



**HAL**  
open science

# Unifying Maximum Likelihood Approaches in Medical Image Registration

Alexis Roche, Grégoire Malandain, Nicholas Ayache

► **To cite this version:**

Alexis Roche, Grégoire Malandain, Nicholas Ayache. Unifying Maximum Likelihood Approaches in Medical Image Registration. RR-3741, INRIA. 1999. inria-00072923

**HAL Id: inria-00072923**

**<https://inria.hal.science/inria-00072923v1>**

Submitted on 24 May 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

***Unifying Maximum Likelihood Approaches  
in Medical Image Registration***

Alexis Roche — Grégoire Malandain — Nicholas Ayache

**N° 3741**

Juillet 1999

THÈME 3



***rapport  
de recherche***



# Unifying Maximum Likelihood Approaches in Medical Image Registration

Alexis Roche\* , Grégoire Malandain , Nicholas Ayache

Thème 3 — Interaction homme-machine,  
images, données, connaissances  
Projet Epidaure

Rapport de recherche n° 3741 — Juillet 1999 — 21 pages

**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 rigid registration between several modalities of images to illustrate the importance of choosing an appropriate similarity measure.

**Key-words:** registration, medical images, likelihood, similarity measure

\* Email: Alexis.Roche@sophia.inria.fr

# Unification des approches par maximum de vraisemblance en recalage d'images médicales

**Résumé :** Les mesures de similarité en intensité, bien qu'elles soient de plus en plus utilisées pour le recalage d'images médicales, relèvent souvent d'hypothèses implicites sur la physique des processus d'imagerie. Notre objectif est ici de déterminer à quelles hypothèses correspondent un certain nombre de mesures courantes, afin de mieux comprendre leurs conditions d'utilisation et finalement aider le choix de la mesure la plus appropriée à un problème donné. Après avoir formulé le recalage d'images à partir de modèles génériques d'acquisition, nous montrons que la recherche d'une mesure de similarité optimale peut être traitée dans le cadre d'un problème d'estimation par maximum de vraisemblance. Nous calculons alors les mesures correspondant à différents modèles et retrouvons quelques mesures bien connues (le coefficient de corrélation, le rapport de corrélation, l'information mutuelle). Nous présentons pour finir des résultats de recalage rigide inter modalités afin d'illustrer l'importance du choix d'une mesure appropriée.

**Mots-clés :** recalage, images médicales, vraisemblance, mesure de similarité

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Formulation</b>	<b>6</b>
2.1	Maximum likelihood registration . . . . .	6
2.2	Estimating the conditional densities . . . . .	8
2.3	Likelihood and similarity measures . . . . .	9
2.4	Practical issues . . . . .	9
<b>3</b>	<b>From modeling assumptions to similarity measures</b>	<b>10</b>
3.1	Gaussian channel . . . . .	10
3.2	Unspecified channel . . . . .	12
3.3	Comparison of measures . . . . .	12
<b>4</b>	<b>Experiments of rigid registration</b>	<b>13</b>
4.1	Vanderbilt database . . . . .	13
4.2	Ultrasound/MR registration . . . . .	14
<b>5</b>	<b>Conclusion</b>	<b>16</b>
<b>A</b>	<b>Polynomial least-squares fit</b>	<b>17</b>
<b>B</b>	<b>A “symmetric” generalization of the correlation ratio</b>	<b>18</b>

# 1 Introduction

Registration is a problem common to many tasks of medical image analysis. It can be necessary to compare images acquired from the same patient at different times or with different sensors. The difficulty of the problem is different according to whether one is interested in images 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 deformations of the tissues or geometrical distortions brought by the imaging process.

Registration is also useful for comparing images acquired from different patients, e.g. to build statistical anatomical atlases. Non-rigid registration is then always used. Reviews of medical image registration methods were written by van den Elsen et al [22], Lavallée [7], and Maintz [10]. Quite recently, a comparison of algorithms based on a retrospective evaluation was published by West et al [25].

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 extract homologous geometrical landmarks in the images; these can be points, lines, surfaces or volumes. Then the problem consists of computing the transformation that “best” matches these landmarks. Depending on the nature of the images, the first step may be very challenging, especially in the case of multimodal images.

Intensity-based techniques circumvent this difficulty, 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 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 [10]. 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 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 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], 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 intensities of 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 [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:

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

where  $\hat{\phi}(J)$  is the least square optimal non-linear approximation of  $I$  in terms of  $J$  [13]. The correlation ratio is closely related to a very popular measure previously proposed by Woods et al [26], and generalized using robust metrics in [12].

