To be published as:

A. Roche, G. Malandain, and N. Ayache.
Unifying Maximum Likelihood Approaches in Medical Image Registration.
International Journal of Imaging Systems and Technology: Special Issue on 3D Imaging, 11:71–80, 2000.

Unifying Maximum Likelihood Approaches in Medical Image Registration

Alexis Roche, Grégoire Malandain, Nicholas Ayache

INRIA, EPIDAURE Project
2004, route des Lucioles, BP 93
06902 Sophia Antipolis Cedex, France

Abstract

While intensity-based similarity measures are increasingly used for medical image registration, they often rely on implicit assumptions regarding the imaging physics. The motivation of this paper is to clarify the assumptions on which a number of popular similarity measures rely. 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 rigid registration between several image modalities to illustrate the importance of choosing an appropriate similarity measure.

1 Introduction

Registration is a problem common to many tasks in medical image analysis. It can be necessary to compare images acquired from the same patient at different times or with different sensors. One usually distinguishes registration problems according to whether the images are from the same modality (monomodal registration) or from different modalities (multimodal registration). In general, intra-patient registration consists of estimating a rigid transformation between the images, but it can also involve a non-rigid transformation in order to compensate for tissue deformations or geometrical distortions inherent to the imaging processes.

Registration is also useful for comparing images acquired from different patients, e.g. to build statistical anatomical atlases. Much effort in this area has been devoted to the geometrical modeling of anatomical variations from one subject to another. In general, inter-patient registration involves non-rigid transformations [24].

Reviews of medical image registration methods were written by van den Elsen et al. [25], Lavallée [6], and Maintz [9]. Quite recently, a comparison of algorithms based on a retrospective evaluation was published by West et al. [28] in the context of intra-patient rigid registration. Registration methods are usually classified as being either feature-based or intensity-based. Methods from the former class proceed in two sequential steps. The first is to segment homologous geometrical landmarks in the images; these can be points, lines, surfaces, or volumes. The problem then reduces to a purely geometrical task which is to evaluate the transformation that "best" matches these landmarks. Because these methods are highly dependent on the algorithms that are used in the segmentation step, they are often restricted to very specific registration problems. Likewise, when dealing with images from different modalities, finding homologous landmarks is a very challenging task due to the lack of redundancy in anatomical information.

Intensity-based techniques circumvent these difficulties because they do not deal with identifying geometrical landmarks. Their basic principle is to search, in a certain space of transformations, the one that maximizes a criterion measuring the intensity similarity of corresponding voxels. This is the class of methods that we will focus on in this paper. Over the last few years, they have been applied to a number of registration problems, including monomodal as well as multimodal and rigid as well as non-rigid registration [9]. Common to the many proposed similarity measures is the idea that, when matched, the image intensities should verify a certain relationship. The similarity measure is intended to quantify "how well" this relationship is verified given a 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 relation between the image intensities. Second, most of the similarity measures rely on imaging assumptions that are not fully explicit. We can roughly classify existing similarity measures according to four main kinds of hypotheses:

Identity relationship. In this category, the basic assumption is the conservation of intensity from one image to the other. 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], etc. 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 on the correlation coefficient [3], defined as the ratio between the images' covariance and the product of individual standard deviations:

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

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 [19] was to assume that, at the registration position, one image could be approximated in terms of the other by applying some intensity function, $I \approx \phi(J)$. Making no assumption regarding the nature of the function, we derived a natural statistical measure, the correlation ratio:

$$\eta^2(I|J) = 1 - \frac{\mathrm{Var}(I - \hat{\phi}(J))}{\mathrm{Var}(I)}, \tag{2}$$

where $\hat{\phi}(J)$ is the least squares optimal non-linear approximation of I in terms of J [12]. The correlation ratio is closely related to a very popular measure previously proposed by Woods et al [29], and generalized using robust metrics in [11].

Statistical relationship. Finally, assuming a functional relationship is sometimes too restrictive. In theses cases, it is more appropriate to use information theoretic measures, from this group mutual information [8, 26] is today probably the most popular:

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

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 [28, 2, 14, 11, 19]. 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. Some authors [10, 5] proposed that image registration could be seen as a maximum likelihood estimation problem, then Viola et al [26, 27] suggested the analogy of this approach with registration based on information theory. Notably, other teams had motivated information-theoretic measures using different arguments [8, 22].

In section 2, we propose to formulate image registration as a general maximum likelihood estimation problem, examining carefully the assumptions that are required. In section 3, deriving optimal similarity measures from specific modeling assumptions, we retrieve the correlation ratio and mutual information. Section 4 proposes to illustrate the practical differences between these two measures with results of rigid multimodal registration of 3D brain 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 the 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 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 in computer vision 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:

- *(A1)* 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.

