

Towards a Better Comprehension of Similarity Measures used in Medical Image Registration

Alexis Roche, Grégoire Malandain, Nicholas Ayache, and Sylvain Prima

INRIA Sophia Antipolis - Epidaure Project
2004 Route des Lucioles BP 93
06902 Sophia Antipolis Cedex, France
{aroch,greg,na,sprima}@sophia.inria.fr

Abstract. While intensity-based similarity measures are increasingly used for medical image registration, they often rely on implicit assumptions regarding the physics of imaging. The motivation of this paper is to determine what are the assumptions corresponding to a number of popular similarity measures in order to better understand their use, and finally help choosing the one which is the most appropriate for a given class of problems. After formalizing registration based on general image acquisition models, we show that the search for an optimal measure can be cast into a maximum likelihood estimation problem. We then derive similarity measures corresponding to different modeling assumptions and retrieve some well-known measures (correlation coefficient, correlation ratio, mutual information). Finally, we present results of registration between 3D MR and 3D Ultrasound images to illustrate the importance of choosing an appropriate similarity measure.

1 Introduction

Over the last years, intensity-based (or iconic) techniques have been applied to a number of registration problems including monomodal as well as multimodal, and rigid as well as non-rigid registration [8]. Their basic principle is to maximize a criterion measuring the intensity similarity of corresponding voxels. Common to the many proposed similarity measures is the idea that, when matched, the images should verify a certain relationship; the similarity measure is intended to quantify how well this relationship is verified depending on the considered transformation between the images.

Choosing one measure adapted to a specific registration problem is not always straightforward for at least two reasons. First, it is often difficult to model the physical relationship that exists between two images. Second, most of the similarity measures rely on imaging assumptions that are not fully explicit. Existing similarity measures may be classified into four main kinds of hypotheses:

Identity relationship. The basic assumption is that when matched the images are identical. This includes a number of popular measures: the sum of squared intensity differences (SSD), the sum of absolute intensity differences, cross-correlation

[3], entropy of the difference image [4],... Although these measures are not equivalent in terms of robustness and accuracy, none of them is able to cope with relative intensity changes from one image to the other.

Affine relationship. The step beyond is to assume that the two images I and J to be registered are related by an affine mapping, i.e. $I \approx \alpha J + \beta$. The measures adapted to this situation are more or less variants from the correlation coefficient [3], defined as the ratio between the images' covariance and the product of individual standard deviations:

$$\rho(I, J) = \frac{\text{Cov}(I, J)}{\sqrt{\text{Var}(I)} \sqrt{\text{Var}(J)}}. \quad (1)$$

The correlation coefficient is generally useful for matching images from the same modality. Nevertheless, the affine hypothesis is hardly valid for images from different modalities, and thus it has not provided convincing results in multimodal registration.

Functional relationship. For multimodal images, more complex relationships are involved. The approach we proposed in [15] was to assume that, at the registration position, one image could be approximated in terms of the other by applying some intensity function, $I \approx f(J)$. Making no assumption regarding the nature of the function, we derived a natural statistical measure, the correlation ratio:

$$\eta^2(I|J) = 1 - \frac{\text{Var}(I - \hat{f}(J))}{\text{Var}(I)}, \quad (2)$$

where $\hat{f}(J)$ is the least square optimal non-linear approximation of I in terms of J [11]. The correlation ratio is closely related to a very popular measure previously proposed by Woods et al.[23], and generalized using robust metrics in [10].

Statistical relationship. Finally, assuming a functional relationship is sometimes too restrictive. Then, it is more appropriate to use information theoretic measures, from which mutual information [7, 20] is today probably the most popular:

$$\mathcal{I}(I, J) = \sum_i \sum_j \log \frac{p(i, j)}{p(i) p(j)}, \quad (3)$$

where $p(i, j)$ is the intensity joint probability distribution of the images, and $p(i)$ and $p(j)$ the corresponding marginal distributions. This category is not fundamentally different from the previous one, as the ideal case is still perfect functional dependence; mutual information is however theoretically more robust to variations with respect to this ideal situation.

A number of comparison studies have shown that similarity measures yield different performances depending on the considered modality combinations [22, 2, 13, 10, 15]. There is probably no universal measure and, for a specific problem,

the point is rather to choose the one that is best adapted to the nature of the images.