**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 [9, 23] 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 [25, 2, 15, 12, 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. After some authors [11, 5] proposed that image registration could be seen as a maximum likelihood estimation problem, Viola et al [23, 24] suggested the analogy of this approach with registration using information theory. Remarkably, other teams have motivated information-theoretic measures by other arguments [9, 20].

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 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  $\Omega_I$  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  $\Omega_I$ .

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

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

where  $T$  is the spatial transformation that relates the coordinate frames of  $\Omega_I$  and  $\Omega_S$ , 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, for having a sense, the transformation  $T$  needs to be defined as a *mapping* from  $\Omega_I$  to  $\Omega_S$ . In section 2.4, we 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:

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

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 [23, 8, 1, 11, 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.

In order to simplify expression (6), we will also need the transformation  $T$  to be an *injection mapping* from  $\Omega_I$  to  $\Omega_J$ , i.e.  $T$  maps distinct voxels from  $\Omega_I$  to distinct voxels from  $\Omega_J$ . Let us denote the subset of matched points,  $\mathcal{A} \equiv T(\Omega_I) = \{y_l \in \Omega_J, \exists x_k \in \Omega_I, T(x_k) = y_l\}$ . Then, the conditional probability of image  $I$  can 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  $i_l^\uparrow \equiv I((T^{-1}(y_l)))$  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 now are in position to rewrite the joint probability of  $(I, J)$ :

$$\begin{aligned} P(I, J|T) &= \int \prod_{y_l \in \mathcal{A}} P(i_l^\uparrow | s_l) \prod_{y_l \in \Omega_J} P(j_l | s_l) P(s_l) \prod_{y_l \in \Omega_J} ds_l, \\ &= \prod_{y_l \in \mathcal{A}} \int P(i_l^\uparrow | s_l) P(j_l | s_l) P(s_l) ds_l \times \prod_{y_l \notin \mathcal{A}} \int P(j_l | s_l) P(s_l) ds_l \\ &= \prod_{y_l \in \mathcal{A}} P(j_l) \int P(i_l^\uparrow | s_l) P(s_l | j_l) ds_l \times \prod_{y_l \notin \mathcal{A}} P(j_l) \\ &= \underbrace{\prod_{y_l \in \Omega_J} P(j_l)}_{P(J)} \times \underbrace{\prod_{y_l \in \mathcal{A}} \int P(i_l^\uparrow | s_l) P(s_l | j_l) ds_l}_{P(I|J, T)}. \end{aligned} \quad (7)$$

In (7), the left term of the product turns out to be the marginal probability  $P(J)$ , and is independent of the transformation  $T$ . We should remark that the right term,  $P(I|J, T)$ , is of the same factored form as  $P(I|S, T)$ ,

the conditional probability of  $I$  in terms of the *scene*:

$$P(I|J, T) = \prod_{y_l \in \mathcal{A}} P(i_l^\uparrow | j_l), \quad \text{with} \quad P(i_l^\uparrow | j_l) = \int P(i_l^\uparrow | s_l) P(s_l | j_l) ds_l,$$

or, equivalently, by rewriting this expression in the coordinate frame of  $\Omega_I$ :

$$P(I|J, T) = \prod_{x_k \in \Omega_I} P(i_k | j_k^\downarrow), \quad \text{with} \quad \begin{cases} j_k^\downarrow \equiv J(T(x_k)) \\ P(i_k | j_k^\downarrow) = \int P(i_k | s_k^\downarrow = s) P(s_k^\downarrow = s | j_k^\downarrow) ds \end{cases} \quad (8)$$

## 2.2 Estimating the conditional densities

Needless to say, the conditional densities  $P(i_k | j_k^\downarrow)$  are important features of the registration criterion since they depend on how we instantiate our 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 conditional densities  $P(i_k | j_k^\downarrow)$ , and given current estimates of the densities, update the transformation by maximizing  $P(I|J, T)$  (8), hoping for convergence.

Within analogous formulations of image registration, several methods have been proposed for the density estimation step. Viola [23] uses the method of Parzen windows to estimate the joint densities  $P(i_k, j_k^\downarrow)$ , from which it is straightforward to obtain the conditional densities  $P(i_k | j_k^\downarrow)$ . Other authors [9, 20] 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 from 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.

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 (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  $\theta$  (regardless, for the moment, of what  $\theta$  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.