- *(A2)* The noise is context-free. In other words, the intensity of a voxel depends only on its homologous in the scene:

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

  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. Of course, to be meaningful, the transformation T needs to be defined as a mapping from Ω_I to Ω_S, i.e. a grid point of I is supposed to match a grid point of S. In section 2.4, we will discuss how this may be achieved in practice.

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

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

We can model the relationship between J and S in the same manner. However, as we are only interested 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,

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

$\Omega_J \equiv \Omega_S$ 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:

$$P(I, J|T) = \int P(I|S,T)\, P(J|S)\, P(S)\, dS. \tag{6}$$

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 [26, 7, 1, 10, 5]. This simply states that the most likely transformation between I and J is the one that maximizes the joint probability of (I, J),

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

Unfortunately, the integral in eq (6) may be intractible unless we assume that the voxels of the scene are independently distributed, so that: $P(S) = \prod_{y_l} P(s_l)$. This appears as a minimal way to introduce prior anatomical information. Notice, however, this does not mean that the voxels are *identically* distributed, so that spatial dependences may still be incorporated into the model. Once S is a coupled field, there might not exist an analytical expression of $P(I, J|T)$. Instead of the present maximum likelihood approach, a maximum a posteriori (MAP) strategy could be employed. This alternative, using for example a Gibbs-Markov random field, would require an explicit estimation of the scene that would be computationally very expensive in the case of 3D images.

In order to simplify expression (6), we will also need the transformation T to be an injection mapping from Ω_I to Ω_J, i.e. T maps distinct points from Ω_I to distinct points from Ω_J. Let us denote the subset of matched points, $\mathcal{A} \equiv T(\Omega_I) = \{y_l \in \Omega_J,\ \exists x_k \in \Omega_I,\ T(x_k) = y_l\}$. Recall that, since T is assumed to be a mapping from Ω_I to Ω_J (assumption *A2*), the matched points $\mathcal{A}$ are assumed to lie entirely within Ω_J. The conditional probability $P(I|S,T)$ can then be evaluated in the same coordinate frame as $P(J|S)$ and $P(S)$:

$$P(I|S,T) = \prod_{x_k \in \Omega_I} P(i_k | s_k^{\downarrow}) = \prod_{y_l \in \mathcal{A}} P(i_l^{\uparrow} | s_l),$$

where we have to be cautious that $i_l^{\uparrow} \equiv I\left((T^{-1}(y_l)\right)$ represents the intensity of the unique voxel x_k such that $T(x_k) = y_l$: it is defined iff $y_l \in \mathcal{A}$. We are now in a position to rewrite the joint probability of (I, J). Starting from eq (6), we have:

$$\begin{aligned} P(I,J|T) &= \int \prod_{y_l \in \mathcal{A}} P(i_l^{\uparrow}|s_l) \prod_{y_l \in \Omega_J} P(j_l|s_l)\, P(s_l) \prod_{y_l \in \Omega_J} ds_l, \\ &= \int \prod_{y_l \in \mathcal{A}} P(i_l^{\uparrow}|s_l)\, P(j_l|s_l)\, P(s_l)\, ds_l \ \times \int \prod_{y_l \notin \mathcal{A}} P(j_l|s_l)\, P(s_l)\, ds_l, \end{aligned}$$

in which we have regrouped the points y_l according to whether they match a point in Ω_I or not. Then, applying the Fubini theorem, we can invert the integral operands and the products, so that:

$$\begin{aligned} P(I,J|T) &= \prod_{y_l \in \mathcal{A}} \int P(i_l^{\uparrow}|s_l)\, P(j_l|s_l)\, P(s_l)\, ds_l \ \times \prod_{y_l \notin \mathcal{A}} \int P(j_l|s_l)\, P(s_l)\, ds_l, \\ &= \prod_{y_l \in \mathcal{A}} P(i_l^{\uparrow}, j_l) \ \times \prod_{y_l \notin \mathcal{A}} P(j_l), \end{aligned}$$

Noting that $P(i_l^{\uparrow}, j_l) = P(i_l^{\uparrow}|j_l) P(j_l)$, we finally get:

$$\begin{aligned} P(I,J|T) &= \prod_{y_l \in \Omega_J} P(j_l) \ \times \prod_{y_l \in \mathcal{A}} P(i_l^{\uparrow}|j_l), \\ &= \underbrace{\prod_{y_l \in \Omega_J} P(j_l)}_{P(J)} \ \times \underbrace{\prod_{x_k \in \Omega_I} P(i_k|j_k^{\downarrow})}_{P(I|J,T)}, \end{aligned} \tag{7}$$

where the last step is only a rewriting of $P(I|J,T)$ in the coordinate frame of Ω_I. In eq (7), the left term of the product is the marginal probability of J and is independent of the transformation T. Only the right term, the conditional probability of I knowing J and T, will play a role in the maximization with respect to T. We should note that $P(I|J,T)$ is of the same factored form as $P(I|S,T)$:

$$P(I|J,T) = \prod_{x_k \in \Omega_I} P(i_k|j_k^{\downarrow}), \qquad \text{with} \quad P(i_k|j_k^{\downarrow}) = \frac{\int P(i_k|s_k^{\downarrow}) P(j_k^{\downarrow}|s_k^{\downarrow}) P(s_k^{\downarrow})\, ds_k^{\downarrow}}{\int P(j_k^{\downarrow}|s_k^{\downarrow}) P(s_k^{\downarrow})\, ds_k^{\downarrow}}. \tag{8}$$