Up to now, the link between explicit modeling assumptions and similarity measures has not been made clear. After some authors [9, 5] proposed that image registration could be seen as a maximum likelihood estimation problem, Viola et al. [20, 21] suggested the analogy of this approach with registration using information theory. Remarkably, other teams had motivated information-theoretic measures by other arguments [7, 17].

In section 2, we propose to formulate image registration as a general maximum likelihood estimation problem. Then, deriving optimal similarity measures from specific modeling assumptions, we retrieve the correlation ratio and mutual information in section 3. Section 4 proposes to illustrate the practical differences between these two measures with results of registration between 3D MR and 3D Ultrasound images.

2 Formulation

2.1 Maximum likelihood registration

Two images I and J to be registered are related through the common anatomical reality that they measure. But the way anatomical structures are represented depends on the physics of imaging involved in each acquisition. Thus, modeling the relationship between the images requires the knowledge of both the underlying anatomy and the image formation processes. A convenient model of the anatomy will be an image called “segmentation” or “scene”: by definition, a scene is any image for which the intensity of a given voxel represents directly the tissue class it belongs to.

Assuming that we know a scene, we can model indirectly the relationship between I and J based on image acquisition models. A standard approach is to interpret an image as being a realization of a random process that corrupts the scene. This means that the relationship between I and S (resp. J and S) is defined in terms of a conditional probability density function $P(I|S)$. The two following assumptions are usually stated:

- The voxels of the image are conditionally independent knowing the scene, that is:

$$P(I|S) = \prod_{x_k \in \Omega_I} P(i_k|S),$$

where Ω_I denotes the voxel grid of I , and $i_k \equiv I(x_k)$ is the intensity of the voxel with coordinates x_k in a given frame attached to the grid Ω_I .

- The noise is context-free. In other words, the intensity of a voxel depends only on its homologous in the scene:

$$P(i_k|S) = P(i_k|s_k^\downarrow), \quad \text{with} \quad s_k^\downarrow \equiv S(T(x_k)) = (S \circ T)(x_k),$$

where T is the spatial transformation that relates the coordinate frames of Ω_I and Ω_S , the grid of S . In the case where I and S are not supposed to be aligned, T has no reason to be the identity.

Under these assumptions, the conditional probability of I knowing the scene, S , and the transformation, T , is easily seen to be:

$$P(I|S, T) = \prod_{x_k \in \Omega_I} P(i_k | s_k^\downarrow). \quad (4)$$

We can model the relationship between J and S in the same manner. However, as we are interested only in the relative displacement between I and J , we will consider J as a “reference” image being already aligned with the scene, meaning that no transformation is involved in the conditional probability,

$$P(J|S) = \prod_{y_l \in \Omega_J} P(j_l | s_l), \quad \text{with } j_l = J(y_l), \quad s_l = S(y_l), \quad (5)$$

$\Omega_J \equiv \Omega_S$ being the voxel grid of J , which coincides with that of S . Without knowledge of the scene, the probability of the image pair (I, J) is obtained by integrating over all possible realizations of S . Assuming that the two acquisitions are independent, we have $P(I, J|S, T) = P(I|S, T) P(J|S)$, and thus:

$$P(I, J|T) = \int P(I|S, T) P(J|S) P(S) dS = P(J) \underbrace{\int P(I|S, T) P(S|J) dS}_{P(I|J, T)}, \quad (6)$$

where the last step relies on Bayes rule. The transformation T appears as a parameter of this joint probability function, and it is natural to invoke the maximum likelihood principle to formulate registration, as already proposed for instance in [20, 6, 1, 9, 5]. It simply states that the most likely transformation between I and J is the one that maximizes the joint probability of (I, J) ,

$$\hat{T} = \arg \max_T P(I, J|T) = \arg \max_T P(I|J, T),$$

the last equivalence holding because J is independent of T .

Evaluating the integral of (6) may be somewhat cumbersome if the prior probability on S is a complex function. For mathematical convenience, we will assume that the voxels of the scene are independently distributed, yielding the factored form: $P(S) = \prod_{y_l} P(s_l)$. Thus, no auto-correlation is assumed for the scene: this is the minimal way to introduce prior anatomical information. Notice that this does not mean that the voxels are identically distributed, such that spatial dependences may be incorporated to the model. If this modeling is still insufficient, a maximum a posteriori (MAP) estimation strategy would be employed. This alternative, using for example a Gibbs-Markov random field prior, requires an explicit estimation of the scene that is computationally very expensive in the case of 3D images.