### 2.3 Likelihood and similarity measures

The maximum likelihood search consists not only in finding the transformation  $T$ , but also the auxiliary parameter vector,  $\theta$ . The link with similarity measures arises from the remark that maximizing the likelihood criterion with respect to  $(T, \theta)$  can be performed by enforcing  $\theta$  to travel through the maxima of the criterion at a fixed transformation  $T$ . This idea is summarized in the mathematical equivalence :

$$\max_{(T, \theta)} P(I|J, T, \theta) = \max_T \left[ \max_{\theta} P(I|J, T, \theta) \right].$$

Then the parameter  $\theta$  “disappears” from the likelihood function, and only the dependence on  $T$  remains. For practical optimization, it is often more convenient to consider the negative log-likelihood (to be minimized); thus, we will define the energy of the transformation  $T$  as:

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

This energy  $U(T)$  is an optimal (in the sense of likelihood) registration measure that depends only on the search space for the conditional densities  $P_{\theta}(i_k | j_k^{\downarrow})$ .

### 2.4 Practical issues

In section 2.1, we have derived the likelihood registration criterion under the assumption that the transformation  $T$  is searched among mappings from the floating image grid,  $\Omega_I$ , to the reference image grid,  $\Omega_J$ . 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 practice, we generally want to take into account continuous spatial transformations, not only for a question of accuracy but also because the motion between the images 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 [9]. Notice that for evaluating the registration criterion (8), we actually do not 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 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

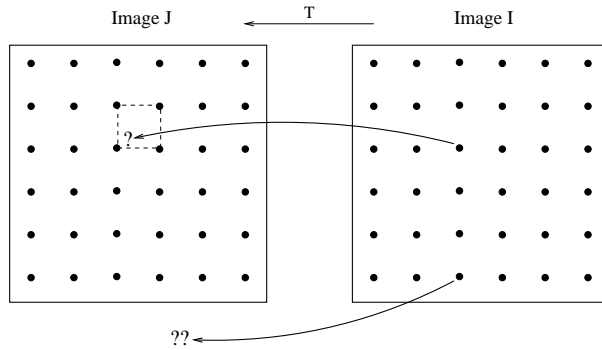


Figure 1: *Effects of applying a continuous spatial transformation. The point marked with '?' can be interpolated, unlike the point marked with '??'. Instead, this latter point may be assigned an arbitrary intensity  $j^*$ .*

by the registration criterion, which necessitates some heuristic normalization to avoid nasty effects such as disconnecting the images [21, 23, 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 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

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 that 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:

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

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  $P(i_k | j_k^\perp)$  are of the Gaussian form:

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

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  $\theta$ , we group the voxels  $x_k$  that match the same class. Letting  $N =$

Card  $\Omega_I$ ,  $\Omega_I^j = \{x_k \in \Omega_I, j_k^\perp = j\}$ , and  $N_j = \text{Card } \Omega_I^j$ , we have:

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

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  $\Omega_I^j$ . The registration energy  $U(T)$  is then obtained by substituting the optimal  $\theta$  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) [19, 18], 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 [19], 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 2.4,  $N$  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 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 loose its sense and we should impose constraints to the intensity function  $f$ . In the appendix A, 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 (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 [8]: it turns out that there is probably nothing much faster than a EM algorithm to provide maximum likelihood estimates of the density parameters (see the appendix B).

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 normalized image 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 2.4. 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 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. [9], we implemented the Powell's method [17] to optimize the measures with respect to the transformation parameters. Partial volume interpolation (PV) was used in all 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  $1.25 \times 1.25 \times 4 \text{ mm}^3$ )
- 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  $2.59 \times 2.59 \times 8 \text{ mm}^3$ )

The gold standard transformations between the modalities were known thanks to a prospective, marker-based registration method [25]. We performed three kinds of registrations: T1 to T2, CT to T1, and PET to T1. In all experiments, the transformation was initially set either to the identity or to the gold standard: this was done to test if the algorithm was sensible to initialization. However, since the results were almost the same for both initializations, we present only those obtained with initializing with the identity.

For each registration, a typical error  $\epsilon$  was computed by taking the average registration error of the eight vertices of a bounding box corresponding approximately to the head volume. Thus,  $\epsilon$  represents the error one can expect in the region of interest. We also computed the intrinsic rotation error,  $\Delta\theta$ , and translation error,  $\Delta t$ , as described in [14]. Table 1 shows RMS of  $\epsilon$ ,  $\Delta\theta$ , and  $\Delta t$  over the ten patients, for each modality combination. These have to be compared to the image resolution, which is quite poor here (4 mm in the  $z$ -axis for the MR and CT datasets, and 8 mm for the PET).