It turns out that the statistical relation between I and J is of the same form as that between I and S. This result is obtained under the assumption that the scene voxels are mutually independent. Therefore, the image J can be considered as a scene for I in the sense that $P(I|J,T)$ verifies the assumptions *(A1)* and *(A2)* stated above. However, it is important to realize that the conditional densities $P(i_k|s_k^{\downarrow})$ and $P(i_k|j_k^{\downarrow})$ may have very different expressions. Since the $P(i_k|s_k^{\downarrow})$ are intended to model acquisition noise, they may generally be chosen as single-mode densities (e.g. Gaussian densities). On the other hand, the $P(i_k|j_k^{\downarrow})$ may have much more complicated forms since they incorporate the noise models corresponding to each image as well as the prior probability on S.

2.2 Estimating the probability densities

Up until now, we have worked under the assumption that all the probability densities involved in our model were perfectly known. We address here the question of how to estimate them.

Since these densities stand for anatomical and image acquisition models, they should vary significantly from one dataset to another, due not only to inter-patient anatomical variability but also to changes in acquisition settings. For example, a tissue such as white matter may have very different ranges of response in two differently acquired brain MR scans. Very often in practice, we cannot use information from previous datasets to model the relationship between the images we want to register.

Therefore, the conditional densities have to be estimated on-line, in the same manner as we have to estimate the transformation T. We notice that estimating the densities would be easier if the images were aligned; on the other hand, the registration process needs density estimates to work. Thus, the trade-off is to alternate from registration to density estimation: given a current estimate of T, estimate the densities, and given current estimates of the densities, update the transformation by maximizing $P(I, J|T)$, hoping for convergence. Notice that since only the conditional densities $P(i_k|j_k^{\downarrow})$ play a role in the maximization with respect to T (see eq (7)), we could forget about estimating the marginal densities $P(j_l)$.

Within analogous formulations of image registration, several methods have been proposed for the density estimation step. Viola [26] uses the method of Parzen windows to estimate the joint densities $P(i_k, j_k^{\downarrow})$, from which it would be straightforward to obtain the conditional densities $P(i_k|j_k^{\downarrow})$. Other authors [8, 22] construct the 2D histogram of the images, which may be seen as a particular case of Parzen windowing. In these techniques, the densities are not constrained by any model of anatomy or image acquisition; this is perhaps both a strength and a weakness. Moreover, we notice that they provide density estimates that are independent of the spatial positions, in the sense that two voxels x_k and $x_{k'}$ are assumed to be identically distributed provided that they have the same intensities in both images, i.e. $i_k = i_{k'}$ and $j_k^{\downarrow} = j_{k'}^{\downarrow}$. A way to incorporate explicit spatial dependence was recently suggested by Bansal et al. [1], who applied the maximum entropy principle to get non-stationary prior probabilities $P(s_l)$ for the tissue classes, implying non-stationary densities $P(i_k|j_k^{\downarrow})$.

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 eq (8). 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, $P(i_k|s_k^{\downarrow})$, $P(j_l|s_l)$, and $P(s_l)$ are themselves chosen as parametric densities. Then, from eq (8), the form of $P(I|J,T,\theta)$ can be known. We will show in section 3 that under some specific modeling assumptions, maximum likelihood density estimates can be computed explicitly.

2.3 Registration energy

By substituting the estimated densities in eq (8), our registration criterion becomes the maximum of $P(I|J,T,\theta)$ under θ at fixed T. Actually this is only a special way to maximize $P(I|J,T,\theta)$ with respect to (T,θ). There is no formal difference between the parameters T and θ except that the latter models the relation between the image intensities whereas, in the context of registration, T is the parameter we are really interested in. For practical optimization, it is often more convenient to consider the negative log-likelihood (to be minimized); thus, we will define the energy of a transformation T as:

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

2.4 Practical issues

In section 2.1, we have derived the likelihood registration criterion under the assumption that the transformation T is searched for among mappings from the floating image grid, Ω_I, to the reference image grid, Ω_J. In other words, a grid point of I was supposed to always match a grid point of J. The spatial resolution of the transformation is thus intrinsically limited by the resolution of the reference grid, and clearly this assumption cannot deal with sub-voxel accurate registration.

