Variational Methods for Multimodal Image Matching

GERARDO HERMOSILLO, CHRISTOPHE CHEFD'HOTEL* AND OLIVIER FAUGERAS
INRIA Sophia Antipolis, Odyssée Lab, 2004 route des Lucioles BP 93 Sophia Antipolis Cedex, France
Gerardo.Hermosillo@inria.fr
Christophe.Chef_d_hotel@inria.fr
Olivier.Faugeras@inria.fr

Abstract. Matching images of different modalities can be achieved by the maximization of suitable statistical similarity measures within a given class of geometric transformations. Handling complex, nonrigid deformations in this context turns out to be particularly difficult and has attracted much attention in the last few years. The thrust of this paper is that many of the existing methods for nonrigid monomodal registration that use simple criteria for comparing the intensities (e.g. SSD) can be extended to the multimodal case where more complex intensity similarity measures are necessary. To this end, we perform a formal computation of the variational gradient of a hierarchy of statistical similarity measures, and use the results to generalize a recently proposed and very effective optical flow algorithm (L. Alvarez, J. Weickert, and J. Sánchez, 2000, Technical Report, and *IJCV* 39(1):41–56) to the case of multimodal image registration. Our method readily extends to the case of locally computed similarity measures, thus providing the flexibility to cope with spatial non-stationarities in the way the intensities in the two images are related. The well posedness of the resulting equations is proved in a complementary work (O.D. Faugeras and G. Hermosillo, 2001, Technical Report 4235, INRIA) using well established techniques in functional analysis. We briefly describe our numerical implementation of these equations and show results on real and synthetic data.

Keywords: image matching, cross correlation, correlation ratio, mutual information, partial differential equations, regularization, variational methods

1. Introduction

The problem of establishing correspondences between two or more images is fundamental in computer vision. It is one of the building blocks for a number of challenging problems such as 3D reconstruction, camera motion estimation, template matching and camera calibration. When images have been acquired through similar sensors, they can be realigned by a direct comparison of their intensities. This results in matching algorithms that essentially look for the geometric transformation between two images which minimizes the sum of the squared differences between their intensity values.

There are several situations in which the hypothesis of the invariance of the intensity is no longer valid. One may consider for instance varying illumination conditions, or sensors with different spectral sensitivities. The same situation is encountered in medical imaging, where several acquisition modalities must be realigned to allow for an accurate fusion of complementary information. To cope with this new difficulty, statistical similarity measures between the image intensities have been proposed. Two such similarity measures are the mutual information, proposed by Viola and Wells (1997) and independently by Maes et al. (1997) and the correlation ratio, proposed by Roche et al. (1998a).

---

*Part of this work was done while the second author was visiting the MIT Artificial Intelligence Laboratory.

Initially, these criteria were used through their maximization over a low-dimensional space of parametric transformations, e.g. rigid or affine. Recent extensions to larger sets of deformations rely on more complex parametric models (Meyer et al., 1998; Rückert et al., 1998) or block-matching strategies (Maintz et al., 1998; Hata et al., 1998; Gaens et al., 1998). Another interesting approach to multimodal, nonrigid matching was proposed by Roche et al. (2000a), where parametric intensity corrections are combined with standard monomodal algorithms.

In this paper, we propose to extend these approaches by analyzing three statistical criteria in a variational setting. The idea is to recover the underlying deformation by estimating a dense displacement field through the maximization of the similarity measures over a suitable, infinite-dimensional functional space. We believe that this approach opens the possibility of generalizing many nonrigid matching algorithms that are expressed under similar formalisms to the case of multimodal matching. Our main contribution in this paper is the derivation of the Euler-Lagrange equations for the statistical criteria, a task which is nontrivial due to their intricate definition. We consider the cases of mutual information, correlation ratio and cross correlation since they provide a suitable hierarchy of statistical criteria going from the most robust (cross correlation) to the most general (mutual information) (Roche et al., 2000b). Our approach relies on smooth estimates of the joint probability density function (pdf) of the intensities in the first and the warped second image. We extend our previous work reported in Chefd'hotel et al. (2001) by also considering locally estimated pdfs, thus accounting for possible spatial non-stationarities in the intensity relations.

The paper is organized as follows. Section 2 discusses the variational formalism and its application to defining the matching problem. Section 3 describes the choice of a suitable functional to perform regularization. In Section 4, we start with a simple registration method which assumes that the joint intensity pdf between different modalities is known (e.g. learned from sample datasets). This idea is extended in Section 5 to the minimization of statistical dissimilarity measures when the joint intensity distribution is unknown. Theoretical results on the existence and uniqueness of a generic matching flow which embodies both the dissimilarity and the regularization components are presented in Section 6. In Section 7 we briefly describe the numerical implementation of the flows and show

results of experiments on synthetic and real images in Section 8. In Section 9, we conclude and provide some perspectives.

2.   Dense Matching and the Variational Framework

We now state our modeling assumptions and define the matching problem in the context of the calculus of variations. We consider two images $I_1^\sigma = I_1 \star G_\sigma$ and $I_2^\sigma = I_2 \star G_\sigma$ at a given scale σ, i.e. resulting from the convolution of two square-integrable functions $I_1 \colon \mathbb{R}^n \to \mathbb{R}$ and $I_2 \colon \mathbb{R}^n \to \mathbb{R}$ with a Gaussian kernel of standard deviation σ (we restrict ourselves to the cases $n = 2, 3$). Given a region of interest Ω, a bounded region of $\mathbb{R}^n$, we look for a function $\mathbf{h} \colon \Omega \to \mathbb{R}^n$ assigning to each point $\mathbf{x}$ in Ω a displacement vector $\mathbf{h}(\mathbf{x}) \in \mathbb{R}^n$. This function is searched for in a set $\mathcal{F}$ of admissible functions such that it minimizes an energy functional $\mathcal{I} \colon \mathcal{F} \to \mathbb{R}$ of the form

$$\mathcal{I}(\mathbf{h}) = \mathcal{J}(\mathbf{h}) + \mathcal{R}(\mathbf{h}).$$

Generally speaking, the set $\mathcal{F}$ is assumed to be a linear subspace of a Hilbert space H, the scalar product of which is denoted $(\cdot, \cdot)_H$.

The term $\mathcal{J}(\mathbf{h})$ is designed to measure the "dissimilarity" between the reference image (I_1^σ) and the $\mathbf{h}$-warped second image ($I_2^\sigma (\mathbf{Id} + \mathbf{h})$), $\mathbf{Id}$ denoting the identity mapping of $\mathbb{R}^n$. The term $\mathcal{R}(\mathbf{h})$ is designed to penalize fast variations of the function $\mathbf{h}$. To summarize, we define the matching problem as that of minimizing $\mathcal{I}$ with respect to $\mathbf{h}$, the matching being a solution $\hat{\mathbf{h}}$ of this minimization problem:

$$\hat{\mathbf{h}} = \arg \min_{\mathbf{h} \in \mathcal{F}} \mathcal{I}(\mathbf{h}) = \arg \min_{\mathbf{h} \in \mathcal{F}} (\mathcal{J}(\mathbf{h}) + \mathcal{R}(\mathbf{h})).$$

Assuming that $\mathcal{I}$ is sufficiently regular, its first variation at $\mathbf{h} \in \mathcal{F}$ in the direction of $\mathbf{k} \in \mathcal{F}$ is defined by (see e.g. Courant, 1946)

$$\delta_{\mathbf{k}} \mathcal{I}(\mathbf{h}) = \left. \frac{\partial \mathcal{I}(\mathbf{h} + \epsilon \mathbf{k})}{\partial \epsilon} \right|_{\epsilon = 0}. \qquad (1)$$

The gradient $\nabla_H \mathcal{I}(\mathbf{h})$ of $\mathcal{I}$ is defined[1] by requiring the equality

$$\delta_{\mathbf{k}} \mathcal{I}(\mathbf{h}) = (\nabla_H \mathcal{I}(\mathbf{h}), \mathbf{k})_H$$

to hold for every $\mathbf{k} \in \mathcal{F}$. If a minimizer $\hat{\mathbf{h}}$ of $\mathcal{I}$ exists, then the set of equations $\delta_{\mathbf{k}} \mathcal{I}(\hat{\mathbf{h}}) = 0$ must hold for every $\mathbf{k} \in \mathcal{F}$, which is equivalent to $\nabla_H \mathcal{I}(\hat{\mathbf{h}}) = 0$.

These are called the Euler-Lagrange equations associated with the energy functional $\mathcal{I}$. Rather than solving them directly (which is usually impossible), the search for a minimizer of $\mathcal{I}$ is done using a "gradient descent" strategy. Given an initial estimate $\mathbf{h}_0 \in H$, a time-dependent, differentiable function (also noted $\mathbf{h}$) from the interval $[0, +\infty[$ into H is computed as the solution of the following initial value problem:

$$\begin{cases} \frac{d\mathbf{h}}{dt} = -(\nabla_H \mathcal{J}(\mathbf{h}) + \nabla_H \mathcal{R}(\mathbf{h})), \\ \mathbf{h}(0)(\cdot) = \mathbf{h}_0(\cdot). \end{cases} \quad (2)$$

The asymptotic state (i.e. when $t \to \infty$) of $\mathbf{h}(t)$ is then chosen as the solution of the matching problem, provided that $\mathbf{h}(t) \in \mathcal{F}$ for all t's.

3. Partial Differential Equations for Image Matching

In this section, we summarize a method for optical flow estimation that was recently proposed by Alvarez et al. (2000) and point out its relation with the previously described framework. We then give some examples of differential operators $\nabla_H \mathcal{R}(\mathbf{h})$.