Under these assumptions, we show in [16] that $P(I|J, T)$ is of the same factored form as $P(I|S, T)$; letting $j_k^\downarrow \equiv J(T(x_k))$, we have:

$$P(I|J, T) = \prod_{x_k \in \Omega_I} P(i_k | j_k^\downarrow), \quad (7)$$

$$\text{with } P(i_k | j_k^\downarrow) = \int P(i_k | s_k^\downarrow = s) P(s_k^\downarrow = s | j_k^\downarrow) ds. \quad (8)$$

2.2 Estimating the conditional densities

In the framework where the transformation is found by maximum likelihood, the most natural way to estimate densities is also to use a maximum likelihood strategy. This means that we can search for the conditional densities $P(i_k|j_k^\downarrow)$ that maximize exactly the same criterion as in (7). Basically, this is a parametric approach: we assume that the $P(i_k|j_k^\downarrow)$ belong to a given class of distributions parameterized by a vector θ (regardless, for the moment, of what θ represents); then their maximum likelihood estimates, for a given estimate of the transformation T , are found by:

$$\hat{\theta}(T) = \arg \max_{\theta} P(I|J, T, \theta) = \arg \max_{\theta} \prod_{x_k \in \Omega_I} P_{\theta}(i_k|j_k^\downarrow).$$

The parametric form of $P(I|J, T, \theta)$ may be derived from the modeling assumptions presented in section 2.1 whenever all the components of the model, $P(I|S, T)$, $P(J|S)$, and $P(S)$ are themselves chosen as parametric densities. Then, from (8), the form of $P(I|J, T, \theta)$ can be known. In section 3, we will show that under some specific modeling assumptions, maximum likelihood density estimates can be computed explicitly.

For practical optimization, it is more convenient to consider the negative log-likelihood (to be minimized); thus, we will define the energy of a transformation T as:

$$U(T) = -\log \max_{\theta} P(I|J, T, \theta) = \min_{\theta} \left[- \sum_{x_k \in \Omega_I} \log P_{\theta}(i_k|j_k^\downarrow) \right]. \quad (9)$$

2.3 Practical issues

In section 2.1, we have derived the likelihood registration criterion under the implicit assumption that the transformation T is searched among mappings from the floating image grid, Ω_I , to the reference image grid, Ω_J . This means that a given voxel of image I is always supposed to match a node of the reference grid. Therefore, the spatial resolution of the transformation is intrinsically limited by the resolution of the reference grid, and clearly this assumption cannot deal with sub-voxel accurate registration.

In order to manipulate continuous transformations, we can oversample the reference image using interpolation techniques [7]. But this is possible only if the transformed position of a voxel falls inside the reference domain. Since this domain has a finite extension in space, other voxels may fall outside, so that there is not enough information to extrapolate the intensity of their correspondent (see figure 1). The problem of how to treat these “outside” voxels plays an important role in voxel-based image registration. They are generally ignored by the registration criterion, which necessitates some heuristic normalization to avoid nasty effects such as disconnecting the images [18, 20, 15].

We discuss in [16] that ignoring these voxels would not be consistent with the maximum likelihood framework. Instead, a natural approach is simply to extend

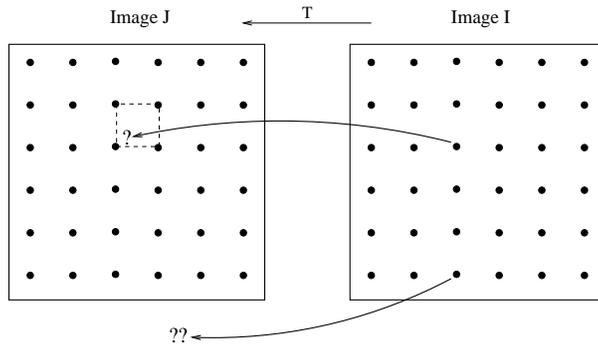


Fig. 1. Effects of applying a continuous spatial transformation. The point marked with '?' can be interpolated, unlike the point marked with '??'. Instead, this latter point may be assigned an arbitrary intensity j^* .

the reference domain by assigning the external points to a constant and arbitrary intensity $J = j^*$. Although this is a computational artifice, this enables us to take into account every voxel of I at each iteration of the registration process.

