

The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration

Alexis Roche, Grégoire Malandain, Xavier Pennec, and Nicholas Ayache

INRIA Sophia Antipolis, EPIDAURE project, France
Alexis.Roche@sophia.inria.fr

Abstract. Over the last five years, new “voxel-based” approaches have allowed important progress in multimodal image registration, notably due to the increasing use of information-theoretic similarity measures. Their wide success has led to the progressive abandon of measures using standard image statistics (mean and variance). Until now, such measures have essentially been based on heuristics. In this paper, we address the determination of a new measure based on standard statistics from a theoretical point of view. We show that it naturally leads to a known concept of probability theory, the *correlation ratio*. In our derivation, we take as the hypothesis the functional dependence between the image intensities. Although such a hypothesis is not as general as possible, it enables us to model the image smoothness prior very easily. We also demonstrate results of multimodal rigid registration involving Magnetic Resonance (MR), Computed Tomography (CT), and Positron Emission Tomography (PET) images. These results suggest that the correlation ratio provides a good trade-off between accuracy and robustness.

1 Introduction

The general principle of voxel-based registration consists of quantifying the quality of matching with respect to a *similarity measure* of the images’ overlapping voxels. As the measure is assumed to be maximal when the images are correctly aligned, these approaches are often implemented using an optimization scheme, or simulating a dynamic process [7].

Many similarity measures have been proposed in the literature (see [3, 15, 2, 6] for reviews). Considering the elementary problem of aligning two similar images, the first idea was to use a least squares criterion. Simple correlation measures were then proposed in order to cope with inter-image bias. Although these similarity measures have been used extensively in medical imaging, they basically assume a linear relationship between the image intensities. Such a hypothesis is generally too crude in multimodal registration.

More recently, Woods et al. [21, 20] have proposed an original criterion which proved itself to be efficient for matching PET with MR. Although the method needs some manual segmentation to work, Nikou et al. [9] have defined a robust version of the criterion that led to a fully automatic algorithm and extended its usage to several modality combinations.

But the currently most popular multimodal measure is probably mutual information [18, 17, 5, 13, 8] since it has been used with success for a large variety of combinations including MR, CT, PET, and SPECT¹. Given two images X and Y , one can define their joint probability density function (joint pdf), $P(i, j)$, by simple normalization of their 2D-histogram (other approaches are possible, see section 2.3). Let $P_x(i)$ and $P_y(j)$ denote the corresponding marginal probability density functions (pdf's). Mutual information between X and Y is given by [1]:

$$I(X, Y) = \sum_{i,j} P(i, j) \log_2 \frac{P(i, j)}{P_x(i)P_y(j)}.$$

The mutual information measure is very general because it makes no assumptions regarding the nature of the relationship that exists between the image intensities (see [16] for an excellent discussion). It does not assume a linear, nor functional correlation but only a *predictable* relationship.

However, one pitfall of mutual information is to treat intensity values in a purely *qualitative* way, without considering any notion of proximity in the intensity space. As one tissue is never represented by a single intensity value, nearby intensities convey a lot of *spatial* information.

Let us illustrate this remark with a synthetic experiment (see figure 1). We consider two artificial images : a binary image A representing a "grey stripe" (40×30 pixels), and a gradation image B of the stripe (30×30 pixels) in which the intensity is uniform in any column but each column has a different intensity.

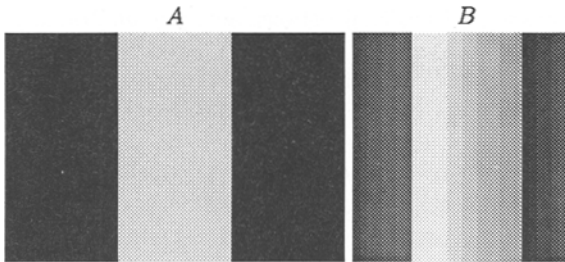


Fig. 1. "Grey stripe" registration experiment.

If we horizontally move B over A , we note that any translation corresponding to an integer number of pixels makes mutual information $I(A, B_T)$ maximal (provided that B_T totally falls into A). Then, $I(A, B_T)$ reaches 1, which is its theoretical upper bound in this case. This is to say that mutual information does not explain how to align the stripes.

