, 2005.

[11] C. Tanner, J. A. Schnabel, D. L. G. Hill, D. J. Hawkes, A. Degenhard, M. O. Leach, D. R. Hose, M. A. Hall-Craggs, and S. I. Usiskin, "Quantitative evaluation of free-form deformation registration for dynamic contrast-enhanced MR mammography.," *Med Phys* **34**, pp. 1221–1233, Apr 2007.

[12] N. D. Cahill, J. A. Noble, D. J. Hawkes, and L. A. Ray, "Fast fluid registration with Dirichlet boundary conditions," in *International Symposium on Biomedical Imaging*, pp. 712–715, 2007.

[13] W. R. Crum, D. L. G. Hill, and D. J. Hawkes, "Information theoretic similarity measures in non-rigid registration.," *Inf Process Med Imaging* **18**, pp. 378–387, Jul 2003.

[14] C. Studholme, C. Drapaca, B. Iordanova, and V. Cardenas, "Deformation-based mapping of volume change from serial brain mri in the presence of local tissue contrast change.," *IEEE Trans Med Imaging* **25**, pp. 626–639, May 2006.

[15] D. Rueckert, M. J. Clarkson, D. L. G. Hill, and D. J. Hawkes, "Non-rigid image registration using higher-order mutual information," in *Proceedings of SPIE*, 2000.

[16] M. I. Miller, A. Trouve, and L. Younes, "On the metrics and euler-lagrange equations of computational anatomy.," *Annu Rev Biomed Eng* **4**, pp. 375–405, 2002.