3 From modeling assumptions to similarity measures

3.1 Gaussian channel

Perhaps the simplest model we can imagine is that the image J be a valid scene ($J = S$) and the image I be a measure of J corrupted with additive and stationary Gaussian white noise:

$$I(x_k) = f(J(T(x_k))) + \epsilon_k,$$

where f is some unknown intensity function: each tissue class j is imaged in I with an average response value $f(j) = f_j$. Then, the conditional densities $P(i_k | j_k^\downarrow)$ are of the Gaussian form:

$$P(i_k = i | j_k^\downarrow = j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(i-f_j)^2}{2\sigma^2}},$$

and the parameter vector $\theta = (f_0, f_1, \dots, \sigma)$ needs to be estimated. In order to minimize the negative log-likelihood (equation 9) with respect to θ , we group the voxels x_k that match the same class. Letting $N = \text{Card } \Omega_I$, $\Omega_I^j = \{x_k \in \Omega_I, j_k^\downarrow = j\}$, and $N_j = \text{Card } \Omega_I^j$, we have:

$$\begin{aligned} -\log P(I|J, T, \theta) &= N \log \sqrt{2\pi}\sigma + \frac{1}{2} \sum_{x_k \in \Omega_I} \frac{(i_k - f(j_k^\downarrow))^2}{\sigma^2} \\ &= N \log \sqrt{2\pi}\sigma + \frac{1}{2} \sum_j \sum_{x_k \in \Omega_I^j} \frac{(i_k - f_j)^2}{\sigma^2}. \end{aligned} \quad (10)$$

The optimal parameters are then easily found by differentiating the log-likelihood:

$$\begin{aligned} -\frac{\partial \log P}{\partial f_j} &= -\frac{1}{\sigma^2} \sum_{\Omega_I^j} (i_k - f_j) \implies \hat{f}_j = \frac{1}{N_j} \sum_{\Omega_I^j} i_k, \\ -\frac{\partial \log P}{\partial \sigma} &= \frac{N}{\sigma} - \frac{1}{\sigma^3} \sum_j \sum_{\Omega_I^j} (i_k - f_j)^2 \implies \hat{\sigma}^2 = \sum_j \frac{N_j}{N} \hat{\sigma}_j^2, \end{aligned}$$

where $\hat{\sigma}_j^2 \equiv \frac{1}{N_j} \sum_{x_k \in \Omega_I^j} (i_k - \hat{f}_j)^2$ is the image variance corresponding to the iso-set Ω_I^j . The registration energy $U(T)$ is then obtained by substituting the optimal θ parameter:

$$U(T) = N \log \left[\sqrt{2\pi} e \sum_j \frac{N_j}{N} \hat{\sigma}_j^2 \right] = N \log \left[\sqrt{2\pi} e \text{Var}(I - \hat{f}(J^\perp)) \right].$$

This result has a satisfying interpretation: $U(T)$ decreases with the variance of the difference image between I and the intensity corrected $\hat{f}(J)$. The intensity function \hat{f} is nothing but a least-square fit of the image I in terms of the reference J : it is in fact the same fitting function as in the definition of the correlation ratio (equation 2) [15], and we see that the registration energy $U(T)$ is related to the correlation ratio $\eta^2(I|J^\perp)$ by:

$$\eta^2(I|J^\perp) = 1 - \frac{1}{k} e^{\frac{U(T)}{N}}, \quad \text{with } k = \sqrt{2\pi} e \text{Var}(I).$$

In the original version of the correlation ratio [15], the quantities N and $\text{Var}(I)$ were computed only in the overlap between the images, and thus they could vary according to the considered transformation. Their role was precisely to prevent the image overlap from being minimized. In the implementation proposed in section 2.3, N and $\text{Var}(I)$ are independent of the considered transformation. Minimizing $U(T)$ is then strictly equivalent to maximizing the correlation ratio, although is not strictly equivalent to maximizing the original version of the correlation ratio. In our experiments, this distinction seemed to have very little impact on the results.

Another remark is that we can impose constraints to the intensity function f , e.g. to be polynomial [16]. Notably, if we constraint f to follow an affine variation with respect to j , i.e. $f(j) = \alpha j + \beta$, we get a similar equivalence with the correlation coefficient defined in (1):

$$\rho^2(I, J^\perp) = 1 - \frac{1}{k} e^{\frac{U(T)}{N}}, \quad \text{with } k = \sqrt{2\pi} e \text{Var}(I).$$

3.2 Unspecified channel

A straightforward extension of the previous model would be to assume the reference image J to be also corrupted with Gaussian noise. Then, having defined

the prior probabilities for the tissue classes, we could derive the analytical form of the conditional densities $P(i_k|j_k^\perp)$ from (8). This case has been investigated by Leventon and Grimson [6]: it turns out that there is probably nothing much faster than a EM algorithm to provide maximum likelihood estimates of the density parameters.