Assuming the images have been acquired through similar sensors (possibly corrupted by a Gaussian noise), a suitable dissimilarity functional is given by the sum of squared differences of the image intensities (Toga, 1998):

$$\mathcal{J}(\mathbf{h}) = \int_\Omega \left(I_1^\sigma(\mathbf{x}) - I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) \right)^2 d\mathbf{x}.$$

Besides, one of the simplest ways of penalizing fast variations of $\mathbf{h}$ is by defining

$$\mathcal{R}(\mathbf{h}) = \alpha \int_\Omega |D\mathbf{h}|^2 \, d\mathbf{x}.$$

In this case, Eq. (2) yields a set of reaction-diffusion partial differential equations (PDEs) with initial condition $\mathbf{h}_0$ and isotropic diffusion:

$$\begin{cases} \frac{\partial \mathbf{h}}{\partial t}(t, \mathbf{x}) = F(\mathbf{h})(\mathbf{x}) + \alpha \Delta \mathbf{h}(\mathbf{x}) \\ \mathbf{h}(0, \mathbf{x}) = \mathbf{h}_0(\mathbf{x}), \end{cases} \quad (3)$$

where

$$F(\mathbf{h})(\mathbf{x}) = \left(I_1^\sigma(\mathbf{x}) - I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) \right) \nabla I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})). \quad (4)$$

Considering a more general regularization functional of the form

$$\mathcal{R}(\mathbf{h}) = \alpha \int_\Omega \text{Tr}(D\mathbf{h} \, \mathbf{T} D\mathbf{h}^T) \, d\mathbf{x}$$

yields a anisotropic diffusion term $\alpha \, \text{div}(\mathbf{T} D\mathbf{h})$. Alvarez et al. (2000) proposed an optical flow estimation method using this type of PDEs. They propose an image-driven anisotropic diffusion term based on previous work by Nagel and Enkelmann (1986). Their diffusion tensor is designed to prevent fast variations of $\mathbf{h}$ across the contours of I_1^σ. We propose here a n-dimensional generalization of this tensor

$$\mathbf{T}_{I_1^\sigma} = \frac{\left(\lambda + \left\| \nabla I_1^\sigma \right\|^2 \right) \mathbf{I} - \nabla I_1^\sigma \nabla I_1^{\sigma T}}{(n-1) \left\| \nabla I_1^\sigma \right\|^2 + n\lambda}. \quad (5)$$

It has one eigenvector equal to ∇I_1^σ, while the remaining eigenvectors span the plane perpendicular to ∇I_1^σ. The corresponding eigenvalues are respectively

$$\kappa_n = \frac{\lambda}{(n-1) \left\| \nabla I_1^\sigma \right\|^2 + n\lambda}$$

and

$$\kappa_{1,\ldots,n-1} = \frac{\lambda + \left\| \nabla I_1^\sigma \right\|^2}{(n-1) \left\| \nabla I_1^\sigma \right\|^2 + n\lambda}.$$

In homogeneous regions, as $\|\nabla I_1^\sigma\| \ll \lambda$, all the eigenvalues tend to $1/n$ and the diffusion is nearly isotropic. Along the contours of I_1^σ, when $\|\nabla I_1^\sigma\| \gg \lambda$, we have $\kappa_n \to 0$, while the remaining eigenvalues tend to $1/(n-1)$. The diffusion becomes anisotropic in this case, taking place mainly along the contours.

As in Christensen et al. (1994), another possible regularization term is obtained by minimizing an approximate strain energy borrowed from linear elasticity. This approach yields a regularization operator of the form

$$\lambda \Delta \mathbf{h} + (\mu + \lambda) \nabla (\nabla \cdot \mathbf{h}). \quad (6)$$

The constants λ and μ are known as the Lamé coefficients.

In the following two sections, we generalize Eq. (3) to the case of nonrigid multimodal registration by focusing on the matching term $F(\mathbf{h})$. We replace it by a set of functions designed for the minimization of statistical dissimilarity measures.

4. Supervised Registration

Our approach relies on regarding the intensity values of two different modalities as samples of two random processes. Within this probabilistic framework, the link between the two modalities is characterized by their joint probability density function (pdf).

In this section, we assume that reference templates, or sets of manually pre-registered images are available to estimate the real joint pdf P, which is assumed to be differentiable and strictly positive. From this knowledge, we can derive a *supervised* registration principle (**SR**). We borrow from information theory the notion of *uncertainty* or *information*, defined for an event with respect to a probability measure. In this case, an event is the co-occurrence of two intensity values $i_1 = I_1^\sigma(\mathbf{x})$ and $i_2 = I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))$ at any point $\mathbf{x} \in \Omega$. For conciseness, we will use the notations $\mathbf{i} = (i_1, i_2)$ and $\mathbf{I_h}(\mathbf{x}) = (I_1^\sigma(\mathbf{x}), I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})))$. The amount of information conveyed by this event is given by $-\log P(\mathbf{i})$. A global dissimilarity measure follows by computing the total amount of information for a given displacement field $\mathbf{h}$:

$$\mathcal{J}_{\mathbf{SR}}(\mathbf{h}) = \int_\Omega -\log P(\mathbf{I_h}(\mathbf{x}))\, d\mathbf{x}.$$

This expression can be viewed as a continuous form of the maximum likelihood principle developed by Leventon and Grimson in Leventon and Grimson (1998). By applying (1), the first variation of $\mathcal{J}_{\mathbf{SR}}$ is readily found to be

$$\delta_{\mathbf{k}} \mathcal{J}_{\mathbf{SR}}(\mathbf{h}) = \int_\Omega -\frac{\partial_2 P(\mathbf{I_h}(\mathbf{x}))}{P(\mathbf{I_h}(\mathbf{x}))} \nabla I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) \cdot \mathbf{k}(\mathbf{x})\, d\mathbf{x},$$

where $\partial_2 P$ denotes the partial derivative of P with respect to its second variable. Choosing H as the set of vector-valued square integrable functions on Ω, this expression is of the form

$$\delta_{\mathbf{k}} \mathcal{J}_{\mathbf{SR}}(\mathbf{h}) = (\nabla_H \mathcal{J}_{\mathbf{SR}}(\mathbf{h}), \mathbf{k})_H.$$

We define the function

$$L_{\mathbf{SR}}(\mathbf{i}) = -\partial_2 P(\mathbf{i})/P(\mathbf{i})$$

so that the gradient $\nabla_H \mathcal{J}_{\mathbf{SR}}$ of our first dissimilarity measure $\mathcal{J}_{\mathbf{SR}}$ is given by

$$\nabla_H \mathcal{J}_{\mathbf{SR}}(\mathbf{h})(\mathbf{x}) = L_{\mathbf{SR}}(\mathbf{I_h}(\mathbf{x}))\nabla I_2^\sigma(\mathbf{x} + \mathbf{h}(x)). \quad (7)$$

The mapping $L_{\mathbf{SR}} : \mathbb{R}^2 \to \mathbb{R}$ plays the role of an intensity comparison function. It is the so-called *score* function, often encountered in the derivation of maximum likelihood principles. Notice that a functional dependence between the intensities is not required. Any intensity in the first image may have several corresponding intensities in the second one, and conversely. The knowledge of the nearest most likely intensity correspondence is implicitly given by this function.

Let us now assume that we do not have any training set available for the learning process. One can develop the following argument: if the initial pose is close to the solution (small deformations), a sufficiently robust estimate of their joint intensity distribution may be a good approximation of the real joint density for these two modalities. The incorrectly matched image values are considered as noise. Subsequently, $\mathbf{h}$ could be recovered from this initial joint density estimate. This is one of the underlying ideas of the methods presented in the following section.

5. Statistical Dissimilarity Functionals

In this section, we drop the assumption of the prior knowledge of the function P and compute the variational gradient of three statistical dissimilarity functionals. Among many possible criteria, the cross correlation, the correlation ratio and the mutual information provide us with a convenient hierarchy because of the relation they can enforce between the intensities of the two images (see Fig. 1). To be able to evaluate these criteria for a given field $\mathbf{h}$, we consider a nonparametric Parzen estimator (Parzen, 1962) for the joint pdf as described below. We call X the random variable associated to the intensity values of I_1^σ and $Y_\mathbf{h}$ that associated to the values of $I_2^\sigma(\mathbf{Id} + \mathbf{h})$.

5.1. Density Estimation

Our estimator is based on a normalized Gaussian kernel of variance β, noted $G_\beta(\mathbf{i})$:

$$P(\mathbf{i}, \mathbf{h}) = \frac{1}{|\Omega|} \int_\Omega G_\beta(\mathbf{I_h}(\mathbf{x}) - \mathbf{i})\, d\mathbf{x}.$$

Note that for each pair of intensities $\mathbf{i} \in \mathbb{R}^2$, the value of the estimated joint pdf is a nonlinear functional of $\mathbf{h}$.
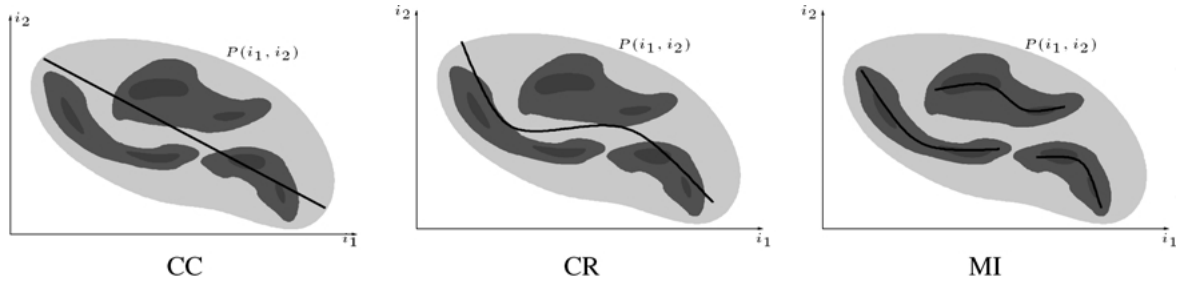
Figure 1. The three criteria provide a hierarchy of dependence-levels between two random variables. The cross correlation (CC) measures their affine dependence so that maximizing this criterion amounts to finding an affine function best fitting the joint pdf. This is illustrated in the left figure by the straight line superimposed over the schematic joint pdf P, where dark values represent high probabilities. The correlation ratio (CR) measures their functional dependence so that the optimal density can have the shape of a nonlinear function (middle). Finally, the mutual information (MI) gives an estimate of the statistical dependence and maximizing this criterion tends to cluster P (right).