Mutual information and the correlation ratio (later explained) have been computed for various horizontal translations of B , using bilinear interpolation for non-integer ones (see figure 2). Unlike mutual information, the correlation

¹ Single Photon-Emission Computed Tomography.

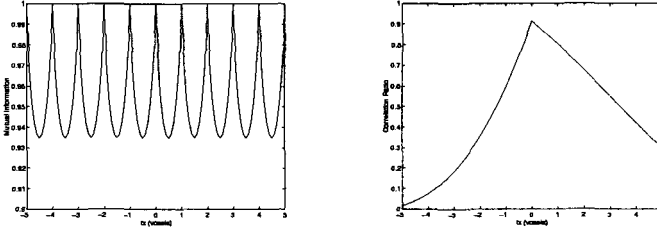


Fig. 2. “Grey stripe” registration experiment. Left, plot of mutual information $I(A, B_T)$ vs. horizontal translation. Right, the correlation ratio. By convention, the null translation corresponds to the case where the stripes completely overlap. Notice that for any integer translation, $I(A, B_T)$ is maximal (for non-integer translations, smaller values are observed due to interpolation).

ratio has an absolute maximum corresponding to the position where the stripes completely overlap.

This example suggests that mutual information may be under-constrained when reasonable assumptions can be made upon the existing relationship between the images. Practically, one often observes its tendency to handle many local maxima. In this paper, we address the case where a *functional* correlation can be assumed, but making minimal assumptions regarding the nature of the function itself. The similarity measure we propose is inherited from probability theory and is known as the *correlation ratio*.

2 Theory

We give an intuitive argument to introduce our approach. Suppose that we have two registered images, X and Y . If we randomly select voxels in their overlapping region, we will observe that the intensity couples we get are statistically consistent: all the voxels having a certain intensity i in X may also have clustered intensities in Y (possibly very different from i). Depending on the images type, any iso-set $X = i$ might project to one or several such clusters. In the case of a single cluster per iso-set, the intensity in Y could be approximately predicted from the intensity in X , by applying a simple function. This argument is valid only if the images are correctly registered. Thus, we could use the degree of *functional* dependence between X and Y as a matching criterion.

How now to measure the functional dependence? In the above thought experiment, images X and Y are considered as random variables. Evaluating the functional dependence between two variables comes down to an unconstrained regression problem. Suppose we want to determine how well X approximates Y . A natural approach is:

- (1) find the function $\phi^*(X)$ that best fits Y among *all* possible functions of X ,
- (2) measure the quality of fitting.

2.1 A solution to problem (1)

One must beforehand determine a cost function in order to perform regression. A convenient choice is variance, which measures a variable's average dispersion around its mean value. Thus, it naturally imposes a constraint of proximity in the sample space. Using variance, our problem is to find

$$\phi^* = \arg \min_{\phi} \text{Var} [Y - \phi(X)]. \quad (1)$$

If no constraint is imposed on the functions ϕ (such as linearity), eq (1) is known to be minimized uniquely by the *conditional expectation* of Y in terms of X [10]. Recall that it is defined by,

$$E(Y|X) = \phi^*(X), \quad \text{with } \phi^*(x) = \int y p(y|x) dy,$$

where $p(y|x)$ denotes the conditional pdf of Y assuming the event $X = x$. To a given event corresponds a given conditional pdf.

2.2 A solution to problem (2)

Now that we have optimally estimated Y in terms of X , we use a result known as the total variance theorem [12, 11], which relies on the orthogonality principle well-known in Kalman filtering:

$$\text{Var}(Y) = \text{Var} [E(Y|X)] + \text{Var} [Y - E(Y|X)]. \quad (2)$$

This may be seen as an energy conservation equation. The variance of Y is decomposed as a sum of two antagonist "energy" terms: while $\text{Var} [E(Y|X)]$ measures the part of Y which is predicted by X , $\text{Var} [Y - E(Y|X)]$ measures the part of Y which is functionally independent of X .