In order to get explicit density estimates, we can relax every formal constraint on the model. Then, the densities $P(i_k|j_k^\perp)$ are totally unspecified, and we will only assume that they are stationary, i.e. $P(i_k = i|j_k^\perp = j)$ is independent of the position x_k . For the sake of simplicity, we consider the case of discrete densities, but the study is similar for continuous densities. The problem is now to minimize

$$-\log P(I|J, T, \theta) = \sum_{x_k \in \Omega_I} -\log f(i_k|j_k^\perp),$$

with respect to $\theta = (f(0|0), f(1|0), \dots, f(1|1), \dots, f(2|0), \dots)$ and under the constraints: $\forall j, C_j = \sum_i f(i|j) - 1 = 0$. We regroup the intensity pairs (i_k, j_k^\perp) that have the same values:

$$\Omega_{i,j} = \{x_k \in \Omega_I, I(x_k) = i, J(T(x_k)) = j\}, \quad N_{i,j} = \text{Card } \Omega_{i,j}.$$

Then, the negative log-likelihood becomes:

$$-\log P(I|J, T, \theta) = - \sum_{i,j} N_{i,j} \log f(i|j).$$

Introducing Lagrange multipliers, there exist constants $\lambda_0, \lambda_1, \dots$, such that for any j :

$$0 = \frac{\partial \log P}{\partial f(i|j)} - \sum_{j'} \lambda_{j'} \frac{\partial C_{j'}}{\partial f(i|j)} = \frac{N_{i,j}}{f(i|j)} - \lambda_j.$$

Thus, taking into account the constraints $\sum_i f(i|j) = 1$, the optimal parameters verify:

$$\hat{f}(i|j) = \frac{N_{i,j}}{N_j} = \frac{p(i,j)}{p(j)},$$

where $p(i,j) \equiv N_{i,j}/N$ is the image normalized 2D histogram and $p(j) \equiv \sum_i p(i,j)$ the corresponding marginal distribution for J^\perp . From (3), we see that $U(T)$ is nothing but a decreasing function of mutual information:

$$U(T) = -N \sum_{i,j} p(i,j) \log \frac{p(i,j)}{p(j)} = N [H(I) - \mathcal{I}(I, J^\perp)],$$

where $H(I)$ is the entropy of image I and is constant in the implementation proposed in section 2.3. The same remark as in 3.1 holds for the distinction between the usual implementation of mutual information and the one considered here.

3.3 Comparison of measures

In the derivation of the correlation ratio (CR), it was assumed that the image to be registered is a measure of the reference corrupted with additive and stationary Gaussian white noise. In contrast, for deriving mutual information (MI), no assumption was made apart from stationarity and, of course, the general assumptions discussed in section 2. Does it make MI necessarily a better registration measure than CR ?

In principle, the answer is no whenever the assumptions of CR are verified by the images. Basically, these are reasonable if the reference image can be considered as a good anatomical model: in practice, this is seldom perfectly true. The problem then becomes to determine what is better between an over-constrained and an under-constrained measure, a question to which experiments can yield some insight, as will be illustrated in the next section.

4 Experiments

In order to illustrate the practical differences between CR and MI, we present results of 3D rigid registration between two brain images of the same patient: an MR, T1 weighted scan ($256 \times 256 \times 124$ voxels of $0.9 \times 0.9 \times 1.1$ mm), and an intra-operative 3D ultrasound (US) image ($180 \times 136 \times 188$ voxels of 0.95^3 mm). As the US image was acquired before opening the duramater, there is essentially a rigid displacement to find. In these experiments, we used a registration algorithm similar to MIRIT [7], using Powell’s method as an optimization strategy and partial volume (PV) interpolation.

The correct registration position was found manually using an interactive matching tool, and then validated by a clinician. The estimated accuracy was 2 degrees in rotation and 2 mm in translation [12]. We took this result as a “ground truth” for subsequent experiments.