In practice, we generally want to take into account continuous spatial transformations, not only for a question of accuracy but also because the motion to be estimated is continuous in nature. Thus, we would like the resolution of the reference grid to be as small as the computer working precision. This is achieved in practice by oversampling the image J using fast interpolation techniques such as trilinear, or partial volume interpolation [8, 21, 20, 16]. Notice that for evaluating the registration criterion eq (8), we do not actually have to interpolate every point in space, but only, for a given transformation, the points that are put into correspondence with voxels of I, i.e. the subset $\mathcal{A}$ defined in section 2.1.

However, interpolation 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 interpolate 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 [23, 26, 19].

Here, to keep consistent with the maximum likelihood framework, we definitely cannot ignore them. Doing so, we would no longer maximize the image likelihood, $P(I|J,T)$, but the likelihood of a part of I which is variable according to the considered transformation. There is always a risk to isolate small image regions that seem very likely to be aligned (typically, in the background), and then the algorithm might converge to an aberrant solution.

To tackle this problem, a natural approach is simply to extend the reference domain by assigning the external points to an arbitrary intensity class $J = j^*$ and defining a specific conditional density corresponding to this class. Although this sounds like a computational artifice, this enables us to take into account every voxel of I at each iteration of the registration process. Each voxel with intensity i_k falling outside the reference domain will have a non-zero contribution $-\log P(i_k|j^*)$ to the registration energy. Thus, we can expect the registration energy to vary little with image overlap as is the effect achieved with classical normalization.

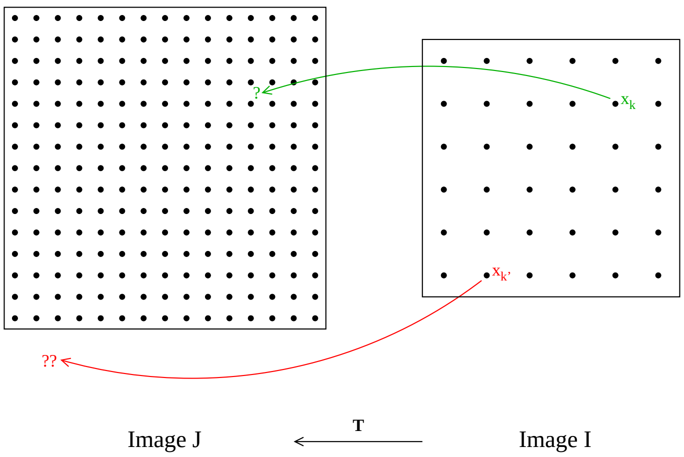

Figure 1: Effects of applying a continuous spatial transformation.

3 From modeling assumptions to similarity measures

The purpose of this section is to demonstrate the link between the general maximum likelihood approach that has been presented, and two popular registration criteria: the correlation ratio and mutual information. We will show these measures can be derived from the above formalism using specific modeling assumptions.

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:

$$I(x_k) = f\left(S\left(T(x_k)\right)\right) + \epsilon_k,$$

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 $P(i_k|j_k^{\downarrow})$ have 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 $\theta = (f_0, f_1, \ldots, \sigma)$ 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 $N = \mathrm{Card}\Omega_I$, $\Omega_I^j = \{x_k \in \Omega_I,\, j_k^{\downarrow} = j\}$, and $N_j = \mathrm{Card}\Omega_I^j$, we have:

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

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

$$-\frac{\partial \log P}{\partial f_j} = -\frac{1}{\sigma^2}\sum_{\Omega_I^j}(i_k - f_j) \quad \Longrightarrow \quad \hat{f}_j = \frac{1}{N_j}\sum_{\Omega_I^j} i_k,$$

$$-\frac{\partial \log P}{\partial \sigma} = \frac{N}{\sigma} - \frac{1}{\sigma^3} \sum_j \sum_{\Omega_I^j} (i_k - f_j)^2 \quad \Longrightarrow \quad \hat{\sigma}^2 = \sum_j \frac{N_j}{N} \hat{\sigma}_j^2,$$

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:

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

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 squares 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) [19, 18], and we see that the registration energy $U(T)$ is related to the correlation ratio $\eta^2(I|J^{\downarrow})$ by:

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

In the original version of the correlation ratio [19], the quantities N and $\operatorname{Var}(I)$ 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 2.4, N and $\operatorname{Var}(I)$ 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. Still there are reasons to believe that differences may be observed in cases where the image overlap is susceptible of rapidly changing with pose. In our experiments, this was not the case. However this question needs to be addressed with further testing.