From eq (2), we remark that $\text{Var} [Y - E(Y|X)]$ can actually be low for two distinct reasons: either Y is well "explained" by X ($\text{Var} [E(Y|X)]$ is high), or Y gives little information ($\text{Var}(Y)$ is low). In a registration problem, $\text{Var}(Y)$ can only be computed in the overlapping region of the images. It may be arbitrarily low depending on the region size. Thus, minimizing $\text{Var} [Y - E(Y|X)]$ would tend to completely disconnect the images. Notice that for exactly the same reasons, mutual information is preferred to conditional entropy [16].

It seems more reasonable to compare the "explained" energy of Y with its total energy. This leads to the definition of the *correlation ratio*:

$$\eta(Y|X) = \frac{\text{Var} [E(Y|X)]}{\text{Var}(Y)} \iff \eta(Y|X) = 1 - \frac{\text{Var} [Y - E(Y|X)]}{\text{Var}(Y)}. \quad (3)$$

The correlation ratio measures the functional dependence between X and Y . It takes on values between 0 (no functional dependence) and 1 (purely deterministic dependence). Due to the use of a ratio instead of a subtraction, $\eta(Y|X)$ is invariant to multiplicative changes in Y , i.e. $\forall k, \eta(kY|X) = \eta(Y|X)$. Also note that the correlation ratio is asymmetrical by nature since the two variables fundamentally do not play the same role in the functional relationship; in general, $\eta(Y|X) \neq \eta(X|Y)$.

2.3 Application to registration

In order to compute the correlation ratio between two images, we must be able to define them as random variables, that is determine their marginal and joint pdf's. A common technique consists of normalizing the image pair 2D-histogram [4, 5, 2]. Then, the images may be seen as discrete random variables [11]. Viola [16] has proposed a continuous approach using Parzen density estimates.

If we choose the discrete approach, there is no need to manipulate explicitly the images 2D-histogram. Instead, the correlation ratio can be computed recursively by accumulating local computations. Let Ω denote the images overlapping region, and $N = \text{Card}(\Omega)$ the total number of voxels it contains. We consider the iso-sets of X , $\Omega_i = \{\omega \in \Omega, X(\omega) = i\}$ and their cardinals $N_i = \text{Card}(\Omega_i)$. The total and conditional moments (mean and variance) of Y are:

$$\begin{aligned}\sigma^2 &= \frac{1}{N} \sum_{\omega \in \Omega} Y(\omega)^2 - m^2, & m &= \frac{1}{N} \sum_{\omega \in \Omega} Y(\omega). \\ \sigma_i^2 &= \frac{1}{N_i} \sum_{\omega \in \Omega_i} Y(\omega)^2 - m_i^2, & m_i &= \frac{1}{N_i} \sum_{\omega \in \Omega_i} Y(\omega).\end{aligned}$$

Starting from eq (3), we obtain a very simple expression for the correlation ratio (the complete proof can be found in [11]):

$$1 - \eta(Y|X) = \frac{1}{N\sigma^2} \sum_i N_i \sigma_i^2. \quad (4)$$

The algorithm derived from these equations does not require the computation of the images 2D-histogram. This makes an important difference with mutual information. Classical algorithms for computing mutual information have an $\mathcal{O}(n_x n_y)$ complexity, n_x and n_y being the number of intensity levels in the X and Y images, respectively. Our computation of the correlation ratio has only an $\mathcal{O}(n_x)$ complexity, and is independent from n_y .

3 Related measures

The correlation ratio generalizes the correlation coefficient, which is a symmetrical measure of *linear* dependence between two random variables:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)^2}{\text{Var}(X)\text{Var}(Y)}.$$

The correlation coefficient is closely related to the various correlation measures that have been used in image registration. The linear dependence being a stronger constraint than the functional dependence, it can be shown that [12, 11],

$$\eta(Y|X) \geq \rho(X, Y), \quad \eta(X|Y) \geq \rho(X, Y).$$

We now analyze two similarity measures which are based on standard statistics but not limited to the case of linear correlation.