In T1/T2 registration, CR and MI give good and similar results. For other combinations, significant differences are observed. MI does a better job than CR for matching CT to T1; this might be due to the fact that



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

Experiment	Measure	RMS		
		$\Delta \theta$ (deg)	$\Delta t$ (mm)	$\epsilon$ (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

functional dependence is a crude hypothesis in the CT/MR case (see [24]). On the other hand, CR seems better suited than MI for PET/T1 registration.

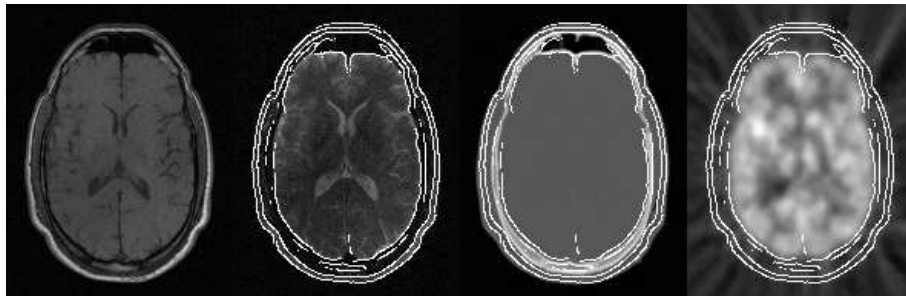


Figure 2: *Multimodal registration by maximization of CR. Images from left to right: MR-T1, MR-T2, CT, and PET. The images are resampled in the same reference frame after registration. Contours extracted from the MR-T1 are superimposed to each other modality in order to better visualize the quality of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

## 4.2 Ultrasound/MR registration

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:

- 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 dura mater, 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 then 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  $\Delta r$  with random direction and constant magnitude  $\|\Delta r\| =$

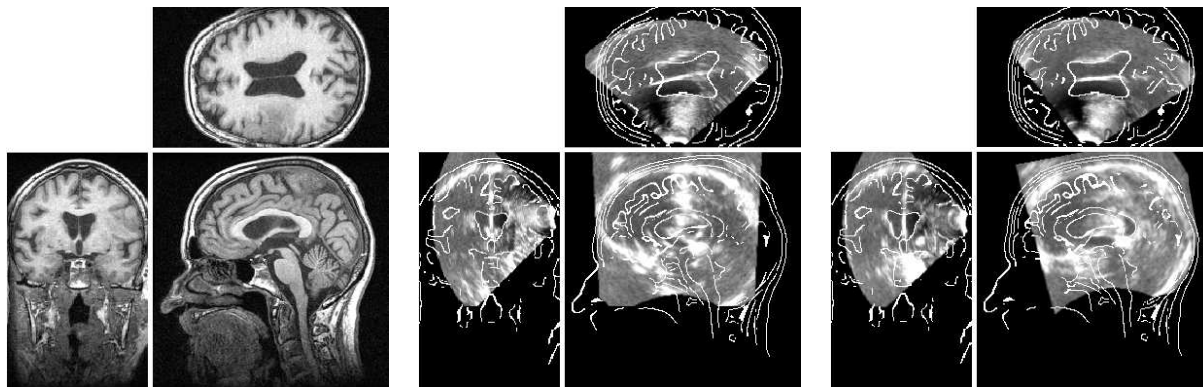


Figure 3: *US/MR registration by maximization of the correlation ratio. Left, three orthogonal views of the MR, T1 weighted image. Middle, corresponding views of the unregistered US image, with contours from the MR overlaid: the initial position corresponds to a random perturbation from the manually found “ground truth” position. Right, the result of automatic registration.*

15 degrees, and a translation  $\Delta t$  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.

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 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].

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
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 [16], 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.



Figure 4: *The different versions of the MR image used in these experiments. Left, original data. Middle, pre-segmentation by anisotropic diffusion. Right, addition of a 10% Gaussian white noise.*

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 rigid registration confirm (if needed) that similarity measures relating to different assumptions yield different performances. The CR measure was shown to be sometimes 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 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

Many thanks to Sébastien Ourselin, Alexandre Guimond and Sylvain Prima, for constant interaction and countless suggestions.