Its first variation is obtained by applying (1):

$$\delta_{\mathbf{k}} P(\mathbf{i}, \mathbf{h}) = \frac{1}{|\Omega|} \int_{\Omega} \partial_2 G_\beta(\mathbf{I_h}(\mathbf{x}) - \mathbf{i})$$
$$\nabla I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) \cdot \mathbf{k}(\mathbf{x}) \, d\mathbf{x}. \quad (8)$$

An interesting generalization is to make the density estimator local, since it allows to take into account non-stationarities in the relation between intensities. To do this, we build an estimate in the neighborhood of each point $\mathbf{x}_0$ in Ω. This is achieved by weighting our previous estimate with a normalized Gaussian of variance γ. This means that to each point $\mathbf{x}_0$ we associate a joint pdf defined by:

$$P(\mathbf{i}, \mathbf{h}, \mathbf{x}_0) = \frac{1}{\mathcal{G}_\gamma(\mathbf{x}_0)} \int_{\Omega} G_\beta(\mathbf{I_h}(\mathbf{x}) - \mathbf{i}) G_\gamma(\mathbf{x} - \mathbf{x}_0) \, d\mathbf{x},$$

where

$$\mathcal{G}_\gamma(\mathbf{x}_0) = \int_{\Omega} G_\gamma(\mathbf{x} - \mathbf{x}_0) \, d\mathbf{x}.$$

This new pdf is along the line of the ideas discussed in Koenderink and van Doorn (1999), except that we now have a bidimensional local histogram at each point.

5.2. Mutual Information

Using these tools, we first consider the maximization of mutual information, a concept which is borrowed from information theory. Given two random variables X and Y, their mutual information is defined as

$$\mathbf{MI}(X, Y) = \mathcal{H}(X) + \mathcal{H}(Y) - \mathcal{H}(X, Y),$$

where $\mathcal{H}$ stands for the differential entropy. The mutual information is positive and symmetric, and measures how the intensity distributions of two images fail to be independent. It can be defined in terms of the joint pdf P and its marginals $p(i_1) = \int_{\mathbb{R}} P(\mathbf{i}, \mathbf{h}) \, di_2$ and $p(i_2, \mathbf{h}) = \int_{\mathbb{R}} P(\mathbf{i}, \mathbf{h}) \, di_1$. The following short notation will be useful:

$$E^{\mathrm{MI}}(\mathbf{i}, \mathbf{h}) = -\log \frac{P(\mathbf{i}, \mathbf{h})}{p(i_1)p(i_2, \mathbf{h})}. \quad (9)$$

The dissimilarity functional based on mutual information is then defined as the expected value of the function E^{MI}:

$$\mathcal{J}_{\mathbf{MI}^g}(\mathbf{h}) = -\mathbf{MI}(X, Y_{\mathbf{h}}) = \int_{\mathbb{R}^2} P(\mathbf{i}, \mathbf{h}) E^{\mathrm{MI}}(\mathbf{i}, \mathbf{h}) \, d\mathbf{i}.$$

We do an explicit computation of its first variation by directly applying (1). Using the fact that

$$\int_{\mathbb{R}^2} \delta_{\mathbf{k}} P(\mathbf{i}, \mathbf{h}) \, d\mathbf{i} = \int_{\mathbb{R}} \delta_{\mathbf{k}} p(i_2, \mathbf{h}) \, di_2 = 0,$$

this yields

$$\delta_{\mathbf{k}} \mathcal{J}_{\mathbf{MI}^g}(\mathbf{h}) = \int_{\mathbb{R}^2} \delta_{\mathbf{k}} P(\mathbf{i}, \mathbf{h}) E^{\mathrm{MI}}(\mathbf{i}, \mathbf{h}) \, d\mathbf{i}. \quad (10)$$

We then apply (8) to obtain

$$\delta_{\mathbf{k}} \mathcal{J}_{\mathbf{MI}^g}(\mathbf{h}) = \frac{1}{|\Omega|} \int_{\mathbb{R}^2} \int_{\Omega} E^{\mathrm{MI}}(\mathbf{i}, \mathbf{h})$$
$$\partial_2 G_\beta(\mathbf{I_h}(\mathbf{x}) - \mathbf{i}) \nabla I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) \cdot \mathbf{k}(\mathbf{x}) \, d\mathbf{x} \, d\mathbf{i}.$$

A convolution with respect to the intensity variable $\mathbf{i}$ appears in this expression. It commutes with the derivative ∂_2 with respect to the second intensity variable i_2,

and therefore

$$\delta_{\mathbf{k}}\mathcal{J}_{\mathbf{MI}^g}(\mathbf{h}) = \frac{1}{|\Omega|} \int_{\Omega} (G_{\beta} \star \partial_2 E^{\mathrm{MI}})(\mathbf{I_h(x)}, \mathbf{h})$$
$$\nabla I_2^{\sigma}(\mathbf{x} + \mathbf{h(x)}) \cdot \mathbf{k(x)}\, d\mathbf{x}.$$

We define the function $\mathbb{R}^2 \times H \to \mathbb{R}$:

$$L_{\mathrm{MI}}^g(\mathbf{i}, \mathbf{h}) = \frac{1}{|\Omega|} G_{\beta} \star \partial_2 E^{\mathrm{MI}}(\mathbf{i}, \mathbf{h})$$
$$= -\frac{1}{|\Omega|} G_{\beta} \star \left(\frac{\partial_2 P(\mathbf{i}, \mathbf{h})}{P(\mathbf{i}, \mathbf{h})} - \frac{p'(i_2, \mathbf{h})}{p(i_2, \mathbf{h})} \right).$$

The variational gradient $\nabla_H \mathcal{J}_{\mathbf{MI}^g}(\mathbf{h})$ of our second dissimilarity functional $\mathcal{J}_{\mathbf{MI}^g}$ is thus given by

$$\nabla_H \mathcal{J}_{\mathbf{MI}^g}(\mathbf{h})(\mathbf{x}) = L_{\mathrm{MI}}^g(\mathbf{I_h(x)}, \mathbf{h}) \nabla I_2^{\sigma}(\mathbf{x} + \mathbf{h(x)}).$$

Similarly to the case of supervised registration (recall equation (7)), the function L_{MI}^g plays the role of an intensity comparison function. Its first term $\partial_2 P(\mathbf{i}, \mathbf{h})/P(\mathbf{i}, \mathbf{h})$ tends to cluster the joint pdf, while the term $-p'(i_2, \mathbf{h})/p(i_2, \mathbf{h})$ tries to prevent the marginal law $p(i_2, \mathbf{h})$ from becoming too clustered; i.e. keeps the intensities of $I_2^{\sigma}(\mathbf{Id} + \mathbf{h})$ as unpredictable as possible.

5.2.1 Extension to Locally Estimated Densities.
The extension of the previous dissimilarity criterion using the local version of the density estimate is carried out in a straightforward manner. All the definitions remain unchanged, up to their dependence on the position $\mathbf{x}$. For example, Eq. (9) becomes

$$E^{\mathrm{MI}}(\mathbf{i}, \mathbf{h}, \mathbf{x}) = -\log \frac{P(\mathbf{i}, \mathbf{h}, \mathbf{x})}{p(i_1, \mathbf{x})\, p(i_2, \mathbf{h}, \mathbf{x})}.$$

The local version of the dissimilarity functional based on mutual information is then defined as

$$\mathcal{J}_{\mathbf{MI}^l}(\mathbf{h}) = \int_{\Omega} \mathcal{J}_{\mathrm{MI}}(\mathbf{h}, \mathbf{x})\, d\mathbf{x}$$
$$= \int_{\Omega} \int_{\mathbb{R}^2} P(\mathbf{i}, \mathbf{h}, \mathbf{x}) E^{\mathrm{MI}}(\mathbf{i}, \mathbf{h}, \mathbf{x})\, d\mathbf{i}\, d\mathbf{x}.$$

The computation of its first variation is carried out in the same way as that of $\mathcal{J}_{\mathbf{MI}^g}$ in the previous section. The only difference is the additional integral in the space variable which, combined with G_{γ}, yields a spatial convolution. The intensity comparison function becomes a function from $\mathbb{R}^2 \times H \times \Omega$ into $\mathbb{R}$:

$$L_{\mathrm{MI}}^l(\mathbf{i}, \mathbf{h}, \mathbf{x})$$
$$= G_{\gamma} \star G_{\beta} \star \frac{1}{\mathcal{G}_{\gamma}(\mathbf{x})} \left(\frac{\partial_2 P(\mathbf{i}, \mathbf{h}, \mathbf{x})}{P(\mathbf{i}, \mathbf{h}, \mathbf{x})} - \frac{p'(i_2, \mathbf{h}, \mathbf{x})}{p(i_2, \mathbf{h}, \mathbf{x})} \right).$$