We then performed 200 automatic registrations by initializing the algorithm with random displacements from the “ground truth” position: a rotation vector Δr with random direction and constant magnitude $\|\Delta r\| = 15$ degrees, and a translation Δt with random direction and constant magnitude $\|\Delta t\| = 20$ mm. These values correspond to the variation between the “ground truth” and the original position. For each random transformation, two registrations were performed using alternatively CR and MI.

Two kinds of results are observed: either the algorithm retrieves the “ground truth” transformation (yielding errors systematically lower than $\|\delta r\| = 2$ degrees and $\|\delta t\| = 2$ mm), or it converges to a local maximum (yielding errors systematically larger than $\|\delta r\| = 10$ degrees and $\|\delta t\| = 10$ mm). The main result is that CR fails in 14% cases while MI fails in 51% cases (see table 1). The RMS errors computed on successful registrations are lower than the expected accuracy of the “ground truth”; thus, they prove nothing but the fact that both CR and MI have a maximum in the neighborhood of the ideal registration transformation (this is probably also a global maximum). However, the percentages

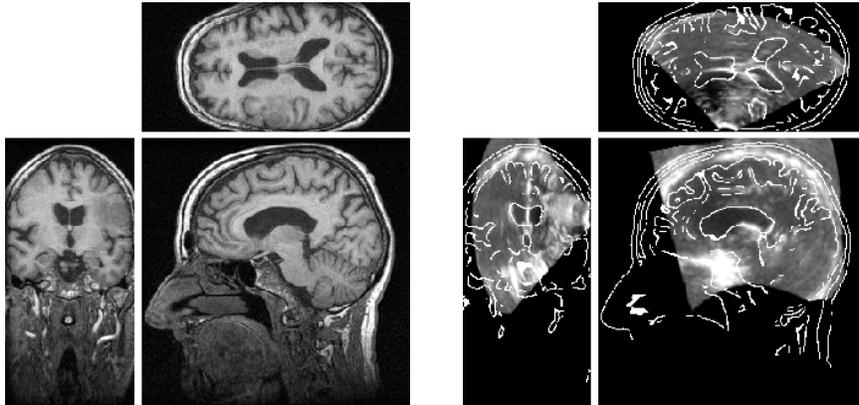


Fig. 2. Result of US/MR registration by maximization of CR. Left, three orthogonal views of a MR, T1 weighted image. Right, corresponding views of the registered US image with contours from the MR overlaid.

of success indicate that CR may have a wider attraction basin, an observation consistent with previous experiments with other modality combinations [15].

Table 1. RMS errors and percentages of failures in 3D US-MR rigid registration.

Reference Image	Similarity Measure	RMS		RMS (successes)		Failures (%)
		$\Delta\theta$ (deg)	Δt (mm)	$\Delta\theta$ (deg)	Δt (mm)	
Original MR	CR	11.49	23.33	1.11	0.42	14.0
	MI	19.07	47.14	1.27	0.64	51.0
Diffused MR	CR	12.64	26.29	0.92	0.52	12.5
	MI	17.35	27.41	1.35	0.82	28.0
Distorted MR ($\sigma = 10\%$)	CR	28.51	18.08	3.21	2.04	36.0
	MI	44.23	45.06	1.84	1.36	90.0

To study the effect of noise in the reference image, we repeated the same experimental protocol twice, using as a reference image the MR pre-segmented by anisotropic diffusion [14], and the MR corrupted with Gaussian noise. The number of failures for both measures are clearly affected by the amount of noise, as can be seen on table 1. This comes as no surprise in the case of CR, since this measure has been derived under the assumption that there is no noise in the reference image (see section 3.1). This is more surprising for MI, as no such assumption was made.

We conclude that the attraction basin of the measures could be extended by denoising the reference image as a preprocessing step. Studying effects on accuracy would have been of great interest too, but this was not possible here because the “ground truth” was not accurate enough.

5 Conclusion

We have formalized image registration as a general maximum likelihood estimation problem and shown that several existing similarity measures may be reinterpreted in this framework. This enables us to better understand the implicit assumptions we make when using a particular measure, and hopefully help the selection of an appropriate strategy given a certain problem.

Experimental results of US/MR registration confirm (if needed) that similarity measures relating to different assumptions yield different performances. The CR measure was shown to be more robust than MI with respect to the initialization of registration. As CR relies on more restrictive hypotheses than MI, this suggests the importance of constraining the relationship between the images. On the other hand, the assumptions should also be founded, and we are aware that CR relies on a model that is simpler than realistic.