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 (UK), INRIA Sophia Antipolis (France), ISM-Salzburg and Kretz Technik (Austria).

Part of this work was supported by la Région Provence-Alpes-Côte d’Azur (France).

## A Polynomial least-squares fit

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

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

Then, we aim at minimizing the log-likelihood,

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

with respect to  $\theta = (\alpha_0, \alpha_1, \dots, \alpha_p, \sigma)$ . By differentiating (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  $AX = B$ , with

$$X = \left( \alpha_0 \quad \alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_d \right)^T$$

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

$$B = \left( \sum i_k \quad \sum j_k^\downarrow i_k \quad \sum j_k^{\downarrow 2} i_k \quad \cdots \quad \sum j_k^{\downarrow d} i_k \right)^T$$

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  $\sigma$ , we differentiate the log-likelihood:

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

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

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

a measure that directly generalizes the correlation ratio.

## B A “symmetric” generalization of the correlation ratio

We consider here the case where both images,  $I$  and  $J$ , are corrupted measures of the same unknown scene,  $S$ :

$$\begin{cases} I(x_k) &= f(S(T(x_k))) + \epsilon_k, \\ J(y_l) &= g(S(y_l)) + \gamma_k, \end{cases}$$

where  $f$  and  $g$  are the imaging functions corresponding respectively to  $I$  and  $J$ . We assume the noise processes,  $\epsilon$  and  $\gamma$ , to be white, stationary, and Gaussian with zero means and standard deviations  $\sigma$  and  $\varphi$ , respectively.

Let  $n_c$  denote the number of distinct tissue classes in the scene, and  $p = 1, \dots, n_c$  the labels corresponding to each class. Our modeling simply states that each tissue class is imaged in  $I$  (resp.  $J$ ) with a mean intensity  $f_p \equiv f(p)$  (resp.  $g_p \equiv g(p)$ ). We will assume a stationary prior probability on the scene, meaning that a point has a spatially constant probability to belong to a given class:

$$P(s_l = p) = \pi_p, \quad p = 1, \dots, n_c.$$

Under these assumptions, the intensity joint probability of the images is seen to be a mixture of bivariate Gaussians (with diagonal covariance matrix):

$$P(i_k, j_k^\downarrow | \theta) = \sum_{p=1}^{n_c} \pi_p N(i_k; f_p, \sigma) N(j_k^\downarrow; g_p, \varphi),$$

where  $\theta$  is the  $(3n_c+2)$ -dimensional parameter vector  $(f_1, f_2, \dots, f_{n_c}, g_1, \dots, g_{n_c}, \pi_1, \dots, \pi_{n_c}, \sigma, \varphi)$ . Equivalently, the conditional densities  $P(i_k | j_k^\downarrow, \theta)$  are mixtures of univariate Gaussians with proportions depending on the

value of  $j_k^\downarrow$ , that is:

$$P(i_k | j_k^\downarrow, \theta) = \sum_p \alpha_p(j_k^\downarrow) N(i_k; f_p, \sigma), \quad \text{with} \quad \alpha_p(j_k^\downarrow) = \frac{\pi_p N(j_k^\downarrow; g_p, \varphi)}{\sum_{p'} \pi_{p'} N(j_k^\downarrow; g_{p'}, \varphi)}. \quad (12)$$

Our problem is to determine the registration energy (9), which necessitates to compute the maximum likelihood  $\theta$  parameter corresponding to a given transformation  $T$  between the images, i.e. maximize  $P(I|J, T, \theta)$  with respect to  $\theta$ . We are faced to the classical problem of finding the maximum likelihood parameters of a mixture of Gaussians. As it is well known, there is no explicit solution to this, and only a numerical optimization can be performed.

The Expectation-Maximization (EM) algorithm is a powerful method that is typically adapted to this situation. The interested reader will find an excellent introduction to the EM algorithm and several references in [6]. Applying the EM algorithm to our problem, the parameters are updated iteratively until convergence according to the following reestimation formulae:

$$\begin{aligned} f_p^{(n+1)} &= \frac{\sum_k w_{pk}^{(n)} i_k}{\sum_k w_{pk}^{(n)}}, \\ g_p^{(n+1)} &= \frac{\sum_k w_{pk}^{(n)} j_k^\uparrow}{\sum_k w_{pk}^{(n)}}, \\ \sigma^{(n+1)} &= \sqrt{\frac{1}{N} \sum_p \sum_k w_{pk}^{(n)} (i_k - f_p^{(n)})^2}, \\ \varphi^{(n+1)} &= \sqrt{\frac{1}{N} \sum_p \sum_k w_{pk}^{(n)} (j_k^\uparrow - g_p^{(n)})^2}, \\ \pi_p^{(n+1)} &= \frac{1}{N} \sum_k w_{pk}^{(n)}, \end{aligned}$$