5.3. Correlation Ratio

We now consider the correlation ratio. This criterion relies on a slightly different notion of similarity. Given two random variables X and Y, the correlation ratio is defined as

$$\mathbf{CR}(X, Y) = \frac{\mathbf{Var}(\mathbf{E}(Y|X))}{\mathbf{Var}(Y)}.$$

From this formula, the correlation ratio can be described as the proportion of energy in Y which is "explained" by X. More formally, this measure is bounded ($0 \leq \mathbf{CR} \leq 1$) and expresses the level of *functional* dependence between Y and X:

$$\begin{cases} \mathbf{CR}(X, Y) = 1 \Leftrightarrow \exists \phi : Y = \phi(X) \\ \mathbf{CR}(X, Y) = 0 \Leftrightarrow \mathbf{E}(Y|X) = \mathbf{E}(Y). \end{cases}$$

To define the dissimilarity functional based on the correlation ratio and compute its variational gradient, we need the conditional expectation $\mathbf{E}(Y_{\mathbf{h}} \,|\, X)$. We note the value of this random variable $\mu_{2|1}(i_1, \mathbf{h})$, indicating that it depends on the intensity value i_1 and on the field $\mathbf{h}$:

$$\mu_{2|1}(i_1, \mathbf{h}) = \int_{\mathbb{R}} i_2 \frac{P(\mathbf{i}, \mathbf{h})}{p(i_1)}\, di_2.$$

We also use the conditional variance $\mathbf{Var}[\mathbf{E}(Y_{\mathbf{h}} \,|\, X)]$, which is again a random variable. Its value is abbreviated $v_{2|1}(i_1, \mathbf{h})$:

$$v_{2|1}(i_1, \mathbf{h}) = \int_{\mathbb{R}} i_2^2 \frac{P(\mathbf{i}, \mathbf{h})}{p(i_1)}\, di_2 - \mu_{2|1}(i_1, \mathbf{h})^2.$$

Finally, the mean and variance of $Y_{\mathbf{h}}$ will also be used:

$$\mu_2(\mathbf{h}) = \int_{\mathbb{R}} i_2\, p(i_2, \mathbf{h})\, di_2,$$
$$v_2(\mathbf{h}) = \int_{\mathbb{R}} i_2^2\, p(i_2, \mathbf{h}) di_2 - \mu_2(\mathbf{h})^2.$$

Instead of working with the original definition of $\mathbf{CR}$, we use the total variance theorem to obtain

$$\mathbf{CR}(X, Y) = 1 - \frac{\mathbf{E}(\mathbf{Var}(Y \,|\, X))}{\mathbf{Var}(Y)}.$$

This transformation was suggested by Roche et al. (1998b). We then define the dissimilarity measure based on the correlation ratio as

$$\mathcal{J}_{\mathbf{CR}^g}(\mathbf{h}) = \frac{w(\mathbf{h})}{v_2(\mathbf{h})},$$

where

$$w(\mathbf{h}) = \int_{\mathbb{R}} v_{2|1}(i_1, \mathbf{h}) p(i_1) \, di_1.$$

The basic apparatus for computing the first variation of $\mathcal{J}_{\mathbf{CR}^g}$ is similar to the one we use for the mutual information. We have

$$\delta_{\mathbf{k}} \mathcal{J}_{\mathbf{CR}^g}(\mathbf{h}) = \frac{\delta_{\mathbf{k}} w(\mathbf{h}) - \mathcal{J}_{\mathbf{CR}^g}(\mathbf{h}) \delta_{\mathbf{k}} v_2(\mathbf{h})}{v_2(\mathbf{h})},$$

where

$$\delta_{\mathbf{k}} w(\mathbf{h}) = \int_{\mathbb{R}^2} \left(i_2^2 - 2 i_2 \mu_{2|1}(\mathbf{h}) \right) \delta_{\mathbf{k}} P(\mathbf{i}, \mathbf{h}) \, d\mathbf{i}$$

and

$$\delta_{\mathbf{k}} v_2(\mathbf{h}) = \int_{\mathbb{R}^2} \left(i_2^2 - 2 i_2 \mu_2(\mathbf{h}) \right) \delta_{\mathbf{k}} P(\mathbf{i}, \mathbf{h}) \, d\mathbf{i}.$$

As in (10), the first variation of $\mathcal{J}_{\mathbf{CR}^g}(\mathbf{h})$ is of the form

$$\delta_{\mathbf{k}} \mathcal{J}_{\mathbf{CR}^g}(\mathbf{h}) = \int_{\mathbb{R}^2} \delta_{\mathbf{k}} P(\mathbf{i}, \mathbf{h}) E^{\mathbf{CR}}(\mathbf{i}, \mathbf{h}) \, d\mathbf{i}.$$

The discussion starting at Eq. (10) remains identical in this case and therefore

$$\nabla_H \mathcal{J}_{\mathbf{CR}^g}(\mathbf{h})(\mathbf{x}) = L_{\mathrm{CR}}^g(\mathbf{I_h}(\mathbf{x}), \mathbf{h}) \nabla I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})),$$

where

$$L_{\mathrm{CR}}^g(\mathbf{i}, \mathbf{h}) = \frac{1}{|\Omega|} G_\beta \star \partial_2 E^{\mathrm{CR}}(\mathbf{i}, \mathbf{h}).$$

The expanded expression of the function L_{CR}^g and its extension using the local version of the joint pdf are given in Table 1. Like in the previous two cases, L_{CR}^g plays the role of an intensity-comparison function. Minimizing $\mathcal{J}_{\mathbf{CR}^g}$ amounts to making i_2 lie as close as possible to $\mu_{2|1}(i_1, \mathbf{h})$ (which gives the "*backbone*" of P), while keeping the value of $\mu_{2|1}(i_1, \mathbf{h})$ away from $\mu_2(\mathbf{h})$.

5.4. Cross Correlation

To conclude this section, we consider the case of the cross correlation, which has been widely used as a robust comparison function for image matching. Within recent energy-minimization approaches relying on the computation of its gradient, we can mention for instance the works of Faugeras and Keriven (1998), Cachier and Pennec (2000), and Netsch et al. (2001). The cross correlation, being a measure of the *affine* dependency between the intensities, is the most constrained of the three criteria. Besides the quantities already introduced, its definition relies on the mean and variance of X:

$$\mu_1 = \int_{\mathbb{R}} i_1 p(i_1) \, di_1, \quad v_1 = \int_{\mathbb{R}} i_1^2 p(i_1) \, di_1 - \mu_1^2,$$

Table 1. Comparison functions.

Method	Global intensity comparison $L^g(\mathbf{i}, \mathbf{h})$
MI	$L_{\mathrm{MI}}^g(\mathbf{i}, \mathbf{h}) = G_\beta \star \frac{1}{|\Omega|} \left(\frac{\partial_2 P(\mathbf{i}, \mathbf{h})}{P(\mathbf{i}, \mathbf{h})} - \frac{p'(i_2, \mathbf{h})}{p(i_2, \mathbf{h})} \right)$
CR	$L_{\mathrm{CR}}^g(\mathbf{i}, \mathbf{h}) = G_\beta \star \dfrac{\mu_2(\mathbf{h}) - \mu_{2|1}(i_1, \mathbf{h}) + (1 - \mathcal{J}_{\mathbf{CR}^g}(\mathbf{h}))(i_2 - \mu_2(\mathbf{h}))}{\frac{1}{2} v_2(\mathbf{h}) |\Omega|}$
CC	$L_{\mathrm{CC}}^g(\mathbf{i}, \mathbf{h}) = -\dfrac{2}{|\Omega|} \left[\dfrac{v_{1,2}(\mathbf{h})}{v_2(\mathbf{h})} \left(\dfrac{i_1 - \mu_1}{v_1} \right) + \mathcal{J}_{\mathbf{CC}^g}(\mathbf{h}) \left(\dfrac{i_2 - \mu_2(\mathbf{h})}{v_2(\mathbf{h})} \right) \right]$

Method	Local intensity comparison $L^l(\mathbf{i}, \mathbf{h}, \mathbf{x})$
MI	$L_{\mathrm{MI}}^l(\mathbf{i}, \mathbf{h}, \mathbf{x}) = G_\gamma \star G_\beta \star \frac{1}{\mathcal{G}_\gamma(\mathbf{x})} \left(\frac{\partial_2 P(\mathbf{i}, \mathbf{h}, \mathbf{x})}{P(\mathbf{i}, \mathbf{h}, \mathbf{x})} - \frac{p'(i_2, \mathbf{h}, \mathbf{x})}{p(i_2, \mathbf{h}, \mathbf{x})} \right)$
CR	$L_{\mathrm{CR}}^l(\mathbf{i}, \mathbf{h}, \mathbf{x}) = G_\gamma \star G_\beta \star \dfrac{\mu_2(\mathbf{h}, \mathbf{x}) - \mu_{2|1}(i_1, \mathbf{h}, \mathbf{x}) + (1 - \mathcal{J}_{\mathbf{CR}}(\mathbf{h}, \mathbf{x}))(i_2 - \mu_2(\mathbf{h}, \mathbf{x}))}{\frac{1}{2} v_2(\mathbf{h}, \mathbf{x}) \mathcal{G}_\gamma(\mathbf{x})}$
CC	$L_{\mathrm{CC}}^l(\mathbf{i}, \mathbf{h}, \mathbf{x}) = -G_\gamma \star \dfrac{2}{\mathcal{G}_\gamma(\mathbf{x})} \left(\dfrac{v_{1,2}(\mathbf{h}, \mathbf{x})}{v_2(\mathbf{h}, \mathbf{x})} \left(\dfrac{i_1 - \mu_1(\mathbf{x})}{v_1(\mathbf{x})} \right) + \mathcal{J}_{\mathrm{CC}}(\mathbf{h}, \mathbf{x}) \left(\dfrac{i_2 - \mu_2(\mathbf{h}, \mathbf{x})}{v_2(\mathbf{h}, \mathbf{x})} \right) \right)$

as well as on the covariance of X and $Y_{\mathbf{h}}$, noted $v_{1,2}(\mathbf{h})$:

$$v_{1,2}(\mathbf{h}) = \int_{\mathbb{R}^2} i_1 i_2 P(\mathbf{i}, \mathbf{h}) d\mathbf{i} - \mu_1 \mu_2(\mathbf{h}).$$