3.1 Woods criterion

The heuristic criterion devised by Woods et al. [21] was originally intended for PET-MR registration, but it has also been used with other modalities [2]. This turns out to be very similar to the correlation ratio. According to the notations introduced in section 2.3, the Woods criterion can be written as follows:

$$W(Y|X) = \frac{1}{N} \sum_i N_i \frac{\sigma_i}{m_i}, \quad (5)$$

where the notation $W(Y|X)$ is used in order to emphasize that the criterion is asymmetrical, such as the correlation ratio. Notice that $W(Y|X)$ has to be *minimized*, just like $1 - \eta(Y|X)$.

Though different, eq (5) and eq (4) express the same basic idea. Even so, we can identify two differences. First, the correlation ratio sums variances, σ_i^2 whereas the Woods criterion sums normalized standard deviations, σ_i/m_i . Second, the multiplicative invariance property is achieved in the correlation ratio via a global division by σ^2 ; in the Woods criterion, every term of the sum is divided by a conditional mean, m_i .

3.2 Weighted neighbor likelihood

In [16], Viola already proposed performing registration by evaluating the degree of functional dependence between two images. This approach is very analogous to that we have proposed in section 2. First, a *weighted neighbor approximator* is used to estimate the Y image in terms of the X image. Second, a similarity measure is obtained by considering the estimation log-likelihood (under hypotheses we won't discuss here).

We have previously shown [11] that the approximator devised by Viola is nothing but the conditional expectation of Y in terms of a variable, \tilde{X} , whose pdf is the Parzen estimate of X . Furthermore, the weighted neighbor likelihood is negatively proportional to the estimation error:

$$L(Y|X) = -k \text{Var} \left[Y - E(Y|\tilde{X}) \right], \quad k > 0.$$

Maximizing the weighted neighbor likelihood is in fact equivalent to minimizing the *numerator* in eq (3) (up to the use of Parzen windowing). However, the correlation ratio involves a division by $\text{Var}(Y)$, which plays a critical role in registration problems since it prevents disconnecting the images (see section 2.2).

4 Results

We tested voxel-based 3D multimodal registration over ten patient brain datasets. For each patient, the following images were available :

- MR, T1 weighted ($256 \times 256 \times 20/26$ voxels of $1.25 \times 1.25 \times 4 \text{ mm}^3$)

- MR, T2 weighted ($256 \times 256 \times 20/26$ voxels of $1.25 \times 1.25 \times 4 \text{ mm}^3$)
- CT ($512 \times 512 \times 28/34$ voxels of $0.65 \times 0.65 \times 4 \text{ mm}^3$)
- PET ($128 \times 128 \times 15$ voxels of $2.59 \times 2.59 \times 8 \text{ mm}^3$)

All images were stored with one byte per voxel. The gold standard transformations between each modality were known thanks to a prospective, marker-based registration method [19]. No preprocessing of the images was done.

We implemented an algorithm similar to that of Maes et al. [5], employing Powell's multidimensional direction set method as a maximization scheme. Four similarity measures were tested: the correlation ratio (CR), mutual information (MI), the correlation coefficient (CC), and the opposite of the Woods criterion (OW). The choice of the opposite is only for consistency: OW has to be maximized like MI, CR, and CC. In all registration experiments, the transformation was initialized as the identity.

We used two different interpolation techniques: trilinear interpolation (TRI) and trilinear partial volume interpolation [5] (PV). The results that are presented here were obtained using PV interpolation; on the whole, they are better than those obtained with TRI interpolation.

After each registration, a "typical" error ϵ was computed in the following way. We selected eight points in the transformed image, approximately situated on the skull surface. Registration errors corresponding to these points were computed according to the marker-based transformation, and then averaged to obtain ϵ .

Table 1. Mean and median of the registration typical errors (based on positions of stereotaxic markers) obtained over ten intra-patient experiments.