Because the presented work allows to systematically derive similarity measures from explicit modeling assumptions, this is a step towards taking into account more realistic models of image acquisition and anatomy. In the future, we plan to develop this approach for the challenging problem of US/MR registration.

Acknowledgments

The images were provided by ISM, Salzburg, Austria, for the US datasets, and the Max Planck Institute for Psychiatry, AG-NMR, Munich, Germany, for the MR datasets, as part of the EC-funded ROBOSCOPE project HC 4018, a collaboration between The Fraunhofer Institute (Germany), Fokker Control System (Netherlands), Imperial College (UK), INRIA Sophia Antipolis (France), ISM-Salzburg and Kretz Technik (Austria).

Part of this work was supported by la Région PACA (France).

Many thanks also to Sébastien Ourselin and Alexandre Guimond for constant interaction and countless suggestions.

References

1. R. Bansal, L. H. Staib, Z. Chen, A. Rangarajan, J. Knisely, R. Nath, and J. S. Duncan. A Novel Approach for the Registration of 2D Portal and 3D CT Images for Treatment Setup Verification in Radiotherapy. In *Proc. MICCAI'98*, pages 1075–1086, Cambridge, MA (USA), October 1998. LNCS.
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. T. M. Buzug and J. Weese. Voxel-Based Similarity Measures for Medical Image Registration in Radiological Diagnosis and Image Guided Surgery. *Journal of Computing and Information Technology*, pages 165–179, 1998.

5. W. L. S. Costa, D. R. Haynor, R. M. Haralick, T. K. Lwellyn, and M. M. Graham. A Maximum-Likelihood Approach to PET Emission/Attenuation Image Registration. In *IEEE Nuclear Science Symposium and Med. Imag. Conference*, 1993.
6. M. E. Leventon and W. E. L. Grimson. Multi-modal Volume Registration Using Joint Intensity Distributions. In *Proc. MICCAI'98*, pages 1057–1066, Cambridge, MA (USA), October 1998. LNCS.
7. 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.
8. J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *MedIA*, 2(1):1–36, 1998.
9. M. S. Mort and M. D. Srinath. Maximum Likelihood Image Registration with Subpixel Accuracy. In *Proc. SPIE*, volume 974, pages 38–45, San Diego, 1988.
10. C. Nikou, F. Heitz, J.-P. Armspach, and I.-J. Namer. Single and multimodal subvoxel registration of dissimilar medical images using robust similarity measures. In *SPIE Conference on Medical Imaging*, volume 3338, pages 167–178, 1998.
11. A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Inc., third edition, 1991.
12. X. Pennec and J.P. Thirion. A framework for uncertainty and validation of 3D registration methods based on points and frames. *IJCV*, 25(3):203–229, 1997.
13. G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes. A Comparison of Similarity Measures for Use in 2D-3D Medical Image Registration. In *Proc. MICCAI'98*, pages 1153–1161, Cambridge, MA (USA), 1998. LNCS.
14. P. Perona and J. Malik. Scale-Space and edge detection using anisotropic diffusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.
15. A. Roche, G. Malandain, X. Pennec, and N. Ayache. The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration. In *Proc. MICCAI'98*, pages 1115–1124, Cambridge, MA (USA), October 1998. LNCS.
16. A. Roche, G. Malandain, and N. Ayache. Unifying Maximum Likelihood Approaches in Medical Image Registration. INRIA Research Report. *In press*.
17. 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.
18. C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 1(32):71–86, 1998.
19. P. Viola. *Alignment by Maximization of Mutual Information*. PhD thesis, M.I.T. Artificial Intelligence Laboratory, 1995.
20. P. Viola and W. M. Wells. Alignment by Maximization of Mutual Information. *Intern. J. of Comp. Vision*, 24(2):137–154, 1997.
21. W. M. Wells, P. Viola, H. Atsumi, and S. Nakajima. Multi-modal volume registration by maximization of mutual information. *MedIA*, 1(1):35–51, 1996.
22. J. West and al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *J. of Comp. Assist. Tomography*, 21:554–566, 1997.
23. R. P. Woods, J. C. Mazziotta, and S. R. Cherry. MRI-PET Registration with Automated Algorithm. *J. of Comp. Assist. Tomography*, 17(4):536–546, 1993.