The dissimilarity functional based on the cross correlation is simply defined in terms of these quantities:

$$\mathcal{J}_{\mathbf{CC}^g}(\mathbf{h}) = -\frac{v_{1,2}(\mathbf{h})^2}{v_1 v_2(\mathbf{h})}.$$

Its gradient has the same general expression

$$\nabla_H \mathcal{J}_{\mathbf{CC}^g}(\mathbf{h})(\mathbf{x}) = L_{\mathbf{CC}}^g(\mathbf{I_h}(\mathbf{x}), \mathbf{h}) \nabla I_2^\sigma (\mathbf{x} + \mathbf{h}(\mathbf{x})),$$

where

$$L_{\mathbf{CC}}^g(\mathbf{i}, \mathbf{h}) = \frac{1}{|\Omega|} G_\beta \star \partial_2 E^{\mathbf{CC}}(\mathbf{i}, \mathbf{h})$$

and

$$E^{\mathbf{CC}}(\mathbf{i}, \mathbf{h}) = \frac{-1}{v_1 v_2(\mathbf{h})} (2 v_{1,2}(\mathbf{h}) i_2 (i_1 - \mu_1)$$
$$+ \mathcal{J}_{\mathbf{CC}^g}(\mathbf{h}) v_1 i_2 (i_2 - 2\mu_2(\mathbf{h}))).$$

The expanded expression of $L_{\mathbf{CC}}^g$ and that of its extension using the local version of the joint pdf ($L_{\mathbf{CC}}^l$) are given in Table 1. Notice that in this case the convolution with respect to the intensity variable yields the same expression and is thus not required, since only linear terms are involved. Again, the $L_{\mathbf{CC}}$ functions play the role of comparing intensities. They show that minimizing $\mathcal{J}_{\mathbf{CC}^g}$ amounts to making the pair of intensities lie on a straight line in $\mathbb{R}^2$ (not necessarily passing through the origin).

6. Well-Posedness

The main idea that we have pushed forward is that many of the methods already developed for dense, nonrigid matching which assume the invariance of the intensities can be generalized to the case where this assumption is no longer valid. This has been achieved by computing the variational gradient of some statistical similarity measures, which have been shown to be very well suited for solving the rigid, multimodal registration problem. We have proposed a general framework for formally computing this variational gradient (i.e. through nonparametric local and global joint pdf estimates) and

carried over these calculations in the case of three statistical dissimilarity measures which, in our opinion, provide a convenient trade-off between flexibility and robustness.

We have used these results to generalize a recently proposed and very effective optical flow estimation method (Alvarez et al., 2000) based on solving a set of partial differential equations of the type of (3). These equations have the general form

$$\begin{cases} \frac{d\mathbf{h}}{dt} + A\mathbf{h}(t) = F(\mathbf{h}(t)) \\ \mathbf{h}(0)(\cdot) = \mathbf{h}_0(\cdot), \end{cases}$$

where the linear operator A corresponds to the regularization term and F is the (nonlinear) matching function. Our study of statistical measures yields a set of matching functions given by

$$F(\mathbf{h})(\mathbf{x}) = L(\mathbf{I_h}(\mathbf{x}), \mathbf{h}, \mathbf{x}) \nabla I_2^\sigma (\mathbf{x} + \mathbf{h}(\mathbf{x}))$$

where $L(\mathbf{i}, \mathbf{h}, \mathbf{x})$ is an intensity comparison term whose expression for the different criteria is summarized in Table 1.

Alvarez et al. showed the well-posedness of their method by proving that the set of PDEs they consider has a unique generalized solution. The proof relies on showing two conditions:

1. A is a maximal monotone operator from a linear subspace of H into H.
2. The function $F : H \rightarrow H$ is bounded, and Lipschitz continuous.

We have extended their work to the cases considered in this paper for the six different matching functions and the two regularization criteria. One of the difficulties of this generalization is the fact that in our case equations (2) are not PDEs since $F(\mathbf{h})(\mathbf{x})$ depends not only of the value of $\mathbf{h}(\mathbf{x})$ as in (4) but on its values in a neighborhood of $\mathbf{x}$ (which can be as large as the whole image in the case of the global criteria). Moreover we have shown that the unique solution of the initial value problem (2) is classical and regular, i.e. more than a generalized solution. Showing these properties (especially the Lipschitz-continuity) for the matching functions that we present here is an important aspect of our work. The proofs are however quite long and technical and we refer the interested reader to Faugeras and Hermosillo (2001).

7. Numerical Implementation

The numerical implementation of the described continuous matching flows involves estimating the matching term, which depends on one of the six intensity-comparison functions of Table 1, and the regularization operator, which in our case is either $\mathbf{div}(\mathbf{T}_{I_1^\sigma} D\mathbf{h})$ with $\mathbf{T}_{I_1^\sigma}$ given by Eq. (5) or the elasticity operator, given by Eq. (6). For the discretization in time, we adopt an explicit forward scheme. Implicit schemes are difficult to devise due to the high nonlinearity of the matching functions.

7.1. Regularization Operators

Alvarez et al. (2000) propose a very efficient scheme for estimating the Nagel-Enkelmann operator which we adopt in our experiments. Concerning that of linear elasticity, we use standard centered finite-difference schemes based on first-order Taylor expansions of $\Delta\mathbf{h}$ and $\nabla(\nabla\cdot\mathbf{h})$. Assuming the pixel size in both directions to be equal to one, the corresponding scheme for $\Delta\mathbf{h}$ in 2D is

$$L_k^{ij} = \frac{1}{4}\big(h_k^{i+1,j} + h_k^{i-1,j} + h_k^{i,j-1} + h_k^{i,j+1} - 4h_i^{i,j}\big),$$

where $L_k^{i,j}$ and $h_k^{i,j}$ denote respectively the components ($k = 1, 2$) of $\Delta\mathbf{h}$ and $\mathbf{h}$ at a grid point (i, j). Using a stencil notation, this scheme may be written as

$$L_k^{i,j} = \frac{1}{4}\underbrace{\begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 1 & -4 & 1 \\ \hline 0 & 1 & 0 \\ \hline \end{array}}_{h_k},$$

where i increases from left to right and j from top to bottom. The labeled under-brace indicates the grid is that of the component h_k of $\mathbf{h}$. Using this notation, our scheme for $\nabla(\nabla\cdot\mathbf{h})$ in 2D, whose components we denote as $A_k^{i,j}$ is:

$$A_1^{i,j} = \underbrace{\begin{array}{|c|c|c|} \hline 0 & 0 & 0 \\ \hline 1 & -2 & 1 \\ \hline 0 & 0 & 0 \\ \hline \end{array}}_{h_1} + \frac{1}{4}\underbrace{\begin{array}{|c|c|c|} \hline 1 & 0 & -1 \\ \hline 0 & 0 & 0 \\ \hline -1 & 0 & 1 \\ \hline \end{array}}_{h_2},$$

$$A_2^{i,j} = \underbrace{\begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 0 & -2 & 0 \\ \hline 0 & 1 & 0 \\ \hline \end{array}}_{h_2} + \frac{1}{4}\underbrace{\begin{array}{|c|c|c|} \hline 1 & 0 & -1 \\ \hline 0 & 0 & 0 \\ \hline -1 & 0 & 1 \\ \hline \end{array}}_{h_1}.$$

7.2. Intensity Comparison Functions

Concerning the intensity comparison functions, we approximate convolutions by a Gaussian kernel through recursive filtering using the smoothing operator introduced in Deriche (1990). Terms of the form $\nabla I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))$ and $I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x}))$ are calculated by tri-linear interpolation. The global functions L_{MI}^g, L_{CR}^g and L_{CC}^g are estimated by explicitly computing the global density estimate $P(\mathbf{i}, \mathbf{h})$ through recursive smoothing of the discrete joint histogram of intensities as detailed in Section 7.3. Values such as $\mu_2(\mathbf{h})$ are then estimated using finite sums on P. For the local functions, special implementations have been developed as detailed in Sections 7.4 and 7.5.