Another remark is that in practice we may compute the correlation ratio using a reference image that is not a valid segmentation: then there are as many "tissue classes" as image iso-intensity sets, typically 256 for a one-byte encoded image. For 2-byte or 4-byte images, this approach may be meaningless and we should impose constraints to the intensity function f. In the appendix, we generalize the notion of correlation ratio to polynomial imaging functions. 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 eq (1):

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

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 $P(i_k|j_k^{\downarrow})$ from eq (8). This case has been investigated by Leventon and Grimson [7]. It turns out that there is probably nothing much faster than an 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 $P(i_k|j_k^{\downarrow})$ are totally unspecified, and we will only assume that they are stationary, i.e. $P(i_k = i | j_k^{\downarrow} = j)$ 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

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

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

$$\Omega_{i,j} = \{x_k \in \Omega_I,\ I(x_k) = i,\ J(T(x_k)) = j\}, \qquad N_{i,j} = \mathrm{Card}\Omega_{i,j},$$

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 exists constants $\lambda_0, \lambda_1, \ldots$, 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 $J^{\downarrow}$. From eq (3), we see that $U(T)$ is nothing but a decreasing function of mutual information:

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

where $H(I)$ is the entropy of image I and is constant in the implementation proposed in section 2.4. The same remark as made in section 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 assumptions *(A1)* and *(A2)* stated 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 often a rough approximation. 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 of rigid registration

This section illustrates the practical differences between the CR and MI measures in the context of 3D rigid registration of brain images acquired from different modalities. Following the ideas proposed by Maes et al [8], we implemented Powell's method [17] to optimize the measures with respect to the transformation parameters. Partial volume interpolation (PV) was used in all the experiments.

4.1 Vanderbilt database

The registration algorithm was tested using image datasets from 10 patients. For each patient, the following images were available:

- MR, T1 weighted ($256 \times 256 \times 20/26$ voxels of $1.25 \times 1.25 \times 4\, mm^3$)
- MR, T2 weighted ($256 \times 256 \times 20/26$ voxels of $1.25 \times 1.25 \times 4\, mm^3$)
- CT ($512 \times 512 \times 28/34$ voxels of $0.65 \times 0.65 \times 4\, mm^3$)
- PET ($128 \times 128 \times 15$ voxels of $2.59 \times 2.59 \times 8\, mm^3$)

The gold standard transformations between the modalities were known thanks to a prospective, marker-based registration method [28]. We performed three kinds of registrations: T1 to T2, CT to T1, and PET to T1. In all the experiments, the transformation was initially set either to the identity or to the gold standard: this was done to test if the algorithm was sensitive to initialization. However, since the results were almost the same for both types of initialization, we present only those obtained when starting from the identity.

After each registration, a typical error ϵ was computed by taking the average registration error of the eight vertices of a bounding box corresponding approximately to the head volume. Thus, ϵ represents the error one can expect in the region of interest. We also computed the *intrinsic* rotation and translation errors [13]. Let R_* and t_* be the "ground truth" rotation matrix and translation vector. For a rotation matrix R and a translation vector t found by the registration algorithm, the intrinsic rotation error $\Delta\,\theta$ is the norm of the rotation vector corresponding to the residual rotation matrix $R_*^t R$, and the translation error is the Euclidean distance between t_* and t, that is $\Delta\, t = \|t - t_*\|$. Intrinsic errors were chosen because these are objective measures that are independent of any region of interest.

Table 1 shows RMS of ϵ, $\Delta\,\theta$, and $\Delta\, t$ over the ten patients, for each modality combination. These have to be compared to the image resolution, which is quite poor (4 mm in the z-axis for the MR and CT datasets, and 8 mm for the PET). The reader may notice that errors reported here are higher than target registration errors (TRE) reported by other groups in the retrospective registration evaluation project [28]. Our guess is that this is simply because our respective methods for computing errors are different.

In T1/T2 registration, CR and MI give good, and similar, results. For the other combinations, more significant differences are observed. MI does a much better job for matching CT to T1. This might be due to the fact that functional dependence is a crude hypothesis in the CT/MR case [27]. On the other hand, CR tends to give slightly better results for PET/T1 registration.

4.2 US/MR

A very challenging registration problem consists of aligning an intra-operative ultrasound (US) image with a pre-operative image such as an MR scan. We tested the registration algorithm with:

Table 1: Rigid registration errors obtained over ten intra-patient experiments.