Experiment	Measure	Mean ϵ (mm)	Median ϵ (mm)
T1-to-T2	MI	4.30	1.48
	CR (X:T2)	1.93	1.46
	OW (X:T2)	2.65	2.00
	CC	2.42	2.37
CT-to-T1	MI	2.52	2.00
	CR (X:T1)	3.27	3.24
PET-to-T1	MI	5.87	5.58
	CR (X:T1)	4.60	3.65
	OW (X:T1)	7.69	7.62

Statistics on typical errors over the ten patients are shown in table 1. We got non sensible results with CC in CT-to-T1 and PET-to-T1 registration, and with OW in CT-to-T1 registration. Conversely, MI and CR demonstrated suitable accuracy levels for every modality combination. MI gave the best results for CT-to-T1 registration, while CR was better for PET-to-T1 registration. In the case of T1-to-T2 registration, CR and MI generally provided the best results, but MI failed in two cases. Notice that for CT-to-T1 registration, the CT images were subsampled by factors $2 \times 2 \times 1$ in the x , y , and z direction, respectively; due to

the large dimensions of the CT images, registration at full resolution was indeed too time consuming.

Several subsampling factors were also tested for every modality combination in order to speed up the registration process with minimal loss in terms of accuracy. A typical drawback of subsampling is to introduce local maxima in the similarity measure so that the global maximum becomes difficult to track.

Table 2. *The correlation ratio performances depending on resolution.*

Experiment	Subsampling	Mean ϵ (mm)	Median ϵ (mm)
T1-to-T2	$(2 \times 2 \times 1)$	1.90	1.48
	$(4 \times 4 \times 1)$	2.01	1.67
CT-to-T1	$(4 \times 4 \times 1)$	4.23	3.55
	$(8 \times 8 \times 1)$	6.65	5.96
PET-to-T1	$(2 \times 2 \times 1)$	6.82	6.20
	$(4 \times 4 \times 1)$	11.65	11.19

The influence of subsampling on the correlation ratio performances was remarkably moderate (see table 2). While CR allowed good registration at relatively low resolutions, other studies [11] (which could not be presented here) qualitatively demonstrated that CR was less sensitive to subsampling than MI and OW.

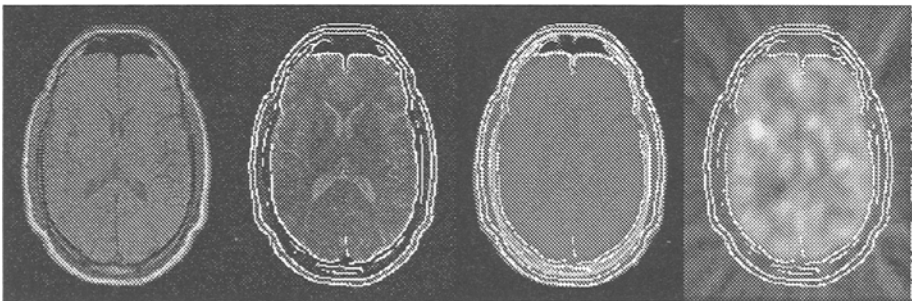


Fig. 3. *Multimodal registration by maximization of CR. Images from left to right : MR-T1, MR-T2, CT, and PET. The images are resampled in the same reference frame after registration. Contours extracted from the MR-T1 are superimposed to each other modality in order to visualize the quality of registration.*

5 Discussion and conclusion

Our experiments tend to show that assuming a functional correlation between certain multimodal images is not critical. Even if this is an approximation, notably in the CT-MR case (see [18] for a discussion), a preprocessing step might validate it. Van den Elsen et al. [14] have proposed a simple intensity mapping to the original CT image so that bone and air appear in the same intensity range as is the case in MR images. Then, low intensities in MR (air and bone) may project to nearby intensities in CT. Another possible strategy for CT-MR registration could be to use the correlation ratio for a quick guessing of the correct transformation (using subsampling), and then mutual information for probably more accurate alignment.

The case of PET images is particular because they are much more distorted than MR or CT. This might explain why mutual information is relatively inaccurate in PET-T1 registration. It is generally admitted that today Woods method is the best one for this specific problem. In some way, our results corroborate this observation, suggesting that taking into account nearby intensities in PET images might be crucial. Mutual information seems to be better adapted to low-noise images.