7.3. Density Estimation

Parzen density estimates are obtained by smoothing the discrete joint histogram of intensities. We define the piecewise constant function $\mathbf{v} : \Omega \to [0, N]^2 \subset \mathbb{N}^2$ by quantification of $\mathbf{I_h}(\mathbf{x})$ into $N + 1$ intensity levels (bins) between O and $\mathcal{A}$:

$$\mathbf{v}(\mathbf{x}) = \begin{pmatrix} \lfloor \zeta I_1^\sigma(\mathbf{x}) \rfloor \\ \lfloor \zeta I_2^\sigma(\mathbf{x} + \mathbf{h}(\mathbf{x})) \rfloor \end{pmatrix}$$

$$= \begin{cases} (0, 0)^T & \text{on} \quad \Omega_{0,0} \\ \vdots \\ (N, N)^T & \text{on} \quad \Omega_{N,N}, \end{cases}$$

where $\zeta = N/\mathcal{A}$, $\lfloor \cdot \rfloor$ denotes the floor operator in $\mathbb{R}^+$, i.e. the function $\mathbb{R}^+ \to \mathbb{N}$ such that $\lfloor x \rfloor = \max\{n \in \mathbb{N} : n \leq x\}$, and $\{\Omega_{k,l}\}_{(k,l)\in[0,N]^2}$ is a partition of Ω. We then compute, setting $\beta' = \zeta^2\beta$,

$$P(\mathbf{i}, \mathbf{h}) = \frac{1}{|\Omega|}\int_\Omega G_\beta(\mathbf{I_h}(\mathbf{x}) - \mathbf{i})\,d\mathbf{x}$$

$$= \frac{\zeta}{|\Omega|}\int_\Omega G_{\beta'}(\zeta(\mathbf{I_h}(\mathbf{x}) - \mathbf{i}))\,d\mathbf{x}$$

$$\simeq \frac{\zeta}{|\Omega|}\int_\Omega G_{\beta'}(\mathbf{v}(\mathbf{x}) - \zeta\mathbf{i})\,d\mathbf{x}$$

$$= \frac{\zeta}{|\Omega|}\sum_{k=0}^N \sum_{l=0}^N \int_{\Omega_{k,l}} G_{\beta'}(k - \zeta i_1, l - \zeta i_2)\,d\mathbf{x}$$

$$= \zeta \sum_{k=0}^N \sum_{l=0}^N \underbrace{|\Omega_{k,l}|/|\Omega|}_{K(k,l)} G_{\beta'}(k - \zeta i_1, l - \zeta i_2)$$

$$= \zeta(K \star G_{\beta'})(\zeta\mathbf{i}),$$

K being the discrete joint histogram. The convolution is performed by recursive filtering. Note that this way of computing P is quite efficient since only one pass on the images is required, followed by the convolution.

7.4.   Implementation of L_{CC}^l

The function L_{CC}^l is estimated as

$$L_{CC}^l(\mathbf{i}, \mathbf{x}) = (G_\gamma \star f_1)(\mathbf{x})i_1 + (G_\gamma \star f_2)(\mathbf{x})i_2 \\ + (G_\gamma \star f_3)(\mathbf{x}),$$

where

$$f_1(\mathbf{x}) = -2v_{1,2}(\mathbf{h}, \mathbf{x})/(\mathcal{G}_\gamma(\mathbf{x})v_1(\mathbf{x})v_2(\mathbf{h}, \mathbf{x})),$$
$$f_2(\mathbf{x}) = -2\mathcal{J}_{\mathbf{CC}}(\mathbf{h}, \mathbf{x})/(\mathcal{G}_\gamma(\mathbf{x})v_2(\mathbf{h}, \mathbf{x}))$$

and

$$f_3(\mathbf{x}) = -(f_1(\mathbf{x})\mu_1(\mathbf{x}) + f_2(\mathbf{x})\mu_2(\mathbf{h}, \mathbf{x})).$$

All the required space dependent quantities like $\mu_1(\mathbf{x})$ are computed through recursive spatial smoothing. This algorithm is similar to the one proposed in Cachier and Pennec (2000).

7.5.   Approximate Implementations of L_{MI}^l and L_{CR}^l

The functions L_{MI}^l and L_{CR}^l are much more difficult to compute than L_{CC}^l because they involve two convolutions, one with respect to the intensity variable $\mathbf{i}$ and the other with respect to the space variable $\mathbf{x}$. Applying a smoothing filter would require a dense data structure of dimension $(n + 1)$ for L_{CR}^l and $(n + 2)$ for L_{MI}^l. With 3D images, it becomes extremely difficult to maintain these four and five-dimensional structures. Our implementations rely on computing "un-smoothed" versions of these functions i.e. on eliminating both convolutions. Although the resulting functions are no longer the gradient of the considered dissimilarity functionals, this approximation makes the computation tractable and has given good results in practice. Quantities like $\mu_2(\mathbf{h}, \mathbf{x})$ are estimated using sums, weighted by gaussian functions, over a large neighborhood around each pixel. This corresponds to a non-stationary filtering procedure which is very computationally expensive. However, the resulting algorithms are very well adapted to parallelization and we have implemented them using the MPI library for parallel execution using a cluster

of up to thirty processors, obtaining up to twenty times faster execution with respect to the sequential version. We refer the interested reader to Hermosillo (2002) for more details concerning implementation issues.

7.6.   Parameters

We now discuss the way in which the different parameters of the algorithms are determined.

- γ: This is the variance of the spatial Gaussian for local density estimates. In the case of the cross correlation, its value does not affect the computation time since the local statistics are calculated using the recursive smoothing filter. Thanks to this, we have conducted some experiments with different values of this parameter, which have shown that the algorithms are not very sensitive to it. Qualitatively speaking, the local window has to be large enough for the statistics to be significant, and small enough to account for non-stationarities of the density. It is fixed to 5 with a local window size of 19×19 for the mutual information and the correlation ratio.
- β: This is the variance of the Gaussian for the Parzen estimates. Unlike γ, determining a good value for β is important for obtaining good results. In our case, it is determined automatically as follows (we refer to Bosq (1998)) for a recent comprehensive study on nonparametric density estimation). We adopt a cross-validation technique based on an empirical maximum likelihood method. We note $\{\mathbf{i}_k\}$ a set of m intensity pair samples ($k = 1 \ldots m$) and take the value of β which maximizes the empirical likelihood:

$$L(\beta) = \prod_{k=1}^{m} \hat{P}_{\beta,k}(\mathbf{i}_k)$$

where

$$\hat{P}_{\beta,k}(\mathbf{i}_k) = \frac{1}{m - n_k} \sum_{\{s:\mathbf{i}_s \neq \mathbf{i}_k\}} G_\beta(\mathbf{i}_k - \mathbf{i}_s)$$

and n_k is the number of data samples for which $\mathbf{i}_s = \mathbf{i}_k$.

- α: This parameter determines the weight of the regularization term in the energy functional. Since the range of the different matching functions varies considerably, we replace it by another one, noted C such that

$$\alpha = C\kappa,$$

where κ is given by

$$\kappa = \|F(\mathbf{h}_0)\|_\infty,$$

$\mathbf{h}_0$ being the initial field and F any of the matching functions.

- σ: This is the scale parameter. We adopt a multi-resolution approach, smoothing the images at each stage by a small amount. Within each stage of the multi-resolution pyramid, the parameter σ is fixed to a small value, typically 0.25 voxels.

Besides these global parameters, one extra parameter is needed for each family of regularization operators.

- ξ: Instead of the Lamé coefficients in the linear elasticity operator (Eq. (6)), we use a single parameter ξ controlling the relative weight between the two operators $\Delta\mathbf{h}$ and $\nabla(\nabla\cdot\mathbf{h})$ as $\xi\Delta\mathbf{h}+(1-\xi)\nabla(\nabla\cdot\mathbf{h})$. For ξ close to 1, the Laplacian operator becomes dominant, while the operator $\nabla(\nabla\cdot\mathbf{h})$ becomes dominant for ξ close to zero. In practice, we fix the value of ξ to 0.5, giving thus the same weight to both operators.
- λ: This is the parameter controlling the anisotropic behavior of the Nagel-Enkelmann tensor. We adopt the method proposed by Alvarez et al. (2000). Given s, which in practice is fixed to 0.1, we take the value of λ such that

$$s = \int_0^\lambda \mathcal{H}_{|\nabla I_1^\sigma|}(z)dz,$$

where $\mathcal{H}_{|\nabla I_1^\sigma|}(z)$ is the normalized histogram of $|\nabla I_1^\sigma|$.

8. Experiments

We present results of experiments using the previously described algorithms. To recover large deformations, we use a multi-resolution approach by applying the gradient descent to a set of smoothed and subsampled images. Since the considered functionals are not convex, this coarse-to-fine strategy helps avoiding irrelevant extrema while reducing the computational cost of the algorithms.

8.1. Supervised Registration Method (Fig. 2)

In this experiment, a synthetic learning set is produced by applying a *sin* function and adding Gaussian noise (zero mean with 0.1 variance) to four images (whose intensity has been normalized on $[0, 2\pi]$). We use the
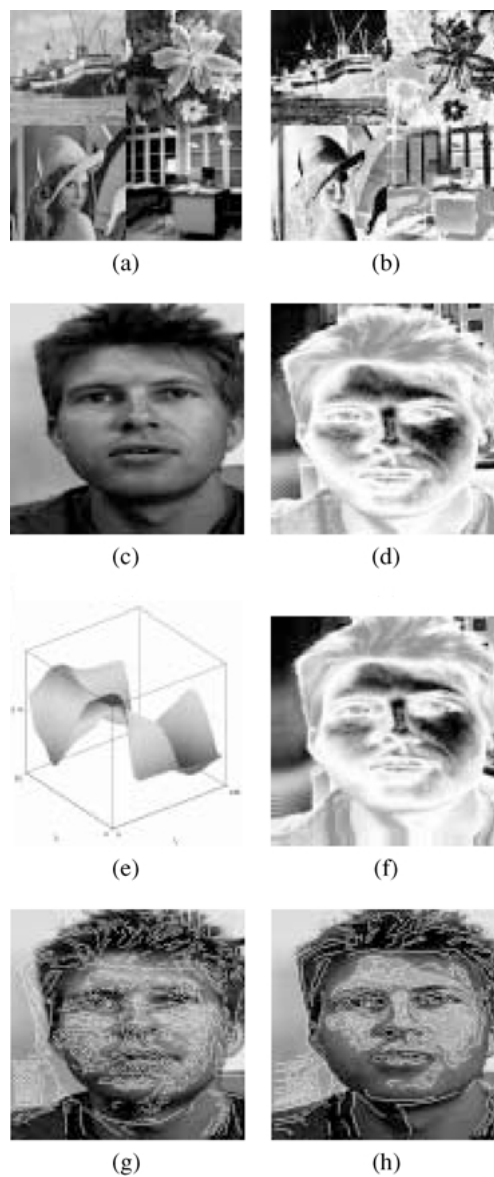

Figure 2. **SR** matching: image mosaics (a) and (b) represent the learning sets: 4 pairs of 256×256 images. (c) is the reference image I_1^σ and (d) the image to register I_2^σ. The surface (e) represents the estimated score function $L_{\mathbf{SR}}(\mathbf{i}) = -\partial_2 P(\mathbf{i})/P(\mathbf{i})$. (f) shows the resulting image $I_2^\sigma(\mathbf{Id} + \hat{\mathbf{h}})$. (g) and (h) represent the contours of I_2^σ and $I_2^\sigma(\mathbf{Id} + \hat{\mathbf{h}})$ superimposed on I_1^σ.

comparison function estimated from this learning set to match two views of a face. The functional dependence between these two intensity maps is clearly recovered from the estimated comparison function, and a visually correct realignment is achieved. Some small artefacts appear, mainly due to some elements in the background which are not present in both views.