Experiment	Measure	RMS		
		$\Delta\,\theta$ (deg)	$\Delta\,t$ (mm)	ϵ (mm)
T1/T2	CR	0.31	2.28	1.91
	MI	0.58	2.19	2.16
CT/T1	CR	2.91	11.27	6.75
	MI	0.77	3.98	3.31
PET/T1	CR	1.53	5.49	5.16
	MI	1.42	7.55	7.84

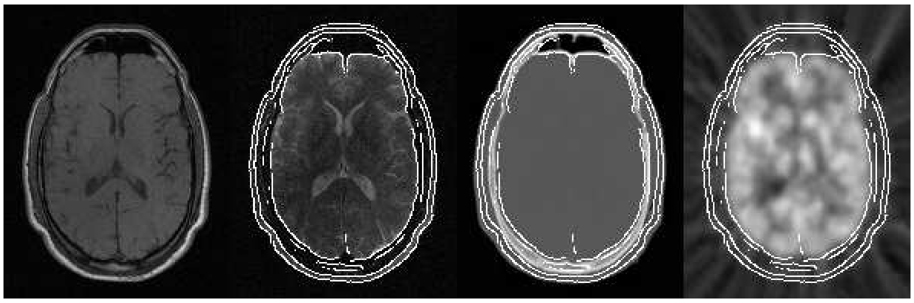

Figure 2: Multimodal registration by maximization of CR. Images from left to right: MR-T1, MR-T2, CT, and PET. The images are resampled in the same reference frame after registration. Contours extracted from the MR-T1 are superimposed on each other modality in order to better visualize the quality of registration.

- an MR, T1 weighted scan ($256 \times 256 \times 124$ voxels of $0.9 \times 0.9 \times 1.1$ mm),
- an intra-operative 3D ultrasound image ($180 \times 136 \times 188$ voxels of 0.95^3 mm)

Since the US image was acquired before opening the duramater, we neglect the brain shift phenomenon. Thus, there is essentially a rigid displacement to find. The correct registration position was found manually using an interactive matching tool, and validated by a clinician. The estimated accuracy was 2 degrees in rotation and 2 mm in translation. We took this first 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. In order to avoid interpolation artifacts due to resampling, the algorithm did *not* take as an input the US resampled by the "ground truth" transformation, but always the original US itself. Another advantage of doing so is that the "ground truth" corresponds to a partial overlap between the MR volume and the original US volume. Therefore, there is no reason to expect the registration results to be biased towards the "ground truth" due to the problems associated with changing the image overlap (see section 2.4).

We observe two kinds of results: 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

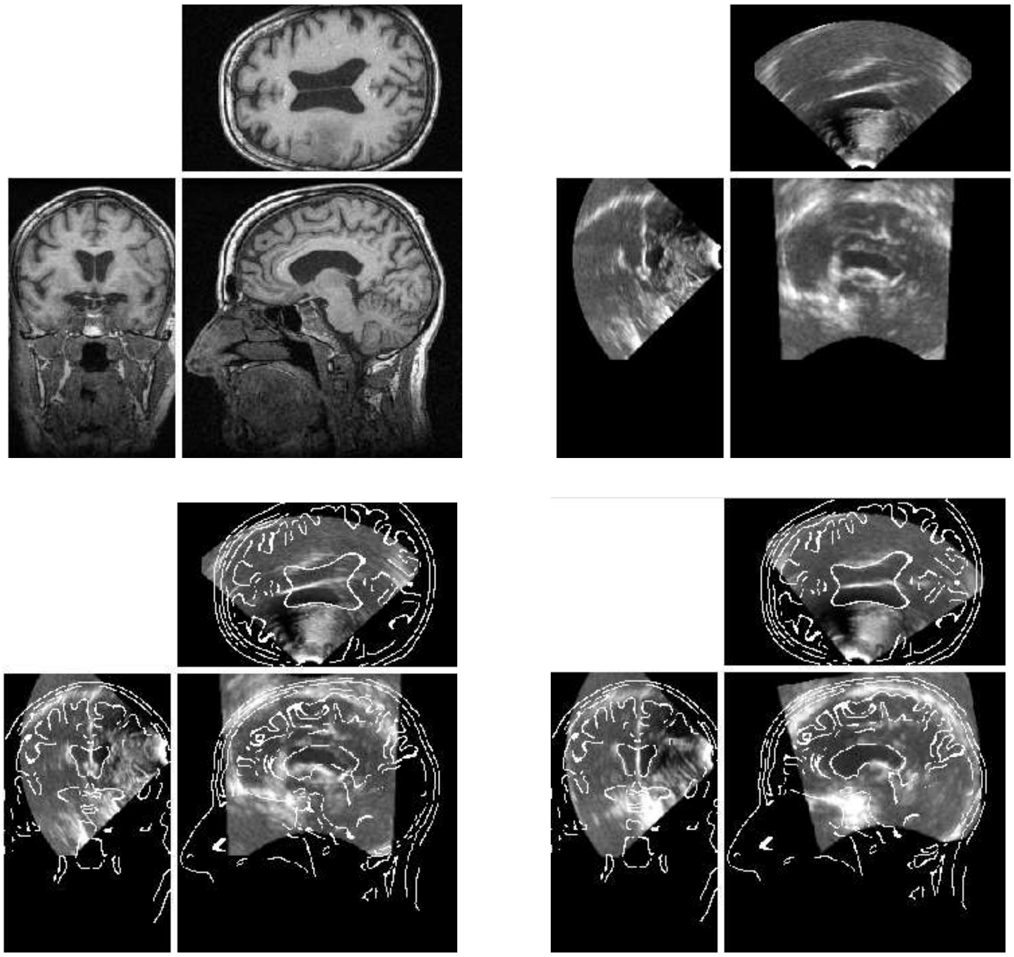

Figure 3: Top left, three orthogonal views of the MR image. Top right, corresponding views of the US image in a random initial position. Bottom left, display of the initial US with contours from the MR superimposed. Bottom right, same display with the registered US.