where the adaptative weights  $w_{pk}^{(n)}$  represent the probability that the voxel with coordinates  $x_k$  belong to the tissue class  $p$ , given the images and the current estimate  $\theta_n$ :

$$w_{pk}^{(n)} = P(s_k^\downarrow = p | i_k, j_k^\downarrow, \theta_n) = \frac{\pi_p^{(n)} N(i_k; f_p^{(n)}, \sigma^{(n)}) N(j_k^\downarrow; g_p^{(n)}, \varphi^{(n)})}{\sum_{p'} \pi_{p'}^{(n)} N(i_k; f_{p'}^{(n)}, \sigma^{(n)}) N(j_k^\downarrow; g_{p'}^{(n)}, \varphi^{(n)})}.$$

Once this estimation step has been completed, the registration energy is computed using (9) via reinjecting the optimal parameters  $\theta$  in the conditional densities (12). This registration measure is symmetric in the sense that the underlying acquisition models are of the same form for both images. This does not mean however that the result of registration is independent of which image is chosen as the reference.

Needless to say, due to the use of an EM algorithm, heavy extra computational cost is required with respect to the measures presented in section 3. In [8], mixture of Gaussians are also used to model the intensity joint density of a training pair of images. This density is then considered valid for subsequent image pairs, and it is used unmodified (non parametrically) within a comparable maximum likelihood registration strategy. In this case, the registration measure is fast to compute since the acquisition parameters are fixed. However, this might only work under the assumption that the novel images to be registered are acquired from the same sensors as the training images and contain equivalent anatomical information.

Two important remarks should be added. First, the EM algorithm does not attempt to maximize the *conditional* probability  $P(I|J, T, \theta)$  as would be consistent with our formulation (see section 2.2), but the following *joint* probability,

$$P(I, J^\perp | \theta) = P(I|J, T, \theta) P(J^\perp | \theta), \quad (13)$$

which is the joint probability of  $I$  and the subset of voxels from  $J$  that are put into correspondence with the voxels from  $I$ , that is:  $J^\perp$ . Since  $P(J^\perp | \theta)$  also depends on  $\theta$ , the value of  $\theta$  that maximizes (13) has no reason to be located at the same position as that of  $P(I|J, T, \theta)$ . Thus, we cannot claim in this case that the estimation of  $T$  and the estimation of  $\theta$  are performed by optimizing a unique criterion.

The second remark is that the EM algorithm is not guaranteed to converge to the global maximum of the likelihood function, but only to a local maximum. In practice, this means that a good initialization of  $\theta$  is crucial.

## 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. *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] C. Couvreur. The EM Algorithm: A Guided Tour. In *Proc. 2d IEEE European Workshop on Computationaly Intensive Methods in Control and Signal Processing*, Prague, Czech Republik, August 1996.
- [7] S. Lavalée. Registration for computer integrated surgery: methodology, state of the art. In R. Taylor, S. Lavalée, G. Burdea, and R. Moesges, editors, *Computer Integrated Surgery*, pages 77–97. MIT Press, 1995.
- [8] 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.
- [9] 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.
- [10] J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *MedIA*, 2(1):1–36, 1998.
- [11] 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.
- [12] 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.

- [13] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Inc., third edition, 1991.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C*. Cambridge 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*, pages 1115–1124, Cambridge, Massachusetts (USA), October 1998. Lecture Notes in Computer Science.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] P. Viola and W. M. Wells. Alignment by Maximization of Mutual Information. *Intern. J. of Comp. Vision*, 24(2):137–154, 1997.
- [24] 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.
- [25] J. West and al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Comp. Assist. Tomography*, 21:554–566, 1997.
- [26] 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.





---

Unité de recherche INRIA Sophia Antipolis

2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Lorraine : Technopôle de Nancy-Brabois - Campus scientifique

615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)

Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot St Martin (France)

Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

---

Éditeur

INRIA - Domaine de Voluceau - Rocquencourt, B.P. 105 - 78153 Le Chesnay Cedex (France)

<http://www.inria.fr>

ISSN 0249-6399