8.2.   *Different Regularization Operators (Fig. 3)*

In this example, we use the global mutual information criterion to compute the displacement field between two synthetic images. Two experiments are carried out, one using linear elasticity regularization (second row of Fig. 3) and the other one using the Nagel-Enkelmann operator (third row). In this case, we clearly see the advantage of the latter regularization method. It preserves the natural discontinuities of the displacement field. This example underlines the role of the prior knowledge which is embedded in the regularization functional. It also illustrates the fact that validating with synthetic deformation fields may not always be appropriate, since the solution depends on assumed properties of the field. Another interesting feature of this experiment is the fact that the discrete histogram (i.e. before Gaussian smoothing) contains only six non-zero entries, which underlines the importance of the Parzen-window regularization.
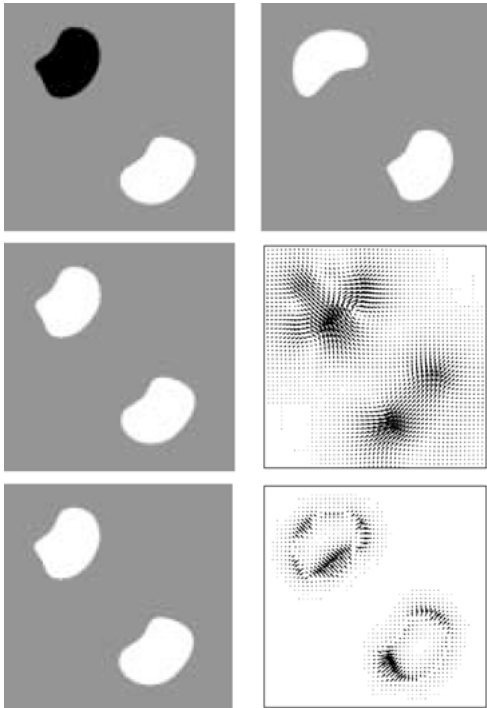

Figure 3.   Global **MI** criterion with a synthetic example. The first row shows the reference image $(256 \times 256) I_1^\sigma$ on the left and the image to register I_2^σ. The second and third rows represent the warped image $I_2^\sigma (\mathbf{Id} + \hat{\mathbf{h}})$, and the corresponding displacement field $\hat{\mathbf{h}}$, for two different types of regularization operators (linear elasticity model for the second row, geometry-driven anisotropic diffusion for the third row).
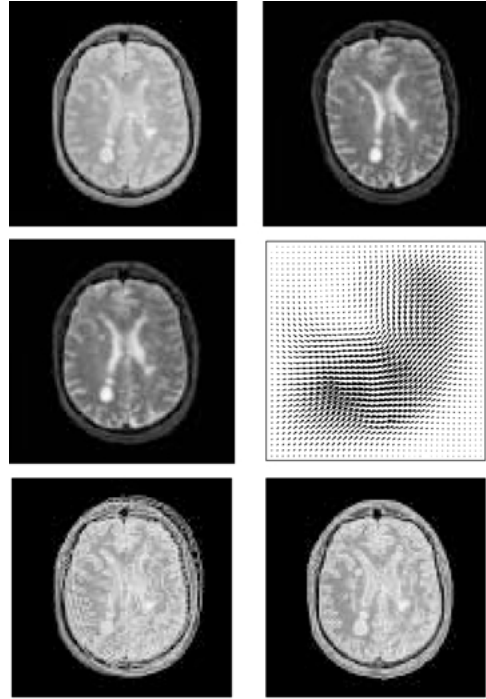

Figure 4.   PD/T2 MRI registration using the global CR comparison function. The first row shows the reference image I_1^σ on the left and the image to register I_2^σ. The second row represents the warped image $I_2^\sigma (\mathbf{Id} + \hat{\mathbf{h}})$, and the corresponding displacement field $\hat{\mathbf{h}}$. The last row represents the contours of I_2^σ and $I_2^\sigma (\mathbf{Id} + \hat{\mathbf{h}})$ superimposed on I_1^σ.

8.3.   *Global Correlation Ratio (Fig. 4)*

We use the global correlation ratio criterion to realign two slices from a Proton Density (PD) and a T2 MRI volume (same patient). An artificial geometric distortion (based on a set of three Gaussian kernels) has been applied to the original pre-registered dataset. In order to evaluate the accuracy of the realignment, we superimposed some contours of the T2 image (initial and recovered pose) over the reference image (PD). It gives a good qualitative indication of the quality of the registration. Most of the anatomical structures seem correctly realigned.

8.4.   *Registration Between MRI and fMRI Modalities Using Global Mutual Information (Fig. 5)*

This example shows an experiment with real MR data of the brain of a macaque monkey. The reference image is a T1-weighted anatomical volume and the image to register is a functional, mion contrast MRI (fMRI). The
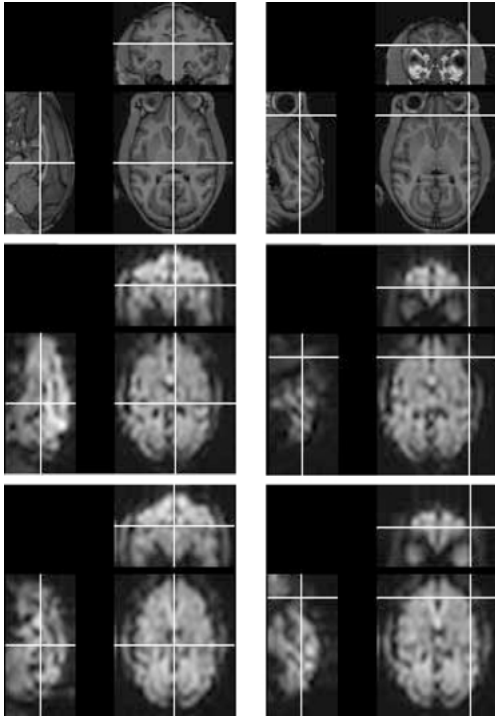
Figure 5. MRI-fMRI registration using the global **MI** criterion. The two columns show two different regions of interest within the image volume. The first and second rows show the reference, anatomical MRI and the initial fMRI volume, respectively, while the third row shows the final, corrected fMRI (to be compared with the second row).

contrast in this modality is related to blood oxygenation level. This registration was obtained using global mutual information. Notice that the alignment of the main axis of the volume has been corrected.

8.5. *Local Mutual Information and Correlation Ratio (Fig. 6)*

This experiment shows results using the dissimilarity measures based on the local mutual information and the local correlation ratio on synthetic data. The reference and deformed image where both taken from the same 2D plane in a MRI data volume. The reference image J was then transformed as:

$$J'(x, y) = \sin\left(2\pi J(x, y)\right) - \cos\left(\frac{2\pi(x + y\sqrt{|\Omega|})}{|\Omega|}\right).$$

A nonrigid smooth deformation was then applied to the second image. As expected, the global similarity criteria failed to align these two images, due to the severe
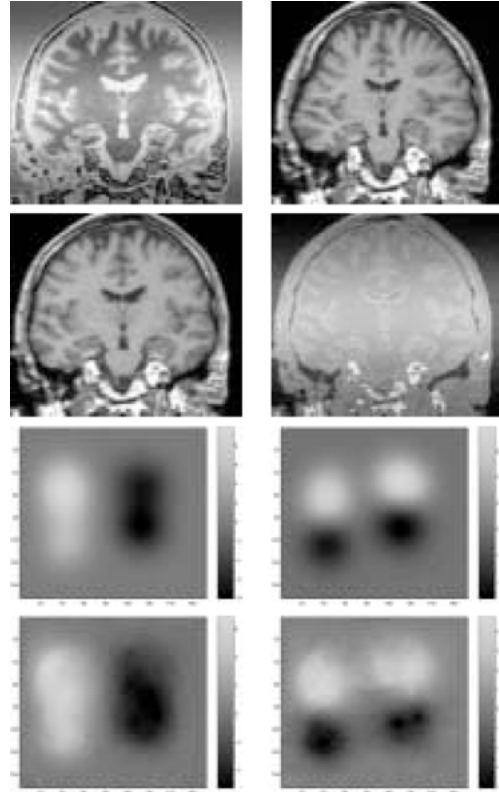

Figure 6. Matching using local mutual information and correlation ratio. The first row shows I_1^σ on the left and I_2^σ on the right. The second row shows $I_2^\sigma(\mathbf{Id} + \mathbf{h})$ on the left and its superposition with I_1^σ on the right. The third and fourth rows show respectively the applied and recovered displacement fields (x component on the left). Similar results were obtained using both criteria.

non-stationarity in the intensity distributions, while a correct realignment was obtained using the local versions. Similar results were obtained using mutual information and correlation ratio.