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 2). 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 of success indicate that CR may have a wider attraction basin, an observation consistent with previous experiments with other modality combinations [19].

To study the effect of noise in the data, we repeated the same experimental protocol twice, using as a reference image the MR pre-segmented by anisotropic diffusion [15], 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 2. 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 MR image in a preprocessing step. Studying the effects on accuracy would have been of great interest

Table 2: RMS errors and percentages of failures in 3D US-MR rigid registration.

Reference Image	Similarity Measure	RMS		RMS (successes)		Failures (%)
		$\Delta\,\theta$ (deg)	$\Delta\,t$ (mm)	$\Delta\,\theta$ (deg)	$\Delta\,t$ (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
Filtered MR (anisotropic diffusion)	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

too, but this was not possible here because the "ground truth" could not been considered accurate enough.

We believe that the registration algorithm would greatly benefit from reducing noise in the US. Unfortunately, applying classical anisotropic diffusion to the US did not provide convincing results. It tended to blur the image while not removing speckle artifacts. Specific filtering tools need to be developed for US images, and this is still an open research track.

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, helps the selection of an appropriate strategy given a certain problem.

Experimental results of rigid registration confirm (if needed) that similarity measures relating to different assumptions yield different performances. The CR measure was shown to be more efficient than MI in the case of PET/MR and US/MR 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 although simpler may not be realistic.

Because the presented work allows to derive systematically the 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 authors wish to thank Sbastien Ourselin, Alexandre Guimond, and Sylvain Prima, for constant interaction and countless suggestions. Also many thanks to Janet Bertot for the proofreading of this article.

The images and the standard transformations used in section 4.1 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 (USA).

The images used in section 4.2 were provided by ISM-Austria, 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 (United Kingdom), INRIA Sophia Antipolis (France), ISM-Salzburg and Kretz Technik (Austria).

Part of this work was supported by la Rgion Provence-Alpes-Côtes d'Azur (France).

Appendix

A Generalization of the correlation ratio

For the problem considered in section 3.1, we could define the unknown imaging function f to be polynomial with degree d:

$$f(j) = \sum_{p=0}^{d} \alpha_p \, j^p.$$

Then, we aim at minimizing the log-likelihood,

$$\begin{aligned} -\log P(I|J,T,\theta) &= N \log \sqrt{2\pi}\,\sigma + \frac{1}{2} \sum_{x_k \in \Omega_I} \frac{(i_k - f(j_k^{\downarrow}))^2}{\sigma^2} \\ &= N \log \sqrt{2\pi}\,\sigma + \frac{1}{2\,\sigma^2} \sum_{x_k \in \Omega_I} \left[i_k - \sum_{p=0}^{d} \alpha_p \, {j_k^{\downarrow}}^p \right]^2, \end{aligned} \tag{11}$$

with respect to $\theta = (\alpha_0, \alpha_1, \ldots, \alpha_p, \sigma)$. By differentiating eq (11), we get:

$$\begin{aligned} -\frac{\partial \log P}{\partial \alpha_q} &= -\frac{1}{\sigma^2} \sum_{x_k \in \Omega_I^j} {j_k^{\downarrow}}^q \left[i_k - \sum_{p=0}^{d} \alpha_p \, {j_k^{\downarrow}}^p \right], \\ &= -\frac{1}{\sigma^2} \left[\sum_{x_k \in \Omega_I^j} {j_k^{\downarrow}}^q \, i_k - \sum_{x_k \in \Omega_I^j} \sum_{p=0}^{d} \alpha_p \, {j_k^{\downarrow}}^{p+q} \right], \\ &= -\frac{1}{\sigma^2} \left[\sum_{x_k \in \Omega_I^j} {j_k^{\downarrow}}^q \, i_k - \sum_{p=0}^{d} \alpha_p \left(\sum_{x_k \in \Omega_I^j} {j_k^{\downarrow}}^{p+q} \right) \right]. \end{aligned}$$