Finally, the discrepancies found experimentally between the correlation ratio and the Woods criterion are surprising since these two measures are formally based on similar considerations (see section 3.1). It seems that the correlation ratio gives not only a theoretical justification to the Woods criterion but also perceptible practical improvements.

Acknowledgments

The images and the standard transformations were provided as part of the project, "Evaluation of Retrospective Image Registration", National Institutes of Health, Project Number 1 R01 NS33926-01, Principal Investigator, J. Michael Fitzpatrick, Vanderbilt University, Nashville, TN.

Many thanks to Frederik Maes, Jean-Pierre Nadal, and Christophoros Nikou for fruitful discussion and to Janet Bertot for the proofreading of this article.

References

1. R. E. Blahut. *Principles and Practice of Information Theory*. Addison-Wesley Pub. Comp., 1987.
2. M. Bro-Nielsen. Rigid Registration of CT, MR and Cryosection Images Using a GLCM Framework. *CVRMed-MRCAS'97*, pages 171-180, March 1997.
3. L. G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325-376, 1992.
4. D. L. G. Hill and D. J. Hawkes. Medical image registration using voxel similarity measures. *AAAI Spring Symposium Series: Applications of Comp. Vision in Med. Im. Proces.*, pages 34-37, 1994.

5. 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(2):187-198, 1997.
6. J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *MedIA*, 2(1):1-36, 1998.
7. G. Malandain, S. Fernández-Vidal, and J.C. Rocchisani. Improving registration of 3-D images using a mechanical based method . *ECCV'94*, pages 131-136, May 1994.
8. C. R. Meyer, J. L. Boes, B. Kim, P. H. Bland, K. R. Zasadny, P. V. Kison, K. Koral, K. A. Frey, and R. L. Wahl. Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate warped geometric deformations. *Medical Image Analysis*, 1(3):195-206, 1996/7.
9. C. Nikou, F. Heitz, J.-P. Armspach, and I.-J. Namer. Single and multimodal subvoxel registration of dissimilar medical images using robust similarity measures. *SPIE Conference on Medical Imaging*, February 1998.
10. A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Inc., third edition, 1991.
11. A. Roche, G. Malandain, X. Pennec, and N. Ayache. Multimodal Image Registration by Maximization of the Correlation Ratio. Technical Report 3378, INRIA, March 1998.
12. G. Saporta. *Probabilités, analyse des données et statistique*. Editions Technip, Paris, 1990.
13. C. Studholme, D. L. G. Hill, and D. J. Hawkes. Automated 3-D registration of MR and CT images of the head. *Medical Image Analysis*, 1(2):163-175, 1996.
14. P. A. van den Elsen, E.-J. D. Pol, T. S. Sumanaweera, P. F. Hemler, S. Napel, and J. R. Adler. Grey value correlation techniques for automatic matching of CT and MR brain and spine images. *Proc. Visualization in Biomed. Comp.*, 2359:227-237, October 1994.
15. P.A. van den Elsen, E.J.D. Pol, and M.A. Viergever. Medical image matching - a review with classification. *IEEE Engineering in Medicine and Biology*, 12(4):26-39, march 1993.
16. P. Viola. *Alignment by Maximization of Mutual Information*. PhD thesis, M.I.T. Artificial Intelligence Laboratory, 1995. also A.I.T.R. No. 1548, available at <ftp://publications.ai.mit.edu>.
17. P. Viola and W. M. Wells. Alignment by Maximization of Mutual Information. *Intern. J. of Comp. Vision*, 24(2):137-154, 1997.
18. W. M. Wells, P. Viola, H. Atsumi, and S. Nakajima. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35-51, 1996.
19. J. West and al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Comp. Assist. Tomography*, 21:554-566, 1997.
20. R. P. Woods, S. R. Cherry, and J. C. Mazziotta. Rapid Automated Algorithm for Aligning and Reslicing PET Images. *Journal of Comp. Assist. Tomography*, 16(4):620-633, 1992.
21. R. P. Woods, J. C. Mazziotta, and S. R. Cherry. MRI-PET Registration with Automated Algorithm. *Journal of Comp. Assist. Tomography*, 17(4):536-546, 1993.