8.6. *Face Template Matching Using Local Cross Correlation (Fig. 7)*

This last experiment shows template matching of human faces. The different albedos of the two skins create a "multimodal" situation and the transformation is truly nonrigid due to the different shapes of the noses and mouths. Notice the excellent matching of the different features. This result was obtained using local cross correlation. The running time was approximately five minutes on a PC at 900 MHz. With the correspondences, one can interpolate the displacement field and the texture to perform fully automatic morphing.
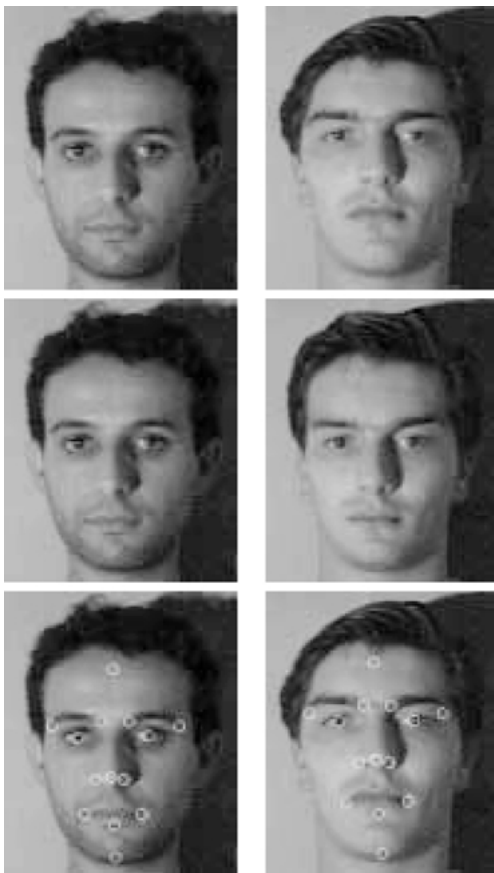
Figure 7. Human face template matching: the left column shows the reference image (I_1^σ) with some reference points at the bottom. The right column shows from top to bottom: I_2^σ, $I_2^\sigma(\mathbf{Id} + \hat{\mathbf{h}})$ and the corresponding reference points according to the found displacement field $\hat{\mathbf{h}}$. Local cross correlation was used as similarity criterion.

9. Conclusion

In this paper, we proposed a variational framework to address the problem of dense matching between images when the hypothesis of intensity preservation is not valid. Our approach relies on the computation of the first variation of a hierarchy of statistical criteria, computed either globally or locally within corresponding regions. We illustrated the efficiency of the algorithms that we propose trough several real and synthetic examples.

Our main contributions consist in:

1. Modeling a real problem as a variational problem.
2. Deriving the necessary equations satisfied by the solutions (if any) of this variational problem.

3. Showing that the resulting equations are well-posed (existence and uniqueness of a solution).
4. Showing that the resulting computational theory provides interesting predictions (results) on real images.

Our undergoing research concerns the application of these results in the context of the diffeomorphic matching framework described by Trouvé (1998), in which large deformations are more naturally handled.

Note

1. The uniqueness of $\nabla_H \mathcal{I}(\mathbf{h})$ is guaranteed by the Riesz representation theorem (Evans, 1998). It depends on the choice of the scalar product $(\cdot, \cdot)_H$, a fact which explains our notation.

References

Alvarez, L., Weickert, J., and Sánchez, J. 2000. Reliable estimation of dense optical flow fields with large displacements. Technical Report, Cuadernos del Instituto Universitario de Ciencias y Tecnologías Cibernéticas. A revised version has appeared at *IJCV* 39(1):41–56.

Bosq, D. 1998. *Nonparametric Statistics for Stochastic Processes.* 2nd edn. vol. 110 of Lecture Notes in Statistics. Springer–Verlag: Berlin.

Cachier, P. and Pennec, X. 2000. 3d non-rigid registration by gradient descent on a gaussian weighted similarity measure using convolutions. In *Proceedings of MMBIA*, pp. 182–189.

Chefd'hotel, C., Hermosillo, G., and Faugeras, O. 2001. A variational approach to multi-modal image matching. In *IEEE Workshop on Variational and Level Set Methods*, University of British Columbia, Vancouver, Canada. IEEE Computer Society, pp. 21–28.

Christensen, G., Miller, M., and Vannier, M. 1994. A 3d deformable magnetic resonance textbook based on elasticity. In *Proceedings of the American Association for Artificial Intelligence, Symposium: Applications of Computer Vision in Medical Image Processing.*

Courant, R. 1946. *Calculus of Variations.* New York.

Deriche, R. 1990. Fast algorithms for low-level vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(12):78–88.

Evans, L. 1998. Partial Differential Equations. In *Proceedings of the American Mathematical Society*. vol. 19 of Graduate Studies in Mathematics.

Faugeras, O. and Keriven, R. 1998. Variational principles, surface evolution, PDE's, level set methods and the stereo problem. *IEEE Transactions on Image Processing*, 7(3):336–344.

Faugeras, O.D. and Hermosillo, G. 2001. Well-posedness of eight problems of multi-modal statistical image-matching. Technical Report 4235, INRIA.

Gaens, T., Vandermeulen, F.M.D., and Suetens, P. 1998. Non-rigid multimodal image registration using mutual information. In *First International Conference on Medical Image Computing and Computer-Assisted Intervention*. G. Goos and J. Hartmanis, (Eds.), vol. 1496 of Lecture Notes in Computer Science. Springer: Berlin.

Hata, N., Dohi, T., Warfield, S., III, W.W., Kikinis, R., and Jolesz, F.A. 1998. Multi-modality deformable regristration of pre- and intra-operative images for mri-guided brain surgery. In *First International Conference on Medical Image Computing and Computer-Assisted Intervention*. G. Goos and J. Hartmanis (Eds.), vol. 1496 of Lecture Notes in Computer Science. Springer: Berlin.

Hermosillo, G. 2002. Variational Methods for Multimodal Image Matching. PhD thesis, INRIA, The document is accessible at ftp://ftp-sop.inria.fr/robotvis/html/Papers /hermosillo:02.ps.gz.

Koenderink, J.J. and van Doorn, A.J. 1999. Blur and disorder. In *Scale-Space Theories in Computer Vision, Second International Conference, Scale-Space'99*, M. Nielsen, P. Johansen, O.F. Olsen, and J. Weickert (Eds.), vol. 1682 of Lecture Note in Computer Science, Springer: Berlin, pp.1–9.

Leventon, M. and Grimson, W. 1998. Multi-modal volume registration using joint intensity distributions. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI'98*, W. Wells, A. Colchester, and S. Delp (Eds.). Cambridge, MA, USA, vol. 1496 in Lecture Notes in Computer Science. Springer: Berlin.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. 1997. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198.

Maintz, J., Meijering, H., and Viergever, M. 1998. General multi-modal elastic registration based on mutual information. In *Medical Imaging 1998—Image Processing*, vol. 3338, SPIE, pp. 144–154.

Meyer, C., Boes, J., Kim, B., and Bland, P. 1998. Evaluation of control point selection in automatic, mutual information driven, 3d warping. In *First International Conference on Medical Image Computing and Computer-Assisted Intervention, Proceedings*, G. Goos and J. Hartmanis (Eds.), vol. 1496 of Lecture Notes in Computer Science.

Nagel, H. and Enkelmann, W. 1986. An investigation of smoothness constraint for the estimation of displacement vector fiels from images sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:565–593.

Netsch, T., Rosch, P., van Muiswinkel, A., and Weese, J. 2001. Towards real-time multi-modality 3d medical image registration. In *Proceedings of the 8th International Conference on Computer Vision*, Vancouver, Canada. IEEE Computer Society, IEEE Computer Society Press.

Parzen, E. 1962. On the estimation of probability density function. *Ann. Math. Statist.*, 33:1065–1076.

Roche, A., Guimond, A., Meunier, J., and Ayache, N. 2000a. Multimodal elastic matching of brain images. In *Proceedings of the 6th European Conference on Computer Vision*, Dublin, Ireland.

Roche, A., Malandain, G., and Ayache, N. 2000b. Unifying maximum likelihood approaches in medical image registration. *International Journal of Imaging Systems and Technology: Special Issue on 3D Imaging*, 11(1):71–80.

Roche, A., Malandain, G., Pennec, X., and Ayache, N. 1998a. The correlation ratio as new similarity metric for multimodal image registration. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI'98*, W. Wells, A. Colchester, and S. Delp (Eds.), Cambridge, MA, USA, vol. 1496 in Lecture Notes in Computer Science. Springer: Berlin, pp. 1115–1124.

Roche, A., Malandain, G., Pennec, X., and Ayache, N. 1998b. Multimodal image registration by maximization of the correlation ratio. Technical Report 3378, INRIA.

Rückert, D., Hayes, C., Studholme, C., Summers, P., Leach, M., and Hawkes, D. 1998. Non-rigid registration of breast mr images using mutual information. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI'98*, W. Wells, A. Colchester, and S. Delp (Eds.), Cambridge, MA, USA, vol. 1496 in Lecture Notes in Computer Science. Springer: Berlin.

Toga, A., (Ed.). 1998. *Brain Warping*. Academic Press.

Trouvé, A. 1998. Diffeomorphisms groups and pattern matching in image analysis. *International Journal of Computer Vision*, 28(3):213–21.

Viola, P. and Wells, III, W. M. 1997. Alignement by maximization of mutual information. *The International Journal of Computer Vision*, 24(2):137–154.