The polynomial coefficients are then seen to be solutions of the linear system $A\,X = B$, with

$$X = \begin{pmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \ldots & \alpha_d \end{pmatrix}^T$$

$$A = \begin{pmatrix} N & \sum j_k^{\downarrow} & \sum {j_k^{\downarrow}}^2 & \ldots & \sum {j_k^{\downarrow}}^d \\ \sum j_k^{\downarrow} & \sum {j_k^{\downarrow}}^2 & \sum {j_k^{\downarrow}}^3 & \ldots & \sum {j_k^{\downarrow}}^{d+1} \\ \sum {j_k^{\downarrow}}^2 & \sum {j_k^{\downarrow}}^3 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \sum {j_k^{\downarrow}}^{2d-1} \\ \sum {j_k^{\downarrow}}^d & \sum {j_k^{\downarrow}}^{d+1} & \ldots & \ldots & \sum {j_k^{\downarrow}}^{2d} \end{pmatrix}$$

$$B = \begin{pmatrix} \sum i_k & \sum j_k^{\downarrow} i_k & \sum {j_k^{\downarrow}}^2 i_k & \ldots & \sum {j_k^{\downarrow}}^d i_k \end{pmatrix}^T$$

In practice, we invert the $(d+1)\times(d+1)$ matrix A by the method of singular value decomposition (SVD). This avoids numerical explosion when A comes close to singularity. To solve for the standard deviation σ, we differentiate the log-likelihood:

$$-\frac{\partial \log P}{\partial \sigma} = \frac{N}{\sigma} - \frac{1}{\sigma^3} \sum_{x_k \in \Omega_I} (i_k - f(j_k^{\downarrow}))^2 \quad \Longrightarrow \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{x_k \in \Omega_I} (i_k - \hat{f}(j_k^{\downarrow}))^2.$$

Thus, the optimal σ is equal to the standard deviation of the difference image between I and the optimally corrected image $J^{\downarrow}$. This yields the registration energy,

$$U(T) = \frac{N}{2} \log \left[2\pi\, e \operatorname{Var}(I - \hat{f}(J^{\downarrow})) \right],$$

a measure that directly generalizes the correlation ratio.

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, Massachusetts (USA), October 1998.

[2] M. Bro-Nielsen. Rigid Registration of CT, MR and Cryosection Images Using a GLCM Framework. In *Proc. 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. Lwellen, 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] S. Lavallee. Registration for computer integrated surgery: methodology, state of the art. In R. Taylor, S. Lavallee, G. Burdea, and R. Moesges, editors, *Computer Integrated Surgery*, pages 77–97. MIT Press, 1995.

[7] M. E. Leventon and W. E. L. Grimson. Multi-modal Volume Registration Using Joint Intensity Distributions. In *Proc. MICCAI'98*, pages 1057–1066, Cambridge, Massachusetts (USA), October 1998.

[8] 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.

[9] J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *MedIA*, 2(1):1–36, 1998.

[10] M. S. Mort and M. D. Srinath. Maximum Likelihood Image Registration with Subpixel Accuracy. In *Proc. SPIE*, volume 974, pages 38–45, San Diego, CA, USA, August 1988.

[11] 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, February 1998.

[12] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Inc., third edition, 1991.

[13] X. Pennec and J.P. Thirion. A framework for uncertainty and validation of 3D registration methods based on points and frames. *Int. Journal of Computer Vision*, 25(3):203–229, 1997.

[14] 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, Massachusetts (USA), October 1998.

[15] P. Perona and J. Malik. Scale-Space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990.

[16] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual information matching and interpolation artefacts. In K. M. Hanson, editor, *Proc. SPIE*, volume 3661, Bellingham, WA, 1999. SPIE Press.

[17] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C*. Cambrige Univ. Press, 2nd edition, 1992.

[18] A. Roche, G. Malandain, N. Ayache, and X. Pennec. Multimodal Image Registration by Maximization of the Correlation Ratio. Technical Report 3378, INRIA, August 1998.

[19] 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*, volume 1496 of *Lecture Notes in Computer Science*, pages 1115–1124, Cambridge, Massachusetts (USA), October 1998.

[20] D. Sarrut and F. Feschet. The Partial Intensity Difference interpolation. In H. R. Arabnia, editor, *International Conference on Imaging Science, Systems and Technologiy*, Las Vegas, USA, July 1999. CSREA Press.

[21] D. Sarrut and S. Miguet. Fast 3D Images Transformations for Registration Procedures. In *10th International Conference on Image Analysis and Processing*, pages 446–452, Venice, Italy, September 1999. IEEE Computer Society.

[22] 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.

[23] 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.

[24] Arthur W. Toga. *Brain Warping*. Academic Press, 1999.

[25] 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.

[26] P. Viola and W. M. Wells. Alignment by Maximization of Mutual Information. *Intern. J. of Comp. Vision*, 24(2):137–154, 1997.

[27] 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.

[28] J. West and al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Comp. Assist. Tomography*, 21:554–566, 1997.

[29] 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.